ÉLÉMENTS DE MATHÉMATIQUES APPLIQUÉES

version 0.01

8 Novembre 2009

Ce document est sous licence Creative Commons version 3.0 France
paternité; pas d'utilisation commerciale; partage des conditions initiales à l'identique
http://creativecommons.org/licenses/by-nc-sa/3.0/deed.fr



Table des matières

Table des matières i
Avant-propos iii
Introduction v

1 I ARITHMÉTIQUE

1 Théorie de la démonstration 3
2 Nombres 29
3 Opérateurs arithmétiques 61
4 Théorie des ensembles 77
5 Théorie des nombres 109
6 Probabilités 125
7 Statistiques 137

233 II ALGÈBRE

8 Calcul Algébrique 235
9 Algèbre et géométrie ensemblistes 259
10 Analyse fonctionnelle 273
11 Topologie 299
12 Théorie de la mesure 311
13 Calcul différentiel et intégral 319
14 Suites et séries 387
15 Calcul vectoriel 433

481 III CHIMIE

16 Chimie quantique 483
17 Chimie moléculaire 497
18 Chimie analytique 505
19 Chimie thermique 509

517 IV ANNEXES

A Repères biographiques 519

525 V RÉFÉRENCES ET INDEXES

Références 527
Index général 529
Index des noms propres 533

Avant-propos

Le présent document a été conçu de façon à ce que les connaissances requises pour le parcourir soient les plus élémentaires. Pour le consulter, il suffit de savoir raisonner, d'observer et d'avoir du temps. À ce titre, il doit pouvoir se substituer à de nombreuses références et permettre ainsi à toute personne curieuse de parfaire ses connaissances dans ce domaine.

Naissance

Force est de constater qu'un ouvrage assez complet, détaillé et pédagogique, si possible d'accès gratuit et portatif et présentant une vue d'ensemble des notions de mathématiques appliquées, comprenant des démonstrations, n'est actuellement pas disponible. Un tel ouvrage offrirait, en plus, une notation homogène ainsi que la possibilité d'améliorations ou de compléments suggérés par le lecteur attentif. Certes, dans le but de partager ce savoir mathématique, il peut paraître paradoxal de vouloir augmenter la liste, déjà longue, des ouvrages disponibles dans les bibliothèques, dans le commerce et sur internet. Il faut donc en mesure de présenter une argumentation solide qui justifie la création d'un tel document. Voici donc les quelques arguments qui paraissent susceptibles d'être énumérés :

1. le partage de la connaissance ;
2. le caractère évolutif et pratique d'un document électronique libre, en fonction de la demande ;
3. une présentation rigoureuse avec des démonstrations simplifiées ;
4. la possibilité pour les étudiants, les professeurs et plus généralement les lecteurs, de réutiliser les sources ;
5. une notation homogène, dans tout l'ouvrage, pour les opérateurs mathématiques, un langage clair et rigoureux sur l'ensemble des sujets abordés ;
6. l'archivage des sources d'un tel document en vue d'une traduction en un nouveau language issu des nouvelles technologies de l'information.

Les informations détaillées dans ce document ne sont pas nouvelles mais malheureusement, elles ne sont pas non plus toujours facilement accessibles. Datant de plusieurs dizaines d'années au minimum, elles ne sont normalement plus soumises à une quelconque propriété intellectuelle et peuvent être librement distribuées.

Le choix de présenter l'ensemble des concepts par ordre logique et non par ordre de nécessité a eu les conséquences suivantes :

1. il faudra parfois admettre provisoirement certaines idées, quitte à les comprendre plus tard ;
2. il sera certainement nécessaire pour le lecteur de parcourir au moins deux fois l'ensemble de l'ouvrage. Lors de la première lecture, on appréhende l'essentiel et lors de la deuxième lecture, on comprend les détails ;
3. il faut accepter certaines répétitions ainsi que les nombreuses références croisées et remarques complémentaires.

Certains savent que pour chaque théorème et modèle mathématique, il existe souvent plusieurs démonstrations. A été choisie celle qui semblait la plus simple.

LaTeX

Parmi les possibilités disponibles pour le développement d'un tel document, la seule qui semble pertinente actuellement est LaTeX. Ce language associe qualité visuelle et honnêteté typographique mais permet aussi, par son principe de fonctionnement même, la création de documents électroniques au sein desquels il est possible de naviguer facilement. Une solution purement basée sur la

technologie internet (html, xml, mathml...) ne semble ne pas avoir atteint un niveau de maturité suffisant quand le contenu scientifique est important. Qui plus est, ce système est libre et par conséquent accessible à tout le monde, et permet de travailler manière complètement autonome. Une attention toute particulière est portée aux illustrations, grâce à l'utilisation de l'extension PSTricks, qui permet de conserver une homogénéité. Il s'agit aussi d'un choix pérenne puisque les sources sont utilisables sous tout type de système d'exploitation et pourront être traduites, au cas où l'avenir l'exige, dans un format plus moderne.

L'un des principes fondamentaux de LaTeX est la séparation du contenu et de la mise en forme. LaTeXcalcule la mise en page, numérote chapitres, sections, tableaux et figures, crée les références bibliographiques, les indexes, la table des matières et les références croisées avec un minimum d'interventions.

Bien que cette caractéristique soit découplée du choix du programme de traitement texte, il s'avère que LaTeXréalise un travail remarquable concernant la prise en compte des règles typographiques relatives à chaque langue. La langue française a donc ses propres conventions, détaillées pour la plupart dans le document :

http://jacques-andre.fr/faqtypo/lessons.pdf

et respectées, dans la mesure du possible, dans cet ouvrage.

Utilisation

Ce document contient des renvois hypertextes qui permettent une lecture efficace et fluide. Il peut s'agir par exemple, au sein du texte principal, d'un renvoi à une référence bibliographique qui contient les détails appropriés pour un approfondissement ultérieur. Il peut s'agir aussi d'un renvoi à une définition expliquée dans une autre page : dans ce cas précis, c'est dans la marge que le lien apparaît afin qu'il conserve un sens pour une version imprimée. Enfin, l'index et la table des matières autorisent un saut directement à la page sélectionnée par le lecteur. Un comportement similaire existe pour les équations, les figures et les tableaux. Pour un ouvrage de cette proportion, il semble important de souligner cet aspect.

On notera l'existence de la commande Alt + ← dans Acrobat Reader qui permet de revenir à la page lue précédente et qui n'existe pas dans tous les lecteurs de documents pdf.

Pour la légèreté du document et lecture améliorée, les polices et les graphiques sont conservés sous forme vectorielle.

Remerciements

Dans l'ordre alphabétique :

- Rafaël Guglielmetti, passionné du développement web et étudiant en mathématiques à l'École Polytechnique Fédérale de Lausanne, pour le développement et l'amélioration du code du forum ;
- Léon Harmel, ingénieur diplômé en électromécanique avec une spécialisation en électronique et automaticité, responsable au laboratoire de recherches physiques chez A.C.E.C, pour l'apport de la documentation qui a servi à la rédaction des chapitres de physique quantique corpusculaire, ondulatoire, de physique quantique des champs et de calcul spinoriel et relativité générale ;
- Vincent Isoz, ingénieur diplômé en physique avec une spécialisation en génie nucléaire. Responsable et créateur/développeur du site http://www.sciences.ch/ et rédacteur de la plus grande partie du contenu.
- Ruben Ricchiuto, ingénieur physicien et mathématicien à l'université de Genève, pour sa précieuse aide et disponibilité en physique des plasmas, électromagnétisme, physique quantique, statistiques, topologie, analyse et chimie quantique, et encore beaucoup d'autres domaines relatifs aux mathématiques pures.

Introduction

Omniprésentes dans l'industrie [1], ou dans les services [2], les mathématiques appliquées apparaissent aussi dans de nombreux autres secteurs. Elles interviennent dans notre vie quotidienne et contribuent à la résolution de problématiques actuelles : énergie, santé, environnement, climatologie, optimisation, développement durable... Leur grand succès prend racine dans leur fantastique dispersion dans le monde réel, évolutif et contingent, et leur intégration croissante à toutes les activités humaines. Nous évoluons lentement donc vers une situation où les mathématiciens n'auront plus le monopole des mathématiques, mais où des économistes, gestionnaires et marchands feront tous des mathématiques.

Ceux qui entrevoient la mathématique appliquée que comme un outil ou comme l'ennemi des croyances religieuses, ou encore comme un domaine scolaire rébarbatif, sont légion. Il est cependant peut-être utile de rappeler que, comme le disait Galilée, « Le livre de la Nature est écrit dans le langage des mathématiques. » C'est dans cet esprit que ce document aborde la mathématique appliquée pour les étudiants en sciences de la nature, de la terre et de la vie, ainsi que pour tous ceux qui exercent une profession liée à ces diverses matières, y compris la philosophie, ou pour toute personne curieuse de s'informer de l'implication des sciences dans la vie quotidienne.

La frontière entre la philosophie et les sciences pures et exactes est très ténue. Comme le relate Platon dans le Phédon, Socrate à ses dernières heures s'est entouré de ses amis et disciples, dont deux pythagoriciens Cébès et Simmias considérés comme interlocuteurs privilégiés. Ce n'est pas un hasard puisqu'il convient alors de philosopher, déjà, en s'inspirant du modèle pythagoricien qui fait des mathématiques une voie nécessaire d'accès à la vérité, seule capable de se frayer un chemin fiable pour aborder des sujets aussi importants que ceux de l'âme et de sa destinée.

Le choix de traiter l'ingénierie ici comme une branche de la mathématique appliquée provient certainement du fait que l'ensemble des domaines de la physique et la mathématique sont à ce jour tellement peu discernables que la médaille de Fields [3] a été décernée en 1990 au physicien E. Witten, qui a utilisé des idées issues de la physique pour redémontrer un théorème mathématique. Cette tendance n'est certainement pas fortuite, car nous pouvons observer que toute science, dès qu'elle cherche à atteindre une compréhension plus détaillée du sujet qu'elle étudie, voit finalement toujours sa course aboutir dans les mathématiques pures. Ainsi pouvons-nous présager dans un futur lointain, la convergence de toutes les sciences (pures, exactes ou sociales) vers la mathématique pour la modélisation [4].

Il peut parfois nous paraître difficile de transmettre le sentiment de beauté mathématique de la nature, de son harmonie la plus profonde et de la mécanique élégante de l'univers, à ceux qui ne connaissent que les rudiments du calcul formel. Le physicien R. Feynmann a parlé une fois de *deux cultures* : les gens qui ont, et ceux qui n'ont pas eu une compréhension suffisante des mathématiques pour apprécier la structure scientifique de la nature. Pour l'anecdote, on prétend qu'un roi ayant demandé à Euclide de lui enseigner la géométrie se plaignit de sa difficulté. Euclide répondit : « Il n'y a pas de voie royale. » Les physiciens et mathématiciens ne peuvent se convertir à un autre langage. Si vous voulez apprendre à connaître la nature, à l'apprécier à sa juste valeur, vous devez comprendre son langage. Elle se révèle sous cette unique forme.

Au même titre, aucune discussion intellectuelle ne vous permettra de communiquer à un sourd ce que vous ressentez vraiment en écoutant de la musique. De même, toutes les discussions du monde resteront impuissantes à transmettre une compréhension intime de la nature à ceux de "l'autre culture". Les philosophes et théologiens peuvent essayer de vous donner des idées qualitatives sur l'Univers. Le fait que la méthode scientifique ne puisse convaincre le monde entier de sa justesse et de sa pureté, trouve peut-être sa cause dans l'horizon limité de certaines gens qui sont amené à s'imaginer que l'homme ou qu'un autre concept intuitif, sentimental ou arbitraire est le centre de l'Univers.

1. On peut par exemple citer l'aérospatiale, l'imagerie médicale, la cryptographie, les transports, la chimie,...
2. banques, assurances, ressources humaines, projets, logistique, architecture, télécommunications...
3. Il s'agit de la plus haute récompense de nos jours dans le domaine de la mathématique.
4. Lire à titre d'exemple le document "L'explosion des mathématiques" disponible dans la rubrique téléchargement du site.

Méthodes

La science est l'ensemble des efforts systématiques pour acquérir des connaissances sur notre environnement monde, pour les organiser et les synthétiser en lois et théories vérifiables et ayant pour principal objectif d'expliquer le "comment" des choses. Les scientifiques doivent soumettre leurs idées et résultats à la vérification et la reproduction indépendante de leurs pairs. Ils doivent abandonner ou modifier leurs conclusions lorsque confrontées à des évidences plus complètes ou différentes. La crédibilité de la science s'appuie sur ce mécanisme d'autocorrection. L'histoire de la science montre que ce système fonctionne depuis très longtemps et ce même très bien par rapport à tous les autres. Dans chaque domaine, les progrès ont été spectaculaires. Toutefois, le système a parfois des ratés qu'il faut corriger avant que les petites dérives ne s'accumulent.

Cet avertissement, et les rappels qui vont suivre, doivent servir le scientifique à se remettre en question en faisant un bon usage de ce que nous pouvons considérer aujourd'hui comme les bonnes méthodes de travail pour résoudre des problèmes ou développer des modèles théoriques.

Dans ce but, le tableau 1 récapitule les différentes étapes du travail en mathématique ou physique théorique.

MATHÉMATIQUE	PHYSIQUE
Poser *l'hypothèse*, la *conjecture* ou la *propriété* à démontrer de manière formelle ou en langage commun	Poser correctement et de manière détaillée le ou les *problèmes* à résoudre de manière formelle ou en langage commun
Définir les *axiomes* qui vont donner les points de départ et établir des restrictions aux développements [5]. Dans la même idée, le mathématicien définit le vocabulaire spécialisé relié à des opérateurs mathématiques [6]	Définir ou énoncer les *postulats* ou *principes* ou encore les *hypothèses* et *suppositions* qui vont donner les points de départ et établir des restrictions aux développements
Des axiomes posés, tirer directement des *lemmes* ou des *propriétés* dont la validité en découle directement et qui préparent au développement du théorème censé valider l'hypothèse ou la conjecture de départ	Une fois le *modèle théorique* développé vérifier les équations dimensionnelles pour déceler une éventuelle erreur dans les développements
Une fois le ou les *théorèmes* démontrés en tirer des *corollaires* et encore des propriétés	Chercher les cas limites, dont les *singularités* font partie, du modèle pour en vérifier la validité intuitive
Tester la force ou l'utilité de sa ou ses conjectures ou hypothèses en démontrant la réciproque du théorème ou en la comparant avec des exemples à d'autres théories mathématiques pour voir si l'ensemble forme un tout cohérent	Tester expérimentalement le modèle théorique obtenu et soumettre le travail à comparaison avec d'autres équipes indépendantes. Le nouveau modèle doit prévoir des résultats expérimentaux observés et jamais observés. Si le modèle est validé alors il prend officiellement le statut de *théorie*
D'éventuelles remarques peuvent être indiquées	D'éventuelles remarques peuvent être indiquées

TABLEAU 1 – Les différentes étapes du travail scientifique

REMARQUES

1. Attention, il est très facile de faire des nouvelles théories physiques en alignant des mots. Cela s'appelle de la "philosophie" et les grecs ont pensé aux atomes comme cela. Ca peut donc mener à une vraie théorie. Par contre il est bien plus difficile de faire une "théorie prédictive", c'est-à-dire avec des équations qui permettent de prédire le résultat d'une expérience.
2. Toutefois ce qui sépare la mathématique de la physique est que, en mathématique, l'hypothèse est toujours vraie. Le discours mathématique n'est pas une démonstration d'une vérité extérieure à chercher, mais vise uniquement la cohérence. Ce qui doit être juste est le raisonnement.

Présentons les quatre principes de la méthode de DESCARTES qui, rappelons-le, est considéré comme le premier scientifique de l'histoire de par sa méthode d'analyse :

5. Parfois par abus, *propriétés*, *conditions* et *axiomes* sont confondus alors que le concept d'axiome est beaucoup plus précis et profond.

6. Il ne faut pas cependant oublier que la validité d'un modèle ne dépend pas du réalisme de ses hypothèses mais bien de la conformité de ses implications avec la réalité.

1. Ne recevoir jamais aucune chose pour vraie que je ne la connusse évidemment être telle. C'est-à-dire, d'éviter soigneusement la précipitation et la prévention, et de ne comprendre rien de plus en mes jugements que ce qui se présenterait si clairement et si distinctement à mon esprit, que je n'eusse aucune occasion de le mettre en doute ;
2. De diviser chacune des difficultés que j'examinerais, en autant de parcelles qu'il se pourrait, et qu'il serait requis pour les mieux résoudre ;
3. De conduire par ordre mes pensées, en commençant par les objets les plus simples et les plus aisés à connaître, pour monter peu à peu comme par degrés jusques à la connaissance des plus composés, et supposant même de l'ordre entre ceux qui ne se précèdent point naturellement les uns les autres ;
4. Et le dernier, de faire partout des dénombrements si entiers et des revues si générales, que je fusse assuré de ne rien omettre.

Sur les sciences

Il est important que soient définis rigoureusement les différents types de sciences développés au cours de l'histoire. Effectivement, il semble qu'au 21^{e} siècle un abus de langage s'instaure et qu'il ne devienne plus possible pour la population de distinguer la *qualité intrinsèque* d'une science d'une autre.

Remarque Étymologiquement le mot *science* vient du latin *scienta* (connaissance) dont la racine est le verbe *scire* qui signifie *savoir*.

Cet abus de langage vient probablement du fait que les sciences pures et exactes perdent leurs illusions d'universalité et d'objectivité, dans le sens où elles s'auto-corrigent. Ceci ayant pour conséquence que certaines sciences sont reléguées au second plan et tentent d'en emprunter les méthodes, les principes et les origines pour créer une confusion quant à leurs distinctions.

En soi, la science cependant ne produit pas de vérité absolue. Par principe, une théorie scientifique est valable tant qu'elle permet de prédire des résultats mesurables et reproductibles. Mais les problèmes d'interprétation de ces résultats font partie de la philosophie naturelle.

Étant donné la diversité des phénomènes à étudier, au cours des siècles s'est constitué un nombre grandissant de disciplines comme la chimie, la biologie, la thermodynamique, etc. Toutes ces disciplines à priori hétéroclites ont pour socle commun la physique, pour langage les mathématiques et comme principe élémentaire la méthode scientifique.

Ainsi, un petit rafraîchissement semble nécessaire :

Définition 0.1 *Nous définissons par* science pure, *tout ensemble de connaissances fondées sur un raisonnement rigoureux valable quel que soit le facteur (arbitraire) élémentaire choisi —nous disons alors* indépendant de la réalité sensible— *et restreint au minimum nécessaire. Il n'y a que la mathématique qui puisse être classifiée dans cette catégorie.*

Définition 0.2 *Nous définissons par* science exacte *ou* science dure, *tout ensemble de connaissances fondées sur l'étude d'une observation, observation qui aura été transcrite sous forme symbolique. Principalement, le but des sciences exactes est non d'expliquer le* pourquoi *mais le* comment.

Remarque Les deux définitions précédentes sont souvent incluses dans la définition de *sciences déductives* ou encore de *sciences phénoménologiques*.

Définition 0.3 *Nous définissons par* science de l'ingénieur, *tout ensemble de connaissances théoriques ou pratiques appliquées aux besoins de la société humaine tels que : l'électronique, la chimie, l'informatique, les télécommunications, la robotique, l'aérospatiale, biotechnologies...*

Définition 0.4 *Nous définissons par* science, *tout ensemble de connaissances fondées sur des études ou observations de faits dont l'interprétation n'a pas encore été retranscrite ni vérifiée avec la rigueur mathématique caractéristique des sciences qui précèdent, mais qui applique des raisonnements comparatifs statistiques. Sont inclues dans cette définition la médecine*[7]*, la sociologie, la psychologie, l'histoire, la biologie...*

7. Il faut cependant prendre garde au fait que certaines parties de la médecine étudient des phénomènes descriptifs sous forme mathématique tels que les réseaux de neurones ou autres phénomènes associés à des causes physiques connues.

Selon le philosophe K. POPPER, une théorie n'est scientifiquement acceptable que si, telle qu'elle est présentée, elle peut être falsifiable, c'est à dire soumise à des tests expérimentaux. La *connaissance scientifique* est ainsi par définition l'ensemble des théories qui ont jusqu'alors résisté à la falsification. La science est donc par nature soumise en permanence à la remise en question.

Définition 0.5 *Nous définissons par* science molle *ou* para-sciences, *tout ensemble de connaissances ou de pratiques qui sont actuellement fondées sur des faits invérifiables (non reproductibles scientifiquement) ni par l'expérience, ni par la mathématique. Sont inclues dans cette définition l'astrologie, la théologie, le paranormal, la graphologie...*

Définition 0.6 *Nous définissons par* sciences phénoménologiques *ou* sciences naturelles, *toute science qui n'est pas inclue dans les définitions précédentes : l'histoire, la sociologie, la psychologie, la zoologie, la biologie...*

Définition 0.7 *Le* scientisme *est la doctrine fondamentale suivant laquelle il n'y a de vérité que dans la science.*

Ce qui est intéressant dans cette doctrine, c'est que c'est certainement une des seules qui demande aux gens de devoir réfléchir par eux-mêmes et de comprendre l'environnement qui les entoure en remettant continuellement tout en question et sans ne jamais rien accepter comme acquis.

Terminologie

Le tableau méthodique 1 contient des termes qui peuvent vous sembler inconnus ou barbares. C'est la raison pour laquelle il nous semble fondamental de présenter les définitions de ces derniers, ainsi que de quelques autres tout aussi importants qui peuvent éviter des confusions malheureuses.

Définition 0.8 *Au-delà de son sens négatif, l'idée de* problème *renvoie à la première étape de la démarche scientifique. Formuler un problème est ainsi essentiel à sa résolution et permet de comprendre correctement ce qui fait problème et de voir ce qui doit être résolu.*

Le concept de problème est intimement relié au concept *d'hypothèse* dont nous allons voir la définition ci-dessous.

Définition 0.9 *Une* hypothèse *est toujours, dans le cadre d'une théorie déjà constituée ou sous-jacente, une supposition en attente de confirmation ou d'infirmation qui tente d'expliquer un groupe de faits ou de prévoir l'apparition de faits nouveaux.*

Ainsi, une hypothèse peut être à l'origine d'un problème théorique qu'il faudra formellement résoudre.

Définition 0.10 *Le* postulat *en physique correspond fréquemment à un principe dont l'admission est nécessaire pour établir une démonstration —nous sous-entendons que cela est une proposition non-démontrable.*

Définition 0.11 *Un* principe *(parent proche du postulat) est donc une proposition admise comme base d'un raisonnement ou une règle générale théorique qui guide la conduite des raisonnements qu'il faudra effectuer. En physique, il s'agit également d'une loi générale régissant un ensemble de phénomènes et vérifiée par l'exactitude de ses conséquences.*

REMARQUE Le mot *principe* est utilisé avec abus dans les petites classes ou écoles d'ingénieurs par les professeurs ne sachant, ne voulant ou ne pouvant pas démontrer une relation.

L'équivalent du postulat ou du principe en mathématiques est l'axiome.

Définition 0.12 *Un* axiome *est une vérité ou proposition évidente par elle-même dont l'admission est nécessaire pour établir une démonstration.*

REMARQUE Les axiomes doivent toujours êtres indépendants entre eux —on ne doit pas pouvoir démontrer l'un à partir de l'autre— et non contradictoires —nous disons également parfois qu'ils doivent être *consistants*.

Définition 0.13 *Le* corollaire *est une proposition résultant d'une vérité déjà démontrée. Nous pouvons également dire qu'un corollaire est une conséquence nécessaire et évidente d'un théorème.*

Définition 0.14 *Un* lemme *constitue une proposition déduite d'un ou de plusieurs postulats ou axiomes et dont la démonstration prépare celle d'un théorème.*

Remarque Le concept de lemme, tout comme celui de corollaire, est malheureusement l'exclusivité des mathématiques.

Définition 0.15 *Une* conjecture *constitue une supposition ou opinion fondée sur la vraisemblance d'un résultat mathématique.*

Remarque Beaucoup de conjectures jouent un rôle un peu comparable à des lemmes, car elles sont des passages obligés pour obtenir d'importants résultats.

Définition 0.16 *Par-delà son sens faible de conjecture, une* théorie *ou* théorème *est un ensemble articulé autour d'une hypothèse et étayé par un ensemble de faits ou développements qui lui confèrent un contenu positif et rendent l'hypothèse bien fondée.*

Définition 0.17 *Une* singularité *est une indétermination d'un calcul qui intervient par l'apparition d'une division par le nombre zéro. Ce terme est aussi bien utilisé en mathématique qu'en physique.*

Définition 0.18 *Une* démonstration *constitue un ensemble de procédures mathématiques à suivre pour démontrer le résultat déjà connu ou non d'un théorème.*

Définition 0.19 *Si le mot* paradoxe *signifie étymologiquement "contraire à l'opinion commune". Le* sophisme *quant à lui, est un énoncé volontairement provocateur, une proposition fausse reposant sur un raisonnement apparemment valide. Ainsi parle-t-on du fameux* paradoxe de Zénon *alors qu'il ne s'agit que d'un sophisme. Le paradoxe ne se réduit pas à de la fausseté, mais implique la coexistence de la vérité et de la fausseté, au point qu'on ne parvient plus à discriminer le vrai et le faux. Le paradoxe apparaît alors problème insoluble ou* aporie.

Remarque Ajoutons que les grands paradoxes, par les interrogations qu'ils ont suscitées, ont fait progresser la science et amené des révolutions conceptuelles de grande ampleur, en mathématique comme en physique théorique (les paradoxes sur les ensembles et sur l'infini en mathématique, ceux à la base de la relativité et de la physique quantique).

Science et foi

Nous verrons qu'en science, une théorie est normalement incomplète, car elle ne peut décrire exhaustivement la complexité du monde réel. Il en est ainsi de toutes les théories, comme celle du Big Bang (cf. chapitre d'Astrophysique) ou de l'évolution des espèces (cf. chapitre de Dynamique Des Populations ou de Théorie Des Jeux).

Il convient de distinguer différents courants scientifiques :

- le *réalisme* est une doctrine selon laquelle les théories physiques ont pour objectif de décrire la réalité telle qu'elle est en soi, dans ses composantes inobservables ;
- l'*instrumentalisme* est une doctrine selon les théories sont des outils servant à prédire des observations mais qui ne décrivent pas la réalité en soi ;
- le *fictionnalisme* est le courant selon lequel le contenu référentiel des théories est un leurre, utile seulement pour assurer l'articulation linguistique des équations fondamentales.

Même si aujourd'hui les théories scientifiques ont le soutien de beaucoup de spécialistes, les théories alternatives ont des arguments valables et nous ne pouvons totalement les écarter. Pour autant, la création du monde en sept jours décrite par la Bible ne peut plus être perçue comme un possible, et bien des croyants reconnaissent qu'une lecture littérale est peu compatible avec l'état actuel de nos connaissances et qu'il est plus sage de l'interpréter comme une parabole. Si la science ne fournit jamais de réponse définitive, il n'est plus possible de ne pas en tenir compte.

La foi, qu'elle soit religieuse, superstitieuse, pseudo-scientifique ou autre, a au contraire pour objectif de donner des vérités absolues d'une toute autre nature puisqu'elle relève d'une conviction personnelle invérifiable. En fait, l'une des fonctions des religions est de fournir du sens à des

phénomènes qui ne sont pas explicables rationnellement. Les progrès de la connaissance entraînent donc parfois une remise en cause des dogmes religieux par la science.

A contrario, sauf à prétendre imposer sa foi aux autres, il faut se défier de la tentation naturelle de qualifier de fait scientifiquement prouvé les extrapolations des modèles scientifiques au-delà de leur champ d'application.

Le mot *science* est de plus en plus utilisé pour soutenir qu'il existe des preuves scientifiques là où il n'y a que croyance. Selon ses détracteurs c'est le cas du mouvement de scientologie. Selon ces derniers, nous devrions plutôt parler de *sciences occultes*.

Les sciences occultes et sciences traditionnelles existent depuis l'Antiquité, elles consistent en un ensemble de connaissances et de pratiques mystérieuses ayant pour but de pénétrer et dominer les secrets de la nature. Au cours des derniers siècles, elles ont été progressivement exclues du champ de la science. Le philosophe K. POPPER s'est longuement interrogé sur la nature de la démarcation entre science et pseudo-science. Après avoir remarqué qu'il est possible de trouver des observations pour confirmer à peu près n'importe quelle théorie, il propose une méthodologie fondée sur la réfutabilité. Une théorie doit selon lui, pour mériter le qualificatif de *scientifique*, pouvoir garantir l'impossibilité de certains événements. Elle devient dès lors réfutable, donc, et alors seulement, apte à intégrer la science. Il suffirait en effet d'observer un de ces événements pour invalider la théorie, et orienter par conséquent sur une amélioration de celle-ci.

ARITHMÉTIQUE

1 Théorie de la démonstration 3
Crise des fondements 4 ● Raisonnement hypothético-déductif 7 ● Calcul propositionnel 7 ● Calcul des prédicats 16 ● Démonstrations 20

2 Nombres 29
Bases numériques 30 ● Types de nombres 31

3 Opérateurs arithmétiques 61
Relations binaires 61 ● Lois fondamentales de l'arithmétique 66 ● Polynômes arithmétiques 73 ● Valeur absolue 73 ● Règles de calcul 74

4 Théorie des ensembles 77
Axiomatique de ZERMELO-FRAENKEL 79 ● Opérations ensemblistes 85 ● Fonctions et applications 90 ● Structures algébriques 94 ● Homomorphismes 105

5 Théorie des nombres 109
Principe du bon ordre 109 ● Principe d'induction 110 ● Divisibilité 111

6 Probabilités 125
Univers des événements 125 ● Analyse combinatoire 130 ● Chaînes de MARKOV 134

7 Statistiques 137
Échantillons 138 ● Moyennes 138 ● Types de variables 148 ● Lois de distributions 156 ● Estimateurs de vraisemblance 198 ● Intervalles de confiance 204 ● Loi faible des grands nombres 210 ● Fonction caractéristique 213 ● Théorème central limite 215 ● Tests d'hypothèse 218 ● Calculs d'erreurs/incertitudes 228

1

Théorie de la démonstration

Nous avons choisi de commencer l'étude de la mathématique appliquée par la théorie qui nous semble la plus fondamentale et la plus importante dans le domaine des sciences pures et exactes. La théorie de la démonstration et du calcul propositionnel, ou logique, a trois objectifs dans le cadre de ce site :

1. apprendre au lecteur comment raisonner et à démontrer et cela indépendamment de la spécialisation étudiée ;
2. démontrer qu'une démonstration (son processus) est indépendante du langage utilisé ;
3. se préparer à la théorie de la logique et au théorème d'incomplétude de Gödel ainsi qu'aux automates (cf. chapitre d'Informatique Théorique).

Le dernier point est le plus passionnant car si nous définissons une religion comme un système de pensée qui contient des affirmations indémontrables, alors elle contient des éléments de foi, et Gödel nous enseigne que les mathématiques sont non seulement une religion, mais que c'est la seule religion capable de prouver qu'elle en est une [1] [2].

Souvent, quand un étudiant arrive dans une classe supérieure, il a surtout appris à calculer, à utiliser des algorithmes. Il a très peu appris à raisonner. Pour tous les raisonnements, le support visuel est fort, et les personnes qui ne voient pas qu'en traçant telle ou telle courbe droite la solution apparaît ou qui ne voient pas dans l'espace sont très pénalisées.

Lors des études secondaires, nous manipulons déjà des objets inconnus, mais c'est surtout pour faire des calculs, et quand nous raisonnons sur des objets représentés par des lettres, nous pouvons remplacer ceux-ci visuellement par un nombre réel, un vecteur, etc. À partir d'un certain niveau, nous demandons aux personnes de raisonner sur des structures plus abstraites, et donc de travailler sur des objets inconnus qui sont des éléments d'un ensemble lui-même inconnu, par exemple les éléments d'un groupe• quelconque. Ce support visuel n'existe alors plus. Chap4

Nous demandons ainsi souvent aux étudiants de raisonner, de démontrer des propriétés, mais personne ne leur a jamais appris à raisonner convenablement, à écrire des preuves. Si nous demandons à un étudiant de licence de mathématiques ce qu'est une démonstration, il a quelque difficulté à répondre. Il peut dire que c'est un texte dans lequel on trouve des *mots clés* : *donc*, *parce que*, *si*, *si et seulement si*, *prenons un x tel que*, *supposons que*, *cherchons une contradiction*, etc. Mais il est incapable de donner la grammaire de ces textes ni même ces rudiments, et d'ailleurs, ses enseignants, s'ils n'ont pas suivi de cours, en seraient probablement incapables aussi.

Pour comprendre cette situation, rappelons que pour parler un enfant n'a pas besoin de connaître la grammaire. Il imite son entourage avec succès : un enfant de six ans sait utiliser des phrases déjà compliquées quant à la structure grammaticale sans avoir jamais appris les règles courantes de grammaire. La plupart des enseignants ne connaissent pas non plus la grammaire du raisonnement leur processus d'imitation a fonctionné correctement et ils raisonnent sans erreur. L'expérience montre que ce processus d'imitation est suffisamment bien intégré par les très bons étudiants mais reste utopique chez la plupart.

1. Il est fortement conseillé de lire en parallèle à ce chapitre, ceux sur la théorie des automates et de l'algèbre de Boole disponibles dans la section d'informatique théorique du site.
2. Il faut prendre cette théorie comme une curiosité sympathique qui n'amène fondamentalement pas grand chose excepté des méthodes de raisonnement. Par ailleurs, son objectif n'est pas de démontrer que tout est démontrable mais que toute démonstration peut se faire sur un langage commun à partir d'un certain nombre de règles.

Tant que le degré de complexité est faible, la grammaire ne sert à rien, mais quand il augmente ou quand on ne comprend pas pourquoi quelque chose est faux, il devient nécessaire de faire un peu de grammaire pour pouvoir progresser. Les enseignants et les étudiants connaissent bien la situation suivante : dans un devoir, le correcteur a barré toute une page d'un grand trait rouge et mis *faux* dans la marge. Quand l'étudiant demande ce qui est faux, le correcteur ne peut que dire des choses du genre *ça n'a aucun rapport avec la démonstration demandée*, *rien n'est juste*, ..., ce qui n'aide évidemment pas l'étudiant à comprendre. Cela vient en partie, du fait que le texte rédigé par l'étudiant utilise les mots voulus mais dans un ordre plus ou moins aléatoire et qu'on ne peut donner de sens à l'assemblage de ces mots. De plus, l'enseignant n'a pas les outils nécessaires pour pouvoir expliquer ce qui ne va pas. Il faut donc les lui donner !

Ces outils existent mais sont assez récents. La théorie de la démonstration est une branche de la logique mathématique dont l'origine est la crise des fondements : il y a eu un doute sur ce que nous avions le droit de faire dans un raisonnement mathématique. Des paradoxes sont apparus, et il a alors été nécessaire de préciser les règles de démonstration et de vérifier que ces règles ne sont pas contradictoires. Cette théorie est apparue il y a environ un siècle, ce qui est très peu puisque l'essentiel des mathématiques enseignées en première moitié de l'université est connu depuis deux ou trois cents ans.

1.1 Crise des fondements

1.1.1 Introduction

Pour les premiers Grecs, la géométrie était considérée comme la forme la plus haute du savoir, une puissante clé pour les mystères métaphysiques de l'Univers. Elle était plutôt une croyance mystique, et le lien entre le mysticisme et la religion était rendu explicite dans des cultes comme ceux des Pythagoriciens. Aucune culture n'a depuis déifié un homme pour avoir découvert un théorème géométrique ! Plus tard, les mathématiques furent considérées comme le modèle d'une connaissance *a priori* dans la tradition aristotélicienne du rationalisme.

L'étonnement des Grecs pour les mathématiques ne nous a pas quitté, on le retrouve sous la traditionnelle métaphore des mathématiques comme *Reine des Science*. Il s'est renforcé avec les succès spectaculaires des modèles mathématiques dans la science, succès que les Grecs n'avaient pas prévus. Depuis la découverte par Isaac NEWTON du calcul intégral et de la loi du carré inverse de la gravité, à la fin des années 1600, les sciences phénoménales et les plus hautes mathématiques étaient restées en étroite symbiose —au point qu'un formalisme mathématique prédictif était devenu le signe distinctif d'une *science dure*.

Après NEWTON, pendant les deux siècles qui suivirent, la science aspira à ce genre de rigueur et de pureté qui semblaient inhérentes aux mathématiques. La question métaphysique semblait simple : les mathématiques possédaient une connaissance *a priori* parfaite, et parmi les sciences, celles qui étaient capables de se mathématiser le plus parfaitement étaient les plus efficaces pour la prédiction des phénomènes . La connaissance parfaite consistait donc dans un formalisme mathématique qui, une fois atteint par la science et embrassant tous les aspects de la réalité, pouvait fonder une connaissance empirique *a posteriori* sur une logique rationnelle *a priori*. Ce fut dans cet esprit que Jean-Antoine Nicolas de CARTITAT, marquis de CONDORCET (philosophe et mathématicien français), entreprit d'imaginer la description de l'Univers entier comme un ensemble d'équation différentielles partielles se résolvant les unes par les autres.

La première faille dans cette image inspiratrice apparut dans la seconde moitié du 19^{e} siècle, quand RIEMANN et LOBACHEVSKY prouvèrent séparément que l'axiome des parallèles d'EUCLIDE pouvait être remplacé par d'autres qui produisaient des géométries *consistantes* (nous reviendrons sur ce terme plus loin). La géométrie de RIEMANN prenait modèle sur une sphère, celle de LOBACHEVSKY, sur la rotation d'un hyperboloïde.

L'impact de cette découverte a été obscurci plus tard par de grands chamboulements, mais sur le moment, il fut un coup de tonnerre dans le monde intellectuel. L'existence de systèmes axiomatiques mutuellement inconsistants, et dont chacun pouvait servir de modèle à l'Univers phénoménal, remettait entièrement en question la relation entre les mathématiques et la théorie physique.

Quand on ne connaissait qu'EUCLIDE, il n'y avait qu'une géométrie possible. On pouvait croire que les axiomes d'EUCLIDE constituaient un genre de connaissance parfaite *a priori* sur la géométrie

dans le monde phénoménal. Mais soudain, nous avons eu trois géométries, embarrassantes pour les subtilités métaphysique.

Pourquoi aurions-nous à choisir entre les axiomes de la géométrie plane, sphérique et hyperbolique comme descriptions de la géométrie du réel ? Parce que toutes les trois sont consistantes, nous ne pouvons en choisir aucune comme fondement *a priori* - le choix doit devenir empirique, basé sur leur pouvoir prédictif dans une situation donnée.

Bien sûr, Les théoriciens de la physique ont longtemps été habitués à choisir des formalismes pour poser un problème scientifique. Mais il était admis largement, si ce n'est inconsciemment, que la nécessité de procéder ainsi ad hoc était fonction de l'ignorance humaine, et, qu'avec de la logique ou des mathématiques assez bonnes, on pouvait déduire le bon choix à partir de premiers principes, et produire des descriptions *a priori* de la réalité, qui devaient être confirmées après coup par une vérification empirique.

Cependant, la géométrie euclidienne, considérée pendant plusieurs centaines d'années comme le modèle de la perfection axiomatique des mathématiques, avait été détrônée. Si l'on ne pouvait connaître *a priori* quelque chose d'aussi fondamental que la géométrie dans l'espace, quel espoir restait-il pour une pure théorie rationnelle qui embrasserait la totalité de la nature ? Psychologiquement, Riemann et Lobachevsky avaient frappé au cœur de l'entreprise mathématique telle qu'elle avait été conçue jusqu'alors.

De plus, Riemann et Lobachevsky remettaient la nature de l'intuition mathématique en question. Il avait été facile de croire implicitement que l'intuition mathématique était une forme de perception, une façon d'entrevoir le noumène platonicien derrière la réalité. Mais avec deux autres géométries qui bousculaient celle d'Euclide, personne ne pouvait plus être sûr de savoir à quoi le noumène ressemblait.

Les mathématiciens répondirent à ce double problème avec un excès de rigueur, en essayant d'appliquer la méthode axiomatique à toutes les mathématiques. Dans la période pré-axiomatique, les preuves reposaient souvent sur des intuitions communément admises de la *réalité* mathématique, qui ne pouvaient plus être considérées automatiquement comme valides.

La nouvelle façon de penser les mathématiques conduisait à une série de succès spectaculaires. Pourtant cela avait aussi un prix. La méthode axiomatique rendait la connexion entre les mathématiques et la réalité phénoménale toujours plus étroite. En même temps, des découvertes suggéraient que les axiomes mathématiques qui semblaient être consistants avec l'expérience phénoménale pouvait entraîner de vertigineuses contradictions avec cette expérience.

La majorité des mathématiciens devinrent rapidement des *formalistes*, soutenant que les mathématiques pures ne pouvaient qu'être considérées philosophiquement comme une sorte de jeu élaboré qui se jouait avec des signes sur le papier [3]. La croyance platonicienne en la réalité nouménale des objets mathématiques, à l'ancienne manière, semblait bonne pour la poubelle, malgré le fait que les mathématiciens continuaient à se sentir comme les platoniciens durant le processus de découverte des mathématiques.

Philosophiquement, donc, la méthode axiomatique conduisait la plupart des mathématiciens à abandonner les croyances antérieures en la spécificité métaphysique des mathématiques. Elle produisit aussi la rupture contemporaine entre les mathématiques pures et appliquées. La plupart des grands mathématiciens du début de la période moderne —Newton, Leibniz, Fourier, Gauss et les autres— s'occupaient aussi de science phénoménale. La méthode axiomatique avait couvé l'idée moderne du mathématicien pur comme un super esthète, insoucieux de la physique. Ironiquement, le formalisme donnait aux purs mathématiciens un mauvais penchant à l'attitude platonicienne. Les chercheurs en mathématiques appliquées cessèrent de côtoyer les physiciens et apprirent à se mettre à leur traîne.

Ceci nous emmène au début du 20^{e} siècle. Pour la minorité assiégée des platoniciens, le pire était encore à venir. Cantor, Frege, Russell et Whitehead montrèrent que toutes les mathématiques pures pouvaient être construites sur le simple fondement axiomatique de la théorie des ensembles. Cela convenait parfaitement aux formalistes : les mathématiques se réunifiaient, du moins en principe, à partir d'un faisceau de petits jeux détachés d'un grand. Les platoniciens aussi étaient satisfaits, s'il en survenait une grande structure, clé de voûte consistante pour toutes les mathématiques, la spécificité métaphysique des mathématiques pouvait encore être sauvée.

3. C'est la théorie qui sous-tend la prophétique qualification des mathématiques de *système à contenu nul* par R. Heinlein.

D'une façon négative, pourtant, un platonicien eut le dernier mot. K. GÖDEL mit son grain de sable dans le programme formaliste d'axiomatisation quand il démontra que tout système d'axiomes assez puissant pour inclure les entiers devait être soit inconsistant —contenir des contradictions— soit incomplet —trop faible pour décider de la justesse ou de la fausseté de certaines affirmations du système. Et c'est plus ou moins où en sont les choses aujourd'hui. Les mathématiciens savent que de nombreuses tentatives pour faire avancer les mathématiques comme une connaissance *a priori* de l'Univers doivent se heurter à de nombreux paradoxes et à l'impossibilité de décider quel système axiomatique décrit les mathématiques réelles. Ils ont été réduits à espérer que les axiomatisations standard ne soient pas inconsistantes mais incomplètes, et à se demander anxieusement quelles contradictions ou quels théorèmes indémontrables attendent d'être découverts ailleurs.

Cependant, sur le front de l'empirisme, les mathématiques étaient toujours un succès spectaculaire en tant qu'outil de construction théorique. Les grands succès de la physique du 20^{e} siècle —la relativité générale et la physique quantique— poussaient si loin hors du royaume de l'intuition physique, qu'ils ne pouvaient être compris qu'en méditant profondément sur leurs formalismes mathématiques, et en prolongeant leurs conclusions logiques, même lorsque ces conclusions semblaient sauvagement bizarres. Quelle ironie. Au moment même où la perception mathématique en venait à paraître toujours moins fiable dans les mathématiques pures, elle devenait toujours plus indispensable dans les sciences phénoménales.

À l'opposé de cet arrière-plan, l'applicabilité des mathématiques à la science phénoménale pose un problème plus épineux qu'il n'apparaît d'abord. Le rapport entre les modèles mathématiques et la prédiction des phénomènes est complexe, pas seulement dans la pratique mais dans le principe. D'autant plus complexe que, comme nous le savons maintenant, il y a des façons d'axiomatiser les mathématiques qui s'excluent !

Mais pourquoi existe-t-il seulement de bons choix de modèle mathématique ? C'est-à-dire, pourquoi y a-t-il un formalisme mathématique, par exemple pour la physique quantique, si productif qu'il prédit réellement la découverte de nouvelles particules observables ?

Pour répondre à cette question nous observerons qu'elle peut, aussi bien, fonctionner comme une sorte de définition. Pour beaucoup de système phénoménaux, de tels formalismes prédictifs exacts n'ont pas été trouvés, et aucun ne semble plausible. Les poètes aiment marmonner sur le cœur des hommes, mais on peut trouver des exemples plus ordinaires : le climat, où le comportement d'une économie supérieure à celle d'un village, par exemple, systèmes si chaotiquement interdépendants que la prédiction exacte est effectivement impossible.

1.1.2 Paradoxes

Dès l'antiquité, certains logiciens avaient constaté la présence de nombreux paradoxes au sein de la rationalité. En fait, nous pouvons dire que malgré leur nombre, ces paradoxes ne sont que les illustrations d'un petit nombre de structures paradoxales. Attardons nous à exposer à titre de culture de générale les plus connus.

Paradoxe 1.1 (Paradoxe de la classe des classes—Russell) *Il existe deux types de classes : celles qui se contiennent elles-mêmes (ou classes réflexives : la classe des ensembles non-vides, la classe des classes...) et celles qui ne se contiennent pas elles-mêmes (ou classes irréflexives : la classe des travaux à rendre, la classe des oranges sanguines...). La question est la suivante : la classe des classes irréflexives est-elle elle même réflexive ou irréflexive ? Si elle est réflexive, elle se contient et se trouve rangée dans la classe des classes irréflexives qu'elle constitue, ce qui est contradictoire. Si elle est irréflexive, elle doit figurer dans la classe des classes irréflexives qu'elle constitue et devient ipso facto réflexive, nous sommes face à une nouvelle contradiction.*

Paradoxe 1.2 (Paradoxe du bibliothécaire—Gonseth) *Dans une bibliothèque, il existe deux types de catalogues. Ceux qui se mentionnent eux-mêmes et ceux qui ne se mentionnent pas. Un bibliothécaire doit dresser le catalogue de tous les catalogues qui ne se mentionnent pas eux-mêmes. Arrivé au terme de son travail, notre bibliothécaire se demande s'il convient ou non de mentionner le catalogue qu'il est précisément en train de rédiger. À ce moment, il est frappé de perplexité. Si ne le mentionne pas, ce catalogue sera un catalogue qui ne se mentionne pas et qui devra dès lors figurer dans la liste des catalogues ne se mentionnant pas eux-mêmes. D'un autre côté, s'il le mentionne, ce catalogue deviendra un catalogue qui se mentionne et qui ne doit donc pas figurer dans ce catalogue, puisque celui-ci est le catalogue des catalogues qui ne se mentionnent pas.*

Paradoxe 1.3 (Paradoxe du menteur (variante)) *Définissons provisoirement le mensonge comme l'action de formuler une proposition fausse. Le poète crétois* Épiménide *affirme : « Tous les Crétois sont des menteurs », soit la proposition P. Comment décider de la valeur de vérité de P ? Si P est vraie, comme* Épiménide *est Crétois, P doit être fausse. Il faut donc que P soit fausse pour pouvoir être vraie, ce qui est contradictoire. P est donc fausse. Remarquons qu'on ne peut pas en déduire, comme dans le véritable paradoxe du menteur, que P doit aussi être vraie.*

1.2 Raisonnement hypothético-déductif

Le raisonnement hypothético-déductif est, nous le savons, la capacité qu'a l'apprenant de déduire des conclusions à partir de pures hypothèses et pas seulement d'une observation réelle. C'est un processus de réflexion qui tente de dégager une explication causale d'un phénomène quelconque (nous y reviendrons lors de nos premiers pas en physique). L'apprenant qui utilise ce type de raisonnement commence par formuler une hypothèse et essaie ensuite de confirmer ou d'infirmer son hypothèse selon le schéma synoptique 1.1. QLa procédure déductive consiste à tenir pour vrai,

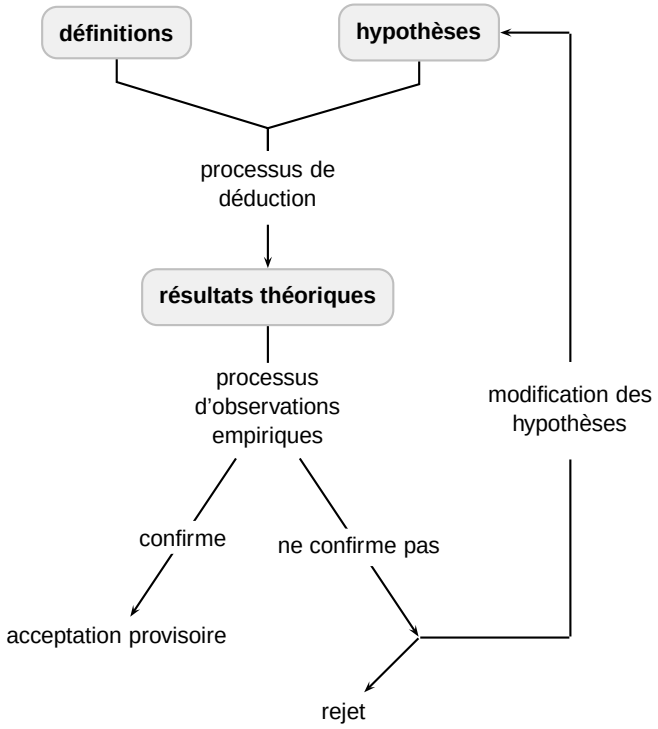


Figure 1.1 – Schéma synoptique de raisonnement hypothético-déductif

à titre provisoire, cette proposition première que nous appelons, en logique *le prédicat* (voir plus bas) et à en tirer toutes les conséquences logiquement nécessaires, c'est-à-dire à en rechercher les implications [4].

Exemple 1.1 *Soit la proposition P : X* est un homme, *elle implique la proposition suivante Q : X est mortel. L'expression P ⇒ Q est un implication prédicative (d'où le terme* prédicat*). Il n'y a pas dans cet exemple de cas où nous puissions énoncer P sans Q. Cet exemple est celui d'une implication stricte, telle que nous la trouvons dans le* syllogisme *(figure logique du raisonnement).*

1.3 Calcul propositionnel

Le calcul propositionnel —ou logique propositionnelle— est un préliminaire absolument indispensable pour aborder une formation en sciences, philosophie, droit, politique, économie, etc. Ce type

4. Des spécialistes ont montré que le raisonnement hypothético-déductif s'élabore progressivement chez l'enfant, à partir de 6-7ans, et que ce type de raisonnement n'est utilisé systématiquement, en partant d'une fonction propositionnelle stricte qu'à partir de 11-12 ans.

de calcul autorise des procédures de décision ou tests. Ceux-ci permettent de déterminer dans quel cas une expression (proposition) logique est vraie et en particulier si elle est toujours vraie [5] [6].

Définition 1.1 *Une expression toujours vraie quel que soit le contenu linguistique des variables qui la composent est appelée une* expression valide, *une* tautologie *ou encore une* loi de la logique propositionnelle.

Définition 1.2 *Un expression toujours fausse est appelée une* contradiction *ou* antologie.

Définition 1.3 *Une expression qui est parfois vraie, parfois fausse est appelée une* expression contingente.

Définition 1.4 *Nous appelons* assertion *une expression dont nous pouvons dire sans ambiguïté s'il elle est vraie ou fausse.*

Définition 1.5 *Le* langage objet *est le langage utilisé pour écrire les expressions logiques.*

Définition 1.6 *Le* métalangage *est le langage utilisé pour parler du langage objet dans la langue courante.*

1.3.1 Propositions

Définition 1.7 *En logique, une* proposition *est une affirmation qui a un sens. Cela veut dire que nous pouvons dire sans ambiguïté si cette affirmation est vraie (V) ou fausse (F). C'est ce que nous appelons le* principe du tiers exclu.

Exemple 1.2 *« Je mens »n'est pas une proposition. Si nous supposons que cette affirmation est vraie, elle est une affirmation de sa propre invalidité, donc nous devrions conclure qu'elle est fausse. Mais si nous supposons qu'elle est fausse, alors l'auteur de cette affirmation ne ment pas, donc il dit la vérité, aussi la proposition serait vraie.*

Définition 1.8 *Une proposition en logique binaire —où les propositions sont soit vraies, soit fausses— n'est donc jamais vraie et fausse à la fois. C'est que nous appelons le* principe de non-contradiction.

Ainsi, une propriété sur l'ensemble E des propositions est une application P de E dans l'ensemble des *valeurs de vérité* :

$$P : E \to (V, F) \tag{1.1}$$

Nous parlons de *sous-ensemble associé* lorsque la proposition engendre uniquement une partie E' de E et inversement.

Exemple 1.3 *Dans $E = \mathbb{N}$, si $P(x)$ s'énonce x* est pair, *alors $P = (0, 2, 4, \ldots, 2k, \ldots)$ ce qui est bien seulement un sous-ensemble associée de E mais de même cardinal*•.

Chap4

Définition 1.9 *Soit P une propriété sur l'ensemble E. Une propriété Q sur E est une* négation *de P si et seulement si, pour tout $x \in E$:*

- *$Q(x)$ est F si $P(x)$ est V ;*
- *$P(x)$ est V si $P(x)$ est F.*

Nous pouvons rassembler ces conditions dans une table dite *table de vérité* dont un exemple est donné dans le tableau 1.1. En d'autres termes, P et Q ont toujours des valeurs de vérité contraires. Nous noterons ce genre d'énoncé *Q est une négation de P* :

$$Q \Leftrightarrow \neg P \tag{1.2}$$

où le symbole $\neg$ est le *connecteur de négation.*

REMARQUE Les expressions doivent être des expressions bien formées (souvent abrégé *ebf*). Par définition, toute variable est une expression bien formée, alors $\neg P$ est une expression bien formée. Si P et Q sont des expressions bien formées, alors $P \Rightarrow Q$ est une expression bien formée (l'expression « je mens » n'est pas bien formée car elle se contredit elle-même).

5. Il existe des expressions qui ne sont effectivement pas des assertions. Par exemple, l'énoncé : « cet énoncé est faux », est un paradoxe qui ne peut être ni vrai, ni faux.

6. Soit un expression logique A. Si celle-ci est une tautologie, nous la notons fréquemment $\vDash A$ et s'il l'expression est une contradiction, nous la notons $A \vDash$.

P	Q
V	F
F	V

TABLEAU 1.1 – Exemple de table de vérité

1.3.2 Connecteurs

Il y a d'autres types de connecteurs en logique. Soit P et Q deux propriétés définies sur le même ensemble E. $P \vee Q$ (P ou Q) est une propriété sur E définie par :

- $P \vee Q$ est vraie si au moins l'une des propriétés P, Q est vraie ;
- $P \vee Q$ est fausse sinon.

Nous pouvons créer la table de vérité du *connecteur OU* ou *connecteur de disjonction* $\vee$:

P	Q	$P \vee Q$
V	V	V
V	F	V
F	V	V
F	F	F

TABLEAU 1.2 – Table de vérité du connecteur OU ($\vee$)

Il est facile de se convaincre que, si les parties P, Q de E sont respectivement associées aux propriétés P, Q que $P \cup Q$ (voir théorie des ensembles) est associé à $P \vee Q$.

$$\begin{aligned} &P \leftrightarrow P \\ &Q \leftrightarrow Q \\ &P \vee Q \leftrightarrow P \vee Q \end{aligned} \tag{1.3}$$

Le connecteur $\vee$ est associatif. Pour s'en convaincre, il suffit de faire une table vérité où nous vérifions que :

$$[P \vee (Q \vee R)] \Leftrightarrow [(P \vee Q) \vee R] \tag{1.4}$$

Il existe également le *connecteur ET* ou *connecteur de conjonction* $\wedge$ pour que que soient P, Q deux propriétés définies sur E, $P \wedge Q$ est une propriété sur E définie par :

- $P \wedge Q$ est vraie si toutes les deux propriétés P, Q sont vraies ;
- $P \wedge Q$ est fausse sinon.

Nous pouvons créer la table de vérité du connecteur $\wedge$:

P	Q	$P \wedge Q$
V	V	V
V	F	F
F	V	F
F	F	F

TABLEAU 1.3 – Table de vérité du connecteur ET ($\wedge$)

Il est également facile de se convaincre que, si les parties P, Q de E sont respectivement associées aux propriétés P, Q que $P \cap Q$ (voir théorie des ensembles) est associé à $P \wedge Q$:

$$\begin{aligned} &P \leftrightarrow P \\ &Q \leftrightarrow Q \\ &P \wedge Q \leftrightarrow P \cap Q \end{aligned} \tag{1.5}$$

Le connecteur $\wedge$ est associatif. Pour s'en convaincre, il suffit aussi de faire une table vérité où nous vérifions que :

$$[P \wedge (Q \wedge R)] \Leftrightarrow [(P \wedge Q) \wedge R] \tag{1.6}$$

Les connecteurs $\vee, \wedge$ sont distributifs l'un sur l'autre. À l'aide d'une simple table de vérité, nous prouvons que :

$$[P \vee (Q \wedge R)] \Leftrightarrow [(P \vee Q) \wedge (P \vee R)] \tag{1.7}$$

ainsi que :

$$[P \wedge (Q \vee R)] \Leftrightarrow [(P \wedge Q) \vee (P \wedge R)] \tag{1.8}$$

Une négation de $P \vee Q$ est $[(\neg P) \wedge (\neg Q)]$ une négation de $P \wedge Q$ est $[(\neg P) \vee (\neg Q)]$ tel que :

$$\begin{array}{l} \neg(P \vee Q) \Leftrightarrow [(\neg P) \wedge (\neg Q)] \\ \neg(P \wedge Q) \Leftrightarrow [(\neg P) \vee (\neg Q)] \end{array} \tag{1.9}$$

À nouveau, ces propriétés peuvent se démontrer par une simple table de vérité[7].

Revenons maintenant sur le *connecteur d'implication logique* noté $\Rightarrow$[8]. Soient P, Q deux propriétés sur E. $P \Rightarrow Q$ est une propriété sur E définie par :

- $P \Rightarrow Q$ est fausse si P est vraie et Q fausse ;
- $P \Rightarrow Q$ est vraie sinon.

En d'autres termes, P implique logiquement Q signifie que Q est vrai pour toute évaluation pour laquelle P est vraie. L'implication représente donc le *si... alors...* La table de vérité de l'implication est indiquée dans le tableau 1.4. Si $P \Rightarrow Q$, pour que Q soit vraie, il suffit que P soit vraie :

P	Q	$P \Rightarrow Q$
V	V	V
V	F	F
F	V	V
F	V	V

TABLEAU 1.4 – Table de vérité de l'implication

effectivement l'implication sera vraie si P est vraie ou fausse selon la table de vérité. Donc P est une condition suffisante de Q. D'un autre côté, $P \Rightarrow Q$ est équivalent à $\neg Q \Rightarrow \neg P$. Donc, si Q est fausse, il est impossible que P soit vraie. Finalement, Q est une condition nécessaire de P.

Exemple 1.4 *Soit la proposition : « Si tu obtiens ton diplôme, je t'achète un ordinateur. » Parmi tous les cas, un seul correspond à une promesse non tenue : celui où l'enfant à son diplôme, et n'a toujours pas d'ordinateur (deuxième ligne dans le tableau). Et le cas où il n'a pas le diplôme, mais reçoit quand même un ordinateur ? Il est possible qu'il ait été longtemps malade et a raté un semestre, et le père a le droit d'être bon.*

Que signifie cette promesse, que nous écrirons aussi : « Tu as ton diplôme $\Rightarrow$ je t'achète un ordinateur » ? Exactement ceci :

- *Si tu as ton diplôme, c'est sûr, je t'achète un ordinateur (je ne peux pas ne pas l'acheter)*
- *Si tu n'as pas ton diplôme, je n'ai rien dit*

Exemple 1.5 *De toute proposition fausse nous pouvons déduire toute propositions (deux dernières lignes). C'est un exemple plutôt anecdotique : dans un cours de RUSSELL portant sur le fait que d'une proposition fausse, toute proposition peut être déduite, un étudiant lui posa la question suivante :*

7. Pour voir les détails de tous les opérateurs logiques, le lecteur devra se rendre dans le chapitre d'Algèbre De BOOLE (cf. section d'Informatique Théorique) où l'identité, la double négation, l'idempotence, l'associativité, la distributivité, les relations de DE MORGAN sont présentées plus formellement.

8. Dans certains ouvrages sur le calcul propositionnel, ce connecteur est noté $\supset$ et dans le cadre de la théorie de la démonstration nous lui préférons souvent le symbole $\rightarrow$.

– *« Prétendez-vous que de 2 + 2 = 5, il s'ensuit que vous êtes le pape ? »*
– *« Oui », fit Russell*
– *« Et pourriez-vous le prouver ? », demanda l'étudiant sceptique*
– *« Certainement ! », réplique Russell, qui proposa sur le champ la démonstration suivante :*
  1. *Supposons que 2 + 2 = 5*
  2. *Soustrayons 2 de chaque membre de l'égalité, nous obtenons 2 = 3*
  3. *Par symétrie, 3 = 2*
  4. *Soustrayant 1 de chaque côté, il vient 2 =1*

Maintenant le pape et moi sommes deux. Puisque 2 = 1, le pape et moi sommes un. Par suite, je suis le pape.

Le connecteur d'implication est essentiel en mathématiques, philosophie, etc. C'est un des fondements de toute démonstration, preuve ou déduction.

Le connecteur d'implication a les propriétés suivantes :

$$\begin{aligned} &P \Rightarrow Q \Leftrightarrow [(\neg Q) \Rightarrow (\neg P)] \\ &P \Rightarrow Q \Leftrightarrow [(\neg P) \vee Q] \end{aligned} \tag{1.10}$$

conséquence de la dernière propriété :

$$\neg(P \Rightarrow Q) \Leftrightarrow [P \wedge (\neg Q)] \tag{1.11}$$

Le *connecteur d'équivalence logique* ou *bi-conditionnel noté* $\Leftrightarrow$ *ou* $\leftrightarrow$ signifiant par définition que :

$$\neg(P \Leftrightarrow Q) \Leftrightarrow (P \Rightarrow Q) \wedge (Q \Rightarrow P) \tag{1.12}$$

en d'autres termes, la première expression a la même valeur pour toute évaluation de la deuxième.

Ce que nous pouvons vérifier à l'aide d'une table de vérité 1.5. $P \Leftrightarrow Q$ signifie bien que P *et* Q

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$P \Leftrightarrow Q$
V	V	V	V	V
V	F	F	V	F
F	V	V	F	F
F	F	V	V	V

Tableau 1.5 – Table de vérité de l'équivalence

ont toujours la même valeur de vérité ou encore P *et* Q *sont équivalents*. C'est vrai si P et Q ont même valeur, faux dans tout cas contraire. Bien évidemment (c'est une tautologie) :

$$\neg(P \Leftrightarrow Q) \Rightarrow [(P) \Rightarrow (\neg Q)] \tag{1.13}$$

La relation $P \Leftrightarrow Q$ équivaut donc à ce que P soit une condition nécessaire et suffisante de Q et à ce que Q soit une condition nécessaire et suffisante de P.

Par conséquent, les conditions de type *nécessaire, suffisant, nécessaire et suffisant* peuvent être reformulées avec les termes *seulement si*, *si*, *si et seulement si*. Ainsi :

1. $P \Rightarrow Q$ traduit le fait que Q est une condition nécessaire pour P ou dit autrement, P est vraie seulement si Q est vraie (dans le table de vérité, lorsque $P \Rightarrow Q$ prend la valeur 1 on constate bien que P vaut 1 seulement si Q vaut 1 aussi). On dit aussi, si P est vraie alors Q est vraie.
2. $P \Leftarrow Q$ ou ce qui revient au même $Q \Rightarrow P$ traduit le fait que Q est une condition suffisante pour P ou dit autrement, P est vraie si Q est vraie (dans le table de vérité, lorsque $Q \Rightarrow P$ prend la valeur 1 on constate bien que P vaut 1 si Q vaut 1 aussi).
3. $P \Leftrightarrow Q$ traduit le fait que Q est une condition nécessaire et suffisante pour P ou dit autrement, P est vraie si et seulement si Q est vraie (dans le table de vérité, lorsque $P \Leftrightarrow Q$ prend la valeur 1 on constate bien que P vaut 1 si Q vaut 1 et seulement si Q vaut 1).

REMARQUE L'expression *si et seulement si* correspond donc à une équivalence logique et ne peut être utilisée pour décrire un implication.

La première étape du calcul propositionnel est donc la formalisation des énoncés du langage naturel. Pour réaliser ce travail, le calcul propositionnel fournit finalement trois types d'outils :

1. Les *variables propositionnelles* $(P, Q, R, \ldots)$ symbolisent des propositions simples quelconques. Si la même variable apparaît plusieurs fois, elle symbolise chaque fois la même proposition.
2. Les cinq opérateurs logiques :$\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$
3. Les signes de ponctuation se réduisent aux seules parenthèses ouvrante et fermante qui organisent la lecture de manière à éviter toute ambiguïté.

Description	Symbole	Utilisation
La *négation* est un opérateur qui ne porte que sur une proposition, il est unaire ou monadique. « Il ne pleut pas » s'écrit $\neg P$. Cet énoncé est vrai si et seulement si P est faux. L'usage classique de la négation est caractérisé par la loi de double négation : $\neg\neg P$ est équivalent à P.	$\neg$	$\neg P$
La *conjonction* ou *produit logique* est un opérateur binaire, elle met en relation deux propositions. « Tout homme est mortel » ET « Ma voiture perd de l'huile »s'écrit $(P \wedge Q)$. Cette dernière expression est vraie si et seulement si P est vrai et Q est vrai.	$\wedge$	$(P \wedge Q)$
La *disjonction* ou *somme logique* est, elle aussi, un opérateur binaire. $(P \vee Q)$ est vrai si et seulement si P est vrai ou Q est vrai. Nous pouvons comprendre ce OU de deux façons : soit de manière inclusive, soit de manière exclusive. Dans le premier cas, $(P \vee Q)$ est vrai si P est vrai, si Q est vrai ou si P et Q sont tous deux vrais. Dans le second cas, $(P \vee Q)$ est vrai si P est vrai ou si Q est vrai mais pas si les deux le sont. La disjonction du calcul propositionnel est le OU inclusif et on donne au OU exclusif le nom *d'alternative*.	$\vee$	$(P \vee Q)$
L'implication est également un opérateur binaire. Elle correspond, en gros, au schéma linguistique *Si...alors...* : « *si j'ai le temps, j'irai au cinéma* » s'écrit $P \Rightarrow Q$. $P \Rightarrow Q$ est faux si P est vrai et Q est faux. Si le conséquent (ici Q) est vrai, l'implication $(P \Rightarrow Q)$ est vraie. Lorsque l'antécédent (ici P) est faux, l'implication est toujours vraie. Cette dernière remarque peut être comprise si l'on se réfère à des énoncés de type : « Si on pouvait mettre Paris en bouteille, on utiliserait la tour Eiffel comme bouchon. » En résumé, une implication est fausse si et seulement si son antécédent est vrai et son conséquent est faux.	$\Rightarrow$	$(P \Rightarrow Q)$
La *bi-implication* est, elle aussi, binaire : elle symbolise les expressions *...si et seulement si...* et *...est équivalent à...* L'équivalence entre deux propositions est vraie si celles-ci ont la même valeur de vérité. La bi-implication exprime donc aussi une forme d'identité et c'est pourquoi elle est souvent utilisée dans les définitions.	$\Leftrightarrow$	$(P \Leftrightarrow Q)$

TABLEAU 1.6 – Récapitulatif des symboles

Il est possible d'établir des équivalences entre ces opérateurs. Nous avons déjà vu comment le bi-conditionnel pouvait se définir comme un produit de conditionnels réciproques, voyons maintenant d'autres équivalences :

$$\begin{aligned}
&(P \Rightarrow Q) \Leftrightarrow \neg(P \wedge \neg Q)\\
&(P \Rightarrow Q) \Leftrightarrow (\neg P \vee Q)\\
&(P \vee Q) \Leftrightarrow (\neg P \Rightarrow Q)\\
&(P \wedge Q) \Leftrightarrow \neg(P \Rightarrow \neg Q)
\end{aligned} \tag{1.14}$$

Les opérateurs classiques $\wedge, \vee, \Leftrightarrow$ peuvent donc être définis à l'aide de $\neg, \Rightarrow$ grâce aux lois d'équivalence entre opérateurs. Sont à noter également les deux relations de De MORGAN (cf. chapitre

d'Algèbre de Boole) :

$$\begin{aligned} \neg(P \vee Q) &\Leftrightarrow (\neg P \wedge \neg Q) \\ \neg(P \wedge Q) &\Leftrightarrow (\neg P \vee \neg Q) \end{aligned} \tag{1.15}$$

Elles permettent de transformer la disjonction en conjonction et vice-versa :

$$\begin{aligned} (P \vee Q) &\Leftrightarrow \neg(\neg P \wedge \neg Q) \\ (P \wedge Q) &\Leftrightarrow \neg(\neg P \vee \neg Q) \end{aligned} \tag{1.16}$$

1.3.3 Procédures de décision

Nous avons introduit précédemment les éléments de base nous permettant d'opérer sur des expressions à partir de propriétés (variables propositionnelles) sans toutefois dire grand chose quant à la manipulation de ces expressions. Alors, il convient maintenant de savoir qu'en calcul propositionnel qu'il existe deux manières d'établir qu'une proposition est un loi de la logique propositionnelle. Nous pouvons [9] :

1. employer des procédures non axiomatisées ;
2. recourir à des procédures axiomatiques et démonstratives.

Procédures de décisions non axiomatisées

Plusieurs de ces méthodes existent mais nous nous limiterons ici à la plus simple et à la plus parlante d'entre elles, celle du calcul matriciel, souvent appelée aussi *méthodes des tables de vérité*.

La procédure de construction est comme nous l'avons vu précédemment assez simple. Effectivement, la valeur de vérité d'une expression complexe est fonction de la valeur vérité des énoncés plus simples qui la composent, et finalement fonction de la valeur de vérité des variables propositionnelles qui la composent. En envisageant toutes les combinaisons possibles des valeurs de vérité des variables de propositionnelles, nous pouvons déterminer les valeurs de vérité de l'expression complexe.

Les tables de vérité permettent donc de décider, à propos de toute proposition, si celle-ci est une tautologie (toujours vraie), une contradiction (toujours fausse) ou une expression contingente (parfois vraie, parfois fausse).

Nous pouvons ainsi distinguer quatre façons de combiner les variables propositionnelles, les parenthèses et les connecteurs définis dans le tableau 1.7. La méthode des tables de vérité permet de

	NOM	DESCRIPTION	EXEMPLE
1	énoncé mal formé	non-sens : ni vrai, ni faux	$(\vee P)Q$
2	tautologie	énoncé toujours vrai	$P \vee \neg Q$
3	contradiction	énoncé toujours faux	$(P \wedge \neg P)$
4	énoncé contingent	énoncé parfois vrai, parfois faux	$(P \vee Q)$

Tableau 1.7 – Variables propositionnelles

déterminer le type d'expression bien formée face auquel nous nous trouvons. Elle n'exige en principe aucune invention, c'est une procédure mécanique. Les procédures axiomatisées, en revanche, ne sont pas entièrement mécaniques. Inventer une démonstration dans le cadre d'un système axiomatisé demande parfois de l'habilité, de l'habitude ou de la chance. Pour ce qui est des tables de vérité, voici la marche à suivre : lorsqu'on se trouve face à un expression bien formée, ou fonction de vérité, nous commençons par déterminer à combien de variables propositionnelles distinctes nous avons affaire. Ensuite, nous examinons les différents arguments qui constituent cette expression. Nous construisons alors un tableau comprenant 2^n rangées (n étant le nombre de variables) et un nombre de colonnes égal au nombre d'arguments plus des colonnes pour l'expression elle-même et ses autres composantes. Nous attribuons alors aux variables les différentes combinaisons de vérité et de fausseté qui peuvent leur être conférées (la vérités est exprimée dans la table par un 1 et la

9. Dans de nombreux ouvrages ces procédures sont présentées avant même la structure du langage propositionnel. Nous avons choisi de faire le contraire pensant que l'approche serait plus aisée.

fausseté par un 0). Chacune des rangées correspond à un monde possible et la totalité des rangées constitue l'ensemble des mondes possibles. Il existe, par exemple, un mode possible dans lequel P est une proposition vraie tandis que Q est fausse.

Procédures de décisions axiomatisées

L'axiomatisation d'une théorie implique, outre la formalisation de celle-ci, que nous partions d'un nombre fini d'axiomes et que, grâce à la transformation réglée de ces derniers, que nous puissions obtenir tous les théorèmes de cette théorie. Nous pardons donc de quelques axiomes dont la vérité est posée (et non démontrée). Nous déterminons des règles de déduction permettant de manipuler les axiomes ou toute expression obtenue à partir de ceux-ci. L'enchaînement de ces déductions est une démonstration qui conduit à un théorème, à une loi.

Nous allons sommairement présenter deux systèmes axiomatiques, chacun étant constitué d'axiomes utilisant deux règles dites *règles d'inférence* (règles intuitives) particulières :

Règle 1.1 (modus ponens) *Si nous avons prouvé A et $A \Rightarrow B$, alors nous pouvons déduire B. A est appelé la* prémisse mineure *et $A \Rightarrow B$ la prémisse majeure de la règle du modus ponens.*

Exemple 1.6 *De $x > y$ et $(x > y) \Rightarrow (y < x)$ nous pouvons déduire $(y < x)$*

Règle 1.2 (substitution) *Nous pouvons dans un schéma d'axiome remplacer une lettre par une formule quelconque, pourvue que toutes les lettres identiques soient remplacées par des formules identiques.*

Donnons à titre d'exemple, deux systèmes axiomatiques : le système axiomatique de WHITEHEAD et RUSSELL, le système axiomatique de LUKASIEWICZ.

Système axiomatique de WHITEHEAD et RUSSELL Il adopte comme symboles primitifs $\neg, \vee$ et définit $\Rightarrow, \Leftrightarrow, \wedge$ à partir de ces derniers de la manière suivante :

$$\begin{aligned} &(A \Rightarrow B) \Leftrightarrow \neg A \vee B \\ &(A \wedge B) \Leftrightarrow \neg(\neg A \vee \neg B) \\ &(A \Leftrightarrow B) \Leftrightarrow (A \Rightarrow B) \Rightarrow (B \Rightarrow A) \end{aligned} \tag{1.17}$$

Nous avions déjà présenté plus haut quelque uns de ces éléments.

Ce système comprend cinq axiomes, assez évidents en soi plus les deux règles d'inférence. Les axiomes sont donnés ici en utilisant des symboles non primitifs, comme le faisaient WHITEHEAD et RUSSELL [10] :

Axiome 1.1 $(A \vee A) \Rightarrow A$

Axiome 1.2 $B \Rightarrow (A \vee B)$

Axiome 1.3 $(A \vee B) \Rightarrow (B \vee A)$

Axiome 1.4 $(A \vee (B \vee C)) \Rightarrow (B \vee (A \vee C))$

Axiome 1.5 $(B \Rightarrow C) \Rightarrow ((A \vee B) \Rightarrow (A \vee C))$

Exemple 1.7 *Pour prouver $\neg A \vee A$, nous pouvons procéder ainsi :*

$B \Rightarrow (A \vee B)$	*axiome A2*	(1.18a)
$A \Rightarrow (A \vee A)$	*(1) et subst.*	(1.18b)
$(B \Rightarrow C) \Rightarrow ((A \vee B) \Rightarrow (A \vee C))$	*axiome A5 (nécessaire)*	(1.18c)
$(B \Rightarrow C) \Rightarrow ((\neg A \vee B) \Rightarrow (\neg A \vee C))$	*(3) et subst.*	(1.18d)
$(B \Rightarrow C) \Rightarrow ((A \Rightarrow B) \Rightarrow (A \Rightarrow C))$	*(4) et définition de* $\Rightarrow$	(1.18e)
$((A \vee A) \Rightarrow A) \Rightarrow ((A \Rightarrow (A \vee A)) \Rightarrow (A \Rightarrow A))$	*(5) et subst.*	(1.18f)
$(A \Rightarrow (A \vee A)) \Rightarrow (A \Rightarrow A)$	*(6) modus ponens*	(1.18g)
$(A \Rightarrow A) \Rightarrow (A \Rightarrow A)$	*(2) et (8)*	(1.18h)
$A \Rightarrow A$	*(8) modus ponens*	(1.18i)
$\neg A \vee A$		(1.18j)

10. Ces cinq axiomes ne sont pas indépendants les uns des autres ; le quatrième peut être obtenu à partir des quatre autres.

Système axiomatique de Lukasiewicz Il comprend les trois axiomes suivants, plus les deux règles d'interférences :

Axiome 1.6 $(A \Rightarrow B) \Rightarrow (B \Rightarrow C) \Rightarrow (A \Rightarrow C))$

Axiome 1.7 $A \Rightarrow (\neg A \Rightarrow B)$

Axiome 1.8 $(\neg A \Rightarrow A) \Rightarrow A$

Voici des preuves des deux premiers axiomes, dans le système de Whitehead et Russell. Ce sont les formules (6) et (16) de la dérivation suivante :

$(A \vee (B \vee C)) \Rightarrow (B \vee (A \vee C))$	axiome A4	(1.19a)
$(\neg(B \Rightarrow C) \vee (\neg(A \vee B) \vee (A \vee)))$ $\Rightarrow (\neg(A \vee B) \vee (\neg(B \Rightarrow C) \vee (A \vee C)))$	(1) et subst.	(1.19b)
$\neg(A \vee B) \vee (\neg(B \Rightarrow C) \vee (A \vee C))$	A4 sur (2) et modus ponens	(1.19c)
$(A \vee B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \vee C))$	(3) par déf. de $\Rightarrow$	(1.19d)
$(\neg A \vee B) \Rightarrow ((B \Rightarrow C) \Rightarrow (\neg A \vee C))$	(4) et subs.	(1.19e)
$(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))$	(5) par déf. de $\Rightarrow$	(1.19f)
$(B \Rightarrow (A \vee B)) \Rightarrow ((A \vee B) \Rightarrow (B \vee A)) \Rightarrow (B \Rightarrow (B \vee A))$	(6) et subs.	(1.19g)
$((A \vee B) \Rightarrow (B \vee A)) \Rightarrow (B \Rightarrow (B \vee A))$	(7) modus ponens	(1.19h)
$B \Rightarrow (B \vee A)$	(8) modus ponens	(1.19i)
$\neg B \Rightarrow (\neg B \vee A)$	(9) et subs.	(1.19j)
$\neg\neg B \vee (\neg B \vee A)$	(10) et déf. de $\Rightarrow$	(1.19k)
$(\neg\neg B \vee (\neg B \vee A)) \Rightarrow (\neg B \vee (\neg\neg B \vee A))$	A4 + subst. avec (11)	(1.19l)
$\neg B \vee (\neg\neg B \vee A)$	(12) et modus ponens	(1.19m)
$B \Rightarrow (\neg\neg B \vee A)$	(13) et déf. de $\Rightarrow$	(1.19n)
$B \Rightarrow (\neg B \Rightarrow A)$	(14) et déf. de $\Rightarrow$	(1.19o)
$A \Rightarrow (\neg A \Rightarrow B)$	(16) et subst.	(1.19p)

Ces axiomatisations permettent de retrouver comme théorème toutes les tautologies ou lois de la logique propositionnelle. De par tout ce qui a été dit jusqu'à maintenant, nous pouvons tenter de définir ce qu'est une preuve.

Définition 1.10 *Une suite finie de formules $B_1, B_2, \ldots, B_n$ est appelée* preuve *à partir des hypothèses $A_1, A_2, \ldots, A_n$ si pour chaque i :*

- *B_i est l'une des hypothèses $A_1, A_2, \ldots, A_n$*
- *ou B_i est une variante d'un axiome*
- *ou B_i est inférée (par application de la règle du modus ponens) à partir de la prémisse majeure B_j et de la prémisse mineure B_k où $j, k < i$*
- *ou B_i est inférée (par application de la règle de substitution) à partir d'une prémisse antérieure B_j, la variable remplacée n'apparaissant pas dans $A_1, A_2, \ldots, A_n$*

Une telle suite de formules, B_m étant la formule finale de la suite, est appelée plus explicitement *preuve de B_m* à partir des hypothèses $A_1, A_2, \ldots, A_n$, ce que nous notons par [11] :

$$A_1, A_2, \ldots, A_n \vdash B_m \tag{1.20}$$

1.3.4 Quantificateurs

Nous devons compléter l'utilisation des connecteurs du calcul propositionnel par ce que nous appelons des *quantificateurs* si nous souhaitons pouvoir résoudre certains problèmes. Effectivement,

11. Il faut noter que lorsque nous essayons de prouver un résultat à partir d'un certain nombre d'hypothèses, nous essayons pas de prouver les hypothèses elles-mêmes.

le calcul propositionnel ne nous permet pas d'affirmer des choses générales sur les éléments d'un ensemble par exemple. Dans ce sens, la logique des propositionnelle ne reflète qu'une partie du raisonnement. Le *calcul des prédicats* au contraire permet de manipuler formellement des affirmations telles qu'*il existe un x tel que [x a une voiture américaine]* ou *pour tous les x [si x est une teckel, alors x est petit]* ; en somme, nous étendons les formules composées afin de pouvoir affirmer des quantifications existentielles (*il existe...*) et des quantification universelle (*pour tout...*). Les exemples que nous venons de donner font intervenir des propositions un peu particulières comme « x a une voiture américaine ». Il s'agit ici de propositions comportant une variable. Ces propositions sont en fait l'application d'une fonction à x. Cette fonction, c'est celle qui associe « x a une voiture américaine » à x. Nous dénoterons cette fonction par « a_b a une voiture américaine » et nous dirons que c'est une fonction propositionnelle, car c'est une fonction dont la valeur est une proposition, ou encore, un *prédicat*.

Les quantificateurs existentiels et universels vont donc de pair avec l'emploi de fonctions propositionnelles. Le calcul des prédicats est cependant limité dans les formules existentielles et universelles. Ainsi, nous nous interdisons des formules comme « il existe une affirmation de x telle que... » En fait, nous ne nous autorisons à quantifier que des *individus*. C'est pour cela que la logique des prédicats est dite une *logique de premier ordre*.

Avant de passer à l'étude du calcul des prédicats nous devons définir :

Définition 1.11 *Le* quantificateur universel *: $\forall$ (pour tout)*

Définition 1.12 *Le* quantificateur existentiel *: $\exists$ (il existe)*

Nous utilisons parfois le symbole $\exists!$ pour dire brièvement *il existe un et un seul* :

$$\forall x \in \mathbb{R}, \exists! y \in \mathbb{R}_+^*, x = ln(y) \tag{1.21}$$

Nous allons voir que la théorie de la démonstration et des ensembles, est l'exacte transcription des principes et résultats de la Logique.

1.4 Calcul des prédicats

Dans un cours de mathématiques, il faut démontrer les propriétés de différents types d'objets (entiers, réels, matrices, suites, fonctions continues, courbes...). Pour pouvoir prouver ces propriétés, il faut bien sûr que les objets sur lesquels nous travaillons soient clairement définis.

En logique du premier ordre et, en particulier, en théorie de la démonstration, les objets que nous étudions sont les formules et leurs démonstrations. Il faut donc donner une définition précise de ce que sont ces notions. Les termes et les formules forment la grammaire d'une langue, simplifiée à l'extrême et calculée exactement pour dire ce que nous voulons sans ambiguïté et sans détour inutile.

1.4.1 Grammaire

Définition 1.13 (Termes) *Les termes désignent les objets dont nous voulons prouver des propriétés (nous reviendrons un peu plus loin beaucoup plus en détail sur ces derniers) :*

- *en algèbre, les termes désignent les éléments d'un groupe (ou anneau, corps, espace vectoriel, etc.). Nous manipulons aussi des ensembles d'objets (sous-groupe, sous-espace vectoriel, etc). Les termes qui désignent ces objets, d'un autre type, seront appelés* termes du second ordre *;*
- *en analyse, les termes désignent les réels ou des fonctions.*

Définition 1.14 (Formules) *Les formules représentent les propriétés des objets que nous étudions :*

- *en algèbre, nous pourrons écrire des formules pour exprimer que deux éléments commutent, qu'un sous-espace vectoriel est de dimension 3, etc.*
- *en analyse, nous écrirons des formules pour exprimer la continuité d'une fonction, la convergence d'une suite, etc.*
- *en théorie des ensembles, les formules pourront exprimer l'inclusion de deux ensembles, l'appartenance d'un élément à un ensemble, etc.*

Définition 1.15 (Démonstrations) *Les démonstrations permettent d'établir qu'une formule est vraie. Le sens précis de ce mot aura lui aussi besoin d'être défini. Plus exactement, elles sont des déductions sous hypothèses, elles permettent de* mener du vrai au vrai, *la question de la vérité de la conclusion étant alors renvoyée à celle des hypothèses, laquelle ne regarde pas la logique mais repose sur la connaissance que nous avons des choses dont nous parlons.*

1.4.2 Langages

En mathématique, nous utilisons, suivant le domaine, différents langages qui se distinguent par les symboles utilisés. La définition 1.16 exprime simplement qu'il suffit de donner la liste de ces symboles pour préciser le langage.

Définition 1.16 (Langage) *Un langage est la donnée d'une famille (pas nécessairement finie) de symboles. Nous en distinguons trois sortes : symboles, termes et formules.*

Nous utilisons quelques fois le mot *vocabulaire* ou le mot *signature* à la place du mot *langage*. Le mot *prédicat* peut être utilisé à la place du mot *relation*. Nous parlons alors de *calcul des prédicats* au lieu de *logique du premier ordre* (ce que nous avons étudié précédemment).

Symboles

Il existe différents types de symboles que nous allons tacher de définir :

- les *symboles de constante* : le n pour l'élément neutre en théorie des ensembles• ; Chap4
- les *symboles de fonction* ou *foncteurs* : à chaque symbole de fonction est associé un entier strictement positif que nous appelons son *arité* : c'est le nombre d'arguments de la fonction. Si l'arité est 1 (resp. $2, \ldots, n$), nous disons que la fonction est unaire (resp. binaire, ..., n-aire)

Le foncteur binaire de multiplication $*$ dans les groupes.

Définition 1.17 (symboles de relation) *De la même manière, à chaque symbole de relation est associé un entier positif ou nul (son arité) qui correspond à son nombre d'arguments et nous parlons de relation unaire, binaire, n-aire (comme par exemple le symbole de relation $=$).*

Définition 1.18 (variables individuelles) *Dans toute la suite, nous nous donnerons un ensemble infini V de variables. Les variables seront notées comme il l'est par tradition x, y, z éventuellement indexées.*

À cela il faut rajouter les connecteurs et quantificateurs que nous avons longuement présenté plus haut et sur lesquels il est pour l'instant inutile de revenir.

Un symbole de constante peut être vu comme un symbole de fonction à 0 argument (d'arité nulle).

Nous considérons (sauf mention contraire) que chaque langage contient le symbole de relation binaire $=$ (lire *égal*) et le symbole de relation à zéro argument dénoté $\bot$ (lire *bottom* ou *absurde*) qui représente le faux. Dans la description d'un langage, nous omettrons donc souvent de les mentionner. Le symbole $\bot$ est souvent redondant. Nous pouvons en effet, sans l'utiliser, écrire une formule qui est toujours fausse. Il permet cependant de représenter le faux d'une manière canonique et donc d'écrire des règles de démonstration générales.

Le rôle des fonctions et des relations est très différent. Comme nous le verrons plus loin, les symboles de fonction sont utilisés pour construire les termes (les objets du langage) et les symboles de relation pour construire les formules (le propriétés de ces objets). fin remarque

Termes

Les termes représentent les objets associés au langage. Soit $\mathcal{L}$ un langage :

Définition 1.19 *L'ensemble $\mathcal{I}$ des termes sur $\mathcal{L}$ est le plus petit ensemble contenant les variables, les constantes et stable (on ne sort pas de l'ensemble) par l'application des symboles de fonction de $\mathcal{L}$ à des termes.*

Définition 1.20 (Terme clos) *Un terme clos est un terme qui ne contient pas de variables (donc par extension, seulement des constantes).*

Définition 1.21 *Pour obtenir une définition plus formelle, nous pouvons écrire :*

$$\mathcal{I}_0 = (t) \tag{1.22}$$

où t est une variable ou un symbole de constante et, pour tout $k \in \mathbb{N}$:

$$\mathcal{I}_{k+1} = \mathcal{I}_k \cup (f(t_1, \ldots, t_n)/t_i \in \mathcal{I}_k) \tag{1.23}$$

où f est une fonction d'arité n (rappelons que l'arité est le nombre d'arguments de la fonction). Ainsi, pour chaque arité, il y a un degré d'ensemble de termes. Nous avons finalement :

$$\mathcal{I} = \underset{k \in \mathbf{N}}{\cup} \mathcal{I}_k \tag{1.24}$$

Définition 1.22 *Nous appellerons* hauteur *d'un terme t le plus petit k tel que $t \in \mathcal{I}_k$*

La définition 1.22 signifie que les variables et les constantes sont des termes et que si f est un symbole de fonction n-aire et $t_1, \ldots, t_n$ sont des termes alors $f(t_1, \ldots, t_n)$ est un terme en soi aussi. L'ensemble $\mathcal{I}$ des termes est défini par la grammaire :

$$\mathcal{I} = V | S_C | S_F(\mathcal{I}, \ldots, \mathcal{I}) \tag{1.25}$$

Cette expression se lit de la manière suivante : un élément de l'ensemble $\mathcal{I}$ que nous sommes en train de définir est soit un élément de V (variables), soit un élément de S_C (l'ensemble des symboles de constantes), soit l'application d'un symboles de fonction $f \in S_F$ à n éléments (constantes ou variables) de $\mathcal{I}$.

Attention, le fait que f soit de la bonne arité est seulement implicite dans cette notation. De plus, l'écriture $S_F(\mathcal{I}, \ldots, \mathcal{I})$ ne signifie pas que tous les arguments d'une fonction sont identiques mais simplement que ces arguments sont des éléments de $\mathcal{I}$.

Il est souvent commode de voir un terme (expression) comme un arbre dont chaque nœud est étiqueté par un symbole de fonction (opérateur ou fonction) et chaque feuille par une variable ou une constante.

Dans la suite, nous allons sans cesse définir des notions (ou prouver des résultats) *par récurrence* sur la structure ou la taille d'un terme.

Définition 1.23 *Pour prouver une propriété P sur les termes, il suffit de prouver P pour les variables et les constantes et de prouver $P(f(t_1, \ldots, t_n))$ à partir de $P(t_1), \ldots, P(t_n)$. Nous faisons ainsi ici une* preuve par induction sur la hauteur *d'un terme. C'est une technique que nous retrouverons dans les chapitres suivants.*

Définition 1.24 *Pour définir une fonction Φ sur les termes, il suffit de la définir sur les variables et les constantes et de dire comment nous obtenons $\Phi(f(t_1, \ldots, t_n))$ à partir de $\Phi(t_1), \ldots, \Phi(t_n)$. Nous faisons ici encore une* définition par induction sur la hauteur d'un terme.

Exemple 1.8 *La taille (nous disons aussi la* longueur*) d'un terme t (notée $\tau(t)$) est le nombre de symboles de fonction apparaissant dans t. Formellement :*

- *$\tau(x) = \tau(c) = O$ si x est une variable et c est une constante*
- *$\tau(f(t_1, \ldots, t_n)) = 1 + \sum_{1 \leqslant i \leqslant n} \tau(t_i)$*

La preuve par induction sur la hauteur d'un terme sera souvent insuffisante. Nous pourrons alors prouver une propriété P sur les termes en supposant la propriété vraie pour tous les termes de taille $p < n$ et en la démontrant ensuite pour les termes de taille n. Il s'agira alors d'une preuve par récurrence sur la taille du terme•.

Chap4

Formules

Les formules sont construites à partir des *formules dites atomiques* en utilisant des connecteurs et des *quantificateurs*. Nous utiliserons les connecteurs et les quantificateurs suivants :

- connecteur unaire de négation $\neg$;
- connecteurs binaires de conjonction et disjonction ainsi que d'implication $\wedge$, $\vee$, $\rightarrow$;
- quantificateurs : $\exists$ qui se lit *il existe* et $\forall$ qui se lit *pour tout*.

Cette notation des connecteurs est standard. Elle est utilisée pour éviter les confusions entre les formules et le langage courant (le métalangage).

Définition 1.25 *Soit $\mathcal{L}$ un langage, les* formules atomiques *de $\mathcal{L}$ sont les formules de la forme $R(t_1, \ldots, t_n)$ où R est un symbole de relation n-aire de $\mathcal{L}$ et $t_1, \ldots, t_n$ sont des termes de $\mathcal{L}$. Nous notons* Atom *l'ensemble des formules atomiques. Si nous notons S_R l'ensemble des symboles de relation, nous pouvons écrire l'ensemble des termes mis en relations par l'expression :*

$$Atom = S_R(\mathcal{I}, \ldots, \mathcal{I}) \tag{1.26}$$

L'ensemble F des formules de la logique du premier ordre de $\mathcal{L}$ est donc défini par la grammaire (où x est une variable) :

$$F = Atom|F \vee F|F \wedge F|F \to F|F\neg F|\exists xF|\forall xF \tag{1.27}$$

où il faut lire : l'ensemble des formules est le plus petit ensemble contenant les formules et tel que si F_1 et F_2 sont des formules alors $F_1 \vee F_2$... sont des formules et qu'elles peuvent être en relation en relation entre elles.

Exemple 1.9 *Les symboles de relation du langage propositionnel sont des relations d'arité 0 (même le symbole $=$ est absent), les quantificateurs sont alors inutiles (puisqu'une formule ne peut pas contenir des variables). Nous obtenons alors le calcul propositionnel défini par :*

$$C_P = V_P|\bot|\neg C_P|C_P \vee C_P|C_P \wedge C_P|C_P \to C_P \tag{1.28}$$

Remarquons la présence du symbole botton *signifiant le* faux *que nous n'avions pas mentionné lors de notre étude de la logique propositionnelle.*

Nous ferons attention à ne pas confondre termes et formules. $\sin(x)$ est un terme (fonction), $x = 3$ est une formule. Mais $\sin(x) \wedge (x = 3)$ n'est rien : nous ne pouvons, en effet, mettre un connecteur entre un terme et une formule (aucun sens).

Pour définir une fonction Φ sur les formules, il suffit de définir Φ sur les formules atomiques et de dire comment on obtient $\Phi(F_1 \oplus F_2)$ (resp. $\Phi(\neg F_1)$, $\Phi(QxF_1)$) à partir de $\Phi(F_1)$ et $\Phi(F_2)$ (resp. $\Phi(F_1)$)

Pour prouver une propriété P sur les formules, il suffit de prouver P pour les formules atomiques et de prouver $P(F_1 \oplus F_2)$ (resp. $P(\neg F_1)$, $P(QxF_1)$) à partir de $P(F_1)$ et $P(F_2)$ (resp. $P(F_1)$).

Pour prouver une propriété P sur les formules, il suffit de supposer la propriété vraie pour toutes les formules de taille $p < n$ et de la démontrer pour les formules de taille n.

Définition 1.26 *Une* sous-formule *d'une formule (ou expression) F est l'un de ses composants, in extenso une formule à partir de laquelle F est construite. Formellement, nous définissons l'ensemble $SF(F)$ des sous-formules F par :*

- *Si F est atomique, $SF(F) = (F)$*
- *Si $F = F_1 \oplus F_2$ avec $\oplus \in (\wedge, \vee, \to)$, $SF(F) = (F) \cup SF(F_1) \cup SF(F_2)$*
- *Si $F = \neg F_1$ ou QxF_1 avec $Q \in (\forall, \exists)$, $SF(F) = (F) \cup SF(F_1)$*

Définition 1.27 *Une formule F de $\mathcal{L}$ n'utilise qu'un nombre fini de symboles de $\mathcal{L}$. Ce sous-ensemble est appelé le* langage de la formule *et noté $\mathcal{L}(F)$.*

Définition 1.28 *La* taille *(ou la longueur) d'une formule F (notée $\tau(F)$) est le nombre de connecteurs ou de quantificateurs apparaissant dans F. Formellement :*

- *$\tau(F) = 0$ si F est une formule atomique*
- *$\tau(F_1 \oplus F_2) = 1 + \tau(F_1) + \tau(F_2)$ où $\oplus \in \{\vee, \wedge, \to\}$*
- *$\tau(\neg F_1) = \tau(QxF_1) = 1 + \tau(F_1)$ avec $Q \in \{\forall, \exists\}$*

Définition 1.29 L'opérateur principal *(nous disons aussi le* connecteur principal*) d'une formule est défini par :*

- *Si A est atomique, alors elle n'a pas d'opérateur principal*
- *Si $A = \neg B$ alors $\neg$ est l'opérateur principal de A*
- *Si $A = B \oplus C$ où $\oplus \in (\wedge, \vee, \to)$, alors $\oplus$ est l'opérateur principal de A*
- *Si $A = QxB$ où $A \in (\forall, \exists)$, alors Q est l'opérateur principal de A*

Définition 1.30 *Soit F une formule. L'ensemble $VL(F)$ des variables libres de F et l'ensemble $VM(F)$ des variables muettes (ou liées) de F sont définis par récurrence sur $\tau(F)$.*

Une occurrence d'une variable donnée est dite *variable liée* ou *variable muette* dans une formule F si dans cette même formule, un quantificateur y fait référence. Dans le cas contraire, nous disons avoir une *variable libre*[12].

Pour préciser les variables libres possibles d'une formule F, nous noterons $F[x_1, \ldots, x_n]$. Cela signifie que les variables libres de F sont parmi $x_1, \ldots, x_n$ in extenso si y est libre dans F, alors y est l'un des x_i mais les x_i n'apparaissent pas nécessairement dans F.

Nous pouvons définir les variables muettes ou libre de manière plus formelle :

1. Si $F = R(t_1, \ldots, t_n)$ est atomique alors $VL(F)$ est l'ensemble des variables apparaissant dans les t_i et nous avons $VM(F) = \varnothing$
2. Si $F = F_1 \oplus F_2$ où $\oplus \in (\wedge, \vee, \rightarrow) : VL(F) = VL(F_1) \cup VL(F_2)$ alors $VM(F) = VM(F_1) \cup VM(F_2)$
3. si $F = \neg F_1$ alors $VL(F) = VL(F_1)$ et $VM(F) = VM(F_1)$
4. si $F = QxF_1$ avec $Q \in \{\forall, \exists\} : VL(F) = VL(F_1) - \{x\}$ et $VM(F) = VM(F_1) \cup \{x\}$

Exemple 1.10 *Soit* $F : \forall x\ (x \cdot y = y \cdot x)$ *alors* $VL(F) = \{y\}$ *et* $VM(F) = \{x\}$

Exemple 1.11 *Soit* $G : \{\forall x\ \exists y\ (x \cdot z = z \cdot x)\} \wedge \{x = z \cdot z\}$ *alors* $VL(F) = \{x, z\}$ *et* $VM(F) = \{x, y\}$

Définition 1.31 *Nous disons que les formules F et G sont α-*équivalentes *si elles sont (syntaxiquement) identiques à un renommage près des occurrences liées des variables.*

Définition 1.32 (formule close) *Une* formule close *est une formule sans variables libres.*

Définition 1.33 *Soit F une formule, x une variable et t un terme. $F[x := t]$ est la formule obtenue en remplaçant dans F toutes les occurrences libres de x par t, après renommage éventuel des occurrences liées de F qui apparaissent libres dans t.*

Nous noterons dans les exemples vus qu'une variable peut avoir à la fois des occurrences libres et des occurrences liées. Nous n'avons donc pas toujours $VL(F) \cap VM(F) = \varnothing$.

Nous ne pouvons pas renommer y en x dans la formule $\forall y(x \cdot y = y \cdot x)$ et obtenir la formule $\forall x(x \cdot x = x \cdot x)$: la variable x serait *capturée*. Nous ne pouvons donc pas renommer des variables liées sans précautions : il faut éviter de capturer des occurrences libres.

La notion de formule définie ici est la notion de formule de logique du premier ordre. Certaines formules ne sont pas des formules du premier ordre (le nombre de quantificateurs ne doivent pas êtres dépendant d'une variable !).

1.5 Démonstrations

1.5.1 Introduction

Les démonstrations que l'on trouve dans les ouvrages de mathématiques sont des assemblages de symboles mathématiques et de phrases contenant des mots clés tels que : *donc*, *parce que*, *si*, *si et seulement si*, *il est nécessaire que*, *il suffit de*, *prenons un x tel que*, *supposons que*, *cherchons une contradiction*, etc. Ces mots sont supposés être compris par tous de la même manière, ce qui n'est en fait, pas toujours le cas.

Dans tout ouvrage, le but d'une démonstration est de convaincre le lecteur de la vérité de l'énoncé. Suivant le niveau du lecteur, cette démonstration sera plus ou moins détaillée : quelque chose qui pourra être considéré comme évident dans un cours de licence pourrait ne pas l'être dans un cours de maturité.

Dans un devoir, le correcteur sait que le résultat demandé à l'étudiant est vrai et il en connaît la démonstration. L'étudiant doit démontrer (correctement) le résultat demandé. Le niveau de détail qu'il doit donner dépend donc de la confiance qu'aura le correcteur : dans une bonne copie, une *preuve par une récurrence évidente* passera bien, alors que dans une copie où il y eu auparavant un *évident*, qui était évidemment... faux, ça ne passera pas !

Pour pouvoir gérer convenablement le niveau de détail, il faut savoir ce qu'est une démonstration complète. Ce travail de formalisation n'a été fait qu'au début de 20^{e} siècle !! Plusieurs choses peuvent paraître surprenantes :

12. Une occurrence d'une variable x dans une formule F est une position de cette variable dans la formule F. Ne pas confondre avec l'objet qu'est la variable elle-même.

- il n'y a qu'un nombre fini de règles : deux pour chacun des connecteurs (et l'égalité) plus trois règles générales. Il n'était pas du tout évident *a priori* qu'un nombre fini de règles soit suffisant pour démontrer tout ce qui est vrai. Nous montrerons ce résultat (c'est essentiellement, le théorème de complétude). La preuve n'en est pas du tout triviale ;
- ce sont les mêmes règles pour toutes les mathématiques et la physique : algèbre, analyse, géométrie, etc. Cela veut dire que nous avons réussi à isoler tout ce qui est général dans un raisonnement. Nous verrons plus loin qu'une démonstration est un assemblage de couples, où Γ est un ensemble de formules (les hypothèses) et A une formule (la conclusion). Quand nous faisons de l'arithmétique, de la géométrie ou de l'analyse réelle, nous utilisons, en plus des règles, des hypothèses que l'on appelle des *axiomes*. Ceux-ci expriment les propriétés particulières des objets que nous manipulons (pour plus de détails sur les axiomes voir la page d'introduction du site).

Nous démontrons donc, en général, des formules en utilisant un ensemble d'hypothèses, et cet ensemble peut varier au cours de la démonstration : quand nous disons *supposons F et montrons G*, F est alors une nouvelle hypothèse que nous pourrons utiliser pour montrer G. Pour formaliser cela, nous introduisons le concept de *séquent* :

Définition 1.34 (séquent) *Un* séquent *est un couple noté $\Gamma \vdash F$ où :*
- Γ *est un ensemble fini de formules qui représente les hypothèses que nous pouvons utiliser. Cet ensemble s'appelle aussi le* contexte du séquent *;*
- F *est une formule. C'est la formule que nous voulons montrer. Nous dirons que cette formule est la* conclusion du séquent.

Remarques
1. Si $\Gamma = (A_1, \ldots, A_n)$ nous pourrons noter $A_1, \ldots, A_n \vdash F$ au lieu de $\Gamma \vdash F$. Le signe $\vdash$ se lit *thèse* ou *démontre*.
2. Nous noterons $\vdash F$ un séquent dont l'ensemble d'hypothèses est vide et $\Gamma_1, \ldots, \Gamma_n \vdash F$ un séquent dont l'ensemble d'hypothèses est $\cup_i \Gamma_i$.
3. Nous noterons que dans le séquent $\Gamma, A \vdash B$, la formule A peut-être dans Γ (elle devient alors un hypothèse).
4. Nous écrirons $\Gamma \nvdash F$ pour dire que $\Gamma \vdash F$ *est non prouvable*.

Définition 1.35 *Un séquent $\Gamma \vdash F$ est* prouvable *(ou démontrable, dérivable) s'il peut être obtenu par une application finie de règles. Une formule F est prouvable si le séquent $\vdash F$ est prouvable.*

1.5.2 Règles de démonstration

Les règles de démonstration sont les briques qui permettent de construire les dérivations. Une dérivation formelle est un assemblage fini (et correct !) de règles. Cet assemblage n'est pas linéaire (ce n'est pas une suite) mais un *arbre*. Nous sommes en effet souvent amenés à faire des branchements.

Nous allons présenter un choix de règles. Nous aurions pu en présenter d'autres (à la place ou en plus) qui donneraient la même notion de prouvabilité. Celle que l'on a choisies sont *naturelles* et correspondent aux raisonnements que l'on fait habituellement en mathématique. Dans la pratique courante nous utilisons, en plus des règles ci-dessous, beaucoup d'autres règles mais celles-ci peuvent se déduire des précédentes. Nous les appellerons *règles dérivées*.

Il est de tradition d'écrire la racine de l'arbre (le séquent conclusion) en bas, les feuilles en haut : la nature est ainsi faite... Pourtant, nous construisons souvent l'arbre en allant de la racine vers les feuilles. Comme il est également de tradition d'écrire, sur une feuille de papier, de haut en bas, il ne serait pas déraisonnable d'écrire la racine en haut et les feuilles en bas. Il faut faire un choix !

Une règle se compose :
- d'un ensemble de *prémisses* : chacune d'elles est un séquent. Il peut y en avoir zéro, un ou plusieurs ;
- du séquent conclusion de la règle ;
- d'une barre horizontale séparant les prémisses (en haut) de la conclusion (en bas). Sur la droit de la barre, nous indiquerons le nom de la règle.

Exemple 1.12

$$\frac{\Gamma \vdash A \to B \; \Gamma \vdash A}{\Gamma \vdash B} \to_e \tag{1.29}$$

Cette règle à deux prémisses ($\Gamma \vdash A \to B$ et $\Gamma \vdash A$) et une conclusion ($\Gamma \vdash B$). Le nom abrégé de cette règle est $\to_e$.

Cette règle peut se lire de deux manières :
- de bas en haut : si nous voulons prouver la conclusion, il suffit par utilisation de la règle de prouver les prémisses. C'est ce qu'on fait quand nous cherchons une démonstration. Cela correspond à *l'analyse*.
- de haut en bas : si nous avons prouvé les prémisses, alors nous avons aussi prouvé la conclusion. C'est ce que nous faisons fait quand nous rédigeons une démonstration. Cela correspond à la *synthèse*.

Pour les démonstrations, il existe dix-sept règles définies ci-après. Les explications qui accompagnent les formules sont données de bas en haut :

1. Axiome :

$$\frac{}{\Gamma, A \vdash A} ax \tag{1.30}$$

   Si la conclusion du séquent est une des hypothèses, alors le séquent est prouvable.

2. Affaiblissement :

$$\frac{\Gamma \vdash A}{\Gamma, B \vdash A} aff \tag{1.31}$$

   - Si nous démontrons A sous les hypothèses Γ, en ajoutant d'autres hypothèses on peut encore démontrer A.
   - Il y a des hypothèses qui peuvent ne pas servir

3. Introduction de l'implication :

$$\frac{\Gamma, A \vdash B}{\Gamma \vdash B \to A} \to_i \tag{1.32}$$

   Pour montre que $A \to B$ nous supposons A (c'est-à-dire que nous l'ajoutons aux hypothèses) et nous démontrons B.

4. Elimination de l'implication :

$$\frac{\Gamma \vdash A \to B \Gamma \vdash A}{\Gamma \vdash B} \to_e \tag{1.33}$$

   Pour démontrer B, si nous connaissons un théorème de la forme $A \to B$ et si nous pouvons démontrer le lemme $A \to B$, il suffit de démontrer A.

5. Introduction à la conjonction :

$$\frac{\Gamma \vdash A \Gamma \vdash B}{\Gamma \vdash A \wedge B} \wedge_i \tag{1.34}$$

   Pour montrer $A \wedge B$, il suffit de montrer A et de montrer B.

6. Elimination de la conjonction :

$$\frac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \wedge_e^g \tag{1.35}$$

   et :

$$\frac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \wedge_e^d \tag{1.36}$$

   De $A \wedge B$, nous pouvons déduire A (élimination gauche) et B (élimination droite).

7. Introduction de la disjonction :

$$\frac{\Gamma \vdash A}{\Gamma \vdash A \vee B} \wedge_i^g \tag{1.37}$$

   ou :

$$\frac{\Gamma \vdash B}{\Gamma A \vee B} \vee_i^d \tag{1.38}$$

   Pour démontrer $A \vee B$, il suffit de démontrer A ou de démontrer B

8. Elimination de la disjonction :

$$\frac{\Gamma \vdash A \vee B \Gamma, A \vdash C \Gamma, B \vdash C}{\Gamma \vdash C} \vee_e \tag{1.39}$$

Si nous voulons montrer C et que nous savons que nous avons $A \vee B$, il suffit de le montrer d'une part en supposant A, d'autre part en supposant B. C'est un raisonnement par cas.

9. Introduction de la négation :

$$\frac{\Gamma, A \vdash \bot}{\Gamma \vdash \neg A} \neg_i \tag{1.40}$$

Pour montrer $\neg A$, nous supposons A et nous démontrons l'absurde ($\bot$)

10. Elimination de la négation :

$$\frac{\Gamma \vdash \neg A \, \Gamma \vdash A}{\Gamma \vdash \bot} \neg_e \tag{1.41}$$

Si nous avons montré $\neg A$ et A, alors nous avons montré l'absurde ($\bot$)

11. Absurdité classique :

$$\frac{\Gamma, \neg A \vdash \bot}{\Gamma \vdash A} \bot_c \tag{1.42}$$

Pour démontrer A, il suffit de démontrer l'absurde en supposant $\neg A$.
Cette règle, est équivalent à dire *A est vraie si et seulement s'il est faux que A soit fausse.* Cette règle ne va pas de soi : elle est nécessaire pour prouver certaines résultats (il y a des résultats que nous ne pouvons pas prouver si nous n'avons pas cette règle). Contrairement, à beaucoup d'autres, cette règle peut par ailleurs être appliquée à tout moment : nous pouvons, en effet, toujours dire : pour prouver A je suppose $\neg A$ et je vais cherche une contradiction.

12. Introduction au quantificateur universel :

$$\frac{\Gamma \vdash A \; x \text{ n'est pas libre dans les formules de } \Gamma}{\vdash \forall x A} \forall_i \tag{1.43}$$

Pour démontrer $\forall x A$, il suffit de montrer A en ne faisant aucune hypothèse sur x [13].

13. Elimination du quantificateur universel :

$$\frac{\Gamma \vdash \forall x A}{\Gamma \vdash A[x := t]} \forall_e \tag{1.44}$$

De $\forall x A$, nous pouvons déduire $A[x := t]$ pour n'import quel terme t. Ce que nous pouvons dire aussi sous la forme : si nous avons prouvé A pour tout x, alors nous pouvons utiliser A avec n'importe quel objet t (!!).

14. Introduction du quantificateur existentiel

$$\frac{\Gamma \vdash A[x := t]}{\Gamma \vdash \exists x A} \exists_i \tag{1.45}$$

Pour démontrer $\exists x A$, il suffit de trouver un objet (in extenso un terme t) pour lequel nous savons montrer $A[x := t]$.

15. Elimination du quantificateur existentiel

$$\frac{\Gamma \vdash \exists x A \; \Gamma, A \vdash C \; x \text{ n'est libre ni dans les formules de } \Gamma, \text{ ni dans } C}{\Gamma \vdash C} \exists_e \tag{1.46}$$

Nous démontrons qu'il existe bien un ensemble d'hypothèses tel que $\exists x A$ et partant de ce résultat comme nouvelle hypothèse, nous démontrons C. Ce formule C hérite alors de la formule $\exists x A$ et dès lors x n'est pas libre dans C car il ne l'était déjà pas dans $\exists x A$.

13. pour des démonstrations cette vérification (aucune hypothèse sur x) est souvent source d'erreur.

16. Introduction de l'égalité :

$$\frac{}{\Gamma \vdash t = t} =_i \tag{1.47}$$

Chap3 Nous pouvons toujours montrer $t = t$. Cette règle signifie que l'égalité est réflexive•.

17. Elimination de l'égalité :

$$\frac{\Gamma \vdash A[x := t] \; \Gamma \vdash t = u}{\Gamma \vdash A[x := u]} =_e \tag{1.48}$$

Si nous avons démontré $\Gamma \vdash A[x := t]$ et $t = u$, alors nous avons démontré $A[x := u]$. Cette règle exprime que les objets égaux ont les mêmes propriétés. Nous noterons cependant que les formules (ou relations) $t = u$ et $u = t$ ne sont pas, formellement, identiques. Il nous faudra démontrer que l'égalité est symétrique (nous en profiterons aussi pour démontrer que l'égalité est transitive).

Exemple 1.13 *Cet exemple montre que l'égalité est symétrique (un petit peu non trivial mais bon pour commencer) :*

$$\dfrac{\dfrac{\dfrac{\dfrac{}{x_1 = x_2 \vdash x_1 = x_1} =_i \quad \dfrac{}{x_1 = x_2 \vdash x_1 = x_2} ax}{x_1 = x_2 \vdash x_2 = x_1} =_e}{\vdash x_1 = x_2 \to x_2 = x_1} \to_i}{\vdash \forall x_1, x_2 (x_1 = x_2 \to x_2 = x_1} \forall_i \times 2 \tag{1.49}$$

Nous introduisons l'égalité $=_i$ et prouvons à partir de l'hypothèse $x_1 = x_2$ la formule $x_1 = x_1$. En même temps, nous définissons l'axiome comme quoi $x_1 = x_2$. Ensuite à partir de ces prémisses, nous éliminons l'égalité $=_e$ en substituant les termes de façon à ce que à partir de la supposition $x_1 = x_2$ (venant de l'axiome) nous obtenions $x_2 = x_1$. Ensuite, l'élimination de l'égalité implique automatiquement sans aucune hypothèse que $x_1 = x_2 \to x_2 = x_1$. Dès lors, il nous suffit d'introduire le quantificateur universel pour chacune des variables (donc deux fois) sans aucune hypothèse afin d'obtenir que l'égalité est symétrique.

Exemple 1.14 *Cet exemple montre que l'égalité est transitive (c'est-à-dire si $x_1 = x_2$ et $x_2 = x_3$ alors $x_1 = x_3$). En notant F la formule $x_1 = x_2 \wedge x_2 = x_3$:*

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{F \vdash x_1 = x_2 \wedge x_2 = x_3} ax}{F \vdash x_1 = x_2} \wedge_e^g \quad \dfrac{\dfrac{}{F \vdash x_1 = x_2 \wedge x_2 = x_3} ax}{F \vdash x_2 = x_3} \wedge_e^d}{F \vdash x_1 = x_3} =_e}{\vdash (x_1 = x_2 \wedge x_2 = x_3) \to x_1 = x_3} \to_i}{\forall x_1, x_2, x_3 ((x_1 = x_2 \wedge x_2 = x_3) \to x_1 = x_3)} \forall_i \times 3 \tag{1.50}$$

Que faisons, nous ici ? Nous introduisons d'abord la formule F deux fois en tant qu'axiome afin de la décortiquer plus tard à gauche et à droite (nous n'introduisons pas l'égalité supposée déjà introduite en tant que règle). Une fois ceci fait, nous éliminons à gauche et à droite la conjonction sur la formule pour travailler sur les termes gauches et droites seuls et introduisons l'égalité sur les deux termes ce qui fait qu'à partir de la formule nous avons l'égalité transitive. Il s'ensuit que sans aucune hypothèse cela implique automatiquement que l'égalité est transitive et finalement nous disons que ceci est valable pour tout valeur des différentes variables (si la formule est vraie, alors l'égalité est transitive).

Chap4 ***Exemple 1.15*** *L'objectif sera de démontrer que toute involution est une bijection•. Soit f est un symbole de fonction unaire (à une variable), nous notons :*

- *Inj[f] la formule :*

$$\forall x, y \; f(x) = f(y) \to x = y \tag{1.51}$$

qui signifie que f est injective.

- *Surj[f] la formule :*

$$\forall y, \exists x \; f(x) = y \tag{1.52}$$

qui signifie que f est surjective

– *Bij*$[f]$ *la formule :*

$$Inj[f] \wedge Surj[f] \tag{1.53}$$

qui signifie que f *est bijective.*
– *Inv*$[f]$ *la formule :*

$$\forall x\ f(f(x)) = x \tag{1.54}$$

qui signifie que f est involution (nous notons également cela $f \circ f = Id$ *c'est-à-dire que la composition de* f *est l'identité)*

Nous aimerions savoir si :

$$\text{Inv}[f] \vdash \text{Bij}[f] \tag{1.55}$$

Nous allons présenter cette démonstration de quatre manières différentes : classique (informelle), classique (pseudo-formelle) et formelle en arbre et formelle en ligne.

Méthode classique

Nous devons montrer que si f est involutive alors elle est donc bijective. Nous avons donc deux propriétés à montrer et les deux doivent être satisfaites en même temps : que la fonction est injective et surjective.

1. Montrons que l'involution est injective : nous supposons pour cela, puisque f est involutive elle est donc injective, tel que :

   $$\forall x, y\ f(x) = f(y) \tag{1.56}$$

   implique :

   $$x = y \tag{1.57}$$

   Or, cette supposition découle automatiquement de la définition de l'involution que :

   $$\forall x\ f(f(x)) = x, \forall y\ f(f(y)) = y \tag{1.58}$$

   et de l'application de f à la relation :

   $$f(x) = f(y) \tag{1.59}$$

   (soit trois égalités) tel que :

   $$f(f(x)) = f(f(y)) \tag{1.60}$$

   nous avons donc :

   $$x = y \tag{1.61}$$

2. Montrons que l'involution est surjective : si elle est surjective, alors nous devons avoir :

   $$\forall y \exists x\ f(x) = y \tag{1.62}$$

   Or, définissons la variable x par définition de l'involution elle-même :

   $$x = f(y) \tag{1.63}$$

   (puisque $y = f(x)$...) un changement de variables après, nous obtenons :

   $$f(f(y)) = y \tag{1.64}$$

   et donc la surjectivité est assurée.

Méthode pseudo-formelle

Nous reprenons la même chose et nous y injecte les règles de la théories de la démonstration.

Nous devons montrer que f involutive est donc bijective. Nous avons donc deux propriétés à montrer ($\wedge_i$) et les deux doivent être satisfaites en même temps : que la fonction est injective et surjective :

$$\text{Inj}[f] \wedge \text{Surj}[f] \tag{1.65}$$

1. Montrons que l'involution est injective : nous supposons pour cela, puisque f est involutive et donc injective, que :

   $$(\forall i)\forall x, y\ f(x) = f(y) \tag{1.66}$$

   implique :

   $$(\rightarrow_i)x = y \tag{1.67}$$

   Or, cette supposition découle automatiquement de la définition de l'involution (donc nous pouvons éliminer le quantificateur universel sous-jacent et introduire l'involution comme un axiome) que :

   $$(\forall_e, ax)\forall x\ f(f(x)) = x, \forall y\, f(f(y)) = y \tag{1.68}$$

   et de l'application de f à la relation :

   $$f(x) = f(y) \tag{1.69}$$

   (soit trois égalités $=_e \times 2$) tel que :

   $$f(f(x)) = f(f(y)) \tag{1.70}$$

   nous avons donc :

   $$x = y \tag{1.71}$$

2. Montrons que l'involution est surjective : si elle est surjective, alors nous devons avoir :

   $$\forall y \exists x f(x) = y (\forall i) \tag{1.72}$$

   Or, définissons la variable x par définition de l'involution elle-même :

   $$(\exists_i)x = f(y) \tag{1.73}$$

   (puisque $y = f(x)...$) un changement de variables après nous obtenons :

   $$f(f(y)) = y \tag{1.74}$$

   et donc :

   $$(\forall_e, ax) \tag{1.75}$$

   la surjectivité est assurée.

Méthode formelle en arbre

Faisons cela avec le méthode graphique que nous avons déjà présenté plus haut.

1. Montrons que l'involution est injective :
   Pour cela, d'abord montrons que $f(x) = f(y) \vdash f(f(x)) = f(f(y))$

$$\dfrac{\dfrac{\overline{\text{Inv}[f] \vdash \forall y(f(f(y)) = y)}^{ax}}{\text{Inv}[f] \vdash \{f(y) = x\}[x := f(y)]}\forall_e \quad \dfrac{}{f(x) = f(y) \vdash f(x) = f(y)}^{ax}}{f(x) = f(y) \vdash \{f(x) = y\}[x := f(y)]} =_e \tag{1.76}$$

$$\dfrac{\dfrac{\dfrac{}{\mathrm{Inv}[f] \vdash \forall x(f(f(x)) = x)}ax}{\mathrm{Inv}[f] \vdash \{f(x) = y\}[x := f(y)]}\forall_e \quad \dfrac{}{f(x) = f(y) \vdash f(x) = f(y)}ax}{f(x) = f(y) \vdash \{f(y) = x\}[y := f(x)]} =_e \tag{1.77}$$

$$\dfrac{(1.76)(1.77)}{f(x) = f(y) \vdash f(f(x)) = f(f(y))} =_e \tag{1.78}$$

Cette dernière relation est abrégée $=_e$ et appelée (comme d'autres existantes) *règle dérivée* car c'est un raisonnement qui est très souvent fait lors de démonstrations et un peu long à développer à chaque fois... Dès lors :

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{Inv[f] \to Inv[f]}ax}{Inv[f] \vdash f(f(x)) = x}\forall_e \quad \dfrac{\dfrac{}{Inv[f] \to Inv[f]}ax}{Inv[f] \vdash f(f(y)) = y}\forall_e}{f(f(x)) = f(f(y)) \vdash x = y} =_e \ (1.78)}{Inv[f], f(x) = f(y) \vdash x = y} =_e}{\dfrac{Inv[f] \vdash f(x) = f(y) \to x = y}{Inv[f] \vdash Inj[f]}\forall_i} \to_i \tag{1.79}$$

2. Montrons que l'involution est surjective :

$$\dfrac{\dfrac{\dfrac{\dfrac{}{Inv[f] \vdash \forall x(f(f(x)) = x)}ax}{Inv[f] \vdash (f(x) = y)[x := f(y)]}\forall_e}{Inv[f] \vdash \exists x f(x) = y}\exists_i}{Inv[f] \vdash Surj[f]}\forall_i \tag{1.80}$$

Il s'ensuit :

$$\dfrac{(5)(6)}{Inv[f] \vdash Inj[f] \wedge Surj[f]}\wedge_i \tag{1.81}$$

Méthode formelle en ligne

Nous pouvons faire la même chose sous une forme un peu moins large et plus tabulée : (1.87)

$$\begin{array}{lll}
\mathrm{Inv}[f] \vdash \mathrm{Bij}[f] & \forall i & (1.82a)\\
\quad 1 := \mathrm{Inv}[f] \vdash \mathrm{Inj}[f] & & \\
\quad 2 := \mathrm{Inv}[f] \vdash \mathrm{Surj}[f] & & \\
(1)\mathrm{Inv}[f] \vdash \mathrm{Inj}[f] & \forall i & (1.82b)\\
\quad \mathrm{Inv}[f] \vdash f(x) = f(y) \to x = y & \to_i & \\
\quad \mathrm{Inv}[f], f(x) = f(y) \vdash x = y & =_e \times 2(i)(ii)(iii) & \\
\quad (i) \quad \mathrm{Inv}[f] \vdash f(f(x)) = x & \forall_e & \\
\qquad \mathrm{Inv}[f] \vdash \mathrm{Inv}[f] & ax & \\
\quad (ii) \quad \mathrm{Inv}[f] \vdash f(f(y)) = y & \forall_e & \\
\qquad \mathrm{Inv}[f] \vdash \mathrm{Inv}[f] & ax & \\
\quad (iii) \quad f(x) = f(y) \vdash f(f(x)) = f(f(x)) & =_c & \\
\qquad (1')f(x) = f(y =\vdash f(x) = f(y) & ax & \\
(2)\mathrm{Inv}[f] \vdash \mathrm{Surj}[f] & \forall i & (1.82c)\\
\quad \mathrm{Inv}[f] \vdash \exists x\{f(x) = y\} & \exists_i & \\
\quad \mathrm{Inv}[f] \vdash \{f(x) = y\}[x := f(y)] & \forall_e & \\
\quad \mathrm{Inv}[f] \vdash \forall x\{f(f(x)) = x\} & ax &
\end{array}$$

2

Nombres

La base des mathématiques, mis à part le raisonnement•, est sans nul doute pour le commun des personnes —mathématiciens mis à part— l'arithmétique. Il est donc obligatoire que nous y fassions étape pour étudier sa provenance, quelques unes de ses propriétés et conséquences. Chap1

Les nombres, comme les figures géométriques, constituent les bases de l'arithmétique. Ce sont aussi les bases historiques car les mathématiques ont certainement commencé par l'étude de ces objets, mais aussi les bases pédagogiques, car c'est en apprenant à compter que nous entrons dans le monde des mathématiques.

L'histoire des nombres, également appelés scalaires, est beaucoup trop longue pour être relatée ici, mais nous ne pouvons que vous conseiller l'« Histoire universelle des chiffres »[1], un des meilleurs ouvrages sur le sujet.

Cependant voici une petite bride de cette dernière qui nous semble fondamentale. Notre système décimal actuel, de base 10, utilise les chiffres de 0 à 9, dits *chiffres arabes*, mais au fait d'origine indienne. En fait, les chiffres arabes, d'origine indienne, sont différents : Il faut lire : 0 « zéro », 1

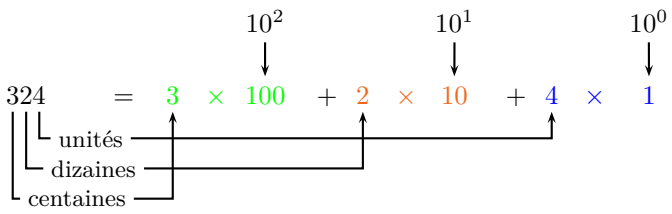
$$0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \tag{2.1}$$

« un », 2 « deux », 3 « trois », 4 « quatre », 5 « cinq », 6 « six », 7 « sept », 8 « huit », 9 « neuf ». Ce système est beaucoup plus efficace que les chiffres romains.

Ces chiffres ne furent introduits en Europe que vers l'an 1 000. Utilisés en Inde, ils furent transmis par les Arabes au monde occidental par le pape G. D'Aurillac lors de son séjour en Andalousie à la fin du 9^{e} siècle. Le mot français « chiffre »est une déformation du mot arabe « sifr »désignant « zéro ». En italien, « zéro »se dit « zero », et serait une contraction de zefiro, on voit là encore la racine arabe. Ainsi nos termes « chiffre » et « zéro »ont la même origine.

L'usage précoce d'un symbole numérique désignant *rien*, au sens de *aucune quantité* ou *absence de quantité*, c'est-à-dire notre zéro, provient du fait que les Indiens utilisèrent un système dit *système positionnel*. Dans un tel système, la position d'un chiffre dans l'écriture d'un nombre exprime la puissance de 10 et le nombre de fois qu'elle intervient. L'absence d'une puissance est notée par un petit rond : c'est le zéro. Notre système actuel est donc le *système décimal et positionnel*.

Exemple 2.1 (Système décimal et positionnel) *Le nombre 324 s'écrit de gauche à droite comme étant trois centaines : 3 fois 100, deux dizaines : 2 fois 10 et quatre unités : 4 fois 1.*



Nous différencions un chiffre d'un nombre : le nombre est composé de chiffres et non inversement.

Nous voyons parfois un séparateur de milliers représenté par un espace fine. Ainsi, nous écrirons 1 034 au lieu de 1034 ou encore 1 344 567 569 au lieu de 1344567569. Les séparateurs de milliers permettent de rapidement quantifier l'ordre de grandeur des nombres.

Tout nombre entier, autre que l'unité, peut être pris pour base d'un système de numérotation. Nous avons ainsi les systèmes de numérotation binaire, ternaire, quaternaire..., décimal, duodécimal qui correspondent respectivement aux bases deux, trois quatre,..., dix, douze.

Une généralisation peut s'écrire sous la forme suivante : tout nombre entier positif peut être représenté dans une base b sous forme de somme, où les coefficients a_i sont multipliés chacun par leur poids respectif b^i, tel que :

$$N = a_{n-1}b^{n-1} + a_{n-1}b^{n-1} + \ldots + a_1 b^1 + a_0 b^0 \tag{2.1}$$

plus élégamment écrit :

$$N = \sum_{i=0}^{n-1} a_i b^i \tag{2.2}$$

avec $a_i \in [0; b-1]$ et $N \in [0; b^n - 1]$.

Comme très fréquemment en mathématique, nous remplacerons l'écriture des chiffres ou des nombres par des lettres latines ou grecques afin de généraliser leur représentation. Ainsi, lorsque nous parlons d'une base b la valeur b peut prendre n'importe quelle valeur entière 1, 2, 3, ...

Lorsque nous prenons la valeur 2 pour b, N aura pour valeur maximale $2^n - 1$. Les nombres qui s'écrivent sous cette forme s'appellent les "nombres de MERSENNE". Ces nombres ne peuvent
Déf2.6|p34 être premiers• que si n est premier. Effectivement, si nous prenons $b = 10$ et $n = 3$, la plus grande valeur que nous pourrons avoir sera alors :

$$(b-1)\cdot 10^2 + (b-1)\cdot 10^1 + (b-1)\cdot 10^0 = 900 + 90 + 9 = 999 = 10^3 - 1 = b^3 - 1 \tag{2.3}$$

Lorsque qu'un nombre est le même lu de gauche à droite ou de droite à gauche, nous parlons de *nombre palindrome*.

2.1 Bases numériques

Pour écrire un nombre dans un système de base b, nous devons commencer par adopter b caractères destinés à représenter les b premiers nombres $\{0, 1, 2, 3, 4, 5, 6, 7, 9\}$. Ces caractères sont comme nous les avons déjà définis, les *chiffres* que nous énonçons comme à l'ordinaire.

Pour la numérotation écrite, nous faisons cette convention, qu'un chiffre, placé à gauche d'un autre représente des unités de l'ordre immédiatement supérieur, ou b fois plus grandes. Pour tenir la place des unités qui peuvent manquer dans certains ordres, nous nous servons du zéro (0) et par suite, le nombre de chiffres employés est toujours égal à la base du système.

Définition 2.1 *Pour la numérotation parlée, nous convenons d'appeler* unité simple, dizaine, centaine, mille, *etc., les unités du premier ordre, du second, du troisième, du quatrième, etc.*

Ainsi les nombres 10, 11, ..., 19 se liront de même dans tous les systèmes de numérotation ; les nombres 1a, 1b, a0, b0, ... se liront dix-a, dix-bé, a-dix, bé-dix, etc. Ainsi, le nombre 5b6a71c se lira : cinq millions bé-cent soixant-a mille sept cent dix-cé.

Cet exemple est pertinent car il nous montre l'expression générale de la langue parlée que nous utilisons quotidiennement et intuitivement en base dix.

REMARQUES

1. Les règles des opérations définies pour les nombres écrits dans le système décimal sont les mêmes pour les nombres écrits dans un système quelconque de numérotation ;
2. Pour opérer rapidement dans un système quelconque de numérotation, il est indispensable de savoir par cœur toutes les sommes et tous les produits de deux nombres d'un seul chiffre ;
3. Le fait que la base décimale ait été choisie est semble-t-il due au fait que l'humain a dix doigts.

Voyons comment nous convertissons un système de numérotation dans un ordre :

Exemple 2.2 *En base dix nous savons que* 142 713 *s'écrit :*

$$1\,427\,713_{10} = 1 \cdot 10^5 + 4 \cdot 10^4 + 2 \cdot 10^3 + 7 \cdot 10^2 + 1 \cdot 10^1 + 3 \cdot 10^0 \tag{2.4}$$

En base deux (base binaire) le nombre 0110 s'écrirait en base 10 :

$$0110_2 = 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 10^0 = 6_{10} \tag{2.5}$$

et ainsi de suite.

L'inverse est toujours un peu plus délicat. Par exemple la conversion du nombre décimal 1 492 en base deux se fait par divisions successives par 2 des restes et expliqué sur la figure 2.1. Ainsi,

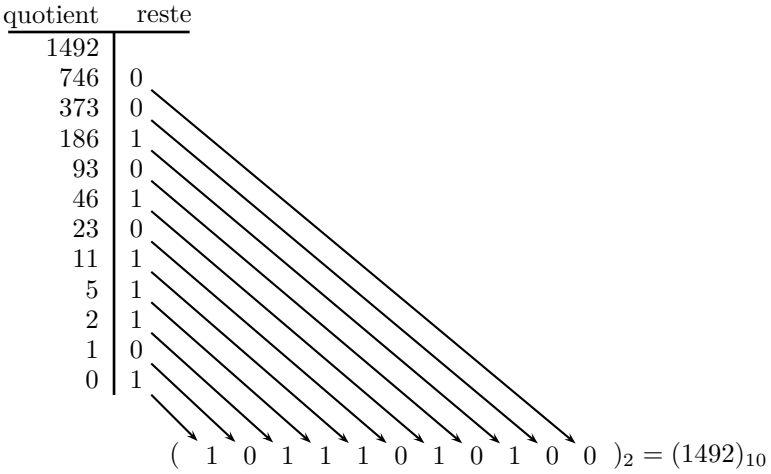


FIGURE 2.1 – Conversion d'un nombre décimal en base deux

pour convertir le nombre 142 713 (base décimale) en base duodécimale (base douze) nous avons (notation : q est le *quotient* et r le *reste*) :

$$\frac{142\,713}{12} \Rightarrow q = 11'892, r = 9 \tag{2.6a}$$

$$\frac{11\,892}{12} \Rightarrow q = 991, r = 0 \tag{2.6b}$$

$$\frac{991}{12} \Rightarrow q = 82, r = 7 \tag{2.6c}$$

$$\frac{82}{12} \Rightarrow q = 6, r = 10 \tag{2.6d}$$

$$\frac{6}{12} \Rightarrow q = 0, r = 6 \tag{2.6e}$$

Ainsi nous avons les restes 6, 10, 7, 0, 9 ce qui nous amène à écrire :

$$142\,713_{10} = 6a709_{10} \tag{2.7}$$

Nous avons choisi pour ce cas particulier la symbolique que nous avions définie précédemment (a-dix) pour éviter toute confusion.

2.2 Types de nombres

Il existe en mathématiques une très grande variété de nombres (naturels, rationnels, réels, irrationnels, complexes, p-adiques, quaternions, transcendants, algébriques, constructibles...) puisque le mathématicien peut à loisirs en créer en ayant uniquement à poser les axiomes (règles) de manipulations de ceux-ci•. Chap4

Cependant, il y en a quelques uns que nous retrouvons plus souvent que d'autres et certains qui servent de base de construction à d'autres et qu'il conviendrait de définir suffisamment rigoureusement (sans aller dans les extrêmes) pour pouvoir savoir de quoi nous parlerons lorsque nous les utiliserons.

2.2.1 Nombres entiers naturels

L'idée du *nombre entier* (nombre pour lequel il n'y a pas de chiffres après la virgule) est le concept fondamental de la mathématique et nous vient à la vue d'un groupement d'objets de même espèce. Lorsque la quantité d'objets d'un groupe est différente de celle d'un autre groupe nous parlons alors de groupe numériquement supérieur ou inférieur quelque soit l'espèce d'objets contenus dans ces groupes. Lorsque la quantité d'objet d'un ou de plusieurs groupes est équivalente, nous parlons alors *d'égalité*. À chaque objet correspond le nombre *un* ou *unité* noté "1".

Pour former des groupements d'objets, nous pouvons opérer ainsi : à un objet, ajouter un autre objet, puis encore un et ainsi de suite ; chacun des groupements, au point de vue de sa collectivité, est caractérisé par un nombre ; il résulte de là qu'un nombre peut être considéré comme représentant un groupement d'unités tel que chacune de ces unités corresponde à un objet de la collection.

Définition 2.2 *Deux nombres sont dits* égaux *si à chacune des unités de l'un nous pouvons faire correspondre une unité de l'autre et inversement. Si ceci ne se vérifie pas alors nous parlons* d'inégalité.

Prenons un objet, puis un autre, puis au groupement formé, ajoutons encore un objet et ainsi de suite. Les groupements ainsi constitués sont caractérisés par des nombres qui, considérés dans le même ordre que les groupements successivement obtenus, constituent la *suite naturelle* notée $\mathbb{N}$ et notée :

$$\mathbb{N} = \{0; 1; 2; 3; 4; \ldots\} \tag{2.8}$$

La présence du 0 (zéro) dans notre définition de $\mathbb{N}$ est discutable étant donné qu'il n'est ni positif ni négatif. C'est la raison pour laquelle dans certains ouvrages vous pourrez trouver une définition de $\mathbb{N}$ sans le 0.

Nous pouvons définir cet ensemble de façon générale et non axiomatique : $\mathbb{N}$ est l'ensemble le plus commun et intuitif d'abstraits quantitatifs arbitraires qui satisfont à des règles subjectives dépendantes de la complexe logique qui en est à l'origine.

Chap4 Les constituants de cet ensemble peuvent être définis par [1] les propriétés• suivantes :

Propriété 2.1 0 *(lire* zéro*) est le nombre d'éléments (défini comme une relation d'équivalence) de tous les ensembles équivalents à (en bijection avec) l'ensemble vide.*

Propriété 2.2 1 *(lire* un*) est le nombre d'éléments de tous les ensembles équivalents à l'ensemble dont le seul élément est 1.*

Propriété 2.3 2 *(lire* deux*) est le nombre d'éléments de tous les ensembles équivalents à l'ensemble dont tous les éléments sont* 0 *et* 1*.*

Propriété 2.4 *En général, un nombre entier est le nombre d'éléments de tous les ensembles équivalents à l'ensemble des nombres entiers le précédent.*

La construction de l'ensemble des entiers naturels s'est faite de la manière la plus naturelle et cohérente qui soit. Les naturels doivent leur nom à ce qu'ils avaient pour objet, aux prémices de leur existence, de dénombrer des quantités et des choses de la nature ou qui intervenaient dans la vie de l'homme. L'originalité de l'ensemble réside dans la manière empirique dont il s'est construit car il ne résulte pas réellement d'une définition mathématique, mais davantage d'une prise de conscience par l'homme du concept de quantité dénombrable, de nombre et de lois qui traduisent des relations entre eux.

La question de l'origine de $\mathbb{N}$ est dès lors la question de l'origine des mathématiques. Et de tout temps des débats confrontant les pensées des plus grands esprits philosophiques ont tenté d'élucider ce profond mystère, à savoir si les mathématiques sont une pure création de l'esprit humain ou si au contraire l'homme n'a fait que redécouvrir une science qui existait déjà dans la nature. Outre les nombreuses questions philosophiques que cet ensemble peut susciter, il n'en est pas moins intéressant d'un point de vue exclusivement mathématique. Du fait de sa structure, il présente des propriétés remarquables qui peuvent se révéler d'une grande utilité lorsque l'on pratique certains raisonnements ou calculs.

1. Nous devons cette définition au mathématicien GOTTLOB

La suite naturelle des nombres entiers est illimitée• mais dénombrable, car, à un groupement d'objets qui se trouve représenté par un certain nombre n, il suffira d'ajouter un objet pour obtenir un autre groupement qui sera défini par un nombre entier immédiatement supérieur equation. Chap4

Définition 2.3 *Deux nombres entiers qui diffèrent d'une unité positive sont dits* consécutifs.

Axiomes de Peano

Lors de la crise des fondements des mathématiques, les mathématiciens ont bien évidemment cherché à axiomatiser l'ensemble $\mathbb{N}$ et nous devons l'axiomatisation actuelle à Peano et à Dedekind.

Les axiomes de ce système comportent les symboles $<$ et $=$ pour représenter les relations *plus petit* et *égal*•. Ils comprennent d'autre part les symboles 0 pour le nombre zéro et les symboles s pour représenter le nombre *successeur*. Dans ce système, 1 est noté $s(0)$ dit *successeur de zéro*, 2 est noté $s(s(0))$. Chap3

Les axiomes de Peano qui construisent $\mathbb{N}$ sont les suivants• : Chap1

Axiome 2.1 *0 est un entier naturel (permet de poser que $\mathbb{N}$ n'est pas vide)*

Axiome 2.2 *Tout entier naturel a un successeur, noté $s(n)$.*

Donc s est une application injective, c'est- à-dire :

$$\forall x, y s(x) = s(y) \Rightarrow x = y \tag{2.9}$$

Si deux successeurs sont égaux, ils sont les successeurs d'un même nombre (cette propriété est présentée souvent à tort comme un axiome sur un grand nombre de sites internet).

Axiome 2.3 *$\forall x \neg(s(x) = 0 =$, le successeur d'un entier naturel n'est jamais égal à zéro (ainsi $\mathbb{N}$ à un premier élément)*

Axiome 2.4 *$\varphi(0) \wedge \forall x \ \ (\varphi(x) \Rightarrow \varphi(s(x))) \Rightarrow \forall x \ \ \varphi(x)$*, axiome de récurrence *qui se doit se lire de la manière suivante : si l'on démontre qu'une propriété est vraie pour un x et son successeur, alors cette propriété est vraie pout tout x.*

Donc l'ensemble de tous les nombres vérifiant les quatre axiomes sont :

$$\mathbb{N} = \{1, 2, 3, \ldots, n, \ldots\} \tag{2.10}$$

Les axiomes de Peano permettent de construire très rigoureusement les deux opérations de base de l'arithmétique que sont l'addition et la multiplication• et ainsi tous les autres ensembles que nous verrons par la suite. Chap3

Nombres pairs, impairs et parfaits

En arithmétique, étudier la parité d'un entier, c'est déterminer si cet entier est ou non un multiple de deux. Un entier multiple de deux est un entier pair, les autres sont les entiers impairs.

Définition 2.4 *Les nombres obtenus en comptant par deux à partir de zéro, (soit 0, 2, 4, 6, 8,...) dans cette suite naturelle sont appelés* nombres pairs.

Le n^{e} nombre impair est donné par :

$$(n + n) = 2n \tag{2.11}$$

Définition 2.5 *Les nombres que nous obtenons en comptant par deux à partir de un (soit 1, 3, 5, 7,...) dans cette suite naturelle s'appellent* nombres impairs.

Le $(n + 1)^{\text{e}}$ nombre impair est donné par :

$$2n + 1 \tag{2.12}$$

Nous appelons *nombres parfaits*, les nombres égaux à la somme de leurs diviseurs entiers strictement plus petits qu'eux mêmes comme par exemple $6 = 1 + 2 + 3$ et $28 = 1 + 2 + 4 + 7 + 14$.

Nombres premiers

Définition 2.6 *Un* nombre premier *est un entier possédant exactement deux diviseurs : "1" et lui-même.*

Il faut noter que la définition de nombre premier exclut le chiffre "1" car il a un unique diviseur (lui-même) et pas deux comme le veut la définition.

Nous pouvons nous demander s'il existe une infinité de nombres premiers ? La réponse est positive et en voici une démonstration par l'absurde.

Preuve Supposons qu'il en existe qu'un nombre fini de nombres premiers qui seraient $p_1, p_2, \ldots, p_n$. Nous formons un nouveau nombre. Le produit de tous les nombres premiers auquel nous ajoutons "1" :

$$N = p_1 p_2 \ldots p_n + 1 \tag{2.13}$$

Selon notre hypothèse initiale ce nouveau nombre devrait être divisible par l'un des nombres premiers existants d'après le théorème fondamental de l'arithmétique• :

Chap5

$$N = q p_i \tag{2.14}$$

Nous pouvons effectuer la division :

$$q = \frac{p_1 p_2 \ldots p_n}{p_i} + \frac{1}{p_i} \tag{2.15}$$

Le premier terme se simplifie, car p_i est dans le produit. On note E cet entier :

$$q = E + \frac{1}{p_i} \tag{2.16}$$

Or q et E sont deux entiers donc p_i doit être un entier. Cependant p_i est par définition supérieur à 1 donc $1/p_i$ n'est pas un entier. Il y a une contradiction et par conséquent, les nombres premiers sont en nombre infini. •

REMARQUES

1. $p_n = p_1 p_2 \ldots p_n$ (le produit des n premiers nombres premiers) est appelé *primorielle n*.
2. Nous renvoyons le lecteur au chapitre de Cryptographie de la section d'Informatique Théorique pour étudier quelques propriétés remarquables des nombres premiers dont la non moins fameuse fonction phi d'EULER (ou appelé aussi *fonction indicatrice*).
3. L'étude des nombres premiers est un sujet immensément vaste et certains théorèmes relatifs sortent largement du cadre d'étude de ce site.

2.2.2 Nombres entiers relatifs

L'ensemble $\mathbb{N}$ à quelques défauts que nous n'avons pas énoncés tout à l'heure. Par exemple, la soustraction de deux nombres dans $\mathbb{N}$ n'a pas toujours un résultat dans $\mathbb{N}$ (les nombres négatifs n'y existent pas). Autre défaut, la division de deux nombres dans $\mathbb{N}$ n'a également pas toujours un résultant dans $\mathbb{N}$ (les nombres fractionnaires n'y existent pas).

Ainsi, nous pouvons dans un premier temps résoudre le problème de la soustraction en ajoutant à l'ensemble des entiers naturels, les entiers négatifs nous obtenons *l'ensemble des entiers relatifs* noté $\mathbb{Z}$ (pour Zahl de l'allemand) :

$$\mathbb{Z} = \{\ldots; -3; -2; -1; 0; 1; 2; 3; \ldots\} \tag{2.17}$$

L'ensemble des entiers naturels est donc inclus dans l'ensemble des entiers relatifs. C'est ce que nous notons sous la forme :

$$\mathbb{N} \subset \mathbb{Z} \tag{2.18}$$

et nous avons par définition :

$$\mathbb{Z}_+ = \mathbb{N} \tag{2.19}$$

Cet ensemble à été crée à l'origine pour faire de l'ensemble des entiers naturels un objet que nous appelons un groupe• par rapport à l'addition.

Chap4

Définition 2.7 *Nous disons qu'un ensemble E est un* ensemble dénombrable, *s'il est équipotent à $\mathbb{N}$, c'est-à-dire s'il existe une bijection• de $\mathbb{N}$ sur E. Ainsi, deux ensembles équipotents ont* autant *d'éléments au sens de leurs cardinaux•, ou tout au moins la même infinité.* Chap4 Chap4

L'objectif de cette remarque est de faire comprendre que les ensembles $\mathbb{N}, \mathbb{Z}$ sont dénombrables.

Démontrons que $\mathbb{Z}$ est dénombrable en posant :

$$x_{2k} = k \quad \text{et} \quad x_{2k+1} = -k - 1 \tag{2.20}$$

pour tout entier $k \geqslant 0$. Ceci donne l'énumération suivante $0, -1, 1, -2, 2, -3, 3, \ldots$ de tous les entiers relatifs.

2.2.3 Nombres rationnels

L'ensemble $\mathbb{Z}$ à aussi un défaut. Ainsi, la division de deux nombres dans $\mathbb{Z}$ n'a également pas toujours un résultat dans $\mathbb{Z}$ (les nombres fractionnaires n'y existent pas). Nous disons alors dans le langage de la théorie des ensembles que la division n'est pas une opération interne dans $\mathbb{Z}$.

Nous pouvons ainsi définir un nouvel ensemble qui contient tous les nombres qui peuvent s'écrire sous forme de fraction (rapport d'un dividende et d'un diviseur entiers) ou autrement dit : les nombres fractionnaires. Cet *ensemble des nombres rationnels* est noté :

$$\mathbb{Q} = \left\{ \frac{p}{q} | (p, q) \in \mathbb{Z} \;\; q \neq 0 \right\} \tag{2.21}$$

et où p et q sont des entiers sans facteurs communs.

Nous supposerons par ailleurs comme évident que :

$$\mathbb{N}, \mathbb{Z} \subset \mathbb{Q} \tag{2.22}$$

La logique de la création de l'ensemble des nombres rationnels est similaire à celle des entiers relatifs. Effectivement, les mathématiciens ont souhaité faire de l'ensemble des nombre relatifs un groupe• par rapport à la loi de multiplication et de division. Chap4|p97

De plus, contrairement à l'intuition, l'ensemble des nombres entiers et nombres rationnels sont équipotents. Nous pouvons nous persuader de cette équipotence en rangeant comme le fit Cantor, les rationnels dans un premier temps de la façon suivante :

equation

Ce tableau est construit de telle manière que chaque rationnel n'apparaît qu'une seule fois (au sens de sa valeur décimale) par diagonale d'où le nom de la méthode : *diagonale de* Cantor.

Nous définissons ainsi une application $f : \mathbb{N} \to \mathbb{Q}$ qui est injective (deux rationnels distincts admettent des rangs distincts) et surjective (à tout place sera inscrit un rationnel : on le cherche par son dénominateur, puis en colonne pour son numérateur).

L'application f est donc bijective : $\mathbb{N}$ et $\mathbb{Q}$ sont donc bien équipotents. La définition un peu plus rigoureuse de $\mathbb{Q}$ se fait à partir de $\mathbb{Z}$ en procédant comme suit :

Définition 2.8 *Sur l'ensemble $\mathbb{N} \to \mathbb{Q}$, qu'il faut lire comme étant l'ensemble construit à partir de deux éléments entiers relatifs dont on exclut le zéro pour le deuxième, on considère la relation R entre deux couples d'entiers relatifs définie par :*

$$(a, b) R (a', b') \Leftrightarrow a \cdot b' = a' \cdot b \tag{2.23}$$

Nous vérifions facilement ensuite que R est une relation d'équivalence• sur $\mathbb{Z} \times \mathbb{Z} \backslash \{0\}$. Chap3

L'ensemble des classes d'équivalence pour cette relation R noté alors $(\mathbb{Z} \times \mathbb{Z} \backslash \{0\})/R$ est par définition $\mathbb{Q}$. C'est-à-dire que nous posons alors plus rigoureusement :

$$\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z} \backslash \{0\})/R \tag{2.24}$$

La classe d'équivalence de $(a, b) \in \mathbb{Z} \times \mathbb{Z} \backslash \{0\}$ est notée :

$$\frac{a}{b} \tag{2.25}$$

conformément à la notation que tout le monde a l'habitude d'employer.

Nous vérifions facilement que l'addition et la multiplication qui étaient des opérations définies sur $\mathbb{Z}$ passent sans problème à $\mathbb{Q}$ en posant :

$$\frac{a}{b}\cdot\frac{p}{q}=\frac{a\cdot p}{b\cdot q}\quad\text{et}\quad\frac{a}{b}+\frac{p}{q}=\frac{aq+bp}{bq} \tag{2.26}$$

Chap4 De plus ces opérations munissent $\mathbb{Q}$ d'une structure de corps• avec $0/1$ comme élément neutre additif et $1/1$ comme élément neutre multiplicatif. Ainsi, tout élément non nul de $\mathbb{Q}$ est inversible, en effet :

$$\frac{a}{b}\cdot\frac{b}{a}=\frac{ab}{ab}=\frac{1}{1} \tag{2.27}$$

ce qui s'écrit aussi plus techniquement $(ab,ab)R(1,1)$. Même si nous aurions envie de définir $\mathbb{Q}$ comme étant l'ensemble $\mathbb{Z}\times\mathbb{Z}\backslash\{0\}$ où $\mathbb{Z}$ représente les numérateurs et $\mathbb{Z}\backslash\{0\}$ les dénominateurs des rationnels, ceci n'est pas possible car autrement nous aurions par exemple $(1,2)\neq(2,4)$ tandis que nous nous attendons à une égalité.

D'où le besoin d'introduire une relation d'équivalence qui nous permet d'identifier, pour revenir à l'exemple précédent, $(1,2)$ et $(2,4)$. La relation R que nous avons définie ne tombe pas du ciel, en effet le lecteur qui a manipulé les rationnels jusqu'à présent sans jamais avoir vu leur définition formelle sait que :

$$\frac{a}{b}=\frac{a'}{b'}\Leftrightarrow ab'=a'b \tag{2.28}$$

Il est donc naturel de définir la relation R comme nous l'avons fait. En particulier, en ce qui concerne l'exemple ci-dessus, $\frac{1}{2}=\frac{2}{4}$ car $(1,2)R(2,4)$ et le problème est résolu.

Outre les circonstances historiques de sa mise en place, ce nouvel ensemble se distingue des ensembles d'entiers relatifs car il induit la notion originale et paradoxale de quantité partielle. Cette notion qui *a priori* n'a pas de sens, trouvera sa place dans l'esprit de l'homme notamment grâce à la géométrie où l'idée de fraction de longueur, de proportion s'illustre plus intuitivement.

2.2.4 Nombres irrationnels

L'ensemble des rationnels $\mathbb{Q}$ est limité et non suffisant lui aussi. Effectivement, nous pourrions penser que tout calcul mathématique numérique avec les opérations communément connues se réduisent à cet ensemble mais ce n'est pas le cas.

Exemple 2.3 *Prenons le calcul de la racine carrée de deux que nous noterons $\sqrt[2]{2}$. Supposons que $\sqrt[2]{2}$ soit rationnel. Alors, nous devrions pouvoir l'exprimer comme a/b, où a et b sont des entiers sans facteurs commun. Pour cette raison, a et b ne peuvent tous les deux êtres pairs. Il y a deux possibilités :*

1. *a est impair (b est alors pair)*
2. *a est pair (b est alors impair)*

En mettant au carré, nous avons :

$$\sqrt[2]{2}=\frac{a}{b} \tag{2.29}$$

qui peut s'écrire :

$$2b^2=a^2 \tag{2.30}$$

Puisque le carré d'un nombre impair est impair et le carré d'un nombre pair est pair, le cas (1) est impossible, car a^2 serait impair et $2b^2$ serait pair.

Le cas (2) est aussi est aussi impossible, car alors nous pourrions écrire $a=2c$, où c est un entier quelconque, et donc si nous le portons au carré, il vient $a^2=4c^2$ où nous avons un nombre pair des deux côtés de l'égalité. En remplaçant dans $2b^2=a^2$ nous obtenons $b^2=2c^2$. Par conséquent, b^2 serait impair alors que $2c^2$ serait pair. Il n'y a pas de solution ; c'est donc que l'hypothèse de départ est fausse et qu'il n'existe pas deux entiers a et b tels que $\sqrt[2]{2}=a/b$.

Exemple 2.4 *Démontrons, aussi par l'absurde, que le fameux nombre d'*Euler *e est irrationnel. Pour cela, rappelons que e (cf. chapitre d'Analyse Fonctionnelle) peut aussi être défini par la série de* Taylor *(cf. chapitre sur les Suites Et Séries) :*

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!} + \ldots \tag{2.31}$$

Alors si e est rationnel, il doit pouvoir s'écrire sous la forme p/q (avec equation, car nous savons que e n'est pas entier). Multiplions les deux côtés de côtés de l'égalité par $q!$:

$$q!e = q! + \frac{q!}{1!} + \frac{q!}{2!} + \frac{q!}{3!} + \ldots + \frac{q!}{q!} + \frac{q!}{(q+1)!} + \frac{q!}{(q+2)!} + \ldots \tag{2.32}$$

Le premier membre, $q!$, est un entier, car, par définition :

$$q! = q \cdot (q-1) \cdot (q-2) \ldots 2 \cdot 1 \tag{2.33}$$

d'où :

$$q!\frac{p}{q} = (q-1)!p \tag{2.34}$$

Les premiers termes du second membre, jusqu'au terme $q!/q! = 1$ sont aussi des entiers (car $q!/m!$ se simplifie si $q > m$). Donc, par soustraction, nous trouvons :

$$q!e - \left(q! + \frac{q!}{1!} + \frac{q!}{2!} + \frac{q!}{3!} + \ldots + \frac{q!}{q!} + \ldots \right) = \frac{q!}{(q+1)!} + \frac{q!}{(q+2)!} + \ldots \tag{2.35}$$

devrait aussi être un entier. Après simplification, le second membre de l'égalité devient :

$$\frac{1}{q+1} + \frac{1}{(q+1)(q+2)} + \ldots \tag{2.36}$$

le premier terme de cette somme est strictement inférieur à $1/2$, le deuxième inférieur à $1/4$, le troisième inférieur à $1/8$, etc. Par conséquent, puisque chaque terme est strictement inférieur aux termes de la série harmonique suivant qui converge vers 1 :

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1 \tag{2.37}$$

Par conséquent, elle n'est pas un entier (étant strictement inférieur à 1), ce qui constitue une contradiction. Ainsi, les nombres rationnels ne satisfont pas à l'expression numérique de $\sqrt[2]{2}$ comme de e. Il faut donc les compléter par l'ensemble de tous les nombres qui ne peuvent s'écrire sous forme de fraction et que nous appelons des nombres irrationnels.

2.2.5 Nombres réels

Définition 2.9 *La réunion des nombres rationnels et irrationnels donne* l'ensemble des nombres réels.

Donc nous pouvons écrire que :

$$\mathbb{Q} \subset \mathbb{R} \tag{2.38}$$

Les mathématiciens dans leur rigueur habituelle ont différentes techniques pour définir les nombres réels. Ils utilisent pour cela des propriétés de la topologie (entre autres) et en particulier les suites de Cauchy mais c'est une autre histoire qui dépasse le cadre formel du présent chapitre.

Nous sommes évidemment amenés à nous demander si $\mathbb{R}$ est dénombrable. La démonstration est assez simple.

Preuve Par définition, nous avons vu plus haut qu'il doit y avoir une bijection entre $\mathbb{Q}$ et $\mathbb{R}$ pour que dire que $\mathbb{R}$ est dénombrable.

Pour simplifier, nous allons montrer que l'intervalle $]0, 1[$ n'est alors pas dénombrable ; ce qui implique bien sûr par extension que $\mathbb{R}$ ne l'est pas. Les éléments de cet intervalle sont représentés par des suites infinies entre 0 et 9 (dans le système décimal) :

$n°1$	x_{11}	x_{12}	x_{13}	x_{14}	$\cdots$	x_{1p}	$\cdots$
$n°2$	x_{21}	x_{22}	x_{23}	x_{24}	$\cdots$	x_{2p}	$\cdots$
$n°3$	x_{31}	x_{32}	x_{33}	x_{34}	$\cdots$	x_{3p}	$\cdots$
$n°4$	x_{41}	x_{42}	x_{43}	x_{44}	$\cdots$	x_{4p}	$\cdots$
$n°5$	x_{51}	x_{52}	x_{53}	x_{54}	$\cdots$	x_{5p}	$\cdots$
$n°6$	x_{61}	x_{62}	x_{63}	x_{64}	$\cdots$	x_{6p}	$\cdots$
	$\cdots$						
		$\cdots$					
			$\cdots$				
$n°k$	x_{k1}	x_{k2}	x_{k3}	x_{k4}	$\cdots$	x_{kp}	$\cdots$
						$\cdots$	

TABLEAU 2.1 – (2.50)

– Certaines de ces suites sont nulles à partir d'un certain rang, d'autres non
– Nous pouvons donc identifier $[0, 1[$ à l'ensemble de toutes les suites (finies ou infinies) d'entiers compris entre 0 et 9

Si cet ensemble était dénombrable, nous pourrions les classer. Ainsi, la x_{11}, x_{12}, x_{13}, $x_{14} \ldots x_{1p} \ldots$ serait classée première et ainsi de suite. On pourrait alors modifier cette matrice infinie de la manière suivante : à chaque élément de la diagonale, ajouter 1, selon la règle : $0 + 1 = 1$, $1 + 1 = 2$, $8 + 1 = 9$ et $9 + 1 = 0$. Considérons alors la suite infinie qui se trouve sur la diagonale :

$n°1$	$x_{11} + 1$	x_{12}	x_{13}	x_{14}	$\cdots$	x_{1p}	$\cdots$
$n°2$	x_{21}	$x_{22} + 1$	x_{23}	x_{24}	$\cdots$	x_{2p}	$\cdots$
$n°3$	x_{31}	x_{32}	$x_{33} + 1$	x_{34}	$\cdots$	x_{3p}	$\cdots$
$n°4$	x_{41}	x_{42}	x_{43}	$x_{44} + 1$	$\cdots$	x_{4p}	$\cdots$
$n°5$	x_{51}	x_{52}	x_{53}	x_{54}	$\cdots$	x_{5p}	$\cdots$
$n°6$	x_{61}	x_{62}	x_{63}	x_{64}	$\cdots$	x_{6p}	$\cdots$
	$\cdots$						
		$\cdots$					
			$\cdots$				
$n°k$	x_{k1}	x_{k2}	x_{k3}	x_{k4}	$\cdots$	x_{kp}	$\cdots$
						$\cdots$	

TABLEAU 2.2 – (2.51)

– elle ne peut être égale à la première car elle s'en distingue au moins par le premier élément ;
– elle ne peut être égale à la deuxième car elle s'en distingue au moins par le deuxième élément ;
– elle ne peut être égale à la troisième car elle s'en distingue au moins par le troisième élément ;
– et ainsi de suite...
– elle ne peut donc être égale à aucune des suites contenues dans ce tableau.

Par conséquent, quel que soit le classement choisi des suites infinies de $0 \ldots 9$, il y en a toujours une qui échappe à ce classement ! C'est donc qu'il est impossible de les numéroter, tout simplement parce qu'elles ne forment pas un ensemble dénombrable. •

La technique qui nous a permis d'arriver à ce résultat est connue sous le nom de *procédé diagonal de* Cantor et l'ensemble des nombres réels est dit avoir *la puissance du continu* de par le fait qu'il est indénombrable.

Nous supposerons intuitif pour le lecteur que tout nombre réel peut être approché infiniment près par un nombre rationnel [2]. Les mathématiciens disent alors que $\mathbb{Q}$ est *dense* dans $\mathbb{R}$ et notent cela :

$$\bar{\mathbb{Q}} = \mathbb{R} \tag{2.39}$$

Nombres transfinis

Nous nous retrouvons donc avec un *infini* des nombres réels qui est différent de celui des nombres naturels. CANTOR osa alors ce que personne n'avait osé depuis ARISTOTE : la suite des entiers

2. Pour les nombres irrationnels il suffit de s'arrêter à un nombre de décimales données et d'en trouver le rationnel correspondant.

positifs est infinie, l'ensemble $\mathbb{N}$, est donc un ensemble qui a une infinité d'éléments, alors il affirma que le cardinal• de cet ensemble était un nombre qui existait comme tel sans que l'on utilise le symbole ∞, il le nota : Chap4

$$\aleph_0 = \text{Card}(\mathbb{N}) \tag{2.40}$$

Ce symbole est la première lettre de l'alphabet hébreu et se prononce *aleph zéro*. Cantor allait appeler ce nombre étrange, un nombre *transfini*.

L'acte décisif est d'affirmer qu'il y a, après le fini, un transfini, c'est-à-dire une échelle illimitée de modes déterminés qui par nature sont infinis, et qui cependant peuvent êtres précisés, tout comme le fini, par des nombres déterminés, bien définis et distinguables les uns des autres.

Après ce premier coup d'audace allant à l'encontre de la plupart des idées reçues depuis plus de deux mille ans, Cantor allait poursuivre sa lancée et établir des règles de calcul, paradoxales à première vue, sur les nombres transfinis. Ces règles se basaient, comme nous l'avons défini tout à l'heure, sur le fait que deux ensembles infinis sont équivalent s'il existe une bijection entre les deux ensembles.

Ainsi, nous pouvons facilement montrer que l'infini des nombres pairs est équivalent à l'infini des nombres entiers : pour cela, il suffit de montrer qu'à chaque nombre entier, nous pouvons associer un nombre pair, son double, et inversement.

Même si les nombres pairs sont inclus dans l'ensemble des nombres entiers, il y en a une infinité $\aleph_0$ égal, les deux ensembles sont équipotents. En affirmant qu'un ensemble peut être égale à une de ses parties, Cantor va à l'encontre ce qui semblait être une évidence pour Aristote et Euclide : l'ensemble de tous les ensembles est infini ! Cela va ébranler la totalité des mathématiques et va amener à l'axiomatisation de Zermelo-Frankel que nous verrons en théorie des ensembles.

À partir de ce qui précède, Cantor établit les règles de calculs suivants sur les cardinaux :

$$\aleph_0 + 1 = \aleph_0; \quad \aleph_0 + \aleph_0 = \aleph_0; \quad \aleph_0^2 = \aleph_0 \tag{2.41}$$

Nous pouvons remarquer que ces règles ne sont pas nécessairement intuitives.

À première vue ces règles semblent non intuitives mais en fait elles le sont bien. En effet, Cantor définit l'addition de deux nombres transfinis comme le cardinal de l'union disjointe des ensembles correspondants.

Exemple 2.5 *En notant donc $\aleph_0$ le cardinal de $\mathbb{N}$ nous avons $\aleph_0 + \aleph_0$ qui est équivalent à dire que nous sommons le cardinal de $\mathbb{N}$ union disjointe $\mathbb{N}$. Or $\mathbb{N}$ union disjointe $\mathbb{N}$ est équipotent à $\mathbb{N}$ donc $\aleph_0 + \aleph_0 = \aleph_0$ (il suffit pour s'en convaincre de prendre l'ensembles des entiers pairs et impairs tout deux dénombrables dont l'union disjointe est dénombrable).*

Exemple 2.6 *$\aleph_0 + 1$ correspond au cardinal de l'ensemble $\mathbb{N}$ union un point. Ce dernier ensemble est encore équipotent à $\mathbb{N}$ donc $\aleph_0 + 1 = \aleph_0$.*

Nous verrons également lors de notre étude de la théorie des ensembles que le concept de produit cartésien de deux ensembles dénombrable est tel que nous ayons :

$$\text{Card}(\mathbb{N} \times \mathbb{N}) = \text{Card}(\mathbb{N}^2) = [\text{Card}(\mathbb{N})]^2 \tag{2.42}$$

et donc :

$$\aleph_0^2 = \aleph_0 \tag{2.43}$$

De même•, puisque $\mathbb{Z} = \mathbb{Z}^+ \cup \mathbb{Z}^-$ et en identifiant $\mathbb{Q}$ à $\mathbb{Z} \times \mathbb{Z}$, nous pouvons aussi écrire : Chap4

$$\aleph_0 + \aleph_0 = \aleph_0 \quad \text{et} \quad \aleph_0 \times \aleph_0 = (\aleph_0)^2 = \aleph_0 \tag{2.44}$$

Nous pouvons d'ailleurs démontrer un énoncé intéressant : si nous considérons le cardinal de l'ensemble de tous les cardinaux, il est nécessairement plus grand que tous les cardinaux, y compris lui-même• : Chap4

Théorème 2.1 *Soit A un ensemble non vide, alors $\text{Card}(A) < \text{Card}(\mathcal{P}(A))$ où $\mathcal{P}(A)$ est l'ensemble des parties de A.*

C'est-à-dire par définition de la relation d'ordre $<$ (strictement inférieur), qu'il faut montrer qu'il n'existe pas d'application surjective $f : A \to \mathcal{P}(A)$, en d'autres termes qu'à chaque élément de l'ensemble des parties de A il ne correspond pas au moins une pré-image dans A. $\mathcal{P}(\mathbb{N})$ est par exemple constitué de l'ensemble des nombres impairs, pairs, premiers, et l'ensemble des naturels, ainsi que l'ensemble vide lui-même, etc. $\mathcal{P}(\mathbb{N})$ est donc l'ensemble de toutes les *patates* possibles qui forment $\mathbb{N}$ (pour emprunter le vocabulaire à la petite école).

Preuve (par l'absurde) L'idée est de supposer que nous pouvons numéroter chacune des patates $\mathcal{P}(A)$ avec au moins un élément de A. En d'autres termes que $f : A \to \mathcal{P}(A)$ est surjective et considérons un sous-ensemble de A tel que :

$$E = \{x \in A | x \notin f(x)\} \tag{2.45}$$

c'est-à-dire l'ensemble d'éléments x de A qui n'appartiennent pas à l'ensemble numéro x (l'élément x n'appartient pas à la patate qu'il numérote en d'autres termes).

Or, si f est surjective il existe alors un $y \in A$ tel que :

$$f(y) = E = \{x \in A | x \notin f(x)\} \tag{2.46}$$

Mais si $f(y) = E$ alors $y \notin f(y)$ et de par la définition de E nous avons alors $y \in E$ et nous avons donc une absurdité de par l'hypothèse de la surjectivité. •

2.2.6 Nombres complexes

Inventés au 18^{e} siècle entre autres par J. CARDAN et R. BOMBELLE, ces nombres permettent de résoudre des problèmes n'ayant pas de solutions dans $\mathbb{R}$ ainsi que de formaliser mathématiquement certaines transformations dans le plan tel que la rotation, la similitude, la translation, etc. Pour les physiciens, les nombres complexes constituent surtout un moyen très commode de simplifier les notations. Il est ainsi très difficile d'étudier les phénomènes ondulatoires, la relativité générale ou la mécanique quantique sans recourir aux nombres et expressions complexes.

Il existe plusieurs manières de construire les nombres complexes. La première est typique de la construction telle que les mathématiciens en ont l'habitude dans le cadre de la théorie des ensembles. Ils définissent un couple de nombres réels et définissent des opérations entre ces couples pour arriver enfin à une signification du concept de nombre complexe. La deuxième est moins rigoureuse mais son approche est plus simple et consiste à définir le nombre imaginaire pur unitaire i et ensuite de construire les opérations arithmétiques à partir de sa définition. Nous allons opter pour cette deuxième méthode.

Définition 2.10 *Nous définissons le* nombre imaginaire unitaire pur *que nous notons i par la propriété suivante :*

$$i^2 = -1 \Leftrightarrow i = \sqrt{-1} \tag{2.47}$$

Définition 2.11 *Un* nombre complexe *est un couple d'un nombre réel a et d'un nombre imaginaire ib et s'écrit généralement sous la forme suivante :*

$$z = a + ib \tag{2.48}$$

a et b étant des nombres appartenant à $\mathbb{R}$. Nous notons l'ensemble des nombres complexes $\mathbb{C}$ et avons donc par construction :

$$\mathbb{R} \subset \mathbb{C} \tag{2.49}$$

Chap?? REMARQUE L'ensemble $\mathbb{C}$ est identifié au plan euclidien orienté $E^{\bullet}$ grâce au choix d'une base orthonormée directe : nous obtenons ainsi le *plan d'*Argand-Cauchy ou plus communément *plan de* Gauss.

Chap4 L'ensemble des nombres complexes qui constitue un corps$^{\bullet}$, et noté equation, est défini, de manière simple pour commencer, dans la notation de la théorie des ensembles par :

$$\mathbb{C} = \{z = (x + iy) | x, y \in \mathbb{R}\} \tag{2.50}$$

En d'autres termes nous disons que le corps $\mathbb{C}$ est le corps $\mathbb{R}$ auquel nous avons *adjoint* le nombre imaginaire i. Ce qui se note formellement $\mathbb{R}[i]$.L'addition et la multiplication de nombres complexes sont des opérations internes à l'ensemble des complexes• et définies par : Chap4

$$\begin{aligned} z_1 + z_2 &= (x_1 + x_2) + i(y_1 + y_2) \\ z_1 \cdot z_2 &= (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) \end{aligned} \tag{2.51}$$

La *partie réelle* de z est traditionnellement notée :

$$\Re(z) = x \tag{2.52}$$

La *partie imaginaire* de z est notée :

$$\Im(z) = x \tag{2.53}$$

Le *conjugué* ou *conjugaison* de z est défini par :

$$\bar{z} = x - iy \tag{2.54}$$

et est aussi parfois noté z^*, en particulier en physique quantique dans certains ouvrages.

À partir d'un complexe et de son conjugué, il est possible de trouver ses parties réelles et imaginaires. Ce sont les relations évidentes suivantes :

$$\Re(z) = \frac{z + \bar{z}}{2} \quad \text{et} \quad \Im(z) = \frac{z - \bar{z}}{2i} \tag{2.55}$$

Le *module* de z, ou *norme*, représente la longueur par rapport au centre du plan de Gauss (voir un peu plus bas ce qu'est le plan de Gauss) et est simplement calculé avec l'aide du théorème de Pythagore :

$$|z| = \sqrt{x^2 + y^2} = \sqrt{z \cdot \bar{z}} \tag{2.56}$$

et est donc toujours un nombre positif ou nul[3]. La division entres deux complexes se calcule comme (le dénominateur étant évidemment non nul) :

$$\frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2} = \frac{x_1 + iy_1}{x_2 + iy_2} \cdot \frac{x_2 - iy_2}{x_2 - iy_2} = \frac{(x_1 x_2 + y_1 y_2) - i(x_1 y_2 - x_2 y_1)}{x_2^2 + y_2^2} \tag{2.57}$$

L'inverse d'un complexe se calculant de façon similaire :

$$\frac{1}{x + iy} = \frac{x - iy}{(x + iy)(x - iy)} = \frac{x - iy}{x^2 + y^2} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2} \tag{2.58}$$

Nous pouvons aussi énumérer huit importantes propriétés du module et du conjugué complexe :

Propriété 2.5 $\forall z \in \mathbb{C}$ *:*

$$|z| = 0 \Leftrightarrow z = 0 \tag{2.59}$$

Preuve Par définition du module $|z| = \sqrt{x^2 + y^2}$, pour que la somme $x^2 + y^2$ soit nulle, la condition nécessaire est que $x = y = 0$. •

Propriété 2.6 $\forall z \in \mathbb{C}$ *:*

$$|z| = |-z| = |\bar{z}| \tag{2.60}$$

Preuve $|\bar{z}| = \sqrt{x^2 + y^2} = \sqrt{(-x)^2 + (-y)^2} = |-z| = \sqrt{x^2 + (-y)^2} = |\bar{z}|$ •

Propriété 2.7 $\forall z \in \mathbb{C}$ *:*

$$\begin{aligned} &|\Re z| \leqslant |z| \quad \textit{avec égalité ssi } z \textit{ est réel} \\ &|\Im z| \leqslant |z| \quad \textit{avec égalité ssi } z \textit{ est imaginaire} \end{aligned} \tag{2.61}$$

3. La notation $|z|$ pour le module n'est pas innocente puisque $|z|$ coïncide avec la valeur absolue de z lorsque z est réel.

Preuve Les deux inégalités (2.61) peuvent s'écrire $|a| \leqslant \sqrt{a^2+b^2}$ et $|b| \leqslant \sqrt{a^2+b^2}$ donc équivalent respectivement à $a^2 \leqslant a^2+b^2$ et $b^2 \leqslant a^2+b^2$ qui sont triviales. Bien entendu $|a| = \sqrt{a^2+b^2}$ si et seulement si $a^2 = a^2+b^2 \Rightarrow b = 0$ avec le même type de raisonnement pour la partie imaginaire. •

Propriété 2.8 *$\forall (z_1, z_2) \in \mathbb{C} \times \mathbb{C}$:*

$$|z_1 z_2| = |z_1||z_2|; \quad z_2 \neq 0 \quad \left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|} \tag{2.62}$$

Preuve Nous avons $|z_1 z_2|^2 = (z_1 z_2)(\bar{z}_1 \bar{z}_2) = (z_1 \bar{z}_1)(z_2 \bar{z}_2) = |z_1|^2 |z_2|^2$ ce qui implique $|z_1 z_2| = |z_1||z_2|$ et $|z_1/z_2|^2 = (z_1 \bar{z}_1)/(z_2 \bar{z}_2) = |z_1|^2/|z_2|^2$. •

Propriété 2.9 *$\forall z \in \mathbb{C}$:*

$$|z|^2 = z \cdot \bar{z} \tag{2.63}$$

Preuve $|z|^2 = \sqrt{x^2+y^2} = (x+iy)(x-iy) = x^2 - ixy + ixy + y^2 = x^2 + y^2 = \bar{Z} \cdot z$ •

Propriété 2.10 *$\forall z, z' \in \mathbb{C}$:*

$$\bar{\bar{z}} = z, \quad \overline{z+z'} = \bar{z} + \bar{z'}, \quad \overline{z \cdot z'} = \bar{z} \cdot \bar{z'} \tag{2.64}$$

Preuve La première identité s'écrit $\bar{\bar{z}} = \overline{\overline{x+iy}} = \overline{x-iy} = x+iy = z$, la seconde $\overline{z+z'} = \overline{(x_1+iy_1)+(x_2+iy_2)} = \overline{(x_1+x_2)+i(y_1+y_2)} = (x_1+x_2)-i(y_1+y_2) = (x_1-iy_1)+(x_2-iy_2) = \bar{z}+\bar{z'}$ et la troisième $\overline{zz'} = \overline{(x_1+iy_1)(x_2+iy_2)} = \overline{(x_1x_2-y_1y_2)+i(x_1y_2+y_1x_2)} = (x_1x_2-y_1y_2)-i(x_1y_2+y_1x_2) = (x_1-iy_1)\cdot(x_2-iy_2) = \bar{z}\bar{z'}$. •

REMARQUES

1. en des termes mathématiques, la première démonstration permet de montrer que la conjugaison complexe est ce que l'on appelle *involutive* dans le sens qu'elle ne fait rien évoluer ;
2. en des termes tout aussi mathématiques, la deuxième démonstration montre que la conjugaison de la somme de deux nombres complexes est ce que nous appelons un *automorphisme du groupe* $(\mathbb{C}, +)$• ; Chap4
3. la troisième démonstration montre que la conjugaison du produit de deux nombres complexes est ce que nous appelons un *automorphisme du corps* $(\mathbb{C}, +, \times)$•. Chap4

Propriété 2.11 *$\forall z, z' \in \mathbb{C}^*$:*

$$\frac{\bar{1}}{z} = \frac{1}{\bar{z}}; \quad \left(\frac{\bar{z}}{z'}\right) = \frac{\bar{z}}{\bar{z'}} \tag{2.65}$$

Preuve Nous restreignons la démonstration à la seconde relation qui est un cas général de la première. La décomposition de la fraction est la suivante :

$$\begin{aligned}\left(\frac{\bar{z}}{z'}\right) &= \overline{\left(\frac{x_1+iy_1}{x_2+iy_2}\right)} = \overline{\left(\frac{x_1+iy_1}{x_2+iy_2}\right)\left(\frac{x_2-iy_2}{x_2-iy_2}\right)} = \frac{\overline{(x_1x_2+y_1y_2)+i(y_1x_2-y_2x_1)}}{x_2^2+y_2^2} \\ &= \frac{x_1x_2+y_1y_2}{x_2^2+y_2^2} - i\frac{y_1x_2-y_2x_1}{x_2^2+y_2^2} = \frac{(x_1x_2+y_1y_2)+i(y_2x_1-y_1x_2)}{x_2^2+y_2^2} \\ &= \frac{(x_1-iy_1)(x_2+iy_2)}{x_2^2+y_2^2} = \frac{x_1-iy_1}{x_2-iy_2}\frac{x_2+iy_2}{x_2+iy_2} = \frac{\bar{z}}{\bar{z'}}\end{aligned} \tag{2.66}$$

•

Propriété 2.12 *$\forall (z_1, z_2) \in \mathbb{C}^* \times \mathbb{C}^*$*

$$|z_1+z_2| \leqslant |z_1| + |z_2| \tag{2.67}$$

De plus l'égalité a lieu si et seulement si z_1 et z_2 sont colinéaires et de même sens, autrement dit s'il existe $\lambda \in \mathbb{R}_+$ tel que $z_1 = \lambda z_2$.

Preuve

$$|z_1+z_2|^2 = |z_1|^2 + 2\Re(z_1\bar{z}_2) + |z_2|^2 \leqslant |z_1|^2 + 2|z_1\bar{z}_2| + |z_2|^2 = (|z_1|^2+|z_2|^2)^2 \tag{2.68}$$

Cette inégalité peut ne pas paraître évidente à tout le monde alors développons un peu et supposons-la vraie :

$$\begin{aligned}
&|z_1+z_2|^2 \overset{?}{\leqslant} (|z_1|+|z_2|)^2 \\
&\sqrt{(x_1+x_2)^2+(y_1+y_2)^2}^2 \overset{?}{\leqslant} \left(\sqrt{x_1^2+y_1^2}+\sqrt{x_2^2+y_2^2}\right)^2 \\
&(x_1+x_2)^2+(y_1+y_2)^2 \overset{?}{\leqslant} (x_1^2+y_1^2)+2\sqrt{x_1^2+y_1^2}\sqrt{x_2^2+y_2^2}+(x_2^2+y_2^2) \\
&x_1^2+2x_1x_2+x_2^2+y_1^2+2y_1y_2+y_2^2 \overset{?}{\leqslant} (x_1^2+y_1^2)+2\sqrt{x_1^2+y_1^2}\sqrt{x_2^2+y_2^2}+(x_2^2+y_2^2)
\end{aligned} \tag{2.69}$$

Après simplification :

$$\begin{aligned}
&x_1x_2+y_1y_2 \overset{?}{\leqslant} \sqrt{x_1^2+y_1^2}\sqrt{x_2^2+y_2^2} \\
&(x_1x_2+y_1y_2)^2 \overset{?}{\leqslant} (x_1^2+y_1^2)(x_2^2+y_2^2) \\
&x_1^2x_2^2+2x_1x_2y_1y_2+y_1^2y_2^2 \overset{?}{\leqslant} x_1^2x_2^2+x_1^2y_2^2+y_1^2x_2^2+y_1^2y_2^2
\end{aligned} \tag{2.70}$$

et encore après simplification :

$$\begin{aligned}
&2x_1x_2y_1y_2 \overset{?}{\leqslant} x_1^2y_2^2+y_1^2x_2^2 \\
&0 \overset{?}{\leqslant} x_1^2y_2^2-2x_1x_2y_1y_2+y_1^2x_2^2 \\
&0 \overset{?}{\leqslant} (x_1y_2-y_1x_2)^2
\end{aligned} \tag{2.71}$$

Cette dernière relation est vérifiée et démontre donc que l'inégalité est vraie. •

Remarque Il existe une forme plus générale de cette inégalité appelée "inégalité de Minkowski" présentée dans le chapitre de Calcul Vectoriel puisque les nombres complexes peuvent effectivement s'écrire sous la forme de vecteurs comme nous allons le voir maintenant.

Interprétation géométrique

Nous pouvons aussi représenter un nombre complexe $a+ib$ ou $a-ib$ dans un plan délimité par deux axes (deux dimensions) de longueur infinie et orthogonaux entres eux. L'axe vertical représente alors la partie imaginaire d'un nombre complexe et l'axe horizontal, la partie réelle comme illustré sur la figure 2.2.

Il y donc bijection entre l'ensemble des nombres complexes et l'ensemble des vecteurs du plan de Gauss. On nomme parfois ce type de représentation "plan de Gauss".

Nous écrivons alors :

$$\text{Aff}(\vec{r}) = a+ib \tag{2.72}$$

Nous voyons sur ce diagramme qu'un nombre complexe a donc une interprétation vectorielle (cf. chapitre de Calcul Vectoriel) donnée par :

$$z = a\begin{pmatrix}1\\0\end{pmatrix} + b\begin{pmatrix}0\\1\end{pmatrix} = a+ib \tag{2.73}$$

où la base canonique est définie telle que :

$$\left\{1 = \begin{pmatrix}1\\0\end{pmatrix}, i = \begin{pmatrix}0\\1\end{pmatrix}\right\} \tag{2.74}$$

avec :

$$r = |z| = \sqrt{a^2+b^2} \tag{2.75}$$

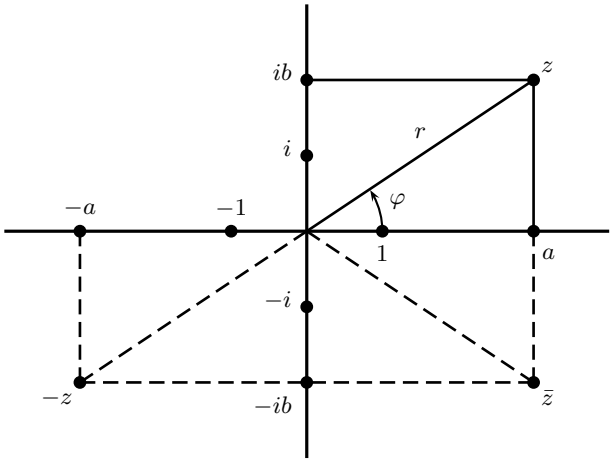


FIGURE 2.2 – Interprétation géométrique des nombres complexes

Ainsi, $(1\ 0)^T$ est le vecteur de la base unitaire porté par l'axe horizontal $\mathbb{R}$ et $(0\ 1)^T$ est le vecteur de la base unitaire porté par l'axe imaginaire $\mathbb{R}_i$ et r est le module (la norme) positif ou nul. Ceci est à comparer avec les vecteurs de $\mathbb{R}^2$ (cf. chapitre de Calcul Vectoriel) :

$$\vec{v} = x\vec{e}_1 + y\vec{e}_2 = x\begin{pmatrix}1\\0\end{pmatrix} + y\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}x\\y\end{pmatrix} \tag{2.76}$$

avec $\| \vec{v} \| = \sqrt{x^2+y^2}$, ce qui fait que nous pouvons identifier le plan complexe avec la plan euclidien.

Par ailleurs, la définition du cosinus et sinus (cf. chapitre de Trigonométrie) nous donne :

$$a = r\cos\varphi \qquad b = r\sin\varphi \tag{2.77}$$

Finalement :

$$r = \sqrt{a^2+b^2} \quad \text{et} \quad \varphi = \cos^{-1} a/r = \sin^{-1} b/r = \tan^{-1} b/a \tag{2.78}$$

Ainsi :

$$z = a + ib = r\cos\varphi + ir\sin\varphi = r(\cos\varphi + i\sin\varphi) = r\text{cis}\varphi \tag{2.79}$$

complexe qui est toujours égal à la lui-même modulo 2π de par les propriétés des fonctions trigonométriques :

$$z = r(\cos\varphi + i\sin\varphi) = r(\cos(\varphi + 2k\pi) + i\sin(\varphi + 2k\pi)) \tag{2.80}$$

avec $k \in \mathbb{N}$ et où φ est appelé *l'argument de* z et est noté traditionnellement $\arg(z)$. Les propriétés du cosinus et du sinus (cf. chapitre de Trigonométrie) nous amènent directement à écrire pour l'argument :

$$\arg\bar{z} = -\arg z \quad \text{et} \quad \arg(-z) = \arg z + \pi \tag{2.81}$$

Nous démontrons entre autres avec les séries de TAYLOR (cf. chapitre des Suites Et Séries) que :

$$\cos\varphi = 1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \ldots + (-1)^k \frac{\varphi^{2k}}{(2k)!} \tag{2.82}$$

et :

$$\sin\varphi = \varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \ldots + (-1)^k \frac{\varphi^{2k+1}}{(2k+1)!} \tag{2.83}$$

dont la somme est semblable à :

$$e^{\varphi} = 1 + \varphi + \frac{\varphi^2}{2!} + \frac{\varphi^3}{3!} + \ldots + (-1)^k \frac{\varphi^k}{k!} \tag{2.84}$$

mais par contre parfaitement identique au développement de Taylor de $e^{i\varphi}$:

$$e^{i\varphi} = 1 + i\varphi - \frac{\varphi^2}{2!} - i\frac{\varphi^3}{3!} + \ldots + i^k \frac{\varphi^k}{k!} = \cos\varphi + i\sin\varphi \tag{2.85}$$

Donc finalement, nous pouvons écrire :

$$z = r(\cos\varphi + i\sin\varphi) = re^{i\varphi} \tag{2.86}$$

relation nommée "formule d'Euler".

En supposant connues les relations trigonométriques de bases (cf. chapitre de Trigonométrie) nous avons les relations suivantes pour la multiplication de deux nombres complexes :

$$z_1 \cdot z_2 = r_1 e^{i\varphi_1} r_2 e^{i\varphi_2} = r_1 r_2 e^{i(\varphi_1+\varphi_2)} \tag{2.87}$$

dès lors :

$$\arg(z_1 \cdot z_2) = \arg(z_1) + \arg(z_2) \tag{2.88}$$

et donc :

$$\arg(nz) = n\arg z \tag{2.89}$$

Pour le module de la multiplication :

$$|z_1 z_2| = |r_1 e^{i\varphi_1} r_2 e^{i\varphi_2}| = |r_1 r_2 e^{i(\varphi_1+\varphi_2)}| = r_1 r_2 = |r_1||r_2| \tag{2.90}$$

d'où :

$$|z^m| = |z|^m \tag{2.91}$$

Pour la division de deux nombres complexes :

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}\frac{e^{i\varphi_1}}{e^{i\varphi_2}} = \frac{r_1}{r_2}e^{i(\varphi_1-\varphi_2)} \tag{2.92}$$

Le module de leur division vient alors immédiatement :

$$\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|} \tag{2.93}$$

dès lors nous avons pour l'argument, d'après l'équation (2.92) :

$$\arg\left(\frac{z_1}{z_2}\right) = \arg\left(\frac{r_1}{r_2}e^{i(\varphi_1-\varphi_2)}\right) = \varphi_1 - \varphi_2 = \arg z_1 - \arg z_2 \tag{2.94}$$

ainsi il vient immédiatement :

$$\arg z^{-1} = -\arg z \tag{2.95}$$

Dans le cas où nous avons un module unité tel que $z = \cos\varphi + i\sin\varphi$ nous avons alors la relation :

$$(\cos\varphi + i\sin\varphi)^m = \left(e^{i\varphi}\right)^m = e^{im\varphi} = \cos(m\varphi) + i\sin(m\varphi) \tag{2.96}$$

appelée "formule de Moivre".

Pour le logarithme népérien d'un nombre complexe, nous avons trivialement la relation suivante sur laquelle nous reviendrons dans le chapitre d'Analyse Complexe :

$$\ln z = \ln\left(re^{i\varphi}\right) = \ln z + i\varphi \tag{2.97}$$

où $\ln z$ est souvent dans le cas complexe écrit $\mathrm{Log} z$.

Toutes les relations précédentes pourraient bien sûr être obtenues avec la forme trigonométrique des nombres complexes mais nécessiteraient alors quelques lignes supplémentaires de développements.

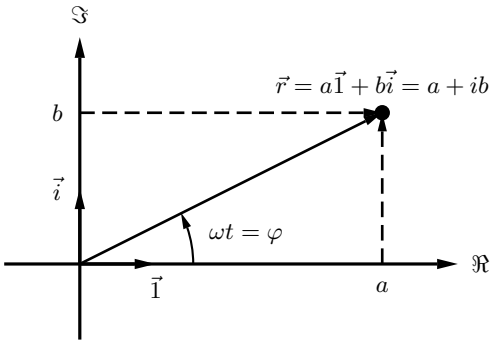


FIGURE 2.3 – Vecteur tournant

REMARQUE Une variation sinusoïdale $f(t) = r \sin \omega t$ peut être représentée comme la projection (cf. chapitre de Trigonométrie) sur l'axe vertical y (axe des imaginaires de l'ensemble $\mathbb{C}$) d'un vecteur $\vec{r}$ tournant à vitesse angulaire ω autour de l'origine dans le plan xOy comme illustré sur la figure 2.3. Un tel vecteur tournant s'appelle "vecteur de FRESNEL" et peut très bien être interprété comme la partie imaginaire du nombre complexe

$$\vec{r} = \Re(\vec{r}) + \Im(\vec{r}) = r(\cos \omega t + i \sin \omega t) == re^{i\omega t} \tag{2.98}$$

Nous retrouverons les vecteurs tournants de façon explicite lors de notre étude de la mécanique ondulatoire et optique géométrique (dans le cadre de la diffraction).

Transformations dans le plan

Il est habituel de représenter les nombres réels comme points d'une droite graduée. Les opérations algébriques y ont leur interprétation géométrique : l'addition est une translation, la multiplication une homothétie centrée à l'origine.

En particulier nous pouvons parler de la *racine carrée d'une transformation*. Une translation d'amplitude a peut être obtenue comme l'itération d'une translation d'amplitude $a/2$. De même, une homothétie de rapport a peut être obtenue comme l'itérée d'une homothétie de rapport $\sqrt{a}$. En particulier une homothétie de rapport 9 est la composée de deux homothéties de rapport 3 (ou -3).

La racine carrée prend alors un sens géométrique. Mais qu'en est-il de la racine carrée de nombres négatifs ? En particulier la racine carrée de -1 ?

Une homothétie de rapport -1 peut être vue comme une symétrie par rapport à l'origine ; toutefois si nous voulons voir cette transformation d'une manière continue, force nous est de placer la droite dans un plan. Dès lors une homothétie de rapport -1 peut être vue comme une rotation de π radians autour de l'origine.

Du coup, le problème de la racine carrée se simplifie. En effet il n'est guère difficile de décomposer une rotation de π radians en deux transformations : nous pouvons répéter soit une rotation de $\pi/2$ soit une rotation de $-\pi/2$. L'image de 1 sera la racine carrée de -1 et i est située sur une perpendiculaire à l'origine à une distance 1 soit vers le haut soit vers le bas.

Ayant réussi à positionner le nombre i il n'est plus guère difficile de disposer les autres nombres complexes dans un plan de GAUSS. Nous pouvons ainsi associer à $2i$ le produit de l'homothétie (cf. chapitre de Géométrie Euclidienne) de rapport 2 par la rotation de centre O et d'angle $\pi/2$, soit une similitude centrée à l'origine. C'est ce que nous allons nous efforcer à montrer maintenant.

Soient $k \in \mathbb{N}$, $z_1 = x_1 + iy_1 = ae^{i\alpha}$ et $z_2 = x_2 + iy_2 = be^{i\beta}$: nous avons les propriétés de transformations géométriques suivantes pour les nombres complexes (voir le chapitre de Trigonométrie pour les propriétés du sinus et cosinus) :

Propriété 2.13 *La multiplication de z_1 par un réel λ dans le plan de GAUSS correspond à une homothétie (agrandissement) de centre O (l'intersection des axes imaginaires et réels), de rapport λ.*

Preuve $\lambda z_1 = (\lambda a)e^{i\alpha}$ •

Propriété 2.14 *La multiplication de z_1 par un nombre complexe de module unitaire :*

$$z_0 = x_0 + iy_0 = \cos\omega + i\sin\omega = e^{i\omega} \tag{2.99}$$

correspond à une rotation de centre O et d'angle du complexe equation.

Preuve $z_1 z_0 = e^{i\omega} a e^{i\alpha} = a e^{i(\alpha+\omega)}$ •

Remarque Nous voyons alors immédiatement, par exemple, que multiplier un nombre complexe par i correspond à une rotation de $\pi/2$.

Il est intéressant d'observer que sous forme vectorielle la rotation de centre O de z_1 par z_0 peut s'écrire à l'aide de la matrice suivante :

$$\begin{bmatrix} x_0 & -y_0 \\ y_0 & x_0 \end{bmatrix} \tag{2.100}$$

Preuve Nous savons que $z_0 z_1$ est une rotation de centre O et d'angle ω. Il suffit de l'écrire comme suit : $z_0 z_1 = (x_0 + iy_0)(x_1 + iy_1) = (x_0x_1 - y_0y_1) + i(x_0y_1 + y_0y_1)$ ce qui donne sous forme vectorielle :

$$z_0 z_1 = \begin{pmatrix} x_0x_1 - y_0y_1 \\ 0 \end{pmatrix} + i \begin{pmatrix} 0 \\ x_0y_1 + y_0y_1 \end{pmatrix} \tag{2.101}$$

donc l'application linéaire est :

$$\begin{bmatrix} x_0 & -y_0 \\ y_0 & x_0 \end{bmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0x_1 - y_0y_1 \\ x_0y_1 + y_0y_1 \end{pmatrix} \tag{2.102}$$

ou encore en utilisant $z = r(\cos\varphi + i\sin\varphi) = r(\cos(\varphi + 2k\pi) + i\sin(\varphi + 2k\pi))$. Dans le cas particulier et arbitraire où r serait unitaire nous avons immédiatement :

$$\begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0x_1 - y_0y_1 \\ x_0y_1 + y_0y_1 \end{pmatrix} \tag{2.103}$$

Remarquons que la matrice de rotation peut aussi s'écrire sous la forme :

$$\begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix} = \cos\omega \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sin\omega \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \cos\omega \cdot I + \sin\omega \cdot J \tag{2.104}$$

de même :

$$\begin{bmatrix} x_0 & -y_0 \\ y_0 & x_0 \end{bmatrix} = x_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + y_0 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = x_0 \cdot I + y_0 \cdot J \tag{2.105}$$

•

Nous remarquons que ces matrices de rotation ne sont pas que des applications mais sont des nombres complexes aussi. Ainsi, nous avons pour habitude de poser en analyse complexe que :

$$1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{et} \quad i = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{2.106}$$

Le corps des nombres complexes est donc isomorphe au corps des matrices réelles carrées de dimension 2. C'est un résultat que nous réutiliserons de nombreuses fois dans divers chapitres pour des études particulières en algèbre, géométrie et en physique quantique relativiste.

Propriété 2.15 *La multiplication de deux complexes correspond à une homothétie ajoutée d'une rotation. En d'autres termes, d'une* similitude directe.

Preuve

$$z_1 z_2 = a\mathrm{e}^{i\alpha} b\mathrm{e}^{i\beta} = (ab)\mathrm{e}^{i(\alpha+\beta)} \tag{2.107}$$

il s'agit donc bien d'une similitude de rapport a et d'angle β. •

Au contrait, l'opération suivante :

$$z_1\bar{z}_2 = a\mathrm{e}^{i\alpha}b\mathrm{e}^{-i\beta} = (ab)\mathrm{e}^{i(\alpha-\beta)} \tag{2.108}$$

sera appelée une *similitude linéaire rétrograde*. Par ailleurs, il en retourne trivialement que :

$$\arg(z_1z_2) = \arg z_1 + \arg z_2 \tag{2.109}$$

REMARQUES

1. La somme de deux nombres $z_1 + z_2$ complexes ne pouvant avoir une écriture mathématique simplifiée sous quelque forme que ce soit, nous disons alors que la somme équivaut à une *translation d'amplitude*.
2. La combinaison d'une similitude linéaire (multiplication de deux nombres complexes) directe et d'une translation d'amplitude (sommation par un troisième nombre complexe) correspond à ce que nous appelons une *similitude linéaire directe*.

Propriété 2.16 *Le conjugué d'un nombre complexe est géométriquement son symétrique par rapport à l'axe* $\mathbb{R}$ *tel que :*

$$\bar{z}_1 = x_1 - iy_1 = r(\cos\alpha - i\sin\alpha) = (\cos(-\alpha) - i\sin(-\alpha)) = r\mathrm{e}^{-i\alpha} \tag{2.110}$$

sans oublier que $\cos\alpha = \cos(\alpha + 2k\pi)$ *et* $\sin\alpha = \sin(\alpha + 2k\pi)$*, ce qui nous donne :*

$$\arg(\bar{z}_1) = -\arg(z_1) \tag{2.111}$$

D'où nous pouvons tirer la propriété suivante :

$$\begin{aligned} &r(\cos(\varphi+\pi) + i\sin(\varphi+\pi)) = r(-\cos\varphi - i\sin\varphi) = -r(\cos\varphi + i\sin\varphi) = -z_1 \\ &\Rightarrow \arg z_1 + \pi = -\arg(-z_1) \end{aligned} \tag{2.112}$$

Propriété 2.17 *La négation du conjugué d'un nombre complexe est géométriquement son symétrique par rapport à l'axe des imaginaires* $\mathbb{R}_i$ *tel que :*

$$-\bar{z}_1 = -x_1 + iy_1 = r(-\cos\alpha + i\sin\alpha) = r(\cos(\pi \pm \alpha) + i\sin\alpha) \tag{2.113}$$

REMARQUES

1. La combinaison des propositions 2.16 et 2.17 est appelée une *similitude rétrograde*.
2. L'opération géométrique qui consiste à prendre l'inverse du conjugué d'un nombre complexe, soit $\bar{z}^{-1}$, est appelé une *inversion de pôle*.

Propriété 2.18 *La rotation de centre* c *et d'angle* φ *est donnée par :*

$$\mathcal{R}(z_1) = c + \mathrm{e}^{i\varphi}(z_1 - c) \tag{2.114}$$

En effet, le complexe c donne un point dans le plan de GAUSS. La différence $z_1 - c$, choisie pour que la rotation se fasse dans le sens des aiguilles d'une montre lorsque $\varphi \geqslant 0$ (sinon on prend $c - z_1$), donne le rayon du cercle. La multiplication par $\mathrm{e}^{i\varphi}$, la rotation du rayon par rapport à l'origine du plan de GAUSS. Finalement, l'addition par c la translation nécessaire pour ramener le rayon r tourné à l'origine du centre c.

Propriété 2.19 *Une homothétie de centre* c*, de rapport* λ *est obtenue par l'opération :*

$$\mathcal{H}(z_1) = c + \lambda(z_1 - c) \tag{2.115}$$

La différence $z_1 - c$ donne toujours le rayon r et c un point dans le centre de GAUSS. $\lambda(z_1 - c)$ donne l'homothétie du rayon par rapport à l'origine du plan de GAUSS et finalement l'addition par c la translation nécessaire pour que l'homothétie soit vue comme étant faite de centre c.

2.2.7 Nombres quaternions

Appelés aussi *hypercomplexes*, ces nombres ont été inventés en 1843 par William Rowan Hamilton pour généraliser les nombres complexes.

Définition 2.12 *Un quaternion est un élément* $(a,b,c,d) \in \mathbb{R}^4$ *dont nous notons* $\mathbb{H}$ *l'ensemble qui le contient et que nous appelons* ensemble des quaternions.

Un *quaternion* peut aussi bien être représenté en ligne ou en colonne tel que :

$$(a,b,c,d) = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \tag{2.116}$$

Nous définissons la somme de deux quaternions (a,b,c,d) et (a',b',c',d') par :

$$(a,b,c,d) + (a',b',c',d') = (a+a', b+b', c+c', d+d') \tag{2.117}$$

Il est évident que $(\mathbb{H},+)$ est un groupe commutatif•, d'élément neutre $(0,0,0,0)$, l'opposé d'un élément (a,b,c,d) étant $(-a,-b,-c,-d)$ [4]. L'associativité se vérifie en appliquant les propriétés correspondantes des opérations sur $\mathbb{R}$. Chap4

$$(a,b,c,d)\cdot(a',b',c',d') = \begin{pmatrix} aa'-bb'-cc'-dd' \\ (ab'+ba')+(cd'-dc') \\ (ac'+ca')-(bd'-db') \\ (da'+ad')+(bc'-cb') \end{pmatrix} \tag{2.118}$$

C'est peut-être difficile à accepter mais nous verrons un peu plus loin qu'il y a un air de famille avec les nombres complexes.

Nous pouvons remarquer que la loi de multiplication n'est pas commutative [5]. Effectivement, en prenant la définition de la multiplication (2.118), nous avons :

$$\begin{aligned} (0,1,0,0)\cdot(0,0,1,0) &= (0,0,0,1) \\ (0,0,1,0)\cdot(0,1,0,0) &= (0,0,0,-1) \end{aligned} \tag{2.119}$$

La multiplication a pour élément neutre $(1,0,0,0)$ puisqu'effectivement $(1,0,0,0)\cdot(a,b,c,d) = (a,b,c,d)\cdot(1,0,0,0) = (a,b,c,d)$. De plus, tout élément $(a,b,c,d) \in \mathbb{H}^* = \mathbb{H} - \{(0,0,0,0)\}$ est inversible. En effet, si (a,b,c,d) est un quaternion non nul, nous avons alors nécessairement $a^2+b^2+c^2+d^2 \neq 0$, sinon les quatre nombres a, b, c, d sont de carré nul, donc tous nuls. Soit alors le quaternion (a_1,b_1,c_1,d_1) défini par :

$$\begin{aligned} a_1 &= \frac{a}{a^2+b^2+c^2+d^2} \\ b_1 &= \frac{-b}{a^2+b^2+c^2+d^2} \\ c_1 &= \frac{-c}{a^2+b^2+c^2+d^2} \\ d_1 &= \frac{-d}{a^2+b^2+c^2+d^2} \end{aligned} \tag{2.120}$$

alors en appliquant machinalement la définition de la multiplication des quaternions, nous vérifions que :

$$(a,b,c,d)\cdot(a_1,b_1,c_1,d_1) = (a_1,b_1,c_1,d_1)\cdot(a,b,c,d) = (1,0,0,0) \tag{2.121}$$

ce dernier quaternion est donc l'inverse pour la multiplication.

Montrons maintenant que le corps des complexes $(\mathbb{C},+,\times)$ est un sous-corps de $(\mathbb{H},+,\times)$. Soit $\mathbb{H}'$ l'ensemble des quaternions de la forme $(a,b,0,0)$. $\mathbb{H}'$ est non vide et si $(a,b,0,0)$, $(a',b',0,0)$ sont des éléments de $\mathbb{H}'$. Effectivement :

4. C'est l'addition naturelle dans $\mathbb{R}^4$ vu comme $\mathbb{R}$-espace vectoriel•. Chap4

5. La loi de multiplication est distributive avec la loi d'addition mais c'est un excellent exemple où il faut quand même prendre garde à démontrer la distributivité à gauche et à droite, puisque le produit n'est pas commutatif !

1. Pour la soustraction :

$$(a,b,0,0)-(a',b',0,0)=(a-a',b-b',0,0)\in\mathbb{H}' \tag{2.122}$$

2. La multiplication :

$$(a,b,0,0)\cdot(a',b',0,0)=(aa'-bb',ab'+ba',0,0)\in\mathbb{H}' \tag{2.123}$$

3. L'élément neutre :

$$(1,0,0,0)\in\mathbb{H}' \tag{2.124}$$

4. Et finalement l'inverse de $(a,b,0,0)$:

$$(a/(a^2+b^2),-b/(a^2+b^2),0,0)\in\mathbb{H}' \tag{2.125}$$

Donc $(\mathbb{H},+,\times)$ est un sous-corps de $\mathbb{H}$. Soit alors l'application :

$$\begin{aligned} f : a+ib &\to (a,b,0,0)\\ \mathbb{C} &\to \mathbb{H} \end{aligned} \tag{2.126}$$

f est bijective, et nous vérifions aisément que pour tous complexes z_1, z_2, nous avons :

$$\begin{aligned} f(z_1+z_2)&=f(z_1)+f(z_2)\\ f(z_1z_2)&=f(z_1)f(z_2) \end{aligned} \tag{2.127}$$

Donc f est un isomorphisme de $(\mathbb{C},+,\times)$ sur $(\mathbb{H}',+,\times)$. Cet isomorphisme a pour intérêt d'identifier $\mathbb{C}$ à $\mathbb{H}'$ et d'écrire $\mathbb{C}\subset\mathbb{H}$, les lois d'addition et de soustraction sur $\mathbb{H}$ prolongeant les opérations déjà connues sur $\mathbb{C}$.

Ainsi, par convention, nous écrirons tout élément de $(a,b,0,0)$ de $\mathbb{H}$ sous la forme complexe $a+ib$. En particulier 0 est l'élément $(0,0,0,0)$, 1 l'élément $(1,0,0,0)$ et i l'élément $(0,1,0,0)$. Nous notons par analogie et extension j l'élément $(0,0,1,0)$ et k l'élément $(0,0,0,1)$. La famille $\{1,i,j,k\}$ forme une base de l'ensemble des quaternions vu comme un espace vectoriel sur $\mathbb{R}$, et nous écrirons ainsi $(a+bi+cj+dk)$ le quaternion (a,b,c,d).

La notation des quaternions sous forme définie avant est parfaitement adaptée à l'opération de multiplication. Pour le produit de deux quaternions nous obtenons en développant l'expression $(a+bi+cj+dk)\cdot(a'+b'i+c'j+d'k)$, seize termes que nous devons identifier à la définition d'origine de la multiplication des quaternions pour obtenir les relations suivantes :

$$\begin{aligned} i\cdot j&=k=-j\cdot i\\ j\cdot k&=i=-k\cdot j\\ k\cdot i&=j=-i\cdot k\\ i^2&=j^2=k^2=-1 \end{aligned} \tag{2.128}$$

Ce qui peut se résumer par :

·	1	i	j	k
1	1	i	j	k
i	i	-1	k	-j
j	j	-k	-1	i
k	k	j	-i	-1

(2.129)

Nous pouvons constater que l'expression de la multiplication (2.118) de deux quaternions ressemble en partie beaucoup à un produit vectoriel et scalaire, respectivement notés $\times$ et $\circ$ dans ce document. Pour illustrer, aidons-nous d'un exemple concret.

Exemple 2.7 *Soient deux quaternions sans partie réelle $p=xi+yj+zk$, $q=x'i+y'j+z'k$ et $\vec{u}$, $\vec{v}$ les vecteurs de $\mathbb{R}^3$ de coordonnées respectives (x,y,z) et (x',y',z'). Alors le produit $p\cdot q=(0,\vec{u})\cdot(0,\vec{v})$ est :*

$$p\cdot q=(-xx'-yy'-zz',yz'-zy',-xz'+zx',xy'-yx')=(-\vec{u}\circ\vec{v},\vec{u}\times\vec{v}) \tag{2.130}$$

Nous pouvons aussi par curiosité nous intéresser au cas général. Soient pour cela deux quaternions $p = (a, \vec{u})$ *et* $q = (b, \vec{v})$*. Nous avons alors :*

$$\begin{aligned} p \cdot q &= (a + (0, \vec{u})) \cdot (b + (0, \vec{v})) \\ &= ab + (0, a\vec{v}) + (0, b\vec{u}) + (-\vec{u} \circ \vec{v}, \vec{u} \times \vec{v}) \\ &= (ab - \vec{u} \circ \vec{v}, a\vec{v} + b\vec{u} + \vec{u} \times \vec{v}) \end{aligned} \tag{2.131}$$

Définition 2.13 *Le centre du corps non-commutatif* $(\mathbb{H}, +, \times)$ *est l'ensemble des éléments de* $\mathbb{H}$ *commutant pour la loi de multiplication avec tous les éléments de* $\mathbb{H}$*.*

Nous allons montrer que le centre de $(\mathbb{H}, +, \times)$ est l'ensemble des réels.

Soit $\mathbb{H}_1$ le centre de $(\mathbb{H}, +, \times)$, et (x, y, z, t) un quaternion. Nous devons avoir les conditions suivantes qui soient satisfaites :

Soit $(x, y, z, t) \in \mathbb{H}_1$ alors pour tout $(a, b, c, d) \in \mathbb{H}$ nous cherchons :

$$(x, y, z, t) \cdot (a, b, c, d) = (a, b, c, d) \cdot (x, y, z, t) \tag{2.132}$$

ce qui donne en développant :

$$\begin{aligned} xa - yb - zc - td &= ax - by - cz - dt \\ xb + ya + zd - tc &= ay + bx + ct - dz \\ xc + za - yd + tb &= az + cx - bt + dy \\ ta + xd + yc - zb &= dx + at + bz - cy \end{aligned} \tag{2.133}$$

La résolution de ce système, nous donne :

$$\begin{aligned} ct - dz &= 0 \\ bt - dy &= 0 \\ bz - cy &= 0 \end{aligned} \tag{2.134}$$

Donc pour que le quaternion (x, y, z, t) soit le centre de $\mathbb{H}$, il doit être réel.

Au même titre que pour les nombres complexes, nous pouvons définir un conjugué des quaternions :

Définition 2.14 *Le conjugué d'un quaternion* $Z = (a, b, c, d)$ *est le quaternion* $\bar{Z} = (a, -b, -c, -d)$

Au même titre que pour les complexes, nous remarquons que :

1. si $Z = \bar{Z}$ alors cela signifie que $Z \in \mathbb{R}$.
2. $Z + \bar{Z} \in \mathbb{R}$
3. par développement du produit $Z \cdot \bar{Z}$, nous avons :

$$\begin{aligned} Z \cdot \bar{Z} &= (a, b, c, d) \cdot (a, -b, -c, -d) \\ &= (a^2 + b^2 + c^2 + d^2, -ab + ba - cd + dc, -ac + ca + bd - db, da - ad - bc + cb) \\ &= (a^2 + b^2 + c^2 + d^2, 0, 0, 0) = a^2 + b^2 + c^2 + d^2 \in \mathbb{R} \end{aligned} \tag{2.135}$$

que nous adopterons, par analogie avec les nombres complexes, comme une définition de la norme (ou module) des quaternions tel que :

$$|Z| = \sqrt{Z \cdot \bar{Z}} \tag{2.136}$$

Dès lors, nous avons $|ZZ'| = \sqrt{(ZZ') \cdot (\overline{ZZ'})}$.

Comme pour les nombres complexes, il est aisé de montrer que la conjugaison est un automorphisme du groupe $(\mathbb{H}, +)$. C'est-à-dire que :

Propriété 2.20 *Soient* $Z = (a, b, c, d)$ *et* $Z' = (a', b', c', d')$ *alors :*

$$\begin{aligned} \overline{Z + Z'} &= (a + a', -b - b', -c - c', -d - d') \\ &= (a, -b, -c, -d) + (a', -b', -c', -d') \\ &= \bar{Z} + \bar{Z'} \end{aligned} \tag{2.137}$$

Ainsi qu'elle est involutive :

Propriété 2.21 *Soient Z, Z' alors :*

$$\bar{\bar{Z}} = (a, -(-b), -(-c), -(-d)) = (a, b, c, d) = Z \tag{2.138}$$

La conjugaison n'est par contre pas un automorphisme multiplicatif du corps equation. En effet, si nous considérons la multiplication de Z, Z' et en prenons le conjugué :

$$ZZ' = \begin{pmatrix} aa' - bb' - cc' - dd' \\ (ab' + ba') + (cd' - dc') \\ (ac' + ca') - (bd' - db') \\ (da' + ad') + (bc' - cb') \end{pmatrix}, \quad \overline{ZZ'} = \begin{pmatrix} aa' - bb' - cc' - dd' \\ -(ab' + ba') - (cd' - dc') \\ -(ac' + ca') + (bd' - db') \\ -(da' + ad') - (bc' - cb') \end{pmatrix} \tag{2.139}$$

nous voyons immédiatement que :

$$ZZ' \neq \overline{ZZ'} \tag{2.140}$$

Revenons maintenant au module. Pour cela, calculons le carré de la norme de $|ZZ'|$:

$$|ZZ'|^2 = (ZZ') \cdot (\overline{ZZ'}) \tag{2.141}$$

Or, en reprenant l'équation (2.118), définissons ce produit comme suit :

$$Z \cdot Z' = (\alpha, \beta, \gamma, \lambda) = Z" \tag{2.142}$$

Nous avons alors :

$$Z" \cdot \overline{Z"} = \alpha^2 + \beta^2 + \gamma^2 + \lambda^2 \tag{2.143}$$

En substituant, il vient :

$$\begin{aligned} (ZZ') \cdot \overline{(ZZ')} = (aa' - bb' - cc' - dd')^2 + (ab' + ba' + cd' - dc')^2 \\ + (ac' + ca' - bd' + db')^2 + (da' + ad' + bc' - cb')^2 \end{aligned} \tag{2.144}$$

après un développement algébrique élémentaire (honnêtement ennuyeux), nous trouvons :

$$(ZZ') \cdot \overline{(ZZ')} = (a'^2 + b'^2 + c'^2 + d'^2)(a^2 + b^2 + c^2 + d^2) = |Z'|^2 |Z|^2 \tag{2.145}$$

Donc :

$$|ZZ'|^2 = |\bar{Z}|^2 |Z|^2 \tag{2.146}$$

REMARQUE La norme est donc un homomorphisme de $(\mathbb{H}, \times)$ dans $(\mathbb{R}, \times)$. Par la suite, nous noterons G l'ensemble des quaternions de norme 1.

Interprétation matricielle

Soit q un quaternion donné, soit l'application $p \to q \cdot p$. La multiplication (à gauche) peut être faite avec une application linéaire (cf. chapitre d'Algèbre Linéaire) sur $\mathbb{H}$.

Si $q = a + bi + cj + dk$, cette application a pour matrice, dans la base $1, i, j, k$:

$$\begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix} \tag{2.147}$$

Ce que nous vérifions bien :

$$ZZ' = \begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix} \begin{pmatrix} a' \\ b' \\ c' \\ d' \end{pmatrix} = \begin{pmatrix} aa' - bb' - cc' - dd' \\ (ab' + ba') + (cd' - dc') \\ (ac' + ca') - (bd' - db') \\ (da' + ad') + (bc' - cb') \end{pmatrix} \tag{2.148}$$

En fait, nous pouvons alors définir les quaternions comme l'ensemble des matrices ayant la structure visible de l"équation (**??**) si nous le voulions. Cela les réduirait alors à un sous espace vectoriel de $M_4(\mathbb{R})$.

En particulier, la matrice de 1 (la partie réelle du quaternion) n'est alors rien d'autre que la matrice de l'identité :

$$M_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbb{1} \tag{2.149}$$

de même :

$$M_i = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad M_j = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \quad M_k = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{2.150}$$

Rotations

Nous allons maintenant voir que la conjugaison par un élément du groupe G (quaternion de norme 1) peut s'interpréter comme une rotation dans l'espace.

Définition 2.15 *La* conjugaison *par un quaternion q non nul et de norme unité est l'application S_q définie sur $\mathbb{H}$ par :*

$$S_q : p \to q \cdot p \cdot q^{-1} = q \cdot p \cdot \bar{q} \tag{2.151}$$

et nous affirmons que cette application est une rotation.

Remarques

1. Comme q est de norme 1, nous avons bien évidemment $|q| = q \cdot \bar{q} = 1$ donc $q^{-1} = \bar{q}$. Ce quaternion peut être vu comme la valeur propre (unitaire) de l'application (matricielle) p sur le vecteur $\vec{q}$ (on se retrouve avec un concept en tout point similaire aux matrices orthogonales de rotation vues en algèbre linéaire).
2. S_q est une application linéaire (donc si c'est bien une rotation, la rotation peut être décomposée en plusieurs rotations). Effectivement, prenons deux quaternions p_1, p_2 et $\lambda_1, \lambda 2$ des réels, alors nous avons :

$$S_q(\lambda_1 p_1 + \lambda_2 p_2) = q(\lambda_1 p_1 + \lambda_2 p_2)\bar{q} = \lambda_1 q p_1 \bar{q} + \lambda_2 q p_2 \bar{q} = \lambda_1 S_q(p_1) + \lambda_2 S_q(p_2) \tag{2.152}$$

Vérifions maintenant que l'application est bien une rotation. Comme nous l'avons vu lors de notre étude de l'algèbre linéaire et en particulier les matrices orthogonales (cf. chapitre d'Algèbre Linéaire), une condition est que l'application conserve la norme.

Vérifions :

$$|S_q(p)| = |qp\bar{q}| = |q||p||\bar{q}| = |p| \tag{2.153}$$

Par ailleurs, nous pouvons vérifier qu'une rotation d'un quaternion purement complexe (tel qu'alors nous nous restreignons à $\mathbb{R}^3$) et la même rotation inverse sommées est nul (le vecteur sommé à son opposé s'annulent) :

$$S_q(p) + \overline{S_q(p)} = qp\bar{q} + \overline{qp\bar{q}} = qp\bar{q} + \overline{q(p\bar{q})} \tag{2.154}$$

nous vérifions trivialement que si nous avons deux quaternions q, p alors $\overline{p \cdot q} = \bar{q} \cdot \bar{p}$ dès lors :

$$\begin{aligned} S_q(p) + \overline{S_q(p)} &= qp\bar{q} + \overline{q(p\bar{q})} = qp\bar{q} + \overline{(p\bar{q})}\bar{q} \\ &= q \cdot p \cdot \bar{q} + \bar{\bar{q}} \cdot \bar{p} \cdot \bar{q} = q \cdot p \cdot \bar{q} + q \cdot \bar{p} \cdot \bar{q} \\ &= S_q(p + \bar{p}) \end{aligned} \tag{2.155}$$

pour que cette opération soit nulle, nous voyons immédiatement que nous devons restreindre p aux quaternions purement complexes. Dès lors :

$$S_q(p + \bar{p}) = S_q(0) = 0 \tag{2.156}$$

Nous en déduisons alors que p doit être purement complexe pour que l'application S_q soit une rotation et que $S_q(p)$ est un quaternion pur. En d'autres termes : un quaternion pur par cette application reste un quaternion pur. S_q restreint à l'ensemble des quaternions est donc une isométrie vectorielle, c'est-à-dire une symétrie ou une rotation.

Nous avons vu également lors de notre étude des matrices de rotation dans le chapitre d'Algèbre Linéaire que l'application A devait être de déterminant 1 pour que nous ayons une rotation. Voyons si c'est le cas de S_q. Pour cela, nous calculons explicitement en fonction de $q = a + bi + cj + dk$, la matrice (dans la base canonique $\{i, j, k\}$) de S_q et nous en calculons le déterminant. Ainsi, nous obtenons les coefficients des colonnes de A en se rappelant de (2.128) et en calculant :

$$\begin{aligned} S_q(i) &= (a+bi+cj+dk)i(a-bi-cj-dk) \\ &= (ai+bi^2+cji+dki)(a-bi-cj-dk) \\ &= (ai-b-ck+dj)(a-bi-cj-dk) \\ &= \dots \\ &= (a^2+b^2-c^2-d^2)i + 2(ad+bc)j + 2(bd-ac)k \end{aligned} \tag{2.157a}$$

$$\begin{aligned} S_q(j) &= (a+bi+cj+dk)j(a-bi-cj-dk) \\ &= (aj+bij+cj^2+dkj)(a-bi-cj-dk) \\ &= (aj+bk-c-di)(a-bi-cj-dk) \\ &= \dots \\ &= 2(bc-ad)i + (a^2-b^2+c^2-d^2)j + 2(ab+cd)k \end{aligned} \tag{2.157b}$$

$$\begin{aligned} S_q(k) &= (a+bi+cj+dk)k(a-bi-cj-dk) \\ &= (ak+bik+cjk+dk^2)(a-bi-cj-dk) \\ &= (ak-bj+ci-d)(a-bi-cj-dk) \\ &= \dots \\ &= 2(ac+bd)i + 2(cd-ab)j + (a^2-b^2-c^2+d^2)k \end{aligned} \tag{2.157c}$$

Il faut alors calculer le déterminant de la matrice :

$$\begin{bmatrix} a^2+b^2-c^2-d^2 & 2(ad+bc) & 2(bd-ac) \\ 2(bc-ad) & a^2-b^2+c^2-d^2 & 2(ab+cd) \\ 2(ac+bd) & 2(cd-ab) & (a^2-b^2-c^2+d^2) \end{bmatrix} \tag{2.158}$$

en se souvenant que $a^2+b^2+c^2+d^2 = 1$ et nous trouvons que le déterminant vaut bien 1. Montrons maintenant que cette rotation est un demi-tour d'axe. D'abord, si $q = a + bi + cj + dk$ nous avons $S_q(q) = q \cdot q \cdot \bar{q} = q$, ce qui signifie que l'axe de rotation (x, y, z) est fixé par l'application S_q elle-même. D'autre part, nous avons vu que si q est un quaternion purement complexe de norme 1 alors $q^{-1} = \bar{q}$ et aussi $\bar{q} = -q$. Ce qui nous donne la relation :

$$q^2 = q \cdot (-\bar{q}) = q \cdot (-(q^{-1})) = -1 \tag{2.159}$$

Ce résultat nous amène à calculer la rotation d'une rotation :

$$S_q(S_q(p)) = q(qp\bar{q})\bar{q} = q^2 p \bar{q}^2 = S_{q^2}(p) = (-1)p(-q)^2 = -pq^2 = (-1)p(-1) = p \tag{2.160}$$

Puisque la rotation d'une rotation est un tour complet, alors S_q est nécessairement un demi-tour $S_q(p) = -p$ par rapport à l'axe (x, y, z).

À ce stade, nous pouvons affirmer que toute rotation de l'espace peut se représenter par S_q (la conjugaison par un quaternion q de norme 1). En effet, les demi-tours engendrent le groupe des rotations, c'est-à-dire que toute rotation peut s'exprimer comme le produit d'un nombre fini de demi-tours, et donc comme la conjugaison par un produit de quaternions de norme 1, produit qui est lui-même un quaternion de norme 1.

Nous allons tout de même donner une forme explicite reliant une rotation et le quaternion qui la représente, au même titre que nous l'avons fait pour les nombres complexes.

Soit $\vec{u}(x,y,z)$ un vecteur unitaire et $\theta \in [0,2\pi]$ un angle. Alors nous affirmons que la rotation d'axe $\vec{u}$ et d'angle θ correspondant à l'application S_q, où q est le quaternion :

$$q = \cos\frac{\theta}{2} + x\sin\frac{\theta}{2}i + y\sin\frac{\theta}{2}j + z\sin\frac{\theta}{2}k = \left(\cos\frac{\theta}{2}, \sin\frac{\theta}{2}\vec{u}\right)\begin{pmatrix}\cos\frac{\theta}{2}\\ x\sin\frac{\theta}{2}\\ y\sin\frac{\theta}{2}\\ z\sin\frac{\theta}{2}\end{pmatrix} \tag{2.161}$$

Pour que cette affirmation soit vérifiée, nous savons qu'il faut que la norme de q soit unitaire, le déterminant de l'application S_q soit égal à l'unité, que l'application S_q conserve la norme, que l'application S_q renvoie tout vecteur colinéaire à l'axe de rotation sur l'axe de rotation :

1. La norme du quaternion proposé précédemment vaut effectivement 1 :

$$\begin{aligned}|q| &= \cos^2\frac{\theta}{2} + x^2\sin^2\frac{\theta}{2} + y^2\sin^2\frac{\theta}{2} + z^2\sin^2\frac{\theta}{2}\\ &= \cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2}(x^2+y^2+z^2)\\ &= \cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2} = 1\end{aligned} \tag{2.162}$$

2. Le fait que q soit un quaternion de norme 1 amène immédiatement à ce que le déterminant de l'application S_q soit unitaire. Nous l'avons déjà montré plus haut dans le cas général de n'importe quel quaternion de norme 1 (condition nécessaire et suffisante).
3. Il en est de même pour la conservation de la norme. Nous avons déjà montré plus haut que c'était de toute façon le cas dès que le quaternion q était de norme 1 (condition nécessaire et suffisante).
4. Voyons maintenant que tout vecteur colinéaire à l'axe de rotation est projeté sur l'axe de rotation. Notons q' le quaternion purement imaginaire et unitaire $xi + yj + zk$. Nous avons alors :

$$q = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}q' \tag{2.163}$$

Alors :

$$S_q(q') = qq'\bar{q} \tag{2.164}$$

mais comme q' est la restriction de q à ces éléments purs qui le constituent, cela revient à écrire :

$$S_q(q') = S_q(q) = qq\bar{q} = q \tag{2.165}$$

Montrons maintenant le choix de l'écriture $\theta/2$. Si $\vec{v}(x_1,y_1,z_1)$ désigne un vecteur unitaire orthogonal à l'axe de rotation $\vec{u}$ et p le quaternion $x_1i + y_1j + z_1k$ alors nous avons :

$$\begin{aligned}S_q(p) &= \left(\cos\frac{\theta}{2} + \sin\frac{\theta}{2}q'\right)p\left(\cos\frac{\theta}{2} - \sin\frac{\theta}{2}q'\right)\\ &\cos^2\frac{\theta}{2}p + \cos\frac{\theta}{2}\sin\frac{\theta}{2}(q'p - pq') - \sin^2\frac{\theta}{2}q'pq'\end{aligned} \tag{2.166}$$

Nous avons montré lors de la définition de la multiplication de deux quaternions que $pq' = -q'p$. Nous obtenons alors :

$$\begin{aligned}S_q(p) &= \cos^2\frac{\theta}{2}p + 2\cos\frac{\theta}{2}\sin\frac{\theta}{2}q'p - \sin^2\frac{\theta}{2}q'pq'\\ &= \cos^2\frac{\theta}{2}p + 2\cos\frac{\theta}{2}\sin\frac{\theta}{2}q'p + \sin^2\frac{\theta}{2}q'p(-q')\\ &= \cos^2\frac{\theta}{2}p + 2\cos\frac{\theta}{2}\sin\frac{\theta}{2}q'p + \sin^2\frac{\theta}{2}q'p\bar{q'}\end{aligned} \tag{2.167}$$

Nous avons également montré que :

$$q'p\bar{q'} = S_{q'}(p) = -p \tag{2.168}$$

(le demi-tour d'axe (x, y, z)). Donc :

$$\begin{aligned} S_q(p) &= \cos^2\frac{\theta}{2}p + 2\cos\frac{\theta}{2}\sin\frac{\theta}{2} - \sin^2\frac{\theta}{2}p \\ &= \left(\cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}\right)p + 2\cos\frac{\theta}{2}\sin\frac{\theta}{2}q'p \\ &= \cos\theta p + \sin\theta q'p \end{aligned} \tag{2.169}$$

Nous savons que p est le quaternion pur assimilé à un vecteur unitaire $\vec{v}$ orthogonal à l'axe de rotation $\vec{u}$, lui-même assimilé à la partie purement imaginaire de q'. Nous remarquons alors de suite que la partie imaginaire du produit des quaternions $q'p$ est alors égal au produit vectoriel $\vec{u} \times \vec{v} = \vec{w}$. Ce produit vectoriel engendre donc un vecteur perpendiculaire à $\vec{u}, \vec{v}$ et donc $(\vec{u}, \vec{v}, \vec{w})$.

Le couple $(\vec{v}, \vec{w})$ forme donc un plan perpendiculaire à l'axe de rotation $\vec{u}$[6] et finalement :

$$S_q(p) = \cos\theta p + \sin\theta q'p = \cos\theta\vec{v} + \sin\theta\vec{w} \tag{2.170}$$

Nous nous retrouvons avec une rotation dans le plan identique à celle présentée plus haut avec les nombres complexes normaux dans le plan de GAUSS.

Nous savons donc maintenant comment faire n'importe quel type de rotation dans l'espace en une seule opération mathématique et ce en plus par rapport à un libre choix de l'axe !

Nous pouvons aussi maintenant mieux comprendre pourquoi l'algèbre des quaternions n'est pas commutative. Effectivement, les rotations vectorielles du plan sont commutatives mais celles de l'espace ne le sont pas comme nous le montre l'exemple illustré des figures 2.4 à 2.6. Les

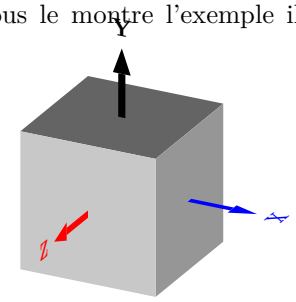
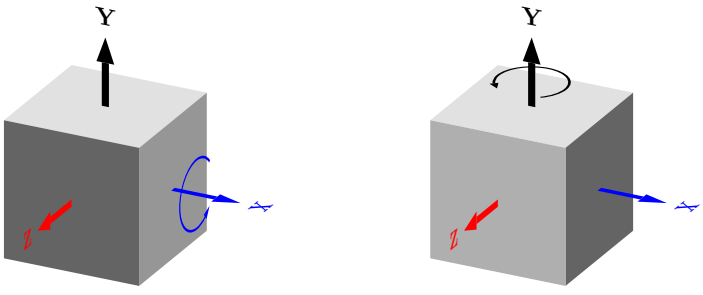


FIGURE 2.4 – configuration initiale

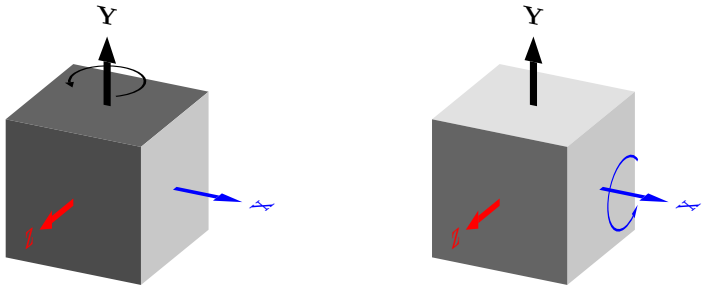


(a) rotation autour de X (b) rotation autour de Y

FIGURE 2.5 – Composition de rotations XY

résultats obtenus seront fondamentaux pour notre compréhension des spineurs (cf. chapitre de Calcul Spinoriel) !

6. C'est comme pour les nombres complexes simples dans lequel nous avons le plan de GAUSS et perpendiculairement à celui-ci un axe de rotation.



(a) rotation autour de Y (b) rotation autour de X

FIGURE 2.6 – Composition de rotations YX

2.2.8 Nombres algébriques et transcendants

Définition 2.16 *Nous appelons* nombre algébrique, *tout nombre qui est solution d'une équation algébrique, à savoir un polynôme (concept que nous aborderons dans la section d'Algèbre) dont les coefficients sont des entiers relatifs et non nuls.*

Un résultat intéressant est qu'un nombre rationnel est un nombre algébrique si et seulement si c'est un entier relatif [7].

Preuve Nous supposons que le nombre p/q, où p et q sont deux entiers premiers entre eux, est une racine du polynôme (cf. chapitre de Calcul Algébrique) :

$$x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 \tag{2.171}$$

où l'égalité avec zéro du polynôme est implicite. Dans ce cas :

$$p^n = -(a_{n-1}p^{n-1} + \ldots + a_1 pq^{n-2} + a_0 q^{n-1})q \tag{2.172}$$

Puisque les coefficients sont tous entiers et leurs multiples aussi dans la parenthèse, alors la parenthèse a une valeur dans $\mathbb{Z}$. Ainsi, q divise une puissance de p, ce qui n'est possible, dans l'ensemble $\mathbb{Z}$, que si q vaut ± 1 puisqu'ils étaient premiers entre eux. Réciproquement, tout entier relatif est évidemment un entier algébrique. •

Par extension, tout nombre rationnel est algébrique, car le quotient p/q de deux entiers est racine de l'équation :

$$qx - p = 0 \tag{2.173}$$

Les nombres qui ne sont pas algébriques sont transcendants. Les transcendants sont donc beaucoup plus nombreux que les algébriques.

Nous avons aussi le nombre réel et irrationnel $\sqrt{2}$ qui est algébrique, car il est racine de :

$$x^2 - 2 = 0 \tag{2.174}$$

et le nombre complexe i est algébrique, car il est racine de l'équation :

$$x^2 + 1 = 0 \tag{2.175}$$

REMARQUE Les transcendants les plus connus sont π et e. Les démonstrations de leur transcendance est en cours de rédaction. Nous devrions pouvoir vous les fournir fin 2010.

7. En termes savants, nous disons que l'anneau $\mathbb{Z}$ est *intégralement clos*.

2.2.9 Nombres abstraits

Le nombre peut être envisagé en faisant abstraction de la nature des objets qui constituent le groupement qu'il caractérise ; et ainsi qu'à la façon de codifier (chiffre arabe, romain, ou autre système universel) on dit alors que le nombre est *abstrait*[8].

Pour les mathématiciens, il n'est pas avantageux de travailler avec ces symboles car ils représentent uniquement des cas particuliers. Ce que cherchent les physiciens théoriciens ainsi que les mathématiciens, se sont des relations applicables universellement dans un cas général et que les ingénieurs puissent en fonction de leurs besoins changer ces nombres abstraits par les valeurs numériques qui correspondent au problème qu'ils ont besoin de résoudre.

Ces nombres abstraits appelés aujourd'hui communément *variables* ou *inconnues* sont très souvent représentés par :

1. L'alphabet latin[9] : a, b, c, d, e,... x, y, z ; A, B, C, D, E... (2.230)
2. L'alphabet grec[10] listé dans le tableau 2.3.
3. L'alphabet hébraïque (à moindre mesure)[11]

Bien que ces symboles puissent représenter n'importe quel nombre il en existe quelques uns qui peuvent représenter en physique des valeurs dites *constantes Universelles* comme la vitesse de la lumière c, la constante gravitationnelle G, la constante de Planck h etc. Nous utilisons très souvent

Aα	Alpha	$\Lambda\lambda$	Lambda
Bβ	Beta	Mμ	Mu
$\Gamma\gamma$	Gamma	Nν	Nu
$\Delta\delta$	Delta	$\Xi\xi$	Xi
E$\epsilon\varepsilon$	Epsilon	Oo	Omicron
Zζ	Zeta	$\Pi\pi$	Pi
Hη	Eta	Pρ	Rho
$\Theta\theta\vartheta$	Theta	$\Sigma\sigma$	Sigma
Iι	Iota	Tτ	Tau
Kκ	Kappa	$\Upsilon\upsilon$	Upsilon
$\Phi\phi\varphi$	Phi	Xχ	chi
$\Psi\psi$	Psi	$\Omega\omega$	Omega

TABLEAU 2.3 – Alphabet grec

encore d'autres symboles que nous introduirons et définirons au fur et à mesure[12].

Domaines de définition

Une variable est un nombre abstrait susceptible de prendre des valeurs numériques différentes. L'ensemble de ces valeurs peut varier suivant le caractère du problème considéré.

Définition 2.17 (domaine de définition) *Nous appelons* domaine de définition *d'une variable, l'ensemble des valeurs numériques qu'elle est susceptible de prendre entre deux valeurs finies ou infinies appelées* bornes.

Soit a et b deux nombres tel que $a < b$. Alors :

Définition 2.18 (intervalle fermé) *Nous appelons* intervalle fermé d'extrémité a et b, *l'ensemble de tous les nombres x compris entre ces deux valeurs comprises et nous le désignons de la façon suivante :*

$$[a, b] = \{x \in \mathbb{R} | a \leqslant x \leqslant b\} \tag{2.176}$$

8. Arbitrairement, l'être humain a adopté un système numérique majoritairement utilisé de par le monde et représenté par les symboles 1, 2, 3, 4, 5, 7, 8, 9 du système décimal et qui seront supposés connus aussi bien en écriture qu'oralement par le lecteur (apprentissage du langage).

9. Les lettres minuscules du début l'alphabet latin (a, b, c, d, e...) sont souvent utilisées pour représenter de manière abstraite des constantes, alors que les lettres minuscules de la fin de l'alphabet latin ($...x, y, z$) sont utilisées pour représenter des entités (variables ou inconnues) dont nous recherchons la valeur.

10. Cet alphabet est particulièrement utilisé pour représenter soit des opérateurs mathématiques plus ou moins complexes (comme la somme indexée Σ, le variationnel δ, l'élément infinitésimal ε, le différentiel partiel ∂, etc.) soit des variables dans le domaine de la physique (comme ω pour la pulsation, la fréquence ν, la densité ρ, etc.).

11. Comme nous l'avons vu, les nombres transfinis sont par exemples donnés par la lettre $\aleph_0$ *aleph*.

12. Les lettres pour représenter les nombres ont été employées pour la première fois par VIÈTE.

Définition 2.19 (intervalle ouvert) *Nous appelons* intervalle ouvert d'extrémité a et b, *l'ensemble de tous les nombres x compris entre ces deux valeurs non comprises et nous le désignons de la façon suivante :*

$$]a, b[= \{x \in \mathbb{R} | a < x < b\} \tag{2.177}$$

Définition 2.20 *Nous appelons* intervalle fermé à gauche, ouvert à droite *la relation suivante :*

$$[a, b[= \{x \in \mathbb{R} | a \leqslant x < b\} \tag{2.178}$$

Définition 2.21 *Nous appelons* intervalle ouvert à gauche, fermé à droite *la relation suivante :*

$$]a, b] = \{x \in \mathbb{R} | a < x \leqslant b\} (2.235) \tag{2.179}$$

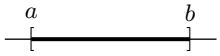
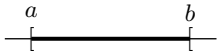
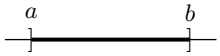
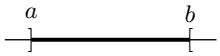
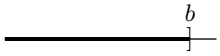
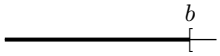
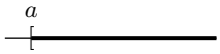
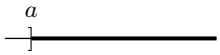
$[a; b]$		$a \leqslant x \leqslant b$	Intervalle fermé borné
$[a; b[$		$a \leqslant x < b$	Intervalle borné semi-fermé en a et semi-ouvert en b (semi-fermé à gauche et semi-ouvert à droite)
$]a; b]$		$a < x \leqslant b$	Intervalle borné semi-ouvert en a et semi-fermé en b (semi-ouvert à gauche et semi-fermé à droite)
$]a; b[$		$a < x < b$	Intervalle ouvert borné
$]-\infty; b]$		$x \leqslant b$	Intervalle non borné fermé en b (à droite)
$]-\infty; b[$		$x < b$	Intervalle non borné ouvert en b (à droite)
$[a; +\infty[$		$a \leqslant x$	Intervalle non borné fermé en a (à gauche)
$]a; +\infty[$		$a < x$	Intervalle non borné ouvert en a (à gauche)

TABLEAU 2.4 – Rappel des définitions

REMARQUES
- La notation $\{x$ tels que $a < x < b$ $\}$ désigne l'ensemble des réels x tels que $a < x < b$ (sous-entendu qui sont strictement plus grand que a et strictement inférieur à b).
- Le fait de dire qu'un intervalle est par exemple ouvert en b signifie que le réel b ne fait pas partie de celui-ci. Par contre, s'il y avait été fermé alors il en aurait fait partie.
- Si la variable peut prendre toutes les valeurs négatives et positives possibles nous écrivons dès lors : $]-\infty, +\infty[$ où le symbole ∞ signifie une *infinité*. Évidemment il peut y avoir des combinaisons d'intervalles ouverts et infinis à droite, fermé et limité gauche et réciproquement.
- Nous rappellerons ces concepts avec une autre approche lorsque nous étudierons l'algèbre (calcul littéral).

Nous disons que la variable x est *ordonnée* si en représentant son domaine de définition par un axe horizontal où chaque point de l'axe représente une valeur de x, alors que pour chaque couple de valeurs, nous pouvons indiquer celle qui est *antécédente* (qui précède) et celle qui est *conséquente* (qui suit). Ici la notion d'antécédente ou de conséquente n'est pas liée au temps, elle exprime juste la façon d'ordonner les valeurs de la variable.

Définition 2.22 (valeur croissante) *Une variable est dite* croissante *si chaque valeur conséquente est plus grande que chaque valeur antécédente.*

Définition 2.23 (valeur décroissante) *Une variable est dite* décroissante *si chaque valeur conséquente est plus petite que chaque valeur antécédente.*

Définition 2.24 (variables monotones) *Les variables croissantes et les variables décroissantes sont appelées* variables à variations monotones *ou simplement* variables monotones.

3

Opérateurs arithmétiques

PARLER des nombres comme nous l'avons fait dans le chapitre précédent amène naturellement à considérer les opérations de calculs. Il est donc logique que nous fassions une description non exhaustive des opérations qui peuvent exister entre les nombres. Dans ce cadre, nous considérerons qu'il existe deux types d'outils fondamentaux en arithmétique : Chap2

opérateurs arithmétiques – il existe deux opérateurs de base, addition et soustraction, à partir desquels nous pouvons construire d'autres opérateurs : la *multiplication* et la *division*. Ces quatre opérateurs sont couramment appelés *opérateurs rationnels*[1].

opérateurs (relations) binaires – il existe six relations binaires fondamentales –égal, différent de, plus grand que, plus petit que, plus grand ou égal, plus petit ou égal– qui permettent de comparer des grandeurs d'éléments se trouvant à gauche et à droite afin d'en tirer certaines conclusions.

Il est bien évidemment essentiel de connaître au mieux ces deux outils et leurs propriétés avant de se lancer dans des calculs plus ardus.

3.1 Relations binaires

Le concept de *relation* est la base de toute la mathématique dont le but est d'étudier —par observation et déduction (raisonnement), calcul et comparaison— des configurations ou relations abstraites ou concrètes de ses objets (nombres, formes, structures) en cherchant à établir les liens logiques, numériques ou conceptuels entre ces objets.

Définition 3.1 *Considérons deux ensembles non vides E et F non nécessairement identiques. Si à certains éléments x de E, nous pouvons associer par une règle mathématique précise R (non ambiguë), un élément y de F, nous définissons ainsi une* relation fonctionnelle *de E vers F qui s'écrit :* Chap4

$$R : E \to F \tag{3.1}$$

Ainsi, de façon plus générale, une relation fonctionnelle R peut être définie comme une règle mathématique qui associe à certains éléments x de E, certains éléments y de F.

Alors, dans ce contexte plus général, si xRy, nous disons que y est une *image* de x par R et que x est un *antécédent* ou *pré-image* de y.

L'ensemble des couples (x, y) tel que xRy soit une assertion vraie forme un *graphe* ou une *représentation* de la relation R. Nous pouvons représenter ces couples dans un repère adéquatement choisi pour en faire une représentation graphique de la relation R.

Il s'agit d'un type de relations sur lequel nous reviendrons dans le chapitre d'Analyse Fonctionnelle et qui ne nous intéresse pas directement dans ce chapitre.

Définition 3.2 *Considérons un ensemble A non vide, si nous associons à cet ensemble des outils permettant de comparer les éléments le composant alors nous parlons de* relation binaire *ou* relation de comparaison *et qui s'écrit pour tout élément x et y composant A :*

$$xRy \tag{3.2}$$

1. Rigoureusement l'addition suffirait si nous considérons l'ensemble commun des réels car dès lors la soustraction n'est que l'addition d'un nombre négatif.

Ces relations ne peuvent aussi être représentées sous forme graphique. Dans le cas des opérateurs binaires de comparaison classiques et où A est l'ensemble des nombres naturels, relatifs, rationnels ou réels, cette forme graphique est représentée par une droite horizontale,le plus souvent, dans le cas de la congruence elle est représentée par des droites dans le plan dont les points sont donnés par la contrainte de la congruence.

Chap5|p117

Comme nous l'avons déjà dit, il existe six relations binaires fondamentales : égal, différent de, plus grand que, plus petit que, plus grand ou égal, plus petit ou égal. Nous verrons un peu plus loin que la définition rigoureuse des relations binaires permet de construire des outils plus abstraits.

Chap5

3.1.1 Égalités

Il est fort difficile de définir la notion d'*égalité* dans un cas général applicable à toute situation. Pour notre part, nous nous permettrons pour cette définition de nous inspirer du théorème d'extensionalité de la théorie des ensembles (que nous verrons plus tard) :

Définition 3.3 *Deux éléments sont égaux si, et seulement si, ils ont les mêmes valeurs. L'égalité est décrite par le symbole* $=$ *qui signifie* égal à.

Propriété 3.1 *Si nous avons* $a = b$, c *un nombre et* $*$ *une opération quelconque*[2] *alors :*

$$a * c = b * c \tag{3.3}$$

Cette propriété est très utilisée pour résoudre ou simplifier des équations de type quelconque.

Définition 3.4 *Si deux éléments ne sont pas égaux (inégaux), nous les relions par le symbole* $\neq$ *et nous disons qu'ils sont* non égaux.

Il existe encore d'autres symboles d'égalités, qui sont une extension des deux que nous avons définis précédemment. Malheureusement, ils sont assez souvent mal utilisés dans la plupart des ouvrages disponibles sur le marché :

$$\approx \quad = \quad \cong \quad \equiv \quad \doteq \quad \propto \tag{3.4}$$

qui correspondent dans l'ordre à :
- presque égal (plutôt utilisé en ingénierie)
- asymptotiquement égal à (utilisé en analyse fonctionnelle)
- approximativement égal (utilisé en physique lors d'approximation de séries)
- identique à (utilisé aussi bien en analyse fonctionnelle qu'en physique)
- tend vers la limite (idem) et enfin proportionnel à (utilisé en physique ou en mathématiques financières).

3.1.2 Comparateurs

Les comparateurs sont des outils qui nous permettent de comparer et d'ordonner tout couple de nombres.

La possibilité d'ordonner des nombres est presque fondamentale en mathématique. Si une telle notion était inexistante, les concepts suivants ne pourraient être définis[3] :
- fonctions monotones et dérivation ;
- approche de zéros d'un polynôme par dichotomie (algorithme classique de recherche dans un ensemble ordonné partagé en deux à chaque itération) ;
- segments et demi-droites, demi-espace, convexité, orientation de l'espace, etc.

Il est donc important de pouvoir ordonner certains éléments méthématiques.

Ainsi, pour tout $a, b, c \in \mathbb{R}$, nous écrivons lorsque a est plus grand ou égal à b :

$$a \geqslant b \tag{3.5}$$

et lorsque a est plus petit ou égal à b :

$$a \leqslant b \tag{3.6}$$

2. Il peut s'agir par exemple de l'addition, la soustraction, la multiplication ou la division.
3. Certains des concepts n'ont pas encore été vus mais nous souhaitons quand même y faire référence.

Remarque Il est utile de rappeler que l'ensemble des réels $\mathbb{R}$ est un groupe totalement ordonné•, sans quoi nous ne pourrions pas définir des relations d'ordre entre ces éléments (ce qui n'est pas le cas des nombres complexes que nous ne pouvons pas ordonner !). Chap4

Définition 3.5 *Le symbole* $\leqslant$ *est une* relation d'ordre• *qui signifie* plus petit que *et inversement le symbole* $\geqslant$ *est aussi une relation d'ordre qui signifie* plus grand que. Déf3.7|p64

Nous avons également, concernant la relation de comparaison stricte [4], les propriétés suivantes qui sont relativement intuitives :

$$\begin{aligned} &a < b \quad \text{et} \quad b < c \quad \Rightarrow \quad a < c \\ &a < b \quad \text{et} \quad c > 0 \quad \Rightarrow \quad ac < bc \end{aligned} \tag{3.7}$$

Soient a, b deux nombres réels quelconques, alors :

$$a < b \Rightarrow a + c < b + c \qquad \text{et} \qquad a < b \Rightarrow a - c < b - c \tag{3.8}$$

inversement :

$$a > b \Rightarrow a + c > b + c \qquad \text{et} \qquad a > b \Rightarrow a - c > b - c \tag{3.9}$$

Nous avons aussi :

$$0 < a < b \Rightarrow \frac{1}{a} > \frac{1}{b} \tag{3.10}$$

et inversement :

$$b < a < 0 \Rightarrow \frac{1}{a} < \frac{1}{b} \tag{3.11}$$

Nous pouvons bien évidemment multiplier, diviser, addition ou soustraire un terme de chaque côté de la relation telle que celle-ci soit toujours vraie. En multipliant les deux membres par un nombre négatif il faudra bien évidemment changer le comparateur tel que si :

$$a < b \quad \text{et} \quad c > 0 \quad \Rightarrow \quad ac < bc \tag{3.12}$$

et inversement :

$$a < b \quad \text{et} \quad c < 0 \quad \Rightarrow \quad ac > bc \tag{3.13}$$

Nous avons aussi :

$$a < b \quad \text{et} \quad p \in \mathbb{R}^{*+} \quad \Rightarrow \quad a^p < b^p \tag{3.14}$$

Soit :

$$b < a < 0 \quad \text{et} \quad p \in \mathbb{N}^{*+} \quad \Rightarrow \quad \begin{cases} 0 < a^p < b^p & \text{si } p \text{ pair} \\ a^p > b^p & \text{sinon} \end{cases} \tag{3.15}$$

Ce résultat provient simplement de la multiplication des signes puisque la puissance lorsqu'elle est non fractionnaire n'est qu'une multiplication.

Finalement :

$$0 < a < b \quad \text{et} \quad n \in \mathbb{N}^{*+} \quad \Rightarrow \quad \sqrt[n]{a} < \sqrt[n]{b} \tag{3.16}$$

Les relations d'ordre :

$$> \quad < \quad \leqslant \quad \geqslant \quad \gg \quad \ll \tag{3.17}$$

correspondent respectivement à : (strictement) plus grand que, (strictement) plus petit que, plus petit ou égal à, plus grand ou égal à, beaucoup plus grand que et enfin beaucoup plus petit que.

Les relations d'ordre peuvent être définies de façon un peu plus subtile et rigoureuse et abstraite et ne s'appliquent pas seulement aux comparateurs•. Chap5|p117

4. La relation de comparaison stricte n'appartient pas à la famille des relations d'ordre pour des raisons que nous préciserons plus loin.

Définition 3.6 *Soit une relation binaire R d'un ensemble A vers lui-même, une relation R dans A est un sous-ensemble du produit cartésien $R \subseteq A \times A$*[5] *avec la propriété d'être :*

1. *une* relation réflexive *si* $\forall x \in A,\ xRx$;
2. *une* relation symétrique *si* $\forall x, y \in A,\ xRy \Rightarrow yRx$;
3. *une* relation antisymétrique *si* $\forall x, y \in A,\ xRy$ *et* $yRx \Rightarrow x = y$;
4. *une* relation transitive *si* $\forall x, y, z \in A,\ xRy$ *et* $yRz \Rightarrow xRz$;
5. *une* relation connexe *si* $\forall x, y \in A,\ \forall (x, y) \in A \Rightarrow (xRy)$ *ou* (yRx) ;

Les mathématiciens ont donné des noms particuliers aux familles de relations satisfaisant certaines de ces propriétés.

Définition 3.7 *Une relation est appelée* relation d'ordre stricte *si et seulement si elle est transitive.*

Définition 3.8 *Une relation est appelée un* pré-ordre *si et seulement si elle y est réflexive et transitive*

Définition 3.9 *Une relation est appelée une* relation ordre partiel *si et seulement si elle y est réflexive, antisymétrique et transitive*

Définition 3.10 *Une relation est appelée* une relation d'équivalence *si et seulement si elle y est réflexive, symétrique, transitive*

Définition 3.11 *Une relation est appelée* relation d'ordre *si et seulement si elle y est réflexive, transitive et antisymétrique*

Définition 3.12 *Une relation est appelée* relation d'ordre total *si et seulement si elle y est réflexive, transitive, connexe et antisymétrique*

REMARQUE Les relations d'ordre binaire ont toutes des propriétés similaires dans les ensembles naturels, rationnels, relatifs et réels (il n'y a pas de relation d'ordre naturelle sur l'ensemble des nombres complexes).

L'ensemble est résumé dans le tableau 3.1.

RELATION BINAIRE	$=$	$\neq$	$>$	$<$	$\leqslant$	$\geqslant$
réflexive	oui	non	non	non	oui	oui
symétrique	oui	oui	non	non	non	non
transitive	oui	non	oui	oui	oui	oui
connexe	non	non	non	non	oui	oui
antisymétrique	oui	non	non	non	oui	oui

TABLEAU 3.1 – Relations d'ordre

Ainsi, nous voyons que les relations binaires $\leqslant, \geqslant$ forment avec les ensembles précités, des relations d'ordre total et qu'il est très facile de voir quelles relations binaires sont des relations d'ordre partiel, total ou d'équivalence.

Définition 3.13 *Si R est une relation d'équivalence sur A alors $\forall x \in A$, la* classe d'équivalence *de x est par définition l'ensemble :*

$$[x] = \{y \in A : xRy\} \tag{3.18}$$

$[x]$ est donc un sous-ensemble de A ($[x] \subseteq A$) que nous noterons aussi par la suite R[6].

Nous disposons ainsi d'un nouvel ensemble qui est *l'ensemble des classes d'équivalences* ou *ensemble quotient* noté A/R. Ainsi :

$$A/R = \{[x] | x \in A\} \tag{3.19}$$

5. C'est-à-dire que la relation binaire engendre un sous-ensemble de par les contraintes qu'elle impose aux éléments de A qui satisfont la relation.

6. Il faut bien évidemment faire attention à ne pas confondre dans ce qui suit la relation d'équivalence et le sous-ensemble.

Il faut savoir que dans A/R nous ne regardons plus $[x]$ comme un sous-ensemble de A mais comme un élément !

Une relation d'équivalence, de manière vulgarisée sert donc à coller une seule étiquette à des éléments qui vérifient une même propriété, et à les confondre avec ladite étiquette.

Exemple 3.1 *Dans l'ensemble des entiers relatifs $\mathbb{Z}$, si nous étudions les restes de la division par 2, nous avons que ceux-ci valent toujours soit 0 soit 1.*

La classe d'équivalence de zéro est alors appelée l'ensemble des nombres entiers pairs, la classe d'équivalence de 1 est appelée l'ensemble des entiers impairs.

Si nous nommons la première 0 et la deuxième 1, nous retrouvons les règles d'opérations entre nombres pairs et impairs :

$$0+0=0 \qquad 0+1=0 \qquad 1+1=0 \tag{3.20}$$

ce qui signifie respectivement que la somme de deux entiers pair est pair, que la somme d'un pair et d'un impair est impair et que la somme de deux impairs est pair.

Et pour la multiplication :

$$0\times 0=0 \qquad 0\times 1=0 \qquad 1\times 1=1 \tag{3.21}$$

ce qui signifie respectivement que le produit de deux pairs est pair, le produit d'un pair et d'un impair est pair et que le produit de deux impairs est impair. Nous avons déplacé les opérations de $\mathbb{Z}$ sur cet ensemble quotient noté $\mathbb{Z}/2\mathbb{Z}$.

Maintenant, pour vérifier que nous avons bien affaire à une relation d'équivalence, il faudrait encore vérifier qu'elle est réflexive (xRx), symétrique (si xRy alors yRx) et transitive (si xRy et yRz alors xRz).

Définition 3.14 *L'application $\pi : A \to A$ définie par $x \to [x]$ est appelée* projection canonique. *Tout élément $z \in [x]$ est alors appelé* représentant de la classe $[x]$.

Proposition 3.1 *Considérons maintenant un ensemble E. Il y a bijection entre l'ensemble des relations d'équivalences sur E et l'ensemble des partitions de E.*

En d'autres termes cette proposition dit qu'une relation d'équivalence sur E n'est rien d'autre qu'une partition de E.

Preuve Soit R une relation d'équivalence sur E. Nous choisissons $I := E/R$ comme ensemble d'indexation des partitions et nous posons pour tout $[x] \in E/R$, $E_{[x]} := [x]$. Il suffit de vérifier les deux propriétés suivantes de la définition des partitions pour montrer que la famille $(E_{[x]})_I$ est une partition de E :

1. Soit $[x], [y] \in E/R$ tels que $[x] \neq [y]$ alors (trivial) $E_{[x]} \cap E_{[y]} = \varnothing$.
2. $E = \cup_{[x]\in E/R} E_{[x]}$ est évident car si $x \in E$ alors $x \in [x] = E_{[x]}$. •

Nous avons donc associé à la relation R une partition de E. Réciproquement si $(E_i)_I$ est une partition de E alors nous vérifions facilement que la relation R définie par xRy si et seulement s'il existe $i \in I$ tel que $x, y \in E_i$, est une relation d'équivalence. Les deux applications ainsi définies sont bijectives et réciproques l'une de l'autre.

Nous allons à présent appliquer sur un exemple un peu moins trivial que le précédent ce que nous venons de voir à la construction des anneaux• $\mathbb{Z}/d\mathbb{Z}$ après quelques rappels : Chap4|p100

1. Soit deux nombres $n, m \in \mathbb{Z}$. Nous disons que n *divise* m et nous écrivons $n|m$ si et seulement s'il existe un entier $k \in \mathbb{Z}$ tel que $m = k \cdot n$•. Chap2
2. Soit $d \geqslant 1$ un entier. Nous définissons la relation R par nRm si et seulement si $d|(n-m)$ ou dit autrement nRm si et seulement s'il existe $k \in \mathbb{Z}$ tel que $n = m + k \cdot d$. Généralement nous écrivons ceci aussi $n \equiv m$ (modulo d) au lieu de nRm et nous disons que n *est congru à m modulo d*. Rappelons aussi que $n \equiv 0$ (modulo d) si et seulement si d divise n•. Chap2

Nous allons maintenant introduire une relation d'équivalence sur $\mathbb{Z}$. Démontrons que pour tout entier $d \geqslant 1$, la congruence modulo d est une relation d'équivalence sur $\mathbb{Z}$.

Preuve (contrôle des trois propriétés de l'équivalence)

1. réflexivité : $n \equiv m$ car $n = n + 0 \cdot d$;
2. symétrie : si $n \equiv m$ alors $n = m + kd$ et donc $m = n + (-k)d$ c'est-à-dire $m \equiv n$;

3. transitivité : si $n \equiv m$ et $m \equiv j$ alors $n = m + kd$ et $m = j + k'd$ donc $n = j + (k + k')d$, soit $n \equiv j$.•

Dans la situation ci-dessus, nous notons $\mathbb{Z}/d\mathbb{Z}$ l'ensemble des classes d'équivalences et noterons $[n]_d$ la classe d'équivalence de la congruence d'un entier n donnée par [7] :

$$[n]_d = \{\ldots, n-2d, n-d, n, n+d, n+2d, n+3d, \ldots\} \tag{3.22}$$

et ainsi :

$$\mathbb{Z}/d\mathbb{Z} = \{[0]_d, [1]_d, [2]_d, \ldots, [d-1]_d\} \tag{3.23}$$

En particulier (trivial car nous obtenons ainsi tout $\mathbb{Z}$) :

$$\mathbb{Z}/2\mathbb{Z} = \{[0]_2, [1]_2\} \tag{3.24}$$

Ainsi, nous voyons que le premier exemple que nous avions donné avec les nombres pairs et impairs est un cas particulièrement simple des classes d'équivalence de congruence.

REMARQUE Les opérations d'addition et de multiplication définies sur $\mathbb{Z}$ définissent des opérations d'addition et de multiplication sur $\mathbb{Z}/d\mathbb{Z}$. Nous disons alors que ces opérations sont compatibles avec la relation d'équivalence et forment alors un anneau•.

Chap4

3.2 Lois fondamentales de l'arithmétique

Comme nous l'avons déjà dit précédemment, il existe un opérateur de base (addition) à partir duquel il possible de définir la multiplication, la soustraction et la division [8] et autour desquels nous pouvons construire toute la mathématique analytique.

3.2.1 Addition

Définition 3.15 *L'addition des nombres entiers est une opération notée "+" qui a pour seul but de réunir en un seul nombre toutes les unités contenues dans plusieurs autres. Le résultat de l'opération se nomme* somme *ou* total. *Les nombres à additionner sont appelés* termes de l'addition. *Les signes d'addition "+ et de soustraction −" sont dus à* WIDMANN *en 1489.*

Ainsi, $A+B+C...$ sont les termes de l'addition et le résultat est la somme des termes de l'addition.

Propriété 3.2 *La somme de plusieurs nombres ne dépend pas de l'ordre des termes. Nous disons alors que l'addition est une* opération commutative. *Ce qui signifie que nous avons fréquemment :*

$$A - B \neq B - A \tag{3.25}$$

Propriété 3.3 *La somme de plusieurs nombres ne change pas si nous remplaçons deux ou plusieurs d'entre eux par leur résultat intermédiaire. Nous disons alors que l'addition est* opération associative.

Propriété 3.4 *Le zéro est l'élément neutre de l'addition car tout nombre additionné à zéro donne ce même nombre.*

Propriété 3.5 *Suivant l'ensemble dans lequel nous travaillons, l'addition peut comporter un terme de telle façon à ce que le total soit nul. Nous disons alors qu'il existe un* opposé *pour l'addition.*

Nous allons définir plus rigoureusement l'addition en utilisant l'axiomatique de PEANO comme nous en avons déjà fait mention dans le chapitre traitant des nombres. Ainsi, avec ces axiomes il est possible de démontrer qu'il existe (existence) une et une seule application (unicité), notée "+", de $\mathbb{N} \times \mathbb{N}$ dans $\mathbb{N}$ vérifiant :

$$\begin{aligned} &\forall n \in \mathbb{N}, \quad n + 0 = n \\ &\forall p \in \mathbb{N}, \forall q \in \mathbb{N}, \quad p + S(q) = S(p + q) \\ &\forall n \in \mathbb{N}, \quad S(n) = n + 1 \end{aligned} \tag{3.26}$$

7. Dans l'équation (3.22), chaque différence de deux valeurs se trouvant dans les accolades est divisible par d et c'est ainsi bien un classe d'équivalence.

8. À condition, pour la soustraction et la division, que l'ensemble de nombres le permettent

REMARQUE Ce document n'ayant pas pour vocation de s'adresser à des mathématiciens, nous nous passerons de la démonstration (relativement longue) et admettrons intuitivement que l'application "+" existe et est unique et qu'il en découle les propriétés susmentionnées.

Soit $x_1, x_2, \ldots, x_n$ des nombres quelconques alors nous pouvons noter également la somme ainsi :

$$x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i \tag{3.27}$$

en définissant des bornes supérieurs et inférieures à la somme (au-dessus et en-dessous de la lettre grecque majuscule *sigma*). Si k est une constante, alors :

$$\sum_{i=1}^{n} kx_i = k\sum_{i=1}^{n} x_i \qquad \sum_{i=1}^{n} k = nk \qquad \sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \tag{3.28}$$

Exemple 3.2 *L'addition de deux nombres relativement petits est assez facile dès que nous avons appris par cœur à compter jusqu'au nombre résultant de cette opération. Ainsi sur la base décimale :*

$$\begin{array}{r} 5 \\ +\,2 \\ \hline 7 \end{array} \qquad \begin{array}{r} 10 \\ +\;3 \\ \hline 13 \end{array} \qquad \begin{array}{r} 1014 \\ +\quad 3 \\ \hline 1017 \end{array} \tag{3.29}$$

Pour les beaucoup plus grands nombres il faut adopter une autre méthode qu'il s'agit d'apprendre par cœur. Ainsi par exemple :

$$\begin{array}{r} 1244 \\ +3475 \\ \hline ???? \end{array} \tag{3.30}$$

Démarche : nous additionnons les colonnes (4 colonnes dans cet exemple) de droite à gauche. Pour la première colonne nous avons donc $4 + 5 = 9$ ce qui nous donne :

$$\begin{array}{r} 1244 \\ +3475 \\ \hline ???9 \end{array} \tag{3.31}$$

et nous continuons ainsi pour la deuxième $4 + 7 = 11$ mais à la différence que comme nous avons un nombre supérieur à la dizaine, nous reportons le premier chiffre (de gauche) sur la colonne suivante de l'addition. Ainsi :

$$\begin{array}{r} {\scriptstyle 1} \\ 1244 \\ +3475 \\ \hline ??19 \end{array} \tag{3.32}$$

La troisième colonne se calcule dès lors comme $4 + 2 + 1 = 7$ ce qui nous donne :

$$\begin{array}{r} {\scriptstyle 1} \\ 1244 \\ +3475 \\ \hline ?719 \end{array} \tag{3.33}$$

Pour la dernière colonne nous avons $1 + 3 = 4$:

$$\begin{array}{r} {\scriptstyle 1} \\ 1244 \\ +3475 \\ \hline 4719 \end{array} \tag{3.34}$$

Voilà comment nous procédons donc pour l'addition de nombres quelconques : nous faisons une addition par colonne de droite à gauche et si le résultat d'une addition est supérieure à la dizaine, nous reportons une unité sur la colonne suivante.

Cette méthodologie d'addition est simple à comprendre et à effectuer. Nous ne l'expliciterons pas plus.

3.2.2 Soustraction

Définition 3.16 *La soustraction du nombre entier A par le nombre entier B notée par le symbole "$-$", c'est trouver le nombre C qui, ajouté à B, redonne A.*

REMARQUE L'opération n'est rigoureusement parlant pas possible dans les entiers naturels $\mathbb{N}$ que si $A > B$.

Nous écrivons la soustraction sous la forme :

$$A - B = C \tag{3.35}$$

qui doit évidemment vérifier :

$$A = C + B \tag{3.36}$$

Propriétés intuitives que nous admettrons sans démonstrations de l'opération de soustraction :

Propriété 3.6 *La soustraction de plusieurs nombres dépend de l'ordre des termes. Nous disons alors que la soustraction est une* opération non-commutative.

Propriété 3.7 *La soustraction de plusieurs nombres ne change pas si l'on remplace deux ou plusieurs d'entre eux par leur résultat intermédiaire. Nous disons alors que la soustraction est une* opération non-associative.

Propriété 3.8 *Le zéro n'est l'élément neutre de la soustraction car tout nombre à qui on soustrait zéro donne ce même nombre, donc le zéro est neutre à droite mais pas à gauche !*

Propriété 3.9 *Suivant l'ensemble dans lequel nous travaillons, la soustraction peut comporter un terme de telle façon à ce que le total soit nul. Nous disons alors qu'il existe un* opposé *pour la soustraction.*

Exemple 3.3 *La soustraction de deux nombres relativement petits est assez facile dès que nous avons appris par cœur à compter jusqu'à au moins le nombre résultant de cette opération. Ainsi :*

$$\begin{array}{r} 5 \\ -\,2 \\ \hline 3 \end{array} \qquad \begin{array}{r} 10 \\ -\,3 \\ \hline 7 \end{array} \qquad \begin{array}{r} 1014 \\ -\quad 3 \\ \hline 1011 \end{array} \tag{3.37}$$

Pour les beaucoup plus grands nombres, il faut adopter un autre méthode qu'il s'agit d'apprendre par cœur. Ainsi par exemple :

$$\begin{array}{r} 4574 \\ -3485 \\ \hline ???? \end{array} \tag{3.38}$$

nous soustrayons les colonnes de droite à gauche. Pour la première colonne, nous avons $4 - 5 = -1 < 0$ *et nous ajoutons et soustrayons* 10 *aux deux nombres, d'où* $14 - 5 = 9$ *:*

$$\begin{array}{r} 457\not{1}4 \\ -34\not{1}85 \\ \hline ???9 \end{array} \tag{3.39}$$

et nous continuons ainsi pour la deuxième $7 - 9 = -2$ *ce qui fait que nous reportons une nouvelle fois 100 aux deux membres d'où* $17 - 9 = 8$ *:*

$$\begin{array}{r} 45\not{1}7\not{1}4 \\ -3\not{1}4\not{1}85 \\ \hline ??89 \end{array} \tag{3.40}$$

Continuons de la même manière au niveau des milliers avec $5 - 5 = 0$ *:*

$$\begin{array}{r} 45\not{1}7\not{1}4 \\ -3\not{1}4\not{1}85 \\ \hline ?089 \end{array} \tag{3.41}$$

Pour la dernière colonne nous avons une situation qui ne pose pas de problème :

$$\begin{array}{r} 4\,5\,{}^{1}\!\not{7}\,{}^{1}\!\not{4} \\ -\,3\,{}^{1}\!\not{4}\,{}^{1}\!\not{8}\,5 \\ \hline 1\,0\,8\,9 \end{array} \tag{3.42}$$

Voilà comment nous procédons donc pour la soustraction de nombres quelconques. Nous faisons une soustraction par colonne de droite à gauche et si le résultat d'une soustraction est inférieure à zéro nous reportons 1 *au niveau du calcul en cours.*

La méthodologie utilisée pour la soustraction se basant sur exactement le même principe que l'addition nous ne l'expliciterons pas plus. Cette méthode est très simple et nécessite bien sûr une certaine habitude de travailler avec les chiffres pour être totalement appréhendée.

3.2.3 Multiplication

Définition 3.17 *La multiplication des nombres est une opération qui a pour but, étant donné deux nombres, l'un appelé* multiplicateur*, et l'autre* multiplicande*, d'en trouver un troisième appelé* produit *qui soit la somme (donc la multiplication d'écoule de la somme !) d'autant de nombres égaux au multiplicande qu'il y a d'unités au multiplicateur. Le multiplicande et le multiplicateur sont appelés les* facteurs du produit*.*

La multiplication s'indique à l'aide du signe "$\times$" (anciennement) ou du point de ponctuation surélevé (notation moderne) ou sans aucun symbole tel que :

$$a \times b = a \cdot b = ab \tag{3.43}$$

Remarque Le signe de croix "$\times$" pour la multiplication se trouve pour la première fois dans l'ouvrage d'Ougtred (1631) quant au point à mi-hauteur (notation moderne pour la multiplication), nous le devons à Leibniz. Dès 1544, Stiefel, dans un de ses ouvrages n'employait aucun signe et désignait le produit de deux nombres en les plaçant l'un après l'autre.

Nous pouvons définir la multiplication en utilisant l'axiomatique de Peano comme nous en avons déjà fait mention dans le chapitre traitant des nombres. Ainsi, avec ces axiomes il est possible de démontrer qu'il existe une et une seule application, notée "$\times$" ou plus souvent "$\cdot$", de $\mathbb{N} \times \mathbb{N}$ dans $\mathbb{N}$ vérifiant :

$$\begin{aligned} &\forall n \in \mathbb{N}, \quad n \cdot 0 = 0 \\ &\forall p \in \mathbb{N}, \forall q \in \mathbb{N}, \quad p \cdot (q+1) = p \cdot q + q \end{aligned} \tag{3.44}$$

La puissance est une notation particulière d'un cas précis de multiplications. Lorsque le(s) multiplicateur(s) et multiplicande(s) sont identique(s) en valeur numérique, on note la multiplication (par exemple) :

$$n \cdot n \cdot n \cdot n \cdot n \cdot n \cdot n \cdot n = n^8 \tag{3.45}$$

c'est ce que nous nommons la *notation en puissance* ou *l'exponentiation*. Le nombre en exposant est la *puissance* ou *l'exposant* du nombre, n en l'occurrence. La notation en exposants se trouve pour la première fois dans l'ouvrage de Chuquet intitulé *Triparty en la science des nombres* (1484).

Les propriétés admises sans démonstration de l'opération de multiplication sont :

Propriété 3.10 *La multiplication de plusieurs nombres ne dépend pas de l'ordre des termes. Nous disons alors que la multiplication est une* opération commutative*.*

Propriété 3.11 *La multiplication de plusieurs nombres ne change pas si l'on remplace deux ou plusieurs d'entre eux par leur résultat intermédiaire. Nous disons alors que la multiplication est* opération associative*.*

Propriété 3.12 *L'unité est l'élément neutre de la multiplication car tout multiplicande multiplié par le multiplicateur 1 est égal au multiplicande.*

Propriété 3.13 *La multiplication peut comporter un terme de telle façon à ce que le produit soit égal à l'unité (l'élément neutre). Nous disons alors qu'il existe un* inverse pour la multiplication.

Propriété 3.14 *La multiplication est* distributive, *c'est-à-dire que :*

$$a \cdot (b + c) = ab + ac \tag{3.46}$$

l'opération inverse s'appelant la factorisation.

Introduisons encore quelques notations particulières relatives à la multiplication :

1. Soit $x_1, x_2, \ldots, x_n$ des nombres quelconques (non nécessairement égaux) alors nous pouvons noter le produit ainsi :

$$x_1 \cdot x_2 \cdot \ldots \cdot x_n = \prod_{i=1}^{n} x_i \tag{3.47}$$

en définissant des bornes supérieurs et inférieures au produit (au-dessus et en-dessous de la lettre grecque majuscule *Pi*).
Rappel des propriétés relatives à cette notation :

$$\prod_{i=1}^{n} kx_i = k^n \prod_{i=1}^{n} x_i \tag{3.48}$$

pour tout nombre k tel que :

$$\prod_{i=1}^{n} k = k^n \tag{3.49}$$

Nous avons aussi par exemple :

$$\prod_{i=1}^{n} (x + y) = (x + y)^n \tag{3.50}$$

2. Nous définissons également la *factorielle* simplement (car il existe aussi un manière complexe de la définir en passant par la fonction Gamma d'EULER comme cela est fait dans le chapitre de Calcul Différentiel Et Intégral) par :

$$1 \times 2 \times 3 \times 4 \times \cdots \times n = n! \tag{3.51}$$

Exemple 3.4 *La multiplication de deux nombres relativement petits est assez facile dès que nous avons appris par cœur à compter jusqu'à au moins le nombre résultant de cette opération. Ainsi :*

$$\begin{array}{rrr} 5 & 10 & 1014 \\ \times\, 2 & \times\, 3 & \times \quad 3 \\ \hline 10 & 30 & 3042 \end{array} \tag{3.52}$$

Pour les beaucoup plus grands nombres il faut adopter une autre méthode qu'il s'agit d'apprendre par cœur. Ainsi par exemple :

$$\begin{array}{r} 4574 \\ - \quad 8 \\ \hline ???? \end{array} \tag{3.53}$$

nous multiplions colonne par colonne et nous additionnons l'ensemble des résultats décalés d'un chiffre comme ci-dessous ($8 \times 4 = 32$, $8 \times 7 = 56$, $8 \times 5 = 40$, $8 \times 4 = 32$) ainsi nous obtenons :

$$\begin{array}{r} 4575 \\ \times \quad 8 \\ \hline 32 \\ 56 \\ 40 \\ +32 \\ \hline 36592 \end{array} \tag{3.54}$$

Cette méthodologie est très logique si vous avez bien compris comment nous construisons un chiffre en base dix. Ainsi, nous avons[9] *:*

$$\begin{aligned} 8 \times 4574 &= 8 \times (4000 + 500 + 70 + 4) \\ &= 8 \times 4000 + 8 \times 500 + 8 \times 70 + 8 \times 4 \\ &= 32000 + 4000 + 560 + 32 \\ &= 36592 \end{aligned} \tag{3.55}$$

Pour ne pas surcharger l'écriture dans la multiplication par la méthode *verticale*, nous ne représentons pas les zéros qui surchargeraient inutilement les calculs.

3.2.4 Division

Définition 3.18 *La division des nombres entiers est une opération, qui a pour but, étant donné deux nombres, l'un appelé* dividende*, l'autre appelé* diviseur*, d'en trouver un troisième appelé* quotient *qui soit le plus grand nombre dont le produit par le diviseur puisse se retrancher (donc la division découle de la soustraction !) du dividende (la différence étant nommé le* reste *ou la* congruence*).*

Remarque Dans les cas des nombre réels il n'y a jamais de reste à la fin de l'opération de division (car le quotient multiplié par le diviseur donne exactement le dividende) !

D'une façon générale dans le cadre des nombres entiers, si nous notons D le dividende, d le diviseur, Q le quotient et R le reste nous avons la relation :

$$D = A \cdot d + R \tag{3.56}$$

en sachant que la division était :

$$D : d = \frac{D}{d} \tag{3.57}$$

Nous désignons également souvent par *fraction* (au lieu de *quotient*), le rapport de deux nombres ou autrement dit, la division du premier par le deuxième.

Remarque Le signe de la division ":" est dû à Leibniz ; la barre de fraction se trouve elle pour la première fois dans les ouvrages de Fibonacci (1202) ; elle est probablement due aux hindous.

Si nous divisons deux entiers et que nous souhaitons un entier comme quotient et reste uniquement, alors nous parlons de *division euclidienne*.

Nous indiquons l'opération en plaçant entre les deux nombres, le dividende et le diviseur le signe " :" ou *une barre de division* "/". Si nous avons :

$$D : d = D \cdot \frac{1}{d} = D \cdot i_D \tag{3.58}$$

on appelle i_D l'inverse du dividende. À tout nombre est associé un inverse qui satisfait cette condition. De cette définition, il vient la notation, $\forall x \neq 0$:

$$\frac{x}{x} = x^1 \cdot x^{-1} = x^{1-1} = x^0 = 1 \tag{3.59}$$

Dans le cas de deux nombres fractionnaires, nous disons qu'ils sont *inverses* ou *réciproques*, lorsque leur produit est égal à l'unité, comme dans (3.59), pour toute valeur de x, positive ou négative.

Remarques

1. Une division par zéro est ce que l'on nomme une *singularité*. C'est-à-dire que le résultat de la division est indéterminé.
2. Lorsque nous multiplions le dividende et le diviseur d'une division par un même nombre, le quotient ne change pas, mais le reste est multiplié par ce nombre.
3. Diviser un nombre par un produit effectué de plusieurs facteurs revient à diviser ce nombre successivement par chacun des facteurs du produit et réciproquement.

9. Nous supposons, à ce niveau, que la distributivité est connue.

Les propriétés des divisions avec les puissances sont les suivantes :

$$\frac{x \cdot x \cdot x}{y \cdot y} = x^3 \cdot y^{-2} \tag{3.60}$$

ou :

$$\frac{x \cdot x \cdot x}{x \cdot x} = \frac{x}{x} \cdot \frac{x}{x} \cdot x = 1 \cdot 1 \cdot x = x^3 \cdot x^{-2} = x^1 = x \tag{3.61}$$

Rappelons qu'un nombre premier (entier relatif) est un nombre qui n'a d'autres diviseurs que lui-même et l'unité. Donc tout nombre qui n'est pas premier a un nombre premier comme diviseur. Le plus petit des diviseurs d'un nombre entier est donc un nombre premier.

Quelques propriétés de la division qui découlent d'un raisonnement logique des propriétés de la multiplication sont :

$$\begin{array}{ll} \frac{a}{b} = \frac{c}{d} \Leftrightarrow a \cdot d = b \cdot c & \frac{a \cdot d}{b \cdot d} = \frac{a}{b} \cdot \frac{d}{d} = \frac{a}{b} \cdot 1 = \frac{a}{b} \\ \frac{a}{-b} = \frac{-a}{b} = -\frac{a}{b} & \frac{a}{b} + \frac{c}{b} = \frac{a+c}{b} \\ \frac{a}{b} + \frac{c}{d} = \frac{a \cdot d + b \cdot c}{b \cdot d} & \frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d} \\ \frac{a}{b} \div \frac{c}{d} = \frac{a}{b} \cdot \frac{d}{c} = \frac{a \cdot d}{b \cdot c} & \end{array} \tag{3.62}$$

accompagnées de ce qui suit :

Propriété 3.15 *La division de plusieurs nombres dépend de l'ordre des termes. Nous disons alors que la division est une* opération non-commutative. *Ce qui signifie que nous avons fréquemment :*

$$\frac{A}{B} \neq \frac{B}{A} \tag{3.63}$$

Propriété 3.16 *La division de plusieurs nombres change si l'on remplace deux ou plusieurs d'entre eux par leur résultat intermédiaire. Nous disons alors que la division est une opération* non-associative.

Propriété 3.17 *L'unité est l'élément neutre à droite de la division car tout dividende divisé par le diviseur 1 est égal au dividende mais l'unité par pas neutre à gauche.*

Propriété 3.18 *La division peut comporter un terme de telle façon à ce que la division soit égale à l'unité (l'élément neutre). Nous disons alors qu'il existe un* symétrique pour la division.

Si a et b sont deux nombres réels positifs et non nuls nous avons :

$$a^p \cdot a^q = a^{p+q} \quad \frac{a^p}{a^q} = a^{p-q} \Rightarrow a^{-q} = \frac{1}{a^q} \quad a^p \cdot b^p = (a \cdot b)^p \Leftrightarrow \frac{a^p}{b^p} = \left(\frac{a}{b}\right)^p \tag{3.64}$$

Nous pouvons maintenant définir la racine q-ème principale d'un nombre quelconque a :

$$a^{\frac{1}{q}} = \sqrt[q]{a} \tag{3.65}$$

la dernière relation n'étant définie que pour $q \in \mathbb{N}^*$. Au niveau de la terminologie, nous avons :

$$a^{\frac{1}{q}} = \sqrt[q]{a} \tag{3.66}$$

qui est une racine, le nombre a est le *radicante* et q est l'indice de la racine. Le symbole $\sqrt{}$ est appelé le *radical*. De ce qui a déjà été dit pour les puissances, nous pouvons conclure aisément que :

$$\sqrt[q]{a} \cdot \sqrt[q]{b} = \sqrt[q]{a \cdot b} \Leftrightarrow \frac{\sqrt[q]{a}}{\sqrt[q]{b}} = \sqrt[q]{\frac{a}{b}} \quad \text{et} \quad (a^p)^q = a^{p \cdot q} = (a^q)^p \tag{3.67}$$

d'où il en ressort que :

$$a^{\frac{p}{q}} = \sqrt[q]{a^p} = (\sqrt[q]{a})^p \quad \text{et} \quad \sqrt[p]{\sqrt[q]{a}} = \sqrt[pq]{a} \tag{3.68}$$

Nous avons également si $a < 0$:

$$\sqrt[q]{a^q} = \begin{cases} a & \text{si} \quad q \in \mathbb{N}^* \text{ est impair} \\ |a| & \text{si} \quad q \in \mathbb{N}^* \text{ est pair} \end{cases} \tag{3.69}$$

Si le dénominateur d'un quotient contient un facteur de la forme $\sqrt[n]{a^k}$ avec $a \neq 0$, en multipliant la numérateur et le dénominateur par $\sqrt[n]{a^{n-k}}$, nous supprimerons la racine au dénominateur, puisque :

$$\frac{x}{\sqrt[n]{a^k}} \frac{\sqrt[n]{a^{n-k}}}{\sqrt[n]{a^{n-k}}} = \frac{x\sqrt[n]{a^{n-k}}}{\sqrt[n]{a^k}\sqrt[n]{a^{n-k}}} = \frac{x\sqrt[n]{a^{n-k}}}{\sqrt[n]{a^k a^{n-k}}} = \frac{x\sqrt[n]{a^{n-k}}}{\sqrt[n]{a^{k+n-k}}} = \frac{x\sqrt[n]{a^{n-k}}}{\sqrt[n]{a^n}} = \frac{x\sqrt[n]{a^{n-k}}}{|a|} \tag{3.70}$$

Nous appelons communément ce procédé *rendre un dénominateur rationnel*. Nous pouvons bien sûr faire de même avec le numérateur.

Remarques

1. *Simplifier* une expression contenant des puissances de nombres réels signifie qu'il faut regrouper les termes ayant des exposants identiques.
2. Si $a < 0$ et $q \in \mathbb{N}^*$ est impair, alors $\sqrt[q]{a}$ est le nombre réel négatif b tel que $b^q = a$. Si $q \in \mathbb{N}^*$ est pair alors bien sûr, comme nous l'avons déjà vu, la racine est complexe•. Chap2

3.3 Polynômes arithmétiques

Définition 3.19 *Un polynôme arithmétique*[10] *est un ensemble de nombres séparés les uns des autres par les opérations d'addition ou de soustraction, + ou −.*

Les composants enfermés dans le polynôme sont appelés *termes* du polynôme. Lorsque le polynôme contient un unique terme, nous parlons alors de *monôme*, s'il y a deux termes nous parlons de *binôme*, et ainsi de suite...

Démontrons que la valeur d'un polynôme arithmétique est égale à l'excès de la somme des termes précédés du signe + sur la somme des termes précédés du signe −.

Preuve

$$n_1 - n_2 + n_3 - n_4 + n_5 - n_6 + \ldots - n_i = (n_1 + n_3 + n_5 + \ldots + n_i) + (-1)\cdot(n_2 + n_4 + n_6 + \ldots + n_{i+1}) \tag{3.71}$$

quelque soit les valeurs des termes. •

Mettre en évidence l'unité négative −1 est ce que nous appelons une *factorisation* ou *mise en facteurs*. L'opération inverse, s'appelant une *distribution*.

Le produit de plusieurs polynômes peut toujours être remplacé par un polynôme unique que nous appelons le *produit effectué*. Nous opérons habituellement comme suit : nous multiplions successivement tous les termes du premier polynôme, en commençant par la gauche, par le premier, le second..., le dernier terme du second polynôme. Nous obtenons ainsi un premier produit partiel ; nous faisons, s'il y a lieu, la réduction des termes semblables. Nous multiplions ensuite chacun des termes du produit partiel successivement par le premier, le second..., le dernier terme du troisième polynôme en commençant par la gauche et ainsi de suite.

Le produit des polynômes A, B, C,... L,... est la somme de tous les produites de n facteurs formés avec un terme de A, un terme de B..., et un terme de L. S'il n'y a aucune réduction, le nombre des termes du produit est égal au produit des nombres des termes des facteurs.

3.4 Valeur absolue

Un nombre réel est constitué de deux parties : un signe + ou - et une valeur absolue.

Exemple 3.5 *+7 est constitué du signe + et de la valeur absolue 7 alors que −5 est constitué du signe - et de la valeur absolue 5. La valeur absolue de +7 est donc 7, la valeur absolue de −5 est donc 5.*

10. Il ne faut pas les confondre avec les polynômes algébriques qui seront étudiés dans la section d'Algèbre.

Définition 3.20 *Pour tout nombre réel x, la* valeur absolue *de x, notée $|x|$ est donnée par :*

$$|x| = \begin{cases} x & si\ x > 0 \\ -x & si\ x < 0 \\ 0 & si\ x = 0 \end{cases} \tag{3.72}$$

Nous remarquons que :

$$|x| = \max(-x, x) \tag{3.73}$$

ainsi que les expressions équivalentes :

$$x \leqslant |x| \quad |-x| = |x| \quad |x| \leqslant y \Leftrightarrow -y \leqslant x \leqslant y \quad |x| \geqslant y \Leftrightarrow x \leqslant -y \vee x \geqslant y \tag{3.74}$$

ces dernières étant souvent utilisées dans le cadre de la résolution des inéquations. Il est aussi utile d'interpréter l'expression $|x - y|$ comme la distance entre les deux nombres x et y sur la droite réelle. *En munissant l'ensemble des nombres réels de la distance valeur absolue, il devient un espace métrique.* La résolution d'une inéquation telle que $|x-3| \leqslant 9$ se résout alors simplement à l'aide de la notion de distance. La solution est l'ensemble des réels dont la distance au réel 3 est inférieure ou égale à 9. C'est l'intervalle de centre 3 et de rayons 9 ou autrement écrit $[3-9, 3+9] = [-6, 12]$.

La valeur absolue a quelques propriétés triviales que nous énoncerons sans démonstrations :

Propriété 3.19 *La valeur absolue de la somme algébrique de plusieurs nombres réels est inférieure ou égale à la somme des valeurs absolues des composantes de la somme :*

$$|x + y| \geqslant |x| + |y| \tag{3.75}$$

ce que les mathématiciens appellent parfois la première inégalité triangulaire.

Propriété 3.20 *La valeur absolue de la différence est supérieure ou égale à la valeur absolue de la différence des valeurs absolues des composantes de la différence :*

$$|x - y| \geqslant \big||x| + |y|\big| \tag{3.76}$$

ce que les mathématiciens appellent parfois la deuxième inégalité triangulaire.

Propriété 3.21 *La valeur absolue du produit (multiplication) est égale au produit des valeurs absolues :*

$$|x \cdot y| = |x| \cdot |y| \tag{3.77}$$

Propriété 3.22 *La valeur absolue du rapport est égale au rapport des valeurs absolues :*

$$\left|\frac{x}{y}\right| = \frac{|x|}{|y|} \tag{3.78}$$

3.5 Règles de calcul

Fréquemment en informatique, nous parlons de *priorité des opérateurs*. En mathématiques nous parlons de *priorité des ensembles d'opérations et des règles des signes*. De quoi s'agit-il exactement ?

Nous avons déjà vu qu'elles étaient les propriétés des opérations d'addition, soustraction, multiplication, mise en puissance et division. Nous tenons donc à ce que le lecteur différencie la notion de *propriété* de celle ce de *priorité* qui sont des choses complètement différentes.

En mathématiques, en particulier, nous définissons les priorités des symboles $\{, [, (,),], \}$:

1. les opérations qui sont entre parenthèses $(\bullet)$ doivent êtres effectuées en premier dans le polynôme.
2. les opérations qui sont entre crochets $[\bullet]$ doivent êtres effectuées en second à partir des résultats obtenus des opérations qui se trouvaient entre les parenthèses $(\bullet)$.
3. à partir des résultats intermédiaires des opérations qui se trouvaient entre parenthèses $(\bullet)$ et crochets $[\bullet]$, sont calculées les opérations qui se situent entre les accolades $\{\bullet\}$.

Exemple 3.6 *Soit à calculer le polynôme :*

$$\{[5 \cdot (8+2) + 3 \cdot [4 + (8+6) \cdot 2]] \cdot (1+9)\} \cdot 7 + 1 \tag{3.79}$$

Selon les règles que nous avons définies tout à l'heure, nous calculons d'abord tous les éléments qui sont entre parenthèses $(\bullet)$*, c'est-à-dire :*

$$8+2=10 \quad (8+6)=14 \quad (1+9)=10 \tag{3.80}$$

ce qui nous donne :

$$\{[5 \cdot 10 + 3 \cdot [4 + 14 \cdot 2]] \cdot 10\} \cdot 7 + 1 \tag{3.81}$$

Toujours selon le règles que nous avons définies tout à l'heure, nous calculons maintenant tous les éléments entre crochets en commençant toujours à calculer les termes qui sont dans les crochets $[\bullet]$ *au plus bas niveau des autres crochets* $[\bullet]$*. Ainsi, nous commence par calculer l'expression* $[4+14\cdot 2]$ *qui se trouve dans le crochet de niveau supérieur :* $[5 \cdot 10 + 3 \ldots]$.

Cela nous donne $[4 + 14 \cdot 2] = 32$ *et donc :*

$$\{[5 \cdot 10 + 3 \cdot 32] \cdot 10\} \cdot 7 + 1 \tag{3.82}$$

Il nous reste à calculer maintenant $[5 \cdot 10 + 3 \cdot 32] = 146$ *et donc :*

$$\{146 \cdot 10\} \cdot 7 + 1 \tag{3.83}$$

Nous calculons maintenant l'unique terme entre accolade, ce qui nous donne $\{146 \cdot 10\} = 1\ 460$ *et finalement il nous reste :*

$$1\ 460 \cdot 7 + 1 = 10\ 221 \tag{3.84}$$

Évidemment il s'agit d'un cas particulier mais le principe est toujours le même.

La priorité des opérateurs arithmétiques est une notion spécifique aux langages informatiques du fait qu'on ne peut dans ces derniers écrire des relations mathématiques que sur une ligne unique.

Ainsi, en informatique l'expression :

$$\frac{-a \cdot (b+c)^d}{e^f} - g \tag{3.85}$$

s'écrit (à peu de choses près) :

$$-a * (b+c)\hat{}d/e\hat{}f - g \tag{3.86}$$

Un non-initié pourrait y lire :

$$\frac{-a \cdot (b+c)^d}{e^f} - g; \quad \frac{[-a \cdot (b+c)]^d}{e^f} - g; \quad \left\{\frac{-a \cdot (b+c)^d}{e}\right\}^f - g; \quad [-a \cdot (b+c)]^{\frac{d}{e^f}} - g \tag{3.87}$$

Ainsi, il a logiquement été défini un ordre de priorité des opérandes tel que les opérations soient effectuées dans l'ordre suivant :

1. $-$ négation
2. $^{(\cdot)}$ puissance
3. $\times$; / multiplication et division
4. $\backslash$ division entière (spécifique à l'informatique)
5. Mod modulo (voir théorie des nombres)
6. $+$; $-$ addition et soustraction

Évidemment les règles des parenthèses ( ), crochets [ ], et accolades { } qui ont été définies en mathématiques s'appliquent à l'informatique. Ainsi, nous obtenons dans l'ordre (nous remplaçons chaque opération effectuée par un symbole) :

D'abord les termes entre parenthèses :

$$-a * (b + c)\hat{}d/e\hat{}f - g = -a * \alpha\hat{}d/e\hat{}f - g \tag{3.88}$$

Ensuite les règles de priorité des opérateurs s'appliquent dans l'ordre défini précédemment, d'abord (1) :

$$-a * \alpha\hat{}d/e\hat{}f - g = \beta * \alpha\hat{}d/e\hat{}f - g \tag{3.89}$$

ensuite (2) :

$$\beta * \alpha\hat{}d/e\hat{}f - g = \beta * \chi/\delta - g \tag{3.90}$$

nous appliquons la multiplication (3) :

$$\beta * \chi/\delta - g = \varepsilon/\delta - g \tag{3.91}$$

et finalement la division (3) :

$$\varepsilon/\delta - g = \phi - g \tag{3.92}$$

Les règles (4) et (5) ne s'appliquent pas à cet exemple particulier. Finalement (6) :

$$\phi - g = \varphi \tag{3.93}$$

En informatique, il existe cependant plusieurs opérateurs que nous ne retrouvons pas en mathématiques et qui changent souvent de propriétés d'un langage informatique à un autre.

Les opérateurs de comparaison $(=, <, >, \ldots)$ possèdent tous une priorité identique. Cependant, les opérateurs les plus à gauche dans une expression, détiennent une priorité plus élevée. Les opérateurs logiques sont évalués dans l'ordre de priorité suivant :

$$\text{1. Not - 2. And - 3. Or - 4. Xor - 5. Eqv - 6. Imp.} \tag{3.94}$$

Il s'agit aussi de connaître les règles des signes en vigueur en mathématiques. D'abord, il faut savoir que ces dernières ne s'appliquent que dans le cas de la multiplication et la division. Soit deux nombres positifs $(+x), (+y)$. Nous avons :

$$(+x) \cdot (+y) = (+x) \cdot (+y) = +(x \cdot y) \tag{3.95}$$

Autrement dit, la multiplication de deux nombres positifs est un nombre positifs et ce pour généralisable à la multiplication de n nombres positifs.

$$(-x) \cdot (+y) = (+x) \cdot (-y) = -(x \cdot y) \tag{3.96}$$

Autrement dit, la multiplication d'un nombre positif par un nombre négatif est négatif. Ce qui est généralisable à un résultat positif de la multiplication s'il y a un nombre pair de nombres négatifs et à un résultat négatif pour un nombre impair de nombres négatifs sur la totalité n des nombres de la multiplication.

$$(-x) \cdot (-y) = (-x) \cdot (-y) = +(x \cdot y) \tag{3.97}$$

Autrement dit, la multiplication de deux nombres négatifs est positif. Ce qui est généralisable à un résultat positif de la multiplication s'il y a un nombre pair de nombre négatifs et à un résultat négatif pour un nombre impair de nombres négatifs.

Pour ce qui est des divisions, le raisonnement est identique :

$$(+x)/(+y) = +(x/y) \quad \text{et} \quad (+y)/(+x) = +(y/x) \tag{3.98}$$

Autrement dit, si le numérateur et le dénominateur sont positifs, alors le résultat de la division sera positif.

$$(+x)/(-y) = (-x)/(+y) = -(x/y) \quad \text{et} \quad (-y)/(+x) = (+y)/(-x) = -(y/x) \tag{3.99}$$

Autrement dit, si soit le numérateur ou le dénominateur est négatif, alors le résultat de la division sera forcément négatif.

$$(-x)/(-y) = +(x/y) \quad \text{et} \quad (-y)/(-x) = +(y/x) \tag{3.100}$$

Autrement dit, si le numérateur et le dénominateur sont positifs, alors le résultat de la division, sera forcément positif. Évidemment, si nous avons une soustraction de termes, il est possible de la récrire sous la forme :

$$x - y = x + (-1)y = -1 \cdot (-x + y) \tag{3.101}$$

4

Théorie des ensembles

Lors de notre étude des nombres, des opérateurs, et de théorie des nombres (dans les chapitres du même nom), nous avons assez souvent utilisé les termes de *groupes*, d'*anneaux*, de *corps*, d'*homomorphisme*, etc. et continuerons par la suite à le faire encore de nombreuses fois. Outre le fait que ces concepts soient d'une extrême importance, permettant de faire des démonstrations ou de construire des concepts mathématiques indispensables à l'étude de la physique théorique contemporaine (physique quantique des champs, théories des cordes, modèle standard,...), ils permettent de comprendre les composants et les propriétés de base de la mathématique et de ses opérateurs en rangeant ceux-ci par catégories distinctes. Ainsi, choisir de mettre la théorie des ensembles en tant que cinquième chapitre de ce site est un choix tout à fait discutable puisque rigoureusement c'est par là que tout commence. Cependant, nous avions besoin d'exposer quand même la théorie de la démonstration ne serait-ce que pour les notations et les méthodes dont il sera fait usage ici.

Par ailleurs, lors de l'enseignement des mathématiques modernes dans le secondaire, on introduisit le langage des ensembles et l'étude préalable des relations binaires pour une approche plus rigoureuse de la notion de fonctions et d'applications et de la mathématique en générale.

Définition 4.1 *Nous parlons de diagramme sagittal*[1] *pour tout schéma représentant une correspondance entre les composantes de deux ensembles reliés totalement ou partiellement par un ensemble de flèches.*

Exemple 4.1 *La représentation graphique d'une fonction définie de l'ensemble* $E = \{-3, -2, -1, 0, 1, 2, 3\}$ *vers l'ensemble* $F = \{0, 1, 2, 3, \ldots 9\}$ *conduirait au diagramme sagittal ci-dessous :*

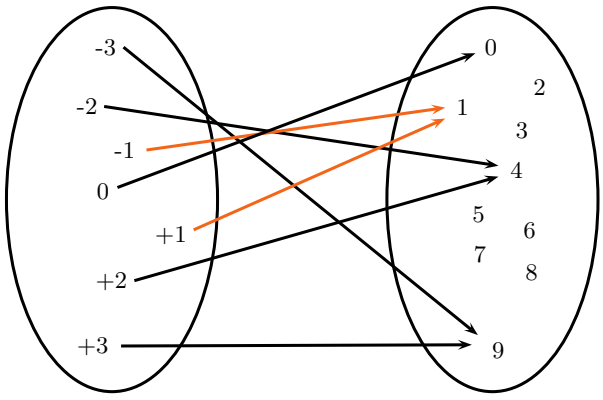


FIGURE 4.1 – Diagramme sagittal de $E \to F$

Le bouclage de chaque élément montrant une *relation réflexive* et la présence systématique d'une flèche retour indiquant une *relation symétrique*.

Cependant le choix d'introduire la théorie des ensembles dans les classes d'école a une raison aussi un peu autre. Au fait, dans un souci de rigueur interne (in extenso : non liées à la réalité), une

1. On parle aussi de schéma sagittal du terme latin sagitta = flèche.

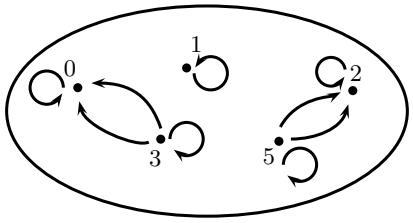


FIGURE 4.2 – Diagramme sagittal de $E \to E$

très grande partie des mathématiques a été reconstruite à l'intérieur d'un seul cadre axiomatique, dénommé donc *théorie des ensembles*, dans le sens où chaque concept mathématique (autrefois indépendant des autres) est ramené à une définition dont tous les constituants logiques proviennent de ce même cadre : elle est considérée comme fondamentale. Ainsi, la rigueur d'un raisonnement effectué au sein de la théorie des ensembles est garantie par le fait que le cadre est *non-contradictoire* ou *consistant*. Voyons les définitions qui construisent ce cadre.

Définition 4.2 *Nous appelons* ensemble *toute liste, collection ou rassemblement d'objets bien définis, explicitement ou implicitement.*

Définition 4.3 *D2. Un* univers U *est un objet dont les constituants sont des ensembles.*

Il faut noter que ce que les mathématiciens appellent *univers* n'est pas un ensemble ! En fait il s'agit d'un modèle qui satisfait aux axiomes des ensembles.

Effectivement, nous verrons que nous ne pouvons pas parler de l'ensemble de tous les ensembles (ce n'est pas un ensemble), pour désigner l'objet qui est constitué de tous les ensembles ainsi, nous parlons d'univers.

Définition 4.4 *Nous appelons* éléments *ou* membres de l'ensemble *les objets appartenant à l'ensemble et nous notons $p \in A$ si p est un élément de l'ensemble A et dans le cas contraire $p \notin A$.*

Si B est une *partie* de A, ou sous-ensemble de A, nous notons cela $B \subset A$ ou $A \supset B$ dès lors que $\forall x, x \in B \Rightarrow x \in A$.

Nous identifions également un ensemble soit en listant ses éléments, soit en donnant la définition de ses éléments.

Exemple 4.2

1. $A = \{1, 2, 3\}$
2. $X = \{x | x \in \mathbb{N}^*\}$

Définition 4.5 *D3. Nous pouvons munir les ensembles d'un certain nombre de relations qui permettent de comparer ses éléments ou de comparer certaines de leurs propriétés. Ces relations sont appelées* relations de comparaisons *ou* relations d'ordre•.

Chap3

REMARQUES

1. La structure d'ensemble ordonnée a été mise en place à la base dans le cadre de la théorie des Nombres par CANTOR et DEDEKIND.
2. Comme nous l'avons démontré dans le chapitre sur les Opérateurs, $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ sont totalement ordonnées par les relations usuelles $\leqslant, \geqslant$. La relation $<$, souvent dite *d'ordre strict*, n'est pas une relation d'ordre car non réflexive et non antisymétrique•. Par exemple, dans $\mathbb{N}$, la relation *a divise b*, souvent notée par le symbole "/", est un ordre partiel.

Chap3

3. Si R est un ordre sur E et F est une partie de E, la restriction à F de la relation R est un ordre sur F, dit *ordre induit par R dans F*.
4. Si R est un ordre sur E, la relation R' définie par $xRy' \Leftrightarrow yRx$ est un ordre sur E, dit *ordre réciproque* de R. L'ordre réciproque de l'ordre usuel $\leqslant$ est l'ordre noté $\geqslant$ ainsi que l'ordre réciproque de l'ordre *a divise b* dans $\mathbb{N}$ est l'ordre *b est multiple de a*.

L'ensemble est l'être mathématique de base, dont l'existence est posée : il n'est pas défini en tant que tel, mais par ses propriétés, données par les axiomes. Il fait appel à une procédure humaine : une sorte de fonction de catégorisation, qui permet à la pensée de distinguer plusieurs éléments qualifiés d'indépendants.

Nous pouvons démontrer à partir des ces concepts, que le nombre de sous-ensembles d'un ensemble de cardinal n (nombre d'éléments) est 2^n :

Il y a d'abord l'ensemble vide $\varnothing$, soit 0 éléments choisi par n, in extenso C_n^0 et ainsi de suite.

Le nombre de sous-ensembles de E correspond donc à la sommation de tous les coefficients binomiaux :

$$\text{Card}(P(E)) = \sum_{k=0}^{n} C_k^n \tag{4.1}$$

Or, nous avons (cf. chapitre de Calcul Algébrique) :

$$(x+y)^n = \sum_{k=0}^{n} C_k^n x + y^{n-k} \tag{4.2}$$

Donc :

$$\text{Card}(P(E)) = \sum_{k=0}^{n} C_k^n 1^k 1^{n-k} = (1+1)^n = 2^n \tag{4.3}$$

Exemple 4.3 *Considérons l'ensemble $S = \{x_1, x_2, x_3\}$. L'ensemble des parties $P(S)$ est constitué par :*

- *ensemble vide :* $\{\} = \varnothing$
- *singletons :* $\{x_1\}, \{x_2\}, \{x_3\}$
- *duets :* $\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_2\}$
- *lui-même :* $\{x_1, x_2, x_3\}$

REMARQUE L'ordre dans lequel sont différenciés les éléments ne rentre pas en compte lors du comptage des parties de l'ensemble de départ.

En mathématique appliquée, nous travaillerons presque exclusivement avec des ensembles de nombres. Nous nous se restreindrons donc à l'étude des définitions et propriétés de ces derniers.

4.1 Axiomatique de ZERMELO-FRAENKEL

L'axiomatique de ZERMELO-FRAENKEL, abrégée *axiomatique ZF*, a été formulée par E. ZERMELO (1908) puis précisée par A. FRAENKEL (1922). Elle est considérée comme la plus naturelle et intuitive dans le cadre de la théorie des ensembles[2].

Strictement parlant, les axiomes de ZFC sont simplement des énoncés du calcul des prédicats du premier ordre• égalitaire dans un langage ayant un seul symbole primitif pour l'appartenance (relation binaire). Ce qui suit doit donc seulement être perçu comme une tentative d'exprimer en français la signification attendue de ces axiomes. Chap1

Axiome 4.1 (Axiome d'extensionalité) *Deux ensembles sont égaux si, et seulement si ils ont les mêmes éléments. C'est ce que nous notons :*

$$A = B \Leftrightarrow (\forall x \in A, x \in B) \wedge (\forall x \in B, x \in A) \tag{4.4}$$

ou :

$$\forall A \forall B \quad (\forall x, x \in A \Leftrightarrow x \in B) \Rightarrow A = B \tag{4.5}$$

Cet axiome exprime seulement le fait qu'un ensemble ne contient rien d'autre que ce qui est spécifié par la donnée ses éléments.

L'unicité de certains ensembles est démontrée en utilisant conjointement l'axiome de sélection et l'axiome d'extensionalité.

2. Il existe bien d'autres axiomatiques, basées sur le concept plus général de *classe*, comme celle développée par VON NEUMANN, BERNAYS et GÖDEL•. Chap1

Axiome 4.2 (Axiome de l'ensemble vide) *L'ensemble vide existe, il n'a aucun élément, son cardinal est noté* 0. *Si* X *est un objet* $\{X\}$ *est un ensemble appelé singleton (single = seul), son cardinal est* 1.

Axiome 4.3 (Axiome de la paire) *Si* A *et* B *sont deux ensembles, alors, il existe un ensemble* C *contenant* A *et* B *et eux seuls comme éléments. Cet ensemble se note* $\{A, B\}$ *:*

$$\forall A \forall B \exists C \quad (\forall X, X \in C \Leftrightarrow (X = A \vee X = B)) \tag{4.6}$$

Axiome 4.4 (Axiome de l'union[3]) *Soient* A *et* B *deux ensembles. Il existe un ensemble* C *dont les éléments sont exactement ceux qui appartiennent à* A *et à* B. *Nous notons cela :*

Axiome 4.5 (Axiome de la somme) *A tout ensemble* A, *nous pouvons associer un ensemble de l'univers* B *qui est la réunion des éléments* x *du premier, in extenso dont les éléments sont exactement les éléments des éléments du premier.*

$$\forall A \exists B \quad (\forall x, x \in B \Leftrightarrow \exists C, x \in C \vee C \in A) \tag{4.7}$$

L'ensemble B *est noté* $\cup A$ *ou bien* $\cup_{x \in A} x$.

Axiome 4.6 (Axiome de parties[4]) *À tout ensemble* A, *nous pouvons associer un ensemble de l'univers* B *qui contient exactement les parties (in extenso les sous-ensembles)* C *du premier :*

$$\forall A \exists B \quad (\forall C, C \in B \Leftrightarrow \exists C \Leftrightarrow C \subset A) \tag{4.8}$$

Axiome 4.7 (Axiome de l'infini) *Il existe un ensemble, dit* ensemble autosuccesseur *K contenant $\varnothing$ (l'ensemble vide) tel que si x appartient à K, alors $x \cup \{x\}$ appartient également à K :*

$$K \text{ est autosuccesseur } \Leftrightarrow (\varnothing \in K) \wedge (x \in K \Rightarrow x \cup \{x\} \in K) \tag{4.9}$$

Cet ensemble permet d'utiliser des ensembles infinis. $\mathbb{N}$ est ainsi le plus petit ensemble autosuccesseur, au sens de l'inclusion $\mathbb{N} = \{\varnothing, \{\varnothing, \{\varnothing, \ldots\}\}\}$ et par convention nous notons (ou nous construisons l'ensemble des naturels) :

$$0 = \varnothing, \quad 1 = \{\varnothing\}, \quad 2 = \{\varnothing, \{\varnothing\}\} \tag{4.10}$$

Axiome 4.8 (Axiome de régularité[5]) *Pour tout ensemble non vide* A, *il existe un ensemble* B, *élément de* A *tel qu'aucun élément de* A *ne soit élément de* B[6] *ce que nous notons :*

$$A \neq \varnothing \Rightarrow \exists B \in A, A \cap B = \varnothing \tag{4.11}$$

Le but principal de cet axiome est d'éliminer la possibilité d'avoir A comme élément de lui-même. En conséquence $\forall A$, $A \notin A$.

Axiome 4.9 (Axiome de l'ensemble des parties)

Preuve En effet, soit A un ensemble tel que $A \in A$. Considérons le singleton $\{A\}$, ensemble dont le seul élément est A. D'après l'axiome de fondation, nous devons avoir un élément de ce singleton qui n'a aucun élément en commun avec lui. Mais le seul élément possible est A lui-même, c'est-à-dire que nous devons avoir :

$$A \cap \{A\} = \varnothing \tag{4.12}$$

Or par hypothèse, $A \in A$ et par construction $A \in \{A\}$. Donc $A \in A \cap \{A\}$, ce qui contredit l'assertion précédente, donc $A \notin A$. •

Axiome 4.10 (Axiome de remplacement) *Quel que soit l'ensemble* A *d'éléments* a *et la relation binaire* f, *il existe un ensemble* B *constitué des éléments* b *tel que* $f(a, b)$ *soit vraie. Si* f *est une fonction, alors* $b = f(a)$ *et* $B = f(A)$.

6. Il faut bien différencier le niveau du langage utilisé, un ensemble et ses éléments n'ont pas le même statut.

Axiome 4.11 (Axiome de sélection (ou de compréhension)) *À tout ensemble A et toute condition ou proposition $S(x)$, il correspond un ensemble B dont les éléments sont exactement les éléments x de A pour lesquels $S(x)$ est vraie. C'est ce que nous notons :*

$$B = \{x \in A : S(x)\} \tag{4.13}$$

Cet axiome est primordial : le respect de ses conditions très strictes d'application permet d'éliminer les paradoxes de la *théorie naïve des ensembles*, comme le paradoxe de RUSSELL ou le paradoxe de CANTOR qui ont invalidé la théorie naïve des ensembles.

Considérons par exemple l'ensemble R de RUSSELL de tous les ensembles qui ne s'auto-contiennent pas (notez bien que nous donnons une propriété de R sans expliciter quel est cet ensemble) $R = \{E : E \notin E\}$. Le problème est de savoir si R se contient ou non. Si $R \in R$, alors, R s'auto-contient, et, par définition $R \notin R$ et inversement. Chaque possibilité est donc contradictoire.

Si maintenant nous désignons par C l'ensemble de tous les ensembles (l'Universel de CANTOR), nous avons en particulier $P(C) \in C$, ce qui est impossible (*i.e.* par exemple avec la puissance du continu de l'ensemble de réels), d'après le théorème de CANTOR.

Ces *paradoxes* (ou antinomies syntaxiques) proviennent d'un non respect des conditions d'application de l'axiome de sélection : pour définir E (dans l'exemple de RUSSELL), il doit exister une proposition $S(x)$ qui porte sur l'ensemble R, qui doit être explicité. La proposition définissant l'ensemble de RUSSELL ou celui de CANTOR n'indique pas quel est l'ensemble E ; elle est donc invalide.

Un exemple fort sympathique et fort connu (c'est la raison pour laquelle nous le présentons) permet de mieux comprendre (il s'agit du paradoxe de RUSSELL) :

Un jeune étudiant se rendit un jour chez son barbier. Il engagea la conversation et lui demanda s'il avait de nombreux concurrents dans sa jolie cité. De manière apparemment innocente, le barbier lui répondit : « Je n'ai aucune concurrence. En effet, de tous les hommes de la cité, je ne rase évidemment pas ceux qui se rasent eux-mêmes, mais j'ai le bonheur de raser tous ceux qui ne se rasent pas eux-mêmes. »

En quoi donc, une telle affirmation si simple peut-elle mettre en défaut la logique de notre jeune étudiant si malin ? La réponse est en effet innocente, jusqu'au moment ou nous décidons de l'appliquer au cas du barbier : se rase-t-il lui-même, oui ou non ?

Supposons qu'il se rase lui-même : il entre dans la catégorie de ceux qui se rasent eux-mêmes, dont le barbier a précisé qu'il ne les rasait évidemment pas. Donc il ne rase pas lui-même. Supposons alors qu'il ne se rase pas lui-même : il entre alors dans la catégorie de ceux qui ne se rasent pas eux-mêmes, dont le barbier a précisé qu'il les rasait tous. Donc il se rase lui-même. Finalement, ce malheureux barbier est dans une position étrange : s'il se rase lui-même, il ne se rase pas, et s'il ne se rase pas lui-même, il se rase. Cette logique est autodestructrice, stupidement contradictoire, rationnellement irrationnelle.

Vient alors l'axiome de sélection : nous excluons le barbier de l'ensemble des personnes auxquelles s'applique la déclaration. Car en réalité, le problème vient du fait que le barbier est un élément de l'ensemble de tous les hommes de la cité. Ainsi, ce qui s'applique à tous les hommes ne s'applique pas au cas individuel du barbier.

Axiome 4.12 (Axiome du choix) *Étant donné un ensemble A d'ensembles non vides mutuellement disjoints, il existe un ensemble B —l'ensemble de choix pour A— contenant exactement un élément pour chaque membre de A.*

Complétons le tout par une note de culture générale : la théorie des ensembles basée sur les axiomes 1 à 7 est dite de ZERMELO (Z). Complétée par l'axiome 8, nous parlons de la théorie ZERMELO-FRAENKEL ou, plus simplement, la théorie ZF. Si nous lui ajoutons l'axiome du choix, elle est dite *axiomatique ZFC* (C comme choix - pour l'axiome du même nom).

Indiquons cependant que la question de l'axiomatisation et donc des fondements se trouva quand même ébranlée de deux questions à l'époque de leur construction : quels axiomes valides doivent être choisis et dans un système d'axiomes les mathématiques sont-elles cohérentes ?

La première question fut soulevée d'abord par l'hypothèse du continu : si nous pouvons mettre deux ensembles de nombres en correspondance terme à terme, ils ont le même nombre d'éléments (cardinal). Nous pouvons mettre en correspondance les entiers avec les rationnels comme nous l'avons démontré dans le chapitre sur les Nombres, ils ont donc même cardinal, nous ne pouvons

par contre mettre en correspondance les entiers avec les réels. La question est alors de savoir s'il y a un ensemble dont le nombre d'éléments serait situé entre les deux ou pas ? Cette question est importante pour construire la théorie classique de l'analyse et les mathématiciens choisissent en général de dire qu'il n'y en a pas, mais nous pouvons aussi dire le contraire.

En fait l'hypothèse du continu est liée de manière plus profonde à l'axiome du choix qui peut aussi être formulé de la manière suivante : si C est une collection d'ensembles non vides alors nous pouvons choisir un élément de chaque ensemble de la collection. Si C a un nombre fini d'éléments ou un nombre dénombrable d'éléments, l'axiome semble assez évident : nous pouvons ranger les ensembles de C en les numérotant, et le choix d'un élément dans chaque ensemble est simple. Là où ça se complique c'est lorsque l'ensemble C a la puissance du continu : comment choisir des éléments s'il n'y pas la possibilité de les numéroter ?

Finalement en 1938, GÖDEL montre que la théorie des ensembles est cohérente sans l'axiome du choix et sans l'hypothèse du continu aussi bien qu'avec ! Et pour clore tout ça P. COHEN montre en 1963 que l'axiome du choix et l'hypothèse du continu ne sont pas liés.

4.1.1 Cardinaux

Définition 4.6 *Des ensembles sont dits* équipotents *s'il existe une bijection entre ces ensembles. Nous disons qu'ils ont alors même* cardinal.

Ainsi, plus rigoureusement, un cardinal, qui quantifie le nombre d'éléments contenus dans l'en-
Chap3 semble, est une classe d'équivalence• pour la relation d'équipotence.

REMARQUE CANTOR est le principal créateur de la théorie des ensembles. Mais, à côté de considérations élémentaires, sa théorie comportait des niveaux d'abstraction élevés. La vraie nouveauté de la théorie de CANTOR, c'est qu'elle permet de parler de l'infini. Par exemple, une idée importante de CANTOR a justement été de définir l'équipotence.

Si nous écrivons $c_1 = c_2$ en tant qu'égalité de cardinaux, nous entendons alors par là qu'il existe deux ensembles équipotents A et B tels que :

$$c_1 = \text{Card}(A) \quad \text{et} \quad c_2 = \text{Card}(B) \tag{4.14}$$

Chap3 Les cardinaux peuvent donc êtres comparés. L'ordre ainsi défini est une relation d'ordre total• entre les cardinaux. La preuve que la relation d'ordre est totale utilise l'axiome du Choix et la preuve qu'elle soit antisymétrique est connue sous le nom de théorème de CANTOR-BERNSTEIN.

Dire que $c_1 < c_2$ signifie dans un vocabulaire simple que A est équipotent à une partie propre de B, mais B n'est équipotent à aucune partie propre de A. Si Les mathématiciens diraient que le $\text{Card}(A)$ est plus petit ou égal au $\text{Card}(B)$ si il existe une injection de A dans B.

Nous avons vu lors de notre étude des nombres, en particulier des nombres transfinis, qu'un ensemble équipotent (ou en bijection) à $\mathbb{N}$ était dit *ensemble dénombrable*.

Voyons cette notion un petit peu plus dans les détails :

Soit A un ensemble, s'il existe un entier n tel qu'il y ait au moins à chaque élément de A un correspondant dans l'ensemble $\{1, 2, \dots n\}$ nous disons alors que le cardinal de A, noté $\text{Card}(A)$, est de *cardinal fini* et vaut n.

Dans le cas contraire, nous disons que l'ensemble A est de *cardinal infini* et nous posons :

$$\text{Card}(A) = +\infty \tag{4.15}$$

Un ensemble A est donc *dénombrable* s'il existe une bijection entre A et $\mathbb{N}$. Un ensemble de nombre A est *au plus dénombrable* s'il existe une bijection entre A et une partie $\mathbb{N}$. Un ensemble au plus dénombrable est donc soit de cardinal fini, soit dénombrable.

Nous vérifions dès lors les propositions suivantes :

Proposition 4.1 *Une partie d'un ensemble dénombrable est au plus dénombrable.*

Proposition 4.2 *Un ensemble contenant un ensemble non-dénombrable n'est lui aussi pas dénombrable*

Proposition 4.3 *Le produit de deux ensembles dénombrables est dénombrable*

REMARQUE Nous pouvons restreindre un ensemble de nombres par rapport à l'élément nul et aux éléments négatifs ou positifs qu'il contient et dès lors nous notons (exemple pour l'ensemble des réels) :

$$\begin{aligned} &\mathbb{R}_+ = \{x | x \in \mathbb{R} \text{ et } x \geqslant 0\}, \quad \mathbb{R}_- = \{x | x \in \mathbb{R} \text{ et } x \leqslant 0\} \\ &\mathbb{R}^* = \mathbb{R}\backslash\{0\}, \quad \mathbb{R}_+^* = \mathbb{R}_+\backslash\{0\}, \quad \mathbb{R}_-^* = \mathbb{R}_-\backslash\{0\} \end{aligned} \tag{4.16}$$

Ces notions étant analogues pour $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ (l'ensemble des nombres complexe n'état pas ordonné, la deuxième et troisième ligne ne s'y applique pas).

Donc tout sous-ensemble infini de $\mathbb{N}$ est équipotent à $\mathbb{N}$ lui-même. En particulier, il y a autant d'entier naturels pairs que d'entiers naturels quelconques (utiliser la bijection $f(n) = 2n$) de $\mathbb{N}$ vers P, où P désigne l'ensemble des entiers naturels pairs), autant d'entiers relatifs que d'entiers naturels, autant d'entiers relatifs que de nombres rationnels (voir le chapitre traitant des nombres pour les démonstrations).

Nous pouvons donc écrire :

$$\text{Card}(\mathbb{N}) = \text{Card}(\mathbb{Z}) = \text{Card}(\mathbb{Q}) = \aleph_0 \tag{4.17}$$

et plus généralement, toute partie infinie de $\mathbb{Q}$ est dénombrable.

Un résultat important : tout ensemble infini possède donc une partie infinie dénombrable.

Puisque nous avons démontré dans le chapitre traitant des nombres que l'ensemble des réels avait la *puissance du continu* et que l'ensemble des nombres naturels était de cardinal transfini $\aleph_0$, CANTOR souleva la question s'il existait un cardinal transfini entre $\aleph_0$ et le cardinal de $\mathbb{R}$? Autrement dit, existe-il un ensemble infiniment grand qui serait intermédiaire entre l'ensemble des nombres entiers et l'ensemble des réels ?

Le problème se posa en notant bien évidemment $\aleph_0$ le cardinal de $\mathbb{N}$ et $\aleph_1$ le cardinal de $\mathbb{R}$ et en proposant de démontrer ou de contredire que :

$$\aleph_1 = 2^{\aleph_0} \tag{4.18}$$

selon la loi combinatoire qui donne le nombre d'éléments de l'ensemble que l'on peut obtenir à partir de tous les sous-ensembles d'un ensemble.

Le reste de sa vie, CANTOR essaya, en vain, de démontrer ce résultat que l'on nomma *l'hypothèse du continu*. Il n'y réussit pas et sombra dans la folie. En 1900, au congrès international des mathématiciens, Hilbert estima qu'il s'agissait là d'un des 23 problèmes majeurs qui devraient êtres résolus au 20^e siècle.

Ce problème se résout d'une façon assez étonnante. D'abord, en 1938, un des plus grands logiciens du 20^e siècle, K. GÖDEL, démontra que l'hypothèse de CANTOR n'était pas réfutable, c'est-à-dire qu'on ne pourrait jamais démontrer qu'elle était fausse. Puis en 1963, le mathématicien P. COHEN boucla la boucle. Il démontra qu'on ne pourrait jamais non plus démontrer qu'elle était vraie. Nous ne pouvons conclure à juste raison que CANTOR avait perdu la raison à chercher à démontrer un problème qui ne pouvait pas l'être.

4.1.2 Produit cartésien

Si E et F sont deux ensembles, nous appelons *produit cartésien de E par F* l'ensemble noté $E \times F$ (à ne pas confondre avec le produit vectoriel) formé de tous les couples possibles (e, f) où e est un élément de E et f un élément de F, autrement écrit :

$$E \times F = \{(e, f) | e \in E \wedge f \in F\} \tag{4.19}$$

Nous remarquons facilement que $E \times F$ et $F \times E$ ne sont pas les mêmes ensembles (sauf bien sur si $E = F$).

Nous notons le produit cartésien de E par lui même :

$$E \times E = E^2 \tag{4.20}$$

et nous disons alors E^2 est *l'ensemble des couples d'éléments de E*.

Nous pouvons effectuer le produit cartésien d'une suite d'ensembles $E_1 \times E_2 \times \ldots \times E_n$ et ainsi obtenir l'ensemble des n-uplets $(e_1, e_2, \ldots, e_n)$ où $e_1 \in E_1, e_2 \in E_2, \ldots, e_n \in E_n$.

Dans le cas où tous les ensembles E_i sont identiques à E, le produit cartésien $E_1 \times E_2 \times \ldots \times E_n$ se note bien évidemment E^n. Nous disons alors que E^n est *l'ensemble des n-uplets d'éléments de* E.

Si E et F sont finis alors le produit cartésien $E \times F$ est fini. De plus :

$$\text{Card}(E \times F) = \text{Card}(E) \cdot \text{Card}(F) \tag{4.21}$$

De là, nous voyons que si les ensembles $E_1, E_2, \ldots, E_n$ sont finis alors le produit cartésien $E_1 \times E_2 \times \ldots \times E_n$ est aussi fini et nous avons :

$$\text{Card}(E_1 \times E_2 \times \ldots \times E_n) = \prod_{i=1}^{n} \text{Card}(E_i) \tag{4.22}$$

En particulier, $\text{Card}(E^n) = [\text{Card}(E)]^n$ si E est un ensemble fini.

Exemple 4.4 *Si $\mathbb{R}$ est l'ensemble des nombres réels, $\mathbb{R}^2$ est alors l'ensemble des couples de réels. Dans le plan rapporté à un repère, tout point M admet des coordonnées qui sont un élément de $\mathbb{R}^2$.*

Exemple 4.5 *Lorsque nous lançons deux dés dont les faces sont numérotées de 1 à 6, chaque dé peut être symbolisé par l'ensemble $E = \{1, 2, 3, 4, 5, 6\}$. Le résultat d'un lancer est alors un élément de $E \times E$. Le cardinal de $E \times E$ est alors 36. Il y a donc 36 résultats possibles quand nous lançons deux dés dont les faces sont numérotées de 1 à 6.*

REMARQUE La théorie de base des ensembles ainsi que le concept de cardinal sont à la base théorique des logiciels de bases de données relationnelles.

4.1.3 Bornes

Soit M un ensemble de nombres quelconques de façon à ce que $M \subset \mathbb{R}$ (exemple particulier mais fréquent) nous avons comme définitions :

Définition 4.7 *$x \in \mathbb{R}$ est appelé* borne supérieure *ou* majorant *de l'ensemble M, si $x \geqslant m$ pour $\forall m \in M$. Inversement, nous parlons de* borne inférieure *ou de* minorant *(il ne faut donc pas confondre le concept de borne avec le concept d'intervalle !).*

Définition 4.8 *Soit $M \subset \mathbb{R}, M \neq \varnothing$, $x \in \mathbb{R}$ est appelé* plus petite borne supérieure *noté $x = \sup M$ de M si x est une borne supérieure de M et si pour toute borne supérieure $y \in \mathbb{R}$ nous avons $x \leqslant y$. Inversement, nous parlons de* plus petite borne inférieure *que nous notons $x = \inf M$.*

Les définitions sont équivalentes dans le cadre de l'analyse fonctionnelle (cf. chapitre du même nom) puisque les fonctions sont définies sur des ensembles.

Effectivement, soit f une fonction dont le domaine de définition I balaie tout $\mathbb{R}$. Ce que nous notons $f : E \to \mathbb{R}$ et soit $x_0 \in \mathbb{R}$.

Définition 4.9 *Nous disons que f présente un* maximum global *en x_0 si :*

$$\forall x \in E, \quad f(x) \leqslant f(x_0) \tag{4.23}$$

Définition 4.10 *Nous disons que f présente un* minimum global *en x_0 si :*

$$\forall x \in E, \quad f(x) \geqslant f(x_0) \tag{4.24}$$

Dans l'un de ces deux cas, nous disons que f présente un *extremum global* en x_0 (c'est un concept que nous retrouverons souvent en mécanique analytique !).

Définition 4.11 *f est* majorée *s'il existe un réel M tel que $\forall x \in I, f(x) \leqslant M$.*

Dans ce cas, la fonction possède une borne supérieure de f sur son domaine de définition I notée traditionnellement :

$$\sup_I f \tag{4.25}$$

et elle représente donc la plus petite borne supérieure (le plus petit majorant).

Définition 4.12 *f est* minorée *s'il existe un réel M tel que $\forall x \in I, f(x) \geqslant M$.*

Dans ce cas, la fonction possède une borne inférieure de f sur son domaine de définition I notée traditionnellement :

$$\inf_I f \tag{4.26}$$

et elle représente la plus grande borne inférieure (le plus grand minorant).

Définition 4.13 *Nous disons que f est* bornée *si elle est à la fois majorée et minorée*[7].

4.2 Opérations ensemblistes

Nous pouvons construire à partir d'au moins trois ensembles A, B, C, l'ensemble des opérations, dont nous devons les notations à DEDEKIND) existant dans la théorie des ensembles.

REMARQUE Certaines des notations présentes ci-dessous se retrouveront fréquemment dans des théorèmes complexes, il est donc nécessaire de bien comprendre de quoi il en retourne.

Ainsi, nous pouvons construire les opérations ensemblistes définies dans ce qui suit.

4.2.1 Inclusions

Dans le cas le plus simple, nous définissons *l'inclusion* par :

$$A \subset B \Leftrightarrow \forall x[x \in A \Rightarrow x \in B] \tag{4.27}$$

En langage non spécialisé voici qu'il faut lire : A est inclus[8] dans B alors pour tout x appartenant à A chacun des ces x appartient aussi à B. De ceci il en découle les propriétés suivantes :

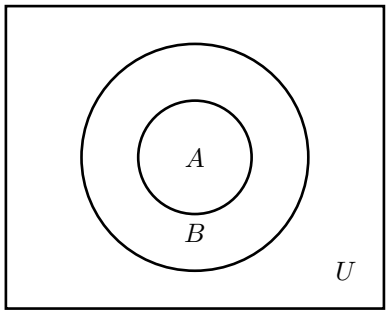


FIGURE 4.3 – $A \subset B$

Propriété 4.1 *P1. Si $A \subset B$ et $B \subset A$ alors cela implique $A = B$ et réciproquement.*

Propriété 4.2 *P2. Si $A \subset B$ et $B \subset C$ alors cela implique $A \subset C$.*

4.2.2 Intersection

Dans le cas le plus simple, nous avons :

$$A \cap B = \{x | x \in A \text{ et } x \in B\} \tag{4.28}$$

En langage non spécialisé voici qu'il faut lire : *L'intersection* des ensembles A et B consiste en l'ensemble des éléments qui se trouvent à la fois dans A et dans B. Plus généralement, si (A_i) est une famille d'ensembles indexés par $i \in I$, l'intersection des $(A_i)_{i \in I}$ est notée $\cap_{i \in I} A_i$ et est définie explicitement par :

$$\bigcap_{i \in I} A_i = \{x | \forall i \in I, x \in A_i\} \tag{4.29}$$

7. À titre d'exemple, mentionnons ici les fonctions trigonométriques.
8. On dit aussi que A *fait partie* ou encore est un sous-ensemble de B.

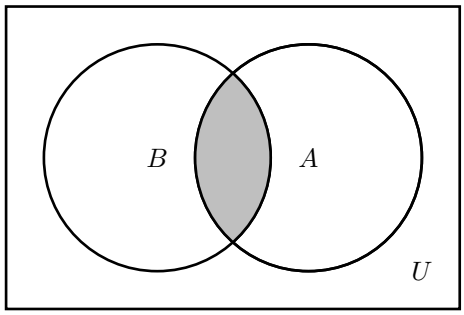


FIGURE 4.4 – $A \cap B$

C'est-à-dire que l'intersection de la famille d'ensembles indexés comprend tous les x qui se trouvent dans chaque ensemble de tous les ensembles de la famille.

Soit deux ensembles A et B, nous disons qu'ils sont *disjoints* si et seulement si :

$$A \cap B = \varnothing \tag{4.30}$$

Par ailleurs, si :

$$E \cap F = \varnothing \Rightarrow \text{Card}(E \cup F) = \text{Card}(E) + \text{Card}(F) \tag{4.31}$$

Les mathématiciens notent cela :

$$E \sqcup F \tag{4.32}$$

et l'appellent *union disjointe*.

Définition 4.14 *Une collection $S = \{S_i\}$ d'ensembles non vides forment une* partition *d'un ensemble A si les propriétés suivantes sont vérifiées :*

1. $\forall S_i, S_j \in S$ *et* $i \neq j \Rightarrow S_i \cap S_j = \varnothing$
2. $A = \cup_{S_i \in S} S_i$

Exemple 4.6 *L'ensemble des nombres pairs et l'ensemble des nombres impairs forment une partition de $\mathbb{Z}$.*

Exemple 4.7 *E2. La loi d'intersection $\cap$ est une loi commutative (voir plus loin la définition du concept de* loi*) telle que :*

$$A \cap B = B \cap A \tag{4.33}$$

4.2.3 Réunion/union

Dans le cas le plus simple, nous avons :

$$A \cup B = \{x | x \in A \text{ ou } x \in B\} \tag{4.34}$$

En langage non spécialisé voici ce qu'il faut lire : La *réunion* ou *union* des ensembles A et B consiste en l'ensemble des éléments qui se trouvent dans A et en plus dans B. Plus généralement, si (A_i) est une famille d'ensembles indexés par $i \in I$, l'union des $(A_i)_{i \in I}$ est notée $\cup_{i \in I} A_i$. Cette réunion est définie par :

$$\bigcup_{i \in I} A_i = \{x | \exists i \in I, x \in A_i\} \tag{4.35}$$

C'est-à-dire que la réunion de la famille d'ensembles indexés comprend tous les x pour lesquels il existe un ensemble indexé par i tel que x soit inclus dans cet ensemble A_i.

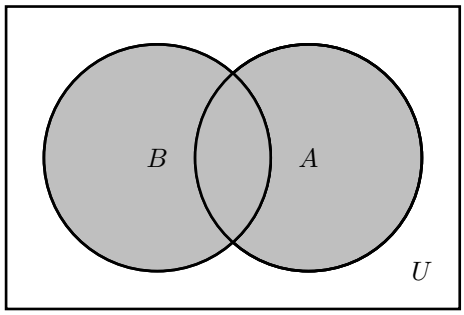


FIGURE 4.5 – $A \cup B$

Nous avons les propriétés de distributivité suivantes :

$$\begin{aligned} \left(\bigcup_{i\in I} A_i\right) \cap B &= \bigcup_{i\in I}(A_i \cap B) \\ \left(\bigcap_{i\in I} A_i\right) \cup B &= \bigcap_{i\in I}(A_i \cup B) \end{aligned} \tag{4.36}$$

La loi de réunion $\cup$ est une loi commutative (voir plus loin la définition du concept de *loi*) telle que :

$$A \cup B = B \cup A \tag{4.37}$$

Nous appelons par ailleurs *lois d'idempotences* les relations (précisons cela pour la culture générale) :

$$\begin{aligned} A \cup A &= A \\ A \cap A &= A \end{aligned} \tag{4.38}$$

et *lois d'absorptions* les lois :

$$\begin{aligned} A \cap (A \cup B) &= A \\ A \cup (A \cap B) &= A \end{aligned} \tag{4.39}$$

Les lois de réunion et d'intersection sont associatives telles que :

$$\begin{aligned} A \cap (B \cap C) &= (A \cap B) \cap C \\ A \cup (B \cup C) &= (A \cup B) \cup C \end{aligned} \tag{4.40}$$

et distributives telles que :

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned} \tag{4.41}$$

4.2.4 Différence

Dans le cas le plus simple, nous avons :

$$A/B = \{x | x \in A \wedge x \notin B\} \tag{4.42}$$

En langage non spécialisé voici ce qu'il faut lire : La *différence* des ensembles A et B consiste en l'ensemble des éléments qui se trouvent uniquement dans A (et qui excluent donc les éléments de B). Si nous nous rappelons du concept de *cardinal* (voir plus haut), nous avons avec les opérations précédemment définies, la relation suivante :

$$\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B) \tag{4.43}$$

d'où si $A \cap B = \varnothing$:

$$\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) \tag{4.44}$$

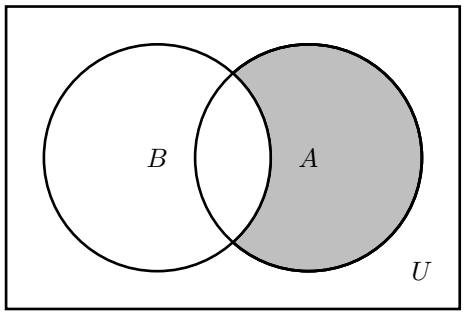


FIGURE 4.6 – A/B

4.2.5 Différence symétrique

Soit E un ensemble. Pour tout $A, B \subseteq E$ nous définissons la différence symétrique $A \triangle B$ entre A et B par :

$$A \triangle B = (A/B) \cup (B/A) \tag{4.45}$$

La *différence symétrique* des ensembles A et B consiste en l'ensemble des éléments qui se trouvent uniquement dans A et de ceux se trouvant uniquement dans B (nous laissons donc de côté les éléments qui sont communs). Les propriétés triviales sont les suivantes :

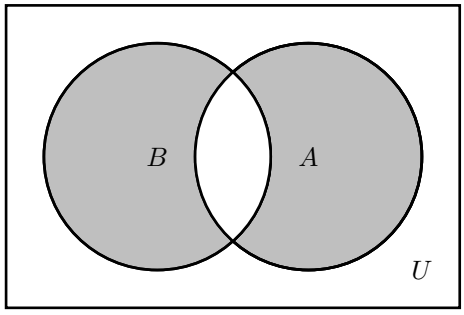


FIGURE 4.7 – $A \triangle B$

1. $A \triangle B = B \triangle A$
2. $A^c \triangle B^c = A \triangle B$
3. $A \triangle B = (A \cup B)/(A \cap B)$

4.2.6 Produit

Dans le cas le plus simple, nous avons :

$$A \times B \tag{4.46}$$

l'ensemble produit[9] de deux ensembles A et B est l'ensemble des couples tels que :

$$(x, y), x \in A, y \in B \tag{4.47}$$

L'ensemble produit des réels $\mathbb{R} \times \mathbb{R}$ par exemple forme le plan où chaque élément est défini par une abscisse et son ordonnée.

4.2.7 Complémentarité

Dans le cas le plus simple, nous avons :

$$\forall A \subset B, \quad \bar{A} = \{x | x \in B, x \notin A\} \tag{4.48}$$

9. Attention à ne pas le confondre avec la multiplication ou le produit vectoriel.

En langage non spécialisé voici que qu'il faut lire : Le *complémentaire* est défini comme en prenant B un ensemble et A un sous-ensemble de B alors le complémentaire de A dans B est l'ensemble des éléments qui sont dans B mais pas dans A.

Par exemple, dans la figure 4.8 nous avons le complémentaire de A par rapport à U qui est indiqué en gris (s'il est seul il s'agit donc de l'univers seul qui l'entoure). Nous avons comme

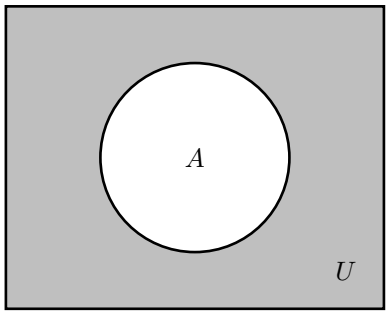


FIGURE 4.8 – A^U

propriétés pour tout A_i inclus dans B :

$$\begin{aligned}\left(\bigcup_{i\in I} A_i\right)^c &= \bigcap_{i\in I} A_i^c \\ \left(\bigcap_{i\in I} A_i\right)^c &= \bigcup_{i\in I} A_i^c\end{aligned} \tag{4.49}$$

Voici quelques lois triviales relatives aux compléments :

$$\bar{\bar{A}} = A, \quad A \cap \bar{A} = \varnothing, \quad A \cup \bar{A} = U \tag{4.50}$$

Il existe d'autres lois très importantes en logique booléenne. Si nous considérons trois ensembles A, B, C comme représentés sur la figure 4.9. Nous avons donc :

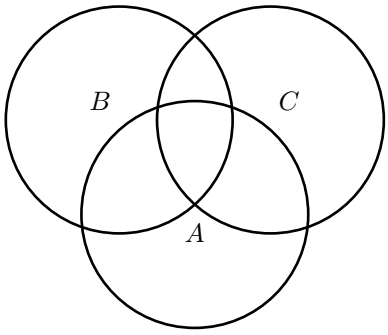


FIGURE 4.9 – Trois ensembles

$$\begin{aligned}A/(B \cap C) &= (A/B) \cup (A/C) \\ A/(B/C) &= (A/B) \cup (A \cap C)\end{aligned} \tag{4.51}$$

et les fameuses *lois de De Morgan* sous forme ensembliste (cf. chapitre de Systèmes Logiques Formels) et qui sont données par les relations :

$$\begin{aligned}\overline{A \cap B} &= \overline{A} \cup \overline{B} \\ \overline{A \cup B} &= \overline{A} \cap \overline{B}\end{aligned} \tag{4.52}$$

4.3 Fonctions et applications

4.3.1 Surjectivité, injectivité, bijectivité et composition

Définition 4.15 *En mathématiques, une* application *(ou* fonction*) notée f ou A est la donnée de deux ensembles, l'ensemble de départ E et l'ensemble d'arrivée F (ou d'image de E), et d'une relation associant à chaque élément x de l'ensemble de départ un et un seul élément de l'ensemble d'arrivée, que nous appelons* image de x par f *et que nous notons $f(x)$.*

Nous appelons *images* les éléments de $f(E)$ et les éléments de E sont appelés les antécédents.

Nous disons alors que f est une application de E dans F notée :

$$f : E \to F \tag{4.53}$$

(se rappeler du premier diagramme sagittal présenté au début de ce chapitre), ou encore une application à arguments dans E et valeurs dans F.

REMARQUE Le terme *fonction* est souvent utilisé pour les applications à valeurs numériques, réelles ou complexes, c'est-à-dire lorsque l'ensemble d'arrivée est $\mathbb{R}$ ou $\mathbb{C}$. Nous parlons alors de *fonction réelle* ou de *fonction complexe.*

Définition 4.16 *D1. Le* graphe *(ou encore* graphique *ou* représentative*) d'une application $f : E \to F$ est le sous-ensemble du produit cartésien $E \times F$ constitué des couples $(x, f(x))$ pour x variant dans E. La donnée du graphe de f détermine son ensemble de départ (par projection sur la première composante souvent notée x) et son image (par projection sur la seconde composante souvent notée y).*

Définition 4.17 *D2. Si le triplet $f(E, F, \Gamma)$ est une fonction où E et F sont deux ensembles et $\Gamma \subset E \times F$ est un graphe et donc E et F sont respectivement la source et le but de f. Le* domaine de définition *ou* ensemble de départ *de f est :*

$$D_f = \{x \in E | \exists y \in \Gamma, (x, y) \in \Gamma\} \tag{4.54}$$

Définition 4.18 *D3. Étant donnés trois ensembles E, F et G (non vides), toute fonction de $E \times F$ vers G est appelée* loi de composition *de $E \times F$ à valeurs dans G.*

Définition 4.19 *D4. Une* loi de composition interne *(ou simplement* loi interne*) dans E est une loi de composition de $E \times E$ à valeurs dans E (cas $E = F = G$).*

REMARQUE La soustraction dans $\mathbb{N}$ n'est pas une loi de composition interne bien qu'elle fasse partie des quatre opérations élémentaires apprises à l'école. Par contre l'addition sur $\mathbb{N}$ en est biens une.

Définition 4.20 *D5. Une* loi composition externe *(ou simplement* loi externe*) dans E est une loi de composition de $F \times E$ à valeurs dans E, où F est un ensemble distinct de E. En général, F est un corps, dit* corps de scalaires

Exemple 4.8 *Dans le cas d'un espace vectoriel (voir définition beaucoup plus bas) la multiplication d'un vecteur (dont les composantes se basent sur un ensemble donné) par un réel est une loi de composition externe.*

REMARQUE Une loi de composition externe à valeurs dans E est aussi appelée *action de F sur E*. L'ensemble F est alors le domaine d'opérateurs. On dit aussi que F opère sur E (ayez en tête l'exemple des vecteurs précédemment cité)

Définition 4.21 *D6. Nous appelons* image de f*, et nous notons* $\mathrm{Im}\,(f)$*, le sous-ensemble défini par :*

$$\mathrm{Im}\,(f) = f(E) = \{y \in F | \exists x \in E, y = f(x)\} \tag{4.55}$$

Ainsi, *L'image* d'une application $f : E \to F$ est la collection des $f(x)$ pour x parcourant E, c'est un sous-ensemble de F.

Et nous appelons *noyau de f*, et nous notons $Ker(f)$, le sous-ensemble très important en mathématiques défini par :

$$\mathrm{Ker}\,(f) = f^{-1}(\{0\}) = \{x \in E, f(x) = 0\} \tag{4.56}$$

Remarques

1. $Ker(f)$ provient de l'allemand *Kern*, signifiant tout simplement *noyau*. En anglais, le noyau se dit aussi *kernel*, signifiant *amande* dans le civil.
2. Normalement les notations Im et Ker sont réservées aux homomorphismes de groupes, d'anneaux, de corps et aux applications linéaires entre espaces vectoriels ou modules etc. (voir plus loin). Nous n'avons normalement pas l'habitude de les utiliser pour des applications quelconques entre ensembles quelconques.

Les applications peuvent avoir une quantité phénoménale de propriétés dont voici celles qui font partie des connaissances générales du physicien (pour plus de renseignements sur ce qu'est une fonction, voir le chapitre traitant de l'Analyse Fonctionnelle).

Soit f une application d'un ensemble E à un ensemble F alors nous avons les propriétés suivantes :

Propriété 4.3 *P1. Une application est dite* surjective *si tout élément y de F est l'image par f d'au moins un élément de E. Nous disons encore que c'est une* surjection *de E dans F. Il découle de cette définition, qu'une application $f : E \to F$ est surjective si et seulement si $F = \mathrm{Im} f$. En d'autres termes, nous écrivons aussi cette définition ainsi :*

$$\forall y \in F, \exists x \in E, \quad y = f(x) \tag{4.57}$$

Cette propriété est illustrée sur la figure 4.10.

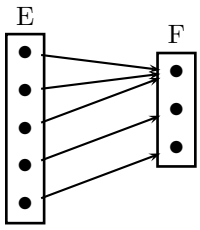


Figure 4.10 – Notion de surjectivité

Propriété 4.4 *P2. Une application est dite* injective *si tout élément y de F est l'image par f d'au plus (nous insistons sur le* au plus*) un seul élément de E. Nous disons encore que f est une injection de E dans F. Il résulte de cette définition, qu'une application $f : E \to F$ est injective si et seulement si les relations $x_1, x_2 \in E$ et $f(x_1) = f(x_2)$ impliquent $x_1 = x_2$ autrement dit : une application pour laquelle deux éléments distincts ont des images distinctes est dite injective. Ou encore, une application est injective si l'une aux moins des propriétés équivalentes suivantes, et illustrées sur la figure 4.11, est vérifiée :*

1. $\forall x, y \in E^2 \quad f(x) = f(y) \Rightarrow x = y$;
2. $\forall x, y \in E^2 \quad x \neq y \Rightarrow f(x) \neq f(y)$;
3. $\forall y \in F$ *l'équation en* x, $y = f(x)$ *a au plus une solution dans* E.

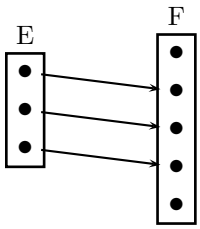


Figure 4.11 – Notion d'injectivité

Propriété 4.5 *P3. Une application est dite* bijective *si une application f de E dans F est à la fois surjective et injective. Dans ce cas, nous avons que pour tout élément y de F de l'équation $y = f(x)$ admet dans E une unique (ni* au plus, *ni* au moins*) pré-image x. Ce que nous écrivons aussi :*

$$\forall y \in F, \exists! x \in E, \quad y = f(x) \tag{4.58}$$

ce qui s'illustre par la figure 4.12.

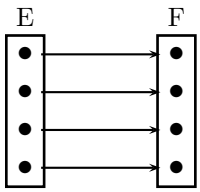


FIGURE 4.12 – Notion de bijectivité

Nous sommes ainsi tout naturellement amené à définir une nouvelle application de F dans E, appelée *fonction réciproque* de f et notée f^{-1}, qui a tout élément y de F, fait correspondre l'élément x de E pré-image (ou solution) unique de l'équation $y = f(x)$. Autrement dit :

$$x = f^{-1}(y) \tag{4.59}$$

L'existence d'une application réciproque implique que le graphique d'une application bijective (dans l'ensemble des réels) et celui de son application réciproque sont symétriques par rapport à la droite d'équation $y = x$.

Effectivement, nous remarquons que si $y = f(x)$ est équivalent à $x = f^{-1}(y)$. Ces équations impliquent que le point (x, y) est sur le graphique de f si et seulement si le point (y, x) est sur le graphique de f^{-1}.

Exemple 4.9 *Prenons le cas d'une station de vacances où un groupe de touristes doit être logé dans un hôtel. Chaque façon de répartir ces touristes dans les chambres de l'hôtel peut être représentée par une application de l'ensemble des touristes vers l'ensemble des chambres (à chaque touriste est associée une chambre).*

- *Les touristes souhaitent que l'application soit injective, c'est-à-dire que chacun d'entre eux ait une chambre individuelle. Cela n'est possible que si le nombre de touristes ne dépasse pas le nombre de chambres.*
- *L'hôtelier souhaite que l'application soit surjective, c'est-à-dire que chaque chambre soit occupée. Cela n'est possible que s'il y a au moins autant de touristes que de chambres.*
- *S'il est possible de répartir les touristes de telle sorte qu'il y en ait un seul par chambre, et que toutes les chambres soient occupées : l'application sera alors à la fois injective et surjective nous dirons qu'elle est bijective.*

REMARQUES

1. Il vient des définitions ci-dessus qu'une application f est bijective (ou *biunivoque*) dans l'ensemble des réels si et seulement si toute droite horizontale coupe la représentation graphique de la fonction en un seul point. Nous pouvons donc amener à faire la seconde remarque suivante :
2. Une application qui vérifie le test de la droite horizontale est continûment croissante ou décroissante en tout point de son domaine de définition.

Propriété 4.6 (fonction composée) *Soit ϕ une application de E dans F et ψ une fonction de F dans G. L'application qui associe à chaque élément x de l'élément de E, $\psi(\phi(x))$ de G s'appelle* application composée *de ϕ et de ψ et se note $\psi \circ \phi$.*

Le symbole $\circ$ est appelé rond. *Ainsi, la relation précédente ce lit* psi rond phi *:*

$$(\psi \circ \phi)(x) = \psi(\phi(x)) \tag{4.60}$$

Soit, de plus, Ξ une application de G dans H. Nous vérifions aussitôt que l'opération de composition est associative :

$$\chi \circ (\psi \circ \phi) = (\chi \circ \psi) \circ \phi \tag{4.61}$$

Cela nous permet d'omettre les parenthèses et d'écrire plus simplement $\chi \circ \psi \circ \phi$. Dans le cas particulier où ϕ serait une application de E dans E, nous notons ϕ^k l'application composée $\phi \circ \phi \circ \ldots \circ \phi$, cela k fois.

Toutes les propriétés définies et énoncées ci-dessus sont applicables aux ensembles de nombres.

4.3.2 Théorème de Cantor-Bernstein

Ce théorème, dont le résultat peut sembler évident, n'est pas forcément simple à aborder. Nous vous conseillons de lire très lentement et de vous imaginer les diagrammes sagittaux dans la tête.

Voici l'hypothèse à démontrer : soient X et Y deux ensembles. S'il existe une injection de X vers Y et une autre de Y vers X, alors les deux ensembles sont en bijection. Il s'agit donc aussi d'une relation antisymétrique.

Ce qui s'illustre par la figure 4.13. Pour la démonstration, nous avons besoin de démontrer

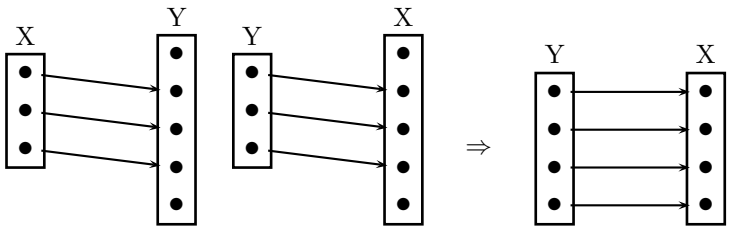


Figure 4.13 – Notion de bijectivité

au préalable un lemme dont l'énoncé est le suivant : soient X, Y, Z trois ensembles tels que $X \subseteq Z \subseteq Y$. Si X et Y sont en bijection, alors X et Z sont en bijection.

Exemple 4.10 *Un exemple d'application de ce lemme est l'ensemble des nombres naturels et des nombres rationnels qui sont en bijection. Dès lors, l'ensemble des entiers relatifs est en bijection avec l'ensemble des nombres naturels puisque :*

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \tag{4.62}$$

Preuve D'abord, au niveau formel, créons une fonction f que telle quelle soit bijective : Nous avons besoin maintenant de définir l'ensemble A par les images de l'union des fonctions des fonctions f (du genre $f(f(f\ldots))))$ des pré-images de l'ensemble Z dont nous excluons les éléments de X (ce que nous notons $Z - X$).

En d'autres termes (si la première forme n'est pas claire...) nous définissons l'ensemble A comme étant l'union des images de $(Z - X)$ par les applications $f \circ f \circ f \ldots$ Ce que nous noterons donc :

$$f : Y \to X \tag{4.63}$$

Nous avons alors :

$$A = \bigcup_{n=1}^{\infty} f^n(Z - X) \tag{4.64}$$

Nous pouvons démontrer élégamment cette dernière relation :

$$f((Z - X) \cup A) = f(Z - X) \cup f(A) = f(Z - X) \cup \bigcup_{n=2}^{\infty} f^n(Z - X) = \bigcup_{n=1}^{\infty} f^n(Z - X) = A \tag{4.65}$$

Comme Z peut être partitionné en $(Z - X) \cup A$ et $(Z - A)$, nous posons comme une définition l'application g telle que :

$$g : Z \to X \tag{4.66}$$

tel que pour toute pré-image a, nous ayons :

$$\forall a \in (Z - X) \cup A \mapsto f(a) \tag{4.67}$$

et :

$$\forall a \in (X - A) \mapsto a \tag{4.68}$$

L'application g est alors bijective car ses restrictions à $(Z-X)\cup A$ et $(X-A)$, (qui forment une partition) sont f et l'identité qui sont par définition bijectives.

Finalement il existe bien, par construction, une bijection entre X et Z.

Reprenons les hypothèses du théorème : soit ϕ une injection de X vers Y et ψ une injection de Y vers X. Nous avons alors :

$$\phi(X) \subset Y \quad \text{et} \quad \psi(Y) \subset X \tag{4.69}$$

donc :

$$\psi(\phi(X)) \subset \psi(Y) \subset X \tag{4.70}$$

Comme ϕ est injective, X et $\phi(X)$ sont par définition en bijection et de même, comme ψ est injective, $\psi(\phi(X))$ et $\phi(X)$ sont en bijection et par conséquent, X et $\psi(\phi(X))$ sont eux aussi en bijection.

En utilisant le lemme sur $\psi(\phi(X))$, $\psi(Y)$ et X, il vient donc que $\psi(\phi(X))$ est en bijection $\psi(Y)$ ce qui nous donne avec ceux que nous avons vu juste précédemment, que puisque aussi $\psi(\phi(X))$ et $\phi(X)$ sont en bijection, alors que $\psi(Y)$ est en bijection avec $\phi(X)$, alors X et Y sont en injection. •

Ce théorème s'interprète de la manière suivante : si nous pouvons compter une partie d'un ensemble avec la totalité des éléments d'un autre ensemble, et réciproquement, alors ils ont le même nombre d'éléments. Ce qui est évident pour des ensembles finis. Ce théorème généralise alors cette notion pour des ensembles infinis.

À partir de là, ce théorème représente l'une des briques de base pour généraliser la notion de tailles d'ensembles à des ensembles infinis.

4.4 Structures algébriques

L'algèbre dite *algèbre moderne* commence avec la théorie des structures algébriques due en partie à C. GAUSS et surtout à É. GALOIS. Ces structures existent en un très grand nombre mais seulement les fondamentales nous intéresseront ici. Avant des les détailler, il faut lire le diagramme synoptique 4.14 des ces principales grandeurs et de leur hiérarchie.

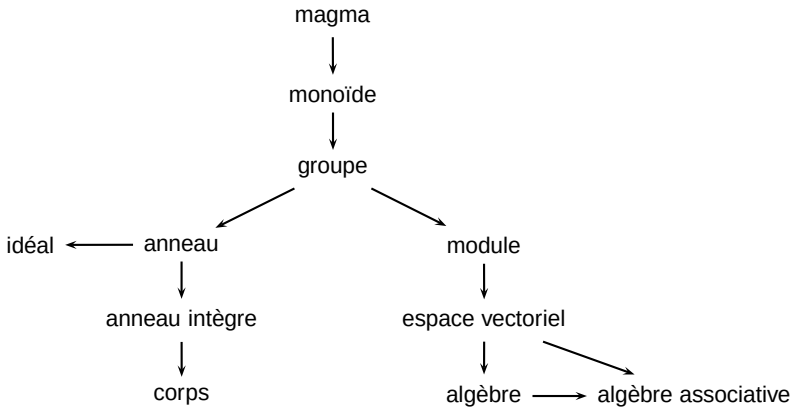


FIGURE 4.14 – Structures algébriques

REMARQUE Tout en haut du diagramme, la structure au nombre minimal de contraintes, en bas, un maximum. Soit, plus nous descendons, plus la structure est en quelque sorte spécialisée.

Soit pour simplifier les écritures, $\star$ une loi de composition.

REMARQUE Cette notation généralisée est parfois appelée *notation stellaire*.

Soient $\star$ et $\circ$ des symboles de lois internes à un ensemble E alors :

Définition 4.22 *$\star$ est une* loi commutative *si :*

$$a \star b = b \star a \tag{4.71}$$

Définition 4.23 *$\star$ est une* loi associative *si :*

$$a \star (b \star c) = (a \star b) \star c \tag{4.72}$$

Définition 4.24 *n est* élément neutre *pour $\star$ si :*

$$a \star n = n \star a = a \tag{4.73}$$

Nous admettrons par ailleurs sans démonstration que s'il existe un élément neutre, il est unique.

Définition 4.25 *a' est* l'élément symétrique [10] *de a pour $\star$ si :*

$$a' \star a = a \star a' = n \tag{4.74}$$

Nous admettrons également et sans démonstration que le symétrique de tout élément est unique.

Définition 4.26 *$\circ$ est une* loi distributive *par rapport à $\star$ si :*

$$\begin{aligned} a \circ (b \star c) &= (a \circ b) \star (a \circ c) \\ (a \star b) \circ c &= (a \circ c) \star (b \circ c) \end{aligned} \tag{4.75}$$

Remarques

1. si a est son propre symétrique par rapport à la loi $\star$, les mathématiciens disent que a est *involutif* ;
2. si un élément b de E vérifie $a \star b = b \star a = b$, alors b est dit *élément absorbant* pour la loi $\star$;
3. il faut toujours vérifier que les neutres et les symétriques le soient *à gauche* et *à droite*. Ainsi, par exemple, dans $(\mathbb{Z}, -)$, l'élément 0 n'est un neutre qu'à droite car $x - 0 = x$ mais $0 - x = -x$.

4.4.1 Magma

Définition 4.27 *Nous désignons un ensemble par le terme* magma *M , si les composants le constituant sont opérables par rapport une loi interne $\star$:*

$$(M, \star) \text{ est un magma si } \begin{cases} \star \text{ est une opération} \\ \star \text{ est une loi interne} \end{cases} \tag{4.76}$$

Remarques

1. si de plus la loi interne $\star$ est commutative, nous parlons de *magma commutatif* ;
2. si de plus la loi interne $\star$ est associative, nous parlons de *magma associatif* ;
3. si de plus la loi interne $\star$ possède un élément neutre, nous parlons de *magma unitaire*.

Il est donc important de se rappeler que si nous désignons une structure algébrique par le terme *magma* tout court, cela ne signifie en aucun cas que la loi interne est commutative, associative ou même qu'elle possède un élément neutre.

Définition 4.28 *Dans un magma $(M, \star)$, un élément x est dit* élément régulier *(ou* élément simplifiable*) à gauche si pour tout couple $(a, b) \in M$ nous avons :*

$$x \star a = x \star b \Rightarrow a = b \tag{4.77}$$

Remarque Nous définissons de même un élément régulier à droite.

Ainsi, un élément est dit *régulier* s'il est régulier à droite et à gauche. Si $\star$ est commutative (ce qui est le cas pour un magma commutatif), les notions d'élément régulier à gauche ou à droite coïncident.

Exemple 4.11 *Dans $(\mathbb{N}, +)$ tout élément est régulier et dans $(\mathbb{N}, \times)$ tout élément non nul est régulier. Un magma $(M, \star)$ est donc une structure algébrique élémentaire. Il existe des structures plus subtiles —monoïdes, groupes, anneaux, corps, espace vectoriels— dans lesquelles un ensemble est muni de plusieurs lois et de différentes propriétés. Nous allons les voir de suite et les utiliser tout au long de ce site.*

10. Dans le sens général de l'opposé par exemple pour l'addition et l'inverse pour la multiplication.

4.4.2 Monoïde

Définition 4.29 *Si la loi $\star$ est associative et possède un élément neutre nous disons alors que le* magma associatif unitaire *est un* monoïde *:*

$$(\ddot{M},\star) \text{ est un monoïde si } \begin{cases} \star \text{ est associative} \\ \star \text{ il existe dans } \ddot{M} \text{ un élément neutre } n \text{ pour } \star \end{cases} \tag{4.78}$$

REMARQUES

1. si de plus la loi interne $\star$ est en plus commutative alors nous disons alors que la structure forme un *monoïde abélien* (ou simplement *monoïde commutatif*) ;
2. dans certains ouvrages nous trouvons aussi comme définition que le monoïde est un demi-groupe avec une loi associative muni d'un élément neutre.

Montrons que l'ensemble des entiers naturels $\mathbb{N}$ est un monoïde abélien totalement ordonné (comme nous l'avons partiellement vu dans le chapitre des opérateurs) par rapport aux lois d'addition et de multiplication :

La loi d'addition $(+)$ est-elle une opération interne telle que $\forall a, b \in \mathbb{N}$ nous ayons :

$$a + b = c \tag{4.79}$$

Nous pouvons démontrer que c'est bien le cas en sachant que 1 appartient à $\mathbb{N}$ tel que :

$$\sum_{i=1}^{a} 1 + \sum_{i=1}^{b} 1 = \sum_{i=1}^{a+b} 1 \tag{4.80}$$

Donc $c \in \mathbb{N}$ et l'addition est bien une loi interne[11] et en même temps associative puisque 1 peut être additionné à lui-même par définition dans n'importe quel ordre sans que le résultat en soit altéré. Si vous vous rappelez que la multiplication est une loi qui se construit sur l'addition, alors la loi de multiplication (x) est aussi une loi interne et associative.

Nous admettrons à partir d'ici qu'il est trivial que la loi d'addition est également commutative et que le zéro "0" en est l'élément neutre (n). Ainsi, la loi de multiplication est elle aussi commutative et il est trivial que "1" en est l'élément neutre (n).

Existe-t-il, en restant dans la lignée de l'exemple précédent pour la loi d'addition $(+)$, un symétrique c tel que $\forall a, b \in \mathbb{N}$ nous ayons :

$$\left(\sum_{i=1}^{a} 1 + \sum_{i=1}^{b} 1\right) + c = \left(\sum_{i=1}^{a+b} 1\right) + c = 0 \tag{4.81}$$

avec $c \in \mathbb{N}$? Il est assez trivial que pour que cette égalité soit satisfaite nous ayons :

$$\left(\sum_{i=1}^{a+b} 1\right) = -c \tag{4.82}$$

soit $a + b = -c$, or les nombres négatifs n'existent pas dans $\mathbb{N}$. Ce qui nous amène aussi à la conclusion que la loi d'addition $(+)$ n'a pas de symétrique et que la loi de soustraction $(-)$ n'existe pas dans $\mathbb{N}$ (la soustraction étant rigoureusement l'addition d'un nombre négatif).

Existe-t-il pour la loi de multiplication $(\times)$ un symétrique a' tel que $\forall a \in \mathbb{N}$ nous ayons :

$$a' \cdot a = n = 1 \tag{4.83}$$

avec $a' \in \mathbb{N}$?

D'abord il est évident que :

$$a' = \frac{1}{a} \tag{4.84}$$

Mais excepté pour $a = 1$, le quotient $1/a$ n'existe pas dans $\mathbb{N}$. Donc nous devons conclure qu'il n'existe pas pour tout élément de $\mathbb{N}$ de symétriques pour la loi de multiplication et ainsi que la loi de division n'existe pas dans $\mathbb{N}$ et qu'elle ne forme pas un monoïde dans cet ensemble. La synthèse est donnée dans le tableau 4.1.

11. Nous disons également que l'ensemble $\mathbb{N}$ est *stable* par rapport à l'addition.

LOIS DE $\mathbb{N}$	$(+)$	$(-)$	$(\times)$	$(/)$
opération interne	oui	non	oui	non
commutative	oui		oui	
élément neutre	oui 0		oui 1	
symétrique	non		non	

TABLEAU 4.1 – Lois dans un monoïde

REMARQUE Le "$-$" signifie que cette propriété n'existe pas dans l'ensemble considéré.

Propriété 4.7 $(\mathbb{N}, \leqslant, \geqslant)$ *est totalement ordonné (attention cette notation est un peu abusive ! il suffit qu'il y ait juste une des deux relations d'ordre R pour que l'ensemble soit totalement ordonné).*

Propriété 4.8 $(\mathbb{N}, +)$ *et* $(\mathbb{N}, \times)$ *sont des monoïdes abéliens.*

Propriété 4.9 *L'élément zéro "0" est l'élément absorbant pour le monoïde* $(\mathbb{N}, \times)$.

Propriété 4.10 *Les lois de soustraction et division n'existent pas dans l'ensemble* $\mathbb{N}$.

Propriété 4.11 $\mathbb{N}$ *est un monoïde abélien totalement ordonné par rapport aux lois d'addition et de multiplication*[12] *:*

$$(\mathbb{N}, +, \times, \leqslant, \geqslant) \tag{4.85}$$

REMARQUES

1. Il est rare d'utiliser les monoïdes ; car souvent, lorsque nous nous trouvons face à une structure trop pauvre pour pouvoir vraiment discuter, nous la prolongeons vers quelque chose de plus riche, comme un groupe, ou un anneau (voir plus loin) tel que l'ensemble des entiers relatifs.
2. Dire qu'une structure algébrique est totalement ordonnée par rapport à certaines lois signifie que soit $\star$ une loi, et R une relation d'ordre et a, b, c, d quatre éléments de la structure intéressée, alors si aRb et cRd implique $(a \star c)R(b \star d)$. Nous notons alors cette structure $(S, \star, R)$ ou simplement (S, R) et en indiquant la (ou les) loi concernée.

4.4.3 Groupes

Définition 4.30 *Nous désignons un ensemble par le terme* groupe, *si les composants le constituant satisfont aux trois conditions de ce que nous nommons la* loi interne de groupe, *définie comme suit :*

$$(G, \star) \text{ est un groupe si } \begin{cases} \text{est associative} \\ \text{il existe dans } G \text{ un élément neutre pour } \star \\ \text{tout élément de } G \text{ possède un symétrique pour } \star \end{cases} \tag{4.86}$$

Dans ce cas, la loi de compositions interne $\star$ sera souvent, mais pas exclusivement, notée "$+$" et appelée l'addition, le neutre e noté 0 et le symétrique de x noté "$-x$".

Insistons sur le fait que la structure de groupe est probablement une des plus importantes dans la pratique de l'ingénieur et de la physique moderne en général.

Si de plus, la loi interne $\star$ est également commutative, nous disons alors que le groupe est un *groupe abélien* ou simplement *groupe commutatif*.

S'il existe dans G au moins un élément a tel que tout élément de G est une puissance de a ou du symétrique a' de a, nous disons que $(G, \star)$ est un *groupe cyclique de générateur* a s'il est fini, sinon nous disons qu'il est *monogène* (nous reviendrons sur les groupes cycliques dans le chapitre d'Algèbre Ensembliste).

Plus généralement un groupe $(G, \star)$ d'élément neutre e, non réduit uniquement à $\{e\}$ sera monogène, s'il existe un élément a de G distinct de e tel que $G = \{e, a^1, a^2, a^3, \ldots, a^n, \ldots\}$. Un tel groupe sera cyclique, s'il existe un entier n non nul pour lequel $a^n = e$. Le plus petit entier non nul vérifiant cette égalité est alors *l'ordre du groupe*.

12. Il faut remarquer que la notation (4.85) est abusive car le monoïde est composé d'une seule loi interne et d'une relation d'ordre R ce qui donnerait au total quatre monoïdes.

Exemple 4.12 *Montrons tout de suite que l'ensemble des entiers relatifs $\mathbb{Z}$ est un groupe abélien totalement ordonné (comme nous l'avons vu dans le chapitre des opérateurs) par rapport aux lois d'addition et de multiplication.*

D'abord pour raccourcir les développements, il est utile de rappeler que l'ensemble $\mathbb{Z}$ est un prolongement de $\mathbb{N}$ par le fait que nous y avons ajouté tous les nombres symétrique de signe négatif, soit $\mathbb{N} \subset \mathbb{Z}$.

Ainsi, en abusant toujours des notations[13]*, $(\mathbb{Z}, +, -, \leqslant, \geqslant)$ forme un groupe abélien totalement ordonné, soit quatre groupes, et $(\mathbb{Z}, \times, \leqslant, \geqslant)$ un monoïde abélien totalement ordonné, autrement dit deux monoïdes.*

Remarquons aussi que la loi de division n'existe pas pour tout élément de l'ensemble $\mathbb{Z}$.

La synthèse est indiquée dans le tableau 4.2.

LOIS DE $\mathbb{Z}$	$(+)$	$(-)$	$(\times)$	$(/)$
opération interne	oui	oui	oui	
associative	oui	non	oui	
commutative	oui	non	oui	non
élément neutre	oui 0	non	oui 1	
symétrique	oui	oui	non	

TABLEAU 4.2 – Lois dans un groupe

Propriété 4.12 *$(\mathbb{Z}, \leqslant, \geqslant)$ est totalement ordonné (attention à nouveau cette notation est un peu abusive ! il suffit qu'il y ait juste une des deux relations d'ordre R pour que l'ensemble soit totalement ordonné).*

Propriété 4.13 *$(\mathbb{Z}, +)$ est un groupe commutatif dont zéro 0 est l'élément absorbant*

Propriété 4.14 *La loi de division n'existe pas dans l'ensemble $\mathbb{Z}$.*

Propriété 4.15 *L'ensemble $\mathbb{Z}$ est un groupe abélien totalement ordonné par rapport aux lois d'addition et de soustraction et de multiplication*[14] *$(\mathbb{Z}, +, \leqslant, \geqslant)$.*

Propriété 4.16 *L'ensemble $\mathbb{Z}$ est un groupe commutatif totalement ordonné par rapport à la loi de multiplication $(\mathbb{Z}, \times, \leqslant, \geqslant)$.*

Nous voyons que $\mathbb{Z}$ a des propriétés trop restreintes, c'est la raison pour laquelle il est intéressant de le prolonger par l'ensemble des rationnels $\mathbb{Q}$ défini de manière très simpliste par (cf. chapitre sur les Nombres) :

$$\mathbb{Q} = \left\{ \frac{p}{q} \middle| (p, q) \in \mathbb{Z}/q = 0 \right\} \tag{4.87}$$

Ce qui signifie que l'ensemble des rationnels et défini par l'ensemble des quotients p et q appartenant chacun à $\mathbb{Z}$ dont nous excluons à q de prendre la valeur nulle (la notation $/q$ signifiant l'exclusion).

Et nous avons évidemment :

$$\mathbb{N}, \mathbb{Z} \subset \mathbb{Q} \tag{4.88}$$

Il est dès lors évident que $(\mathbb{Q}, \leqslant, \geqslant)$ est aussi totalement ordonné et aussi que $\mathbb{Q}$ est un groupe abélien totalement ordonné par rapport à la loi d'addition seulement :

$$(\mathbb{Q}, +, \leqslant, \geqslant) \tag{4.89}$$

Ce qui devient intéressant avec $\mathbb{Q}$, c'est que la de multiplication devient une loi interne et forme un groupe abélien commutatif dit *groupe multiplicatif* par rapport à $\mathbb{Q}$.

13. Normalement un groupe n'a qu'une seule loi et une seule relation d'ordre R suffit à l'ordonner.
14. La notation suivante est encore une fois abusive car le groupe est composé que d'une seule loi interne et d'une relation d'ordre R ce qui donnerait au total quatre groupes

Preuve Démontrons donc que le symétrique existe pour la loi de multiplication $(\cdot)$ tel que :

$$a \cdot a' = n = 1 \tag{4.90}$$

Puisque dans $\mathbb{Q}^*$ tout nombre peut se mettre sous la forme :

$$a = \frac{p}{q} \tag{4.91}$$

avec $p \in \mathcal{F}, q \in \mathbb{Z}^*$. Alors puisque :

$$a' = \frac{1}{a} = \frac{1}{\frac{p}{q}} = \frac{q}{p} \tag{4.92}$$

Il existe donc un symétrique à tout rationnel dans $\mathbb{Q}^*$ pour la loi de multiplication. •

Par définition, ou par construction, la division existe dans $\mathbb{Q}^*$ et est une opération interne. Mais est-elle associative telle que $\forall p, q, r \in \mathbb{Q}^*$ nous ayons :

$$(p/q)/r = p/(q/r) \tag{4.93}$$

Preuve Au fait, la démonstration est assez triviale si nous nous rappelons que la division se définit à partir de la loi de multiplication par l'inverse et que cette dernière loi est (elle !) associative. Ainsi, il vient :

$$(p/q)/r = \left(p \cdot \frac{1}{q}\right) \cdot \frac{1}{r} = p \cdot \left(\frac{1}{q} \cdot \frac{1}{r}\right) = p \cdot \left(\frac{1}{q \cdot r}\right) \neq p/(q/r) = p/\left(q \cdot \frac{1}{r}\right) = p \cdot \frac{r}{q} \tag{4.94}$$

Donc la loi de division n'est pas associative dans $\mathbb{Q}^*$. •

Nous pouvons aussi nous demander si la loi de division $(/)$ est commutative tel que $\forall a, b \in \mathbb{Q}^*$:

$$a/b = b/a \tag{4.95}$$

Nous voyons très bien que cela n'est pas le cas puisque nous pouvons écrire cette dernière relation sous la forme :

$$\frac{a}{b} = \frac{b}{a} = \pm 1 \Rightarrow a = \pm b \tag{4.96}$$

La synthèse est indiquée dans le tableau 4.3.

LOIS DE $\mathbb{Q}$	$(+)$	$(-)$	$(\times)$	$(/)$
opération interne	oui	oui	oui	oui
associative	oui	non	oui	non
commutative	oui	non	oui	non
élément neutre	oui 0	non	oui 1	oui
symétrique	oui	oui	non	non

TABLEAU 4.3 – Lois dans un groupe

Propriété 4.17 *$(\mathbb{Q}, \leqslant, \geqslant)$ est totalement ordonné.*

Propriété 4.18 *$(\mathbb{Q}, +)$, $(\mathbb{Q}^*, \times)$ sont indépendamment des groupes abéliens totalement ordonnés.*

Propriété 4.19 *Zéro "0" est l'élément absorbant par rapport au groupe $(\mathbb{Q}^*, \times)$.*

Propriété 4.20 *L'ensemble $\mathbb{Q}$ est un groupe abélien totalement ordonné par rapport aux lois d'addition et de multiplication que nous notons :*

$$(\mathbb{Q}, +, \leqslant, \geqslant) \quad et \quad (\mathbb{Q}^*, \times, \leqslant, \geqslant) \tag{4.97}$$

Les mêmes propriétés sont applicables à $\mathbb{R}$ et $\mathbb{C}$ mais à la différence que ce dernier n'est pas ordonnable. Dans ce dernier cas, nous devons nous assurer que la somme, la différence, le produit et le quotient de deux nombres de la forme $x + \mathrm{i}y$ donne quelque chose d'encore de cette forme.

Additionnons les nombres $a + \mathrm{i}b$ et $c + \mathrm{i}d$ où a, b, c et d sont des réels :

$$(a + \mathrm{i}b) + (c + \mathrm{i}d) = a + \mathrm{i}b + c + \mathrm{i}d = (a + c) + \mathrm{i}(b + d) \tag{4.98}$$

Donc l'addition est bien une loi interne commutative et associative pour laquelle il existe un élément neutre et symétrique dans l'ensemble des complexes.

Soustrayons les nombres $a + \mathrm{i}b$ et $c + \mathrm{i}d$ où a, b, c et d sont ici encore, des réels :

$$(a + \mathrm{i}b) - (c + \mathrm{i}d) = a + \mathrm{i}b - c - \mathrm{i}d = (a - c) + \mathrm{i}(b - d) \tag{4.99}$$

Donc la soustraction est une opération interne elle n'est ni commutative, ni associative elle n'a pas d'élément neutre à gauche et pas de symétrique.

Multiplions maintenant les nombres $a + \mathrm{i}b$ et $c + \mathrm{i}d$ où a, b, c et d là toujours, des réels. Pour parvenir à nos fins, nous emploierons la distributivité de la multiplication par rapport à l'addition.

$$\begin{aligned}(a + \mathrm{i}b) \cdot (c + \mathrm{i}d) &= a \cdot c + a \cdot \mathrm{i} \cdot d + \mathrm{i} \cdot b \cdot c + \mathrm{i} \cdot b \cdot \mathrm{i} \cdot d \\ &= a \cdot c + \mathrm{i} \cdot a \cdot d + \mathrm{i} \cdot b \cdot c + \mathrm{i}^2 \cdot b \cdot d \\ &= a \cdot c + \mathrm{i} \cdot a \cdot d + \mathrm{i} \cdot b \cdot c - b \cdot d \\ &= (a \cdot c - b \cdot d) + \mathrm{i} \cdot (a \cdot d + b \cdot c)\end{aligned} \tag{4.100}$$

Donc la loi de multiplication est bien une opération interne commutative, associative et distributive pour laquelle il existe un élément neutre et symétrique dans $\mathbb{C}^*$.

Une division est avant tout une multiplication par l'inverse. Prouver qu'il existe un inverse c'est prouver qu'il existe un symétrique pour la multiplication. Inversons donc le nombre $x + \mathrm{i}y$ où x et y sont des réels (différents de zéro) :

$$\frac{1}{x + \mathrm{i}y} = \frac{x - \mathrm{i}y}{(x + \mathrm{i}y)(x - \mathrm{i}y)} = \frac{x - \mathrm{i}y}{x^2 - (\mathrm{i}y)^2} = \frac{x - \mathrm{i}y}{x^2 + y^2} = \frac{x}{x^2 + y^2} - \mathrm{i}\frac{y}{x^2 + y^2} \tag{4.101}$$

Donc l'inverse d'un nombre complexe est bien une opération interne non associative et non commutative pour laquelle il existe un élément neutre, et elle est symétrique. Il en est de même pour la division, qui correspond au produit par l'inverse d'un nombre complexe.

Voyons un exemple de groupe cyclique : Dans $\mathbb{C}$, considérons $G = \{1, \mathrm{i}, -1, -\mathrm{i}\}$ muni de la multiplication usuelle des nombres complexes. Alors $(G, \times)$ est évidemment un groupe abélien. Un tel groupe est aussi monogène car engendré par les puissances d'un de ses éléments : i (ou bien $-\mathrm{i}$). Ce groupe monogène étant fini, il s'agit alors d'un groupe cyclique.

4.4.4 Anneaux

L'anneau est le cœur de l'algèbre commutative qui est la structure algébrique correspondant aux concepts collégiens d'addition, de soustraction, et de multiplication.

Définition 4.31 *Un groupe commutatif (ou* groupe abélien*) A est un* anneau *s'il est muni d'une seconde loi de composition interne vérifiant les propriétés suivante :*

$$(A, +, \star) \text{ est un anneau si } \begin{cases} (A, +) \text{ est un groupe abélien} \\ \star \text{ est associative} \\ \star \text{ est distributive par rapport à la loi } + \end{cases} \tag{4.102}$$

Comme nous le savons déjà, l'élément neutre de la première loi de composition interne $+$ est noté "0" et appelé *zéro* de l'anneau. La deuxième loi interne est souvent notée par un point à demi-hauteur et appelée la *multiplication*.

Remarques

1. Si de plus, la deuxième loi interne de composition $\star$ est également commutative, l'anneau est dit *anneau commutatif*. Nous rencontrons aussi des anneaux non-commutatifs dans lesquels la relation de commutativité n'est pas imposée, il faut alors renforcer la propriété de l'élément neutre de cette

deuxième loi en imposant à 1 d'être un élément neutre à la fois à droite et à gauche tel que $1a = a1 = a$ [15].

2. Si de plus, il existe dans A un élément neutre pour la deuxième loi de composition interne $\star$, et que cet élément neutre est l'unité 1 nous disons alors que l'anneau est un *anneau unitaire* et 1 est appelé *unité* de l'anneau. Si l'anneau est commutatif et possède un élément neutre pour la deuxième loi de composition interne alors nous parlons *d'anneau commutatif unitaire*
3. Si $a \star b = 0 \Rightarrow (a = 0 \text{ ou } b = 0)$, quels que soient les éléments a, b de A, l'anneau est dit *anneau intègre* ou *anneau sans diviseurs de zéro* (dans le cas contraire il est bien évidemment *non intègre*).
4. Un *anneau factoriel* est un anneau commutatif unitaire et intègre dans lequel le théorème fondamental de l'arithmétique (cf. chapitre de Théorie des Nombres) est vérifié.

Définition 4.32 *Un élément a d'un anneau A est un* élément unité *s'il existe $b \in A$ tel que $ab = ba = 1$. Si un tel b existe, il est unique* [16].

Définition 4.33 *Soit A un anneau. Nous disons que A possède des diviseurs de zéro s'il existe $a, b \in A$ avec $a \neq 0$, $b \neq 0$ et $a \cdot b = 0$. Les éléments a et b sont appelés des* diviseurs de zéro.

Remarques

1. Il est clair qu'un anneau est intègre si et seulement si il ne possède aucun diviseur de zéro.
2. Les notions d'unité et de diviseurs de zéro sont incompatibles mais un élément d'un anneau peut être ni l'un ni l'autre. C'est le cas, par exemple, de tous les entiers $\neq \{0, -1, 1\}$ dans $\mathbb{Z}$. Ce ne sont ni des unités, ni des diviseurs de zéro.

Nous verrons un exemple important d'anneau lors du cadre de notre étude des polynômes (cf. chapitre de Calcul Algébrique) mais nous en avons déjà vu de très importants lors de notre étude des classes de congruences dans le chapitre de théorie des nombres.

Voyons quelques exemples d'anneaux : Lors de notre étude des groupes nous avons trouvé que les structures :

$$(\mathbb{Z}, +), (\mathbb{Q}, +), (\mathbb{R}, +), (\mathbb{C}, +) \tag{4.103}$$

sont tous les quatre des groupes abéliens et les trois premiers sont en plus totalement ordonnés.

La loi de division n'étant en aucun cas associative, nous pouvons nous restreindre à étudier pour chacun des groupes précités, le couple de lois : $(+)$ et $(\times)$. Ainsi :

$$(\mathbb{Z}, +, \times), (\mathbb{Q}, +, \times), (\mathbb{R}, +, \times), (\mathbb{C}, +, \times) \tag{4.104}$$

constituent des anneaux commutatifs unitaires et intègres.

Remarque Nous considérerons comme évident, à ce niveau du discours, que le lecteur aura remarqué que $\mathbb{Z}$ est un *sous-anneau* de $\mathbb{Q}$ dans le sens où les opérations définies sont internes à chacun des ensembles et que les éléments neutres et identité sont identiques et qu'il existe pour chaque élément de ces ensembles un opposé qui est dans le même ensemble. Nous allons approfondir le concept de sous-anneau un peu plus loin.

Soit A un anneau. Nous avons les propriétés suivantes :

Propriété 4.21 $a + b = a + c \Rightarrow b = c, \quad \forall a, b, c \in A$

Preuve La propriété 4.21 découle de la définition D4 des structures algébriques à savoir que tout élément possède un opposé/symétrique. En effet, nous pouvons additionner à l'égalité $a + b = a + c$ l'élément $-a$. Nous obtenons alors $(-a) + a + b = (-a) + a + c$ par l'existence de l'opposé cela donne $0 + b = 0 + c$ d'où $b = c$. •

Propriété 4.22 $0 \cdot a = 0, \quad \forall a \in A$

Preuve La propriété 4.22 découle des définitions D3 (existence de l'élément neutre), D4 (existence de l'opposé/symétrique), D5 (distributivité par rapport à l'autre loi) ainsi que de la propriété 4.21. En effet, nous avons :

$$0 \cdot a + a = 0 \cdot a + 1 \cdot a = (0 + 1) \cdot a = 1 \cdot a = a \tag{4.105}$$

Nous avons donc $0 \cdot a + a = a$. La propriété 4.21 permet de conclure que $0 \cdot a = 0$ [17]. •

15. Un exemple d'anneau non-commutatif est fourni par l'ensemble des matrices $n \times n$ à coefficients dans un anneau A, par exemple $M_n(\mathbb{R})$- voir chapitre d'Algèbre Linéaire.
16. Nous en avons vu un exemple lors de notre étude des classes de congruence en théorie des nombres.
17. Nous pourrions aussi discuter de la pertinence de ce genre de démonstration.

Propriété 4.23 $-1 \cdot a = -a, \quad \forall a \in A$

Preuve La propriété 4.23 se montre à l'aide de la propriété 4.22. Nous avons :

$$0 = 0 \cdot a = (1 + (-1)) \cdot a = 1 \cdot a + (-1) \cdot a \tag{4.106}$$

en ajoutant $-a$ à cette dernière égalité, nous avons $-a = (-1) \cdot a$. •

Sous-anneau

Définition 4.34 *Soit A un anneau et $S \subset A$ un sous-ensemble de A. Nous disons que S est un sous-anneau de A si :*

1. $n \in S$ *(élément neutre de A est aussi dans S)*
2. $a \in S \Rightarrow -a \in S$
3. $a, b \in S \Rightarrow a + b \in S$
4. $a, b \in S \Rightarrow a \cdot b \in S$

Exemple 4.13 *L'anneau $\mathbb{Z}$ est un sous-anneau de $\mathbb{Q}, \mathbb{R}$.*

4.4.5 Corps

Définition 4.35 *Nous désignons un ensemble de nombres par le terme* corps *si :*

$$(C, +, \star) \text{ est un corps si } \begin{cases} (C, +, \star) \text{ est un anneau unitaire} \\ (C - \{0\}, \star) \star \text{ est un groupe} \end{cases} \tag{4.107}$$

Donc un corps est un anneau non nul dans lequel tout élément non nul est inversible ou en d'autres termes : un anneau dont tous les éléments non nuls sont des unités est un corps.

REMARQUES
1. si la loi interne $\star$ est également commutative, le corps est dit *corps commutatif* ;
2. les quaternions (cf. chapitre sur les Nombres) forment un corps non commutatif pour l'addition et la multiplication.

Voyons des exemples de corps parmi les anneaux unitaires (**??**). Il nous faut d'abord déterminer lesquels ne constituent pas des groupes par rapport à la loi interne de multiplication ($\times$).

Comme nous l'avons déjà vu dans notre étude des groupes précédemment, il est évident qu'il nous faut éliminer $(\mathbb{Z}, +, \times)$ à cause de l'existence des inverses qui n'est pas assurée dans cet ensemble. Ainsi, les corps fondamentaux de l'arithmétique sont :

$$(\mathbb{Q}, +, \times), (\mathbb{R}, +, \times), (\mathbb{C}, +, \times) \tag{4.108}$$

et puisque la loi de multiplication ($\times$) est commutative dans ces ensembles, nous pouvons affirmer que ces corps sont également des corps commutatifs. L'ensemble est rappelé sur le schéma 4.15.

Ainsi, nous appellerons *corps* un système C de nombres réels ou complexes a tels que la somme, la différence, le produit et le quotient de deux quelconques de ces nombres a appartiennent au même système C.

Nous énonçons également cette propriété de la manière suivante : les nombres d'un corps se reproduisent par les opérations rationnelles (addition, soustraction, multiplication, division). Ainsi, il est évident que le nombre zéro ne pourra jamais former le dénominateur d'un quotient et l'ensemble des entiers ne peut former un corps car la division dans l'ensemble des nombres entiers ne donne pas nécessairement un résultat dans ce même ensemble.

4.4.6 Espaces vectoriels

Lorsque nous définissons un *vecteur* (cf. chapitre de Calcul Vectoriel), nous faisons habituellement référence à un *espace euclidien* (cf. aussi chapitre de Calcul Vectoriel) de n dimensions de $\mathbb{R}^n$. Cependant, la notion d'espace vectoriel est beaucoup beaucoup plus vaste que ce dernier qui ne représente qu'un cas particulier.

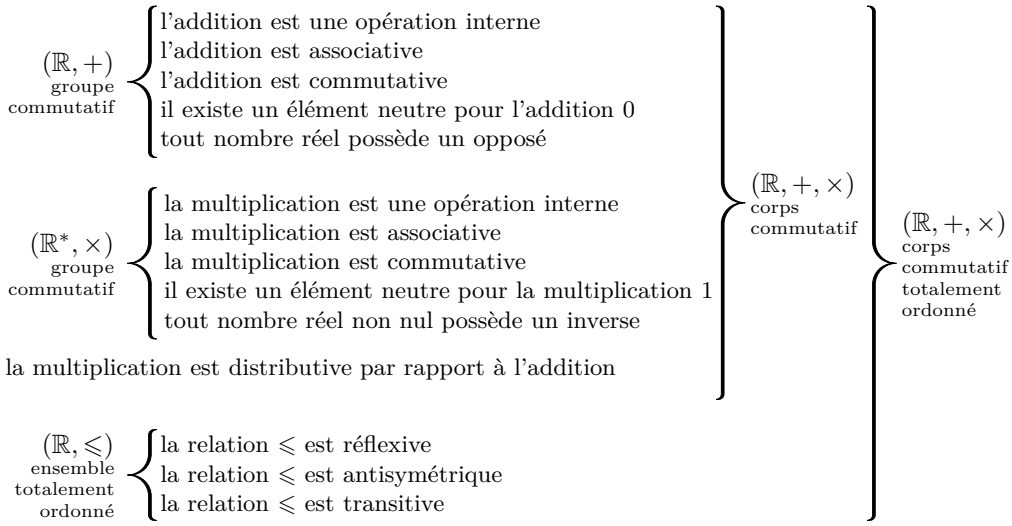


FIGURE 4.15 – Classification des corps

Définition 4.36 *Un* espace vectoriel *(EV) ou* K-espace vectoriel *(abrégé : K-ev) sur le corps K* [18] *est un ensemble $(E,+,\star)$ se $(E,+)$ est un groupe abélien et $\star$ est une loi externe définie par :*

$$\begin{aligned} &E\times K\to E\\ &(x,a)\to a\star x \end{aligned} \tag{4.109}$$

Nous avons donc deux lois de composition en prenant les notations traditionnelles des vecteurs :

1. une loi de composition interne, l'addition notée $+$ qui vérifie :
   (a) associativité : $\forall\vec{x},\vec{y},\vec{z}\in E:(\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z})$
   (b) commutativité : $\forall\vec{x},\vec{y}\in E:\vec{x}+\vec{y}=\vec{y}+\vec{x}$
   (c) élément neutre : $\exists\vec{0}\in E|\forall\vec{x}\in E:\vec{x}+\vec{0}=\vec{x}$
   (d) élément opposé : $\forall\vec{x}\in E,\exists\vec{x}'\in E:\vec{x}+\vec{x}'=\vec{0}$
2. une loi de composition externe : la multiplication par un scalaire, notée $\cdot$, qui vérifie :
   (a) associativité : $\forall\lambda,\mu\in K:\forall\vec{x}\in E,\ \lambda\cdot(\mu\cdot\vec{x})=(\lambda\cdot\mu)\cdot\vec{x}$
   (b) distributivité : $\forall\lambda,\mu\in K:\forall\vec{x}\in E,\ (\lambda+\mu)\cdot\vec{x})=\lambda\cdot\vec{x}+\mu\cdot\vec{x}$
   (c) distributivité : $\forall\lambda\in K:\forall\vec{x},\vec{y}\in E,\ \lambda\cdot(\vec{x}+\vec{y})=\lambda\cdot\vec{x}+\lambda\cdot\vec{y}$
   (d) élément neutre (de K sur E) : $\forall\vec{x}\in E,\ 1\cdot\vec{x}=\vec{x}$

REMARQUES

1. R1. Nous disons alors que l'espace vectoriel à une *structure algébrique vectorielle* et que ces éléments sont des *vecteurs*, les éléments de K des *scalaires*.
2. R2. Les opérations respectives s'utilisent fréquemment comme l'addition et la multiplication que nous connaissons bien sur $\mathbb{R}$, ce qui est bien commode pour nos habitudes...
3. R3. Dorénavant, pour distinguer les éléments du corps K et de l'ensemble E, nous noterons ceux de K par des lettres grecques et ceux de E par des lettres latines majuscules.
4. R4. Outre les cinq propriétés énumérées ci-dessus, il ne faut pas oublier d'ajouter les cinq autres propriétés du groupe abélien (opération interne, commutativité, associativité, élément neutre, élément inverse). Ce qui nous fait donc au total dix propriétés à respecter.

Il est inutile de démontrer que ces propriétés sont respectées pour $\mathbb{R}^n$ et, par conséquent pour $\mathbb{R}^2$. Nous pouvons cependant nous poser la question à propos de certains sous-ensembles de $\mathbb{R}^n$.

Exemple 4.14 *Considérons la région rectangulaire de la figure 4.16(a). Ce sous-ensemble de* $\mathbb{R}^2$

18. Nous prendrons fréquemment pour ce corps $\mathbb{R}$ ou $\mathbb{C}$.

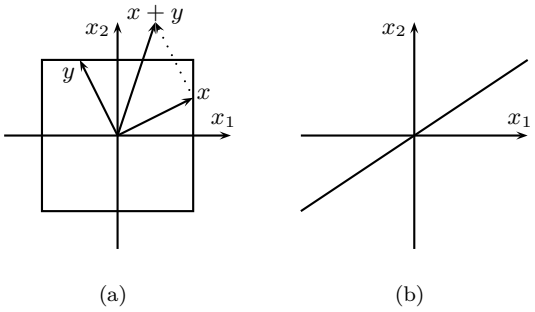


FIGURE 4.16 – Illustration de la notion d'espace vectoriel

n'est pas un espace vectoriel car, entre autres, la propriété d'opération interne du groupe abélien n'est pas satisfaite. En effet, si nous prenons deux vecteurs à l'intérieur du rectangle et que nous les additionnons, il se peut que le résultat sorte du rectangle. Par contre, il est facile de voir que la droite (infinie) illustrée dans la figure 4.16(b) respecte toutes les propriétés énumérées précédemment et, par conséquent, définit un espace vectoriel. Notons bien, cependant, que cette droite se doit de passer par l'origine, sinon la propriété d'élément neutre du groupe abélien ne serait pas respectée (l'élément neutre n'existant plus).

Exemple 4.15 *Un autre exemple d'un espace vectoriel est l'ensemble P^2 des polynômes de degré deux ou moins (cf. chapitre de Calcul Algébrique). Par exemple, deux éléments de cet espace sont :*

$$x = t^2 + 2t + 3 \qquad y = -t + 5 \tag{4.110}$$

Cet ensemble respecte les dix propriétés d'un espace vectoriel. En effet, si nous additionnons deux polynômes de degré deux ou moins, nous obtenons un autre polynôme de degré deux ou moins. Nous pouvons aussi multiplier un polynôme par un scalaire sans changer l'ordre de celui-ci, etc. Nous pouvons donc représenter un polynôme par des vecteurs dont les termes sont les coefficients du polynôme.

Mentionnons que nous pouvons aussi former des espaces vectoriels avec des ensembles de fonctions plus générales que des polynômes. Il importe seulement de respecter les dix propriétés fondamentales d'un espace vectoriel.

Ainsi défini, un espace vectoriel E sur K est une action de $K\star$ sur $(E,+)$ qui est compatible avec
Déf4.46|p106 la loi de groupe, soit, par extension, un *automorphisme*• sur $(E,+)$.

Définition 4.37 *Soit E un espace vectoriel, nous appelons* sous-espace vectoriel *(SEV) F de E un sous-ensemble de E si et seulement si :*

$$\begin{aligned} &F \neq \varnothing \\ &\forall (X,Y) \in F^2, X+Y \in F \\ &\forall \lambda \in K, \forall X \in F, \lambda X \in F \end{aligned} \tag{4.111}$$

4.4.7 Algèbres

Une *C-algèbre* A où C est un corps commutatif, est un ensemble A muni de deux lois de composition internes $+$ (addition) et $\times$ (produit) et d'une loi externe $\cdot$ (multiplication) à domaine d'opérateurs dans C (produit par un scalaire) si et seulement si :

$$(A, C, +, \cdot, \times) \text{ est une } C\text{-algèbre si } \begin{cases} (A,+,\cdot) \text{ est un } C\text{-espace vectoriel} \\ (A,+,\times) \text{ est un anneau unitaire} \\ \forall \lambda \in C, \forall A, b \in A, (\lambda a)\cdot b = a \cdot (\lambda b) = \lambda(a \cdot b) \end{cases} \tag{4.112}$$

Exemple 4.16 *Pour reprendre un exemple dans la lignée de celui sur les exemples vectoriels, l'espace euclidien $\mathbb{R}^3$ muni de l'addition $(+)$, de la multiplication $(\cdot)$ et du produit vectoriel $\times$ est une $\mathbb{R}$-algèbre non associative et non commutative notée $(\mathbb{R}^3, \mathbb{R}, +, \cdot, \times)$.*

Exemple 4.17 *$\mathbb{C}$ est une $\mathbb{R}$-algèbre puisqu'un nombre complexe peut être vu comme un vecteur à deux composantes.*

4.5 Homomorphismes

Le concept d'homomorphismes (du grec *homoios = semblable* et *morphê = forme*) a été défini par les mathématiciens car permettant de mettre en évidence des propriétés remarquables des fonctions en particulier avec leurs structures, leur noyau, et de ce que nous appelons les *idéaux*. Ils nous permettront ainsi d'identifier une structure algébrique d'une autre.

Définition 4.38 *Si $(A, \star)$ et $(B, \circ)$ sont deux magmas*[19]*, une application f de A dans B est un* homomorphisme de magma *ou* morphisme de magma *si :*

$$f(a \star b) = f(a) \circ f(b), \quad \forall a, b \in A \tag{4.113}$$

en d'autres termes, si l'image d'un composé dans A est le composé des images dans B.

Définition 4.39 *Si $(A, \star)$ et $(B, \circ)$ sont deux monoïdes, une application f de A dans B est un* homomorphisme de monoïde *si :*

$$f(a \star b) = f(a) \circ f(b), \quad \forall a, b \in A \qquad \text{et} \quad f(1_A) = 1_B \tag{4.114}$$

où $1_A, 1_B$ sont les éléments neutres respectifs des monoïdes A, B.

Définition 4.40 *Si A, B sont deux anneaux, un* homomorphisme d'anneaux *(très important pour le chapitre de Cryptographie!) de A dans B est une application $f : A \to B$ telle que nous ayons pour tout $a, a' \in A$:*

$$f(1_A) = 1_B; \qquad f(a + a') = f(a) + f(a'); \qquad f(a \cdot a') = f(a) \cdot f(a') \tag{4.115}$$

où $1_A, 1_B$ sont les éléments neutres des anneaux A, B par rapport à la multiplication.

Soit $f : A \to B$ un homomorphisme d'anneaux, alors :

Propriété 4.24 $f(0) = 0$

Preuve Par $f(a+a') = f(a)+f(a')$, nous avons $f(a+0) = f(a)+f(0)$. En ajoutant $-f(a)$ des deux côtés de l'égalité, nous obtenons $0 = f(0)$. •

Propriété 4.25 $f(-a) = -f(a)$

Preuve La propriété 4.25 découle aussi de $f(a + a') = f(a) + f(a')$ et de la propriété 4.24. En effet, nous avons $f(0) = f(a + (-a)) = f(a) + f(-a) = 0$. En additionnant $-f(a)$ aux deux côtés de la dernière égalité, nous obtenons $f(-a) = -f(a)$. •

Propriété 4.26 *Si a est une unité de A, alors $f(a)$ est une unité de B et $f(a)^{-1} = f(a^{-1})$.*

Preuve Soient $a, b \in A$ tel que $a \cdot b = b \cdot a = 1$. Alors par $f(a \cdot b) = f(a) \cdot f(b)$ et $f(1_A) = 1_B$, nous avons $1 = f(1) = f(a \cdot b) = f(a) \cdot f(b)$ et de même, $1 = f(b) \cdot f(a)$ ce qui montre que $f(b)$ est l'inverse de $f(a)$ si b est l'inverse de a. •

Théorème 4.1 *Un homomorphisme d'anneaux $f : A \to B$ est injectif si et seulement si l'élément 0 est la seul pré-image de 0, autrement dit si :*

$$\mathrm{Ker}(f) = \{0\} \tag{4.116}$$

Preuve La condition est clairement nécessaire. Montrons qu'elle est suffisante. Nous supposons donc que $\mathrm{Ker}(f) = \{0\}$. Soit $a, a' \in A$ tel que $f(a) = f(a')$. Alors $0 = f(a) - f(a') = f(a - a')$ ceci implique que $a - a' = 0$ donc $a = a'$, ce qui montre que f est injectif. •

Définition 4.41 *Soient $(A, +)$ et $(B, \star)$, deux groupes et une application $f : A \to B$. Nous disons que f est un* homomorphisme de groupe *si :*

$$f(x + y) = f(x) \star f(y) \quad \forall x, y \in A \tag{4.117a}$$
$$f(1_A) = 1_B \tag{4.117b}$$
$$f(x^{-1}) = (f(x))^{-1} \quad \forall x \in A \tag{4.117c}$$

où $1_A, 1_B$ sont les éléments neutres respectifs des groupes A, B.

19. La notation utilisée pour les lois internes importe peu.

Ceci dit, la proposition (4.117c) est en fait une conséquence de (4.117a) et (4.117b). Effectivement, considérons un homomorphisme f entre les groupes $(A,+)$ et $(B,\star)$ avec $1_A, 1_B$ respectivement les éléments neutres de A et B. Nous avons alors :

$$\forall x \in A : 1_B = f(1_A))f(x + x^{-1}) = f(x) \star f(x^{-1}) \tag{4.118}$$

d'où :

$$\forall x \in A : f(x) = \frac{1_B}{f(x^{-1})} = (f(x^{-1}))^{-1} \tag{4.119}$$

Définition 4.42 *Soit f une application $f : A \to B$ d'un corps vers un autre. Nous disons que f est un* homomorphisme de corps *si f est un homomorphisme d'anneaux.*

REMARQUE Le fait que l'homomorphisme de corps soit le même que celui d'un anneau tient juste au fait que la différence entre les deux est que les éléments du corps sont inversibles (aucune loi ou propriété de loi ne diffère entre les deux selon leur définition).

Montrons maintenant que tout homomorphisme de corps est injectif.

Preuve Soit $f : A \to B$ un homomorphisme de corps et $\text{Ker}(f)$ un idéal (voir plus bas pour la définition d'un idéal). Nous avons forcément $\text{Ker}(f) \neq A$ car $f(1) = 1$ du fait de la définition de l'homomorphisme d'anneau. Il en résulte donc que $\text{Ker}(f) = \{0\}$ et par suite f est injectif.

Ainsi, un homomorphisme de corps est donc simplement homomorphisme d'anneaux. •

Preuve (Stratégie 2) Si a est différent de 0 alors :

$$1 = f(1) = f(a \cdot b) = f(a) \cdot f(b) \tag{4.120}$$

où $b = a^{-1}$. Donc $f(a) \neq 0$ ce qui prouve que $\text{Ker}(f) \neq 0$ et donc que f est injective. •

Définition 4.43 *Soient A et B deux K-ev et $f : A \to B$ une application de A dans B. Nous disons que f est une* application linéaire *ou* homomorphisme d'espaces vectoriels *si :*

$$\begin{aligned} &f(x+y) = f(x) + f(y) && \forall (x,y) \in A \\ &f(\lambda x) = \lambda f(x) && \forall x \in A, \forall \lambda \in K \end{aligned} \tag{4.121}$$

et nous notons $\mathcal{L}(A,B)$ l'ensemble des applications linéaires.

REMARQUES

1. Nous avions déjà défini plus haut le concept d'application linéaire mais n'avions pas précisé que les deux ensembles A et B étaient des K-ev.
2. L'application linéaire est appelée *forme linéaire* si et seulement si $B = K$.

4.5.1 Isomorphisme, endomorphisme, automorphisme

Définition 4.44 *Si l'homomorphisme est bijectif nous dirons alors que f est un* isomorphisme. *S'il existe un isomorphisme entre A et B, nous disons que A et B sont* isomorphes *et nous noterons cela $A \simeq B$.*

REMARQUE L'isomorphisme permet au fait d'identifier deux ensembles munis d'une structure algébrique identique (que ce soit groupe, anneau, etc.) mais dont les éléments sont nommés d'une façon différente.

Définition 4.45 *Si l'homomorphisme f est une application uniquement interne, nous dirons alors que f est un* endomorphisme *—en d'autres termes, nous avons un endomorphisme si dans la définition de l'homomorphisme nous avons $A = B$.*

REMARQUE Si nous avons un endomorphisme f de E, f est donc restreint à $\text{Im}\,(f)$. Donc le terme *endomorphisme* veut juste dire que l'application f arrive dans E et pas qu'elle touche tous les éléments de E. Nous avons $f(E) \subset E$ et pas forcément $f(E) = E$ car dans ce dernier cas nous disons que f est surjective comme nous l'avons déjà vu.

Définition 4.46 *Si l'endomorphisme f est en plus bijectif (donc en d'autres termes si l'homomorphisme est un endomorphisme et un isomorphisme), nous dirons alors que f est un* automorphisme.

4.5.2 Idéal

Définition 4.47 *Soit A un anneau commutatif. Un sous-ensemble $S \subset A$ est un* idéal *si :*

1. $a + a' \in S$ *pour tout* $a, a' \in S$;
2. $r \cdot a \in S$ *pour tout* $a \in S$ *et tout* $r \in A$.

En d'autres termes, un idéal est un sous-ensemble fermé pour l'addition et stable par multiplication par un élément quelconque de A.

Exemple 4.18 *L'ensemble des nombres pairs est un exemple d'idéal de l'ensemble des nombres naturels.*

Remarque Les idéaux $S = \{0\}$ et $S = A$ sont appelés *idéaux triviaux*.

Pour savoir si un idéal est égal à tout l'anneau, il est utile d'utiliser la propriété suivante qui spécifie que si A est un anneau et I un idéal de A, alors si $1 \in I$ nous avons $I = A$.

Preuve Ceci résulte de la deuxième entrée de la définition d'un idéal : pour tout $r \in A$, nous avons $r = r \cdot 1 \in I$ car $1 \in I$. •

Un premier exemple d'idéal est donné par le noyau d'un homomorphisme d'anneaux. Effectivement, démontrons que le noyau d'un homomorphisme $f : R \to S$ est un idéal de R.

Preuve Soient $a, a' \in \mathrm{Ker}(f)$, alors :

$$f(a + a') = f(a) + f(a') = 0 + 0 = 0 \tag{4.122}$$

ce qui montre que $a + a' \in \mathrm{Ker}(f)$. Soit $r \in R$, alors :

$$f(ra) = f(r)f(a') = f(r) \cdot 0 = 0 \tag{4.123}$$

ce qui montre que $ra \in \mathrm{Ker}(f)$. •

Proposition 4.4 *Soit A un anneau et soit $a \in A$. Le sous-ensemble $\{ax | x \in A\}$ noté (a) ou aA, est un idéal.*

Définition 4.48 *Un idéal $I \neq A$ d'un anneau A est dit* idéal principal *s'il existe $a \in A$ tel que $I = (a)$.*

Définition 4.49 *Un anneau dont tous les idéaux sont principaux est dit* anneau principal.

Montrons maintenant que l'anneau $\mathbb{Z}$ est principal (car tous ses idéaux sont principaux).

Preuve Soit I un idéal de $\mathbb{Z}$ [20]. Soit $r \in I$ le plus petit entier positif non nul de I. Nous allons montrer que $I = r\mathbb{Z}$. Soit a un élément quelconque de I. La division euclidienne nous permet d'écrire $a = qr + r'$ avec $0 \leqslant r' \leqslant r$. Puisque $r' = a - qr$ et que $a, r \in I$, par la définition d'un idéal, nous avons $r' \in I$ —la somme ou différences des éléments d'un idéal appartenant à l'idéal. Par choix de r, ceci implique que $r' = 0$ et donc que $a = rq$.

Ainsi tout élément de I est un multiple de r et donc $I = r\mathbb{Z}$. •

Cette démonstration n'utilise que la division euclidienne sur $\mathbb{Z}$. Nous pouvons alors généraliser ce résultat aux anneaux qui possèdent une division euclidienne. Ainsi, par exemple, l'anneau $k[X]$ des polynômes (cf. chapitre de Calcul Algébrique) à coefficients dans un corps k est un anneau principal car il possède une division euclidienne.

Preuve Soit I un idéal de $k[X]$. Notons d le plus petit degré que puisse avoir un polynôme non nul de I. Si $d = 0$ alors $1 \in I$ et donc $I = 1 \cdot k[X] = k[X]$. Autrement, soit $a(X)$ un polynôme de degré d. Si $u(X) \in I$ alors on peut diviser $u(X)$ par $a(X)$. Il existe $q(X), r(X) \in k[x]$ tels que $\deg(r) < \deg(a) = d$ et $u(X) = q(X) \cdot a(X) + r(X)$. Donc $r(X) \in I$ ce qui entraîne $r = 0$ ce qui est en contradiction avec la minimalité de d. Par suite, $u(X) = q(X) \cdot a(X)$ d'où $I = a(X)k[X]$. •

Ainsi, les seuls idéaux de $\mathbb{Z}$ sont ceux de la forme $m\mathbb{Z}$. De plus, si nous avons d et m qui sont des entiers supérieurs à 1 alors $m\mathbb{Z} \subset d\mathbb{Z}$ si et seulement si $d|m$.

20. Il est facile d'en choisir un : par exemple, tous les multiples de 2 ou de 3, etc.

Preuve Si $d|m$ alors il existe n avec $m = d \cdot n$. Soit $m \cdot a$ un élément de $m\mathbb{Z}$ alors $m \cdot a = d \cdot n \cdot a \in d\mathbb{Z}$, ce qui montre que ù$\mathbb{Z} \subset d\mathbb{Z}$.

Réciproquement, $m \in d\mathbb{Z}$ implique que m est de la forme $d \cdot n$ et ceci prouve que d divise m. •

Démontrons aussi qu'un anneau R est un corps si et seulement s'il ne possède que les idéaux triviaux $\{0\}, R$.

Preuve Montrons que la condition est nécessaire. Soit I un idéal non nul de R et $r \in I$ un élément non nul. Par hypothèse (qu'il s'agit d'un corps), il est inversible, c'est-à-dire qu'il existe $t \in R$ tel que $r \cdot t = t \cdot r = 1$. Ceci implique que $1 \in I$ et donc, par un résultat obtenu plus haut $I = R$.

Réciproquement, supposons que tout idéal $I \neq R$ soit l'idéal nul. Alors si $r \in R$ est un élément non nul de R, l'idéal principal (r) doit être égal à R. Mais ceci implique que $1 \in (r)$ et dont qu'il existe $x \in R$ avec $r \cdot x = 1$ ce qui montre que r est inversible. L'anneau R est donc un corps. •

Cette caractérisation va nous permettre de démontrer facilement que tout homomorphisme partant d'un corps est injectif. Autrement dit, si $f : R \to S$ est un homomorphisme où R est un corps, alors f est injectif.

Preuve Nous mettons ensemble ce qui a été vu jusque-là. Nous avons démontré plus haut que le noyau $\text{Ker}(f)$ d'un homomorphisme est un idéal. Mais nous avons également démontré que nous avons soit $\text{Ker}(f) = \{0\}$ soit $\text{Ker}(f) = R$ (car l'anneau R est un corps si et seulement s'il ne possède que les idéaux triviaux).

Mais puisque $f(1) = 1 \neq 0$ —du fait de la définition d'un homomorphisme— il s'ensuit qu'il reste $\text{Ker}(f) = \{0\}$ [21]. Ceci implique par le théorème 4.1 que f est injective. •

Étudions maintenant les homomorphismes dont l'anneau de départ est $\mathbb{Z}$. Soit A un anneau et $f : \mathbb{Z} \to A$, un homomorphisme. Par définition d'un homomorphisme et par ses propriétés, il faut que $f(0) = 0$ et $f(1) = 1$ et aussi :

$$f(n) = f(\underbrace{1 + 1 + \ldots + 1}_{n \text{ fois}}) = \underbrace{1_A + 1_A + \ldots + 1_A}_{n \text{ fois}} = n \cdot 1_A \tag{4.124}$$

pour tout $n \in \mathbb{Z}$. Ainsi f est complètement déterminé par la donnée de $f(1)$ et est donc unique. Réciproquement, nous montrons que l'application $f : \mathbb{Z} \to A$ définie par :

$$f(n) = n \cdot 1_A \tag{4.125}$$

est un homomorphisme d'anneaux. En résumé, il existe un et un seul homomorphisme de $\mathbb{Z}$ dans un anneau quelconque A.

Définition 4.50 *Soit A un anneau et $f : \mathbb{Z} \to A$ l'unique homomorphisme défini précédemment. Si f est injectif, nous dirons que A est de* caractéristique nulle. *Sinon,* $\text{Ker}(f)$ *est un idéal non trivial de $\mathbb{Z}$ et comme $\mathbb{Z}$ est dès lors principal, il est de la forme $m\mathbb{Z}$ avec $m > 0$. L'entier m est appelé la* caractéristique de A.

REMARQUE Moins formellement, la caractéristique d'un anneau est le plus petit entier positif m tel que $m \cdot 1_A = 0$. S'il n'y en a pas, alors la caractéristique est nulle.

Exemple 4.19 *L'anneau $\mathbb{Z}$ est de caractéristique nulle car l'unique homomorphisme $f : \mathbb{Z} \to \mathbb{Z}$ est l'identité. Il est donc injectif. Les injections $\mathbb{Z} \to \mathbb{Q}$ et $\mathbb{Z} \to \mathbb{R}$ montrent que $\mathbb{Q}$ et $\mathbb{R}$ (et $\mathbb{C}$ également) sont des corps de caractéristique nulle.*

Nous nous proposons maintenant de démontrer que la caractéristique d'un anneau intègre (et en particulier d'un corps) est égale 0 ou à un premier p.

Preuve Nous montrons la contraposée. Soit A un anneau de caractéristique $m \neq 0$ avec m non premier. Il existe alors des entiers naturels $n, r < m$ tels que $m = n \cdot r$. Soit $f : \mathbb{Z} \to A$ l'unique homomorphisme. Par définition de l'idéal de m, nous avons $f(m) = 0$ mais $f(r) \neq 0 \neq f(n)$. Mais alors $f(r) \cdot f(n) = f(r \cdot n) = f(m) = 0$ ce qui montre que A n'est pas intègre. •

REMARQUE La réciproque du théorème n'est pas vraie comme le montre l'exemple de l'anneau $\mathbb{R} \times \mathbb{R}$ où l'addition et la multiplication se font composante par composante. C'est un anneau de caractéristique nulle mais avec des diviseurs de zéro $(1, 0) \cdot (0, 1) = 0$.

21. Il a en effet été démontré plus haut que si A est un anneau, I un idéal de A et $1 \in I$ alors $I = R$.

5

Théorie des nombres

La théorie des nombres est une branche des mathématiques qui s'occupe des propriétés des nombres entiers, qu'ils soient entiers naturels ou entiers relatifs, et contient beaucoup de problèmes ouverts qu'il est facile relativement de comprendre, même par les non-mathématiciens.

Plus généralement, le champ d'étude de cette théorie concerne une large classe de problèmes qui proviennent naturellement de l'étude des entiers. La théorie des nombres peut être divisée en plusieurs champs d'étude (théorie algébrique des nombres, théorie calculatoire des nombres, etc.) en fonction des méthodes utilisées et des questions traitées.

Remarque Le terme *arithmétique* est aussi utilisé pour faire référence à la théorie des nombres. C'est un terme assez ancien, qui n'est plus aussi populaire que par le passé.

Nous avons choisi de ne présenter dans cet exposé, que les sujets qui sont indispensables à l'étude de la mathématique et de la physique théorique ainsi que ceux devant faire absolument partie de la culture générale de l'ingénieur.

5.1 Principe du bon ordre

Nous tiendrons acquis ce principe qui s'énonce ainsi :

Théorème 5.1 *Tout ensemble non vide $S \subset \mathbb{N}$ contient un plus petit élément.*

Nous pouvons utiliser ce théorème pour démontrer une propriété importante des nombres appelée *propriété archimédienne* ou *axiome d'*Archimède qui s'énonce :

Axiome 5.1 *Pour $\forall a, b \in \mathbb{N}$ où a est non nul, il existe au moins un entier positif n tel que :*

$$n \cdot a \geqslant b \tag{5.1}$$

En d'autres termes, pour deux grandeurs inégales, il existe toujours un multiple entier de la plus petite, supérieur à la plus grande. Nous appelons *archimédien* des structures dont les éléments vérifient une propriété comparable•. Chap4

Même si cela est trivial à comprendre faisons la démonstration car elle permet de voir le type de démarches utilisés par les mathématiciens quand ils doivent démontrer des éléments triviaux de ce type.

Preuve Supposons le contraire. Il se formule en disant que pour $\forall n \in \mathbb{N}$ nous avons :

$$n \cdot a < b \tag{5.2}$$

Si nous démontrons que cela est absurde pour tout n alors nous aurons démontré la propriété archimédienne.

Considérons alors l'ensemble $S = \{b - n \cdot a | n \in \mathbb{N}$. En utilisant le principe du bon ordre, nous déduisons qu'il existe $s_0 \in S$ tel que $s_0 \leqslant s$ pour tout $s \in S$. Posons donc :

$$s_0 = b - n_0 \cdot a \tag{5.3}$$

Comme $n \cdot a < b$ impose également $b - (n_0 + 1) \cdot a \in S$, nous devons avoir :

$$b - (n_0 + 1) \cdot a \geqslant b - n_0 \cdot a \tag{5.4}$$

ce qui voudrait dire que $n_0 + 1 \leqslant n_0$, d'où une contradiction évidente. Cette contradiction amène que l'hypothèse initiale comme quoi $n \cdot a < b$ pour tout n alors est fausse et donc que la propriété archimédienne est démontrée par l'absurde. •

5.2 Principe d'induction

Soit S un ensemble de nombres naturels qui possède les deux propriétés suivantes :

- $1 \in S$ (5.5a)
- si $k \in S$ alors $k+1 \in S$ (5.5b)

alors :

$$S = \mathbb{N}\backslash\{0\} = \mathbb{N}^* \tag{5.6}$$

Nous construisons ainsi l'ensemble des nombres naturels. Soit $B = \mathbb{N}^*\backslash S$, le symbole $\backslash$ signifiant *excluant*. Nous voulons donc démontrer que :

$$B = \varnothing \tag{5.7}$$

À nouveau, même si cela est trivial à comprendre faisons la démonstration car elle permet de voir le type de démarches utilisées par les mathématiciens quand ils doivent démontrer des éléments triviaux de ce type.

Preuve Supposons le contraire de (5.7). Par le principe du bon ordre, puisque alors $B \subset \mathbb{N}$, B doit posséder un plus petit élément b_0. Mais d'après (5.5a), nous avons que $b_0 > 1 \notin B$, c'est-à-dire aussi $b_0 - 1 \in S$. En faisant appel à (5.5b), nous avons finalement que $b_0 \in S$, c'est-à-dire que $b_0 \notin B$, donc une contradiction.•

Exemple 5.1 *Nous souhaitons montrer à l'aide du principe d'induction, que la somme des n premiers carrés est égale à $n(n+1)(2n+1)/6$, c'est-à-dire que pour $n \geqslant 1$, nous aurions (cf. chapitre de Suites Et Séries) :*

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6} \tag{5.8}$$

D'abord la relation (5.8) est facilement vérifiée pour $n = 1$. Nous allons montrer que $n = k+1$ vérifie aussi cette relation. En vertu de l'hypothèse d'induction :

$$\sum_{i=1}^{k+1} i^2 = \frac{k(k+1)(2k+1)}{6} + (k+1)^2 = \frac{(k+1)(k+2)(2k+3)}{6} \tag{5.9}$$

nous retrouvons bien l'hypothèse de la validité de (5.8) d'où le résultat.

Exemple 5.2 *La suite de Fibonacci (cf. chapitre sur les Suites Et Séries) possède la propriété que la somme des premiers termes consécutifs de la suite, augmentée de 1, est égale au terme qui suit de deux rangs le terme auquel nous nous sommes arrêtés.*

Ainsi, nous avons :

$$1 + u_0 + u_1 + u_2 + \ldots + u_n = u_{n+2} \tag{5.10}$$

Pour démontrer cette propriété, nous constatons que cette dernière est aisément vérifiée pour les premières valeurs $0, 1, 2, 3, 4$ de n. Supposons que la propriété ait été démontrée pour n et pour $n+1$. Nous avons :

$$\begin{aligned} 1 + u_0 + u_1 + u_2 + \ldots + u_n &= u_{n+2} \\ 1 + u_0 + u_1 + u_2 + \ldots + u_n + u_{n+1} &= u_{n+3} \end{aligned} \tag{5.11}$$

par conséquent, en ajoutant ces deux égalités et tenant compte de la loi de formation nous obtenons :

$$1 + u_0 + u_1 + u_2 + \ldots + u_{n+2} = u_{n+4} \tag{5.12}$$

par conséquent la deuxième relation ne diffère que de la première par le changement de n en $n+2$ puisque u_0 est nul. Le théorème est démontré.

Ce procédé de démonstration est d'une très grande importance dans l'étude de l'arithmétique ; souvent l'observation et l'induction ont permis de soupçonner des lois qu'il eût été plus difficile de trouver par a priori. Nous nous rendons compte de l'exactitude des formules par la méthode précédente qui a donné naissance à l'algèbre moderne par les études de FERMAT et de PASCAL sur le triangle de PASCAL (voir la section d'algèbre)

5.3 Divisibilité

Définition 5.1 *Soient $A, B \in \mathbb{Z}$ avec $A \neq 0$. Nous disons que A* divise *B* (sans restes) *s'il existe un entier q (le quotient) tel que :*

$$B = A \cdot q \tag{5.13}$$

auquel cas nous écrivons $A|B$. Dans le cas contraire, nous écrivons $A \nmid B$ et nous lisons A ne divise pas *B.*

Remarque Il ne faut pas oublier que le symbole | est une relation alors que le symbole / est une opération.

Par ailleurs, Si $A|B$, nous dirons aussi que B *est divisible par* A ou que B *est un multiple de* A. Dans le cas où $A|B$ et que $1 \leqslant 1 < B$, nous dirons que A est un *diviseur propre* de B. De plus, il est clair que $A|0$ quel que soit $A \in \mathbb{Z}\backslash\{0\}$ sinon quoi nous avons une singularité.

Voici maintenant quelques théorèmes élémentaires se rattachant à la divisibilité :

Théorème 5.2 *Si $A|B$ alors $A|BC$ quel que soit $C \in \mathbb{Z}$.*

Preuve Si $A|B$ alors il existe un entier q tel que $B = Aq$. Alors, $BC = A(qC)$ et ainsi $A|BC$. •

Théorème 5.3 *Si $A|B$ et $B|C$ alors $A|C$.*

Preuve Si $A|B$ et $B|C$ alors il existe des entiers q et r tels que $B = Aq$ et $C = Br$. Par conséquent $C = A(qr)$ et ainsi $A|C$. •

Théorème 5.4 *Si $A|B$ et $A|C$, alors $A|(Bx + Cy)$, $\forall x, y \in \mathbb{Z}$*

Preuve Si $A|B$ et $A|C$, alors il existe des entiers q et r tels que $B = Aq$ et $C = Ar$. Il s'ensuit que $Bx + Cy = (Aq)x + (Ar)y = A(qx + ry)$ et ainsi que $A|(Bx + Cy)$, $\forall x, y \in \mathbb{Z}$. •

Théorème 5.5 *Si $A|B$ et $B|A$, alors $A = \pm B$*

Preuve Si $A|B$ et $B|A$, alors il existe des entiers q et r tels que $B = Aq$ et $A = Br$. Nous avons donc $B = B(qr)$ et ainsi $qr = 1$; c'est pourquoi nous pouvons avoir $q = \pm 1$ si $r = \pm 1$ et qu'ainsi $a = \pm b$. •

Théorème 5.6 *Si $A|B$ et $B \neq 0$ alors $|A| \leqslant |B|$.*

Preuve Si $A|B$ et $B \neq 0$, alors il existe un entier $q \neq 0$ tel que $B = Aq$. Mais alors $|B| = |A||q| \geqslant |A|$ puisque $|q| \geqslant 1$. •

5.3.1 Division euclidienne

La division euclidienne est une opération qui, à deux entiers naturels appelés dividende et diviseur, associe deux entiers appelés quotient et reste. Initialement définie aux entiers naturels non nuls, elle se généralise aux entiers relatifs et aux polynômes, par exemple.

Définition 5.2 *Nous appelons* division euclidienne *ou* division entière *de deux nombres A et B l'opération consistant à diviser B par A en s'arrêtant quand le reste devient strictement inférieur à A.*

Rappelons que tout nombre qui n'a pas de diviseur euclidien est dit *nombre premier* et que tout couple de nombres qui n'ont que 1 comme diviseur euclidien commun sont dits *premiers entre eux*.

Théorème 5.7 (Division euclidienne) *Soit $A, B \in \mathbb{Z}$ avec $A > 0$; alors, il existe des entiers uniques q et r tels que $B = Aq + r \Leftrightarrow r = B - Aq$, où $0 \leqslant r < A$. De plus, si $A \nmid B$, alors $0 < r < A$.*

Preuve Considérons l'ensemble :

$$S = \{b - ma | m \in \mathbb{Z}, b - ma \geqslant 0\} \tag{5.14}$$

Il est facile de voir que $S \subset \mathbb{N} \cup \{0\}$ et que $S \neq \varnothing$, d'où nous concluons, d'après le principe du bon ordre, que S contient un plus petit élément $r \geqslant 0$. Soit q l'entier satisfaisant à $r = B - Aq$. Il reste à montrer que $r < A$. Supposons le contraire à nouveau, c'est-à-dire que $r \geqslant a$. Alors, dans ce cas, nous avons $B - aQ = r \geqslant A$, ce qui est équivalent à $B - (q + 1)A = r - A \geqslant 0$; mais $B - (q + 1)A \in S$ et $B - (q + 1)A \geqslant B - qA$, ce qui contredit le fait que $r = B - qA$ est le plus petit élément de S, donc $r < A$. Enfin, il est clair que si $r = 0$, nous avons $A|B$, d'où la seconde affirmation du théorème. •

REMARQUE Dans l'énoncé de la division euclidienne, nous avons supposé que $A > 0$. Qu'obtenons-nous lorsque $A < 0$? Dans cette situation, $-A$ est positif, et alors nous pouvons appliquer la division euclidienne à B et $-A$. Par conséquent, il existe des entiers q et r tels que :

$$B = q(-A) + r \quad \text{où} \quad 0 \leqslant r < |A| \tag{5.15}$$

Or, cette relation peut s'écrire $B = q(-A) + r$, où bien sûr, $-q$ est un entier. La conclusion est que la division euclidienne peut s'énoncer sous la forme plus générale suivante :

Soit $A, B \in \mathbb{Z}$, alors il existe des entiers q et r tels que $B = Aq + r$, où $0 \leqslant r < |A|$. De plus, si $A \nmid B$, alors $0 < r < |A|$.

Les entiers q et r sont dans la division euclidienne uniques. En effet, s'il existe deux autres entiers q_1 et r_1 tels que $B = Aq_1 + r_1$ avec toujours $0 \leqslant r_1 < |A|$, alors $A(q_1 - q) = r - r_1$ et ainsi $A|(r - r_1)$. En vertu de 5.6 nous avons, si $r - r_1 \neq 0$, $|r - r_1| \geqslant A$. Or, cette dernière inégalité est impossible puisque $-A < r - r_1 < A$. Donc, $r = r_1$ et, puisque $A \neq 0$, alors $q = q_1$; d'où l'unicité.

Plus grand commun diviseur

Soient $a, b \in \mathbb{Z}$ tels que $ab \neq 0$. Le *plus grand commun diviseur*, noté pgcd, de a et b, noté (a, b) est l'entier positif d qui satisfait aux deux propriétés suivantes :

1. $d|a$ et $d|b$
2. si $c|a$ et $c|b$ et $c \leqslant d$ alors $c|d$ (par définition)

REMARQUE Notons que 1 est un diviseur commun de deux entiers arbitraires. Cependant, il n'est pas évident que le pgcd autre qu'unitaire de deux entiers a et b existe toujours. Ce fait est démontré dans le théorème suivant dit *théorème de* Bézout.

Preuve Soient $a, b \in \mathbb{Z}$ tels que $ab \neq 0$. Alors, il existe des entiers x et y tels que :

$$d = (a, b) = ax + by \tag{5.16}$$

Cette relation est appelée *identité de* Bézout et il s'agit aussi d'une équation diophantienne linéaire (cf. chapitre de Calcul Algébrique).

Considérons l'ensemble $S = \{ax + by | x, y \in \mathbb{Z}, ax + by > 0\}$. Comme $S \subset \mathbb{N}$ et $S \neq \varnothing$, nous pouvons utiliser le principe du bon ordre et conclure que S possède un plus petit élément d. Nous pouvons alors écrire $d = ax_0 + by_0$ pour un certain choix $x_0, y_0 \in \mathbb{Z}$. Il suffit donc de montrer que $d = (a, b)$.

Supposons que $d \nmid a$. Alors, d'après la division euclidienne, il existe $q, r \in \mathbb{Z}$ tels que $a = qd + r$, où $0 < r < d$. Mais alors :

$$r = a - qd = q - q(ax_0 + by_0) = a(1 - qx_0) + b(-qy_0) \tag{5.17}$$

et donc $r \in S$ et $r < d$, ce qui contredit le fait que d est le plus petit élément possible de S. Donc $d|a$ et, de la même façon, nous démontrons que $d|b$. •

Corollaire 5.8 *Si $a, b \in \mathbb{Z}$ tels que $ab \neq 0$, alors :*

$$S = \{ax + by | x, y \in \mathbb{Z}\} \tag{5.18}$$

constitue l'ensemble de tous les multiples de $d = (a, b)$.

Preuve Comme $d|a$ et $d|b$ alors $d|ax + by$ pour tout $x, y \in \mathbb{Z}$. Soit $M = \{nd | n \in \mathbb{Z}\}$. Nous voulons montrer que $S = M$. Supposons $s \in S$, ce qui signifie que $d|s$ et implique $s \in M$. Soit un $m \in M$, alors $m = nd$ pour un certain $n \in \mathbb{Z}$. Comme $d = ax_0 + by_0$, pour un choix d'entiers quelconques $x_0, y_0 \in \mathbb{Z}$, nous avons bien $m = nd = n(ax_0 + by_0) = a(nx_0) + b(ny_0)$. •

REMARQUE Si au lieu de définir le pgcd de deux entiers non nuls, nous permettons à l'un d'entre eux d'être nul, disons $a \neq b$, $b = 0$ alors, nous avons $a|b$ et, selon notre définition du pgcd, il est clair que $(a, 0) = |a|$.

Soit $d = (a, b)$ et soit $m \in \mathbb{Z}$, alors nous avons les propriétés suivantes du pgcd :

Propriété 5.1 $(a, b + ma) = (a, b) = (a, -b)$

Propriété 5.2 $(am, bm) = |m|(a, b)$ *où* $m \neq 0$.

Propriété 5.3 $\left(\frac{a}{d}, \frac{b}{d}\right) = 1$

Propriété 5.4 *Si $g \in \mathbb{Z}\backslash\{0\}$ tel que $g|a$ et $g|b$ alors* $\left(\frac{a}{g}, \frac{b}{g}\right) = \frac{(a, b)}{|g|}$

Élaborons maintenant une méthode qui s'avérera très importante pour calculer le plus grand commun diviseur de deux entiers (utile en informatique parfois).

Algorithme d'Euclide

L'algorithme d'Euclide est un algorithme permettant de déterminer le plus grand commun diviseur de deux entiers.

Pour aborder cette méthode de manière intuitive, il faut savoir que vous devez comprendre un nombre entier comme une longueur, un couple d'entiers comme un rectangle et leur pgcd comme la taille du plus grand carré permettant de paver ce rectangle.

L'algorithme décompose le rectangle initial en carrés, de plus en plus petits, par divisions euclidiennes successives, de la longueur par la largeur, puis de la largeur par le reste, jusqu'à un reste nul. Il faut bien comprendre cette démarche géométrique pour comprendre ensuite l'algorithme.

Exemple 5.3 *Considérons que nous cherchons le pgcd (a, b) où $b = 21$ et $a = 15$ et gardons à l'esprit que le pgcd, outre le fait qu'il divise a et b, doit laisser un reste nul. En d'autres termes il doit pouvoir diviser le reste de la division de b par a aussi. Nous avons donc le rectangle de la figure 5.1. D'abord nous regardons si 15 est le pgcd (on commence toujours par le plus petit). Nous*

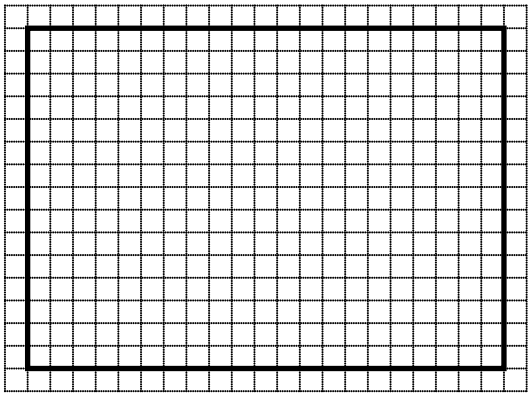


Figure 5.1 – Rectangle de 21×15

divisons alors 21 par 15 ce qui équivaut géométriquement à la figure 5.2. 15 n'est donc pas le pgcd

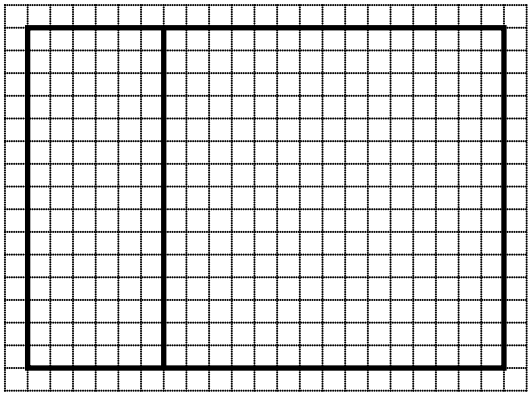


Figure 5.2 – Rectangle de 21×15 et division par 15

(on s'en doutait...). Nous voyons immédiatement que nous n'arrivons pas à paver le rectangle avec un carré de 15 par 15.

Nous avons donc un reste de 6 (rectange de gauche). Le pgcd comme nous le savons doit, s'il existe, par définition pouvoir diviser ce reste et laisser un reste nul.

Il nous reste donc un rectangle de 15 par 6. Nous cherchons donc maintenant à paver ce nouveau rectangle car nous savons que le pgcd est par construction inférieur ou égal à 6. Nous avons alors la figure 5.3. Et nous divisons donc 15 par le reste 6 (ce résultat sera inférieur à 6 et permet immédiatement de tester si le reste sera le pgcd). Nous obtenons la figure 5.4. À nouveau, nous n'arrivons pas à paver ce rectangle rien qu'avec des carrés. En d'autres termes, nous avons un reste non nul qui vaut 3. Soit un rectangle de 6 par 3. Nous cherchons donc maintenant à paver ce nouveau rectangle car nous savons que le pgcd est par construction inférieur ou égal à 3 et qu'il

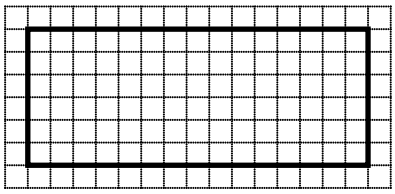


FIGURE 5.3 – Rectangle de 15×6

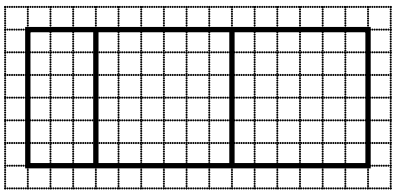


FIGURE 5.4 – Rectangle de 15×6 et division par 6



FIGURE 5.5 – Rectangle de 6×3

laissera un reste nul si il existe. Nous avons alors géométriquement la figure 5.5. Nous divisons 6 par 3 (ce qui sera inférieur à 3 et permet immédiatement de tester si le reste sera le pgcd) pour avoir la figure 5.6 et c'est terminé. Nous avons 3 qui laisse donc un reste nul et divise le reste 6 il

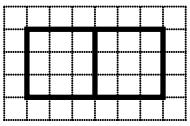


FIGURE 5.6 – Rectangle de 6×3 et division par 3

s'agit donc du pgcd. Nous avons donc au final la figure 5.7.

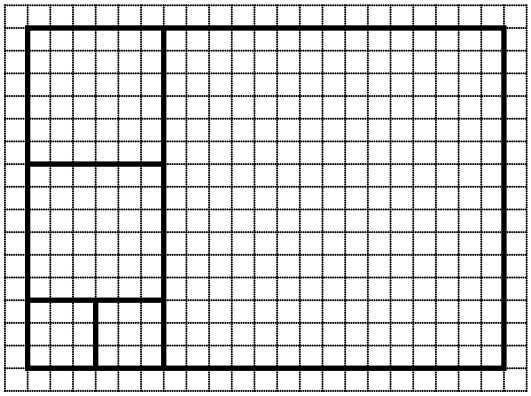


FIGURE 5.7 – Pavage final

Maintenant, voyons l'approche formelle équivalente : soit $a, b \in \mathbb{Z}$, où $a > 0$. En appliquant

successivement la division euclidienne (avec $b > a$), nous obtenons la suite d'équations :

$$\begin{aligned} b &= qa_1 + r_1 & 0 < r_1 < a \\ a &= r_1 q_2 + r_2 & 0 < r_2 < r_1 \\ r_1 &= r_1 2q_3 + r_3 & 0 < r_3 < r_2 \\ &\dots \\ r_{j-2} &= r_{j-1} q_j + r_j & 0 < r_j < r_{j-1} \\ r_{j-1} &= r_j q_{j+1} \end{aligned} \tag{5.19}$$

Si $d = (a, b)$, alors $d = r_j$.

Preuve Nous voulons d'abord montrer que $r_j = (a, b)$. Or, d'après la propriété $(a, b + ma) = (a, b) = (a, -b)$, nous avons $(a, b) = (a, r_1) = (r_1, r_2) = \dots = (r_{j-1}, r_j)$. Pour démontrer la deuxième propriété de l'algorithme d'EUCLIDE, nous écrivons l'avant dernière équation du système sous la forme $r_j = r_{j-2} - q_j r_{j-1}$. Or, en utilisant l'équation qui précède cette avant dernière équation du système, nous avons :

$$r_j = r_{j-2} - q_j(r_{j-3} - q_{j-1} r_{j-2}) = (1 + q_j q_{j-1}) r_{j-2} + (-q_j) r_{j-3} \tag{5.20}$$

En continuant ce processus, nous arrivons à exprimer r_j comme une combinaison linéaire de a et b. •

Exemple 5.4 *Calculons le plus grand commun diviseur de* 966 429 *et exprimons ce nombre comme une combinaison linéaire de* 966 *et de* 429. *Nous appliquons l'algorithme d'*EUCLIDE *:*

$$\begin{aligned} 966 &= 429 \times 2 + 108 \\ 429 &= 108 \times 3 + 105 \\ 108 &= 105 \times 1 + 3 \\ 105 &= 35 \times 3 \end{aligned} \tag{5.21}$$

Nous en déduisons donc que :

$$r_j = d = (966, 429) = 3 \tag{5.22}$$

et, de plus, que :

$$\begin{aligned} 3 &= 108 - 105 = 108 - (429 - 108 \times 3) = 108 \times 4 - 429 \\ &= (966 - 429 \times 2) \times 4 - 429 \\ &= 966 \times 4 - 429 \times 9 = 966 \times 4 + 429 \times (-9) \end{aligned} \tag{5.23}$$

Donc le pgcd est bien exprimé comme une combinaison linéaire de a et b.

Définition 5.3 *Nous disons que les entiers $a_1, a_2, \dots, a_n$ sont* relativement premiers *si :*

$$(a_1, a_2, \dots, a_n) = 1 \tag{5.24}$$

Plus petit commun multiple

Définition 5.4 *Soit $a_1, a_2, \dots, a_n \in \mathbb{Z}\backslash\{0\}$, nous disons que m est un* commun multiple *de $a_1, a_2, \dots, a_n$ si $a_i | m$ pour $i = 1, 2, \dots, n$.*

Définition 5.5 *Soit $a_1, a_2, \dots, a_n \in \mathbb{Z}\backslash\{0\}$, nous appelons* plus petit commun multiple *(ppcm) de $a_1, a_2, \dots, a_n$, noté $[a_1, a_2, \dots, a_n]$, le plus petit entier positif par tous les communs multiples de $a_1, a_2, \dots, a_n$.*

REMARQUE Soit $a_1, a_2, \dots, a_n \in \mathbb{Z}\backslash\{0\}$, alors, le plus petit commun multiple existe. En effet, considérons l'ensemble :

$$E = \{m \in \mathbb{N} | a_i | m, i = 1, 2, \dots, n\} \tag{5.25}$$

Puisque $|a_1 a_2 \dots a_n| \in E$, alors l'ensemble est non vide et, d'après l'axiome du bon ordre, l'ensemble E contient un plus petit élément positif.

Voyons maintenant quelques théorèmes relatifs au ppcm :

Théorème 5.9 *T1. Si m est un commun multiple de $a_1, a_2, \ldots, a_n$ alors $[a_1, a_2, \ldots, a_n]|m$.*

Preuve Soit $M = [a_1, a_2, \ldots, a_n]$. Alors, d'après la division euclidienne, il existe des entiers q et r tels que :

$$m = qL + r, \quad 0 \leqslant r < M \tag{5.26}$$

Il suffit de montrer par l'absurde que $r = 0$. Supposons $r \neq 0$. Puisque $a_i|m$ et $a_i|M$, alors on a $a_i|r$ et cela pour $i = 1, 2, \ldots, n$. Donc r est un commun multiple de $a_1, a_2, \ldots, a_n$ plus petit que le ppcm On vient d'obtenir une contradiction, ce qui prouve le théorème. •

Théorème 5.10 *T2. Si $k > 0$, alors $[ka_1, ka_2, \ldots, ka_n] = k[a_1, a_2, \ldots, a_n]$.*

La démonstration est évidente.

Théorème 5.11 $[a, b] \cdot (a, b) = |ab|$

Preuve Pour la démonstration, nous allons utiliser le *lemme d'*Euclide qui dit que si $a|bc$ et $(a, b) = 1$ alors $a|c$.

Effectivement cela se vérifie aisément car nous avons vu qu'il existe $a, y \in \mathbb{Z}$ tels que $1 = ax + by$, alors $c = acx + bcy$. Mais $a|ac$ et $a|bc$ impliquent que $a|acx + bcy$, c'est-à-dire également que $a|c$.

Revenons à notre théorème :

Puisque $(a, b) = (a, -b)$ et $[a, b] = [a, -b]$, il suffit de prouver le résultat pour des entiers positifs a et b. En tout premier lieu, considérons le cas où $(a, b) = 1$. L'entier $[a, b]$ étant un multiple de a, nous pouvons écrire $[a, b] = ma$. Ainsi, nous avons $b|ma$ et, puisque $(a, b) = 1$, il s'ensuit, d'après le lemme d'EUCLIDE, que $b|m$. Donc, $b \leqslant m$ et alors $ab \leqslant am$. Mais ab est un commun multiple de a et b qui ne peut être plus petit que le ppcm ; c'est pourquoi $ab = ma = [a, b]$.

Pour le cas général, c'est-à-dire $(a, b) = d > 1$, nous utilisons la propriété 5.3 couplée au résultat précédent, soit :

$$\left[\frac{a}{d}, \frac{b}{d}\right] \cdot \left(\frac{a}{d}, \frac{b}{d}\right) = \frac{a}{d}\frac{b}{d} \tag{5.27}$$

Lorsque nous multiplions des deux côtés de l'équation par d^2, le résultat suit et la démonstration est effectuée. •

5.3.2 Théorème fondamental de l'arithmétique

Le théorème fondamental de l'arithmétique dit que tout nombre naturel $n > 1$ peut s'écrire comme un produit de nombres premiers, et cette représentation est unique, à part l'ordre dans lequel les facteurs premiers sont disposés.

Le théorème établit l'importance des nombres premiers. Essentiellement, ils sont les briques élémentaires de construction des entiers positifs, chaque entier positif contenant des nombres premiers d'une manière unique.

REMARQUE Ce théorème est parfois appelé *théorème de factorisation* un peu à tort puisque d'autres théorèmes portent le même nom.

Preuve Si n est premier, alors la preuve est terminée. Supposons que n n'est pas premier et considérons l'ensemble $D = \{d|\ d|n \text{ et } 1 < d < n\}$. Alors, $D \subset \mathbb{N}$ et, puisque n est composé, nous avons $D \neq \varnothing$. D'après le principe du bon ordre, D possède un plus petit élément p_1 qui est premier, sans quoi le choix minimal de p_1 serait contredit. Nous pouvons donc écrire $n = p_1 n_1$. Si n_1 est premier, alors la preuve est terminée. Si n_1 est composé, alors nous répétons le même argument que précédemment et nous en déduisons l'existence d'un nombre premier p_2 et d'un entier $n_2 < n_1$ tels que $n = p_1 p_2 n_2$. En poursuivant ainsi nous arrivons forcément à la conclusion que n_k sera premier.

Donc finalement nous avons bien démontré qu'un nombre quelconque est décomposable en facteurs de nombres premiers à l'aide du principe du bon ordre. •

Nous ne connaissons pas à ce jour de loi simple qui permette de calculer le n^{e} facteur premier (p_n). Ainsi, pour savoir si un entier m est premier, il est pratiquement plus facile à ce jour de vérifier sa présence dans une table de nombres premiers.

En fait, nous utilisons aujourd'hui la méthode suivante : soit un nombre m, si nous voulons déterminer s'il est premier ou non, nous calculons s'il est divisible par les nombres premiers p_n qui appartiennent à l'ensemble :

$$\{p_n \in \mathbb{N} | p_n < \sqrt{m}\} \tag{5.28}$$

Exemple 5.5 *L'entier* 223 *n'est divisible par* 2, *ni par* 3, *ni par* 5, *ni par* 7, *ni par* 11, *ni par* 13. *Il est inutile de continuer avec le prochain nombre premier, car* $17^2 = 289 > 223$. *Nous en déduisons dès lors que le nombre* 223 *est premier.*

5.3.3 Congruences

Définition 5.6 *Soit* $m \in \mathbb{Z}\backslash\{0\}$. *Si* a *et* b *ont même reste dans la division euclidienne par* m *nous disons que* a est congru à b modulo m *et nous écrivons :*

$$a = b \bmod(m) \tag{5.29}$$

ou de manière équivalente il existe un nombre entier relatif k *tel que :*

$$a = b + km \tag{5.30}$$

Le lecteur pourra vérifier que cela impose que

$$m|(a - b) \tag{5.31}$$

soit que m *divise la différence entre* a *et* b. *Dans le cas contraire, nous disons que* a est non congru à b modulo m.

L'étude de ces propriétés qui relient trois nombres entre eux est appelée communément *l'arithmétique modulaire.*

Remarques

1. la congruence implique un reste nul pour la division ;
2. sont exclues des valeurs prises par m dans la définition de la congruence, en plus de 0, les valeurs 1 et -1 ;
3. derrière le terme de congruence se cachent des notions semblables mais de niveaux d'abstraction différents :
   - en arithmétique modulaire, nous disons donc que *deux entiers relatifs* a *et* b *sont congrus modulo* m *s'ils ont même reste dans la division euclidienne par* m. Nous pouvons aussi dire qu'ils sont congrus modulo m si leur différence est un multiple de m ;
   - dans la mesure des angles orientés, nous disons que *deux mesures sont congrues modulo* 2π *rad si et seulement si leur différence est un multiple de* 2π *rad.* Cela caractérise deux mesures d'un même angle (cf. chapitre de Trigonométrie) ;
   - en algèbre, nous parlons de congruence modulo I dans un anneau commutatif• dont I est un idéal : *x est congru à y modulo I si et seulement si leur différence appartient I*. Cette congruence est une relation d'équivalence, compatible avec les opérations d'addition et multiplication et permet de définir un anneau quotient de l'ensemble parent avec son idéal I ; Chap4
   - nous trouvons parfois, dans l'étude de la géométrie (cf. chapitre de Géométrie Euclidienne) le terme de congru mis à la place de semblable. Il s'agit alors d'une simple relation d'équivalence sur l'ensemble des figures planes.

La relation de congruence $\equiv$ est une relation d'équivalence•, en d'autres termes, soient $a, b, c, m \in \mathbb{Z}$, $m > 1$ alors la relation de congruence est : Chap3

Propriété 5.5 *Réflexive :* $a \equiv a \bmod(m)$

Propriété 5.6 *Symétrique :* $a \equiv b \bmod(m) \Leftrightarrow b \equiv a \bmod(m)$

Propriété 5.7 *Transitive :* $a \equiv b \bmod(m)$, $b \equiv c \bmod(m) \Rightarrow a \equiv c \bmod(m)$

Preuve Les propriétés 5.5 et 5.6 sont évidentes. Concernant la propriété 5.7, les hypothèses impliquent que $b = a + km$ et $c = b + lm$ d'où :

$$c = b + lm = (a + km) + lm = a + (k + l)m \tag{5.32}$$

ce qui montre que a et c sont congrus modulo m. •

La relation de congruence $\equiv$ est compatible avec la somme et le produit. Effectivement, soient $a, b, c, m \in \mathbb{Z}$, $m > 1$ tels que $a = b \bmod(m)$ et $a' = b' \bmod(m)$ alors :

Propriété 5.8 $a + a' \equiv b + b' \bmod(m)$

Propriété 5.9 $a \cdot a' \equiv b \cdot b' \bmod(m)$

Preuve Nous avons :

$$a = b + km, \quad a' = b' + lm \tag{5.33}$$

par hypothèse, d'où :

$$a + a' = b' + b + (k + l)m \tag{5.34}$$

ce qui démontre la propriété 5.8. Nous avons également :

$$aa' = bb' + blm + b'km + klm^2 = bb' + sm \tag{5.35}$$

ce qui démontre la propriété 5.9. •

REMARQUE La relation de congruence se comporte sur de nombreux points comme la relation d'égalité. Néanmoins une propriété de la relation d'égalité n'est plus vraie pour celle de congruence, à savoir la simplification : si $ab \equiv ac \bmod(m)$, nous n'avons pas nécessairement $b \equiv c \bmod(m)$.

Exemple 5.6 $2 \times 1 \equiv 2 \times 3 \bmod(4)$ *mais* $1 \not\equiv 3 \bmod(4)$.

Jusqu'ici, nous avons vu des propriétés des congruences faisant intervenir un seul modulus. Nous allons maintenant étudier le comportement de la relation de congruence lors d'un changement de modulus.

Propriété 5.10 *Si* $a \equiv b \bmod(m)$ *et* $d|m$ *alors* $a \equiv b \bmod(d)$.

Propriété 5.11 *Si* $a \equiv b \bmod(r)$ *et* $a \equiv b \bmod(s)$ *alors* a *et* b *sont congrus modulo* $[r, s]$.

Preuve Ces deux propriétés sont évidentes et il est inutile d'aller dans les détails pour 5.10. Pour 5.11, puisque $b - a$ est un multiple de r et de s puisque par hypothèse :

$$\frac{b-a}{r} = k, \quad \frac{b-a}{s} = l \quad \Rightarrow \quad b - a = rk = sl \tag{5.36}$$

$b - a$ est donc un multiple du ppcm de r et s, ce qui démontre la propriété 5.11. •

De ces propriétés il vient que si nous désignons par $f(x)$ un polynôme à coefficient entiers (positifs ou négatifs) :

$$f(x) = Ax^n + B^{n-1} + Cx^{n-2} + \ldots + Kx + L \tag{5.37}$$

La congruence $a \equiv b \bmod(m)$ donnera aussi $f(a) \equiv f(b) \bmod(m)$.

Si nous remplaçons x successivement par tous les nombres entiers dans un polynôme $f(x)$ à coefficients entiers, et si nous prenons les résidus pour le module m, ces résidus se reproduisent de m en m (dans le sens où la congruence se vérifie), puisque nous avons, quel que soit l'entier m et x $f(x) \equiv f(x + m) \bmod(m)$. Nous en déduisons alors l'impossibilité de résoudre la congruence $f(x) \equiv r \bmod(m)$ en nombres entiers, si r désigne un des non-résidus —un résidu qui ne satisfait pas la congruence.

Classes de congruence

Définition 5.7 *Nous appelons* classe de congruence modulo m, *le sous-ensemble de l'ensemble* $\mathbb{Z}$ *défini par la propriété que deux éléments* a *et* b *de* $\mathbb{Z}$ *sont dans la même classe si et seulement si* $a \equiv b \bmod(m)$ *ou qu'un ensemble d'éléments entre eux sont congrus par ce même modulo.*

REMARQUE Nous avons vu dans le chapitre traitant des opérateurs qu'il s'agit en fait d'une classe d'équivalence car la congruence modulo m est, comme nous l'avons démontré plus haut, une relation d'équivalence.

Exemple 5.7 *Soit $m = 3$. Nous divisons l'ensemble des entiers en classes de congruence modulo 3. Voici trois ensembles dont tous les éléments sont congrus entre eux sans reste —observez bien ce que donne l'ensemble des classes :*

$$\begin{aligned}&\{\ldots,-9,-6,-3,0,3,6,9,\ldots\}\\&\{\ldots,-8,-5,-2,1,4,7,10,\ldots\}\\&\{\ldots,-7,-4,-1,2,5,8,11,\ldots\}\end{aligned}\tag{5.38}$$

Ainsi, nous voyons que pour chaque couple d'élément d'une classe de congruence, la congruence modulo 3 existe. Cependant, nous voyons que nous ne pouvons pas prendre $-9 \equiv -8 \bmod(3)$ où -9 se trouve dans le première classe et -8 dans le seconde.

Le plus petit nombre non négatif de la première classe est 0, celui de la deuxième est 1 et celui de la dernière est 2. Ainsi, nous noterons ces trois classes respectivement $[0]_3$, $[1]_3$ et $[2]_3$, le chiffre 3 en indice indiquant le modulus.

Il est intéressant de noter que si nous prenons un nombre quelconque de la première classe et un nombre quelconque de la deuxième, alors leur somme est toujours dans la deuxième classe. Ceci se généralise et permet de définir une somme sur les classes modulo 3 en posant :

$$\begin{aligned}&[0]_3+[0]_3=[0]_3;\quad [0]_3+[1]_3=[1]_3\\&[0]_3+[2]_3=[2]_3;\quad [1]_3+[1]_3=[2]_3\\&[1]_3+[2]_3=[0]_3;\quad [2]_3+[2]_3=[1]_3\end{aligned}\tag{5.39}$$

Ainsi que :

$$\begin{aligned}&[0]_3\times[0]_3=[0]_3;\quad [0]_3\times[1]_3=[0]_3\\&[0]_3\times[2]_3=[0]_3;\quad [1]_3\times[1]_3=[1]_3\\&[1]_3\times[2]_3=[2]_3;\quad [2]_3\times[2]_3=[1]_3\end{aligned}\tag{5.40}$$

Ainsi, pour tout $m > 1$, la classe de congruence de $a \bmod(m)$ est l'ensemble des entiers congrus à a modulo m (et congrus entre eux modulo m). Cette classe est notée $[a]_m$.

Remarque Le fait d'avoir mis entre parenthèse l'expression *et congrus entre eux modulo m* est du au fait que la congruence, étant une relation d'équivalence nous avons comme nous l'avons démontré plus haut que si $b \equiv a \bmod(m)$ et $c \equiv a \bmod(m)$ alors $b \equiv c \bmod(m)$.

Définition 5.8 *L'ensemble des classes de congruences $[a]_m$ —qui forment, par le fait que la congruence est une relation d'équivalence, des* classes d'équivalences*—, pour un m fixe donne un* ensemble quotient•*. Plus rigoureusement, nous parlons de* l'ensemble quotient de $\mathbb{Z}$ par la relation de congruence *dont les éléments sont les classes de congruences (ou : classes d'équivalences) et qui forment alors l'anneau $\mathbb{Z}/m\mathbb{Z}$.* Chap3

Nous déduisons de la définition les deux propriétés triviales suivantes :

Propriété 5.12 *Le nombre b est dans la classe $[a]_m$ si et seulement si $a \equiv b \bmod(m)$.*

Propriété 5.13 *Les classes $[a]_m$ et $[b]_m$ sont égales si et seulement si $a \equiv b \bmod(m)$.*

Montrons maintenant qu'il y a exactement m différentes classes de congruence modulo m, à savoir $[0]_m, [1]_m, \ldots, [m-1]_m$.

Preuve Soit $m > 1$, alors tout nombre entier a est congru modulo m à un et un seul entier r de l'ensemble $\{0, 1, 2, \ldots, m-1\}$ [1]. De plus, cet entier r est exactement le reste de la division de a par m. En d'autres termes, si $0 \leqslant r < m$, alors $a \equiv r \bmod(m)$ si et seulement si $a = qm + r$ où q est le quotient de a par m et r le reste. La démonstration est donc une conséquence immédiate de la définition de la congruence et de la division euclidienne. •

Définition 5.9 *Un entier b dans une classe de congruence modulo m est appelé* représentant de cette classe [2].

1. Il est important de remarquer que nous nous restreignons aux entiers positifs ou nuls sans prendre en compte les négatifs.
2. Il est clair que par la relation d'équivalence que deux représentants d'une même classe sont donc congrus entre eux modulo m.

Nous allons pouvoir maintenant définir une addition et une multiplication sur les classes de congruences. Pour définir la somme de deux classes $[a]_m$ et $[b]_m$, il suffit de prendre un représentant de chaque classe, de faire leur somme et de prendre la classe de congruence du résultat. Ainsi, pour la somme :

$$[a]_m + [b]_m = [a+b]_m \tag{5.41}$$

et pour la multiplication :

$$[a]_m \cdot [b]_m = [a \cdot b]_m \tag{5.42}$$

Par définition de la somme et du produit, nous constatons que la classe de 0 est l'élément neutre pour l'addition :

$$[a]_m + [0]_m = [a]_m, \quad \forall a \in \mathbb{Z}, \forall m \in \mathbb{N} \tag{5.43}$$

et la classe de l'entier 1 est l'élément neutre pour la multiplication :

$$[a]_m \cdot [1]_m = [a]_m, \quad \forall a \in \mathbb{Z}, \forall m \in \mathbb{N} \tag{5.44}$$

Définition 5.10 *Un élément $[a]_m$ de $\mathbb{Z}/m\mathbb{Z}$ est* une unité *s'il existe un élément $[b]_m \in \mathbb{Z}/m\mathbb{Z}$ tel que $[a]_m \cdot [b]_m = 1$.*

Le théorème suivant permet de caractériser les classes modulo m qui sont des unités dans $\mathbb{Z}/m\mathbb{Z}$:

Théorème 5.12 *Soit $[a]$ un élément de $\mathbb{Z}/m\mathbb{Z}$, alors $[a]$ est une unité si et seulement si $(a,m) = 1$.*

Preuve Supposons d'abord que $(a,m) = 1$. Alors par BÉZOUT, nous avons son identité $as + mr = 1$. Autrement dit, as est congru à 1 modulo m. Mais ceci est équivalent à écrire par définition que $[a][s] = [1]$, ce qui montre que $[a]$ est une unité. Réciproquement, si $[a]$ est une unité, alors il existe une classe $[s]$ telle que $[a][s] = [1]$.

Ainsi, nous venons de démontrer que $\mathbb{Z}/m\mathbb{Z}$ constitue bien un anneau puisqu'il possède une addition, une multiplication, un élément neutre et un inverse. •

5.3.4 Fractions continues

La notion de fraction continue remonte à l'époque de FERMAT et atteint son apogée avec les travaux de LAGRANGE et LEGENDRE vers la fin du 18^{e} siècle. Ces fractions sont importantes en physique car nous les retrouvons en acoustique ainsi que dans la démarche intellectuelle qui a amené GALOIS à créer sa théorie des groupes.

Considérons dans un premier temps le nombre rationnel a/b avec $(a,b) = 1$, $b > 0$ et $a > 0$. Nous savons que tous les quotients q_i et les restes r_i sont dans le cadre de la division euclidienne des entiers positifs.

Rappelons l'algorithme d'EUCLIDE :

$$\begin{aligned}
&\frac{a}{b} = q_1 + \frac{r_1}{b} \\
&\frac{b}{r_1} = q_2 + \frac{r_2}{r_1} \\
&\frac{r_1}{r_2} = q_3 + \frac{r_3}{r_2} \\
&\dots \\
&\frac{r_{n-2}}{r_{n-1}} = q_n
\end{aligned} \tag{5.45}$$

Par substitutions successives, nous obtenons :

$$\begin{aligned}\frac{a}{b} &= q_1 + \frac{1}{b/r_1} = q_1 + \frac{1}{q_2 + \frac{1}{r_1/r_2}} \\ &= \dots \\ &= q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3 + \cfrac{1}{\ddots \cfrac{1}{q_{n-1} + \cfrac{1}{q_n}}}}}\end{aligned} \tag{5.46}$$

Ce qui est aussi parfois noté :

$$\frac{a}{b} = q_1 + \frac{1}{1/\varepsilon_1} = q_1 + \frac{1}{q_2 + 1/\varepsilon_2} = \dots \tag{5.47}$$

Ainsi, tout nombre rationnel positif peut s'exprimer comme une fraction continue finie où $q_n \in \mathbb{N}$.

Exemple 5.8 *Cherchons l'expression de* $17/49$. *Nous savons déjà que* $0 < 17/49 < 1$ *donc que* $q_1 = 0$. *Nous avons alors :*

$$\frac{17}{49} = 0 + \frac{1}{49/17} = 0 + \frac{1}{2 + \frac{1}{17/15}} = 0 + \frac{1}{2 + \frac{1}{1 + \frac{1}{15/2}}} = 0 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{7 + \frac{1}{2/1}}}} \tag{5.48}$$

Nous voyons bien dans cet exemple que nous avons effectivement $q_n \in \mathbb{N}$. *Nous pouvons également remarquer que par construction :*

$$q_n = \left[\frac{1}{\varepsilon_{n-1}}\right] \tag{5.49}$$

où les crochets représentent la partie entière et nous avons aussi :

$$\frac{1}{\varepsilon_{n-1}} = q_n + \varepsilon_n \tag{5.50}$$

Exemple 5.9 *Voyons comment extraire la racine carrée d'un nombre* A *par la méthode des fractions continues. Soit* a *le plus grand nombre entier dont le carré* a^2 *est plus petit que* A. *On le soustrait de* A. *Il y a donc un reste de :*

$$r = A - a^2 = (\sqrt{A} - a)(\sqrt{A} + a) \tag{5.51}$$

En divisant les deux membres par la deuxième parenthèse, nous avons :

$$\sqrt{A} - a = \frac{r}{\sqrt{A} + a} \tag{5.52}$$

Soit :

$$\sqrt{A} = a + \frac{r}{\sqrt{A} + a} \tag{5.53}$$

Dans le dénominateur de l'équation 5.53, nous remplaçons $\sqrt{A}$ *par le membre de droite :*

$$\sqrt{A} = a + \cfrac{r}{2a + \cfrac{r}{\sqrt{A} + a}} \tag{5.54}$$

etc. On voit ainsi que le système est simple pour déterminer l'expression d'une racine en termes de fraction continue.

Le développement du nombre a/b s'appelle le *développement du nombre a/b en fraction continue finie* et est condensé sous la notation $[q_1, q_2, q_3, \ldots, q_n]$.

Nous considérerons comme intuitif que tout nombre rationnel peut s'exprimer comme fraction continue finie et inversement que toute fraction continue finie représente un nombre rationnel. Par extension, un nombre irrationnel est représenté par une fraction continue infinie.

Considérons maintenant $[q_1, q_2, q_3, \ldots, q_n]$ une fraction continue finie. La fraction continue $C_k = [q_1, q_2, q_3, \ldots, q_k]$ où $k = 1, 2, \ldots, n$ est appelée la k^e *réduite* ou la k^e *convergente* ou encore le k^e *quotient partiel*. Avec cette notation, nous avons :

$$\begin{aligned}
C_1 &= [q_1] = q_1 \\
C_2 &= [q_1, q_2] = q_1 + \frac{1}{q_2} = \frac{q_1 q_2 + 1}{q_2} \\
C_3 &= [q_1, q_2, q_3] = q_1 + \frac{1}{q_2 + \dfrac{1}{q_3}} = \frac{(q_1 q_2 + 1) \cdot q_3 + q_1}{q_2 q_3 + 1} \\
C_4 &= [q_1, q_2, q_3, q_4] = \frac{((q_1 q_2 + 1) \cdot q_3 + q_1) \cdot q_4 + (q_1 q_2 + 1)}{(q_2 q_3 + 1) \cdot q_4 + q_2}
\end{aligned} \tag{5.55}$$

Pour simplifier les expressions ci-dessus, nous introduisons les suites $\{n_i\}$, $\{q_i\}$ (n pour numérateur et d pour dénominateur) définies par :

$$\begin{aligned}
&n_0 = 1, \quad n_1 = q_1, \ldots, \quad n_i = q_i n_{i-1} + n_{i-2} \\
&d_0 = 0, \quad d_1 = 1, \ldots, \quad d_i = q_i d_{i-1} + d_{i-2}
\end{aligned} \tag{5.56}$$

à l'aide de cette construction, nous avons une petite inégalité immédiate intéressante :

$$0 = d_0 < d_1 < d_2 < d_3 < \ldots \tag{5.57}$$

Avec la définition ci-dessus, nous constatons que :

$$C_1 = \frac{n_1}{d_1}, \quad C_2 = \frac{n_2}{d_2}, \quad C_3 = \frac{n_3}{d_3}, \ldots, C_i = \frac{n_i}{d_i} \tag{5.58}$$

Soit en généralisant :

$$C_k = [q_1, q_2, q_3, \ldots, q_k] = \frac{n_k}{d_k} \tag{5.59}$$

Maintenant, montrons pour un usage ultérieur que pour $i \geqslant 1$, nous avons :

$$n_i d_{i-1} - d_i n_{i-1} = (-1)^i \tag{5.60}$$

Preuve Le résultat est immédiat pour $n = 1$. En supposant que le résultat est vrai pour i, montrons qu'il est aussi vrai pour $i + 1$:

$$n_{i+1} d_i - d_{i+1} n_i = (q_{i+1} n_i + n_{i-1}) d_i - (q_{i+1} d_i + d_{i-1}) n_i = n_{i-1} d_i - d_{i-1} n_i = (-1)^{i+1} \tag{5.61}$$

alors en utilisant l'hypothèse d'induction, nous obtenons le résultat. •

Nous pouvons maintenant établir une relation indispensable pour la suite : montrer que si C_k est la k^{e} réduite de la fraction continue simple finie $[q_1, q_2, q_3, \ldots, q_n]$ alors :

$$C_1 < C_3 < C_5 < \ldots < C_6 < C_4 < C_2 \tag{5.62}$$

Preuve Soit :

$$\begin{aligned}
C_{k+2} - C_k &= (C_{k+2} - C_{k+1}) + (C_{k+1} - C_k) = \left(\frac{n_{k+2}}{d_{k+2}} - \frac{n_{k+1}}{d_{k+1}}\right) + \left(\frac{n_{k+1}}{d_{k+1}} - \frac{n_k}{d_k}\right) \\
&= \left(\frac{n_{k+2} d_{k+1} - n_{k+1} d_{k+2}}{d_{k+2} d_{k+1}}\right) + \left(\frac{n_{k+1} d_k - n_k d_{k+1}}{d_{k+1} d_k}\right) = \frac{(-1)^{k+2}}{d_{k+2} d_{k+1}} + \frac{(-1)^{k+1}}{d_{k+1} d_k} \\
&= \frac{(-1)^{k+2} d_k + (-1)^{k+1} d_{k+2}}{d_{k+2} d_{k+1} d_k} = \frac{(-1)^{k+1}(-1) d_k + (-1)^{k+1} d_{k+2}}{d_{k+2} d_{k+1} d_k} \\
&= \frac{(-1)^{k+1}(d_{k+2} - d_k)}{d_{k+2} d_{k+1} d_k}
\end{aligned} \tag{5.63}$$

Or, d'après (5.57), $d_{k+2} - d_k > 0$, ce qui nous indique que le signe $C_{k+2} - C_k$ est le même que celui de $(-1)^{k+1}$. Il en résulte que $C_{k+2} > C_k$ pour k impair, et que $C_{k+2} < C_k$ pour k pair. Il s'ensuit que :

$$C_1 < C_3 < C_5 < \ldots \quad \text{et} \quad C_2 > C_4 > C_6 > \ldots \tag{5.64}$$

Ensuite, puisque :

$$C_k - C_{k-1} = \frac{n_k}{d_k} - \frac{n_{k-1}}{d_{k-1}} = \frac{n_k d_{k-1} - n_{k-1} d_k}{d_k d_{k-1}} = \frac{(-1)^k}{d_k d_{k-1}} \tag{5.65}$$

Donc pour k pair, nous avons $c_k > C_{k-1}$, nous en déduisons (5.62). •

Montrons maintenant que toute fraction continue infinie peut représenter un nombre irrationnel quelconque.

En termes formels, si $\{q_n\}$ est une suite d'entiers tous positifs et que nous considérons $C_n = [q_1, q_2, q_3, \ldots, q_n]$ alors celui-ci converge nécessairement vers un nombre réel quand $n \to \infty$. Effectivement, il n'est pas difficile d'observer avec un exemple pratique que nous avons $C_n - C_{n-1} \to 0$ lorsque $n \to \infty$.

Preuve Notons x un nombre réel quelconque et $q_1 = [x]$, la partie entière de ce nombre réel. Alors nous avons vu tout au début de notre étude des fractions continues que $x = q_1 + \varepsilon_1$. Il vient donc $\varepsilon_1 = x - q_1 +$. Attardons nous pour les nécessités du chapitre d'Acoustique sur le calcul d'une fraction continue d'un logarithme en utilisant la relation précédente. D'abord rappelons que :

$$\frac{a}{b} = q_1 + \frac{1}{1/\varepsilon_1} = q_1 + \frac{1}{q_2 + 1/\varepsilon_2} = \ldots \tag{5.66}$$

Soit (relation démontrée dans le chapitre d'Analyse fonctionnelle) :

$$x = \log_a(u) = \frac{\ln(u)}{\ln(a)} \tag{5.67}$$

avec $1 < a < u$ et $(a, u) = 1$. Soit y_n défini par :

$$y_{-1} = u \quad y_0 = a \quad y_{n+1} = \frac{y_{n-1}}{y_n^{q_1}} \tag{5.68}$$

Alors montrons que :

$$\varepsilon_n = \frac{\ln(y_{n-2})}{\ln(y_{n-1})} - q_n \tag{5.69}$$

En effet, pour $n = 1$ nous avons :

$$\varepsilon_1 = x - q_1 = \frac{\ln(u)}{\ln(a)} - q_1 = \frac{\ln(y_{n-2})}{\ln(y_{n-1})} - q_n \tag{5.70}$$

pour $n = 2$ nous avons :

$$\begin{aligned} \varepsilon_1 = x - q_1 = \frac{\ln(u)}{\ln(a)} - q_1 = \frac{\ln(u) - q_1 \ln(a)}{\ln(a)} = \frac{\ln(u) - \ln(a^{q_1})}{\ln(a)} \\ = \frac{\ln\left(\frac{u}{a^{q_1}}\right)}{\ln(a)} = \frac{\ln\left(\frac{y_{-1}}{y_0^{q_1}}\right)}{\ln(a)} = \frac{\ln(y_1)}{\ln(a)} \end{aligned} \tag{5.71}$$

donc :

$$\frac{1}{\varepsilon_1} = \frac{\ln(a)}{\ln(y_1)} \tag{5.72}$$

et puisque nous avions montré que :

$$\frac{1}{\varepsilon_{n-1}} = q_n + \varepsilon_n \Rightarrow \varepsilon_2 = \frac{1}{\varepsilon_1} - q_2 = \frac{\ln(a)}{\ln(y_1)} - q_2 = \frac{\ln(y_0)}{\ln(y_1)} - q_2 \tag{5.73}$$

et ainsi de suite par récurrence, ce qui démontre notre droit d'utiliser ce changement d'écriture. •

Exemple 5.10 *Cherchons l'expression de la fraction continue de* $x = \log_a(u) = \log_2(3)$. *Nous savons, en jouant avec la définition du logarithme, que* $2^1 < 3 < 2^2$ *donc :*

$$\log_2(2^1) < \log_2(3) < \log_2(2^2) \Rightarrow \log_2(2) < \log_2(3) < 2\log_2(2) \Rightarrow 1 < \log_2(3) < 2 \tag{5.74}$$

donc $q_1 = 1$. *Nous avons alors :*

$$\varepsilon_1 = \frac{\ln(y_{-1})}{\ln(y_0)} - q_1 \tag{5.75}$$

et puisque $y_{-1} = u = 3$ *et* $y_0 = a = 2$, *il vient :*

$$\varepsilon_1 = \frac{\ln(3)}{\ln(2)} - 1 \tag{5.76}$$

Donc nous avons le premier quotient partiel :

$$\frac{a}{b} = \log_2(3) = q_1 + \frac{1}{1/\varepsilon_1} = 1 + \frac{1}{\frac{1}{\frac{\ln(3)}{\ln(2)} - 1}} = \dots \tag{5.77}$$

Et in extenso nous avons déjà :

$$y_{0+1} = y_1 = \frac{y_{-1}}{y_0^{q_1}} = \frac{3}{2^1} = \frac{3}{2} \tag{5.78}$$

Simplifions :

$$\left(\frac{\ln(3)}{\ln(2)} - 1\right)^{-1} = \frac{\ln(2)}{\ln\left(\frac{3}{2}\right)} = \frac{\ln(2)}{\ln(y_1)} \tag{5.79}$$

Donc le premier quotient partiel peut s'écrire :

$$\log_2(3) = 1 + \frac{1}{\frac{\ln(2)}{\ln(y_1)}} = \dots \tag{5.80}$$

et passons au deuxième quotient partiel. Nous savons déjà pour cela que :

$$1 < \frac{\ln(2)}{\ln\left(\frac{3}{2}\right)} < 2 \tag{5.81}$$

donc il est immédiat que $q_2 = 1$ *et alors :*

$$\varepsilon_2 = \frac{\ln(y_{n-2})}{\ln(y_{n-1})} - q_n = \frac{\ln(y_0)}{\ln(y_1)} - q_2\frac{\ln(2)}{\ln\left(\frac{3}{2}\right)} - 1 = \frac{\ln(2) - \ln\left(\frac{3}{2}\right)}{\ln\left(\frac{3}{2}\right)} = \frac{\ln\left(\frac{4}{2}\right)}{\ln\left(\frac{3}{2}\right)} = \tag{5.82}$$

Il vient alors :

$$\log_2(3) = q_1 + \frac{1}{q_2 + 1/\varepsilon_2} = 1 + \frac{1}{1 + \frac{1}{\frac{\ln\left(\frac{4}{3}\right)}{\ln\left(\frac{3}{2}\right)}}} = 1 + \frac{1}{1 + \frac{\ln\left(\frac{3}{2}\right)}{\ln\left(\frac{4}{3}\right)}} \tag{5.83}$$

etc.

6

Probabilités

Le calcul des probabilités s'occupe des phénomènes aléatoires aussi dits *processus stochastiques* lorsqu'ils sont dépendants du temps, c'est-à-dire de phénomènes qui ne mènent pas toujours à la même issue et qui peuvent êtres étudiés grâce aux nombres et à leurs conséquences et apparitions. Néanmoins, même si ces phénomènes ont des issues variées, dépendant du hasard, nous observons cependant une certaine régularité statistique.

Il existe plusieurs manières de définir une probabilité.

Définition 6.1 (Probabilité expérimentale ou inductive) *Probabilité déduite de toute la population concernée.*

Définition 6.2 (Probabilité théorique ou déductive) *Probabilité connue grâce à l'étude du phénomène sous-jacent sans expérimentation. Il s'agit donc d'une connaissance* a priori *par opposition à la définition précédente qui faisait plutôt référence à une notion de probabilité* a posteriori.

Comme il n'est pas toujours possible de déterminer des probabilités *a priori*, nous sommes souvent amenés à réaliser des expériences. Il faut donc pouvoir passer de la première à la deuxième solution. Ce passage est supposé possible en termes de limite [1].

La modélisation formelle par le calcul des probabilités a été inventée par A. Kolmogorov dans un livre paru en 1933. Cette modélisation est faite à partir de l'espace de probabilités (U, A, P) que nous définirons de manière un peu complète plus loin et que nous pouvons relier à la théorie de la mesure.

6.1 Univers des événements

Définition 6.3 *L'univers des événements, ou des* observables, *U est l'ensemble de toutes les issues (résultats) possibles, appelées* événements élémentaires, *qui se présentent au cours d'une épreuve aléatoire déterminée.*

L'univers peut être fini —dénombrable— si les événements élémentaires sont en nombre fini ou continu —non dénombrable— s'ils sont infinis.

Définition 6.4 *Un* événement *quelconque A est un ensemble d'événements élémentaires et constitue une partie de l'univers des possible U. Il est possible qu'un événement ne soit constitué que d'un seul événement élémentaire.*

Exemple 6.1 *Considérons l'univers de tous les groupes sanguins possible, alors l'événement A* l'individu est de rhésus positif *est représenté par :*

$$A = \{A+, B+, AB+, O+\} \subset U \tag{6.1}$$

alors que l'événement B l'individu est donneur universel *est représenté par :*

$$B = \{O-\} \subset U \tag{6.2}$$

qui constitue donc un événement élémentaire.

1. Ceci est possible si on considère une population dont la taille tend vers la taille de la population réelle.

Définition 6.5 *Soit U un univers et A un événement, nous disons que l'événement A* a lieu ou se réalise *si lors du déroulement de l'épreuve se présente l'issue i $(i \in U)$ et que $(i \in A)$. Dans le cas contraire, nous disons que A* n'a pas lieu.

Définition 6.6 *Le sous-ensemble vide $\varnothing$ de U s'appelle* événement impossible. *En effet, si lors de l'épreuve l'issue i se présente, nous avons toujours $i \notin \varnothing$ et donc l'événement $\varnothing$ n'a donc jamais lieu.*

Si U est fini, ou infini dénombrable, tout sous-ensemble de U est un événement, ce n'est plus vrai si U est non dénombrable (nous verrons dans le chapitre de Statistique pourquoi).

Définition 6.7 *L'ensemble U s'appelle aussi* événement certain. *En effet, si lors de l'épreuve l'issue i se présente, nous avons toujours $i \in U$ (car U est l'univers des événements). L'événement U a donc toujours lieu.*

Définition 6.8 *Soit A et B deux sous-ensembles de U. Nous savons que les événements $A \cup B$ et $A \cap B$ sont tous deux des sous-ensembles de U donc des événements aussi respectivement conjoints et disjoints.*

Si deux événements A et B sont tels que :

$$A \cap B = \varnothing \tag{6.3}$$

les deux événements ne peuvent pas êtres réalisables pendant la même épreuve, nous disons alors qu'ils sont des *événements incompatibles*.

Sinon, si :

$$A \cap B \neq \varnothing \tag{6.4}$$

les deux événements peuvent êtres réalisables dans la même épreuve (possibilité de voir un chat noir au moment où on passe sous une échelle par exemple), nous disons inversement qu'ils sont des *événements indépendants*.

6.1.1 Axiomatique de KOLMOGOROV

La probabilité d'un événement sera en quelque sorte le répondant de la notion de fréquence d'un phénomène aléatoire, en d'autres termes, à chaque événement nous allons attacher un nombre réel, appartenant à l'intervalle $[0, 1]$, qui mesurera sa probabilité (chance) de réalisation. Les propriétés des fréquences que nous pouvons mettre en évidence lors d'épreuves diverses nous permettent de fixer les propriétés des probabilités.

Soit U un univers. Nous disons que nous définissons une probabilité sur les événements de U si à tout événement A de U nous associons un nombre ou une mesure $P(A)$, appelé *probabilité a priori de l'événement A* ou *probabilité marginale de A*.

Axiome 6.1 *Pour tout événement A :*

$$1 \geqslant P(A) \geqslant 0 \tag{6.5}$$

Ainsi, la probabilité de tout événement est un nombre réel compris entre 0 et 1 inclus.

Axiome 6.2 *La probabilité de l'événement certain ou de l'ensemble des événements possibles est égale à 1 :*

$$P(U) = 1 \tag{6.6}$$

Axiome 6.3 *Si $A \cup B = \varnothing$ sont deux événements incompatibles, alors :*

$$P(A \cup B) = P(A) + P(B) \tag{6.7}$$

la probabilité de la réunion de deux événements incompatibles est donc égale à la somme de leurs probabilités. Nous parlons alors de *probabilité disjointe*.

Par exemple, si nous considérons qu'il est impossible d'avoir les cheveux totalement blonds et bruns en même temps et que chaque état à une probabilité de 50%, alors la probabilité d'être l'un

ou l'autre des couleurs est la somme des probabilités. Nous retrouverons ce genre de probabilité dans le chapitre de Génie Industriel dans la méthode AMDEC des systèmes à structure complexe pour un exemple pratique.

Autrement dit sous forme plus générale si $(A_i)_{i\in\mathbb{N}}$ est une suite d'événements disjoints deux à deux [2] alors :

$$P\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} P(A_i) \tag{6.8}$$

Nous parlons alors de *σ-additivité* car si nous regardons de plus près, les trois axiomes ci-dessus la mesure P forme une σ-algèbre (cf. chapitre de Théorie de la Mesure).

Une conséquence immédiate des axiomes 6.2 et 6.3 est la relation entre les probabilités d'un événement A et son complémentaire, noté $\bar{A}$:

$$P(\bar{A}) = 1 - P(A) \tag{6.9}$$

Définition 6.9 *Si A et B sont indépendants, nous savons que $A \cap B = \varnothing$, alors :*

$$P(A \cap B) = P(A) \cdot P(B) \tag{6.10}$$

la probabilité de l'intersection de deux événements indépendants est égale au produit de leurs probabilités. Nous parlons alors de probabilité conjointe.

Autrement dit sous forme plus générale, les événements $A_1, \ldots, A_n$ sont indépendants si la probabilité de l'intersection est le produit des probabilités :

$$P\left(\bigcap_{i\in\mathbb{N}} A_i\right) = \prod_{i\in\mathbb{N}} P(A_i) \tag{6.11}$$

REMARQUE Il faut bien faire attention à ne pas confondre *indépendants* et *incompatibles*.

Soit U un univers comportant un nombre fini n d'issues possibles $U = \{i_1, i_2, \ldots, i_n\}$. Les événements $I_1 = \{i_1\}$, $I_2 = \{i_2\}, \ldots, I_n = \{i_n\}$ sont donc appelés *événements élémentaires*. Lorsque ces événements ont même probabilité, nous disons qu'ils sont *équiprobables*. Dans ce cas, il est très facile de calculer leur probabilité. En effet, ces événements étant par définition incompatibles entre eux à ce niveau de notre discours, nous avons en vertu de l'axiome 6.3 des probabilités :

$$P(I_1 \cup I_2 \cup \ldots \cup I_n) = P(I_1) + P(I_2) + \ldots + P(I_n) \tag{6.12}$$

mais puisque $P(I_1 \cup I_2 \cup \ldots \cup I_n) = P(U) = 1$ et que les probabilités du membre de droite sont par hypothèse équiprobables, nous avons :

$$P(I_1) = P(I_2) = P(I_n) = \frac{1}{n} \tag{6.13}$$

6.1.2 Probabilités conditionnelles

Que pouvons-nous déduire sur la probabilité d'un événement B sachant qu'un événement A est réalisé ? En d'autres termes, nous voulons savoir s'il est possible de définir la probabilité d'un événement conditionnellement à un autre événement.

Ce type de probabilité est appelée *probabilité conditionnelle* ou *probabilité a posteriori* de B sachant A, et se note dans le cadre de l'étude des probabilités conditionnelles $P(B/A)$ et souvent dans la pratique pour éviter la confusion avec une possible division $P(B|A)$. Nous avons aussi le cas : $P(A|B)$ qui est appelé *fonction de vraisemblance de A* ou encore *probabilité a priori* de A sachant B.

2. A_i et A_j ne peuvent pas se produire en même temps si $i \neq j$.

Historiquement, le premier mathématicien à avoir utilisé correctement la notion de probabilité conditionnelle fut T. BAYES (1702-1761). Aussi parlons-nous souvent de BAYES ou de bayésien dès que des probabilités conditionnelles sont en jeu : formule de BAYES, statistique bayésienne.

La notion de probabilité conditionnelle que nous allons introduire est beaucoup moins simple qu'elle ne paraît *a priori* et les problèmes de conditionnement sont une source inépuisable d'erreurs en tout genre (il existe de fameux paradoxes sur le sujet).

Commençons d'abord par un exemple simpliste. Supposons que nous ayons deux dès. Imaginons maintenant que nous ayons lancé seulement le premier dé. Nous voulons savoir quelle est la probabilité qu'en lançant le second dé, la somme des deux chiffres vaille une certaine valeur minimale. Ainsi, la probabilité d'obtenir cette valeur minimale fixée sachant la valeur du premier dé est totalement différente de la probabilité d'obtenir cette même valeur minimale en lançant les deux dès en même temps. Comment calculer cette nouvelle probabilité ?

Formalisons la démarche. Après le lancer du premier dé, nous avons :

$$A = \{\text{résultat du premier lancer}\} \tag{6.14}$$

Soit l'hypothèse que $B \subset A$, nous pressentons que $P(B|A)$ doit être proportionnel à $P(B)$, la constante de proportionnalité étant déterminée par la normalisation :

$$P(A|A) = 1 \tag{6.15}$$

Soit maintenant $B \subset A^c$ (B est inclus dans le complémentaire de A donc les événements sont incompatibles). Il est assez intuitif que sous l'hypothèse précédente nous ayons :

$$P(A|B) = 0 \tag{6.16}$$

Ceci nous mène aux définitions suivantes des probabilités *a posteriori* et respectivement *a priori* :

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{et} \quad P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{6.17}$$

Ainsi, le fait de savoir que B est réalisé réduit l'ensemble des résultats possibles de U à B. A partir de là, seules les éventualités de $A \cap B$ ont une importance. La probabilité de A sachant B et inversement (par symétrie) doit donc être proportionnelle à $P(A \cap B)$.

Le coefficient de proportionnalité qui est le dénominateur permet d'assurer l'événement certain. Effectivement, si les deux événements A et B sont incompatibles, nous avons donc :

$$P(A \cap B) = P(A) \cdot P(B) \tag{6.18}$$

et nous voyons alors $P(B|A)$ qui vaut $P(B)$ et donc A n'apporte rien sur B et réciproquement.

Une autre façon assez intuitive pour voir les choses est de se représenter la mesure de probabilité P comme une mesure d'aires de sous-ensembles de $\mathbb{R}^2$.

En effet, si A et B sont deux sous-ensembles de $\mathbb{R}^2$ d'aires respectives $P(A)$ et $P(B)$ alors à la question de savoir qu'elle est la probabilité qu'un point du plan appartienne à B sachant qu'il appartient à A il est assez évident de répondre que cette probabilité est donnée par :

$$\frac{\text{Surface}(A) \cap \text{Surface}(B)}{\text{Surface}(A)} \tag{6.19}$$

Indiquons aussi que la définition des probabilités conditionnelles s'utilise souvent sous la forme suivante :

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \tag{6.20}$$

appelée *formule des probabilités composées*. Ainsi, la probabilité *a posteriori* de B sachant A peut donc aussi s'écrire sous la forme :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{6.21}$$

Exemple 6.2 *Supposons une maladie comme la méningite. La probabilité de l'avoir sera noté* $P(M) = 0{,}001$ *et un signe de cette maladie comme le mal de tête sera noté* $P(S) = 0{,}1$. *Supposons connue la probabilité* a posteriori *d'avoir mal à la tête si nous avons une méningite :*

$$P(S|M) = 0{,}9 \tag{6.22}$$

Le théorème de Bayes *donne alors la probabilité* a priori *d'avoir une méningite si nous avons mal à la tête :*

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = 0{,}9 \tag{6.23}$$

Notons que nous avons aussi :

$$\begin{aligned} P(A) &= P((B_1 \cup B_2 \cup \ldots \cup B_n) \cap A) = P((B_1 \cap A) \cup (B_2 \cap A) \cup \ldots \cup (B_n \cap A)) \\ &= P(B_1 \cap A) + P(B_2 \cap A) + \ldots + P(B_n \cap A) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_n)P(B_n) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \end{aligned} \tag{6.24}$$

qui est appelée la *formule des probabilités totales* ou *théorème des probabilités totales*. Mais aussi, pour tout j, nous avons le corollaire suivant en utilisant les résultats précédents :

$$P(B_j|A) = \frac{P(B_j \cap A}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)} \tag{6.25}$$

qui est la forme générale de la *formule de* Bayes ou *théorème de* Bayes que nous utiliserons un tout petit peu en Mécanique Statistique et dans le cadre de l'étude de la théorie des files d'attentes (cf. chapitre de Techniques De Gestion). Il faut savoir que les implications de ce théorème sont cependant considérables dans le quotidien, dans la médecine, dans l'industrie et dans le domaine du Data Mining informatique.

Exemple 6.3 *Deux machines* M_1 *et* M_2 *produisent respectivement 100 et 200 pièces.* M_1 *produit 5% de pièces défectueuses et* M_2 *en produit 6% (ces valeurs proviennent d'une loi exponentielle !). Quelle est la probabilité pour qu'un objet défectueux ait été fabrique par la machine* M_1 *?*

L'événement constaté A est donc la présence d'une pièce défectueuse et la probabilité recherchée est la probabilité a priori *que celle-ci provienne de la machine* M_1. *Nous avons alors, d'après* (**??**) *:*

$$P(M_1|A) = \frac{P(A|M_1)P(M_1)}{P(A|M_1)P(M_1) + P(A|M_2)P(M_2)} = \frac{5/100 \times 1/3}{5/100 \times 1/3 + 6/100 \times 2/3} \simeq 0{,}29 \tag{6.26}$$

L'analyse bayésienne fournit donc un outil puissant de formalisation du raisonnement dans l'incertain et les exemples que nous avons montrés illustrent surtout à quel point cet outil est délicat à employer.

6.1.3 Martingales

Une martingale en probabilités[3] est une technique permettant d'augmenter les chances de gain aux jeux de hasard tout en respectant les règles de jeu. Le principe dépend complètement du type de jeu qui en est la cible, mais le terme est accompagné d'une aura de mystère qui voudrait que certains joueurs connaissent des techniques secrètes mais efficaces pour tricher avec le hasard. Par exemple, de nombreux joueurs cherchent *la* martingale qui permettra de battre la banque dans les jeux les plus courants dans les casinos[4].

De nombreuses martingales ne sont que le rêve de leur auteur, certaines sont en fait inapplicables, quelques-unes permettent effectivement de tricher un peu. Les jeux d'argent sont en général inéquitables : quel que soit le coup joué, la probabilité de gain du casino est plus importante que celle du joueur. Dans ce type de jeu, il n'est pas possible d'inverser les chances, seulement de minimiser la probabilité de ruine du joueur.

3. Il est nécessaire de mentionner ici qu'il existe une autre martingale dans les processus stochastiques.
4. Les casinos sont des institutions dont la rentabilité repose presque entièrement sur la différence —même faible— qui existe entre les chances de gagner et celles de perdre.

L'exemple le plus courant est la martingale de la roulette, elle consiste à jouer une chance simple à la roulette —noir ou rouge, paire ou impaire— de façon à gagner, par exemple, une unité dans une série de coups en doublant sa mise si l'on perd, et cela jusqu'à ce que l'on gagne. Le joueur mise 1 unité sur le rouge, si le rouge sort, il arrête de jouer et il a gagné une unité, si le noir sort, il double sa mise en pariant deux unités sur le rouge et ainsi de suite jusqu'à ce qu'il gagne.

Ayant une chance sur deux de gagner, il peut penser qu'il va finir par gagner ; quand il gagne, il est forcément remboursé de tout ce qu'il a joué, plus une fois sa mise de départ. Cette martingale semble être sûre en pratique. Il faut noter que sur le plan théorique, pour être sûr de gagner, il faudrait avoir la possibilité de jouer au cas où un nombre de fois illimité. Ce qui présente des inconvénients majeurs : cette martingale est en fait limitée par les mises que le joueur peut faire car il faut doubler la mise à chaque coup tant que l'on perd : deux fois la mise de départ, puis 4, 8, 16... s'il perd dix fois de suite, il doit pouvoir avancer 1 024 fois sa mise initiale pour la 11^{e} partie. Il faut donc beaucoup d'argent pour gagner peu.

Les roulettes comportent un 0 qui n'est ni rouge ni noir. Le risque de perdre lors de chaque coup est ainsi plus grand que 1/2.

De plus, pour paralyser cette stratégie, les casinos proposent des tables de jeu par tranche de mise : de 1 à 100.-, de 2 à 200.-, de 5 à 500.-,... Impossible donc d'utiliser cette méthode sur un grand nombre de coups, ce qui augmente le risque de tout perdre.

Le black jack est un jeu qui possède des stratégies gagnantes : plusieurs techniques de jeu, qui nécessitent généralement de mémoriser les cartes, permettent de renverser les chances en faveur du joueur. Le mathématicien Edward Thorp a ainsi publié en 1962 un livre qui fut à l'époque un véritable best-seller. Mais toutes ces méthodes demandent de longues semaines d'entraînement et sont facilement décelables par le croupier [5]. Le casino a alors tout loisir d'écarter de son établissement les joueurs en question.

Il faut noter qu'il existe des méthodes assez évoluées. L'une d'elles repose sur les combinaisons les moins jouées. Dans les jeux où le gain dépend du nombre de joueurs gagnants, jouer les combinaisons les moins jouées optimisera les gains. C'est ainsi que certaines personnes vendent des combinaisons qui seraient statistiquement très rarement utilisées par les autres joueurs.

Partant de ce raisonnement, on peut encore conclure qu'un joueur qui aurait réussi à déterminer ainsi les combinaisons statistiquement les moins jouées, afin d'optimiser son espérance de gain ne sera en fait certainement pas le seul joueur à avoir obtenu par l'analyse ces fameuses combinaisons, et tous ces joueurs risquent donc finalement d'être très déçus par leurs gains s'il s'avérait que cette combinaison équiprobable sorte au tirage ! Cela revient à dire que les numéros en théorie les moins joués sont en fait surjoués par combinaisons, le mieux serait peut-être de réaliser un savant mélange de numéros sous-joués et de numéros surjoués pour obtenir les combinaisons idéales, qui peuvent par ailleurs être observées dans les tirages passés lorsqu'il n'y a pas eu de gagnant. Une autre conclusion à tout cela est peut-être que le mieux est encore de jouer des combinaisons aléatoires qui ont finalement moins de chance d'être également choisies par les joueurs qui incorporent un facteur humain et harmonieux dans le choix de leurs nombres.

6.2 Analyse combinatoire

L'analyse combinatoire est le domaine de la mathématique qui s'occupe de l'étude de l'ensemble des issues, événements ou faits —distinguables ou non tous distinguables— avec leurs arrangements —combinaisons— ordonnés ou non selon certaines contraintes données.

Définition 6.10 *Une suite d'objets (événements, issues, objets,...) est dite* ordonnée *si chaque suite composée d'un ordre particulier des objets est comptabilisée comme une configuration particulière.*

Définition 6.11 *Une suite est* non ordonnée *si et seulement si nous intéresse la fréquence d'apparition des objets indépendamment de leur ordre.*

Définition 6.12 *Des objets d'une suite sont dits* distincts *si leurs caractéristiques ne permettent pas de les confondre avec d'autres objets.*

REMARQUE Nous avons choisi de mettre l'analyse combinatoire dans ce chapitre car lorsque nous calculons des probabilités, nous avons également assez souvent besoin de savoir quelle est la probabilité de tomber sur une combinaison ou un arrangement d'événements donnés sous certaines contraintes.

5. Les brusques changements de montant des mises sont caractéristiques.

Il existe plusieurs types d'arrangements selon les contraintes et les propriétés des éléments arrangés. Nous allons présenter et démontrer ci-dessous les cinq cas les plus répandus à partir desquels nous pouvons trouver tous les autres.

6.2.1 Arrangements avec répétition

Définition 6.13 *Un* arrangement avec répétition *est une suite ordonnée de longueur m de n objets distincts non nécessairement tous différents dans la suite.*

Soient A et B deux ensembles finis de cardinaux respectifs m, n tels que trivialement il y ait m façons de choisir un objet dans A (de type a) et n façons de choisir un objet dans B (de type b).

Nous avons vu en théorie des ensemble que si A et B sont disjoints, que :

$$\text{Card}(A \cup B) = m + n \tag{6.27}$$

Nous en déduisons donc les propriétés suivantes :

Propriété 6.1 *Si un objet ne peut être à la fois de type a et de type b et s'il y a m façons de choisir un objet de type a et n façons de choisir un objet de type b, alors il y a $m+n$ façons de choisir un objet de type a ou de type b.*

Propriété 6.2 *Si nous pouvons choisir un objet de type a de m façons puis un objet de type b de n façons, alors il y a selon le produit cartésien de deux ensembles• :* Chap4

$$\text{Card}(A \cdot B) = \text{Card}(A) \cdot \text{Card}(B) = m \cdot n \tag{6.28}$$

de manière choisir un seul et unique objet de type a puis un objet de type b.

Avec les mêmes notations, choisir une fonction de A dans B, c'est choisir (dans le cas général) pour chaque élément de A, son unique image parmi les n éléments de B. Il y a donc n façons de choisir l'image b_i du premier élément a_1 de A, puis aussi n façons de choisir l'image b_i du deuxième..., puis n façons de choisir b_i l'image du m-ème. Le nombre d'applications totales possibles de A dans B est donc égal au produit de m égaux à n. Ainsi, nous avons :

$$\text{Card}(B^A) = \bar{A}_n^m = n^m \tag{6.29}$$

où B^A est l'ensemble des applications de A dans B. La progression du nombre de possibilités est donc géométrique.

Ce résultat mathématique est assimilable au résultat non-ordonné (un arrangement $\bar{A}_n^m$ dont l'ordre des éléments de la suite n'est pas est pris en compte) de m tirages dans un sac contenant n boules différentes avec remise après chaque tirage.

Exemple 6.4 *Combien de* mots ordonnés *de 7 lettres pouvons-nous former à partir d'un alphabet de 24 lettres distinctes ? La solution est :*

$$\bar{A}_7^{24} = 24^7 = 4\,586\,471\,424 \tag{6.30}$$

Exemple 6.5 *Combien de groupes d'individus aurons-nous lors d'une votation sur 5 sujets et où chacun peut être soit accepté, soit rejeté ? La solution est :*

$$\bar{A}_5^2 = 2^5 = 32 \tag{6.31}$$

Une généralisation simple de ce dernier résultat peut consister dans l'énoncé du problème suivant :

Si nous disposons de m objets $k_1, \ldots, k_m$ tels que k_i peut prendre n_i états différents alors le nombre de combinaisons possibles est :

$$A_{mn} = n_1 n_2 \ldots n_i \ldots n_m \tag{6.32}$$

Et si nous avons $n_1 = n_2 = \ldots = n_m$ alors nous retombons sur :

$$A_{mn} = \bar{A}_n^m = n_1 n_2 \ldots n_i \ldots n_m = n^m \tag{6.33}$$

6.2.2 Permutations simples

Définition 6.14 *Une* permutation simple *de n objets distincts est une suite ordonnée (différente) de ces n objets par définition tous différents dans la suite (sans répétition).*

Le nombre d'arrangements de n éléments peut être calculé par récurrence : il y a n places pour un premier élément, $n-1$ pour un deuxième élément... et une place pour le dernier élément restant. Il est dès lors trivial que nous aurons un nombre d'arrangements donné par :

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-(n-1)) \tag{6.34}$$

autrement dit *factorielle de n* et nous la notons $n!$ pour $n \in \mathbb{N}$. Il y a donc pour n éléments distinguables $A_n = n!$ arrangements possibles.

Exemple 6.6 *Combien de* mots ordonnés *de 7 lettres distinctes sans répétition pouvons-nous former ?*

$$A_7 = 7! = 5\,040 \tag{6.35}$$

Ce résultat nous amène à l'assimiler au résultat ordonné —un arrangement A_n dont l'ordre des éléments de la suite est pris en compte— du tirage de toutes les boules différentes d'un sac contenant n boules distinguables sans remise.

6.2.3 Permutations avec répétition

Définition 6.15 *Lorsque nous considérons le nombre de permutations ordonnées (différentes) d'une suite de n objets distincts tous nécessairement non différents dans une quantité donnée dans la suite nous parlons de* permutation avec répétition.

REMARQUE Il ne faut pas confondre cette dernière définition avec *l'arrangement avec répétition.*

Lorsque certains éléments éléments ne sont pas distinguables dans une suite d'objets (ils sont répétitifs dans la suite), alors le nombre d'arrangements (permutations) que nous pouvons constituer se réduit alors assez trivialement à un nombre plus petit que si tous les éléments étaient distinguables.

Soit n_i le nombre d'objets du type i, avec $n_1 + n_2 + n_3 + \ldots + n_k = n$ alors, nous notons $\bar{A}_n(n_1, \ldots, n_i, \ldots, n_k)$ avec $i = 1, 2, \ldots, k$ le nombre d'arrangements possibles avec répétition [6].

Si chacune des n_i places occupées par des éléments identiques était occupée par des éléments différents, le nombre de permutations serait alors à multiplier par chacun des $n_i!$.

Il vient alors que nous retombons sur la factorielle telle que :

$$\bar{A}_n(n_1, \ldots, n_i, \ldots, n_k) \cdot n_1! \cdot n_2! \cdot n_3! \cdot \ldots \cdot n_k! = n! \tag{6.36}$$

alors :

$$\bar{A}_n(n_1, \ldots, n_i, \ldots, n_k) = \frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdot \ldots \cdot n_k!} \tag{6.37}$$

Si les n objets sont tous différentes dans la suite, nous avons alors :

$$n_1! = n_2! = n_3! = \ldots = n_k! = 1! = 1 \tag{6.38}$$

et nous nous retrouvons bien avec une permutation simple (sans répétition) telle que :

$$\bar{A}_n(1, \ldots, 1, \ldots, 1) = \frac{n!}{1! \times 1! \times \ldots \times 1!} = n! \tag{6.39}$$

Les permutations avec répétition sont en plus petit nombre que celles sans répétition puisque nous ne prenons pas en compte les permutations des éléments identiques entre eux.

Exemple 6.7 *Combien de mots ordonnés pouvons-nous former avec les lettres du mot* mississippi *:*

$$\bar{A}_4(\underbrace{1}_{m}, \underbrace{2}_{pp}, \underbrace{4}_{iiii}, \underbrace{4}_{ssss}) = \frac{11!}{1! \times 2! \times 4! \times 4!} = 34\,560 \tag{6.40}$$

Ce résultat nous amène à l'assimiler au résultat ordonné (un arrangement $\bar{A}_n$ dont l'ordre des éléments de la suite est pris en compte) du tirage de n boules non toutes distinguables d'un sac contenant $k \geqslant n$ boules avec remise limitée pour chaque boule.

6. Un ou plusieurs éléments répétitifs dans une suite d'éléments sont non distinguables par permutation.

6.2.4 Arrangements simples sans répétition

Définition 6.16 *Un* arrangement simple sans répétition *est une suite ordonnée de p objets tous distincts pris parmi n objets distincts avec $n \geqslant p$.*

Nous nous proposons donc maintenant de dénombrer les arrangements possibles de p objets parmi n. Nous noterons A_n^p le nombre des ces arrangements.

Il est aisé de calculer $A_n^1 = n$ et de vérifier que $A_n^2 = n(n-1)$. Effectivement, il existe n façons de choisir le premier objet et $(n-1)$ façons de choisir le deuxième lorsque nous avons déjà le premier.

Pour déterminer A_n^p, nous raisonnons alors par récurrence. Nous supposons A_n^{p-1} connu et nous en déduisons $A_n^p = A_n^{p-1}(n-p+1)$. Dès lors $A_n^p = n(n-1)(n-2)(n-3)\dots(n-p+1)$ d'où :

$$A_n^p = \frac{n!}{(n-p)!} \tag{6.41}$$

Ce résultat nous amène à l'assimiler au résultat ordonné [7] du tirage de p boules d'un sac contenant n boules différentes sans remise.

Exemple 6.8 *Soit les 24 lettres de l'alphabet, combien de mots ordonnés de 7 lettres distinctes pouvons-nous former ?*

$$A_{24}^7 = \frac{24!}{(24-7)!} = 1\,744\,364\,160 \tag{6.42}$$

Le lecteur aura peut-être remarqué que si nous prenons $p = n$ nous nous retrouvons avec :

$$A_n^n = \frac{n!}{0!} = n! \tag{6.43}$$

Une permutation simple est donc un arrangement simple sans répétition avec $n = p$.

6.2.5 Combinaisons simples

Définition 6.17 *Une* combinaison simple *ou* choix *est une suite non-ordonnée* [8] *de p éléments tous différents choisis parmi n objets distincts est notée C_p^n et appelée la* binomiale.

Si nous permutons les éléments de chaque arrangement simple de p éléments parmi n, nous obtenons toutes les permutations simples et nous savons qu'il y en a $p!$ d'où en utilisant la convention d'écriture :

$$C_p^n = \binom{n}{p} = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!} \tag{6.44}$$

C'est une relation très souvent utilisée dans les jeux de hasard mais également dans l'industrie via la loi hypergéométrique (cf. chapitre de Techniques De Gestion).

Remarques

1. Nous avons nécessairement par définition $C_p^n \leqslant A_p^n$;
2. Selon les auteurs nous inversons l'indice ou le suffixe de C il faut donc être prudent.

Exemple 6.9 *Soit un alphabet de 24 lettres, combien avons-nous de choix de prendre 7 lettres parmi les 24 sans prendre en compte l'ordre dans lequel sont triées les lettres :*

$$C_7^{24} == \frac{24!}{7!(24-7)!} = 346\,104 \tag{6.45}$$

Ce résultat nous amène à l'assimiler au résultat non ordonné [9] du tirage de p boules d'un sac contenant n boules différentes sans remise.

7. Autrement dit, un arrangement A_n^p dont l'ordre des éléments de la suite est pris en compte.
8. Le terme "non-ordonné" signifie ici que l'ordre ne nous intéresse pas.
9. Autrement dit, un arrangement C_p^n dont l'ordre des éléments de la suite n'est pas pris en compte.

Il existe, relativement à la binomiale, une autre relation très souvent utilisée dans de nombreux cas d'études ou également de manière plus globale en physique ou analyse fonctionnelle. Il s'agit de la *formule de Pascal* :

$$C_{p-1}^{n-1} + C_p^{n-1} = C_p^n \tag{6.46}$$

Preuve Nous partons de :

$$\begin{aligned} C_{p-1}^{n-1} + C_p^{n-1} &= \frac{(n-1)!}{((n-1)-(p-1))!(p-1)!} + \frac{(n-1)!}{(n-1-p)!p!} \\ &= \frac{(n-1)!}{(n-p)!(p-1)!} + \frac{(n-1)!}{(n-1-p)!p!} \end{aligned} \tag{6.47}$$

Or $p! = p(p-1)!$ et $(n-p)(n-p-1)! = (n-p)!$ d'où :

$$\begin{aligned} C_{p-1}^{n-1} + C_p^{n-1} &= \frac{(n-1)!}{((n-1)-(p-1))!(p-1)!} + \frac{(n-1)!}{(n-1-p)!p!} = \frac{(n-1)!p!}{(n-p)!p!} + \frac{(n-1)!(n-p)}{(n-p)!p!} \\ &= \frac{(n-1)!}{(n-p)!p!}(p+(n-p)) = \frac{(n-1)!n}{(n-p)!p!} = \frac{n!}{(n-p)!p!} = C_p^n \end{aligned} \tag{6.48}$$

d'où (6.46). •

6.3 Chaînes de MARKOV

Les chaînes de MARKOV sont des outils statistiques et probabilistes simples mais dont la forme de présentation mathématique prête parfois à l'horreur. Nous allons tenter ici de simplifier un maximum les notations pour introduire cet outil formidable très utilisé au sein des entreprises pour gérer la logistique, les files d'attentes aux centrales d'appel ou aux caisses de magasins jusqu'à la théorie de la défaillance pour la maintenance préventive, en physique statistique ou en génie biologique.

Définition 6.18 *Nous noterons $\{X(t)\}_{t\in T}$ un processus probabiliste fonction du temps —c'est donc un processus stochastique— dont la valeur à chaque instant dépend de l'issue d'une expérience aléatoire. Ainsi, à chaque instant t, $X(t)$ est donc une variable aléatoire.*

Si nous considérons un temps discret, nous notons alors $\{X_n\}_{n\in\mathbb{N}}$ un processus stochastique à temps discret. Si nous supposons que les variables aléatoires X_n ne peuvent prendre qu'un ensemble discret de valeurs. Nous parlons alors de *processus à temps discret et à espace discret.*

REMARQUE Il est tout à fait possible comme dans l'étude du télétrafic d'avoir un processus à temps continu et à espace d'état discret.

Définition 6.19 *$\{X_n\}_{n\in\mathbb{N}}$ est une* chaîne de MARKOV *si et seulement si :*

$$P(X_n = j | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) = P(X_n = j | X_{n-1} = i_{n-1}) \tag{6.49}$$

En d'autres termes, la probabilité pour que la chaîne soit dans un certain état à la n-ème étape du processus ne dépend que de l'état du processus à l'étape $n-1$ et pas des étapes précédentes.

REMARQUE En probabilité un processus stochastique vérifie la propriété markovienne si et seulement si la distribution conditionnelle de probabilité des états futurs, étant donné l'instant présent, ne dépend que de ce même état présent et pas des états passés. Un processus qui possède cette propriété est donc appelé *processus de* Markov.

Définition 6.20 *Une* chaîne de MARKOV homogène *est une chaîne telle que la probabilité qu'elle a pour passer dans un certain état à la n-ème soit indépendante du temps. En d'autres termes, la loi de probabilité caractérisant la prochaine étape ne dépend pas du temps, et en tout temps la loi de probabilité à tout moment de la chaîne est toujours la même pour caractériser la transition à l'étape en cours.*

Nous pouvons alors définir la loi de *probabilité de transition* d'un état i vers un état j par :

$$p_{ij} = P(X_n = j | X_{n-1} = i) \tag{6.50}$$

Il est alors naturel de définir la *matrice de transition* :

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \vdots & p_{32} & \ddots & \\ \vdots & \vdots & & \ddots \end{bmatrix} \tag{6.51}$$

Les chaînes de MARKOV peuvent être représentées graphiquement sous la forme d'un graphe orienté (cf. chapitre de Théorie Des Graphes). Nous associons alors à chaque composante un arc orienté et sa la probabilité de transition.

Exemple 6.10 *Ainsi, les seules transitions permises par les quatre états —matrice 4×4— sont celles indiquées par les flèches. Ce qui fait que la matrice de transition s'écrit alors :*

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} = P = \begin{bmatrix} 0 & p_{12} & 0 & p_{14} \\ 0 & p_{22} & p_{23} & 0 \\ p_{31} & 0 & p_{33} & 0 \\ p_{41} & 0 & 0 & 0 \end{bmatrix} \tag{6.52}$$

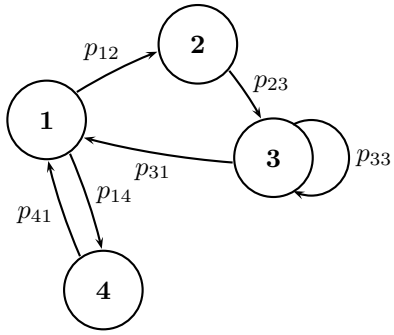


FIGURE 6.1 – Exemple de graphe

L'analyse du régime transitoire d'une chaîne de MARKOV consiste à déterminer la matrice-colonne (vecteur) $p(n)$ d'être dans un état j à l'étape n :

$$p(n) = \begin{pmatrix} p_1(n) \\ p_2(n) \\ \vdots \\ p_{\mathrm{Card}(E)}(n) \end{pmatrix} \tag{6.53}$$

Ce vecteur de probabilités dépend de la matrice de transition P et du vecteur de probabilités initiales $p(0)$.

Nous en verrons un exemple trivial dans le chapitre de Théorie des Graphes qui sera redéveloppé sous forme détaillée et complète ainsi que dans le chapitre de Théorie Des Jeux Et De La Décision dans le cadre de la pharmaco-économie. Mais signalons que les chaînes de MARKOV sont également utilisées en météorologie par exemple, comme illustré sur la figure 6.2, ou dans le domaine médical, financier, des transports du marketing, etc.

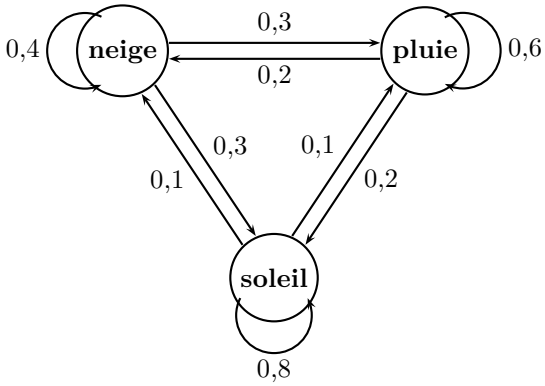


FIGURE 6.2 – Graphe météorologique

7

Statistiques

La statistique est une science qui a pour objet le groupement méthodique de faits ou événements répétitifs qui se prêtent à une évaluation numérique dans le temps suivant une loi donnée.

Il faut savoir que parmi tous les domaines de la mathématique, celle qui est utilisée à la plus large échelle dans un cadre professionnel dans les entreprises est bien la statistique. Raison pour laquelle ce chapitre est un des plus gros alors que seuls les concepts élémentaires y sont présentés.

Il est peut être inutile de préciser que la statistique est beaucoup utilisée en ingénierie, physique théorique, en économétrie, en gestion de projets, dans l'industrie des processus, dans les domaines des assurances vies et non vies, dans l'actuariat ou dans la simple analyse de banque de données et la liste est encore longue. Par ailleurs, nous rencontrerons les outils présentés ici assez souvent dans les chapitres de Mécanique des Fluides, de Thermodynamique, des Techniques de Gestion, du Génie Industriel et d'Économétrie (en particulier dans ces deux dernières). Le lecteur pourra donc s'y reporter pour avoir des applications pratiques concrètes des quelques-uns des éléments théoriques les plus importants qui seront vus ici.

Définition 7.1 *Le but principal de la statistique est de déterminer les caractéristiques d'une population donnée à partir de l'étude d'une partie de cette population, appelée* échantillon *ou* échantillon représentatif.

Remarque Le traitement des données concerne la *statistique descriptive*. L'interprétation des données à partir des estimateurs s'appelle *l'inférence statistique* (ou *statistique inférentielle*), et l'analyse de données en masse la *statistique fréquentielle* (en opposition à l'inférence bayesienne).

Lorsque nous observons un événement prenant en compte certains facteurs, il peut arriver qu'une deuxième observation ait lieu dans des conditions qui semblent identiques. En répétant ces mesures plusieurs fois sur différents objets supposés similaires, nous pouvons constater que les résultats observables sont distribués statistiquement autour d'une valeur moyenne qui est, finalement le résultat possible le plus probable. Dans la pratique, nous n'effectuons cependant parfois qu'une seule mesure et il s'agit alors de déterminer la valeur de l'erreur que nous commettons en adoptant celle-ci comme moyenne mesurée. Cette détermination nécessite de connaître le type de distribution statistique auquel nous avons à faire et c'est ce que nous allons nous attarder (entre autres) à étudier ici. Il existe cependant plusieurs approches méthodologiques courantes (les moins courantes n'étant pas citées pour l'instant) face au hasard :

1. Une toute première consiste à ignorer purement et simplement les éléments aléatoires, pour la bonne raison que l'on ne sait pas comment les intégrer. Nous utilisons alors la *méthode des scénarios* appelé aussi *simulation déterministe*. C'est typiquement un outil utilisé par les financiers ou gestionnaires non diplômés.
2. Une seconde approche envisageable, quand nous ne savons pas associer des probabilités précises aux futurs événements aléatoires, est la théorie des jeux (cf. chapitre de la Théorie Des Jeux Et De La Décision) où l'on utilise des critères de sélection semi-empiriques comme le critère du maximax, du minimax, de Laplace, de Savage, etc.
3. Enfin, quand nous pouvons lier des probabilités aux événements aléatoires, soit que ces probabilités découlent de calculs ou de mesures, soit qu'elles reposent sur une expérience acquise auprès de situations antérieurs de même nature que la situation actuelle, nous pouvons faire appel aux statistiques descriptives et inférentielles (contenu du présent chapitre) pour tirer des informations exploitables et pertinentes de cette masse de données acquises.

4. Une dernière approche, quand nous avons connaissance de probabilités relatives aux issues intervenantes faisant suite à des choix stratégiques, est l'utilisation de la théorie de la décision (cf. chapitre de la Théorie Des Jeux Et De La Décision).

REMARQUE Sans la statistique mathématique, un calcul sur des données (par exemple une moyenne), n'est qu'un *indicateur*. C'est la statistique mathématique qui lui donne le statut d'estimateur dont on maîtrise le biais, l'incertitude et autres caractéristiques statistiques. On cherche en général à ce que l'estimateur soit sans biais, convergeant et efficace.

Introduisons avant de continuer quelques définitions qui vont nous être utiles pour la suite sur le concept d'échantillons et de moyennes.

7.1 Échantillons

Lors de l'étude statistique d'ensembles d'informations, la façon de sélectionner l'échantillon est aussi importante que la manière de l'analyser. Il faut que l'échantillon soit représentatif de la population. Pour cela, l'échantillonnage aléatoire est le meilleur moyen d'y parvenir.

Le statisticien part toujours de l'observation d'un ensemble fini d'éléments, que nous qualifions de *population*. Les éléments observés, en nombre n, sont tous de même nature, mais cette nature peut être fort différente d'une population à l'autre.

Définition 7.2 *Nous sommes en présence d'un* caractère quantitatif *lorsque chaque élément observé fait explicitement l'objet d'une même mesure. A un caractère quantitatif donné, nous associons une* variable quantitative *qui synthétise toutes les valeurs possibles que la mesure considérée est susceptible de prendre (ce type d'information étant représenté par des courbes de Gauss, de Bêta, de POISSON, etc.)*

Définition 7.3 *Nous sommes en présence d'un*caractère qualitatif *lorsque chaque élément observé fait explicitement l'objet d'un rattachement unique à une* modalité *(nombre d'occurrences dans l'observation) choisie dans un ensemble de modalités exclusives permettant de classer tous les éléments de l'ensemble étudié selon un certain point de vue (ce type d'information étant représenté par des diagrammes à barre, fromages, diagrammes à bulles, etc.).*

Définition 7.4 *Un* échantillon aléatoire *est un échantillon tiré au hasard dans lequel tous les individus d'une population ont la même chance, ou* équiprobabilité *(et nous insistons sur le fait que cette probabilité doit être égale), de se retrouver dans l'échantillon.*

Définition 7.5 *Dans le cas contraire d'un échantillon dont les éléments n'ont pas été pris au hasard, nous disons alors que l'échantillon est* biaisé. *Dans le cas inverse, nous disons qu'il est* non-biaisé.

REMARQUE Un petit échantillon représentatif est, de loin, préférable à un grand échantillon biaisé.

En fait, lorsque la taille des échantillons utilisés est petite, le hasard peut donner un résultat moins bon que celui qui est biaisé.

7.2 Moyennes

La notion de *moyenne* ou *tendance centrale* (les financiers appellent cela aussi une *mesure de localisation*) est avec la notion de variable à la base des statistiques.

Cette notion nous semble très familière et nous en parlons beaucoup sans nous poser trop de questions. Pourtant il existe divers qualificatifs (nous insistons sur le fait que ce ne sont que des qualificatifs) pour distinguer la forme de la résolution d'un problème consistant à calculer la moyenne.

Il faut donc être très très prudent quand aux calculs des moyennes car nous avons une fâcheuse tendance à nous précipiter et à utiliser systématiquement la moyenne arithmétique sans réfléchir, ce qui peut amener à de graves erreurs.

Nous verrons ci-dessous différentes moyennes avec des exemples relatifs à l'arithmétique, au dénombrement, à la physique, à l'économétrie, à la géométrie. Le lecteur trouvera d'autres exemples en parcourant l'ensemble du site.

Définition 7.6 *Soient des nombres x_i réels, alors la* moyenne arithmétique *ou* moyenne empirique *est définie par :*

$$\mu_a = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7.1}$$

et très souvent notée $\bar{x}$ ou encore $\hat{\mu}$.

La moyenne arithmétique est une notion connue de tous comme étant dans les cas simple le quotient de la somme des n valeurs observées x_i par l'effectif n.

Si plusieurs valeurs occurrent plus d'une fois dans les mesures, la moyenne arithmétique sera alors bien évidemment donnée par :

$$\mu_a = \frac{1}{n}\sum_{i=1}^{n} n_i x_i \tag{7.2}$$

Remarque Le *mode*, noté Mod, est par définition la valeur qui apparaît le plus grand nombre de fois dans une série de valeurs.

Mais, pour calculer cette moyenne, nous pouvons tout aussi bien utiliser les fréquences d'apparition des valeurs observées (fréquence des classes).

Avant de continuer, indiquons que dans la pratique il est parfois difficile de choisir la largeur des intervalles de classe de l'analyse. Il faut souvent faire plusieurs essais même s'il existe des formules semi-empiriques pour choisir le nombre de classes lorsque nous avons n valeurs à disposition.

Une règle empirique utilisée par de nombreux praticiens consiste à retenir un nombre de classes k égal à l'entier le plus faible tel que :

$$2^k \geqslant n \tag{7.3}$$

La largeur de l'intervalle de classe étant alors obtenue en divisant l'étendue par k.

Indiquons aussi que par convention et en toute rigueur, un intervalle de classe est fermé à gauche et ouvert à droite : $[\ldots, \ldots[$.

Maintenant, puisque par définition $f_i = n_i/n$ et que l'une des propriétés de l'opérateur Σ est telle que, pour une constante k quelconque :

$$k\sum_i x_i = \sum_i k x_i \tag{7.4}$$

nous pouvons alors écrire avec $k = 1/n$:

$$\mu_a = \frac{1}{n}\sum_{i=1}^{r} x_i = \sum_{i=1}^{r} \frac{n_i}{n} x_i = \sum_{i=1}^{r} f_i x_i \tag{7.5}$$

Cette relation est un cas particulier de la relation ci-dessous :

$$\mu_P = \frac{1}{n}\sum_{i=1}^{r} \alpha_i x_i \tag{7.6}$$

définissant la *moyenne pondérée* ($\alpha_i = f_i$) qui pondère les valeurs x_i par un coefficient α_i tel que la somme des ces coefficients soit égale à 1.

Plus loin, nous verrons deux propriétés extrêmement importantes de la moyenne arithmétique qu'il vous faudra absolument comprendre (moyenne pondérée des écarts à la moyenne et la moyenne des écarts à la moyenne est nulle).

Définition 7.7 *La* médiane *ou* moyenne milieu *définie par la valeur qui coupe une population en deux parties égales.*

Autrement dit, soit une série de valeurs ordonnées $x_1, x_2, \ldots, x_i, \ldots, x_n$, la médiane notée M_e (ou plus simplement M) est par définition la valeur de la variable telle que l'on ait autant d'éléments qui ont une valeur qui lui est supérieure ou égale, que d'éléments qui ont une valeur qui lui est inférieure ou égale, soit, plus rigoureusement :

– Si le nombre de termes est impair, de la forme $2n+1$, la médiane de la série est le terme de rang $n+1$.
– Si le nombre de termes est pair, de la forme $2n$, la médiane de la série est la demi-somme (moyenne arithmétique) des valeurs des termes de rang n et $n+1$.

Dans tous les cas, de par cette définition, il découle qu'il y a au moins 50 % des termes de la série inférieurs ou égaux à la médiane, et au moins 50% des termes de la série supérieurs ou égaux à la médiane.

Définition 7.8 *Soit donné une série statistique $x_1, x_2, \ldots, x_i, \ldots, x_n$, nous appelons* dispersion des écarts absolus *autour de x le nombre $\varepsilon'(x)$ défini par :*

$$\varepsilon'(x) = \sum_i |x_i - x| \tag{7.7}$$

$\varepsilon'(x)$ est minimum pour une valeur de x la plus proche d'une valeur donnée x_i au sens de l'écart absolu. La médiane est la valeur qui réalise ce minimum (extremum). L'idée va alors consister à étudier les variations de la fonction pour trouver le rang de cet extremum.

En effet, nous pouvons écrire :

$$\forall x \in [x_r, x_{r+1}], r \in \{1, 2, 3, \ldots, n-1\} \quad \varepsilon'(x) = \sum_{i=1}^{r} |x_i - x| + \sum_{i=r+1}^{n} |x_i - x| \tag{7.8}$$

donc, par définition de la valeur x :

$$\begin{aligned} \varepsilon'(x) &= \sum_{i=1}^{r} |x_i - x| + \sum_{i=r+1}^{n} |x_i - x| \\ &= [rx - (x_1 + x_2 + \ldots + x_n)] + [(x_{r+1} + \ldots + x_n) - (n-r)x] \\ &= (2r - n)x + (x_{r+1} + \ldots + x_n) - (x_1 + x_2 + \ldots + x_r)] \end{aligned} \tag{7.9}$$

Ce qui nous permet donc de faire sauter les valeurs absolues est simplement le choix de l'indice r qui est pris de telle manière que la série de valeurs peut en pratique toujours être coupé en deux parties : tout ce qui est inférieur à un élément de la série indexé par r et tout ce qui lui est supérieur (la médiane donc par anticipation).

$\varepsilon'(x)$ est donc une fonction affine (assimilable à l'équation d'une droite pour r et n fixés) par morceaux (discrète) où l'on peut assimiler le facteur $2r - n$ à la pente et $(x_{r+1} + \ldots + x_n) - (x_1 + x_2 + \ldots + x_r)$ à l'ordonnée à l'origine. La fonction est donc décroissante (pente négative) tant que r est inférieur à $n/2$ et croissante quand r est supérieur à $n/2$. Plus précisément, nous distinguons deux cas qui nous intéressent particulièrement puisque n est un entier, elle passe donc par un extremum :

– Si n est pair, nous pouvons poser $n = 2n'$, alors la pente peut s'écrire $2(r - n')$ et elle est nulle si $r = n'$ et dès lors puisque ce résultat n'est valable par construction que pour $\forall x \in [x_r, x_{r+1}]$ alors $\varepsilon'(x)$ est constante sur $[x_{n'}, x_{n'+1}]$ et nous avons un extremum obligatoirement au milieu de cet intervalle (moyenne arithmétique des deux termes).
– Si n est impair, nous pouvons poser $n = 2n' + 1$ (nous coupons la série en deux parties égales), alors le pente peut s'écrire $2r - 2n' - 1$ et elle est donc nulle si $r = n' + 1/2$ et dès lors puisque ce résultat n'est valable que pour $\forall x \in [x_r, x_{r+1}]$ alors il est immédiat que la valeur du milieu sera la médiane $x_{n'+1}$.

Nous retrouvons donc bien la médiane dans les deux cas. Nous verrons aussi plus loin comment la médiane est définie pour une variable aléatoire continue.

Il existe un autre cas pratique où le statisticien n'a à sa disposition que des valeurs regroupées sous forme d'intervalles de classes statistiques. La procédure pour déterminer la médiane est alors différente :

Lorsque nous avons à notre disposition uniquement une variable classée, l'abscisse du point de la médiane se situe en général à l'intérieur d'une classe. Pour obtenir alors une valeur plus précise de la médiane, nous procédons à une interpolation linéaire. C'est ce que nous appelons la *méthode d'interpolation linéaire de la médiane.*

La valeur de la médiane peut être lue sur le graphique ou calculée analytiquement. Effectivement, considérons le graphique représentant la probabilité cumulée $F(x)$ en intervalles de classe

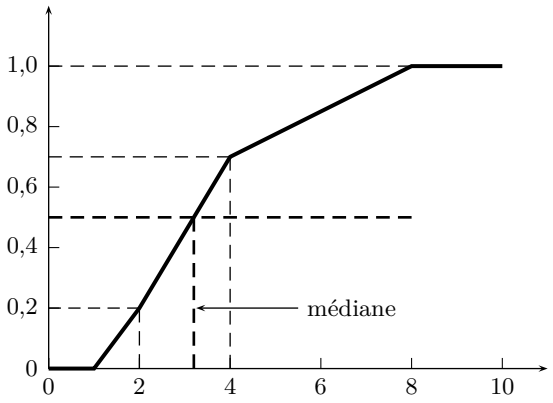


FIGURE 7.1 – Probabilité cumulée $F(x)$ en intervalles de classe et médiane

de la figure 7.1 où les bornes des intervalles ont été reliées par des droites. La valeur de la médiane M se trouve au croisement évidemment entre la probabilité de 50% (0.5) et l'abscisse. Si nous prenons dans le cadre particulier de l'exemple ci-dessus la borne supérieure de l'intervalle de classe précédant celle contenant la médiane nous avons 2 et 4 pour la borne inférieure de l'intervalle suivant. Nous avons alors en calculant la pente la relation suivante :

$$\frac{\Delta x}{\Delta y} = \frac{M-2}{0{,}5-0{,}2} = \frac{4-2}{0{,}7-0{,}2} \tag{7.10}$$

Ce que nous écrivons fréquemment :

$$\frac{M_e - a}{0{,}5 - F(a)} = \frac{b-a}{F(b)-F(a)} \tag{7.11}$$

d'où la valeur de la médiane :

$$M_e = a + (b-a)\frac{0{,}5-a}{F(b)-F(a)} \tag{7.12}$$

Prenons le tableau 7.1 que nous retrouverons bien plus tard dans le présent chapitre. Nous voyons

Montant des tickets	Nombre de tickets	Nombre cumulés de tickets	Fréquences relatives cumulées
[0 ;50[	668	668	0,068
[50,100[	919	1 587	0,1587
[100,150[	1 498	3 085	0,3085
[150,200[	1 915	5 000	0,5000
[200,250[	1 915	6 915	0,6915
[250,300[	1 498	8 413	0,8413
[300,350[	919	9 332	0,9332
[350,400[	440	9 772	0,9772
[400 et +	228	10 000	1

TABLEAU 7.1 – (7.16)

que la *classe médiane* est dans l'intervalle $[150, 200]$ car la valeur cumulée de 0.5 s'y trouve mais la médiane a elle, en utilisant la relation établie précédemment, précisément une valeur de :

$$M_e = 150 + (200-150)\frac{0{,}5-0{,}3085}{0{,}5-0{,}3085} = 200 \tag{7.13}$$

et nous pouvons faire de même avec n'importe quel autre centile.

La question qui se pose ensuite est celle de la pertinence du choix de la moyenne ou de la médiane. Un bon exemple reste celui du marché du travail où de façon générale, le salaire moyen et le salaire médian sont relativement différents.

REMARQUE Pour éviter d'obtenir une moyenne arithmétique ayant peu de sens, nous calculons souvent une *moyenne élaguée*, c'est à dire une moyenne arithmétique calculée après avoir enlevé des valeurs aberrantes à la série.

Les *quantiles* généralisent la notion de médiane en coupant la distribution en des ensembles données de parties égales ou autrement dit en intervalles réguliers. Nous définissons ainsi les *quartiles*, les *déciles* et les *centiles* (ou *percentiles*) sur la population, ordonnée dans l'ordre croissant, que nous divisons en 4, 10 ou 100 parties de même effectif.

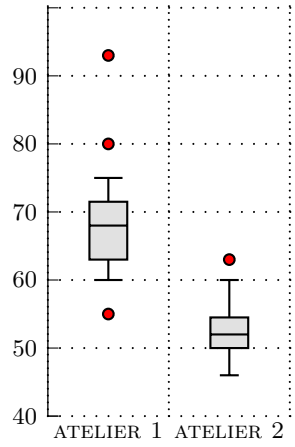
Nous parlerons ainsi du centile 90 pour indiquer la valeur séparant les premiers 90% de la population des 10% restant.

Sec7.6|p204

Ce concept est très important dans le cadre des intervalles de confiance• et très utile dans le domaine de la qualité avec l'utilisation des boîtes à moustaches permettant de comparer rapidement deux population de données et surtout d'éliminer les valeurs aberrantes. Une autre représentation

VALEURS	ATELIER 1	ATELIER 2
Q_1	63	50
min	55	45
médiane	68	52
max	93	63
Q_3	72	54

(b) grandeurs statistiques



(a) représentation graphique

atelier 1	atelier 2
65	60
68	45
72	46
55	50
64	47
70	55
66	49
74	52
75	50
70	46
93	50
64	52
60	52
62	48
70	52
80	63
61	53
60	54
62	54
68	54
70	53
72	55
71	57

(c) nombre de défaillances

FIGURE 7.2 – Analyse des défaillances techniques dans deux ateliers

mentale très importante des boîtes à moustache est indiquée sur les figures 7.3(a) et 7.3(b).

Définition 7.9 *La* moyenne quadratique *parfois simplement notée Q est définie par :*

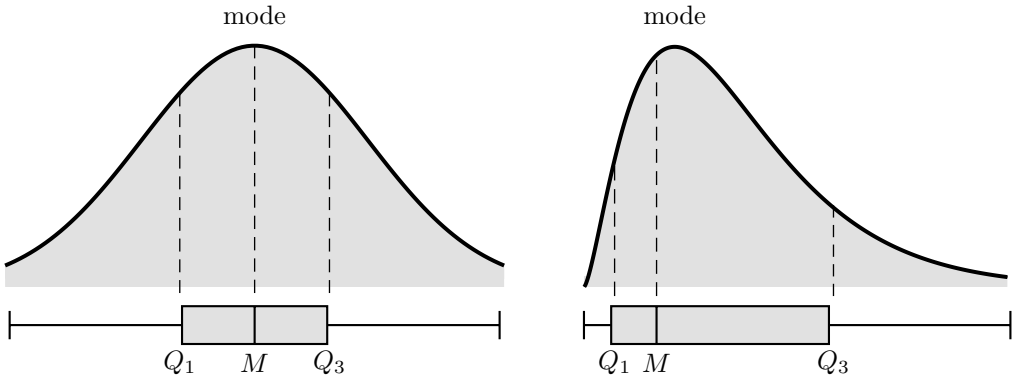
$$\mu_q = \sqrt[m]{\frac{1}{n}\sum_{i=1}^{n} {x_i}^m} \tag{7.14}$$

avec $m = 2$.

REMARQUE C'est une des moyennes les plus connues en statistiques car l'écart-type est une moyenne quadratique.

Exemple 7.1 *Soit un carré de côté a et un autre carré de côté b. La moyenne des aires des deux carrés est égale au carré de côté :*

$${\mu_q}^2 = \frac{a^2 + b^2}{2} \quad \Rightarrow \quad \mu_q = \sqrt{\frac{a^2 + b^2}{2}} \tag{7.15}$$



(a) (7.18) (b) (7.19)

FIGURE 7.3 – Représentation des boîtes à moustache

Définition 7.10 *La* moyenne harmonique *parfois simplement notée H est définie par :*

$$\frac{1}{\mu_h} = \frac{1}{n}\sum_{i=1}^{n} {x_i}^{-1} \tag{7.16}$$

peu connue mais qui découle souvent de raisonnements simples et pertinents [1].

Exemple 7.2 *Soit une distance d parcourue dans un sens à la vitesse v_1 et dans l'autre à la vitesse v_2. La vitesse moyenne s'obtiendra en divisant la distance totale $2d$ par le temps mis à la parcourir :*

$$v = \frac{2d}{t} \tag{7.17}$$

Si nous calculons le temps mis lorsqu'on parcourt d avec une vitesse v_i c'est tout simplement le quotient :

$$t_i = \frac{d}{v_i} \tag{7.18}$$

Le temps total vaut donc :

$$t = \frac{2d}{v} = t_1 + t_2 = \frac{d}{v_1} + \frac{d}{v_2} \tag{7.19}$$

La vitesse moyenne, ou son inverse pour être exacte, sera donc bien du type harmonique :

$$\frac{1}{v} = \frac{1}{2d}\left(\frac{d}{v_1} + \frac{d}{v_2}\right) = \frac{1}{2}\left(\frac{1}{v_1} + \frac{1}{v_2}\right) \tag{7.20}$$

Définition 7.11 *La* moyenne géométrique *parfois notée simplement G est définie par :*

$$\mu_g = \sqrt[n]{\prod_{i=1}^{n} x_i} \tag{7.21}$$

Cette moyenne est souvent oubliée mais néanmoins très connue dans le domaine de l'économétrie et de la finance d'entreprise (cf. chapitre Techniques De Gestion).

Exemple 7.3 *Supposons qu'une banque offre une possibilité de placement et prévoit pour la première année un intérêt de $(X-Y)\%$, mais pour la deuxième année un intérêt de $(X+Y)\%$ Au même moment une autre banque offre un intérêt constant pour deux ans $X\%$. C'est pareil, dirons-nous un peu rapidement. En fait les deux placements n'ont pas la même rentabilité.*

Dans la première banque, un capital C_0 deviendra au bout de la première année $(X-Y)\% \cdot C_0$ et la seconde année $(X+Y)\%[(X-Y)\% \cdot C_0]$. Dans l'autre banque nous aurons au bout d'un an $X\% \cdot C_0$ et après la seconde année $X\%(X\% \cdot C_0)$.

1. Il s'agit ici typiquement la résistance équivalente d'un circuit électrique ayant plusieurs résistances en parallèle.

Comme vous pouvez le voir le placement ne sera pas identique si $Y \neq 0$. Donc $X\%$ n'est donc pas la moyenne de $(X-Y)\%$ et $(X+Y)\%$.

Posons maintenant $r_1 = (X+Y)\%$ et $r_2 = (X-Y)\%$. Quelle est en fait la valeur moyenne r ?

Au bout de deux ans le capital est multiplié par $r_1 \cdot r_2$. Si la moyenne vaut r il sera alors multiplié par r^2. Nous avons donc la relation :

$$r^2 = r_1 \cdot r_2 \Leftrightarrow r = \sqrt{r_1 \cdot r_2} \tag{7.22}$$

C'est un exemple d'application où nous retrouvons donc la moyenne géométrique. L'oubli de la moyenne harmonique une erreur fréquente dans les entreprises lorsque certains employés calculent le taux moyen d'augmentation d'une valeur de référence.

Définition 7.12 *La* moyenne mobile, *appelée aussi* moyenne glissante *est définie par :*

$$MM_s = \frac{M_1 + M_2 + \ldots + M_n}{n} \tag{7.23}$$

La moyenne mobile est particulièrement utilisée en économie, où elle permet de représenter une courbe de tendance d'une série de valeurs, dont le nombre de points est égal au nombre total de points de la série de valeurs moins le nombre que vous spécifiez pour la période.

Une Moyenne Mobile (MM) en finance est calculée à partir des moyennes des cours d'une valeur, sur une période donnée : chaque point d'une moyenne mobile sur 100 séances est la moyenne des 100 derniers cours de la valeur considérée. Cette courbe, affichée simultanément avec la courbe d'évolution des cours de la valeur, permet de lisser les variations journalières de la valeur, et de dégager des tendances.

Les moyennes mobiles peuvent être calculées sur différentes périodes, ce qui permet de dégager des tendances à court terme MMC (20 séances selon les habitudes de la branche), moyen terme (50-100 séances) ou long terme MML (plus de 100 séances).

Les croisements des moyennes mobiles par la courbe des cours (découpée avec une certaine granularité) de la valeur génèrent des signaux d'achat ou de vente (selon les professionnels) suivant le cas :

- Signal d'achat : lorsque la courbe des cours franchit la MM.
- Signal de vente : lorsque la courbe des cours franchit la MM vers le bas.

Outre la moyenne mobile, précisons qu'il existe une quantité d'autres indicateurs artificiels souvent utilisés en finance comme par exemple le *upside/downside ratio*.

L'idée est la suivante : si vous avec un produit financier (cf. chapitre d'Économétrie) actuellement de prix courant P_c pour lequel vous avez un objectif de gain haut à un prix haut correspondant que noterons P_h et inversement le potentiel de perte que vous estimez à un prix P_l. Dans ce cadre, le rapport :

$$\frac{P_h - P_c}{P_c - P_l} = UD_R \tag{7.24}$$

donne le Upside/Downside Ratio.

Certaines associations boursières recommandent de refuser les UD_R inférieurs à 3. Les investisseurs ont tendance à rejeter les UD_R trop élevés pouvant être un signe de gonflage artificiel.

Définition 7.13 *La* moyenne pondérée *est définie par :*

$$\mu_p = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i} \tag{7.25}$$

La moyenne pondérée est utilisée par exemple en géométrie pour localiser le barycentre d'un polygone, en physique pour déterminer le centre de gravité ou en statistiques pour calculer une espérance (le dénominateur étant toujours égal à l'unité en probabilités) et en gestion de projets pour estimer les durées des tâches.

Dans le cas général le poids p_i représente l'influence de l'élément x_i par rapport aux autres.

Définition 7.14 *La* moyenne fonctionnelle *ou* moyenne intégrale *est définie par :*

$$\mu_f = \frac{1}{b-a}\int_a^b f(x)dx \tag{7.26}$$

où μ_f dépend d'une fonction f d'une variable réelle intégrable (cf. chapitre de Calcul Différentiel Et Intégral) sur un intervalle $[a, b]$.

7.2.1 Propriétés des moyennes

Voyons maintenant quelques propriétés pertinentes qui relient quelques-unes de ces moyennes ou qui sont propres à une moyenne donnée.

Propriété 7.1 *Le calcul des moyennes arithmétique, quadratique et harmonique peut être généralisé à l'aide de la relation suivante :*

$$\mu = \sqrt[m]{\frac{1}{n}\sum_{i=1}^{n} {x_i}^m} \tag{7.27}$$

où nous retrouvons pour $m = 1$, la moyenne arithmétique, pour $m = 2$, la moyenne quadratique et pour $m = -1$, la moyenne harmonique.

Propriété 7.2 *La moyenne arithmétique a une propriété de linéarité, c'est-à-dire que :*

$$\overline{\lambda x + \mu} = \lambda\bar{x} + \mu \tag{7.28}$$

C'est la version statistique de la propriété de l'espérance en probabilité que nous verrons plus loin.

Propriété 7.3 *La somme pondérée des écarts à la moyenne arithmétique est nulle.*

Preuve D'abord, par définition, nous savons que :

$$n = \sum_{i=1}^{r} n_i \quad \text{et} \quad \mu = \frac{1}{n}\sum_{i=1}^{r} n_i x_i \tag{7.29}$$

il s'ensuit que :

$$\sum_{i=1}^{r} n_i(x_i - \mu) = \sum_{i=1}^{r} n_i x_i - \mu\sum_{i=1}^{r} n_i = \sum_{i=1}^{r} n_i x_i - \left(\frac{1}{n}\sum_{i=1}^{r} n_i x_i\right) n = \sum_{i=1}^{r} n_i x_i - \sum_{i=1}^{r} n_i x_i = 0 \tag{7.30}$$

Ainsi, cet outil ne peut être utilisé comme mesure de dispersion. Par extension la moyenne des écarts à la moyenne pondérée par les effectifs est nulle aussi :

$$\frac{1}{n}\sum_{i=1}^{r} n_i(x_i - \mu) = 0 \tag{7.31}$$

d'où la propriété 7.3. •

Ce résultat est relativement important car il permettra plus loin de mieux saisir le concept d'écart-type et de variance.

Propriété 7.4 $\mu_h < \mu_g < \mu_a < \mu_q$

Preuve Prenons deux nombres réels non nuls x_1 et x_2 tels que $x_2 > x_1 > 0$:

1. La moyenne arithmétique :

$$\mu_a = \frac{x_1 + x_2}{2} \tag{7.32}$$

2. La moyenne géométrique :

$$\mu_g = \sqrt{x_1 x_2} \tag{7.33}$$

3. La moyenne harmonique :

$$\frac{1}{\mu_h} = \frac{\frac{1}{x_1} + \frac{1}{x_2}}{2} \Leftrightarrow \mu_h = \frac{2x_1x_2}{x_1 + x_2} \tag{7.34}$$

4. La moyenne quadratique :

$$\mu_q^2 = \frac{x_1^2 + x_2^2}{2} \tag{7.35}$$

Prouvons déjà que $\mu_g > \mu_h$ par l'absurde en posant $\mu_g - \mu_h < 0$:

$$\begin{aligned} &\mu_g - \mu_h = \sqrt{x_1x_2} - \frac{2x_1x_2}{x_1 + x_2} = \frac{\sqrt{x_1x_2}x_1 + \sqrt{x_1x_2}x_2 - 2x_1x_2}{x_1 + x_2} < 0 \\ \Rightarrow &\sqrt{x_1x_2}x_1 + \sqrt{x_1x_2}x_2 - 2x_1x_2 < 0 \\ \Rightarrow &\sqrt{x_1x_2}x_1 + \sqrt{x_1x_2}x_2 < 2x_1x_2 \\ \Rightarrow &\frac{\sqrt{x_1x_2}}{x_1} + \frac{\sqrt{x_1x_2}}{x_2} < 2 \Rightarrow \sqrt{\frac{x_1}{x_2}} + \sqrt{\frac{x_2}{x_1}} < 2 \end{aligned} \tag{7.36}$$

Par commodité posons $y = \sqrt{x_1/x_2}$, nous savons que $y > 1$ et :

$$\sqrt{\frac{x_1}{x_2}} + \sqrt{\frac{x_2}{x_1}} = \frac{1}{y} + y \tag{7.37}$$

Nous cherchons à montrer que la dernière relation de (7.36) n'est pas possible. Ceci découle des équivalences suivantes :

$$\frac{1}{y} + y = \frac{y^2 + 1}{y} < 2 \Leftrightarrow y^2 - 2y + 1 < 0 \Leftrightarrow (y - 1)^2 < 0 \tag{7.38}$$

Il y donc contradiction et ce qui vérifie $\mu_g > \mu_h$.

Regardons maintenant si $\mu_a > \mu_g$. Sous l'hypothèse $x_2 > x_1 > 0$, nous cherchons maintenant à montrer que :

$$\frac{x_1 + x_2}{2} > \sqrt{x_1x_2} \tag{7.39}$$

Or nous avons les équivalences suivantes en partant de (7.39) :

$$(x_1 + x_2)^2 > 4x_1x_2 \Leftrightarrow (x_1 - x_2)^2 > 0 \tag{7.40}$$

et la dernière expression est évidement correcte d'où $\mu_a > \mu_g$.

Nous prouvons maintenant $\mu_q > \mu_a$ et démontrons-le par l'absurde en posant $\mu_q - \mu_a < 0$:

$$\begin{aligned} &\mu_q - \mu_a = \sqrt{\frac{x_1^2 + x_2^2}{2}} - \frac{x_1 + x_2}{2} < 0 \Rightarrow \sqrt{\frac{x_1^2 + x_2^2}{2}} < \frac{x_1 + x_2}{2} \\ \Rightarrow &\frac{x_1^2 + x_2^2}{2} < \frac{x_1^2 + 2x_1x_2 + x_2^2}{4} \\ \Rightarrow &2x_1^2 + 2x_2^2 < x_1^2 + 2x_1x_2 + x_2^2 \Rightarrow x_1^2 - 2x_1x_2 + 2x_2^2 < 0 \Rightarrow (x_1 - x_2)^2 < 0 \end{aligned} \tag{7.41}$$

or, le carré d'un nombre est toujours positif ce qui vérifie $\mu_q > \mu_a$ Nous avons donc la propriété 7.4. •

REMARQUE Les comparaisons entre les moyennes précitées et la médiane ou encore les moyennes glissantes et pondérées n'ont pas de sens c'est pour cela que nous nous abstenons à les faire.

Démontrons par l'absurde que $MM_p > MM_s$ en posant $x_2 > x_1$ et que $MM_p < MM_s$.

Preuve Nous avons alors :

$$\frac{x_1 + 2x_2}{3} < \frac{x_1 + x_2}{2} \Leftrightarrow \frac{2x_1 + 4x_2}{3} < x_1 + x_2 \Leftrightarrow \frac{4}{3}x_2 - x_2 < x_1 - \frac{2}{3}x_1 \Leftrightarrow \frac{x_2}{3} < \frac{x_1}{3} \Leftrightarrow x_2 < x_1 \tag{7.42}$$

Il y a donc contradiction avec l'hypothèse initiale et nous avons donc bien :

$$MM_p > MM_s = \mu_a \tag{7.43}$$

•

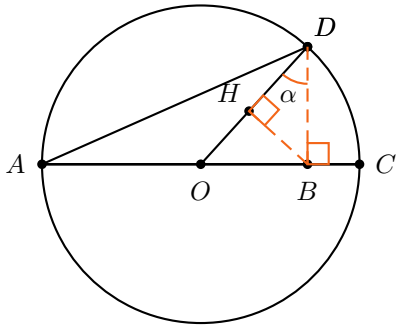


FIGURE 7.4 – Moyennes selon ARCHIMÈDE

Ces inégalités démontrées, nous pouvons alors passer à la figure 7.4 que nous attribuons à ARCHIMÈDE pour placer trois de ces moyennes. L'intérêt de cet exemple est de montrer qu'il existe des relations remarquables parfois entre la statistique et la géométrie. Nous allons d'abord poser $a = \overline{AB}$, $b = \overline{BC}$ et O est le milieu de $\overline{AC}$. Ainsi, le cercle Ω est de centre O et de rayon $\overline{OA}$. D est l'intersection de la perpendiculaire à $\overline{AC}$ passant par B et du cercle Ω. H est quant à lui le projeté orthogonal de B sur $\overline{OD}$.

ARCHIMÈDE affirme que $\overline{OA}$ est la moyenne arithmétique de a et b et que $\overline{BD}$ est la moyenne géométrique de a et b, et $\overline{OH}$ la moyenne harmonique de a et b.

$$\overline{OA} = \frac{\overline{AC}}{2} = \frac{a+b}{2} \tag{7.44}$$

Par définition $\overline{OA} = \overline{AC}/2 = (a+b)/2$, donc $\overline{OA}$ est bien la moyenne arithmétique μ_a de a et b. Ensuite nous avons dans le triangle rectangle ADB :

$$\overline{AD}^2 = \overline{DB}^2 + \overline{BA}^2 \tag{7.45}$$

Puis dans le triangle rectangle $\overline{BDC}$ nous avons :

$$\overline{DC}^2 = \overline{BC}^2 + \overline{DB}^2 \tag{7.46}$$

Nous additionnons alors ces deux égalités, et nous trouvons :

$$2\overline{DB}^2 + \overline{BA}^2 + \overline{BC}^2 = \overline{AD}^2 + \overline{DC}^2 \tag{7.47}$$

Nous savons que D est sur un cercle de diamètre $\overline{AC}$, donc ADC est rectangle en D, donc :

$$\overline{AD}^2 + \overline{DC}^2 = \overline{AC}^2 \tag{7.48}$$

Puis nous remplaçons $\overline{BA}$ et $\overline{BC}$ par a et b :

$$2\overline{DB}^2 + a^2 + b^2 = \overline{AC}^2 = (a+b)^2 \Leftrightarrow 2\overline{DB}^2 = 2ab \Leftrightarrow \overline{DB} = \sqrt{ab} \tag{7.49}$$

Et donc, $\overline{DB}$ est bien la moyenne géométrique μ_g de a et b. Il nous reste à prouver alors que $\overline{OH}$ est la moyenne harmonique de a et b. Nous avons dans un premier temps (projection orthogonale) :

$$\vec{DO} \cdot \vec{DB} = \overline{DO} \cdot (\overline{DB}\cos\alpha) = \overline{DO} \cdot \overline{DH} = \frac{a+b}{2}\overline{DH} \tag{7.50}$$

Puis nous avons aussi (projection orthogonale aussi) :

$$\vec{DO} \cdot \vec{DB} = (\overline{DO}\cos\alpha) \cdot \overline{DB} = \overline{DB} \cdot \overline{DB} = \overline{DB}^2 \tag{7.51}$$

Nous avons donc :

$$\overline{DB}^2 = \frac{a+b}{2}\overline{DH} \tag{7.52}$$

et comme $\overline{DB} = \sqrt{ab}$, nous avons donc :

$$\overline{DH} = \frac{2ab}{a+b} \tag{7.53}$$

$\overline{DH}$ est donc bien la moyenne harmonique de a et b, ARCHIMÈDE ne s'était pas trompé.

7.3 Types de variables

Lorsque nous avons parlé des échantillons au début de ce chapitre, nous avons fait mention de deux types d'informations : les variables quantitatives et qualitatives. Nous n'avons cependant pas précisé qu'il existait deux types de variables quantitatives très importantes qu'il convient absolument de différencier : les variables discrètes et les variables continues.

7.3.1 Variables discrètes

Soit X une variable indépendante (un élément d'un échantillon dont la propriété est indépendante des autres éléments) qui peut prendre les valeurs aléatoires discrètes $x_1, x_2, x_3 \ldots$ dans $\mathbb{R}$ avec les probabilités respectives $p_1, p_2, p_3 \ldots$ où, de par l'axiomatique des probabilités :

$$p_i \in [0,1] \quad \sum_i p_i = 1 \tag{7.54}$$

Définition 7.15 L'espérance mathématique *de la variable X est définie par :*

$$\mu_X = E(X) = \sum_i p_i x_i \tag{7.55}$$

En d'autres termes, nous savons qu'à chaque événement de l'espace des échantillons est associé une probabilité à laquelle nous associons également une valeur (donnée par la variable aléatoire). La question étant alors de savoir quelle valeur, à long terme, nous pouvons obtenir. La valeur espérée, (l'espérance mathématique donc...) est alors la moyenne pondérée, par la probabilité, de toutes les valeurs des événements de l'espace des échantillons.

Si la probabilité est donnée par une fonction de distribution (voir les définitions des fonctions de distribution plus bas) de la variable aléatoire, nous avons :

$$p_i = f(x_i) \Rightarrow \mu_X = E(X) = \sum_i x_i f(x_i) \tag{7.56}$$

REMARQUE μ_X peut être notée μ s'il n'y pas de confusion possible.

Voici les propriétés mathématiques les plus importantes de l'espérance pour toute variable aléatoire ou pour toute série de variables aléatoires et que nous utiliserons souvent tout au long de ce chapitre :

Propriété 7.5 *Multiplication par une constante :*

$$E(\alpha X) = \alpha E(X) \tag{7.57}$$

Propriété 7.6 *Somme de deux variables aléatoires :*

$$E(X+Y) = E(X) + E(Y) \tag{7.58}$$

Preuve où nous avons utilisé dans la 4ème ligne, la propriété vue dans le chapitre de Probabilités :
equation
Nous en déduisons que pour n variables aléatoires equation, définies sur une même loi de distribution :
equation (7.77) •

Propriété 7.7 *Espérance d'une constante C :*

$$E(C) = C \tag{7.59}$$

Après avoir traduit la tendance par l'espérance, il est intéressant de traduire la dispersion ou *déviation standard* autour de l'espérance par une valeur appelée *variance de X*, notée $V(X)$ ou σ_X^2 (lire sigma-deux) et donnée sous sa forme discrète par :

$$\sigma_X^2 = V(X) = E[(X - \mu_X)^2] = \sum_i (x_i - \mu_X)^2 p_i \tag{7.60}$$

La variance n'est cependant pas comparable directement à la moyenne, car l'unité de la variance est le carré de l'unité de la variable, ce qui découle directement de sa définition. Pour que l'indicateur de dispersion puisse être comparé aux paramètres de tendance centrale (moyenne, médiane et... mode), il suffit d'en prendre la racine carrée.

Par commodité, nous définissons ainsi l'*écart-type* de X, noté $\sigma(X)$, par :

$$\sigma(X) = \sigma_X = \sqrt{V(X)} \tag{7.61}$$

L'écart-type est donc la moyenne quadratique des écarts entre les observations et leur moyenne.

Remarques

1. L'écart-type σ_X de la variable aléatoire X peut être noté σ s'il n'y pas de confusion possible ;
2. L'écart-type et la variance sont, dans la littérature, souvent appelés *paramètres de dispersion*.

Définition 7.16 *Le rapport σ/μ (exprimé en %) parfois utilisé dans les entreprises comme comparaison de la moyenne et de l'écart-type est appelée le* coefficient de variation.

Pourquoi trouvons-nous un carré (réciproquement une racine) dans cette définition de la variance ? La raison intuitive est simple (la rigoureuse l'est nettement moins...). Nous avons démontré plus haut que la somme des écarts à la moyenne pondéré par les effectifs, est toujours nulle :

$$\sum_{i=1}^{r} n_i(x_i - \mu) = 0 \tag{7.62}$$

Or, si nous assimilons les effectifs par la probabilité en normalisant ceux-ci par rapport à n, nous tombons sur une relation qui est la même que la variance à la différence que le terme entre parenthèse n'est pas au carré. Et nous voyons alors immédiatement le problème : la mesure de dispersion serait toujours nulle d'où la nécessité de porter cela au carré.

Nous pourrions imaginer cependant d'utiliser la valeur absolue des écarts à la moyenne, mais pour un certain nombre de raisons que nous verrons plus loin lors de notre étude des estimateurs le choix de porter au carré intervient s'impose assez naturellement.

Dans le cas où nous avons à disposition une série de mesures, nous pouvons estimer la valeur moyenne (l'espérance) et la variance des mesures par les estimateurs suivants (il s'agit simplement au fait de l'espérance et l'écart-type d'un échantillon dont les événements sont tous équiprobables) dont la notation est particulière :

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \tag{7.63}$$

Preuve Par développement, nous obtenons :

$$\begin{aligned}
\mu = E(X) &= \sum_{i=1}^{n} p_i x_i = \sum_{i=1}^{n} \frac{1}{n} x_i = \frac{1}{n}\sum_{i=1}^{n} x_i = \hat{\mu} \\
\sigma^2 = E(X - E(X))^2 &= E[(X - \mu_X)^2] \\
&= \sum_{i=1}^{n}(x_i - \mu_X)^2 p_i = \sum_{i=1}^{n}(x_i - \mu_X)^2 \frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \hat{\sigma}^2
\end{aligned} \tag{7.64}$$

d'où (7.63). •

Le terme de la somme se trouvant dans l'expression de la variance (écart-type) est appelée *somme des carrés des écarts à la moyenne*. Nous l'appelons aussi la *somme des carrés totale*, ou encore la *variation totale*.

Remarque Il est important que le lecteur comprenne que dans ce cas l'espérance se calcule simplement en utilisant la moyenne arithmétique.

La variance peut également s'écrire sous la forme de la formule de Huyghens :

$$V(X) = E(X^2) - E(X)^2 \tag{7.65}$$

que nous réutiliserons plusieurs fois.

Preuve Voyons de quoi il s'agit :

$$\begin{aligned} V(X) &= E[(X-\mu_X)^2] = \sum_i (x_i-\mu_X)^2 f(x_i) \\ &= \sum_i x_i^2 f(x_i) - 2\mu \sum_i x_i f(x_i) + \mu_X^2 \sum_i f(x_i) = \sum_i x_i^2 f(x_i) - 2\mu_X^2 + \mu_X^2 \\ &= \sum_i x_i^2 f(x_i) - \mu_X^2 = E(X^2) - \mu_X^2 = E(X^2) - E(X)^2 \end{aligned} \tag{7.66}$$

d'où la relation (7.65). •

Soit X une variable aléatoire d'espérance μ (valeur constante et déterminée) et de variance σ^2 (valeur constante et déterminée), nous définissons la *variable centrée réduite* par la relation :

$$Y = \frac{X-\mu}{\sigma} \tag{7.67}$$

et l'on démontre de façon très simple en utilisant la propriété de linéarité de l'espérance et la propriété de multiplication par un scalaire de la variance que :

$$E(Y) = 0; \qquad V(Y) = 1 \tag{7.68}$$

Preuve Par expansion et en utilisant (7.67) et (7.65), il vient :

$$E(Y) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}(E(X) - E(\mu)) = \frac{1}{\sigma}(\mu - \mu E(1)) = \mu_Y = 0 \tag{7.69}$$

et en utilisant la formule de HUYGHENS :

$$\begin{aligned} V(Y) &= E(Y^2) - \mu_Y^2 = E(Y^2) = E\left(\frac{1}{\sigma^2}(X-\mu)^2\right) = \frac{1}{\sigma^2}(E(X^2) - 2E(X)\mu + \mu^2) \\ &= \frac{1}{\sigma^2}(E(X^2) - 2\mu^2 + \mu^2) = \frac{1}{\sigma^2}(E(X^2) - \mu^2) = \frac{1}{\sigma^2}(V(X) + \mu^2 - \mu^2) = \frac{V(X)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1 \end{aligned} \tag{7.70}$$

d'où (7.68). •

Ainsi, toute répartition statistique définie par une moyenne et un écart-type peut être transformée en une autre distribution statistique souvent plus simple à analyser.

Voici quelques propriétés mathématiques importantes de la variance :

Propriété 7.8 (Multiplication par une constante)

$$V(nX) = \sum_i f(x_i)[nx_i - n\mu_X]^2 = n^2 \sum_i f(x_i)[X-\mu_X]^2 = n^2 V(X) \tag{7.71}$$

Propriété 7.9 (Somme de deux variables aléatoires)

$$V(X+Y) = V(X) + V(Y) + 2\text{cov}(X,Y) \tag{7.72}$$

où nous introduisons le concept de covariance.

Preuve Développons le terme $V(X+Y)$:

$$\begin{aligned} V(X+Y) &= \sum_i \sum_j \sum_k \big((x_i + y_j + z_k) - (\mu_X + \mu_Y + \mu_Z)\big)^2 f(x_i, y_j) \\ &= \sum_i \sum_j \big((x_i - \mu_X) + (y_j - \mu_Y)\big)^2 f(x_i, y_j) \\ &= \sum_i \sum_j \big((x_i-\mu_X)^2 + 2(x_i-\mu_X)(y_j-\mu_Y) + (y_j-\mu_Y)^2\big) f(x_i, y_j) \\ &= V(X) + V(Y) + 2\sum_{i,j} (x_i-\mu_X)(y_j-\mu_Y) f(x_i, y_j) \\ &= V(X) + V(Y) + 2E[(x_i-\mu_X)(y_j-\mu_Y)] = V(X) + V(Y) + 2\text{cov}(X,Y) \end{aligned} \tag{7.73}$$

ce qui démontre la propriété 7.9. Introduisons une forme plus générale et extrêmement importante dans de nombreux domaines :

$$\begin{aligned}
V(X+Y+Z) &= \sum_i \sum_j ((x_i+y_j)-(\mu_X+\mu_Y))^2 f(x_i,y_j) \\
&= \sum_i \sum_j ((x_i-\mu_X)+(y_j-\mu_Y))^2 f(x_i,y_j) \\
&= \sum_i \sum_j ((x_i-\mu_X)^2+2(x_i-\mu_X)(y_j-\mu_Y)+(y_j-\mu_Y)^2) f(x_i,y_j) \\
&= V(X)+V(Y)+2\sum_{i,j}(x_i-\mu_X)(y_j-\mu_Y)f(x_i,y_j) \\
&= V(X)+V(Y)+2E[(x_i-\mu_X)(y_j-\mu_Y)] = V(X)+V(Y)+2\text{cov}(X,Y)
\end{aligned} \tag{7.74}$$

ce qui entraîne aussi la propriété 7.10. •

Propriété 7.10 (Somme de variables aléatoires) *Donc dans le cas général :*

$$V\left(\sum_i^n X_i\right) = \sum_i^n X_i + 2\sum_{i<j}^n \text{cov}(X_i,X_j) \tag{7.75}$$

En utilisant la linéarité de l'espérance et l'équation (7.59), nous avons pour la covariance :

$$\begin{aligned}
\text{cov}(X,Y) &= E[(X-E(X))(Y-E(Y))] \\
&= E[XY-E(X)Y-XE(Y)+E(X)E(Y)] \\
&= E(XY)-E(E(X)Y)-E(XE(Y))+E(E(X)E(Y)) \\
&= E(XY)-E(X)E(Y)-E(Y)E(X)+E(X)E(Y)E(1) \\
&= E(XY)-2E(X)E(Y)+E(X)E(Y)E(1) \\
&= E(XY)-E(X)E(Y)
\end{aligned} \tag{7.76}$$

et donc nous obtenons la relation très utilisée en statistiques et finance :

$$C_{X,Y} = \text{cov}(X,Y) = E(XY)-E(X)E(Y) \tag{7.77}$$

Indiquons également que si $X=Y$, nous retrouvons la formule de Huyghens :

$$C_{X,Y} = E(X^2)-\mu_X^2 \tag{7.78}$$

Ainsi, le terme de covariance $C_{X,Y}$ est défini par l'expression :

$$C_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)] \tag{7.79}$$

appelée *forme bilinéaire de la variance* ou *forme multivariée*.

Remarque Les statistiques peuvent être découpées selon le nombre de variables aléatoires que nous étudions. Ainsi, lorsqu'une seule variable aléatoire est étudiée, nous parlons de *statistique univariée*, pour deux variables aléatoires de *statistique bivariée* et en général, de *statistique multivariée*.

Si la covariance est univariée, nous avons dès lors :

$$C_{X,X} = E[(X-\mu_X)^2] = V(X) = \sigma_X^2 \tag{7.80}$$

Si les variables sont équiprobables, nous la retrouvons dans la littérature sous la forme suivante qui découle de calculs que nous avons déjà fait ultérieurement avec l'espérance :

$$C_{X,X} = \frac{1}{n}\sum_{i=1}^n (y_i-\mu_Y)(x_i-\mu_X) \tag{7.81}$$

La covariance est un indicateur de la variation simultanée de X et Y. En effet si, en général X et Y croissent simultanément, les produits $(y_i-\mu_Y)(x_i-\mu_X)$ seront positifs (corrélés positivement), tandis que si Y décroît lorsque X croît, ces même produits seront négatifs (corrélés négativement).

Soit $\vec{X}$ un vecteur de composantes (x_i) et $\vec{Y}$ un autre vecteur de composantes (y_i), tous deux étant des variables aléatoires, le calcul de la covariance des composantes deux à deux donnent ce que l'on appelle la *matrice des covariances*.

Effectivement, si nous notons :

$$c_{X_n,X_m} = E[(x_n - \mu_X)(x_m - \mu_X)] = c_{n,m} \tag{7.82}$$

Nous pouvons dès lors écrire une matrice symétrique (le plus souvent dans la pratique elle est carrée) sous la forme :

$$c_{n,m} = \begin{bmatrix} c_{11} & \cdots & \cdots & c_{1n} \\ \vdots & c_{22} & & \vdots \\ \vdots & & c_{33} & \vdots \\ c_{m1} & \cdots & \cdots & c_{mm} \end{bmatrix} \tag{7.83}$$

Cette matrice a comme propriété remarquable que si nous prenons deux vecteurs identiques (dont les composantes sont les mêmes variables aléatoires) et que nous calculons la matrice, alors la diagonale de cette dernière donnera les variances des composantes de vecteurs (voir les exemples dans le chapitre d'économétrie). Raisons pour laquelle cette matrice est souvent appelée *matrices des variances-covariances*.

REMARQUE Cette matrice est très importante et nous la retrouverons fréquemment dans le chapitre d'Économétrie lors de notre étude da la théorie du portefeuille et dans les techniques de fouille de données (data mining, clustering) dans le chapitre de Méthodes numériques (l'analyse par composantes principales).

Rappelons maintenant que nous avions un axiome en probabilités (cf. chapitre de Probabilités) qui énonçait que deux événements A, B sont indépendants si :

$$P(A \cap B) = P(A) \cdot P(B) \tag{7.84}$$

De la même façon, par extension, nous définissons l'indépendance des variables aléatoires discrètes.

Définition 7.17 *Soit X, Y deux variables aléatoires discrètes. Nous disons que X, Y sont indépendantes si :*

$$\forall x, y \in \mathbb{R}, P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \tag{7.85}$$

Plus généralement, les variables discrètes $X_1, \ldots, X_n$ sont indépendantes (en bloc) si :

$$\forall x_1 \ldots x_n \in \mathbb{R}, P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) \tag{7.86}$$

L'indépendance de deux variables aléatoires implique que leur covariance est nulle (la réciproque est fausse). Prouvons ceci dans le cas où les variables aléatoires ne prennent qu'un nombre fini de valeurs $\{x_i\}_I$ et $\{y_j\}_J$ respectivement, avec I, J des ensembles finis :

$$\begin{aligned} E(X \cdot Y) &= \sum_{i,j} P(X = x_i, Y = y_j) x_i y_j = \sum_{i,j} P(X = x_i) \cdot P(Y = y_j) x_i y_j \\ &= \left(\sum_i P(X = x_i) x_i \right) \cdot \left(\sum_j P(X = y_j) y_j \right) = E(X) \cdot E(Y) \end{aligned} \tag{7.87}$$

et donc :

$$c_{X,Y} = E(X \cdot Y) - E(X) \cdot E(Y) = 0 \tag{7.88}$$

REMARQUE Donc plus la covariance est faible, plus les séries sont indépendantes. À l'inverse, plus la covariance est élevée, plus les séries sont liées.

Étant donné que :

$$V(X+Y) = V(X) + V(Y) + 2c_{X,Y} \tag{7.89}$$

si X, Y sont indépendantes alors $c_{X,Y} = 0$ et :

$$V(X+Y) = V(X) + V(Y) \tag{7.90}$$

De manière plus générale si $X_1, \ldots, X_n$ sont indépendantes (en bloc) alors pour toute loi statistique, nous avons :

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) \tag{7.91}$$

Souvent en statistique, il est utile de déterminer l'écart-type de la moyenne empirique (ou en d'autres termes, l'erreur quadratique moyenne). Voyons de quoi il s'agit. Soit la moyenne d'une série de termes déterminés chacun par la mesure de plusieurs valeurs (il s'agit au fait de son estimateur dans un cas particulier comme nous le verrons beaucoup plus loin) :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{7.92}$$

alors en utilisant les propriétés de l'espérance :

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \tag{7.93}$$

et si toutes les variables aléatoires sont identiquement distribuées et indépendantes nous avons alors :

$$E(\bar{X}) = \frac{1}{n} n\mu = \mu \tag{7.94}$$

Pour la variance, le même raisonnement s'applique :

$$V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 \tag{7.95}$$

et si les variables aléatoires sont toutes identiquement distribuées :

$$V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \tag{7.96}$$

d'où l'écart-type de la moyenne appelé aussi *erreur type*, *erreur standard* ou encore *variation non systématique* :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \tag{7.97}$$

et il s'agit rigoureusement de l'écart-type de l'estimateur de la moyenne.

Signalons que la dernière relation peut-être utilisée même si la moyenne des n variables aléatoires n'est pas identique. La condition suffisante étant juste que les écarts-types soient tous égaux et c'est le cas de l'industrie.

Nous avons donc :

$$E(S_n) = n\mu \qquad E(M_n) = \mu \qquad V(S_n) = \sigma^2(S_n) = n\sigma^2 \qquad V(M_n) = \sigma^2(M_n) = \frac{\sigma^2}{n} \tag{7.98}$$

où S_n désigne la somme des n variables aléatoires et M_n leur moyenne. La variable centrée réduite (7.67) peut alors s'écrire de plusieurs manières :

$$Y = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{M_n - \mu}{\sigma/\sqrt{n}} \tag{7.99}$$

Par ailleurs, en supposant connue la loi normale $N(\mu, \sigma)$, nous démontrerons plus loin en détails que la loi de probabilité de la variable aléatoire $\bar{X}$, moyenne de n variables aléatoires identiquement distribuées et linéairement indépendantes, est alors la loi :

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \tag{7.100}$$

Maintenant, considérons X et Y deux variables aléatoires ayant pour covariance $c_{X,Y}$ alors :

$$(c_{X,Y})^2 \leqslant V(X) \cdot V(Y) \tag{7.101}$$

nous allons démontrer cette relation immédiatement car l'utilisation de la covariance seule pour l'analyse des données n'est pas géniale car elle n'est pas à proprement parler bornée et simple d'usage (au niveau de l'interprétation). Nous allons donc construire un indicateur plus facile d'usage en entreprise.

Preuve Choisissons une constante a quelconque et calculons la variance de $aX + Y$. Nous pouvons alors immédiatement écrire à l'aide des propriétés de la variance et de l'espérance :

$$V(aX + Y) = a^2 V(X) + 2ac_{X,Y} + V(Y) \tag{7.102}$$

La quantité de droite est positive et nulle en tout a par construction de la variance (de gauche). Donc le discriminant de l'expression, vue comme un trinôme en a du type :

$$P(x) = ax^2 + bx + C = a\left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a^2}\right] \tag{7.103}$$

soit :

$$P(a) = V(X)\left[\left(a + \frac{2c_{X,Y}}{2V(X)}\right)^2 - \frac{(2c_{X,Y})^2 - 4V(X)V(Y)}{4V(X)^2}\right] \tag{7.104}$$

Donc pour que $P(a)$ soit positif pour tout a nous avons comme seule possibilité que :

$$(2c_{X,Y})^2 - 4V(X) \cdot V(Y) \leqslant 0 \tag{7.105}$$

Soit après simplification la condition (7.101). •

La condition (7.101) implique :

$$\frac{(c_{X,Y})^2}{V(X) \cdot V(Y)} \leqslant 1 \Leftrightarrow \frac{|c_{X,Y}|}{\sqrt{V(X) \cdot V(Y)}} \leqslant 1 \tag{7.106}$$

Finalement nous obtenons une forme de l'inégalité statistique dite *inégalité de* Cauchy-Schwarz :

$$1 \leqslant \frac{c_{X,Y}}{\sqrt{V(X) \cdot V(Y)}} \leqslant 1 \tag{7.107}$$

Si les variances de X et Y sont non nulles, la corrélation entre X et Y est définie par le *coefficient de corrélation linéaire* :

$$R_{X,Y} = \frac{c_{X,Y}}{\sqrt{V(X) \cdot V(Y)}} \tag{7.108}$$

ce qui peut aussi s'écrire sous forme développée (en utilisant la formule de HUYGHENS) :

$$R_{X,Y} = \frac{c_{X,Y}}{\sigma(X)\sigma(Y)} \tag{7.109}$$

Quels que soient l'unité et les ordres de grandeur, le coefficient de corrélation est un nombre sans unité, compris entre -1 et 1. Il traduit la plus ou moins grande dépendance linéaire de X et Y ou, géométriquement, le plus ou moins grand aplatissement. Un coefficient de corrélation nul ou proche de 0 signifie qu'il n'y a pas de relation linéaire entre les caractères. Mais il n'entraîne aucune notion d'indépendance plus générale.

Quand le coefficient de corrélation est proche de 1 ou -1, les caractères sont dits fortement corrélés. Il faut prendre garde à la confusion fréquente entre corrélation et causalité. Cependant, que deux phénomènes soient corrélés n'implique en aucune façon que l'un soit cause de l'autre :

- si $R_{X,Y} = -1$, nous avons affaire à une corrélation négative dite *corrélation négative parfaite*[2] ;
- si $-1 < R_{X,Y} < 1$, nous avons affaire à une corrélation négative ou positive dite *corrélation imparfaite* ou la relation linéaire sera respectivement décroissante ou croissante ;
- si $R_{X,Y} = 0$, la corrélation est nulle ;
- si $R_{X,Y} = 1$, nous avons affaire à une corrélation positive dite *corrélation positive parfaite*[3].

L'analyse de régression et de corrélation poursuit donc deux objectifs :

1. Déterminer le degré d'association entre les différentes variables : celui-ci est exprimé par le coefficient de détermination, qui est le carré du coefficient de corrélation. Le coefficient de détermination mesure la contribution d'une des variables à l'explication de la seconde ;
2. Déterminer les caractéristiques de cette association, c'est-à-dire des paramètres α et β de la droite de régression (voir la section d'analyse numérique du site au chapitre des algorithmes traitant de la régression linéaire). Si l'on peut faire valablement l'hypothèse de la stabilité du processus générateur des couples de valeurs des deux variables, la connaissance de ces paramètres permettrait de prédire le comportement du phénomène étudié.

En utilisant les expressions de la moyenne et de l'écart-type de variables équiprobables tel que démontré plus haut, nous passons de :

$$R_{X,Y} = \frac{E(X,Y) - E(X)E(Y)}{\sqrt{(E(X^2) - E(X)^2)(E(Y^2) - E(Y)^2)}} \tag{7.110}$$

à :

$$R_{X,Y} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_k y_k - \frac{1}{n^2}\sum_{i=1}^{n} x_k \sum_{i=1}^{n} y_k}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n} x_k^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_k\right)^2\right]\left[\frac{1}{n}\sum_{i=1}^{n} y_k^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} y_k\right)^2\right]}} \tag{7.111}$$

où nous voyons que la covariance devient alors la moyenne des produits moins le produit des moyennes, soit après simplification :

$$R_{X,Y} = \frac{\sum_{i=1}^{n} x_k y_k - \frac{1}{n}\sum_{i=1}^{n} x_k \sum_{i=1}^{n} y_k}{\sqrt{\left[\sum_{i=1}^{n} x_k^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_k\right)^2\right]\left[\sum_{i=1}^{n} y_k^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_k\right)^2\right]}} \tag{7.112}$$

Remarques

1. Dans la littérature le coefficient de corrélation est souvent appelée *coefficient d'échantillonnage de* Pearson (dans le cas équiprobable) ou *test de* Bravais-Pearson (dans le cas non équiprobable) et lorsque nous le portons au carré, nous parlons alors de *coefficient de détermination.*
2. Souvent le carré de ce coefficient est un peu abusivement interprété comme le % de variation expliqué de la variable étudiée Y par la variable explicative X.

7.3.2 Variables continues

Définition 7.18 *Nous disons que X est une variable continue si sa* fonction de répartition *est continue. La fonction de répartition de X étant définie par :*

$$F(x) = P(X \leqslant x) \quad \forall x \in \mathbb{R} \tag{7.113}$$

soit la probabilité cumulée que la variable aléatoire X soit plus petite ou égale à la valeur x fixée. Nous avons aussi bien évidemment $P(X) \geqslant 0$.

2. Dans ce cas de figure, tous les points de mesures sont situés sur une droite de régression de pente négative.
3. Dans ce cas de figure, tous les points de mesures sont situés sur une droite de régression de pente positive.

Définition 7.19 *Si de plus la fonction de répartition F de X est continûment dérivable de dérivée ρ appelée* fonction de densité *ou* fonction de masse *ou encore* fonction de distribution *alors nous disons que X est absolument continue et dans ce cas nous avons :*

$$P(x_1 \leqslant X \leqslant x_2) = \int_{x_1}^{x_2} \rho(x)dx \tag{7.114}$$

avec la condition de normalisation :

$$P(X \leqslant +\infty) = \int_{-\infty}^{+\infty} \rho(x)dx = 1 \tag{7.115}$$

Toute fonction de distribution de probabilité doit satisfaire l'intégrale de normalisation dans son domaine de définition.

REMARQUE Il est intéressant de remarquer que la définition amène à ce que la probabilité qu'une variable aléatoire totalement continue prenne une valeur donnée est nulle !. Donc ce n'est pas parce qu'un événement à une probabilité nulle qu'il ne peut arriver.

La moyenne ayant été définie par la somme pour une variable discrète, elle devient une intégrale pour une variable continue :

$$E(X) = \int_{-\infty}^{+\infty} x\rho(x)dx = 1 \tag{7.116}$$

et la variance s'écrit donc :

$$E(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2\rho(x)dx = 1 \tag{7.117}$$

Nous avons alors aussi la médiane qui est logiquement redéfinie dans le cas d'une variable aléatoire continue par :

$$M_e = \int_{-\infty}^{M_e} dF(x) = \int_{-\infty}^{M_e} \rho(x)dx = 1/2 \tag{7.118}$$

et elle coïncide rarement avec la moyenne.

Souvent les statisticiens utilisent les mêmes notations pour l'espérance mathématique d'une variable continue $E(X), M(X), \mu_X, \mu$ et pour la variance $V(X), \sigma_X, \sigma^2$ que pour une variable discrète.

Par la suite, nous calculerons ces différents termes avec développements uniquement dans les cas les plus usités.

7.4 Lois de distributions

Lorsque nous observons des phénomènes probabilistes, et que nous prenons note des valeurs prises par ces derniers et que nous les reportons graphiquement, nous observons toujours que les différentes mesures obtenues suivent une caractéristique courbe ou droite typique fréquemment reproductible.

Dans le domaine des probabilités et statistiques, nous appelons ces caractéristiques des *fonctions de distribution* car elles indiquent la fréquence avec laquelle la variable aléatoire apparaît avec certaines valeurs.

REMARQUE Nous utilisons aussi simplement le terme *fonction* ou encore *loi* pour désigner ces caractéristiques.

Ces fonctions sont en pratique bornées par ce que nous appelons *l'étendue de la distribution* qui correspond à la différence entre la donnée maximale (à droite) et la donnée minimale (à gauche) des valeurs des données observées :

$$E = Max - Min \tag{7.119}$$

Si les valeurs observées se distribuent d'une certaine manière c'est qu'elles ont alors une probabilité d'avoir une certaine valeur de la fonction de distribution.

Définition 7.20 *La relation mathématique qui donne la probabilité qu'a une variable aléatoire d'avoir une valeur donnée de la fonction de distribution est appelée* fonction de densité, fonction de masse *ou encore* fonction marginale.

Définition 7.21 *La relation mathématique qui donne la probabilité cumulée qu'a une variable aléatoire d'être inférieure ou égale à une certaine valeur est nommée la* fonction de répartition *ou* fonction cumulée.

Définition 7.22 *Des variables aléatoires sont dites* indépendantes et identiquement distribuées *(i.i.d.) si elles suivent toutes la même fonction de distribution et qu'elles sont indépendantes...*

Remarque Le lecteur pourra trouver la fonction de distribution de Weibull (ou *loi de Weibull*) dans le chapitre traitant du Génie Industriel (section sur l'Ingénierie).

De telles fonctions étant très nombreuses dans la nature, nous proposons au lecteur une étude détaillée des plus connues seulement.

7.4.1 Loi discrète uniforme

Si nous admettons qu'il est possible d'associer une probabilité à un événement, nous pouvons concevoir des situations où nous pouvons supposer *a priori* que tous les événements élémentaires sont équiprobables (c'est-à-dire qu'ils ont même probabilité). Nous utilisons alors le rapport entre le nombre de cas favorables et le nombre de cas possibles pour calculer la probabilité de tous les événements de l'Univers des événements U. Plus généralement si U est un ensemble fini d'événements équiprobables et A une partie de U nous avons sous forme ensembliste :

$$P(A) = \text{Card}(A)/\text{Card}(U) \tag{7.120}$$

Plus communément, soit e un événement pouvant avoir N issues équiprobables possibles. Alors la probabilité d'observer l'issue donnée de l'événement suit une loi discrète uniforme donnée par la relation :

$$P_e = \frac{1}{N} \tag{7.121}$$

Ayant pour espérance (ou moyenne) :

$$\mu = E(X) = \sum_i p_i x_i = P_e \sum_i x_i \tag{7.122}$$

Si nous nous mettons dans le cas particulier où $x_i = i$ avec $i = 1, \ldots, N$, nous avons alors (cf. chapitre de Suites et Séries) :

$$E(X) = \frac{1}{N} \sum_{i=1}^{N} i = \frac{N+1}{2} \tag{7.123}$$

Et pour variance :

$$\begin{aligned} V(X) &= \sum_{i=1}^{N} \frac{1}{N}(i-\mu)^2 = \frac{1}{N} \sum_{i=1}^{N} (i-\mu)^2 = \frac{1}{N} \left(\sum_{i=1}^{N} i^2 - 2\mu \sum_{i=1}^{N} i + \sum_{i=1}^{N} i\mu^2 \right) \\ &= \frac{1}{N} \left(\frac{N(N+1)(2N+1)}{6} \right) - \frac{N+1}{N} \sum_{i=1}^{N} i + \left(\frac{N+1}{2} \right)^2 \\ &= \frac{1}{N} \left(\frac{N(N+1)(2N+1)}{6} \right) - \frac{N+1}{N} \frac{N(N+1)}{2} + \left(\frac{N+1}{2} \right)^2 = \frac{N^2-1}{12} \end{aligned} \tag{7.124}$$

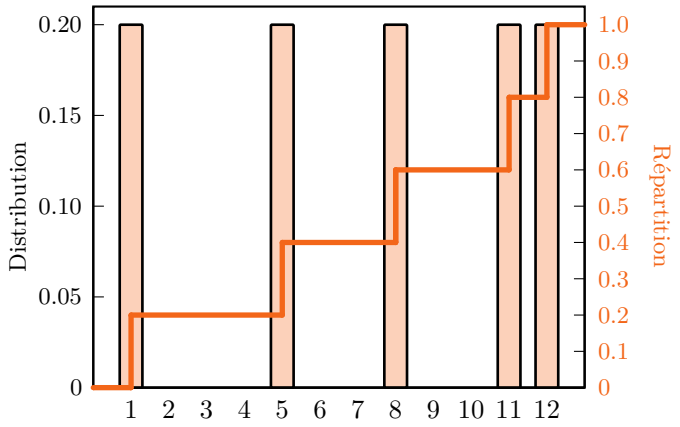


FIGURE 7.5 – Loi discrète uniforme de paramètres $1, 5, 8, 11, 12$: chaque valeur a une probabilité équiprobable.

7.4.2 Loi de BERNOULLI

Si nous avons affaire à une observation binaire alors la probabilité d'un événement reste constant d'une observation à l'autre s'il n'y a pas d'effet mémoire[4].

Nous appelons ce genre d'observations où la variable aléatoire à valeurs 0 ou 1, avec probabilité $(1-p)$, p respectivement, des essais de BERNOULLI.

Ainsi, une variable aléatoire X suit une loi de BERNOULLI si elle ne peut prendre que les valeurs 0 ou 1, associées aux probabilités q et p de sorte que $q+p=1$ et :

$$P(X=0)=q \qquad P(X=1)=p=1-q \tag{7.125}$$

L'exemple classique d'un tel processus est le jeu de pile de face ou de tirage avec remise. Il est inutile de vérifier formellement que la probabilité cumulée est unitaire.

Remarquons que par extension, si nous considérons N événements où nous obtenons dans un ordre particulier k fois une des issues possible (réussite) et $N-k$ l'autre (échec), alors la probabilité d'obtenir une telle série (de k réussites et $N-k$ échecs ordonnées dans un ordre particulier) sera donnée par :

$$P=p^k(1-p)^{N-k}=p^kq^{N-k} \tag{7.126}$$

conformément à ce que nous avions vu obtenu en combinatoire dans le chapitre de Probabilités.

La fonction de BERNOULLI a donc pour espérance (moyenne) :

$$\mu=E(X)=\sum p_ix_i=1\cdot\mathrm{p}+0\cdot(1-p)=p \tag{7.127}$$

et pour variance (nous utilisons la formule de HUYGHENS démontrée plus haut) :

$$\begin{aligned} V(X)=\sigma^2&=E[X-E(X)]=E(X^2)-E(X)^2\\ &=E(X^2)-p^2=p-p^2=p(1-p)=pq \end{aligned} \tag{7.128}$$

REMARQUE L'exemple ci-dessus n'est certes par pertinent mais nous verrons dans le chapitre de Techniques De Gestion que la fonction de BERNOULLI apparaît naturellement au début de notre étude des files d'attentes.

7.4.3 Fonction géométrique

La loi géométrique ou *loi de* Pascal consiste dans une épreuve de type BERNOULLI[5] que nous renouvelons de manière indépendante jusqu'au premier succès.

4. Autrement dit, une somme de variables de BERNOULLI, deux à deux indépendantes.
5. On rappelle qu'une épreuve de type BERNOULLI a une probabilité de succès p et d'échec $q=1-p$ constantes.

Si nous appelons X la variable aléatoire donnant le rang du premier succès la probabilité que X=k est alors (cas particulier de la fonction de Bernoulli) :

$$P(k) = q^{k-1}p \quad \text{avec } k \in \mathbb{N}^* \tag{7.129}$$

Cette loi a pour espérance :

$$\mu = E(X) = \sum p_i x_i = \sum_k kq^{k-1}p = \sum_k k(1-p)^{k-1}p = p\sum_k k(1-p)^{k-1} \tag{7.130}$$

Or, cette dernière relation s'écrit aussi (car c'est une simple série géométrique) :

$$\sum_k \mathrm{k}q^{k-1} = \tfrac{1}{(1-q)^2} \tag{7.131}$$

Effectivement, nous avons démontré dans le chapitre sur les Suites et Séries que :

$$\sum_k q^k = \frac{1-q^{k+1}}{1-q} \tag{7.132}$$

En prenant la limite lorsque $k \to +\infty$ nous obtenons :

$$\sum_k q^k = \frac{1}{1-q} \tag{7.133}$$

car $0 \leqslant q < 1$. Ensuite, il suffit de dériver les deux membres de l'égalité par rapport à q et nous obtenons (7.131). Nous avons donc le nombre moyen d'essais X qu'il faut faire pour arriver au premier succès :

$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} kpq^{k-1} = \frac{p}{(1-q)^2} = \frac{1}{p} \tag{7.134}$$

Calculons maintenant la variance en rappelant comme à chaque fois que (formule de Huyghens) :

$$V(X) = E(X^2) - E(X)^2 \tag{7.135}$$

Commençons donc par calculer equation :

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X=k) = p\sum_{k=0}^{\infty} k^2 q^{k-1} = p\sum_{k=0}^{\infty} k(k-1)q^{k-1} + p\sum_{k=0}^{\infty} kq^{k-1} \tag{7.136}$$

Le dernier terme de cette expression est l'équivalent de l'espérance calculée précédemment, soit :

$$p\sum_{k=0}^{\infty} \mathrm{k}q^{k-1} = \tfrac{1}{p} \tag{7.137}$$

Il reste à calculer le terme précédent or en dérivant l'égalité (7.131), nous obtenons :

$$\sum_{k=1}^{\infty} \mathrm{k}(k-1)q^{k-2} = \tfrac{2}{(1-q)^3} \tag{7.138}$$

Par conséquent :

$$p\sum_{k=1}^{\infty} k(k-1)q^{k-1} = pq\sum_{k=2}^{\infty} k(k-1)q^{k-2} = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2} \tag{7.139}$$

donc :

$$E(X^2) = \frac{2q}{p^2} + \frac{1}{p} \tag{7.140}$$

Pour finir :

$$V(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2} \tag{7.141}$$

Exemple 7.4 *Vous essayez, tard dans la nuit et dans l'obscurité, d'ouvrir une serrure au moyen d'un trousseau de cinq clés, sans porter attention, car vous êtes un peu fatigué, vous essayez chaque clé. Sachant qu'une seule convient, quelle est la probabilité d'utiliser la bonne clé au k^e essai ?*

$$P(k) = \left(\frac{4}{5}\right)^{k-1}\left(\frac{1}{5}\right) \tag{7.142}$$

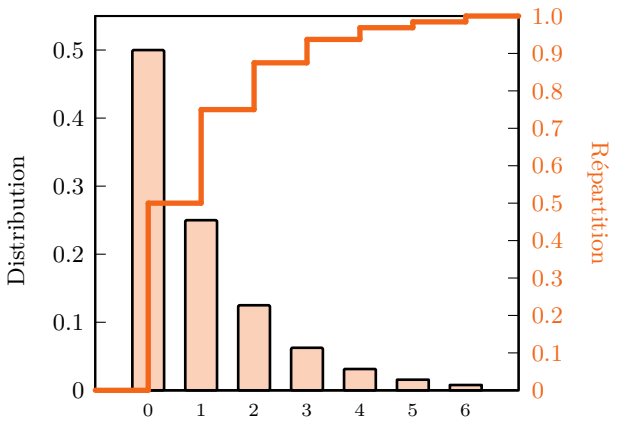


FIGURE 7.6 – Loi discrète géométrique pour $p = 0{,}5$

Déterminons maintenant la fonction de répartition de la loi géométrique en partant de :

$$P(k) = q^{k-1}p \tag{7.143}$$

Par définition, la probabilité que l'expérience réussisse dans les n premiers essais s'écrit :

$$P(X \leqslant n) = 1 - \sum_{j=n+1}^{+\infty} q^{j-1}p = 1 - p\sum_{j=n+1}^{+\infty} q^{j-1} \tag{7.144}$$

avec $n = 0, 1, 2, \ldots$ Posons $j - 1 = n + k$, soit $k = n - j + 1$, nous avons alors :

$$\begin{aligned} P(X \leqslant n) &= 1 - p\sum_{k=0}^{+\infty} q^{n+k} = 1 - p\sum_{k=0}^{+\infty} q^n q^k = 1 - pq^n\sum_{k=0}^{+\infty} q^k \\ &= 1 - pq^n\frac{1}{1-q} = 1 - (1-q)q^n\frac{1}{1-q} = 1 - q^n \end{aligned} \tag{7.145}$$

7.4.4 Fonction binomiale

Si nous revenons maintenant à notre épreuve de BERNOULLI. Plus généralement, tout N-uplet particulier formé de k succès et de $N - k$ échecs aura pour probabilité [6] :

$$P(N, k) = p^k q^{N-k} \tag{7.146}$$

d'être tiré (ou d'apparaître) quel que soit l'ordre d'apparition des échecs et réussites.

Mais, nous savons que la combinatoire permet de déterminer le nombre de N-uplets de ce type [7].
Chap6 Le nombre d'arrangements possibles étant donné par la binomiale• :

$$C_k^N = \frac{N!}{k!(N-k)!} \tag{7.147}$$

Puisque la probabilité d'obtenir une série de k succès et $N - k$ échecs particuliers est toujours identique, il suffit de multiplier la probabilité d'une série particulière par la combinatoire :

$$P(N, k) = C_k^N \cdot P = \frac{N!}{k!(N-k)!}p^k q^{N-k} \tag{7.148}$$

pour avoir la probabilité totale d'obtenir une quelconque de ces séries possibles.

6. Dans le cadre d'un tirage avec remise ou sans remise si la population est grande en première approximation
7. Il s'agit en fait du nombre de manières d'ordonner les apparitions d'échecs et de réussites.

REMARQUE Cela équivaut à l'étude d'un tirage avec remise• simple avec contrainte sur l'ordre ou à l'étude d'une série de succès ou d'échecs. Nous utiliserons cette relation dans le cadre de la théorie des files d'attentes ou en fiabilité. Il faut noter que dans le cas de grandes populations, même si le tirage n'est pas avec remise il est considéré comme tel. Chap6

Écrite autrement ceci donne la *fonction Binomiale* connue aussi sous la forme de la fonction de distribution suivante :

$$P(N,k) = C_k^N p^k (1-p)^{N-k} = C_k^N p^k q^{N-k} \tag{7.149}$$

et parfois notée $X \to \beta(n,p)$. Nous disons aussi que la loi Binomiale est non exhaustive car la taille de la population initiale n'est pas apparente dans l'expression de la loi. Nous avons bien évidemment pour la fonction de répartition (très utile dans la pratique comme le contrôle de lots de fournisseurs ou la fiabilité) :

$$\sum_{k=0}^{N} C_k^N p^k (1-p)^{N-k} = (p + (1-p))^N = 1 \tag{7.150}$$

puisque d'après le chapitre de Calcul Algébrique :

$$(x+y)^n = \sum_{k=0}^{N} C_k^N x^k y^{N-k} \tag{7.151}$$

L'espérance mathématique (moyenne) de $P(N,k)$ est :

$$E(X) = \sum_{k=0}^{N} P(X=k)k = \sum_{k=0}^{N} C_k^N p^k (1-p)^{N-k} k = \sum_{k=1}^{N} C_k^N p^k (1-p)^{N-k} k \tag{7.152}$$

Or :

$$C_k^N = \frac{N}{k} C_{k-1}^{N-1} \Leftrightarrow k C_k^N = N C_{k-1}^{N-1} \tag{7.153}$$

d'où :

$$\begin{aligned} E(X) &= N \sum_{k=1}^{N} C_{k-1}^{N-1} p^k (1-p)^{N-k} = Np \sum_{k=1}^{N} C_{k-1}^{N-1} p^{k-1} (1-p)^{(N-1)-(k-1)} \\ &= Np \sum_{k=0}^{N-1} C_k^{N-1} p^k (1-p)^{N-1-k} = Np(p + (1-p))^{N-1} = Np \end{aligned} \tag{7.154}$$

donne le nombre moyen de fois que l'on obtiendra l'issue souhaitée de probabilité p après N essais.

Avant de calculer la variance, introduisons la relation suivante :

$$\sum_{s=0}^{N-1} s \frac{(N-1)!}{s!(N-1-s)!} p^s q^{N-1-s} = (N-1)p \tag{7.155}$$

En effet, en utilisant les développements précédents :

$$\begin{aligned} (7.155) &= \sum_{s=1}^{N-1} \frac{(N-1)!}{(s-1)!(N-1-s)!} p^s q^{N-1-s} = (N-1) \sum_{s=1}^{N-1} \frac{(N-2)!}{(s-1)!(N-1-s)!} p^s q^{N-1-s} \\ &= (N-1) \sum_{s=0}^{N-2} \frac{(N-2)!}{s!(N-2-s)!} p^{s+1} q^{N-2-s} = (N-1)p \sum_{s=0}^{N-2} C_s^{N-2} p^s q^{N-2-s} \\ &= (N-1)p(p+q)^{N-2} = (N-1)p \end{aligned} \tag{7.156}$$

puisque $p+q=1$. Commençons maintenant le (long) calcul de la variance dans lequel nous allons utiliser les résultats précédents :

$$\begin{aligned}
V(X) &= \sum_{k=0}^{N} \frac{N!}{k!(N-k)!} p^k q^{N-k} (k - N\cdot p)^2 = \sum_{k=0}^{N} \frac{N!}{k!(N-k)!} p^k q^{N-k} (k-\mu)^2 \\
&= \sum_{k=0}^{N} k^2 \frac{N!}{k!(N-k)!} p^k q^{N-k} + \mu^2 \sum_{k=0}^{N} \frac{N!}{k!(N-k)!} p^k q^{N-k} - 2\mu \sum_{k=0}^{N} k \frac{N!}{k!(N-k)!} p^k q^{N-k} \\
&= \sum_{k=0}^{N} k^2 \frac{N!}{k!(N-k)!} p^k q^{N-k} + \mu^2 \underbrace{\sum_{k=0}^{N} C_k^N p^k q^{N-k}}_{=1} - 2\mu \sum_{k=0}^{N} k C_k^N p^k q^{N-k} \\
&= \sum_{k=0}^{N} k^2 \frac{N!}{k!(N-k)!} p^k q^{N-k} + \mu^2 - 2\mu N \underbrace{\sum_{k=0}^{N} k C_{k-1}^{N-1} p^k q^{N-k}}_{=Np=\mu} \\
&= \sum_{k=0}^{N} k^2 \frac{N!}{k!(N-k)!} p^k q^{N-k} - \mu^2 = -\mu^2 + N \sum_{k=1}^{N} k^2 \frac{(N-1)!}{(k-1)!(N-k)!} p^k q^{N-k} \\
&\underset{s=k-1}{=} -\mu^2 + N \sum_{s=0}^{N-1} (s+1) \frac{(N-1)!}{s!(N-1-s)!} p^{s+1} q^{N-s-1} \\
&= -\mu^2 + Np \sum_{s=0}^{N-1} (s+1) \frac{(N-1)!}{s!(N-1-s)!} p^s q^{N-s-1}
\end{aligned} \tag{7.157}$$

$$\begin{aligned}
&= -\mu^2 + Np \sum_{s=0}^{N-1} s \frac{(N-1)!}{s!(N-1-s)!} p^s q^{N-s-1} + Np \sum_{s=0}^{N-1} \frac{(N-1)!}{s!(N-1-s)!} p^s q^{N-s-1} \\
&= -\mu^2 + Np \sum_{s=0}^{N-1} s C_s^{N-1} p^s q^{N-s-1} + Np \sum_{s=0}^{N-1} C_s^{N-1} p^s q^{N-s-1} \\
&= -\mu^2 + Np(N-1)p + Np(p+(1-p))^{N-1} = -\mu^2 + Np^2(N-1) + Np \\
&= -\mu^2 + N^2p^2 - Np^2 + Np = -\mu^2 + \mu^2 - Np^2 + Np = Np(1-p) = Npq
\end{aligned}$$

L'écart-type étant $\sigma^2 = V(k)$, nous avons :

$$\sigma = \pm\sqrt{Np(1-p}} \tag{7.158}$$

7.4.5 Fonction hypergéométrique

Nous considérons pour approche à cette fonction un exemple simple concernant une urne contenant n boules dont m sont noires et les autres m' blanches (pour un exemple concret utilisé dans l'industrie se reporter au chapitre de Génie Industriel). Nous tirons successivement, et sans les remettre dans l'urne, p boules. Quelle est la probabilité que parmi ces p boules, il y en ait k qui soient noires (dans cet énoncé l'ordre du tirage ne nous intéresse donc pas.).

Nous parlons souvent de *tirage exhaustif* avec la loi hypergéométrique car contrairement à la loi binomiale, la taille du lot qui sert de base au tirage va apparaître dans la loi.

Chap6 REMARQUE Cela équivaut à l'étude non ordonnée d'un tirage sans remise• avec contrainte sur les occurrences appelé parfois *tirage simultané*. Nous utiliserons cette relation souvent dans le domaine de la qualité ou de la fiabilité ou les boules noires sont associées à des éléments avec défauts et les blanches à des éléments sans défauts.

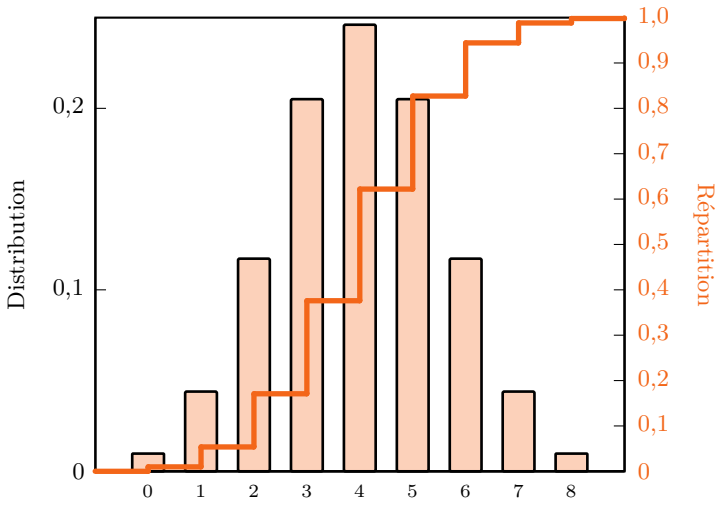


FIGURE 7.7 – Loi discrète binomiale $\beta(10; 0{,}5)$

Les p boules peuvent être choisies parmi les n boules de C_n^p façons (représentant donc le nombre de tirages différents possibles) avec pour rappel[•] : Chap6

$$C_n^p = \binom{n}{p} = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!} \tag{7.159}$$

Les k boules noires peuvent être choisies parmi les m noires de C_k^m façons. Les $p-k$ boules blanches peuvent être elles choisies de C_{p-k}^{n-m} façons. Il y a donc $C_k^m C_{p-k}^{n-m}$ tirages qui donnent k boules noires et $p-k$ boules blanches.

La probabilité recherchée vaut donc :

$$P(n,p,m,k) = H(k)\frac{C_k^m C_{p-k}^{n-m}}{C_n^p} \tag{7.160}$$

Il n'est pas interdit de faire le calcul direct de l'espérance et de la variance la fonction hypergéométrique mais le lecteur pourra sans trop de peine imaginer que ce calcul va être relativement indigeste. Alors nous pouvons utiliser une méthode indirecte qui de plus est intéressante.

D'abord le lecteur aura peut-être remarqué qu'au fait l'expérience de la loi hypergéométrique est une série d'essais de BERNOULLI.

Alors, nous allons tricher en utilisant dans un premier temps la propriété de linéarité de l'espérance. Définissons pour cela une nouvelle variable correspondant implicitement au fait à l'expérience da la fonction hypergéométrique, soit k essais de BERNOULLI de suite :

$$X = \sum_{i=1}^{k} X_i \tag{7.161}$$

où X_i représente la réussite d'obtenir au i-ème tirage une boule noire (soit 0 ou 1). Or, nous savons que pour tout i la variable aléatoire X_i suit une fonction de BERNOULLI pour laquelle nous avons démontré lors de notre étude de la loi de BERNOULLI que $E(X_i) = p$. Dès lors, de par la propriété de linéarité de l'espérance nous avons :

$$E(X) = E\left(\sum_{i=1}^{k} X_i\right) = k{\cdot}\mathrm{p} \tag{7.162}$$

Or, dans l'essai de BERNOULLI, p est la probabilité d'obtenir l'élément recherché. Dans la loi hypergéométrique, ce qui nous intéresse est la probabilité d'avoir une boule noire par rapport à la quantité totale de boules n. Le rapport nous donne évidemment cette probabilité. Ainsi, nous avons :

$$\mu = E(X) = k \cdot \frac{m}{m'+m} = k \cdot \frac{m}{n} \tag{7.163}$$

où k est le nombre de tirages. Cette moyenne donne donc le nombre moyen de boules noires lors d'un tirage de k boules.

Pour déterminer la variance, nous allons utiliser la variance de la fonction de BERNOULLI la propriété 7.9, les équations (7.76) et (7.161) et il vient :

$$V\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} V(X_i) + 2 \sum_{1 \leqslant i < \leqslant n}^{k} \left(E(X_i \cdot X_j) - E(X_i)E(X_j)\right) \tag{7.164}$$

Or, pour la loi de BERNOULLI, nous avons :

$$V(X) = p \cdot q = \frac{m}{m'+m}\frac{m'}{m'+m} = \frac{m \cdot m'}{n^2} \tag{7.165}$$

Alors nous avons déjà :

$$\sum_{i=1}^{k} V(X_i) = k\frac{m \cdot m'}{(m'+m)^2} = k\frac{m \cdot m'}{n^2} \tag{7.166}$$

Ensuite, nous avons facilement :

$$E(X_i)E(X_j) = \frac{m}{m'+m}\frac{m}{m'+m} = \left(\frac{m}{m+m'}\right)^2 \tag{7.167}$$

Le calcul de $E(X_iX_j)$ nécessite une bonne compréhension des probabilités.

L'espérance $E(X_iX_j)$ est donnée (implicitement) par la somme pondérée des probabilités que deux événements aient lieu en même temps comme nous le savons. Or, nos événements sont binaires : soit c'est une boule noire (1) soit c'est une boule blanche (0). Donc tous les termes de la somme n'ayant pas deux boules noirs consécutivement seront nuls.

Le problème est alors de calculer la probabilité d'avoir deux boules noires consécutives et celle-ci s'écrit donc :

$$\begin{aligned} P(X_iX_j = 1) &= P\big((X_i = 1) \cap (X_j = 1)\big) = P(X_i = 1)P_{X_i=1}(X_j = 1) \\ &= \frac{m}{m'+m}\frac{m-1}{m+m'-1} \end{aligned} \tag{7.168}$$

Donc nous avons finalement :

$$E(X_iX_j) = \frac{m}{m'+m}\frac{m-1}{m+m'-1} \tag{7.169}$$

Soit :

$$\operatorname{cov}(X_i, X_j) = \frac{m}{m'+m}\frac{m-1}{m+m'-1} - \left(\frac{m}{m+m'}\right)^2 = -\frac{mm'}{(m+m')^2(m+m'-1)} \tag{7.170}$$

Finalement :

$$\begin{aligned} V(X) &= \sum_{i=1}^{k} V(X_i) + \sum_{1 \leqslant i < j \leqslant n}^{k} \big(E(X_iX_j) - E(X_i)E(X_j)\big) \\ &= k\frac{mm'}{(m'+m)^2} - \frac{2mm'}{(m+m')^2(m+m'-1)} \sum_{1 \leqslant i < j \leqslant n}^{k} 1 \\ &= k\frac{mm'}{(m'+m)^2} - \frac{2mm'}{(m+m')^2(m+m'-1)}\frac{(k-1)k}{2} = kpq\frac{m+m'-k}{m+m'-1} \end{aligned} \tag{7.171}$$

où nous avons utilisé le fait que :

$$\sum_{i<j}^{n} \operatorname{cov}(X_i, X_j) \tag{7.172}$$

est composé de :

$$C_k^2 = \binom{\mathrm{k}}{2} = \frac{(k-1)k}{2} \tag{7.173}$$

terme puisqu'il correspond au nombre de façons qu'il y a de choisir le couple (i, j) avec $i < j$. Donc finalement :

$$V(X) = kpq\frac{n-k}{n-1} \tag{7.174}$$

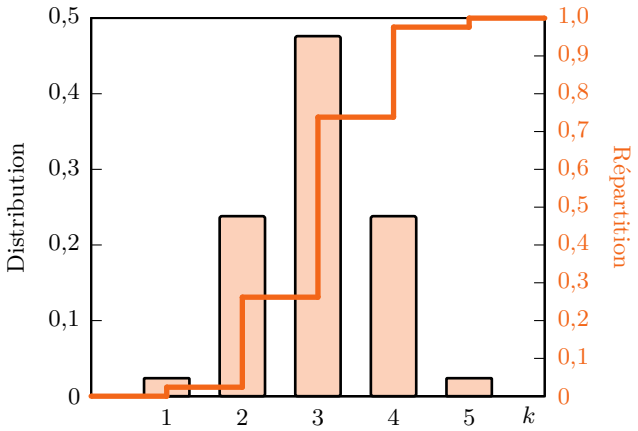


FIGURE 7.8 – Loi discrète hypergéométrique pour $(n, p, m) = (10, 6, 5)$

7.4.6 Fonction multinomiale

La loi binomiale concerne le nombre de succès dans N épreuves de BERNOULLI indépendantes donnant chacune un résultat binaire, comme dans le jeu de pile ou face. La loi multinomiale est une généralisation de celle-ci, applicable par exemple à N jets d'un dé à six faces. Contrairement à ces exemples simples, les différentes possibilités ne sont en plus généralement pas équiprobables.

Soit l'espace des événements $\Omega = \{1, 2, \ldots, m\}$ muni d'une probabilité $P(\{i\}) = p_i$, $i = 1, 2, \ldots, n$. Nous tirons n fois de suite avec remise un élément de Ω avec la probabilité p_i, $i = 1, 2, \ldots, n$. Quelle est la probabilité d'obtenir le nombre 1, k_1 fois le nombre 2, k_2 fois, sur une suite d'un tirage de n éléments.

REMARQUE Cela équivaut à l'étude d'un tirage avec remise (cf. chapitre de Probabilités) avec contraintes sur les occurrences. Donc sans contraintes nous verrons par l'exemple que nous retombons sur un tirage avec remise simple.

Nous avons vu• que si nous prenons un ensemble d'événements ayant plusieurs issues, alors les différentes combinaisons de suites que nous pouvons obtenir en prenant p éléments choisis parmi n est donnée C_p^n [équation (7.159)]. Il y a donc : Chap6

$$C_{k_1}^n = \frac{n!}{k_1!(n-k_1)!} \tag{7.175}$$

façons différentes d'obtenir k_1 fois un certain événement. Ensuite, en associant ce que nous avons vu pour la loi binomiale il vient :

$$C_{k_2}^{n-k_1} = \frac{(n-k_1)!}{k_2!((n-k_1)-k_2)!} \tag{7.176}$$

façons différentes d'obtenir k_2 un second événement puisque dans l'ensemble de la suite, de n éléments déjà k_1 on été tirés ce qui fait qu'il n'en reste plus $(n-k_1)!$ sur lesquels nous pouvons obtenir les k_2 voulus.

Nous avons alors selon la loi binomiale :

$$\begin{aligned} P(n,k_1) &= C_{k_1}^{n} p_1^{k_1} q_1^{n-k_2-k_1} = C_{k_1}^{n} p_1^{k_1} (1-p_1)^{n-k_2-k_1} \\ P(n-k_1,k_2) &= C_{k_2}^{n-k_1} p_2^{k_2} q_2^{n-k_2-k_1} = C_{k_2}^{n-k_1} p_2^{k_2} (1-p_2)^{n-k_2-k_1} \end{aligned} \tag{7.177}$$

Alors nous avons dans le cas particulier de deux séries d'uplets :

$$\begin{aligned} P(n-k_1,k_2) \cap P(n,k_1) &= C_{k_1}^{n} p_1^{k_1} (1-p_1)^{n-k_2-k_1} C_{k_2}^{n-k_1} p_2^{k_2} (1-p_2)^{n-k_2-k_1} \\ &= C_{k_1}^{n} C_{k_2}^{n-k_1} p_1^{k_1} (1-p_1)^{n-k_2-k_1} p_2^{k_2} (1-p_2)^{n-k_2-k_1} \\ &= \frac{n!}{k_1!(n-k_1)!} \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} p_1^{k_1} (1-p_1)^{n-k_2-k_1} p_2^{k_2} (1-p_2)^{n-k_2-k_1} \\ &= \frac{n!}{k_1!k_2!} \frac{1}{(n-k_1-k_2)!} p_1^{k_1} p_2^{k_2} (1-p_1)^{n-k_2-k_1} (1-p_2)^{n-k_2-k_1} \end{aligned} \tag{7.178}$$

et comme $-k_1 - k_2 = -n$, il vient :

$$P = \frac{n!}{k_1!k_2!} \frac{1}{0!} p_1^{k_1} p_2^{k_2} (1-p_1)^0 (1-p_2)^0 = \frac{n!}{k_1!k_2!} p_1^{k_1} p_2^{k_2} \tag{7.179}$$

Ainsi, par récurrence nous avons la probabilité probabilité P recherchée appelée loi Multinomiale et donnée par :

$$P = \frac{n!}{\prod_{i=1}^{m} k_i!} \prod_{i=1}^{m} p_i^{k_i} = \frac{\left(\sum_{i=1}^{m} k_i\right)!}{\prod_{i=1}^{m} k_i!} \prod_{i=1}^{m} p_i^{k_i} \tag{7.180}$$

Exemple 7.5 *Nous lançons un dé non-pipé 12 fois. Quelle est la probabilité que les six faces apparaissent le même nombre de fois (mais pas nécessairement consécutivement) :*

$$P = \frac{n!}{\prod_{i=1}^{m} k_i!} \prod_{i=1}^{m} p_i^{k_i} = \frac{12!}{\prod_{i=1}^{6} 2!} \prod_{i=1}^{6} \left(\frac{1}{6}\right)^2 = \frac{12!}{2^6 6^{12}} = 0{,}34\% \tag{7.181}$$

où nous voyons bien que m correspond au nombre de groupes de réussites.

Exemple 7.6 *Nous lançons un dé non-pipé 12 fois. Quelle est la probabilité qu'une seule et unique face apparaisse 12 fois :*

$$P = \frac{12!}{\prod_{i=1}^{1} 12!} \prod_{i=1}^{1} \left(\frac{1}{6}\right)^{12} = \frac{1}{6^{12}} \tag{7.182}$$

Nous retrouvons donc avec ce dernier exemple un résultat connu de la binomiale.

7.4.7 Fonction de POISSON

Pour certains événements forts rares, la probabilité p est très faible et tend vers zéro. Toutefois la valeur moyenne $n \cdot p$ tend vers une valeur fixe lorsque n tend vers l'infini.

Nous partirons donc d'une distribution binomiale de moyenne $\mu = n \cdot p$ que nous supposerons finie lorsque n tend vers l'infini.

La probabilité de k réussites lors de n épreuves vaut (loi Binomiale) :

$$P_n(k) = \binom{n}{p} p^k q^{n-k} \tag{7.183}$$

En posant $p = m/n$ (où m est temporairement la nouvelle notation pour la moyenne selon $\mu = n{\cdot}p$), cette expression peut s'écrire :

$$\frac{n!}{k!(n-k)!}\left(\frac{m}{k}\right)^k\left(1-\frac{m}{n}\right)^{n-k} \tag{7.184}$$

En regroupant les termes, nous pouvons mettre la valeur de $P_n(k)$ sous la forme :

$$\frac{m^k}{k!}\left(1-\frac{m}{n}\right)^n\frac{n!}{(n-k)!n^k\left(1-\frac{m}{n}\right)^k} \tag{7.185}$$

Nous reconnaissons que, lorsque n tend vers l'infini, le deuxième facteur du produit a pour limite equation.

Quant au troisième facteur, puisque nous nous intéressons aux petites valeurs de k —la probabilité de réussite est très faible— sa limite pour n tendant vers l'infini vaut 1. Ces techniques de passage à la limite sont parfois appelées *théorème limite de* Poisson.

Nous obtenons ainsi la *loi de* Poisson :

$$P_k = \frac{\mu^k}{k!}\mathrm{e}^{-\mu} \tag{7.186}$$

Il s'agit bien d'une loi de probabilité puisque en utilisant les séries de Taylor, nous montrons que :

$$\sum_{k=0}^{+\infty}\frac{\mu^k}{k!}\mathrm{e}^{-\mu} = \mathrm{e}^{-\mu}\sum_{k=0}^{+\infty}\frac{\mu^k}{k!} = \mathrm{e}^{-\mu}\mathrm{e}^{\mu} = 1 \tag{7.187}$$

Remarque Nous retrouverons fréquemment cette loi dans différents chapitres du site comme par exemple lors de l'étude du Génie Industriel en maintenance préventive ou encore dans le même chapitre lors de l'étude des théories des files d'attentes (le lecteur peut s'y reporter pour un exemple intéressant et pragmatique).

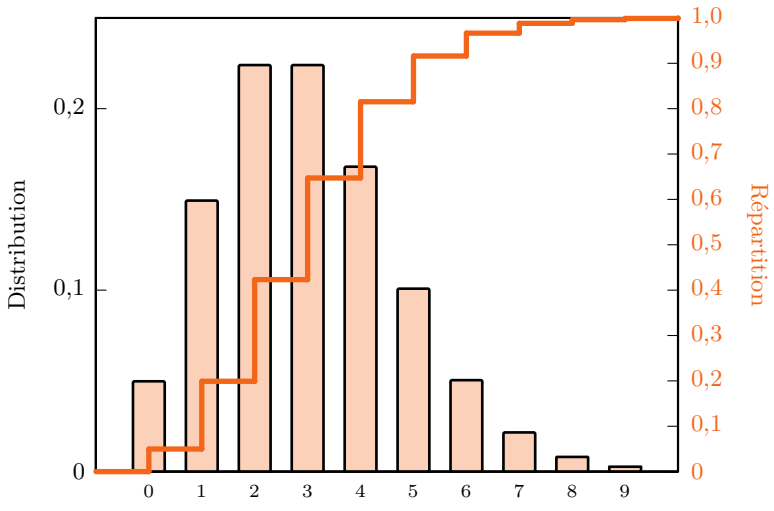


Figure 7.9 – Loi discrète de Poisson pour $\mu = 3$

Cette distribution est importante car elle décrit beaucoup de processus dont la probabilité est petite et constante. Elle est souvent utilisée dans la théorie du temps d'attente, test d'acceptabilité et fiabilité, et contrôles statistiques de qualité. Entre autres, elle s'applique aux processus tels que l'émission des quanta de lumière par des atomes excités, le nombre de globules rouges observés au microscope, le nombre d'appels arrivant à une centrale téléphonique. La distribution de Poisson est valable pour de nombreuses observations faites en physique nucléaire ou corpusculaire.

L'espérance moyenne de la fonction de POISSON est écrite par développement de TAYLOR de la fonction l'exponentielle :

$$\mu = E(k) = \sum_{k=0}^{+\infty} kP_k = \sum_{k=0}^{+\infty} k\frac{\mu^k}{k!}\mathrm{e}^{-\mu} = \mu\sum_{k=1}^{+\infty}\frac{\mu^{k-1}}{(k-1)!}\mathrm{e}^{-\mu} = \mathrm{e}^{-\mu}\mu\mathrm{e}^{\mu} = \mu \tag{7.188}$$

et donne le nombre moyen de fois que l'on obtiendra l'issue souhaitée. Ce résultat peut paraître déroutant puisque la moyenne s'exprime par la moyenne. Il ne faut simplement pas oublier que celle-ci est donnée au début par $\mu = np$.

Sec7.5|p198 REMARQUE Pour plus de détails le lecteur peut aussi se reporter à la partie concernant les *estimateurs*• dans le présent chapitre.

La variance de la fonction de distribution de POISSON est elle donnée par (en utilisant à nouveau les séries de TAYLOR) :

$$\begin{aligned}
\mu = V(k) &= \sum_{k=0}^{+\infty}(k_\mu)^2 P_k = \sum_{k=0}^{+\infty}(k-\mu)^2\frac{\mu^k}{k!}\mathrm{e}^{-\mu} \\
&= \sum_{k=0}^{+\infty}k^2\frac{\mu^k}{k!}\mathrm{e}^{-\mu} - 2\mu\sum_{k=0}^{+\infty}k\frac{\mu^k}{k!}\mathrm{e}^{-\mu} + \mu^2\sum_{k=0}^{+\infty}\frac{\mu^k}{k!}\mathrm{e}^{-\mu} \\
&= \sum_{k=0}^{+\infty}k^2\frac{\mu^k}{k!}\mathrm{e}^{-\mu} - 2\mu\mu\sum_{k=0}^{+\infty}\frac{k\mu^{k-1}}{k(k-1)!}\mathrm{e}^{-\mu} + \mu^2\sum_{k=0}^{+\infty}\frac{\mu^k}{k!}\mathrm{e}^{-\mu} \\
&= \sum_{k=0}^{+\infty}k^2\frac{\mu^k}{k!}\mathrm{e}^{-\mu} - 2\mu^2\sum_{k=0}^{+\infty}\frac{\mu^{k-1}}{(k-1)!}\mathrm{e}^{-\mu} + \mu^2 = \sum_{k=0}^{+\infty}k^2\frac{\mu^k}{k!}\mathrm{e}^{-\mu} - 2\mu^2 + \mu^2 \\
&= \mu\mathrm{e}^{-\mu}\sum_{k=1}^{+\infty}k\frac{\mu^{k-1}}{(k-1)!} - \mu^2 \underset{s=k-1}{=} \mu\mathrm{e}^{-\mu}\sum_{s=0}^{+\infty}(s+1)\frac{\mu^s}{s!} - \mu^2 \\
&= \mu\mathrm{e}^{-\mu}\left(\sum_{s=0}^{+\infty}s\frac{\mu^s}{s!} + \sum_{s=0}^{+\infty}\frac{\mu^s}{s!}\right) - \mu^2 = \mu\mathrm{e}^{-\mu}\left(\sum_{s=1}^{+\infty}\frac{\mu^{s-1}}{(s-1)!} + \mathrm{e}^{\mu}\right) - \mu^2 \\
&= \mu\mathrm{e}^{-\mu}\left(\mu + \mathrm{e}^{\mu}\right) - \mu^2 = \mu
\end{aligned} \tag{7.189}$$

toujours avec $\mu = np$. Les lois théoriques de distribution statistiques sont établies en supposant la réalisation d'un nombre infini de mesures. Il est évident que nous ne pouvons en effectuer qu'un nombre fini N. D'où la nécessité d'établir des correspondances entre les valeurs utiles théoriques et expérimentales. Pour ces dernières nous n'obtenons évidemment qu'une approximation dont la validité est toutefois souvent admise comme suffisante.

7.4.8 Loi de GAUSS-LAPLACE

Cette loi est la plus importante fonction de distribution en statistiques suite au résultat d'un théorème connu appelé *théorème central limite* qui comme nous le verrons, permet de démontrer, entre autres, que toute suite de variables aléatoires indépendantes de même loi ayant une espérance et un écart-type fini et non nécessairement égaux converge vers une loi de GAUSS-LAPLACE.

REMARQUE Cette loi est aussi appelée Loi Normale.

Il est donc très important de focaliser particulièrement sont attention sur les développements qui vont être faits ici.

Partons d'une fonction Binomiale et faisons tendre le nombre n d'épreuves vers l'infini. Si p est fixé au départ, la moyenne $\mu = np$ tend également vers l'infini, de plus l'écart-type npq tend également vers l'infini.

REMARQUE Le cas où p varie et tend vers 0 tout en laissant fixe la moyenne $\mu = np$ ayant été étudié lors de la présentation de la fonction de POISSON.

Si nous voulons calculer la limite de la fonction Binomiale, il s'agira donc de faire un changement d'origine qui stabilise la moyenne, en 0 par exemple, et un changement d'unité qui stabilise l'écart-type, à 1 par exemple.

Voyons tout d'abord comment varie $P_n(k)$ en fonction de k (nombre de réussites) et calculons la différence :

$$\begin{aligned} P_n(k+1) - P_n(k) &= \binom{n}{k+1} p^{k+1} q^{n-k-1} - \binom{n}{k} p^k q^{n-k} \\ &= \binom{n}{k} p^k q^{n-k} \left(\frac{(n-k)p}{(k+1)q} - 1 \right) \\ &= P_n(k) \frac{np - k - q}{(k+1)q} \end{aligned} \tag{7.190}$$

Nous en concluons que $P_n(k)$ est une fonction croissante de k, tant que $np - k - q$ est positif (pour n, p et q fixés). Pour le voir, il suffit de prendre quelques valeurs ou d'observer la distribution graphique de la fonction Binomiale en se souvenant bien que $\mu = np$. Comme $q < 1$ il est par conséquent évident que la valeur de k voisine de la moyenne μ constitue le maxima de $P_n(k)$.

D'autre part la différence $P_n(k+1) - P_n(k)$ est le taux d'accroissement de la fonction $P_n(k)$. Nous pouvons alors écrire :

$$\frac{\Delta P_n(k)}{\Delta k} = \frac{P_n(k+1) - P_n(k)}{(k+1) - k} \tag{7.191}$$

comme étant la pente de la fonction.

Définissons maintenant une nouvelle variable aléatoire telle que sa moyenne soit nulle et son écart-type unitaire [8]. Nous avons alors :

$$x \frac{k - np}{\sqrt{npq}} \tag{7.192}$$

Nous avons alors aussi avec cette nouvelle variable :

$$\Delta x = \frac{(k+1) - np}{\sqrt{npq}} - \frac{k - np}{\sqrt{npq}} = \frac{((k+1) - np) - (k - np)}{\sqrt{npq}} \tag{7.193}$$

Appelons $F(x)$ l'expression de $P_n(k)$ calculée en fonction de la nouvelle variable de moyenne nulle et d'écart-type unitaire dont nous recherches l'expression quand n tend vers l'infini. Reprenons :

$$\frac{\Delta P_n(k)}{\Delta k} = \frac{np - k - q}{(k+1)q} P_n(k) \tag{7.194}$$

Afin de simplifier l'étude de cette relation quand n tend vers l'infini et k vers l'espérance, multiplions des deux côtés par $npq/\sqrt{npq}$:

$$\frac{\Delta P_n(k)}{\Delta k} \frac{npq}{\sqrt{npq}} = \frac{np - k - q}{(k+1)q} \frac{npq}{\sqrt{npq}} P_n(k) = \frac{(np - k - q)np}{(k+1)\sqrt{npq}} P_n(k) \tag{7.195}$$

Et maintenant récrivons le terme de gauche de la relation (7.195). Il vient :

$$\begin{aligned} \frac{\Delta P_n(k)}{\Delta k} \frac{npq}{\sqrt{npq}} &= \frac{\Delta P_n(k)}{(k+1) - k} \frac{npq}{\sqrt{npq}} = \frac{\Delta P_n(k)}{((k+1) - np - (k - np))} \frac{npq}{\sqrt{npq}\sqrt{npq}} \\ &= \frac{\Delta P_n(k)}{((k+1) - np - (k - np))} = \frac{\Delta P_n(k)}{\Delta x} \end{aligned} \tag{7.196}$$

puisque $(k+1) - np - (k - np) = \sqrt{npq}$. Après un passage à la limite pour n tendant vers l'infini nous avons dans un premier temps pour le terme (7.195) :

$$(k+1)\sqrt{npq} \underset{n \to +\infty}{=} k\sqrt{npq} \tag{7.197}$$

Donc :

$$-\frac{((k - np) + q)np}{k\sqrt{npq}} P_n(k) = -\frac{k - np}{k\sqrt{npq}} np P_n(k) - \frac{qnp}{k\sqrt{npq}} P_n(k) \tag{7.198}$$

8. Il s'agit donc d'une variable centrée-réduite.

et dans un second temps, tenant compte du fait que les valeurs de k considérées se trouvent alors au voisinage de l'espérance np, nous obtenons :

$$\frac{k-np}{k\sqrt{npq}}np = \frac{np}{k}\frac{k-np}{k\sqrt{npq}} \underset{n\to+\infty}{\simeq} \frac{k-np}{\sqrt{npq}} = x \tag{7.199}$$

et :

$$\frac{qnp}{k\sqrt{npq}} \simeq \frac{qnp}{np\sqrt{npq}} = \frac{q}{\sqrt{npq}} \underset{n\to+\infty}{\simeq} 0 \tag{7.200}$$

Donc :

$$\frac{-((k-np)+q)np}{k\sqrt{npq}} \underset{n\to+\infty}{\simeq} \mathrm{x} \tag{7.201}$$

et comme $P_n(k) \underset{n\to+\infty}{=} F(x)$, nous avons finalement :

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = -xF(x) \tag{7.202}$$

Cette relation peut encore s'écrire en réarrangeant les termes :

$$\frac{\mathrm{d}F(x)}{F(x)} = -x\mathrm{dx} \tag{7.203}$$

et en intégrant les deux membres de cette égalité nous obtenons (cf. chapitre de Calcul Différentiel Et Intégral) :

$$\ln(F(x)) = -\frac{x^2}{2} + k \tag{7.204}$$

où k est une constante, soit :

$$F(x) = \lambda \mathrm{e}^{-\frac{x^2}{2}} \tag{7.205}$$

La constante est déterminée par normalisation, soit :

$$\int_{-\infty}^{+\infty} F(x)\mathrm{dx} = 1 \tag{7.206}$$

autrement dit :

$$\lambda = \frac{1}{\sqrt{2\pi}} \tag{7.207}$$

Preuve Nous avons :

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x = \int_{-\infty}^{+\infty} \mathrm{e}^{-\left(\frac{x}{\sqrt{2}}\right)^2}\mathrm{d}x = \sqrt{2}\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z \tag{7.208}$$

Donc concentrons-nous sur le dernier terme de l'égalité (7.208). Ainsi :

$$I = \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z = 2\int_{0}^{+\infty} \mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z \tag{7.209}$$

puisque $\mathrm{e}^{-\frac{z^2}{2}}$ est une fonction paire (cf. chapitre d'Analyse Fonctionnelle). Écrivons maintenant le carré de l'intégrale de la manière suivante :

$$I^2 = 4\lim_{R\to+\infty}\left[\left(\int_0^R \mathrm{e}^{-x^2}\mathrm{d}x\right)\left(\int_0^R \mathrm{e}^{-y^2}\mathrm{d}y\right)\right] = 4\lim_{R\to+\infty}\int_0^R\int_0^R \mathrm{e}^{-(x^2+y^2)}\mathrm{d}x\mathrm{d}y \tag{7.210}$$

et faisons un changement de variable en passant en coordonnées polaires, dès lors nous faisons aussi usage du Jacobien dans ses mêmes coordonnées (cf. chapitre de Calcul Différentiel Et Intégral) :

$$\begin{aligned} I^2 &= 4 \lim_{R\to+\infty} \left(\iint \mathrm{e}^{-r^2} r \mathrm{d}r\mathrm{d}\phi \right) \\ &= 4 \lim_{R\to+\infty} \int_0^{2\pi} \int_0^R \mathrm{e}^{-r^2} r \mathrm{d}r\mathrm{d}\phi = 2\pi \lim_{R\to+\infty} \int_0^R \mathrm{e}^{-r^2} r \mathrm{d}r = 2\pi \left(0 + \frac{1}{2}\right) = \pi \end{aligned} \tag{7.211}$$

soit $I = \sqrt{\pi}$. Par extension pour le terme $\mathrm{e}^{-\frac{x^2}{2}}$, nous avons $I = \lambda^{-1} = \sqrt{2\pi}$ d'où l'égalité (7.207). •

Nous obtenons donc :

$$F(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \tag{7.212}$$

Pour information, une variable suivant une loi Normale centrée réduite est très souvent par tradition notée Z (pour *Zentriert* en allemand).

En revenant aux variables non normées :

$$x = \frac{k - np}{\sqrt{npq}} = \frac{k - \mu}{\sigma} \tag{7.213}$$

nous obtenons donc la *loi de* Gauss-Laplace également appelée *loi Normale* :

$$P(k, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(k-\mu)^2}{2\sigma^2}} \tag{7.214}$$

La probabilité cumulée étant bien évidemment donnée par :

$$P(X \leqslant x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{k} \mathrm{e}^{-\frac{(k-\mu)^2}{2\sigma^2}} \mathrm{d}k \tag{7.215}$$

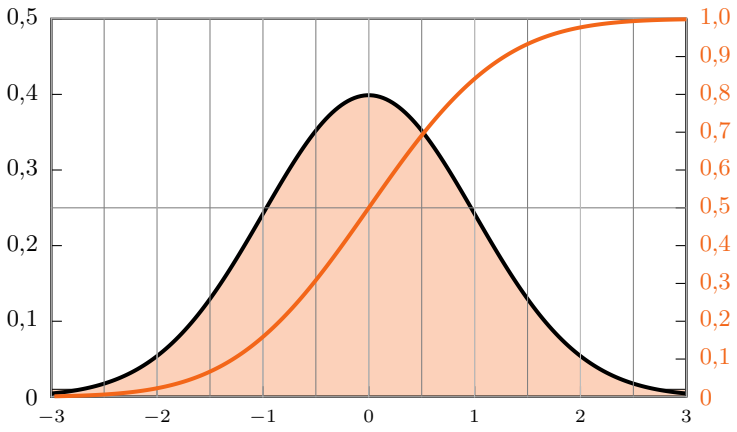


FIGURE 7.10 – Loi Normale pour $(\sigma, \mu) = (1, 0)$

Cette loi régit sous des conditions très générales, et souvent rencontrées, beaucoup de phénomènes aléatoires. Elle est par ailleurs symétrique par rapport à la moyenne μ.

Montrons maintenant que μ représente bien l'espérance mathématique de x :

$$E(X) = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \tag{7.216}$$

Posons $u = (x-\mu)/\sigma$. Nous avons dès lors :

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma u + \mu)\mathrm{e}^{-\frac{u^2}{2}}\sigma\mathrm{d}u \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u \end{aligned} \tag{7.217}$$

Calculons :

$$J = \int_{-\infty}^{+\infty} u\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u = u\mathrm{e}^{-\frac{u^2}{2}}\Big|_{-\infty}^{+\infty} = 0 - 0 = 0 \tag{7.218}$$

Donc :

$$E(X) = \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u = \sqrt{2\pi}\frac{\mu}{\sqrt{2\pi}} = \mu \tag{7.219}$$

REMARQUE Le lecteur pourrait trouver déroutant, dans un premier temps, que le paramètre d'une fonction soit un des résultats que nous cherchons de la fonction. Ce qui dérange est la mise en pratique d'une telle chose. Tout s'éclairera lors de l'étude des concepts d'*estimateurs de vraisemblance*.

Montrons aussi que σ représente bien l'écart type de X [9] et pour cela, utilisons l'expression de HUYGHENS (7.65). Commençons alors par calculer $E(X^2)$:

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu^2}{\sigma}\right)}\mathrm{d}x \tag{7.220}$$

Posons $y = (x-\mu)/\sqrt{2}\sigma$ qui conduit dès lors à :

$$\begin{aligned} E(X^2) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} (y\sqrt{2}\sigma + \mu)^2\mathrm{e}^{-y^2}\mathrm{d}y \\ &= \frac{2}{\sqrt{\pi}}\sigma^2 \int_{-\infty}^{+\infty} y^2\mathrm{e}^{-y^2}\mathrm{d}y + 2\mu\sigma\sqrt{\frac{2}{\pi}} \int_{-\infty}^{+\infty} y\mathrm{e}^{-y^2}\mathrm{d}y + \frac{\mu^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2}\mathrm{d}y \end{aligned} \tag{7.221}$$

or, nous savons :

$$\int_{-\infty}^{+\infty} y\mathrm{e}^{-y^2}\mathrm{d}y = 0; \qquad \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2}\mathrm{d}y = \sqrt{\pi} \tag{7.222}$$

p348 Il reste donc à calculer la première intégrale. Procédons par une intégration par parties• :

$$\int_{-\infty}^{+\infty} y^2\mathrm{e}^{-y^2}\mathrm{d}y = \int_{-\infty}^{+\infty} y\left(y\mathrm{e}^{-y^2}\right)\mathrm{d}y = -y\frac{\mathrm{e}^{-y^2}}{2}\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2}\mathrm{d}y = \frac{\sqrt{\pi}}{2} \tag{7.223}$$

D'où :

$$E(X^2) = \frac{2}{\sqrt{\pi}}\sigma^2\frac{\sqrt{\pi}}{2} + 2\mu\sigma\sqrt{\frac{2}{\pi}}0 + \frac{\mu^2}{\sqrt{\pi}}\sqrt{\pi} = \mu^2 + \sigma^2 \tag{7.224}$$

Il vient finalement :

$$V(X) = E(X^2) - E(X)^2 = \sigma^2 \tag{7.225}$$

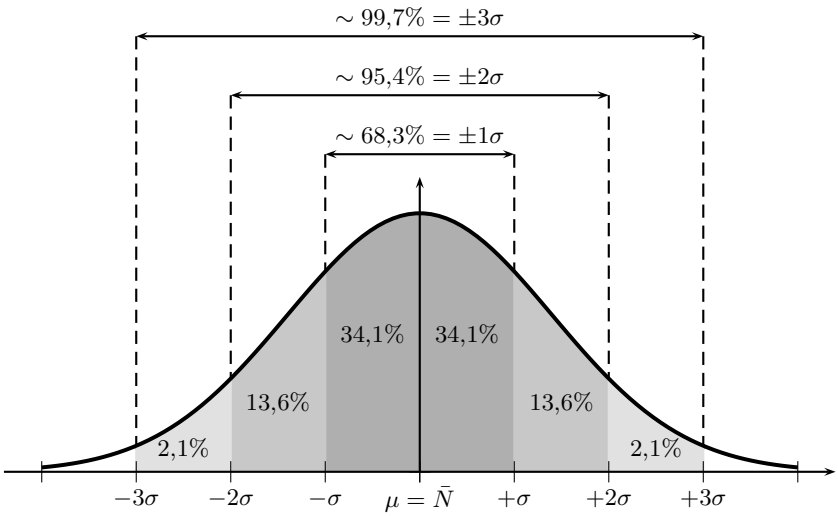


FIGURE 7.11 – Loi de GAUSS-LAPLACE. Les formes exactes ne sont pas conservées à des fins d'illustration

Une signification supplémentaire de l'écart-type dans la loi de GAUSS-LAPLACE est une mesure de la largeur de la distribution telle [10] que toute moyenne et pour tout écart-type non nul nous obtenons la figure 7.11. La largeur de l'intervalle a une très grande importance dans l'interprétation des incertitudes d'une mesure. La présentation d'un résultat comme $\bar{N} \pm \sigma$ signifie que la valeur moyenne a environ 68,3% de chance (probabilité) de se trouver entre les limites de $\bar{N} - \sigma$ et $\bar{N} + \sigma$, ou qu'elle a 95,5% de se trouver entre $\bar{N} - 2\sigma$ et $\bar{N} + 2\sigma$ etc.

REMARQUE Ce concept est beaucoup utilisé en gestion de la qualité en entreprise particulièrement avec le concept industriel anglo-saxon Six Sigma (cf. chapitre de Génie Industriel) qui impose une maîtrise de 6σ autour de chaque côté de la moyenne des côtés des pièces fabriquées.

La loi de GAUSS-LAPLACE n'est par ailleurs pas qu'un outil d'analyse de données mais également de génération de données. Effectivement, cette loi est une des plus importantes dans le monde des multinationales qui recourent aux outils statistiques pour la gestion du risque, la gestion de projets et la simulation lorsqu'un grand nombre de variables aléatoires sont en jeu.

Dans ce cadre d'application, il est par ailleurs très souvent fait usage de la somme —durée des tâches— ou le produit de variables aléatoires —facteur d'incertitude du client— suivant des lois de GAUSS-LAPLACE.

niveau de qualité sigma	taux de non-défection assuré en %	taux de défection en parties par million
1σ	68,26894	317 311
2σ	95,4499	45 500
3σ	99,73002	2 700
4σ	99,99366	63,4
5σ	99,999943	0,57
6σ	99,9999998	0,002

TABLEAU 7.2 – (7.274)

Somme de deux variables aléatoires

Soit X, Y deux variables aléatoires indépendantes. Supposons que X suit la loi $N(\mu_1, \sigma_1)$ et que Y suit la loi $N(\mu_2, \sigma_2)$. Alors, la variable aléatoire $Z = X + Y$ aura une densité égale au produit

9. Il convient, en d'autres termes de montrer que $V(x) = \sigma^2$.
10. Toutes ces grandeurs ne peuvent se vérifier qu'à l'aide d'intégration numérique.

de convolution de f_X et f_Y. C'est-à-dire :

$$f_Z(s) = \int_{-\infty}^{+\infty} f_X(x) f_Y(s-x)\mathrm{d}x = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \mathrm{e}^{-\frac{(s-x-\mu_2)^2}{2\sigma_2^2}} \mathrm{d}x \tag{7.226}$$

ce qui équivaut à faire le produit conjoint des probabilités d'apparition des deux variables continues[11]. Pour simplifier l'expression, faisons le changement de variable $t = x - \mu_1$ et posons $a = \mu_1 + \mu_2 - s$, $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Puisque :

$$(s - x - \mu^2)^2 = (-s + x + \mu^2)^2 = (t + a)^2 \tag{7.227}$$

nous obtenons :

$$\begin{aligned} f_Z(s) &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{t^2}{2\sigma_1^2}} \mathrm{e}^{-\frac{(t+a)^2}{2\sigma_2^2}} \mathrm{d}t = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{\left(\sigma t + \frac{a\sigma_1^2}{\sigma}\right)^2 + \frac{\sigma_1^2 a^2 \sigma_2^2}{\sigma^2}}{2\sigma_1^2\sigma_2^2}} \mathrm{d}t \\ &= \frac{\mathrm{e}^{\frac{a^2}{2\sigma^2}}}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{\left(\sigma t + \frac{a\sigma_1^2}{\sigma}\right)^2}{2\sigma_1^2\sigma_2^2}} \mathrm{d}t \end{aligned} \tag{7.228}$$

Posons :

$$u = \frac{\sigma t + \frac{a\sigma_1^2}{\sigma}}{\sqrt{2}\sigma_1\sigma_2} \Rightarrow \frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\sigma}{\sqrt{2}\sigma_1\sigma_2} \Rightarrow \mathrm{d}t = \sqrt{2}\sigma_1\sigma_2 \frac{\mathrm{d}u}{\sigma} \tag{7.229}$$

alors :

$$f_Z(s) = \frac{\mathrm{e}^{\frac{a^2}{2\sigma^2}}}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \mathrm{e}^{-u^2} \mathrm{d}u \tag{7.230}$$

or, d'après l'égalité (7.211) et $a^2 = (s - \mu_1 - \mu_2)^2$, l'expression (7.230) devient :

$$f_Z(s) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{\frac{(s-\mu_1-\mu_2)^2}{2\sigma^2}} \tag{7.231}$$

Nous reconnaissons l'expression de la loi de GAUSS-LAPLACE de moyenne $\mu_1 + \mu_2$ et d'écart type $\sigma = \sqrt{\sigma_1^2 + \sigma_1^2}$. Par conséquent, $X + Y$ suit la loi $N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_1^2})$.

Ce résultat est ce que nous nommons en statistiques la *stabilité par la somme* de la loi de GAUSS-LAPLACE. Nous retrouverons ce type de propriétés pour d'autres lois que nous étudierons plus loin.

Dans l'industrie, nous avons in-extenso l'expression suivante qui est très importante lorsque possédant plusieurs échantillons de mesure d'une même population, nous souhaitons savoir la loi de toute la population à partir de la connaissance de la moyenne et de l'écar-type de chacune :

$$N(\mu, \sigma) = N\left(\frac{1}{n}\sum_{i=1}^{n}\mu_i, \frac{1}{n}\sqrt{\sum_{i=1}^{n}\sigma_i^2}\right) \tag{7.232}$$

Produit de deux variables aléatoires

Soit X, Y deux variables aléatoires indépendantes réelles. Nous désignerons par f_X et f_Y, les densités correspondantes et nous cherchons à déterminer la densité de la variable $Z = X \cdot Y$.

Chap6 Notons f la fonction de densité du couple (X, Y). Puisque X, Y sont indépendantes• :

$$f(x, y) = f_X(x) \cdot f_Y(y) \tag{7.233}$$

La fonction de répartition de Z est :

$$F(z) = P(Z \leqslant z) = P(X \cdot Y \leqslant z) = \iint_D f(x, y)\mathrm{d}x\mathrm{d}y = \iint_D f_X(x) f_Y(y)\mathrm{d}x\mathrm{d}y \tag{7.234}$$

11. Le lecteur peut essayer à titre d'exercice de se souvenir du même genre de calcul sous forme discrète.

où $D = \{(x,y)|x \cdot y \leqslant z\}$. D peut se réécrire comme union disjointe [12] $D = D_1 \sqcup D_2 \sqcup D_3$ avec :

$$\begin{aligned} D_1 &= \{(x,y) \in \mathbb{R}^2 | x \cdot y \leqslant z \text{ et } x > 0\} \\ D_2 &= \{(x,y) \in \mathbb{R}^2 | x \cdot y \leqslant z \text{ et } x < 0\} \\ D_3 &= \{(x,y) \in \mathbb{R}^2 | x \cdot y \leqslant z \text{ et } x = 0\} \end{aligned} \tag{7.235}$$

Nous avons :

$$F(z) = \iint_{D_1} f_X(x) f_Y(y) \mathrm{d}x\mathrm{d}y + \iint_{D_2} f_X(x) f_Y(y) \mathrm{d}x\mathrm{d}y + \underbrace{\iint_{D_3} f_X(x) f_Y(y) \mathrm{d}x\mathrm{d}y}_{=0} \tag{7.236}$$

La dernière intégrale vaut zéro car D_3 est de mesure (épaisseur) nulle pour l'intégrale selon x.

Nous effectuons ensuite le changement de variable suivant $x = x$ et $t = xy$ de jacobien :

$$\text{dét} \begin{bmatrix} 1 & 0 \\ -t/x^2 & 1/x \end{bmatrix} = \frac{1}{|x|} \tag{7.237}$$

Donc :

$$\begin{aligned} F(z) &= \int_{-\infty}^{z} \int_{0}^{+\infty} \frac{f_X(x) f_Y(t/x)}{|x|} \mathrm{d}x\mathrm{d}t + \int_{-\infty}^{z} \int_{-\infty}^{0} \frac{f_X(x) f_Y(t/x)}{|x|} \mathrm{d}x\mathrm{d}t \\ &= \int_{-\infty}^{z} \int_{-\infty}^{+\infty} \frac{f_X(x) f_Y(t/x)}{|x|} \mathrm{d}x\mathrm{d}t \end{aligned} \tag{7.238}$$

Notons f_Z la densité de la variable Z. Par définition :

$$F(z) = \int_{-\infty}^{z} f_Z(t) \mathrm{d}t \tag{7.239}$$

Par conséquent, d'après l'équation (7.238) :

$$f_Z(t) = \int_{-\infty}^{+\infty} \frac{f_X(x) f_Y(t/x)}{|x|} \mathrm{d}x \tag{7.240}$$

Dans le cas d'une loi de Gauss-Laplace (loi Normale), cette intégrale ne peut être calculée que numériquement : il faut alors faire appel à des méthodes d'intégration du type Monte-Carlo (cf. chapitre de Méthodes Numériques).

Loi normale centrée réduite

La fonction de Gauss-Laplace n'est pas tabulée puisqu'il faudrait autant de tables numériques que de valeurs possibles pour la moyenne μ et l'écart-type σ. C'est pourquoi, en opérant un changement de variable, la loi Normale devient la *loi Normale centrée réduite* où :

1. *Centrée* signifie soustraire la moyenne μ ;
2. *Réduite* signifie, diviser par l'écart-type σ.

Par ce changement de variable, la variable k est remplacée par la variable aléatoire centrée réduite :

$$k^* = \frac{k - \mu}{\sigma} \tag{7.241}$$

Si la variable k a pour moyenne μ et pour écart-type σ alors la variable k^* a pour moyenne 0 et pour écart-type 1 et la relation (7.214) s'écrit plus simplement :

$$P(x,0,1) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} = N(0,1) \tag{7.242}$$

qui n'est d'autre que l'expression de la loi Normale centrée réduite souvent notée $N(0,1)$.

Remarque Nous retrouverons très fréquemment cette distribution en physique et en gestion de la qualité.

12. Nous faisons cette opération pour anticiper lors du futur changement de variables une division par zéro

Droite de HENRY

Souvent, dans les entreprises c'est la loi de GAUSS-LAPLACE (Normale) qui est analysée mais des logiciels courants et facilement accessibles sont incapables de vérifier que les données mesurées suivent une loi Normale lorsque nous faisons de l'analyse fréquentielle (aucun outil intégré par défaut ne permet de le faire) et que nous n'avons pas les données d'origines non groupées.

L'astuce consiste alors à utiliser la variable centré réduite k^*. L'idée de la droite d'HENRY est alors d'utiliser la relation linéaire entre k et k^* donnée par l'équation :

$$k^* = f(k) = \frac{k}{\sigma} - \frac{\mu}{\sigma} \tag{7.243}$$

et qui peut être tracée pour déterminer la moyenne et l'écart-type de la loi Normale.

Exemple 7.7 *Supposons que nous ayons l'analyse fréquentielle suivante de 10 000 tickets de caisse dans un supermarché selon le tableau 7.3. Le tracé du nombre de tickets en fonction du prix est*

montant des tickets	nombre de tickets	nombre cumulés tickets	fréquences relatives cumulées	correspondance pour k^* de $N(0,1)$
$[0;50[$	668	668	0,068	$-1,5$
$[50,100[$	919	1 587	0,1587	-1
$[100,150[$	1 498	3 085	0,3085	$-0,5$
$[150,200[$	1 915	5 000	0,5000	0
$[200,250[$	1 915	6 915	0,6915	0,5
$[250,300[$	1 498	8 413	0,8413	1
$[300,350[$	919	9 332	0,9332	1,5
$[350,400[$	440	9 772	0,9772	2
[400 et $+[$	228	10 000	1	-

TABLEAU 7.3 – (7.301)

indiqué sur la figure 7.13 et qui ressemble terriblement à une loi Normale d'où l'autorisation, sans trop de risques, d'utiliser dans cet exemple la technique de la droite d'HENRY. Mais que faire maintenant ? Eh bien connaissant les fréquences cumulées, il ne nous reste plus qu'à calculer pour chacune d'entre elles k^* à l'aide de tables numériques.

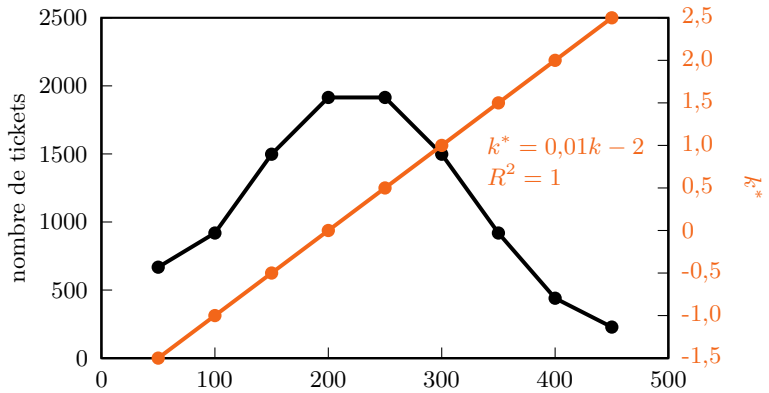


FIGURE 7.12 – Répartition du prix des tickets, loi Binomiale pour $\beta = (10, 0{,}5)$ et droite de HENRY

Ceci nous donnera les valeurs de la loi Normale centrée réduite $N(0,1)$ de ces mêmes fréquences respectives cumulées (fonction de répartition). D'après (7.243), nous avons sous forme discrète, la fonction suivante :

$$k_i^* = f(k_i) = \frac{k_i}{\sigma} - \frac{\mu}{\sigma} \tag{7.244}$$

tracée sur la figure 7.13. À l'aide de la régression, il vient :

$$k^* = f(k) = \frac{k}{\sigma} - \frac{\mu}{\sigma} = 0{,}01k - 2 \tag{7.245}$$

soit, immédiatement $\sigma = 100$ et $\mu = 200$. Il s'agit donc d'une technique particulière pour une distribution particulière. Des techniques similaires plus ou moins simples existent pour nombre de distributions.

7.4.9 Loi log-normale

Nous disons qu'une variable aléatoire positive X suit une loi log-normale de paramètres $\mu, \sigma > 0$, si et seulement si en posant :

$$y = \ln(X) \tag{7.246}$$

nous voyons que y suit une fonction de probabilité cumulée de type loi Normale de moyenne μ et de variance σ^2 (moments de la loi Normale).

La fonction de densité de X pour $x \geqslant 0$ est alors (cf. chapitre de Calcul Intégral) :

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} \mathrm{e}^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \tag{7.247}$$

Ce type de scénario se retrouve fréquemment en physique, dans les techniques de maintenance ou encore en finance des marchés dans le modèle de pricing des options (voir ces chapitres respectifs du site pour des exemples concrets). Il y a par ailleurs une remarque importante relativement à la loi log-normale dans le traitement plus loin du théorème central limite.

Montrons que la fonction de probabilité cumulée correspond bien à une loi Normale si nous faisons le changement de variable mentionné précédemment en posant :

$$y = \ln(x) \Rightarrow \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{x} \Rightarrow \mathrm{d}x = x\mathrm{d}y \quad \text{et} \quad x = \mathrm{e}^y \tag{7.248}$$

nous avons bien :

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma^2}} \mathrm{d}y \tag{7.249}$$

L'espérance (moyenne) de X est donnée alors par (le logarithme népérien n'étant pas défini pour $x < 0$ nous bornons l'intégrale à partir de zéro) :

$$E(X) = \int_{0}^{+\infty} xf(x)\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma^2}+y} \mathrm{d}y \tag{7.250}$$

où nous avons effectué le changement de variable (7.248). Grâce à l'identité suivante :

$$\frac{(y-\mu)^2}{2\sigma^2} + y = -\frac{1}{2\sigma^2}((y-(\mu+\sigma^2))^2 - (\mu+\sigma^2)^2 + \mu^2) \tag{7.251}$$

l'intégrale (7.250) devient donc :

$$E(X) = \frac{\mathrm{e}^{\frac{(\mu+\sigma^2)^2-\mu^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{(y-(\mu+\sigma^2))^2}{2\sigma^2}} \mathrm{d}y = \mathrm{e}^{\frac{(\mu+\sigma^2)^2-\mu^2}{2\sigma^2}} = \mathrm{e}^{\left(\mu+\frac{\sigma^2}{2}\right)} \tag{7.252}$$

Afin d'obtenir l'expression de la variance de X, calculons $E(X^2)$ en procédant de manière similaire aux développements précédents :

$$E(X^2) = \frac{\mathrm{e}^{\frac{(\mu+2\sigma^2)^2-\mu^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{(y-(\mu+2\sigma^2))^2}{2\sigma^2}} \mathrm{d}y = \mathrm{e}^{\frac{(\mu+2\sigma^2)^2-\mu^2}{2\sigma^2}} = \mathrm{e}^{2\left(\mu+\sigma^2\right)} \tag{7.253}$$

où nous avons encore une fois le changement de variable (7.248) associé à la transformation :

$$-\frac{(y-\mu)^2}{2\sigma^2} + 2y = -\frac{1}{2\sigma^2}((y-(\mu+2\sigma^2))^2 - (\mu+2\sigma^2)^2 + \mu^2) \tag{7.254}$$

Finalement :

$$V(X) = E(X^2) - E(X)^2 = \mathrm{e}^{2\mu+2\sigma^2} - \mathrm{e}^{\left(\mu+\frac{\sigma^2}{2}\right)} = \mathrm{e}^{2\mu+2\sigma^2} - \mathrm{e}^{2\mu+\sigma^2} = \mathrm{e}^{2\mu+\sigma^2}\left(\mathrm{e}^{\sigma^2}-1\right) \tag{7.255}$$

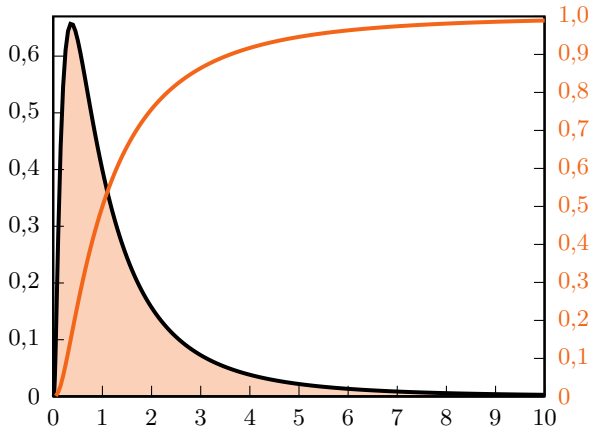


FIGURE 7.13 – Loi Log-Normale pour $(\sigma, \mu) = (1, 0)$

7.4.10 Loi uniforme continue

Soit $a < b$. Nous définissons la fonction de distribution de la loi uniforme par la relation :

$$P_{a,b}(x) = \frac{1}{b-a} 1_{[a,b]} \tag{7.256}$$

Nous avons donc pour fonction de répartition :

$$P(X \leqslant x) = \int_a^x \frac{1}{b-a} 1_{[a,b]} \mathrm{d}u = \frac{1}{b-a} 1_{[a,b]} \int_a^x \mathrm{d}u = \frac{a_{[a,b]} - a}{b-a} \tag{7.257}$$

Il s'agit bien d'une fonction de distribution car elle vérifie (intégrale simple) :

$$\int_{-\infty}^{+\infty} P_{a,b}(x)\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{1_{[a,b]}}{b-a}\mathrm{d}x = \frac{1_{[a,b]}}{b-a} \int_a^b \mathrm{d}x = \frac{1_{[a,b]}}{b-a} x|_a^b = 1_{[a,b]} \frac{b-a}{b-a} = 1 \tag{7.258}$$

La fonction uniforme a par ailleurs pour espérance (moyenne) :

$$\begin{aligned} \mu = E(X) &= \int x f(x)\mathrm{d}x = \frac{1}{b-a} \int_a^b x\mathrm{d}x = \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b = \frac{1}{b-a} \frac{b^2-a^2}{2} \\ &= \frac{1}{b-a} \frac{(b+a)(b-a)}{2} = \frac{a+b}{2} \end{aligned} \tag{7.259}$$

et pour variance en utilisant la formule de HUYGHENS :

$$\begin{aligned} V(X) = E(X^2) - E(X)^2 = \mu &= \int_a^b x^2 f(x)\mathrm{d}x - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{1}{b-a} \int_a^b x^2 \mathrm{d}x - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12} \end{aligned} \tag{7.260}$$

$1_{[a,b]}$ signifie qu'en dehors du domaine de définition $[a, b]$ la fonction de distribution est nulle. Nous retrouverons ce type de notation dans certaines autres fonctions de distribution.

REMARQUE Cette fonction est souvent utilisée en simulation dans les entreprises pour signaler que la variable aléatoire a des probabilités égales d'avoir une valeur comprise dans un certain intervalle (typiquement dans les rendements de portefeuilles ou encore dans l'estimation des durées des projets).

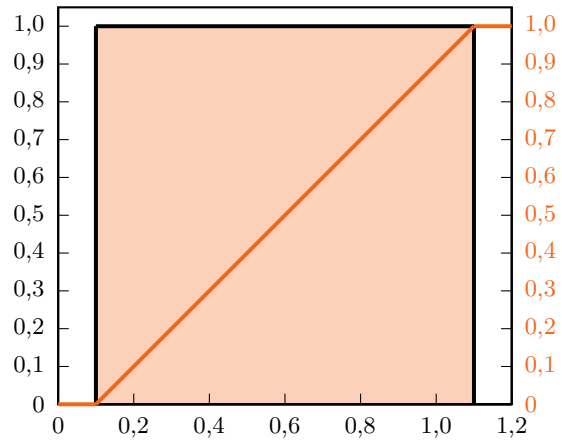


FIGURE 7.14 – Loi uniforme de paramètres $(a, b) = (0, 1)$

Voyons un résultat intéressant de la loi Uniforme. Souvent j'entends des gestionnaires de haut niveau dire que comme une mesure à une probabilité égale d'avoir lieu dans un intervalle fermé donné, alors la somme de deux variables aléatoires indépendantes du même type aussi. Or nous allons démontrer ici que ce n'est pas.

Considérons deux variables aléatoires indépendantes X et Y suivante une loi uniforme dans un intervalle fermé $[0, a]$. Nous cherchons donc la densité de leur somme qui sera notée :

$$Z = X + Y \tag{7.261}$$

Nous avons alors :

$$f_X(x) = f_Y(y) = \begin{cases} 1 & \text{si } 0 \leqslant x, y \leqslant a \\ 0 & \text{sinon} \end{cases} \tag{7.262}$$

avec la variable $0 \leqslant z \leqslant 2a$. Pour calculer la loi de la somme, rappelons que nous savons qu'en termes discrets cela équivaut faire le produit conjoint des probabilités (cf. chapitre de Probabilités) d'apparition des deux variables continues, c'est-à-dire :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z - y) f_Y(y) \mathrm{d}y \tag{7.263}$$

Comme $f_Y(y) = 1$ si $0 \leqslant y \leqslant a$ et sinon 0 alors le produit de convolution précédent se réduit à :

$$f_Z(z) = \int_0^a f_X(z - y) \mathrm{d}y \tag{7.264}$$

L'intégrant vaut par définition 0 sauf lorsque par construction $0 \leqslant z - y \leqslant a$ où il vaut alors 1.

Intéressons nous alors aux bornes de l'intégrale dans ce dernier cas. Faisons d'abord un changement de variables en posant $u = z - y$ d'où $\mathrm{d}u = -\mathrm{d}y$. L'intégrale s'écrit alors dans alors dans cet intervalle après ce changement de variable :

$$f_Z(z) = \int_0^a f_X(z - y) \mathrm{d}y = -\int_z^{z-a} f_X(u) \mathrm{d}u = \int_{z-a}^{z} \mathrm{d}u \tag{7.265}$$

En se rappelant comme vu au début que $0 \leqslant z \leqslant 2a$, alors nous avons immédiatement si $z < 0$ et $z > 2a$ que l'intégrale est nulle.

Nous allons considérer deux cas pour cet intervalle car la convolution de ces deux fonctions rectangulaires peuvent se distingueront à la situation où dans un premier temps elles se croisent, c'est-à-dire où $0 \leqslant z \leqslant a$, et ensuite s'éloignent l'une de l'autre, c'est-à-dire $a < z \leqslant 2a$:

– cas où $0 \leqslant z \leqslant a$:

$$f_Z(u) = \int_{z-a}^{z} \mathrm{d}u = \int_0^z \mathrm{d}u = u|_0^z = z \tag{7.266}$$

où nous avons changé la borne inférieure à 0 car de toute façon $f_X(u)$ est nulle pour toute valeur négative (et lorsque $0 \leqslant z \leqslant a$, $z - a$ est justement négatif ou nul) ;

– cas où $a < z \leqslant 2a$:

$$f_Z(u) = \int_{z-a}^{z} \mathrm{d}u = \int_{z-a}^{a} \mathrm{d}u = u|_{z-a}^{a} = a - (z-a) = 2a - z \tag{7.267}$$

où nous avons changé la borne supérieur à a car de toute façon $f_X(u)$ est nulle pour toute valeur supérieure (et lorsque $a < z \leqslant 2a$, z est justement plus grand que a).

Donc au final, nous avons :

$$f_Z(z) = \begin{cases} z & \text{si} \quad 0 \leqslant z \leqslant a \\ 2a - z & \text{si} \quad a < z \leqslant 2a \\ 0 & \text{sinon} \end{cases} \tag{7.268}$$

Il s'agit d'un cas particulier, volontairement simplifié, de la loi triangulaire.

7.4.11 Loi triangulaire

Soit $a < c < b$. Nous définissons la loi triangulaire par construction selon les deux fonctions de distribution suivantes :

$$P_{a,c}(x) = \frac{2(x-a)}{(b-a)(c-a)} \cdot 1_{[a,c]} \quad \text{et} \quad P_{c,b}(x) = \frac{2(b-x)}{(b-a)(b-c)} \cdot 1_{]c,b]} \tag{7.269}$$

où a est souvent assimilé à la valeur optimiste, c la valeur attendue (le mode) et b la valeur pessimiste.

C'est effectivement la seule manière de l'écrire si le lecteur garde à l'esprit que le triangle de base $c-a$ doit avoir une hauteur h valant $2/(c-a)$ telle que sa surface totale soit égale à l'unité.

La pente de la première droite (croissante de gauche) est donc bien évidemment :

$$\frac{2a}{(b-a)(c-a)} \tag{7.270}$$

et la pente de la deuxième droite (décroissante à droite) :

$$\frac{-2}{(b-a)(b-c)} \tag{7.271}$$

Cette fonction est une fonction de distribution si elle vérifie :

$$\rho = \int_{-\infty}^{+\infty} (P_{a,c}(x) + P_{c,b}(x))\mathrm{d}x = 1 \tag{7.272}$$

Il s'agit dans ce cas de l'aire du triangle qui rappelons-le est simplement la base multipliée par la hauteur le tout divisé par 2 (cf. chapitre sur les Formes Géométriques) :

$$\rho = \frac{1}{2}\left((b-a) \cdot \frac{2}{b-a}\right) = 1 \tag{7.273}$$

REMARQUE Cette fonction est beaucoup utilisée en gestion de projet dans le cadre de l'estimation des durées des tâches ou encore en simulations industrielles. La valeur a correspondant à la valeur optimiste, la valeur c à la valeur attendue (mode) et la valeur b à la valeur pessimiste.

La fonction triangulaire a par ailleurs une espérance (moyenne) :

$$\mu = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \int_{a}^{c} x\frac{2(x-a)}{(b-a)(c-a)}\mathrm{d}x + \int_{c}^{b} x\frac{2(b-x)}{(b-a)(b-c)}\mathrm{d}x = \frac{a+b+c}{3} \tag{7.274}$$

et pour variance :

$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{+\infty}(x-\mu)^2 f(x)\mathrm{d}x = \int_a^c (x-\mu)^2\frac{2(x-a)}{(b-a)(c-a)}\mathrm{d}x + \int_c^b (x-\mu)^2\frac{2(b-x)}{(b-a)(b-c)}\mathrm{d}x \\ &= \frac{a^2+b^2+c^2-ab-ac-bc}{18}\end{aligned} \tag{7.275}$$

après introduction de l'égalité (7.274). Nous pouvons montrer que la somme de deux variables aléatoires indépendantes chacune de loi uniforme sur $[a,b]$ suit une loi uniforme sur $[2a,2b]$ mais si elles n'ont pas les mêmes bornes, alors leur somme donne une loi triangulaire.

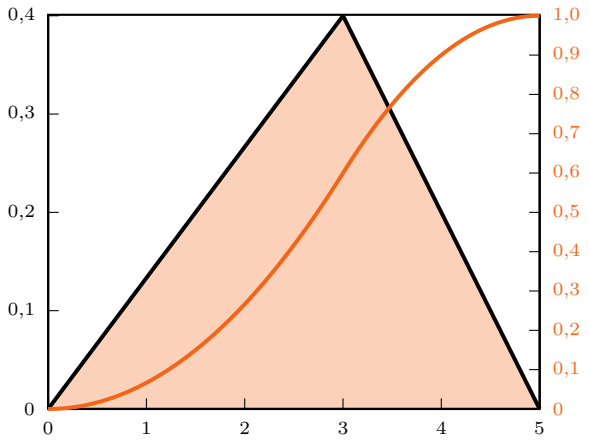


FIGURE 7.15 – Loi triangulaire de paramètres $(a,c,b)=(0,3,5)$

7.4.12 Loi de Pareto

La loi de PARETO est la formalisation du principe des 80-20. Cet outil d'aide à la décision détermine les facteurs (environ 20%) cruciaux qui influencent la plus grande partie (80%) de l'objectif.

REMARQUE Cette loi est un outil fondamental et basique en gestion de la qualité (cf. chapitre de Génie Industriel et Techniques de Gestion). Elle est aussi utilisée en réassurance. La théorie des files d'attente s'est intéressée à cette distribution, lorsque des recherches des années quatre-vingt dix ont montré que cette loi régissait aussi au nombre de grandeurs observées dans le trafic internet (et plus généralement sur tous les réseaux de données à grande vitesse).

Une variable aléatoire est dite par définition suivre une loi de PARETO si sa fonction de répartition est donnée par :

$$P(X \leqslant x) = 1 - \left(\frac{x_m}{x}\right)^k \tag{7.276}$$

avec x qui doit être supérieur ou égal à x_m. La fonction de densité (fonction de distribution) de PARETO est alors :

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(1-\left(\frac{x_m}{x}\right)^k\right) = -\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x_m}{x}\right)^k = -x_m^k\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{x^k} = k\frac{x_m^k}{x^{k+1}} \tag{7.277}$$

avec $k\in\mathbb{R}_+$ et $x \geqslant x_m \geqslant 0$ (donc $x>0$).

La distribution de PARETO est donc définie par deux paramètres, x_m et k (nommé *index de Pareto*). C'est par ailleurs bien une fonction de distribution puisque étant connue sa fonction de répartition :

$$\int_0^{+\infty} f(x)\mathrm{d}x = 1-\left(\frac{x_m}{x}\right)^k\Big|_{x_m}^{+\infty} = (1-0)-(1-1^k) = 1 \tag{7.278}$$

L'espérance (moyenne) est donnée par :

$$\mu = E(X) = \int_{x_m}^{+\infty} xf(x)\mathrm{d}x = \int_{x_m}^{+\infty} xk\frac{x_m^k}{x^{k+1}}\mathrm{d}x = \frac{kx_m}{k-1} \tag{7.279}$$

si $k > 1$. Si $k \leqslant 1$, l'espérance n'existe pas. Pour calculer la variance, nous avons :

$$E(X^2) = \int_{x_m}^{+\infty} x^2 f(x)\mathrm{d}x = kx_m^k \int_{x_m}^{+\infty} \frac{1}{x^{k-1}}\mathrm{d}x = \frac{kx_m^2}{k-2} \tag{7.280}$$

si $k > 2$. Si $k \leqslant 2$, $E(X^2)$ n'existe pas. Donc si $k > 2$:

$$\sigma^2 = V(X) = \frac{kx_m^2}{k-2} - \left(\frac{kx_m}{k-1}\right)^2 = \frac{kx_m^2}{(k-1)^2(k-2)} \tag{7.281}$$

Si $k \leqslant 2$, la variance n'existe pas. Les fonctions respectives sont indiquées sur la figure 7.16.

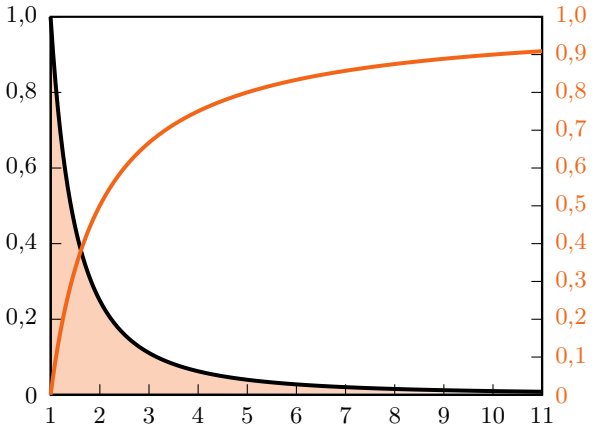


FIGURE 7.16 – Loi de PARETO de paramètres $(x_m, k) = (2, 1)$

REMARQUE Lorsque $k \to +\infty$, la distribution s'approche de $\delta(x - x_m)$ où δ est la fonction Delta de Dirac.

7.4.13 Loi exponentielle

Nous définissons la loi exponentielle par la relation de fonction de distribution suivante :

$$P(x) = \lambda \mathrm{e}^{-\lambda x} 1_{[0,+\infty]} \tag{7.282}$$

avec $\lambda > 0$ qui comme nous allons de suite le montrer n'est au fait que l'inverse de la moyenne.

REMARQUE Cette fonction se retrouve fréquemment en physique nucléaire (désintégrations) ou encore en physique quantique ainsi qu'en fiabilité (maintenance préventive).

Il s'agit par ailleurs bien d'une fonction de distribution car elle vérifie :

$$\int_{-\infty}^{+\infty} P_\lambda(x)\mathrm{d}x = \lambda \int_0^{+\infty} \mathrm{e}^{-\lambda x}\mathrm{d}x = \mathrm{e}^{-\lambda x}\Big|_0^{+\infty} = 1 \tag{7.283}$$

La fonction exponentielle a pour espérance (moyenne) en utilisant l'intégration par parties :

$$\mu = \int_{-\infty}^{+\infty} xP_\lambda(x)\mathrm{d}x = \lambda \int_0^{+\infty} x\mathrm{e}^{-\lambda x}\mathrm{d}x = -x\mathrm{e}^{-\lambda x}\Big|_0^{+\infty} - \int_0^{+\infty} -\mathrm{e}^{-\lambda x}\mathrm{d}x = \int_0^{+\infty} \mathrm{e}^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda} \tag{7.284}$$

et pour variance, il ne nous reste plus qu'à calculer :

$$E(X^2) = \lambda \int_0^{+\infty} x^2 \mathrm{e}^{-\lambda x} \mathrm{d}x = \frac{2}{\lambda^2} \tag{7.285}$$

après une double intégration par parties. Il vient dès lors :

$$V(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \tag{7.286}$$

Donc l'écart-type et la moyenne ont exactement la même expression. Déterminons maintenant la fonction de répartition de la loi exponentielle :

$$P(X \leqslant x) = \int_0^x \lambda \mathrm{e}^{-\lambda t} \mathrm{d}t = 1 - \mathrm{e}^{-\lambda x} \tag{7.287}$$

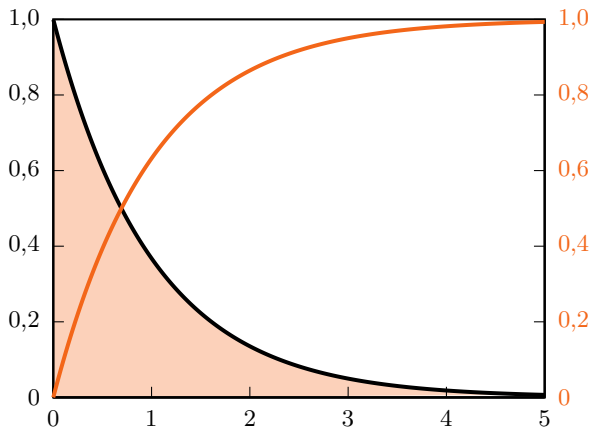


FIGURE 7.17 – Loi exponentielle de paramètre $\lambda = 1$

REMARQUE Nous verrons plus loin que la fonction de distribution exponentielle n'est qu'un cas particulier d'une fonction plus générale qui est la fonction du Khi-Deux, cette dernière aussi n'étant qu'un cas particulier d'une fonction encore plus générale qui est la fonction Gamma.

7.4.14 Loi de CAUCHY

Soient X, Y deux variables aléatoires indépendantes suivant des lois Normales centrées réduites (variance unité et espérance nulle). La fonction de densité est donc donnée par :

$$f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \tag{7.288}$$

La variable aléatoire $T = X/|Y|$ [13] suit une caractéristique appelée loi de CAUCHY ou loi de Lorentz.

Déterminons sa fonction de densité f grâce à la relation :

$$\forall t \in \mathbb{R}, P(T \leqslant t) = \int_{-\infty}^{t} f(x) \mathrm{d}x \tag{7.289}$$

soit, par dérivation élémentaire :

$$f(t) = \frac{\mathrm{d}}{\mathrm{d}t} P(T \leqslant t) \tag{7.290}$$

13. La valeur absolue intervient dans l'intégrale lors du changement variable.

dans le cas où f est continue.

Étant donné que X et Y sont indépendantes, la fonction de densité du vecteur aléatoire est donnée par la définition 6.9 :

$$(x, y) \to f_X(x) \cdot f_Y(x) \tag{7.291}$$

Donc :

$$P(T \leqslant t) = P\left(\frac{X}{|Y|} < t\right) = P(X \leqslant t|Y|) = \int_D f_X(x) \cdot f_Y(x)\mathrm{d}x\mathrm{d}y \tag{7.292}$$

où donc $D = \{(x, y)|x < t \cdot |y|\}$. Cette dernière intégrale devient :

$$\int_D f_X(x) \cdot f_Y(x)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty}\int_{-\infty}^{t\cdot|y|} f_X(x) \cdot f_Y(x)\mathrm{d}x\mathrm{d}y \tag{7.293}$$

Faisons le changement de variable $x = u \cdot |y|$ dans l'intégrale intérieure. Nous obtenons :

$$P(T \leqslant t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{t} f_X(u \cdot |y|) \cdot f_Y(y)|y|\mathrm{d}u\mathrm{d}y = \int_{-\infty}^{t}\int_{-\infty}^{+\infty} f_X(u \cdot |y|) \cdot f_Y(y)|y|\mathrm{d}y\mathrm{d}u \tag{7.294}$$

Donc :

$$f(t) = \frac{\mathrm{d}}{\mathrm{d}t}P(T \leqslant t) = \int_{-\infty}^{+\infty} f_X(t \cdot |y|) \cdot f_Y(y)|y|\mathrm{d}y = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{y^2(t^2+1)}{2}}|y|\mathrm{d}y \tag{7.295}$$

C'est maintenant que la valeur absolue va nous être utile pour écrire :

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{y^2(t^2+1)}{2}}|y|\mathrm{d}y = -\frac{1}{2\pi}\int_{-\infty}^{0} \mathrm{e}^{-\frac{y^2(t^2+1)}{2}}y\mathrm{d}y + \frac{1}{2\pi}\int_{0}^{+\infty} \mathrm{e}^{-\frac{y^2(t^2+1)}{2}}y\mathrm{d}y \tag{7.296}$$

Pour la première intégrale nous avons :

$$\frac{1}{2\pi}\int_{0}^{+\infty} \mathrm{e}^{-\frac{y^2(t^2+1)}{2}}y\mathrm{d}y = \left.\frac{\mathrm{e}^{-\frac{y^2(t^2+1)}{2}}}{t^2+1}\right|_{-\infty}^{0} = -\frac{1}{t^2+1} + \frac{1}{t^2+1} = 0 \tag{7.297}$$

Il ne reste donc plus que la seconde intégrale et en faisant le changement de variable $\nu = y^2$, nous obtenons :

$$f(t) = \frac{1}{2\pi}\int_{0}^{+\infty} \mathrm{e}^{-\frac{\nu(t^2+1)}{2}}\mathrm{d}\nu = -\left.\frac{\mathrm{e}^{-\frac{\nu(t^2+1)}{2}}}{\pi(t^2+1)}\right|_{0}^{+\infty} = \frac{1}{\pi(t^2+1)} \tag{7.298}$$

Ce que nous noterons par la suite, afin de respecter les notations optées jusqu'à présent :

$$P(x) = \frac{1}{\pi(x^2+1)} \tag{7.299}$$

et qui n'est d'autre que la fonction de CAUCHY, tracée sur la figure 7.18. Il s'agit par ailleurs bien d'une fonction de distribution car elle vérifie (cf. chapitre de Calcul Différentiel et Intégral) :

$$\int_{-\infty}^{+\infty} P(x)\mathrm{d}x = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\mathrm{d}x}{x^2+1} = \frac{1}{\pi}(\arctan(+\infty) - \arctan(-\infty)) = 1 \tag{7.300}$$

L'espérance de la fonction se calcule comme suit :

$$\begin{aligned}\mu &= \int_{-\infty}^{+\infty} xP(x)\mathrm{d}x = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x}{1+x^2}\mathrm{d}x = \frac{1}{\pi}\left(\int_{-\infty}^{0}\frac{x}{1+x^2}\mathrm{d}x + \int_{0}^{+\infty}\frac{x}{1+x^2}\mathrm{d}x\right)\\ &= \frac{1}{2\pi}\left(\ln(1+x^2)\big|_{-\infty}^{0} + \ln(1+x^2)\big|_{0}^{+\infty}\right) = \frac{1}{2\pi}(-\ln(\infty)+\ln(\infty))\end{aligned} \tag{7.301}$$

Les calculs précédents ne donnent pas zéro au fait car la soustraction d'infinis est indéterminée : la loi de CAUCHY n'admet donc pas d'espérance rigoureusement parlant. De même, une variance est absurde.

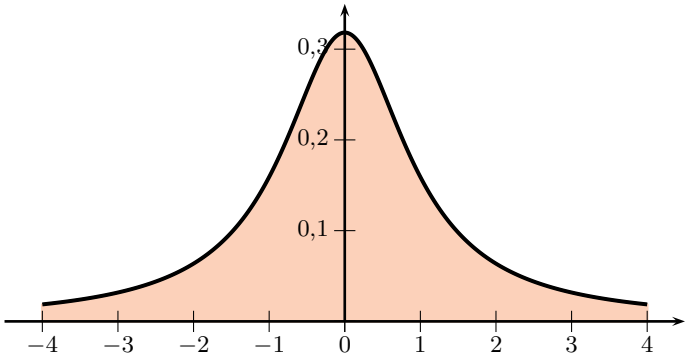


FIGURE 7.18 – Loi de CAUCHY

7.4.15 Loi bêta

Rappelons d'abord que la fonction Gamma d'EULER est définie par la relation (cf. chapitre de Calcul Différentiel Et Intégral) :

$$\Gamma(z) = \int_{0}^{+\infty} \mathrm{e}^{-x}x^{z-1}\mathrm{d}x \tag{7.302}$$

Nous avons démontré (cf. chapitre de Calcul Différentiel Et Intégral) qu'une propriété non triviale de cette fonction est que :

$$\Gamma(z+1) = z\Gamma(z) \tag{7.303}$$

Posons maintenant :

$$\Gamma(a)\Gamma(b) = \lim_{R\to+\infty}\iint_{A_R} \mathrm{e}^{-x-y}x^{a-1}y^{b-1}\mathrm{d}x\mathrm{d}y \tag{7.304}$$

où $A_R = \{(x,y)|x \geqslant 0, y \geqslant 0, x+y \leqslant R\}$. En faisant le changement de variables :

$$\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}u-v\\v\end{pmatrix} \tag{7.305}$$

nous obtenons :

$$\Gamma(a)\Gamma(b) = \lim_{R\to+\infty}\iint_{A_R} \mathrm{e}^{-x-y}x^{a-1}y^{b-1}\mathrm{d}x\mathrm{d}y = \lim_{R\to+\infty}\int_{0}^{R}\mathrm{e}^{-u}\left(\int_{0}^{u}(u-v)^{a-1}v^{b-1}\mathrm{d}v\right)\mathrm{d}u \tag{7.306}$$

Pour l'intégrale interne nous utilisons maintenant la substitution $v = ut$ $(0 \leqslant t \leqslant 1)$ et nous trouvons alors :

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \lim_{R\to+\infty}\int_0^R \mathrm{e}^{-u}\left(\int_0^u (u-v)^{a-1}v^{b-1}\mathrm{d}v\right)\mathrm{d}u \\ &= \lim_{R\to+\infty}\int_0^R \mathrm{e}^{-u}u^{a+b-1}\mathrm{d}u\int_0^1 (1-t)^{a-1}t^{b-1}\mathrm{d}t \overset{\text{déf}}{=} B(a,b)\int_0^{+\infty}\mathrm{e}^{-u}u^{a+b-1}\mathrm{d}u \\ &= B(a,b)\Gamma(a+b)\end{aligned} \tag{7.307}$$

La fonction $B(a,b)$ qui apparaît dans l'expression ci-dessus est appelée Loi bêta et nous avons donc :

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{7.308}$$

Maintenant que nous avons défini ce qu'était la fonction bêta, considérons deux paramètres $a > 0, b > 0$ et considérons la relation particulière ci-dessous comme étant la loi bêta [14] :

$$P_{a,b}(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \tag{7.309}$$

Nous vérifions d'abord que que $P_{a,b}(x)$ est bien une fonction de distribution :

$$\begin{aligned}\int_{-\infty}^{+\infty} P_{a,b}(x)\mathrm{d}x &= \int_{-\infty}^{+\infty}\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}\cdot 1_{]0,1[}\mathrm{d}x = \int_{-\infty}^{+\infty}\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}\cdot 1_{]0,1[}\mathrm{d}x \\ &= \frac{1}{B(a,b)}\int_{-\infty}^{+\infty} x^{a-1}(1-x)^{b-1}\mathrm{d}x = \frac{B(a,b)}{B(a,b)} = 1\end{aligned} \tag{7.310}$$

Maintenant, nous calculons qu'elle est son espérance (moyenne) :

$$\begin{aligned}\mu &= \int_{-\infty}^{+\infty} xP_{a,b}(x)\mathrm{d}x = \frac{1}{B(a,b)}\int_0^1 x^a(1-x)^{b-1}\mathrm{d}x \\ &= \frac{B(a+1,b)}{B(a,b)} = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}\cdot\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{a}{a+b}\end{aligned} \tag{7.311}$$

en utilisant la relation (7.303). Sa variance s'écrit :

$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{+\infty}(x-\mu)^2 f(x)\mathrm{d}x = \frac{1}{B(a,b)}\int_0^1 (x-\mu)^2 x^{a-1}(1-x)^{b-1}\mathrm{d}x \\ &= \frac{1}{B(a,b)}\left(\int_0^1 x^{a+1}(1-x)^{b-1}\mathrm{d}x - 2\mu\int_0^1 x^a(1-x)^{b-1}\mathrm{d}x + \mu^2\int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x\right) \\ &= \frac{B(a+2,b) - 2\mu^2 B(a,b) + \mu^2 B(a+1,b)}{B(a,b)} = \frac{B(a+2,b) - \mu^2 B(a,b)}{B(a,b)}\end{aligned} \tag{7.312}$$

À l'aide des conditions (7.303) et (7.308), nous trouvons :

$$B(a+2,b) = \frac{\mu\cdot(a+1)}{a+b+1} \quad \text{et} \quad \sigma^2 = \frac{\mu\cdot(a+1)}{a+b+1} - \mu^2 = \frac{ab}{(a+b)^2(a+b+1)} \tag{7.313}$$

Dans le cas de la figure 7.20, nous avons :

$$P_{2,3}(x) = 12x\,(1-x)^2 \quad \text{et} \quad P_{2,3}(X \leqslant x) = 3x^4 - 8x^3 + 6x^2 \tag{7.314}$$

14. Il existe plusieurs formulations de la loi bêta dont une très importante qui est étudiée en détails dans le chapitre de Techniques de Gestion.

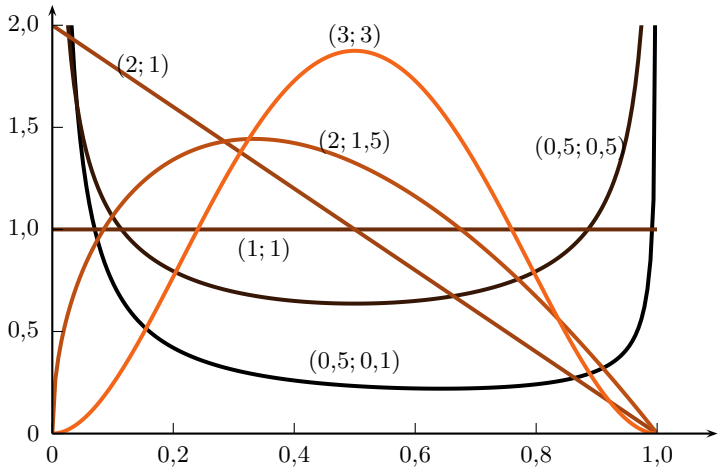


FIGURE 7.19 – Loi bêta pour plusieurs couples $(\beta; \alpha)$

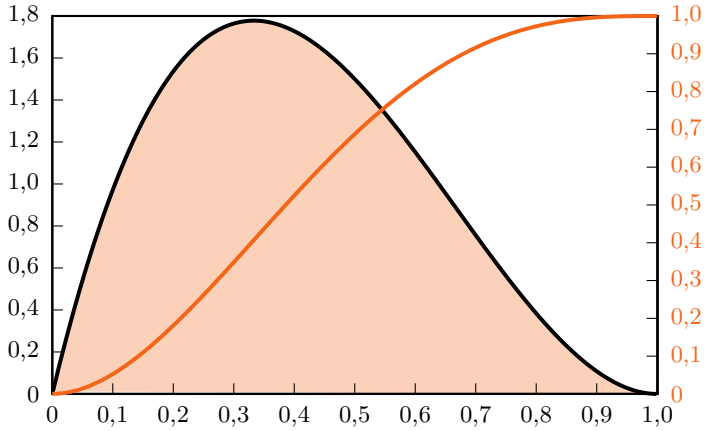


FIGURE 7.20 – Loi bêta pour $(a, b) = (2, 3)$

7.4.16 Loi gamma

La fonction Gamma d'EULER étant connue, considérons deux paramètres $a > 0$ et $\lambda > 0$ et définissons la loi Gamma comme étant donnée par la relation :

$$P_{a,\lambda}(x) = \frac{x^{a-1}\mathrm{e}^{-\lambda x}}{\int_0^{+\infty} x^{a-1}\mathrm{e}^{-\lambda x}\mathrm{d}x} \cdot 1_{]0,+\infty[} \tag{7.315}$$

En faisant le changement de variables $t = \lambda x$ nous obtenons :

$$\int_0^{+\infty} x^{a-1}\mathrm{e}^{-\lambda x}\mathrm{d}x = \frac{\Gamma(a)}{\lambda^a} \tag{7.316}$$

et pouvons alors écrire la relation sous une forme plus classique que nous trouvons fréquemment dans les ouvrages :

$$P_{a,\lambda}(x) = \frac{x^{a-1}\mathrm{e}^{-\lambda x}}{\int_0^{+\infty} x^{a-1}\mathrm{e}^{-\lambda x}\mathrm{d}x} \cdot 1_{]0,+\infty[} \tag{7.317}$$

REMARQUES
1. Si $a = 1$ alors $\Gamma(a) = 1$ et $x^{a-1} = 1$ et nous retombons sur la loi exponentielle.
2. Si $a \in \mathbb{N}$ la distribution s'appelle alors la *fonction d'Erlang*.

Ensuite, nous vérifions avec un raisonnement similaire en tout point celui de la loi bêta que $P_{a,\lambda}(x)$ est une fonction de distribution :

$$\int_{-\infty}^{+\infty} P_{a,\lambda}(x)\mathrm{d}x = 1 \tag{7.318}$$

La fonction Gamma a par ailleurs pour espérance :

$$\mu = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \frac{\lambda^a}{\Gamma(a)}\int_{-\infty}^{+\infty} x^a \mathrm{e}^{-\lambda x}\mathrm{d}x = \frac{\lambda^a}{\Gamma(a)}\frac{\Gamma(a+1)}{\lambda^{a+1}}\frac{\lambda^a}{\Gamma(a)}\frac{a\Gamma(a)}{\lambda^{a+1}} = \frac{a}{\lambda} \tag{7.319}$$

et pour variance :

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\mathrm{d}x = \frac{\lambda^a}{\Gamma(a)}\int_{-\infty}^{+\infty} (x-\mu)^2 x^{a-1}\mathrm{e}^{-\lambda x}\mathrm{d}x \\
&= \frac{\lambda^a}{\Gamma(a)}\left(\int_{-\infty}^{+\infty} x^{a+1}\mathrm{e}^{-\lambda x}\mathrm{d}x - 2\mu\int_{-\infty}^{+\infty} x^a\mathrm{e}^{-\lambda x}\mathrm{d}x + \mu^2\int_{-\infty}^{+\infty} x^{a-1}\mathrm{e}^{-\lambda x}\mathrm{d}x\right) \\
&= \frac{\lambda^a}{\Gamma(a)}\left(\frac{\Gamma(a+2)}{\lambda^{a+2}} + \mu^2\frac{\Gamma(a)}{\lambda^a} - 2\mu\frac{\Gamma(a+1)}{\lambda^{a+1}}\right) \\
&= \frac{1}{\lambda^2\Gamma(a)}\left(\Gamma(a+2) + a^2\Gamma(a) - 2a\Gamma(a+1)\right) \\
&= \frac{1}{\lambda^2\Gamma(a)}\left((a+1)a\Gamma(a) + a^2\Gamma(a) - 2a^2\Gamma(a)\right) = \frac{a}{\lambda^2}
\end{aligned} \tag{7.320}$$

Démontrons une propriété de la fonction Gamma qui nous servira à démontrer plus tard dans ce

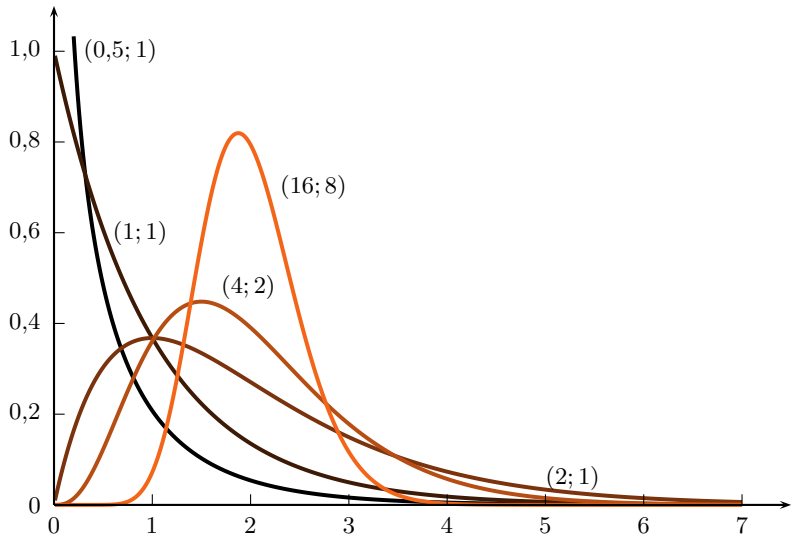


FIGURE 7.21 – Distributions de la loi Gamma pour différents couples (a, λ)

chapitre lors de notre étude de l'analyse de la variance et des intervalles de confiance sur des petits échantillons une autre propriété extrêmement importante de la loi du khi-deux.

Quand une variable aléatoire suite une fonction Gamma nous la notons :

$$X = \gamma(a, \lambda) \tag{7.321}$$

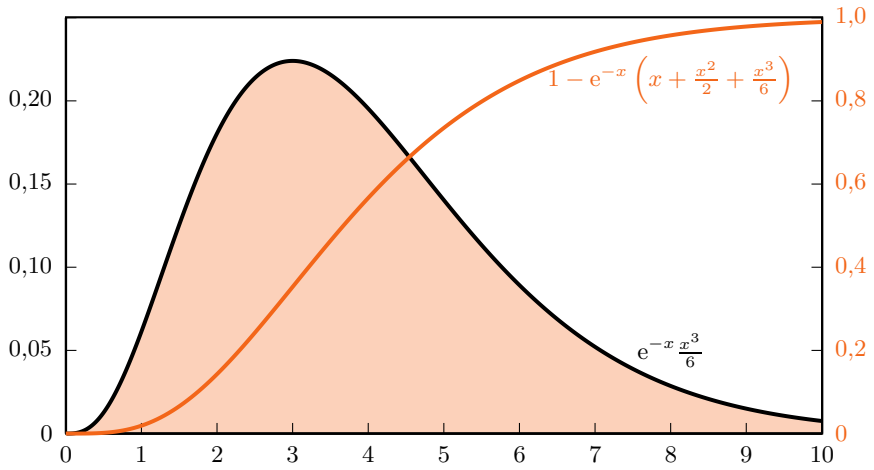


FIGURE 7.22 – Loi Gamma pour $(a, \lambda) = (4, 1)$

Soit X, Y deux variables indépendantes. Montrons que si $X = \gamma(p, \lambda)$ et $X = \gamma(q, \lambda)$ alors :

$$X + Y = \gamma(p + q, \lambda) \tag{7.322}$$

Notons f la fonction de densité du couple (X, Y), f_X la fonction de densité de X et f_Y la fonction de densité de Y. Vu que X, Y sont indépendantes, nous avons :

$$f(x, y) = f_X(x) \cdot f_Y(y) \tag{7.323}$$

pour tout $x, y \geqslant 0$.

Soit $Z = X + Y$. La fonction de répartition de Z est alors :

$$F(z) = P(Z \leqslant z) = P(X + Y \leqslant z) = \iint_D f(x, y)\mathrm{d}x\mathrm{d}y \tag{7.324}$$

où $D\{(x, y)|x + y \leqslant z\}$.

REMARQUE Nous appelons un tel calcul une *convolution* et les statisticiens ont souvent à manipuler de telles entités ayant à travailler sur des nombreuses variables aléatoires qu'il faut sommer ou même multiplier.

En simplifiant :

$$F(z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \tag{7.325}$$

Nous effectuons le changement de variable $x = x$ et $y = s - x$ dont le jacobien est (cf. chapitre de Calcul Différentiel Et Intégral) :

$$J = \begin{vmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial s} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial s} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 0 \end{vmatrix} = 1 \tag{7.326}$$

Donc avec la nouvelle borne d'intégration $s = x + y = x + (z - x) = z$ nous avons :

$$F(z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{z} f_X(x)f_Y(s - x)\mathrm{d}s\mathrm{d}x = \int_{-\infty}^{z}\int_{-\infty}^{+\infty} f_X(x)f_Y(s - x)\mathrm{d}x\mathrm{d}s \tag{7.327}$$

Si nous notons g la fonction de densité de Z nous avons :

$$F(z) = \int_{-\infty}^{z}\int_{-\infty}^{+\infty} f_X(x)f_Y(s - x)\mathrm{d}x\mathrm{d}s = \int_{-\infty}^{z} g(s)\mathrm{d}s \tag{7.328}$$

Par suite :

$$g(s) = \int_{-\infty}^{+\infty} f_X(x) f_Y(s-x) \mathrm{d}x \tag{7.329}$$

f_X et f_Y étant nulles lorsque leur argument est négatif, nous pouvons changer les bornes d'intégration :

$$g(s) = \int_{0}^{s} f_X(x) f_Y(s-x) \mathrm{d}x \quad \text{pour} \quad s \geqslant 0 \tag{7.330}$$

Calculons g :

$$g(s) = \frac{\lambda^{p+q} \mathrm{e}^{-\lambda s}}{\Gamma(p)\Gamma(q)} \int_{0}^{s} x^{p-1}(s-x)^{q-1} \mathrm{d}x \tag{7.331}$$

Après le changement de variable $x = st$ nous obtenons :

$$g(s) = \frac{\lambda^{p+q} \mathrm{e}^{-\lambda s}}{\Gamma(p)\Gamma(q)} s^{p+q-1} \int_{0}^{1} t^{p-1}(1-t)^{q-1} \mathrm{d}t = \frac{\lambda^{p+q} \mathrm{e}^{-\lambda s}}{\Gamma(p)\Gamma(q)} s^{p+q-1} B(p,q) \tag{7.332}$$

où B est la fonction bêta que nous avons vu plus haut dans notre étude la fonction de distribution bêta. Or nous avons aussi démontré la relation :

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \tag{7.333}$$

Donc :

$$B(p,q) = \frac{\lambda^{p+q} \mathrm{e}^{-\lambda s}}{\Gamma(p+q)} s^{p+q-1} \tag{7.334}$$

Ce qui finalement nous donne :

$$P(Z \leqslant z) = P(X+Y \leqslant z) = \int_{0}^{s} g(s) \mathrm{d}s = \int_{0}^{s} \frac{\lambda^{p+q} \mathrm{e}^{-\lambda s}}{\Gamma(p+q)} s^{p+q-1} \mathrm{d}s \tag{7.335}$$

Ce qui montre que bien que si deux variables aléatoires suivent une fonction Gamma alors leur somme aussi tel que :

$$X + Y = \gamma(p+q, \lambda) \tag{7.336}$$

donc la fonction Gamma est stable par addition de même que le sont toutes les lois qui découlent de la loi gamma et que nous allons aborder ci-après.

7.4.17 Loi de khi-deux

La loi de Khi-Deux, ou encore loi de PEARSON, n'est qu'un cas particulier de la fonction de distribution Gamma dans le cas où $a = k/2$ et $\lambda = 1/2$, avec k entier positif :

$$P_k(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(k/2)} x^{\frac{k}{2}-1} \mathrm{e}^{-\frac{x}{2}} \cdot 1_{[0,+\infty[} \tag{7.337}$$

Tous les calculs faits auparavant s'appliquent et nous avons alors immédiatement $\mu = k$ et $\sigma^2 = 2k$.
Dans la littérature, il est de tradition de noter $X = \chi^2_k$ ou $X = \chi^2(k)$ pour indiquer que la distribution de la variable aléatoire X est la loi du khi-deux. Par ailleurs il est courant de nommer

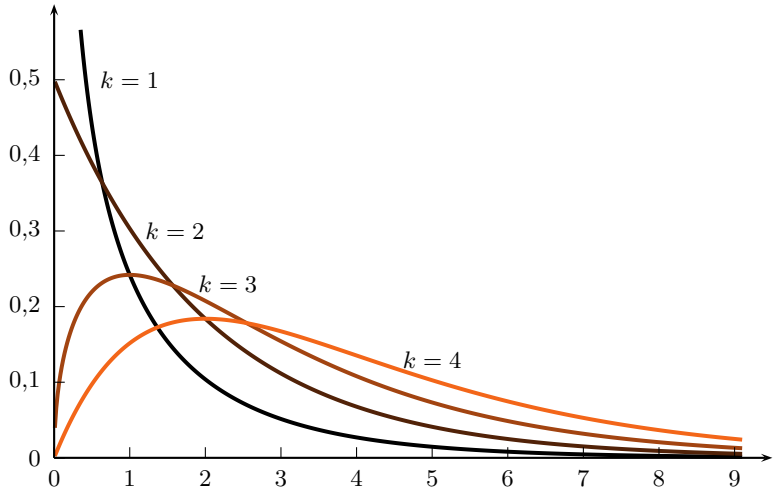


FIGURE 7.23 – Distributions de loi khi-deux

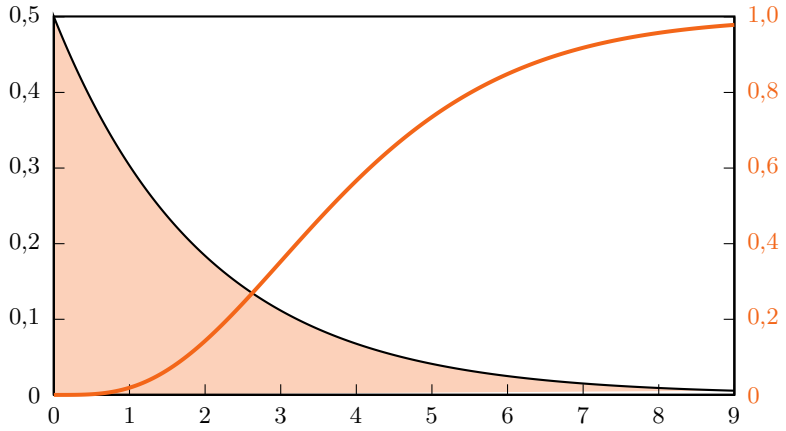


FIGURE 7.24 – Loi khi-deux pour $k = 2$

le paramètre k *degré de liberté* et de l'abréger *ddl*. La fonction khi-deux découle donc de la loi gamma et par ailleurs en prenant $k = 2$, nous retrouvons aussi la loi exponentielle pour $\lambda = 1/2$:

$$P_2(x) = \frac{1}{2\Gamma(1)} \mathrm{e}^{-x/2} \cdot 1_{[0,+\infty[} = \frac{1}{2} \mathrm{e}^{-x/2} \cdot 1_{[0,+\infty[} \tag{7.338}$$

Par ailleurs, puisque (cf. chapitre de Calcul Différentiel Et Intégral) $\Gamma(1/2) = \sqrt{\pi}$ la loi du khi-deux avec k égal à l'unité peut s'écrire sous la forme :

$$P_1(x) = \frac{1}{2^{1/2}\Gamma(1/2)} x^{1/2-1} \mathrm{e}^{-x/2} \cdot 1_{[0,+\infty[} = \frac{1}{\sqrt{2\pi x}} \mathrm{e}^{-x/2} \cdot 1_{[0,+\infty[} \tag{7.339}$$

7.4.18 Fonction de student

La loi de Student de paramètre k est définie par la relation :

$$P(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} \tag{7.340}$$

avec k étant le degré de liberté de la loi du khi-deux sous jacente à la construction de la fonction de Student comme nous allons le voir.

Il s'agit bien d'une fonction de distribution car elle vérifie également :

$$\int_{-\infty}^{+\infty} P_k(x)\mathrm{d}x = 1 \tag{7.341}$$

Voyons la démonstration la plus simple pour justifier la provenance de la loi de Student et qui nous sera en même temps très utile dans l'inférence statistique et l'analyse de la variance plus loin. Tout d'abord, rappelons que :

1. Si X, Y sont deux variables aléatoires indépendantes de densités respectives f_X, f_Y, la loi du couple (X,Y) possède une densité f vérifiant $f(x,y) = f_X(x)f_Y(x)$
2. La loi $N(0,1)$ est donnée par :
$$f(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \tag{7.342}$$
3. La loi χ^2_n est donnée par :
$$f(y) = \frac{1}{2^{n/2}\Gamma(n/2)}y^{n/2-1}\mathrm{e}^{-y/2} \tag{7.343}$$
pour $y \geqslant 0$ et $n \geqslant 1$.
4. La fonction Γ• est définie pour tout $\alpha > 0$ par :
$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x \tag{7.344}$$
et vérifie (cf. chapitre de Calcul Différentiel et Intégral) :
$$\Gamma(\alpha-1) = \frac{\Gamma(\alpha)}{\alpha-1} \tag{7.345}$$
pour $\alpha \geqslant 2$.

Sec13.391|p370

Ces rappels étant faits, considérons maintenant X une variable aléatoire suivant la loi $N(0,1)$ et Y une variable aléatoire suivant la loi χ^2_n.

Nous supposons X et Y indépendantes et nous considérons la variable aléatoire[15] :

$$T = \sqrt{n}\frac{X}{\sqrt{Y}} = \frac{X}{\sqrt{Y/n}} \tag{7.346}$$

Nous allons montrer que T suit une loi de Student de paramètre n.

Preuve Notons F et f les fonctions de répartition et de densité de T et f_X, f_Y, f, les fonctions de densité de X, Y et (X,Y) respectivement. Nous avons alors pour tout $t \in \mathbb{R}$:

$$F(t) = P(T \leqslant t) = P\left(\sqrt{n}\frac{X}{\sqrt{Y}} \leqslant t\right) = \iint_D f(x,y)\mathrm{d}x\mathrm{d}y = \iint_D f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \tag{7.347}$$

où $D = \{(x,y) \in \mathbb{R}\times\mathbb{R}^* | x \leqslant t\sqrt{y}/\sqrt{n}\}$. La valeur imposée positive et non nulle de y étant due au fait qu'elle est sous une racine et en plus au dénominateur. Ainsi :

$$F(t) = \int_0^{+\infty} f_Y(y)\mathrm{d}y \int_{-\infty}^{t\sqrt{y}/\sqrt{n}} f_X(x)\mathrm{d}x = \int_0^{+\infty} f_Y(y)\Phi(t\sqrt{y}/\sqrt{n})\mathrm{d}y \tag{7.348}$$

où, puisque X suit une loi $N(0,1)$:

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}\mathrm{d}x \tag{7.349}$$

15. C'est à l'origine l'étude historique de la loi de Student dans le cadre de l'inférence statistique qui a amené à poser cette variable dont nous justifierons l'origine plus loin.

qui est la fonction de répartition de la loi Normale centrée réduite. Nous obtenons alors la fonction de densité de T en dérivant F :

$$f(t) = F'(t) = \frac{1}{\sqrt{n}} \int_0^{+\infty} f_Y(y) f_X(t\sqrt{y}/\sqrt{n})\sqrt{y}\mathrm{d}y \tag{7.350}$$

donc :

$$\begin{aligned} f(t) &= \frac{1}{\sqrt{n}} \int_0^{+\infty} \frac{1}{2^{n/2}\Gamma(n/2)} \mathrm{e}^{-y/2} y^{n/2-1} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(t\sqrt{y}/\sqrt{n})^2/2} \sqrt{y}\mathrm{d}y \\ &= \frac{1}{2^{n/2}\Gamma(n/2)\sqrt{2\pi n}} \int_0^{+\infty} \mathrm{e}^{-\frac{y(1+t^2/n)}{2}} y^{n/2-1}\mathrm{d}y \end{aligned} \tag{7.351}$$

En faisant le changement de variable :

$$u = \frac{y}{2}\left(1 + \frac{t^2}{n}\right) \quad \Rightarrow \quad \mathrm{d}y = \frac{2}{1+\frac{t^2}{n}}\mathrm{d}u \tag{7.352}$$

nous obtenons :

$$\begin{aligned} f(t) &= \frac{1}{2^{n/2}\Gamma(n/2)\sqrt{2\pi n}} \left(\frac{2}{1+t^2/n}\right)^{(n+1)/2} \int_0^{+\infty} \mathrm{e}^{-u} u^{(n+1)/2}\mathrm{d}u \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)\sqrt{\pi n}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \end{aligned} \tag{7.353}$$

ce qui est bien la loi de Student de paramètre n. •

Voyons maintenant quelle est l'espérance de la loi de Student. Nous avons :

$$E(T) = E(\sqrt{n}X)E\left(\frac{1}{\sqrt{Y}}\right) \tag{7.354}$$

Mais $E\left(1/\sqrt{Y}\right)$ existe si et seulement si $n \geqslant 2$. En effet :

$$E\left(\frac{1}{\sqrt{Y}}\right) = \int_0^{+\infty} \frac{f_Y(y)}{\sqrt{y}}\mathrm{d}y = \frac{1}{2^{\frac{n}{2}}\Gamma(n/2)} \int_0^{+\infty} \mathrm{e}^{-\frac{y}{2}} y^{\frac{n-3}{2}}\mathrm{d}y \tag{7.355}$$

et pour $n = 1$:

$$\int_0^{+\infty} \mathrm{e}^{-\frac{y}{2}} y^{\frac{1-3}{2}}\mathrm{d}y = \int_0^{+\infty} \frac{\mathrm{e}^{-\frac{y}{2}}}{y}\mathrm{d}y \geqslant \int_0^{1} \frac{\mathrm{e}^{-\frac{y}{2}}}{y}\mathrm{d}y \geqslant \mathrm{e}^{-\frac{1}{2}} \int_0^{+\infty} \frac{1}{y}\mathrm{d}y \to +\infty \tag{7.356}$$

et, par conséquent, l'espérance n'existe pas. Par contre, pour $n = 2$ nous avons :

$$\int_0^{+\infty} \mathrm{e}^{-\frac{y}{2}} y^{\frac{n-3}{2}}\mathrm{d}y = 2^{\frac{n-1}{2}} \int_0^{+\infty} \mathrm{e}^{-u} u^{\frac{n-3}{2}}\mathrm{d}u = 2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right) < +\infty \tag{7.357}$$

soit :

$$E(T) = \sqrt{n}\underbrace{E(X)}_{=0} E\left(\frac{1}{\sqrt{Y}}\right) = 0 \tag{7.358}$$

Voyons maintenant la valeur de la variance :

$$V(T) = E(T^2) - E(T)^2 \tag{7.359}$$

Discutons de l'existence de $E(T^2)$. Nous avons trivialement :

$$E(T^2) = nE\left(\frac{X^2}{Y}\right) = nE(X^2)E\left(\frac{1}{Y}\right) \tag{7.360}$$

or X suit une loi normale centrée réduite donc :

$$V(X) = 1 = E(X^2) - E(X)^2 = E(X^2) \quad \Rightarrow \quad E(X^2) = 1 \tag{7.361}$$

Pour ce qui est de $E(1/Y)$, nous avons :

$$E\left(\frac{1}{Y}\right) = \int_0^{+\infty} \frac{f_Y(y)}{y}\mathrm{d}y = \frac{1}{2^{\frac{n}{2}}\Gamma(n/2)} \int_0^{+\infty} \mathrm{e}^{-\frac{y}{2}} y^{\frac{n}{2}-2}\mathrm{d}y = \frac{\Gamma\left(\frac{n}{2}-1\right)}{2\Gamma(n/2)} \tag{7.362}$$

Mais l'intégrale définissant $\Gamma(n/2-1)$ converge seulement si $n \geqslant 3$, donc $E(T^2)$ existe si et seulement si $n \geqslant 3$ et vaut alors selon les propriétés de la loi Gamma d'EULER démontrées dans le chapitre de Calcul Différentiel et Intégral :

$$nE\left(T\right) = n\frac{\Gamma\left(\frac{n}{2}-1\right)}{2\Gamma(n/2)} = \frac{n}{n-2} \tag{7.363}$$

Ainsi pour $n \geqslant 3$:

$$V(T) = \frac{n}{n-2} \tag{7.364}$$

Il est par ailleurs important de remarque que cette loi est symétrique par rapport à 0. Dans le cas

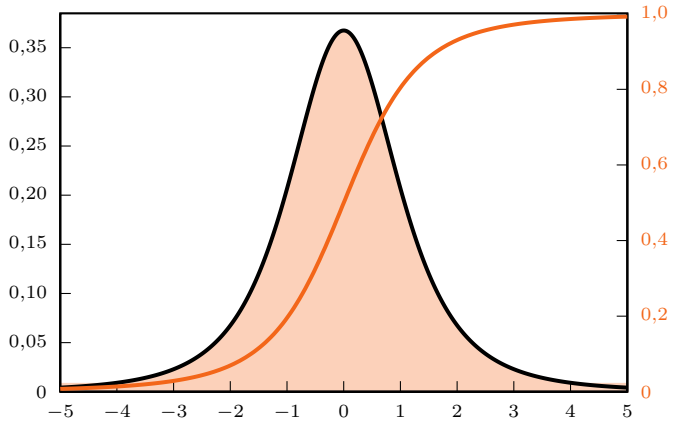


FIGURE 7.25 – Loi Student pour $k = 3$

de la figure 7.25 :

$$\begin{aligned} P(x) &= \frac{2}{\sqrt{3}\pi\left(1+1/3x^2\right)^2} \\ Q(x) &= \frac{3\pi + \pi x^2 + 2x\sqrt{3} + 6\arctan\left(1/3x\sqrt{3}\right) + 2\arctan\left(1/3x\sqrt{3}\right)x^2}{2\pi\left(3+x^2\right)} \end{aligned} \tag{7.365}$$

7.4.19 Fonction de FISHER

La loi de FISHER-SNEDECOR de paramètres k et l est définie par la relation :

$$P(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)\left(\frac{k}{l}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{l}{2}\right)} x^{\frac{k}{2}-1}\left(1+\frac{kx}{l}\right)^{-\frac{k+l}{2}} \tag{7.366}$$

si $x \geqslant 0$. Les paramètres k et l sont des entiers positifs et correspondent aux degrés de liberté des deux lois du khi-deux sous-jacentes. Cette distribution est souvent notée $F_{k,l}$ ou $F(k,l)$.

Il s'agit bien d'une fonction de distribution car elle vérifie également :

$$\int_0^{+\infty} F_{k,l}(x)\mathrm{d}x = 1 \tag{7.367}$$

Voyons la démonstration la plus simple pour justifier la provenance de la loi de Fisher et qui nous sera en même temps très utile dans l'inférence statistique et l'analyse de la variance.

Preuve Rappelons que :

1. La loi χ^2_n est donnée par l'équation (7.337) pour $y \geqslant 0$ et $n \geqslant 1$.
2. La fonction Γ est définie pour tout $\alpha > 0$ par l'équation (13.391).

Soient X, Y deux variables aléatoires indépendantes suivant respectivement les lois χ^2_n et χ^2_m. Nous considérons la variable aléatoire :

$$F = \frac{X/n}{Y/m} \tag{7.368}$$

Nous allons donc montrer que la loi de T est la loi de Fisher-Snedecor de paramètres n, m.

Notons pour cela F et f les fonctions de répartition et de densité de T et f_X, f_Y, f les fonctions de densité de X, Y et (X,Y) respectivement. Nous avons pour tout $t \in \mathbb{R}$:

$$F(t) = P(T \leqslant t) = P\left(\frac{X/n}{Y/m} \leqslant t\right) = \iint_D f(x,y)\mathrm{d}x\mathrm{d}y = \iint_D f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \tag{7.369}$$

où $D = \{(x,y) \in \mathbb{R}^* \times \mathbb{R}^* | x \leqslant nt/my\}$ et les valeurs positives imposées proviennent de l'origine d'une loi du khi-deux pour x et y. Ainsi :

$$F(t) = \int_0^{+\infty} f_Y(y)\mathrm{d}y \left[\int_0^{\frac{nty}{m}} f_X(x)\mathrm{d}x\right] \mathrm{d}y \tag{7.370}$$

Nous obtenons la fonction de densité de T en dérivant F. D'abord la dérivée intérieure :

$$f(t) = F'(t) = \frac{n}{m}\int_0^{+\infty} f_Y(y) f_X\left(\frac{nty}{m}y\right) y\mathrm{d}x\mathrm{d}y \tag{7.371}$$

Ensuite en explicitant puisque :

$$f(y) = \frac{1}{2^{m/2}\Gamma(m/2)}\mathrm{e}^{-y/2}y^{m/2-1} \quad \text{et} \quad f\left(\frac{nty}{m}y\right) = \frac{1}{2^{n/2}\Gamma(n/2)}\mathrm{e}^{-\frac{nt}{2m}y}\left(\frac{nt}{m}y\right)^{n/2-1} \tag{7.372}$$

nous avons alors :

$$\begin{aligned}
f(t) &= \frac{n}{m}\int_0^{+\infty} \frac{1}{2^{m/2}\Gamma(m/2)}\mathrm{e}^{-y/2}y^{\frac{m}{2}-1}\frac{1}{2^{\frac{n}{2}}\Gamma(n/2)}\mathrm{e}^{-\frac{nt}{2m}y}\left(\frac{nt}{m}y\right)^{\frac{n}{2}-1} y\mathrm{d}y \\
&= \frac{1}{2^{n/2}\Gamma(n/2)2^{m/2}\Gamma(m/2)}\frac{n}{m}\int_0^{+\infty} \mathrm{e}^{-y/2}y^{m/2-1}\mathrm{e}^{-\frac{nt}{2m}y}\left(\frac{nt}{m}y\right)^{\frac{n}{2}-1} y\mathrm{d}y \\
&= \frac{1}{2^{n/2}\Gamma(n/2)2^{m/2}\Gamma(m/2)}\frac{n}{m}\left(\frac{nt}{m}\right)^{n/2-1}\int_0^{+\infty} \mathrm{e}^{-y/2}y^{m/2-1}\mathrm{e}^{-\frac{nt}{2m}y}y^{n/2-1}y\mathrm{d}y \\
&= \frac{1}{2^{\frac{n+m}{2}}\Gamma(n/2)\Gamma(m/2)}\frac{n}{m}\left(\frac{nt}{m}\right)^{\frac{n}{2}-1}\int_0^{+\infty} y^{\frac{n+m}{2}-1}\mathrm{e}^{-\frac{y}{2}\left(1+\frac{nt}{m}\right)}\mathrm{d}y
\end{aligned} \tag{7.373}$$

En faisant le changement de variable :

$$u = \frac{y}{2}\left(1+\frac{nt}{m}\right) \Rightarrow y = \frac{2u}{1+\frac{nt}{m}} \Rightarrow \mathrm{d}y = \frac{2\mathrm{d}u}{1+\frac{nt}{m}} \tag{7.374}$$

nous obtenons :

$$\begin{aligned} f(t) &= \frac{n}{m}\frac{1}{2^{\frac{n+m}{2}}\Gamma(n/2)\Gamma(m/2)}\left(\frac{nt}{m}\right)^{\frac{n}{2}-1}\left(\frac{2}{1+\frac{nt}{m}}\right)^{\frac{n+m}{2}}\underbrace{\int_0^{+\infty} u^{\frac{n+m}{2}-1}\mathrm{e}^{-u}\mathrm{d}u}_{\Gamma\left(\frac{n+m}{2}\right)} \\ &= \frac{\frac{n}{m}\Gamma\left(\frac{n+m}{2}\right)}{\Gamma(n/2)\Gamma(m/2)}\left(\frac{nt}{m}\right)^{\frac{n}{2}-1}\left(1+\frac{nt}{m}\right)^{-\frac{n+m}{2}} \\ &= \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}}\Gamma\left(\frac{n+m}{2}\right)}{\Gamma(n/2)\Gamma(m/2)}t^{\frac{n}{2}-1}\left(1+\frac{nt}{m}\right)^{-\frac{n+m}{2}} \end{aligned} \tag{7.375}$$

•

7.4.20 Loi de Benford

Cette distribution aurait été découverte une première fois en 1881 par Simon Newcomb, un astronome américain, après qu'il se fut aperçu de l'usure (et donc de l'utilisation) préférentielle des premières pages des tables de logarithmes (alors compilées dans des ouvrages). Frank Benford, aux alentours de 1938, remarqua à son tour cette usure inégale, crut être le premier à formuler cette loi qui porte indûment son nom aujourd'hui et arriva aux même résultats après avoir répertorié des dizaines de milliers de données (longueurs de fleuves, cours de la bourse, etc).

Seule explication possible : nous avons plus souvent besoin d'extraire le logarithme de chiffres commençant par 1 que de chiffres commençant par 9, ce qui implique que les premiers sont *plus nombreux* que les seconds.

Bien que cette idée lui paraisse tout à fait invraisemblable, Benford entreprend de vérifier son hypothèse. Rien de plus simple : il se procure des tables de valeurs numériques, et calcule le pourcentage d'apparition du chiffre le plus à gauche (première décimale). Les résultats qu'il obtient confirment son intuition et sont exposés dans le tableau 7.4. À partir de ces données, Benford trouve

chiffre initial	1	2	3	4	5	6	7	8	9
probabilité d'apparition (%)	30,1	17,6	12,5	9,7	7,9	6,7	5,8	5,1	4,6

TABLEAU 7.4 – (7.470)

expérimentalement que la probabilité qu'un nombre commence par le chiffre n (excepté 0) est (nous allons le démontrer plus loin) donnée par la relation :

$$P(n) = \log_{10}\left(1+\frac{1}{n}\right) \tag{7.376}$$

appelée loi de Benford et illustrée sur la figure 7.26. Il convient de préciser que cette loi ne s'applique

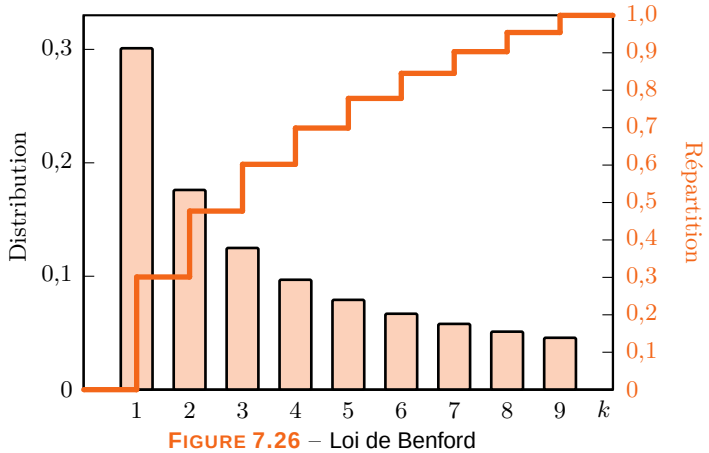


FIGURE 7.26 – Loi de Benford

qu'à des listes de valeurs *naturelles*, c'est-à-dire à des chiffres ayant une signification physique. Elle ne fonctionne évidemment pas sur une liste de chiffres tirés au hasard.

La loi de Benford a été testée sur toutes sortes de tables : longueur des fleuves du globe, superficie des pays, résultat des élections, liste des prix de l'épicerie du coin. Elle se vérifie à presque tous les coups.

Elle est évidemment indépendante de l'unité choisie. Si l'on prend par exemple la liste des prix d'un supermarché, elle fonctionne aussi bien avec les valeurs exprimées en Francs qu'avec les mêmes prix convertis en Euros.

Cet étrange phénomène est resté peu étudié et inexpliqué jusqu'à une époque assez récente. Puis une démonstration générale en a été donnée en 1996, qui fait appel au théorème de la limite centrale.

Aussi surprenant que cela puisse paraître, cette loi a trouvé une application : le fisc l'utilise aux États-Unis pour détecter les fausses déclarations. Le principe est basé sur la restriction vue plus haut : la loi de Benford ne s'applique que sur des valeurs ayant une signification physique.

S'il existe une distribution de probabilité universelle $P(n)$ sur de tels nombres, ils doivent êtres invariants sous un changement d'échelle tel que :

$$P(kn) = f(k)P(n) \tag{7.377}$$

Si :

$$\int P(n)\mathrm{d}n = 1 \tag{7.378}$$

alors :

$$\int P(kn)\mathrm{d}n = \frac{1}{k} \tag{7.379}$$

et la normalisation de la distribution donne :

$$f(k) = \frac{1}{k} \tag{7.380}$$

si nous dérivons $P(kn) = f(k)P(n)$ par rapport à k nous obtenons :

$$\frac{\mathrm{d}}{\mathrm{d}k}P(kn) = \frac{\mathrm{d}}{\mathrm{d}k}\frac{1}{k}P(n) \Rightarrow nP'(kn) = P(n)\frac{\mathrm{d}}{\mathrm{d}k}\frac{1}{k} \Rightarrow nP'(kn) = -P(n)\frac{1}{k^2} \tag{7.381}$$

soit, en posant $k = 1$:

$$nP'(n) = -P(n) \tag{7.382}$$

Cette équation différentielle a pour solution :

$$P(n) = \frac{1}{n} \tag{7.383}$$

Cette fonction, n'est pas en premier lieu à proprement parler une fonction de distribution de probabilité (elle diverge) et deuxièmement, les lois de la physique et humaines imposent des limites.

Nous devons donc comparer cette distribution par rapport à une référence arbitraire. Ainsi, si le nombre décimal étudié contient plusieurs puissance de 10 (10 au total : $0, 1, 2, 3, 4, 5, 6, 7, 9$) la probabilité que le premier chiffre non nul (décimal) soit D est donné par la distribution logarithmique :

$$P_D = \frac{\int_D^{D+1} P(n)\mathrm{d}n}{\int_1^{10} P(n)\mathrm{d}n} \tag{7.384}$$

L'intégrale est calculée de 1 à 10 puisque la valeur nulle est interdite.

L'intégrale du dénominateur donne :

$$\int_1^{10} P(n)\mathrm{d}n = \int_1^{10} \frac{n}{\mathrm{d}n} = \ln(10) - \ln(1) = \ln(10) \tag{7.385}$$

L'intégrale du numérateur donne :

$$\int_{D}^{D+1} P(n)\mathrm{d}n = \int_{D}^{D+1} \frac{n}{\mathrm{d}n} = \ln(D+1) - \ln(D) = \ln\left(\frac{D+1}{D}\right) \tag{7.386}$$

Ce qui nous donne finalement, du fait des propriétés des logarithmes (Analyse fonctionnelle) :

$$P_D = \frac{\ln\left(1+\frac{1}{D}\right)}{\ln 10} = \log_{10}\left(1+\frac{1}{D}\right) \tag{7.387}$$

Cependant, la loi de Benford ne s'applique pas uniquement aux données invariantes par changement d'échelle mais également à des nombres de provenant de sources quelconques. Expliquer ce cas implique une investigation plus rigoureuse en utilisant le théorème de la limite centrale. Cette démonstration a été effectuée seulement en 1996 par T. Hill par une approche utilisant la distribution des distributions.

7.5 Estimateurs de vraisemblance

Ce qui va suivre est d'une extrême importance en statistiques et est utilisé énormément en pratique. Il convient donc d'y accorder une attention toute particulière.

Nous supposons que nous disposons d'observations $x_1, x_2, x_3, \ldots, x_n$ qui sont des réalisations de variables aléatoires non biaisées (dans le sens qu'elles sont choisies aléatoirement parmi un lot) indépendantes $X_1, X_2, X_3, \ldots, X_n$ de loi de probabilité inconnue mais identique.

Nous allons chercher à estimer cette loi de probabilité P inconnue à partir des observations $x_1, x_2, x_3, \ldots, x_n$.

Supposons que nous procédons par tâtonnement pour estimer la loi de probabilité P inconnue. Une manière de procéder est de se demander si les observations $x_1, x_2, x_3, \ldots, x_n$ avaient une probabilité élevée ou non de sortir avec cette loi de probabilité arbitraire P.

Nous devons pour cela calculer la probabilité conjointe qu'avaient les observations $x_1, x_2, x_3, \ldots, x_n$ de sortir avec $p_1, p_2, p_3, \ldots, p_m$. Cette probabilité vaut (cf. chapitre de Probabilités) :

$$\prod_{i=1}^{n} P(X_i = x_i) \tag{7.388}$$

en notant P la loi de probabilité supposée associée à $p_1, p_2, p_3, \ldots, p_m$. Il faut avouer qu'il serait alors particulièrement maladroit de choisir une loi de probabilité (avec ses paramètres.) qui minimise cette quantité...

Au contraire, nous allons chercher la probabilité $p_1, p_2, p_3, \ldots, p_m$ qui maximise (7.388), c'est-à-dire qui rende les observations $x_1, x_2, x_3, \ldots, x_n$ le plus vraisemblable possible.

Nous sommes donc amené à chercher le (les) paramètre(s) θ qui maximise(nt) la quantité :

$$L_n(\theta) = \prod_{i=1}^{n} P_\theta(X_i = x_i) \tag{7.389}$$

Cette quantité L porte le nom de *vraisemblance*. C'est une fonction du ou des paramètres θ et des observations $x_1, x_2, x_3, \ldots, x_n$.

La ou les valeurs du paramètre θ qui maximisent la vraisemblance $L_n(\theta)$ sont appelées *estimateurs du maximum de vraisemblance* (estimateur MV).

Examinons quand même trois petits exemples très classiques avec dans l'ordre d'importance (donc pas forcément dans l'ordre de facilité...) la fonction de distribution de GAUSS-LAPLACE (Normale), la fonction de distribution de POISSON et finalement Binomiale.

REMARQUE Ces trois exemples sont importants car utilisés dans les SPC (maîtrise statistiques de processus) dans différentes multinationales à travers le monde (cf. chapitre de Génie Industriel).

7.5.1 Estimateurs de la loi normale

Soit $x_1, x_2, x_3, \ldots, x_n$ un n-échantillon de variables aléatoires identiquement distribuées supposées suivre une loi de Gauss-Laplace (loi Normale) de paramètres μ et σ^2.

Nous recherchons quelles sont les valeurs des estimateurs de maximum de vraisemblance θ qui maximisent la vraisemblance $L_n(\theta)$ de la loi Normale ?

Remarque Il va de soit que les estimateurs de maximum de vraisemblance θ sont ici :

$$\theta = (\mu, \sigma^2) \tag{7.390}$$

Nous avons démontré plus haut que la densité d'une variable aléatoire gaussienne était donnée par :

$$P(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{7.391}$$

La vraisemblance est alors donnée par :

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} P(x_i, \mu, \sigma) = \frac{1}{\sigma^n \sqrt{2\pi}^n} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{7.392}$$

Maximiser une fonction ou maximiser son logarithme est équivalent et par conséquent la *log-vraisemblance* sera :

$$\ln[L(\mu, \sigma^2)] = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{7.393}$$

Pour déterminer les deux estimateurs de la loi Normale, fixons d'abord l'écart-type. Pour cela, dérivons l'équation (7.393) par rapport à μ et regardons pour quelle valeur de la moyenne la fonction s'annule. Il nous reste après simplification :

$$\sum_{i=1}^{n}(x_i - \mu) = 0 \tag{7.394}$$

Ainsi, l'estimateur de maximum de vraisemblance de la moyenne (espérance) de la loi Normale est donc après réarrangement :

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7.395}$$

et nous voyons qu'il s'agit simplement de la moyenne arithmétique [16]. Fixons maintenant la moyenne. L'annulation de la dérivée de l'équation (7.393) en σ conduit à :

$$\frac{\partial}{\partial\sigma}\ln[L(\mu, \sigma^2)] = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \tag{7.396}$$

Ce qui nous permet d'écrire l'estimateur de maximum de vraisemblance pour l'écart-type (la variance lorsque la moyenne est connue selon la loi de distribution supposée elle aussi connue.) :

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \tag{7.397}$$

Cependant, nous n'avons pas encore défini ce qu'était un bon estimateur. Ce que nous entendons par là :

- si l'espérance d'un estimateur est égale à elle-même, nous disons que cet estimateur est *sans biais* et c'est bien évidemment ce que nous cherchons ;

16. Cette moyenne prend parfois le nom de *moyenne empirique*.

– si l'espérance d'un estimateur n'est pas égale à elle-même, nous disons alors que cet estimateur est *biaisé* et c'est forcément moins bien.

Dans l'exemple précédent, la moyenne est donc non biaisée (trivial car la moyenne de la moyenne arithmétique est égale à elle même). Mais qu'en est-il de la variance (in extenso de l'écart-type) ?

Un petit calcul simple par linéarité de l'espérance (puisque les variables aléatoires sont identiquement distribuées) va nous donner la réponse dans le cas où la moyenne théorique est approchée comme dans la pratique (industrie) par l'estimateur de la moyenne (cas le plus fréquent).

Nous avons donc le calcul de l'espérance de la *variance empirique* :

$$\begin{aligned} E(\hat{\sigma}^2) = E(V(X)) &= E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)\right) \\ &= E\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{2\mu}{n}\sum_{i=1}^{n}x_i+\mu^2)\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2-2\mu^2+\mu^2)\right) \\ &= E\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2-\mu^2)\right) = \frac{1}{n}\sum_{i=1}^{n}E\left(x_i^2\right) - E\left(\mu^2\right) \end{aligned} \tag{7.398}$$

Or, comme les variables sont équidistribuées :

$$E(\hat{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n}E\left(x^2\right) - E\left(\mu^2\right) = E\left(x^2\right) - E\left(\mu^2\right) \tag{7.399}$$

Et nous avons (formule de HUYGHENS) :

$$V(x) = E(x^2) - E(x)^2 \tag{7.400}$$

ainsi que :

$$V(\hat{\mu}) = E(\hat{\mu}^2) - E(\hat{\mu})^2 = E(\hat{\mu}^2) - E(x)^2 \tag{7.401}$$

où la deuxième relation ne peut s'écrire que parce que nous utilisons l'estimateur de maximum de vraisemblance de la moyenne (moyenne empirique). D'où :

$$E(\hat{\sigma}^2) = E(x^2) - E(\hat{\mu})^2 = (V(x)+E(x)^2) - (V(\hat{\mu})+E(x)^2) = V(x) - V(\hat{\mu}) \tag{7.402}$$

et comme :

$$V(x) = \sigma^2 \quad \text{et} \quad V(\hat{\mu}) = \frac{\sigma^2}{n} \tag{7.403}$$

Nous avons finalement :

$$E(\hat{\sigma}^2) = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2\left(1-\frac{\sigma^2}{n}\right) = \frac{n-1}{n}\sigma^2 \tag{7.404}$$

nous avons donc un biais de -1 fois l'erreur-standard :

$$-\frac{\sigma^2}{n} \tag{7.405}$$

Nous noterons également que l'estimateur tend vers un estimateur sans biais (E.S.B.) lorsque le nombre d'échantillons tend vers l'infini $n \to \infty$. Nous disons alors que nous avons un *estimateur asymptotiquement non biaisé*.

REMARQUE Un estimateur est aussi dit *estimateur consistant* s'il converge en probabilité, lorsque equation, vers la vraie valeur du paramètre.

De par les propriétés de l'espérance, nous avons alors :

$$E(E(\hat{\sigma}^2)) = E(\hat{\sigma}^2) = E\left(\frac{n-1}{n}\sigma^2\right) \tag{7.406}$$

il vient alors :

$$\sigma = \sqrt{\frac{n-1}{n}\sigma^2} = \sqrt{\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2} \tag{7.407}$$

Nous avons donc finalement deux résultats importants :

1. L'estimateur de maximum de vraisemblance biaisé ou appelé également *variance empirique* ou encore *variance échantillonnale* et donc donné par :

$$\hat{\mu} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2} \tag{7.408}$$

   lorsque equation.

2. *L'estimateur de maximum vraisemblance non biaisé* :

$$\mu = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2} \tag{7.409}$$

   deux relations que nous retrouvons souvent dans les tables et dans de nombreux logiciels et que nous utiliserons plus bas dans les développements des intervalles de confiance et des tests d'hypothèses.

Au total, cela nous fait donc trois estimateurs pour la même quantité. Comme dans l'écrasante majorité des cas de l'industrie la moyenne théorique n'est pas connue, nous utilisons le plus souvent les deux dernières relations encadrées ci-dessus. Maintenant, c'est la que c'est le plus vicieux : lorsque nous calculons le biais des deux estimateurs, le premier est biaisé, le second ne l'est pas. Donc nous aurions tendance à utiliser que le second. Nous pourrions aussi parler de la variance et de la précision d'un estimateur, qui sont aussi des critères importants pour juger de la qualité d'un estimateur par rapport à un autre. Si nous faisions le calcul de la variance des deux estimateurs, alors le premier, qui est biaisé, a une variance plus petite que le second qui est sans biais. Tout ça pour dire que le critère du biais n'est pas (et de loin) le seul à étudier pour juger de la qualité d'un estimateur.

Enfin, il est important de se rappeler que le facteur -1 du dénominateur de l'estimateur de maximum de vraisemblance non biaisé provient du fait qu'il fallait corriger l'espérance de l'estimateur biaisé à la base minoré de une fois l'erreur-standard.

In extenso, ils est possible de démontrer (mais c'est long) que si la variable aléatoire suivant une loi normale dont nous cherchons l'expression de l'estimateur non biaisé est la somme de k variables aléatoires linéairement indépendantes alors nous avons :

$$\mu = \sqrt{\frac{1}{n-k}(x_i - \hat{\mu})^2} \tag{7.410}$$

7.5.2 Estimateur de la loi de Poisson

En utilisant la même méthode que pour la loi Normale, nous allons donc rechercher l'estimateur de maximum de vraisemblance la loi de Poisson définie par (7.186). Dès lors, la vraisemblance est donnée par :

$$L(\mu) = \prod_{i=1}^{n} P(x_i, \mu) = \frac{\mu^{\sum_{i=1} x_i}}{\prod_{i=1} x_i!}\mathrm{e}^{-\mu n} \tag{7.411}$$

Maximiser une fonction ou maximiser son logarithme est équivalent donc :

$$\ln[L(\mu)] = \ln(\mu)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\ln(x_i!) - \mu n \tag{7.412}$$

Nous cherchons maintenant à la maximiser :

$$\frac{\partial}{\partial\sigma}\ln[L(\mu)] = \frac{1}{\mu}\sum_{i=1} x_i - n = 0 \tag{7.413}$$

et obtenons donc son unique estimateur de maximum de vraisemblance qui sera :

$$\hat{\mu} = \frac{1}{n}\sum_{i=1} x_i \tag{7.414}$$

Il est tout à fait normal de retrouver dans cet exemple didactique la moyenne empirique, car c'est le meilleur estimateur possible pour le paramètre de la loi de POISSON (qui représente aussi l'espérance d'une loi de POISSON).

Sachant que l'écart type de la distribution particulière (voir plus haut) n'est que la racine carrée de la moyenne, nous avons alors pour l'écart-type de maximum de vraisemblance biaisé :

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1} x_i} \tag{7.415}$$

7.5.3 Estimateur de la loi binomiale

En utilisant la même méthode que pour la loi Normale et la loi de POISSON, nous allons donc rechercher l'estimateur de maximum de vraisemblance la loi Binomiale définie par l'équation (7.148). Dès lors, la vraisemblance est donnée par :

$$L(p) = \prod_{i=1}^{n} P(x_i, p) = C_k^N p^k (1-p)^{N-k} \tag{7.416}$$

Il convient de se rappeler que le facteur qui suit le terme combinatoire exprime déjà les variables successives selon ce que nous avons vu lors de notre étude de la fonction de distribution de BERNOULLI et de la fonction bin0miale.

Maximiser une fonction ou maximiser son logarithme est équivalent donc :

$$\ln[L(p)] = \ln(C_k^N) + k\ln p + (N-k)\ln(1-p) \tag{7.417}$$

Nous cherchons maintenant à la maximiser :

$$\frac{\partial}{\partial\sigma}\ln[L(p)] = \frac{k}{p} - \frac{N-k}{1-p} \tag{7.418}$$

Ce qui donne :

$$k(1-p) - p(N-k) = k - kp - pN + pk = k - pN = 0 \tag{7.419}$$

d'où nous tirons l'estimateur de maximum de vraisemblance biaisé qui sera :

$$\hat{p} = \frac{k}{N} \tag{7.420}$$

Ce résultat est assez intuitif si l'on considère l'exemple classique d'une pièce de monnaie qui à une chance sur deux de tomber sur une des ces faces. La probabilité p étant le nombre de fois k où une face donnée a été observée sur le nombre d'essais total (toutes faces confondues).

REMARQUE Dans la pratique, il n'est pas aussi simple d'appliquer ces estimateurs. Il faut bien réfléchir pour comprendre lesquels sont les plus adaptés à une expérience donnée et idéalement, il faut également calculer l'erreur quadratique moyenne de chacun des estimateurs de la moyenne.

7.5.4 Estimateur de la loi WEIBULL

Nous avons vu dans le chapitre de Génie Industriel une étude très détaillée de la loi de WEIBULL à trois paramètres avec son écart-type et son espérance car nous avions précisée qu'elle était assez utilisée dans le domaine de l'ingénierie de la fiabilité.

Malheureusement les trois paramètres de cette loi nous sont en pratique inconnus. À l'aide des estimateurs nous pouvons cependant déterminer l'expression de deux des trois en supposant γ comme étant nul. Cela nous donne donc la loi de WEIBULL dite *à deux paramètres* suivante :

$$P(x,\beta,\eta) = \frac{\beta}{\eta}\frac{x}{\eta}^{\beta-1}\mathrm{e}^{-\left(\frac{x}{\eta}\right)^{\beta}} \tag{7.421}$$

avec pour rappel $\beta > 0$ et $\eta > 0$.

Dès lors la vraisemblance est donnée par :

$$L(\beta,\eta) = \prod_{i=1}^{n} P(x_i,\beta,\eta) = \prod_{i=1}^{n} \frac{\beta}{\eta}\left(\frac{x_i}{\eta}\right)^{\beta-1} \mathrm{e}^{-\left(\frac{x_i}{\eta}\right)^{\beta}} = \frac{\beta}{\eta}^{n} \mathrm{e}^{-\sum_{i=1}^{n}\left(\frac{x_i}{\eta}\right)^{\beta}} \left(\frac{x_i}{\eta}\right)^{\beta-1} \tag{7.422}$$

Maximiser une fonction ou maximiser son logarithme est équivalent donc :

$$\begin{aligned} \ln\left(L(\beta,\eta)\right) &= n\ln\left(\frac{\beta}{\eta}\right) - \frac{1}{n^{\beta}}\sum_{i=1}^{n}\left(x_i^{\beta}\right) + \sum_{i=1}^{n}\ln\left[\left(\frac{x_i}{\eta}\right)^{\beta}\right] \\ &= n\ln\left(\frac{\beta}{\eta}\right) - n^{-\beta}\sum_{i=1}^{n}\left(x_i^{\beta}\right) + (\beta-1)\sum_{i=1}^{n}\ln\left(\frac{x_i}{\eta}\right) \end{aligned} \tag{7.423}$$

Cherchons maintenant à maximiser cela en se rappelant que (cf. chapitre de Calcul Différentiel et Intégral) :

equation et equation (7.39)

d'où :

$$\begin{aligned} \frac{\partial\ln\left(L(\beta,\eta)\right)}{\partial\beta} &= \frac{n}{\beta} + \frac{1}{\eta^{\beta}}\ln\eta\sum_{i=1}^{n}x_i^{\beta} - \frac{1}{\eta^{\beta}}\ln\eta\sum_{i=1}^{n}x_i^{\beta}\ln(x_i) + \sum_{i=1}^{n}\ln\left(\frac{x_i}{\eta}\right) \\ &= \frac{n}{\beta} + \frac{1}{\eta^{\beta}}\left(\sum_{i=1}^{n}x_i^{\beta}(\ln\eta - \ln(x_i))\right) + \sum_{i=1}^{n}\ln\left(\frac{x_i}{\eta}\right) \\ &= \frac{n}{\beta} - \frac{1}{\eta^{\beta}}\sum_{i=1}^{n}x_i^{\beta}\ln\left(\frac{x_i}{\eta}\right) + \sum_{i=1}^{n}\ln\left(\frac{x_i}{\eta}\right) = 0 \end{aligned} \tag{7.424}$$

Et nous avons pour le deuxième paramètre :

$$\frac{\partial\ln\left(L(\beta,\eta)\right)}{\partial\eta} = -\frac{n}{\eta} - \frac{n}{\eta^{\beta}\eta}\sum_{i=1}^{n}x_i^{\beta} + (1-\beta)\frac{n}{\eta} = 0 \tag{7.425}$$

d'où :

$$-\frac{1}{\eta^{\beta}}\sum_{i=1}^{n}x_i^{\beta} - n = 0 \tag{7.426}$$

Finalement avec les écritures correctes (et dans l'ordre de résolution dans la pratique) :

$$\frac{n}{\hat{\beta}} - \frac{1}{\hat{\eta}^{\hat{\beta}}}\sum_{i=1}^{n}x_i^{\hat{\beta}}\ln\left(\frac{x_i}{\hat{\eta}}\right) + \sum_{i=1}^{n}\ln\left(\frac{x_i}{\hat{\eta}}\right) = 0 \quad \text{et} \quad -\frac{1}{\hat{\eta}^{\hat{\beta}}}\sum_{i=1}^{n}x_i^{\hat{\beta}} - n = 0 \tag{7.427}$$

On prend alors une approche différente en écrivant notre loi de WEIBULL à deux paramètres ainsi :

$$P(x,\beta,\theta) = \frac{\beta}{\theta}x^{\beta-1}\mathrm{e}^{-\frac{x^{\beta}}{\theta}} \tag{7.428}$$

avec pour rappel $\beta > 0$ et $\theta > 0$. Dès lors la vraisemblance est donnée par :

$$L(\beta,\eta) = \prod_{i=1}^{n}P(x_i,\beta,\eta) = \prod_{i=1}^{n}\frac{\beta}{\theta}x_i^{\beta-1}\mathrm{e}^{-\frac{x_i^{\beta}}{\theta}} = \left(\frac{\beta}{\eta}\right)^{n}\mathrm{e}^{-\sum_{i=1}^{n}\frac{x_i^{\beta}}{\eta}}\prod_{i=1}^{n}x_i^{\beta-1} \tag{7.429}$$

Maximiser une fonction ou maximiser son logarithme est équivalent donc :

$$\begin{aligned}\ln(L(\beta,\eta)) &= n\left(\frac{\beta}{\eta}\right) - \frac{1}{\theta}\sum_{i=1}^{n} x_i^{\beta} + \sum_{i=1}^{n}\ln(x_i^{\beta-1}) \\ &= n\ln\left(\frac{\beta}{\theta}\right) - \frac{1}{\theta}\sum_{i=1}^{n} x_i^{\beta} + (\beta-1)\sum_{i=1}^{n}\ln x_i\end{aligned} \tag{7.430}$$

Cherchons maintenant à maximiser cette expression en se rappelant que (cf. chapitre de Calcul Différentiel et Intégral) :

equation et equation (7.47)

d'où :

$$\frac{\partial \ln\left(L(\beta,\theta)\right)}{\partial\beta} = \frac{n}{\beta} + \frac{1}{\theta}\sum_{i=1}^{n} x_i^{\beta}\ln x_i + \sum_{i=1}^{n}\ln x_i = 0 \tag{7.431}$$

Et nous avons pour le deuxième paramètre :

$$\frac{\partial \ln\left(L(\beta,\theta)\right)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^{\beta} = 0 \tag{7.432}$$

Il est alors immédiat que :

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\beta} = 0 \tag{7.433}$$

injecté dans la relation (7.431), il vient :

$$\frac{n}{\beta} - \frac{\sum_{i=1}^{n} x_i^{\beta}\ln x_i}{\frac{1}{n}\sum_{i=1}^{n} x_i^{\beta}} + \sum_{i=1}^{n}\ln x_i = 0 \tag{7.434}$$

en simplifiant :

$$\frac{\sum_{i=1}^{n} x_i^{\hat{\beta}\ln x_i}}{\sum_{i=1}^{n} x_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\ln x_i \tag{7.435}$$

7.6 Intervalles de confiance

Jusqu'à maintenant nous avons toujours déterminé les différents estimateurs de vraisemblance ou estimateurs simples (variance, écart-type) à partir de lois (fonctions) statistiques théoriques ou mesurées sur toute une population de données.

Nous allons maintenant aborder une approche un peu différente et importante dans l'industrie en se demandant maintenant quelles doivent être les tailles d'échantillons pour avoir une certaine validité (intervalle de confiance I.C.) pour les données mesurées ou encore quel écart-type ou fractile dans une loi Normale centrée réduite (grand nombre d'échantillons), du Khi-deux, de Student ou de FISHER correspond à un certain intervalle de confiance (nous verrons ces deux derniers cas de faibles échantillons dans la partie traitant de l'analyse de la variance ou ANOVA) lorsque la variance ou la moyenne est connue ou respectivement inconnue sur l'ensemble ou une partie de la population de donnée.

Indiquons que ces intervalles de confiance utilisent le théorème central limite démontré plus loin (afin d'éviter toute frustration) et que les développements que nous allons faire maintenant nous seront également utiles dans le domaine des Tests d'Hypothèse qui ont une place majeure en statistique.

7.6.1 Sur la moyenne avec variance théorique connue

Commençons par le cas le plus simple et le plus courant qui est la détermination du nombre d'échantillons pour avoir une certaine confiance dans la moyenne des mesures effectuées d'une variable aléatoire supposée suivre une loi Normale.

D'abord souvenons-nous de l'expression (7.97). Maintenant, considérons X comme une variable aléatoire suivant une loi Normale de moyenne μ et d'écart-type σ. Nous souhaiterions déterminer à combien de sigma correspond un intervalle de confiance de 95%. Pour déterminer cela, nous écrivons d'abord :

$$P(\mu - \delta \leqslant X \leqslant \mu + \delta) = 0{,}95 \tag{7.436}$$

Remarque Donc avec un intervalle de confiance de 95% vous aurez raison 19 fois sur 20, ou n'importe quel autre niveau de confiance ou niveau de risque (1-niveau de confiance) que vous vous serez fixé à l'avance. En moyenne, vos conclusions seront donc bonnes, mais nous ne pourrons jamais savoir si une décision particulière est bonne.

En centrant et réduisant la variable aléatoire :

$$P\left(-\frac{\delta}{\sigma} \leqslant \frac{X-\mu}{\sigma} \leqslant \frac{\delta}{\sigma}\right) = 0{,}95 \tag{7.437}$$

Notons maintenant Y la variable centrée réduite :

$$P\left(-\frac{\delta}{\sigma} \leqslant Y\right) - P\left(Y \geqslant \frac{\delta}{\sigma}\right) = 0{,}95 \tag{7.438}$$

Puisque la loi Normale centrée réduite est symétrique :

$$1 - 2P\left(Y \geqslant \frac{\delta}{\sigma}\right) = 0{,}95 \tag{7.439}$$

D'où :

$$P\left(Y \geqslant \frac{\delta}{\sigma}\right) = 0{,}025 \tag{7.440}$$

$$\frac{\delta}{\sigma} \simeq 1{,}96 \quad \Leftrightarrow \quad \delta \simeq 1{,}96\sigma \tag{7.441}$$

Ce qui est noté de façon traditionnelle dans le cas général autre que 95% par (Z n'est pas une variable aléatoire c'est juste le facteur qui est la variable suivante) :

$$\delta \simeq Z\sigma \tag{7.442}$$

Or, considérons que la variable X sur la quelle nous souhaitons faire de l'inférence statistique est justement la moyenne (et nous démontrerons plus loin que celle-ci suit une loi Normale centrée réduite). Dès lors :

$$\delta \simeq Z\sigma = Z\frac{\sigma}{\sqrt{n}} \tag{7.443}$$

autrement dit :

$$n \simeq \frac{Z^2\sigma^2}{\delta^2} \tag{7.444}$$

Ainsi, nous pouvons maintenant savoir le nombre d'échantillons à avoir pour s'assurer un intervalle de précision δ autour de la moyenne et pour qu'un pourcentage donné des mesures se trouvent dans cet intervalle et en supposant l'écart-type expérimental σ connu d'avance.

Cependant, en réalité, la variable Z provient du théorème central limite (voir plus bas) qui donne pour un échantillon de grande taille :

$$Z = Z_n = \frac{M_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{7.445}$$

En réarrangeant nous obtenons :

$$\mu = \bar{X} - Z\frac{\sigma}{\sqrt{n}} \tag{7.446}$$

et comme Z peut être négatif ou positif alors il est plus censé d'écrire cela sous la forme :

$$\mu = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}} \tag{7.447}$$

Soit :

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + Z\frac{\sigma}{\sqrt{n}} \tag{7.448}$$

que les ingénieurs notent :

$$LCL \leqslant \mu \leqslant UCL \tag{7.449}$$

où LCL représente la *lower confidence limit* et UCL, la *upper confidence limit*. C'est de la terminologie Six Sigma (cf. chapitre de Génie Industriel).

Nous venons de voir qu'un intervalle de confiance à 95% nécessite $Z = 1{,}96$ et puisque la loi Normale est symétrique, il est possible d'écrire :

$$95\% = 1 - 5\% = 1 - \alpha = 1 - 2 \times 2{,}5\% = 1 - 2 \times 0{,}025 \tag{7.450}$$

Cela se note finalement :

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{7.451}$$

soit dans le cas d'un intervalle de confiance à 95% :

$$\bar{X} - Z_{0{,}025}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + Z_{0{,}025}\frac{\sigma}{\sqrt{n}} \tag{7.452}$$

Nous sommes ainsi capables maintenant d'estimer des tailles de population nécessaires à obtenir un certain niveau de confiance dans un résultat, soit d'estimer dans quel intervalle de confiance se trouve la moyenne théorique par rapport à la moyenne expérimentale (empirique).

7.6.2 Sur la variance avec moyenne théorique connue

Commençons à démontrer une propriété fondamentale de la loi du khi-deux :

Si une variable aléatoire X suit une loi Normale centrée réduite $X = N(0,1)$ alors son carré suit une loi du khi-deux de degré de liberté 1 :

$$X^2 = \chi^2(1) \tag{7.453}$$

Preuve Pour démontrer cette propriété, il suffit de calculer la densité de la variable aléatoire X^2 avec $X = N(0,1)$. Dans ce cadre, en posant $Y = X^2$, alors pour tout $y \geqslant 0$ nous obtenons :

$$P(Y \leqslant y) = P(X^2 \leqslant y) = P(-\sqrt{y} \leqslant Y \leqslant \sqrt{y}) \tag{7.454}$$

Puisque la loi Normale centrée réduite est symétrique par rapport à 0 pour la variable aléatoire X, nous pouvons écrire :

$$P(Y \leqslant y) = 2P(0 \leqslant X \leqslant \sqrt{y}) \tag{7.455}$$

En notant Φ la fonction de répartition de la loi Normale centrée réduite, nous avons :

$$P(Y \leqslant y) = 2\left(\Phi(\sqrt{y}) - \Phi(0)\right) \tag{7.456}$$

et comme :

$$\Phi(0) = P(0,0,1) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \mathrm{e}^{-k/2}\mathrm{d}k \tag{7.457}$$

alors :

$$P(Y \leqslant y) = 2\Phi(\sqrt{y}) - 2 \times 0{,}5 = 2\Phi(\sqrt{y}) - 1 \tag{7.458}$$

qui représente fonction de répartition de la variable aléatoire $Y = X^2$ si $y \geqslant 0$ et nulle sinon. Nous noterons cette réparation $F_Y(y)$ pour la suite des calculs.

Puisque la fonction de distribution est la dérivée de la fonction de répartition et que X suit une loi Normale centrée réduite alors nous avons pour la variable aléatoire X :

$$P(X = x) = \Phi'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\Phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \tag{7.459}$$

alors nous avons pour la loi de distribution de Y :

$$\frac{\mathrm{d}}{\mathrm{d}y}f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}(2\Phi(\sqrt{y}) - 1) = 2\Phi'(\sqrt{y})\frac{\mathrm{d}}{\mathrm{d}y}\sqrt{y} = \frac{1}{\sqrt{y}}\Phi'(\sqrt{y}) = \frac{1}{\sqrt{2\pi y}}\mathrm{e}^{-y/2} \tag{7.460}$$

cette dernière expression correspond exactement à la relation que nous avions obtenu lors de notre étude de la loi du khi-deux en imposant un degré de liberté unité.

Le théorème est donc bien démontré tel que si X suit une loi Normale centrée réduite alors son carré suit une loi du khi-deux à 1 degré de liberté tel que :

$$\left(X_{N(0,1)}\right)^2 = \chi^2(1) \tag{7.461}$$

•

Ce type de relation est utilisé dans les processus industriels et leur contrôle (cf. chapitre de Génie Industriel).

Nous allons maintenant utiliser un résultat démontré lors de notre étude de la loi Gamma. Nous avons effectivement dans l'expression (7.322) que la somme de deux variables aléatoires suivant une loi Gamma suit aussi une loi Gamma dont les paramètres s'additionnent. Comme la loi du khi-deux n'est qu'un cas particulier de la loi Gamma, le même résultat s'applique. Par conséquent, si $X_1, \ldots, X_k$ sont des variables aléatoires indépendantes et identiquement distribuées selon $N(0,1)$, alors par extension de (7.322) et de la propriété d'addition de la loi Gamma, la somme de leurs carrés suit une loi du khi-deux de degré k telle que :

$$\chi^2(k) = X_1^2 + X_2^2 + \ldots + X_k^2 \tag{7.462}$$

Ainsi, la loi du χ^2 à k degrés de liberté est la loi de probabilité de la somme des carrés de k variables normales centrées réduites linéairement indépendantes entre elles. Il s'agit de la propriété de linéarité de la loi du Khi-deux et implicitement de celle de la loi Gamma.

Maintenant voyons une autre propriété importante de la loi du khi-deux. Si $X_1, \ldots, X_k$ sont des variables aléatoires indépendantes et identiquement distribuées selon $N(\mu, \sigma^2)$ et si nous notons l'estimateur de maximum de vraisemblance de la variance :

$$S_*^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \tag{7.463}$$

alors, le rapport de la variable aléatoire S_*^2 sur l'écart-type supposé connu de l'ensemble de la population multiplié par le nombre d'échantillons n de la population suit une loi du khi-deux de degré n telle que :

$$\chi^2(n) = \frac{nS_*^2}{\sigma^2} \tag{7.464}$$

Remarques

1. En laboratoire, les $X_1, \ldots, X_k$ peuvent être vues comme une classe d'échantillons d'un même produit étudié identiquement par différentes équipes de recherche avec des instruments de même précision (écart-type de mesure nul).
2. S_*^2 est la *variance interclasse* également appelée *variance expliquée*. Donc elle donne la variance d'une mesure ayant eu lieu dans les différents laboratoires.

Ce qui est intéressant c'est qu'à partir du calcul de la loi du khi-deux en connaissant n et l'écart-type σ^2 il est possible d'estimer cette variance (écart-type) interclasse. Pour voir que cette dernière propriété est une généralisation élémentaire de la relation (7.462), il suffit de constater que la variable aléatoire nS_*^2/σ^2 est une somme de n carrés de $N(0,1)$ indépendants les uns des autres. Effectivement, rappelons qu'une variable aléatoire centrée réduite est donnée par l'expression (7.67). Dès lors :

$$\frac{nS_*^2}{\sigma^2} = \frac{n}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \tag{7.465}$$

Or, puisque les variables aléatoires $X_1,\ldots,X_k$ sont indépendantes et identiquement distribuées selon une loi Normale, les variables aléatoires :

$$\frac{X_1-\mu}{\sigma},\ldots,\frac{X_n-\mu}{\sigma} \tag{7.466}$$

sont aussi indépendantes et identiquement distribuées mais selon une loi Normale centrée réduite. En réarrangeant (7.467), nous obtenons :

$$\sigma^2 = \frac{nS_*^2}{\chi^2(n)} \tag{7.467}$$

Donc sur la population de mesures, l'écart-type vrai suit la relation donnée ci-dessus. Il est donc possible de faire de l'inférence statistique sur l'écart-type lorsque la moyenne théorique est connue.

Puisque la fonction du khi-deux n'est pas symétrique, la seule possibilité pour faire l'inférence, c'est de faire appel au calcul numérique et nous noterons alors l'intervalle de confiance à 95% de la manière suivante :

$$\frac{nS_*^2}{\chi^2_{2,5\%}(n)} \leqslant \sigma^2 \leqslant \frac{nS_*^2}{\chi^2_{97,5\%}(n)} \tag{7.468}$$

Soit en notant $95\% = 1-\alpha$:

$$\frac{nS_*^2}{\chi^2_{\alpha/2}(n)} \leqslant \sigma^2 \leqslant \frac{nS_*^2}{\chi^2_{1-\alpha/2}(n)} \tag{7.469}$$

le dénominateur étant alors bien évidemment la probabilité cumulée. Cette relation est rarement utilisée dans la pratique car la moyenne théorique n'est pas connue. Voyons donc le cas le plus courant.

7.6.3 Sur la variance avec moyenne empirique connue

Cherchons maintenant à faire de l'inférence statistique lorsque la moyenne théorique de la population μ n'est pas connue. Pour cela, considérons maintenant la somme :

$$\chi^2(n) = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left((X_i-\bar{X})+(\bar{X}-\mu)\right)^2 \tag{7.470}$$

où pour rappel $\bar{X}$ est la moyenne empirique (arithmétique) de l'échantillon. En continuant le développement nous avons :

$$\chi^2(n) = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2 + 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\bar{X}) + n(\bar{X}-\mu)^2\right] \tag{7.471}$$

Or, nous avons démontré au début de ce chapitre que la somme des écarts à la moyenne était nulle, d'où :

$$\begin{aligned}\chi^2(n) &= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2 + 2(\bar{X}-\mu)\cdot 0 + n(\bar{X}-\mu)^2\right] \\ &= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2\right] \\ &= \sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2} = \sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2} + \left(\frac{\bar{X}-\mu}{\sigma^2/\sqrt{n}}\right)^2\end{aligned} \tag{7.472}$$

et reprenons l'estimateur sans biais de la loi Normale [17] :

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{7.473}$$

Dès lors, en reprenant (7.472) :

$$\chi^2(n) = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X}-\mu}{\sigma^2/\sqrt{n}}\right)^2 \tag{7.474}$$

ou autrement écrit :

$$\chi^2(n) = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{M_n-\mu}{\sigma^2/\sqrt{n}}\right)^2 \tag{7.475}$$

Puisque le deuxième terme suit une loi Normale centrée réduite aussi, en le supprimant nous obtenons de par la propriété démontrée plus haut de la loi du Khi-deux :

$$\chi^2(n-1) = \frac{(n-1)S^2}{\sigma^2} \tag{7.476}$$

Ces développements nous permettent cette fois-ci de faire aussi de l'inférence sur la variance σ^2 d'une loi $N(\mu,\sigma)$ lorsque les paramètres μ et σ sont tous les deux inconnus pour l'ensemble de la population. C'est ce résultat qui nous donne, par exemple, l'intervalle de confiance :

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)} \tag{7.477}$$

lorsque la moyenne théorique μ est donc inconnue.

7.6.4 Sur la moyenne avec moyenne empirique connue

Nous avons démontré beaucoup plus haut que la loi de Student provenait de la relation suivante :

$$T(k) = \frac{Z}{\sqrt{U/k}} \tag{7.478}$$

si Z et U sont des variables aléatoires indépendantes et si Z suit une loi Normale centrée réduite $N(0,1)$ et U une loi du khi-deux $\chi^2(k)$ tel que $F_T(t) = P(t)$ où $P(t)$ est défini par l'équation (7.340).

Voici une application très importante de ce résultat.

Supposons que $X_1,\ldots,X_n$ constituent un échantillon aléatoire de taille n issu de la loi $N(\mu,\sigma)$. Alors nous pouvons déjà écrire :

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \tag{7.479}$$

Si U suit une loi $\chi^2(n)$ et si nous posons $k = n-1$, nous retrouvons les résultats obtenus plus haut :

$$U = \frac{(n-1)S^2}{\sigma^2} = \chi^2(n-1) \tag{7.480}$$

autrement dit, après quelques simplifications triviales :

$$\frac{Z}{\sqrt{U/k}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \tag{7.481}$$

Or, d'après (7.478) :

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = T(n-1) \tag{7.482}$$

17. Dans ce cas précis, nous changeons de notation pour respecter les traditions et bien différencier la moyenne empirique de la moyenne théorique.

suit aussi une loi de Student de paramètre $n-1$ et qui nous donne aussi :

$$\mu = \bar{X} - \frac{S}{\sqrt{n}} T(n-1) \tag{7.483}$$

Ce qui nous permet de faire de l'inférence sur la moyenne μ d'une loi Normale d'écart-type inconnu mais dont l'estimateur sans biais de l'écart-type est connu (donc l'écart-type théorique est inconnu). C'est ce résultat qui nous donne l'intervalle de confiance :

$$\bar{X} - \frac{S}{\sqrt{n}} T_{\alpha/2}(n-1) \leqslant \mu \leqslant \bar{X} + \frac{S}{\sqrt{n}} T_{\alpha/2}(n-1) \tag{7.484}$$

où nous retrouvons les mêmes indices que pour l'inférence statistique sur la moyenne d'une variable aléatoire d'écart-type connu puisque la loi de Student est symétrique.

REMARQUE Le résultant précédent fut obtenu par William S. GOSSET vers 1910. GOSSET, qui avait étudié les mathématiques et la chimie, travaillait comme statisticien pour la brasserie Guinness en Angleterre. À l'époque, on savait que si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes et identiquement distribuées alors :

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = N(0,1) \tag{7.485}$$

Toutefois, dans les applications statistiques, on s'intéressait bien évidemment plutôt à la quantité :

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = N(0,1) \tag{7.486}$$

on se contentait alors de supposer que cette quantité suivait à peu près une loi Normale centrée réduite ce qui n'était pas une mauvaise approximation.

Suite à de nombreuses simulations, GOSSET arriva à la conclusion que cette approximation était valide seulement lorsque n est suffisamment grand (donc cela lui donnait l'indication comme quoi il devait y avoir quelque part derrière le théorème central limite). Il décida de déterminer l'origine de la distribution et après avoir suivi un cours de statistique avec Karl PEARSON il obtint son fameux résultat qu'il publia sous le pseudonyme de Student. Ainsi, on appelle loi de Student la loi de probabilité qui aurait dû être appelée la loi ou fonction de GOSSET.

Signalons enfin que le test de student est très utilisée pour identifier si des variations (progressions ou l'inverse) de la moyenne de chiffres de deux populations identiques sont significatives.

7.7 Loi faible des grands nombres

Nous allons maintenant nous attarder sur une relation très intéressante en statistique qui permet de dire pas mal de choses tout en ayant peu de données et ce quelque soit la loi considérée. C'est une propriété très utilisée en simulation statistique, par exemple, dans le cadre de l'utilisation de Monte-Carlo.

Soit une variable aléatoire à valeurs dans $\mathbb{R}^+$. Alors nous allons démontrer la relation suivante appelée *inégalité de Markov* :

$$\forall \lambda > 0 : P(X \geqslant \lambda) \leqslant \frac{E(X)}{\lambda} \tag{7.487}$$

avec $E(X) \leqslant \lambda$ dans le contexte particulier des probabilités.

En d'autres termes, nous proposons de démontrer que la probabilité qu'une variable aléatoire soit plus grande ou égale qu'une valeur λ est inférieure ou égale à son espérance divisée par la valeur considérée λ et ce quelle que soit la loi de distribution de la variable aléatoire X.

Preuve Notons les valeurs de X par $\{x_1, \ldots, x_n\}$, où $0 \leqslant x_1 < x_2 < \ldots < x_n$ et posons $x_0 = 0$. Nous remarquons d'abord que l'inégalité est triviale au cas où $\lambda \geqslant x_n \geqslant 0$. Effectivement, comme X ne peut être compris qu'entre 0 et x_n par définition, alors la probabilité qu'il soit supérieure à x_n est nulle. En d'autres termes :

$$P(X \geqslant \lambda) = 0 \tag{7.488}$$

et X puisque est positif, $E(X)$ l'est aussi, d'où l'inégalité pour ce cas particulier. Sinon $0 \leqslant \lambda \leqslant x_n$ et $\exists k \in \{1, \ldots, n\}$ tel que $x_{k-1} \leqslant \lambda \leqslant x_k$ et par conséquent :

$$E(X) = \sum_{i=1}^{n} x_i p_i \geqslant \sum_{i=k}^{n} x_i p_i \geqslant \sum_{i=k}^{n} p_i = \lambda P(X \geqslant \lambda) \tag{7.489}$$

•

Exemple 7.8 *Nous supposons que le nombre de pièces sortant d'une usine donnée en l'espace d'une semaine est une variable aléatoire d'espérance 50. Si nous souhaitons estimer la probabilité cumulée que la production dépasse 75 pièces nous appliquerons simplement :*

$$P(X \geqslant 75) \leqslant \frac{E(X)}{\lambda} = \frac{50}{75} = 0{,}6666 = 66{,}66\% \tag{7.490}$$

Considérons maintenant une généralisation de cette inégalité. Soit une variable aléatoire X afin de démontrer l'inégalité de Bienaymé-Tchebychev suivante :

$$\forall \varepsilon > 0 : P(|X - E(X)| \geqslant \varepsilon) \leqslant \frac{V(X)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} \tag{7.491}$$

qui exprime le fait que plus l'écart-type est petit, plus la probabilité que la variable aléatoire X s'éloigne de sont espérance est faible.

Nous obtenons cette inégalité en écrivant d'abord :

$$\forall \varepsilon > 0 : P(|X - E(X)| \geqslant \varepsilon) = P\left([X - E(X)]^2 \geqslant \varepsilon^2\right) \tag{7.492}$$

et le choix du carré va nous servir pour une simplification future.

Puis en appliquant l'inégalité de Markov à la variable aléatoire $Y = [X - E(X)]^2$ avec $\lambda = \varepsilon^2$, il vient automatiquement :

$$P(Y \geqslant \varepsilon^2) = P\left([X - E(X)]^2 \geqslant \varepsilon^2\right) \leqslant \frac{E(Y)}{\varepsilon^2} = \frac{E\left([X - E(X)]^2\right)}{\varepsilon^2} \tag{7.493}$$

Ensuite, en utilisant la définition de la variance :

$$E\left([X - E(X)]^2\right) = E\left([X - \mu_X]^2\right) = V(X) = \sigma^2 \tag{7.494}$$

Nous obtenons bien :

$$P(Y \geqslant \varepsilon^2) \leqslant \frac{E\left([X - E(X)]^2\right)}{\varepsilon^2} = \frac{V(x)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} \tag{7.495}$$

Si nous posons $\varepsilon = t\sigma$, l'inégalité (7.495) s'écrit :

$$P(|X - E(X)| \geqslant t\sigma) \leqslant \frac{1}{t^2} \tag{7.496}$$

et exprime que la probabilité que pour que X s'éloigne de son espérance de plus que t fois son écart-type, est inférieure à $1/t^2$. Il y a, en particulier, moins de une chance sur neuf pour que X s'éloigne de son espérance de plus de trois fois l'écart-type.

Exemple 7.9 *Nous reprenons l'exemple où le nombre de pièces sortant d'une usine donnée en l'espace d'une semaine est une variable aléatoire d'espérance 50. Nous supposons en plus que la variance de la production hebdomadaire est de 25. Nous cherchons à calculer la probabilité que la production de la semaine prochaine soit comprise entre 40 et 60 pièces.*

Pour calculer ceci il faut d'abord se souvenir que l'inégalité de BT est basée en parties sur le terme $|X - E(X)|$ donc nous avons $|40 - 50| = |50 - 60| = 10$ donc l'inégalité de BT nous permet bien de travailler sur des intervalles égaux en valeur absolue, ce qui s'écrit aussi $-10 \leqslant X - E(X) \leqslant 10$. Ensuite, ne reste plus qu'à appliquer simplement l'inégalité (7.495) numériquement :

$$P(|X - E(X)| \geqslant \varepsilon) = P(|X - E(X)| \geqslant 10) = \frac{V(x)}{\varepsilon^2} = \frac{25}{10^2} = 0{,}25 = 25\% \tag{7.497}$$

Ces deux dernières inégalités vont nous permettre d'obtenir une relation très importante et puissante que nous appelons la loi faible des grands nombres *(L.F.G.N.) ou encore* théorème de Khintchine.

Considérons une variable aléatoire X admettant une variance et $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes (donc non corrélées deux-deux) de même loi que X et ayant toutes les mêmes espérances μ et les mêmes écarts-types σ.

Ce que nous allons montrer est que si nous mesurons une même quantité aléatoire X_n de même loi au cours d'une suite d'expériences indépendantes (alors dans ce cas, nous disons techniquement que la suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires sont définies sur le même espace probabilisé), alors la moyenne arithmétique des valeurs observées va se stabiliser sur l'espérance de X quand le nombre de mesures est infiniment élevée. De manière formelle ceci s'exprime sous la forme :

$$\forall\varepsilon>0,\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\geqslant\varepsilon\right)\leqslant\frac{\sigma^2}{n\varepsilon^2}\underset{n\to+\infty}{\to}0 \tag{7.498}$$

En d'autres termes, la probabilité cumulée que la différence entre la moyenne arithmétique et l'espérance des variables aléatoires observées soit compris dans un intervalle autour de la moyenne tend vers zéro quand le nombre de variables aléatoires mesurées tend vers l'infini.

Ce résultat nous permet d'estimer l'espérance mathématique en utilisant la moyenne empirique calculée sur un très grand nombre d'expériences.

Preuve Nous utilisons l'inégalité de BIENAYMÉ-TCHEBYCHEV pour la variable aléatoire [18] :

$$Y=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{7.499}$$

Et nous calculons d'abord en utilisant les propriétés mathématiques de l'espérance :

$$E(Y)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\sum_{i=1}^{n}\mu=\mu \tag{7.500}$$

et dans un deuxième temps en utilisant les propriétés mathématiques de la variance :

$$V(Y)=V\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}V\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\left(\sum_{i=1}^{n}V(X_i)+2\sum_{1\leqslant i<j\leqslant n}^{n}\mathrm{cov}(X_i,X_j)\right) \tag{7.501}$$

et puisque nous avons supposé les variables non corrélées entre elles alors la covariance est nulle, dès lors :

$$V(Y)=\frac{1}{n^2}\left(\sum_{i=1}^{n}V(X_i)\right)=\frac{1}{n^2}\left(\sum_{i=1}^{n}\sigma^2\right)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n^2} \tag{7.502}$$

Donc en injectant cela dans l'inégalité BT (7.491), nous avons alors :

$$P\left([Y-E(Y)]^2\geqslant\varepsilon^2\right)\leqslant\frac{V(Y)}{\varepsilon^2} \tag{7.503}$$

qui devient :

$$P\left(\left[\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right]^2\geqslant\varepsilon^2\right)\leqslant\frac{\sigma^2}{n\varepsilon^2} \tag{7.504}$$

et l'inégalité tend bien vers zéro quand n au numérateur tend vers l'infini. •

Signalons que la relation (7.504) est souvent notée dans certains ouvrages et conformément à ce que nous avons vu au début de ce chapitre :

$$P\left(|M_n-m_X|^2\geqslant\varepsilon\right)\leqslant\frac{\sigma_X^2}{n\varepsilon^2} \tag{7.505}$$

ou encore :

$$P\left(|M_n-m_X|^2<\varepsilon\right)\geqslant1-\frac{\sigma_X^2}{n\varepsilon^2} \tag{7.506}$$

de laquelle, pour $\forall\varepsilon>0$, il vient :

$$\lim_{n\to+\infty}P\left(|M_n-m_X|^2<\varepsilon\right)=1 \tag{7.507}$$

18. Cette relation s'interprète difficilement mais permet d'avoir le résultat escompté.

7.8 Fonction caractéristique

Avant de donner une démonstration à la manière ingénieur du théorème central limite, introduisons d'abord la conception de *fonction caractéristique* qui tient une place centrale en statistiques.

D'abord, rappelons que la transformée de Fourier est donnée dans sa version physicienne par (cf. chapitre de Suites et Séries) la relation :

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x \tag{7.508}$$

Rappelons que la transformation de Fourier est un analogue de la théorie des séries de Fourier pour les fonctions non périodiques, et permet de leur associer un spectre en fréquence.

Nous souhaitons maintenant démontrer que :

$$g(x) = f'(x) \quad \Rightarrow \quad G(k) = \mathrm{i}kF(k) \tag{7.509}$$

En d'autres termes, nous cherchons une expression simplifiée de la transformée de Fourier de la dérivée de $f(x)$.

Preuve Nous partons donc de :

$$G(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f'(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x \tag{7.510}$$

Une intégration par parties donne :

$$\int_{-\infty}^{+\infty} f'(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x = f(x) \mathrm{e}^{-\mathrm{i}xk} \Big|_{-\infty}^{+\infty} + \mathrm{i}k \int_{-\infty}^{+\infty} f(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x \tag{7.511}$$

En imposant que, $f(x)$ tend vers zéro à l'infini, nous avons alors :

$$f(x) \mathrm{e}^{-\mathrm{i}xk} \Big|_{-\infty}^{+\infty} = 0 \tag{7.512}$$

et finalement :

$$G(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f'(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x = \frac{\mathrm{i}k}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x = \mathrm{i}kF(k) \tag{7.513}$$

C'est la premier résultat dont nous avions besoin. •

Maintenant, démontrons que :

$$g(x) = xf(x) \quad \Rightarrow \quad G(k) = iF'(k) \tag{7.514}$$

Preuve Nous partons de :

$$\begin{aligned} G(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} xf(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x = \mathrm{i} \frac{\mathrm{d}}{\mathrm{d}k} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x \right) = \mathrm{i}F'(k) \\ &= \mathrm{i} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) \frac{\mathrm{d}}{\mathrm{d}k} \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x = \mathrm{i} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) - \mathrm{i}x \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} xf(x) \mathrm{e}^{-\mathrm{i}xk} \mathrm{d}x \end{aligned} \tag{7.515}$$

C'est le deuxième résultat dont nous avions besoin. •

Maintenant effectuons le calcul de la transformée de Fourier de la loi Normale centrée-réduite :

$$y = f_X(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \tag{7.516}$$

Nous savons que cette dernière relation est trivialement solution de l'équation différentielle :

$$y' = -xy \tag{7.517}$$

en prenant la transformée de Fourier des deux côté de l'égalité et en utilisant les deux résultats (7.509) et (7.514), nous avons :

$$\mathrm{i}kF(k) = -\mathrm{i}F'(k) \tag{7.518}$$

Ou encore :

$$-k\mathrm{d}k = \frac{\mathrm{d}F(k)}{F(k)} \tag{7.519}$$

soit, après intégration :

$$F(k) = F(0)\mathrm{e}^{-k^2/2} \tag{7.520}$$

Nous avons :

$$\begin{aligned} F(0) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}\mathrm{d}x \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2/2}\mathrm{d}x \end{aligned} \tag{7.521}$$

Nous avons démontré lors de notre étude de la loi Normale que :

$$\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2/2}\mathrm{d}x = \sqrt{2\pi} \tag{7.522}$$

et donc $F(0) = 1/\sqrt{2\pi}$. Nous avons alors le résultat important suivant :

$$F(k) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-k^2/2} \tag{7.523}$$

Introduisons maintenant la fonction caractéristique telle que définie par les statisticiens :

$$\phi_X(\omega) = E\left(\mathrm{e}^{\mathrm{i}kX}\right) = \int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}kx} f_X(x)\mathrm{d}x \tag{7.524}$$

qui est un outil analytique important et puissant permettant d'analyser une somme de variables aléatoires indépendantes. De plus, cette fonction contient toutes les informations caractéristiques de la variable aléatoire X.

REMARQUE La notation de (7.524) n'est pas innocente puisque le $E(\bullet)$ représente une espérance de la fonction de densité par rapport à l'exponentielle complexe.

Finalement, la fonction caractéristique de la variable aléatoire normale centrée réduite de distribution :

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \tag{7.525}$$

devient simple à déterminer car :

$$\phi_X(k) = \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}kx} f_X(x)\mathrm{d}x = \sqrt{2\pi}F(-k) \tag{7.526}$$

raison pour laquelle la fonction caractéristique de la loi Normale centrée réduite est souvent assimilée à une simple transformée de Fourier. Or, d'après (7.523), $F(k) = F(-k)$ et :

$$\phi_X(k) = \frac{\sqrt{2\pi}}{\sqrt{2\pi}}\mathrm{e}^{-k^2/2} = \mathrm{e}^{-k^2/2} \tag{7.527}$$

qui est le résultat dont nous avons besoin pour le théorème central limite.

Mais avant cela, regardons d'un peu plus près cette fonction caractéristique (7.524). Par un développement de MacLaurin, nous avons (cf. chapitre Suites et Séries) en changeant un peu les notations :

$$\phi_X(\omega) = \int_{-\infty}^{+\infty} \left[\sum_{k=0}^{+\infty} \frac{(\mathrm{i}\omega)^k}{k!} x^k f_X(x)\mathrm{d}x\right] \tag{7.528}$$

et en intervertissant la somme et l'intégrale, nous avons :

$$\phi_X(\omega) = \sum_{k=0}^{+\infty} \frac{(\mathrm{i}\omega)^k}{k!} \left[\int_{-\infty}^{+\infty} x^k f_X(x)\mathrm{d}x\right] = \sum_{k=0}^{+\infty} \frac{(\mathrm{i}\omega)^k}{k!} M_X^k \tag{7.529}$$

Cette fonction caractéristique contient donc tous les moments (terme général utilisé pour l'écart-type et l'espérance) de X.

7.9 Théorème central limite

Le théorème central limite est un ensemble de résultats du début du 20ème siècle sur la convergence faible d'une suite de variables aléatoires en probabilité. Intuitivement, d'après ces résultats, toute somme (implicitement : la moyenne de ses variables) de variables aléatoires indépendantes et identiquement distribuées tend vers une certaine variable aléatoire. Le résultat le plus connu et le plus important est simplement appelé *théorème central limite* qui concerne une somme de variables aléatoires dont le nombre tend vers l'infini et c'est celui-ci que nous allons démontrer de manière heuristique ici.

Dans le cas le plus simple, considéré ci-dessous pour la démonstration du théorème, ces variables sont continues, indépendantes et possèdent la même moyenne et la même variance. Pour tenter d'obtenir un résultat fini, il faut centrer cette somme en lui soustrayant sa moyenne et la réduire en la divisant par son écart-type. Sous des conditions assez larges, la loi de probabilité (de la moyenne) converge alors vers une loi Normale centrée réduite. L'omniprésence de la loi Normale s'expliquant par le fait que de nombreux phénomènes considérés comme aléatoires sont dus à la superposition de causes nombreuses.

Ce théorème de probabilités possède donc une interprétation en statistique mathématique. Cette dernière associe une loi de probabilité à une population. Chaque élément extrait de la population est donc considéré comme une variable aléatoire et, en réunissant un nombre n de ces variables supposées indépendantes, nous obtenons un échantillon. La somme de ces variables aléatoires divisée par n donne une nouvelle variable nommée la moyenne empirique. Celle-ci, une fois réduite, tend vers une variable Normale réduite lorsque n tend vers l'infini comme nous le savons.

Le théorème central limite nous dit à quoi il faut s'attendre en matière de sommes de variables aléatoires indépendantes. Mais qu'en est-il des produits ? Eh bien, le logarithme d'un produit (à facteurs strictement positifs) est la somme des logarithmes des facteurs, de sorte que le logarithme d'un produit de variables aléatoires (à valeurs strictement positives) tend vers une loi Normale, ce qui entraîne une loi log-Normale pour le produit lui-même.

En elle-même, la convergence vers la loi Normale de nombreuses sommes de variables aléatoires lorsque leur nombre tend vers l'infini n'intéresse que le mathématicien. Pour le praticien, il est

intéressant de s'arrêter un peu avant la limite : la somme d'un grand nombre de ces variables est presque gaussienne, ce qui fournit une approximation souvent plus facilement utilisable que la loi exacte.

En s'éloignant encore plus de la théorie, on peut dire que bon nombre de phénomènes naturels sont dus à la superposition de causes nombreuses, plus ou moins indépendantes. Il en résulte que la loi Normale les représente de manière raisonnablement efficace.

À l'inverse, on peut dire qu'aucun phénomène concret n'est vraiment gaussien car il ne peut dépasser certaines limites, en particulier s'il est à valeurs positives.

Preuve Soit $\{X_i\}_{i=1,\ldots,+\infty}$ une suite (échantillon) de variables aléatoires continues, indépendantes[19] et identiquement distribuées, dont la moyenne μ_X et l'écart-type σ_X existent.

Replaçons Y par Z_n dans (7.99) qui est la même expression d'une variable centrée réduite générée à l'aide d'une suite de n variables aléatoires identiquement distribuées qui par construction a donc une moyenne nulle et une variance unitaire :

$$E(Z_n) = 0 \quad \text{et} \quad V(Z_n) = 1 \tag{7.530}$$

Développons la première forme de l'égalité anté-précédente (elles sont de toute façon égales) :

$$Z_n = \frac{1}{\sqrt{n}\sigma_{\bar{X}}}\left[\sum_{i=1}^{n}(X_i - \mu_{\bar{X}})\right] = \frac{1}{\sqrt{n}\sigma_{\bar{X}}}\left[\sum_{i=1}^{n} X_i - n\mu\right] \tag{7.531}$$

Maintenant utilisons la fonction caractéristique de la loi Normale centrée-réduite :

$$\phi_{Z_n}(\omega) = E[\mathrm{e}^{i\omega Z_n}] = E\left[\exp\left(i\omega\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n}\sigma_X}\right)\right] \tag{7.532}$$

Comme les variables aléatoires X_i sont indépendantes et identiquement distribuées, il vient :

$$\phi_{Z_n}(\omega) = E\left[\exp\left(i\omega\frac{X_1 - \mu}{\sqrt{n}\sigma_{\bar{X}}}\right)\right]\cdot\ldots\cdot E\left[\exp\left(i\omega\frac{X_n - \mu}{\sqrt{n}\sigma_{\bar{X}}}\right)\right] = \left(E\left[\exp\left(i\omega\frac{X - \mu}{\sqrt{n}\sigma_{\bar{X}}}\right)\right]\right)^n \tag{7.533}$$

Un développement de TAYLOR du dernier terme donne au troisième ordre :

$$E\left[\exp\left(i\omega\frac{X - \mu}{\sqrt{n}\sigma_{\bar{X}}}\right)\right] \simeq E\left[1 + i\frac{\omega}{\sqrt{n}\sigma_{\bar{X}}}(X - \mu) - \frac{\omega^2}{n\sigma_{\bar{X}}^2}(X - \mu)^2\right] \tag{7.534}$$

Finalement :

$$\begin{aligned}
\phi_{Z_n}(\omega) &\simeq \left(E\left[1 + i\frac{\omega}{\sqrt{n}\sigma_{\bar{X}}}(X - \mu) - \frac{\omega^2}{n\sigma_{\bar{X}}^2}(X - \mu)^2\right]\right)^n \\
&= \left[\int_{-\infty}^{+\infty}\left(1 + i\frac{\omega}{\sqrt{n}\sigma_{\bar{X}}}(X - \mu) - \frac{\omega^2}{n\sigma_{\bar{X}}^2}(X - \mu)^2\right) f_X(x)\mathrm{d}x\right]^n \\
&= \left[\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x + \int_{-\infty}^{+\infty}\frac{iX\omega}{\sqrt{n}\sigma_{\bar{X}}} f_X(x)\mathrm{d}x - \int_{-\infty}^{+\infty}\frac{i\mu\omega}{\sqrt{n}\sigma_{\bar{X}}} f_X(x)\mathrm{d}x - \frac{\omega^2}{n\sigma_{\bar{X}}^2}\int_{-\infty}^{+\infty}(X - \mu)^2 f_X(x)\mathrm{d}x\right]^n \\
&= \left[1 + \frac{i\omega}{\sqrt{n}\sigma_{\bar{X}}}\int_{-\infty}^{+\infty} X f_X(x)\mathrm{d}x - \frac{i\mu\omega}{\sqrt{n}\sigma_{\bar{X}}}\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x - \frac{\omega^2}{n\sigma_{\bar{X}}^2}\int_{-\infty}^{+\infty}(X - \mu)^2 f_X(x)\mathrm{d}x\right]^n \\
&= \left[1 + \frac{i\omega}{\sqrt{n}\sigma_{\bar{X}}}\mu - \frac{i\mu\omega}{\sqrt{n}\sigma_{\bar{X}}} - \frac{\omega^2}{n\sigma_{\bar{X}}^2}\int_{-\infty}^{+\infty}(X - \mu)^2 f_X(x)\mathrm{d}x\right]^n \\
&= \left[1 - \frac{\omega^2}{n\sigma_{\bar{X}}^2}\int_{-\infty}^{+\infty}(X - \mu)^2 f_X(x)\mathrm{d}x\right]^n = \left[1 - \frac{\omega^2}{n\sigma_{\bar{X}}^2}V(X)\right]^n = \left[1 - \frac{\omega^2}{2n}\right]^n
\end{aligned} \tag{7.535}$$

19. Il peut s'agir alors de mesures de phénomènes physiques ou mécaniques indépendants par exemple.

Posons :

$$-\frac{\omega^2}{2n} = \frac{1}{x} \tag{7.536}$$

Nous avons alors :

$$\phi_{Z_n}(\omega) \simeq \left(1 - \frac{1}{x}\right)^{-\frac{\omega^2 x}{2}} = \left(\left(1 - \frac{1}{x}\right)^{x}\right)^{-\frac{\omega^2}{2}} \tag{7.537}$$

Nous avons donc quand n tend vers l'infini (cf. chapitre d'Analyse fonctionnelle) :

$$\phi_{Z_n}(\omega) \simeq \lim_{n\to+\infty} \left(1 - \frac{1}{x}\right)^{-\frac{\omega^2 x}{2}} = \left(\lim_{n\to+\infty} \left(1 - \frac{1}{x}\right)^{x}\right)^{-\frac{\omega^2}{2}} = \mathrm{e}^{-\frac{\omega^2}{2}} \tag{7.538}$$

Nous retrouvons donc la fonction caractéristique de la loi Normale centrée réduite. •

De manière simple, le Théorème Central Limite (TCL) dit que pour de grands échantillons, la somme centrée et réduite de n variables aléatoires identiquement distribuées suit une loi Normale centrée et réduite. Et donc nous avons in extenso pour la moyenne empirique :

$$\frac{M_n - \mu}{\sigma/\sqrt{n}} = N(0,1) \quad \Leftrightarrow \quad M_n = N(\mu, \sigma/\sqrt{n}) \tag{7.539}$$

Malgré l'immensité de son champ d'applications, le TCL n'est pas universel. Dans sa forme la plus simple, il impose en particulier à la variable considérée d'avoir des moments du premier et du deuxième ordre (moyenne et variance). Si tel n'est pas le cas, il ne s'applique plus.

L'exemple le plus simple d'échec du TLC est donné par la distribution de Cauchy, qui n'a ni moyenne, ni variance, et dont la moyenne empirique a toujours la même distribution (Cauchy) quelle que soit la taille de l'échantillon.

Maintenant, nous allons illustrer le théorème central limite dans le cas d'une suite (X_n) de variables aléatoires indépendantes discrètes suivant une loi de Bernoulli de paramètre $1/2$.

Nous pouvons imaginer que X_n représente le résultat obtenu au n^{e} lancé d'une pièce de monnaie (en attribuant le nombre 1 pour pile et 0 pour face). Notons :

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{7.540}$$

la moyenne. Nous avons pour tout n bien évidemment $E(X_n) = 1/2$ et $V(X_n) = 1/4$ et donc :

$$E(\bar{X}_n) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2} = \frac{1}{2} \qquad V(\bar{X}_n) = \frac{1}{n^2}\sum_{k=1}^{n}\frac{1}{4} = \frac{1}{4n} \tag{7.541}$$

Après avoir centré et réduit $\bar{X}_n$ nous obtenons :

$$\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = 2\sqrt{n}\left(\bar{X}_n - \frac{1}{2}\right) \tag{7.542}$$

Notons Φ la fonction de répartition de la loi Normale centrée réduite. Le théorème central limite nous dit que pour tout $t \in \mathbb{R}$:

$$\lim_{n\to+\infty} P\left(2\sqrt{n}\left(\bar{X}_n - \frac{1}{2}\right) \leqslant t\right) = \Phi(t) \tag{7.543}$$

A l'aide de Maple nous avons tracé en bleu quelques graphiques de la fonction :
equation (7.185)
pour différentes valeurs de n. Nous avons représenté en rouge la fonction equation.
equation :
equation (7.186)
equation :
equation (7.187)

equation
equation (7.188)
equation
equation (7.189)

Ces graphiques nous montrent bien la convergence de F_n vers Φ. En fait nous remarquons que la convergence est carrément uniforme ce qui est confirmé par le *théorème central limite de* Moivre-Laplace :

Théorème 7.1 (théorème central limite de* Moivre-Laplace*) *Soit X_n une suite de variables aléatoires indépendantes de même loi de* Bernoulli *de paramètre $0 < p < 1$. Alors :*

$$F_n(t) = P\left(2\sqrt{n}\left(\bar{X}_n - \frac{1}{2}\right) \leqslant t\right) \tag{7.544}$$

tend uniformément vers $\Phi(t)$ sur $\mathbb{R}$ lorsque $n \to +\infty$.

7.10 Tests d'hypothèse ou d'adéquation

Lors de notre étude des intervalles de confiance, rappelons nous sommes arrivées aux relations (7.451), (7.469), (7.477) et (7.484) qui permettent de faire de l'inférence statistique en fonction de la connaissance ou non de la moyenne ou de la variance vraie sur la totalité ou sur un échantillon de la population. En d'autres termes, de savoir dans quelles bornes se situait un moment (moyenne ou variance) en fonction d'un certain niveau de confiance imposé. Nous avions vu que le deuxième intervalle (7.469) ne peut être que difficilement utilisé dans la pratique (suppose la moyenne théorique connue) et nous lui préférons donc (7.477).

Nous allons également démontrer en détails plus loin les deux intervalles suivants :

$$\frac{1}{F_{\alpha/2}(n,m)}\frac{S_{*X}^2}{S_{*Y}^2} \leqslant \frac{\sigma_X^2}{\sigma_Y^2} \leqslant \frac{1}{F_{1-\alpha/2}(n,m)}\frac{S_{*X}^2}{S_{*Y}^2} \tag{7.545}$$

et :

$$\frac{1}{F_{\alpha/2}(n-1,m-1)}\frac{S_X^2}{S_Y^2} \leqslant \frac{\sigma_X^2}{\sigma_Y^2} \leqslant \frac{1}{F_{1-\alpha/2}(n-1,m-1)}\frac{S_X^2}{S_Y^2} \tag{7.546}$$

Le premier intervalle ci-dessus ne peut être lui aussi que difficilement utilisé dans la pratique (suppose la moyenne théorique connue) et nous lui préférons donc le deuxième.

Lorsque nous cherchons à savoir si nous pouvons faire confiance à la valeur d'un moment ou d'une variable aléatoire en général avec une certaine confiance, nous parlons de *test d'hypothèse* ou *test d'adéquation* ou encore de *test de conformité*.

Les tests d'hypothèses sont destinés à vérifier si un échantillon peut être considéré comme extrait d'une population donnée ou représentatif de cette population, vis-à-vis d'un paramètre comme la moyenne, la variance ou la fréquence observée. Ceci implique que la loi théorique du paramètre est connue au niveau de la population.

Par exemple, si nous souhaitons savoir avec une certaine confiance si une moyenne donnée d'un échantillon de population est réaliste par rapport à la vraie moyenne théorique inconnue, nous utiliserons le *test-Z* qui est simplement (7.451) si la moyenne de toute la population se trouve bien dans les bornes pour la confiance donnée, la moyenne de l'échantillon test de taille n avec l'écart-type de toute la population connue.

Maintenant rappelons que nous avons démontré que si nous avions deux variables aléatoires de loi $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ alors la soustraction (différencier) des moyennes donne $N(\mu_2 - \mu_1, \sqrt{\sigma_1^2 + \sigma_2^2})$. Par conséquent, pour la différence de deux moyennes de variables aléatoires provenant de deux échantillons de population nous obtenons directement :

$$N\left(\bar{X}_2 - \bar{X}_1, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \tag{7.547}$$

Nous pouvons alors adapter le test-Z sous la forme :

$$Z^{\text{calc}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} N = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{7.548}$$

Cette relation est très utile lorsque pour deux échantillons de deux populations de données, nous voulons vérifier s'il existe une différence significative des différences des moyennes théoriques à un niveau de confiance donné et la probabilité associée pour avoir cette différence par exemple donné par :

$$\bar{X}_2 - \bar{X}_1 - Z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leqslant \mu_2 - \mu_1 \leqslant \bar{X}_2 - \bar{X}_1 + Z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{7.549}$$

Donc :

$$|Z^{\text{calc}}| \leqslant |Z_{\alpha/2}| \tag{7.550}$$

Nous parlons du *test-Z de la moyenne à deux échantillons* et il est beaucoup utilisé dans l'industrie pour vérifier l'égalité de la moyenne de deux populations de mesures.

Si l'écart-type théorique n'est pas connu, nous utiliserons les *test-T* de Student (7.484). Dans la même idée pour l'écart-type, nous utiliserons le *test du khi-deux* (7.477). Lorsque nous voulons tester l'égalité de la variance de deux populations nous utilisons le *test-F* de Fisher (7.546).

Le fait que nous obtenions alors l'ensemble des valeurs satisfaisant à ce test borné à droite et à gauche est ce que nous appelons dans le cas général un *test bilatéral* car il comprend le test unilatéral à gauche et unilatéral à droite. Ainsi, tous les tests susmentionnés sont dans une forme bilatérale mais nous pourrions en faire une analyse unilatérale aussi.

Signalons aussi que les tests d'hypothèses sur l'écart-type (variance), la moyenne ou la corrélation sont appelés des *tests paramétriques* à l'inverse des tests non-paramétriques que nous verrons plus loin.

Remarque Il existe également une autre définition du concept de test paramétrique et non-paramétrique (complétement différente).

Enfin, de nombreux logiciels calculent ce que nous appelons la *p-value* qui est généralement le risque limite pour lequel nous passons de l'état d'hypothèse acceptée à l'état refusée.

Pour un test , le 5% de risque est le risque de rejeter l'hypothèse alors même qu'elle est vraie. Si le risque est 5% et que la p-value est supérieure, le test échoue (rejet de l'hypothèse). Nous rejetons l'hypothèse si la p-value est plus faible que 5%.

Remarque Nous ne devrions jamais dire que nous *acceptons* une hypothèse ou encore qu'elle soit *vraie* ou *fausse* car ces termes sont trop forts. Nous devons dire si nous *rejetons* ou non l'hypothèse et qu'elle est éventuellement *correcte* ou *non correcte*.

Nous allons dans ce qui suit démontrer l'origine du test F de Fisher et par la même occasion nous introduirons deux autres tests qui sont le *test-T homoscédastique* et le *test-T hétéroscédastique*.

7.10.1 Analyse de la variance

L'objectif de l'analyse de la variance est de comparer les moyennes de deux populations ou plus. Cette méthode, néanmoins, doit son nom au fait qu'elle utilise des mesures de variance afin de déterminer le caractère significatif, ou non, des différences de moyenne mesurées sur les populations.

Plus précisément, la vraie signification est de savoir si le fait que des moyennes d'échantillons sont (légèrement) différentes peut être attribué au hasard de l'échantillonnage ou provient du fait que les échantillons sont significativement différents (si nous avons les valeurs de toute la population, nous n'avons rien à faire.).

Pour l'analyse de la variance abrégée anova (ANalysis Of VAriance) ou anavar (ANAlyse de la VARiance) nous allons d'abord rappeler, comme nous l'avons démontré, que la loi de Fisher-Snedecor est donnée par le rapport de deux variables aléatoires indépendantes suivant une loi

du khi-deux et divisée par leur degré de liberté tel que :

$$\frac{U/k}{V/l} = F_{k,l}(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)\left(\frac{k}{l}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{l}{2}\right)} x^{\frac{k}{2}-1}\left(1+\frac{kx}{l}\right)^{-\frac{k+l}{2}} \tag{7.551}$$

et nous allons voir maintenant son importance.

Considérons un échantillons aléatoire de taille n, disons $X_1, \ldots, X_n$ issu de la loi $N(\mu_X, \sigma_X)$ et un échantillon aléatoire de taille m, disons $Y_1, \ldots, Y_n$ issu de la loi $N(\mu_Y, \sigma_Y)$.

Considérons les estimateurs de maximum de vraisemblance de l'écart-type de la loi Normale traditionnellement notée dans le domaine de l'analyse de la variance par :

$$S^2_{*X} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)^2 \quad \text{et} \quad S^2_{*Y} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_Y)^2 \tag{7.552}$$

Les statistiques ci-dessus sont les statistiques que nous utiliserions pour estimer les variances si les moyennes théoriques μ_X, μ_Y sont connues. Donc nous pouvons utiliser un résultat démontré plus haut lors de notre étude des intervalles de confiance :

$$\chi^2(n) = \frac{nS^2_{*X}}{\sigma^2_X} \quad \text{et} \quad \chi^2(m) = \frac{mS^2_{*Y}}{\sigma^2_Y} \tag{7.553}$$

Comme les Y_j sont indépendantes des X_i par hypothèse, les variables :

$$\frac{nS^2_{*X}}{\sigma^2_X} \quad \text{et} \quad \frac{mS^2_{*Y}}{\sigma^2_Y} \tag{7.554}$$

sont indépendantes l'une de l'autre.

Nous pouvons donc appliquer la loi de FISHER-SNEDECOR avec :

$$U = \frac{nS^2_{*X}}{\sigma^2_X};\ k = n \quad \text{et} \quad V = \frac{mS^2_{*Y}}{\sigma^2_Y};\ l = m \tag{7.555}$$

Nous avons donc :

$$\frac{U/k}{V/l} = \frac{\frac{nS^2_{*X}}{\sigma^2_X}/n}{\frac{mS^2_{*Y}}{\sigma^2_Y}/m} = \frac{\frac{S^2_{*X}}{\sigma^2_X}}{\frac{S^2_{*Y}}{\sigma^2_Y}} = \frac{S^2_{*X}\sigma^2_Y}{S^2_{*Y}\sigma^2_X} = F_{n,m}(x) \tag{7.556}$$

Soit :

$$\frac{\sigma^2_X}{\sigma^2_Y} = \frac{1}{F_{n,m}(x)}\frac{S^2_{*X}}{S^2_{*Y}} \tag{7.557}$$

Ce théorème nous permet de déduire l'intervalle de confiance du rapport de deux variances lorsque la moyenne théorique est connue. Puisque la fonction du FISHER n'est pas symétrique, la seule possibilité pour faire l'inférence c'est de faire appel au calcul numérique et nous noterons alors pour un intervalle de confiance donné le test de la manière suivante :

$$\frac{1}{F_{\alpha/2}(n,m)}\frac{S^2_{*X}}{S^2_{*Y}} \leqslant \frac{\sigma^2_X}{\sigma^2_Y} \leqslant \frac{1}{F_{1-\alpha/2}(n,m)}\frac{S^2_{*X}}{S^2_{*Y}} \tag{7.558}$$

Dans le cas où les moyennes μ_X, μ_Y sont inconnues, nous utilisons l'estimateur sans biais de la variance traditionnellement notée dans le domaine de l'analyse de la variance par :

$$S^2_X = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{et} \quad S^2_Y = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \bar{Y})^2 \tag{7.559}$$

Pour estimer les variances théoriques, nous utilisons le résultat démontré plus haut :

$$\frac{(n-1)S^2_X}{\sigma^2_X} = \chi^2(n-1) \quad \text{et} \quad \frac{(m-1)S^2_Y}{\sigma^2_Y} = \chi^2(m-1) \tag{7.560}$$

Comme les Y_j sont indépendantes des X_i par hypothèse, les variables :

$$\frac{(n-1)S_X^2}{\sigma_X^2} \quad \text{et} \quad \frac{(m-1)S_Y^2}{\sigma_Y^2} \tag{7.561}$$

sont indépendantes l'une de l'autre. Nous pouvons donc appliquer la loi de Fisher-Snedecor avec :

$$U = \frac{(n-1)S_X^2}{\sigma_X^2};\ k = n-1 \quad \text{et} \quad V = \frac{(m-1)S_Y^2}{\sigma_Y^2};\ l = m-1 \tag{7.562}$$

Nous avons donc :

$$\frac{U/k}{V/l} = \frac{\frac{(n-1)S_X^2}{\sigma_X^2}/(n-1)}{\frac{(m-1)S_Y^2}{\sigma_Y^2}/(m-1)} = \frac{\frac{S_X^2}{\sigma_X^2}}{\frac{S_Y^2}{\sigma_Y^2}} = \frac{S_X^2\sigma_Y^2}{S_Y^2\sigma_X^2} = F_{n-1,m-1}(x) \tag{7.563}$$

Soit :

$$\frac{\sigma_X^2}{\sigma_Y^2} = \frac{1}{F_{n-1,m-1}(x)}\frac{S_X^2}{S_Y^2} \tag{7.564}$$

Ce théorème nous permet de déduire l'intervalle de confiance du rapport de deux variances lorsque la moyenne empirique est connue. Puisque la fonction du Fisher n'est pas symétrique, la seule possibilité pour faire l'inférence c'est de faire appel au calcul numérique et nous noterons alors pour un intervalle de confiance donné le test de la manière suivante :

$$\frac{1}{F_{\alpha/2}(n-1,m-1)}\frac{S_X^2}{S_Y^2} \leqslant \frac{\sigma_X^2}{\sigma_Y^2} \leqslant \frac{1}{F_{1-\alpha/2}(n-1,m-1)}\frac{S_X^2}{S_Y^2} \tag{7.565}$$

R. Fisher (1890-1962) est, comme K. Pearson, l'un des principaux fondateurs de la théorie moderne de la statistique. Fisher étudia à Cambridge où il obtint en 1912 un diplôme en astronomie. C'est en étudiant la théorie de l'erreur dans les observations astronomiques que Fisher s'intéressa à la statistique. Fisher est l'inventeur de la branche de la statistique appelée l'analyse de la variance.

Au début du 20^{e} siècle, R. Fischer développe donc la méthodologie des plans d'expérience. Pour valider l'utilité d'un facteur, il met au point un test permettant d'assurer que des échantillons différents sont de natures différentes. Ce test est basé sur l'analyse de la variance des échantillons, et nommé anova (analyse normalisée de la variance).

Prenons k échantillons de n valeurs aléatoires chacun, appelé *facteur explicatif* dans l'analyse de la variance. Chacune des valeurs étant considérée comme une observation ou une mesure de quelque chose. Nous aurons donc un nombre total de N d'observations donnée par $N = k \cdot n$.

Si chacun des échantillons a un nombre identique de valeurs tel que $n_1 = n_2 = \ldots = n_{k!}$. Nous considérerons que chacun des k échantillons est issu (suit) d'une variable aléatoire suivant une loi Normale.

En termes de test, nous voulons tester si les moyennes des k échantillons sont égales sous l'hypothèse que leurs variances sont égales. Ce que nous écrivons sous forme d'hypothèse de la manière suivante :

1. $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$
2. $\exists i, j\ \mu_i = \mu_j$

Autrement dit, les échantillons sont représentatifs d'une même population (d'une même loi statistique). C'est-à-dire que les variations constatées entre les valeurs des différents échantillons sont dues essentiellement au hasard. Pour cela nous étudions la variabilité des résultats dans les échantillons et entre les échantillons.

Nous noterons i l'indice d'échantillon (de 1 à k) et j l'indice de l'observation (de 1 à n). Donc x_{ij} sera la valeur de la j^{e} observation de l'échantillon de données numéro i.

Selon l'hypothèse susmentionnée, nous avons :

$$x_{ij} = N(\mu_i, \sigma^2) \tag{7.566}$$

Nous noterons par $\bar{x}_i$ la moyenne empirique/estimée (arithmétique) de l'échantillon i :

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \quad \Rightarrow \quad = \sum_{j=1}^{n_i} x_{ij} = n_i \bar{x}_i \tag{7.567}$$

et $\bar{\bar{x}}$ la moyenne empirique/estimée des N valeurs (soit la moyenne des $\bar{x}_i$) donnée donc par :

$$\bar{\bar{x}} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} = \frac{1}{N}\sum_{i=1}^{k}(n_i \bar{x}_i) \tag{7.568}$$

En utilisant les propriétés de l'espérance et de la variance déjà démontrées plus haut nous savons que :

$$\bar{x}_i = N(\mu_i, \sigma_i^2/n_i) \quad \text{et} \quad \bar{\bar{x}} = N(\mu, \sigma^2/n) \tag{7.569}$$

avec μ qui est la moyenne des moyennes vraies μ_i :

$$\mu = \frac{1}{N}\sum_{i=1}^{k} \mu_i \tag{7.570}$$

Maintenant, introduisons la *variance totale* comme étant la variance estimée sans biais en considérant l'ensemble des N observations comme un seul échantillon :

$$V_T = \frac{\sum_{ij}(x_{ij} - \bar{\bar{x}})^2}{N-1} \tag{7.571}$$

où rappelons que le terme au numérateur est appelé *variation totale*.

La variance entre échantillons (c'est-à-dire entre les moyennes des échantillons) est l'estimateur de la variance des moyennes des échantillons :

$$V_E = \frac{\sum_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} \tag{7.572}$$

Comme nous avons démontré que si toutes les variables sont identiquement distribuées (même variance) la variance des individus vaut n fois celle de la moyenne :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad V(\bar{X})n = V(X) \tag{7.573}$$

alors la variance des observations (variables aléatoires dans un échantillon) est donnée par :

$$V_A = nV_E = \frac{n\sum_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} \tag{7.574}$$

Nous avons donc ci-dessus l'hypothèse de l'égalité des variances qui est exprimée sous forme mathématique pour les développements à suivre.

La variance résiduelle est l'effet des facteurs dits non contrôlés. C'est par définition la moyenne des variances des échantillons.

$$V_R = \frac{1}{k}\sum_i \frac{\sum_j(x_{ij} - \bar{x}_i)^2}{n-1} = \frac{\sum_{ij}(x_{ij} - \bar{x}_i)^2}{k(n-1)} \tag{7.575}$$

Au final, ces indicateurs sont parfois résumés sous la forme suivante :

$$\begin{aligned}
&Q_T = \sum_{ij}(x_{ij} - \bar{\bar{x}})^2 \quad && ddl_T = N - 1 = kn - 1 \quad && V_T = \frac{Q_T}{ddl_T} \\
&Q_A = n\sum_i(\bar{x}_i - \bar{\bar{x}})^2 \quad && ddl_A = k - 1 && V_A = \frac{Q_A}{ddl_A} \\
&Q_R = \sum_{ij}(x_{ij} - \bar{x}_i)^2 \quad && ddl_R = k(n-1) && V_R = \frac{Q_R}{ddl_R}
\end{aligned} \tag{7.576}$$

Remarquons que si les échantillons n'ont pas la même taille (ce qui est rare), nous avons alors :

$$\begin{aligned} Q_T &= \sum_{ij}(x_{ij}-\bar{\bar{x}})^2 \quad ddl_T = N-1 \qquad V_T = \frac{Q_T}{ddl_T} \\ Q_A &= n\sum_{i}(\bar{x}_i-\bar{\bar{x}})^2 \quad ddl_A = k-1 \qquad V_A = \frac{Q_A}{ddl_A} \\ Q_R &= \sum_{ij}(x_{ij}-\bar{x}_i)^2 \quad ddl_R = \sum_{i}(n_i-1) \quad V_R = \frac{Q_R}{ddl_R} \end{aligned} \tag{7.577}$$

Remarques

1. Le terme Q_T est souvent indiqué dans l'industrie par l'abréviation SST signifiant en anglais *Sum of Squares Total* ou plus rarement TSS pour *Total Sum of Square*.
2. Le terme Q_A est souvent indiqué dans l'industrie par l'abréviation SSB signifiant en anglais "Sum of Squares Between" (samples) ou plus rarement SSk pour "Sum of Squared Between treatments".
3. Le terme Q_R est souvent indiqué dans l'industrie par l'abréviation SSW signifiant en anglais "Sum of Squares Within" (samples) ou plus rarement SSE pour "Sum of Squared due to Errors".

Indiquons que nous voyons souvent dans la littérature (nous réutiliserons un peu plus loin cette notation) :

$$\begin{aligned} Q_R &= \sum_{ij}(x_{ij}-\bar{x}_i)^2 = \frac{n_i-1}{n_i-1}\sum_{ij}(x_{ij}-\bar{x}_i)^2 = \sum_{i}^{k}\sum_{j}^{n_i}(x_{ij}-\bar{x}_i)^2 \\ &= \sum_{i}^{k}(n_i-1)\sum_{j}^{n_i}\frac{(x_{ij}-\bar{x}_i)^2}{n_i-1} = \sum_{i}^{k}(n_i-1)S_i^2 \end{aligned} \tag{7.578}$$

avec donc l'estimateur sans biais de la variance des observations :

$$S_i^2 = \frac{1}{n_i-1}\sum_{j}^{n_i}(x_{ij}-\bar{x}_i)^2 \tag{7.579}$$

Avant d'aller plus loin, arrêtons-nous sur la variance résiduelle. Nous avons donc pour des échantillons qui ne sont pas de même taille :

$$V_R = \frac{\sum_{ij}(x_{ij}-\bar{x}_i)^2}{\sum_i(n_i-1)} \tag{7.580}$$

Ouvrons maintenant une petite parenthèse. Prenons le cas particulier deux échantillons seulement. Nous pouvons alors écrire :

$$V_R = \frac{\sum_j(x_{1j}-\bar{x}_1)^2+\sum_j(x_{2j}-\bar{x}_2)^2}{(n_1-1)+(n_2-1)} \tag{7.581}$$

Soit en introduisant l'estimateur de maximum de vraisemblance de la variance :

$$V_R = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \tag{7.582}$$

Nous pouvons d'ailleurs observer que dans le cas particulier où $n_1 = n_2 = n$ alors :

$$V_R = \frac{nS_1^2-S_1^2+nS_2^2-S_2^2}{2n-2} = \frac{S_1^2+S_2^2}{2} \tag{7.583}$$

Donc :

$$S_1^2 \leqslant V_R \leqslant S_2^2 \tag{7.584}$$

Supposons maintenant que nous souhaitions comparer avec un certain intervalle de confiance la moyenne de deux populations ayant une variance différente pour savoir si elles sont de natures différentes ou non.

Nous connaissons pour le moment deux tests pour vérifier les moyennes. Le test-Z et le test-T. Comme dans l'industrie il est rare que nous ayons le temps de prendre des grands échantillons, concentrons-nous sur le deuxième, soit (7.469) et rappelons aussi que :

$$S_{\bar{X}} = \frac{S}{\sqrt{n}} \tag{7.585}$$

Maintenant rappelons que nous avons démontré que si nous avions deux variables aléatoires de lois $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$ alors la soustraction des moyennes donne $N(\mu_2 - \mu_1, \sqrt{\sigma_1^2 + \sigma_2^2})$. Donc pour la différence de deux moyennes de variables aléatoires provenant de deux échantillons de population nous obtenons directement :

$$N(\bar{X}_2 - \bar{X}_1, \sqrt{S_{\bar{X}_1}^2 + S_{\bar{X}_2}^2}) = N\left(\bar{X}_2 - \bar{X}_1, \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) \tag{7.586}$$

Et maintenant l'idée est de prendre l'approximation (sous l'hypothèse que les variances sont égales) :

$$N\left(\bar{X}_2 - \bar{X}_1, \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) \simeq N\left(\bar{X}_2 - \bar{X}_1, \sqrt{\frac{V_R}{n_1} + \frac{V_R}{n_2}}\right) = N\left(\bar{X}_2 - \bar{X}_1, \sqrt{V_R\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right) \tag{7.587}$$

Cette approximation est appelée *hypothèse homoscédastique*.

Nous avons alors l'intervalle de confiance :

$$(\bar{X}_2 - \bar{X}_1) - \sqrt{V_R\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} T_{\alpha/2}(n-1) \leqslant \mu_2 - \mu_1 \leqslant (\bar{X}_2 - \bar{X}_1) + \sqrt{V_R\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} T_{\alpha/2}(n-1) \tag{7.588}$$

Comme l'idée est de tester l'égalité des moyennes théoriques à partir des estimateurs connus alors :

$$(\bar{X}_2 - \bar{X}_1) - \sqrt{V_R\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} T_{\alpha/2}(n-1) \leqslant 0 \leqslant (\bar{X}_2 - \bar{X}_1) + \sqrt{V_R\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} T_{\alpha/2}(n-1) \tag{7.589}$$

avec $n = n_1 + n_2$.

Dans la plupart des logiciels disponibles sur le marché, le résultat est uniquement donné à partir du fait que le $T_{\alpha/2}$ que nous avons est compris dans le $T_{\alpha/2}$ correspondant à l'intervalle de confiance donné rappelons-le par :

$$T^{\text{calc}}(n-1) = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{V_R\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{7.590}$$

dans le cas de l'hypothèse homoscédastique (égalité des variances) sinon par :

$$T^{\text{calc}}(n-1) = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{7.591}$$

dans l'hypothèse hétéroscédasticité (non égalité des variances).

Donc :

$$|T^{\text{calc}}(n-1)| \leqslant |T_{\alpha/2}(n-1)| \tag{7.592}$$

Si nous faisons ce test avec deux échantillons à variances égales, nous parlons du *t-test homoscédastique*, sinon du *test-t hétéroscédastique*.

Reprenons l'équation (7.576) où nous avons donc le cas d'échantillons de même taille $ddl_T = ddl_1 + ddl_R = k - 1 + k(n-1) = kn - 1$. Ainsi que la variance totale qui est la somme de la variance

des moyennes (interclasses) et de la variance résiduelle (intra-classes) et ce que les échantillons soient de même taille ou non :

$$Q_T = Q_A + Q_R \tag{7.593}$$

Effectivement :

$$\begin{aligned} Q_T &= \sum_{ij}(x_{ij} - \bar{\bar{x}})^2 = \sum_i \sum_j \left((x_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{\bar{x}})\right)^2 \\ &\sum_i \sum_j (x_{ij} - \bar{x}_i)^2 + \sum_i \sum_j \left(\bar{x}_i - \bar{\bar{x}}\right)^2 + 2\sum_i \sum_j (x_{ij} - \bar{x}_i)\left(\bar{x}_i - \bar{\bar{x}}\right) \end{aligned} \tag{7.594}$$

Or, nous avons :

$$\sum_i \sum_j (x_{ij} - \bar{x}_i)\left(\bar{x}_i - \bar{\bar{x}}\right) = \sum_i \left[\left(\bar{x}_i - \bar{\bar{x}}\right)\sum_j (x_{ij} - \bar{x}_i)\right] = 0 \tag{7.595}$$

car :

$$\sum_j (x_{ij} - \bar{x}_i) = 0 \quad \forall i \tag{7.596}$$

Donc :

$$Q_T = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 + \sum_i \sum_j \left(\bar{x}_i - \bar{\bar{x}}\right)^2 = Q_R + Q_A \tag{7.597}$$

Sous les hypothèses mentionnées au début (égalité des moyennes entre échantillons) nous avons :

$$\frac{Q_R}{\sigma^2} = \frac{\sum_{ij}(x_{ij} - \bar{x}_i)^2}{\sigma^2} = \chi^2_{k-1} = Q_R + Q_A \tag{7.598}$$

ce qui découle immédiatement de la démonstration que nous avions fait lors de notre étude de l'inférence statistique (7.476). Ce que nous souhaitons faire c'est voir s'il y a une différence entre la variance des moyennes (interclasses) et de la variance résiduelle (intra-classes). Pour comparer deux variances lorsque les moyennes vraies sont inconnues nous avons vu que le mieux était d'utiliser le test de Fisher.

De même, nous avons :

$$\frac{Q_A}{\sigma^2} = \frac{\sum_i n_i \left(\bar{x}_i - \bar{\bar{x}}\right)^2}{\sigma^2} = \chi^2_{N-k} \tag{7.599}$$

Effectivement, l'identité (7.476) entraîne :

$$\chi^2_{n_i-1} = \frac{(n_i - 1)S_i^2}{\sigma^2} \tag{7.600}$$

Donc de par la linéarité de la loi du Khi-deux :

$$\sum_{i=1}^{k} \frac{(n_i - 1)S_i^2}{\sigma^2} = \chi^2_{N-k} \tag{7.601}$$

puisque :

$$\chi^2_{N-k} = \chi^2_{n_1-1} + \chi^2_{n_2-1} + \ldots + \chi^2_{n_k-1} \tag{7.602}$$

et parce que :

$$\sum_{i=1}^{k}(n_i - 1) = N - k \tag{7.603}$$

Donc pour résumer nous avons :

$$\frac{Q_A}{\sigma^2} = \chi^2_{N-k} \quad \text{et} \quad \frac{Q_R}{\sigma^2} = \chi^2_{k-1} \tag{7.604}$$

et d'après (7.597), nous avons alors :

$$\frac{Q_T}{\sigma^2} = \chi^2_{N-k} + \chi^2_{k-1} = \chi^2_{N-1} \tag{7.605}$$

C'est maintenant qu'intervient la loi de FISHER dans l'hypothèse où les variances sont égales ! Puisque :

$$\frac{(N-k)S_A^2}{\sigma^2} = \chi^2_{N-k} \quad \text{et} \quad \frac{(k-1)S_R^2}{\sigma^2} = \chi^2_{k-1} \tag{7.606}$$

Or, nous avons démontré dans notre étude de la loi de FISHER le résultat (7.564) où dans notre cas d'étude :

$$\frac{\sigma_A^2}{\sigma_R^2} = 1 \Rightarrow F_{k-1,N-k} = \frac{S_A^2}{S_R^2} = \frac{V_A}{V_R} = \frac{\dfrac{\sum_i n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1}}{\dfrac{\sum_{ij}(x_{ij} - \bar{x}_i)^2}{\sum_i (n_i - 1)}} = \frac{\dfrac{Q_A}{k-1}}{\dfrac{Q_R}{N-k}} \tag{7.607}$$

et souvent indiquée dans la littérature sous la formation suivante :

$$F_{k-1,N-k} = \frac{MSK}{MSE} \tag{7.608}$$

où MSK est appelé "Mean Square for treatments" et MSE "Mean Square for Error".

REMARQUE S'il y a seulement deux populations, il faut bien comprendre qu'à ce moment l'utilisation du test-T de Student suffit amplement.

Tout les calculs que nous avons fait sont très souvent représentés dans les logiciels sous la forme d'une table standardisée donc voici la forme et le contenu :

Source	Somme des carrés	ddl du χ^2	Moyenne des carrés	F	Valeur critique F
Inter-Classe	$Q_A = \sum_i n_i \left(\bar{x}_i - \bar{\bar{x}}\right)^2$	$k-1$	$MSK = \frac{Q_A}{k-1}$	$\frac{MSK}{MSE}$	$P(F > F_{k-1,N-k})$
Intra-Classe	$Q_R = \sum_{ij} \left(x_{ij} - \bar{x}_i\right)^2$	$N-k$	$MSE = \frac{Q_R}{k-1}$		
Total	$Q_T = \sum_{ij} \left(x_{ij} - \bar{\bar{x}}\right)^2$	$N-1$			

TABLEAU 7.5 – (7.279)

ainsi, pour que l'hypothèse soit acceptée, il faut que la valeur de (7.608) soit plus petite ou égale à au centile de la même loi F avec une probabilité cumulée à l'intervalle de confiance imposé.

7.10.2 Test d'ajustement du khi-deux

Nous allons étudier ici notre premier test d'hypothèse non-paramétrique, un des plus connus certainement et des plus simples.

Supposons qu'une variable statistique suive une loi de probabilité P. Si nous tirons un échantillon dans la population correspondant à cette loi, la distribution observée s'écartera toujours plus ou moins de la distribution théorique, compte tenu des fluctuations d'échantillonnage.

Généralement, nous ne connaissons ni la forme de la loi P, ni la valeur de ses paramètres. C'est la nature du phénomène étudié et l'analyse de la distribution observée qui permettent de choisir une loi susceptible de convenir et d'en estimer les paramètres.

Les écarts entre la loi théorique et la distribution observée peuvent être attribués soit aux fluctuations d'échantillonnage, soit au fait que le phénomène ne suit pas, en réalité, la loi supposée.

En gros, si les écarts sont suffisamment faibles, nous admettrons qu'ils sont imputables aux fluctuations aléatoires et nous accepterons la loi retenue ; au contraire, s'ils sont trop élevés, nous en conclurons qu'ils ne peuvent pas être expliqués par les seules fluctuations et que le phénomène ne suit pas la loi retenue.

Pour évaluer ces écarts et pouvoir prendre une décision, il faut :

1. Définir la mesure de la distance entre distribution empirique et distribution théorique résultant de la loi retenue.
2. Déterminer la loi de probabilité suivie par cette variable aléatoire donnant la distance
3. Énoncer une règle de décision permettant de dire, d'après la distribution observée, si la loi retenue est acceptable ou non.

Nous aurons pour cela besoin du théorème central limite et deuxièmement rappelons que lors lors de la construction de la loi Normale, nous avons montré que la variable :

$$x = \frac{k - np}{\sqrt{npq}} = \frac{k - np}{\sqrt{np(1-p)}} \tag{7.609}$$

suivait une loi Normale centrée réduite lorsque n tendait vers l'infini (condition de Laplace).

En pratique, l'approximation est tout à fait acceptable... dans certaines entreprises... lorsque $np > 5$ et $p \leqslant 0{,}5$ soit (c'était un des termes qui devait tendre vers zéro quand nous avions fait la démonstration) :

$$\frac{q}{\sqrt{npq}} \simeq 0{,}3 \tag{7.610}$$

Par exemple dans les deux figures ci-dessous où nous avons représenté les lois binomiales approchées par les lois Normales associées, nous avons à gauche equation et à droite equation :

equation (7.283)

Rappelons enfin, que nous avons démontré que le sommes des carrées de n variables aléatoires normales centrées réduites linéairement indépendantes suit une loi du khi-deux à n degrés de liberté noté $\chi^2(n)$.

Considérons maintenant une variable aléatoire X suivant une fonction de distribution théorique (continue ou discrète) P et tirons un échantillon de taille n dans la population correspondant à cette loi P.

Les n observations seront réparties suivant k modalités (classes de valeurs) C_1, C_2, ..., C_k, dont les probabilités p_1, p_2, ..., p_k sont déterminées par la fonction de distribution P (se référer à l'exemple de la droite de Henry).

Pour chaque modalité C_i, l'effectif empirique est lui une variable aléatoire k_i de loi binomiale $\beta(n, p_i)$. Cet effectif k_i correspond en effet au nombre de succès *résultat égal à la modalité* C_i de probabilité p_i, obtenus au cours des n tirages d'un lot expérimental (et non dans la population de la loi théorique comme avant).

Nous avons démontré lors de l'étude de la loi binomiale que son espérance :

$$E(k_i) = np_i \tag{7.611}$$

représente l'effectif théorique de la modalité C_i et sa variance est :

$$V(k_i) = np_iq_i = np_i(1 - p_i) \simeq np_i \tag{7.612}$$

car p_i est relativement petite, ce qui donne $1 - p_i$ qui est assez proche de 1. Son écart-type est donc :

$$\sigma(k_i) = \sqrt{np_i} \tag{7.613}$$

Dans ces conditions, pourvu que la modalité C_i ait un effectif théorique np_i au moins égal à 5, l'écart réduit :

$$E_i = \frac{k_i - np_i}{\sqrt{np_i}} = \frac{k_i - \mu_i}{\sigma_i} \tag{7.614}$$

entre effectif empirique et effectif théorique peut être approximativement considéré comme une variable normale centrée réduite comme nous l'avons vu plus haut.

Nous définissons alors la variable :

$$D = \sum_{i=1}^{N} E_i^2 = \sum_{i=1}^{N} \frac{(k_i - np_i)^2}{np_i} \tag{7.615}$$

où k_i est souvent nommée *fréquence expérimentale* et np_i *fréquence théorique*.

Signalons que cette variable est aussi parfois (un peu malheureusement) notée :

$$D = \sum_{i=1}^{N} \frac{(n_i - n'_i)^2}{n'_i} \tag{7.616}$$

ou le plus souvent :

$$D = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} \tag{7.617}$$

Cette variable D, somme des carrés des variables E_i, nous donne une mesure de ce que nous pourrions appeler une *distance* ou *différence* ou *écart* entre distribution empirique et distribution théorique. Notons bien cependant qu'il ne s'agit cependant pas d'une distance au sens mathématique habituel (topologique).

Rappelons que D peut donc aussi s'écrire :

$$D = \sum_{i=1}^{N} \frac{(k_i - np_i)^2}{np_i} = \sum_{i=1}^{N} \frac{(k_i - \mu_i)^2}{\sigma_i^2} \tag{7.618}$$

D est donc la somme des carrés de N variables aléatoires normales centrées réduites liées par la seule relation linéaire $k_1 + k_2 + \ldots + k_N = n$ où n est donc la taille de l'échantillon. Donc D suit une loi khi-deux mais à $N - 1$ degrés de liberté, donc un degré de moins à cause de l'unique relation linéaire qui les lie. Effectivement, rappelons que le degré de liberté indique le nombre de variables indépendantes dans la somme et non pas juste le nombre de termes sommés.

Donc :

$$\chi^2(N-1) = \sum_{i=1}^{N} \frac{(k_i - np_i)^2}{np_i} \tag{7.619}$$

Nous appelons ce test un *test non-paramétrique du khi-2* ou test du khi-2 de PEARSON ou encore *test d'ajustement du khi-2*.

Ensuite, l'habitude est de déterminer la valeur de la loi du khi-deux à $N - 1$ degrés de liberté ayant 5% de probabilité d'être dépassée. Donc dans l'hypothèse où le phénomène étudié suit la loi théorique P, il y a donc 95% de probabilité cumulée que la variable D prenne une valeur inférieur à celle donnée par la loi du khi-deux.

Si la valeur de la loi du khi-deux obtenu à partir de l'échantillon prélevé est inférieure à celle correspondant aux 95% de probabilité cumulée, nous acceptons l'hypothèse selon laquelle le phénomène suit la loi P.

REMARQUES

1. Le fait que l'hypothèse de la loi P soit acceptée ne signifie pas pour autant que cette hypothèse soit vraie, mais simplement que les informations données par l'échantillon ne permettent pas de la rejeter. De même, le fait que l'hypothèse de la loi P soit rejetée ne signifie pas nécessairement que cette hypothèse soit fausse mais que les informations données par l'échantillon conduisent plutôt à conclure à l'inadéquation d'une telle loi.
2. Pour que la variable D suive une loi du khi-deux, il est nécessaire que les effectifs théoriques np_i des différentes modalités C_i soient au moins égaux à 5, que l'échantillon soit tiré au hasard (pas d'autocorrélation) et qu'aucune des probabilités p_i soit trop proche de zéro.

7.11 Calculs d'erreurs/incertitudes

Il est impossible de connaître (mesurer) la valeur exacte d'une grandeur physique expérimentalement, il est très important donc d'en déterminer l'incertitude.

Nous appelons bien évidemment *erreur*, la différence entre la valeur mesurée et la valeur exacte. Cependant, comme nous ignorons la valeur exacte, nous ne pouvons pas connaître l'erreur commise quand même.... Le résultat est donc toujours incertain. C'est la raison pour laquelle nous parlons des *incertitudes de mesure*.

Nous distinguons deux types d'incertitudes :

1. les *erreurs systématiques* : elles affectent le résultat constamment et dans le même sens (erreurs des appareils de mesures, limites de précision, etc.). Il faut alors éliminer, ou corriger le résultat, si possible ;
2. les *erreurs accidentelles* (statistiques) : il faut alors répéter les mesures, calculer la moyenne et évaluer l'incertitude en utilisant les outils de statistique.

Le deuxième type d'erreurs faits un très gros usage de tous les outils statistiques que nous avons présentés jusqu'à maintenant. Nous ne reviendrons donc pas dessus et nous concentrerons alors uniquement sur quelques nouveaux concepts.

7.11.1 Incertitudes absolues et relatives

Si la vraie valeur d'une grandeur est x (connue théoriquement) et la valeur mesurée est x_0, Δx_0 est *l'incertitude absolue* (l'incertitude dû aux appareils de mesure) telle que :

$$x_0 - \Delta x_0 \leqslant x \leqslant x_0 + \Delta x_0 \tag{7.620}$$

Le résultat s'écrit alors :

$$x = x_0 \pm \Delta x_0 \tag{7.621}$$

L'incertitude relative est quant à elle définie par :

$$\frac{\Delta x_0}{x_0} \tag{7.622}$$

L'incertitude absolue permet de savoir l'approximation du dernier chiffre significatif de celle-ci. Par contre, lorsque nous désirons comparer deux mesures ayant des incertitudes absolues afin de déceler lequel a la plus grande marge d'erreur, nous calculons l'incertitude relative de ce nombre en divisant l'incertitude absolue par le nombre, et transformé en pourcentage.

En d'autres termes, l'incertitude relative permet d'avoir une idée de la précision de la mesure en %. Si nous faisons une mesure avec une incertitude absolue de 1 [mm], nous ne saurons pas si c'est une bonne mesure ou non. Ça dépend si nous avons mesuré la taille d'une pièce de monnaie, de notre voisin, de la distance Paris-Marseille ou de la distance Terre-Lune. Bref, ça dépend de l'incertitude relative (c'est-à-dire du rapport de l'incertitude absolue sur la mesure).

7.11.2 Erreurs statistiques

Dans la plupart des mesures, nous pouvons estimer l'erreur due à des phénomènes aléatoires par une série de n mesures $x_1, x_2, \ldots, x_i, \ldots, x_n$. Comme nous l'avons vu plus haut, la valeur moyenne arithmétique sera alors :

$$\bar{x} = \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7.623}$$

et l'écart moyen (estimateur biaisé démontré plus haut) :

$$\hat{\sigma}^2 = \bar{\Delta x} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{7.624}$$

et l'écart quadratique moyen ou écart-type (estimateur sans biais) :

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{7.625}$$

et nous avions démontré que l'écart-type de la moyenne était donné par :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{7.626}$$

et comme nous l'avons vu, après un grand nombre de mesures indépendantes, la distribution des erreurs sur une mesure suit une gaussienne tel que nous puissions écrire :

$$\begin{aligned} &\bar{x}_i - \sigma < x_i < \bar{x}_i + \sigma \quad 68\% \\ &\bar{x}_i - 2\sigma < x_i < \bar{x}_i + 2\sigma \quad 95{,}5\% \end{aligned} \tag{7.627}$$

bref nous pouvons réutiliser tous les outils statistiques vus jusqu'à maintenant.

7.11.3 Propagation des erreurs

Soit une mesure $x + \pm\Delta x$ et $y = f(x)$ une fonction de x. Quelle est l'incertitude sur y ? Lorsque Δx est petit, $f(x)$ est remplacée au voisinage de x par sa tangente :

$$\Delta y = \left| \frac{\mathrm{d}f(x)}{\mathrm{d}x} \right| \Delta x \tag{7.628}$$

mais si y dépend de plusieurs grandeurs x, z, t mesurées avec les incertitudes $\Delta x, \Delta z, \Delta t$, soit $y = f(x, z, t)$ alors l'erreur maximale possible est alors la différentielle totale exacte (cf. chapitre de Calcul Différentiel Et Intégral) :

$$\Delta y = \left| \frac{\partial f(x,z,t)}{\partial x} \right| \Delta x + \left| \frac{\partial f(x,z,t)}{\partial z} \right| \Delta z + \left| \frac{\partial f(x,z,t)}{\partial t} \right| \Delta t \tag{7.629}$$

REMARQUE Le résultat d'une multiplication, d'une division, d'une soustraction ou d'une addition est arrondi à autant de chiffres significatifs que la donnée qui en comporte le moins.

7.11.4 Chiffres significatifs

Dans les petites écoles (et aussi les plus grande parfois), il est demandé de transformer une mesure exprimée en une certaine unité en une autre unité.

Par exemple, en prenant les tables, nous pouvons avoir le type de conversion suivante :

$$140\ [\mathrm{lb}] = 140 \times 0{,}45359237\ [\mathrm{kg/lb}] = 63{,}5029318\ [\mathrm{kg}] \tag{7.630}$$

Vient alors la question suivante : au départ d'une mesure dont la précision est de l'ordre de 1 [lb], une simple conversion d'unité pourrait-elle amener à une précision au 1/10 [mg] près ?

De cet exemple, il faut donc retenir qu'une marge d'incertitude est associée à toute valeur mesurée et à toute valeur calculée à partir de valeurs mesurées.

Dans les sciences exactes, tout raisonnement, toute analyse doit prendre cette incertitude en compte.

Mais pourquoi des chiffres sont-ils significatifs et d'autres pas ? Parce qu'en sciences, nous ne rapportons que ce qui a objectivement été observé (principe d'objectivité). En conséquence, nous limitons l'écriture d'un nombre aux chiffres raisonnablement fiables en dépit de l'incertitude : les chiffres significatifs. La précision que des chiffres supplémentaires sembleraient apporter est alors illusoire.

Il faut alors savoir arrondir selon des règles et conventions. Par exemple, lorsque le chiffre de rang le plus élevé qu'on laisse tomber est supérieur à 5, le chiffre précédent est augmenté de 1 (exemple : 12,66 s'arrondit à 12,7). Lorsque le chiffre de rang le plus élevé qu'on laisse tomber est inférieur à 5, le chiffre précédent reste inchangé (exemple 12,64 s'arrondit à 12,6). Lorsque le chiffre de rang le plus élevé qu'on laisse tomber est égal à 5, si un des chiffres qui le suivent n'est pas nul, le chiffre précédent est augment de 1 (exemple : 12,6502 s'arrondit à 12,7). Si le chiffre de rang le plus élevé que nous laissons tomber est un 5 terminal (qui n'est suivi d'aucun chiffre) ou qui n'est suivi que de zéros, nous augmentons de 1 le dernier chiffre du nombre arrondi s'il est impair, sinon nous le laissons inchangé (exemples : 12,75 s'arrondit à 12,8 et 12,65 à 12,6). Dans ce dernier cas, le dernier chiffre du nombre arrondi est toujours un chiffre pair.

Les chiffres significatifs d'une valeur comprennent tous ses chiffres déterminés avec certitude ainsi que le premier chiffre sur lequel porte l'incertitude (ce dernier significatif occupe le même rang que l'ordre de grandeur de l'incertitude).

Souvent, les sources de données ne mentionnent pas d'intervalle de confiance (c'est-à-dire une indication +/-). Par exemple, lorsque nous écrivons $m = 25{,}4$ [kg], nous considérons conventionnellement que l'incertitude est du même ordre de grandeur que le rang du dernier chiffre significatif (soit le chiffre incertain).

En fait, seul le rang décimal de l'incertitude est implicite : sa marge réelle n'est pas précisée.

ALGÈBRE

8 Calcul Algébrique 235
Équations et inéquations 236 ● Identités remarquables 241 ● Polynômes 244

9 Algèbre et géométrie ensemblistes 259
Algèbre et géométrie corporelle 259 ● Groupes de transformations 262 ● Groupes de symétries 266 ● Groupe des permutations 268

10 Analyse fonctionnelle 273
Représentations 273 ● Fonctions 282 ● Logarithmes 291 ● Produit scalaire fonctionnel 296

11 Topologie 299
Espace topologique 299

12 Théorie de la mesure 311
Espaces mesurables 311

13 Calcul différentiel et intégral 319
Calcul différentiel 319 ● Fonction de Dirac 369 ● Fonction Gamma d'Euler 370 ● Equations différentielles 374

14 Suites et séries 387
Suites 387 ● Séries 392 ● Critère de convergence 428

15 Calcul vectoriel 433
Notion de flèche 433 ● Ensemble des vecteurs 433 ● Espaces vectoriels 436 ● Espace Affine 442 ● Espaces vectoriels euclidiens 444 ● Espaces vectoriels fonctionnels 455 ● Espace vectoriels hermitiens 456 ● Systèmes de coordonnées 458 ● Opérateurs différentiels 466

8

Calcul Algébrique

Dans la section d'arithmétique de ce site, nous avons beaucoup écrit sur différentes théories utilisant les nombres abstraits afin de généraliser l'étendue de la validité de ces dernières. Nous avons cependant peu abordé la façon dont nous devions manipuler avec exactitude ces nombres abstraits. C'est ce que nous allons voir maintenant.

Comme vous le savez peut-être déjà, le nombre peut être envisagé en faisant abstraction de la nature des objets qui constituent le groupement qu'il caractérise ; et ainsi qu'à la façon de codifier (chiffre arabe, romain, ou autre système...). Nous disons alors que le nombre est une "nombre abstrait" et lorsque nous manipulons ces types de nombres nous disons que nous faisons du "calcul algébrique" ou encore du "calcul littéral".

Pour les mathématiciens il n'est pas avantageux de travailler avec des valeurs numériques $(1, 2, 3...)$ car ils représentent uniquement des cas particuliers. Ce que cherchent les physiciens théoriciens ainsi que les mathématiciens, se sont des relations applicables universellement dans un cadre le plus général possible.

Ces nombres abstraits appelés aujourd'hui communément "variables" sont très souvent représentés par :

1. L'alphabet latin $(a, b, c, d, e \ldots ; A, B, C, D, E)$
2. Par l'alphabet grec, qui est lui particulièrement utilisé pour représenter des opérateurs mathématiques plus ou moins complexes (comme le sont le $+$ pour l'addition, le $-$ pour la soustraction par exemple).
3. Par l'alphabet hébraïque (à moindre mesure)

Bien que ces symboles puissent représenter n'importe quel nombre, il en existe quelques uns qui peuvent représenter des constantes dites Universelles (vitesse de la lumière c, la constante gravitationnelle G, ...).

Remarque Il semblerait que les lettres pour représenter les nombres ont été employées pour la première fois par Viète.

Habituellement, les premières lettres de l'alphabet latin $a, b, c, \ldots$ désignent les nombres connus, et les dernières $x, y, z, \ldots$ les nombres inconnus.

Une variable est donc susceptible de prendre des valeurs numériques différentes. L'ensemble de ces valeurs peut varier suivant le caractère du problème considéré.

Rappels (nous avions déjà défini cela dans le chapitre traitant des nombres dans la section d'arithmétique) :

R1. Nous appelons "domaine de définition" d'une variable, l'ensemble des valeurs numériques qu'elle est susceptible de prendre entre deux bornes.

Soient a et b deux nombres tel que $a < b$. Alors :

R2. Nous appelons "intervalle fermé d'extrémité a et b", l'ensemble de tous les nombres x compris entre ces deux valeurs et nous le désignons de la façon suivante :

$$[a, b] = \{x \in \mathbb{R} \| a \leq x \leq b\} \tag{8.1}$$

R3. Nous appelons "intervalle ouvert d'extrémité a et b", l'ensemble de tous les nombres x compris entre ces deux valeurs non comprises et nous le désignons de la façon suivante :

$$]a, b[= \{x \in \mathbb{R} \| a < x < b\} \tag{8.2}$$

R4. Nous appelons "intervalle fermé à gauche, ouvert à droite" la relation suivante :

$$]a,b] = \{x \in \mathbb{R} \| a < x \leq b\} \tag{8.3}$$

R5. nous appelons "intervalle ouvert à gauche, fermé à droite" la relation suivante :

$$[a,b[= \{x \in \mathbb{R} \| a \leq x < b\} \tag{8.4}$$

REMARQUE Si la variable peut prendre toutes les valeurs négatives et positives possibles nous écrivons dès lors : $]-\infty,+\infty[$ où le symbole "∞" signifie "infini". Evidemment il peut y avoir des combinaisons d'intervalles ouverts et infinis à droite, fermé et limité gauche et réciproquement.

Définition 8.1 *Nous appelons "voisinage de a", tout intervalle ouvert de contenant a (c'est un concept simple que nous reprendrons pour définir ce qu'est une fonction continue).*

Ainsi :

$$\forall \varepsilon > 0,]a-\varepsilon, a+\varepsilon[\tag{8.5}$$

est un voisinage de a.

8.1 Équations et inéquations

L'algèbre élémentaire consiste à partir des définitions de l'addition, soustraction, multiplication, et puissance et de leurs propriétés (associativité, distributivité, commutativité, élément neutre, inverse...) - ce qui constitue selon l'ensemble sur lequel nous travaillons un corps ou un groupe commutatif abélien ou non (voir chapitre de théorie des ensembles) - de manipuler selon un but fixé des "équations algébriques".

Nous allons définir de suite après ce qu'est une équation et une inéquation mais nous souhaitons d'abord définir certaines de leurs propriétés :

Soit A et B deux polynômes (ou monômes) quelconques - voir définitions un peu plus loin - les expressions :

$$A > B, A = B, A < B, A \leq B, A \geq B \tag{8.6}$$

Vérifient les propriétés suivantes :

Propriété 8.1 *P1. Nous pouvons toujours ajouter ou ôter aux deux membres d'une inéquation ou équation un même polynôme en obtenant une inéquation ou équations équivalente (c'est à dire avec les mêmes solutions ou réductions).*

Propriété 8.2 *P2. Si nous multiplions ou si nous divisons les deux membres d'une inéquation par un même nombre positif nous obtenons également une inéquation ou équation équivalente (nous avons déjà vu cela).*

Propriété 8.3 *P3. Si nous multiplions ou si nous divisons les deux membres d'une inéquation par un même nombre négatif et si nous inversons le sens de l'inégalité, nous obtenons alors une inéquation ou équation équivalente.*

8.1.1 Équations

Définition 8.2 *Une "équation" est une relation d'égalité entre des valeurs toutes abstraites (autrement dit : deux expressions algébriques) ou non toutes abstraites (dès lors nous parlons d'équations à une inconnue, deux inconnues, trois inconnues, ...) reliées entre elles par des opérateurs divers.*

La maîtrise parfaite de l'algèbre élémentaire est fondamentale en physique-mathématique. Comme il existe une infinité de types d'équations, nous ne les présenterons pas ici. C'est le rôle de l'enseignant dans les classes d'entraîner le cerveau de ses élèves pendant plusieurs années (2 à 3 ans en moyenne) à résoudre énormément de configurations différentes d'équations algébriques (exposées

sous forme de problèmes de tous les jours, géométriques ou purement mathématiques) et ce afin que les élèves manipulent ces dernières sans erreurs en suivant un raisonnement logique et rigoureux.

Comme nous n'aimons pas les exemples pratiques, nous avons tenté, ci-dessous, de faire une généralisation simpliste des règles de base de l'algèbre élémentaire. Cette généralisation sera d'autant plus simple à comprendre que le lecteur aura l'habitude de manipuler des quantités abstraites.

Ainsi, soient $a, b, c, d, e, \ldots, x, y$ des nombres abstraits pouvant prendre n'importe quelle valeur numérique. Soit x représentant un ou plusieurs nombres abstraits quelconques opérants entre eux d'une façon quelconque tel que nous ayons des monômes (un seul nombre abstrait) ou polynômes (poly = plusieurs) algébriques différents distinguables ou non (nous faisons un abstrait de l'abstraction).

Propriété 8.4 *P1. Nous aurons toujours si et seulement si le terme à gauche de l'égalité représente le même terme que celui qui est à droite de l'égalité. Si cette condition est satisfaite nous avons alors :*

$$x - x = 0 \tag{8.7}$$

Propriété 8.5 *P2. Sinon :*

$$x - x = x \tag{8.8}$$

où nous excluons le cas particulier où tous les x sont nuls.

Propriété 8.6 *P3. Nous avons :*

$$\frac{x}{x} = 1 \Leftrightarrow 1 = 1 \Leftrightarrow x = x \tag{8.9}$$

qui vérifie la symbolique de l'équation $x = x$ dans le cas seulement où les éléments sont identiques entre eux[1].

Propriété 8.7 *P4. Nous avons sinon (dans le cas où tous les x sont strictement différents) :*

$$\frac{x}{x} \neq \frac{x}{x} \tag{8.10}$$

mais nous pouvons quand même avoir :

$$\frac{x}{x} = \frac{x}{x} \tag{8.11}$$

dans le cas où une simplification des termes contenus dans les x amènent à une identité de la relation binaire non nécessairement égale à l'unité.

Propriété 8.8 *P5. Si tous les x sont strictement identiques, alors :*

$$x \cdot x \cdot \ldots \cdot x = \prod_{n} x = x^n \tag{8.12}$$

Propriété 8.9 *P6. Sinon nous avons :*

$$x \cdot x \cdot \ldots \cdot x \tag{8.13}$$

qui ne peut s'écrire autrement mais il peut arriver que $x \cdot x \cdot \ldots \cdot x = x$ avec le x à droite de l'égalité identique à aucun, un ou encore plusieurs x du membre gauche de l'égalité.

Propriété 8.10 *P7. Nous pouvons avoir :*

$$\frac{x^x}{x^x} = 1 \tag{8.14}$$

sans que nécessairement les exposants du numérateur ou dénominateur soient égaux (nous excluons le dénominateur nul)

1. Nous excluons le dénominateur nul.

Propriété 8.11 *P8. Sinon nous pouvons avoir :*

$$\frac{x^x}{x^x} = x \text{ ou } \frac{x^x}{x^x} = x^{\Xi - \Xi} \tag{8.15}$$

mais il n'est cependant bien évidemment pas impossible d'avoir quand même $x = 1$ ou $x^{x-x} = 1$ (nous excluons le dénominateur nul)

Propriété 8.12 *P9. Nous avons si tous les sont strictement identiques :*

$$x/x/x \ldots x = x/x^n = x^{1-n} \tag{8.16}$$

Mais il est également possible que dans l'expression précédente certains différents s'annulent cependant entre eux dès que leur divisions mutuelle est égale à l'unité (nous excluons le dénominateur nul)

Propriété 8.13 *P10. Sinon nous avons :*

$$x/x/x \ldots x = x/x^n = x^{1-n} \tag{8.17}$$

mais il n'est cependant pas impossible d'avoir quand même $x/x/x \ldots x = x$ avec le x à droite de l'égalité identique à aucun, un ou plusieurs x du membre gauche de l'égalité ou même $x/x/x \ldots x = x/x^n$

Propriété 8.14 *P11. Soit $\perp$ représentant indifféremment soit exclusivement l'addition ou exclusivement la soustraction nous avons (au signe près) :*

$$x \perp x \perp x \perp \ldots \perp x = n \cdot x \tag{8.18}$$

si tous les x sont identiques entre eux ou si la combinaison d'un nombre indéterminés de x sont égaux au x présent à droite de l'égalité.

Propriété 8.15 *P11'. Sinon quoi nous aurons :*

$$x \perp x \perp x \perp \ldots \perp x = x \tag{8.19}$$

il peut cependant arriver que le x à droite de l'égalité soit identique à aucun, un ou plusieurs x du membre gauche de l'égalité.

Propriété 8.16 *P12. Nous pouvons également avoir :*

$$x^x \cdot x^x \cdot \ldots x^x \cdot x^x = x^{\sum x} \tag{8.20}$$

si et seulement si les x sont tous égaux et les puissances x non nécessairement égales.

À partir de ces treize règles 8.4 à 8.16, nous pouvons résoudre, simplifier ou montrer qu'une équation possède des solutions ou non par rapport à un problème ou énoncé donné.

Ainsi, soit Ψ une opérande ou une suite d'opérations quelconques sur une ou des abstractions d'abstrait x et parmi tous les x, une (ou plusieurs) dont la ou les valeurs numériques est ou sont inconnues (les autres étant connues). Alors, nous devons pouvoir trouver ou démontrer qu'une équation du type :

$$\Psi x = \Psi x \tag{8.21}$$

possède ou non des solutions.

Dans le cas d'une équation avec la valeur absolue du type :

$$\mid \Psi x \mid = \Psi x \tag{8.22}$$

avec le deuxième membre strictement positif (sinon la relation précédente serait un non sens) cela équivaut bien sûr d'après la définition de la valeur absolue à écrire :

$$\Psi x = \Psi x \quad \text{et} \quad \Psi x = -\Psi x \tag{8.23}$$

Remarques

1. La présence de la valeur absolue dans une équation algébrique dont nous cherchons les solutions double souvent le nombre de solutions.
2. Une équation est dite "équation conditionnelle", s'il y a des nombres dans l'ensemble de définition des expressions qui ne sont pas solutions (ce qui est en fait le cas le plus fréquent). Inversement, si tout nombre de l'ensemble de définition est solution de l'équation alors l'équation est dit "équation identité".

Nous pouvons parfois avoir à résoudre (et non à simplifier) un "système d'équations". Qu'est-ce que c'est ? : C'est un ensemble d'au moins 2 équations à résoudre (et non à simplifier). La particularité du système ? : L'ensemble des solutions du système est l'intersection des solutions de toutes les équations à résoudre. Quelle est leur utilité ? : Elle est sans fin, ces systèmes permettent de résoudre des problèmes faisant intervenir des applications des mathématiques à d'autres domaines. A cause de la variété illimitée des applications, il est difficile d'établir des règles précises pour trouver des solutions. La marche à suivre que voici peut-être utile pour autant bien sûr que le problème puisse être formulé sous forme d'équations :

1. Si le problème est posé par écrit, le lire plusieurs fois soigneusement, réfléchir aux faits donnés ainsi qu'à la quantité d'inconnues à trouver (résumer l'énoncé sur une feuille de papier est souvent plus qu'utile pour les gros problèmes).
2. Choisir une lettre qui représente la quantité inconnue. C'est l'un des pas décisifs dans la recherche de la solution. Des phrases contenant des mots comme : trouver, quoi, combien, où, quand ; devraient vous renseigner sur la quantité inconnue
3. Faire éventuellement un dessin (de tête ou sur papier) avec des légendes
4. Dresser une liste des faits connus et des relations concernant les quantités inconnues. Une relation peut être décrite par une équation dans laquelle apparaissent d'un seul ou des deux côtés du signe égal des énoncés écrits à la place des lettres ou des nombres.
5. Après avoir analyse la liste de l'étape 4, formuler une équation qui décrive précisément ce qui est énoncé avec des mots.
6. Résoudre l'équation formulée à l'étape 5.
7. Contrôler les solutions obtenues à l'étape 6 en se reportant à l'énoncé de départ du problème. Vérifier que la solution concorde avec les conditions de l'énoncé.

Les méthodes de résolutions des systèmes d'équations sont traités en détails dans le chapitre traitant des algorithmes dans la section d'analyse numérique du site (vous y verrez la méthode) et également dans le chapitre d'algèbre linéaire de la présente section (vous comprenez pourquoi la méthode est telle quelle).

8.1.2 Inéquations

Une "équation" est une relation d'égalité entre des valeurs toutes abstraites (autrement dit : deux expressions algébriques) ou non toutes abstraites (dès lors nous parlons d'équations à une inconnue, deux inconnues, trois inconnues, ...) reliées entre elles par des opérateurs divers.Précédemment nous avons vu qu'une équation était une égalité composée de différents calculs avec différents termes (dont au moins une "inconnue" ou "un chiffre abstrait"), et que "résoudre" une équation revenait à calculer la valeur de l'inconnue de l'égalité, alors que la "simplifier" revenait à minimiser mathématiquement le nombre de termes (en factorisant ou autre..) et que développer revenait à mettre à plat tous les termes.

Pourquoi avons-nous besoin de rappeler la définition d'une équation ? Tout simplement parce que pour l'inéquation, c'est la même système. La différence ? Si l'équation est une égalité, l'inéquation est une inégalité : comme l'équation, l'inéquation est composée de différents calculs avec différents termes reliés entre eux par des opérateurs quelconques, dont au moins une inconnue.

Différence entre égalité et inégalité :

Egalité nous savons qu'elle est symbolisée par le signe $=$

Inégalité l'inégalité est l'utilisation des symboles de relation d'ordre symbolisée par les signes d'égalités strictes et larges $<, \leq, >, \geq$.

Lorsque nous résolvons une inéquation, notre inconnue peut-avoir plusieurs valeurs qui satisfont à l'inéquation. Nous disons alors que la solution de l'inéquation est un "ensemble de valeurs". C'est la différence fondamentale entre une égalité et une inégalité !

Rappelons les signes que nous pouvons rencontrer dans une inéquation sont :

$<$ Se lit "strictement inférieur à" ou "strictement plus petit que". Dans ce cas, le plus souvent, la valeur butoir numérique n'est pas comprise dans le domaine et nous pouvons représenter alors le domaine avec un crochet ouvert à gauche]... ou à droite ...[selon que la valeur butoir est positive ou négative.

$>$ Se lit "strictement supérieur à" ou "strictement plus grand que". Dans ce cas, le plus souvent, la valeur butoir numérique n'est également pas comprise dans le domaine et nous pouvons représenter alors le domaine avec un crochet ouvert à gauche]... ou à droite ...[selon que la valeur butoir est positive ou négative.

REMARQUE Attention cependant pour les deux cas précités, il existe des cas où le domaine est imposé par l'ensemble de nombres sur lequel nous travaillons (penser par exemple à une inéquation où pour certaines valeurs les solutions appartient à l'ensemble des complexes). Dans ce cas, les valeurs butoirs à l'ensemble de nombres sur lequel nous travaillons peuvent imposer des crochets fermés.

$\leq$ Se lit "inférieur ou égal à " ou "plus petit ou égal à". Dans ce cas, la valeur butoir numérique est comprise dans le domaine et nous pouvons représenter alors le domaine avec un crochet fermé à gauche [... ou à droite ...] (mais pas nécessairement les deux !) selon que la valeur butoir est positive ou négative.

$\geq$ Se lit "supérieur ou égal à ou "plus grand ou égal à". Dans ce cas, la valeur butoir numérique est également comprise dans le domaine t nous pouvons représenter alors le domaine avec un crochet fermé à gauche [... ou à droite ...] (mais pas nécessairement les deux !) selon que la valeur butoir est positive ou négative.

REMARQUE Nous renvoyons le lecteur au début de ce chapitre où nous avions définit la manière d'écrire des domaines de définition.

L'objectif des inéquations est la plupart du temps (excepté le côté esthétique) d'avoir au moins parmi l'ensemble des termes une valeur numérique qui permet de définir le domaine de solution (de tous les termes abstraits de l'inéquation) qui satisfait l'inéquation.

Il existe plusieurs façons de représenter les domaines de définition des variables qui satisfont à l'inéquation. Nous allons voir à travers un petit exemple quelles sont ces possibilités :

Soit une inéquation linéaire (du premier degré) en x à une seule inconnue à laquelle nous imposons un contrainte particulière arbitraire pour l'exemple (évidemment l'expression peut contenir plus de termes...) :

$$x < 2 \tag{8.24}$$

nous avons dans l'inéquation ci-dessus déjà simplifié tous les termes qui étaient superflus.

Résoudre l'inégalité revient à chercher la valeur de x inférieure à 2. Bien sûr, il n'existe pas un seule solution dans $\Re$ mais un ensemble de solutions et c'est cela même le principe des inéquations.

Pour résoudre l'inéquation, nous observons d'abord le type d'inégalité imposée ("stricte" ou "égal"). Ensuite, dans les petites classes (et pas seulement parfois...) nous représentons l'ensemble $\Re$ par un tableau tel que :

$-\infty$	0	$+\infty$
...	... | ...	...

(8.25)

Nous savons intuitivement que la solution de notre inéquation regroupe toutes les valeurs inférieures à 2 (2 exclu des solutions) et ce jusqu'à $-$. Nous écrivons alors cet intervalle ou domaine sous la forme suivante :

$$S =]-\infty, 2[\tag{8.26}$$

Ensuite, nous pouvons représenter graphiquement l'ensemble des solutions (cela aide à comprendre et prépare l'étudiant à la résolution de systèmes d'équations et d'inéquations et aux variations de

fonctions). Pour cela, nous reprenons le modèle de schéma du système numérique, et y plaçons notre notre valeur butoir (nous n'en avons qu'une dans cet exemple mais parfois il peut y en avoir plusieurs dû au fait qu'il y un singularité ou des racines pour certaines valeurs du domaine de définition), soit 2 :

$-\infty$	0	2	$+\infty$
...	... | ...	... | ...	...

(8.27)

et enfin, nous délimitons au stylo de couleur (...) l'ensemble des solutions de $-\infty$ à 2 exclu :

$-\infty$	0	2	$+\infty$
...	... | ...	...[...	...

(8.28)

A la valeur 2, nous n'oublions pas de marquer le signe ...[pour montrer que cette valeur est exclue des solutions. Et voilà, le tour est joué et le concept est extrapolable a des inéquation beaucoup plus complexes.

Remarque Parfois au lieu de représenter les tableaux comme nous l'avons fait, certains professeurs (c'est un choix complètement artistique) demandent à leur élèves d'hachurer les cases du tableau et d'y dessiner de petits ronds, ou encore se servent de petites flèches, ou encore de dessiner le graphique des fonctions de l'inéquation (cette dernière méthode est certes esthétique mais prend du temps..).

Remarque Dans le cadre d'inéquations de degré supérieur à 1, il faut (voir plus loin ce que cela signifie exactement) d'abord déterminer les racines de l'inéquation qui permettent de déterminer les intervalles et ensuite par essais successifs, déterminer quels intervalles sont à rejeter ou a conserver.

Nous pouvons également (au même titre que les équations) parfois avoir à résoudre un "système d'inéquations". Qu'est-ce que c'est ? : C'est un ensemble d'au moins 2 inéquations à résoudre. La particularité du système ? : L'ensemble des solutions du système est l'intersection des solutions des toutes les inéquations à résoudre.

Autrement dit, la méthode est la même que la précédente, à la différence que notre tableau (représentant les domaines de solutions) comportera une ligne supplémentaire par inéquation supplémentaire dans le système plus une ligne de synthèse qui est la projection des domaines de solutions possibles du système.

Ainsi, un système à n inéquations aura un tableau récapitulatif à $n+1$ lignes.

Mathématiquement, les domaines (car il peuvent y en avoir plusieurs qui sont disjoints) peuvent s'écrire comme un ensemble des domaines :

$$S = \{]..,...], [...,...], [...,...[\} \tag{8.29}$$

Les systèmes d'inéquations sont très fréquents dans beaucoup de problèmes de la mathématique, physique, économétrie, etc... Il est donc important de s'entraîner à les résoudre pendant vos études avec l'aide de votre professeur. Par exemple, voici une possible représentation du domaine de solutions d'un système d'inéquations pris du chapitre traitant de l'algorithmique dans la section d'informatique théorique où nous étudions la "recherche opérationnelle".

8.2 Identités remarquables

Les identités remarquables sont des sortes formules magiques, qui nous servent le plus souvent pour la factorisation ou la résolution d'équations algébriques.

Rappelons certaines notions qui ont déjà été vues dans le chapitre de théorie des ensembles de la section d'arithmétique (nous supposons le concept d'élément neutre connu puisque déjà défini) :

Commutativité :

$$a+b=b+a \quad \text{et} \quad a\cdot b=b\cdot a \tag{8.30}$$

Associativité :

$$a+(b+c)=(a+b)+c \quad \text{et} \quad a\cdot(b\cdot c)=(a\cdot b)\cdot c \tag{8.31}$$

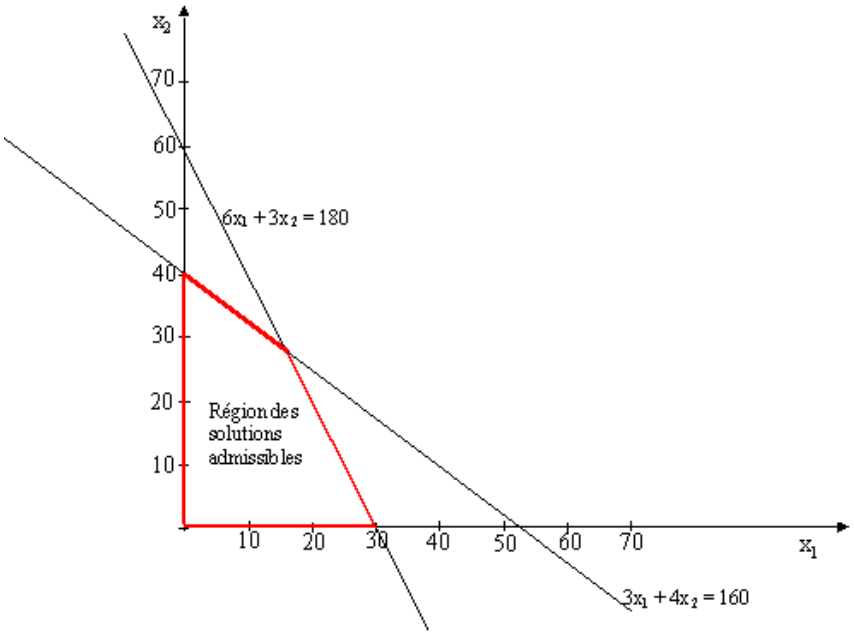


FIGURE 8.1 – Système d'inéquation

Distributivité :

$$a \cdot (b + c) = a \cdot b + a \cdot c \tag{8.32}$$

Les mêmes observations sont valables avec l'opération de soustraction bien évidemment dans les domaines de définition adéquats.

Nous pouvons vérifier avec des valeurs numériques (en remplaçant chaque nombre abstrait par un nombre choisi au hasard), ou par distribution (ce serait mieux, ainsi vous êtes sûr d'avoir compris ce dont quoi nous parlions), que les identités algébriques suivantes sont vérifiées (ce sont les plus connues) :

1. Identité du second degré :

$$(a + b)^2 = a^2 + 2ab + b^2 \tag{8.33}$$

2. Identité du troisième degré :

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3 \tag{8.34}$$

REMARQUE Nous pouvons très bien poser que $(a + (c + d))^2 = (a + b)^2$ où nous avons bien évidemment posé que $b = c + d$ (nous faisons un "abstrait d'abstraction" ou plus couramment : un "changement de variable")... :

$$(a + b + c)^2 = a^2 + 2a(b + c) + (b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2bc + 2ac \tag{8.35}$$

Nous pouvons remarquer que pour calculer le développement de $(a + b)^n$, nous utilisons le développement de $(a + b)^n - 1$, c'est-à-dire calculé avec la valeur précédente de n.

Nous remarquons les propriétés suivantes pour a et b :

P1. Les puissances de a décroissent de n à 0 (, donc il n'est pas noté dans le dernier terme)

P2. Les puissances de b croissent de 0 à n (, donc il n'est pas noté dans le dernier terme)

P3. Dans chaque terme, la somme des puissances de a et b est égal à n

P4. Les coefficients multiplicateurs devant chaque terme se calculent en faisant la somme des coefficients multiplicateurs de deux termes du développement obtenu avec la valeur précédente de b (voir la figure ci-dessous).

8.2.1 Triangle de Pascal

Les coefficients binomiaux peuvent alors êtres obtenus par construction du triangle de Pascal" :

$$\begin{array}{llcccccccccccccccc}
 & & & & & & & & & \nearrow^{p=0} & & & & & & & \\
n=0 & \rightarrow & & & & & & & \mathbf{1} & & \nearrow^{p=1} & & & & & & \\
n=1 & \rightarrow & & & & & & \mathbf{1} & & \mathbf{1} & & \nearrow^{p=2} & & & & & \\
n=2 & \rightarrow & & & & & \mathbf{1} & & \mathbf{2} & & \mathbf{1} & & \nearrow^{p=3} & & & & \\
n=3 & \rightarrow & & & & \mathbf{1} & & \mathbf{3} & & \mathbf{3} & & \mathbf{1} & & \nearrow^{p=4} & & & \\
n=4 & \rightarrow & & & \mathbf{1} & & \mathbf{4} & & \mathbf{6} & & \mathbf{4} & & \mathbf{1} & & \nearrow^{p=5} & & \\
n=5 & \rightarrow & & \mathbf{1} & & \mathbf{5} & & \mathbf{10} & & \mathbf{10} & & \mathbf{5} & & \mathbf{1} & & \nearrow^{p=6} & \\
n=6 & \rightarrow & \mathbf{1} & & \mathbf{6} & & \mathbf{15} & & \mathbf{20} & & \mathbf{15} & & \mathbf{6} & & \mathbf{1} & & \nearrow^{p=7} \\
n=7 & \rightarrow \ \mathbf{1} & & \mathbf{7} & & \mathbf{21} & & \mathbf{35} & & \mathbf{35} & & \mathbf{21} & & \mathbf{7} & & \mathbf{1} &
\end{array} \tag{8.36}$$

Dont chaque élément est donné par (cf. chapitre de Probabilités) :

$$C_p^n = \frac{n!}{p!(n-p)!} \tag{8.37}$$

avec $n, p \in \mathbb{N}$. Nous pouvons alors facilement vérifier que :

$$(a+b)^n = \sum_{p=0}^{n} C_p^n a^{n-p} b^p \tag{8.38}$$

ce qui constitue le fameux "binôme de Newton".

Preuve Cette relation se démontre simplement par récurrence en supposant la relation précédente vraie et en la calculant pour le rang 1 :

$$(a+b)^1 = a+b = C_0^1 a^1 b^0 + C_1^1 a^0 b^1 = a+b \tag{8.39}$$

Montrons que si elle est vraie pour n alors elle est vraie pour $n+1$:

$$\begin{aligned}
(a+b)^{n+1} = (a+b)(a+b)^n = (a+b)\sum_{p=0}^{n} C_p^n a^{n-p} b^p &= \sum_{p=0}^{n} C_p^n a^{n+1-p} b^p + \sum_{p=0}^{n} C_p^n a^{n-p} b^{b+1} \\
&= C_0^n a^{n+1} + \sum_{p=1}^{n} C_p^n a^{n+1-p} b^p + \sum_{p=0}^{n-1} C_p^n a^{n-p} b^{b+1} + C_n^n b^{p+1} \\
&= C_0^n a^{n+1} + \sum_{p=1}^{n} C_p^n a^{n+1-p} b^p + \sum_{p=1}^{n} C_{p-1}^n a^{n+1-p} b^b + C_n^n b^{p+1} \\
&= C_0^n a^{n+1} + \sum_{p=1}^{n} (C_p^n + C_{p-1}^n) a^{n+1-p} b^p + C_n^n b^{p+1} \\
&= C_0^n a^{n+1} + \sum_{p=1}^{n} C_p^{n+1} a^{n+1-p} b^p + C_n^n b^{p+1} = \sum_{p=0}^{n+1} C_p^{n+1} a^{n+1-p} b^p + C_n^n b^p
\end{aligned}$$

La relation est vraie au rang $n+1$, elle est donc vraie pour tout n. •

Pour ce qui est des identités remarquables avec des valeurs négatives, il est inutile d'apprendre par cœur l'emplacement du signe "$-$". Il suffit de faire un changement de variable et une fois le développement fait de refaire le changement de variable dans l'autre sens.

Exemple 8.1

$$(a-b)^2 = (a+c)^2 = a^2 + 2ac + c^2 = a^2 + 2a(-b) + (-b)^2 = a^2 - 2ab + b^2 \quad \textit{avec} \quad c = -b \tag{8.40}$$

et ainsi de suite pour toute puissance n.

Nous pouvons bien sûr mélanger les genres tels que (fameux exemple particulier) :

$$(a+b)(a-b) = a(a-b) + b(a-b) = a^2 - b^2 \tag{8.41}$$

et quelques relations remarquables pratiques supplémentaires qui sont souvent utilisées dans les petites classes pour les exercices :

$$\begin{aligned} a^3 + b^3 &= (a+b)(a^2 - ab + b^2) \\ a^3 - b^3 &= (a-b)(a^2 + ab + b^2) \\ (x+a)(x+b) &= x^2 + (a+b)x + ab \\ (x+a)(x-b) &= x^2 + (a-b)x - ab \\ (x-a)(x-b) &= x^2 - (a+b)x + ab \end{aligned} \tag{8.42}$$

et autre cas très fréquent :

$$(ax+b)(cx+d) = acx^2 + (ad+bc)x + bd \tag{8.43}$$

REMARQUE Lorsqu'à partir du terme de droite (sous forme numérique simplifiée) le professeur demande à ses élèves en tant qu'exercice d'obtenir la factorisation à gauche de l'égalité, il n'existe pas d'autres moyens que de procéder par essais successifs.

Bien sûr, il y en a encore un beaucoup plus grand nombre de relations utiles (dont une partie découlent d'une généralisation de celle présentées ci-dessus) que le lecteur découvrira par ses propres raisonnements et en fonction de sa pratique.

REMARQUE Il est bien sûr possible de multiplier des polynômes entre eux et de distribuer les termes multiplicatifs. Inversement, il est souvent demandé aux élèves des petites classes de faire la procédure inverse ("factoriser" ou "décomposer" un polynôme) afin qu'ils s'habituent à la manipulation des identités remarquables. Décomposer en un produit de facteurs est une opération importante en mathématiques, puisqu'il est ainsi possible de réduire l'étude d'expressions compliquées à l'étude de plusieurs expressions plus simples.

8.3 Polynômes

Définition 8.3 (Simpliste) *Nous appelons "polynôme algébrique $P(x)$" une fonction de degré qui s'écrit :*

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \tag{8.44}$$

ou de façon plus condensée par :

$$P_n(x) = \sum_{k=0}^{n} a_k x^k \tag{8.45}$$

REMARQUES

1. Le n en indice du $P(x)$ est parfois omis car explicitement défini dans l'énoncé.
2. Le lecteur qui aura parcouru le chapitre de théorie des ensembles, se rappellera certainement que l'ensemble des polynômes de degré n ou inférieurs forment un structure d'espace vectoriel !

Définition 8.4 (ensembliste) *Soit k un anneau (cf. chapitre de Théorie Des Ensembles) et $n \in \mathbb{N}^0$, "l'anneau des polynômes" en n indéterminées (ou variables)$k[X_1, \ldots, X_n]$ est construit à partir d'un polynôme élémentaire, appelé "monôme" de la forme :*

$$\lambda X_1^{m_1} \ldots X_n^{m_n} \tag{8.46}$$

où $\lambda \in k$ et $m_1, \ldots, m_n$ sont des entiers $\geqslant 0$.

Ainsi, un polynôme est une somme d'un nombre fini de monômes. Le cas particulier commun utilisé dans les petites classes et présenté au début est $k[X]$, c'est-à-dire l'anneau des polynômes à une variable à coefficients dans k. Tout élément de $k[X]$ s'écrit donc :

$$\sum_{i=0}^{n} a_i X^i \tag{8.47}$$

avec $a_i \in K, i = 0 \ldots n$ et $n \in \mathbb{N}$

REMARQUE Les puissances sont toujours positives (ou nulles) dans $k[X]$.

Définition 8.5 *Nous nommons "racine" ou "zéro de polynôme", la ou les valeurs x telles que "l'équation polynomiale" soit satisfaite à la condition qu'au moins un des a_n avec $n > 0$ soit non nul.*

Si le polynôme admet une ou plusieurs racines nous pouvons alors factoriser ce dernier sous la forme :

$$P(x) = (x + r_n) \cdot (x + r_{n-1} \cdot \ldots \cdot (x + r_1) \cdot (x + r_0) \tag{8.48}$$

Les identités algébriques sont des exemples de types particuliers de fonctions polynomiales. Considérons une constante c et une variable x et :

$$\begin{aligned}(c + x)^n &= \sum_{k=0}^{n} C_k^n c^k \cdot x^{n-k} \\ &= C_0^n c^0 \cdot x^n + C_1^n c^1 \cdot x^{n-1} + \ldots + C_{n-1}^n c^{n-1} \cdot x^1 + C_n^n c^n\end{aligned} \tag{8.49}$$

Nous voyons que si nous posons :

$$C_0^n = a_n, C_1^n c^1 = a_{n-1}, C_{n-1}^n c^{n-1} = a_1 \cdot x^1, C_n^n c^n = a_0, \tag{8.50}$$

nous retrouvons :

$$(c + x)^n = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x^1 + a_0 \tag{8.51}$$

Définition 8.6 *Le "coefficient dominant" d'un polynôme est le coefficient de son monôme de plus haut degré.*

8.3.1 Division euclidienne des polynômes

Plaçons nous à présent dans l'anneau $k[X]$. Si $P(X) \in k[X]$, nous notons $deg(P)$ le degré du polynôme $P(X)$ à coefficients dans un anneau k (les réels ou les complexes... peu importe !)

REMARQUE Par convention, $deg(0) = -\infty$

Soit :

$$u(X) = \sum_{i=0}^{k} u_i X^i, v(X) = \sum_{i=0}^{m} v_i X^i \in k[X] \tag{8.52}$$

avec $m > 0$. Alors il existe deux polynômes uniques $q(X), r(X) \in k[X]$ tels que :

$$deg(r) < deg(v) = m \tag{8.53}$$

et :

$$u(X) = q(X) \cdot v(X) + r(X) \tag{8.54}$$

Preuve Si $u(X) = 0$ le résultat est évident. Supposons que $u(X) \neq 0$ et montrons l'existence par récurrence sur le degré k de $u(X)$.

1. Si $k = 0$ alors $q(X) = 0$ (puisque $m > 0$) et donc $r(X) = u(X)$ fait l'affaire.
2. Supposons l'affirmation vraie pour tout $k \leq n$:
3. Soit $u(X)$ de degré $k = n + 1$. Si $m > n + 1$ alors $q(X) = 0$ et $r(X) = u(X)$ font l'affaire. Sinon, si $m \leq n + 1$ alors en écrivant :

$$u(X) - \frac{u_{n+1}}{v_m} X^{n+1-m} \cdot v(X) \tag{8.55}$$

nous réduisons $u(X)$ à polynôme de degré $\leq n$ puisque $v(X)$ est de degré m. Effectivement, le terme :

$$\frac{1}{v_m} X^{n+1-m} \cdot v(X) \tag{8.56}$$

élimine (au moins) le terme de plus grand degré $u_{n+1}X^{n+1}$. Par hypothèse de récurrence, il existe $f(X), g(X)$ tels que :

$$u(X) - \frac{u_{n+1}}{v_m} X^{n+1-m} \cdot v(X) = f(X) \cdot v(X) + g(X) \tag{8.57}$$

avec $deg(g) < m$. Donc :

$$u(X) = \left(\frac{u_{n+1}}{v_m} X^{n+1-m} + f(X) \right) \cdot v(X) + g(X) \tag{8.58}$$

et :

$$q(X) = \frac{u_{n+1}}{v_m} X^{n+1-m} + f(X), r(X) = g(X) \tag{8.59}$$

font l'affaire.

Par conséquent, la division euclidienne existe dans l'anneau des polynômes $k[X]$. •

REMARQUE Cette démonstration nous a permis dans le chapitre de théorie des ensembles de montrer que cet anneau est "principal".

8.3.2 Théorème de factorisation des polynômes

Nous allons maintenant démontrer une propriété importante qui est au fait à l'origine illustré (entre autres) par les identités remarquables que nous avons vues plus haut :

Si une fonction polynôme $P(X) \in k[X]$ à coefficients dans k de degré $n \geq 1$ a une racine dans $x = r$ l'anneau k, alors nous pouvons factoriser $P(x)$ par $(x - r)$ tel que :

$$P(x) = (x - r)Q(x) \tag{8.60}$$

où Q est une fonction polynôme de degré $n - 1$.

Preuve L'idée consiste à effectuer la division euclidienne de P par $(x-r)$. D'après le théorème, il existerait un couple (Q, R) de polynôme tels que :

$$P(x) = (x - r)Q + R \tag{8.61}$$

et selon le résultat obtenu du théorème précédent sur la division euclidienne :

$$deg(R) < deg(x - r) \tag{8.62}$$

Or, $deg(x - r) = 1$, donc $deg(R) = 0$ (ou $-\infty$). R est donc une fonction polynôme constante. Par ailleurs, r est une racine de P. Nous avons donc :

$$0 = P(r) = (r - r)Q(r) + R(r) \tag{8.63}$$

Donc $R(r) = 0$. Donc R est la fonction polynôme nulle et le théorème est pratiquement démontré. Il reste encore à prouver que $deg(Q) = n - 1$, ce qui est une conséquence immédiate de la relation :

$$deg(P) = deg(x - r) + deg(Q) \tag{8.64}$$

D'où :

$$n = 1 + deg(Q) \tag{8.65}$$

•

De cette propriété de factoriser un polynôme, appelée "théorème de factorisation", nous pouvons donner un avant goût d'un théorème plus important :

Montrons que si nous avons une fonction polynôme $P(X) \in [X]$ de degré $n \in \mathbb{N}$ à coefficients dans k, alors elle possède au plus n racines (certaines étant éventuellement confondues) dans k.

Preuve Puisque P a un degré, P n'est pas la fonction polynôme nulle. Raisonnons par l'absurde. Si la fonction P possède p racine avec $p > n$, en notant $r_1, \ldots, r_p$ ces racines, nous avons, d'après le théorème de factorisation précédant (appliqué p fois) :

$$P(x) = (x - r_1)(x - r_2) \ldots (x - r_p)Q(x) \tag{8.66}$$

où Q est donc une fonction polynôme de degré $n - p < 0$. Or, comme par définition un polynôme en est un si seulement son degré appartient à $\mathbb{N}$, le polynôme Q doit donc être le polynôme nul tel que $Q(x) = 0$. Il s'ensuite que $P(x) = 0$, ce qui contredit l'hypothèse initiale comme quoi P n'est la fonction polynôme nulle d'où $p \leq n$. •

8.3.3 Équations diophantiennes

Si nous généralisons le concept de polynôme avec plusieurs variables tel que :

$$P = P(x, y, z) \tag{8.67}$$

nous appelons alors "équation diophantienne" une équation de la forme :

$$P(x, y, z) = 0 \tag{8.68}$$

où P est un polynôme à coefficients entiers (ou rationnels) dont nous cherchons les radicaux strictement dans $\mathbb{N}$ ou $\mathbb{Q}$. Des exemples classiques d'équations diophantiennes sont :

- Les triplets pythagoriciens (ou triades) tel que :

$$a^2 + b^2 = c^2 \tag{8.69}$$

- Le grand théorème de Fermat dont la conjecture dit que si n est supérieur à 2, il n'existe pas d'entiers x, y, z non nuls pour lesquels :

$$x^n + y^n = z^n \tag{8.70}$$

Pour la démonstration il faudra attendre un peu que les auteurs du site aient le temps de la comprendre également (...).

8.3.4 Polynômes du premier degré

Soit :

$$P(x) = ax + b = a\left(x + \frac{b}{a}\right) \tag{8.71}$$

Si $a \neq 0$ alors le polynôme admet une unique racine :

$$r = -\frac{b}{a} \tag{8.72}$$

tel que $P(r) = 0$.

Remarques

1. Il faut toujours prendre l'habitude de vérifier l'existence de la solution dans l'équation d'origine pour s'assurer de la validation du domaine de définition de la solution. Effectivement, il existe des solutions aux développement de résolution d'une équation qui ne vérifient pas l'équation d'origine et c'est ce que nous nommons des "solutions étrangères" ou encore "racines étrangères".
2. Si les coefficients du polynômes de degré 1 sont tous réels alors la racine est réelle.
3. Si un des coefficients est complexe alors la racine est nécessairement complexe.
4. Si les deux coefficients sont complexes, alors la racine est soit complexe soit réelle.

8.3.5 Polynômes du second degré

Soit le polynôme à coefficients réels :

$$P(x) = ax^2 + bx + c = a\left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a^2}\right] \tag{8.73}$$

Si alors une des racines satisfera :

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2} \Rightarrow x + \frac{b}{2a} = \pm\frac{\sqrt{b^2 - 4ac}}{2a} \Rightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{8.74}$$

Il peut donc y avoir deux racines que l'on note :

$$r_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{8.75}$$

tel que $P(r) = 0$et où nous définissons un nouveau terme appelé "déterminant du polynôme" ou "discriminant" qui allège souvent les écritures :

$$\Delta = b^2 - 4ac \tag{8.76}$$

REMARQUE Il faut aussi toujours prendre l'habitude de vérifier l'existence de la solution dans l'équation d'origine pour s'assurer de la validation du domaine de définition de la solution au cas où la solution serait "étrangère".

Si le polynôme du deuxième degré en x comporte deux racines, nous pouvons alors factoriser de manière irréductible (selon le théorème fondamental de factorisation des polynômes vus plus haut) de la manière suivante :

$$P(x) = a(x - r_1)(x - r_2) \tag{8.77}$$

Nous démontrons, à partir de l'expression des racines, sans trop de peine les relations dites "relations de Viète" :

$$r_1 + r_2 = -\frac{b}{a} \quad \text{et} \quad r_1 \cdot r_2 = \frac{c}{a} \tag{8.78}$$

1. si $\Delta < 0$ le polynôme n'admet pas de zéros réels et ne se décompose pas en un produit de facteurs réels du premier degré mais de facteurs complexes. Ainsi (il est nécessaire d'avoir lu la partie traitant des nombres complexes dans le chapitre des nombres de la section d'arithmétique du site) :

$$r_{1,2} = \frac{-b \pm \sqrt{(-1)(b^2 - 4ac)}}{2a} = \frac{-b \pm \sqrt{i^2(b^2 - 4ac)}}{2a} = \frac{-b \pm i\sqrt{(b^2 - 4ac)}}{2a} \tag{8.79}$$

   et nous savons que nous pouvons écrire tout nombre complexe sous une forme condensée (formule d'Euler) et comme les racines complexes d'un polynôme du second degré sont conjugées (nous connaissons ce terme) nous avons :

$$r_{1,2} = re^{\pm i(\varphi + 2kz)} \tag{8.80}$$

   où (rappel) r est le module des racines complexes (module égal pour les deux racines) et φ l'argument des racines complexes (égales en valeur absolue).
2. Si $\Delta = 0$ alors le polynôme possède une seule solution qui est bien évidemment :

$$r_{1,2} = \frac{-b}{2a} \tag{8.81}$$

3. Si $\Delta > 0$alors le polynôme possède deux solution définies par les relations générales que nous avons déjà données précédemment.

Évidemment de ce qui a été vu jusqu'à maintenant on en tire que si un polynôme admet une ou plusieurs racines alors ce même polynôme est divisible par $(x - r_n)$.

REMARQUE Il existe un polynôme de degré deux dont la solution est fameuse de par le monde. Ce nombre est appelé la "divine proportion" et se retrouve en architecture, esthétique ou encore en phyllotaxie (c'est-à-dire dans la disposition des feuilles autour de la tige des plantes). Ce nombre vaut :

$$1 + \frac{\sqrt{5}}{2} \cong 1{,}06103398 \tag{8.82}$$

et appartient à l'ensemble des nombres irrationnels car il ne peut pas s'écrire sous forme de fraction entière, mais c'est un nombre algébrique puisqu'il est la solution positive de l'équation :

$$x^2 - x - 1 = 0 \tag{8.83}$$

8.3.6 Polynômes du troisième degré

Bien que rare à résoudre en physique théorique ou lors de ses études, la résolution d'un polynôme du 3ème degré est assez récréative et montre un bon exemple d'un raisonnement mathématique déjà mature (nous devons ces développements à Scipione del Ferro et Jérome Cardan mathématiciens du 16ème siècle...).

Soit l'équation :

$$ax^3 + bx^2 + cx + d = 0 \tag{8.84}$$

avec les coefficients tous dans $\mathbb{R}$. Dans un premier temps, le lecteur pourra voir que les raisonnements que nous avons appliqués pour les polynômes de degrés inférieurs coincent rapidement.

Nous allons contourner le problème par des changements de variables subtils mais tout à fait justifiés.

Ainsi, rien ne nous empêche de poser que :

$$\exists X \Rightarrow x = X - \frac{b}{3a} \tag{8.85}$$

et que en divisant le polynôme de degré 3 par a d'écrire :

$$x^3 + \frac{b}{a}x^2 + \frac{c}{a}x + \frac{d}{a} = 0$$

$$\left(X - \frac{b}{3a}\right)^3 + \frac{b}{a}\left(X - \frac{b}{3a}\right)^2 + \frac{c}{a}\left(X - \frac{b}{3a}\right) + \frac{d}{a} = 0 \tag{8.87}$$

$$\left(X^3 - X^2\frac{b}{a} + X\frac{b^2}{3a^2} - \frac{b^3}{27a^3}\right) + \frac{b}{a}\left(X^2 - 2X\frac{b}{3a} + \frac{b^2}{9a^2}\right) + \frac{c}{a}\left(X - \frac{b}{3a}\right) + \frac{d}{a} = 0$$

En regroupant les termes de même ordre :

$$X^3 + X^2\left(-\frac{b}{a} + \frac{b}{a}\right) + X\left(\frac{b^2}{3a^2} - 2\frac{b^2}{3a^2} + \frac{c}{a}\right) + \left(-\frac{b^3}{27a^3} + \frac{b^3}{9a^3} - \frac{cb}{3a^2} + \frac{d}{a}\right) = 0$$

$$X^3 + X\left(\frac{c}{a} - \frac{b^2}{3a^2}\right) + \left(\frac{2b^3}{27a^3} - \frac{cb}{3a^2} + \frac{d}{a}\right) = 0 \tag{8.90}$$

et posons (rien, mais alors absolument rien ne nous l'interdit) :

$$\begin{aligned} x &= X - \frac{b}{3a} \\ p &= \frac{c}{a} - \frac{b^2}{3a^2} \\ q &= \frac{2b^3}{27a^3} - \frac{cb}{3a^2} + \frac{d}{a} \end{aligned} \tag{8.91}$$

où (1) est connu si et seulement si X est connu et où p, q sont de tout façon connus.

Le polynôme :

$$P = X^3 + pX + q \tag{8.92}$$

étant de degré impair, il admet comme permet de le constater tout tracé visuel d'un tel polynôme à coefficient réels aux moins un racine réelle, appelée "racine certaine" (vérifiez vous verrez bien par une représentation graphique d'un polynôme de degré impair que cela est trivial).

Maintenant, nous faisons un autre changement de variable (nous en avons tout à fait le droit) subtil : $X = u + v$en imposant la condition que u,v doivent êtres tels que $3uv = -p$ (rien ne nous empêche d'imposer une telle contrainte) et nous alors :

$$\begin{aligned} &P = X^3 + pX + q = (u+v)^3 - 3uv(u+v) + q = 0 \\ &\Rightarrow u^3 + v^3 = -q \end{aligned} \tag{8.93}$$

Dès lors nous avons :

$$\begin{aligned} u^3 + v^3 &= -q \\ u^3v^3 &= \frac{-p^3}{27} \Leftrightarrow 3uv = -p \end{aligned} \tag{8.94}$$

Nous pouvons très bien faire une analogie entre les deux relations (1') et (2') et les relations de Viète que nous avions obtenues pour le polynôme de degré 2 qui rappelons le étaient :

$$r_1 + r_2 = -\frac{b}{a} \quad \text{et} \quad r_1 \cdot r_2 = \frac{c}{a} \tag{8.95}$$

à la différence que nous avons maintenant (nous adoptons une autre notation pour ces racines intermédiaires) $z_1 = u^3$ et $z_2 = v^3$ ce qui nous donne pour le polynôme P en imposant (toujours par analogie) $a = 1$ une nouvelle équation :

$$Z^2 + qZ - \frac{p^3}{27} = 0 \tag{8.96}$$

dont z_1, z_2 sont les racines. Cette dernière équation à pour discriminant :

$$\Delta = q^2 + 4\frac{p^3}{27} \tag{8.97}$$

Prenons maintenant le cas par cas :

Si $\Delta > 0$ l'équation en Z admet deux solutions z_1, z_2 dont la somme va nous donner indirectement la valeur de X puisque par définition $X = u + v$ et $z_1 = u^3$ et $z2 = v^3$. Nous voyons que nous avons tous les ingrédients pour trouver la première racine de l'équation initiale qui sera la racine certaine (ou "zéro certain"). Ainsi :

$$X_1 = u + v = \sqrt[3]{z_1} + \sqrt[3]{z_2} = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \tag{8.98}$$

comme $\Delta > 0$ et que les racine supérieurs sont cubiques nous avons nécessairement $X_1 \in \mathbb{R}$ si tous les coefficients de l'équation originale sont bien dans $\mathbb{R}$.

Si $\Delta = 0$ nous le savons, l'équation en Z admet une racine double et puisque le discrimant comporte une puissance carrée de q cela signifie nécessairement que p est négatif.

Le polynôme P admet donc lui aussi une racine double et de même pour l'équation d'origine. Nous avons vu par ailleurs que pour un polynôme du second degré si le discriminant est nul les racines sont :

$$r_{1,2} = -\frac{b}{2a} \tag{8.99}$$

alors par analogie :

$$X_1 = X_2 = u + v = \sqrt[3]{z_1} + \sqrt[3]{z_2} = 2\sqrt[3]{z_{1,2}} = 2 \cdot \sqrt[3]{\frac{-q}{2}} \tag{8.100}$$

Si $\Delta < 0$ nous devons à nouveau utiliser les nombres complexes comme nous l'avons fait lors de notre étude du polynôme de degré 3. Ainsi, nous savons que l'équation en Z admet deux solutions complexes telles que :

$$z_{1,2} = -\frac{q}{2} \pm i\sqrt{-\frac{q^2}{4} - \frac{p^3}{27}} \tag{8.101}$$

et à nouveau comme les racines sont conjuguées nous pouvons écrire sous la forme condensé :

$$z_{1,2} = re^{\pm i(\varphi + 2kz)} \tag{8.102}$$

et comme :

$$X = u + v = \sqrt[3]{z_1} + \sqrt[3]{z_2} \tag{8.103}$$

nous avons donc :

$$\begin{aligned} X_k &= u_k + v_k \\ &= u_k + \overline{u}_k \\ &= \sqrt[3]{re^{i(\varphi+2kz)}} + \sqrt[3]{re^{-i(\varphi+2kz)}} \\ &= \sqrt[3]{r}e^{\frac{i(\varphi+2kz)}{3}} + \sqrt[3]{r}e^{-\frac{i(\varphi+2kz)}{3}} \end{aligned} \tag{8.104}$$

Comme u_k, v_k sont conjugués, nous avons nécessairement $X_k \in \mathbb{R}$.

8.3.7 Polynômes de degré 4

L'équation polynômiale à résoudre ici est :

$$ax^4 + bx^3 + cx^2 + dx + e = 0 \tag{8.105}$$

avec $a \neq 0$.

REMARQUE Nous devons cette méthode de résolution à l'italien Lodovico Ferrari mathématicien italien du 16ème siècle également.

Quitte à diviser par a nous avons :

$$x^4 + b'x^3 + c'x^2 + d'x + e' = 0 \tag{8.106}$$

Puis, en posant :

$$x = y - \frac{b'}{4} \tag{8.107}$$

l'équation se réduit à :

$$\begin{aligned}
&\left(y' - \frac{b'}{4}\right)^4 + b'\left(y - \frac{b'}{4}\right)^3 + c'\left(y' - \frac{b'}{4}\right)^2 + d'\left(y' - \frac{b'}{4}\right) + e' \\
&= \left(y^4 - 4y^3\frac{b'}{4} + 6y^2\frac{b'^2}{4^2} - 4y\frac{b'^3}{4^3} + \frac{b'^4}{4^4}\right) + b'\left(y^3 - 3y^2\frac{b'}{4} + 3y\frac{b'^2}{4^2} - \frac{b'^3}{4^3}\right) \\
&+ c'\left(y^2 - 2y^{\frac{b'}{4}} + \frac{b'^2}{4^2}\right) + d'\left(y - 3\frac{b'}{4}\right) + e' \\
&= y^4 + (b' - b')y^3 + \left(-\frac{3}{4}b^2 + c + \frac{3}{8}b^2\right)y^2 + \left(\frac{3}{16}b^3 - \frac{1}{2}cb - \frac{1}{16}b^3 + d\right)y \\
&\qquad + \left(\frac{b^4}{256} - \frac{1}{4}db - \frac{1}{64}b^4 + \frac{1}{16}cb^2 + e\right)
\end{aligned} \tag{8.108}$$

où nous voyons que le coefficient devant y^3 s'annule. Ainsi, tout polynôme du type :

$$x^4 + bx^3 + cx^2 + dx + e = 0 \tag{8.109}$$

peut être écrit sous la forme suivante :

$$y^4 + c'y^2 + d'y + e' = 0 \tag{8.110}$$

En posant :

$$\begin{aligned}
x &= y - \frac{b}{4} \\
c' &= \left(-\frac{3}{4}b^2 + c + \frac{3}{8}b^2\right) \\
d' &= \left(\frac{3}{16}b^3 - \frac{1}{2}cb - \frac{1}{16}b^3 + d\right) \\
e' &= \left(\frac{1}{256}b^4 - \frac{1}{4}db - \frac{1}{64}b^4 + \frac{1}{16}cb^2 + e\right)
\end{aligned} \tag{8.111}$$

REMARQUE Si $d' = 0$, l'équation à résoudre est en réalité une "équation bicarrée". Le changement de variable $X = y^2$ permet alors de se ramener à une équation polynomiale du deuxième degré (ce que nous savons facilement résoudre).

Nous introduisons maintenant un paramètre t (que nous choisirons judicieusement par la suite) et nous réécrivons l'équation polynomiale sous la forme suivante :

$$(x^2 + t)^2 - \left[(2t - c)x^2 - dx + (t^2 - e)\right] = 0 \tag{8.112}$$

REMARQUE Si le lecteur développe et distribue tous les termes de la relation précédente il retombera bien évidemment sur $x^4 + cx^2 + dx + e = 0$.

L'idée sous-jacente est d'essayer de faire en sorte que la partie entre crochets de l'expression précédente puisse s'écrire comme un carré tel que :

$$(2t - c)x^2 - dx + (t^2 - e) = (x - \lambda)^2 \tag{8.113}$$

Car dans ce cas, en utilisant :

$$a^2 - b^2 = (a + b)(a - b) \tag{8.114}$$

Notre équation polynomiale peut s'écrire :

$$(x^2 + t)^2 - (x - \lambda)^2 = \Big((x^2 + t) + (x - \lambda)\Big) \cdot \Big((x^2 + t) - (x - \lambda)\Big) = 0 \tag{8.115}$$

et nous n'aurions plus qu'à résoudre deux équations polynomiales du deuxième degré (ce que nous savons déjà faire).

Or, pour que nous puissions écrire :

$$(2t - c)x^2 - dx + (t^2 - e) = (x - \lambda)^2 \tag{8.116}$$

Il faudrait que l'expression du deuxième degré à gauche de l'égalité n'ait qu'une seule racine. Or, nous avons vu dans notre étude des équations polynomiales du deuxième degré que cela signifiait dès lors que le déterminant est nul :

$$\Delta = b^2 - 4ac = 0 \tag{8.117}$$

et que la racine s'exprimait par :

$$x = -\frac{b}{2a} \tag{8.118}$$

Ce qui correspond dans notre cas à :

$$\Delta = d^2 - 4(2t - c)(t^2 - e) = 0 \tag{8.119}$$

et donc que :

$$4(2t - c)(t^2 - e) = d^2 \tag{8.120}$$

avec :

$$\lambda = -\frac{-d}{2(2t - c)} = \frac{d}{4c - 2c} \tag{8.121}$$

Donc finalement, si t est tel que $4(2t - c)(t^2 - e) = d^2$, alors nous avons :

$$(2t - c)x^2 - dx + (t^2 - e) = (2t - c)(x - \lambda) = (2t - c)^2\left(x - \frac{}{4t - 2c}\right)^2 \tag{8.122}$$

puisque le théorème fondamental des polynômes nous donne pour un polynôme du deuxième degré n'ayant qu'une seule racine :

$$P(x) = ax^2 + bx + c = a(x - r_1)(x - r_2) = a(x - r_1)^2 \tag{8.123}$$

Pour conclure, il suffit de voir que trouver un nombre t vérifiant la relation :

$$4(2t - c)(t^2 - e) = d^2 \tag{8.124}$$

est un problème de degré 3 que nous savons déjà résoudre par la méthode de Cardan.

De telles méthodes générales n'existent plus pour les degrés égaux ou supérieurs à 5 comme nous le verrons à l'aide de la théorie de Galois (cf. chapitre d'Algèbre Ensembliste).

8.3.8 Polynômes trigonométriques

Définition 8.7 *Nous appelons "polynôme trigonométrique" de degré N toute somme finie :*

$$S_N(\theta) = \sum_{n=-N}^{N} c_n e^{in\theta} \tag{8.125}$$

où $c_n \in \mathbb{C}$.

Un polynôme trigonométrique peut aussi être écrit en utilisant les fonctions trigonométriques usuelles grâce aux transformations suivantes :

$$S_N(\theta) = c_0 + \sum_{n=1}^{N} c_n e^{in\theta} + \sum_{n=1}^{N} c_{-n} e^{-in\theta} \tag{8.126}$$

Soit en utilisant le formule d'Euler (cf. chapitre sur les Nombres) :

$$S_N(\theta) = c_0 + \sum_{n=1}^{N} c_n(\cos(n\theta) + i\sin(n\theta)) + \sum_{n=1}^{N} c_{-n}(\cos(n\theta) - i\sin(n\theta)) \tag{8.127}$$

Ce que nous pouvons récrire aussi sous la forme :

$$S_N(\theta) = c_0 + \sum_{n=1}^{N} (c_n + c_{-n})\cos(n\theta) + \sum_{n=1}^{N} i(c_n - c_{-n})\sin(n\theta) \tag{8.128}$$

En posant alors :

$$a_0 = 2c_0, a_n = c_n + c_{-n}, b_n = i(c_n - c_{-n}) \tag{8.129}$$

Il vient :

$$S_N(\theta) = \frac{a_0}{2} + \sum_{n=1}^{N} a_n \cos(n\theta) + \sum_{n=1}^{N} b_n \sin(n\theta) \tag{8.130}$$

Nous verrons longuement dans le chapitre des suites et séries comment utiliser ces polynômes dans le cadre de l'étude des séries de Fourier.

8.3.9 Polynômes cyclotomiques

Si n est un entier naturel, nous appelons "polynôme cyclotomique" ce que nous notons traditionnellement Φ_n et définissons comme étant le produit de tous les monômes en $x - \alpha$ où α est un racine primitive n^{e} de l'unité de $\mathbb{C}$.

Ainsi, l'ensemble des racines n^{e}s de l'unité est l'ensemble :

$$G_n = \{x \in \mathbb{C} \; tel \; que \; x^n = 1\} \tag{8.131}$$

qui est un groupe cyclique (voir la théorie des ensembles dans la section d'arithmétique du site).

Nous appelons alors "racine primitive n^{e} de l'unité" ou "R.P.N." tout élément de ce groupe l'engendrant.

De plus les éléments de G_n sont du type $e^{\frac{2ikz}{n}}$ avec $k \in \mathbb{Z}$. Nous pouvons même écrire :

$$G_n = \left\{ e^{\frac{2ikz}{n}} \text{ avec } k \in \{0 \dots n-1\} \right\} \tag{8.132}$$

Un petit exemple de polynôme cyclotomique :

$$\Phi_4(x) = (x-i)(x+i)(x-1)(x+1) = x^4 - 1 \tag{8.133}$$

Les racines quatrième de l'unité sont $1, -1, i$ et $-i$ (autrement dit chacun de ces nombres mis à la puissance 4 donne 1). Elle forment le groupe G_4 et celui-ci peut-être engendré que par i et $-i$.

Donc un polynôme cyclotomique est le produit de facteurs qui s'écrit :

$$\Phi_n(x) = \prod_x \left(x - e^{\frac{2ikz}{n}} \right) \tag{8.134}$$

avec $k \in \{0 \dots n-1\}$ et k étant premier par rapport à n.

Les polynômes ont un grand nombre de propriétés que nous n'aborderons pas ici puisque ce site ne se veut pas être un ouvrage de mathématiques supérieure.

8.3.10 Polynômes de Legendre

Définition 8.8 *les polynômes de Legendre sont définis par (lire de préférence les chapitres de calcul différentiel et intégral ainsi que d'analyse fonctionnelle avant) :*

$$P_n(x) = \frac{(-1)^n}{n! \cdot 2^n} \cdot \frac{d^n}{dx^n}(1 - x2)^n \tag{8.135}$$

où P_n est donc un polynôme de degré n.

Nous retrouverons ces polynômes dans la résolution d'équations différentielles en physique (propagation de la chaleur, physique quantique, chimie quantique, etc.).

Démontrons que selon la définition du produit scalaire fonctionnel (cf. chapitre d'Analyse Fonctionnelle et de Calcul Vectoriel) que les polynômes de Legendre sont orthogonaux.

Preuve Soit P un polynôme de degré $\leq n-1$. Il suffit de montrer que $\langle P_n|P\rangle = 0$, c'est-à-dire que $P_n$est orthogonal à l'espace des polynômes de degré inférieur à n. Nous avons en effet :

$$\langle P_n|P\rangle = \int_{-1}^{1} P_n(x)P(x)dx = \frac{(-1)^n}{n! \cdot 2^n} \cdot \int_{-1}^{1} \frac{d^n}{dx^n}(1 - x^2)^n P(x)dx \tag{8.136}$$

en intégrant par parties nous obtenons :

$$\begin{aligned}\langle P_n|P\rangle &= \underbrace{\frac{(-1)^n}{n! \cdot 2^n} \cdot \frac{d^{n-1}}{dx^{n-1}}(1 - x^2)^n \cdot P(x)}_{=0} - \frac{(-1)^n}{n! \cdot 2^n} \cdot \int_{-1}^{1} \frac{d^{n-1}}{dx^{n-1}}(1 - x^2)^n \frac{d}{dx}P(x)dx \\ &= -\frac{(-1)^n}{n! \cdot 2^n} \cdot \int_{-1}^{1} \frac{d^{n-1}}{dx^{n-1}}(1 - x^2)^n \frac{d}{dx}P(x)dx\end{aligned} \tag{8.137}$$

Attention pour le terme nul ci-dessus, seulement le terme $(1 - x2)^n$ y est dérivé . Donc puisque x est au carré, quelque soit la dérivée la valeur sera toujours la même. Ce qui justifie que le terme soit nul.

En continuant de la sorte nous obtenons après n intégrations par parties :

$$\langle P_n|P\rangle = (-1)^n \frac{(-1)^n}{n! \cdot 2^n} \cdot \int_{-1}^{1} (1 - x^2)^n \cdot \underbrace{\frac{d^n}{dx^n}P(x)dx}_{=0} = 0 \tag{8.138}$$

•

REMARQUE Le terme dérivé est nul puisque le polynômes dérivé est de degré $n - 1$

Voici quelques propriétés utiles en chimie quantique des P_n :

P1. $P_n(1) = 1$

Preuve

$$P_n(x) = \frac{(-1)^n}{n! \cdot 2^n} \cdot \frac{d^n}{dx^n}(1 - x^2)^n = \frac{(-1)^n}{n! \cdot 2^n} \cdot \frac{d^n}{dx^n}\big((1 + x)^n \cdot (1 - x)^n\big) \tag{8.139}$$

et par la formule de Leibniz (cf. chapitre de Calcul Différentiel Et Intégral) nous avons :

$$\frac{(-1)^n}{n! \cdot 2^n} \cdot \frac{d^n}{dx^n}\big((1+x)^n \cdot (1-x)^n\big) = \frac{(-1)^n}{n! \cdot 2^n} \sum_{j=0}^{n} \binom{n}{j} \cdot \frac{n!}{(n-j)!}(1+x)^{n-j} \cdot \frac{(-1)^{n-j} \cdot n!}{j!}(1-x)^j \tag{8.140}$$

d'où :

$$P_n(1) = \frac{(-1)^n}{n! \cdot 2^n} 2^n \cdot (-1)^n \cdot n! = 1 \tag{8.141}$$

•

P2. $P(_n(x) = P_n(-x)$ si n est pair :

Preuve Si n est pair, $\frac{d^n}{dx^n}(1 - x2)^n$ est une fonction paire et donc :

$$P_n(x) = \frac{(-1)^n}{n! \cdot 2^n} \frac{d^n}{dx^n}(1 - x^2)^n \tag{8.142}$$

est paire. •

P3.$P(_n(-x) = -P_n(x)$ si n est impair.

Preuve Si n est impair, $\frac{d^n}{dx^n}(1-x2)^n$ est impaire et donc :

$$P_n(x) = \frac{(-1)^n}{n! \cdot 2^n} \frac{d^n}{dx^n}(1-x^2)^n \tag{8.143}$$

est impaire. •

Nous allons à présent démontrer la validité de la relation de récurrence suivante pour les (relations que nous utiliserons en physique) :

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \tag{8.144}$$

pour $n \geq 1$.

Preuve xP_n est un polynôme de degré $n+1$, il existe dès lors des $a_j \in \mathbb{R}$ tel que ce polynôme peut s'exprimer comme combinaisons linéaire de la famille de polynômes constituant la base orthonormale (base qui permet donc d'engendrer xP_n) :

$$xP_n(x) = \sum_{j=0}^{n+1} a_j P_j \tag{8.145}$$

nous pouvons dès lors écrire $k \leq n+1$:

$$\big\langle xP_n | P_k \big\rangle = \left\langle \sum_{j=0}^{n+1} a_j P_j \middle| P_k \right\rangle = \sum_{j=0}^{n+1} a_j \big\langle P_j | P_k \big\rangle = a_k \|P_k\|^2 \tag{8.146}$$

mais nous choisissons $k \leq n-2$ (parce que xP_k est dès lors de degré $n-1$:

$$\big\langle xP_n | P_k \big\rangle = \int_{-1}^{1} xP_n(x)P_k(x)dx = \big\langle P_n | xP_k \big\rangle \tag{8.147}$$

Donc $a_k \|P_k\|^2 = 0$ c'est-à-dire que nous devons avoir $a_k = 0$. Par suite :

$$xP_n = a_{n+1}P_{n+1} + a_n P_n + a_{n-1}P_{n-1} \tag{8.148}$$

Par les propriétés des polynômes de Legendre vues précédemment, nous pouvons écrire les égalités :

$$xP_n(x), x = 1 \;\; 1 = a_{n+1} + a_n + a_{n-1} \tag{8.149}$$

et :

$$xP_n(x), x = -1 \;\; 1 = a_{n+1} - a_n + a_{n-1} \tag{8.150}$$

d'où :

$$a_n = 0 \quad \text{et} \quad 1 = a_{n+1} + a_{n-1} \tag{8.151}$$

Le coefficient dominant de P_n est défini (rappelons-le) par le coefficient du monôme du plus grand degré. Ainsi, il est donné par :

$$\begin{aligned} &P_n(x) = \frac{(-1)^n}{n!2^n} \frac{d^n}{dx^n}(1-x^2)^n \Rightarrow P_n(x) = \frac{(-1)^n}{n!2^n} \frac{d^n}{dx^n} x^{2n} \\ &\Rightarrow P_n(x) = \frac{(-1)^n}{n!2^n} \frac{d^{n-1}}{dx^{n-1}}(2nx^{2n-1}) = \frac{(-1)^n}{n!2^n} \frac{d^{n-n}}{dx^{n-n}}\big(2n(2n-1)\dots x^{2n-n}\big) \\ &\quad = \frac{(-1)^n}{n!2^n}\big(2n(2n-1)\dots(2n-(n-1))x^n\big) = \frac{(-1)^n}{n!2^n}\big(2n(2n-1)\dots(n+1)x^n\big) = \frac{(2n)!}{(n!)^2 2^n}x^n \end{aligned} \tag{8.152}$$

Donc :

$$dom\big(P_n(x)\big) = \frac{(2n)!}{(n!)^2 2^n}x^n \tag{8.153}$$

Remarque Le lecteur vérifiera au besoin pour un n donné que :

$$2n(2n-1)\dots(n+1) = \frac{(2n)!}{n!} \tag{8.154}$$

La relation :

$$xP_n = a_{n+1}P_{n+1} + a_nP_n + a_{n-1}P_{n-1} \tag{8.155}$$

que nous avons obtenu ci-dessus nous impose que le coefficient dominant du polynôme de la combinaison linéaire soit égal au coefficient dominant du polynôme xP_n(nous avons éliminé le $(-1)^n$ qui se simplifie) :

$$(-1)^n \frac{(2n)!}{(n!)^2 \cdot 2^n} = a_{a+1}(-1)^{n+2} \frac{(2n+2)!}{(n+1)!^2 \cdot 2^{n+1}} \tag{8.156}$$

après simplification, nous obtenons :

$$a_{n+1} = \frac{(2n)!(n+1)!^2 2^{n+1}}{(n!)^2 2^n (2n+2)!} = \frac{(n+1)(n+1)^2}{(2n+1)(2n+2)} = \frac{n+1}{2n+1} \tag{8.157}$$

et ce qui donne finalement facilement :

$$a_{n-1} = \frac{n}{2n+1} \tag{8.158}$$

La relation :

$$xP_n = a_{n+1}P_{n+1} + a_nP_n + a_{n-1}P_{n-1} \tag{8.159}$$

devient dès lors :

$$((2n+1)xP_n = (n+1)P_{n+1} + nP_{n-1} \tag{8.160}$$

•

Voici les six premiers polynômes de Legendre :

$$\begin{aligned}
P_0 &= 1 \\
P_1 &= x \\
P_2 &= \frac{3}{2}x^2 \\
P_3 &= \frac{5}{2}x^3 - \frac{3}{2}x \\
P_4 &= \frac{35}{8}x^4 - \frac{15}{4}x^2 + \frac{3}{8}x \\
P_5 &= \frac{63}{8}x^5 - \frac{35}{4}x^3 + \frac{15}{8}
\end{aligned} \tag{8.161}$$

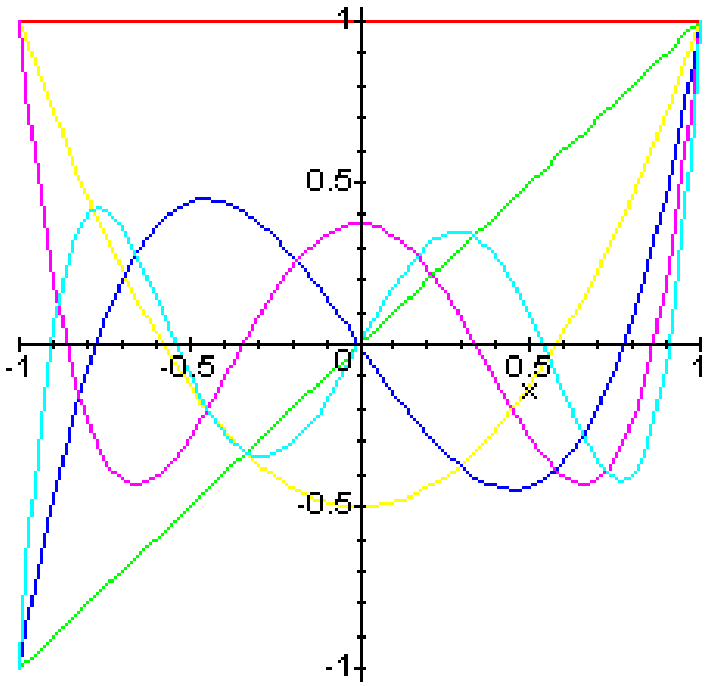


FIGURE 8.2 – Polynômes de Legendre

9

Algèbre et géométrie ensemblistes

Nous allons aborder sur ce site l'étude des structures ensemblistes de manière très pragmatique (puisque rappelons que ce site est dédié aux ingénieurs). Ainsi, il sera fait usage du minimum de formalisme et seulement les démonstrations des éléments que nous considérons comme absolument essentiels à l'ingénieur seront présentées. Par ailleurs, de nombreuses démonstrations seront faites par l'exemple et nous nous focaliserons en grande partie sur la théorie algébrique des groupes car elle a une place prédominant en physique plus que pour les autres structures ensembliste.

9.1 Algèbre et géométrie corporelle

Les symétries des figures géométriques, des cristaux et de tous les autres objets de la physique macroscopique font l'objet depuis des siècles d'observations et d'études. En termes modernes, les symétries d'un objet donné forment un groupe.

Depuis le milieu du 19ème siècle, la théorie des groupes a pris une extension énorme, et ses applications à la mécanique quantique et à la théorie des particules élémentaires se sont développées tout au long du 20ème siècle.

Dans une lettre de 1877 au mathématicien Adolph Mayer, Sophus Lie écrit qu'il a crée la théorie des groupes en janvier 1873. Il s'agit bien sûr des groupes qu'il appelait "groupes continus" et qui sont appelés aujourd'hui "groupes de Lie". Lie cherchait à étendre l'usage des groupes du domaine des équations algébriques, où Galois les avait introduits, à celui des équations différentielles.

Dès 1871, la notion de générateur infinitésimal d'un groupe à un paramètre de transformations était apparue dans son œuvre. C'est l'ensemble des générateurs infinitésimaux des sous-groupes à un paramètre d'un groupe continu qui forme ce que nous appelons aujourd'hui une algèbre de Lie.

Ce furent Wigner et Weyl qui montrèrent le rôle prééminent de la théorie des groupes, et de leurs représentations en particulier, dans la nouvelle mécanique quantique que développaient Heisenberg et Dirac. L'idée générale de la théorie des représentations est d'essayer d'étudier un groupe en le faisant agir sur un espace vectoriel de manière linéaire : nous essayons ainsi de voir le groupe comme un groupe de matrices (d'où le terme "représentation"). Nous pouvons ainsi, à partir des propriétés relativement bien connues du groupe des automorphismes de l'espace vectoriel (cf. chapitre de Théorie des Ensembles), arriver à déduire quelques propriétés du groupe qui nous intéressait.

Nous pouvons considérer la théorie des représentations de groupes comme une vaste généralisation de l'analyse de Fourier. Son développement est continu et elle a, depuis le milieu du vingtième siècle, des application innombrables en géométrie différentielle, en théorie ergodique, en théorie des probabilités, en théorie des nombres, dans la théorie des formes automorphes, dans celle des systèmes dynamiques ainsi qu'en physique, chimie, biologie moléculaire et traitement du signal. À l'heure actuelle, des branches entières des mathématiques et de la physique en dépendent.

Avant de commencer, nous renvoyons le lecteur au chapitre traitant de la théorie des ensembles pour qu'il se rappelle de la structure et des propriétés fondamentales qui constituent le groupe et également au chapitre d'algèbre linéaire (car nous en utiliseront quelques résultats).

9.1.1 Groupes cycliques

Le groupe cyclique (dont la définition a déjà été vue dans le chapitre de Théorie des Ensembles) va nous servir de base dans le cadre de l'étude des groupes finis. Par ailleurs, plutôt que de faire des développements généralisés nous avons préféré prendre des exemples particuliers afin de présenter l'idée de groupe cyclique (approche plus adaptée à l'ingénieur).

Nous allons donc prendre l'exemple fort sympathique des heures de la montre... avec trois approches différentes qui successivement (!) permettront d'abord un groupe cyclique simple.

- Première approche :

Imaginons donc une horloge avec une aiguille qui peut prendre 12 positions possibles (mais pas de positions intermédiaires). Nous noterons de manière spéciale les 12 positions possibles : $\overline{0}, \overline{2}, \overline{3} \ldots \overline{11}$(le trait au-dessus des nombres n'est pas innocent !).

Rien ne nous empêche sur l'ensemble de ces positions de définir une addition, par exemple :

$$\overline{2} + \overline{3} = \overline{5},\ \overline{3} + \overline{2} = \overline{5},\ \overline{5} + \overline{7} = \overline{0},\ \overline{9} + \overline{5} = \overline{2},\ \overline{1} - \overline{8} = \overline{5}, \tag{9.1}$$

ce qui est similaire aux résultat que nous obtenons lorsque dans notre quotidien nous faisons des calculs avec notre montre.

- Deuxième approche (première extension)

Si nous observons bien notre montre, nous remarquons qu'à chaque fois que nous rajoutons 12 (ou retirons...) à une valeur des heures de notre montre alors nous tombons sur un ensemble de nombres bien déterminé qui sont aussi dans $\mathbb{Z}$. Ainsi (évidemment dans le cadre d'une montre seules les premières valeurs positives nous intéressent la plupart du temps mais ici nous faisons des maths alors nous généralisons un peu...) :

$$\begin{aligned} \overline{0} &:= 0 + 12\mathbb{Z} &&= \{\ldots, -12, 0, 12, 24, \ldots\} \\ \overline{0} &:= 0 + 12\mathbb{Z} &&= \{\ldots, -12, 0, 12, 24, \ldots\} \\ \overline{1} &:= 1 + 12\mathbb{Z} &&= \{\ldots, -11, 1, 13, 25, \ldots\} \\ \overline{2} &:= 2 + 12\mathbb{Z} &&= \{\ldots, -10, 2, 14, 26, \ldots\} \end{aligned} \tag{9.2}$$

Nous retrouvons ici un concept que nous avions déjà vu dans le chapitre de théorie des nombres. Il s'agit de classes de congruences et l'ensemble des ces classes forment l'ensemble quotient $12\mathbb{Z}$. Si nous munissons cet ensemble quotient d'une loi d'addition, il est normalement facile d'observer que celle-ci est un interne à l'ensemble quotient, qu'elle est associative, qu'il existe un élément neutre et chaque élément possède un symétrique (inverse).

Ainsi, cet ensemble quotient muni de uniquement de la loi d'addition (sinon en ajoutant la multiplication nous pouvons former un anneau) est un groupe commutatif.

- Troisième approche (deuxième et dernière extension) :

Voyons une troisième et dernière approche qui explique pourquoi le groupe quotient est cyclique.

Si nous projetons la rotation des aiguilles de notre montre dans $\mathbb{C}$ et que nous définissons :

$$x := e^{\frac{2m}{12}} \tag{9.3}$$

Nous avons alors $x^{12} = x^0 = 1$ et :

$$\forall a, k \in \mathbb{Z}\ x\hat{a} = x^{a+12k} \tag{9.4}$$

ce qui explique pourquoi le groupe quotient $(\mathbb{Z}/12\mathbb{Z}, +)$ est appelé "groupe cyclique (par isomorphisme de groupe selon ce qui a été vu en théorie des ensembles). Son isomporphe est noté .

Si nous représentons dans $\mathbb{C}$ l'ensemble isomorphe $C_1 2$ nous obtenons alors sur le cercle unité un polygone ayant n sommets comme le montre la figure 9.1

Par ailleurs, le nombre d'éléments composants $\mathbb{Z}/12\mathbb{Z}$ étant fini, $(\mathbb{Z}/12\mathbb{Z}, +)$est fini. Contrairement au groupe $(\mathbb{Z}, +)$ qui est lui un groupe discret infini.

Ce concept de finitude sera peut-être plus évident avec l'exemple que nous ferons de suite après avec $(\mathbb{Z}/4\mathbb{Z})$ où le lecteur observera que cet ensemble a le même nombre d'éléments que C_4.

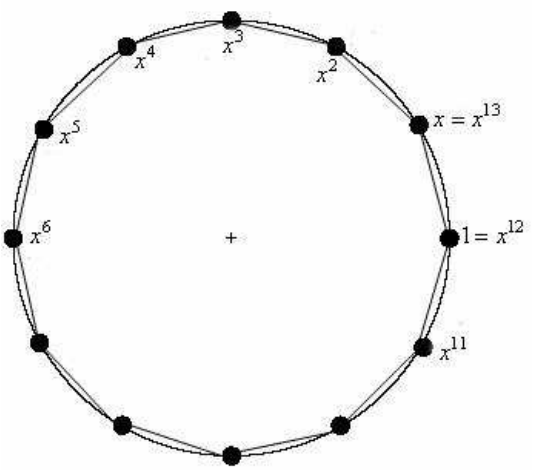


FIGURE 9.1 – groupe cyclique

REMARQUE Les mathématiciens appellent le "groupe des racines n-ièmes de l'unité". Une racine n-ième de l'unité (parfois appelée "nombre de de Moivre") est donc un nombre complexe dont la puissance n-ième vaut 1. Par ailleurs, pour un entier n donné, toutes les racines n-ième de l'unité sont situées sur le cercle unité et sont les sommets d'un polygone régulier à n côtés ayant un sommet d'affixe 1.

Ce qui intéresse les physiciens particulièrement dans un premier temps ce sont les représentations des groupes finis (aussi les groupes continus que vous verrons plus loin). Ainsi, la représentative de $\mathbb{Z}/n\mathbb{Z}$ nous est connue puisque la rotation dans le plan complexe est donnée comme nous l'a montrée notre étude des complexes dans le chapitre sur les Nombres :

$$\begin{bmatrix} \cos(\frac{2k\pi}{p}) & -\sin(\frac{2k\pi}{n}) \\ \sin(\frac{2k\pi}{n}) & \cos(\frac{2k\pi}{n}) \end{bmatrix} \tag{9.5}$$

avec $k \in [0, n-1]$. Cette représentatives est un sous-groupe du groupe des rotations $O(2)$ sur lesquelles nous reviendrons plus loin. Le groupe des rotations du plan étant lui-même un sous-groupe du groupe linéaire $GL(2)$ (nous en donnerons une définition précise et un exemple plus loin).

Au fait, les mathématiciens sont capables de démontrer que tous les groupes quotients sont cycliques à isomorphisme près avec . Les mathématiciens disent aussi que est un quotient fini du groupe monogène .

Cette approche est par contre peut-être un peu abstraite. Alors, si le lecteur se rappelle du chapitre de théorie des ensembles nous avons vu une définition bien précise de que qu'était la cyclicité d'un groupe : Un groupe G est dit cyclique si G est engendré par la puissance d'au moins un de ses éléments $\alpha \in G$ appelé générateur tel que $G = \{a^n, n \in \mathbb{Z}\}$.

Vérifions que ce soit bien le cas pour le groupe $G = (\mathbb{Z}/4\mathbb{Z})$ qui constitue un cas scolaire.

Nous noterons les éléments qui constituent ce groupe $\{0, 1, 2, 3\}$. Ceci étant fait, il convient de faire attention que dans la définition ensembliste du groupe cyclique nous parlons de "puissance" si la loi interne du groupe est la multiplication mais si la loi interne est l'addition, nous avons alors : $G = \{n \cdot a, n \in \mathbb{Z}\}$.

Le premier élément générateur du groupe est l'élément 1. Effectivement :

$$\begin{aligned} 1 &= 1 \cdot 1 = 1 \\ 1+1 &= 2 \cdot 1 = 2 \\ 1+1+1 &= 3 \cdot 1 = 3 \\ 1+1+1+1 &= 4 \cdot 1 = 1 \end{aligned} \tag{9.6}$$

Le deuxième élément générateur du même groupe est 3 :

$$\begin{aligned} 3 &= 1 \cdot 3 = 3 \\ 3+3 &= 2 \cdot 3 = 2 \\ 3+3+3 &= 3 \cdot 3 = 1 \\ 3+3+331 &= 4 \cdot 3 = 0 \end{aligned} \tag{9.7}$$

Par contre, le lecteur pourra vérifier que 2 n'est pas générateur de ce groupe !

Au fait, en ce qui concerne les groupes $G = (\mathbb{Z}/n\mathbb{Z}, +)$ les mathématiciens arrivent à démontrer que seules les éléments du groupe qui sont premiers avec n sont générateurs (c'est-à-dire les éléments de le plus grand commun diviseur est 1).

Voilà pour notre introduction aux groupes cycliques. Passons maintenant à une autre catégorie des groupes.

9.2 Groupes de transformations

/

LE GROUPE des rotations est celui qui intéresse le plus les physiciens surtout dans les domaines des matériaux, de la chimie, de la physique quantique et de l'art... Les mathématiciens apprécient eux l'étude des groupes de rotation dans le cadre de la géométrie bien évidemment (mais pas seulement) et les informaticiens tout autant les groupes linéaires. Nous en avons d'ailleurs vu un exemple de groupe de rotation juste précédemment.

9.2.1 Groupe linéaire

Définition 9.1 *Nous appelons "groupe linéaire d'ordre n" et nous notons $GL(n)$ les matrices inversibles (donc le déterminant est non nul selon ce que nous avons vu dans le chapitre d'algèbre linéaire) dont les coefficients sont dans un corps quelconque :*

$$GL(n) = \{A \in M_n | det(A) \neq 0\} \tag{9.8}$$

Un exemple simple et important de groupe linéaire est celui du sous-"groupe des transformations affines" du plan qui sont traditionnellement notées (c'est intuitif) :

$$F : M(x,y) \to M'(x',y') : \begin{cases} x' = ax + by + \alpha \\ y' = cx + dy + \beta \end{cases} \tag{9.9}$$

avec $a, b, c, d, \alpha, \beta \in \mathbb{R}, ad - bc \neq 0$ (nous verrons le pourquoi du comment de l'inégalité un peu plus loin).

Prenons un exemple pratique :

$$F : M(x,y) \to M'(x',y') : \begin{cases} x' = 0.5x - 0.4y + 0.5 \\ y' = 0.8x + 0.2y - 1.5 \end{cases} \tag{9.10}$$

ce qui appliqué à un cercle donnerait la figure 9.2 :

Cette transformation est une manière de définir les ellipses comme images d'un cercle par une transformation affine.

Les coefficients α, β sont sans importance pour la forme de l'image. En fait, ils induisent bien évidemment des translations sur les figures. Nous pouvons donc nous en passer si nous cherchons seulement à la déformer.

Ainsi, il nous reste :

$$F : M(x,y) \to M'(x',y') : \begin{cases} x' = ax + by \\ y' = cx + dy \end{cases} \tag{9.11}$$

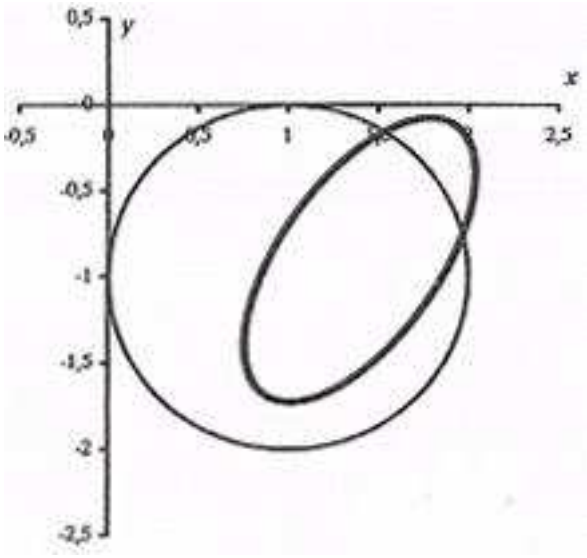


FIGURE 9.2 – groupe cyclique

ce qui peut s'écrire sous forme matricielle :

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \tag{9.12}$$

La transformation se réduit donc à la matrice :

$$F = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{9.13}$$

et comme nous l'avons vu en algèbre linéaire, la multiplication matricielle est associative n'est pas commutative, donc la transformation linéaire ne l'est pas non plus.

L'élément neutre est la matrice :

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{9.14}$$

et l'inverse de F est :

$$F^{-1} = \begin{bmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{9.15}$$

et comme nous avons imposé $ad - bc \neq 0$ tout élément y possède donc un inverse. Ainsi, le groupe linéaire affine est non commutatif et... forme bien un groupe...

9.2.2 Groupe spécial linéaire

Définition : Nous appelons "groupe spécial linéaire d'ordre n" et nous notons $SL(n)$ les matrices inversibles dont les coefficients sont dans un corps quelconque et dont le déterminant est égal à l'unité :

$$SL(n) = \{A \in M_n | det(A) = 1\} \tag{9.16}$$

Il s'agit évidemment d'un sous-groupe de $GL(n)$.

En reprenant l'exemple précédant et en se rappelant que le déterminant d'une matrice carré bidimensionnelle est (cf. chapitre d'Algèbre Linéaire) :

$$det(F) = ad - bc = 1 = det(F')\} \tag{9.17}$$

nous remarquons bien géométrique ce que signifie d'avoir un déterminant unitaire dans ce cas ! Effectivement nous avons vu dans le chapitre d'Algèbre Linéaire lors de notre interprétation géométrique qu'avoir un déterminant équivaut à une surface. Ainsi, le fait d'avoir $ad - bc$ unitaire permet montre donc que quelque soit l'ordre de la transformation, nous avons l'aide qui vaut toujours 1. Ainsi, le groupe spécial linéaire conserve les surfaces.

9.2.3 Groupe orthogonal

Définition : Nous appelons "groupe orthogonal réel d'ordre n" et notons $O(n)$ les matrices orthogonales (cf. chapitre d'Algèbre Linéaire) données par :

$$O(n) = \{A \in M_n(\mathbb{R}) | A_T A = I_n\} \tag{9.18}$$

Par ailleurs, nous avons démontré dans le chapitre d'Algèbre Linéaire lors de notre étude des matrices de rotation que $A_T A = I_n$ implique $det(A) = \pm 1$.

C'est le cas par exemple de la matrice de $O(2)$ vue précédemment (elle appartient au groupe orthogonal mais aussi au groupe des rotations que nous verrons plus loin) :

$$\begin{bmatrix} \cos(\frac{2k\pi}{n}) & -\sin(\frac{2k\pi}{n}) \\ \sin(\frac{2k\pi}{n}) & \cos(\frac{2k\pi}{n}) \end{bmatrix} \tag{9.19}$$

qui est orthogonale comme il est facile de le vérifier.

REMARQUE $O(1)$ est constitué de l'ensemble des matrices $[1], [-1]$

9.2.4 Groupe spécial orthogonal

Définition : Si $A \in O(n)$ et que nous avons $det(A) = 1$ alors nous obtenons un sous-groupe de $O(n)$ appelé "groupe spécial orthogonal réel d'ordre n" et noté $SO(n)$:

$$O(n) = \{A \in M_n(\mathbb{R}) | A_T A = I_n, det(A) = 1\} \tag{9.20}$$

La matrice de rotation donnée précédemment fait partie de ce groupe puisque son déterminant est égal à l'unité ! Par ailleurs, ce groupe occupe une place très spéciale en physique et nous le retrouverons de maintes fois.

9.2.5 Groupe cercle

Le sous-groupe $SO(2)$, appelé aussi parfois "groupe cercle" et noté S^1, que nous avions aussi étudié dans le chapitre de Géométrie Euclidienne :

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{9.21}$$

occupe une place à part dans la famille des groupes $SO(n)$ avec n supérieur à l'unité. Effectivement il est le seul à être commutatif. Par ailleurs, il est isomorphe à $e^{i\theta}$.

Le sous-groupe $SO(3$ donné par la matrice (cf. chapitre de Géométrie Euclidienne) :

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{9.22}$$

pour la rotation autour de l'axe X n'est pas commutatif. Par ailleurs les quaternions, dont la représentative est donc $SO(3)$, forment une groupe non commutatif aussi (par rapport à la loi de multiplication) comme nous l'avons vu dans le chapitre sur les Nombres.

Par rapport à un vecteur unitaire on se rend facilement compte visuellement parlant que $SO(3)$ est un sous-groupe fermé de $GL(3)$, c'est-à-dire de l'ensemble des groupes linéaires de dimension 3.

Remarque $SO(1)$ est constitué de la matrice [1].

9.2.6 Groupe unitaire

Définition : Nous appelons "groupe unitaire d'ordre n" et nous notons $U(n)$ les matrices dont les coefficients sont complexes (dans le cadre de ce site le plus souvent) ou réels et qui sont orthogonales :

$$U(n) = \{A \in M_n(\mathbb{C}) | A_\dagger A = I_n\} \tag{9.23}$$

Remarquons par ailleurs que toute matrice unitaire à coefficients complexes et à une dimension... (de $U(n)$ donc...) est un nombre complexe de module unitaire, qui peut toujours s'écrire sous la forme $e^{i\mathbb{R}}$.

Nous en avons déjà vu un exemple aussi sur le site lors de notre étude des spineurs dans le chapitre de Calcul Spinoriel. Il s'agit des matrices de Pauli (utilisées dans le chapitre de Physique Quantique Relativiste) données par :

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_y = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix} \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{9.24}$$

9.2.7 Groupe spécial unitaire

Définition : Nous appelons "groupe spécial unitaire d'ordre n" et nous notons $SU(n)$ les matrices dont les coefficients sont complexes et qui sont orthogonales et dont le déterminant est unitaire :

$$SU(n) = \{A \in M_n(\mathbb{C}) | A_\dagger A = I_n, det(A) = 1\} \tag{9.25}$$

Remarque $U(1)$ est égal à $SU(1)$ et il s'agit donc du cercle unité complexe égal à $e^{i\mathbb{R}}$. Par ailleurs, $SO(2)$ est commutatif et isomorphe à $U(1)$ car c'est l'ensemble des rotations du plan.

Un exemple connu est toujours celui des matrices de Pauli mais simplement écrites sous la forme utilisée en Physique Quantique Relativiste (voir chapitre du même nom) :

$$i\sigma_x = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, i\sigma_y = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, i\sigma_z = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \tag{9.26}$$

qui font partie de $SU(2)$ et qui comme nous l'avons montré (implicitement) au début du chapitre de Calcul Spinoriel est isomorphe au groupe des quaternions $SO(3)$ de module 1 sur la sphère de dimension 3 (notée S^3).

Remarque Le groupe spécial unitaire possède une importance particulière en physique des particules. Si le groupe unitaire $U(1)$ est le groupe de jauge de l'électromagnétisme (pensez au nombre complexe apparaissant dans les solutions de l'équation d'onde !), $SU(2)$ est le groupe associé à l'interaction faible, et $SU(3)$ celui de l'interaction forte. C'est par exemple grâce à la structure des représentations de $SU(3)$ que Gellman a conjecturé l'existence des quarks.

9.2.8 Groupe unitaire

Définition : Nous appelons "groupe spécial linéaire d'ordre n" et nous notons $SL(n)$ les matrices dont les coefficients sont dans un corps quelconque et dont le déterminant est égal à l'unité :

$$SL(n) = \{A \in M_n | det(A) = 1\} \tag{9.27}$$

9.3 Groupes de symétries

Le groupe de symétrie d'un objet noté X (image, signal, etc., e.g. en 1D, 2D ou 3D) est le groupe de toutes les isométries sous lesquelles il est invariant avec la composition en tant qu'opération.

Tout groupe de symétrie dont les éléments ont un point fixe commun, ce qui est vrai pour tous les groupes de symétrie de figures limitées, peut être représenté comme un sous-groupe du groupe orthogonal $O(n)$ en choisissant l'origine pour point fixe. Le groupe de symétrie propre est alors un sous-groupe du groupe orthogonal spécial $SO(n)$, et par conséquent, il est aussi appelé le groupe de rotation de la figure.

Dans ce qui suit, nous allons interpréter la composée de deux opérations de symétrie comme une multiplication au même titre que pour les permutations.

Définition : Le groupe des symétries de X est l'ensemble des symétries de X, muni de la structure de multiplication donnée par composition.

Exemples :

E1. Le cœur ♡ à un groupe de symétries a deux éléments (d'ordre deux), à savoir l'application identité id, et l'application r_v : réflexion dans l'axe vertical. Nous observons que le symétrique via la relation $r_v \circ r_v = id$.

E2. La lettre Φ à un groupe de symétries à quatre éléments, c'est donc un groupe de symétrie d'ordre 4, à savoir l'application identité id, les deux réflexions r_h et r_v et la rotation par l'angle π que nous noterons t_π.

Dans ce groupe nous avons par exemple $r_v \circ r_v = t_\pi$(et c'est commutatif), $r_v \circ r_v$ est la rotation par un angle 2π , ce qui est la même application que l'application identique, donc $t_\pi \circ t_\pi = id$.

Ainsi, le groupe de symétrie de cette lettre est commutatif et la loi de composition est bien interne. C'est donc bien un groupe.

E3. Le pentagone régulier à un groupe de symétries à 10 éléments, c'est donc un groupe de symétrie d'ordre 10, à savoir les 5 rotations $id, t_{2\pi/5}, t_{4\pi/5}, t_{6\pi/5}, t_{8\pi/5}$ ainsi que les 5 réflexions dans les 5 axes de symétrie.

Les règles de multiplication sont un peu compliquées dans cet exemple mais nous pouvons néanmoins observer que la composition est toujours une opération interne (le produit de deux réflexions est toujours une rotation par exemple).

9.3.1 Groupe diédral

Plus généralement, le groupe de symétries d'un n-gone régulier (si n est impaire) a exactement $2n$ éléments. Ce groupe s'appelle le "groupe diédral d'ordre n" est est noté D_{2n}.

E4. Nous reviendrons sur cet exemple lorsque nous introduirons un peu plus loin le concept de groupe distingué lors de notre étude des groupes de permutations et la définition des groupes distingués.

Le groupe diédral D_6 d'ordre 3 des isométries d'un triangle équilatéral à 6 éléments que nous noterons (afin que l'écriture soit moins lourd) :

$$D_6 = \{id, t_{2\pi/3}, t_{4\pi/3}, \sigma_1, \sigma_2, \sigma_3\} \tag{9.28}$$

où $\sigma_1, \sigma_2, \sigma_3$ sont les symétries par rapports au trois bissectrices (respectivement médiatrices).

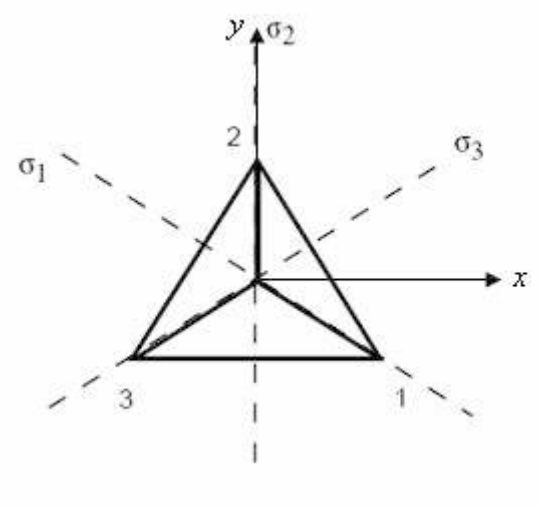


FIGURE 9.3 – Tétraèdre NH_3

La table de composition de ce groupe diédral montre aussi que ce groupe est non-commutatif :

$\nearrow D_6$	id	$t_{2\pi/3}$	$t_{4\pi/3}$	σ_1	σ_2	σ_3
id	id	$t_{2\pi/3}$	$t_{4\pi/3}$	σ_1	σ_2	σ_3
$t_{2\pi/3}$	$t_{2\pi/3}$	$t_{4\pi/3}$	id	σ_3	σ_1	σ_2
$t_{\pi/3}$	$t_{4\pi/3}$	id	$t_{2\pi/3}$	σ_2	σ_3	σ_1
σ_1	σ_1	σ_2	σ_3	id	$t_{2\pi/3}$	$t_{4\pi/3}$
σ_2	σ_2	σ_3	σ_1	$t_{4\pi/3}$	id	$t_{2\pi/3}$
σ_3	σ_3	σ_1	σ_2	$t_{2\pi/3}$	$t_{4\pi/3}$	id

(9.29)

E5. Regardons un dernier exemple appliqué à la chimie en énumérant les opérations de symétrie qui laissent la molécule NH_3 (tétraèdre).

Le groupe de transformation contient 6 éléments : l'identité id, qui est la rotation de $2\pi/3$, C_3^2 la rotation de $4\pi/3$ (que nous noterons par la suite C_{-3}) toutes deux selon l'axe Z (perpendiculaire au plan XY donc...) et 3 axes $\sigma_1, \sigma_2, \sigma_3$ de symétrie/réflexion passant chacun par le milieu d'un des arêtes de base au milieu de l'arête opposée comme le montre la figure 9.3 (pyramide vue du dessus) :

La combinaison des différents éléments de symétrie montre que la table de composition est (ce qui prouve que la loi est interne et que nous travaillons donc bien dans un groupe) :

$\nearrow$	id	σ_1	σ_2	σ_3	C_3	C_{-3}
id	id	σ_1	σ_2	σ_3	C_3	C_{-3}
σ_1	σ_1	id	C_{-3}	C_3	σ_3	σ_2
σ_2	σ_2	C_3	id	C_{-3}	σ_1	σ_3
σ_3	σ_3	C_{-3}	C_3	id	σ_2	σ_1
C_3	C_3	σ_2	σ_3	σ_1	C_{-3}	id
C_{-3}	C_{-3}	σ_3	σ_1	σ_2	id	C_3

(9.30)

Attention à l'ordre, nous appliquons d'abord l'élément de ligne puis l'élément de colonne. Nous constatons que le groupe n'est donc pas commutatif.

9.3.2 Orbite et stabilisateur

Nous allons voir maintenant deux définitions que nous retrouverons en cristallographie (leur nom n'est pas innocent !).

Définition : L'orbite d'un élément x de E est donnée par :

$$Orb_x = \{g \cdot x, g \in G\} \tag{9.31}$$

L'orbite de x est l'ensemble des positions (dans E) susceptibles d'être occupées par l'image de x sous l'action de G. Les orbites forment évidemment une partition de E.

Exemple :

Considérons un ensemble E sur lequel agit un groupe G, par :

$$E = \{S_0, S_1, S_2, S_3, S_4, S_5\} \tag{9.32}$$

l'ensemble des 6 sommets d'un hexagone sur lequel nous faisons agir le groupe $G = \{id, t_{2\pi/3}, t_{4\pi/3}\}$. Nous observons déjà trivialement que G est bien un groupe !

Maintenant, prenons un élément de E, par exemple S_0.

Son orbite va donc être par définition :

$$Orb_{S_0} = \{id(S_0) = S_0, t_{2\pi/3}(S_0) = S_0, t_{4\pi/3}(S_0) = S_4\} \tag{9.33}$$

Définition 9.2 *Le stabilisateur x d'un élément de E est l'ensemble :*

$$Stab_x = \{g \in G \| g \cdot x = x\} \tag{9.34}$$

des éléments qui laissent x invariant sous leur action. C'est un sous-groupe de G.

Pour reprendre notre exemple précédant. Son stabilisateur va être réduit à :

$$Stab_{S_0} = \{id\} \tag{9.35}$$

9.4 Groupe des permutations

LES GROUPES symétriques ont une importance non négligeable dans certains domaines de la physique quantique mais aussi en mathématiques dans le cadre de la théorie de Galois. Il convient donc d'y porter aussi une attention toute particulière.

Rappelons d'abord (cf. chapitre de Probabilités) que dans un ensemble $\{1, \ldots, n\}$ il y a $n!$ permutations possibles. Les mathématiciens disent, à juste titre, qu'il y a $n!$ bijections et appellent ce nombre "ordre du groupe de permutations".

Prenons par exemple l'ensemble $\{1, 2, 3\}$. Cet ensemble à 3! permutations possibles qui sont :

$$\{(1), (12), (13), (23), (123), (132)\} \tag{9.36}$$

Ce qui se lit dans l'ordre : application identitité id, 1 amène sur 2 ou 2 sur un 1, 1 amène sur 3 ou 3 sur 1, 2 amène sur 3 ou 3 sur 2, 1 amène sur 2 qui amène sur 3 qui amène sur 1, 1 amène sur 3 qui amène sur 2 qui amène sur 1.

Nous pouvons observer facilement que la composition de deux permutations n'est pas commutative :

$$\begin{aligned}(1\ 2) \circ (1\ 3) &= \{1, 2, 3\} \\ (1\ 3) \circ (1\ 2) &= \{3, 1, 2\}\end{aligned} \tag{9.37}$$

et que la composition de deux permutations est un loi interne :

$$(1\ 2) \circ (1\ 3) \circ (1\ 2\ 3) = \{3, 2, 1\} = (1\ 3) \tag{9.38}$$

avec un élément neutre qui est bien l'identité id. Nous avons donc bien un groupe non commutatif. Rappelons également au lecteur que certains éléments du groupe, s'ils sont bien choisis, peuvent former un sous-groupe. C'est l'exemple de :

$$\{(1),(12)\} \tag{9.39}$$

qui est un sous-groupe de S_3 (il est facile de vérifier qu'il possède toutes les propriétés d'un groupe).

Définition : Un sous-groupe H d'un groupe G est appelé "groupe distingué" si, pour tout g de G et tout h de H, nous avons ghg^{-1} qui est élément de H. Les mathématiciens appellent cela un "automorphisme intérieur" ...

Voyons d'abord un exemple géométrique parlant après quoi nous reviendrons à cette définition avec S_3.

Exemple :

Nous avons vu plus haut les éléments du groupe de symétrie diédral d'ordre 3 du triangle équilatéral. Géométriquement ils correspondent tous à des déplacements du plan dans lequel se trouve le triangle. Nous avions obtenu pour rappel le tableau de composition suivant :

$\nearrow D_6$	id	$t_{2\pi/3}$	$t_{4\pi/3}$	σ_1	σ_2	σ_3
id	id	$t_{2\pi/3}$	$t_{4\pi/3}$	σ_1	σ_2	σ_3
$t_{2\pi/3}$	$t_{2\pi/3}$	$t_{4\pi/3}$	id	σ_3	σ_1	σ_2
$t_{\pi/3}$	$t_{4\pi/3}$	id	$t_{2\pi/3}$	σ_2	σ_3	σ_1
σ_1	σ_1	σ_2	σ_3	id	$t_{2\pi/3}$	$t_{4\pi/3}$
σ_2	σ_2	σ_3	σ_1	$t_{4\pi/3}$	id	$t_{2\pi/3}$
σ_3	σ_3	σ_1	σ_2	$t_{2\pi/3}$	$t_{4\pi/3}$	id

(9.40)

D'abord, nous constatons facilement à l'aide de ce tableau que nous avons :

- Le sous-groupe formé de $\{id\}$ d'ordre 1
- Le sous-groupe formé de $\{id, t_{2\pi/3}, t_{4\pi/3}\}$ d'ordre 3
- Le sous-groupe formé de $\{id, \sigma_1\}$ d'ordre 2
- Le sous-groupe formé de $\{id, \sigma_2\}$ d'ordre 2
- Le sous-groupe formé de $\{id, \sigma_3\}$ d'ordre 2

Parmi ces 5 sous-groupes, voyons lesquels sont distingués (cela est relativement facile à visualiser à l'aide du tableau de composition) :

- Le sous-groupe formé de {id}
- Le sous-groupe formé de $\{id, t_{2\pi/3}, t_{4\pi/3}\}$

Nous allons voir maintenant une chose remarquable ! En numérotant par 1,2 et 3 les sommets du triangle équilatéral, nous pouvons identifier les éléments de D_6 aux éléments suivants de S_3 :

$$\begin{aligned} id = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, t_{2\pi/3} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, t_{4\pi/3} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} \\ \sigma_1 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}, \sigma_2 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}, \sigma_3 \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \end{bmatrix} \end{aligned} \tag{9.41}$$

et reconstruire la même table de composition !

Bon... ce petit interlude fermé, revenons au groupe distingué de S_3 (car il va être important pour notre introduction aux groupe de Galois) et rappelons d'abord que :

$$G = \big\{(1),(1\ 2),(1\ 3),(2\ 3),(1\ 2\ 3),(1\ 3\ 2)\big\} \tag{9.42}$$

Nous construisons joyeusement la table de composition (copie de la précédente.. hé hé !) :

$\nearrow S_3$	(1)	(123)	(1 3 2)	(1 2)	(1 3)	(2 3)
(1)	(1)	(1 2 3)	(1 3 2)	(1 2)	(1 3)	(2 3)
(1 2 3)	(1 2 3)	(1 3 2)	(1)	(2 3)	(1 2)	(1 3)
(1 3 2)	(1 3 2)	(1)	(1 2 3)	(1 3)	(2 3)	(1 2)
(1 2)	(1 2)	(1 3)	(2 3)	(1)	(1 2 3)	(1 3 2)
(1 3)	(1 3)	(2 3)	(1 2)	(1 3 2)	(1)	(1 2 3)
(2 3)	(2 3)	(1 2)	(1 3)	(1 2 3)	(1 2 3)	(1)

(9.43)

et nous voyons que le sous-groupe distingué est formé de :

$$H = \big\{(1), (1\ 2\ 3), (1\ 3\ 2)\big\} \tag{9.44}$$

Définition : Pour tout sous-groupe H stable par les automorphismes intérieurs d'un groupe G, nous appelons "indice de H dans G" le quotient de l'ordre du groupe G par l'ordre du sous-groupe H et nous l'écrivons $[G/H]$.

Par exemple, l'indice du sous-groupe $\big\{(1), (1\ 2)\big\}$dans le groupe S_3 est 6/2 c'est-à-dire 3. Ce concept nous sera très utile lors de notre introduction aux corps de Galois plus loin.

Considérons maintenant, la permutation particulière σ pour aborder le sujet sous un angle différent mais équivalent :

$$\{1, 2, 3, 4\} \rightarrow^{\sigma} \{3, 2, 4, 1\} \tag{9.45}$$

Les mathématiciens ont pour habitude de noter cela, dans un premier temps, sous la forme :

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix} \tag{9.46}$$

avec :

$$\sigma(1) = 3, \sigma(2) = 2, \sigma(3) = 4, \sigma(4) = 1 \tag{9.47}$$

Etant donné σ etτ, deux permutations, il est naturel de regarder leur $\tau \circ \sigma$ composition (rappelons que cela signifie d'abord σ, puisτ comme pour la composition de fonctions).

Ainsi, si :

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix} \text{ et } \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} \tag{9.48}$$

Alors :

$$\tau \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix} \tag{9.49}$$

et :

$$\sigma \circ \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix} \tag{9.50}$$

Maintenant, l'idée est d'interpréter la composition comme une multiplication de permutations. Cette multiplication est alors non-commutative comme nous venons de le constater dans l'exemple précédent. Nous avons en général $\sigma \circ \tau \neq \tau \circ \sigma$.

Chaque bijection a un inverse (une fonction réciproque). Dans notre exemple il s'agit de évidemment de :

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix} \tag{9.51}$$

Géométrique, pour calculer l'inverse σ^{-1} d'un élément σ, il suffit de prendre la réflexion du dessin de σ dans un axe horizontal comme le montre la partie gauche de la figure 9.3 :

Définitions :

D1. L'ensemble des permutations d'un ensemble avec n éléments, muni de cette structure de multiplication, s'appelle le "groupe des permutations d'ordre n" ou "groupe des substitutions d'ordre n", et se note S_n ou encore $S(n)$.

D2. Nous disons qu'un élément σ de S_n est un "cycle d'ordre k", ou un "k-cycle", s'il existe $a_1, a_2, \ldots, a_k \in \{1, \ldots, n\}$ tel que :

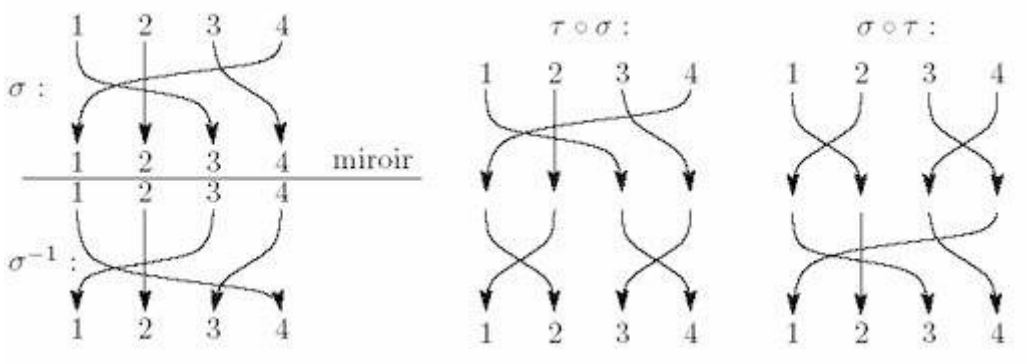


FIGURE 9.4 – Calcul de σ^{-1}

– σ envoie a_1 sur a_2,a_2 sur a_3, $a_3, \ldots, a_{k-1}$ sur a_k et a_k sur a_1
– σ fixe tous les autres éléments de S_n
et nous notons le cycle de la manière $\sigma = (a_1, a_2, \ldots, a_k)$.
Pour mieux comprendre reprenons notre exemple de :

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix} \tag{9.52}$$

Ce groupe symétrique est un 3-cycle noté $\sigma = (1\ 2\ 3)$ car dans l'ordre : 1 envoie sur 3, 3 envoie sur 4 et 4 envoie sur 1 (et le 2 n'étant pas mentionné il reste fixe). Nous pouvons noter cela aussi des façons suivantes équivalents : $\sigma = (3\ 4\ 1)$ ou encore $\sigma = (4\ 1\ 3)$.
Définition : L'ordre d'un k-cycle est k (d'où le nom !).
Effectivement si nous reprenons $\sigma = (1\ 3\ 4)$, nous avons alors :

$$\sigma^2 = (1\ 4\ 3) \text{ et } \sigma^3 = id \tag{9.53}$$

Définition : Nous disons qu'une permutation σ est un "cycle" s'il existe $k \in \mathbb{N}$ tel que σ est un k-cycle.
Attention ! Toute permutation doit s'écrire comme un produit de cycles disjoints (c'est-à-dire qu'un nombre qui apparaît dans un cycle ne doit pas apparaître dans un autre cycle). Par exemple, dans S_9, nous avons :

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 5 & 2 & 8 & 3 & 7 & 1 & 6 & 4 & 9 \end{pmatrix} = (1\ 5\ 7\ 6)(3\ 8\ 4) \tag{9.54}$$

Donc cette permutation est un produit d'un 4-cycle et d'un 3-cycle disjoint.
Nous laisserons d'ailleurs le lecteur vérifier par lui-même que l'ordre de ce groupe S_9 est 12...

REMARQUE Les mathématiciens peuvent démontrer que si σ est un élément qui a une décomposition en c cycles disjoints de longueur $n_1, n_2, \ldots, n_c$ alors l'ordre de σ est le plus petit commun multiple des ordres de tous les cycles disjoints qui le compose.

Nous supposerons également intuitif que S_n dans le vocabulaire commun, un 2-cycle dans s'appelle aussi une "transposition".
Allons un petit peu plus loin. Nous nous proposons de montrer par l'exemple que l'ensemble des transpositions engendre S_n. Autrement, dit, toute permutation s'écrit comme un produit de transpositions.
Reprenons notre exemple (il s'agit d'un permutation paire) :

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix} = (1\ 3\ 4) = (1\ 3)(4\ 1) \tag{9.55}$$

En général, un k-cycle s'écrit donc comme produit de $k-1$ transpositions.

Définition : Soit $\sigma \in S_n$ une permutation. Nous disons que est "permutation paire" si, dans une écriture de σ comme produit de transposition, il y a un nombre paire de transpositions. Nous disons que σ est "permutation impaire" si, dans une écriture de σ comme produit de transpositions, il y a un nombre impaire de transpositions.

Finissons par un petit complément... Nous avons que S_3 est un groupe des permutations d'ordre 3 avec donc $3! = 6$ permutations possibles.

Si nous énumérons les 6 permutations nous avons vu que nous obtenons :

$$\{(1), (1\ 2), (1\ 3), (2\ 3), (1\ 2\ 3), (1\ 3\ 2)\} \tag{9.56}$$

Parmi ceux-ci seulement certains peuvent être écrit comme un produit pair de transpositions :

$$(1\ 2\ 3) = (1\ 2)(3\ 1) \text{ et } (1\ 3\ 2) = (1\ 3)(2\ 1) \tag{9.57}$$

9.4.1 Groupe alterné

Les permutations paires forment avec la permutation identité id, un sous-groupe (non commutatif) que nous appelons le "groupe alterné d'ordre n" et que nous notons A_n. C'est facile de le vérifier avec l'exemple précédent.

10

Analyse fonctionnelle

L'ANALYSE fonctionnelle est la branche des mathématiques et plus particulièrement de l'analyse qui est en rapport avec l'étude des espaces de fonctions. Elle prend ses racines historiques dans l'étude des transformations telles que la transformation de Fourier et dans l'étude des équations différentielles et des intégrales. A ce titre elle englobe tellement de domaines qu'il est difficile de justifier qu'elle fasse l'objet d'un chapitre car il s'agit plutôt d'un domaine d'études. Par ailleurs, c'est à cause de cette difficulté de cerner exactement le domaine qu'elle touche que le lecteur trouvera le théorème fondamental de l'analyse dans le chapitre de calcul différentiel et intégral plutôt qu'ici.

Pourquoi ceci dit utilisons nous le terme "analyse" dans le cas particulier des fonctions. La raison vient pour des raisons historiques à l'étude des divers phénomènes de la nature et la résolution de divers problèmes techniques et par conséquent de mathématiques, qui nous amènent souvent à considérer la variation d'une grandeur en corrélation avec la variation d'une autre ou de plusieurs autres grandeurs. Pour étudier ces variations, de nombreux outils sont à la disposition de tout à chacun :

- L'ingénieur physicien a par exemple fréquemment recours à la représentation graphique (système d'axes cartésien, polaire, logarithmique... concepts sur lesquels nous reviendrons plus en détail) pour déterminer la relation (ou "loi") mathématique qui lient les différentes grandeurs entre elles. Certes ce genre de méthode est (parfois...) esthétique mais les étudiants savent bien combien il est pénible en laboratoire de devoir porter des points sur une feuille de papier ou à l'ordinateur. C'est malheureusement une étape nécessaire (mais dont il faudrait éviter de faire une utilisation abusive) pour comprendre comment nos prédécesseurs travaillaient et ont obtenu les résultats qui nous aident aujourd'hui dans nos avancées en physique théorique.
- L'ingénieur physicien a par exemple fréquemment recours à la représentation graphique (système d'axes cartésien, polaire, logarithmique... concepts sur lesquels nous reviendrons plus en détail) pour déterminer la relation (ou "loi") mathématique qui lient les différentes grandeurs entre elles. Certes ce genre de méthode est (parfois...) esthétique mais les étudiants savent bien combien il est pénible en laboratoire de devoir porter des points sur une feuille de papier ou à l'ordinateur. C'est malheureusement une étape nécessaire (mais dont il faudrait éviter de faire une utilisation abusive) pour comprendre comment nos prédécesseurs travaillaient et ont obtenu les résultats qui nous aident aujourd'hui dans nos avancées en physique théorique.
- Le mathématicien et le physicien théoricien ont habituellement horreur d'avoir recours aux méthodes papier-crayon-gribouillage. Quoiqu'il en soit, le rôle du mathématicien ou du physicien est de développer de nouvelles théories à l'aide d'axiomes ou de principes mathématiques ce qui ne devrait nécessiter aucunement le recours à la représentation graphique et à l'accès aux mesures expérimentales qui y sont souvent rattachées.

REMARQUE Avant de commencer la lecture de ce qui va suivre, il peut être utile de rappeler au lecteur que la définition du concept de "fonction" (et les propriétés élémentaires y relatives) sont données dans le chapitre de Théorie Des Ensembles.

10.1 Représentations

NOUS ALLONS VOIR dans ce qui va suivre, dans un premier temps, comment représenter différentes grandeurs liées de façon tabulaire et graphique (eh oui ! il faut bien car cela aide à

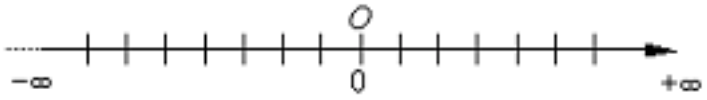


FIGURE 10.1 – Représentation graphique des nombres

comprendre) et ensuite comment analyser mathématiquement les propriétés de ces représentations uniquement à l'aide d'outils mathématiques abstraits.

Définition 10.1 *Une fonction est dite "fonction univalente", si le nombre de ses arguments (paramètres ou variables) est égal à un. Dans le cas d'une fonction à deux arguments, nous parlons de "fonction bivalente", etc.*

10.1.1 Représentation tabulaire

Parmi le mode de représentation visuel des fonctions, la plus intuitive et la plus ancienne est celle où nous disposons dans la colonne ou la ligne d'un tableau de façon ordonnée les valeurs de la variable indépendante $x_1, x_2, \ldots, x_n$ et les valeurs correspondantes, dites "variables transformées" de la fonction $y_1, y_2, \ldots, y_n$ dans une autre colonne ou ligne alignée.

Telles sont par exemple, les tables des fonctions trigonométriques, les tables logarithmiques, etc. et au cours de l'étude expérimentale de certains phénomènes des tables qui expriment la dépendance fonctionnelle existant entre des grandeurs physiques mesurées tel que les relevés de la température de l'air enregistrés dans une station météorologique durant une journée.

Bien évidemment, ce concept est généralisable à toute fonction multivalente quelque soit son ensemble de définition.

Cependant, cette méthode elle est laborieuse et ne permet pas de voir directement le comportement de la fonction et donc une analyse visuelle simple et intéressante de ses propriétés. Elle a pour avantage quand même de ne pas nécessiter d'outils spéciaux ou de connaissances mathématiques poussées.

10.1.2 Représentation graphiques

Les nombres naturels, relatifs, réels ou purement complexes (cf. chapitre sur les Nombres) peuvent tous êtres représentés le plus simplement du monde par des points sur un axe numérique (ligne droite) infini.

Pour ce faire, nous choisissons sur cet axe :

1. Un point O appelé "origine"
2. Un sens positif, que nous indiquons par une flèche horizontale
3. Une unité de mesure (représenté habituellement par un petit trait vertical : la "graduation")

Tel que :

Le plus souvent nous disposons (par tradition) l'axe horizontalement et choisissons la direction de gauche à droite.

Remarque : Le point (lettre) O, représente très fréquemment le nombre zéro en mathématique mais nous pourrions très bien choisir de mettre l'origine ailleurs. Par exemple, en physique le point O est souvent positionné à l'emplacement du barycentre d'un système.

Il est évident que le fait que les ensembles de nombres dont nous avons parlé soient ordonnés implique que tout nombre est représenté par un seul point de l'axe numérique. Ainsi, deux nombre réels distincts correspondent deux points différents de l'axe numérique.

Ainsi, il existe une correspondance biunivoque entre tous les nombres et tous les points de l'axe numérique (dans le cas des nombres réels ou complexes, il correspond non pas un nombre à chaque graduation mais un nombre à chaque point de l'axe). Ainsi, à chaque nombre correspond un point ou une graduation unique et inversement à chaque point ou graduation correspond un seul nombre dont il est l'image.

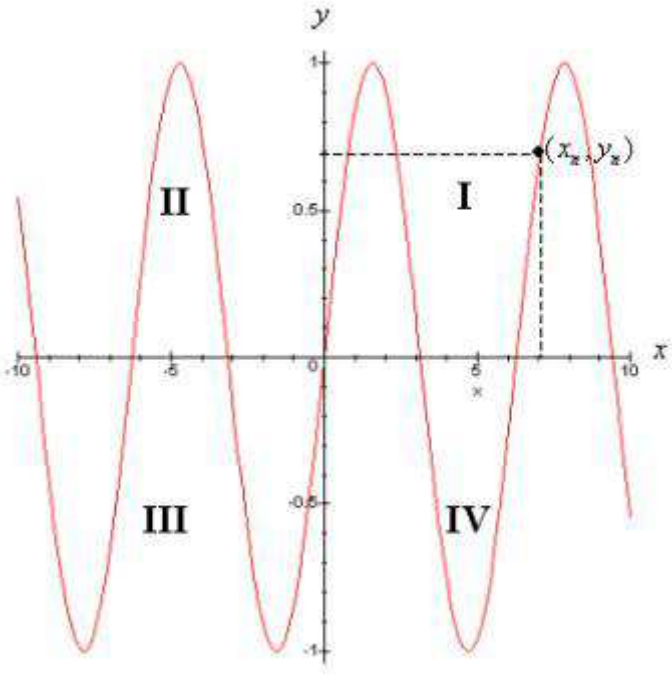


FIGURE 10.2 – Représentation graphique des nombres

Représentation planes

Il existe outre les représentations unidimensionnelle d'autres de dimensions supérieurs (ouf !) qui nous permettent de tracer non plus que des simples points sur une droite unidimensionnelle mais des fonctions d'une variable. Voyons de quoi il s'agit :

Nous pouvons à chaque valeur d'une variable x reportée sur un axe horizontal, appelé "axe des abscisses" ou "axe des x", faire correspondre une valeur y au travers d'une fonction f :

$$y = f(x) \tag{10.1}$$

reportée sur un axe vertical, appelé "axe des ordonnées" ou "axe des y" qui passe par le croisement défini par l'origine O tel que représenté à la figure 10.2(exemple arbitraire).

L'ensemble des points du plan noté sous les variantes XOY, XY ou encore xOy, Oxy, xy, dont les abscisses représentent par tradition les valeurs de la variable indépendante et les ordonnées les valeurs correspondantes de la fonction, est appelé "graphique plan" de cette fonction. S'il n'y a pas de confusion possible on dira simplement "graphique".

Dans le cas d'une représentation par un système de coordonnées rectangulaires (cartésien, polaire ou logarithmique) comme la figure 10.2, nous pouvons observer que l'ensemble du plan des coordonnées est séparé en quatre surfaces que nous avons pour habitude d'appeler "quadrants".

REMARQUE Lorsque nous souhaitons mettre en évidence un point particulier de la fonction représentée, nous y dessinons un petit rond tel que présenté dans la figure 10.2.

Un autre cas classique de représentation graphique plane connu par un grand nombre d'étudiants est le tracé des polynômes (cf. chapitre de Calcul Algébrique) à coefficients réels.

Effectivement, pour résoudre les équations polynomiales du second degré (cf. chapitre de Calcul Algébrique), il est fréquent dans les petites classes que le professeur demande en plus à ses élèves de donner un expression algébrique des racines de :

$$y = f(x) = ax^2 + bx + c = 0 \tag{10.2}$$

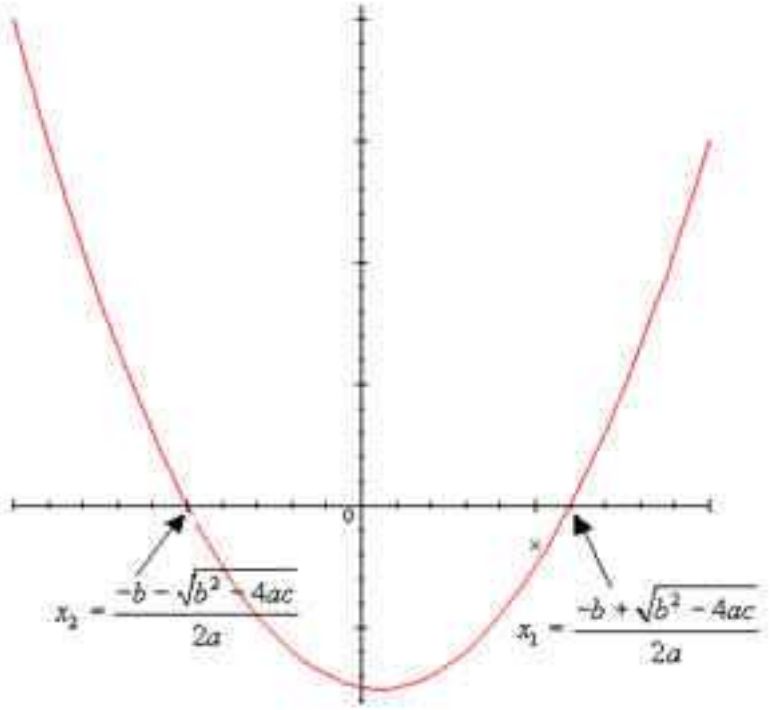


FIGURE 10.3 – Racines équation second degré

données par, rappelons-le :

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{10.3}$$

une résolution graphique où les deux racines (dans le cas où il y en a deux distinctes réelles) sont données par l'intersection de la parabole avec l'axe des abscisses (bien évidemment, si l'équation n'a pas de solutions, il n'y a pas d'intersections...) comme présenté à la figure 10.3

La représentation graphique étant généralisable aux équations polynomiales du 3ème, 4ème et 5ème degré (nous démontrerons bien plus loin, à l'aide de la théorie de Galois qu'il n'est pas possible d'obtenir une expression algébrique générale des racines d'une équation polynômiale du 5ème degré et supérieur).

Représentations 3D

Bien évidemment, dans le cas d'une fonction trivalente (tri-dimensionnelle), c'est-à-dire dont un paramètre dépend de deux autres, le principe reste le même à la différence que le nombre de quadrants double. Par exemple la figure 10.4 (c'est toujours joli à regarder... même si cela ne représente rien de concret) :

Cette méthode de représentation et d'analyse d'une fonction trivalente était longue à mettre en place il y a une dizaine d'années mais avec l'aide des ordinateurs en ce 21ème siècle ce problème (de temps) est assez bien résolu...

Représentations vectorielles

Il est aussi fréquemment fait usage des représentations graphiques dans le cadre de la géométrie analytique pour simplifier les analyses ou faire des démonstration de théorèmes connus sous forme visuelles (il faut cependant pas en abuser !).

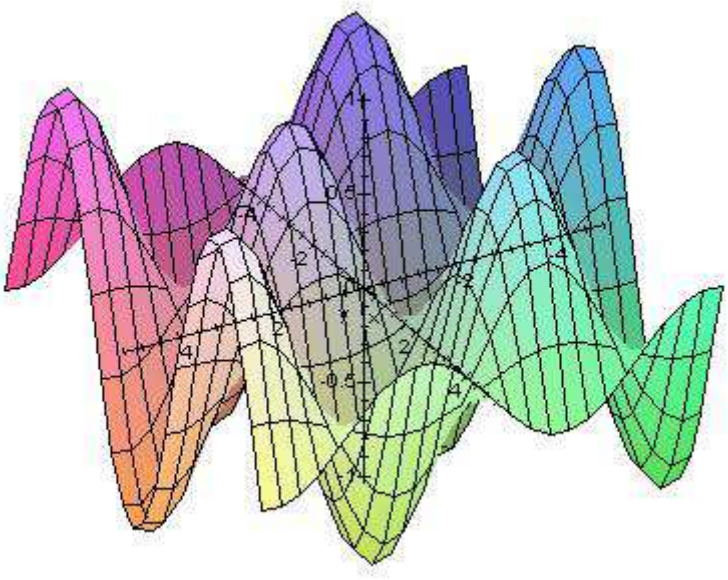


FIGURE 10.4 – Racines équation second degré

Ainsi, nous pouvons introduire par exemple le concept de norme (cf. chapitre de Calcul Vectoriel) de manière simpliste en représentant graphiquement la distance entre deux points et en appliquant le théorème de Pythagore (cf. chapitre de Géométrie Euclidienne) qui sera supposé connu.

Ainsi, représentons trois points P_1, P_2, P_3 sur un graphique plan dans lequel a été défini un repère tel que présenté dans le graphique de la figure 10.5.

Si $x_1 \neq x_2$ et $y_1 \neq y_2$ (comme sur la figure 10.5), les points P_1, P_2, P_3 sont les sommets d'un triangle rectangle. Par application du théorème de Pythagore (cf. chapitre de Géométrie Euclidienne) :

$$[d(P_1, P_2)]^2 = [d(P_1, P_3)]^2 + [d(P_3, P_2)]^2 \tag{10.4}$$

Sur la figure, nous voyons que :

$$d(P_1, P_3)] = |x_2 - x_1| \text{ et } d(P_3, P_2) = |y_2 - y_1| \tag{10.5}$$

Puisque $|x|^2 = x^2, \forall x \in \mathbb{R}$, nous pouvons écrire :

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{10.6}$$

Si $x_1 = y_1 = 0$, nous nous retrouvons avec une relation appelée "norme", "module" ou encore "distance" que nous avions déjà défini dans le cadre de note étude de l'analyse vectorielle (cf. chapitre de Calcul Vectoriel).

Bien évidemment, si nous considérons deux points $P_1(x_1, y_1), P_2(x_2, y_2)$, nous pouvons déterminer si un troisième point $P_3(x_3, y_3)$ est sur la médiatrice (cf. chapitre de Géométrie Euclidienne) des deux premiers et qu'il suffit pour cela que bien évidemment (par définition même de la médiatrice) :

$$d(P_1, P_3) = d(P_2, P_3) \tag{10.7}$$

Comme $(x_1, y_1), (x_2, y_2)$sont connus, nous pouvons facilement exprimer une "expression analy-

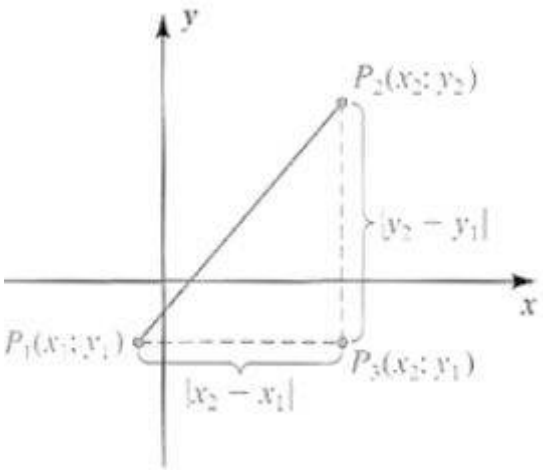


FIGURE 10.5 – Racines équation second degré

tique" de la médiatrice du type :

$$\begin{aligned}
\sqrt{(x_1-x_3)^2+(y_1-y_3)^2} &= \sqrt{(x_2-x_3)^2+(y_2-y_3)^2} \\
(x_1-x_3)^2+(y_1-y_3)^2 &= (x_2-x_3)^2+(y_2-y_3)^2 \\
x_1^2-2x_1x_3+x_3^2+y_1^2-2y_1y_3+y_3^2 &= x_2^2-2x_2x_3+x_3^2+y_2^2-2y_2y_3+y_3^2 \\
x_1^2-2x_1x_3+y_1^2-2y_1y_3 &= x_2^2-2x_2x_3+y_2^2-2y_2y_3 \\
2y_2y_3-2y_1y_3 &= x_2^2-2x_2x_3+y_2^2-y_1^2+2x_1x_3-x_1^2 \\
y_3 2(y_2-y_1) &= 2(x_1-x_2)x_3+(x_2^2-x_2^2+y_2^2-y_1^2) \\
y_3 &= \frac{(x_1-x_2)}{(y_2-y_1)}x_3+\frac{(x_2^2-x_1^2+y_2^2-y_1^2)}{2(y_2-y_1)} \\
y_3 &= ax_3+b=f(x_3)
\end{aligned} \tag{10.8}$$

où a, b sont des constantes et où tout point qui satisfait cette relation, qui est en l'occurrence l'équation d'une droite, se trouve sur la médiatrice.

Par ailleurs, il est aisé de visualiser que le point milieu du segment de droite est donné par :

$$M = \left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right) \tag{10.9}$$

Donc nous voyons qu'avec une simple représentation graphique, nous pouvons obtenir des résultats qui sont parfois (...) plus évident pour les étudiants.

Profitons de cet exemple pour définir quelques concepts sur lesquels nous reviendrons et faire quelques rappels.

Définition 10.2 *Toute fonction de la forme d'un polynôme (cf. chapitre de Calcul Algébrique) de degré 1 à coefficients réels constants :*

$$f(x) = y = ax + b \tag{10.10}$$

est l'expression analytique de ce que nous appelons une "droite" de "pente" a et "d'ordonnée à l'origine" b (quand $x = 0$) .

Bien évidemment, si :

$$f(x) = y = b \tag{10.11}$$

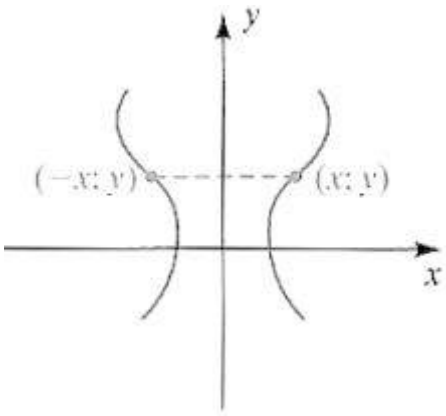


FIGURE 10.6 – Fonction symétrique par rapport à l'axe des ordonnées

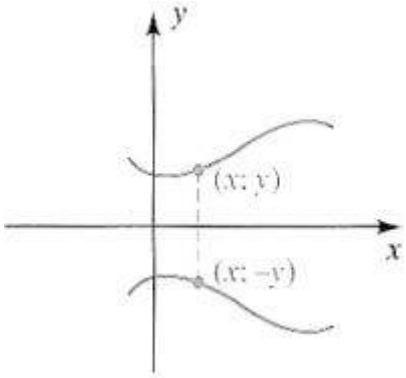


FIGURE 10.7 – Fonction symétrique par rapport à l'axe des abcisses

la droite est horizontale si nous la représentons graphique puisque y est constant pour tout x et vaut alors b. Inversement, si :

$$f(x) = y = a \tag{10.12}$$

la droite est une verticale.

Propriétés des représentations graphiques

Selon le type de graphique que nous visualisons (en particulier les graphiques plans) il est possible d'extraire certaines propriétés de base. Voyons les plus importantes à connaître pour les graphiques plan d'une fonction à une variable :

$$y = f(x) \tag{10.13}$$

P1. Le graphique d'une fonction est symétrique par rapport à l'axe des ordonnées si le changement de x en $-x$ ne modifie pas l'équation tel que représenté à la figure 10.6.

P2. Le graphique d'une fonction est symétrique par rapport à l'axe des abscisses si le changement de y en $-y$ ne modifie pas l'équation tel que représenté à la figure 10.7.

P3. Le graphique d'une fonction est symétrique par rapport à l'origine si le changement simultané de y en $-y$ et de x en $-x$ ne modifie pas l'équation tel que représenté à la figure 10.8.

P4. Soit une fonction $y = f(x)$, si nous ajoutons une constante $c^{te} > 0$ à cette fonction tel que nous écrivions :

$$y = f(x) + c^{te} \tag{10.14}$$

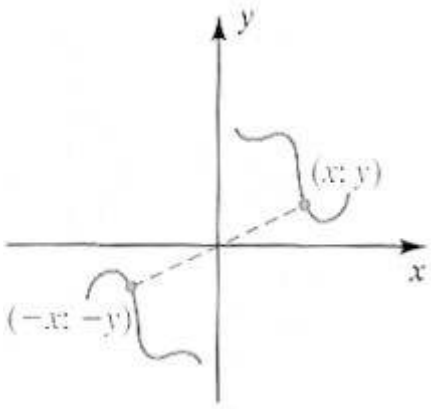


FIGURE 10.8 – Fonction symétrique par rapport à l'axe des ordonnées

alors le graphique de f est déplacé (ou "translaté") verticalement vers le haut d'une distance c^{te} et inversement si $c^{te} > 0$ mais que :

$$y = f(x) - c^{te} \tag{10.15}$$

alors le graphique est bien évidemment translaté verticalement vers le bas.

Nous pouvons aussi envisager des translations horizontales de graphiques. Précisement, si $c^{te} > 0$, alors $y = f(x)$ est translaté horizontalement vers la droite si nous écrivons :

$$y = f(x - c^{te}) \tag{10.16}$$

et inversement, translaté horizontalement vers la gauche, si nous écrivons :

$$y = f(x + c^{te}) \tag{10.17}$$

Pour étirer ou comprimer verticalement un graphique, il suffit de multiplier la fonction $y = f(x)$ par une constante $c^{te} > 1$ et respectivement $0 < c^{te} < 1$tel que :

$$y = c^{te} f(x) \tag{10.18}$$

et pour étirer ou comprimer horizontalement un graphique, il suffit de même, de multiplier la fonction $y = f(x)$ par une constante $c^{te} > 1$ et respectivement $0 < c^{te} < 1$tel que :

$$y = f(c^{te} x) \tag{10.19}$$

REMARQUE Translater, étirer, comprimer un graphique ou lui faire subit une symétrie c'est le transformer. Le graphique résultant de ces transformations est appelé le "transformé" du graphique de départ.

Définition 10.3 *Nous disons qu'une fonction f est :*

Une "fonction croissante" *ou "fonction croissante au sens large" sur I si pour tout couple (x_1, x_2), d'éléments de I tels que $x_1 \leq x_2$, nous avons $f(x_1) \leq f(x_2)$. Ce que nous notons de manière condensée :*

$$\{\forall (x, y) \in I^2 \wedge x \leq y\} \Rightarrow f(x) \leq f(y) \tag{10.20}$$

Une "fonction décroissante" *ou "fonction décroissante au sens large" sur I si pour tout couple (x_1, x_2), d'éléments de I tels que $x_1 \leq x_2$, nous avons $f(x_1) \geq f(x_2)$. Ce que nous notons de manière condensée :*

$$\{\forall (x, y) \in I^2 \wedge x \leq y\} \Rightarrow f(x) \geq f(y) \tag{10.21}$$

Remarque *Une "fonction monotone" ou "fonction monotone au sens large" sur I si elle est croissante sur I ou décroissante.*

Une "fonction strictement croissante" *sur I si pour tout couple (x_1, x_2), d'éléments de I tels que $x_1 < x_2$, nous avons $f(x_1) < f(x_2)$. Ce que nous notons de manière condensée :*

$$\{\forall (x, y) \in I^2 \wedge x < y\} \Rightarrow f(x) < f(y) \tag{10.22}$$

Une "fonction strictement décroissante" *sur I si pour tout couple (x_1, x_2), d'éléments de I tels que $x_1 < x_2$, nous avons $f(x_1) > f(x_2)$. Ce que nous notons de manière condensée :*

$$\{\forall (x, y) \in I^2 \wedge x < y\} \Rightarrow f(x) > f(y) \tag{10.23}$$

Remarque *Nous disons qu'une "fonction strictement monotone" sur I si elle est strictement croissante sur I ou strictement décroissante sur I.*

10.1.3 Représentations analytiques

Le mode de représentation analytique est de loin le plus utilisé et consiste à représenter toute fonction en une "expression analytique" qui est la notation mathématique symbolique et abstraite de l'ensemble des opérations mathématiques connues que l'on doit appliquer dans un certain ordre à des nombres et des lettres exprimant des grandeurs constantes ou variables que nous cherchons à analyser.

Remarquons que par ensemble des opérations mathématiques connues, nous envisageons non seulement les opérations mathématiques vues dans la section arithmétique (addition, soustraction, extraction de la racine, etc.) mais également toutes les opération qui seront définies au fur et à mesure dans le présent site internet.

Si la dépendance fonctionnelle $y = f(x)$ est telle que f est une expression analytique, nous disons alors que la "fonction y de x" est "donnée analytiquement". Voici quelques exemples d'expressions analytiques simples :

$$y = x^4 - 2, y = \frac{x+1}{x-1}, y = \sqrt{1-x^2} \tag{10.24}$$

Lorsque nous avons déterminé l'équation de la médiatrice, nous avons obtenu une expression analytique de la droite visuelle qui l'a caractérise sous la forme d'une fonction du type :

$$\boxed{f(x) = y = ax + b} \tag{10.25}$$

qui rappelons-le, est donc l'expression analytique l'équation d'une droite, appelée également "équation linéaire" ou "fonction affine", sur un plan dont si deux points $P_1(x_1, y_1), P_2(x_2, y_2)$ sont connus, la pente est donnée par le rapport de l'accroissement vertical sur l'accroissement horizontal tel que :

$$a = \frac{y_2 - y_1}{x_2 - x_1} \equiv \frac{\Delta y}{\Delta x} \tag{10.26}$$

Une application sympathique et triviale consiste à démontrer analytiquement que deux droites non verticales sont parallèles si et seulement si elles ont la même pente. Ainsi, soit deux droites données par les équations :

$$y_1 = ax_1 + b_1, y_2 = a_2 + b_2 \tag{10.27}$$

Les droites se coupent en un point (x, y) si et seulement si les valeurs de y sont égales pour un certain x, c'est-à-dire :

$$a_1 x + b_1 = a_2 x + b \tag{10.28}$$

La dernière équation peut être résolue par rapport à x si et seulement si $a_1 - a_2 \neq 0$. Nous avons donc montré que les droites y_1, y_2 se coupent si et seulement si $a_1 \neq a_2$. Donc, elles ne se coupent pas (elles sont parallèles) si et seulement si $a_1 = a_2$.

De façon assez simple en appliquant le théorème de Pythagore, il n'est pas compliqué de déterminer que l'équation d'un cercle de centre $C(h, k)$ à pour équation (nous avons pour habitude en mathématique de ne pas expliciter y pour l'équation du cercle ainsi, l'équation de ce dernier est visuellement beaucoup plus esthétique et parlante)

$$r = \sqrt{(x-h)^2 + (y-k)^2} \tag{10.29}$$

Dans ces exemples les fonctions sont exprimées analytiquement par une seule formule (égalité entre deux expressions analytiques) qui définit dans un même temps le "domaine naturel de définition" des fonctions.

Définition 10.4 *Le "domaine naturel de définition" d'une fonction donnée par une expression analytique est l'ensemble des valeurs x pour lesquelles l'expression du membre droit a une valeur bien déterminée.*

Par exemple, la fonction :

$$y = \frac{x+1}{x-1} \tag{10.30}$$

est définie pour toutes les valeurs de x, excepté la valeur $x = 1$ où nous avons une singularité (division par zéro).

REMARQUE Il existe une infinité de fonction et nous ne pouvons toutes les exposer ici, cependant nous en rencontrerons plus d'un millier sur l'ensemble du site et cela devrait amplement suffire à se faire une idée de leur étude.

10.2 Fonctions

Définition 10.5 *Nous disons que y est une fonction de x et nous écrirons $y = f(x), y = \varphi(x)$, etc., si à chaque valeur de la variable x appartenant à un certain domaine de définition (ensemble) D, correspond une valeur de la variable y dans un autre domaine de définition (ensemble) E. Ce que nous notons :*

$$D \xrightarrow{f} E \; f : D \to E \tag{10.31}$$

La variable x est appelée "variable indépendante" ou "variable d'entrée" et y "variable dépendante".

La dépendance entre les variables x et y s'appelle une "dépendance fonctionnelle". La lettre f, qui entre dans la notation symbolique de la dépendance fonctionnelle, indique qu'il faut appliquer certaines opérations à x pour obtenir la valeur correspondant y. Nous écrivons parfois $y = y(x)$ au lieu de $y = f(x)$; dans ce cas la lettre y exprime en même temps la valeur de la fonction et le symbole des opérations appliquées à x.

REMARQUE Comme nous l'avons vu lors de notre étude de la théorie des ensembles, une application (ou fonction) peut-être injective, bijective ou surjective. Il convient donc que le lecteur dont ces notions ne sont pas connues aille en priorité lire ces définitions.

Définition 10.6 *L'ensemble des valeurs x pour lesquelles la valeur de la fonction y est donnée par la fonction $f(x)$ est appelé "domaine d'existence" de la fonction (ou domaine de définition de la fonction).*

Définition 10.7 *La fonction $y = f(x)$ est dite "fonctions croissante" si à une plus grande valeur de la variable indépendante correspond une plus grande valeur de la fonction (de l'image). Nous définissons de manière analogue mais inverse la "fonction décroissante". Une "fonction constante" est une fonction pour laquelle à toute valeur de la variable indépendante correspond toujours une même image constante.*

Définition 10.8 *La fonction $y = f(x)$ est dit "fonction périodique" s'il existe un nombre constant c^{te} tel que la valeur de la fonction ne change pas quand nous ajoutons c^{te} (ou que retranchons) le nombre à la variable indépendante tel que :*

$$y = f(x + c^{te}) \tag{10.32}$$

Ce qui correspondant à une translation selon x. La plus petite constante satisfaisant à cette condition est appelée "période" de la fonction. Elle est fréquemment notée T en physique.

Définition 10.9 *En calcul différentiel et intégral, l'expression :*

$$\frac{f(x+h) - f(x)}{h} \tag{10.33}$$

avec $h \neq 0$ est d'un intérêt particulier. Nous l'appelons un "quotient d'accroissement" (nous reviendrons beaucoup plus en détail sur ce sujet lors de notre étude du calcul différentiel et intégral).

Définition 10.10 *Nous utilisons certaines propriétés des fonctions pour faciliter leur représentation graphique et leur analyse. En particulier, une fonction $f(x)$ est dite "fonction paire" si :*

$$f(x) = f(-x) \tag{10.34}$$

pour tout x dans son domaine de définition. Une fonction est dite "fonction impaire" si :

$$f(-x) = -f(x) \tag{10.35}$$

pour tout x dans son domaine de définition.

Ainsi, pour résumer une fonction paire est une fonction qui ne dépend pas du signe de la variable et une fonction impaire change de signe quand on change le signe de la variable (la spirale de Cornus dans le chapitre de Génie Civil est un bon exemple pratique de fonction impaire). Ce concept nous sera très utile pour simplifier certaines expressions très utiles en physique (comme les transformées de Fourier des fonctions paires ou impaires par exemple !).

Montrons maintenant que toute fonction $f(x)$ est la somme d'une fonction paire $g(x)$ et d'une fonction impaire $h(x)$.

REMARQUE Ce type de théorème qui consiste à relier un concept général par un cas particulier et son opposé se retrouve souvent en mathématiques. Nous retrouverons de tels exemples en calcul tensoriel avec les tenseurs symétriques et anti-symétriques (cf. chapitre de Calcul Tensoriel) ou encore en physique quantique avec les opérateurs hermitiques et anti-hermitiques (cf. chapitre de Physique Quantique Ondulatoire).

Preuve Posons :

$$f(x) = g(x) + h(x) \tag{10.36}$$

alors :

$$f(-x) = g(x) - h(x) \tag{10.37}$$

Si nous sommons, nous avons dès lors :

$$g(x) = \frac{f(x) + f(-x)}{2} \tag{10.38}$$

et en soustrayant :

$$h(x) = \frac{f(x) - f(-x)}{2} \tag{10.39}$$

Il existe donc bien une décomposition paire et impaire de toute fonction. •

Définition 10.11 *De façon générale, si $f(x)$ et $g(x)$ sont des fonctions quelconques, nous utilisons la terminologie et les notations données dans le tableau suivant suivant :*

Terminologie	*Valeur de la fonction*
Somme $f+g$	$(f+g)(x) = f(x)+g(x)$
Différence $f-g$	$(f-g)(x) = f(x)-g(x)$
Produit $f \cdot g$	$(f \cdot g)(x) = f(x)g(x)$
Quotient $\frac{f}{g}$	$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$

Les domaines de définition de $f+g$, $f-g$, $f \cdot g$ sont l'intersection I des domaines de définition de $f(x)$ et de $g(x)$, c'est-à-dire les nombres qui sont communs aux deux domaines de définition. Le domaine de définition de $f \cdot g^{-1}$ est quant à lui le sous-ensemble de I comprenant tous les x de I tels que$g(x) \neq 0$.

Définition 10.12 *Soit y une fonction de u et u une fonction de la variable x, alors y dépend alors de x et nous avons ce que nous appelle une "fonction composée" et que nous notons :*

$$y = f(g(x)) \text{ ou } (f \circ g)(x) \tag{10.40}$$

Pour la dernière notation, il faut lire "f rond g" et ne pas confondre le "rond" avec la notation du produit scalaire que nous verrons lors de notre étude du calcul vectoriel (cf. chapitre de Calcul Vectoriel).

Le domaine de définition de la fonction composée est soit identique au domaine tout entier de définition de la fonction $u = g(x)$, soit à la partie de ce domaine dans laquelle les valeurs de u sont telles que les valeurs correspondantes $f(u)$ appartiennent au domaine de définition de cette fonction.

Le principe de fonction composée peut être appliquée non seulement une fois, mais un nombre arbitraire de fois.

Si x ne dépend pas d'une autre variable (ou qu'elle n'est pas elle même une fonction composée), on dit alors que $\varphi(x)$ est une "fonction élémentaire".

Les principales fonctions élémentaires sont des fonctions dont l'expression analytique est l'une des suivantes :

La "fonction puissance"

$$f(x) = x^m \tag{10.41}$$

où m est un nombre positif différent de 1 (sinon il s'agit d'une fonction linéaire) (cfr figure 10.9).

La "fonction exponentielle"

$$f(x) = a^x \tag{10.42}$$

où a est un nombre positif différent de 1.

La "fonction logarithmique"

$$f(x) = \log_a(x) \tag{10.43}$$

où la base du logarithme est un nombre positif a différent de l'unité (cette fonction sera définie rigoureusement un peu plus loin).

REMARQUE Les fonctions exponentielles et logarithmiques sont appelées parfois des "fonctions transcendantes".

Les "fonctions trigonométriques" (cf. chapitre de Trigonométrie)

$$y = \sin(x), y = \cos(x), y = \tan(x), y = \cot(x), y = \sec(x), y = cosec(x), \ldots \tag{10.44}$$

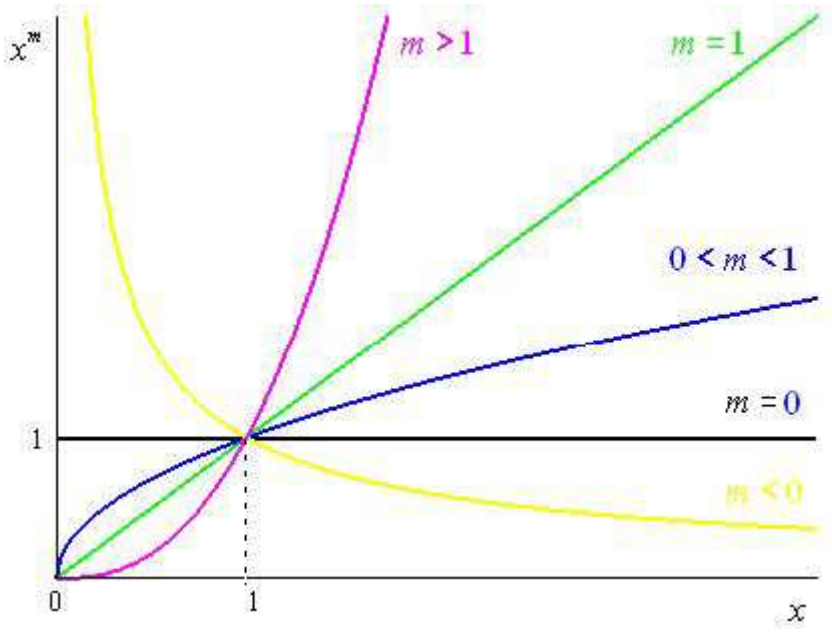


FIGURE 10.9 – Fonction puissance

Les "fonctions polynomiales" Les

$$y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x^1 + a_0 \tag{10.45}$$

où $a_n, a_{n-1}, \ldots, a_1, a_0$sont des nombres constants appelés coefficients et n est un entier positif que nous appelons "degré du polynôme" (cf. chapitre de Calcul Algébrique). Il est évident que cette fonction est définie pour toutes les valeurs de x, c'est-à-dire qu'elle est définie dans un intervalle infini.

Les "fractions rationnelles" qui sont des divisions de polynômes (cf. chapitre de Calcul Algébrique) :

$$y = \frac{a_0 x^n + a_1 x^{n-1} + \cdots + a_n}{b_0 x^n + b_1 x^{n-1} + \cdots + b_n} \tag{10.46}$$

Les "fonctions algébriques" sont définies par le fait que la fonction est le résultat d'opérations d'addition, de soustraction, de multiplication, de division, de variables élevées à une puissance rationnelle non entière.

REMARQUE Il existe cependant un très grand nombre de fonction que nous rencontrerons dans les différents chapitre du site. Citons par exemple les "fonctions de Bessel" (cf. chapitre des Suites Et Séries), les "fonctions lipschitziennes" (cf. chapitre de Topologie), les "fonctions de Dirac" (cf. chapitre de Calcul Différentiel et Intégral), les "fonctions de répartition et de distribution" (cf. chapitre de Statistiques), la "fonction gamma d'Euler" (cf. chapitre de Calcul Différentiel Et Intégral), etc.

10.2.1 Limites et continuité des fonctions

Nous allons considérer maintenant des variables ordonnées d'un type spécial, que nous définissons par la relation "la variable tend vers une limite". Dans la suite de ce cours, la notion de limite d'une variable va jouer un rôle fondamental, étant intimement liée aux notions de base de l'analyse mathématique, la dérivée, l'intégrale, etc.

Définition 10.13 *Le nombre a est appelé la "limite" de la grandeur variable x, si, pour tout nombre arbitrairement petit $\varepsilon > 0$ avons :*

$$|x - a| < \varepsilon \tag{10.47}$$

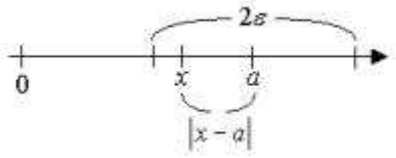


FIGURE 10.10 – Présentation géométrique de la limite

Si le nombre a est la limite de la variable x, nous disons que "x tend vers la limite a".

Nous pouvons définir également la notion de limite en partant de considérations géométriques (cela peut aider à mieux comprendre... quoique pas toujours...) :

Le nombre constant a est la limite de la variable x, si pour tout voisinage donné, aussi petit qu'il soit, de centre a et de rayon ε, nous pouvons trouver un valeur x telle que tous les points correspondant aux valeurs suivantes de la variable appartiennent à ce voisinage (notions que nous avons défini précédemment). La figure 10.10 représente cela géométriquement.

REMARQUE Il devrait être trivial que la limite d'une grandeur constante est égale à cette constante, puisque l'inégalité $|x - c| = |c - c| = 0 < \varepsilon$ est toujours satisfaite pour ε arbitraire.

Il découle également de la définition de la limite qu'une grandeur variable ne peut pas avoir n'importe comment deux limites. En effet, si $\lim_{x\to a} x = a$ et $\lim_{x\to b} x = b$ avec $a < b$, x doit satisfaire simultanément aux deux inégalités suivantes :

$$|x - a| < \varepsilon \text{ et } |x - b| < \varepsilon \tag{10.48}$$

pour ε arbitrairement choisi. Mais si nous faisons une représentation géométrique identique à la précédente, nous voyons assez aisément que cela est impossible si :

$$\varepsilon < \frac{b - a}{2} \tag{10.49}$$

Il ne faut également pas s'imaginer que chaque variable doit nécessairement avoir une limite.

Définition 10.14 *La variable x tend vers l'infini, si pour chaque nombre positif donné M nous indiquons une valeur de x à partir de laquelle toutes les valeurs conséquentes de la variable (valeurs de la variable appartenant dans le voisinage défini à partir de valeur indiquée précédemment) x vérifient l'inégalité $|x| > M$.*

Nous pouvons vérifier ce genre de cas dans les suites arithmétiques, géométriques, ou harmoniques où chaque terme de la progression est une valeur que prend la variable x.

La variable x "tend vers plus l'infini", ou $x \to +\infty$ si pour $M > 0$ arbitraire, à partir d'une certaine valeur, toutes les valeurs conséquentes de la variable vérifient l'inégalité $M < x$. C'est typiquement le genre de considération que nous avons pour des progressions divergentes vers l'infini ou à partir d'un certain terme de valeur égale à M tous les termes suivant sont supérieurs à M. La variable x "tend vers moins l'infini" ou $x \to -\infty$ si pour $M > 0$ arbitraire, à partir d'une certaine valeur, toutes les valeurs suivantes de la variable vérifient l'inégalité $x < -M$.

Définition 10.15 *Soit $y = f(x)$ une fonction définie dans un voisinage du point a ou en certains points de ce voisinage. La fonction $y = f(x)$ tend vers la limite $b (y \to b)$ lorsque x tendant vers $a (x \to a)$, si pour chaque nombre positif ε, aussi petit qu'il soit, nous pouvons indiquer un nombre positif δ tel que tous les x différents de a et vérifiant l'inégalité $|x - a| < \delta$ satisfont également :*

$$|f(x) - b| < \varepsilon \tag{10.50}$$

L'inégalité $|x-a|<\delta$ permet d'exprimer le côté (ou le sens) depuis lequel nous venons avec notre x. Car sur le système d'axe représentant des valeurs ordonnées, nous pouvons, pour une valeur donnée, venir de sa gauche ou de sa droite pour se rapprocher d'elle (imaginez-vous au besoin, un bus qui peut venir depuis un côté ou un autre de la route tant que la distance qui le sépare à l'arrêt qui nous intéresse est inférieur à δ).

Si b est la limite de la fonction $f(x)$ quand $x \to a$, nous écrivons alors sur ce site en tout les cas :

$$\lim_{x\to a} f(x) = b \tag{10.51}$$

Pour définir le côté depuis lequel nous venons en appliquant la limite, nous utilisons une notation particulière (rappelons que cela permet de connaître de quel côté de la route vient notre bus).

Ainsi, si $f(x)$ tend vers la limite b_1 quand x tend vers un nombre a en ne prenant que des valeurs plus petites que a, nous écrirons alors :

$$\lim_{x\to a_-} f(x) = b_1 \tag{10.52}$$

(remarquez le petit - en indice) et nous appellerons la "limite à gauche" de la fonction $f(x)$ au point a (car rappelez vous que l'axe des ordonnées va de $-\infty$ à $+\infty$, donc les petites valeurs par rapport à une valeur donnée, se trouvent à gauche). Si x prend des valeurs plus grande que a, nous écrirons alors :

$$\lim_{x\to a_+} f(x) = b_2 \tag{10.53}$$

(remarquez le petit + en indice) et nous appellerons la "limite à droite" de la fonction au point a.

Définition 10.16 *La fonction $f(x)$ tend vers la limite b quand $x \to \infty$ si pour chaque nombre positif ε aussi petit qu'il soit on peut indiquer un nombre positif N tel que pour toutes les valeurs de x vérifiant l'inégalité $|x|>N$, l'inégalité $|f(x)-b|<\varepsilon$*

Exemple :

Montrons que (nous supposons le résultat connu pour l'instant) :

$$\lim_{x\to\infty}\left(\frac{x+1}{x}\right) = 1 \tag{10.54}$$

Il faut démontrer que, quel que soit ε, l'inégalité $\left|1+\frac{1}{x}-1\right|<\varepsilon$ sera satisfaite dès que $|x|>N$, où N est défini par le choix de ε. L'inégalité précédente est évidemment équivalente à $\left|\frac{1}{x}\right|<\varepsilon$, qui est satisfait si nous avons x :

$$|x| > \left(\frac{1}{\varepsilon}\right) = N \tag{10.55}$$

Nous admettons que l'exemple et la méthode sont discutables mais nous verrons plus tard les outils mathématiques adéquats pour arriver rigoureusement, sans magouilles et hypothèses de départ, au résultat obtenu précédemment.

La signification des symboles $x \to +\infty$ et $x \to -\infty$, rend évidente celle des expressions :

$$f(x) \text{ tend vers } b \text{ quand } x \to +\infty \tag{10.56}$$

et :

$$f(x) \text{ tend vers } b \text{ quand } x \to -\infty \tag{10.57}$$

que nous notons symboliquement par :

$$\lim_{x\to+\infty} f(x) = b \text{ et } \lim_{x\to-\infty} f(x) = b \tag{10.58}$$

Nous avons étudié le cas où la fonction $f(x)$ tend vers une certaine limite b quand $x \to a$ ou $x \to \infty$. Considérons maintenant le cas où la fonction $y = f(x)$ tend vers l'infini quand la variable x varie d'une certaine manière.

Définition 10.17 *La fonction $f(x)$ tend vers l'infini quand $x \to a$, autrement dit $f(x)$ est infiniment grande quand $x \to a$, si pour chaque nombre positif M, aussi grand qu'il soit, nous pouvons trouver un nombre $\delta > 0$ tel que pour toutes les valeurs de x différentes de a et vérifiant la condition $|x - a| < \delta$, l'inégalité $|f(x)| > M$ est satisfaite.*

Si $f(x)$ tend vers l'infini quand $x \to a$, on écrit :

$$\lim_{x\to a} f(x) = \infty \tag{10.59}$$

Si $f(x)$ tend vers l'infini quand $x \to a$, en ne prenant que des valeur positives ou que des valeurs négatives, on écrit respectivement :

$$\lim_{x\to a} f(x) = +\infty \text{ et } \lim_{x\to a} f(x) = -\infty \tag{10.60}$$

Si la fonction $f(x)$ tend vers l'infini quand $x \to \infty$ on écrit :

$$\lim_{x\to\infty} f(x) = \infty \tag{10.61}$$

et en particulier, on peut avoir :

$$\lim_{x\to+\infty} f(x) = \infty, \quad \lim_{x\to-\infty} f(x) = \infty, \quad \lim_{x\to+\infty} f(x) = -\infty, \quad lim_{x\to-\infty} f(x) = -\infty \tag{10.62}$$

Il peut arriver que la fonction $f(x)$ ne tende ni vers une limite finie, ni vers l'infini quand $x \to a$ (par exemple $f(x) = \sin(x)$), la fonction est alors bornée (cf. chapitre de Théorie des Ensembles).

Maintenant que nous avons grosso modo eu un aperçu du concept de limite, nous allons donner une définition extrêmement importante qui est un des piliers de beaucoup de domaines de la mathématique et de la physique.

Définition 10.18 *Soit f une fonction définie sur $I \subseteq \mathbb{R}$. Soit $x_0 \in I$, nous disons que nous avons une "fonction continue" en x_0 si et seulement si :*

$$\lim_{x\to\infty} f(x) = f(x_0) \tag{10.63}$$

c'est-à-dire si (il faut pouvoir arriver à y lire le fait qu'on s'approche de manière infiniment petite d'une limite ce qui permet d'assurer le continuité) que $\forall\varepsilon > 0, \exists\alpha > 0$ tel que $x \in I$ alors :

$$|x - x_0| < \alpha \Rightarrow |f(x) - f(x_0)| \leq \varepsilon \tag{10.64}$$

REMARQUE f est "continue à droite" (resp. à gauche) si nous rajoutons la condition $x \geq x_0$ (resp. $x \leq x_0$).

Nous avons les corollaires triviaux suivants :

Corollaire 10.1 *f est continue en x_0 si et seulement si f est continue à droite et à gauche en x_0*

Corollaire 10.2 *f est continue sur I si et seulement si f est continue en tout point de I.*

Asymptotes

Le terme d'asymptote est utilisé en mathématiques pour préciser des propriétés éventuelles d'une branche infinie de courbe à accroissement tendant vers l'infinitésimal.

L'étude du comportement asymptotique est particulièrement développé dans les études de fonctions. Dans le domaine scientifique, il arrive fréquemment d'étudier des fonctions dépendant du temps (évolution de populations, réaction chimique ou nucléaire, graphique de température, oscillation d'un amortisseur). Un des objectifs du chercheur est alors de connaître l'état à la fin de l'expérience, c'est à dire lorsqu'un grand intervalle de temps s'est écoulé. L'objectif n'est alors pas de connaître les variations intermédiaires mais de déterminer le comportement stable, à l'infini du phénomène mesuré. Le chercheur étudie donc le comportement asymptotique de sa fonction avec les outils que les mathématiques lui offrent.

Définition 10.19 *Lorsque la limite d'une fonction $f(x)$ tends vers une constante c^{te} quand $x \to \pm\infty$, alors la représentation graphique de cette fonction nous amène à dessiner une droite horizontale que nous appelons "asymptote horizontale" et dont l'équation est :*

$$A(x) = c^{te} \tag{10.65}$$

Définition 10.20 *Lorsque la limite d'une fonction $f(x)$ tends vers /pm∞ quand $x \to a$, alors la représentation graphique de cette fonction nous amène à dessiner une droite verticale que nous appelons "asymptote verticale" et dont l'équation est :*

$$A(y) = a_{\pm} \tag{10.66}$$

Exemple : La courbe représentative de la fonction $f(x) = 1/(x-1)$ (figure 10.11 admet la droite d'équation $x = 1$ comme asymptote verticale et $y = 0$ comme asymptote horizontale.

Définition 10.21 *La droite d'équation est une "asymptote oblique" à la courbe de la fonction $f(x)$ si :*

$$\lim_{x \to \pm\infty} \frac{f(x)}{x} \tag{10.67}$$

les valeurs de a et de b peuvent se retrouver facilement à l'aide des relations suivantes :

$$\begin{aligned} a &= \lim_{x \to \pm\infty} \frac{f(x)}{x} \\ b &= \lim_{x \to \pm\infty} f(x) - ax \end{aligned} \tag{10.68}$$

Remarque Attention une courbe peut admettre deux asymptotes obliques distinctes en $+\infty$ et en $-\infty$.

Pour rechercher une asymptote oblique éventuelle, il faut déjà être sûr que la fonction f admet une limite infinie en $+\infty$ ou en $-\infty$ ensuite nous cherchons la limite en $+\infty$ ou en de f(x)/x .

Trois cas sont à considérer :

C1. La courbe représentative de f a pour direction asymptotique la droite d'équation $y = ax$:

$$\lim_{x \to \pm\infty} \frac{f(x)}{x} = a \tag{10.69}$$

Exemple : La fonction $f(x) = x + \frac{1}{x}$ possède une asymptote d'équation $y = x$ (cfr figure 10.12).

C2. La courbe représentative de f admet une branche infinie (cette branche infinie n'admet pas d'asymptote) et l'axe des abscisses en est la direction asymptotique (fonction racine carrée par exemple)

$$\lim_{x \to \pm\infty} \frac{f(x)}{x} = 0 \tag{10.70}$$

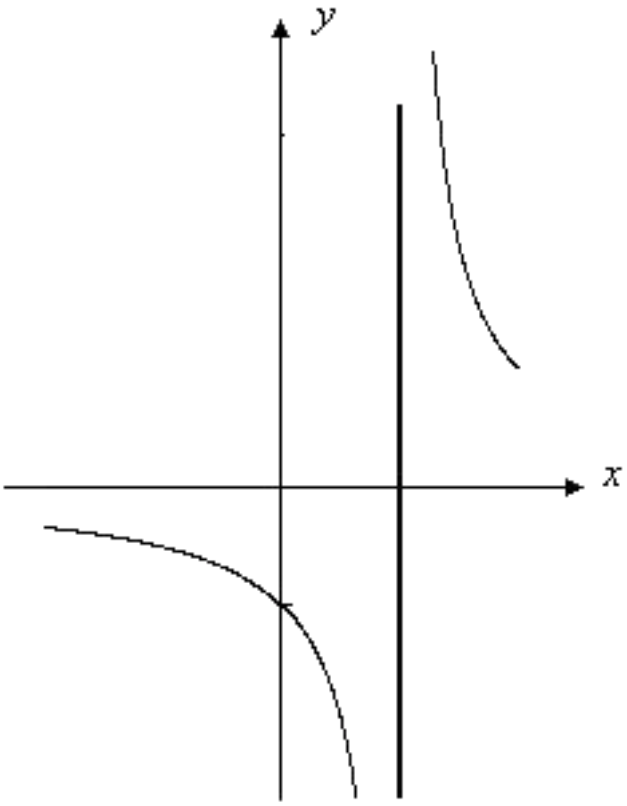


FIGURE 10.11 – Asymptotes de la fonction $f(x) = \frac{1}{1-x}$

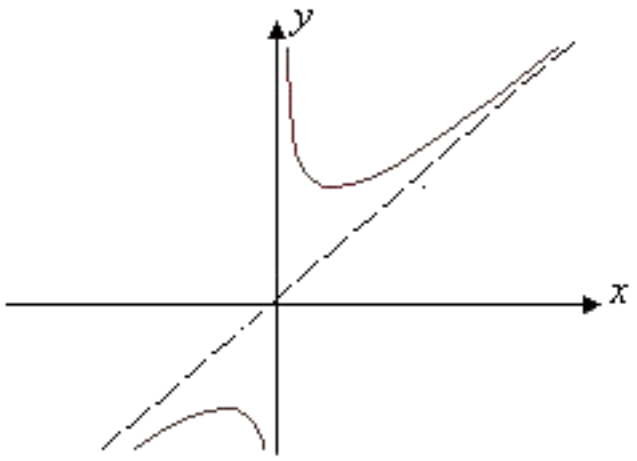


FIGURE 10.12 – Asymptote d'équation $y = x$

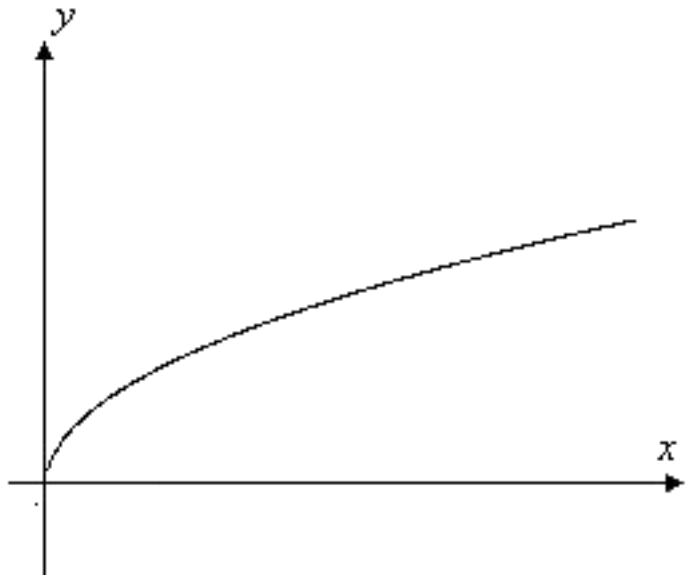


FIGURE 10.13 – Branche parabolique de direction Ox

Exemple : La fonction $f(x) = \sqrt{x}$ ou $\ln(x)$ ont une limite $f(x)/x$ nulle et possèdent donc toutes deux une "branche parabolique" de direction Ox (cfr figure 10.13).

C3. La courbe représentative de f admet une branche infinie (cette branche infinie n'admet pas d'asymptote) et l'axe des ordonnées en est la direction asymptotique. (nous parlons aussi de "branche parabolique" voir de "fonction carrée")

$$\lim_{x \to \pm\infty} \frac{f(x)}{x} = \pm\infty \tag{10.71}$$

Exemple : La fonction $f(x) = x^2$ à une limite $f(x)/x$ infinie et possède donc une "branche parabolique" de direction Oy (cfr figure 10.14).

10.3 Logarithmes

NOUS AVONS LONGUEMENT HÉSITÉ à mettre la définition des logarithmes dans le chapitre traitant du calcul algébrique. Après un moment de réflexion, nous avons décidé qu'il valait mieux la mettre ici car pour bien la comprendre, il faut avoir connaissance des concepts de limite, domaine de définition et fonction exponentielle. Nous espérons que notre choix vous conviendra au mieux.

Soit la fonction exponentielle (bijective) de base quelconque a, où $a \in \mathbb{R}_+^*/1$ notée :

$$f(x) = a^x \tag{10.72}$$

pour laquelle il correspond à chaque nombre réel x, exactement un nombre positif a^x (l'ensemble image de la fonction est dans $\mathbb{R}$) tel que les règles de calcul des puissances soient applicables (cf. chapitre de Calcul Algébrique).

Nous savons que pour une telle fonction, que si $a > 1$, alors $f(x)$ est croissante et positive dans $\mathbb{R}$, et si $0 < a < 1$, alors $f(x)$ est décroissante et positive dans $\mathbb{R}$.

REMARQUE Si $a > 1$, lorsque x décroît vers des valeurs négatives, le graphique de $f(x)$ tend vers l'axe des x. Ainsi, l'axe des x est une asymptote horizontale. Lorsque x croît par valeurs positives, le graphique monte rapidement. Ce type de variation est caractéristique de la "loi de croissance exponentielle" et $f(x)$ est quelque fois appelée "fonction de croissance". Si , lorsque x croît, le graphique tend asymptotiquement vers l'axe des x. Ce type de variation est connue sous le nom de "décroissance exponentielle".

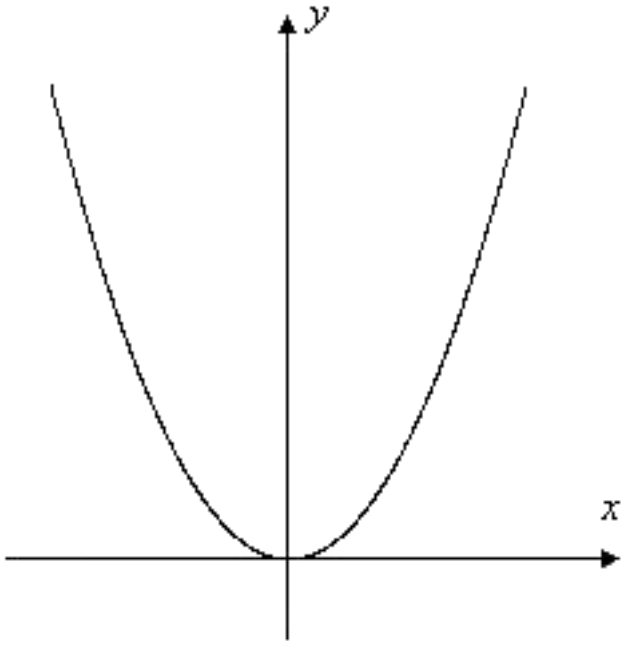


FIGURE 10.14 – Branche parabolique de direction Ox

REMARQUE En étudiant a^x, nous excluons le cas $a \leq 0$ et $a = 1$. Notons que si $a < 0$, alors a^x n'est pas un nombre réel pour de nombreuses valeurs de x (nous rappelons que l'ensemble image est contraint à $\mathbb{R}$). Si $a = 0$, $a^0 = 0^0$ n'est pas défini. Enfin, si $a = 1$, alors $a^x = 1$ pour tout x et le graphique de $f(x)$ est une droite horizontale.

Puisque la fonction exponentielle $f(x)$ est bijective alors il existe une fonction réciproque $f^{-1}(x)$ et appelée "fonction logarithme" de base a notée :

$$f^{-1}(y) = log_a(y) \tag{10.73}$$

Et donc :

$$x = log_a(y) \tag{10.74}$$

si et seulement si $y = a^x$.
En considérant $log_a(x)$ comme un exposant, nous avons les propriétés suivantes :

Propriétés	Jusitification
$log_a 1 = 0$	$a^0 = 1$
$log_a a = a$	$a^1 = a$
$log_a a^x$	$a^x = a^x$
$a^{log_a(x)} = x$	$a^{log_a(x)} = a^y = x$

REMARQUE Le mot "logarithme" signifie "nombre du logos", "logos" signifiant "raison" ou "rapport".

REMARQUE Les fonctions logarithme et exponentielle sont définies par leur base (le nombre a). Lorsqu'on utilise un base de puissance de 10 (10, 100, 1000,...) on parle alors de "système vulgaire" car ils ont pour logarithme des nombres entiers successifs.

REMARQUE La partie entière du logarithme s'appelle la "caractéristique".

Il existe deux types de logarithmes que nous retrouvons presque exclusivement en mathématique et en physique : le logarithme en base dix et le logarithme en base e (ce dernier étant fréquemment

appelé "logarithme naturel" ou plus exactement pour des raisons historiques justifiées "logarithme népérien").

D'abord celui en base 10 :

$$f(x) = \log_{10}(x) \tag{10.75}$$

abusivement noté :

$$f(x) = \log(x) \tag{10.76}$$

et celui en base (eulérienne) e :

$$f(x) = \log_e(x) \tag{10.77}$$

historiquement noté :

$$f(x) = \ln(x) \tag{10.78}$$

le "n" signifiant "népérien".

Remarque Historiquement, c'est à John Napier (1550-1617) dont le nom latinisé est "Neper" que l'on doit l'étude des logarithmes et le nom aux "logarithmes néperien".

En français pour la fonction logarithmique en base 10 il faut pour calculer $f(x) = log(x)$, se poser la question suivante : à quelle puissance $n \in \mathbb{R}$ doit on élever 10 pour obtenir x ?

Formellement, cela consiste à résoudre l'équation : $10^n = x$ ou sinon $10 = \sqrt[n]{x}$ avec x étant connu et donc $log(x) = n$.

Pour la fonction logarithmique en base eulérienne e (ou dite appelée "base nepérienne") il faut pour calculer $f(x) = ln(x)$ se poser aussi la question suivante : à quelle puissance $n \in \mathbb{R}$ devons nous élever le nombre e pour obtenir x ?

Formellement, cela consiste à résoudre l'équation : $e^n = x$ ou sinon $e = \sqrt[n]{x}$ avec x étant connu et donc $ln(x) = n$. Techniquement, nous disons alors que le fonction exponentielle (voir plus bas les détails) :

$$f(x) = e^x \tag{10.79}$$

est la bijection réciproque de la fonction $ln(x)$ comme montré à la figure 10.15.

Mais quel est donc ce nombre "eulérien" appelé également "nombre d'Euler" ? Pourquoi le retrouve-t-on si souvent en physique et en mathématiques ? D'abord déterminons l'origine de sa valeur :

Pour cela, il nous faut déterminer la limite de :

$$\sqrt[n]{1+n} = (1+n)^{\frac{1}{n}} = \left(1 + \frac{1}{\alpha}\right)^{\alpha} \tag{10.80}$$

avec $\alpha \in \mathbb{N}$ et quand $\alpha \to \infty$

L'intérêt que nous avons à poser le problème ainsi c'est que si nous faisons tendre $\alpha \to \infty$ la fonction écrite précédemment tend vers e et cette fonction a pour propriété particulière de pouvoir se calculer plus ou moins facilement pour des raisons historiques à l'aide du binôme de Newton.

Donc d'après le développement du binôme de Newton (cf. chapitre de Calcul Algébrique) nous pouvons écrire :

$$\begin{aligned}\left(1 + \frac{1}{\alpha}\right)^{\alpha} &= 1 + \frac{\alpha}{1!}\frac{1}{\alpha} + \frac{\alpha(\alpha-1)}{2!}\frac{1}{\alpha^2} + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}\frac{1}{\alpha^3} \\ &\quad + \cdots + \frac{\alpha(\alpha-1)\ldots(\alpha-(\alpha-1))}{\alpha!}\frac{1}{\alpha^{\alpha}}\end{aligned} \tag{10.81}$$

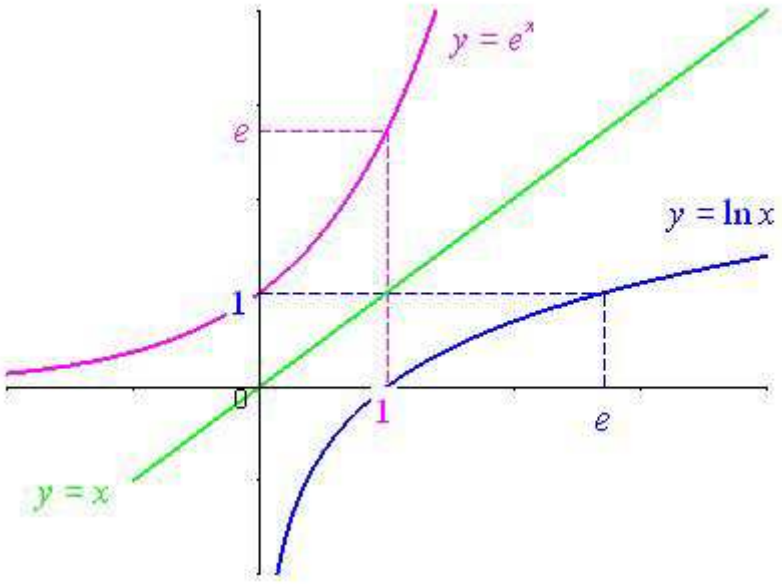


FIGURE 10.15 – Logarithme nepérien et sa bijection exponentielle

Ce développement, est similaire au développement de Taylor de certaines fonctions pour des cas particuliers de valeurs de développement (d'où la raison pour laquelle nous retrouvons ce nombre eulérien dans beaucoup d'endroits que nous découvrirons au fur et à mesure).

En effectuant certaines transformations algébriques évidentes, nous trouvons :

$$\begin{aligned}\left(1+\frac{1}{\alpha}\right)^{\alpha} &= 1+1+\frac{1}{2!}\left(1-\frac{1}{\alpha}\right)+\frac{1}{3!}\left(1-\frac{1}{\alpha}\right)\left(1-\frac{2}{\alpha}\right)\\ &\quad+\cdots+\frac{1}{\alpha!}\left(1-\frac{1}{\alpha}\right)\left(1-\frac{2}{\alpha}\right)\ldots\left(1-\frac{\alpha-1}{\alpha}\right)\end{aligned} \tag{10.82}$$

Nous voyons de cette dernière égalité que la fonction $\left(1+\frac{1}{\alpha}\right)^{\alpha}$ est croissante quand α croît. En effet, quand nous passons de la valeur α à la valeur $\alpha+1$ chaque terme de cette somme augmente :

$$\frac{1}{2!}\left(1-\frac{1}{\alpha}\right)<\frac{1}{2!}\left(1-\frac{1}{\alpha+1}\right)\ \text{etc.} \tag{10.83}$$

Montrons que la grandeur variable $\left(1+\frac{1}{\alpha}\right)^{\alpha}$ est bornée. En remarquant que :

$$\left(1-\frac{1}{\alpha}\right)<1,\ \left(1-\frac{1}{\alpha}\right)\left(1-\frac{2}{\alpha}\right)<1\ \text{etc.} \tag{10.84}$$

Nous obtenons donc :

$$\left(1-\frac{1}{\alpha}\right)^{\frac{1}{\alpha}}<1+\frac{1}{2!}+\frac{1}{3!}+\cdots+\frac{1}{\alpha!} \tag{10.85}$$

D'autre part :

$$\frac{1}{3!}<\frac{2}{2!},\ \frac{1}{4!}<\frac{2}{3!},\ \frac{1}{\alpha!}<\frac{2}{2^{\alpha-1}} \tag{10.86}$$

Nous pouvons donc écrire l'inégalité :

$$\left(1-\frac{1}{\alpha}\right)^{\frac{1}{\alpha}} < 1+\underline{1+\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^{\alpha-1}}} \tag{10.87}$$

Les termes soulignés constituent une progression géométrique de raison $q=\frac{1}{2}$ et dont le premier terme est 1. Par suite :

$$\begin{aligned}\left(1+\frac{1}{\alpha}\right)^{\alpha} < 1+\left[1+\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^{\alpha-1}}\right] &= 1+\frac{1-(1/2)^{\alpha}}{1-(1/2)} \\ &= 1+\left[2-\left(\frac{1}{2}\right)^{\alpha-1}\right] < 3\end{aligned} \tag{10.88}$$

Par conséquent nous avons :

$$2 \leq \left(1+\frac{1}{\alpha}\right)^{\alpha} < 3 \tag{10.89}$$

Nous avons donc prouvé que la fonction $\left(1+\frac{1}{\alpha}\right)^{\alpha}$ est bornée.
La limite :

$$\boxed{e = lim_{\alpha\to\infty}\left(1+\frac{1}{\alpha}\right)^{\alpha}} \tag{10.90}$$

tend donc vers le nombre e dont la valeur est :

$$e = 2.7182818284\ldots \tag{10.91}$$

REMARQUE Comme nous l'avons démontré dans le chapitre traitant des Nombres ce nombre est irrationnel.

Nous pouvons alors définir la "fonction exponentielle naturelle" (réciproque de la fonction logarithme néperienne) par :

$$f(x) = e^x \tag{10.92}$$

ou également parfois notée :

$$f(x) = \exp(x) \tag{10.93}$$

Le nombre e et la fonction qui permet de le déterminer sont très utiles. Nous les retrouvons dans tous les domaines de la mathématique et de la physique.
Les logarithmes ont plusieurs propriétés. Les voici (nous nous référons à une base X donnée) :

$$\log_X(X^m \cdot X^n) = \log_X(X^{m+n}) = \log_X(X^m) + \log_X(X^n) = m+n \tag{10.94}$$

Si nous posons $X^m = a$ et $X^n = b$ nous avons donc :

$$\log_X(a \cdot b) = \log_X(a) + \log_X(b) \tag{10.95}$$

Si nous avons le cas particulier $a = b$ alors :

$$\log_X(a \cdot a) = \log_X(a) + \log_X(a) = 2 \cdot \log_X(a) \tag{10.96}$$

Cherchons à exprimer $\log_X\left(\frac{a}{b}\right)$ sous une forme différente. Posons $t = \frac{a}{b} \Leftrightarrow a = t \cdot b$ ce qui nous amène au développement :

$$\begin{aligned} \log_X(a) &= \log_X(t \cdot b) = \log_X(t) + \log_X(b) \\ \Rightarrow \log_X(t) &= \log_X(a) - \log_X(b) = \log_X\left(\frac{a}{b}\right) \end{aligned} \tag{10.97}$$

Cherchons à exprimer maintenant $\log_X\left(x^{\frac{1}{n}}\right)$ avec $n \in \mathbb{N}^*$ sous un forme différente. Posons :

$$t = x^{\frac{1}{n}} \Leftrightarrow x = t^n \tag{10.98}$$

ce qui nous amène à :

$$\begin{aligned} \log_X(x) &= \log_X(t^n) = n \cdot \log_X(t) \\ \Rightarrow \log_X(t) &= \log_X\left(x^{\frac{1}{n}}\right) = \frac{1}{n}\log_X(x) \end{aligned} \tag{10.99}$$

Il y a une relation assez utilisée en physique relativement aux changements de bases logarithmiques. La première relation est triviale et découle des propriétés algébriques des logarithmes :

$$a^x = e^{x\ln(a)} = 10^{x\log_{10}(a)} = b^{x\log_b(a)} \tag{10.100}$$

La seconde relation :

$$\log_a(u) = \frac{\ln(u)}{\ln(a)} = \frac{\log_{10}(u)}{\log_{10}(a)} = \frac{\log_b(u)}{\log_b(a)} \tag{10.101}$$

est un peu moins triviale et nécessite peut-être une démonstration.

Preuve Nous avons d'abord les équations équivalentes (de la première relation ci-dessus) :

$$w = \log_a(u) \text{ et } a^w = u \tag{10.102}$$

et nous procédons comme suit :

$$\begin{aligned} a^w &= u \\ \log_b(a^w) &= \log_b(u) \\ w \cdot \log_b(a) &= \log_b(u) \end{aligned} \tag{10.103}$$

Ce qui nous amène finalement à :

$$w = \frac{\log_b(u)}{\log_b(a)} = \log_a(u) \tag{10.104}$$

•

10.4 Produit scalaire fonctionnel

LE PRODUIT SCALAIRE FONCTIONNEL (analogie très forte avec le produit scalaire vectoriel vu dans le chapitre de calcul vectoriel) peut paraître inutile lorsqu'il est étudié pour la première fois hors d'un contexte appliqué mais il connaît au fait de nombreuses application pratiques. Nous en ferons par exemple directement usage dans le chapitre de physique quantique ondulatoire et de chimie quantique ou encore dans le cadre plus important encore des polynômes trigonométriques via les séries et transformées de Fourier (cf. chapitre sur les Suites Et Séries) que nous retrouvons partout dans la physique contemporaine.

Cependant, si le lecteur n'a pas encore parcouru le chapitre de calcul vectoriel et la partie y traitant du produit scalaire vectoriel, nous ne serions que trop recommander sa lecteur sans quoi de ce qui va suivre risque d'être un peu incompréhensible.

Nous nous plaçons dans l'espace $\mathcal{C}([a,b],\mathbb{R})$ des fonctions continues de l'intervalle $[a,b]$ dans muni du produit scalaire définit par (nous retrouvons la notation spécifique du produit scalaire à sa version fonctionnelle comme nous en avions fait mention lors de notre définition du produit scalaire vectoriel) :

$$\boxed{\langle f|g\rangle = \int_{-1}^{1} f(x)\cdot g(x)dx} \tag{10.105}$$

Une famille de polynômes orthogonale, comme nous pouvons en faire l'analogie avec le produit scalaire vu dans le chapitre de calcul vectoriel, est donc une famille de polynômes $(p_0,\ldots,p_n,\ldots)$ tels que :

$$\langle p_j|p_k\rangle \text{ si } j\neq k \tag{10.106}$$

Nous rappelons qu'une famille orthogonale est libre (cf. chapitre de Calcul Vectoriel).

Le développement suivant va nous rappeler le procédé de Gram-Schmidt (cf. chapitre de Calcul Tensoriel) pour construire une famille orthogonale :

Soit $(p_0,\ldots,p_n,\ldots)$ une famille de polynômes linéairement indépendants définis sur $[a,b]$ et V l'espace vectoriel engendré par cette famille. La famille $(y_0,\ldots,y_n,\ldots)$ définie par récurrence de la manière suivante :

$$y_n = p_n - \sum_{j=0}^{n-1}\langle p_n|y_j\rangle\frac{y_j}{\|y_j\|^2} \text{ pour } n>1 \tag{10.107}$$

et $y_0 = p_0$ est orthogonale et engendre V.

Preuve Montrons par récurrence sur n que $(y_0,\ldots,y_n,\ldots)$ est une famille orthogonale qui engendre le même espace que $(p_0,\ldots,p_n,\ldots)$. L'assertion est vérifiée pour $n=0$. Supposons l'assertion vérifiée pour $n\geq 0$, pour $0\leq k\leq n$ nous avons :

$$\begin{aligned}
\langle y_{n+1}|y_k\rangle &= \left\langle p_{n+1} - \sum_{j=0}^{n}\langle p_{n+1}|y_j\rangle\cdot\frac{y_j}{\|y_j\|^2}\middle| y_k\right\rangle \\
&= \langle p_{n+1}|y_k\rangle - \sum_{j=0}^{n}\langle p_{n+1}|y_j\rangle\cdot\langle y_j|y_k\rangle\cdot\frac{1}{\|y_j\|^2} \\
&= \langle p_{n+1}|y_k\rangle - \langle p_{n+1}|y_k\rangle\cdot\langle y_k|y_k\rangle\cdot\frac{1}{\|y_j\|^2} \\
&= \langle p_{n+1}|y_k\rangle - \langle p_{n+1}|y_k\rangle \\
&= 0
\end{aligned} \tag{10.108}$$

$(y_0,\ldots,y_n,\ldots)$ est donc orthogonale. Pour finir, l'égalité :

$$y_{n+1} = p_{n+1} - \sum_{j=0}^{n}\langle p_{n+1}|y_j\rangle\cdot\frac{y_j}{\|y_j\|^2} \tag{10.109}$$

montre que $(y_0,\ldots,y_{n+1},\ldots)$ et $(p_0,\ldots,p_{n+1},\ldots)$ engendrent le même espace. $(y_0,\ldots,y_{n+1},\ldots)$ est donc bien une famille orthogonale qui engendre V. •

Exemple :

Considérons l'exemple très important en physique moderne qui est l'ensemble $P_{2\pi}$ des fonctions continues 2π-périodiques qui forme un espace vectoriel (cf. chapitre de Calcul Vectoriel).

Nous définissons donc le produit scalaire de deux fonctions de cet ensemble par :

$$\forall f, g \in P_{2\pi} \quad \langle f|g\rangle = \int_{-\pi}^{\pi} f(x)g(x)dx = \int_{0}^{-\pi} f(x)g(x)dx \tag{10.110}$$

Le but de cette étude est de construire une base de $P_{2\pi}$ sur laquelle nous pouvons décomposer tout fonction 2π-périodique.

L'idée la plus simple est alors de se servir des fonctions trigonométriques sinus et cosinus :

$$\begin{aligned} &\cos(x), \cos(2x), \cos(3x), \ldots, \cos(nx) \\ &\sin(x), \sin(2x), \sin(3x), \ldots, \sin(nx) \end{aligned} \tag{10.111}$$

Les relations ci-dessous montrent que les bases choisies ci-dessus sont orthogonaux et forment donc une famille libre, de plus c'est une famille génératrice de l'espace vectoriel $P_{2\pi}$ car comme nous le démontrerons le démontrerons lors de notre étude des séries de Fourier (cf. chapitre sur les Suites Et Séries), nous avons les valeurs suivantes :

$$\begin{aligned} &\int_{0}^{-\pi} \cos(kx)\cos(mx)dx = \delta_{km}\pi \\ &\int_{0}^{-\pi} \sin(kx)\sin(mx)dx = \delta_{km}\pi \\ &\int_{0}^{-\pi} \cos(kx)\sin(mx)dx = 0 \end{aligned} \tag{10.112}$$

où δ_{km} est le symbole de Kronecker (cf. chapitre de Calcul Tensoriel)

11

Topologie

La topologie (du grec : discours du lieu) est un domaine extrêmement vaste des mathématiques dont il est difficile de définir avec exactitude l'objet dont elle fait l'étude. Ce que nous pouvons dire dans un premier temps, c'est que la topologie est très intimement liée à la théorie des ensembles, aux probabilités, à l'analyse fonctionnelle, aux suites et séries, au calcul intégral et différentiel, au calcul vectoriel, à la géométrie et encore beaucoup d'autres domaines.

L'origine de la topologie semble provenir des problèmes qu'ont posés les progrès de l'analyse fonctionnelle dans l'étude des fonctions continues, de leur dérivabilité, de leurs limites en un point (fini ou non), de l'existence d'extremums, etc... dans des espaces de dimensions supérieures (au fait, implicitement la topologie a pour objectif de créer des outils qui permettent facilement d'étudier les propriétés des fonctions dans toutes les dimensions). Tout ces concepts, demandaient pour le mathématicien une définition rigoureuse de l'idée intuitive de proximité, tout particulièrement lors d'opérations sur ces fonctions.

Nous allons essayer dégager les structures qui permettent de parler de limite et de continuité. L'exemple fondamental est le cas de et de . La théorie générale englobe bien sûr cet exemple qu'il faut garder en tête mais conduit parfois des situations moins intuitives.

Voici cependant un essai de définition de la topologie que nous pourrons toujours compléter :

Définition 11.1 *La topologie s'intéresse aux propriétés morphologiques des ensembles, par l'intermédiaire de l'étude des relations entre leurs différents éléments (la distance entre deux objets d'un ensemble est un exemple), indépendamment de leur forme ou de leur taille, c'est-à-dire indépendamment des propriétés quantitatives des ensembles. Intuitivement, (mais ce n'est qu'un cas particulier), deux objets sont topologiquement équivalents s'il est possible de passer de l'un à l'autre par déformation élastique, sans déchirure ni recollement.*

Remarque Nous pensons que la topologie devrait être introduite dans le cursus scolaire dès le moment où l'étudiant a de solides connaissances en théorie des ensembles, calcul vectoriel, algèbre linéaire, calcul tensoriel et géométrie différentielle. Ce n'est effectivement qu'à ce moment que le niveau de l'étudiant permet à celui-ci de comprendre toute la subtilité, l'utilité et les répercussions majeures des développements et définitions qui vont suivre.

Remarque Souvent les personnes qui commencent à étudier la topologie sont frappées par le fait qu'il y ait autant de définitions dans un seul domaine (ce qui gâche, il est vrai, un peu la beauté mathématique de la topologie). Ils prennent alors pleinement conscience de la citation d'Erdös : "les mathématiciens sont des gens qui font des définitions avec du café...".

Remarque Certains concepts définis ici deviendront aussi plus clair dès le moment où le lecteur aura étudié le chapitre de Théorie De La Mesure (...)

11.1 Espace topologique

Définition 11.2 *Soit un ensemble non vide X. Une "topologie" sur X est une famille F de parties de X appelées "ouverts" tel que les axiomes suivants soient vérifiés :*

A1. L'ensemble vide et X sont des ouverts : $\oslash \in F$ et $X \in F$

A2. Toute intersection finie d'ouverts est un ouvert : $V_1, V_2, \ldots, V_n \in F$ implique $\bigcap_{i=0}^{n} V_i \in F$

A3. Toute réunions d'ouverts est un ouvert : $F' \subset F$ implique $\bigcup_{V_i \in F'} V \in F$

REMARQUE Les mathématiciens notent fréquemment par la lettre O la famille des ouverts et F la famille des fermés. Convention que nous ne suivrons donc pas ici.

REMARQUE Les "fermés" d'une topologie sont les complémentaires des ouverts. Par conséquent, la famille des fermés contient entre autres X et l'ensemble vide...

Définition 11.3 *Etant donné une telle famille F, nous appelons X, ou plus précisément le couple (X, F) un "espace topologique" et F une "topologie".*

Le couple (X, F) forme un "espace de Hausdorff" ou "espace séparé" si de plus la propriété suivante dit "axiome d'Hausdorff" est vérifiée :

A4. $\forall p, q \in X$ avec $p \neq q$, $\exists V_p, V_q$ tels que $p \in V_p$, $q \in V_q$ et $V_p \cap V_q = \oslash$

REMARQUE Un exemple bien connu d'espace topologique est $\mathbb{R}$ muni de l'ensemble F engendré par les intervalles ouverts (par la loi d'union), c'est-à-dire les intervalles $]a, b[$.

REMARQUE Nous verrons une application très concrète des espaces de Hausdorff lors de notre études des fractales en informatique théorique.

Définition 11.4 *Si nous notons (X, F) un espace topologique, F désignant les ouverts de X, une "base", au sens topologique, de (X, F) est une partie B de F telle que tout ouvert de F soit réunion d'ouverts de B (c'est la même idée que les espace vectoriels au fait mais appliqué à des ensembles... rien de bien méchant).*

11.1.1 Espace métrique et Distance

Définition 11.5 *Un "espace métrique" noté (X, d) ou encore X_d est par définition un ensemble X muni d'une application $d : X \times X \to \mathbb{R}$, appelée "distance" ou "métrique", qui satisfait les axiomes suivants :*

A1. $\forall x \neq y$ $d(x, y) > 0$ (positivité)
A2. $\forall x, y \in X d(x, y) = 0 \Leftrightarrow x = y$ (axiome de séparation)
A3. $\forall x, y, z \in X d(x, z) \leq d(x, y) + d(y, z)$ (inégalité triangulaire)
A4. $\forall x, y \in X d(x, y) = d(y, x)$(axiome de symétrie)

REMARQUE Certains lecteurs verront de suite que certaines de ces propriétés ont déjà été vues dans d'autres chapitres du site lors de l'étude des distances entre points fonctionnels et lors de l'étude des normes (inégalité triangulaire démontrée dans le chapitre de Calcul Vectoriel - la symétrie, la nullité, la positivité, la séparation dans le chapitre d'Analyse Fonctionnelle).

REMARQUE Certains auteurs omettent l'axiome A1 ce qui est rigoureusement juste car découle trivialement de A3.

REMARQUE Un espace métrique sera en général noté (X, d) ou bien encore X_d. Nous pouvons également le noter simplement X si la distance d ne peut être confondue.

La "fonction distance" de $\forall x, y \in X$ est donc notée habituellement dans le plus général qui soit en mathématique :

$$d(x, y) \tag{11.1}$$

REMARQUE Si nous n'imposons pas l'axiome A2, nous disons que d est une "semi-distance" sur X.

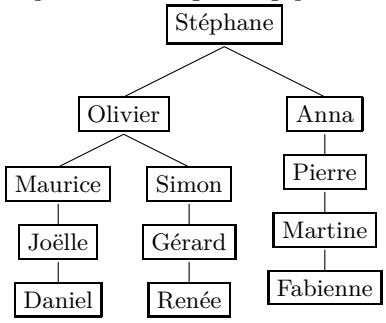
REMARQUE Si nous autorisons une semi-distance d à prendre la valeur $+\infty$, nous préférons dire que d est un "écart"

REMARQUE Si une distance d vérifie la propriété :

$$\forall x, y, z \in X d(x, y) \leq \max\{d(x, y), d(y, z)\} \tag{11.2}$$

propriété plus contraignante que l'inégalité triangulaire dans certains espaces, nous disons que d est "ultra-métrique".

Un exemple de distance ultra-métrique est l'arbre généalogique :



Nous avons les distances suivantes :

$$\begin{aligned} d(\text{Daniel}, \text{Stéphane}) &= 4 \\ d(\text{Renée}, \text{Stéphane}) &= 4 \\ d(\text{Renée}, \text{Fabienne}) &= 8 \\ d(\text{Daniel}, \text{Fabienne}) &= 8 \\ d(\text{Daniel}, \text{Renée}) &= 6 \end{aligned} \tag{11.3}$$

Nous remarquons que les distances ne s'ajoutent pas mais que nous avons par contre :

$$d(x,z) \leq \max\{d(x,y), d(y,z)\} \tag{11.4}$$

Ainsi :

$$\begin{aligned} d(\text{Daniel}, \text{Fabienne}) &\leq \max\{d(\text{Daniel}, \text{Renée}), d(\text{Renée}, \text{Fabienne})\} \\ 8 &\leq \max\{6, 8\} \end{aligned} \tag{11.5}$$

Remarque Soit (E,d) un espace métrique et soit $F \neq \oslash$ une partie de l'ensemble E. L'espace métrique (F,δ) où δ désigne la restriction $d_{F\times F}$ de d à $F \times F \subset E \times E$ est appelé "sous-espace métrique" de (E,d) (il convient de vérifier que la distance d est équivalente à la distanceδ). Dans ce cas, nous disons aussi que F est muni de la distance induite par celle de E. Nous notons simplement d la distance induite.

Exemples :

E1. Si nous prenons pour X le plan, ou bien l'espace à trois dimensions de la géométrie euclidienne et une unité de longueur, la "distance" au sens usuel du terme est bien une distance au sens des 5 axiomes précédemment cités. Dans ces espaces, les trois points A, B, C satisfont comme nous l'avons démontré en Calcul Vectoriel et comme il l'est intuitivement que :

$$d(A,B) \leq d(A,C) + d(C,B) \quad d(A,B) \geq |d(A,C) - d(C,B)| \tag{11.6}$$

avec les autres inégalités obtenues par permutation circulaire de A, B, C. Ces inégalités sont bien connues par exemple entre les longueurs des côtés d'un triangle.

E2. Si nous prenons $X = \mathbb{R}^n, n \in \mathbb{N} \geq 1$, et que nous dotons $\mathbb{R}^n$ d'une structure d'espace vectoriel euclidienne (et non pas non-euclidienne) et que nous prenons deux points :

$$x = (x_1, x_2, \ldots, x_n) \quad y = (y_1, y_2, \ldots, y_n) \tag{11.7}$$

dans $\mathbb{R}^n$.

La distance est donnée nous le savons par (nous avons déjà démontré cela en Analyse Fonctionnelle et Calcul Vectoriel) :

$$\begin{aligned} d(x,y) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \\ &= \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \end{aligned} \tag{11.8}$$

qui satisfait aux 5 axiomes de la distance et que nous appelons la "distance euclidienne". Nous pouvons prendre (c'est une propriété intéressante pour la culture générale), que toute relation de la forme :

$$d(x,y) = \sqrt[p]{\sum_{i=1}^{n}(x_i - y_i)^p} \tag{11.9}$$

est aussi une distance dans $\mathbb{R}^n$ (sans démonstration). Les mathématiciens font encore plus fort en généralisant encore plus (la démonstration à peu d'intérêt pour l'instant) cette dernière relation (en prenant en compte la définition même de la distance) sous la forme :

$$\forall p \in [1,+\infty] \subset \mathbb{R} : \delta_p(x,y) = \sqrt[p]{\sum_{i=1}^{n}(d(x_i,y_i))^p} \tag{11.10}$$

qui est appelée "distance hölderienne".

REMARQUE Suite à l'intervention d'un internaute nous précisons qu'en toute rigueur l'inclusion ci-dessus devrait être notée $[1,+\infty] \subset \overline{\mathbb{R}}$ où $\overline{\mathbb{R}}$ est la droite achevée (précision également valable pour l'inégalité de Minkowski ci-dessous).

Au même titre pour l'inégalité triangulaire, donnée alors par (cf. chapitre de Calcul Vectoriel) :

$$\overrightarrow{x},\overrightarrow{y} \in \mathbb{R}^3 : ||\overrightarrow{x}+\overrightarrow{y}|| \leq ||\overrightarrow{x}|| + ||\overrightarrow{y}|| \Leftrightarrow \sqrt[2]{\sum_{i=1}^{3}(x_i+y_i)^2} \leq \sqrt[2]{\sum_{i=1}^{3}x_i^2} + \sqrt[2]{\sum_{i=1}^{3}y_i^2} \tag{11.11}$$

La généralisation, de par la vérification de l'existence de la distance de hölderienne, nous donne la vraie "inégalité de Minkowski" :

$$\forall x,y \in \mathbb{R}, p \in [1,+\infty] \subset \mathbb{R} : \sqrt[p]{\sum_{i=1}^{n}(x_i+y_i)^p} \leq \sqrt[p]{\sum_{i=1}^{n}x_i^p} + \sqrt[p]{\sum_{i=1}^{n}y_i^p} \tag{11.12}$$

Dans le cas particulier avec , nous avons bien évidemment :

$$d(x,y) = |x-y| = \max\{x-y, y-x\} \tag{11.13}$$

qui est la distance usuelle sur $\mathbb{R}$.
E3. Si nous prenons $X = \mathbb{C}$, nous considérerons la distance :

$$d(z,z') = |z-z'| \tag{11.14}$$

Ainsi, si $z = a+ib = (a,b)$ et $z' = a'+ib' = (a',b')$ nous avons le module qui de même manière que la norme dans $\mathbb{R}^2$, forme une distance :

$$d(z,z') = \sqrt{(a-a')^2+(b-b')^2} \tag{11.15}$$

E4. Considérons aussi $E \neq \oslash$ un ensemble arbitraire. Posons :

$$\delta(x,y) = 1 \text{ si } x \neq y \text{ et } \delta(x,y) = 0 \text{ si } x = y \tag{11.16}$$

Il est très facile de vérifier que cette distance vérifie les 5 axiomes et qu'elle est de plus ultramétrique. Cette distance est appelée "distance discrète" et le lecteur remarquera que nous avons par analogie opté pour cette distance le symbole de la fonction Dirac δ (ce n'est pas innocent !!) plutôt que la traditionnel d.

Distances équivalentes

Parfois, deux distances différentes d et δ sur un même ensemble E sont assez ressemblantes pour que les espaces métriques liés $(E,d),(E,\delta)$ possèdent les mêmes propriétés pour certains objets mathématiques définis par d d'une part, par δ. Il existe plusieurs notions de ressemblances dont voici une première (avant les autres qui nécessitent des outils mathématiques que nous n'avons pas encore définis) :

Définition 11.6 *Soient d et δ deux distances sur un même ensemble E, d et δ sont dites "distances équivalentes" s'il existe deux constantes réelles $c>0, C>0$ telles que*

$$\forall x,y\in E : c\cdot d(x,y)\leq \delta(x,y)\leq C\cdot d(x,y) \tag{11.17}$$

soit :

$$c\leq \frac{\delta(x,y)}{d(x,y)}\leq C \tag{11.18}$$

avec $c\leq C$. Nous noterons par ailleurs cette équivalence $d\sim\delta$.

L'intérêt de cette définition est le suivant : si nous avons convergence pour l'une des métriques, alors nous aurons la convergence pour l'autre aussi. Plus clairement :

$$d(x_n,x)\to 0 \Leftrightarrow \delta(x_n,x)\to 0 \tag{11.19}$$

in extenso :

$$x_n \xrightarrow{d} x \Leftrightarrow x_n \xrightarrow{\delta} x \tag{11.20}$$

Fonctions Lipschitziennes

Relativement aux définitions précédentes, nous pouvons maintenant définir quelques propriétés supplémentaires aux fonctions telles que nous les avions énoncées dans le chapitre de Théorie Des Ensembles :

Soient (E,d) et (F,δ)des espaces métriques, et soit $f:E\to F$ une fonction. Nous définissons les propriétés suivantes :

P1. Nous disons que f est une "isométrie" si (c'est plutôt intuitif... !) :

$$\forall x,y\in E : \delta[f(x),f(y)]=d(x,y) \tag{11.21}$$

Si nous prenons la distance usuelle, la fonction k-lipschitzienne s'écrit alors :

$$\forall x,x'\in E : |f(x)-f(x')|\leq k|x-x'| \tag{11.22}$$

ou ce que nous pouvons écrire également :

$$\frac{|f(x)-f(x')|}{|x-x'|}\leq k \tag{11.23}$$

Ou ce qui revient au même : toutes les cordes tracées entre 2 points quelconques du graphe ont un coefficient directeur (dérivée) compris entre $-k$ et k.

Par exemple, la fonction $\sin(x)$ est 1-lipschitzienne (car la dérivée du cosinus est en valeur absolue compris entre 0 et 1).

P2. Nous disons que deux espaces métriques sont "isométriques" s'il existe une isométrie surjective de l'un sur l'autre (ce qui est assez rassurant en géométrie...).

P3. f est dite "L-lipschitzienne" de constante (ou "de rapport") L s'il existe $L \geq 0 \in \mathbb{R}$ tel que :

$$\forall x, y \in E : \delta[f(x), f(y)] = Ld(x, y) \tag{11.24}$$

Si $L = 1$, nous disons que f est "contractante" (ou une "contraction"), et si $L < 1$, nous disons que f est strictement contractante.

P4. Toute fonction f lipchitzienne est uniformément continue (voir plus loin voir plus loin le concept "d'uniforme continue") si elle vérifie :

$$\forall \varepsilon > 0, \exists \lambda > 0, \forall x, y \in E : d(x, y) \leq \lambda \Rightarrow \delta[f(x), f(y)] \leq \varepsilon \tag{11.25}$$

avec $\lambda = \varepsilon / L$ et $L \neq 0$ (la réciproque n'est pas vraie : toute fonction uniformément continue n'est pas nécessairement continue). En d'autre termes, si nous pouvons rapprocher deux points aussi près que nous voulons dans un espace, nous le pouvons aussi dans l'autre (ce qui assure en quelque sorte la dérivation).

REMARQUE Une isométrie est toujours injective car :

$$f(x) = f(y) \Rightarrow \delta[f(x), f(y)] = 0 = d(u, v) \Rightarrow u = v \tag{11.26}$$

mais elle n'est pas en général surjective.

REMARQUE Si (E, d) et(F, δ) sont isométriques, du point de vue de la théorie des espaces métriques ils sont indiscernables, puisque toutes leurs propriétés sont les mêmes, mais leurs éléments peuvent être de nature très différente (suites dans l'un et fonctions dans l'autre par exemple).

11.1.2 Ensembles ouverts et fermés

Définition 11.7 *Considérons un ensemble E muni d'une distance d. Un sous-ensemble U de E est dit "sous-ensemble ouvert" si, pour chaque élément de U, il existe une distance r non nulle pour laquelle tous les éléments de E dont la distance à cet élément est inférieure ou égale à r, appartiennent à U, ce qui ce traduit en langage mathématique :*

$$U \text{ ouvert de } E \Leftrightarrow \forall x \in U, \exists r \in \mathbb{R}^+ / \forall y \in E : d(x, y) \leq r \Rightarrow y \in U \tag{11.27}$$

REMARQUE Le symbole / signifie dans ce contexte "satisfait le propriété"

Cette définition peut sembler complexe mais en fait, sa signification concrète est plus simple qu'il n'y paraît. En fait, selon cette définition, un ensemble ouvert dans un espace topologique n'est rien d'autre qu'un ensemble de points contiguës et sans bords.

L'absence de bord découle de la condition $r \neq 0$. En effet, en raisonnant par l'absurde, si un ensemble ouvert U avait un bord, alors pour chaque point situé sur celui-ci (le bord) il serait toujours possible de trouver un point n'appartenant pas à U aussi proche que l'on veut de lui. Il s'ensuit que la distance r nécessaire devient donc nulle.

Définition 11.8 *Un "sous-ensemble fermé" est un "ouvert avec bord".*

Définition 11.9 *Un "voisinage" d'un point de E est une partie de E contenant un ouvert contenant ce point.*

La définition d'un ensemble ouvert peut être simplifiée en introduisant une notion supplémentaire, celle de "boule ouverte" :

Boules

Soit x un élément de E :

Définition 11.10 *Une "boule ouverte de centre x et de rayon $r > 0$" ou "boule métrique de rayon r centrée en x" est le sous-ensemble de tous les points de E dont la distance à x est inférieure à r, ce que nous écrivons :*

$$ {}_oB_x^r(E) = \{y \in E | d(x,y) < r\} \tag{11.28}$$

Un ensemble ouvert peut également être défini comme un ensemble pour lequel il est possible de définir une boule ouverte en chaque point.

REMARQUE Les ouverts ainsi définis, forment ce que nous appelons une "topologie induite" par la distance d ou aussi "topologie métrique".

REMARQUE Nous appelons une "couverture ouverte" U de E, un ensemble d'ouverts de E dont la réunion est E.

Définition 11.11 *Une "boule fermée" est similaire à une boule ouverte mais diffère dans le sens que nous y incluons les éléments situés à la distance r du centre :*

$$ {}_fB_x^r(E) = \{y \in E | d(x,y) \leq r\} \tag{11.29}$$

REMARQUE Pour $0 < r < r'$ les inclusions ${}_oB_x^r \subset_f B_x^r \subset B(x,r')$ sont des conséquences directes de la définition de boule ouvert et fermée.

Exemple :
La distance usuelle dans $\mathbb{R}$ est donnée par $d(x,y) = |x-y|$. Les boules sont des intervalles. Pour $x \in \mathbb{R}$ et $r \in \mathbb{R}_+^*$, nous avons :

$$ {}_oB_x^r =]x-r, x+r[\text{ et } {}_fB_x^r = [x-r, x+r] \tag{11.30}$$

Définition 11.12 *Une "sphère" est donnée par :*

$$ S_x^r = \{y \in E | d(x,y) = r \tag{11.31}$$

REMARQUE Puisque par définition, $r > 0$, les boules ouvertes et fermées ne sont pas vides car elle contiennent au moins leur centre. Par contre, une sphère peut être vide.

Exemple :
Avec $\mathbb{R}^n, \mathbb{C}^n$ nous avons vu dans les exemples précédents que nous pouvions définir différentes distances. Pour les distinguer, nous les notons :

$$\begin{aligned} d_1(x,y) &= \sum_{i=1}^{n} |y_i - x_i| \\ d_2(x,y) &= \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2} \\ d_\infty &= \max\{y_i - x_i, x_i - y_i\} \end{aligned} \tag{11.32}$$

Alors, dans $\mathbb{R}^2$ les boules fermées de centre O et de rayon unité équivalentes aux trois formulations précédentes, ont la forme suivante (rappel : $0 < r \leq 1$ dans l'exemple fig 11.1).

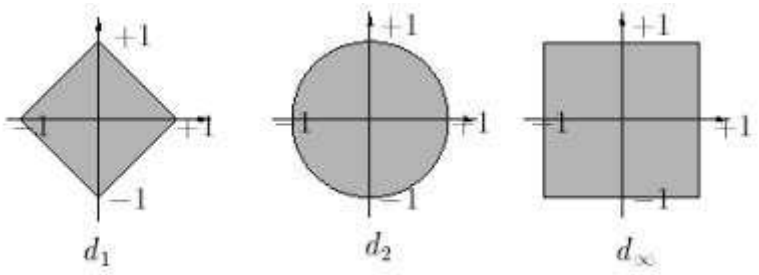


FIGURE 11.1 – Exemple de boules

Parties

Maintenant que nous avons défini les concepts de boules, nous pouvons enfin définir rigoureusement les concepts d'intervalles ouverts et fermées (qui dans un espace à plus d'une dimension sont nommées "parties") dont nous avons fait si souvent usage en Analyse Fonctionnelle et Calcul Intégral Et Différentiel.

Définition 11.13 *Soit (X, d) un espace métrique. Nous disons qu'une partie A de X est "bornée" s'il existe une boule fermée ${}_fB^r_{x_0}(X)$ telle que $A \subset_f B^r_{x_0}(X)$:*

$$\forall x \in A, d(x_0, x) \leq r \tag{11.33}$$

Compte tenu de la remarque précédente sur les inclusions des boules, il est clair que nous pouvons remplacer l'adjectif "fermée" par "ouverte". De plus l'inégalité triangulaire entraîne que le caractère borné de A ne dépend pas du choix de x_0 (avec un x'_0 il suffit de remplacer r par $r' = r + d(x_0, x'_0)$).

Définition 11.14 *Soit X un ensemble et (Y, d) un espace métrique. Si X est un ensemble, nous disons qu'une fonction $f : X \to Y$ est "bornée" si son image $f(X)$ est bornée (cas de la fonction sinus ou cosinus par exemple).*

Définition 11.15 *Soit (E, d) un espace métrique, et soit A une partie non vide de E. Pour tout $u \in E$ nous notons $d(u, A)$ et nous appelons "distance de u à A", le nombre réel positif non nul :*

$$d(u, A) = \inf_{a \in A} d(u, a) = \inf \{d(u, a) | a \in A\} \tag{11.34}$$

Nous prolongeons la notion en posant :

$$\forall u \in E : d(u, \oslash) = +\infty \tag{11.35}$$

Si A et B sont deux parties de E nous avons respectivement (c'est peut-être plus compréhensible ainsi...) :

$$d(A, B) = \inf \{d(x, y) | x \in A, y \in B\} \tag{11.36}$$

Exemple :

Si nous prenons $A = \{0\} \subset \mathbb{R}$ et $B = \left\{\frac{1}{n+1}, n \in \mathbb{N}\right\}$ nous avons $d(A, B) = 0$ quand $n \to +\infty$ tandis que $A \neq B$. Ainsi, la distance entre les parties ne définit pas vraiment une distance sur $\mathcal{P}(\mathbb{R})$ dans cet exemple. Il s'agit donc d'un abus de notation et il faut bien interpréter $d(A, B)$ comme l'infinimum de la distance entre A et B

REMARQUE Si le lecteur à bien compris la définition du concept de "parties" il remarquera qu'il n'existe pas nécessairement toujours un $a \in A$ tel que $d(u, A) = d(u, a)$. En conséquence, nous écrivons trivialement :

$$\forall \varepsilon > 0, \exists \alpha = \alpha(\varepsilon) \in A : d(u, A) \leq d(u, \alpha) < d(u, A) + \varepsilon \tag{11.37}$$

De plus, si un tel α existe, il n'est bien évidemment pas nécessairement unique.

REMARQUE Il convient peut être de rappeler que cette distance satisfait également les 5 axiomes des distances.

Définition 11.16 *Soit (E, d) un espace métrique, et soit A une partie de E. Nous appelons "adhérence" de A et notons $adh(A)$ le sous-ensemble de E défini par :*

$$adh(A) = \{u \in E | d(u, A) = 0\} \tag{11.38}$$

En particulier, puisque $\forall u \in A : d(u, \oslash) = +\infty$, nous avons $adh(\oslash) = \oslash$, et puisque $\forall u \in E : d(u, E) = 0$, nous avons $adh(E) = E$.

REMARQUE Tout élément de l'ensemble $adh(A)$ est dit "point adhérent" à A

REMARQUE Nous disons qu'une partie A de E est une "partie fermée" si elle est égale à son adhérence

REMARQUE Nous disons qu'une partie A de E est une "partie ouverte" si elle est son complémentaire par rapport à E :

$$\complement_E A = A \setminus E \tag{11.39}$$

est fermé.

Il s'ensuite que (de par les définitions) :

$$\begin{aligned} u \in adh(A) &\Leftrightarrow \forall r > 0 :_f B_u^r \cap A = \oslash \\ A \text{ est ouverte } &\Leftrightarrow \forall u \in A, \exists \rho > 0 :_f B_u^r \subset A \end{aligned} \tag{11.40}$$

avec quelques propriétés :
P1. (triviale) Si $A \subset E$ et $B \subset E$ vérifient $A \subset B$, nous avons : $\forall u \in E : d(u, B) \leq d(u, A)$
P2. (triviale) Pour tout $A \subset E$, tout $u \in E : d(u, A) = d(u, adh(A))$
Dernière propriété qui a pour corroloaire (trivial) :
Si pour tout $u \in E$ nous avons $d(u, A) = d(u, B)$, $A, B \neq \oslash$, nous avons : $adh(A) = adh(B)$

Boules généralisées

La notion de distance d'un point à un ensemble permet d'étendre les notions de boule et de sphère.

Définition 11.17 *Soit $A \neq \oslash$ et soit un $r > 0$. Nous appelons "boule ouverte généralisée" de centre A et de rayon r, l'ensemble suivant :*

$$_oB_A^r = \{y \in E | d(y, A) < r\} \tag{11.41}$$

Respectivement "boule fermée généralisée" :

$$_fB_A^r = \{y \in E | d(y, A) \leq r\} \tag{11.42}$$

Respectivement "sphère généralisée" :

$$S_A^r = \{y \in E | d(y, A) = r\} \tag{11.43}$$

Définition 11.18 *Soit $(E, d$ un espace métrique et soient A, B deux parties non vides de E. Nous notons $g(A, B)$ et appelons "gap" (qui signifie "écartement" ou "espacement" en français) de A à B, le nombre réel supérieur ou égal à zéro :*

$$\begin{aligned} g(A,B) &= \inf_{a\in A} g(a,B) = \inf_{b\in B} g(b,A) \\ &= \inf_{a\in A}\inf_{b\in B} g(a,b) = \inf_{b\in A}\inf_{a\in B} g(a,b) \\ &= \inf\{d(a,b)|(a,b)\in A\times B\} \end{aligned} \tag{11.44}$$

REMARQUE L'inégalité triangulaire $g(A, B) \leq g(A, C) + g(C, B)$ n'est pas valide dans le cadre des gap (ceci étant également valable). Il suffit pour le démontrer, d'un seul et unique exemple qui contredirait l'inégalité.

Exemple :
Dans $\mathbb{R}$ prenons $A = \{0, 1\}, B = \{2, 3\}, C = \{1, 3\}$ nous avons alors :

$$g(A,B) = 1 > g(A,C) + g(C,B) = 0 + 0 = 0 \tag{11.45}$$

Il y a donc bien contradiction.

Diamètre

Définition 11.19 *Soit (E, d) un espace métrique et soit A une partie non vide de E. Nous notons $diam(A)$ et nous appelons "diamètre" de A, le nombre réel positif non nul :*

$$diam(A) = \sup_{a\in A}\sup_{b\in B} d(a,b) = \sup\{d(a,b)|(a,b)\in A\times A\} \tag{11.46}$$

Tout partie non vide A d'un espace métrique vérifiant $diam(A) < +\infty$ sera aussi dite "bornée".

REMARQUE Nous considérons la partie vide $\oslash$ comme un borné de diamètre A

Si l'espace métrique (E, d) tout entier est borné, nous disons que la distance d est bornée. Par exemple, la distance discrète est bornée, la distance usuelle sur $\mathbb{R}$ ne l'est pas.

Nous avons aussi les propriétés suivantes :

P1. (triviale) $diam(A) = 0 \Leftrightarrow A = \{a\}$ ou $A = \oslash$

P2. (triviale) $A \subset B \Rightarrow diam(A) \leq diam(B)$

P3. De par la définition du diamètre : $diam({}_fB_x^r) \leq 2r, diam({}_oB_x^r) \leq 2r, diam(S_x^r) \leq 2r$

Exemple :

Pour $diam(S_x^r) \leq 2r$ il suffit pour se convaincre de prendre la distance discrète. Ainsi, dans un espace métrique où nous prenons S_x^r avec $r = 1$, nous avons $diam(S_x^r) = 1$ (c'est un cas intéressant car complètement contre-intuitif).

P4. Nous avons $diam(A \cup B) \leq diam(A) + g(A, B) + diam(B)$

Exemple :

Dans $\mathbb{R}$ prenons $A = B$, nous avons alors (infériorité stricte triviale) :

$$(diam(A\cup B) = diam(A)) \leq diam(A) + g(A,B) + diam(B) \tag{11.47}$$

P5. A est borné si et seulement si $\exists r > 0, \exists x \in E : A \subset_o B_x^r$

Définition 11.20 *Nous appelons "excès de Hausdorff" de A sur B :*

$$e(A,B) = \sup\{d(a,B)|a\in A\} \tag{11.48}$$

REMARQUE Nous avons en général $e(B, A) \neq e(A, B)$ et ces quantités peuvent ne pas être finies.

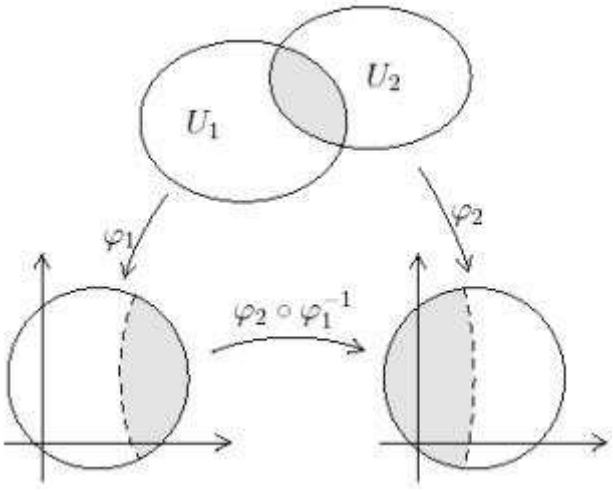


FIGURE 11.2 – Homéomorphisme d'une application de changement de carte

11.1.3 Variétés

Nous introduisons maintenant les "variétés". Ce sont des espaces topologiques qui sont "localement comme $\mathbb{R}^n$" (notre espace par exemple..).

Définition 11.21 *Une "variété topologique de dimension n" est un espace de Hausdorff M tel que pour tout $p \in M$ il existe un voisinage ouvert $U \subset M$ avec $p \in U$, un voisinage ouvert $U' \subset \mathbb{R}^n$ et un homéomorphisme :*

$$\varphi : U \to U' \tag{11.49}$$

Définition 11.22 *Un "homéomorphisme" entre deux espaces est une bijection continue dont l'inverse est également continu.*

Définition 11.23 *Les couples (U, φ) sont appelés des "cartes", U étant le "domaine de la carte" et "l'application de coordonnées". Au lieu de "carte nous disons parfois aussi "système de coordonnées".*

REMARQUE Nous noterons par $dimM$ la dimension d'une variété topologique. Ainsi :

$$dimM^n = n \tag{11.50}$$

Définition 11.24 *Soit M une variété topologique de dimension n. Une famille A de cartes de M est appelée un "atlas" si pour tout $x \in M$, il existe une carte $(U, \varphi) \in A$ telle que $x = U$.*

REMARQUE Notons que si $(U_1, \varphi_1), (U_2, \varphi_2)$ sont deux cartes de M telles que (ne vérifiant pas l'axiome de Hausdorff) $U_1 \cap U_2 \neq 0$, alors l'application de changement de cartes (cfr fig 11.2 :

$$\varphi_2 \circ \varphi_1^{-1} : \varphi_1(U_1 \cap U_2) \to \varphi_2(U_1 \cap U_2) \tag{11.51}$$

est un homéomorphisme.

Définition 11.25 *Une "variété différentiable" est un espace topologique M où les applications $\varphi_2 \circ \varphi_1^{-1}$ sont des fonctions de classe C^∞.*

Définition 11.26 *Un "difféomorphisme" est une application $f : \Omega \to \Omega'$ où Ω, Ω' sont des domaines ouverts de $\mathbb{R}^n$ et si f est un homéomorphisme et en plus si Ω, Ω' sont différentiables.*

REMARQUE "différentiable" dans ce contexte signifiera toujours différentiable de classe C^∞.

Variétés différentiables

Définition 11.27 *Soit une variété topologique $M = M^n$ (pour simplifier l'écriture), deux cartes $(U_1, \varphi_1), (U_2, \varphi_2)$ de M sont des "cartes compatibles" (plus précisément, compatibles de classe C^∞), si l'une des deux propriétés suivantes est vérifiée :*

P1. $U_1 \cap U_2 \neq 0$ et l'application $\varphi_2 \circ \varphi_1$ de changement de cartes est un difféomorphisme

P2. $U_1 \cap U_2 \neq 0$

Un atlas A de M est différentiable si toutes les cartes de A sont compatibles entre elles.

Définition 11.28 *Une "variété différentiable" est un couple (M, A) où M est une variété topologique et A un atlas différentiable de M.*

REMARQUE Etant donné un atlas différentiable, il est parfois nécessaire de le compléter : nous disons qu'une carte de M est compatible avec un atlas différentiable si elle est compatible avec chaque carte de A. Un atlas de A est une "atlas maximal" si toute carte compatible avec A appartient déjà à A. Un atlas maximal est appelé une "structure différentiable".

12

Théorie de la mesure

La mesure, au sens topologique, va nous permettre de généraliser la notion élémentaire de mesure d'un segment, ou d'une aire (au sens de Riemann, par exemple) et est indissociable de la nouvelle théorie de l'intégration que Lebesgue mettra en place de 1901 à 1902 et que nous allons aborder ici afin de construire des outils mathématiques beaucoup plus puissant que l'intégrale simple de Riemann (cf. chapitre de Calcul Différentiel Et Intégral).

La théorie de la mesure va également nous permettre de définir avec rigueur le concept de mesure (peu importe la mesure de quoi) et ainsi de revenir sur des résultats importants de l'étude des probabilités (cf. chapitre de Probabilités). Effectivement, nous allons voir (nous définirons le vocabulaire qui suit plus loin) pourquoi (U, A, P) est un "espace de probabilités" où A est au fait une tribu sur U et P une mesure sur l'espace mesurable (U, A).

Avertissement : Le niveau de connaissance requis pour la lecture de ce chapitre est assez élevé. Il faut être à l'aise avec les notions vues en théorie des ensembles ainsi qu'en topologie. Un très bon niveau d'abstraction et beaucoup de volonté de compréhension est aussi souhaité.

12.1 Espaces mesurables

Définition 12.1 *Soit E un ensemble. Une "tribu" (ou "σ-algèbre") sur E est une famille $\mathcal{A}$ de parties de E vérifiant les axiomes suivants :*

A1. $E \in \mathcal{A}$ (voir exemples plus bas - E étant considéré comme élément)

A2. Si A est un élément d'une tribu alors $A \in \mathcal{A} \Rightarrow A^{\complement} \in \mathcal{A}$. Ce qui signifie que $\mathcal{A}$ est "stable par passage au complémentaire". Cet axiome implique que l'ensemble vide $\varnothing$ est toujours un élément d'une tribu.

A3. Pour toute suite (A_n) d'éléments de $\mathcal{A}$ nous avons $\bigcup_{n=0}^{\infty} A_n \in \mathcal{A}$. Nous disons alors que $\mathcal{A}$ est "stable par union dénombrable".

Remarque Nous écrivons $E \in \mathcal{A}$ car nous ne considérons E non plus comme un sous-ensemble de $\mathcal{A}$ mais comme un élément de $\mathcal{A}$!

Remarque Les cas non-dénombrables sont typiques de la topologie, de la statistique ou du calcul intégral.

Définition 12.2 *Le couple $(E, \mathcal{A})$ est appelé "espace mesurable" et nous disons que les éléments de $\mathcal{A}$ sont des "ensembles mesurables". Si dans la troisième condition nous imposons seulement que $\mathcal{A}$ soit stable par union finie nous imposons la notion plus générale "d'algèbre". Ainsi, une tribu est nécessairement contenu dans une algèbre (mais le contraire n'est pas vrai car justement l'axiome est plus fort).*

Exemples :

E1. Soit $E = \{1, 2\}$ un ensemble de cardinal 2. Les deux seules tribus $\mathcal{A}$ qui satisfont les trois axiomes sont :

$$\mathcal{A}_1 = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\} \quad \mathcal{A}_2 = \{\varnothing, \{1, 2\}\} \tag{12.1}$$

Il n'y a pas d'autres tribus pour l'ensemble E donné que ces deux (la grossière, et la maximale), car il ne faut pas oublier que l'union de chacun des éléments de la tribu doit aussi être dans la tribu (A3), ainsi que le complémentaire d'un élément (A2).

E2. Soit E un ensemble, $\{\oslash, E\}$ est bien une tribu

E3. L'ensemble des parties de E, noté $\mathcal{P}(E)$ est aussi une tribu (dixit l'exemple 1).

Une tribu $\mathcal{A}$ est aussi "stable par union des complémentaires dénombrables". En effet si (A_n) est une suite d'éléments de $\mathcal{A}$ nous avons (trivial en prenant comme référence l'exemple E1) :

$$\bigcup_n A_i^c \in \mathcal{A} \text{ et } \left(\bigcup_n A_i^c\right)^c \in \mathcal{A} \tag{12.2}$$

Une tribu est aussi "stable par intersections dénombrables" (trivial en prenant comme référence l'exemple E1) :

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{A} \tag{12.3}$$

ce qui amène à ce qu'une tribu est stable par unions et intersections dénombrables. En particulier, si nous prenons deux éléments d'une tribu $A, B \in \mathcal{A}$ alors $A \setminus B \in \mathcal{A}$. Avec pour rappel (cf. chapitre de Théorie Des Ensembles) :

$$\boxed{A \setminus B = A \cap B^c} \tag{12.4}$$

REMARQUE Nous voyons aisément avec l'exemple E1 que si $(\mathcal{A}_i)_I$ une famille de tribus sur E (comme l'exemple E1). Alors est $\bigcap_I \mathcal{A}_i$ une tribu (la vérification des trois axiomes est immédiate)

Définition 12.3 *Soit E un ensemble et $\mathcal{B}$ une famille de sous-ensemble de $\mathcal{P}(E)$ tel que $\mathcal{B} \subseteq \mathcal{P}(E)$. Soit $\mathcal{S}$ la famille de toutes les tribus contenant (entre autres) $\mathcal{B}$. $\mathcal{S}$ n'est évidemment pas vide car $\mathcal{P}(E) \in \mathcal{S}$. Nous notons par définition :*

$$\boxed{\sigma(\mathcal{B}) = \bigcap_{\mathcal{A} \in \mathcal{S}} \mathcal{A}} \tag{12.5}$$

la "tribu engendrée" par $\mathcal{B}$. $\sigma(\mathcal{B})$ est donc par définition la plus petite tribu contenant $\mathcal{B}$ (et par extension la plus petite tribu de E).

Voici deux exemples qui permettent de vérifier si ce qui précède à été compris et qui permettent de mettre en évidence des résultats importants pour la suite.

Exemples :

E1. Soit E un ensemble, $A \subset E, A \neq E$ et $\mathcal{B} = \{A\}$alors (lorsque A est vu comme un sous-ensemble de E comme le précise l'énoncé et non comme une famille de sous-ensembles !) :

$$\sigma(\mathcal{B}) = \{\oslash, E, A, A^c\} \tag{12.6}$$

E2. Si $\mathcal{A}$ est une tribu sur E alors :

$$\sigma(\mathcal{A}) = \mathcal{A} \tag{12.7}$$

E3. Soit $E = \{1, 2, 3, 4\}$ et $A = 1, 2, 3$ nous avons dès lors (prenez bien garde car maintenant A est une famille de sous-ensemble et non simplement un unique sous-ensemble !) la tribu engendrée suivante :

$$\sigma(A) = \{\{\oslash\}, \{3\}, \{4\}, \{1, 2\}, \{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}\} \tag{12.8}$$

Plutôt que de déterminer cette tribu en cherchant la plus petite tribu $\mathcal{P}(E)$ de contenant A (ce qui serait laborieux) nous jouons avec les axiomes définissant une tribu pour facilement trouver celle-ci.

Ainsi, nous trouvons donc bien dans $\sigma(A)$ au moins l'ensemble vide obligatoire $\{\oslash\}$ ainsi que ($E = \{1,2,3,4\}$ selon l'axiome A1) et $A = \{\{1,2\},\{3\}\}$ lui-même (par définition de$\sigma(A)$) et les complémentaires de $\{\{3\}^c = \{1,2,4\}, \{1,2\}^c = \{3,4\}, \{\{1,2\} \cup \{3\}\}^c = \{4\}\}$ (selon l'axiome A2) ainsi que les unions $\bigcup_i A_n = \{\{1,3\} \cup \{3\} = \{1,2,3\}, \{1,2\} \cup \{1,2,3\} = \{1,2,3\}, \ldots\}$(selon l'axiome A3).

Définition 12.4 *Soit E un espace topologique (cf. chapitre de Topologie). Nous notons $\mathcal{B}(E)$ la tribu engendrée par les ouverts de E. $\mathcal{B}(E)$ est appelée la "tribu borélienne" sur E. Les éléments de $\mathcal{B}(E)$ sont appelés les "boréliens" de E.*

Remarque La notion de borélien est surtout intéressante car elle est nécessaire à la définition de la "tribu de Lebesgue" et par suite à "la mesure de Lebesgue" qui nous amènera à définir "l'intégrale de Lebesgue".

Remarque La tribu $\mathcal{B}(E)$ étant stable par passage au complémentaire, elle contient aussi tous les fermés.

Remarque Si E est un espace topologique à base dénombrable, $\mathcal{B}(E)$ est engendrée par les ouverts de la base.

Exemple :

Si $\mathbb{R}$ désigne l'espace des réels muni de la topologie euclidienne (cf. chapitre de Topologie), la famille des intervalles ouverts à extrémités rationnelles est une "base dénombrable" (étant donné les extrémités) de $\mathbb{R}$ et donc engendre $\mathcal{B}(\mathcal{R})$. Même remarque pour $\mathbb{R}^d$, $d \in \mathbb{N}^*$, avec comme base dénombrable la famille des pavés ouverts à extrémités rationnelles.

Considérons maintenant $\mathcal{S}$ un ensemble dense (cf. chapitre de Topologie) dans $\mathbb{R}$. Les familles suivantes engendrent $\mathcal{B}(\mathbb{R})$:

$$\begin{aligned} &\{]-\infty, a[\ , a \in \mathcal{S}\} \\ &\{]-\infty, a]\ , a \in \mathcal{S}\} \\ &\{]a, +\infty[\ , a \in \mathcal{S}\} \\ &\{[a, +\infty[\ , a \in \mathcal{S}\} \end{aligned} \tag{12.9}$$

Preuve Soit (la famille des ouverts) :

$$\mathcal{F} := \{]-\infty, a]\ , a \in \mathcal{S}\} \tag{12.10}$$

Nous avons évidement :

$$\sigma(\mathcal{F}) \subseteq \mathcal{B}(\mathbb{R}) \tag{12.11}$$

De plus :

$$[a, +\infty[\ \cap\ ]-\infty, b[= [a, b[\text{ si } a < b \tag{12.12}$$

Donc les intervalles du type $[a, b[$ avec a et b dans $\mathcal{S}$ appartiennent aussi à $\sigma(\mathcal{S})$. Donc, si nous généralisons, avec $x < y$, il existe une suite (a_n) d'éléments de $\mathcal{S}$ décroissant vers x et une suite (b_n) d'éléments de $\mathcal{S}$ croissant vers y tel que :

$$]x, y[= \bigcup_n [a_b, b_n[\in \sigma(\mathcal{F}) \tag{12.13}$$

ce qui entraîne au même titre (que $E \in \mathcal{A}$) que $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{F})$. Les autres cas se traitent de manière analogue. •

Soit $(E, \mathcal{A})$ un espace mesurable et $A \subseteq E$ (et $A \in \mathcal{A}$) (où A est donc considéré comme un sous-ensemble et non comme un élément !). La famille $\{A \cap B | B \in \mathcal{A}\}$ est une tribu sur A appelée "tribu trace" de $\mathcal{A}$ sur A, nous la noterons $A \cap \mathcal{A}$. De plus, si $A \in \mathcal{A}$, la tribu trace est formée par les ensembles mesurables contenus dans A.

Preuve Nous allons faire une démonstration par l'exemple (...). Nous vérifions les trois points de la définition d'une tribu :

1. $A = E \cap A \Rightarrow A \in A \cap \mathcal{A}$
2. Soit $B \in \mathcal{A}, \mathcal{A} \setminus (A \cap B) = A \setminus B = A \cap B^c$ et donc $A \setminus (A \cap B) \in A \cap \mathcal{A}$

Exemple :

Soit $E = \{1, 2, 3\}$ alors (une tribu parmi d'autres - ne pas oublier la stabilité par union !) :

$$\mathcal{A} = \{\oslash, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\} \tag{12.14}$$

Choisissons $A = \{1, 2\}, B = \{2, 3\}$(il est évident que $\{A \cap B | B \in \mathcal{A}\}$ est une tribu sur A)
Dès lors :

$$\begin{aligned} A \setminus (A \cap B) = \{1, 2\} - (\{1, 2\} \cup \{2, 3\}) = \{1, 2\} - \{2\} = \{1\} \\ A \setminus B = \{1, 2\} - \{2, 3\} = \{1\} \\ A \cap B^c = \{1, 2\} \cap \{2, 3\}^c = \{1, 2\} \cap \{1\} = \{1\} \end{aligned} \tag{12.15}$$

et nous avons bien $\{1\} \in A$ ainsi que $A \in A \cap \mathcal{A}$.

3. Soit $(A \cap B_n)$ une suite d'éléments de $A \cap \mathcal{A})(B_n \in \mathcal{A})$ alors :

$$\bigcup_n (A \cap B_n) = A \cap \underbrace{\left(\bigcup_n B_n \right)}_{\in \mathcal{A}} \in A \cap \mathcal{A} \tag{12.16}$$

La dernière assertion de la proposition sera supposée évidente. •

Soit maintenant$E$ un ensemble, $\mathcal{C}$ une famille de parties de E et $A \subseteq E$ non vide. Nous notons $A \cap \mathcal{C}$ la trace de $\mathcal{C}$ sur A et $\sigma_a(A \cap \mathcal{C})$ la tribu engendrée par $A \cap \mathcal{C}$ sur A. Alors :

$$A \cap \sigma(\mathcal{C}) = \sigma_A(A \cap \mathcal{C}) \tag{12.17}$$

Exemple :

Soit l'ensemble $E = \{1, 2, 3, 4\}, \mathcal{C} = \big\{\{1, 2\}, \{3\}\big\}, A = \{3, 4\}$ alors :

$$A \cap \mathcal{C} = \{\oslash, \{3\}\} \; \sigma_A(A \cap \mathcal{C}) = \sigma_A\big(\{\oslash, \{3\}\}\big) = \big\{\oslash, \{3\}, \{4\}, \{3, 4\}\big\} \tag{12.18}$$

Et vérifions $A \cap \sigma(\mathcal{C}) = \sigma_A(A \cap \mathcal{C})$:

$$\begin{aligned} A \cap \sigma(\mathcal{C}) &= \{3, 4\} \cap \sigma\big\{\{1, 2\}, \{3\}\big\} \\ &= \{3, 4\} \cap \big\{\oslash, \{3\}, \{4\}, \{1, 2\}, \{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}\big\} \\ &= \big\{\oslash, \{3\}, \{4\}, \{3, 4\}\big\} \end{aligned} \tag{12.19}$$

Donc l'égalité est vérifiée.

Un corollaire trivial de cette égalité est que si nous considérons un espace topologique E et $A \subseteq E$ muni de la topologie induite. Alors :

$$A \cap \mathcal{B}(E) = \mathcal{B}(A) \tag{12.20}$$

Rappelons (cf. chapitre de Théorie Des Ensembles) que si nous avons E qui est un ensemble, alors pour tout $A, B \subseteq E$ nous définissons la différence symétrique $A \Delta B$ entre A et B par :

$$A \Delta B = (A \setminus B) \cup (B \setminus A) \tag{12.21}$$

Les propriétés triviales sont les suivantes :

P1. Une algèbre est stable par différence symétrique ($A, B \in \mathcal{A}$ nous avons $A \Delta B \in \mathcal{A}$)

P2. $A\Delta B = B\Delta A$
P3. $A^c\Delta B^c = A\Delta B$
P3. $A\Delta B = (A \cup B) \setminus (A \cap B)$

Si $\mathcal{B}$ est une algèbre sur E, alors $(\mathcal{B}, \Delta, \cap)$ est un "anneau de Boole" (ou algèbre de Boole mais attention avec le terme "algèbre" qui peut prêter à confusion avec la théorie des ensembles) avec $\oslash$ et E comme élément neutre "additif" (Δ) respectivement "multiplicatif" ($\cap$).

Pour des rappels sur les éléments cités dans le paragraphe précédent, le lecteur pourra se reporter au chapitre de théorie des ensembles et le chapitre d'algèbre de Boole (cf. chapitre de Systèmes Logiques Formels)

Preuve Δ ("addition") est associative car en développant nous obtenons (cela se vérifie en faisant un diagramme saggital au besoin - les "patates") :

$$A\Delta(B\Delta C) = (A \cap B \cap C) \cup (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C) \tag{12.22}$$

et cette dernière expression est stable par permutation (commutation) de A et C (même méthode de vérification) . Donc :

$$A\Delta(B\Delta C) = C\Delta(B\Delta A) = (B\Delta A)\Delta C = (A\Delta B)\Delta C \tag{12.23}$$

Nous vérifions que vérifie que $\oslash$ est bien neutre par rapport la différence symétrique (la démonstration que E est neutre par rapport à l'inclusion est évidente). Il est trivial que :

$$A\Delta\oslash = \oslash\Delta A \text{ et que } A\Delta A = \oslash \tag{12.24}$$

•

$\mathcal{B}, \Delta, \oslash)$ est donc bien un groupe abélien par rapport à la loi Δ (différence symétrique).

Pour finir $\cap$ est distributif par rapport à Δ. En effet :

$$\begin{aligned} A \cap (B\Delta C) &= A \cap (B \setminus C \cup C \setminus B) \\ &= (A \cap B \setminus C) \cup (A \cap C \setminus B) \\ &= \big(A \cap B \setminus (A \cap C)\big) \cup \big(A \cap C \setminus (A \cap B)\big) \\ &= (A \cap B)\Delta(A \cap C) \end{aligned} \tag{12.25}$$

Ce qui fait bien de $\mathcal{B}, \Delta, \oslash)$ un anneau (qui de plus est un anneau commutatif).

Théorème de la classe monotone

Définition 12.5 *Soit E un ensemble. Une "classe monotone" sur E est une famille $\mathcal{C}$ de parties de E vérifiant :*
A1. $E \in \mathcal{C}$
A2. $A, B \in \mathcal{C}$ et $A \subseteq B \Rightarrow B \setminus A \in \mathcal{C}$
A3. Si (A_n) est une suite croissante (attention au terme "croissant") d'éléments de $\mathcal{C}$ alors $\bigcup_{i=0}^{\infty} a_i \in \mathcal{C}$ (stable par union dénombrable croissant)

Remarque Une suite croissante d'ensembles c'est : $A_1 \subseteq A_2 \subseteq A_3 \ldots$

Remarque Les deux premiers points impliquent que $\mathcal{C}$ est stable par passage au complémentaire.

Remarque Les axiomes (2) et (3) (plus le (1)) amènent que la classe monotone est stable par intersection décroissante. Une manière de vérifier c'est de prendre le complémentaire de chaque élément de la suite croissante pour tomber sur la suite décroissante et inversement.

Remarque L'axiome 3 des classes monotones étant un peu plus restrictif (plus "fort") que l'axiome 3 des tribus (puisque nous y imposons une suite croissante). Cela implique que toute tribu est une classe monotone (tout union dénombrable de la tribu étant dans la tribu ce qui est une condition plus forte que la suite croissante) !

De la même manière que pour les tribus, si nous considérons une famille $(\mathcal{C}_i)_I$ de classes monotones sur E. Alors $\bigcap_I \mathcal{C}_i$ est une classe monotone (la démonstration se vérifie immédiatement par les trois axiomes précédemment cités).

Exemple :

Si E est un ensemble, $\mathcal{P}(E)$ est une classe monotone sur E. Plus généralement, une tribu est une classe monotone.

De manière équivalente aux tribus, considérons un ensemble E et $\mathcal{C} \subset \mathcal{P}(E)$. Soit $\mathcal{S}$ la famille de toutes les classes monotones contenant $\mathcal{C}$. $\mathcal{S}$ n'est pas vide car $\mathcal{P}(E) \in \mathcal{S}$. Nous notons :

$$\mathcal{M}(\mathcal{C}) = \bigcap_{\mathcal{T} \in \mathcal{S}} \mathcal{T} \tag{12.26}$$

la classe monotone engendrée par $\mathcal{C}$. $\mathcal{M}(\mathcal{C})$est la plus petite classe monotone contenant $\mathcal{C}$ (et satisfaisant bien évidemment aux axiomes).

REMARQUE Si E est un ensemble et $\mathcal{C} \subseteq \mathcal{P}(E)$ alors $\mathcal{M}(\mathcal{C}) \subseteq \sigma(\mathcal{C})$, car $\sigma(\mathcal{C})$ est une classe monotone (et aussi une tribu) contenant $\mathcal{C}$ et donc elle contient aussi $\mathcal{M}(\mathcal{C})$ (voir les exemples avec les tribus).

Le théorème (de la classe monotone) s'énonce ainsi : soit E un ensemble. Si $\mathcal{C}$ est une famille de parties de E que nous imposons stable par intersections finies alors $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$ (nous devons donc prouver que la tribu minimale de $\mathcal{C}$ est égale à la classe monotone minimale de $\mathcal{C}$). Si nous imposons pas que $\mathcal{C}$ soit pas stable par intersections finies nous aurions pas nécessairement l'égalité.

Preuve $\mathcal{M}(\mathcal{C}) \subseteq \sigma(\mathcal{C})$ comme déjà dit (c'est quasiment trivial). Nous allons montrer dans un premier temps que $\mathcal{M}(\mathcal{C})$ est donc aussi une tribu sur E. Pour ceci il suffit de montrer que $\mathcal{M}(\mathcal{C})$ est (aussi) stable par union dénombrable (et non nécessairement par une suite croissante d'éléments !).

Considérons les familles suivantes pour la démonstration :

$$\mathcal{M}_1 = \left\{ A \in \mathcal{M}(\mathcal{C}) | A \cap B \in \mathcal{M}(\mathcal{C}) \text{ pour tout } B \in \mathcal{C} \right\} \tag{12.27}$$

$$\mathcal{M}_2 = \left\{ A \in \mathcal{M}(\mathcal{C}) | A \cap B \in \mathcal{M}(\mathcal{C}) \text{ pour tout } B \in \mathcal{M}(\mathcal{C}) \right\} \tag{12.28}$$

Par les définitions précédentes $\mathcal{M}_1 \subseteq \mathcal{M}(\mathcal{C})$ mais $\mathcal{C}$ étant (imposé) stable par intersections finies entraîne $\mathcal{C} \subseteq \mathcal{M}_1$ et donc (c'est le même raisonnement que pour les tribus) :

$$\mathcal{C} \subseteq \mathcal{M}_1 \subseteq \mathcal{M}(\mathcal{C}) \tag{12.29}$$

$\mathcal{M}_1$ est une classe monotone en effet $E \in \mathcal{M}_1$, si $A_1, A_2 \in \mathcal{M}_1$ et que $A_1 \subseteq A_2$ (axiome 2) alors :

$$\forall B \in \mathcal{C}, \underbrace{A_1 \cap B}_{\in \mathcal{M}(\mathcal{C})} \subseteq \underbrace{A_2 \cap B}_{\in \mathcal{M}(\mathcal{C})} \tag{12.30}$$

et donc (ce qui appuie que les autre éléments (A_n) satisfont la relation précédente) :

$$\forall B \in \mathcal{C}, (A_2 \cap B) \setminus (A_1 \cap B) \in \mathcal{M}(\mathcal{C}) \Rightarrow (A_2 \setminus A_1) \cap B \in \mathcal{M}(\mathcal{C}) \tag{12.31}$$

Si (A_n) est une suite croissante d'éléments de $\mathcal{M}_1$ alors :

$$\forall B \in C, \left(\bigcup A_n \right) \cap B = \bigcup \underbrace{(A_n \cap B)}_{\in \mathcal{M}(\mathcal{C})} \in \mathcal{M}(\mathcal{C}) \tag{12.32}$$

car $(A_n \cap B)$ est une suite croissante.

Ainsi $\mathcal{M}_1$ est bien une classe monotone et par $C \subseteq \mathcal{M}_1 \subseteq \mathcal{M}(\mathcal{C})$, nous avons donc :

$$\mathcal{M}_1 = \mathcal{M}(\mathcal{C}) \tag{12.33}$$

Cette dernière égalité implique $\mathcal{C} \subseteq \mathcal{M}_2$. Comme pour $\mathcal{M}_1$, nous montrons montrons que $\mathcal{M}_2$ est une classe monotone et donc $\mathcal{M}(\mathcal{C}) = \mathcal{M}_2$, ce qui veut dire par extension que $\mathcal{M}(\mathcal{C})$ est est donc stable par intersections finies.

$\mathcal{M}(\mathcal{C})$ étant stable par passage au complémentaire ceci entraîne que $\mathcal{M}(\mathcal{C})$ est, nous venons de le montrer, stable par unions finies (alors que nous voulons démontrer que c'est stable par union dénombrable).

Soit à présent (A_n) une suite d'éléments de $\mathcal{M}(\mathcal{C})$. Nous considérons la suite :

$$B_n = \bigcup_{i=1}^{n} A_i \tag{12.34}$$

(B_n) est une suite croissante d'éléments de $\mathcal{M}(\mathcal{C})$, donc :
mais

$$\bigcup_{n=1}^{\infty} B_n \in \mathcal{M}(\mathcal{C}) \text{ mais } \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n \tag{12.35}$$

Donc :

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}(\mathcal{C}) \tag{12.36}$$

•

Ainsi $\mathcal{M}(\mathcal{C})$ est stable par union dénombrable et enfin $\mathcal{M}(\mathcal{C})$ est une tribu. Or comme $C \subseteq \mathcal{M}(\mathcal{C}) \subseteq \sigma(\mathcal{C})$ cela nous amène donc à $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.

Nous verrons plus tard quelques applications importantes de ce théorème.

13

Calcul différentiel et intégral

Le calcul différentiel est un des domaines les plus passionnants et vastes de la mathématique, et il existe une littérature considérable (colossale) sur le sujet. Les résultats retrouvent des implications dans absolument tous les domaines de la physique, de l'informatique, de l'électronique, de la chimie, biologie et de la mathématique elle-même.

Les mathématiciens ont rédigé une telle quantité de théorèmes sur le sujet que la validation de certains est délicate car certains à eux-seuls nécessiteraient la vie d'un homme pour être parcourus (c'est un problème que la communauté des mathématiciens reconnaît) et vérifiés (ce qui fait que personne ne les vérifie...).

Ce constat fait, nous avons choisi de ne présenter ici que les points absolument nécessaires à la compréhension des outils fondamentaux du physicien et du mathématicien (niveau licence à peu près). Les puristes nous excuseront donc pour l'instant de ne pas présenter certains théorèmes qui peuvent leur sembler indispensables mais que nous rédigerons une fois le temps venu...

Nous allons principalement étudier dans ce qui va suivre ce que les mathématiciens aiment bien préciser (et ils ont raison) : les cas généraux des fonctions réelles à une variable réelle. Les fonctions plus complexes (à plusieurs variables réelles ou complexes, continues ou discrètes) viendront une fois cette partie terminée.

Remarque : Nous ne nous attarderons pas à démontrer les dérivées et primitives de toutes les fonctions car comme il y a une infinité de fonctions possibles, il y a également une infinité de dérivées et de primitives. C'est le rôle des professeurs dans les instituts scolaires d'entraîner les élèves à appliquer et à comprendre le raisonnement de dérivation et d'intégration par des applications sur des fonctions connues (l'internet ne remplacera très probablement jamais l'école à ce niveau).

13.1 Calcul différentiel

Soit une fonction f réelle à une variable réelle x notée $f(x)$ (nous nous limitons à ce cas de figure pour l'instant et étudierons les dérivées partielles dans des espaces à un nombre de dimensions quelconques plus loin) continue au moins dans un intervalle où se situe l'abscisse a.

Définition 13.1 *Nous appelons "pente moyenne", ou encore "coefficient directeur" le rapport de la projection orthogonale de deux points $x_1 \neq x_2$ de la fonction f non nécessairement continue sur l'axe des abscisses et des ordonnées tel que :*

$$\frac{\Delta f}{\Delta x} = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \frac{f(x_1 + h) - f(x_1)}{(x_1 + h) - (x_1)} =_{x_1 := a} \frac{f(a + h) - f(a)}{h} \tag{13.1}$$

Remarque Δ signifiant "un delta" exprime le fait que nous sous-entendons une différence d'une même quantité.

Définition 13.2 *Nous appelons "nombre dérivé en a" ou "pente instantanée" ou encore "dérivée première", la limite quand h tend vers 0 (si elle existe) du rapport de la projection orthogonale de deux points $x_1 \neq x_2$ infiniment proches de la fonction f continue (dans le sens qu'elle ne contient pas de "trous") sur l'axe des abscisses et des ordonnées tel que :*

$$\frac{df_a}{dx} = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h} = f'(a) \tag{13.2}$$

Une interprétation graphique donne que $f'(a)$ est le coefficient directeur de la tangente au point d'abscisse a.

REMARQUE d signifiant un "différentiel" exprime le fait que nous sous-entendons une différence infiniment petite d'une même quantité.

REMARQUE Nous renvoyons le lecteur au chapitre d'Analyse Fonctionnelle pour la définition de ce qu'est une fonction continue.

Définition 13.3 *Soit f une fonction définie sur un intervalle I et dérivable en tout point a de I, la fonction qui à tout réel a de I associe le nombre $f'(a)$ est appelée "fonction dérivée de f sur I" et est notée f'.*

REMARQUE Au niveau des notations les physiciens adoptent suivant leur humeur différentes notations possibles pour les dérivées. Ainsi, considérons la fonction réelle à une variable $f(x)$, vous trouverez dans la littérature ainsi que dans le présent site les notations suivantes pour la dérivée première :

$$\frac{d}{dx}f(x) = \frac{df(x)}{dx} = d_x f(x) = \dot{f}(x) = f'(x) \tag{13.3}$$

ou encore en considérant implicitement que f est fonction de x (ceci permet d'alléger un petit peu la tailles des développements) :

$$\frac{d}{dx}f = \frac{df}{dx} = d_x f = \dot{f} = f' \tag{13.4}$$

Nous pouvons de la même manière définir les dérivées d'ordre 2 (dérivée d'une dérivée), les dérivées d'ordre 3 (dérivée d'une dérivée d'ordre 2) et ainsi de suite. Nous rencontrerons par ailleurs très fréquemment de telles dérivées en physique (et même en maths pour l'analyse fonctionnelle).

Maintenant, suite à un problème de compréhension dans un des chapitres d'un internaute, précisons une technique utilisée fréquemment par les physiciens. Considérons une dérivée d'ordre 2 telle que :

$$\ddot{f} = f'' = \frac{d^2 f(x)}{dx^2} = \frac{d}{dx}\left(\frac{df(x)}{dx}\right) = \frac{d}{dx}\left(\frac{d}{dx}f(x)\right) \tag{13.5}$$

Si nous regardons le d/dx comme un opérateur différentiel nous pouvons bien évidemment écrire :

$$\frac{d}{dx}\left(\frac{d}{dx}f(x)\right) = \frac{d}{dx}\frac{d}{dx}(f(x)) = \left(\frac{d}{dx}\right)^2 f(x) \tag{13.6}$$

Finalement nous avons :

$$\frac{d^2 f(x)}{dx^2} = \left(\frac{d}{dx}\right)^2 f(x) \tag{13.7}$$

et donc il vient après simplification par f(x) :

$$\frac{d^2}{dx^2} = \left(\frac{d}{dx}\right)^2 \tag{13.8}$$

Cela peut paraître évident pour certains mais parfois moins pour d'autres et il était visiblement utile de préciser cela car c'est souvent utilisé en relativité et en physique quantique.

Indiquons et démontrons maintenant deux propriétés intuitivement évidente des dérivées et qui nous seront plusieurs fois indispensables pour certaines démonstrations sur ce site (comme par exemple dans le chapitre de méthodes numériques ou ici même...).

Considérons d'abord deux nombres réels $a < b$ et f une fonction à valeurs réelles continue sur $[a, b]$ et dérivable sur $]a, b[$ telle que $f(a) = f(b)$. Alors nous voulons démontrer qu'il existe bien évidemment au moins un élément c de $]a, b[$ tel que $f'(c) = 0$ (c'est typiquement le cas des fonctions polynômial !).

Cette propriété est appelée "théorème de Rolle" et donc explicitement elle montre qu'il existe au moins un élément où la dérivée de f est nulle si en la parcourant nous revenons à la même valeur des images pour deux valeurs distinctes des abscisses, c'est-à-dire qu'il existe au moins un point où la tangente est horizontale.

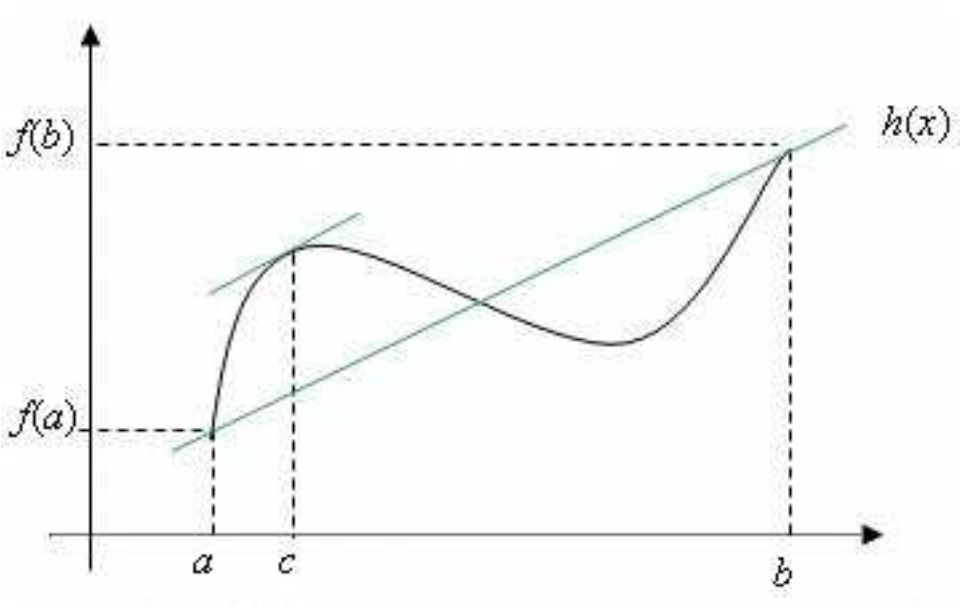


FIGURE 13.1 – le coefficient directeur de $c = (f(b) - f(q))/(b - a)$

Preuve Si f est constante, le résultat est immédiat. Dans le cas contraire, comme f est continue sur l'intervalle fermé borné $[a, b]$ elle admet au moins un minimum global ou maximum global compte tenu que nous nous basons sur l'hypothèse que $f(a) = f(b)$ et que f n'est pas constante. L'extremum est atteint en un point c appartenant à l'intervalle ouvert $]a, b[$ (le fait de prendre l'intervalle ouvert permet dans certains cas d'éviter d'avoir un extremum à nouveau en a ou en b). Supposons comme premier cas que $f(c)$ est maximum global. La dérivée de la fonction f entre c et un deuxième point ont alors un signe connu. Pour h strictement positif et tel que $c + h$ appartienne à l'intervalle $[a, b]$:

$$\frac{f(c+h) - f(c)}{h} < 0 \tag{13.9}$$

En considérant la limite quand h tend vers 0, le nombre dérivé $f'(c)$ est négatif. Pour h strictement négatif et tel que $c + h$ appartienne à l'intervalle $[a, b]$:

$$\frac{f(c+h) - f(c)}{h} > 0 \tag{13.10}$$

En considérant la limite quand h tend vers 0, le nombre dérivé $f'(c)$ est positif. Au bout du compte, la dérivée de f est nulle au point c. La démonstration est analogue si $f(c)$ est un minimum global, avec les signes des dérivées qui sont les opposés. •

Maintenant, considérons deux réels $a < b$ et $f(x)$ une fonction continue sur $[a, b]$ et dérivable sur $]a, b[$. Alors, nous nous proposons de montrer qu'il existe au moins un réel $c \in]a, b[$ tel que :

$$f(b) - f(a) = f'(c)(b - a) \tag{13.11}$$

Ce qui peut aussi s'écrire sous la forme suivante :

$$f(x_0 + k) - f(x_0) = f'(x_0 + s \cdot k)k \tag{13.12}$$

avec $s \in]0, 1[$.

Géométriquement cela signifie qu'en au moins un point c du graphe de la fonction $f(x)$, il existe une tangente de coefficient directeur (cfr fig. 13.1 :

$$\frac{f(b) - f(a)}{b - a} \tag{13.13}$$

Preuve Nous avons d'abord :

$$h(x) = (x-a)\cdot\frac{f(b)-f(a)}{b-a} + f(a) \tag{13.14}$$

car le pente de $h(x)$ est bien évidemment $(f(b)-f(q))/(b-a)$ et comme lorsque $x=a$ nous devons avoir $f(a)$ il s'ensuit donc la relation donnée précédemment.

Ensuite, pour démontrer qu'un tel point c existe, l'idée est de rapporter les deux points a et b à la même ordonnée ce qui en fait nous ramène au théorème de Rolle et pour cela, nous définissons une fonction g par :

$$g(x) = f(x) - h(x) = f(x) - \left[(x-a)\cdot\frac{f(b)-f(a)}{b-a} + f(a)\right] \tag{13.15}$$

qui est telle que effectivement $g(b) = g(a)\ldots$ et en l'occurrence égal à 0 (mais cette valeur importe peu). Dès lors, le théorème de Rolle vu précédemment nous indique qu'il existe un point entre a et b où la dérivée de $g(x)$ est nulle tel que $g'(c) = 0$. Et en constatant que :

$$g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a} \tag{13.16}$$

nous obtenons :

$$g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a} = 0 \tag{13.17}$$

Soit après simplification :

$$f(b)-f(a) = f'(c)(b-a) \tag{13.18}$$

•

Puisque le terme de gauche représente un accroissement fini du terme de droit, alors ce résultat est appelée "théorème des accroissements finis" (TAF).

À l'aide de ce petit théorème et des outils mathématiques introduits précédemment, nous pouvons construire un petit théorème fort utile et puissant en physique.

Définition 13.4 *Nous appelons "règle de L'Hôpital" (également appelée "règle de l'Hospital" ou "règle de Bernoulli") utilise la dérivée dans le but de déterminer les limites difficiles à calculer de la plupart des quotients et qui apparaissent souvent en physique.*

Preuve Considérons deux fonctions $f(x)$ et $g(x)$ et telles que $f(a) = g(a) = 0$ alors nous pouvons écrire :

$$\begin{aligned}\frac{f(x)}{g(x)} &= \frac{f(x)}{g(x)}\cdot\frac{x-a}{x-a}\\ &= \frac{f(x)}{x-a}\cdot\frac{x-a}{g(x)}\\ &= \frac{f(x)-f(a)}{x-a}\cdot\frac{x-a}{g(x)-g(a)}\end{aligned} \tag{13.19}$$

Alors selon la définition de la dérivée :

$$\begin{aligned}\lim_{x\to a}\frac{f(x)}{g(x)} &= \lim_{x\to a}\frac{f(x)-f(a)}{x-a}\cdot\frac{x-a}{g(x)-g(a)}\\ &= \frac{f(a+h)-f(a)}{(a+h)-a}\cdot\frac{(a+h)-a}{g(a+h)-g(a)}\\ &= \frac{f'(a)}{g'(a)}\end{aligned} \tag{13.20}$$

•

Nous pouvons généraliser ce résultat initialement basé sur la contrainte $f(a) = g(a) = 0$.

Preuve Rappelons donc que selon le théorème des accroissement finis, si $f(x)$ est dérivable sur un intervalle $]a,b[$ et continue sur $[a,b]$ alors il existe un réel c dans l'intervalle $[a,b]$ tel que :

$$f'(c) = \frac{f(b)-f(a)}{b-a} \tag{13.21}$$

Si le théorème se vérifie pour deux fonctions satisfaisant aux mêmes contraintes alors nous avons deux fonctions telles que :

$$f'(c) = \frac{f(b) - f(a)}{b - a} \text{ et } g'(c) = \frac{g(b) - g(a)}{b - a} \tag{13.22}$$

Si $g'(c)$ est non nul nous avons alors tout à fait le droit d'écrire le rapport (certains appellent cela le "théorème des accroissements fini généralisé"...) :

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{b - a} \cdot \frac{b - a}{g(b) - g(a)} = \frac{f(b) - f(a)}{g(b) - g(a)} \tag{13.23}$$

ce qui sans perdre en validité tant que c est dans l'étau $[a, x]$ peut s'écrire :

$$\frac{f'(c)}{g'(c)} = \frac{f(x) - f(a)}{g(x) - g(a)} \tag{13.24}$$

Ainsi, lorsque $x \to a$ ce qui implique que l'étau $[a, x]$ se referme et donc $c \to a$ nous avons :

$$\begin{aligned} \lim_{x \to a} \frac{f'(c)}{g'(c)} &= \lim_{x \to a} \frac{f(x) - f(a)}{g(x) - g(a)} \\ &= \frac{f(a + h) - f(a)}{g(a + h) - g(a)} \\ &= \frac{f(a + h) - f(a)}{(a + h) - a} \cdot \frac{(a + h) - a}{g(a + h) - g(a)} \\ &= \frac{f'(a)}{g'(a)} \end{aligned} \tag{13.25}$$

Ainsi, nous venons de prouver quand dans la démonstration précédente de la règle de l'Hôpital la relation :

$$\frac{f(a + h) - f(a)}{(a + h) - a} \cdot \frac{(a + h) - a}{g(a + h) - g(a)} = \frac{f'(a)}{g'(a)} \tag{13.26}$$

que nous avions est vraie en toute généralité et qu'il n'est pas nécessaire que $f(a) = g(a) = 0$ soit vrai pour que le résultat soit juste ! •

13.1.1 Différentielles

Nous avons indiqué précédemment ce qu'était un différentiel d. Mais il existe au fait plusieurs types de sortes de différentielles d'une fonction (remarquez que nous distinguons le genre masculin et féminin du terme) :

1. Les différentiels
2. Les différentielles partielles
3. Les différentielles totales exactes
4. Les différentielles totales inexactes

Rappelons que nous appelons "différentiel df" d'une fonction f à une variable la relation donnée par (voir texte précédent) :

$$df = f'dx \tag{13.27}$$

Cependant, pour exprimer l'effet d'un changement de toutes les variables d'une fonction f de plusieurs variables, nous devons utiliser un autre type de différentiel que nous appelons la "différentielle totale" (dérivée en deux sous-familles : différentielle totale exacte et différentielle totale inexacte).

Soit par exemple, une fonction $f(x, y)$ des deux variables x et y. L'accroissement df de la fonction f, pour un accroissement fini de x à $x + \Delta x$ et de y à $y + \Delta y$ est :

$$\Delta f = f(x + \Delta x, y + \Delta y) - f(x, y) \tag{13.28}$$

que nous pouvons aussi écrire :

$$\Delta f = f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y) + f(x, y + \Delta y) - f(x, y) \tag{13.29}$$

ou encore :

$$\Delta f = \frac{f(x+\Delta x, y+\Delta y) - f(x, y+\Delta y)}{\Delta x}\Delta x + \frac{f(x, y+\Delta y) - f(x,y)}{\Delta y}\Delta y \tag{13.30}$$

Pour des accroissements infiniment petits de x et y :

$$\begin{aligned} df = &\lim_{\Delta x \to 0}\left(\frac{f(x+\Delta x, y+\Delta y) - f(x, y+\Delta y)}{\Delta x}\Delta x\right) \\ &+ \lim_{\Delta y \to 0}\left(\frac{f(x, y+\Delta y) - f(x,y)}{\Delta y}\Delta y\right) \end{aligned} \tag{13.31}$$

Intéressons nous dès lors aux deux termes :

$$\frac{f(x+dx, y+dy) - f(x, y+dy)}{dx} \text{ et } \frac{f(x, y+dy) - f(x,y)}{dy} \tag{13.32}$$

Le premier terme de gauche, nous le voyons, ne donne finalement que la variation en x de la fonction $f(x, y)$ en ayant y constant sur la variation. Nous notons cela dès lors (si la connaissance des variables constantes est triviale, nous ne les indiquons plus) :

$$\left(\frac{\partial f}{\partial x}\right)_y = \frac{\partial f}{\partial x} = \frac{f(x+dx, y+dy) - f(x, y+dy)}{dx} \tag{13.33}$$

et de même :

$$\left(\frac{\partial f}{\partial y}\right)_x = \frac{\partial f}{\partial y} = \frac{f(x, y+dy) - f(x,y)}{dyx} \tag{13.34}$$

Les deux expressions :

$$\left(\frac{\partial f}{\partial x}\right)_y \text{ et } \left(\frac{\partial f}{\partial y}\right)_x \tag{13.35}$$

sont ce que nous appelons des "différentielles partielles". Il vient dès lors :

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \tag{13.36}$$

qui est la "différentielle totale exacte" de df. Il est important de se rappeler de la forme de cette relation que nous retrouverons partout dans des opérateurs particuliers en physique, dans la mécanique des fluides, dans la thermodynamique, etc.

REMARQUE De la même manière, pour une fonction de plus de deux variables, par exemple $f(x, y, z)$, la différentielle totale df est :

$$\begin{aligned} df &= \left(\frac{\partial f}{\partial x}\right)_{y,z} + \left(\frac{\partial f}{\partial y}\right)_{x,z} + \left(\frac{\partial f}{\partial z}\right)_{x,y} \\ &= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \end{aligned} \tag{13.37}$$

Dans l'équation ci-dessus, la différentielle df a été calculée à partir de l'expression de la fonction f. Puisqu'il existe une fonction f qui vérifie l'expression de df, la différentielle df est dite alors aussi "totale exacte".

Profitons pour faire une indication importante sur l'utilisation des dérivées partielles par les phyisiciens (et donc dans les nombreux chapitres y relatifs du site). Nous avons vu plus haut que si f dépend de deux variables x, y nous avons toujours :

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \tag{13.38}$$

et s'il ne dépend que d'un variable nous avons alors :

$$df = \frac{\partial f}{\partial x}dx \tag{13.39}$$

et alors :

$$\frac{df}{dx} = \frac{\partial f}{\partial x}dx \tag{13.40}$$

raison pour laquelle les physiciens mélangent allègrement les deux notations...

Maintenant, il faut cependant savoir qu'il existe également des différentielles totales exactes, qu'aucune fonction ne vérifie. Dans ce cas, nous parlons de "différentielle totale inexacte" et pour déterminer si une différentielle totale est exacte ou inexacte, nous utilisons les propriétés des dérivées partielles.

Soit la forme différentielle :

$$dz = M(x,y)dx + N(x,y)dy \tag{13.41}$$

où $M(x,y)$ et $N(x,y)$ sont des fonctions des variables x et y. Si dz est une différentielle totale exacte, alors :

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy \tag{13.42}$$

Il faut donc que :

$$M(x,y) = \frac{\partial z}{\partial x} \text{ et } N(x,y) = \frac{\partial z}{\partial y} \tag{13.43}$$

ou encore, en effectuant une seconde dérivation, que :

$$\begin{aligned} &\frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 z}{\partial y \partial x} = \frac{\partial}{\partial y}\big(M(x,y)\big) = \frac{\partial M}{\partial y} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{et} \\ &\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}\big(N(x,y)\big) = \frac{\partial N}{\partial x} \end{aligned} \tag{13.44}$$

Avant de continuer, nous avons besoin d'un résultat donné par le "théorème de Schwarz" qui s'énonce de la manière suivante : Soit une fonction f, si :

$$f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial^2 f}{\partial x \partial y} \text{ et } \frac{\partial^2 f}{\partial y \partial x} \tag{13.45}$$

sont continues alors (il faut vraiment vraiment vérifier que ce soit le cas) :

$$\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) \tag{13.46}$$

pour tout $(x_0, y_0) \in U$.

Preuve Nous considérons l'expression :

$$H = f(x_0 + k, y_0 + h) - f(x_0 + k, y_0) - f(x_0, y_0 + h) + f(x_0, y_0) \tag{13.47}$$

Posons :

$$g(x) :== f(x, y_0 + h) - f(x, y_0) \text{ et } w(y) := f(x_0 + k, y) - f(x_0, y) \tag{13.48}$$

Nous avons alors :

$$H = g(x_0 + k) - g(x_0) = w(y_0 + h) - w(y_0) \tag{13.49}$$

Par le théorème des accroissements finis :

$$\begin{aligned} H &= g(x_0 + k) - g(x_0) = k \cdot g'(x_0 + sk) \\ H &= w(y_0 + h) - w(y_0) = h \cdot w'(y_0 + th) \end{aligned} \tag{13.50}$$

avec $s, t \in]0, 1[$. En reprenant les définitions de g et w nous obtenons :

$$\begin{aligned} H &= k \cdot g'(x_0 + sk) = k \cdot \left(\frac{\partial f}{\partial x}(x_0 + sk, y_0 + h) - \frac{\partial f}{\partial x}(x_0 + sk, y_0) \right) \\ H &= h \cdot w'(y_0 + th) = h \cdot \left(\frac{\partial f}{\partial y}(x_0 + k, y_0 + th) - \frac{\partial f}{\partial y}(x_0, y_0 + th) \right) \end{aligned} \tag{13.51}$$

en appliquant à nouveau le théorème des accroissements finis aux deux membres entre parenthèses nous trouvons :

$$\begin{aligned} H &= k \cdot h \cdot \frac{\partial^2 f}{\partial y \partial x}(x_0 + sk, y_0 + \tilde{s}h) \\ H &= h \cdot k \cdot \frac{\partial^2 f}{\partial x \partial y}(x_0 + \tilde{t}k, y_0 + th) \end{aligned} \tag{13.52}$$

avec $\tilde{s}, \tilde{t} \in]0, 1[$. Pour finir :

$$\frac{\partial^2 f}{\partial y \partial x}(x_0 + sk, y_0 + \tilde{s}h) = \frac{\partial^2 f}{\partial x \partial y}(x_0 + \tilde{t}k, y_0 + th) \tag{13.53}$$

et par continuité lorsque $k, h \to 0$, nous avons :

$$\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) \tag{13.54}$$

Plus simplement écrit :

$$\boxed{\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}} \tag{13.55}$$

•

Par récurrence sur le nombre de variables nous pouvons démontrer le cas général (c'est long mais c'est possible, nous le ferons si besoin il y a).

Donc finalement pour en revenir à notre problème initial, nous avons donc :

$$\frac{\partial^2 z}{\partial y \partial x} = \frac{\partial^2 z}{\partial y \partial z} \tag{13.56}$$

Ce qui nous donne finalement :

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial z} \tag{13.57}$$

C'est la condition que doit satisfaire une différentielle total pour être une différentielle totale exacte et la condition qu'elle ne doit pas satisfaire pour être une différentielle totale inexacte.

Afin de ne pas confondre les deux types de différentielles, nous utilisons le symbole ∂ pour représenter une différentielle totale inexacte et d pour une différentielle totale exacte. La distinction est extrêmement importante car seules les différentielles totale exactes ont une intégrale qui ne dépend que des bornes d'intégration (puisque toutes les variables changent en même temps) :

$$\int_{z_1}^{z_2} \partial z \neq z_2 - z_1 \text{ mais } \int_{z_1}^{z_2} dz = z_2 - z_1 = \Delta z \tag{13.58}$$

Autrement dit, la variation d'une fonction dont la différentielle est totale exacte, ne dépend pas du chemin suivi, mais uniquement des états initiaux et finaux. Nous appelons une telle fonction qui satisfait à une différentielle totale exacte, une "fonction d'état", c'est-à-dire une fonction dont la valeur ne dépend que de l'état présent et futur, et non de son histoire.

Cette distinction est très importante et particulièrement en thermodynamique où il convient de déterminer si une quantité physique est une différentielle totale exacte (une "fonction d'état" donc) ou non afin de savoir comment évoluent les systèmes.

Exemple 13.1 *Un exemple important de forme différentielle en thermodynamique, ∂W est le travail élémentaire d'une force exercée sur un corps en mouvement dans le plan Oxy, nous avons :*

$$\partial W = \overrightarrow{F} \circ d\overrightarrow{r} = F_n dx + F_y dy \tag{13.59}$$

F_x et F_y ne dérivent pas nécessairement d'un même potentiel $U(x,y)$ tel que :

$$\frac{\partial U}{\partial x} = \frac{\partial U}{\partial y} \tag{13.60}$$

13.1.2 Dérivées usuelles

Nous allons démontrer ici les dérivées les plus fréquentes (une petite trentaine) que nous puissions rencontrer en physique théorique et mathématique ainsi que certaines de leurs propriétés. La liste est pour l'instant non exhaustive mais les démonstrations étant généralisées, elles peuvent s'appliquer à un grand nombre d'autres cas (que nous appliquerons/rencontrerons tout au long de ce site).

Dérivée de x^n

Partons d'abord d'un cas particulier, la dérivée de x^3 :
Soit donc a un réel quelconque fixé, alors :

$$\begin{aligned}
\lim_{h\to 0} \frac{(a+h)^3 - a^3}{h} &= \lim_{h\to 0} \frac{a^3 + 3a^2h + 3ah^2 + h^3 - a^3}{h} \\
&= \lim_{h\to 0} \frac{3a^2h + 3ah^2 + h^3}{h} \\
&= \lim_{h\to 0} 3a^2 + 3ah + h^2 \\
&= 3a^2
\end{aligned} \tag{13.61}$$

Le nombre dérivé en a de la fonction cube est donc $f'(a) = 3a^2$.

Nous pouvons généraliser ce résultat pour tout entier naturel positif ou négatif n et nous allons voir que la fonction f définie sur $\mathbb{R}$ par $f(x) = x^n$ est dérivable et que sa dérivée f' est définie par $f'(x) = nx^{n-1}$.

$$\begin{aligned}
\lim_{h\to 0} \frac{(a+h)^n - a^n}{h} &= \lim_{h\to 0} \frac{\sum_{k=0}^{n} C_n^k a^k h^{n-k} - a^n}{h} \\
&= \lim_{h\to 0} \frac{\sum_{k=0}^{n-1} C_n^k a^k h^{n-k} + C_n^n a^n - a^n}{h} \\
&= \lim_{h\to 0} \frac{\sum_{k=0}^{n-1} C_n^k a^k h^{n-k} + a^n - a^n}{h} \\
&= \lim_{h\to 0} \frac{\sum_{k=0}^{n-1} C_n^k a^k h^{n-k}}{h} \\
&= \lim_{h\to 0} \sum_{k=0}^{n-1} C_n^k a^k h^{n-k-1} \\
&= \lim_{h\to 0} \left(\sum_{k=0}^{n-2} C_n^k a^k h^{n-k-1} + C_n^{n-1} a^{n-1} \right) \\
&= C_n^{n-1} a^{n-1} \\
&= \frac{n!}{(n-1)!} a^{n-1} \\
&= na^{n-1}
\end{aligned} \tag{13.62}$$

Ainsi, nous avons (quelques exemples peuvent êtres utiles pour comprendre la portée de ce résultat) :

$$\begin{aligned} f(x) = x^2 &\Rightarrow f'(x) = 2x^1 \\ f(x) = x^{-1} &\Rightarrow f'(x) = -x^{-2} \\ f(x) = \sqrt{x} &\Rightarrow f'(x) = \frac{1}{2}x^{-\frac{1}{2}} \end{aligned} \tag{13.63}$$

Nous voyons donc qu'en ayant déterminé la dérivée d'une fonction de la forme x^n, nous avons également déterminé la dérivée de toute fonction qui est mise sous cette forme tel que :

$$f(x) = \frac{1}{x^n} = x^{-n} \text{ et } f(x) = \sqrt[n]{x} = x^{\frac{1}{n}} \tag{13.64}$$

Cependant, les fonctions :

$$f(x) = \frac{1}{x^n} \text{ et } f(x) = \sqrt[n]{x} \tag{13.65}$$

ne sont pas dérivables en $a = 0$ puisque la fonction n'y est plus définie (division par zéro). De plus, en ce qui concerne la fonction comportant la racine (puissance non entière), la dérivée n'est pas définie dans $\mathbb{R}^*_-$.

Cependant, le résultat précédent donne un résultat intéressant pour les fonctions constantes telle que :

$$f(x) = c^{te} = c^{te}x^0 \tag{13.66}$$

il n'est alors pas difficile de déterminer la dérivée qui vaut simplement :

$$f'(x) = c^{te}0x^{0-1} \tag{13.67}$$

Donc la dérivée de toute fonction constante est nulle (il est important de se souvenir de ce résultat quand nous étudierons les propriétés des intégrales) ! ! !

Dérivée de $f(x) = \cos(x)$

Soit donc a un réel quelconque fixé, alors (attention ! il est utile de connaître les relations trigonométriques remarquables que nous démontrons dans le chapitre de trigonométrie dans la section de géométrie) :

$$\begin{aligned} \lim_{h\to 0} \frac{\cos(a+h) - \cos(a)}{h} &= \lim_{h\to 0} \frac{-2\sin\left(\frac{a+h+a}{2}\right) sin\left(\frac{a+h-a}{2}\right)}{h} \\ &= \lim_{h\to 0} \frac{-2\sin\left(\frac{2a+h}{2}\right) sin\left(\frac{h}{2}\right)}{h} \\ &= \lim_{h\to 0} \left(-2\sin\left(\frac{2a+h}{2}\right)\right) \cdot \frac{1}{2} \lim_{h\to 0} \left(\frac{sin\left(\frac{h}{2}\right)}{\frac{h}{2}}\right) \\ &= -2\sin(a)\frac{1}{2} \\ &= -\sin(a) \end{aligned} \tag{13.68}$$

Puisque :

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1 \tag{13.69}$$

Effectivement, rappelons que la fonction $sin(x)$ est assimilable (visuellement et mathématiquement) à une droite de fonction $f(x) = x$ au voisinage de $x \approx 0$.

Donc pour résumer :

$$f(x) = \cos(x) \Rightarrow f'(x) = -\sin(x) \tag{13.70}$$

Dérivée de $f(x) = \sin(x)$

Soit donc a un réel quelconque fixé, alors (attention ! il est utile de connaître les relations trigonométriques remarquables que nous démontrons dans le chapitre de trigonométrie dans la section de géométrie) :

$$\begin{aligned}
\lim_{h\to 0} \frac{\sin(a+h) - \sin(a)}{h} &= \lim_{h\to 0} \frac{2\cos\left(\frac{a+h+a}{2}\right) sin\left(\frac{a+h-a}{2}\right)}{h} \\
&= \lim_{h\to 0} \frac{2\cos\left(\frac{2a+h}{2}\right) sin\left(\frac{h}{2}\right)}{h} \\
&= \lim_{h\to 0} \left(2\cos\left(\frac{2a+h}{2}\right)\right) \cdot \frac{1}{2} \lim_{h\to 0} \left(\frac{sin\left(\frac{h}{2}\right)}{\frac{h}{2}}\right) \\
&= 2\cos(a)\frac{1}{2} \\
&= \cos(a)
\end{aligned} \tag{13.71}$$

Donc pour résumer :

$$f(x) = \sin(x) \Rightarrow f'(x) = \cos(x) \tag{13.72}$$

Dérivée de $f(x) = \log_a(x)$

La dérivée de la fonction $f(x) = \log_a(x)$ est égale à $x{-}1log_a(e)$, c'est-à-dire si :

$$y = f(x) = \log_a(x) \tag{13.73}$$

alors :

$$y' = f'(x) = \frac{1}{x}\log_a(e) \tag{13.74}$$

Preuve Si Δy est l'accroissement de la fonction $y = \log_a(x)$ pour un accroissement correspondant Δx de la variable x, alors :

$$\begin{aligned}
y + \Delta y &= \log_a(x + \Delta x) \\
\Delta y &= \log_a(x + \Delta x) - log_a(x) \\
&= \log_a \frac{x + \Delta x}{x} \\
&= log_a\left(1 + \frac{\Delta x}{x}\right)
\end{aligned} \tag{13.75}$$

et nous pouvons écrire :

$$\frac{\Delta y}{\Delta x} = \frac{1}{\Delta x}\log_a\left(1 + \frac{\Delta x}{x}\right) \tag{13.76}$$

Multiplions et divisons par x l'expression figurant dans le membre droit de la dernière égalité :

$$\frac{\Delta y}{\Delta x} = \frac{1}{x}\frac{x}{\Delta x}\log_a\left(1+\frac{\Delta x}{x}\right) = \frac{1}{x}\log_a\left(1+\frac{\Delta x}{x}\right)^{\frac{x}{\Delta x}} \tag{13.77}$$

Désignons la quantité $\frac{\Delta x}{x}$ par α. Il est évident que $\alpha \to 0$ quand Δx tend vers zéro pour un x donné. Par conséquent :

$$\frac{\Delta y}{\Delta x} = \frac{1}{\Delta x}\log_a(1+\alpha)^{\frac{1}{\alpha}} \tag{13.78}$$

Or, nous retrouvons ici une autre provenance historique de la constante d'Euler (cf. chapitre d'Analyse Fonctionnelle) où :

$$\lim \alpha \to 0(1+\alpha)^{\frac{1}{\alpha}} = e \tag{13.79}$$

Ainsi :

$$y' = \lim_{\Delta x \to 0}\frac{\Delta y}{\Delta x} = \lim_{\alpha \to 0}\frac{1}{x}\log_a(1+\alpha)^{\frac{1}{\alpha}} = \frac{1}{x}\log a(e) \tag{13.80}$$

•

Une cas particulier important est le cas où $a = e$. Nous avons alors :

$$y = \ln(x) \Rightarrow y' = \frac{1}{x} \tag{13.81}$$

Dérivée d'une somme de fonctions

Soient u et v deux fonctions. La fonction somme $s = u + v$ est dérivable sur tout intervalle où u et v sont dérivables, sa dérivée est la fonction s' somme des fonctions dérivées u' et v' de u et v.

Ce résultat se généralise pour une somme d'un nombre quelconque fixé de fonctions.

Preuve Soit a un réel fixé et u et v deux fonctions définies et dérivables en a :

$$\begin{aligned}\lim_{h\to 0}\frac{(u+v)(a+h)-(u+v)(a)}{h} &= \lim_{h\to 0}\frac{u(a+h)+v(a+h)-u(a)-v(a)}{h}\\ &= \lim_{h\to 0}\frac{u(a+h)-u(a)+v(a+h)-v(a)}{h}\\ &= \lim_{h\to 0}\frac{u(a+h)-u(a)}{h}+\lim_{h\to 0}\frac{v(a+h)-v(a)}{h}\\ &= u'(a)+v'(a)\end{aligned} \tag{13.82}$$

•

Donc la dérivée d'une somme est la somme des dérivées.

Dérivée d'un produit de fonctions

Soient u et v deux fonctions. La fonction produit $p = \cdot v$ est dérivable sur tout intervalle où u et v sont dérivables, sa dérivée première est la fonction p' telle que :

$$p' = u'v + uv' \tag{13.83}$$

Preuve Soit a un réel fixé et u et v deux fonctions définies et dérivables en a :

$$\lim_{h\to 0}\frac{(uv)(a+h)-(uv)(a)}{h} = \lim_{h\to 0}\frac{u(a+h)v(a+h)-u(a)v(a)}{h} \tag{13.84}$$

Nous rajoutons à cette dernière relation deux termes dont la somme est nulle tel que :

$$\begin{aligned}
&\lim_{h\to 0}\frac{u(a+h)v(a+h)-u(a)v(a)}{h}\\
&\quad=\lim_{h\to 0}\frac{u(a+h)v(a+h)-u(a)v(a+h)+u(a)v(a+h)-u(a)v(a)}{h}\\
&\quad=\lim_{h\to 0}\frac{u(a+h)v(a+h)-u(a)v(a+h)}{h}+\lim_{h\to 0}\frac{u(a)v(a+h)-u(a)v(a)}{h}\\
&\quad=\lim_{h\to 0}\left(\frac{u(a+h)-u(a)}{h}v(a+h)\right)+\lim_{h\to 0}\left(\frac{u(a)v(a+h)-u(a)v(a)}{h}u(a)\right)\\
&\quad=\lim_{h\to 0}\left(\frac{u(a+h)-u(a)}{h}\right)\lim_{h\to 0}v(a+h)+\lim_{h\to 0}\left(\frac{u(a)v(a+h)-u(a)v(a)}{h}\right)\\
&\quad=u'(a)v(a)+u(a)v'(a)
\end{aligned}\tag{13.85}$$

•

Mais il existe une formulation plus générale que la dérivée première d'un produit :

Considérons pour cela toujours nos deux fonctions u et v n fois dérivables sur un intervalle I. Alors le produit uv est n fois dérivable sur I et :

$$(uv)^{(n)}=\sum_{k=0}^{n}C_k^n u^{(k)}v^{(n-k)}\tag{13.86}$$

et ceci constitue la "formule de Leibniz" que nous avons utilisé dans le chapitre de calcul algébrique pour l'étude des polynômes de Legendre (qui nous sont eux-mêmes indispensables pour l'étude de la chimie quantique).

La démonstration de la formule est très proche de celle fait pour le binôme de Newton (cf. chapitre de Calcul Algébrique).

Preuve Soit :

$$(uv)^{(0)}=uv\tag{13.87}$$

D'autre part :

$$(uv)^{(0)}=\sum_{k=0}^{0}C_k^0 u^{(k)}v^{(0-k)}=u^{(0)}v^{(0)}=uv\tag{13.88}$$

La formule est ainsi bien initialisée.

La démonstration se fait par récurrence. Ainsi, le but est de montrer que pour $\forall n\geq 0\in\mathbb{N}$ que si :

$$(uv)^{(n)}=\sum_{k=0}^{n}C_k^n u^{(k)}v^{(n-k)}\tag{13.89}$$

alors :

$$(uv)^{(n+1)}=\sum_{k=0}^{n+1}C_k^{n+1} u^{(k)}v^{(n+1-k)}\tag{13.90}$$

Nous avons donc :

$$\begin{aligned}
(uv)^{(n+1)}&=\left((uv)^{(n)}\right)'=\sum_{k=0}^{n}C_k^n\left(u^{(k)}v^{(n-k)}\right)'\\
&=\sum_{k=0}^{n}C_k^n\left(\left(u^{(k)}\right)'v^{(n-k)}+u^{(k)}\left(v^{(n-k)}\right)'\right)'\\
&=\sum_{k=0}^{n}C_k^n\left(u^{(k+1)}v^{(n-k)}+u^{(k)}v^{(n+1-k)}\right)\\
&=\sum_{k=0}^{n}C_k^n u^{(k+1)}v^{(n-k)}+\sum_{k=0}^{n}C_k^n u^{(k)}v^{(n+1-k)}
\end{aligned}\tag{13.91}$$

Nous allons procéder à un changement de variable dans la première somme pour ne plus avoir le terme en $k+1$. Nous posons pour cela $j = k+1$:

$$\sum_{k=0}^{n} C_k^n u^{(k+1)} v^{(n-k)} = \sum_{j=1}^{n+1} C_{j-1}^n u^{(j)} v^{(n-(j-1))} \tag{13.92}$$

Si nous revenons à la lettre k, nous avons donc :

$$\sum_{k=0}^{n} C_k^n u^{(k+1)} v^{(n-k)} = \sum_{k=1}^{n+1} C_{k-1}^n u^{(k)} v^{(n+1-k))} \tag{13.93}$$

Nous avons donc :

$$(uv)^{(n+1)} = \sum_{k=1}^{n+1} C_{k-1}^n u^{(k)} v^{(n+1-k))} + \sum_{k=0}^{n} C_k^n u^{(k)} v^{(n+1-k))} \tag{13.94}$$

Nous voulons réunir les deux sommes. Pour cela, nous écartons les termes en trop dans chacun d'elles :

$$\begin{aligned}(uv)^{(n+1)} =& C_{n+1-1}^n u^{(n+1)} v^{(n+1-(n+1))} \\ &+ \sum_{k=1}^{n} C_{k-1}^n u^{(k)} v^{(n+1-k))} + \sum_{k=1}^{n} C_k^n u^{(k)} v^{(n+1-k))} \\ &+ C_0^n u^{(0)} v^{(n+1-0)}\end{aligned} \tag{13.95}$$

Ce qui donne donc :

$$\begin{aligned}(uv)^{(n+1)} &= u^{(n+1)} + \sum_{k=1}^{n} C_{k-1}^n u^{(k)} v^{(n+1-k))} + \sum_{k=1}^{n} C_k^n u^{(k)} v^{(n+1-k))} + v^{(n+1)} \\ &= u^{(n+1)} + \sum_{k=1}^{n} \left(C_{k-1}^n + C_k^n\right) u^{(k)} v^{(n+1-k))} + v^{(n+1)}\end{aligned} \tag{13.96}$$

D'après la formule de Pascal (cf. chapitre de Probabilités), nous avons :

$$C_{k-1}^n + C_k^n = C_k^{n+1} \tag{13.97}$$

Donc :

$$(uv)^{(n+1)} = u^{(n+1)} + \sum_{k=1}^{n} C_k^{n+1} u^{(k)} v^{(n+1-k))} + v^{(n+1)} \tag{13.98}$$

Or :

$$u^{(n+1)} = C_{n+1}^{n+1} u^{(n+1)} v^{(n+1-(n+1))} \text{ et } v^{(n+1)} = C_0^{n+1} u^{(0)} v^{(n+1-0)} \tag{13.99}$$

Donc :

$$(uv)^{(n+1)} = \sum_{k=0}^{n+1} C_k^{n+1} u^{(k)} v^{(n+1-k))} \tag{13.100}$$

•

Dérivée d'une fonction composée

Soit la fonction composée $f = g \circ u$ de deux fonctions u et g dérivables, la première en $u(x)$, la seconde en x, la fonction dérivée f' est définie par $f' = (g' \circ u)u'$, c'est-à-dire :

$$f'(x) = g'(u(x)) \cdot u'(x) \tag{13.101}$$

Preuve Soit a un réel fixé et u une fonction définie et dérivable en a et g une fonction définie et dérivable en $u(a)$:

$$\lim_{h\to 0} \frac{(g \circ u)(a+h) - (g \circ u)(a)}{h} = \lim_{h\to 0} \frac{g(u(a+h)) - g(u)(a))}{h} \tag{13.102}$$

posons $k = u(a+h) - u(a)$, nous avons alors :

$$\lim_{h\to 0} k = 0 \tag{13.103}$$

continuons notre développement précédent :

$$\begin{aligned}
\lim_{h\to 0} \frac{g(u(a+h)) - g(u)(a))}{h} &= \lim_{h\to 0} \frac{g(u(a)+k) - g(u)(a))}{h} = \lim_{h\to 0} \frac{g(u(a)+k) - g(u)(a))}{k} \frac{k}{h} \\
&= \lim_{h\to 0} \frac{g(u(a)+k) - g(u)(a))}{k} \lim_{h\to 0} \frac{k}{h} \\
&= \lim_{h\to 0} \frac{g(u(a)+k) - g(u)(a))}{k} \lim_{h\to 0} \frac{u(a+h) - u(a)}{h} \\
&= g'(u(a)) \cdot u'(a)
\end{aligned} \tag{13.104}$$

et nous retrouvons bien (13.101). •

Donc la dérivée d'une fonction composée est donnée par la dérivée de la fonction multipliée par la "dérivée intérieure". Par ailleurs, ce type de dérivation est très important car souvent utilisé en physique sous la dénomination de "dérivation en chaîne".

Voyons de quoi il s'agit. La dernière relation obtenu peut être écrite sous une autre forme si nous posons $y = f(g(x))$ et $u = g(x)$:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} \tag{13.105}$$

Ce qui peut s'étendre à des cas plus compliqués par exemple si $z = f(x, y), x = g(t), y = h(t)$ alors :

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} \tag{13.106}$$

Dérivée d'une fonction réciproque

Si la fonction f est continue, strictement monotone sur un intervalle I, dérivable sur I, alors la fonction réciproque f^{-1} est dérivable sur l'intervalle $f(I)$ et admet pour fonction dérivée :

$$\forall y \in f(I) \quad \left(f^{-1}\right)'(y) = \frac{1}{f'\left(f^{-1}(y)\right)} \tag{13.107}$$

En effet, nous pouvons écrire :

$$x \mapsto^{f} y = y = f(x) \mapsto^{f^{-1}} x = f^{-1}(y) \tag{13.108}$$

C'est-à-dire (application identité) :

$$x = f^{-1}(y) = f^{-1}(f(x)) = f^{-1} \circ f(x) \tag{13.109}$$

Par application de la dérivation des fonctions composées :

$$f^{-1}(f(x)) = f^{-1}(y)' \cdot f'(x) = (x)' = 1 \tag{13.110}$$

d'où :

$$f^{-1}(y)' = \frac{1}{f'(x)} = \frac{1}{f'\left(f^{-1}(y)\right)} \tag{13.111}$$

Pour une variable x, nous poserons pour la dérivée de la fonction réciproque :

$$f^{-1}(x)' = \frac{1}{f'(y)} \tag{13.112}$$

Dérivée de la fonction $\arccos(x)$

En utilisant le résultat précédent de la fonction réciproque, nous pouvons calculer la dérivée de la fonction $\arccos(x)$:

$$\begin{aligned} f^{-1}(x)' &= \big(\arccos(x)\big)' = \frac{1}{f'(y)} = \frac{1}{-\sin(y)} \\ &= \frac{-1}{\sin(y)} = \frac{-1}{\sqrt{1-\cos^2(y)}} \\ &= \frac{-1}{\sqrt{1-x^2}} \end{aligned} \tag{13.113}$$

Dérivée de la fonction $\arcsin(x)$

En utilisant le résultat précédent de la fonction réciproque, nous pouvons calculer la dérivée de la fonction $\arcsin(x)$:

$$\begin{aligned} f^{-1}(x)' &= \big(\arcsin(x)\big)' = \frac{1}{f'(y)} \\ &= \frac{1}{\cos(y)} = \frac{1}{\sqrt{1-\sin^2(y)}} \\ &= \frac{1}{\sqrt{1-x^2}} \end{aligned} \tag{13.114}$$

Dérivée d'un quotient de deux fonctions

La fonction $f = u/v$ est dérivable sur tout intervalle où les fonctions u et v sont dérivable et où la fonction v est non nulle et :

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2} \tag{13.115}$$

Preuve La fonction f peut être considérée comme le produit de deux fonctions : la fonction u et la fonction $1/v$. Une produit de deux fonctions est dérivable si chacune d'elle est dérivable, il faut donc que la fonction u soit dérivable et que la fonction $1/v$ soit également dérivable ce qui est le cas quand v est dérivable non nulle.

$$\begin{aligned} \left(\frac{u}{v}\right)' &= \left(u \cdot \frac{1}{v}\right)' = u' \cdot \frac{1}{v} + u \cdot \left(\frac{1}{v}\right)' \\ &= \frac{u'}{v} + u \cdot \left(\frac{-v'}{v^2}\right) = \frac{u'v - uv'}{v^2} \end{aligned} \tag{13.116}$$

•

Dérivée de la fonction $\tan(x)$

Par définition (cf. chapitre de Trigonométrie) nous avons :

$$f(x) = \tan(x) = \frac{\sin(x)}{\cos(x)} \tag{13.117}$$

et en appliquant donc la dérivée d'un quotient vu précédemment, nous avons :

$$f'(x) = \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)} \tag{13.118}$$

ou encore :

$$\begin{aligned} f'(x) &= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)} \\ &= 1 + \frac{\sin^2(x)}{\cos^2(x)} = 1 + \tan^2(x) \end{aligned} \tag{13.119}$$

Dérivée de la fonction $\cot(x)$

Par définition (cf. chapitre de Trigonométrie), $\forall x \neq k \cdot \pi (k \in \mathbb{Z})$:

$$f(x) = \cot(x) = \frac{1}{\tan(x)} \tag{13.120}$$

et donc (dérivée d'un quotient à nouveau) :

$$f'(x) = -\frac{1 + \tan^2(x)}{\tan^2(x)} = -\frac{1}{\tan^2(x)} - 1 = -(\cot^2(x) + 1) \tag{13.121}$$

ou encore :

$$\begin{aligned} \cot'(x) &= -\frac{1 + \tan^2(x)}{\tan^2(x)} = -\frac{1}{\tan^2(x)} - 1 = -\frac{\cos^2(x)}{\sin^2(x)} - 1 \\ &= -\frac{\cos^2(x) + sin^2(x)}{\sin^2(x)} = -\frac{1}{\sin^2(x)} \end{aligned} \tag{13.122}$$

Dérivée de la fonction $\arctan(x)$

Nous utilisons les propriétés dérivée des fonctions réciproques :

$$\forall x \in \mathbb{R}\ f'(x) = \frac{1}{1 + \tan^2(\arctan(x))} = \frac{1}{1 + x^2} \tag{13.123}$$

Dérivée de la fonction $arccot(x)$

Selon la même méthode que précédemment :

$$\forall x \in \mathbb{R}\ f'(x) = \frac{1}{1 + \cot^2(arccot(x))} = -\frac{1}{1 + x^2} \tag{13.124}$$

Dérivée de la fonction e^x

Nous verrons lors de notre études des méthodes numérique (cf. chapitre de Méthodes Numériques) que le "nombre d'Euler" peut être calculé selon la série :

$$e^x = \sum_{k=0}^{+\infty} \frac{x^k}{k!} \tag{13.125}$$

qui converge sur $\mathbb{R}$. En dérivant terme à terme cette série qui converge, il vient :

$$\forall x \in \mathbb{R}\ f'(x) = \sum_{k=0}^{+\infty} k \cdot \frac{x^{k-1}}{k!} = \sum_{k=1}^{+\infty} \frac{x^{k-1}}{(k-1)!} = e^x \tag{13.126}$$

Ainsi l'exponentielle est sa propre dérivée. Nous pouvons nous permettre d'étudier les dérivées de quelques fonction trigonométriques hyperboliques (cf. chapitre de Trigonométrie).

Dérivée de la fonction $\sinh(x)$

Rappel :

$$\sinh(x) = \frac{1}{2}\left(e^x - e^{-x}\right) \tag{13.127}$$

Donc trivialement :

$$\sinh'(x) = \frac{1}{2}\left(e^x + e^{-x}\right) = \cosh(x) \tag{13.128}$$

Dérivée de la fonction $\cosh(x)$

Rappel :

$$\cosh(x) = \frac{1}{2}\left(e^x + e^{-x}\right) \tag{13.129}$$

Donc trivialement :

$$\cosh'(x) = \frac{1}{2}\left(e^x - e^{-x}\right) = \sinh(x) \tag{13.130}$$

Dérivée de la fonction $\tanh(x)$

Puisque par définition :

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} \tag{13.131}$$

Donc en appliquant la dérivée d'un quotient nous obtenons :

$$\tanh'(x) = \frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)} = \frac{1}{\cosh^2(x)} \tag{13.132}$$

Ou encore :

$$\tanh'(x) = \frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)} = 1 - \frac{\sinh^2(x)}{\cosh^2(x)} = 1 - \tanh^2(x) \tag{13.133}$$

Dérivée de la fonction $\coth(x)$

Rappel :

$$\forall x \neq 0,\ \coth(x) = \frac{1}{\tanh(x)} \tag{13.134}$$

et donc :

$$\coth'(x) = -\frac{1 - \tanh^2(x)}{\tanh^2(x)} = 1 - \frac{1}{\tanh^2(x)} = 1 - \coth^2(x) \tag{13.135}$$

Dérivée de la fonction $arcsinh(x)$

Nous appliquons les propriétés des dérivées des fonctions réciproques :

$$arcsinh'(x) = \frac{1}{\cosh(arcsinh(x))} \tag{13.136}$$

Or (voir à nouveau le chapitre de trigonométrie) :

$$1 = \cosh^2(arcsinh(x)) - \sinh^2(arcsinh(x)) = \cosh^2(arcsinh(x)) - x^2 \tag{13.137}$$

et donc :

$$1 + x^2 = \cosh^2(arcsinh(x)) \tag{13.138}$$

Etant donné que cosh ne prend que des valeurs positives, nous avons :

$$\cosh(arcsinh(x)) = \sqrt{1 + x^2} \tag{13.139}$$

Donc finalement :

$$arcsinh(x)' = \frac{1}{\sqrt{1 + x^2}} \tag{13.140}$$

Dérivée de la fonction $arccosh(x)$

Nous appliquons les propriétés des dérivées des fonctions réciproques :

$$\forall x > 1, arccosh'(x) = \frac{1}{\sinh(arccosh(x))} \tag{13.141}$$

Or selon la même méthode que précédemment :

$$1 = \cosh^2(arccosh(x)) - \sinh^2(arccosh(x)) = x^2 - \sinh^2(arccosh(x)) \tag{13.142}$$

d'où :

$$-1 + x^2 = \sinh^2(arccosh(x)) \tag{13.143}$$

Etant donné que ne prend que des valeurs positives nous avons alors :

$$\sinh(arccosh(x)) = \sqrt{x^2 - 1} \tag{13.144}$$

Donc :

$$\forall x > 1, arccosh(x)' = \frac{1}{\sqrt{x^2 - 1}} \tag{13.145}$$

Dérivée de la fonction $arctanh(x)$

En appliquant les propriétés des dérivées des fonctions réciproques) :

$$\forall x \in]-1, 1[,\ arctanh(x)' = \frac{1}{1 - \tanh^2(arctanh(x))} = \frac{1}{1 - x^2} \tag{13.146}$$

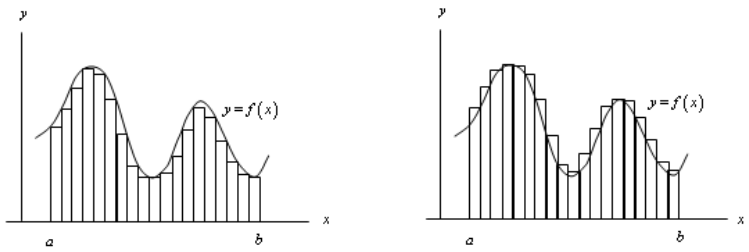


FIGURE 13.2 – Représentation graphique de l'intégrale

Dérivée de la fonction $arccoth(x)$

En appliquant les propriétés des dérivées des fonctions réciproques) si $|x| > 1$:

$$arcoth(x)' = \frac{1}{1 - \coth^2(arcoth(x))} = \frac{1}{1 - x^2} \tag{13.147}$$

Dérivée de la fonction a^x

Avec $a > 0$:

$$a^x = \exp(\ln(a^x)) = \exp(x \cdot \ln(a)) \tag{13.148}$$

Donc (dérivée d'une fonction composée) :

$$f'(x) = \exp(x \cdot \ln(a)) \cdot \ln(a) = a^x \cdot \ln(a) \tag{13.149}$$

13.1.3 Calcul intégral

Nous allons aborder ici les principes élémentaires et de base du calcul intégral. La suite (avec plus de rigueur) viendra en fonction du temps qui est la disposition des responsables du site.

Intégrale définie

Une valeur approchée de l'aire sous une courbe peut être obtenue par un découpage en n bandes rectangulaires verticales de même largeur. En particulier on peut réaliser un encadrement de cette aire à l'aide d'une somme majorante A_M et d'une somme minorante A_m pour un découpage donné comme présenté à la fig. 13.2.

Supposons que le nombre n de bandes tende vers l'infini. Comme les bandes sont de même largeur, la largeur de chaque bande tend vers 0.

Si les sommes AM et Am ont toutes deux une limite lorsque, le nombre n de bandes, tend vers l'infini, alors l'aire A sous la courbe est comprise entre ces deux limites.

Nous avons :

$$\lim_{n\to\infty} A_m \leq A \lim_{n\to\infty} A_M \tag{13.150}$$

Si ces deux limites sont égales, leur valeur est celle de l'aire sous la courbe.

D'où une première définition de l'intégrale définie ou dite "intégrale de Riemann" :

Soit un intervalle $[a,b]$, divisé en n parties égales, soit f une fonction continue sur l'intervalle $[a,b]$, soit A_m, la somme algébrique minorante et soit A_M, la somme algébrique majorante. Nous appelons "intégrale définie" de f, depuis a jusqu'à b, notée :

$$\int_a^b f(x)dx \tag{13.151}$$

le nombre A tel que $A = \lim_{n\to\infty} A_m = \lim_{n\to\infty} A_M$ pourvu que cette limite existe.
Intuitivement, il est évident que lorsque , nous étendons la définition ainsi :

$$\int_a^a f(x)dx \tag{13.152}$$

Remarque Pour calculer l'aire majorante et l'aire minorante, il n'est pas nécessaire que la largeur des sous-intervalles du découpage soit la même partout.

Remarque Si cette limite existe, alors nous disons que f est "intégrable" sur $[a,b]$ et l'intégrale définie $\int_a^b f(x)dx$ existe.

Remarque Le fait de chercher cette limite s'appelle "calculer l'intégrale".

Remarque Les nombres a et b sont appelés les "bornes d'intégration", a est la "borne inférieure", b est la "borne supérieure".

Remarque D'autres lettres que x peuvent être employées dans la notation de l'intégrale définie. Ainsi si f est intégrable sur $[a,b]$, alors $\int_a^b f(x)dx = \int_a^b f(t)dt = \int_a^b f(s)ds$ etc. C'est la raison pour laquelle la variable x de la définition est dite "variable muette".

Remarque Comme nous le verrons plus loin, il est essentiel de ne pas confondre "intégrale définie" et "intégrale indéfinie". Ainsi, une intégrale indéfinie, notée $\int f(x)dx$ est une fonction, ou, plus précisément, une famille de fonctions appelées aussi "primitives de f" (voir plus bas) alors qu'une intégrale définie, notée $\int_a^b f(x)dx$ est une constante.

Intégrale indéfinie

Nous avons vu précédemment lors de notre études des dérivées, le problème suivant : étant donnée une fonction $F(x)$, trouver sa dérivée, c'est-à-dire la fonction .

Définition : Nous disons que la fonction $F(x)$ est une "primitive" ou "intégrale indéfinie" de la fonction $f(x)$ sur le segment $[a,b]$, si en tout point de ce segment nous avons l'égalité $F'(x) = f(x)$.

Une autre manière de voire le concept d'intégrale indéfinie est de passer par le théorème fondamental du calcul intégral (et différentiel) appelé aussi parfois "théorème fondamental de l'analyse" qui s'énonce ainsi :

Soit f une fonction continue sur un intervalle fermé $[a,b]$.

P1. Si A est la fonction définie par $A(X) = \int_a^X f(t)dt$ pour tout X dans $[a,b]$, alors A est la primitive de f sur $[a,b]$ qui s'annule en a.

P2. Si F est une primitive de f sur $[a,b]$, alors $\int_a^b f(t)dt = F(b) - F(a)$.

Preuve P1. Soit la fonction :

$$A(X) = \int_a^X f(t)dt \tag{13.153}$$

Si f est positive et $h > 0$ (la démonstration dans le cas où $h < 0$ est proposée similaire) et comme $X > a$, nous pouvons nous représenter $A(X)$ comme l'aire sous la courbe de f depuis $t = a$ jusqu'à $t = X$ (cfr fig 13.3.

Pour démontrer que A est une primitive de f , nous allons prouver que $A' = f$. Selon la définition de la dérivée :

$$A'(X) = \lim_{h\to 0} \frac{A(X+h) - A(X)}{h} \tag{13.154}$$

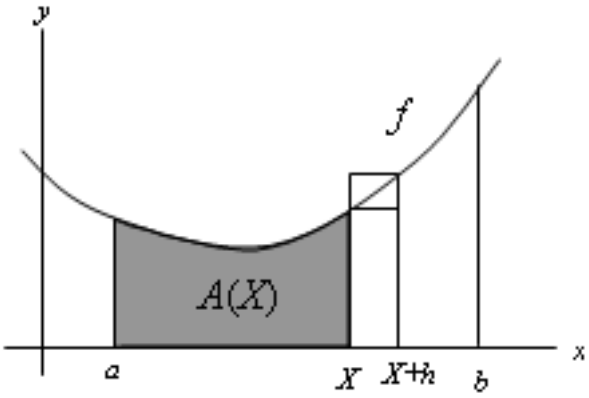


FIGURE 13.3 – $\int_a^X f(t)dt$

Etudions ce quotient $A(X+h)-A(X)$: est représentée par l'aire de la bande de largeur h, prise en sandwich entre deux rectangles de largeur h.

Soit M le maximum de f sur l'intervalle $[X, X+h]$ et m le minimum de f sur ce même intervalle. Les aires respectives des deux rectangles sont Mh et mh.

Nous avons alors la double inégalité suivante :

$$mh \leq A(X+h)-A(X) \leq Mh \tag{13.155}$$

Comme h est positif, on peut diviser par h sans changer le sens des inégalités :

$$m \leq \frac{A(X+h)-A(X)}{h} \leq M \tag{13.156}$$

Lorsque $h \to 0^+$ et si f est une fonction continue, alors M et m ont pour limite $f(x)$, et le rapport $\frac{A(X+h)-A(X)}{h}$, qui est compris entre m et M, a bien pour limite $f(x)$.

Comme $A'(X) = f(X)$ pour tout X, ceci nous montre que la dérivée de la fonction aire est f. Ainsi A est une primitive de f. Comme $A(a) = 0$, est bien la primitive de f qui s'annule en a. •

Avant de commencer la démonstration de la deuxième propriété du théorème fondamental, donnons et démontrons le théorème suivant qui va nous être indispensable : Si $F_1(x)$ et $F_2(x)$ sont deux primitives de la fonction $f(x)$ sur le segment $[a,b]$, leur différence est une constante (ce théorème est très important en physique pour ce qui est de l'étude de ce que nous appelons les "conditions initiales").

Preuve Nous avons en vertu de la définition de la primitive :

$$\begin{aligned} F_1'(x) &= f(x) \\ F_2'(x) &= f(x) \end{aligned} \tag{13.157}$$

pour $\forall x \in [a,b]$.

Posons :

$$F_1(x) - F_2(x) = \varphi(x) \tag{13.158}$$

Nous pouvons écrire :

$$F_1'(x) - F_2'(x) = \big[F_1(x) - F_2(x)\big]' = \varphi'(x) = f(x) - f(x) = 0 \tag{13.159}$$

Il vient donc de ce que nous avons vu pendant notre étude des dérivées que $\varphi(x) = c^{te}$.

Nous avons alors :

$$F_1(x) - F_2(x) = c^{te} \tag{13.160}$$

•

Il résulte de ce théorème que si nous connaissons une primitive quelconque $F(x)$ de la fonction $f(x)$, toute autre primitive de cette fonction sera de la forme :

$$F(x) + c^{te} \tag{13.161}$$

Donc finalement, nous appelons "intégrale indéfinie" de la fonction $f(x)$ et nous notons :

$$\int f(x)dx \tag{13.162}$$

tout expression de la forme $F(x)+c^{te}$ où $F(x)$ est une primitive de $f(x)$. Ainsi, par convention d'écriture :

$$\boxed{\int f(x)dx = F(x) + c^{te}} \tag{13.163}$$

si et seulement si $F'(x) = f(x)$.

Dans ce contexte, $f(x)$ est également appelée "fonction à intégrer" et $f(x)dx$, "fonction sous le signe somme".

Géométriquement, nous pouvons considérer l'intégrale indéfinie comme un ensemble (famille) de courbes telles que nous passons de l'une à l'autre en effectuant une translation dans le sens positif ou négatif de l'axe des ordonnés.

Revenons-en à la démonstration du point (2) du théorème fondamental de l'analyse :

Preuve Soit F une primitive de f.

Puisque deux primitives diffèrent d'une constante, nous avons bien :

$$A(X) = F(X) + c^{te} \tag{13.164}$$

ce que nous pouvons écrire aussi :

$$\int_a^X f(t)dt = F(X) + c^{te} \tag{13.165}$$

pour tout X dans $[a, b]$. Le cas particulier $X = a$ donne $\int_a^a f(t)dt = 0$ et donc $F(a) + c^{te} = 0$ et $c^{te} = -F(a)$. En remplaçant, nous obtenons :

$$\int_a^X f(t)dt = F(X) - F(a) \tag{13.166}$$

Comme cette identité est valable pour tout X de l'intervalle $[a, b]$, elle est vraie en particulier pour $X = b$. D'où :

$$\boxed{\int_a^b f(t)dt = F(b) - F(a)} \tag{13.167}$$

•

Remarque Le théorème fondamental qui montre le lien entre primitive et intégrale a conduit à utiliser le même symbole $\int$ pour écrire une primitive, qui est une fonction, et une intégrale, qui elle, est un nombre.

Voici quelques propriétés triviales de l'intégration qu'il est bon de se rappeler car souvent utilisée ailleurs sur le site (si cela ne vous semble pas évident, contactez-nous et nous le détaillerons) :

P1. La dérivée d'une intégrale indéfinie est égale à la fonction à intégrer :

$$\left(\int f(x)dx\right)' = \left(F(x) + c^{te}\right)' = f(x) \tag{13.168}$$

P2. La différentielle d'une intégrale indéfinie est égale à l'expression sous le signe somme :

$$d\left(\int f(x)dx\right) = f(x) \tag{13.169}$$

P3. L'intégrale indéfinie de la différentielle d'une certaine fonction est égale à la somme de cette fonction et d'une constante arbitraire :

$$\int F(x)dx = F(x) \tag{13.170}$$

P4. L'intégrale indéfinie de la somme (ou soustraction) algébrique de deux ou plusieurs fonctions est égale à la somme algébrique de leurs intégrales (ne pas oublier que l'on travaille avec l'ensemble des primitives et non des primitives particulières !) :

$$\int [f_1(x) + f_2(x)]\, dx = \int f_1(x)dx + \int f_2(x)dx \tag{13.171}$$

Preuve Pour démontrer cela nous allons prouver que la dérivée du membre de gauche permet de trouver le membre de droit et inversement (réciproque) à l'aides des propriétés précédentes.

D'après P1 nous avons :

$$\left(\int [f_1(x) + f_2(x)]\, dx\right)' = f_1(x) + f_2(x) \tag{13.172}$$

Vérifions s'il en est de même avec le membre de droite (nous supposons connues les propriétés des dérivées que nous avons démontrées au début de ce chapitre) :

$$\begin{aligned}\left(\int f_1(x)dx + \int f_2(x)dx\right)' &= \left(\int f_1(x)dx\right)' + \left(\int f_2(x)dx\right)' \\ &= f_1(x) + f_2(x)\end{aligned} \tag{13.173}$$

•

P5. Nous pouvons sortir un facteur constant de sous le signe somme, c'est-à-dire :

$$\int c^{te} f(x)dx = c^{te} \int f(x)dx \tag{13.174}$$

Nous justifions cette égalité en dérivant les deux membres (et d'après les propriétés des dérivées) :

$$\begin{aligned}&\left(\int c^{te} f(x)dx\right)' = c^{te} f(x) \\ &\left(c^{te} \int f(x)dx\right)' = c^{te} \left(\int f(x)dx\right)' = c^{te} f(x)\end{aligned} \tag{13.175}$$

P6. Nous pouvons sortir un facteur constant de l'argument de la fonction intégrée (plutôt rarement utilisée) :

$$\int f(ax)dx = \frac{1}{a}F(ax) + c^{te} \tag{13.176}$$

En effet, en dérivant les deux membres de l'égalité nous avons d'après les propriétés des dérivées :

$$\left(\int f(ax)dx\right)' = \left(\frac{1}{a}F(ax) + c^{te}\right)' = f(ax) \tag{13.177}$$

P7. L'intégration d'une fonction dont l'argument est sommé (ou soustrait) algébriquement est la primitive de l'argument sommé (respectivement soustrait) :

$$\int f(x+b)dx = F(x+b) + c^{te} \tag{13.178}$$

Cette propriété ce démontre également identiquement à la précédente à l'aide des propriétés des dérivées.

P8. La combinaison des propriétés P6 et P7 nous permettent d'écrire :

$$\int f(ax+b)dx = \frac{1}{a}F(ax+b) + c^{te} \tag{13.179}$$

P9. Soit f une fonction continue sur $[a,b]$, nous avons pour $\forall c \cup [a,b]$:

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \tag{13.180}$$

Ce théorème découle immédiatement de la définition de l'intégrale indéfinie. F étant une primitive de f sur $[a,b]$ nous avons :

$$F(c) - F(a) + F(b) - F(c) = F(b) - F(a) \tag{13.181}$$

P10. Voilà une propriété souvent utilisée dans le chapitre de Statistiques du site (nous ne trouvons pas de moyen d'exprimer cette propriété par le langage courant donc...) :

$$\begin{aligned}\frac{d}{dy}\int_{g(x)}^{h(y)} f(t)dt &= \frac{d}{dy}\left(F(h(y)) - F(g(x))\right) \\ &= f(h(y))\frac{dh(y)}{dy} + f(g(x))\frac{dg(x)}{dy} \\ &= f(h(y))h'(y)\end{aligned} \tag{13.182}$$

13.1.4 Intégration par changements de variables

Lorsque nous ne pouvons facilement déterminer la primitive d'une fonction donnée, nous pouvons nous débrouiller par un changement de variable astucieux (parfois même très subtile) à contourner la difficulté. Cela ne marche pas à tous les coups (car certaines fonctions ne sont pas intégrables formellement) mais il vaut la peine d'essayer avant d'avoir recours à l'ordinateur.

A nouveau, nous ne donnons que la forme générale de la méthode. C'est le rôle des professeurs dans les écoles d'entraîner les élèves à comprendre et maîtriser ce genre de techniques. De plus, les chapitres traitant des sciences exactes sur le site (physique, informatique, astrophysique, chimie, ...) regorgent d'exemples utilisant cette technique et servent ainsi implicitement d'exercices de style.

Soit à calculer l'intégrale (non bornée pour l'instant) :

$$\int f(x)dx \tag{13.183}$$

bien que nous ne sachions pas calculer directement la primitive de cette fonction $f(x)$ (en tout cas nous imaginons être dans une telle situation) nous savons (d'une manière ou d'une autre) qu'elle existe (nous ne traitons pas encore des intégrales impropres à ce niveau).

La technique consiste alors dans cette intégrale à effectuer le changement de variable :

$$x = \varphi(t) \tag{13.184}$$

où $\varphi(t)$ est une fonction continue ainsi que sa dérivée, et admettant une fonction inverse. Alors $dx = \varphi'(t)$, démontrons que dans ce cas l'égalité :

$$\int f(x)dx = \int f(\varphi(t))\varphi'(t)dt \tag{13.185}$$

est satisfaite.

Preuve Nous sous-entendons ici que la variable t sera remplacée après intégration du membre droit par son expression en fonction de x. Pour justifier l'égalité en ce sens, il suffit de montrer que les deux quantités considérées dont chacune n'est définie qu'à une constant arbitraire près ont la même dérivée par rapport à x. La dérivée du membre gauche est :

$$\left(\int f(x)dx\right)' = f(x) \tag{13.186}$$

Nous dérivons le membre droit par rapport à x en tenant compte que t est une fonction de x. Nous savons que :

$$\frac{dx}{dt} = \varphi'(t) \tag{13.187}$$

Nous avons par conséquent :

$$\begin{aligned}\frac{d}{dx}\left(\int f(\varphi(t))\varphi'(t)dt\right) &= \frac{d}{dt}\left(\int f(\varphi(t))\varphi'(t)dt\right)\frac{dt}{dx}\\ &= f(\varphi(t))\varphi'(t)\frac{1}{\varphi'(t)}\\ &= f(\varphi(t))\\ &= f(x)\end{aligned} \tag{13.188}$$

Les dérivées par rapport à x des deux membres de l'égalité de départ sont donc égales. •

Bien évidemment, la fonction $x = \varphi(x)$ doit être choisie de manière à ce que nous sachions calculer l'intégrale indéfinie figurant à droite de l'égalité.

REMARQUE Il est parfois préférable de choisir le changement de variable sous la forme $t = \varphi(t)$ au lieu de $x = \psi(t)$ car cela à une large tendance à simplifier la longueur de l'équation au lieu de l'allonger.

Jacobien

Considérons un domaine D du plan u, v limité par une courbe L. Supposons que les coordonnées x, y soient des fonctions des nouvelles variables u,v (toujours dans le cadre d'un changement de variables donc) par les relations de transformations :

$$\begin{aligned}&x = \varphi(u,v) \quad y = \phi(u,v)\\ &(x,y) \in D \xrightarrow[\varphi,\phi]{(} \varphi,\phi) \in D'\end{aligned} \tag{13.189}$$

où les fonctions $\varphi(u,v)$ et $\phi(u,v)$ sont univoques, continues et possèdent des dérivées continues dans un certain domaine D' que nous définirons par la suite. Il correspond alors d'après les relations précédentes à tout couple de valeurs u, v un seul couple de valeur x, y et réciproquement.

Il résulte de ce qui précède qu'à tout point du plan Oxy correspond univoquement un point $P(u,v)j$ du plan Ouv de coordonnées u, v définies par les relations précédentes. Les nombres v et $u$eront appelées "coordonnées curvilignes" de P et nous verrons des exemples concrets et schématisé de ceux-ci dans le chapitre de calcul vectoriel.

Si dans le plan Oxy le point P décrit la courbe fermée L délimitant le domaine D, le point correspondant décrit dans le plan Ouv un certain domaine D'. Il correspond alors à tout point de

D' un point de D. Ainsi, les relations de transformations établissent une correspondance biunivoque entre les points des domaines D et D'.

Considérons maintenant dans D' une droite $u = c^{te}$. En général, les relations de transformation lui font correspondre dans le plan Oxy une ligne courbe (ou inversement). Ainsi, découpons le domaine D' par des droites $u = c^{te}$ et $v = c^{te}$ en de petits domaines rectangulaires (nous ne prendrons pas en compte dans la limite, les rectangles empiétant sur la frontière de D'). Les courbes correspondantes du domaine D découpent alors ce dernier en quadrilatère (définis par des courbes donc). Evidemment, l'inverse est applicable.

Considérons dans le plan Ouv le rectangle $\Delta s'$ limité par les droites :

$$\begin{aligned} &u = c^{te}, u + \Delta u = c^{te} \\ &v = c^{te}, v + \Delta v = c^{te} \end{aligned} \tag{13.190}$$

et le quadrilatère curviligne correspondant Δs dans le plan Oxy. Nous désignerons les aires de ces domaines partiels également par $\Delta s'$ et Δs. Nous avons évidemment :

$$\Delta s' = \Delta u \Delta v \tag{13.191}$$

Les aires $\Delta s'$ et Δs peuvent êtres en générales différentes.

Supposons donne dans D une fonction continue $z = f(x, y)$. Il correspond à toute valeur de cette fonction du domaine D la même valeur $z = F(u, v)$(ce qu'il faut vérifier) dans D', où :

$$F(u, v) = f(x, y) = f\big(\varphi(u, v), \phi(u, v)\big) \tag{13.192}$$

Considérons les sommes intégrales de la fonction z dans le domaine D. Nous avons évidemment l'égalité suivante :

$$F(u, v) = f(x, y) = f\big(\varphi(u, v), \phi(u, v)\big) \tag{13.193}$$

Calculons Δs, c'est-à-dire l'aire du quadrilatère curviligne P_1, P_2, P_3, P_4 dans le plan Oxy :

Déterminons les coordonnes de ses sommets :

$$\begin{array}{lll} P_1(x_1, y_1) & x_1 = \varphi(u, v) & y_1 = \phi(u, v) \\ P_2(x_2, y_2) & x_2 = \varphi(u + \Delta u, v) & y_2 = \phi(u + \Delta u, v) \\ P_2(x_3, y_3) & x_3 = \varphi(u + \Delta u, v + \Delta v) & y_3 = \phi(u + \Delta u, v + \Delta v) \\ P_2(x_4, y_4) & x_4 = \varphi(u, v + \Delta v) & y_4 = \phi(u, v + \Delta v) \end{array} \tag{13.194}$$

Nous assimilerons dans le calcul de l'aire du quadrilatère $P_1P_2P_3P_4$ les arcs $P_1P_2, P_2P_3, P_3P_4, P_4P_1$ à des segments de droites parallèles. Nous remplacerons en outre les accroissement des fonctions par leurs différentielles. C'est dire que nous faisons abstraction des infiniment petits d'ordre plus élevé que Δu et Δv. Les relations précédentes deviennent alors :

$$\begin{aligned} x_1 &= \varphi(u, v) & y_1 &= \phi(u, v) \\ x_2 &= \varphi(u, v) + \frac{\partial \varphi}{\partial u}\Delta u & y_2 &= \phi(u, v) + \frac{\partial \phi}{\partial u}\Delta u \\ x_3 &= \varphi(u, v) + \frac{\partial \varphi}{\partial u}\Delta u + \frac{\partial \varphi}{\partial v}\Delta v & y_3 &= \phi(u, v) + \frac{\partial \phi}{\partial u}\Delta u + \frac{\partial \phi}{\partial v}\Delta v \\ x_4 &= \varphi(u, v) + \frac{\partial \varphi}{\partial v}\Delta v & y_4 &= \phi(u, v) + \frac{\partial \phi}{\partial v}\Delta v \end{aligned} \tag{13.195}$$

Sous ces hypothèses, le quadrilatère curviligne $P_1P_2P_3P_4$ peut être assimilé à un parallélogramme. Son aire Δs est approximativement égale au double de l'aire du triangle $P_1P_2P_3$, aire que nous pouvons calculer en utilisant les propriétés du déterminant (comme nous le démontrerons

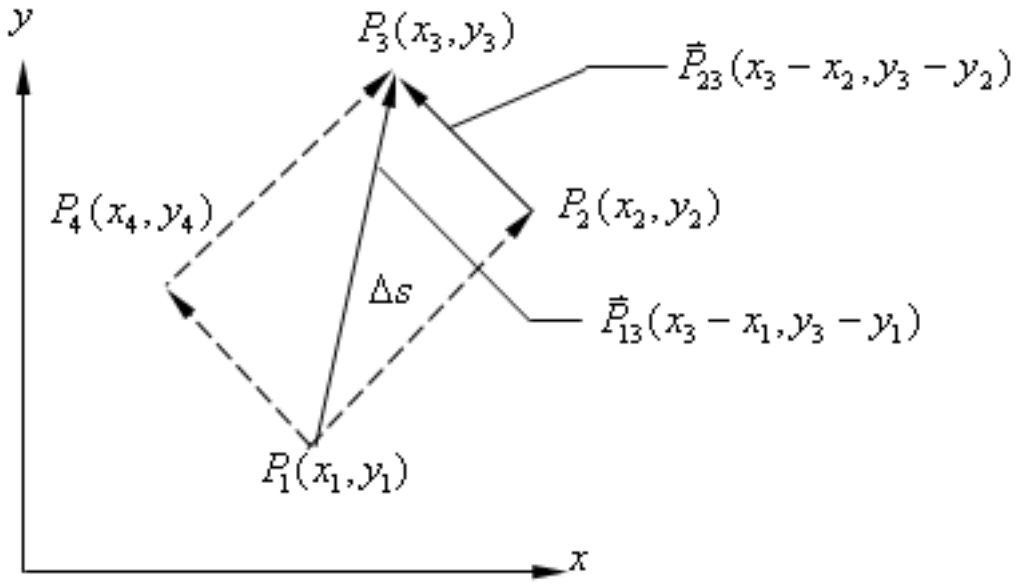


FIGURE 13.4

dans le chapitre d'algèbre linéaire, le déterminant dans $\mathbb{R}^2$ représente un parallélogramme alors que dans celui-ci représente le volume d'un parallélépipède) :

$$\begin{aligned}\Delta s \cong |det(\vec{v_1},\vec{v_2})| &= |det(M)| \\ &= \left|det\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}\right| = |a_{11}a_{22}-a_{21}a_{12}|\end{aligned} \tag{13.196}$$

Tel que (c'est là qu'il faut faire le meilleur choix pour que l'expression final soit la plus simple et la plus esthétique, nous procédons par essais successifs et faisons enfin le choix comme à la fig. 13.4) :

Ainsi, nous avons :

$$\begin{aligned}\Delta s \cong |det(\vec{v_1},\vec{v_2})| &= |D(\vec{P}_{13},\vec{P}_{23})| \\ &= \left|det\begin{bmatrix}(x_3-x_1) & (x_3-x_2)\\ (y_3-y_1) & (y_3-y_2)\end{bmatrix}\right| \\ &= |(x_3-x_1)(y_3-y_2)-(x_3-x_2)(y_3-y_2)| \\ &= \left|\left(\frac{\partial\varphi}{\partial u}\Delta u+\frac{\partial\varphi}{\partial v}\Delta v\right)\frac{\partial\phi}{\partial v}\Delta v-\left(\frac{\partial\phi}{\partial u}\Delta u+\frac{\partial\phi}{\partial v}\Delta v\right)\frac{\partial\varphi}{\partial v}\Delta v\right| \\ &= \left|\frac{\partial\varphi}{\partial u}\frac{\partial\phi}{\partial v}\Delta u\Delta v-\frac{\partial\varphi}{\partial v}\frac{\partial\phi}{\partial u}\Delta u\Delta v\right| \\ &= \left|\frac{\partial\varphi}{\partial u}\frac{\partial\phi}{\partial v}-\frac{\partial\varphi}{\partial v}\frac{\partial\phi}{\partial u}\right|\Delta u\Delta v \\ &= |det(J)|\Delta u\Delta v \\ &= \left|det\begin{bmatrix}\frac{\partial\varphi}{\partial u} & \frac{\partial\varphi}{\partial v}\\ \frac{\partial\phi}{\partial u} & \frac{\partial\phi}{\partial v}\end{bmatrix}\right|\Delta s'\end{aligned} \tag{13.197}$$

Par conséquent :

$$\Delta s \cong= \left|det\begin{bmatrix}\frac{\partial\varphi}{\partial u} & \frac{\partial\varphi}{\partial v}\\ \frac{\partial\phi}{\partial u} & \frac{\partial\phi}{\partial v}\end{bmatrix}\right|\Delta s' \tag{13.198}$$

Avec :

$$J = \begin{bmatrix} \frac{\partial \varphi}{\partial u} & \frac{\partial \varphi}{\partial v} \\ \frac{\partial \phi}{\partial u} & \frac{\partial \phi}{\partial v} \end{bmatrix} \tag{13.199}$$

qui est la "matrice jacobienne" (alors que son déterminant est appelé le "jacobien" (tout court)) de la transformation de coordonnées de $\mathbb{R}^2 \to \mathbb{R}^2$. En appliquant exactement le même raisonnement pour $\mathbb{R}^3$, alors la matrice jacobienne s'écrit alors (en changeant un peu les notations car sinon cela devient illisible) :

$$J = \begin{bmatrix} \frac{\partial P}{\partial u} & \frac{\partial P}{\partial v} & \frac{\partial P}{\partial w} \\ \frac{\partial Q}{\partial u} & \frac{\partial Q}{\partial v} & \frac{\partial Q}{\partial w} \\ \frac{\partial R}{\partial u} & \frac{\partial R}{\partial v} & \frac{\partial R}{\partial w} \end{bmatrix} \tag{13.200}$$

Bref, à quoi cela sert-il concrètement ? Eh bien revenons à notre relation :

$$\Delta s \cong |det(J)| \Delta s' \tag{13.201}$$

qui n'est finalement qu'approximative étant donné que dans les calculs de l'aire Δs nous avons négligé les infiniment petits d'ordre supérieur. Toutefois, plus les dimensions des domaines élémentaires Δs et $\Delta s'$ sont petites, et plus nous nous approchons de l'égalité. L'égalité ayant finalement lieu quand nous passons à la limite (finalement en maths aussi on fait des approximations... hein !), les surfaces des domaines élémentaires tendant vers zéro :

$$|det(J)| = \lim_{\Delta s \to 0} \frac{\Delta s}{\Delta s'} \tag{13.202}$$

Appliquons maintenant l'égalité obtenue au calcul de l'intégral double (nous pouvons faire de même avec la triple bien sûr). Nous pouvons donc finalement écrire (c'est la seule manière de poser la chose qui a un sens) :

$$\begin{aligned} &\sum f(x,y)\Delta s = \sum F(u,v)\Delta s \\ &\qquad\qquad \Downarrow \\ &\sum f(x,y)\Delta s \cong \sum F(u,v)|det(J)|\Delta s' \end{aligned} \tag{13.203}$$

Passant à la limite, nous obtenons l'égalité rigoureuse :

$$\iint_D f(x,y)dxdy = \iint_{D'} F(u,v)|det(J)|dudv \tag{13.204}$$

Telle est la relation de transformation des coordonnées dans une intégrale double. Elle permet de ramener le calcul d'une intégrale double dans le domaine D au domaine D', ce qui peut simplifier le problème.

De même, pour une intégrale triple, nous écrirons :

$$\iiint_D f(x,y,z)dxdydz = \iiint_{D'} F(u,v,w)|det(J)|dudvdw \tag{13.205}$$

Déterminons maintenant le Jacobien pour les systèmes de coordonnées les plus courants (nous renvoyons à nouveau le lecteur au chapitre de calcul vectoriel pour plus d'information concernant ces systèmes) :

1. Coordonnes polaires :

$$\left| det \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} \end{bmatrix} \right| = \left| det \begin{bmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{bmatrix} \right| = |r| \tag{13.206}$$

Comme r est toujours positif, nous écrivons simplement :

$$|det(J)| = r \tag{13.207}$$

2. Coordonnées cylindriques $r = r\cos\phi, y = r\sin\phi, z = z$ (cf. chapitre d'Algèbre Linéaire pour le calcul du déterminant) :

$$\begin{aligned}\left|det\begin{bmatrix}\frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} & \frac{\partial x}{\partial z}\\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} & \frac{\partial y}{\partial z}\\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \phi} & \frac{\partial z}{\partial z}\end{bmatrix}\right| &= \left|det\begin{bmatrix}\cos\phi & -r\sin\phi & 0\\ \sin\phi & r\cos\phi & 0\\ 0 & 0 & 1\end{bmatrix}\right|\\ &= \left|\begin{bmatrix}\cos\phi r\cos\phi\cdot 1 - \cos\phi\cdot 0\cdot 0\\ +r\sin\phi\sin\phi\cdot 1 - r\sin\phi\cdot 0\cdot 0\\ +0\cdot\sin\phi\cdot 0 - 1\cdot r\cos\phi\cdot 0\end{bmatrix}\right|\\ &= |r|\end{aligned} \tag{13.208}$$

Comme r est toujours positif, nous écrivons simplement :

$$|det(J)| = r \tag{13.209}$$

3. En coordonnées sphériques $x = r\sin\theta\cos\phi, y = r\sin\theta\sin\phi, z = r\cos\theta$ (cf. chapitre d'Algèbre Linéaire pour le calcul du déterminant) :

$$\begin{aligned}\left|det\begin{bmatrix}\frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} & \frac{\partial x}{\partial \theta}\\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} & \frac{\partial y}{\partial \theta}\\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \phi} & \frac{\partial z}{\partial \theta}\end{bmatrix}\right| &= \left|det\begin{bmatrix}\sin\theta\cos\phi & -r\sin\theta\sin\phi & r\cos\theta\cos\phi\\ \sin\theta\sin\phi & r\sin\theta\cos\phi & r\cos\theta\sin\phi\\ \cos\theta & 0 & -r\sin\theta\end{bmatrix}\right|\\ &= \left|\begin{matrix}-\sin\theta\cos\phi r\sin\theta\cos\phi r\sin\theta - -\sin\theta\cos\phi r\cos\theta\sin\phi\cdot 0\\ r\sin\theta\sin\phi\sin\theta\sin\phi r\sin\theta - r\sin\theta\sin\phi r\cos\theta\sin\phi\cos\theta\\ +r\cos\theta\cos\phi\sin\theta\sin\phi\cdot 0 + r\cos\theta\cos\phi r\sin\theta\cos\phi r\sin\theta\end{matrix}\right|\\ &= \left|\begin{matrix}-(\sin^3\theta\cos^2\phi r^2 + \sin^3\theta\sin^2\phi r^2)\\ -(r^2\sin\theta\sin^2\phi\cos^2\theta + r^2\cos^2\theta\cos^2\phi\sin\theta)\end{matrix}\right|\\ &= |-\sin^3\theta r^2 - r^2\sin\theta\cos^2\theta|\\ &= |-r^2\sin\theta(\sin^2\theta + \cos^2)|\\ &= |r|\end{aligned} \tag{13.210}$$

Comme r^2 est toujours positif, nous écrivons simplement :

$$|det(J)| = r^2\sin\theta \text{ avec } \sin\theta \geq 0 \tag{13.211}$$

13.1.5 Intégration par parties

Lorsque nous cherchons à effectuer des intégrations, il est très fréquent que nous ayons à utiliser un outil (ou méthode de calcul) appelé "intégration par parties". Voici la démonstration de la validité de ce dernier.

Soit f, g deux applications de classe C^n (dérivables n fois) de $[a, b]$ dans $\mathbb{C}$, alors (voir la version plus light dans la démo...) :

$$\begin{aligned}\int_a^b f(t)\cdot g^{(n)}(t)dt = &\left[\sum_{k=0}^{n-1}(-1)^k f^{(k)}\cdot g^{(n-1-k)}\right]_a^b\\ &+ (-1)^n\int_a^b f^{(n)}(t)\cdot g(t)dt\end{aligned} \tag{13.212}$$

Preuve Procédons par par récurrence sur n. Nous supposons la formule vraie pour n et nous la démontrons pour $n+1$:

$$\begin{aligned}
\int_a^b f(t)\cdot g^{(n+1)}(t)dt &= f(t)\cdot g^{(n)}(t)\Big|_a^b - \int_a^b f'(t)\cdot g^{(n)}(t)dt \\
&= f(t)\cdot g^{(n)}(t)\Big|_a^b - \left[\sum_{k=0}^{n-1}(-1)^k f^{(k+1)}\cdot g^{(n-1-k)}\right]_a^b \\
&\quad + (-1)^{n+1}\int_a^b f^{(n+1)}(t)\cdot g(t)dt \\
&= f(t)\cdot g^{(n)}(t)\Big|_a^b - \left[\sum_{k=0}^{n}(-1)^k f^{(k)}\cdot g^{(n-k)}\right]_a^b \\
&\quad + (-1)^{n+1}\int_a^b f^{(n+1)}(t)\cdot g(t)dt \\
&= \left[\sum_{k=0}^{n}(-1)^k f^{(k)}\cdot g^{(n-k)}\right]_a^b + (-1)^{n+1}\int_a^b f^{(n+1)}(t)\cdot g(t)dt
\end{aligned} \tag{13.213}$$

Pour $n=1$ nous retrouvons la formule bien connue :

$$\boxed{\int_a^b f(t)\cdot g'(t)dt = f(t)\cdot g(t)\Big|_a^b - \int_a^b f'(t)\cdot g(t)dt} \tag{13.214}$$

•

13.1.6 Primitives usuelles

Il existe en mathématique et en physique un grand nombre de primitives ou de fonctions définies sur des intégrales que nous retrouvons assez fréquemment (mais pas exclusivement). Comme dans n'importe quel formulaire, nous vous proposons les primitives connues mais avec les démonstrations.

Cependant, nous omettrons les primitives qui découlent déjà des dérivées que nous avons démontrées plus haut.

Sinon voici déjà une liste de quelques intégrales fréquentes (le lecteur en rencontrera de toute façon bien d'autres - développées dans les détails - lors de son parcours du site) :

Primitive de $f(x)=\tan(x)$

Par définition nous avons donc :

$$\int \tan(x)dx = \int \frac{\sin(x)}{\cos(x)}dx \tag{13.215}$$

Nous utilise le changement de variable $u=\cos(x)$ et $du=-\sin(x)dx$ ainsi :

$$\int \frac{\sin(x)}{\cos(x)}dx = -\int frac1udu = -\ln|u| = -\ln|\cos(x)| \tag{13.216}$$

Donc :

$$F(x) = -\ln|\cos(x)| \tag{13.217}$$

Primitive de $f(x) = \cot(x)$

Par définition nous avons donc :

$$\int \cot(x)dx = \int \frac{\cos(x)}{\sin(x)}dx \tag{13.218}$$

Nous utilisons le changement de variable $u = \sin(x)$ et $du = \cos(x)dx$:

$$\int \frac{\cos(x)}{\sin(x)}dx = \int frac1udu = \ln|u| = -\ln|\sin(x)| \tag{13.219}$$

Donc :

$$F(x) = -\ln|\sin(x)| \tag{13.220}$$

Primitive de $f(x) = \arcsin(x)$

Nous intégrons par parties :

$$\int 1 \cdot \arcsin(x)dx = x \cdot \arcsin(x) - \int \frac{x}{\sqrt{1-x^2}}dx \tag{13.221}$$

Si nous posons $u = 1 - x^2$ et $du = -2xdx$ nous obtenons :

$$\int \frac{x}{\sqrt{1-x^2}}dx = \int \frac{1}{2\sqrt{u}}du = -\sqrt{u} = -\sqrt{1-x^2} \tag{13.222}$$

Donc :

$$F(x) = x \cdot \arcsin(x) + \sqrt{1-x^2} \tag{13.223}$$

Primitive de $f(x) = \arccos(x)$

Nous intégrons à nouveau par parties :

$$\int 1 \cdot \arccos(x)dx = x \cdot \arccos(x) + \int \frac{x}{\sqrt{1-x^2}}dx \tag{13.224}$$

Si nous posons $u = 1 - x^2$ et $du = -2xdx$ nous obtenons :

$$\int \frac{x}{\sqrt{1-x^2}}dx = \int \frac{1}{2\sqrt{u}}du = -\sqrt{u} = -\sqrt{1-x^2} \tag{13.225}$$

Donc :

$$F(x) = x \cdot \arccos(x) - \sqrt{1-x^2} \tag{13.226}$$

Primitive de $f(x) = \arctan(x)$

Nous intégrons encore une fois par parties :

$$\int 1 \cdot \arctan(x)dx = x \cdot \arctan(x) - \int \frac{x}{\sqrt{1+x^2}}dx \tag{13.227}$$

Si nous posons $u = 1 + x^2$ et $du = 2xdx$ nous obtenons :

$$\int \frac{x}{\sqrt{1+x^2}}dx = \int \frac{1}{2u}du = \frac{1}{2} \cdot \ln|u| = \frac{1}{2} \cdot \ln(1+x^2) \tag{13.228}$$

Donc :

$$F(x) = x \cdot \arctan(x) - \frac{1}{2} \cdot \ln(1+x^2) \tag{13.229}$$

Primitive de $f(x) = arccot(x)$

Nous intégrons encore une fois par parties :

$$\int 1 \cdot arccot(x)dx = x \cdot arccot(x) + \int \frac{x}{\sqrt{1+x^2}}dx \tag{13.230}$$

Si nous posons $u = 1 + x^2$ et $du = 2xdx$ nous obtenons :

$$\int \frac{x}{\sqrt{1+x^2}}dx = \int \frac{1}{2u}du = \frac{1}{2} \cdot \ln|u| = \frac{1}{2} \cdot \ln(1+x^2) \tag{13.231}$$

Donc :

$$F(x) = x \cdot arccot(x) + \frac{1}{2} \cdot \ln(1+x^2) \tag{13.232}$$

Primitive de $f(x) = xe^{ax}$ avec $a \in \mathbb{R}\backslash\{0\}$

Une intégration par parties nous donne :

$$\int xe^{ax}dx = \frac{x}{a}e^{ax} - \frac{1}{a}\int e^{ax}dx = \frac{x}{a}e^{ax} - \frac{1}{a^2}e^{ax} = e^{ax}\left(\frac{1}{a}x - \frac{1}{a^2}\right) \tag{13.233}$$

Donc :

$$F(x) = e^{ax}\left(\frac{1}{a}x - \frac{1}{a^2}\right) \tag{13.234}$$

Remarque Une autre intégrale très importante avec l'exponentielle en physique est celle que nous avions démontrée lors de notre étude de la loi de Gauss-Laplace en statistiques et probabilités (détermination de la moyenne).

Primitive de $f(x) = \ln(x)$

$$\int \ln(x)dx = \int 1 \cdot \ln(x)dx \tag{13.235}$$

en intégrant par parties nous trouvons :

$$\begin{aligned}\int \ln(x)dx &= x\ln(x) - \int x\frac{1}{x}dx \\ &= x\ln(x) - x \\ &= x \cdot (\ln(x) - 1)\end{aligned} \tag{13.236}$$

Donc :

$$F(x) = x \cdot (\ln(x) - 1) \tag{13.237}$$

Primitive de $f(x) = x\ln(ax)$ avec $a \in \mathbb{R}\backslash\{0\}$

Une intégration par parties nous donne :

$$\int x\ln(ax)dx = \frac{1}{2}x^2\ln(ax) - \frac{1}{2}\int x^2\frac{1}{x}dx = \frac{1}{2}x^2\ln(ax) - \frac{1}{4}x^2 \tag{13.238}$$

Donc :

$$F(x) = \frac{1}{2}x^2\ln(ax) - \frac{1}{4}x^2 \tag{13.239}$$

Primitive de $f(x) = a^x$ pour $a > 0, a \neq 1$

$$f(x) = a^x \Leftrightarrow f'(x) = a^x \cdot \ln(a) \tag{13.240}$$

Ainsi il vient :

$$\frac{f'(x)}{\ln(a)} = a^x \Rightarrow \int \frac{f'(x)}{\ln(a)}dx = \int a^x dx \tag{13.241}$$

Il vient :

$$\frac{1}{\ln(a)}\int f'(x)dx = \frac{f(x)}{\ln(a)} \text{ et } \int a^x dx = \int f(x)dx = F(x) \tag{13.242}$$

d'où :

$$F(x) = \frac{f(x)}{\ln(a)} = \frac{a^x}{\ln(a)} \tag{13.243}$$

Primitive de $f(x) = \log_a(x)$

Pour $(a > 0, a \neq 1)$ sachant que (voir les propriétés des logarithmes dans le chapitre d'analyse fonctionnelle) :

$$\log_a(x) = \frac{\ln(x)}{\ln(a)} = \frac{1}{\ln(a)} \ln(x) \tag{13.244}$$

nous avons en utilisant la primitive de $\ln(x)$:

$$F(x) = \frac{x}{\ln(a)} \cdot \big(\ln(x) - 1\big) \tag{13.245}$$

Primitive de $f(x) = \tanh(x)$

Nous avons :

$$\int \tanh(x)dx = \int \frac{\sinh(x)}{\cosh(x)}dx \tag{13.246}$$

Nous utilisons le changement de variable $u = \cosh(x), du = \sinh(x)dx$ et obtenons :

$$\int \frac{\sinh(x)}{\cosh(x)}dx = \int \frac{1}{u}du = \ln|u| = \ln|\cosh(x)| \tag{13.247}$$

Donc :

$$F(x) = \ln\big(|\cosh(x)|\big) \tag{13.248}$$

Primitive de $f(x) = \coth(x)$

Nous avons :

$$\int \coth(x)dx = \int \frac{\cosh(x)}{\sinh(x)}dx \tag{13.249}$$

Nous utilisons le changement de variable $u = \sinh(x), du = \cosh(x)dx$ et obtenons :

$$\int \frac{\cosh(x)}{\sinh(x)}dx = \int \frac{1}{u}du = \ln|u| = \ln|\sinh(x)| \tag{13.250}$$

Donc :

$$F(x) = \ln\big(|\sinh(x)|\big) \tag{13.251}$$

Primitive de $f(x) = arcsinh(x)$

Nous intégrons par parties :

$$\int 1 \cdot arcsinh(x)dx = x \cdot arcsinh(x) - \int \frac{x}{\sqrt{1+x^2}}dx \tag{13.252}$$

Si nous posons $u = 1 + x^2$ et $du = 2xdx$ nous obtenons :

$$\int \frac{x}{\sqrt{1+x^2}}dx = \int \frac{1}{2\sqrt{u}}du = \sqrt{u} = \sqrt{1+x^2} \tag{13.253}$$

Donc :

$$F(x) = x \cdot arcsinh(x) - \sqrt{1+x^2} \tag{13.254}$$

Primitive de $f(x) = arccosh(x)$

Nous intégrons par parties :

$$\int 1 \cdot arccosh(x)dx = x \cdot arccosh(x) - \int \frac{x}{\sqrt{x^2-1}}dx \tag{13.255}$$

Si nous posons $u = x^2 - 1$ et $du = 2xdx$ nous obtenons :

$$\int \frac{x}{\sqrt{x^2-1}}dx = \int \frac{1}{2\sqrt{u}}du = \sqrt{u} = \sqrt{x^2-1} \tag{13.256}$$

Donc :

$$F(x) = x \cdot arccosh(x) - \sqrt{x^2-1} \tag{13.257}$$

Primitive de $f(x) = arctanh(x)$

Nous intégrons par parties :

$$\int 1 \cdot arctanh(x)dx = x \cdot arctanh(x) - \int \frac{x}{1-x^2}dx \tag{13.258}$$

Si nous posons $u = 1 - x^2$ et $du = -2xdx$ nous obtenons :

$$\int \frac{x}{1-x^2}dx = -\int \frac{1}{2u}du = -\frac{1}{2}\ln|u| = -\frac{1}{2}\ln|1-x^2| \tag{13.259}$$

Donc :

$$F(x) = x \cdot arctanh(x) + \frac{1}{2}\ln|1-x^2| \tag{13.260}$$

Primitive de $f(x) = arccoth(x)$

Nous intégrons par parties :

$$\int 1 \cdot arccoth(x)dx = x \cdot arccoth(x) - \int \frac{x}{1-x^2}dx \tag{13.261}$$

Si nous posons $u = 1 - x^2$ et $du = -2xdx$ nous obtenons :

$$\int \frac{x}{1-x^2}dx = -\int \frac{1}{2u}du = -\frac{1}{2}\ln|u| = -\frac{1}{2}\ln|1-x^2| \tag{13.262}$$

Donc :

$$F(x) = x \cdot arccoth(x) + \frac{1}{2}\ln|1-x^2| \tag{13.263}$$

Primitive de $f(x) = \sin^n(x)$ avec $n \geq 2$

Posons $I_n = \int \sin x \cdot dx$. Une intégration par partie donne :

$$I_n = \int \sin x \cdot \sin^{n-1} \cdot dx = -\cos x \cdot \sin^{n-1} x + (n+1) \int \cos^2 \cdot \sin^{n-2} x \cdot dx \tag{13.264}$$

en remplaçant $\cos^2 x$ par $1 - \sin^2 x$ dans la dernière intégrale, nous obtenons :

$$I_n = -\cos x \cdot \sin^{n-1} x + (n-1)(I_{n-2} - I_n) \tag{13.265}$$

et donc :

$$I_n = \frac{1}{n}\big(-\cos x \cdot \sin^{n-1} x + (n-1) I_{n-2}\big) \tag{13.266}$$

Primitive de $f(x) = \cos^n(x)$ avec $n \geq 2$

Dans ce cas nous avons la formule de récurrence

$$\int \cos^n x \cdot dx = \frac{1}{n}\big((n-1) \int \cos^{n-2} x \cdot dx + \sin x \cdot \cos^{n-1} x\big) \tag{13.267}$$

qui se démontre de la même façon que la relation de récurrence précédente.

Primitive de $f(x) = \tan^2(x)$

Sachant que $\tan'(x) = 1 + \tan^2(x)$, nous avons :

$$\int \tan^2(x) dx = \int \big(\tan'(x) - 1\big) dx = \tan(x) - x \tag{13.268}$$

Donc :

$$F(x) = \tan(x) - x \tag{13.269}$$

Primitive de $f(x) = \cot^2(x)$

Sachant que $\cot'(x) = -1 - \cot^2(x)$, nous avons :

$$\int \cot^2(x) dx = \int \big(-\cot'(x) - 1\big) dx = -\cot(x) - x \tag{13.270}$$

Donc :

$$F(x) = -\cot(x) - x \tag{13.271}$$

Primitive de $f(x) = \sin^{-2}(x)$

En utilisant les relations trigonométriques remarquables, nous avons :

$$\begin{aligned}\int \frac{1}{\sin^2(x)}dx &= \int \frac{\sin^2(x)+\cos^2(x)}{\sin^2(x)}dx \\ &= \int \big(1+\cot^2(x)\big)dx \\ &= -\cot(x)\end{aligned} \tag{13.272}$$

Selon la primitive $\cot^2(x)$. Donc :

$$F(x) = -\cot(x) \tag{13.273}$$

Primitive de $f(x) = \cos^2(x)$

En utilisant encore une fois les relations trigonométriques remarquables, nous avons :

$$\begin{aligned}\int \frac{1}{\cos^2(x)}dx &= \int \frac{\sin^2(x)+\cos^2(x)}{\cos^2(x)}dx \\ &= \int \big(1+\tan^2(x)\big)dx \\ &= \tan(x)\end{aligned} \tag{13.274}$$

Selon la primitive $\tan^2(x)$. Donc :

$$F(x) = \tan(x) \tag{13.275}$$

Primitive de $f(x) = \sin^{-1}(x)$

Nous faisons la substitution $x = 2\arctan(t)(t = \tan\left(\frac{x}{2}\right))$. Sachant que :

$$\sin(x) = \frac{2\tan\left(\frac{x}{2}\right)}{1+\tan^2\left(\frac{x}{2}\right)} \tag{13.276}$$

(cf. chapitre de Trigonométrie) nous obtenons alors :

$$\sin(x) = \frac{2t}{1+t^2} \text{ et } dx = \frac{2}{1+t^2}dt \tag{13.277}$$

(selon la dérivée de $\arctan(x)$). Donc :

$$\int \frac{1}{\sin(x)}dx = \int \frac{1}{t}dt = \ln|t| = \ln\left|\tan\frac{x}{2}\right| \tag{13.278}$$

et :

$$F(x) = \ln\left|\tan\frac{x}{2}\right| \tag{13.279}$$

Primitive de $f(x) = \cos^{-1}(x)$

Sachant que $\cos x = \sin\left(x + \frac{\pi}{2}\right)$ (cf. chapitre de Trigonométrie) nous avons :

$$\int \frac{1}{\cos(x)} dx = \frac{1}{\sin\left(x + \frac{\pi}{2}\right)} \tag{13.280}$$

Nous faisons le changement de variable $x + \frac{\pi}{2} = u(dt = dx)$:

$$\begin{aligned} \int \frac{1}{\cos(x)} dx = \frac{1}{\sin\left(x + \frac{\pi}{2}\right)} dx = \int \frac{1}{\sin(u)} du \\ = \ln\left|\tan\left(\frac{u}{2}\right)\right| = \ln\left|\tan\left(\frac{x}{2} + \frac{\pi}{4}\right)\right| \end{aligned} \tag{13.281}$$

(selon la primitive de $\sin^{-1}(x)$. Donc :

$$F(x) = \ln\left|\tan\left(\frac{x}{2} + \frac{\pi}{4}\right)\right| \tag{13.282}$$

Primitive de $f(x) = \frac{1}{1+\cos(x)}$

Nous faisons la substitution $x = 2\arctan(t)(t = \tan\left(\frac{x}{2}\right))$. Sachant que (cf. chapitre de Trigonométrie) :

$$\cos(x) = \frac{1 - \tan\left(\frac{x}{2}\right)}{1 + \tan\left(\frac{x}{2}\right)} \tag{13.283}$$

nous obtenons :

$$\cos x = \frac{1 - t^2}{1 + t^2} \text{ et } dx = \frac{2}{1 + t^2} dt \tag{13.284}$$

(selon la dérivée de $arctan(x)$). Donc :
(287)

$$\int \frac{1}{1 + \cos x} dx \int 1dt = t = \tan\left(\frac{x}{2}\right) \tag{13.285}$$

et :

$$F(x) = \tan\left(\frac{x}{2}\right) \tag{13.286}$$

Primitive de $f(x) = \frac{1}{1-\cos(x)}$

Nous faisons à nouveau la substitution (comme précédemment). Nous trouvons alors :

$$\int \frac{1}{1 - \cos x} dx = \int \frac{1}{t^2} dt = -\frac{1}{t} = -\cot\left(\frac{x}{2}\right) \tag{13.287}$$

et donc :

$$F(x) = -\cot\left(\frac{x}{2}\right) \tag{13.288}$$

Primitive de $f(x) = \frac{1}{1+\sin(x)}$

Sachant que :

$$\sin x = \cos\left(x - \frac{\pi}{2}\right) \tag{13.289}$$

Nous avons alors :

$$\int \frac{1}{1+\sin x}dx = \int \frac{1}{1+\cos\left(x-\frac{\pi}{2}\right)}dx \tag{13.290}$$

En faisant le changement de variable :

$$u = x - \frac{\pi}{2} \text{ avec } du = dx \tag{13.291}$$

nous obtenons :

$$\int \frac{1}{1+\sin x}dx = \int \frac{1}{1+\cos(u)}du = \tan\frac{u}{2} = \tan\left(\frac{x}{2} - \frac{\pi}{4}\right) \tag{13.292}$$

D'où :

$$F(x) = \tan\left(\frac{x}{2} - \frac{\pi}{4}\right) \tag{13.293}$$

Primitive de $f(x) = \frac{1}{1-\sin(x)}$

Par le même raisonnement que précédemment en utilisant le cosinus nous obtenons :

$$F(x) = -\cot\left(\frac{x}{2} - \frac{\pi}{4}\right) \tag{13.294}$$

Primitive de $f(x) = \sinh^n(x)$ avec $n \geq 2$

Posons :

$$I_n = \int \sinh^n(x)dx \tag{13.295}$$

Une intégration par partie donne (nous avons démontré lors des dérivées usuelles que la primitive du sinus hyperbolique était le cosinus hyperbolique) :

$$\begin{aligned} I_n = \int \sinh^{n-1}(x)dx =& \cosh(x)\sinh(x)^{n-1}(x) \\ &- (n-1)\int \cosh^2(x)\cdot\sinh^{n-2}(x)dx \end{aligned} \tag{13.296}$$

en remplaçant $\cosh^2(x)$ par $1+\sinh^2(x)$ dans la dernière intégrale, nous obtenons :

$$I_n = \cosh(x)\sinh(x)^{n-1}(x) - (n-1)(I_{n-2} + I_n) \tag{13.297}$$

et donc :

$$I_n = \int \sinh^n(x) = \frac{1}{n}\big(\cosh(x)\sinh(x)^{n-1}(x) - (n-1)I_{n-2}\big) \tag{13.298}$$

Ainsi :

$$\int \sinh^2(x) = \frac{1}{2}(\cosh(x)\sinh(x) - x) \tag{13.299}$$

Primitive de $f(x) = \cosh^n(x)$ avec $n \geq 2$

Dans ce cas nous avons aussi la relation récurrence :

$$I_n = \int \cosh(x)^n x dx = \frac{1}{n}\big(\sinh(x)\cosh(x)^{n-1}(x) - (n-1)I_{n-2}\big) \tag{13.300}$$

qui se démontre de la même façon que ci-dessus. Ainsi :

$$\int \cosh^2(x)dx = \frac{1}{2}(\cosh(x)\sinh(x) + x) \tag{13.301}$$

Primitive de $f(x) = \tanh^2(x)$

Sachant que (démontré lors des dérivées usuelles) :

$$\tanh'(x) = 1 - \tanh^2(x) \tag{13.302}$$

nous avons :

$$\int \tanh^2(x)dx = \int \big(1 - \tanh^2(x)\big)dx = x - \tanh(x) \tag{13.303}$$

Donc :

$$F(x) = x - \tanh(x) \tag{13.304}$$

Primitive de $f(x) = \coth^2(x)$

Sachant que (démontré lors des dérivées usuelles) :

$$\coth'(x) = 1 - \coth^2(x) \tag{13.305}$$

nous avons :

$$\int \coth^2(x)dx = \int \big(1 - \coth^2(x)\big)dx = x - \coth(x) \tag{13.306}$$

Donc :

$$F(x) = x - \coth(x) \tag{13.307}$$

Primitive de $f(x) = \frac{1}{\sinh^2(x)}$

Nous avons en utilisant la primitive de $\coth^2(x)$:

$$\begin{aligned}\int \sinh^2(x)dx &= \int \frac{\cosh^2(x) - \sinh^2(x)}{\sinh^2(x)}dx \\ &= \int (\coth^2(x) - 1)dx \\ &= -\coth(x)\end{aligned} \tag{13.308}$$

Donc :

$$F(x) = \coth(x) \tag{13.309}$$

Primitive de $f(x) = \frac{1}{\cosh^2(x)}$

Nous avons en utilisant la primitive de $\tanh^2(x)$:

$$\begin{aligned}\int \cosh^2(x)dx &= \int \frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)}dx \\ &= \int (1 - \tanh^2(x))dx \\ &= \tanh(x)\end{aligned} \tag{13.310}$$

Donc :

$$F(x) = \tanh(x) \tag{13.311}$$

Primitive de $f(x) = \frac{1}{\sinh(x)}$

Nous faisons la substitution :

$$t = e^x \text{ avec } dx = \frac{dt}{t} \tag{13.312}$$

Nous obtenons en utilisant la dérivée $arctanh(x)$:

$$\begin{aligned}\int \frac{1}{\sinh(x)}dx &= \int \frac{2}{e^x - e^{-x}}dx \\ &= \int \frac{2}{t\left(t - \frac{1}{t}\right)}dt \\ &= \int \frac{2}{t^2 - 1}dt \\ &= -2arctanh(t) \\ &= -2arctanh(e^x)\end{aligned} \tag{13.313}$$

et :

$$F(x) = -2arctanh(e^x) \tag{13.314}$$

Primitive de $f(x) = \frac{1}{\cosh(x)}$

Nous faisons la substitution :

$$t = e^x \text{ avec } dx = \frac{dt}{t} \tag{13.315}$$

Nous obtenons en utilisant la dérivée $arctan(x)$:

$$\begin{aligned}\int \frac{1}{\cosh(x)}dx &= \int \frac{2}{e^x + e^{-x}}dx \\ &= \int \frac{2}{t\left(t + \frac{1}{t}\right)}dt \\ &= \int \frac{2}{t^2 + 1}dt \\ &= 2arctanh(t) \\ &= 2arctanh(e^x)\end{aligned} \tag{13.316}$$

et donc :

$$F(x) = 2arctanh(e^x) \tag{13.317}$$

Primitive de $f(x) = \frac{1}{1+\cosh(x)}$

Nous faisons la substitution :

$$t = e^x \text{ avec } dx = \frac{dt}{t} \tag{13.318}$$

Nous obtenons en utilisant la dérivée $arctan(x)$:

$$\begin{aligned}\int \frac{1}{1+\cosh(x)}dx &= \int \frac{2}{2+e^x+e^{-x}}dx \\ &= \int \frac{2}{t\left(2+t+\frac{1}{t}\right)}dt \\ &= \int \frac{2}{(t+1)^2}dt \\ &= -\frac{2}{1+t} + c^{te} \\ &= -\frac{2}{1+e^x} + c^{te}\end{aligned} \tag{13.319}$$

Nous obtenons donc la primitive :

$$F(x) = -\frac{2}{1+e^x} + 1 = \frac{e^x-1}{e^x+1} = \tanh\left(\frac{x}{2}\right) \tag{13.320}$$

Primitive de $f(x) = \frac{1}{1-\cosh(x)}$

Nous faisons la substitution :

$$t = e^x \text{ avec } dx = \frac{dt}{t} \tag{13.321}$$

Nous obtenons :

$$\begin{aligned}\int \frac{1}{1-\cosh(x)}dx &= \int \frac{2}{2-e^x-e^{-x}}dx \\ &= \int \frac{2}{t\left(2-t-\frac{1}{t}\right)}dt \\ &= \int \frac{2}{(t-1)^2}dt \\ &= \frac{2}{t-1} + c^{te} \\ &= \frac{2}{e^x-1} + c^{te}\end{aligned} \tag{13.322}$$

Nous obtenons donc la primitive :

$$F(x) = \frac{2}{e^x-1} + 1 = \frac{e^x+1}{e^x-1} = \coth\left(\frac{x}{2}\right) \tag{13.323}$$

Primitive de $f(x) = \frac{1}{1+\sinh(x)}$

Nous faisons la substitution :

$$t = e^x \text{ avec } dx = \frac{dt}{t} \tag{13.324}$$

Nous obtenons :

$$\int \frac{1}{1+\sinh(x)}dx = \int \frac{2}{2+e^x-e^{-x}}dx \qquad = \int \frac{2}{2t+t^2-1}dt \tag{13.325}$$

Or :

$$\begin{aligned}\frac{2}{2t+t^2-1} &= \frac{2}{(t+1-\sqrt{2})(t+1+\sqrt{2})} \\ &= \frac{1}{\sqrt{2}}\left(\frac{1}{(t+1-\sqrt{2})} - \frac{1}{(t+1+\sqrt{2})}\right)\end{aligned} \tag{13.326}$$

D'où :

$$\begin{aligned}\int \frac{2}{2t+t^2-1}dt &= \frac{1}{\sqrt{2}}\left(\int \frac{1}{(t+1-\sqrt{2})}dt - \int \frac{1}{(t+1+\sqrt{2})}dt\right) \\ &= \frac{1}{\sqrt{2}}\left(\ln\left|t+1-\sqrt{2}\right| - \ln\left|t+1+\sqrt{2}\right|\right) \\ &= \frac{1}{\sqrt{2}}\ln\frac{\left|t+1-\sqrt{2}\right|}{\left|t+1+\sqrt{2}\right|} \\ &= \frac{1}{\sqrt{2}}\ln\frac{\left|e^x+1-\sqrt{2}\right|}{\left|e^x+1+\sqrt{2}\right|}\end{aligned} \tag{13.327}$$

Donc :

$$F(x) = \frac{1}{\sqrt{2}}\ln\frac{\left|e^x+1-\sqrt{2}\right|}{\left|e^x+1+\sqrt{2}\right|} \tag{13.328}$$

Primitive de $f(x) = \frac{1}{1-\sinh(x)}$

Nous faisons la substitution habituelle :

$$t = e^x \text{ avec } dx = \frac{dt}{t} \tag{13.329}$$

Nous obtenons :

$$\int \frac{1}{1-\sinh(x)}dx = \int \frac{2}{2-e^x-e^{-x}}dx \qquad = -\int \frac{2}{-2t+t^2-1}dt \tag{13.330}$$

Or :

$$\begin{aligned}\frac{2}{-2t+t^2-1} &= \frac{2}{(t-1-\sqrt{2})(t-1+\sqrt{2})} \\ &= \frac{1}{\sqrt{2}}\left(\frac{1}{(t-1-\sqrt{2})} - \frac{1}{(t-1+\sqrt{2})}\right)\end{aligned} \tag{13.331}$$

D'où :

$$\begin{aligned} -\int \frac{2}{-2t+t^2-1}dt &= -\frac{1}{\sqrt{2}}\left(\int \frac{1}{(t-1-\sqrt{2})}dt + \int \frac{1}{(t-1+\sqrt{2})}dt\right) \\ &= \frac{1}{\sqrt{2}}\left(\ln\left|t-1+\sqrt{2}\right| - \ln\left|t-1-\sqrt{2}\right|\right) \\ &= \frac{1}{\sqrt{2}}\ln\frac{\left|t-1+\sqrt{2}\right|}{\left|t-1-\sqrt{2}\right|} \\ &= \frac{1}{\sqrt{2}}\ln\frac{\left|e^x-1+\sqrt{2}\right|}{\left|e^x-1-\sqrt{2}\right|} \end{aligned} \tag{13.332}$$

Donc :

$$F(x) = \frac{1}{\sqrt{2}}\ln\frac{\left|e^x-1+\sqrt{2}\right|}{\left|e^x-1-\sqrt{2}\right|} \tag{13.333}$$

Primitive de $f(x) = e^{ax}\sin(bx)$ avec $a, b \in \mathbb{R}, a^2+b^2 \neq 0$

Une première intégration par parties donne :

$$\int e^{ax}\sin(bx)dx = \frac{1}{a}e^{ax}\sin(bx) - \frac{b}{a}\int e^{ax}\cos(bx)dx \tag{13.334}$$

Une deuxième intégration par parties donne :

$$\int e^{ax}\sin(bx)dx = \frac{1}{a}e^{ax}\cos(bx) + \frac{b}{a}\left(\frac{1}{a}e^{ax}\sin(bx) - \frac{b}{a}\int e^{ax}\cos(bx)dx\right) \tag{13.335}$$

d'où l'égalité :

$$\int e^{ax}\sin(bx)dx = \frac{1}{a}e^{ax}\sin(bx) - \frac{b}{a^2}e^{ax}\cos(bx) - \frac{b^2}{a^2}e^{ax}\int e^{ax}\sin(bx)dx \tag{13.336}$$

Ainsi en redistribuant la relation précédente :

$$F(x) = \int e^{ax}\sin(bx)dx = \frac{e^{ax}}{a^2+b^2}\big(a\sin(bx) - b\cos(bx)\big) \tag{13.337}$$

Primitive de $f(x) = e^{ax}\cos(bx)$ avec $a, b \in \mathbb{R}, a^2+b^2 \neq 0$

Un raisonnement analogue à celui d'avant montre que :

$$F(x) = \frac{e^{ax}}{a^2+b^2}\big(a\cos(bx) + b\sin(bx)\big) \tag{13.338}$$

Primitive de $f(x) = x\sin(ax)$ avec $a \in \mathbb{R}^*$

Une intégration par parties nous donne :

$$\begin{aligned} F(x) = \int x\sin(ax)dx &= -\frac{1}{a}x\cos(ax) + \frac{1}{a}\int \cos(ax)dx \\ &= -\frac{1}{a}x\cos(ax) + \frac{1}{a^2}x\sin(ax) \end{aligned} \tag{13.339}$$

Primitive de $f(x) = x\cos(ax)$ avec $a \in \mathbb{R}^*$

Une intégration par parties nous donne :

$$\begin{aligned} F(x) = \int x\cos(ax)dx &= \frac{1}{a}x\sin(ax) - \frac{1}{a}\int \sin(ax)dx \\ &= \frac{1}{a}x\sin(ax) + \frac{1}{a^2}x\cos(ax) \end{aligned} \tag{13.340}$$

Primitive de $f(x) = \frac{1}{(x-a)(x-b)}$ avec $a \neq b$

Nous avons la relation suivante :

$$\frac{1}{(x-a)(x-b)} = \frac{1}{a-b}\left(\frac{1}{(x-a)} - \frac{1}{(x-b)}\right) \tag{13.341}$$

Par suite :

$$\begin{aligned} \int \frac{1}{(x-a)(x-b)}dx &= \frac{1}{a-b}\int \left(\frac{1}{(x-a)} - \frac{1}{(x-b)}\right)dx \\ &= \frac{1}{a-b}\left(\int \frac{1}{(x-a)}dx - \int \frac{1}{(x-b)}dx\right) \\ &= \frac{1}{a-b}\left(\ln|x-a| - \ln|x-b|\right) \\ &= \frac{1}{a-b}\ln\left|\frac{x-a}{x-b}\right| \end{aligned} \tag{13.342}$$

Ainsi :

$$F(x) = \frac{1}{a-b}\ln\left|\frac{x-a}{x-b}\right| \tag{13.343}$$

Primitive de $f(x) = \frac{1}{a^2-x^2}$ avec $a \neq 0$

Nous avons en utilisant le résultat précédent :

$$\begin{aligned} \int \frac{1}{a^2-x^2}dx &= \int \frac{1}{(a+x)(a-x)}dx \\ &= -\int \frac{1}{(x+a)(x-a)}dx \\ &= -\frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| \\ &= \frac{1}{2a}\ln\left|\frac{x+a}{x-a}\right| \end{aligned} \tag{13.344}$$

Donc :

$$F(x) = \frac{1}{2a}\ln\left|\frac{x+a}{x-a}\right| \tag{13.345}$$

Primitive de $f(x) = \frac{1}{a^2+x^2}$ avec $a \neq 0$

En faisant le changement de variable :

$$x = au \text{ avec } dx = a \cdot du \tag{13.346}$$

Nous obtenons en utilisant la dérivée de $\arctan(x)$:

$$\begin{aligned} F(x) = \int \frac{1}{a^2+x^2}dx &= \frac{1}{a}\int \frac{1}{1+u^2}du \\ &= \frac{1}{a}\arctan(u) \\ &= \frac{1}{a}\arctan(\frac{x}{a}) \end{aligned} \tag{13.347}$$

Primitive de

$$I_{n+1} = \int \frac{1}{(1-x^2)^{n+1}}dx \tag{13.348}$$

avec $n \in \mathbb{N}$. Nous avons :
Nous obtenons en utilisant la dérivée de $\arctan(x)$:

$$\begin{aligned} I_{n+1} &= \int \frac{1}{(1-x^2)^{n+1}}dx \\ &= \int \left(\frac{1-x^2}{(1-x^2)^{n+1}} + \frac{x^2}{(1-x^2)^{n+1}}\right)dx \\ &= \int \frac{1}{(1-x^2)^{n}}dx + \int \frac{x^2}{(1-x^2)^{n+1}}dx \\ &= I_n + \int \frac{x^2}{(1-x^2)^{n+1}}dx \end{aligned} \tag{13.349}$$

Or cette dernière intégrale se résout par parties :

$$\begin{aligned} \int \frac{x^2}{(1-x^2)^{n+1}}dx &= \int \frac{x}{(1-x^2)^{n+1}}xdx \\ &= \frac{1}{2n(1-x^2)^n}x - \int \frac{1}{2n(1-x^2)^n}dx \\ &= \frac{1}{2n(1-x^2)^n}x - \frac{1}{2n}I_n \end{aligned} \tag{13.350}$$

Donc :

$$\begin{aligned} I_{n+1} &= I_n + \frac{1}{2n(1-x^2)^n}x - \frac{1}{2n}I_n \\ &= \frac{1}{2n(1-x^2)^n}x - \frac{(2n-1)}{2n}I_n \end{aligned} \tag{13.351}$$

Que nous retrouvons plus fréquemment dans la littérature sous la forme :

$$\boxed{2nI_{n+1} = \frac{x}{(1-x^2)^n} + (2n-1)I_n} \tag{13.352}$$

Identiquement au développement suivant, nous avons pour (le signe change) :

$$I_{n+1} = \int \frac{1}{(1+x^2)^{n+1}} dx \tag{13.353}$$

la relation suivante :

$$\boxed{2nI_{n+1} = \frac{x}{(1+x^2)^n} + (2n-1)I_n} \tag{13.354}$$

Vous pourrez trouver une application de ces deux primitives dans le modèle cosmologique newtonien de l'univers dans le chapitre d'Astrophysique !

Primitive de $f(x) = \frac{1}{(1-x^2)^2}$

Nous avons en utilisant les primitives de $f(x) = \frac{1}{(1-x^2)^n}$ (vu avant) et $f(x) = \frac{1}{1-x^2}$ (vu plus haut) :

$$F(x) = \int \frac{1}{(1-x^2)^2} dx = \frac{1}{2}\left(\frac{x}{1-x^2} + \frac{1}{2}\ln\left|\frac{x+1}{x-1}\right|\right) \tag{13.355}$$

Primitive de $f(x) = \frac{1}{(1+x^2)^2}$

Nous avons en utilisant les primitives de $f(x) = \frac{1}{(1-x^2)^n}$ (vu avant) et $f(x) = \frac{1}{1-x^2}$ (vu plus haut) :

$$F(x) = \int \frac{1}{(1+x^2)^2} dx = \frac{1}{2}\left(\frac{x}{1+x^2} + \arctan(x)\right) \tag{13.356}$$

Primitive de $f(x) = \sqrt{(x^2-a^2)}$

Nous pouvons sans perte de généralité supposer $a > 0$. Remarquons que le domaine de définition de f est $]-\infty, -a] \cup [a, +\infty[$. Dans un premier temps nous allons déterminer une primitive de f sur l'intervalle $[a, +\infty[$.

Faisons le changement de variable :

$$x = a \cdot \cosh t \tag{13.357}$$

avec donc :

$$t = arccosh\left(\frac{1}{a}x\right), \quad dx = a\sinh(t)dt \tag{13.358}$$

où nous considérons la fonction cosh : $\mathbb{R}_+ \to [1, +\infty[$ avec pour réciproque la fonction $arccosh$: $[1, +\infty[\to \mathbb{R}_+$ donnée par (cf. chapitre de Trigonométrie) :

$$arccosh(y) = \ln\left(y + \sqrt{y^2-1}\right) \tag{13.359}$$

Nous obtenons alors en utilisant la primitive de $\sinh^2 x$:

$$\begin{aligned} \int \sqrt{x^2-a^2} dx &= a^2 \int \sinh^2 \cdot dt = \frac{a^2}{2}(\cosh t \sinh t - t) \\ &= \frac{a^2}{2} \cdot \left(\frac{1}{a}x \sinh\left(arccosh\left(\frac{1}{a}x\right) - arccosh\left(\frac{1}{a}x\right)\right)\right) \end{aligned} \tag{13.360}$$

or (cf. chapitre de Trigonométrie) comme :

$$\sinh^2 u = \cosh^2 u - 1 \tag{13.361}$$

Donc :

$$\sinh\left(arccosh\left(\frac{1}{a}\right)\right) \tag{13.362}$$

et en utilisant un autre résultat du chapitre de Trigonométrie :

$$arccoshh(x) = \ln\left(x + \sqrt{x^2+1}\right) \tag{13.363}$$

nous avons alors :

$$\begin{aligned}\int \sqrt{x^2-a^2}dx &= \frac{a^2}{2}\left(\frac{x}{a}\sqrt{\frac{x^2}{a^2}-1} - \ln\left(\frac{x}{a} + \sqrt{\frac{x^2}{a^2}-1}\right)\right)\\ &= \frac{a^2}{2}\left(\frac{x}{a^2}\sqrt{x^2-a^2} - \ln\left(\frac{x}{a} + \frac{1}{a}\sqrt{x^2-a^2}\right)\right)\\ &= \frac{a^2}{2}\left(\frac{x}{a^2}\sqrt{x^2-a^2} - \ln\left(x + \sqrt{x^2-a^2}\right) + \ln(a)\right)\end{aligned} \tag{13.364}$$

étant donné que les primitives sont données à une constant près, nous pouvons écrire :

$$F(x) = \frac{x}{2}\sqrt{x^2-a^2} - \frac{a^2}{2}\ln\left(x + \sqrt{x^2-a^2}\right) \tag{13.365}$$

pour $x \geq a$. F est donc une primitive de $\sqrt{x^2-a^2}$ sur $[a, +\infty[$.

Primitive de $f(x) = \sqrt{(a^2-x^2)}$ avec $a \in \mathbb{R}^*$

Nous pouvons sans perte de généralité supposer $a > 0$. Remarquons que le domaine de définition de f est $[-a, a]$.
Nous faisons la substitution :

$$x = \sin(t) \text{ avec } dx = a\cos(t)dt \tag{13.366}$$

Nous obtenons :

$$\begin{aligned}\int \sqrt{a^2-x^2} &= a^2\int \sqrt{1-\sin^2(t)}\cos(t)dt = a^2\int \cos^2(t)dt\\ &= \frac{a^2}{2}\left(t + \sin(t)\cos(t)\right)\\ &= \frac{a^2}{2}\left[\arcsin\left(\frac{x}{a}\right) + \frac{x}{a}\cos\left(\arcsin\left(\frac{x}{a}\right)\right)\right]\end{aligned} \tag{13.367}$$

où nous avons utilisé la primitive de $\cos^n(x)$ avec $n = 2$ démontrée plus haut.Or nous avons :

$$\cos\big(\arcsin(y)\big) = \sqrt{1-\sin^2(\arcsin(y))} = \sqrt{1-y^2} \tag{13.368}$$

Donc :

$$\begin{aligned}\int \sqrt{a^2-x^2}dx &= \frac{a^2}{2}\left[\arcsin\left(\frac{x}{a}\right)+\frac{x}{a}\cos\left(\arcsin\left(\frac{x}{a}\right)\right)\right]\\ &= \frac{a^2}{2}\left[\arcsin\left(\frac{x}{a}\right)+\frac{x}{a}\sqrt{1-\frac{x^2}{a^2}}\right]\\ &= \frac{a^2}{2}\arcsin\left(\frac{x}{a}\right)+\frac{x}{2}\sqrt{a^2-x^2}\end{aligned} \tag{13.369}$$

et :

$$F(x) = \frac{a^2}{2}\arcsin\left(\frac{x}{a}\right)+\frac{x}{2}\sqrt{a^2-x^2} \tag{13.370}$$

Primitive de $f(x)=\sqrt{(a^2+x^2)}$ avec $a\in\mathbb{R}^*$

Nous pouvons sans perte de généralité supposer $a>0$. Faisons le changement de variable :

$$x = a\cdot\sinh(t) \tag{13.371}$$

avec donc :

$$t = arcsinh(\frac{1}{a}x),\quad dx = a\cosh(t)dt \tag{13.372}$$

Nous obtenons :

$$\int \sqrt{x^2+a^2}dx = a^2\int \cosh^2 tdt = \frac{a^2}{2}(t+\sinh t\cosh t) \tag{13.373}$$

en ayant utilisé la primitive de $\cosh^2$ démontrée plus haut.
Ainsi :

$$\int \sqrt{x^2+a^2}dx = \frac{a^2}{2}(t+\sinh t\cosh t) = \frac{a^2}{2}\left(arcsinh(\frac{x}{a})+\frac{x}{a}\cosh(arcsinh(\frac{x}{a}))\right) \tag{13.374}$$

Mais comme nous avons vu dans le chapitre de trigonométrie :

$$arcsinh\left(\frac{x}{a}\right) = \ln\left(\frac{x}{a}+\frac{1}{a}\sqrt{x^2+a^2}\right) = \ln(x+\sqrt{x^2+a^2})-\ln(a) \tag{13.375}$$

et :

$$\cosh\left(arcsinh(\frac{x}{a})\right) = \sqrt{1+\sinh^2\left(arcsinh\left(\frac{x}{a}\right)\right)} = \sqrt{1+\frac{x^2}{a^2}} = \frac{1}{a}\sqrt{a^2+x^2} \tag{13.376}$$

Donc :

$$F(x) = \frac{a^2}{2}\ln\left(x+\sqrt{x^2+a^2}\right)+\frac{x}{2}\sqrt{a^2+x^2} \tag{13.377}$$

Primitive de $f(x) = \sqrt{(a^2 - x^2)}$ avec $a \in \mathbb{R}^*$

Nous pouvons sans perte de généralité supposer $a > 0$.
Nous faisons la substitution :

$$x = a\sin(t) \text{ avec } dx = a\cos(t)dt \tag{13.378}$$

Nous obtenons :

$$\begin{aligned} F(x) &= \int \frac{1}{\sqrt{a^2 - x^2}}dx = \int \frac{1}{\sqrt{a^2 - a^2\sin^2 t}}a\cos ddt \\ &= \int \frac{1}{a\sqrt{1 - \sin^2 t}}a\cos tdt = \int \frac{1}{a\cos t}a\cos tdt = \int 1 \cdot dt = t \\ &= \arcsin\frac{x}{a} \end{aligned} \tag{13.379}$$

Primitive de $f(x) = \frac{1}{\sqrt{(x^2+a^2)}}$ avec $a \in \mathbb{R}^*$

Nous pouvons sans perte de généralité supposer $a > 0$.
Faisons le changement de variable :

$$x = a \cdot \sinh t \text{ avec } t = arcsinh\left(\frac{1}{a}x\right), \quad dx = a\cosh(t)dt \tag{13.380}$$

Nous obtenons de la même manière que précédemment. :

$$\int \frac{1}{\sqrt{(x^2 + a^2)}} = \int 1 \cdot dt = t = arcsinh\frac{x}{a} \tag{13.381}$$

Sachant que (cf. chapitre de Trigonométrie) :

$$arcsinh(y) = \ln(y + \sqrt{y^2 + 1}) \tag{13.382}$$

Nous obtenons la primitive :

$$\boxed{F(x) = \ln(x + \sqrt{x^2 + a^2})} \tag{13.383}$$

qui est une primitive importante que nous retrouverons en Mécanique Analytique et en Génie Civil.

13.2 Fonction de Dirac

La fonction de Dirac ou "fonction delta" joue un rôle pratique très important aussi bien en électronique et informatique qu'en physique quantique ondulatoire et physique quantique des champs (cela permet de discrétiser un continuum). Pour l'introduire simplement, considérons la fonction définie par :

$$\begin{aligned} f(x) &= \frac{1}{a} \quad si \quad -\frac{a}{2} < x < \frac{a}{2} \\ f(x) &= 0 \quad si \quad |x| > \frac{a}{2} \end{aligned} \tag{13.384}$$

La représentation de est un rectangle de largeur a, de hauteur 1/a et de surface unité. La fonction de Dirac peut être considérée comme la limite, lorsque de la fonction f(x). On a donc :

$$\boxed{\begin{aligned} \delta(x) &= 0 \quad si \quad x \neq 0 \\ \delta(x) &= \infty \quad si \quad x = 0 \end{aligned}} \tag{13.385}$$

avec :

$$\int_{-\infty}^{+\infty} \delta(x)dx = \int_{-\varepsilon}^{+\varepsilon} \delta(x)dx = 1 \tag{13.386}$$

où ε est un nombre plus grand que 0 aussi petit que nous le voulons. Pour une fonction $g(x)$ continue en $x = 0$ on a :

$$\int_{-\infty}^{+\infty} g(x)\delta(x)dx = g(0) \tag{13.387}$$

Par extension nous avons :

$$\boxed{\begin{array}{lll} \delta(x - x_0) = 0 & si & x \neq x_0 \\ \delta(x - x_0) = \infty & si & x = x_0 \end{array}} \tag{13.388}$$

et pour une fonction $g(x)$ continue en x_0 :

$$\int_{-\infty}^{+\infty} g(x)\delta(x - x_0)dx = g(x_0) \tag{13.389}$$

Il est alors assez aisé de définir la fonction de Dirac dans l'espace à 3 dimensions par :

$$\delta(\vec{r} - \vec{r}_0) = \delta(x - x_0)\delta(y - y_0)\delta(y - y_0) \tag{13.390}$$

13.3 Fonction Gamma d'Euler

Nous définissons la fonction Gamma d'Euler (intégrale Eulérienne de deuxième espèce) par l'intégrale suivante :

$$\Gamma_{-1}(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt \tag{13.391}$$

avec x appartenant à l'ensemble des nombres complexes.

La figure 13.5 représente un tracé graphique pour x parcourant un intervalle des nombres réels. La même fonction mais dans le plan complexe avec en ordonnée le module de x est présenté à la figure 13.6.

REMARQUE Nous avons déjà rencontré cette intégrale et certaines de ses propriétés lors de notre étude des fonctions de distribution Bêta, Gamma, Khi-deux, Student et Fisher en statistiques (cf. chapitre de Statistiques). Nous utiliserons également cette intégrale en maintenance (cf. chapitre de Techniques De Gestion), en théorie des cordes (cf. chapitre de Théorie Des Cordes) et dans d'autres domaines de l'ingénierie (voir la section correspondante).

Cette fonction est intéressante si nous imposons que la variable x appartienne aux entiers positifs et que nous l'écrivons sous la forme suivante :

$$\Gamma_0(x) = \int_0^{+\infty} t^x e^{-t}dt \tag{13.392}$$

Intégrons par partie cette dernière fonction :

$$\begin{aligned} \Gamma_0(x) &= \int_0^{+\infty} t^x e^{-t}dt \\ &= -t^x e^{-t}\Big|_0^{+\infty} + x\int_0^{+\infty} t^{x-1}e^{-t}dt \\ &= -t^x e^{-t}\Big|_0^{+\infty} + x\Gamma_0(x-1) \end{aligned} \tag{13.393}$$

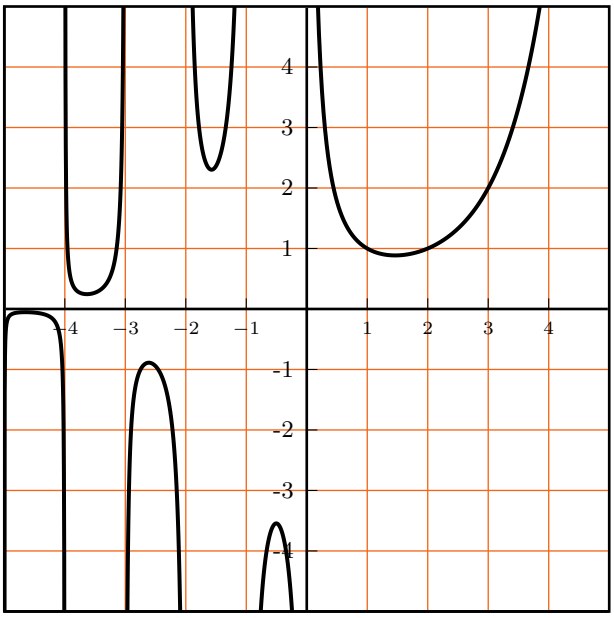


FIGURE 13.5 – Fonction $\Gamma_0(x)$ dans $\mathbb{R}$

Comme la fonction exponentielle décroît beaucoup plus vite que nous avons alors :

$$\Gamma_0(x) = x \cdot \Gamma_0(x-1) \tag{13.394}$$

Dans la littérature, nous retrouvons fréquemment les notations suivantes (qui portent alors à confusion) :

$$\begin{aligned} \Gamma_{-1}(x) = \Gamma_0(x-1) = \Gamma(x) &= \int_0^{+\infty} t^{x-1} e^{-t} dt \\ \Gamma_0(x) = \Gamma_0(x+1) = \Gamma(x) &= \int_0^{+\infty} t^{x} e^{-t} dt \end{aligned} \tag{13.395}$$

Ce qui nous amène à récrire le résultat sous une forme plus classique :

$$\boxed{\Gamma(x+1) = x\Gamma(x)} \tag{13.396}$$

De la relation $\Gamma_0(x) = x \cdot \Gamma_0(x-1)$, il vient par récurrence :

$$\Gamma_0(x) = x(x-1)\Gamma_0(x-2) = \ldots = \Gamma_0(0)x! \tag{13.397}$$

Or :

$$\Gamma_0(0) = \Gamma(1) = \int_0^{+\infty} t^0 e^{-t} dt = \int_0^{+\infty} e^{-t} dt = -e^{-t}\Big|_0^{\infty} = -e^{-\infty} - (-e^{-0}) = 1 \tag{13.398}$$

ce qui donne :

$$\Gamma_0(x) = x \cdot (x-1) \cdot \Gamma_0(x-2) = \ldots = (0)x! \tag{13.399}$$

Donc :

$$\Gamma_0(x+1) = x! \tag{13.400}$$

ou autrement écrit pour $x > 0, \in \mathbb{N}$:

$$\Gamma(x+1) = x! \tag{13.401}$$

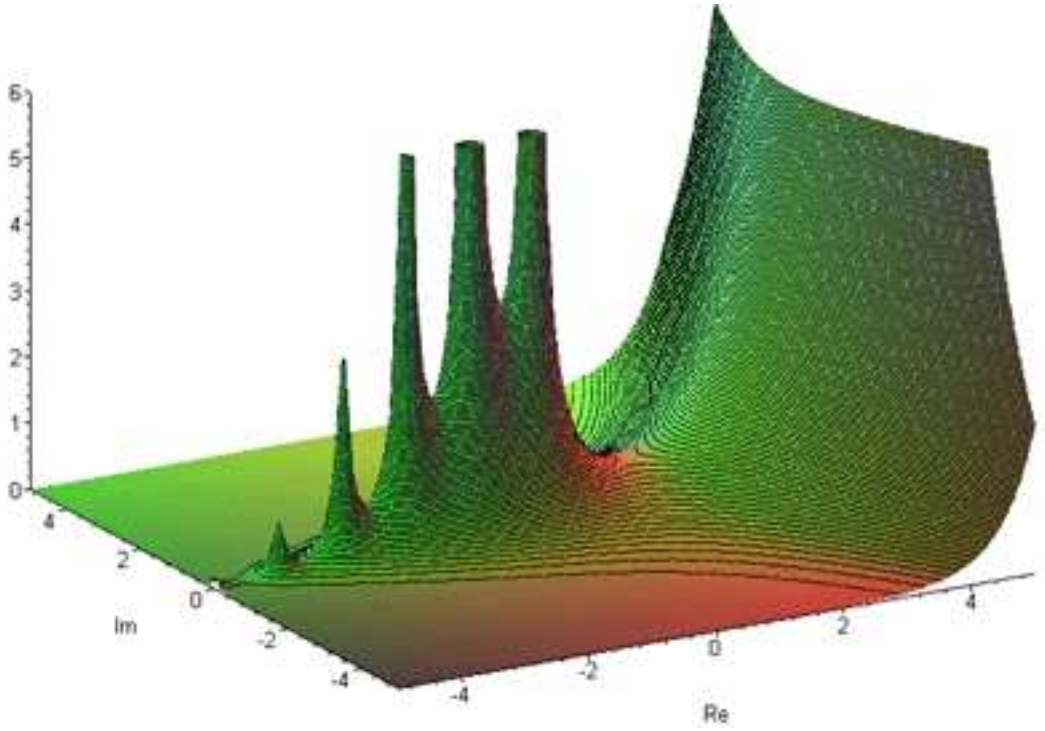


FIGURE 13.6 – Fonction Gamma d'Euler dans $\mathbb{R}$

Un autre résultat intéressant de la fonction gamma d'Euler est obtenu lorsque nous remplaçons t par y^2 et calculons celle-ci pour $x = \frac{1}{2}$.

D'abord, nous avons :

$$\frac{dt}{dy} = \frac{dy^2}{dy} = 2y \Rightarrow dt = 2ydy \tag{13.402}$$

ensuite :

$$\begin{aligned}\Gamma_{-1}\left(\frac{1}{2}\right) &= \int_0^{+\infty} t^{\frac{1}{2}-1}e^{-t}dt \\ &= \int_0^{+\infty} y^{-1}e^{-y^2}2ydy \\ &= 2\int_0^{+\infty} e^{-y^2}dy\end{aligned} \tag{13.403}$$

Or, comme nous l'avons démontré dans le chapitre de statistiques lors de notre étude de loi de de Gauss-Laplace, cette dernière intégrale vaut :

$$\Gamma_{-1}\left(\frac{1}{2}\right) = 2\int_0^{+\infty} e^{-y^2}dy = \sqrt{\pi} \tag{13.404}$$

13.3.1 Constante d'Euler-Mascheroni

Ce petit texte fait juste office de curiosité relativement à la constante d'Euler e et à presque tous les outils de calcul différentiel et intégral que nous avons vu jusqu'à maintenant. C'est un très joli exemple (presque artistique) de ce que nous pouvons faire avec les mathématiques dès que nous avons suffisamment d'outils à notre disposition.

De plus, cette constante est utile dans certaines équations différentielles où nous la retrouverons.

Nous avions vu dans le chapitre d'analyse fonctionnelle que la constante d'euler e est définie

par la limite :

$$e = \lim_{\alpha \to +\infty} \left(1 + \frac{1}{\alpha}\right)^2 \tag{13.405}$$

Dans un cas plus général nous pouvons très facilement démontrer de la même façon que :

$$\lim_{n \to +\infty} \left(1 - \frac{t}{n}\right)^n = e^{-t} \tag{13.406}$$

Cela suggère évidemment :

$$\Gamma_{-1}(x) = \int_0^\infty t^{x-1} e^{-t} dt = \lim_{n \to +\infty} \int_0^\infty t^{x-1} \left(1 - \frac{t}{n}\right)^n dt \tag{13.407}$$

par changement $t = n \cdot u$ de variable nous écrivons :

$$\begin{aligned}
\Gamma_{-1}(x) &= \lim_{n \to +\infty} \int_0^1 (n \cdot u)^{x-1} (1-u)^n \cdot n \cdot du \\
&= \lim_{n \to +\infty} n^x \int_0^1 u^{x-1} (1-u)^n du \\
B(x-1, n) &= \int_0^1 u^{x-1} (1-u)^n du \\
&= \frac{u^x}{x} (1-u)^n \Big|_0^1 + \frac{1}{x} n \int_0^1 u^x (1-u)^{n-1} du \\
&= \frac{n}{x} \cdot B(x, n-1) \\
&= \frac{n(n-1)}{x(x+1)} \cdot B(x+1, n-2) \\
&= \frac{n!}{x(x+1)\dots(x+n)} \\
&\Downarrow \\
\Gamma_{-1}(x) &= \lim_{n \to +\infty} \frac{n! \cdot n^x}{x(x+1)\dots(x+n)}
\end{aligned} \tag{13.408}$$

Pour transformer cette expression nous pouvons écrire :

$$\begin{aligned}
n^x &= e^{\ln(n^x)} = e^{x \cdot \ln(n)} \\
&= e^{x \cdot \ln(n) - 1 + 1 - \frac{1}{2} + \frac{1}{2} - \dots - \frac{1}{n} + \frac{1}{n}} \\
&= e^{-x \cdot (1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln(n))} \cdot e^x \cdot e^{\frac{x}{2}} \cdot \dots \cdot e^{\frac{x}{n}}
\end{aligned} \tag{13.409}$$

Or la quantité :

$$1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln(n) \tag{13.410}$$

tend vers la limite $\gamma \cong 0,5772\dots$, appelée "constante d'Euler-Mascheroni" ou également "constante Gamma d'Euler", lorsque n tend vers l'infini.

D'où :

$$\begin{aligned}
\Gamma_{-1}(x) &= \lim_{n \to +\infty} \frac{n! \cdot e^{-\gamma x} \cdot e^x \cdot e^{\frac{x}{2}} \cdot \dots \cdot e^{\frac{x}{n}}}{x(x+1)\dots(x+n)} \\
&= \lim_{n \to +\infty} \frac{n! \cdot e^{-\gamma x} \prod_{k=1}^{n} e^{\frac{x}{k}}}{x(x+1)\dots(x+n)}
\end{aligned} \tag{13.411}$$

Divisons chacun des termes du produit $(x+1)\dots(x+n)$ par l'entier correspondant pris dans $n!$, nous obtenons donc :

$$\begin{aligned}\frac{1}{\Gamma_{-1}(x)} &= x\cdot e^{-\gamma x}\lim_{n\to+\infty}\frac{x(x+1)\dots(x+n)}{n!}\prod_{k=1}^{n}e^{-\frac{x}{k}}\\ &= x\cdot e^{-\gamma x}\lim_{n\to+\infty}\prod_{k=1}^{n}\left(1+\frac{x}{k}\right)e^{-\frac{x}{k}}\\ &= x\cdot e^{-\gamma x}\prod_{k=1}^{+\infty}\left(1+\frac{x}{k}\right)e^{-\frac{x}{k}}\end{aligned} \tag{13.412}$$

13.4 Equations différentielles

Définition 13.5 *En mathématique, une "équation différentielle" (E.D.) est une relation entre une ou plusieurs fonctions inconnues et leurs dérivées jusqu'à l'ordre n. "L'ordre" d'une équation différentielle correspond au degré maximal de différentiation auquel une des fonctions inconnues y a été soumise.*

Par rapport à notre objectif d'essayer de voir comment les mathématiques décrivent la réalité, les équations différentielles remportent un franc succès, mais sont également la source de bien des soucis. D'abord des difficultés de modélisation —par exemple, le système d'équation différentielles de la relativité générale, des difficultés de résolution —il n'existe pas de méthode générale, puis des difficultés proprement mathématiques, enfin des difficultés liées au fait que certaines équations différentielles ne sont pas stables par nature et donnent des solutions chaotiques[1].

REMARQUE Les équations différentielles sont utilisées pour construire des modèles mathématiques de phénomènes physiques et biologiques, par exemple pour l'étude de la radioactivité ou la mécanique céleste. Par conséquent, les équations différentielles représentent un immense champ d'étude, aussi bien en mathématiques pures qu'appliquées

L'équation différentielle d'ordre n la plus générale peut toujours s'écrire sous la forme :

$$F\left(x,y,y',\dots,y^{(n)}\right)=0 \tag{13.413}$$

Nous ne considérons sur ce site que le cas où x et y sont à valeur dans $\mathbb{R}$. Une solution à une telle E.D. sur l'intervalle $I\subset\mathbb{R}$ est une fonction $y\in C^n(I,\mathbb{R})$ (une fonction $y:I\to\mathbb{R}$ qui est n fois continûment dérivable) telle que pour tout $x\in I$, nous ayons :

$$F\left(x,y(x),y'(x),\dots,y^{(n)}(x)\right)=0 \tag{13.414}$$

REMARQUE Pour des raisons qui seront développés par la suite, nous disons aussi "intégrer l'E.D." au lieu de "trouver une solution à l'E.D.".

REMARQUE Etant donné que tout le site internet est bourré d'exemples d'équations différentielles et de méthodes de résolutions dans les chapitres sur la mécanique, la physique atomique, la cosmologie, l'économétrie, les suites et séries, etc., nous ne ferons pas d'exemples ici et nous intéresserons donc qu'à l'aspect théorique minimal.

13.4.1 Équations différentielles du premier ordre

Une équation différentielle du 1$^{\text{er}}$ ordre est donc une E.D. qui ne fait intervenir que la première dérivée y'.

Définition 13.6 *Une équation différentielle du 1er ordre est dite "E.D. d'ordre 1 à variables séparées" si elle peut s'écrire sous la forme :*

$$f(y)y'=g(x) \tag{13.415}$$

1. Dans le chapitre concernant la dynamique des populations, il exist des des exemples simples flagrants de solutions chaotiques.

Une telle équation différentielle peut s'intégrer facilement : En effet, nous écrivons :

$$y' = \frac{dy}{dx} \tag{13.416}$$

Puis symboliquement :

$$f(y)dy = g(x)dx \Leftrightarrow \int f(y)dy = \int g(x)dx + c^{te} \tag{13.417}$$

REMARQUE Nous écrivons ici explicitement la constante d'intégration arbitraire, qui est implicitement présente dans les intégrales indéfinies, pour ne pas l'oublier.

Il s'agit donc d'abord de trouver des primitives F et G de f et de g, et ensuite d'exprimer y en terme de x (et de C) :

$$F(y) = G(x) + c^{te} \Leftrightarrow y = F^{-1}\big(G(x) + c^{te}\big) \tag{13.418}$$

La constante d'intégration est fixée lorsqu'on demande que pour un donnée, nous ayons une valeur donnée de . Nous parlons alors de "problème aux valeurs initales".

13.4.2 Equation différentielles linéaires

Définition 13.7 *Une équation différentielle d'ordre n est dite "E.D. linéaire" (E.D.L.) si et seulement si elle est de la forme :*

$$L(y) = f(x) \tag{13.419}$$

avec $L(y) = a_0(x)y + a_1(x)y' + a_2(x)y'' + \ldots + a_n(x)y^{(n)}$.

Montrons que L est une application linéaire :

$$L(y+z) = \sum_{i=0}^{n} a_i(x)(y+z)^{(i)} = \sum_{i=0}^{n} a_i(x)y^{(i)} + \sum_{i=0}^{n} a_i(x)z^{(i)} = L(y) + L(z) \tag{13.420}$$

Et pour tout $\lambda \in \mathbb{R}$:

$$L(\lambda y) = \sum_{i=0}^{n} a_i(x)(\lambda y)^{(i)} = \lambda \sum_{i=0}^{n} a_i(x)y^{(i)} = \lambda L(y) \tag{13.421}$$

Définition 13.8 *L'équation différentielle (c'est la plus courante en physique) :*

$$L(y) = 0 \tag{13.422}$$

s'appelle "équation homogène" (E.H.) ou "équation sans second membre" (ESSM) associée à :

$$L(y) = f(x) \tag{13.423}$$

Nous allons maintenant démontrer une propriété importantes des E.H. : l'ensemble $\{S_0\}$ des solutions de E.H. est le noyau de l'application linéaire L (ce qui rappelons-le signifie : $L(S_0) = 0$) et l'ensemble $\{S\}$ des solutions à $L(y) = f(x)$ est donné par :

$$S = y_p + S_0 = \big\{y_p + y_h \big| y_h \in \{S_0\}\big\} \text{ avec } L(y_p) = f(x) \tag{13.424}$$

c'est-à-dire que les solutions de la forme $y = y_p + y_h$, où y_p est une "solution particulière" de $L(y) = f(x)$ et y_h la "solution homogène", parcourent toutes les solutions de l'E.H.

Preuve La première affirmation sera supposée évidente. En ce qui concerne la 2ème partie, toute fonction de la forme $y_p + y_h$ est solution de $L(y) = f(x)$. En effet c'est trivial et cela découle de la définition du concept de noyau (cf. chapitre de Théorie Des Ensembles) :

$$L(y_p + y_h) = L(y_p) + L(y_h) = f(x) + 0 = f(x) \tag{13.425}$$

•

Il existe de nombreuses manières de résoudre les équations différentielles linéaires ou non linéaires de manière exacte ou approchée. Citons les quelques méthodes que nous analyserons plus loin par l'exemple (mais qui se trouvent déjà de nombreuses fois dans le chapitres de physique) :

- La méthode du polynôme caractéristique (voir plus bas)
- La méthode des perturbations (voir plus bas)
- La méthode de variation de la constante ne sera pas présentée car basée sur une hypothèse empirique elle est dangereuse d'usage en physique !

13.4.3 Méthode du polynôme caractéristique

La résolution des équations différentielles simples (à coefficients constants et sans seconde membre la plupart du temps...) utilise une technique faisant appel à un polynôme caractéristique de l'équation différentielle dont nous verrons les détails dans les développements à suivre sur quelques cas particuliers courants en physique.

C'est une méthode relativement simple à mettre en place lorsque nous cherchons les solutions homogènes de l'équation sans second membre (ESSM). Dans le cas contraire, celui de la présence d'un seconde membre, nous additionnons les solutions de l'équation homogènes aux solutions particulières.

Résolution l'équation homogène du premier ordre coefficients constants

Considérons l'E.D.L. à coefficient constant suivante :

$$\frac{dy}{dx} + ay = f(x) \tag{13.426}$$

Nous écrivons son équation homogène associée :

$$\frac{dy}{dx} + ay = 0 \tag{13.427}$$

Ce qui peut s'écrire :

$$\frac{dy}{y} = -adx \tag{13.428}$$

d'où :

$$\ln(y) = -ax + c^{te} \Rightarrow y_h = e^{-ax+c^{te}} = e^{c^{te}} a^{-ax} = Ce^{-ax} \tag{13.429}$$

Il y a derrière cette solution homogène une infinité de solutions : à chaque valeur donnée à C correspond une solution.

Il faut encore à cette solution homogène ajouter la solution particulière y_p et nous disposons pour cela d'une collection de recettes, qui dépendent du type de la fonction $f(x)$ du second membre de l' équation. Nous les verrons au cas par cas dans les différents chapitres de Physique.

Résolution l'équation homogène du second ordre coefficients constants

Considérons l'E.D.L. à coefficient constant suivante :

$$\frac{d^2y}{dx^2} + a\frac{dy}{dx} + by = f(x) \tag{13.430}$$

Nous écrivons son équation homogène associée :

$$\frac{d^2y}{dx^2} + a\frac{dy}{dx} + by = 0 \tag{13.431}$$

dans laquelle la fonction second membre est nulle. Nous pouvons immédiatement entrevoir une solution du type (en s'inspirant de la forme des solutions des E.D. du 1er ordre) :

$$y = e^{K\cdot x+\tau} \tag{13.432}$$

où τ est une constante, ce qui nous donne alors :

$$\lambda K^2 e^{K\cdot x+\tau} + a\lambda K e^{K\cdot x+\tau} + b\lambda e^{K\cdot x+\tau} = 0 \tag{13.433}$$

Ce que nous pouvons simplifier en :

$$K^2 + aK + b = 0 \tag{13.434}$$

Si notre hypothèse de départ est bonne, nous n'avons qu'à résoudre en K cette "équation caractéristique" (ECAR) ou "polynôme caractéristique" de l'équation homogène pour trouver la solution homogène :

$$K_{1,2} = \frac{-a \pm \sqrt{a^2 - 4b}}{2} \tag{13.435}$$

dont les solutions dépendent du signe du discriminant du polynôme caractéristique :

$$\Delta = a^2 - 4b \tag{13.436}$$

Discriminant strictement positif $\Delta > 0$Alors nous savons que le polynôme caractéristique possède deux racines réelles distinctes et nous avons alors :

$$y_h = Ae^{K_1(x+\tau)} + Be^{K_2(x+\delta)} \tag{13.437}$$

où $K_1\tau = c^{te}$ et $K_2\delta = c^{te}$. Nous disons alors que la solution est "retardée" ou "avancée" selon les valeurs de ces constantes. Mais l'essentiel est de remarque que si $y_n(x)$ est solution, alors $y_h(x \pm \Delta x)$ est toujours solution !

Nous parlons alors de "solution générale de l'équation homogène". Il y a derrière ce résultat une infinité de solutions : à chaque valeur donnée aux constantes A, B correspond une solution.

Les physicien écrivent aussi parfois cela sous une forme particulière en posant d'abord :

$$\alpha = \frac{K_1 + K_2}{2} \quad \beta = \frac{K_1 - K_2}{2} \tag{13.438}$$

avec donc :

$$\alpha + \beta = K_1 \quad \alpha - \beta = K_2 \tag{13.439}$$

Et en utilisant les fonctions de trigonométrie hyperbolique (cf. chapitre de Trigonométrie) :

$$\begin{aligned} \cosh(\beta x) &= \frac{e^{\beta x} + e^{-\beta x}}{2} \\ \sinh(\beta x) &= \frac{e^{\beta x} - e^{-\beta x}}{2} \end{aligned} \tag{13.440}$$

d'où finalement la possibilité d'écrire la solution homogène sous la forme (lorsque nous omettons l'avance ou le retard $\delta = \tau = 0$) :

$$y_h = e^{\alpha x}\big(C_1 \cosh(\beta t) + C_2 \sinh(\beta t)\big) \tag{13.441}$$

Par ailleurs, montrons que les solutions de l'ESSM forment un espace vectoriel de dimension 2 [2], puisqu'en effet :

- la fonction nulle $y = 0$ est solution de l'ESSM ;
- la somme ou soustraction des solutions reste solution ;
- les éléments de la base de l'espace vectoriel sont linéairement indépendants.

Posons :

$$y_h = \lambda_1 f_1(x) + \lambda_2 f_2(x) \tag{13.442}$$

Alors :

$$\begin{aligned} y_h' &= \lambda_1 f_1'(x) + \lambda_2 f_2'(x) \\ y_h'' &= \lambda_1 f_1''(x) + \lambda_2 f_2''(x) \end{aligned} \tag{13.443}$$

2. La dimension de l'espace vectoriel correspond donc à l'ordre de l'équation différentielle considérée.

Donc l'équation différentielle à coefficients constants :

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0 \tag{13.444}$$

s'écrit alors :

$$\begin{aligned} ay_h'' + by_h' + cy_h &= a[\lambda_1 f_1'' + \lambda_2 f_2''] + b[\lambda_1 f_1' + \lambda_2 f_2'] + c[\lambda_1 f_1 + \lambda_2 f_2] \\ &= \lambda_1(af_1'' + bf_1' + cf_1) + \lambda_2(af_2'' + bf_2' + cf_2) \\ &= \lambda_1 \cdot 0 + \lambda_2 \cdot 0 \\ &= 0 \end{aligned} \tag{13.445}$$

Donc nous avons bien une structure d'espace vectoriel.

Rappelons que inversement deux fonction sont linéaire dépendantes si :

$$\forall t f_1(x) = \lambda \cdot f_2(x) \tag{13.446}$$

Discriminant nul $\Delta = 0$L' équation caractéristique possède une racine double réelle K.

En allant un peu vite nous dirons alors :

$$y_h = Ae^{K(x+\tau)} \tag{13.447}$$

Il ne faut pas oublier que la base vectorielle doit être formée de deux solutions indépendantes. Donc la deuxième solution est probablement de la forme :

$$y_{h2} = z(x)e^{K(x+\tau)} \tag{13.448}$$

Alors :

$$\begin{aligned} y_{h2}' &= Ke^{K(x+\tau)}z(x) + z'(x)e^{K(x+\tau)} \\ y''h2 &= K^2e^{K(x+\tau)}z(x) + Ke^{K(x+\tau)}z'(x) + Kz'(x)e^{K(x+\tau)} + z''(x)e^{K(x+\tau)} \\ &= K^2e^{K(x+\tau)}z(x) + 2Ke^{K(x+\tau)}z'(x) + z''(x)e^{K(x+\tau)} \end{aligned} \tag{13.449}$$

Si nous l'injectons dans l'ESSM :

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0 \tag{13.450}$$

alors :

$$\begin{aligned} a\left(K^2e^{K(x+\tau)}z + 2Ke^{K(x+\tau)}z' + z''e^{K(x+\tau)}\right) + \\ b\left(e^{K(x+\tau)}z + z'e^{K(x+\tau)}\right) + cze^{K(x+\tau)} = 0 \\ e^{K(x+\tau)}(aK^2 + bK + c)z + e^{K(x+\tau)}(2aK + b)z' + ae^{K(x+\tau)}z'' = 0 \\ (aK^2 + bK + c)z + (2aK + b)z' + az'' = 0 \end{aligned} \tag{13.451}$$

Or :

$$\begin{aligned} aK^2 + bK + c = 0 \\ 2aK + b = 0 \end{aligned} \tag{13.452}$$

Donc :

$$z'' = 0 \Rightarrow z' = c^{te} \Rightarrow z = Bx + c^{te} \tag{13.453}$$

Donc finalement :

$$y_{h2} = \left(Bx + c^{te}\right)e^{K(x+\tau)} \tag{13.454}$$

Ce qui donne pour la solution générale de l'ESSM :

$$y_h = (C_1x + C_2)e^{K(x+\tau)} \tag{13.455}$$

Discriminant négatif $\Delta < 0$L'équation caractéristique possède deux racines complexes conjuguées (cf. chapitre d'Algèbre) :

$$K_1 = \alpha + i\beta, \quad K_2 = \alpha - i\beta \tag{13.456}$$

Dès lors :

$$\begin{aligned} y_h =& Ae^{(\alpha+i\beta)(x+\tau)} + Be^{(\alpha-i\beta)(x+\delta)} \\ =& Ae^{\alpha(x+\tau)}e^{i\beta(x+\tau)} + Be^{\alpha(x+\delta)}e^{-i\beta(x+\delta)} \\ =& e^{\alpha(x+\tau)}Ae^{i\beta(x+\tau)} + e^{\alpha(x+\delta)}Be^{-i\beta(x+\delta)} \\ =& Ae^{\alpha(x+\tau)}\big(\cos(\beta x+\tau\beta) + i\sin(\beta x+\tau\beta)\big) \\ &+ Be^{\alpha(x+\delta)}\big(\cos(\beta x+\delta\beta) - i\sin(\beta x+\delta\beta)\big) \\ =& y_{h1} + y_{y2} \end{aligned} \tag{13.457}$$

Or, si nous cherchons plutôt des solutions réelles, nous pouvons toujours poser A et B égaux tels que :

$$\begin{aligned} y'_h =& \frac{1}{2}(y_{h1} + y_{y2}) + \frac{1}{2i}(y_{h1} - y_{y2}) \\ =& e^{\alpha(x+\tau)}\cos(\beta x+\tau\beta) + e^{\alpha(x+\delta)}\sin(\beta x+\delta\beta) \\ =& e^{\alpha x}e^{\alpha\tau}\cos(\beta x+\tau\beta) + e^{\alpha x}e^{\alpha\delta}\sin(\beta x+\delta\beta) \\ =& e^{\alpha x}\big(A'\cos(\beta x+\tau\beta) + B'\sin(\beta x+\delta\beta)\big) \end{aligned} \tag{13.458}$$

Et si nous posons que le retard, ou l'avance est nulle ($\delta = \tau = 0$), alors nous retrouvons la relation disponible dans la plupart des livres :

$$y'_h = e^{\alpha x}\big(A'\cos(\beta x) + B'\sin(\beta x)\big) \tag{13.459}$$

13.4.4 Méthode régulière des perturbations

Très fréquemment en physique (de pointe), un problème mathématique ne peut pas être résolu de manière exacte. Si la solution est connue il y a parfois une telle dépendance de paramètres que la solution est difficile à utiliser en tant que tel.

Il peut être le cas, cependant, qu'un paramètre identifié, disons par tradition, tel que la solution est disponible est raisonnablement simple pour $\varepsilon = 0$.

Le souci ensuite est de savoir comme la solution est altérée pour un non-nul mais petit quand même. Cette étude est le centre de la théorie des perturbations que nous utilisons par exemple dans le chapitre de relativité générale pour calculer la précession du périhélie de Mercure.

Comme la théorie dans le cadre général est trop complexe par rapport aux objectifs du site, nous nous proposons une approche par l'exemple d'abord avec une simple équation algébrique et ensuite avec ce qui nous intéresse : une E.D.

Théorie perturbative des équations algébriques

Considérons l'équation polynômiale suivante :

$$x^2 - 1 = \varepsilon x \tag{13.460}$$

Nous savons de par notre étude du chapitre d'analyse fonctionnelle, que cette équation polynômiale admet deux racines qui sont trivialement :

$$x_{1,2} = \frac{\varepsilon \pm \sqrt{\varepsilon^2+4}}{2} = \frac{\varepsilon}{2} \pm \sqrt{1+\frac{\varepsilon^2}{4}} \tag{13.461}$$

Pour ε petit, ces racines peuvent être approximées par le premier terme en développement de série de Taylor (cf. chapitre de Suites Et Séries) :

$$x_1 = 1 + \frac{\varepsilon}{2} + \frac{\varepsilon^2}{8} + O(\varepsilon^3) \quad x_2 = -1 + \frac{\varepsilon}{2} - \frac{\varepsilon^2}{8} + O(\varepsilon^3) \tag{13.462}$$

La question et de savoir si nous pouvons obtenir les deux relations précédentes sans à priori de connaissances sur la solution exacte de l'équation polynômiale initiale ? La réponse est bien évidemment affirmative avec l'aide de la théorie des perturbations.

La technique se base en quatre étapes :

1. Dans la première étape, nous assumons que la solution de l'équation polynômiale est un expression du type série de Taylor en . Nous avons alors :

$$x = X_0 + \varepsilon X_1 + \varepsilon^2 X_2 + O(\varepsilon^3) \tag{13.463}$$

où X_0, X_1, X_2 sont bien évidemment à déterminer.

2. Dans la deuxième étape, nous injectons la solution hypothétique dans notre équation polynômiale :

$$\big(X_0 + \varepsilon X_1 + \varepsilon^2 X_2 + O(\varepsilon^3)\big)^2 - 1 = \varepsilon\big(X_0 + \varepsilon X_1 + \varepsilon^2 X_2 + O(\varepsilon^3)\big) \tag{13.464}$$

Comme :

$$\begin{aligned} x^2 =& X_0^2 + 2\big(\varepsilon X_1 + \varepsilon^2 X_2 + O(\varepsilon^3)\big)X_0 + \big(\varepsilon X_1 + \varepsilon^2 X_2 + O(\varepsilon^3)\big)^2 \\ =& X_0^2 + 2\big(\varepsilon X_1 + \varepsilon^2 X_2 + O(\varepsilon^3)\big)X_0 \\ &+ \Big(\big(\varepsilon X_1 + \varepsilon^2 X_2\big)^2 + 2\big(\varepsilon X_1 + \varepsilon^2 X_2\big)O(\varepsilon^3)\varepsilon^3)\Big) \\ \cong& X_0^2 + 2\big(\varepsilon X_1 + \varepsilon^2 X_2\big)X_0 + \big(\varepsilon X_1 + \varepsilon^2 X_2\big)^2 + O(\varepsilon^3) \\ =& X_0^2 + 2\big(\varepsilon X_1 + \varepsilon^2 X_2\big)X_0 + \big(\varepsilon^2 X_1^2 + 2\varepsilon^3 X_1 X_2 + \varepsilon^4 X_2^2\big) + O(\varepsilon^3) \\ \cong& X_0^2 + 2\big(\varepsilon X_1 + \varepsilon^2 X_2\big)X_0 + \big(\varepsilon^2 X_1^2\big) + O(\varepsilon^3) \\ =& X_0^2 + 2\varepsilon X_0 X_1 + \varepsilon^2\big(X_1^2 + 2X_0 X_2\big) + O(\varepsilon^3) \end{aligned} \tag{13.465}$$

et :

$$\varepsilon x = \varepsilon X_0 + \varepsilon^2 X_1 + O(\varepsilon^3) \tag{13.466}$$

Il vient finalement que l'équation polynômiale s'écrit :

$$X_0^2 - 1 + \varepsilon(2X_0 X_1 - X_0) + \varepsilon^2\big(X_1^2 + 2X_0 X_2 - X_1\big) + O(\varepsilon^3) = 0 \tag{13.467}$$

3. Dans la troisième étape nous égalisons successivement les termes avec 0 tel que :

$$\begin{aligned} O(\varepsilon^0) :& X_0^2 - 1 = 0 \\ O(\varepsilon^1) :& 2X_0 X_1 - X_0 = 0 \\ O(\varepsilon^2) :& X_1^2 + 2X_0 X_2 - X_1 = 0 \\ O(\varepsilon^3) :& \ldots \end{aligned} \tag{13.468}$$

4. Quatrième et dernière étape, nous résolvons successivement les équations polynômiales ci-dessus pour obtenir :

$$X_0 = \pm 1 \Rightarrow X_1 = \frac{1}{2} \Rightarrow X_2 = \pm\frac{1}{8} \tag{13.469}$$

En injectant ces résultants dans la solution hypothétique :

$$x = X_0 + \varepsilon X_1 + \varepsilon^2 X_2 + O(\varepsilon^3) \tag{13.470}$$

il est évident d'observer que nous retombons sur la solution certaine :

$$x_1 = 1 + \frac{\varepsilon}{2} + \frac{\varepsilon^2}{8} + O(\varepsilon^3) \quad x_2 = -1 + \frac{\varepsilon}{2} - \frac{\varepsilon^2}{8} + O(\varepsilon^3) \tag{13.471}$$

Théorie perturbative des équations différentielles

La théorie des perturbations est aussi souvent utilisée pour résoudre un bon nombre d'équations différentielles. C'est le cas par exemple en mécanique des fluides, en relativité générale ou en physique quantique.

À nouveau, plutôt que de faire une théorie ultra abstraite et générale, voyons le concept sur un exemple tel que précédemment.

Considérons l'équation différentielle suivante :

$$y'' = -\varepsilon y' - 1 \tag{13.472}$$

ou autrement écrit :

$$y'' + \varepsilon y' = -1 \tag{13.473}$$

avec les conditions aux limites $y(0) = 0, y'(0) = 1$. La résolution exacte est relativement facile à obtenir. D'abord nous commençons par l'équation homogène :

$$y'' + \varepsilon y' = 0 \tag{13.474}$$

C'est donc une équation différentielle linéaire d'ordre 2 avec des coefficients constants, équation qu'il est relativement aisé de résoudre dans le cas général. Soit l'équation :

$$y'' + p\varepsilon y' + qy = 0 \tag{13.475}$$

Supposons que la fonction y qui satisfait cette équation différentielle soit de la forme $y = e^{Kx}$ où K peut être un nombre complexe. Nous avons alors :

$$K^2 e^{Kx} + pK^2 e^{Kx} + qe^{Kx} = 0 \text{ ou } K^2 + pK + q = 0 \tag{13.476}$$

pourvu, bien sûr, que $e^{Kx} \neq 0$. Cette dernière relation est donc l'équation quadratique auxiliaire de l'équation différentielle (polynôme caractéristique). Elle a deux solutions/racines (c'est une simple résolution d'un polynôme du deuxième degré) que nous noterons dans le cas général : K_1, K_2. Ce qui signifie que :

$$K_1^2 e^{K_1 x} + pK_1 e^{K_1 x} + qK_1 e^{K_1 x} = 0 \text{ et } K_2^2 e^{K_2 x} + pK_2 e^{K_2 x} + qK_2 e^{K_2 x} = 0 \tag{13.477}$$

est satisfait pour les deux racines. Si nous faisons la somme puisque les deux sont égales à la même constante :

$$\left(K_1^2 e^{K_1 x} + K_2^2 e^{K_2 x}\right) + p\left(K_1 e^{K_1 x} + K_2 e^{K_2 x}\right) + q\left(K_1 e^{K_1 x} + K_2 e^{K_2 x}\right) = 0 \tag{13.478}$$

Ainsi, il est immédiat que la solution générale de l'équation homogène de y est du type :

$$y_h = Ae^{K_1 x} + Be^{K_2 x} \tag{13.479}$$

où A, B sont bien évidemment des constantes à déterminer. Nous résolvons maintenant le polynôme caractéristique :

$$K^2 + \varepsilon K = 0 \tag{13.480}$$

Il vient immédiatement que :

$$K_{1,2} = \frac{-\varepsilon \pm \sqrt{\varepsilon^2}}{2} \Rightarrow \begin{cases} K_1 = -\varepsilon \\ K_2 = 0 \end{cases} \tag{13.481}$$

Donc :

$$y_h = Ae^{-\varepsilon x} \tag{13.482}$$

Maintenant une solution particulière à :

$$y'' + \varepsilon y' = -1 \tag{13.483}$$

est relativement trivialement une solution du type :

$$y_p = Bx \tag{13.484}$$

où B est bien évidemment une constante à déterminer et qui vaut simplement une fois injectée dans l'équation différentielle :

$$y'' + \varepsilon y' = -1 \Rightarrow 0 + \varepsilon B = 1 \Rightarrow B = -\frac{1}{\varepsilon} \tag{13.485}$$

Soit :

$$y_p = -\frac{x}{\varepsilon} + c^{te} \tag{13.486}$$

D'où finalement la solution générale :

$$y = y_p + y_h = Ae^{-ex} - \frac{x}{\varepsilon} + c^{te} \tag{13.487}$$

Ensuite, avec les conditions initiales $y(0) = 0, y'(0) = 1$ il est très facile de trouver A :

$$y(0) = 0 \Rightarrow Ae^{-e0} - \frac{0}{\varepsilon} + c^{te} = A + c^{te} = 0 \qquad \Rightarrow c^{te} = -A \tag{13.488}$$

et :

$$y'(0) = 0 \Rightarrow -\varepsilon Ae^{-\varepsilon 0} - \frac{1}{\varepsilon} + c^{te} = -\varepsilon A - \frac{1}{\varepsilon} + c^{te} = 0 \Rightarrow A = -\frac{\varepsilon + c^{te}}{\varepsilon^2} \tag{13.489}$$

Il est loisible de choisir $c^{te} = 1$. Donc :

$$\begin{aligned} y =& -\frac{\varepsilon + 1}{\varepsilon^2} e^{-\varepsilon x} - \frac{x}{\varepsilon} + c^{te} \\ =& -\frac{1 + \varepsilon}{\varepsilon^2} e^{-\varepsilon x} - \frac{x}{\varepsilon} + \frac{1 + \varepsilon}{\varepsilon^2} \\ =& \frac{1 + \varepsilon}{\varepsilon^2} \left(1 - e^{-\varepsilon x}\right) - \frac{x}{\varepsilon} \end{aligned} \tag{13.490}$$

Maintenant que nous avons la solution générale, si ε est petit nous pouvons prendre le développement d'ordre 4 en série de MacLaurin de l'exponentielle (cf. chapitre de Suites Et Séries). Tel que :

$$e^{-\varepsilon x} = 1 - \varepsilon x + \frac{1}{2}\varepsilon^2 x^2 - \frac{1}{6}\varepsilon^3 x^3 + O(\varepsilon^4 x^4) \tag{13.491}$$

Injecté dans y cela donne :

$$\begin{aligned} y =& \frac{1 + \varepsilon}{\varepsilon^2} \left(1 - e^{-\varepsilon x}\right) - \frac{x}{\varepsilon} \\ =& \frac{1 + \varepsilon}{\varepsilon^2} \left(\varepsilon x - \frac{1}{2}\varepsilon^2 x^2 + \frac{1}{6}\varepsilon^3 x^3 - O(\varepsilon^4 x^4)\right) - \frac{x}{\varepsilon} \\ =& \frac{1}{\varepsilon^2} \left(\varepsilon x - \frac{1}{2}\varepsilon^2 x^2 + \frac{1}{6}\varepsilon^3 x^3 - O(\varepsilon^4 x^4)\right) \\ & + \frac{1}{\varepsilon} \left(\varepsilon x - \frac{1}{2}\varepsilon^2 x^2 + \frac{1}{6}\varepsilon^3 x^3 - O(\varepsilon^4 x^4)\right) - \frac{x}{\varepsilon} \\ =& \frac{1}{\varepsilon} x - \frac{1}{2} x^2 + \frac{1}{6}\varepsilon x^3 - \frac{1}{24}\varepsilon^2 x^2 + x - \frac{1}{2}\varepsilon x^2 + \frac{1}{6}\varepsilon^2 x^3 - O(\varepsilon^3 x^3) - \frac{x}{\varepsilon} \\ =& x - \frac{1}{2} x^2 + \varepsilon \left(-\frac{1}{2} x^2 + \frac{1}{6} x^3\right) + \varepsilon^2 \left(\frac{1}{6} x^3 - \frac{1}{24} x^2\right) - O(\varepsilon^3 x^3) \end{aligned} \tag{13.492}$$

Maintenant que nous avons ce développement, ce que nous souhaitons montrer c'est qu'à partir d'un développement perturbatif nous pouvons retrouver le même résultat en série et ce sans aucune connaissance préalable sur la solution.

A nouveau, le développement pour cela ce fait en 4 étapes :

1. Dans la première étape, nous assumons que la solution de l'équation différentielle est un expression du type série de Taylor en ε. Nous avons alors :

$$y = y_0 + \varepsilon y_1 + \varepsilon^2 y_2 + O(\varepsilon^3) \tag{13.493}$$

où y_0, y_1, y_2 sont bien évidemment à déterminer.

2. Dans la deuxième étape, nous injectons la solution hypothétique dans notre équation différentielle dans celle-ci avec les conditions initiales et nous développons le tout. D'abord l'équation différentielle :

$$\begin{aligned} y'' + \varepsilon y' + 1 &= \big(y_0'' + \varepsilon y_1'' + \varepsilon^2 y_2'' + O(\varepsilon^3)\big) + \varepsilon\big(y_0' + \varepsilon y_1' + \varepsilon^2 y_2' + O(\varepsilon^3)\big) + 1 \\ &= (y_0'' + 1) + \varepsilon(y_1'' + y_0') + \varepsilon^2(y_2'' + y_1') + O(\varepsilon^3) = 0 \end{aligned} \tag{13.494}$$

ensuite les conditions initiales :

$$\begin{aligned} y(0) &= y_0(0) + \varepsilon y_1(0) + \varepsilon^2 y_2(0) + O(\varepsilon^3) = 0 \\ y'(0) &= \big(y_0'(0) - 1\big) + \varepsilon y_1'(0) + \varepsilon^2 y_2'(0) + O(\varepsilon^3) = 0 \end{aligned} \tag{13.495}$$

3. Dans la troisième étape nous égalisons successivement les termes avec 0 tel que :

$$\begin{array}{lllll} O(\varepsilon^0): & y_0'' + 1 = 0 & y_0(0) = 0 & y_0'(0) - 1 = 0 \\ O(\varepsilon^1): & y_1'' + y_0' = 0 & y_1(0) = 0 & y_1'(0) = 0 \\ O(\varepsilon^2): & y_2'' + y_1' = 0 & y_2(0) = 0 & y_2'(0) = 0 \\ O(\varepsilon^3): & \cdots & & \end{array} \tag{13.496}$$

4. Dans la quatrième étant nous résolvons les équations différentielles listées précédemment (si vous ne voyez pas comment nous les résolvons n'hésitez pas à nous contacter !) :

$$\begin{aligned} y_0 &= x - \frac{1}{2}x^2 \\ y_1 &= -\frac{1}{2}x^2 + \frac{1}{6}x^3 \\ y_2 &= \frac{1}{6}x^3 - \frac{1}{24}x^4 \end{aligned} \tag{13.497}$$

En injectant ces relations dans la solution supposée développée en série de Taylor et injectée dans l'équation différentielle :

$$y = y_0 + \varepsilon y_1 + \varepsilon^2 y_2 + O(\varepsilon^3) \tag{13.498}$$

Nous retombons sur :

$$y = x - \frac{1}{2}x^2 + \varepsilon\big(-\frac{1}{2}x^2 + \frac{1}{6}x^3\big) + \varepsilon^2\big(\frac{1}{6}x^3 - \frac{1}{24}x^4\big) - O(\varepsilon^3 x^3) \tag{13.499}$$

13.4.5 Systèmes d'équations différentielles

Voyons maintenant des développements qui vont aussi bien être utiles en physique quantique que dans la résolution de systèmes d'équations différentielles (et particulièrement une qui est connue en théorie du chaos !).

Avant cela, il va nous falloir introduire le concept d'exponentialisation d'une matrice :

L'ensemble des matrices $n \times n$ à coefficients $\mathbb{C}$ dans noté $M_n(\mathbb{C})$ est un espace vectoriel pour l'addition des matrices et la multiplication par un scalaire. Nous notons I la matrice identité.

Nous admettrons qu'une suite de matrices A_n convergent vers une matrice A si et seulement si les suites de coefficients des matrices convergent vers les coefficients correspondent de A.

Exemple 13.2 *Dans $M_2\mathbb{C}$ la suite de matrices :*

$$\begin{pmatrix} \frac{1}{n} & 2 \\ 0 & \frac{n+1}{n} \end{pmatrix} \tag{13.500}$$

converge vers :

$$\begin{pmatrix} 0 & 2 \\ 0 & 1 \end{pmatrix} \tag{13.501}$$

lorsque $n \to \infty$.

Si $x \in \mathbb{C}$, nous avons vus lors de notre étude des nombres complexes (cf. chapitre sur les Nombres) que la série :

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{13.502}$$

converge et sa limite est notée e^x. En fait ici il n'y a aucune difficulté à remplacer x par une matrice A puisque nous savons (nous l'avons montré lors de notre étude des nombres complexes) que tout nombre complexe peut s'écrire sous la forme (le corps des nombres complexes est isomorphe au corps des matrices réelles carrées 2×2) :

$$x = a + ib = a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} b = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{13.503}$$

et qu'un nombre complexe au carré est équivalent à mettre sa forme matricielle au carré :

$$x^2 = (a+ib)(a+ib) = a^2 + aib + aib - b^2 = (a^2 - b^2) + 2abi = \begin{bmatrix} a^2 - b^2 & 2ab \\ 2ab & a^2 - b^2 \end{bmatrix} \tag{13.504}$$

Effectivement :

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}^2 = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} a^2 - b^2 & 2ab \\ 2ab & a^2 - b^2 \end{bmatrix} \tag{13.505}$$

Nous définissons alors l'exponentielle d'une matrice $A \in M_n(\mathbb{C}$ comme la matrice limite de la suite :

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} \tag{13.506}$$

Si la matrice A est diagonale il est évident que son exponentielle est facile à calculer. En effet, si :

$$A = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \Rightarrow A^k = \begin{bmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{bmatrix} \tag{13.507}$$

Par suite :

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} \begin{bmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{\infty} \frac{\lambda_1^k}{k!} & & 0 \\ & \ddots & \\ 0 & & \sum_{k=0}^{\infty} \frac{\lambda_n^k}{k!} \end{bmatrix} = \begin{bmatrix} e^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n} \end{bmatrix} \tag{13.508}$$

Or, il apparaît évident qu'une matrice non diagonale va être beaucoup plus compliquée à traiter ! Nous allons alors utiliser la technique de diagonalisation soit une réduction des endomorphismes (cf. chapitre d'Algèbre Linéaire).

Alors, remarquons que si $S \in M_n(\mathbb{C})$ est inversible et si $A \in M_n(\mathbb{C})$ alors :

$$e^{SAS^{-1}} = Se^AS^{-1} \tag{13.509}$$

Ceci découle du fait que (penser au changement de base d'une application linéaire comme ce qui a été étudié dans le chapitre d'Algèbre Linéaire) :

$$\left(SAS^{-1}\right)^k = SA\underbrace{S^{-1}S}_{=1}A\underbrace{S^{-1}S}_{=1}AS^{-1}\ldots SAS^{-1} = SA^kS^{-1} \tag{13.510}$$

Donc :

$$e^{SAS^{-1}} = \sum_{k=0}^{\infty} \frac{SA^kS^{-1}}{k!} = S\left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right)S^{-1} \tag{13.511}$$

Ce développement va nous permettre de ramener le calcul de l'exponentielle d'une matrice diagonalisable à la recherche de ses valeurs propres et de ses vecteurs propres.

Exemple 13.3 *Calculons e^A où :*

$$A = \begin{pmatrix} 3 & -4 \\ -6 & 1 \end{pmatrix} \tag{13.512}$$

Les valeurs propres de A sont $\lambda_1 = -3, \lambda_2 = 7$, et les vecteurs propres associés sont :

$$\vec{v}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad \vec{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{13.513}$$

Effectivement :

$$\begin{pmatrix} 3 & -4 \\ -6 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = -3\begin{pmatrix} 2 \\ 3 \end{pmatrix} \text{ et } \begin{pmatrix} 3 & -4 \\ -6 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 7\begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{13.514}$$

En posant :

$$S = \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} \tag{13.515}$$

Nous avons :

$$A = S\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}S^{-1} = S\begin{pmatrix} -3 & 0 \\ 0 & 7 \end{pmatrix}S^{-1} \tag{13.516}$$

avec :

$$S^{-1} = S\begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{3}{5} & -\frac{2}{5} \end{pmatrix} \tag{13.517}$$

Par conséquent :

$$e^A = e^{S\begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix}S^{-1}} = Se^{\begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix}}S^{-1} = S\begin{pmatrix} e^3 & 0 \\ 0 & e^7 \end{pmatrix}S^{-1} = \frac{1}{5}\begin{pmatrix} 2e^{-3}+3e^7 & 2e^{-3}-2e^7 \\ 2e^{-3}-3e^7 & 2e^{-3}+2e^7 \end{pmatrix} \tag{13.518}$$

Maintenant, rappelons que dans le cas des nombres réels nous savons que si $x, y \in \mathbb{R}$ alors $e^{x+y} = e^x e^y$. Dans le cas des matrices nous pouvons que si $A, B \in M_n(\mathbb{C})$ sont deux matrices qui commutent entre-elles c'est-à-dire telles que $AB = BA$. Alors $e^{A+B} = e^A e^B$.

La condition de commutativité vient au fait que l'addition dans l'exponentielle est elle commutative. La démonstration est donc intuitive.

Un corollaire important de cette proposition est que pour toute matrice $A \in M_n(\mathbb{C})$, e^A est inversible. En effet les matrices A et $-A$ commutent, par conséquent :

$$e^A e^{-A} = e^{A-A} = e^0 = I \tag{13.519}$$

Nous rappelons qu'une matrice A à coefficients complexes est unitaire si :

$$AA^\dagger = I \tag{13.520}$$

La proposition suivante nous servira par la suite.

Montrons que si A est une matrice hermitienne (dite aussi "autoadjointe") (cf. chapitre d'Algèbre Linéaire) alors pour tout $t \in \mathbb{R}, e^{itA}$ est unitaire.

Preuve

$$\left(e^{itA}\right)^\dagger = e^{(itA)^\dagger} = e^{-itA^\dagger} = e^{-itA} \tag{13.521}$$

Donc :

$$\left(e^{itA}\right)^\dagger \cdot e^{itA} = e^{-itA}e^{itA} = e^{it(A-A)} = I \tag{13.522}$$

•

Rappelons que cette condition pour une matrice autoadjointe est liée à la définition de groupe unitaire d'ordre n (cf. chapitre d'Algèbre Ensembliste).

Une des premières applications de l'exponentielle de matrices est la résolution des équations différentielles ordinaires. En effet, de l'équation différentielle linéaire ci-dessous avec comme condition initiale $y(0) = y_0$ et où A est une matrice :

$$y'(t) + Ay(t) = 0 \tag{13.523}$$

la solution est donnée (cf. chapitre de Calcul Différentiel et Intégral) par :

$$y(t) = Ce^{-At} \tag{13.524}$$

Nous retrouvons fréquemment ce genre de systèmes d'équations différentielles en biologie (dynamique des populations), en astrophysique (étude des plasmas) ou en mécanique des fluides (théorie du chaos) ainsi que mécanique classique (systèmes couplés), en astronomie (orbites couplées), en électrotechnique, etc.

Exemple 13.4 *Supposons que nous ayons le système d'équations différentielles suivant :*

$$\begin{aligned} x' &= 3x - 4y \\ y' &= -6x + y \end{aligned} \tag{13.525}$$

La matrice associée est alors :

$$A = \begin{pmatrix} 3 & 4 \\ -6 & 1 \end{pmatrix} \tag{13.526}$$

et son exponentielle (voir les développement faits plus haut) :

$$e^{-At} = \frac{1}{5}\begin{pmatrix} 2e^{-3t} + 3e^{7t} & 2e^{-3t} - 2e^{7t} \\ 2e^{-3t} - 3e^{7t} & 2e^{-3t} + 2e^{7t} \end{pmatrix} \tag{13.527}$$

La solution générale du système est donc :

$$\begin{pmatrix} x \\ y \end{pmatrix} = C_1\begin{pmatrix} 2e^{-3t} + 3e^{7t} \\ 2e^{-3t} - 3e^{7t} \end{pmatrix} + C_2\begin{pmatrix} 2e^{-3t} - 2e^{7t} \\ 2e^{-3t} + 2e^{7t} \end{pmatrix} \tag{13.528}$$

Nous avons donc :

$$\begin{aligned} x &= C_1(2e^{-3t} + 3e^{7t}) + C_2(2e^{-3t} - 2e^{7t}) = A'e^{-3t} + B'e^{7t} \\ y &= C_1(2e^{-3t} - 3e^{7t}) + C_2(2e^{-3t} + 2e^{7t}) = D'e^{-3t} + E'e^{7t} \end{aligned} \tag{13.529}$$

Après recherche des constantes nous trouvons :

$$\begin{aligned} E' - 6B' = 7E' &\Rightarrow -6E' = 6B' \Rightarrow E' = -B' \\ -6A' + F' = -3F' &\Rightarrow 4F' = 6A' \Rightarrow F' = \frac{3}{2}A' \end{aligned} \tag{13.530}$$

ce qui nous donne finalement :

$$\begin{aligned} x &= A'e^{-3t} + B'e^{7t} \\ y &= \frac{3}{2}A'e^{-3t} - B'e^{7t} \end{aligned} \tag{13.531}$$

14

Suites et séries

Les suites et séries ont une très grande importance dans les mathématiques pures c'est la raison pour laquelle nous y consacrons un chapitre entier. Nous les retrouverons par ailleurs souvent en physique lorsque nous aurons besoin de faire quelques approximations mineures (...) ainsi qu'en économétrie pour le calcul des rentes. Il conviendra cependant de la part du lecteur de ne pas confondre dans ce qui va suivre le concept de "suite" de celui de "série" qui tout en étant similaires sur le fond ne s'analysent mathématiquement pas toujours de la même manière.

Nous avons souhaité dans ce chapitre rester dans des choses simples sans trop partir dans les concepts topologiques des suites et séries. Cependant, la personne intéressée par des définitions plus rigoureuses pourra se reporter dans le chapitre traitant des fractales (section d'informatique) et de topologie ou de nombreux concepts sur le suites sont définis (supremum, infimum, sous-suite, théorème de Bolzano-Weierstrass, etc.).

14.1 Suites

Définition 14.1 *Une "suite" d'un ensemble est une famille d'éléments indexée par l'ensemble des entiers naturels (cf. chapitre sur les Nombres) ou par une partie de celui-ci. De manière vulgarisée, nous disons qu'une suite est une liste d'objets mis en ordre, chacun ayant un numéro d'ordre. Nous notons classiquement une suite par :*

$$(u_n) \text{ ou } (u_n)_{n\in\mathbb{N}} \tag{14.1}$$

Avant de voir quelques exemples de familles de suites qui seront utilisées dans les différents chapitres du site (dynamiques des populations, économétrie, physique nucléaire, etc.) voyons un petit paquet de définitions comme il est de tradition en mathématique...

Définition 14.2 *Des nombres (en suite) sont en "progression arithmétique" si la différence de deux termes consécutifs est constante.*

Définition 14.3 *Des nombres (en suite) sont en "progression géométrique" si le rapport de deux termes consécutifs est constant.*

Définition 14.4 *Des nombres (en suite) sont en "progression harmonique" si les inverses de deux termes consécutifs sont en progression arithmétique.*

Dès lors, une "suite" est arithmétique, géométrique, harmonique si ses termes sont respectivement en progression arithmétique, géométrique, harmonique et c est la moyenne arithmétique, géométrique, harmonique de a et b si les nombres a, b, c sont en progression arithmétique, géométrique, harmonique.

Remarque Pour les définitions des moyennes citées ci-dessus voir le chapitre de Statistiques

Définition 14.5 *Une "suite majorée", est une suite tel qu'il existe un réel M tel que $\forall n \, in \mathbb{N}, u_n \leq M$*

Définition 14.6 *Une "suite minorée", est une suite tel qu'il existe un réel M tel que $\forall n \, in \mathbb{N}, u_n \geq M$*

Définition 14.7 *Une "suite bornée", est une suite tel qu'elle est à la fois majorée et minorée.*

Définition 14.8 *Une suite est appelée "suite croissante" si* $\forall n\ in\mathbb{N}, u_{n+1} - u_n \geq 0$

Définition 14.9 *Une suite est appelée "suite décroissante" si* $\forall n\ in\mathbb{N}, u_{n+1} - u_n \leq 0$

Définition 14.10 *Une suite est "suite constante stationnaire" si* $\forall n\ in\mathbb{N}, u_{n+1} = u_n$

14.1.1 Suites arithmétiques

Définition 14.11 *Nous disons que des nombres ou que des "termes" en progression forment une "suite arithmétique" lorsque leurs valeurs numériques différent d'une valeur r appelée la "raison" de la suite.*

Ainsi, la suite :

$$n, n+1, n+2, n+3, \cdots, n+i \tag{14.2}$$

est une suite arithmétique de raison $r = 1$.
La suite :

$$n + 0 \cdot x, n + 1 \cdot x, n + 2 \cdot x, n + 3 \cdot x, \cdots, n + i \cdot x \tag{14.3}$$

est une suite arithmétique de raison $r = x$, etc.
Ainsi, si nous notons par u_n un terme quelconque de la suite (u_n) de raison r, nous avons :

$$u_0 = c^{te}, u_1 = u_0 + r, u_2 = u_1 + r = u_0 + 2r, \cdots, u_n = u_0 + n \cdot r \tag{14.4}$$

Nous avons les propriétés suivantes pour un tel type de suite :
P1. Un terme dont le rang est la moyenne arithmétique des rangs de deux autre termes est la moyenne arithmétique de ces deux termes.

Preuve Considérons maintenant (u_n) une suite arithmétique de raison r donné selon le développement précédent :

$$\boxed{u_n = u_0 + n \cdot r} \tag{14.5}$$

et soient $a, b, k \in \mathbb{N}$ tels que $a + b = 2k$, nous avons alors :

$$u_a + u_b = u_0 + (a)r + u_0 + (b)r = 2u_0 + 2\left(\frac{a+b}{2}\right)r = 2\left(u_0 + \left(\frac{a+b}{2}\right)r\right) = 2u_{\frac{a+b}{2}} = 2u_k \tag{14.6}$$

et donc :

$$u_k = \frac{u_a + u_b}{2} \text{ avec } k = \frac{a+b}{2} \tag{14.7}$$

•

P2. Pour trois termes consécutifs en progression arithmétique, le deuxième terme est la moyenne arithmétique des deux autres.

Preuve

$$u_k = \frac{u_a + u_b}{2} = u_{n+1} = \frac{u_n + u_{n+2}}{2} \text{ avec } k = \frac{(a+b)}{2} = \frac{(n)+(n+2)}{2} = \frac{2n+2}{2} = n+1 \tag{14.8}$$

•

14.1.2 Suites harmonique

Définition 14.12 *Nous disons que des nombres $(\frac{1}{a}, \frac{1}{b}, \frac{1}{c}, \cdots)$ forment une "suite harmonique" lorsque leurs inverses sont en progression arithmétique. Nous représentons cette progression par :*

$$\frac{1}{a}, \frac{1}{b}, \frac{1}{c}, \cdots, \frac{1}{h}, \frac{1}{k}, \frac{1}{l} \tag{14.9}$$

où $a, b, c, ..., h, k, l$ désignent des termes au dénominateur en progression arithmétique de raison r. D'ailleurs, nous supposerons, dans ce qui suit, qu'il n'y a aucun dénominateur nul.

En partageant cette série en groupes renfermant successivement 2^n termes, nous observons que chacun de ceux-ci est plus grand que le dernier de son groupe :

$$\frac{1}{1} > \frac{1}{2} + \frac{1}{3} > \frac{1}{4} + \frac{1}{5} > \frac{1}{6} + \frac{1}{7} \tag{14.10}$$

et que la somme des termes de chaque groupe est plus grande que $\frac{1}{2}$. La somme des termes de la série augmente donc indéfiniment ; nous disons alors que la série est une "série divergente" (nous reviendrons plus en détail sur ces concepts de convergence et divergence plus bas).

14.1.3 Suites géométriques

Définition 14.13 *Une "suite géométrique" est une suite de nombres tels que chacun d'eux est égal au précédent n multiplié par un nombre constant q que nous appelons la "raison" de la progression. Nous désignerons par :*

$$n, \cdot n, n^2 \cdot n, q^3 \cdot n, \cdots \tag{14.11}$$

Ainsi, si nous notons par u_n un terme quelconque de la suite (u_n), nous avons (trivial) :

$$\boxed{u_n = u_0 \cdot q^n} \tag{14.12}$$

Voici quelques propriétés pour un tel type de suite (sans démonstration pour l'instant... sauf demande car triviales pour la plupart) :

P1. (triviale) Le quotient de deux termes d'une même suite est une puissance de la raison dont l'exposant égale la différence des rangs des deux termes (simple rapport de termes de puissance).

P2. (triviale) Si nous multiplions ou divisons terme à terme deux suites géométriques, nous obtenons une troisième suite géométrique dont la raison égale le produit (respectivement le quotient) des raisons des progressions données (simple opération avec les raisons des deux séries d'origine).

P3. Dans une suite géométrique, un terme dont le rang est la moyenne arithmétique des rangs de deux autre termes est la moyenne géométrique (cf. chapitre de Statistiques) de ces deux termes (relisez plusieurs fois au besoin).

Preuve Soit une suite géométrique réelle positive de raison q, nous avons :

$$u_n = u_0 \cdot q^n \tag{14.13}$$

Soit a, b deux termes de la suite géométrique, nous avons alors :

$$u_a \cdot u_b = u_0^2 \cdot q^a \cdot q^b = \left(u_0 q^{\frac{a+b}{2}}\right)^2 = u_{\frac{a+b}{2}}^2 = u_k^2 \tag{14.14}$$

et ainsi :

$$\boxed{u_k \sqrt{u_a \cdot u_b}} \tag{14.15}$$

•

Nous avons comme corollaire que pour trois termes consécutifs en progression géométrique, le deuxième terme est la moyenne géométrique des deux autres.

Preuve

$$u_k = \sqrt{u_a \cdot u_b} = u_{n+1} = \sqrt{u_n \cdot u_{n+2}} \tag{14.16}$$

avec :

$$k = \frac{(a+b)}{2} = \frac{(n)+(n+2)}{2} = \frac{2n+2}{2} = n+1 \tag{14.17}$$

•

Il existe cependant quelques suites particulières qui ont des propriétés particulières que nous retrouvons très fréquemment en mathématique ou physique théorique. Sans trop entrer dans les détails, voici une petite liste (non exhaustive de ces dernières) :

14.1.4 Suite de Cauchy

Il est souvent intéressant pour le mathématicien, autant que pour le physicien, de connaître les propriétés d'une suite ayant une type de progression donnée. La propriété la plus importante étant la limite vers laquelle elle tend.

REMARQUE Le lecteur qui n'est pas à l'aise avec la topologie peut sauter le texte qui va suivre en attendant... et celui qui souhaite en savoir plus sur les suites de Cauchy peut se reporter au chapitre de topologie et particulièrement dans le chapitre consacré aux fractales (section d'informatique théorique).

Définition 14.14 *Soit (X, d) un espace métrique (cf. chapitre de Topologie), nous disons que la suite :*

$$(x_n)_{n\in\mathbb{N}}, x_n \in X \forall n \in \mathbb{N} \tag{14.18}$$

converge vers $x \in X$ si par définition :

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ tel que } n \geq N \Rightarrow d(x_n, x) \leq \varepsilon \tag{14.19}$$

En d'autres termes plus nous avançons dans la suite, plus les points sont proches (au sens de la métrique d) les uns des autres.

Nous disons également par définition que la suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de X est une "suite de Cauchy" si :

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ tel que } \forall n, m \geq N \text{ on a } d(x_n, x_m) \leq \varepsilon \tag{14.20}$$

Il est clair alors que toute suite convergente est une suite de Cauchy (bon il y a quelques subtilités auxquelles nous ne ferons pas référence pour l'instant).

REMARQUE Ce critère facilite certaines démonstrations car il permet de montrer l'existence d'une limite sans faire intervenir sa valeur, en général inconnue.

Maintenant, montrons qu'une suite convergente est de Cauchy.

Preuve Soit une suite u_n convergeant vers l (qui nous est inconnu donc !) et $\varepsilon > 0$ (choisi au hasard). Il existe alors selon la définition d'une suite convergente, $N \in \mathbb{N}$ tel que :

$$n \geq N \Rightarrow d(u_n, l) \leq \frac{\varepsilon}{2} \tag{14.21}$$

le choix d'écrire $\frac{\varepsilon}{2}$ est complètement arbitraire mais au fait nous anticipons juste le résultat de la démonstration afin que celui-ci soit plus esthétique.

Alors pour $p, q \geq N$ (au fait connaître le N en question importe peu puisque cela doit marcher pour n'importe lequel... bon n'oublions pas quand même que N dépend de $\frac{\varepsilon}{2}$) nous avons selon l'inégalité triangulaire :

$$d(u_n, u_q) \leq d(u_p, l) + d(l, u_q) \tag{14.22}$$

et puisque $d(u_p, l) \leq \frac{\varepsilon}{2}$:

$$d(u_n, u_q) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \tag{14.23}$$

ce qui revient à écrire :

$$d(u_n, u_q) \leq \varepsilon \tag{14.24}$$

C'est peut être un peu abstrait alors voyons un exemple avec le suite harmonique (divergente comme nous le savons déjà) $u_n = \sum_{k=1}^{n} \frac{1}{k}$. D'abord, rien ne nous interdit de prendre $n \geq 2$(sinon cela va être dur de faire une différence entre deux termes...).

Dès lors nous prenons la distance euclidienne :

$$d(u_p, u_q) = d(u_{2n}, u_n) = u_{2n} - u_n = \sum_{k=1}^{2n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k} = \sum_{k=n+1}^{2n} \frac{1}{k} \tag{14.25}$$

D'abord le lecteur remarquera que dans tous les cas $k \leq 2$ puisque compris entre $n+1$ et $2n$. Ce qui nous amène à pouvoir écrire :

$$\frac{1}{k} \geq \frac{1}{2n} \quad \forall n+1 \leq k \leq 2n \tag{14.26}$$

Donc à partir de cette égalité il vient automatique que chaque terme de la somme de gauche ci-dessous sera plus grand que chaque terme de la somme de droite suivant :

$$\sum_{k=n+1}^{2n} \frac{1}{k} \geq \sum_{k=n+1}^{2n} \frac{1}{2n} \text{ avec } \sum_{k=n+1}^{2n} \frac{1}{2n} = \frac{1}{2} \tag{14.27}$$

maintenant l'idée est de voir que la somme de gauche est donc plus grande ou égale à $\varepsilon = \frac{1}{2}$ et cela quelque soit n. Ainsi, l'idée c'est que nous ayons trouvé un epsilon pour lequel le critère de Cauchy est mis en défaut. Car dans le cas contraire nous aurions du avoir :

$$\sum_{k=n+1}^{2n} \frac{1}{k} \leq \frac{1}{2} \tag{14.28}$$

dont la suite n'est pas convergente. •

Donc, ce n'est pas parce que des points se rapprochent les uns des autres qu'ils convergent vers un point, car ce point n'existe peut-être pas.

Exemple 14.1 *Le meilleur exemple est certainement le suivant : prenons $X = \mathbb{Q}$ et $d(p, q) = |p - q|$. Soit z un nombre irrationnel et $q_j \in \mathbb{Q}$, avec $|z - q_j| \leq \frac{1}{j}$.*

Les $q_1, q_2, \cdots$ forment une suite de Cauchy. En effet :

$$|q_m - q_n| \leq |z - q_m| + |z - q_n| \leq \frac{1}{m} + \frac{1}{n} \tag{14.29}$$

et donc $|q_m - q_n|$ si $m, n \geq \frac{1}{2\varepsilon}$. Nous avons donc trouvé un N qui satisfait à notre définition d'une suite de Cauchy. Or cette suite ne converge pas dans $\mathbb{Q}$ sinon z serait rationnel.

Remarque Les mathématiciens utilisent ce fait pour définir l'ensemble des irrationnels en utilisant quelques concepts topologique supplémentaires.

Nous venons de voir qu'une suite de Cauchy n'est pas forcément une suite convergente dans X. La réciproque toutefois est vraie : toute suite convergente est une suite de Cauchy.

14.1.5 Suite de Fibonacci

Si nous calculons une suite de nombres commençant par 0 et 1, de telle sorte que chaque terme soit égale à la somme des deux précédents, nous pouvons former la suite :

$$0, 1, 1, 2, 3, 5, 8, 13, 21 \cdots \tag{14.30}$$

par conséquent, si nous désignons les différents termes par :

$$u_0, u_1, u_2, \cdots \tag{14.31}$$

nous avons la loi de formation :

$$u_{n+2} = u_{n+1} + u_n \tag{14.32}$$

La suite de Fibonacci possède des propriétés nombreuses fort intéressantes, qui seront développées ultérieurement. Cette suite a été étudiée par Albert Girard (1633) dans la dernière annotation de sa traduction du 5ème et du 6ème livre de l'Arithmétique de Diophante. Il démontre que le rapport de deux termes consécutifs de la suite s'approche de plus en plus du rapport du côté du décagone régulier inscrit dans la circonférence au rayon de cette circonférence. Dans un mémoire présenté à l'Académie des Sciences de Paris en 1844, Lamé indique l'application que l'on peut faire de cette suite à la détermination d'une limite supérieure du nombre des divisions à faire dans la recherche du plus grand commun diviseur de deux nombres entiers ; l'année précédente Binet en avait montré l'emploi pour le dénombrement des combinaisons discontigües. Cette même suite a encore été étudiée par Plana en 1859. Il s'agit cependant de la première "suite récurrente" connue (d'où le fait que nous en parlons aussi).

Nous utilisons également ce genre de suite pour montrer l'utilité du principe d'induction présenté dans le chapitre de la théorie des nombres se trouvant dans la section d'arithmétique.

14.2 Séries

Le physicien a souvent besoin pour résoudre simplement et formellement des problèmes, d'approximer certains "termes" (cf. chapitre de Théorie De La Démonstration) de ses équations. Pour cela, il utilisera les propriétés de certaines séries.

Il existe, une quantité phénoménale de séries et de théories gravitant autour de ces dernières, mais nous citerons en particulier les séries de Taylor (utilisées un peu partout), les séries de Fourier (théorie du signal et en mécanique ondulatoire) et les séries ou fonctions de Bessel (physique nucléaire) dont nous ferons une étude sommaire ici.

Définition 14.15 *Soit donnée une suite numérique infinie :*

$$a_1, a_2, a_3, \cdots, a_n, \cdots \tag{14.33}$$

L'expression :

$$a_1 + a_2 + a_3 + \cdots + a_n + \cdots \tag{14.34}$$

est appelée "série numérique".

Définition 14.16 *La somme partielle des n premiers termes de la série est appelée "somme partielle" et notée s_n :*

$$s_n = a_1 + a_2 + a_3 + \cdots + a_n \tag{14.35}$$

Si la limite notée S suivante existe et est finie :

$$S = \lim_{n\to\infty} S_n \tag{14.36}$$

nous l'appelons la "somme de la série" et nous disons que la "série converge" (elle est donc de Cauchy). Cependant, si la limite n'existe pas, nous disons que la "série diverge" et n'a pas de somme (pour plus de détails voir le sous-chapitre plus loin traitant des critères de convergence).

Montrons par ailleurs que si $S = \sum_{n\geq 0} u_n$ est une série numérique convergente alors :

$$\lim_{n\to\infty} u_n = 0 \tag{14.37}$$

Preuve Nous supposons d'abord que $S = \sum_{k\geq 0} u_k$ est bien une série convergente et notons par S sa limite. Posons :

$$s_n = \sum_{k=0}^{n} u_k \qquad s_{n-1} = \sum_{k=0}^{n-1} u_k \tag{14.38}$$

Alors :

$$s_n - s_{n-1} = u_n \tag{14.39}$$

Or, si la série est convergente :

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} s_{n-1} = S \tag{14.40}$$

Donc :

$$\lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = \lim_{n\to\infty} (s_n - s_{n-1}) = S - S = 0 \tag{14.41}$$

•

Voyons comment calculer la somme partielle des quelques séries classiques :

14.2.1 Série de Gauss

Les séries arithmétiques de Gauss sont l'expression de la somme de n premiers entiers non nuls élevés à une puissance donnée sous une forme condensée. L'application de cette forme condensée de série à une utilité pratique en physique lorsque l'on souhait simplifier l'expression de certains résultats.

Gauss avait trouvé une méthode séduisante en 1786 pour déterminer cette expression lorsqu'il avait 9 ans (...) :

$$\begin{aligned} &S_1 = 1 + 2 + 3 + \cdots + (n-1) + n \\ &S_1 = n + (n-1) + (n-2) + \cdots + 2 + 1 \\ +&\overline{2S_1 = (n+1) + (n+1) + (n+1) + \cdots + (n+1) + (n+1)} \end{aligned} \tag{14.42}$$

En simplifiant, nous trouvons facilement :

$$\boxed{S_1 = \frac{n(n+1)}{2}} \tag{14.43}$$

pour $n \geq 0$.

Nous pouvons continuer ainsi pour des ordres supérieurs (nous les présentons non en tant qu'exercices mais parce que ces relations sont utiles !) :

Calculons maintenant la somme des n premiers carrés (toujours non nuls). Posons :

$$\begin{aligned} S_1 &= 1+2+3+\cdots+n \\ S_2 &= 1^2+2^2+\cdots+n^2 \end{aligned} \tag{14.44}$$

nous savons que (binôme de Newton) :

$$(k+1)^3-k^3 = 3k^2+3k+1 \tag{14.45}$$

nous pouvons donc écrire et ajouter membre à membre les n égalités suivantes :

$$\begin{aligned} 2^3-1^3 &= 3\cdot 1^2+3\cdot 1+1 \\ 3^3-2^3 &= 3\cdot 2^2+3\cdot 2+1 \\ 4^3-3^3 &= 3\cdot 3^2+3\cdot 3+1 \\ &\cdots \\ (n+1)^3-n^3 &= 3\cdot n^2+3\cdot n+1 \\ \hline (n+1)^3-1 &= 3\cdot(1^2+2^3+\cdots+n^2)+3\cdot(1+2+\cdots+n)+n \\ (n+1)^3-1 &= 3S_2+3\cdot S_1+n \end{aligned} \tag{14.46}$$

Avec quelques manipulations algébriques élémentaires :

$$\begin{aligned} 3S_2 &= (n+1)^3-1-3S_1-n = (n+1)^3-1-3\frac{n(n+1)}{2}-n \\ &= (n+1)^3-(n+1)-3\frac{n(n+1)}{2} = \frac{(n+1)}{2}\left[2(n+1)^2-2-3n\right] \end{aligned} \tag{14.47}$$

d'où :

$$\begin{aligned} S_2 &= \frac{(n+1)}{6}\left[2n^2+4n+2-2-3n\right] \\ &= \frac{(n+1)}{6}(2n^2+n) \end{aligned} \tag{14.48}$$

Finalement :

$$\boxed{S_2 = \frac{n(n+1)(2n+1)}{6}} \tag{14.49}$$

Terminons avec la somme des n premiers cubes (non nuls). Le principe étant le même que précédemment, nous posons :

$$\begin{aligned} S_1 &= 1+2+\cdots+n \\ S_2 &= 1^2+2^2+\cdots+n^2 \\ S_3 &= 1^3+2^3+\cdots+n^3 \end{aligned} \tag{14.50}$$

Nous savons par ailleurs que (binôme de Newton) :

$$(k+1)^4-k^4 = 4k^3+6k^2+4k+1 \tag{14.51}$$

Nous obtenons en faisant varier k de 1 à n, n relations que nous pouvons ajouter membre à membre :

$$\begin{aligned} 2^4 - 1^4 &= 4 \cdot 1^2 + 4 \cdot 1 + 1 \\ 3^4 - 2^4 &= 4 \cdot 2^2 + 4 \cdot 2 + 1 \\ &\cdots \\ (n+1)^4 - n^4 &= 4 \cdot n^2 + 4 \cdot n + 1 \\ + \overline{(n+1)^4 - 1} &= 4 \cdot S_2 + 6 \cdot S_2 + 4 \cdot S_1 + n \end{aligned} \tag{14.52}$$

Nous avons donc :

$$(n+1)^4 - 1^4 = 4 \cdot S_3 + 6 \cdot S_2 + 4 \cdot S_1 + n \tag{14.53}$$

Ce qui donne après développement :

$$(n+1)^4 - 1^4 = 4 \cdot S_3 + 6 \cdot \frac{n(n+1)(2n+1)}{6} + 4 \cdot \frac{n(n+1)}{2} + n \tag{14.54}$$

Et après une première simplification :

$$(n+1)^4 - 1 = 4 \cdot S_3 + n(n+1)(2n+1) + 2nn(n+1) + n \tag{14.55}$$

et une deuxième :

$$\begin{aligned} 4 \cdot S_3 =& - n(n+1)(2n+1) - 2n(n+1) - n + (n+1)^4 - 1 \\ =& - \big((n^2+n)(2n+1)\big) - (2n^2+2n) - n + (n^4 + 4n^3 + 6n^2 + 4n + 1) - 1 \\ =& - (2n^3 + n^2 + 2n^2 + n) + n^4 + 4n^3 + 4n^2 + n \\ =& n^4 + 2n^3 + n^2 \\ =& n^2(n^2 + 2n + 1) \\ =& n^2(n+1)^2 \end{aligned} \tag{14.56}$$

La résultat final est donc :

$$S_3 = \left(\frac{n(n+1)}{2}\right)^2 \tag{14.57}$$

ou écrit autrement :

$$\boxed{S_3 = \frac{n^2(n+1)^2}{4}} \tag{14.58}$$

Evidemment, nous pouvons continuer ainsi longtemps mais à partir du certaine valeur de l'élévation de la puissance les choses se compliquent un petit peu (de plus, la méthode est un peu longue). Ainsi, un des membre de la famille des Bernoulli (c'était une famille de mathématiciens assez doués...) a montré une relation générale fonctionnant pour n'importe quelle puissance en définissant ce que nous appelons le "polynôme de Bernoulli".

Nombres et polynômes de Bernoulli

Comme nous venons de le voir plus haut il est possible d'exprimer somme de n premiers entiers non nuls élevés à une puissance donnée selon (les quatre premiers ont été démontrés précédemment) les relations suivantes où nous avons posé $n = n' + 1$ avec n' le nombre de termes dont nous voulons la somme 0 non compris (d'où le signe négatif que nous n'avions pas plus haut) :

$$\begin{aligned}
S_0 &= n \\
S_1 &= \frac{1}{2}n^2 - \frac{1}{2}n = \frac{n(n-1)}{2} \\
S_2 &= \frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n = \frac{n(n-1)(2n-1)}{2} \\
S_3 &= \frac{1}{4}n^4 - \frac{1}{2}n^3 + \frac{1}{4}n^2 = \frac{n^2(n-1)^2}{4} \\
S_4 &= \frac{1}{5}n^5 - \frac{1}{2}n^4 + \frac{1}{3}n^3 - \frac{1}{30}n = \frac{n(n-1)(2n-1)(3n^2-3n-1)}{30} \\
S_5 &= \frac{1}{6}n^6 - \frac{1}{2}n^5 + \frac{5}{12}n^4 - \frac{1}{12}n^2 = \frac{n^2(2n^2-2n-1)(n-1)^{12}}{12} \\
&\cdots
\end{aligned} \tag{14.59}$$

Jacob Bernoulli remarqua ensuite que les polynômes S_p avaient la forme :

$$S_p = \frac{1}{p+1}n^{p+1} - \frac{1}{2}n^p + \frac{p}{12}n^{p-1} + 0 \cdot n^{p-2} + \cdots \tag{14.60}$$

Dans cette expression, les nombres $(1, -\frac{1}{2}, \frac{1}{12}, 0, \cdots)$ semblent ne pas dépendre de p. Plus généralement, après tâtonnement on remarque que le polynôme peut être écrit sous la forme :

$$\begin{aligned}
S_p &= \sum_{k=0}^{p} \frac{B_k}{k!} \frac{p!}{(p+1-k)!} n^{p+1-k} \\
&= \frac{B_0}{0!}\frac{n^{p+1}}{p+1} + \frac{B_1}{1!}n^p + \frac{B_2}{2!}pn^{p-1} + \frac{B_3}{3!}p(p-1)n^{p-2} + \cdots + \frac{B_p}{1!}n
\end{aligned} \tag{14.61}$$

Ce qui donne par identification les "nombres de Bernoulli" :

$$B_0 = 1, B_1 = -\frac{1}{2}, B_2 = \frac{1}{6}, B_3 = 0, \cdots \tag{14.62}$$

Par la suite, les mathématiciens dans leurs recherches sont tombés au hasard sur le fait que les nombres de Bernoulli pouvaient être exprimés par la série :

$$G(z) = \frac{z}{e^z - 1} = \sum_{k=0}^{\infty} B_k \frac{z^k}{k!} \text{ avec } |z| < 2\pi \tag{14.63}$$

En d'autres termes, la fonction génératrice des nombres de Bernoulli serait $G(z)$. Si nous développons les premiers termes de cette série :

$$\begin{aligned}
G(z) &= \left(\frac{e^z - 1}{z}\right)^{-1} = \left(1 + \frac{z}{2!} + \frac{z^2}{3!} + \frac{z^3}{4!} + \cdots\right)^{-1} \\
&= 1 - \frac{z}{2} + \frac{z^2}{12} - \frac{z^4}{720} + \frac{z^6}{30240} - \frac{z^8}{1209600} + \cdots
\end{aligned} \tag{14.64}$$

Preuve Nous avons vu dans notre étude des nombres complexes (cf. chapitre sur les Nombres) que :

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots \tag{14.65}$$

Dès lors :

$$G(z) = \left(\frac{z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots}{z} \right)^{-1} = \left(z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \right)^{-1} \tag{14.66}$$

Posons maintenant :

$$G(z) = a_0 + a_1 z + a_2 z^2 + \cdots \tag{14.67}$$

Nous avons alors :

$$\left(z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \right) \cdot (a_0 + a_1 z + a_2 z^2 + \cdots) = 1 \tag{14.68}$$

Nous voyons (en distribuant) que :

$$\left(z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \right) \cdot (a_0 + a_1 z + a_2 z^2 + \cdots) = a_0 + \left(a_1 + \frac{a_0}{2!} \right) z + \left(a_2 + \frac{a_1}{2!} + \frac{a_0}{3!} \right) z^2 + \cdots \tag{14.69}$$

par suite pour que tout cela soit égal à l'unité il faut que :

$$\begin{aligned} a_0 &= 1 \\ a_1 + \frac{a_0}{2!} &= 0 \\ a_2 + \frac{a_1}{2!} + \frac{a_0}{3!} &= 0 \\ &\cdots \end{aligned} \tag{14.70}$$

De la deuxième équation nous tirons :

$$a_1 = -\frac{1}{2} \tag{14.71}$$

De la troisième équation nous tirons :

$$a_2 = -\frac{1}{6} + \frac{1}{4} = \frac{1}{12} \tag{14.72}$$

etc.
En continuant ainsi nous montrons que :

$$a_3 = 0, a_4 = -\frac{1}{720}, \cdots \tag{14.73}$$

Il est évident que cette méthode nous permet de calculer à la main que les premiers termes de cette série.
Ainsi, en se basant sur :

$$a_3 = 0, a_4 = -\frac{1}{720}, \cdots \tag{14.74}$$

•

Nous trouvons que les premiers nombres de Bernoulli sont les suivants :

k	B_k
0	1
1	$-\frac{1}{2}$
2	$\frac{1}{6}$
3	0
4	$-\frac{1}{30}$
5	0
6	$\frac{1}{42}$
7	0
8	$-\frac{1}{30}$
9	0
10	$\frac{5}{66}$
11	0
12	$-\frac{691}{2730}$
13	0
14	$\frac{7}{6}$

REMARQUE Il est possible de démontrer que $B_k = 0$ lorsque n est impair et différent de 1.

Nous voyons bien par ailleurs, que la valeur des nombres de Bernoulli ne peuvent pas être décrits simplement. En fait, ce sont essentiellement des valeurs de la fonction ζ de Riemann (voir plus bas) pour des valeurs entières négatives de la variable, et sont associés à des propriétés théoriques profondes qui dépassent le cadre de ce site. Par ailleurs, les nombres de Bernoulli apparaissent également dans le développement en série de Taylor des fonctions tangentes circulaire et hyperbolique, dans la formule d'Euler-MacLaurin ainsi (voir plus bas).

Avec une petite modification, il est possible de définir les "polynômes de Bernoulli" $B_k(x)$ par :

$$G(z,x) = \frac{ze^{zx}}{e^z - 1} = \sum_{k=0}^{\infty} B_k(x)\frac{z^k}{k!} \tag{14.75}$$

avec donc $G(z,0) = G(z)$.
Par ailleurs, il est aisé de remarquer que :

$$\frac{\partial G(z,x)}{\partial z} = z\frac{ze^{zx}}{e^z - 1} = zG(z,k) = \sum_{k=0}^{\infty} \frac{dB_k(x)}{dx}\frac{z^k}{k!} \tag{14.76}$$

et donc il est facile d'en déduire :

$$\sum_{k=0}^{\infty} \frac{dB_k(x)}{dx}\frac{z^k}{k!} = \sum_{k=1}^{\infty} B_{k-1}(x)\frac{z^k}{(k-1)!} \tag{14.77}$$

Preuve D'un côté nous avons :

$$zG(z,k) = z\sum_{k=0}^{\infty} B_k(x)\frac{z^k}{k!} = \sum_{k=0}^{\infty} B_k(x)\frac{z^{k+1}}{k!} = \sum_{k=1}^{\infty} B_{k-1}(x)\frac{z^k}{(k-1)!} \tag{14.78}$$

et d'un autre nous avons :

$$zG(z,k) = \sum_{k=0}^{\infty} \frac{dB_k(x)}{dx}\frac{z^k}{k!} \tag{14.79}$$

Donc :

$$\sum_{k=0}^{\infty} \frac{dB_k(x)}{dx}\frac{z^k}{k!} = \sum_{k=1}^{\infty} B_{k-1}(x)\frac{z^k}{(k-1)!} \tag{14.80}$$

•

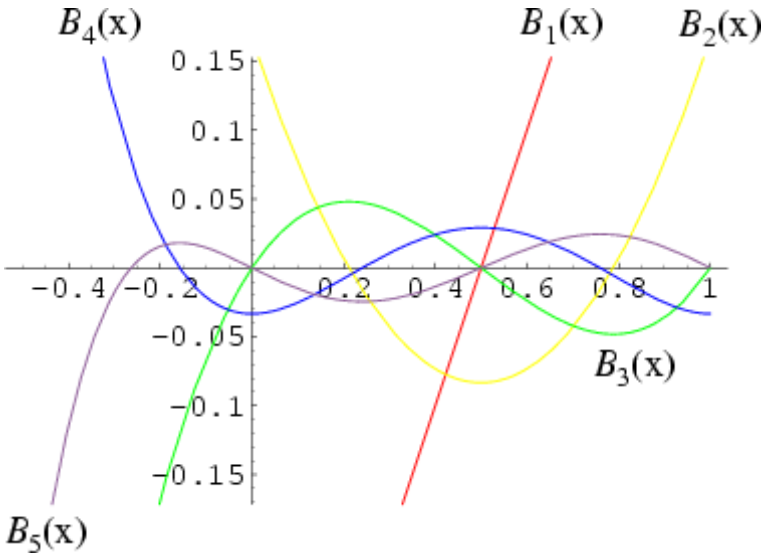


FIGURE 14.1 – Polynômes de Bernoulli

Et par identification des coefficients nous en déduisons :

$$\frac{dB_0(x)}{dx} = 0 \tag{14.81}$$

et pour $k \geq 1$:

$$\frac{dB_k(x)}{dx} = k \cdot B_{k-1}(x) \tag{14.82}$$

Il est alors aisé de déduire que les $B_k(x)$ sont des polynômes de degré k :

$$\begin{aligned}
B_0(x) &= 1 \\
B_1(x) &= x - \frac{1}{2} \\
B_2(x) &= x^2 - x + \frac{1}{6} \\
B_3(x) &= x^3 - 2x^2 + \frac{1}{2}x \\
B_4(x) &= x^4 - 2x^3 + x^2 - \frac{1}{30} \\
B_5(x) &= x^5 - \frac{5}{2}x^4 + \frac{5}{3}x^3 - \frac{1}{6}x \\
&\dots
\end{aligned} \tag{14.83}$$

La figure 14.1 en donne une représentation.

Ce qui est remarquable ce qu'à l'aide des polynômes de Bernoulli, nous voyons qu'il est possible d'écrire les S_n sous la forme suivante :

$$S_p = \sum_{k=0}^{n-1} k^p = \frac{1}{p+1}\left(B_{p+1}(n) - B_{p+1}\right) \tag{14.84}$$

Certains écrivent cette relation encore autrement. Effectivement, de la relation précédente, nous

pouvons écrire :

$$(p+1)\sum_{k=0}^{n-1} k^p = B_{p+1}(n) - B_{p+1} \tag{14.85}$$

Et en utilisant :

$$\frac{dB_k(x)}{dx} = k \cdot B_{k-1}(x) \tag{14.86}$$

Il vient :

$$\begin{aligned} &\frac{dB_{p+1}(x)}{dx} = (p+1)\cdot B_p(x) \\ \Rightarrow &B_p(x) = \frac{1}{p+1}B'_{p+1}(x) \\ \Rightarrow &\int_0^k B_p(x)dx = \frac{1}{p+1}\int_0^k B'_{p+1}(x) = \frac{1}{p+1}B_{p+1}(x)\Big|_0^k = \frac{1}{p+1}\big[B_{p+1}(x) - B_{p+1}(0)\big] \end{aligned} \tag{14.87}$$

Donc nous venons de démontrer :

$$\sum_{k=0}^{n-1} k^p = \int_0^k B_p(x)dx \tag{14.88}$$

Cependant, nous pouvons maintenant nous demander ce qu'il advient de la somme partielle de suites arithmétiques et géométriques telles que présentées au début de ce chapitre.

14.2.2 Séries arithmétiques

Nous avons démontré plus haut que la somme partielle de la série de Gauss (analogue à la somme des termes d'une suite arithmétique de raison $r = 1$) s'écrivait donc :

$$S_1 = \frac{n(n+1)}{2} \tag{14.89}$$

si nous notons non pas n la valeur n-ième terme mais u_n, le développement que nous avions fait pour la série de Gauss nous amène alors à :

$$S_n = \frac{n(u_n+1)}{2} \tag{14.90}$$

et si nous notons le premier terme 1 de la Série de Gauss par u_0, nous avons alors :

$$\boxed{S_n = \frac{n(u_n+u_0)}{2}} \tag{14.91}$$

ce qui nous donne la somme partielle des n-termes d'une suite arithmétique de raison r quelconque (ou plus simplement : la somme partielle de la série arithmétique de raison r)

REMARQUE Le lecteur aura observé que la raison r n'apparaît pas dans la relation. Effectivement, en reprenant (toujours) le même développement fait que pour la série de Gauss, le terme r se simplifie.

14.2.3 Séries géométriques

De même, avec un somme géométrique, nous avons donc (rappel : $u_n = u_0 q^n$) :

$$\begin{aligned} S_n &= u_0 + u_1 + \cdots + u_{n-1} + u_n \\ qS_n &= u_1 + u_2 + \cdots + u_n + u_{n+1} \\ \hline s_n - qS_n &= u_0 + U_1 - u_1 + u_2 - u_2 + \cdots + u_n - u_n - u_{n+1} \end{aligned} \tag{14.92}$$

La dernière relation s'écrit (après simplification) :

$$S_n - qS_n = (1-q)S_n = u_0 - u_{n+1} \tag{14.93}$$

et si $q \neq 0$, nous avons :

$$S_n = \frac{u_0 - u_{n+1}}{(1-q)} \tag{14.94}$$

ce qui peut s'écrire en factorisant u_0 :

$$\boxed{s_n = u_0 \frac{(1-q^{n+1})}{(1-q)}} \tag{14.95}$$

Exemple 14.2 *Soit la suite de raison $q = 2$ suivante :*

$$1 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3 + \cdots + 1 \cdot 2^n = 1 + 2 + 4 + 8 + \cdot + 2^n \tag{14.96}$$

pour calculer la somme des quatre premiers termes $\{1, 2, 4, 8\}$, nous prenons la puissance de 2 équivalent $n = 3$(le zéro n'étant pas pris en compte). Nous obtenons alors bien $S_3 = 15$.

La fonction zêta et identité d'Euler

L'allemand Riemann a baptisé "zêta" une fonction déjà étudiée avant lui, mais qu'il examine lorsque la valeur est un nombre complexe (cf. chapitre sur les Nombres). Cette fonction se présente comme une série de puissances inverses de nombres entiers. C'est la série :

$$\zeta(s) = \frac{1}{1} + \frac{1}{2^5} + \frac{1}{3^5} + \cdots \tag{14.97}$$

REMARQUE Il est traditionnel de noter s la variable dont dépend cette série.

Cette série a une propriété intéressante mais si l'on reste dans le cadre des puissance entières positives et non nulles :

$$\frac{1}{x^0} + \frac{1}{x^1} + \cdots + \frac{1}{x^n} + \cdots = u_0 \frac{(1-q^{n+1})}{(1-q)} = 1 \frac{1 - \left(\frac{1}{x^{n+1}}\right)}{1 - \frac{1}{x}} \tag{14.98}$$

quand nous avons alors :

$$\sum_{n=0}^{+\infty} = \frac{1}{x^n} = \frac{1}{x^0} + \frac{1}{x^1} + \cdots + \frac{1}{x^n} + \cdots = \frac{1}{1 - \frac{1}{x}} \tag{14.99}$$

Si nous faisons $x = 2^5$, nous obtenons la somme des puissances inverses de 2 et de mêmes avec $x = 3^5$ tel que :

$$\begin{aligned} 1 + \frac{1}{2^5} + \frac{1}{4^5} + \frac{1}{8^5} + \cdots + \frac{1}{2^{a5}} + \cdots &= \frac{1}{1 - \frac{1}{2^5}} \\ 1 + \frac{1}{3^5} + \frac{1}{9^5} + \frac{1}{27^5} + \cdots + \frac{1}{3^{b5}} + \cdots &= \frac{1}{1 - \frac{1}{3^5}} \end{aligned} \tag{14.100}$$

Si nous faisons le produit de ces deux expressions, nous obtenons la somme des puissances de toutes les fractions dont le dénominateur est un nombre produit de 2 et de 3 :

$$\frac{1}{1 - \frac{1}{2^5}} \cdot \frac{1}{1 - \frac{1}{3^5}} = 1 + \frac{1}{2^5} + \frac{1}{3^5} + \frac{1}{4^5} + \frac{1}{5^5} + \frac{1}{6^5} + \cdots + \frac{1}{(2^a 3^b 5^c)^5} + \cdots \tag{14.101}$$

Si nous prenons tous les nombres premiers à gauche, nous obtiendrons à droite tous les nombres entiers, puisque tout entier est produit de nombres premiers selon le théorème fondamental de l'arithmétique (cf. chapitre de Théorie Des Nombres), et c'est l'identité fondamentale d'Euler : ce que nous appelons maintenant la "fonction zêta de Riemann" est à la fois un produit fini et la somme des puissances inverse de tous les entiers :

$$\zeta(s) = \frac{1}{1 - \frac{1}{2^5}} \cdot \frac{1}{1 - \frac{1}{3^5}} \cdot \frac{1}{1 - \frac{1}{5^5}} = 1 + \frac{1}{2^5} + \frac{1}{3^5} + \frac{1}{4^5} + \frac{1}{5^5} + \frac{1}{6^5} + \cdots + \frac{1}{(2^a 3^b 4^c)^5} \tag{14.102}$$

En notation condensée, "l'identité d'Euler" est :

$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^5} = \prod_p \frac{1}{1 - \frac{1}{p^5}} \tag{14.103}$$

où p sont les nombres premiers.

14.2.4 Série de Taylor et MacLaurin

Soit un polynôme :

$$P_n(x) = \sum_{k=0}^{n} a_k x^k \tag{14.104}$$

Nous avons trivialement pour ce dernier :

$$P(0) = a_0 \tag{14.105}$$

Soit maintenant la dérivée du polynôme $P(x)$:

$$P(x) = \sum_{k=1}^{n} k a_k x^{k-1} \tag{14.106}$$

donc :

$$P'(0) = 1 \cdot a_1 \tag{14.107}$$

et ainsi de suite avec $P''(x), P'''(x), ...$ tel que :

$$\begin{aligned} P''(x) &= \sum_{k=2}^{n} k(k-1)a_k x^{k-2} \Rightarrow P''(0) = 1 \cdot 2 \cdot a_2 \\ P'''(x) &= \sum_{k=3}^{n} k(k-1)(k-2)a_k x^{k-3} \Rightarrow P''(0) = 1 \cdot 2 \cdot \cdot a_3 a_3 \\ &\dots \\ \frac{d^n P(x)}{dx^n} &= \sum_{k=n}^{n} k(k-1)\cdots(k-(n-1))a_k x^{k-n} \Rightarrow P''(0) = n! \cdot a_n \end{aligned} \tag{14.108}$$

Il s'ensuit que :

$$a_0 = P(0), a_1 = \frac{1}{1!}\frac{dP(0)}{dx}, a_2 = \frac{1}{2!}\frac{d^2P(0)}{dx^2}, a3 = \frac{1}{3!}\frac{d^3P(0)}{dx^3}, \cdots, a_n = \frac{1}{n!}\frac{d^nP(0)}{dx^n} \tag{14.109}$$

Donc finalement notre polynôme peut s'écrire :

$$\boxed{P_n(x) = \sum_{k=0}^{n} \frac{1}{k!}\frac{d^kP(0)}{dx^k}x^k} \tag{14.110}$$

relation que l'on appelle "série de MacLaurin limitée" ou tout simplement "série de MacLaurin".

En appliquant maintenant le même raisonnement mais en centrant le polynôme sur la valeur , nous avons :

$$P_n(x) = \sum_{k=0}^{n} a_k(x - x_0)^k \tag{14.111}$$

et ainsi le développement précédent devient :

$$\boxed{P_n(x) = \sum_{k=0}^{n} \frac{1}{k!}\frac{d^kP(x_0)}{dx^k}(x - x_0)^k} \tag{14.112}$$

qui n'est d'autre que l'expression générale d'un polynôme exprimé sous une forme dite de "série de Taylor limitée".

Ainsi, certaines fonctions $f(x)$ pouvant être approximées par un polynôme $P(x)$ autour d'une valeur x_0 peuvent êtres exprimées sous la forme d'un polynôme tel que :

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!}\frac{d^kf(x_0)}{dx^k}(x - x_0)^k = \sum_{k=0}^{n} \frac{d^{(k)}f(x_0)}{k!}(x - x_0)^k \tag{14.113}$$

Ce qui s'écrit de manière plus conventionnelle... :

$$f(x) = \sum_{k=0}^{n} \frac{D^kf(x_0)}{k!}(x - x_0)^k \tag{14.114}$$

Reste de Lagrange

Il peut y avoir un intérêt dans certaines application numérique (cf. chapitre de Méthodes Numériques) à connaître l'erreur d'approximation du polynôme $P_n(x)$ par rapport à la fonction $f(x), \forall x$.

Définissons pour cela un "reste" $R_n(x)$, tel que :

$$R_n(x) = f(x) - P_n(x) \tag{14.115}$$

La fonction $R_n(x)$ est appelée "reste de Lagrange".

Considérons maintenant une fonction $f(x)$ qui est$(n+1)$ fois dérivable sur un intervalle qui contient x_0. Pour une valeur x de l'intervalle, différente de x_0, nous nous proposons de démontrer qu'il existe un nombre z situé entre x_0 et x tel que :

$$\boxed{R_n(x) = \frac{f^{(n+1)}(z)}{(n+1)!}(x-x_0)^{n+1}} \tag{14.116}$$

Preuve Soit une fonction $g(t)$ une fonction définie par la différence d'une fonction $f(x)$ supposé connue et une approximation de Taylor de cette même fonction :

$$\begin{aligned} g(t) =& f(x) - \left(f(t) + \frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}(x-t)^2 + \cdots + \frac{f^{(n)}(t)}{n!}(x-t)^n + \frac{f^{(n+1)}(t)}{(n+1)!}(x-t)^{(n+1)} \right) \\ =& (x) - \left(f(t) + \frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}(x-t)^2 + \cdots + \frac{f^{(n)}(t)}{n!}(x-t)^n - R_n(x)\frac{(x-t)^{n+1}}{(x-x_0)^{n+1}} \right) \end{aligned} \tag{14.117}$$

avec bien sûr :

$$R_n(x)\frac{(x-t)^{n+1}}{(x-x_0)^{n+1}} = \frac{f^{(n+1)}(t)}{(n+1)!}(x-t)^{(n+1)} \tag{14.118}$$

Nous voyons que $g(t)$ s'annule bien pour la valeur $t = x$.

Dérivons maintenant $g(t)$ par rapport à t, nous trouvons :

$$\begin{aligned} g'(t) =& -f'(t) - f'(t)(-1) - f''(t)(x-t) - f''(t)(x-t)(-1) + \frac{f''(t)}{2}(x-t)^2 + \\ &+ \cdots - \frac{f^{(n+1)}(t)}{n!}(x-t)^n + R_n(x)(n+1)\frac{(x-t)^n}{(x-x_0)^{n+1}} \end{aligned} \tag{14.119}$$

Après simplification :

$$g'(t) = -\frac{f^{(n+1)}(t)}{n!}(x-t)^n + R_n(x)(n+1)\frac{(x-t)^n}{(x-x_0)^{n+1}} \tag{14.120}$$

Selon le théorème de Rolle (cf. chapitre de Calcul Différentiel Et Intégral), il existe une valeur $t = z$ pour lequel la dérivée $g'(t)$ s'annule. Donc :

$$0 = -\frac{f^{(n+1)}(z)}{n!}(x-z)^n + R_n(x)(n+1)\frac{(x-z)^n}{(x-x_0)^{n+1}} \tag{14.121}$$

Nous pouvons simplifier l'équation par $(x-z)^n$:

$$R_n(x)\frac{(n+1)}{(x-x_0)^{n+1}} - \frac{f^{(n+1)}(z)}{n!} = 0 \tag{14.122}$$

ce qui s'écrit aussi :

$$R_n(x) - \frac{f^{(n+1)}(z)}{(n+1)!}(x-x_0)^{n+1} = 0 \tag{14.123}$$

et nous trouvons donc pour maximum de :

$$R_n(x) = \frac{f^{(n+1)}(z)}{(n+1)!}(x - x_0)^{n+1} \tag{14.124}$$

•

Nous voyons que plus le polynôme est de degré élevé, plus il approxime la fonction $f(x)$ avec exactitude. Que se passe-t-il lorsque $n \to \infty$?

$$R_n(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \tag{14.125}$$

Supposons que $f(x)$ admette des dérivées de tout ordre (ce que nous notons C^∞) pour toutes les valeurs d'un intervalle quelconque contenant x_0 et soit R_n le reste de Lagrange de $f(x)$ en x_0. Si, quel que soit x dans l'intervalle :

$$\lim_{n\to\infty} R_n(x) = 0 \tag{14.126}$$

alors $f(x)$ est exactement représentée par $P(x)$ sur l'intervalle.

Preuve Elle découle simplement de l'expression de $P_n(x)$ lorsque $n \to \infty$

Le polynôme :

$$P(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \tag{14.127}$$

est appelé "polynôme de Taylor" ou "série de Taylor". Si $x_0 = 0$, il est appelé "polynôme de MacLaurin" ou "série de MacLaurin".

14.2.5 Série de Fourier

Nous appelons par définition "série trigonométrique" une série de la forme :

$$a_0 \cdot \cos(0 \cdot x) + b_0 \cdot \sin(0 \cdot x) + a_1 \cdot \cos(x) + b_1 \cdot \sin(x) + \cdots a_n \cdot \cos(nx) + b_0 \cdot \sin(nx) \tag{14.128}$$

ou sous une forme plus compacte :

$$\sum_{n=0}^{\infty} \big(a_n \cdot \cos(nx) + b_0 \cdot \sin(nx)\big) \tag{14.129}$$

Les constantes $a_n, b_n (n = 1, 2, \cdots \in \mathbb{N}^*)$ sont les coefficients de la série trigonométrique plus souvent nommés "coefficients de Fourier".

REMARQUE Nous avions déjà fait mention de ce type de série lors de notre étude des types de polynômes existants puisque les séries de Fourier ne sont au fait que des polynômes trigonométriques. (cf. chapitre de Calcul Algébrique). Par ailleurs, nous avons vu comme exemple dans le chapitre d'Analyse Fonctionnelle lors de notre étude du produit scalaire fonctionnel que les fonctions sinus et cosinus constituaient les bases d'un espace vectoriel.

Si la série converge, sa somme est une fonction périodique $f(x)$ de période $T = 2\pi$, étant donné que $\sin(nx)$ et $\cos(nx)$ sont des fonctions périodiques de période . De sorte que :

$$f(x) = f(x + 2\pi) \tag{14.130}$$

Posons maintenant le problème suivant : Nous nous donnons une fonction connue, périodique quelconque $f(x)$ continue par morceaux de période 2π. Nous demandons s'il existe une série trigonométrique convergeant vers $f(x)$ moyennant des conditions sur cette série.

Supposons maintenant que la fonction $f(x)$, périodique et de période 2π, puisse être effectivement représentée par une série trigonométrique convergent vers $f(x)$ dans l'intervalle $[0, T]$, c'est-à-dire qu'elle soit la somme de cette série :

$$f(x) = \sum_{n=0}^{\infty} \big(a_n \cdot \cos(nx) + b_n \cdot \sin(nx)\big) \tag{14.131}$$

Supposons que l'intégrale de la fonction du premier membre de cette égalité soit égale à la somme des intégrales des termes de la série ci-dessus. Ceci aura lieu, par exemple, si nous supposons que la série trigonométrique proposée converge absolument, c'est-à-dire que converge la série numérique suivante (de par la propriété bornée des fonctions trigonométriques) :

$$|a_0| + |b_0| + |a_1| + |b_1| + |a_2| + |b_2| + \cdot + |a_n| + |b_n| \tag{14.132}$$

La série :

$$f(x) = \sum_{n=0}^{\infty} \big(a_n \cdot \cos(nx) + b_n \cdot \sin(nx)\big) \tag{14.133}$$

est alors majorable et peut être intégrée terme à terme de 0 à T (où $T = 2\pi$) ce qui nous permet de déterminer les différents coefficients de Fourier. Mais avant de commencer exposons les intégrales suivantes qui nous seront utiles par la suite :

$$\text{(I)} \begin{cases} \int_0^T \cos(nx) \cdot \cos(kx)dx = 0 \\ \int_0^T \cos(nx) \cdot \sin(kx)dx = 0 \\ \int_0^T \sin(nx) \cdot \sin(kx)dx = 0 \end{cases} \qquad \text{(II)} \begin{cases} \int_0^T \cos^2(kx)dx = \frac{T}{2} \\ \int_0^T \cos(kx) \cdot \sin(kx)dx = 0 \\ \int_0^T \sin^2(kx)dx = \frac{T}{2} \end{cases} \tag{14.134}$$

$$\text{Avec } n, k \in \mathbb{N} \text{ et } n \neq k \qquad\qquad \text{Avec } n, k \in \mathbb{N} \text{ et } n = k$$

Avant de continuer, démontrons la valeur que prennent ces six intégrales (suite à la demande des internautes). Mais d'abord, rappelons que commen, $k \in \mathbb{N}$ alors $(n+k) = w \in \mathbb{N}$ et $(n-k) = v \in \mathbb{N}$

1. Nous procédons en utilisant les relations trigonométrique remarquables (cf. chapitre de Trigonométrie) et les primitives des fonctions trigonométriques élémentaires (cf. chapitre de calcul Différentiel Et Intégral) :

$$\begin{aligned} \int_0^T \cos(nx)\cos(kx)dx =& \frac{1}{2}\int_0^T \big(\cos(nx+kx) + \cos(nx-kx)\big)dx \\ =& \frac{1}{2}\left(\frac{\sin(nx+kx)}{n+k} + \frac{\sin(-nx+kx)}{-n+k}\right)\Bigg|_0^T \\ =& \frac{1}{2}\left(\frac{\sin(wx)}{w} + \frac{\sin(vx)}{v}\right)\Bigg|_0^T \\ =& \frac{1}{2}\left(\frac{\sin(wT)}{w} + \frac{\sin(vT)}{v}\right) - \frac{1}{2}\left(\frac{\sin(w0)}{w} + \frac{\sin(v0)}{v}\right) \\ =& 0 \end{aligned} \tag{14.135}$$

car comme nous l'avons vu en trigonométrie $\sin(k\pi) = 0, k \in \mathbb{Z}$ et comme $T = 2\pi$, les deux différences précédentes ont tous les termes qui sont nuls tel que :

$$\int_0^T \cos(nx)\cos(kx)dx = 0 \tag{14.136}$$

2. Pour la deuxième intégrale, nous procédons selon les mêmes techniques et mêmes propriétés des fonctions trigonométrique :

$$\begin{aligned}\int_0^T \cos(nx)\sin(kx)dx =& \frac{1}{2}\int_0^T \big(\sin(nx+kx)+\sin(nx-kx)\big)dx\\ =& \frac{1}{2}\left(-\frac{\cos(nx+kx)}{n+k}+\frac{\cos(nx-kx)}{n-k}\right)\Bigg|_0^T\\ =& \frac{1}{2}\left(-\frac{\cos(wx)}{w}+\frac{\cos(vx)}{v}\right)\Bigg|_0^T\\ =& \frac{1}{2}\left(-\frac{\cos(wT)}{w}+\frac{\cos(vT)}{v}\right)-\frac{1}{2}\left(-\frac{\cos(w0)}{w}+\frac{\cos(v0)}{v}\right)\\ =& \frac{1}{2}\left(-\frac{1}{w}+\frac{1}{v}\right)-\frac{1}{2}\left(-\frac{1}{w}+\frac{1}{v}\right)\\ =& \frac{1}{2}\left[\left(-\frac{1}{w}+\frac{1}{v}\right)-\left(-\frac{1}{w}+\frac{1}{v}\right)\right]\\ =& 0\end{aligned} \tag{14.137}$$

3. Et nous continuons ainsi pour la troisième, toujours selon les même propriétés :

$$\begin{aligned}\int_0^T \sin(nx)\sin(kx)dx =& \frac{1}{2}\int_0^T \big(\cos(nx-kx)-\cos(nx+kx)\big)dx\\ =& \frac{1}{2}\left(\frac{\sin(nx-kx)}{n-k}-\frac{\sin(nx+kx)}{n+k}\right)\Bigg|_0^T\\ =& \frac{1}{2}\left(\frac{\sin(vx)}{v}-\frac{\sin(wx)}{w}\right)\Bigg|_0^T\\ =& \frac{1}{2}\left(\frac{\sin(vT)}{v}-\frac{\sin(wT)}{w}\right)-\frac{1}{2}\left(\frac{\sin(v0)}{v}-\frac{\sin(w0)}{w}\right)\\ =& 0\end{aligned} \tag{14.138}$$

4. Encore une fois selon les mêmes méthodes (cela devient routinier...) pour $k \neq 0$ d'abord :

$$\begin{aligned}\int_0^T \cos^2(kx)dx =& \int_0^T \left(\frac{1+\cos(2kx)}{2}\right)dx\\ =& \frac{1}{2}\int_0^T dx+\frac{1}{2}\int_0^T \cos(2kx)dx\\ =& \frac{1}{2}x\Big|_0^T+\frac{1}{4k}\sin(2kx)\Big|_0^T\\ =& \frac{1}{2}(T-0)+\frac{1}{4k}\big(\sin(2kT)-\sin(2k\cdot 0)\big)\\ =& \frac{1}{2}T+\frac{1}{4k}(0-0)\\ =& \frac{T}{2}\end{aligned} \tag{14.139}$$

et pour $k=0$ il vient immédiatement :

$$\int_0^T \cos^2(0\cdot x)dx = \int_0^T dx = T \tag{14.140}$$

5. Encore une fois... (bientôt au bout...) pour $k \neq 0$ d'abord :

$$\begin{aligned}\int_0^T \cos(kx)\sin(kx)dx =& \frac{1}{2}\int_0^T \big(\sin(kx+kx)+\sin(kx-kx)\big)dx\\ =& \frac{1}{2}\int_0^T \sin(2kx)dx\\ =& \frac{1}{2}\left(-\frac{\cos(2kx)}{2k}\right)\Bigg|_0^T\\ =& -\frac{1}{4}\left(\frac{\cos(2kx)}{k}\right)\Bigg|_0^T\\ =& -\frac{1}{4}\left(\frac{\cos(2kT)}{k}-\frac{\cos(2k\cdot 0)}{k}\right)\\ =& -\frac{1}{4}\left(\frac{1}{k}-\frac{1}{k}\right)\\ =& 0\end{aligned} \tag{14.141}$$

et pour $k = 0$ il vient immédiatement :

$$\int_0^T \cos(kx)\sin(kx)dx = 0 \tag{14.142}$$

6. Et enfin la dernière (...) :

$$\begin{aligned}\int_0^T \sin^2(kx)dx =& \int_0^T \big(\frac{1-\cos(2kx)}{2}\big)dx\\ =& \frac{1}{2}\int_0^T dx - \frac{1}{2}\int_0^T \cos(2kx)dx\\ =& \frac{1}{2}x\Big|_0^T - \frac{1}{2k}\sin(2kx)\Big|_0^T\\ =& \frac{1}{2}(T-0) - \frac{1}{2k}\big(\sin(2kT)-\sin(2k\cdot 0)\big)\\ =& \frac{1}{2}T + \frac{1}{2k}(0-0)\\ =& \frac{T}{2}\end{aligned} \tag{14.143}$$

Ce petit travail fait, revenons maintenant à nos moutons... Pour déterminer les coefficients a_n multiplions les deux membres de l'égalité :

$$f(x) = \sum_{n=0}^{+\infty}\big(a_n\cdot\cos(nx)+b_n\cdot\sin(nx)\big) \tag{14.144}$$

par $\cos(n\cdot x)$:

$$f(x)\cos(nx) = \sum_{n=0}^{+\infty}\big(a_n\cdot\cos^2(nx)+b_n\cdot\sin(nx)\cos(nx)\big) \tag{14.145}$$

La série du second membre de l'égalité est majorable, étant donné que ses termes ne sont pas supérieurs en valeur absolue aux termes de la série positive convergente. Nous pouvons donc l'intégrer terme à terme sur tout segment borné de 0 à T :

$$\int_0^T f(x)\cos(nx)dx = \sum_{n=0}^{+\infty}\left(a_n\int_0^T\cos^2(nx)dx + b_n\int_0^T\sin(nx)\cos(nx)dx\right) \tag{14.146}$$

Et nous avons démontré plus haut les valeurs de ces intégrales. Il reste donc :

$$\int_0^T f(x)\cos(nx)dx = \sum_{n=0}^{+\infty}\left(a_n\frac{T}{2}\right) \tag{14.147}$$

D'où nous tirons aisément que :

$$a_n = \frac{2}{T}\int_O^T f(x)\cos(nx)dx \tag{14.148}$$

Mais le cas $n = 0$ doit être traité à part. Effectivement, dans ce cas :

$$\int_O^T f(x)dx = a_0\int_O^T dx = A_0T \tag{14.149}$$

Soit :

$$a_0 = \frac{1}{T}\int_O^T f(x)dx \tag{14.150}$$

Il est évident que le coefficient a_0 représente donc la moyenne du signal ou la composante continue si elle existe.

Pour déterminer les coefficients b_n multiplions les deux membres de l'égalité par $\sin(nx)$:

$$f(x)\sin(nx) = \sum_{n=0}^{+\infty}\big(a_n\cos(nx)\sin(nx) + b_n\sin^2(nx)\big) \tag{14.151}$$

La série du second membre de l'égalité est majorable, étant donné que ses termes ne sont pas supérieurs en valeurs absolues aux termes de la série positive convergente. Nous pouvons donc l'intégrer terme à terme sur tout segment borné de 0 à T :

$$\int_0^T f(x)\sin(nx)dx = \sum_{n=0}^{+\infty}\left(a_n\int_0^T\cos(nx)\sin(nx)dx + b_n\int_0^T\sin^2(nx)dx\right) \tag{14.152}$$

Et nous avons démontré plus haut les valeurs de ces intégrales. Il reste donc :

$$\int_0^T f(x)\sin(nx)dx = \sum_{n=0}^{+\infty}\left(b_n\frac{T}{2}\right) \tag{14.153}$$

D'où nous tirons aisément que :

$$b_n = \frac{2}{T}\int_0^T f(x)\sin(nx)dx \tag{14.154}$$

le cas $n = 0$ ne nécessite pas d'être traité à part avec le sinus.

Donc finalement les coefficients de Fourier sont donc déterminés par les intégrales :

$$a_0 = \frac{1}{T}\int_0^T f(x)dx \qquad a_n = \frac{2}{T}\int_0^T f(x)\cos(nx)dx \qquad b_n = \frac{2}{T}\int_0^T f(x)\sin(nx)dx \tag{14.155}$$

Mais comme c'est embêtant d'avoir trois résultats pour les coefficients nous allons jouer un peu avec la définition de la série de Fourier.

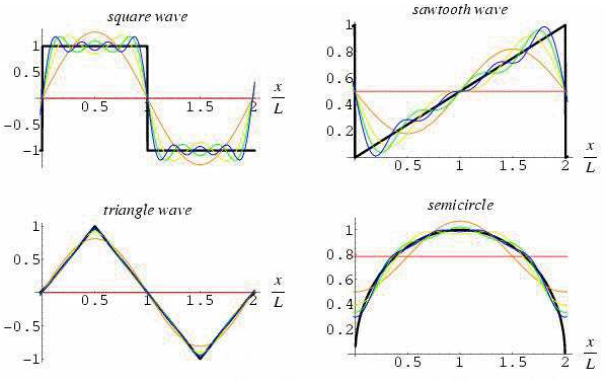


FIGURE 14.2 – Source : Mathworld

Effectivement en sommant de 1 à l'infini plutôt que de 0 à l'infini nous avons :

$$\boxed{f(x) = \frac{a_0}{2} + \sum_{n=0}^{+\infty} \left(a_n \cos(nx) + b_n \sin(nx)\right)} \tag{14.156}$$

Ce qui permet alors de n'avoir qu'à se rappeler de (a_0 inclus donc !) :

$$\boxed{a_n = \frac{2}{T}\int_0^T f(x)\cos(nx)dx} \qquad \boxed{b_n = \frac{2}{T}\int_0^T f(x)\sin(nx)dx} \tag{14.157}$$

Les physiciens ont eux pour habitude de noter ces relations sous la forme suivante :

$$\begin{aligned} a_k =& \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\cos\left(\frac{2\pi}{T}kt\right)dt = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\cos(\omega kt)dt \\ b_k =& \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\sin\left(\frac{2\pi}{T}kt\right)dt = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\sin(\omega kt)dt \end{aligned} \tag{14.158}$$

Cette décomposition possible de toute fonction périodique continue par morceaux approchée par une somme infinie de fonctions trigonométriques (sinus ou cosinus) consistant en une fonction fondamentale et ses harmoniques est appelé "théorème de Fourier" ou encore "théorème de Fourier-Dirichlet".

La série de Fourier permet donc implicitement de représenter toutes les fréquences contenues dans un signal périodique dont la fonction est connue mathématiquement. On se demande bien pourquoi parler des séries de Fourier quand, dans la pratique, nous ne connaissons pas vraiment la représentation mathématique de ce signal ? Cela nous amènera à mieux comprendre le concept de la transformée de Fourier à temps discret (DTFT), que nous verrons un peu plus loin, qui n'a nul besoin d'une représentation mathématique d'un signal continu échantillonné dans le temps.

Nous constatons par ailleurs que si $f(x)$, soit la fonction périodique dont nous cherchons l'expression en série trigonométrique de Fourier, est paire alors la série devra être paire aussi et donc ne comporter que des termes en cosinus (le cosinus étant pour rappel une fonction paire) ce qui implique que $b_n = 0$ et dans le cas contraire d'une fonction impaire $a_n = 0$ (le sinus étant pour rappel une fonction impaire).

Il convient de noter, et c'est important pour la suite, que comme nous l'avons vu dans le chapitre de Calcul Algébrique lors de notre étude des polynômes trigonométriques, que les séries de Fourier pouvaient donc s'écrire sous la forme complexe suivante (en changeant un peu les notations et en

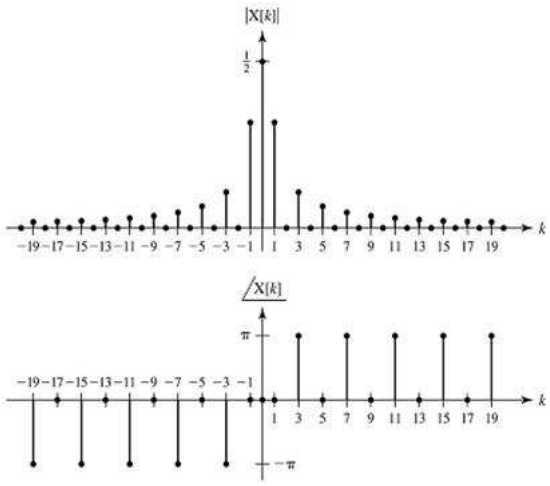


FIGURE 14.3 – Spectre de phase

passant la somme à l'infini) :

$$f(x) = \sum_{n=-\infty}^{+\infty} c_n e^{inx} \tag{14.159}$$

et nous avions vus que :

$$a_n = c_n + c_{-n} \qquad b_n = i(c_n - c_{-n}) \tag{14.160}$$

Soit :

$$c_n = \frac{a_n - ib_n}{2} \tag{14.161}$$

Ce qui nous donne :

$$\begin{aligned} c_n &= \frac{\frac{2}{T}\int_0^T f(x)\cos(nx)dx - \frac{2}{T}i\int_0^T f(x)\sin(nx)dx}{2} \\ &= \frac{1}{T}\int_0^T f(x)\big(\cos(nx) - i\sin(nx)\big)dx \\ &= \frac{1}{T}\int_0^T f(x)e^{-inx}dx \end{aligned} \tag{14.162}$$

Exemple :

Exemple 14.3 *Lors de la décomposition d'un signal continu, nous disons abusivement que les coefficients a_n, b_n représentent chacun (implicitement) une fréquence distincte associée à une amplitude que nous visualisons sur un graphique par des lignes verticales. Ce graphique représente le spectre en fréquence du signal décomposé. Nous pouvons également adjoindre une autre représentation qui se nomme "spectre de phase". Ce spectre nous donne la phase du signal harmonique (en avance ou en retard de phase) comme représenté à la figure 14.3.*

Voyons maintenant comment décomposer un signal périodique connu en plusieurs signaux d'amplitudes et de fréquences distinctes.

Prenons comme exemple, un signal à onde carrée périodique défini sur une période $T = 2$ et d'amplitude A tel que :

$$f(t) = \begin{cases} -A & -1 < t < 0 \\ A & 0 < t < 1 \end{cases} \tag{14.163}$$

A la période T=2 correspond comme nous le savons une pulsation :

$$\omega = \frac{2\pi}{T} = \pi \tag{14.164}$$

Calculons en premier lieu les coefficients c_k à l'aide de l'intégrale permettant de déterminer ces coefficients (le choix des bornes de l'intégrale supposent donc que le signal est périodique par construction !) :

$$\begin{aligned} c_k &= \frac{1}{T}\int_T f(t)e^{-ikwt}dt = \frac{1}{2}\int_{-1}^{1} f(t)e^{-ik\pi t}dt = \frac{1}{2}\left(\int_{-1}^{0} -Ae^{-ik\pi t}dt + \int_0^1 Ae^{-ik\pi t}dt\right) \\ &= \frac{-A}{2}\left(\int_{-1}^{0} e^{-ik\pi t}dt - \int_0^1 e^{-ik\pi t}dt\right) = \frac{-A}{2}\left(\frac{-1}{ik\pi}\right)\left[e^{-ik\pi t}\Big|_{-1}^{0}\right] + \frac{A}{2}\left(\frac{-1}{ik\pi}\right)\left[e^{-ik\pi t}\Big|_{0}^{1}\right] \\ &= \frac{A}{2ik\pi}\left[1 - e^{-ik\pi}\right] - \frac{A}{2ik\pi}\left[e^{-ik\pi} - 1\right] = \frac{A}{2}\left[\frac{1 - e^{-ik\pi}}{ik\pi} + \frac{e^{-ik\pi} - 1}{-ik\pi}\right] \\ &= \frac{A}{ik\pi}\left[1 - \frac{1}{2}\left(e^{ik\pi} + e^{-ik\pi}\right)\right] \end{aligned} \tag{14.165}$$

En prenant $k = 2$, nous avons :

$$c_2 = \frac{A}{i2\pi}\left[1 - \frac{1}{2}\left(e^{i2\pi} + e^{-i2\pi}\right)\right] \tag{14.166}$$

De même pour $k = 4, 6, 8$ ainsi que tout nombre pair.
Pour ce qui est des nombres impairs, nous aurons :

$$c_1 = \frac{A}{i\pi}\left[1 - \frac{1}{2}\left(e^{i\pi} + e^{-i\pi}\right)\right] = \frac{2A}{i\pi}, \quad c_3 = \frac{A}{i3\pi}\left[1 - \frac{1}{2}\left(e^{i3\pi} + e^{-i3\pi}\right)\right] = \frac{2A}{i3\pi} \tag{14.167}$$

Les coefficients seront alors :

$$c_k = \begin{cases} 0 & k \text{ pair} \\ \frac{2A}{ik\pi} & k \text{ impair} \end{cases} \tag{14.168}$$

Il y a un seul hic dans cette relation, le coefficient c_0 ne peut être calculé selon cette relation car on peut voir que si $k = 0$ dans le résultat ci-haut, nous aurons une valeur infinie et c'est du moins impossible. Le coefficient c_0 est soit nul ou non nul mais jamais infini.

Pour trouver le coefficient c_0, nous devons calculer l'intégrale pour $k = 0$. Le coefficient est alors déterminé par :

$$\begin{aligned} c_0 &= \frac{1}{T}\int_T f(t)e^{-ik\omega t}dt = \frac{1}{2}\int_{-1}^{1} f(t)dt = \frac{1}{2}\int_{-1}^{0} f(t)dt + \frac{1}{2}\int_0^1 f(t)dt \\ &= \frac{1}{2}\int_{-1}^{0}(-A)dt + \frac{1}{2}\int_0^1 (A)dt = 0 \end{aligned} \tag{14.169}$$

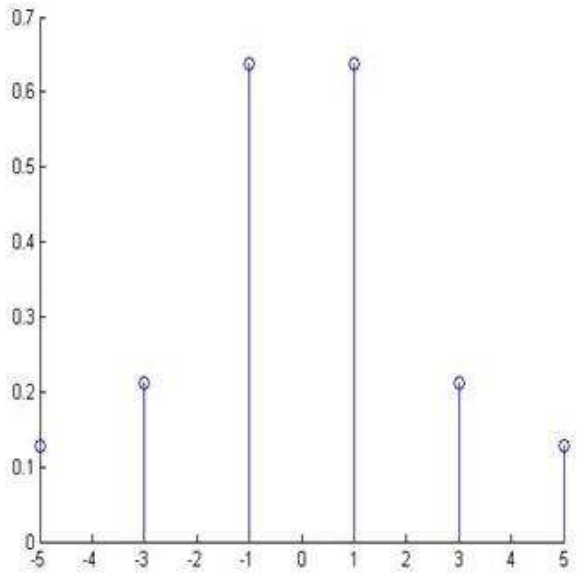


FIGURE 14.4 – Spectre en fréquence

Le spectre en "fréquence" (attention à l'abus de langage) et en amplitude sera alors de la forme de la figure 14.4 pour $k = -5, \ldots, +5$ et $A = 1$ les fréquences nulles n'étant pas représentées :

L'abus de parler de fréquence pour les coefficients de Fourier amène dont à avoir des fréquences négatives... mais ce n'est qu'une question de vocabulaire auquel il faut s'habituer.

Le spectre d'amplitude et de phase se calcule selon les relations :

$$|c_k| = \sqrt{(\Re(c_k))^2 + (\Im(c_k))^2} \text{ et } \angle\theta = \tan^{-1}\left(\frac{\Im(c_k)}{\Re(c_k)}\right) \tag{14.170}$$

Il est alors relativement aisé de remarquer que si T tend vers un nombre de plus en plus grand, les pics du spectre se rapprochent de plus en plus. Ainsi, lorsque T tend vers l'infini le spectre devient continu.

La figure 14.5 donne le spectre de phase pour les valeurs impaires.

Il est même possible pour l'exemple d'obtenir très facilement le spectre des fréquences dans MS Excel!!!

Effectivement, il suffit pour cela d'échantillonner par exemple notre signal 128 fois (MS Excel a besoin de 2^n échantillons). Nous divisons alors l'intervalle $-1 < t < 0$ en 64 échantillons et idem pour l'intervalle (crf figure 14.6.

Ce qui donne sous forme graphique la figure **??**.

Ensuite, il suffit d'aller dans le menu Outils/Utilitaire d'analyse et choisir l'option Analyse de Fourier (cfr figure 14.7).

Vient ensuite la boîte de dialogue suivante qu'il faut remplir comme indiqué à la figure 14.8.

Vient alors le résultant suivant pour les coefficients le tableau représenté à la figure 14.9.

Il reste à calculer le module des nombres complexes avec la fonction `MODULE.COMPLEXE()` de MS Excel et de diviser le résultat par 128 pour chacun des coefficients mais nous voyons déjà que chaque coefficient pair est nul ce qui correspond bien au résultat théorique.

Nous avons alors (cfr figure 14.10) en mettant l'indice n en face de chaque module.

En en traçant un graphique MS Excel à points un peu personnalisé des colonnes D et E, nous obtenons finalement le diagramme de la figure **??**.

Exemple 14.4 *Prenons un autre exemple identique au précédent mais sous une autre approche.*

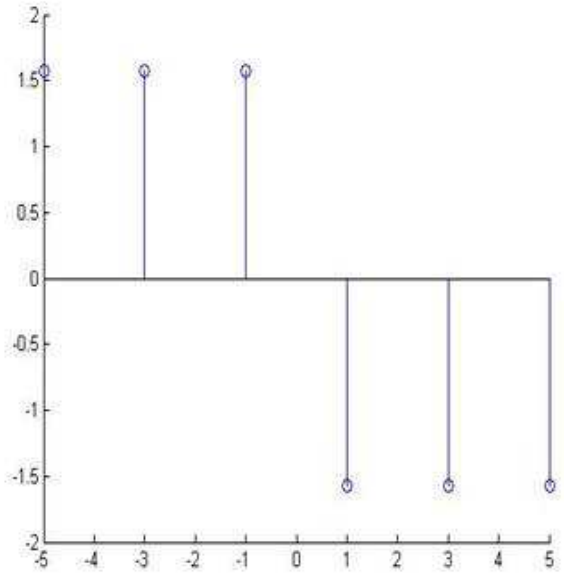


FIGURE 14.5 – Spectre de phase pour les valeurs impaires

	A	B
1	t	A
2	-1	-1
3	-0.9844	-1
4	-0.9688	-1
5	-0.9531	-1
6	-0.9375	-1
7	-0.9219	-1
8	-0.9063	-1
9	-0.8906	-1
10	-0.875	-1
11	-0.8594	-1

58	-0.125	-1
59	-0.1094	-1
60	-0.0938	-1
61	-0.0781	-1
62	-0.0625	-1
63	-0.0469	-1
64	-0.0313	-1
65	-0.0156	-1
66	0.01563	1
67	0.03125	1
68	0.04688	1
69	0.0625	1
70	0.07813	1
71	0.09375	1
72	0.10938	1

117	0.8125	1
118	0.82813	1
119	0.84375	1
120	0.85938	1
121	0.875	1
122	0.89063	1
123	0.90625	1
124	0.92188	1
125	0.9375	1
126	0.95313	1
127	0.96875	1
128	0.98438	1
129	1	1

FIGURE 14.6 – Spectre de fréquence dans une feuille MS Excel

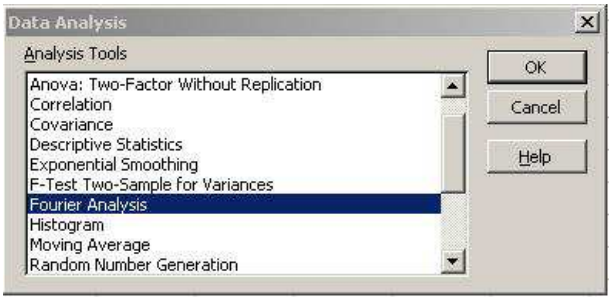


FIGURE 14.7

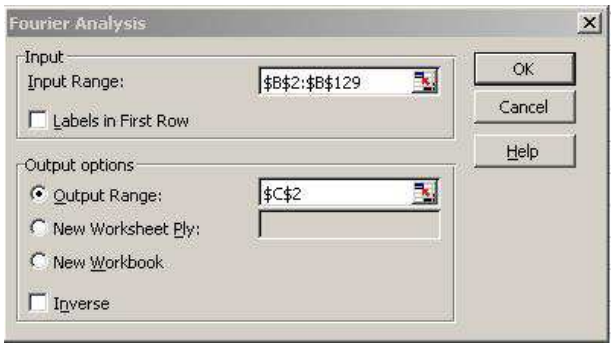


FIGURE 14.8

	A	B	C
1	t	A	C_n
2	-1	-1	0
3	-0.9844	-1	-1.99999999999994+81.4709677441666i
4	-0.9688	-1	0
5	-0.9531	-1	-1.99999999999996+27.1133384847048i
6	-0.9375	-1	0
7	-0.9219	-1	-1.99999999999997+16.2155716073538i
8	-0.9063	-1	0
9	-0.8906	-1	-1.99999999999997+11.5262840102376i
10	-0.875	-1	0
62	-0.0625	-1	0
63	-0.0469	-1	-1.99999999999998+0.147528863044911i
64	-0.0313	-1	0
65	-0.0156	-1	-1.99999999999993+4.90972442178759E-
66	0.01563	1	0
67	0.03125	1	-2-4.90972442178546E-002i
68	0.04688	1	0
69	0.0625	1	-2-0.147528863044895i
70	0.07813	1	0
71	0.09375	1	-2-0.246676472273474i
120	0.85938	1	0
121	0.875	1	-2.00000000000003-8.9064044488288i
122	0.89063	1	0
123	0.90625	1	-2.00000000000003-11.5262840102376i
124	0.92188	1	0
125	0.9375	1	-2.00000000000005-16.2155716073538i
126	0.95313	1	0
127	0.96875	1	-2.00000000000009-27.1133384847049i
128	0.98438	1	0
129	1	1	-2.00000000000032-81.4709677441667i

FIGURE 14.9

61	-0.0781	-1	-1.99999999999999+0.246676472273487i	0.0157434	-59
62	-0.0625	-1	0	0	-60
63	-0.0469	-1	-1.99999999999998+0.147528863044911i	0.0156675	-61
64	-0.0313	-1	0	0	-62
65	-0.0156	-1	-1.99999999999993+4.90972442178759E-	0.0156297	-63
66	0.01563	1	0	0	64
67	0.03125	1	-2-4.90972442178546E-002i	0.0156297	63
68	0.04688	1	0	0	62
69	0.0625	1	-2-0.147528863044895i	0.0156675	61
70	0.07813	1	0	0	60

FIGURE 14.10

Nous définissons une fonction périodique de période 2π comme suit :

$$f(x) = \begin{cases} -1 & -\pi < x < 0 \\ 1 & 0 \leq x \leq \pi \end{cases} \tag{14.171}$$

Calculons les coefficients de Fourier :

$$\begin{aligned} a_n &= \frac{2}{2\pi}\int_0^{2\pi} f(x)\cos(nx)dx = \frac{1}{\pi}\left(\int_0^{\pi} 1\cos(nx)dx + \int_{\pi}^{2\pi} -1\cos(nx)dx\right) \\ &= \frac{1}{\pi}\left(\int_0^{\pi}\cos(nx)dx - \int_{\pi}^{2\pi}\cos(nx)dx\right) \\ &= \frac{1}{\pi}\left(\frac{\sin(nx)}{n}\Big|_0^{\pi} - \frac{\sin(nx)}{n}\Big|_{\pi}^{2\pi}\right) \\ &= 0 \end{aligned} \tag{14.172}$$

et :

$$\begin{aligned} b_n &= \frac{2}{2\pi}\int_0^{2\pi} f(x)\sin(nx)dx = \frac{1}{\pi}\left(\int_0^{\pi} 1\sin(nx)dx + \int_{\pi}^{2\pi} -1\sin(nx)dx\right) \\ &= \frac{1}{\pi}\left(\int_0^{\pi}\sin(nx)dx - \int_{\pi}^{2\pi}\sin(nx)dx\right) = \frac{1}{\pi}\left(-\frac{\cos(nx)}{n}\Big|_0^{\pi} + \frac{\cos(nx)}{n}\Big|_{\pi}^{2\pi}\right) \\ &= \frac{1}{\pi}\left(\frac{\cos(nx)}{n}\Big|_{\pi}^{2\pi} - \frac{\cos(nx)}{n}\Big|_0^{\pi}\right) = \frac{1}{n\pi}\left(\cos(nx)\Big|_{\pi}^{2\pi} - \cos(nx)\Big|_0^{\pi}\right) \\ &= \frac{1}{n\pi}\big(\cos(2\pi n) - \cos(\pi n) - \cos(/pin) + 1\big) = \frac{1}{n\pi}\big(\cos(2\pi n) - 2\cos(\pi n) + 1\big) \end{aligned} \tag{14.173}$$

Nous remarquons que b_n vaut 0 pour n pair et vaut $\frac{4\pi}{n}$ pour n impair.
La série de Fourier de la fonction considérée s'écrit donc :

$$\begin{aligned} f(x) &= \sum_{n=0}^{+\infty}\big(a_n\cos(nx) + b_n\sin(nx)\big) = \sum_{n=0}^{+\infty} b_n\sin(nx) \\ &= 4\pi\left[\frac{\sin(x)}{1} + \frac{\sin(x3)}{3} + \frac{\sin(5x)}{5} + \ldots\frac{\sin(2x+1)}{2x+1} + \ldots\right] \end{aligned} \tag{14.174}$$

Ce qui en Maple s'écrit :

```
S:=(4/Pi)*Sum(sin((2*n+1)*x)/(2*n+1),n=0..N);
```

et que nous pouvons tracé à l'aide de la fonction :

```
plot(subs(N=4,S),subs(N=8,S),subs(N=16,S),
x=-Pi..Pi,color=[red,green,blue],numpoints=200);
```

Ce qui donne à la figure 14.12 trois traces pour 4, 8 et 16 termes de la série en rouge, vert et bleu.

Pour 50 termes nous obtenons (cfr figure 14.12) :

```
plot(subs(N=50,S),x=-Pi..Pi,numpoints=800);
```

Nous voyons les effets de bord appelés "phénomène de Gibbs". Il est possible de montrer avec pas mal de développements que ceux-ci arrivent à la valeur de l'abscisse correspondant à $x = \frac{\pi}{2n}$ et que le dépassement et que le pic va à 1.179 pour toute valeur de n.

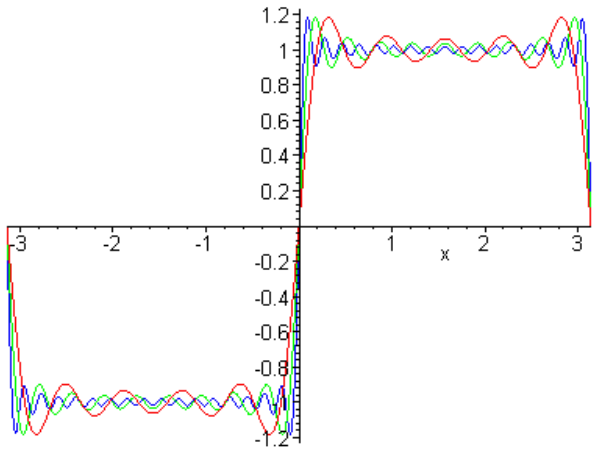


FIGURE 14.11 – Tracé avec 4, 8 et 16 termes (rouge, vert et bleu)

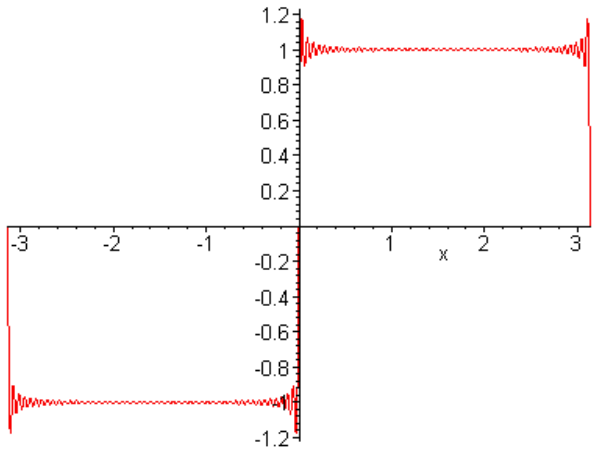


FIGURE 14.12 – Tracé avec 50 termes

Puissance d'un signal

Un signal périodique possède une énergie infinie et une puissance moyenne finie (cf. chapitre d'Électrocinétique). Sa puissance moyenne sur une période est définie par :

$$P_f = \frac{1}{T}\int_T |f(t)|dt \tag{14.175}$$

Si nous développons cette équation, nous avons :

$$\begin{aligned} P_f &= \frac{1}{T}\int_T f(t)\bar{f}(t)dt = \frac{1}{T}\int_T f(t)\left(\sum_{k=-\infty}^{\infty}\bar{c}_k e^{-ik\omega_0 t}\right)dt \\ &= \sum_{k=-\infty}^{\infty}\bar{c}_k\int_T f(t)e^{-ik\omega_0 t}dt = \sum_{k=-\infty}^{\infty}\bar{c}_k c_k = \sum_{k=-\infty}^{\infty}|c_k|^2 \end{aligned} \tag{14.176}$$

Cela signifie que la puissance d'un signal à temps continu périodique est égale à la somme des coefficients de Fourier au carré. C'est ce que l'on nomme le "théorème de Parseval". Cela signifie que si nous avons un signal quelconque que nous pouvons décomposer en série de Fourier, nous pouvons connaître la puissance de ce signal uniquement à l'aide des coefficients spectraux.

Dans la réalité, comme nous ne pouvons déterminer mathématiquement l'expression de ce signal, nous utilisons la discrétisation ou l'échantillonnage et ensuite à l'aide d'une transformée de Fourier discrète, nous pouvons calculer la puissance de ce signal en utilisant uniquement les coefficients spectraux. Cela nous donne une caractéristique du signal.

Transformée de Fourier

Les séries de Fourier sont un outil très puissant pour l'analyse de signaux périodiques par exemple, mais l'ensemble des fonctions périodiques est petit comparé à l'ensemble des fonctions que nous rencontrons dans les problèmes physiques. Ainsi, nous allons introduire un nouvel outil d'analyse extrêmement puissant qui s'étend à une classe de fonctions plus générale.

La transformée de Fourier est ainsi utilisée autant pour les signaux périodiques que pour les signaux apériodiques.

Pour cela, nous repartons de notre étude sur les séries de Fourier en notation complexe d'une fonction périodique de période T quelconque et nous faisons tendre $T \to +\infty$.

Ainsi, reprenons les expressions démontrées avant :

$$f(x) = \sum_{n=-\infty}^{+\infty} c_n e^{inx} \qquad c_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)e^{-inx}dx \tag{14.177}$$

que nous pouvons écrire de manière équivalente sous la forme :

$$f(t) = \frac{1}{T}\sum_{k=-\infty}^{+\infty} c_k e^{i\frac{2\pi k}{T}t} \qquad c_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)e^{-i\frac{2\pi k}{T}t}dt \tag{14.178}$$

et écrivons encore cela pour des besoins ultérieurs sous la forme suivante :

$$f(t) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} \frac{2\pi}{T}c_k e^{i\frac{2\pi k}{T}t} \qquad c_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)e^{-i\frac{2\pi k}{T}t}dt \tag{14.179}$$

et posons :

$$\omega_k = \frac{2\pi k}{T} \tag{14.180}$$

Ainsi, quand $T \to +\infty$, $\omega_k \to \omega$ nous passons de valeur discrète à valeur continue qui parcourt l'ensemble des réels (pour tous les k). Donc de :

$$f(t) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \frac{2\pi}{T} c_k e^{i\omega_k t} \qquad c_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) e^{-i\omega_k t} dt \tag{14.181}$$

nous passons à la limite soit :

$$T \to +\infty \qquad \omega_k \to \omega \qquad \sum_{k=-\infty}^{+\infty} \frac{2\pi}{T} \to \int_{-\infty}^{+\infty} d\omega \tag{14.182}$$

et obtenons ainsi pour les coefficients (nous changeons de notation car l'ancienne est inadaptée)

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \tag{14.183}$$

et pour la série infinie :

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{-i\omega t} d\omega \tag{14.184}$$

Attention!!! Pour faire la différence entre la fonction donnée et son équivalent dont nous cherchons l'expression en somme infinie, nous les noterons dorénavant différemment. Ainsi, il vient :

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{-i\omega t} d\omega \qquad F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \tag{14.185}$$

Définition 14.17 *Nous appelons "transformée de Fourier" de f la relation*

$$\boxed{\mathcal{F}\big(f(t)\big)(\omega) = F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt} \tag{14.186}$$

Définition 14.18 *Nous appelons "transformée de Fourier inverse" de F la relation*

$$\boxed{\mathcal{F}^{-\infty}\big(f(\omega)\big)(t) = f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{-i\omega t} d\omega} \tag{14.187}$$

Remarque Il existe de nombreuses manière d'écrire la transformée de Fourier en fonction du choix de la valeur initiale de T .

Certains physiciens préfèrent symétriser ces deux expressions en mettant le même coefficient dans les deux sens, qui sera par exemple $\frac{1}{\sqrt{2\pi}}$. Cela donnera :

$$\begin{aligned} \mathcal{F}\big(f(t)\big)(\omega) = F(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \\ \mathcal{F}^{-\infty}\big(f(\omega)\big)(t) = f(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(\omega) e^{-i\omega t} d\omega \end{aligned} \tag{14.188}$$

Donnons également la forme tridimensionnelle qui nous servira de nombreuses fois en mécanique ondulatoire, électrodynamique, optique ondulatoire ou encore dans les divers chapitres de physique quantique :

$$\mathcal{F}(\vec{k}) = \frac{1}{(2\pi)^{\frac{3}{2}}} \iiint f(\vec{r}) e^{-i\vec{k}\circ\vec{r}} d^3r \qquad f(\vec{r}) = \frac{1}{(2\pi)^{\frac{3}{2}}} \iiint F(\vec{k}) e^{i\vec{k}\circ\vec{r}} d^3k \tag{14.189}$$

Pour que les choses soient peut-être plus claires, montrons de manière générale que la transformée de Fourier $\mathcal{F}$ précédemment écrite est une isométrie (conserve la norme).

Remarquons tout d'abord que pour tout f, g nous avons le produit scalaire fonctionnel :

$$f \circ \mathcal{F}(g) = \mathcal{F}^{-1}(f) \circ g \tag{14.190}$$

Mais puisque les fonctions sont dans l'espace des complexes, comme nous l'avons vu dans le chapitre de calcul vectoriel, nous devons alors utiliser la notation du produit hermitien :

$$\langle f|\mathcal{F}(g)\rangle = \langle \mathcal{F}^{-1}(f)|g\rangle \tag{14.191}$$

Rappelons quand même que :

$$\langle f|g\rangle = \int_{\mathbb{R}^n} f(\vec{r})\overline{g(\vec{r})}d^n r \tag{14.192}$$

Preuve D'abord, nous avons donc :

$$\langle f|\mathcal{F}(g)\rangle = \int_{\mathbb{R}^n} f(\vec{r})\overline{\mathcal{F}(g)}d^n r \tag{14.193}$$

Mais les variables à intégrer doivent être les mêmes et pour que $\mathcal{F}(g)$ soit implicitement dépendant de $\vec{r}$ il faut donc prendre la transforme de Fourier en $\vec{k}$. Tel que :

$$\mathcal{F}(\vec{r}) = \frac{1}{(2\pi)^{\frac{3}{2}}} \iiint f(\vec{k})e^{-i\vec{k}\circ\vec{r}}d^3 k \tag{14.194}$$

Ainsi :

$$\begin{aligned}\langle f(\vec{r})|\mathcal{F}(g)(\vec{r})\rangle &= \int_{\mathbb{R}^n} f(\vec{r})\overline{\mathcal{F}(g)}d^n r \int_{\mathbb{R}^n} f(\vec{r})\overline{\left(\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} g(\vec{k})e^{-i\vec{k}\circ\vec{r}}d^n k\right)}d^n r \\ &= \int_{\mathbb{R}^n} f(\vec{r})\left(\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n}\overline{g(\vec{k})}e^{-i\vec{k}\circ\vec{r}}d^n k\right)d^n r\end{aligned} \tag{14.195}$$

Soit en utilisant le théorème de Fubini (cf. chapitre de Calcul Différentiel Et Intégral) :

$$\begin{aligned}\langle f|\mathcal{F}(g)\rangle &= \int_{\mathbb{R}^n}\overline{g(\vec{k})}\left(\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} f(\vec{r})e^{i\vec{k}\circ\vec{r}}d^n r\right)d^n k \\ &= \int_{\mathbb{R}^n}\overline{g(\vec{k})}\mathcal{F}^{-1}(f)(\vec{k})d^n k \\ &= \langle \mathcal{F}^{-1}(f), g\rangle\end{aligned} \tag{14.196}$$

A l'aide de ce résultat, nous avons donc démontré que :

$$\langle \mathcal{F}(f), \mathcal{F}(g)\rangle = \langle \mathcal{F}^{-1}(\mathcal{F}(f)), g\rangle = \langle f, g\rangle \tag{14.197}$$

Nous n'avons pas précisé les bornes : elles sont infinies dans chaque définition (nous intégrons sur tous les $\vec{k}$ ou $\vec{r}$ possibles). •

Voyons maintenant deux propriétés intéressantes de la transformée de Fourier :

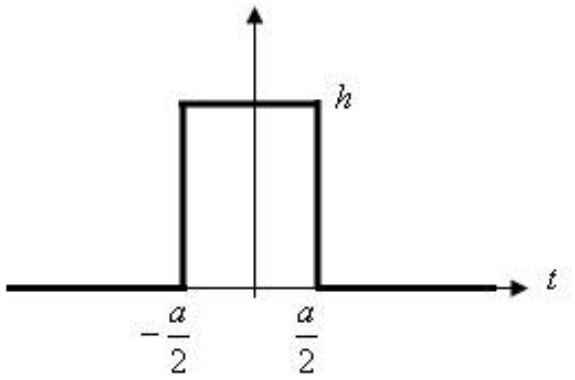


FIGURE 14.13 – Exemple de fonction

P1. Si f est paire, il vient une simplification de la transformée telle que :

$$\begin{aligned}
\mathcal{F}\big(f(x)\big)(\omega) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{0} f(t)e^{-i\omega t}dt + \int_{0}^{+\infty} f(t)e^{-i\omega t}dt\right)\\
&= \frac{1}{\sqrt{2\pi}}\left(-\int_{-\infty}^{0} f(t)e^{i\omega t}dt + \int_{0}^{+\infty} f(t)e^{-i\omega t}dt\right)\\
&= \frac{1}{\sqrt{2\pi}}\left(\int_{0}^{+\infty} f(t)e^{i\omega t}dt + \int_{0}^{+\infty} f(t)e^{-i\omega t}dt\right)\\
&= \frac{1}{\sqrt{2\pi}}\int_{0}^{+\infty} f(t)\left(e^{i\omega t} + e^{-i\omega t}\right)dt\\
&= \frac{2}{\sqrt{2\pi}}\int_{0}^{+\infty} f(t)\cos(\omega t)dt
\end{aligned} \tag{14.198}$$

P2. Si f est impaire, nous procédons de la même manière que ci-dessus et nous obtenons :

$$\mathcal{F}\big(f(x)\big)(\omega) = -\frac{2i}{\sqrt{2\pi}}\int_{0}^{+\infty} f(t)\sin(\omega t)dt \tag{14.199}$$

REMARQUE La branche de "l'analyse harmonique", ou "analyse de Fourier 2D", est la branche des mathématiques qui étudie la représentation des fonctions ou des signaux comme superposition d'ondes de base. Elle approfondit et généralise les notions de série de Fourier et de transformée de Fourier. Les ondes de base s'appellent les harmoniques, d'où le nom de la discipline. Durant ces deux derniers siècles, elle a eu de nombreuses applications en physiques sous le nom d'analyse spectrale, et connaît des applications récentes notamment en traitement des signaux, mécanique quantique, neurosciences, stratigraphie...

Exemple 14.5 *Voyons donc un exemple (parmi les deux fondamentaux) d'une transformée de Fourier que nous retrouvons en physique quantique aussi bien qu'en optique ondulatoire.*

Nous allons calculer la transformée de Fourier de la fonction suivante :

Nous avons donc :

$$\begin{aligned}
\mathcal{F}\big(f(t)\big) &= \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt = \int_{-\infty}^{+\infty} f(t)e^{-i2\pi f\cdot t}dt = \int_{-\frac{a}{2}}^{+\frac{a}{2}} h\cdot e^{-i2\pi f\cdot t}dt\\
&= h\cdot\int_{-\frac{a}{2}}^{+\frac{a}{2}} e^{-i2\pi f\cdot t}dt = -\frac{h}{i2\pi f}\left[e^{-i2\pi f\cdot t}\right]_{-\frac{a}{2}}^{+\frac{a}{2}} = -\frac{h}{i2\pi f}\left(e^{-i\pi f\cdot a} - e^{i\pi f\cdot a}\right)\\
&= -\frac{h}{i2\pi f}\left(e^{i(-\pi f\cdot a)} - e^{i(\pi f\cdot a)}\right) = -\frac{h}{i2\pi f}\big(-2i\sin(\pi f a)\big) = h\cdot\frac{\sin(\pi f a)}{\pi f}\\
&= h\cdot a\cdot\frac{\sin(\pi f a)}{\pi f a} = h\cdot a\cdot sinc(\pi f a)
\end{aligned} \tag{14.200}$$

où sinc est le sinus cardinal. Nous retombons donc sur le sinus cardinal (si nous prenons le module au carré) de la décomposition d'une onde monochromatique diffractée par une fente rectangulaire. Ainsi, il semble possible d'étudier les phénomènes de diffraction en utilisant la transformée de Fourier et ce domaine se nomme "l'optique de Fourier".

14.2.6 Fonctions de Bessel

Les fonctions de Bessel sont très utiles dans de nombreux domaines de pointe de la physique faisant intervenir des équations différentielles délicates à résoudre. Les domaines dans lesquelles nous les trouvons le plus souvent sont la calorimétrie (conduction de la chaleur), la physique nucléaire (physique de réacteurs), et la mécanique des fluides.

Ces séries sont cependant très peu détaillées dans les écoles universitaires et il est souvent du rôle de l'élève de chercher les compléments d'informations dont il a besoin sur le sujet dans la bibliothèque de son école. Nous avons voulu présenter ici les développements permettant d'éviter cette démarche tout en restant chez soi devant son ordinateur (de plus les livres sur le sujet sont assez rares...).

REMARQUE Nous parlons habituellement par abus de langage des "fonctions de Bessel" au lieu des "séries de Bessel".

Il existe une quantité non négligeable de fonctions de Bessel mais nous allons nous restreindre à l'étude de celles qui sont les plus utilisées en physique.

Fonction de Bessel d'ordre zéro

La fonction connue sous le nom de "fonction de Bessel d'ordre zéro", est définie par la série de puissances :

$$\boxed{J_0(x) = 1 - \frac{x^2}{2^2} + \frac{x^4}{2^2 \cdot 4^2} - \frac{x^6}{2^2 \cdot 4^2 \cdot 6^2} + \dots} \tag{14.201}$$

C'est lors de l'étude des propriétés de dérivation et d'intégration que Bessel a trouvé que cette série de puissance est une solution à une équation différentielle que l'on retrouve assez fréquemment en physique. C'est pourquoi elle porte son nom.

Si u_r représente le r-ième terme de la série, nous voyons aisément que :

$$\frac{u_{r+1}}{u_r} = \frac{-x^2}{(2r)^2} \tag{14.202}$$

qui tend vers zéro quand $r \to \infty$, quelque soit la valeur de x. Cela a pour conséquence que la série converge pour toutes les valeurs de x. Comme il s'agit d'une série de puissance positive, la fonction $J_0(x)$ et toutes ses dérivées sont continues pour toutes valeurs de x, réelles ou complexes.

Fonction de Bessel d'ordre n

La fonction , connue sous le nom de "fonction de Bessel d'ordre n", est définie, lorsque n est un entier positif, par la série de puissance :

$$\boxed{J_n(x) = \frac{-x^n}{2^n n!}\left(1 - \frac{x^2}{2 \cdot (2n+2)} + \frac{x^4}{2 \cdot 4(2n+2)(2n+4)} - \dots\right)} \tag{14.203}$$

qui converge pour toutes valeurs de x, réelles ou complexes (cfr figure 14.14).

En particulier, pour nous avons :

$$J_1(x) = \frac{x}{2} - \frac{x^3}{2^2 \cdot 4} + \frac{x^5}{2^2 \cdot 4^2 \cdot 6} - \frac{x^7}{2^2 \cdot 4^2 \cdot 6^2 \cdot 8} + \dots \tag{14.204}$$

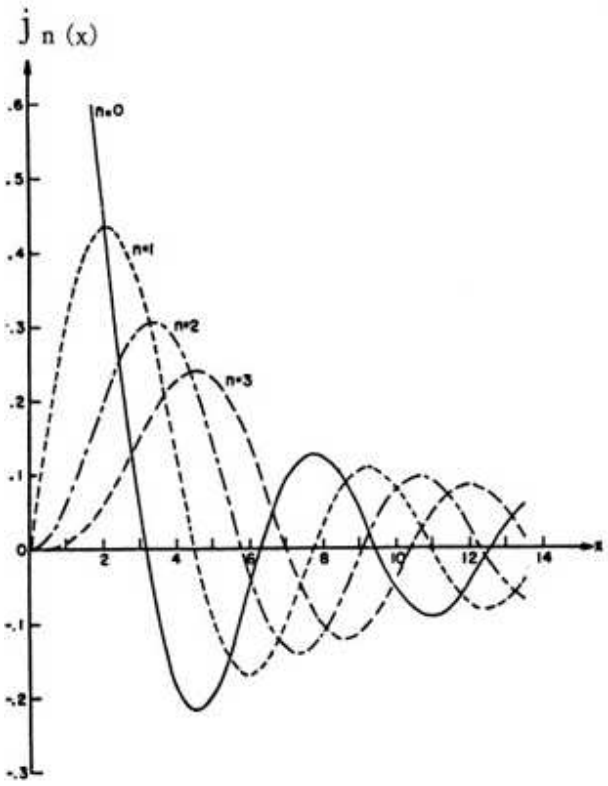


FIGURE 14.14 – Fonction de bessel de différents ordres

et quand :

$$J_2(x) = \frac{x^2}{2 \cdot 4} - \frac{x^4}{2^2 \cdot 4 \cdot 6} + \frac{x^6}{2^2 \cdot 4^2 \cdot 6 \cdot 8} - \frac{x^8}{2^2 \cdot 4^2 \cdot 6^2 \cdot 8 \cdot 10} + \dots \tag{14.205}$$

Nous pouvons noter que $J_n(x)$ est une fonction paire de x quand n est paire, et impaire quand n est impaire (cf. chapitre d'Analyse Fonctionnelle).

En dérivant la fonction $J_0(x)$ et en comparant le résultat avec la série $J_1(x)$, nous voyons sans trop de peine que :

$$\frac{dJ_0(x)}{dx} = -J_1(x) \tag{14.206}$$

Nous trouvons également sans trop de difficulté, la relation suivante :

$$\frac{d}{dx} x J_1(x) = x J_0(x) \tag{14.207}$$

En utilisant le fait que :

$$\frac{dJ_0(x)}{dx} = -J_1(x) \tag{14.208}$$

et en l'incluant dans la précédente relation, nous trouvons :

$$\frac{d}{dx}\left(x \frac{dJ_0(x)}{dx}\right) + x J_0(x) = 0 \tag{14.209}$$

ou écrit autrement :

$$x\frac{d^2J_0(x)}{dx^2}+\frac{dJ_0(x)}{dx}+xJ_0(x)=0 \tag{14.210}$$

$y=J_0(x)$ est donc une solution de l'équation différentielle du second ordre :

$$\frac{d}{dx}\left(x\frac{dy}{dx}\right)+xy=0 \tag{14.211}$$

ou écrit autrement :

$$x\frac{d^2y}{dx^2}+\frac{dy}{dx}+xy=0 \tag{14.212}$$

ou encore :

$$\frac{d^2y}{dx^2}+\frac{1}{x}\frac{dy}{dx}+y=0 \tag{14.213}$$

Une solution à une équation de Bessel de paramètre n qui n'est pas un multiple de $J_n(x)$ est appelé "fonction de Bessel du second type". Supposons que u est une telle fonction et posons $v=J_0(x)$; alors d'après la relation :

$$x\frac{d^2y}{dx^2}+\frac{dy}{dx}+xy=0 \tag{14.214}$$

nous avons :

$$xu''+u'+xu=0 \text{ et } xv''+v'+xv=0 \tag{14.215}$$

En multipliant la première relation par v et la seconde par u et après soustraction, nous obtenons :

$$x(u''v-uv'')+u'v-uv'=0 \tag{14.216}$$

nous avons donc également :

$$\frac{d}{dx}(u'v-uv')=u''v+u'v'-u'v'-uv''=u''v-uv'' \tag{14.217}$$

nous pouvons donc écrire :

$$\frac{d}{dx}[x(u'v-uv')]=0 \tag{14.218}$$

effectivement car si nous développons, nous trouvons :

$$\begin{aligned}\frac{d}{dx}[x(u'v-uv')]&=1\cdot u'v+xu''v+xu'v'-1\cdot uv'-xu'v'-xuv''\\&=x(u''v+u'v'-u'v'-uv'')+u'v-uv'\\&=x(u''v-uv'')+u'v-uv'=0\end{aligned} \tag{14.219}$$

Pour que l'égalité :

$$\frac{d}{dx}[x(u'v-uv')]=0 \tag{14.220}$$

soit satisfaite, nous avons :

$$x(u'v - uv') = c^{te} = B \tag{14.221}$$

En divisant par xv^2, nous avons :

$$\frac{(u'v - uv')}{v^2} = \frac{B}{xv^2} \tag{14.222}$$

ce qui est équivalent à :

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{B}{xv^2} \tag{14.223}$$

de suite, par intégration il vient :

$$\frac{u}{v} = A + B\int \frac{1}{xv^2}dx \tag{14.224}$$

où A est une constante. Consécutivement nous avons, puisque $v = J_0(x)$:

$$u = AJ_0(x) + BJ_0(x)\int \frac{1}{xJ_0^2(x)}dx \tag{14.225}$$

où rappelons-le, A et B sont des constantes, et $B \neq 0$ si u n'est pas un multiple de $J_0(x)$ par définition.

Si dans la dernière relation, $J_0(x)$ est remplacé par son expression en termes de série nous avons :

$$\frac{1}{xJ_0^2(x)} = \frac{1}{x} + \frac{x}{2} + \frac{5x^3}{32} + \dots \tag{14.226}$$

dès lors :

$$\begin{aligned} J_0(x)\int \frac{1}{xJ_0^2(x)}dx &= J_0(x)\left[\log(x) + \frac{x^2}{4} + \frac{5x^4}{128} + \dots\right] \\ &= J_0(x)\log(x) + \left(1 - \frac{x^2}{2^2} + \dots\right)\left(\frac{x^2}{4} + \frac{5x^4}{128} + \dots\right) \\ &= J_0(x)\log(x) + \frac{x^2}{4} - \frac{3x^4}{128} + \dots \end{aligned} \tag{14.227}$$

consécutivement si nous posons :

$$\boxed{Y_0(x) = J_0(x)\log(x) + \frac{x^2}{4} - \frac{3x^4}{128} + \dots} \tag{14.228}$$

où $Y_0(x)$ est une fonction de Bessel particulière du second type appelée "fonction de Bessel-Neumann du second type d'ordre nul".

Identiquement au fait que $J_0(x) \to 1$ quand $x \to 0$, l'expression $Y_0(x)$ à cause du terme $\log(x)$ quand x est petit tend vers $Y_0(x) \to -\infty$ quand $x \to +0$.

Finalement, il vient de ce que nous avons vu précédemment que $J_0(x)$ et $Y_0(x)$ sont des solutions indépendantes de l'équation différentielle :

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + y = 0 \tag{14.229}$$

La solution générale étant donc :

$$y = AJ_0(x) + BY_0(x) \tag{14.230}$$

où A, B sont des constantes arbitraires et $x > 0$ afin que $Y_0(x)$ soit réel.
Si nous remplaçons x par kx, où k est une constante, l'équation différentielle devient :

$$\frac{1}{k^2}\frac{d^2y}{dx^2} + \frac{1}{kx}\frac{dy}{kdx} + y = 0 \tag{14.231}$$

en multipliant le tout par k^2, nous trouvons la forme générale de l'équation différentielle :

$$\boxed{\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + k^2y = 0} \tag{14.232}$$

dont la solution générale est :

$$\boxed{y = AJ_0(kx) + BY_0(kx)} \tag{14.233}$$

où $k > 0$ afin que $Y_0(kx)$ soit réel quand $x > 0$.
Au fait, les fonctions de Bessel viennent des solutions de l'équation différentielle étudiée précédemment et solutionnées par la méthode de Frobenius. Posons :

$$L_y \equiv \left(x\frac{d^2}{dx^2} + \frac{d}{dx} + x\right) y \tag{14.234}$$

et faisons la substitution :

$$y = x\rho + c_1x^{\rho+1} + c_2x^{\rho+2} + c_3x^{\rho+3} + \dots \tag{14.235}$$

en substituant dans Ly, nous obtenons

$$Ly = \rho^2x^{\rho-1} + (\rho+1)^2c_1x^{\rho} + \left[(\rho+2)^2c_2 + 1\right]x^{\rho+1} + \left[(\rho+3)^2c_3 + c_1\right]x^{\rho+1} \tag{14.236}$$

Choisissons maintenant les afin de satisfaire l'équation différentielle tel que :

$$\begin{aligned} (\rho+1)^2c_1 &= 0 \\ (\rho+2)^2c_1 + 1 &= 0 \\ (\rho+3)^2c_3 + c_1 &= 0 \\ &\dots \end{aligned} \tag{14.237}$$

Dès lors, à moins que soit un entier négatif, nous avons :

$$\begin{aligned} &c_1 = c_3 = c_5 = c_7 = \dots = 0 \\ &c_2 = -\frac{1}{(\rho+2)^2} \\ &c_4 = -\frac{c_2}{(\rho+4)^2} = \frac{1}{(\rho+2)^2(\rho+4)^2} \\ &\dots \end{aligned} \tag{14.238}$$

En substituant ces valeurs dans la relation $y = x\rho + c_1 x^{\rho+1} + c_2 x^{\rho+2} + c_3 x^{\rho+3} + \ldots$, nous obtenons :

$$\boxed{y = x^\rho \left(1 - \frac{x^2}{(\rho+2)^2} + \frac{1}{(\rho+2)^2(\rho+4)^2} - \cdots\right)} \tag{14.239}$$

dès lors :

$$x\frac{d^2y}{dx^2} + \frac{dy}{dx} + xy = Ly = \rho^2 x^{\rho-1} \tag{14.240}$$

si nous posons $\rho = 0$ dans l'avant-dernière relation, nous obtenons :

$$y = x^\rho \left(1 - \frac{x^2}{(\rho+2)^2} + \frac{1}{(\rho+2)^2(\rho+4)^2} - \cdots\right) = J_0(x) \tag{14.241}$$

Equation différentielle de Bessel d'ordre n

Nous avons défini les séries de Bessel comme étant :

$$J_n(x) = \frac{x^n}{2^n n!}\left(1 - \frac{x^2}{2\cdot 2n+2} + \frac{x^4}{2\cdot 4\cdot 2n+2\cdot 2n+4} - \cdots\right) \tag{14.242}$$

Posons :

$$y = J_n(x)2^n n! = x^n - \frac{x^{n+2}}{2\cdot 2n+2} + \frac{x^{n+4}}{2\cdot 4\cdot 2n+2\cdot 2n+4} - \cdots \tag{14.243}$$

et différencions ainsi :

$$x\frac{d}{dx}\left(x\frac{dy}{dx}\right) = n^2x^n - \frac{(n+2)^2x^{n+2}}{2\cdot 2n+2} + \frac{(n+4)^2x^{n+4}}{2\cdot 4\cdot 2n+2\cdot 2n+4} - \cdots \tag{14.244}$$

Mais nous avons aussi :

$$n^2y = n^2x^n - \frac{n^2x^{n+2}}{2\cdot 2n+2} + \frac{n^2x^{n+4}}{2\cdot 4\cdot 2n+2\cdot 2n+4} - \cdots \tag{14.245}$$

Par soustraction :

$$x\frac{d}{dx}\left(x\frac{dy}{dx}\right) - n^2y = -x^{n+2} + \frac{x^{n+4}}{2\cdot 2n+2} - \cdots = -x^2y \tag{14.246}$$

Ce qui donne finalement :

$$x\frac{d}{dx}\left(x\frac{dy}{dx}\right) + (x^2 - n^2)y = 0 \tag{14.247}$$

Ce qui s'écrit également :

$$x^2\frac{d^2y}{dx} + x\frac{dy}{dx} + (x^2 - n^2)y = 0 \tag{14.248}$$

Qui est appelé "l'équation différentielle de Bessel d'ordre n" ou plus simplement "équation de Bessel". Au fait, la plupart des écoles ou sites Internet donnent cette équation différentielle comme une définition et pourtant il est clair qu'il y a raisonnement rigoureux derrière cette équation.

La solution est donc du type :

$$y = 2^n n! J_n(x) \tag{14.249}$$

ce qui s'écrit encore parfois en utilisant la fonction gamma d'Euler :

$$y = 2^n \Gamma_0(n) J_n(x) = c^{te} J_n(x) \tag{14.250}$$

Il s'ensuite que :

$$x^2 \frac{d^2 J_n(x)}{dx} + x \frac{dJ_n(x)}{dx} + (x^2 - n^2) J_n(x) = 0 \tag{14.251}$$

et donc que $y = J_n(x)$ est solution de cette équation différentielle.

14.3 Critère de convergence

Lorsque nous étudions une série, l'une des questions fondamentales est celle de la convergence ou de la divergence de cette série. Si une série converge, son terme général tend vers zéro lorsque n tend vers l'infini :

$$\lim_{n \to \infty} a_n = 0 \tag{14.252}$$

Ce critère est nécessaire mais non suffisant pour établir la convergence d'une série. Par contre, si ce critère n'est pas rempli, on est absolument sûr que la série ne converge pas (donc elle diverge !).

Trois méthodes sont proposées pour approfondir le critère de convergence :

1. Le test de l'intégrale
2. La règle d'Alembert
3. La règle de Cauchy

Dans les paragraphes suivants, nous admettrons des séries à terme positifs. Le cas de la série alternée sera vu ultérieurement.

14.3.1 Test de l'intégrale

Soit la série à termes positifs décroissants :

$$a_1 + a_2 + a_3 + \cdots + a_n + \cdots \tag{14.253}$$

c'est-à-dire :

$$a_1 \geq a_2 \geq a_3 \geq \cdots \geq a_n \geq \cdots \tag{14.254}$$

et soit une fonction continue décroissante telle que :

$$f(1) = a_1, f(2) = a_2, \cdots f(n) = a_n \tag{14.255}$$

nous pouvons alors affirmer que :

1. Si l'intégrale :

$$\int_1^\infty f(x)dx \tag{14.256}$$

converge, la série converge également.

2. Si l'intégrale :

$$\int_1^\infty f(x)dx \tag{14.257}$$

diverge, la série diverge également.

REMARQUE En aucun cas l'intégrale ne donne la valeur de la somme de la série ! Le test de l'intégrale donne simplement une indication sur la convergence de la série. Avant de faire le test de l'intégrale, il est important de vérifier que les termes de la série soient strictement décroissants afin de remplir la condition $a_1 \geq a_2 \geq a_3 \geq \cdots \geq a_n \geq \cdots$.

14.3.2 Règle d'Alembert

Si dans une série à termes positifs :

$$a_1 + a_2 + a_3 + \cdots + a_n + \cdots \tag{14.258}$$

le rapport $\frac{a_{n+1}}{a_n}$ (assimilable à une fonction prise en son entier) a une limite finie L lorsque $n \to \infty$:

$$L = \lim_{n\to\infty} \frac{a_{n+1}}{a_n} \tag{14.259}$$

1. Si $L < 1$, la série converge
2. Si $L > 1$, la série diverge
3. Si $L = 1$, on ne peut rien dire

et nous définissons le "rayon de convergence" comme :

$$R = \frac{1}{L} \tag{14.260}$$

14.3.3 Règle de Cauchy

Si dans une série à termes positifs :

$$a_1 + a_2 + a_3 + \cdots + a_n + \cdots \tag{14.261}$$

la quantité $\sqrt[n]{a_n}$ a une limite finie L lorsque $n \to \infty$ telle que :

$$L = \lim_{n\to\infty} \sqrt[n]{a_n} \tag{14.262}$$

avec à nouveau les mêmes considération que pour la règle d'Alembert :

1. Si $L < 1$, la série converge
2. Si $L > 1$, la série diverge
3. Si $L = 1$, on ne peut rien dire

14.3.4 Théorème de Leibniz

Nous avons considéré jusqu'à présent des séries à termes positifs. Nous allons considérer dans cette partie des séries dont les termes sont alternés, c'est-à-dire des séries de la forme :

$$a_1 - a_2 + a_3 - a_4 + \cdots \quad (a_n > 0) \tag{14.263}$$

Définition 14.19 *Une série est dite "série alternée" si deux termes consécutifs de cette série sont de signe contraire*

Si dans une série alternée les termes en valeur absolue vont en décroissant :

$$a_1 > a_2 > a_3 > \cdots > a_n > \cdots \tag{14.264}$$

et si :

$$\lim_{n\to\infty} a_n = 0 \tag{14.265}$$

alors la série converge, sa somme est positive et n'est pas supérieure au premier terme.
Si S est la somme de la série et S_n une somme partielle, alors :

$$|S - S_n| \leq a_{n+1} \tag{14.266}$$

REMARQUE Il est important de vérifier que les valeurs absolues des termes de la série soient strictement décroissants afin de remplir la condition précédente.

14.3.5 Convergence absolue

Définition 14.20 *Une série à termes variables est dite absolument convergent si la série formée avec la valeur absolue de ses termes converge :*

$$a_1 + |-a_2| + a_3 + |-a_4| + \cdots + |-a_n| + \cdots \quad (a_n > 0) \tag{14.267}$$

Si une série alternée de termes est absolument convergente, la série absolue qui en découle converge aussi.
Nous pouvons généraliser la règle d'Alembert au cas des séries à termes quelconques :

$$a_1 \pm a_2 \pm a_3 \pm a_4 \pm \cdots \pm a_n \pm \cdots \quad (a_n > 0) \tag{14.268}$$

Ainsi, le rapport $\left|\frac{a_{n+1}}{a_n}\right|$ a une limite finie L lorsque pour $n \to \infty$ nous avons :

$$L = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| \tag{14.269}$$

toujours avec les mêmes conclusions que pour la règle d'Alembert normale.

14.3.6 Théorème du point fixe

Le théorème du point fixe n'est pas vraiment utile en physique (implicitement il est indispensable mais les physiciens utilisent souvent des outils mathématiques dont les propriétés ont déjà été validées au préalable par des mathématiciens), cependant nous le retrouvons en théorie du chaos (les vortex, tourbillons, etc...) ainsi qu'en informatique théorique (voir chapitre traitant des fractales en particulier le triangle de Sierpinski). Nous ne saurions donc que recommander au lecteur de prendre le temps de lire et de comprendre les explications qui vont suivre.

Soit (X,d), un espace métrique complet (cf. chapitre de Topologie ou des Fractales) et soit $T : X \to X$ une application strictement contractante de constante L (voir les fonctions lipschitziennes chapitre de topologie), alors il existe un unique point $\omega \in X$ tel que $T(\omega) = \omega \cdot \omega$ est alors dit le "point fixe" de T. De plus si nous notons par :

$$T^n(x) = \underbrace{T(T(T\ldots T}_{n\,fois}(x)\ldots)) \tag{14.270}$$

l'image de x par le n-ième itéré de T, nous avons alors :

$$\omega = \lim_{t\to\infty}\big(T^n(x)\big) \quad \forall x \in X \tag{14.271}$$

et la vitesse de convergence peut d'ailleurs être estimée par :

$$d\big(\omega, T^n(x)\big) \leq d\big(x, T(x)\big)\frac{L^n}{1-L} \tag{14.272}$$

Preuve Soit $x \in X$ nous considérons la suite $\big(T^n(x)\big)_{n\in\mathbb{N}}$ définie comme ci-dessus. Nous allons d'abord montrer que cette suite est une suite de Cauchy (voir plus haut sur la présente page ce qu'est une suite de Cauchy).

En appliquant l'inégalité triangulaire (cf. chapitre d'Analyse Vectorielle) plusieurs fois nous avons :

$$\begin{aligned} d\big(T^{n+k}(x), T^n(x)\big) \leq & d\big(T^{n+k}(x), T^{n+k-1}(x)\big) + d\big(T^{n+k-1}(x), T^{n+k-2}(x)\big) \\ & + \cdots + d\big(T^{n+1}(x), T^n(x)\big) \end{aligned} \tag{14.273}$$

Or :

$$d\big(T^{n+1}(x), T^n(x)\big) \leq L \cdot d\big(T^n(x), T^{n-1}(x)\big) + \leq \cdots \leq L^n \cdot d\big(T(x), x\big) \tag{14.274}$$

donc :

$$\begin{aligned} d\big(T^{n+k}(x), T^n(x)\big) &\leq (L^{n+k-1} + L^{n+k-2} + \cdots + L^n) \cdot d\big(T(x), x\big) \\ &\leq L^n \cdot d\big(T(x), x\big) \cdot (1 + L + L^2 + \ldots + L^{k-1}) \\ &\text{et} \\ (1 + L + L^2 + \ldots + L^{k-1}) &\leq \frac{1-L^k}{1-L} \leq \frac{1}{1-L} \\ &\text{ainsi} \\ d\big(T^{n+k}(x), T^n(x)\big) &\leq \frac{L^n}{1-L} \cdot d\big(T(x), x\big) \end{aligned} \tag{14.275}$$

pour finir :

$$0 \leq L \leq 1 \Rightarrow \lim_{n\to\infty}\frac{L^n}{1-L} = 0 \tag{14.276}$$

c'est-à-dire que dans un premier temps $\big(T^n(x)\big)_{n\in\mathbb{N}}$ est bien une suite de Cauchy.

(X,d) étant un espace complet nous avons que $\big(T^n(x)\big)_{n\in\mathbb{N}}$ converge, et nous posons :

$$\omega = \lim_{n\to\infty} T^n(x) \tag{14.277}$$

A présent, nous vérifions que ω est bien un point fixe de T. En effet T est uniformément continue (car lipschitzienne - voir le chapitre de topologie) donc à fortiori continue ainsi :

$$\omega = \lim_{n\to\infty} T^n(x) = \lim_{n\to\infty} T\big(T^{n-1}(x)\big) \underbrace{=}_{\text{car T est continue}} T\big(\lim_{n\to\infty} T^{n-1}(x)\big) = T(\omega) \tag{14.278}$$

Il reste à vérifier que ω est l'unique point fixe (du coup nous aurons démontré que ω ne dépend pas du choix de x). Supposons que nous ayons aussi $T(y) = y$ alors :

$$d(y,\omega) = d\big(T(y),T(\omega)\big) \leq L\cdot d(y,\omega) \Rightarrow d(y,\omega) = 0 \Rightarrow y = \omega \tag{14.279}$$

Une estimation de la vitesse de convergence est donnée par :

$$d\big(T^{n+k}(x),T^n(x)\big) \leq \frac{L^n}{1-L}\cdot d\big(T(x),x\big) \tag{14.280}$$

$d(,)$ est continue par rapport à chacune des variables donc :

$$\lim_{k\to\infty} d\big(T^{n+k}(x),T^n(x)\big) = d\big(\lim_{k\to\infty} T^{n+k}(x),T^n(x)\big) = d\big(\omega,T^n(x)\big) \tag{14.281}$$

et les limites préservent les inégalités (non strictes) donc :

$$d\big(\omega,T^n(x)\big) \leq \frac{L^n}{1-L}\cdot d\big(T(x),x\big) \tag{14.282}$$

•

15

Calcul vectoriel

Le calcul vectoriel ou "analyse vectorielle" est une branche des mathématiques qui étudie les champs de scalaires et de vecteurs suffisamment réguliers des espaces euclidiens (voir définition plus loin).

L'importance du calcul vectoriel provient de son utilisation intensive en physique et dans les sciences de l'ingénieur. C'est de ce point de vue que nous la présenterons, et c'est pourquoi nous nous limiterons le plus souvent au cas de l'espace usuel à trois dimensions. Dans ce cadre, un champ de vecteurs associe à chaque point de l'espace un vecteur (à trois composantes réelles), tandis qu'un champ de scalaires y associe un réel.

Remarque Imaginons par exemple l'eau d'un lac. La donnée de sa température en chaque point forme un champ de scalaires, celle de sa vitesse en chaque point, un champ de vecteurs (voir définition plus loin).

Des notions physiques telles que la force ou la vitesse sont caractérisées par une direction, un sens et une intensité. Ce triple caractère est mis en évidence par les flèches. Celles-ci sont à l'origine de la notion de vecteur et en constituent l'exemple le plus suggestif. Bien que leur nature soit essentiellement géométrique, c'est leur aptitude à se lier les unes aux autres, donc leur comportement algébrique, qui retiendra principalement notre attention. Partagé en classes d'équivalence l'ensemble qu'elles forment représente le modèle classique d'un "espace vectoriel". Un de nos premiers objectifs est la description détaillée de ce modèle.

Remarque Avant de lire ce qui va suivre, nous conseillons au lecteur d'avoir au moins parcouru en diagonale le chapitre traitant de la théorie des ensembles dans la section d'arithmétique. Nous y définissons ce qu'est un "espace vectoriel" en utilisant les outils de la théorie des ensembles. Ce concept bien que non absolument indispensable vaut la peine quand même de s'y attarder pour voir comment deux domaines des mathématiques s'imbriquent et aussi histoire... d'aborder les choses au moins un peu rigoureusement.

Remarque L'analyse vectorielle contient beaucoup de termes et de définitions qu'il faut apprendre par cœur. Ce travail est pénible mais malheureusement nécessaire.

15.1 Notion de flèche

Nous désignerons par U l'espace ordinaire de la géométrie élémentaire et par $P, Q, \ldots$ ses points. Nous appellerons "flèche" tout segment de droite orienté (dans l'espace).

La flèche d'origine P et d'extrémité Q sera notée $\vec{PQ}$ ou abrégé par une lettre unique (latine ou grecque) choisie arbitrairement tel que par exemple : $\vec{F}$.

Nous considérerons comme évident que toute flèche est caractérisée par sa direction, son sens, son intensité ou grandeur et ainsi que son origine.

15.2 Ensemble des vecteurs

Définition 15.1 *Nous disons que deux flèches sont "équivalentes" si elles ont la même direction, le même sens et la même intensité.*

Définition 15.2 *Nous disons que deux flèches sont "colinéaires" si elles ont seulement la même direction*

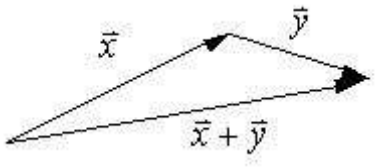


FIGURE 15.1 – Addition vectorielle

Partageons l'ensemble des flèches en classes d'équivalences : deux flèches appartiennent à une même classe si et seulement si elles sont équivalentes.

Définition 15.3 *Chaque classe d'équivalence de flèches constitue un "vecteur".*

Rangeons, en outre, les flèches dégénérées (c'est-à-dire de la forme $\vec{PP}$) en une classe distinguée que nous appellerons "vecteur nul" et noterons $\vec{0}$.

L'ensemble des vecteurs sera lui désigné par V. Il faut souligner que les éléments de V sont des classes de flèches et non pas des flèches individuelles. Il est cependant clair qu'une flèche quelconque suffit cependant à déterminer la classe à laquelle elle appartient et il est donc naturel de l'appeler "représentant de la classe" du vecteur.

Traçons le représentant d'un vecteur $\vec{y}$ à partir de l'extrémité d'un représentant d'un vecteur $\vec{x}$. La flèche dont l'origine est celle du représentant de $\vec{x}$ et l'extrémité celle du représentant de $\vec{y}$ détermine un vecteur que nous noterons $\vec{x}+\vec{y}$. L'opération qui associe à tout couple de vecteurs leur somme s'appelle "addition vectorielle". À l'aide de la figure 15.1, il est facile de montrer que l'opération d'addition vectorielle est associative et commutative, autrement dit, que :

$$(\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z}) \tag{15.1}$$

et :

$$\vec{x}+\vec{y}=\vec{y}+\vec{x} \tag{15.2}$$

Il est en outre évident que le vecteur nul est l'élément neutre de l'addition vectorielle, autrement dit, que :

$$\vec{x}+\vec{0}=\vec{x} \quad \text{et} \quad \vec{x}+(-\vec{x})=\vec{0} \tag{15.3}$$

où $-\vec{x}$ désigne le vecteur opposé de $\vec{x}$, c'est-à-dire le vecteur dont les représentants ont la même direction et la même intensité que ceux de $\vec{x}$, mais le sens opposé.

L'opération inverse de l'addition vectorielle est la soustraction vectorielle. Soustraire un vecteur revient à additionner le vecteur opposé.

REMARQUE L'addition s'étend, par récurrence, au cas d'une famille finie quelconque de vecteurs. En vertu de l'associativité, ces additions successives peuvent être effectuées dans n'importe quel ordre, ce qui justifie l'écriture sans parenthèses.

REMARQUE La multiplication entre deux vecteurs est un concept qui n'existe pas. Par contre, comme nous le verrons un peu plus loin, nous pouvons multiplier les vecteurs par certaines propriétés d'autres vecteurs que nous appelons la "norme" et encore d'autres petites choses...

15.2.1 Pseudo-vecteurs

En physique, lors de l'énoncé de ce que nous appelons le "principe de Curie", les physiciens font mention de ce qu'ils appellent des "pseudo-vecteurs". Il s'agit du vocabulaire simple pour parler de quelque chose de tout aussi trivial mais fondamentalement peu de gens en font vraiment usage. Mais il peut quand même être utile de présenter de quoi il s'agit.

Au fait, vecteurs et pseudo-vecteurs se transforment de la même manière dans une rotation ou une translation (nous verrons plus tard dans ce chapitre comment effectuer mathématiquement ces transformations). Il n'en est pas de même dans la symétrie par rapport à un plan ou à un point. Dans ces transformations nous avons par définition les propriétés suivantes :

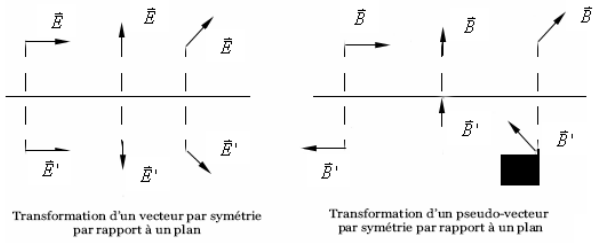


FIGURE 15.2 – Addition vectorielle

P1. Un vecteur est transformé en son symétrique,
P2. Un pseudo-vecteur est transformé en l'opposé du symétrique.
La figure **??** représente des exemples types (le choix des lettre représentant les vecteurs ne sont pas du au hasard, elle sont un clin d'oeil aux propriétés des champs électriques et magnétique étudiés en physique) :

15.2.2 Multiplication par un scalaire

Le vecteur $\alpha \cdot \vec{x}$ appelé "produit du nombre α par $\vec{x}$", est défini de la manière suivante :
Prenons une flèche représentative de $\vec{x}$ et construisons un flèche de même direction, de même sens ou de sens opposé, suivant que α est positif ou négatif, et d'intensité $|\alpha|$ fois l'intensité de la flèche initiale ; la flèche ainsi obtenue est un représentant du vecteur $\alpha \cdot \overrightarrow{x}$; si $\alpha = 0$ ou $\vec{x} = \vec{0}$, nous posons $\alpha \cdot \overrightarrow{x} = \vec{0}$.
L'opération qui consiste à effectuer le produit d'un nombre par un vecteur est appelé "multiplication par un scalaire".
Nous vérifions aisément que la multiplication par un scalaire est associative et distributive par rapport à l'addition numérique vectorielle, autrement dit que :

$$\begin{aligned} \alpha \cdot (\beta \cdot \vec{x}) &= (\alpha \cdot \beta) \cdot \vec{x} \\ (\alpha + \beta) \cdot \vec{x} &= \alpha \cdot \vec{x} + \beta \cdot \overrightarrow{x} \\ \alpha \cdot (\vec{x} + \vec{y}) &= \alpha \cdot \vec{x} + \alpha \cdot \overrightarrow{y} \end{aligned} \tag{15.4}$$

Voyons de suite un exemple concret mondialement connu des vecteurs :

Règle de trois

Revenons un peu sur la "règle de trois" (appelée parfois "règles des rapports et proportions" ou encore "méthode de réduction à l'unité") souvent définie dans les petites classes de manière intuitive mais sans démonstration digne de ce nom. Cette règle est certainement l'algorithme le plus usité de par le monde qui sert à identifier un quatrième nombre quand trois sont donnés et que les quatre nombres sont linéairement dépendants.
La règle de trois est est dérivée sous deux versions :
V1. Simple et directe si les grandeurs sont directement proportionnelles
V2. Simple et inverse si les grandeurs sont inversement proportionnelles
et lorsque deux variables X et Y sont proportionnelles nous le notons :

$$X \propto Y \tag{15.5}$$

Supposons maintenant que X puisse prendre les valeurs x_1, x_2 et Y prendra les valeurs linéairement dépendantes y_1, y_2 alors :
- Le rapport proportionnel suivant :

$$\boxed{\frac{x_1}{x_2} = \frac{y_1}{y_2} = c^{te}} \tag{15.6}$$

est dit "rapport simple et directe".

Preuve Soient deux vecteurs $\vec{X} = (x_1, x_2), \vec{Y} = (y_1, y_2)$ colinéaires et donc proportionnels à un facteur près tel que :

$$\vec{X} = \lambda\vec{Y} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \Leftrightarrow \lambda = \frac{x_1}{y_1} = \frac{x_2}{y_2} \Leftrightarrow \frac{y_1}{y_2} = \frac{x_1}{x_2} \tag{15.7}$$

•

REMARQUE Si ce rapport n'est pas égal, alors il faut passer à l'utilisation d'autres outils tel que la régression et in extenso l'extrapolation.

- Le rapport proportionnel suivant :

$$\boxed{\frac{x_1}{x_2} = \frac{y_2}{y_1} = c^{te}} \tag{15.8}$$

est dit "rapport simple et inverse".

Preuve Soient deux vecteurs $\vec{X} = (x_1, x_2), \vec{Y} = (y_1, y_2)$ colinéaires et donc proportionnels à un facteur près tel que :

$$\vec{X} = \lambda\vec{Y} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} \frac{1}{y_1} \\ \frac{1}{y_2} \end{pmatrix} \Leftrightarrow \lambda = x_1 y_1 = x_2 y_2 \Leftrightarrow \frac{x_1}{x_2} = \frac{y_2}{y_1} \tag{15.9}$$

•

REMARQUE Si ce rapport n'est pas égal, alors il faut passer à la régression linéaire (cf. chapitre de Méthodes Numériques).

En gros, il suffit que nous connaissions trois variables sur les quatre pour résoudre cette simple équation du premier degré.

Les conversions de monnaies ou d'unités de mesure se font à l'aide de la règle de trois simple directe ou indirecte. Les calculs de parités (calcul prévisionnel fait par un importateur d'un certain pays ayant ses propres unités de mesures et de monnaie, qui recherche parmi plusieurs offres étrangères (dans des systèmes d'unités de mesures et de monnaies qui diffèrent de l'importateur), laquelle est la plus avantageuse ou inversement) se font également avec la règle de trois.

REMARQUE Nous appelons également "règle conjointe simple ou inverse", une série de règle de trois directes ou indirectes.

Dans de tels calculs, les agents du marché d'échange ont remarqué que la plupart du temps, les rapports étaient des valeurs proche de l'unité. Ils ont été ainsi naturellement amené à définir "le pourcentage" comme étant la proportion d'une quantité ou d'une grandeur par rapport à une autre, évaluée à la centaine (en général du moins ...) :

Soit un nombre $x \in \mathbb{R}$ alors sa notation en pourcentage sera :

$$x'\% = x \cdot 100 \tag{15.10}$$

Soit un nombre $x \in \mathbb{R}$ alors sa notation en pour-mille sera

$$x'^0/_{00} = x \cdot 1000 \tag{15.11}$$

15.3 Espaces vectoriels

Définition 15.4 *Nous appelons "espace vectoriel" un ensemble E d'éléments désignés par $\vec{x}, \vec{y}, \cdots$ et appelés "vecteurs", muni d'une "structure algébrique vectorielle" définie par la donnée de l'addition (soustraction) vectorielle et la multiplication par un scalaire. Ces deux opérations satisfaisant les lois associativité, de commutativité, de distributivité, d'élément neutre et d'opposé comme nous l'avons déjà vu dans le chapitre de théorie des ensembles.*

Pour plus d'informations sur ce qu'est un espace vectoriel dans le sens "ensembliste" le lecteur devra se reporter au chapitre de théorie des ensembles où ce concept est défini avec plus de rigueur.

REMARQUE Muni de ces deux opérations, un espace vectoriel est dit "vectorialisé".

Pour tout entier positif n, $\mathbb{R}^n$ désignera l'ensemble des n-uplets de nombres disposés en colonne :

$$\begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ \cdots \\ a_n \end{pmatrix} \tag{15.12}$$

et $\mathbb{R}^n$ est à l'évidence muni d'une structure d'espace vectoriel. Les vecteurs de cet espace seront appelés "vecteurs-colonne". Il seront souvent désignés plus brièvement par :

$$(a_i)_{1\leq i\leq n} \tag{15.13}$$

ou encore plus simplement par :

$$(a_i) \tag{15.14}$$

Le nombre est parfois appelé "terme" ou "composante d'indice i" de (a_i).

15.3.1 Combinaisons linéaires

Dorénavant, sauf mention explicite du contraire, les vecteurs seront les éléments d'un espace vectoriel E.

Définition 15.5 *Nous appelons "combinaison linéaire" des vecteurs $\vec{x_1}, \vec{x_2}, \cdots, \vec{x_k}$ tout vecteur de la forme :*

$$\alpha_1\vec{x_1} + \alpha_2\vec{x_2} + \cdots + \alpha_k\vec{x_k} \tag{15.15}$$

ou $\alpha_1, \alpha_2, \cdots, \alpha_k$ sont des nombres appelés "coefficients de la combinaison linéaire".

Le vecteur nul est combinaison linéaire de $\alpha_1\vec{x_1} + \alpha_2\vec{x_2} + \cdots + \alpha_k\vec{x_k}$ avec tous les coefficients égaux à zéro. Nous parlons dès lors de "combinaison linéaire triviale".

Définition 15.6 *Nous appelons "combinaison convexe", toute combinaison linéaire dont les coefficients sont non négatifs et de somme égale à 1. L'ensemble des combinaisons convexes de deux points P et Q d'un espace ponctuel P_0 (ayant un origine) est le segment de droite P et Q. Pour s'en rendre compte, il suffit d'écrire :*

$$\alpha P + (1-\alpha)Q = Q + \alpha(P-Q) \tag{15.16}$$

de faire varier α de 0 à 1 et de constater que tous les points du segment sont ainsi obtenu.

Si le vecteur $\vec{x}$ est combinaison linéaire des vecteurs $\vec{x_1}, \vec{x_2}, \cdots \vec{x_k}$ et chacun de ces vecteurs est combinaison linéaire des vecteurs $\vec{y_1}, \vec{y_2}, \cdots \vec{y_k}$, alors $\vec{x}$ est combinaison linéaire de $\vec{y_1}, \vec{y_2}, \cdots \vec{y_k}$.

15.3.2 Sous-espaces vectoriels

Définition 15.7 *Nous appelons "sous-espace vectoriel de E" tout sous-ensemble de E qui est lui-même un espace vectoriel pour les opérations d'addition et de multiplication par un scalaire définies dans E.*

Un sous-espace vectoriel, en tant qu'espace vectoriel, ne peut être vide puisqu'il comprend au moins un vecteur, à savoir son vecteur nul, celui-ci étant d'ailleurs forcément le vecteur nul de E. En outre, en même temps que les vecteurs $\vec{x}$ et $\vec{y}$ (s'il en contient d'autres que le vecteur nul), il comprend également toutes leurs combinaisons linéaires $\alpha\vec{x} + \beta\vec{y}$.

Inversement, nous voyons aussitôt que tout sous-ensemble jouissant de ces propriétés est un sous-espace vectoriel. Nous avons ainsi établi la proposition suivante :

Un sous ensemble S de E est un sous-espace vectoriel de E si et seulement si S est non vide et $\alpha\vec{x} + \beta\vec{y}$ appartient à S pour tout couple $(\vec{x}, \vec{y})$ de vecteurs de S et tout couple $\alpha, \beta \in \mathbb{R}$.

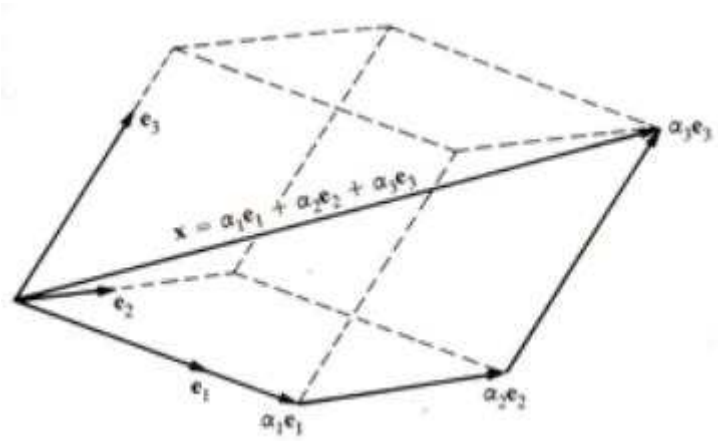


FIGURE 15.3 – Combinaison linéaire de trois vecteurs

15.3.3 Familles Génératrices

Il en découle que si nous avons une famille de vecteurs $(\vec{x_1}, \vec{x_2}, \cdots, \overrightarrow{x_k})$ l'ensemble des combinaisons linéaires de $\vec{x_1}, \vec{x_2}, \cdots, \vec{x_k}$ peut être un sous-espace vectoriel S de E, plus précisément le plus petit sous-espace vectoriel de E comprenant $\vec{x_1}, \vec{x_2}, \cdots, \overrightarrow{x_k}$.

Les vecteurs qui satisfont à la condition ci-dessus sont appelés "générateurs" de S et la famille $(\vec{x_1}, \vec{x_2}, \cdots, \vec{x_k})$, famille génératrice de S. Nous disons aussi que ces vecteurs ou cette famille engendrent S.

REMARQUE Le sous-espace vectoriel engendré par un vecteur non nul est formé de tous les multiples de ce vecteur. Nous appelons un tel sous-espace "droite vectorielle". Un sous-espace vectoriel engendré par deux vecteurs non multiples l'un de l'autre est appelé "plan vectoriel".

15.3.4 Dépendances et indépendances linéaires

Ce qui va suivre est très important en physique nous conseillons donc au futur physicien de prendre vraiment le temps de bien lire les développements qui vont suivre.

Si $\vec{e_1}, \vec{e_2}, \vec{e_3}$ sont trois vecteurs de V^3 dont les représentants ne sont pas parallèles à un même plan (par convention une flèche d'intensité nulle est parallèle à tout plan), alors tout vecteur $\vec{x}$ de V^3 peut s'écrire de manière unique sous la forme :

$$\vec{x} = \alpha_1 \vec{e_1} + \alpha_2 \vec{e_2} + \alpha_3 \vec{e_3} \tag{15.17}$$

où $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ sont des nombres. La figure 15.3 en donne la représentation graphique. En particulier, la seule possibilité d'obtenir le vecteur nul comme combinaison linéaire de $\vec{e_1}, \vec{e_2}, \vec{e_3}$ est d'attribuer la valeur triviale 0 à $\alpha_1, \alpha_2, \alpha_3$.

Réciproquement, si pour trois vecteurs $\vec{e_1}, \vec{e_2}, \vec{e_3}$ de V^3 la relation :

$$\vec{x} = \alpha_1 \vec{e_1} + \alpha_2 \vec{e_2} + \alpha_3 \vec{e_3} = 0 \tag{15.18}$$

implique $\alpha_1 = \alpha_2 = \alpha_3 = 0$, aucun des vecteurs ne peut être combinaison linéaire des deux autres, autrement dit, leurs représentants ne sont pas parallèles à un même plan.

Sur la base de ces observations, nous allons étendre la notion d'absence de parallélisme à un même plan au cas d'un nombre quelconque de vecteurs d'un espace vectoriel E.

Nous disons que les vecteurs $\vec{x_1}, \vec{x_2}, \cdots, \vec{x_k}$ sont "linéairement indépendants" si la relation :

$$\alpha_1 \vec{e_1} + \alpha_2 \vec{e_2} + \cdots + \alpha_k \vec{e_k} = 0 \tag{15.19}$$

implique nécessairement $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$, autrement dit, si la combinaison linéaire triviale est la seule combinaison linéaire de $\vec{x_1}, \vec{x_2}, \cdots, \overrightarrow{x_k}$ qui soit nulle. Dans le cas contraire, nous disons que les vecteurs $\vec{x_1}, \vec{x_2}, \cdots, \vec{x_k}$ sont "linéairement dépendants".

Si l'attention est fixée sur la famille $(\vec{x_1}, \vec{x_2}, \cdots, \overrightarrow{x_k})$ plutôt que sur les termes dont elle est constituée, nous disons que celle-ci est une "famille libre" ou "famille liée" suivant que les vecteurs $\vec{x_1}, \vec{x_2}, \cdots, \vec{x_k}$ sont linéairement indépendants ou dépendants.

15.3.5 Bases d'un espace vectoriel

Définition 15.8 *Nous disons qu'une famille finie de vecteurs est une base de E si et seulement si :*

1. *Elle est libre*
2. *Elle engendre E*

D'après cette définition, toute famille libre $\vec{x_1}, \vec{x_2}, \cdots, \vec{x_k}$ est une base du sous-espace vectoriel qu'elle engendre.

Exemple 15.1 *Si nous considérons $\mathbb{C}$ comme $\mathbb{R}$-espace vectoriel (cf. chapitre de Théorie des Ensembles), alors puisque tous les éléments de $\mathbb{C}$ s'écrivent $a+ib$, les éléments qui engendrent $\mathbb{C}$ sont 1 et i (les deux sont libres).*

Une base de $\mathbb{C}$ (qui est de dimension 2) comme $\mathbb{R}$-espace vectoriel est donc la famille finie libre $\{1, i\}$.

Pour qu'une famille de vecteurs $(\vec{e_1}, \vec{e_2}, \cdots, \vec{e_n})$ soit une base de E, il faut et il suffit donc que tout vecteur $\vec{x}$ de E s'exprime de manière unique sous la forme d'une combinaison linéaire des vecteurs $\vec{e_1}, \vec{e_2}, \cdots, \vec{e_n}$:

$$\vec{x} = x_1\vec{e_1} + x_2\vec{e_2} + \cdots + x_n\vec{e_n} \tag{15.20}$$

La relation ci-dessus est une décomposition de $\vec{x}$ suivant la base $(\vec{e_1}, \vec{e_2}, \cdots, \vec{e_n})$ où les coefficients $x_1, x_2, \cdots, x_n$ sont les composantes de $\vec{x}$ dans cette base. En présence d'une base, tout vecteur est donc entièrement déterminé par ses composantes.

Proposition :

Si $x_1, x_2, \cdots, x_n$ sont les composantes de $\vec{x}$ et $y_1, y_2, \cdots, y_n$ celles de $\vec{y}$, alors :

$$x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n \tag{15.21}$$

sont les composantes de $\vec{x} + \vec{y}$.

En d'autres termes, additionner deux vecteurs revient à additionner leurs composantes et multiplier un vecteur par un scalaire revient évidemment à multiplier ses composantes par ce même scalaire. La base est donc un outil important car elle permet d'effectuer les opérations sur les vecteurs au moyen d'opérations sur les nombres.

Exemple 15.2 *Les vecteurs colonnes de $\mathbb{R}^n$:*

$$\begin{pmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ . \\ . \\ . \\ 0 \end{pmatrix}, \cdots \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 1 \end{pmatrix} \tag{15.22}$$

forment un base que nous appelons "base canonique" de $\mathbb{R}^n$ (nous travaillerons dans les espaces complexes dans un autre chapitre).

Remarque Dans le cadre de l'espace à trois dimensions, les bases sont très souvent assimilées à un trièdre (effectivement si vous reliez les extrémités des trois vecteurs par des traits vous obtiendrez une trièdre imaginaire).

Angles directeurs

Il est clair qu'un seul angle ne peut décrire la direction d'un vecteur dans l'espace. Nous utilisons alors la notion "d'angles directeurs". Il s'agit de mesurer l'angle du vecteur $\vec{U}$ par rapport à chacun des axes positifs de la base comme présenter à la figure 15.4.

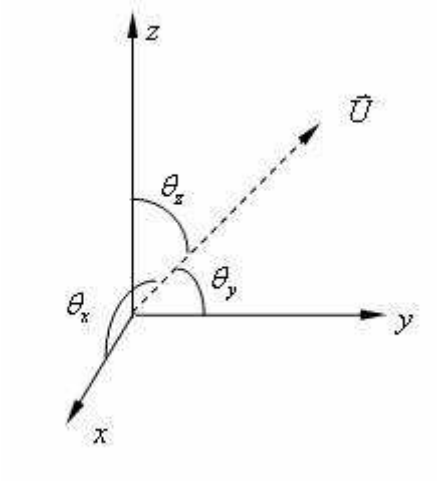


FIGURE 15.4 – Combinaison linéaire de trois vecteurs

Si :

$$\vec{U} = (u_1, u_2, u_3) = u_1 \cdot \vec{i} + u_2 \cdot \vec{j} + u_3 \cdot \vec{k} \tag{15.23}$$

alors :

$$\begin{aligned} u_1 &= ||\vec{U}|| \cos(\theta_x) \\ u_2 &= ||\vec{U}|| \cos(\theta_y) \\ u_3 &= ||\vec{U}|| \cos(\theta_z) \end{aligned} \tag{15.24}$$

Les valeurs :

$$\cos(\theta_x), \cos(\theta_y), \cos(\theta_z) \tag{15.25}$$

sont appelées les "cosinus directeurs" de $\vec{U}$.

Les 3 angles mentionnées ne sont pas complètement indépendants. En effet, 2 suffisent pour déterminer complètement la direction d'un vecteur dans l'espace, le troisième pouvant se déduire la relation suivante (obtenue à partir du calcul de la norme du vecteur, concept que nous verrons un peu plus loin) :

$$\cos^2(\theta_x) + \cos^2(\theta_y) + \cos^2(\theta_z) = 1 \tag{15.26}$$

De plus, les cosinus directeurs sont les composantes scalaires d'un vecteur de norme unitaire ayant la même direction que $\vec{U}$:

$$\frac{\vec{U}}{||\vec{U}||} = \vec{u} = \big(\cos(\theta_x), \cos(\theta_y), \cos(\theta_z)\big) \tag{15.27}$$

15.3.6 Dimension d'un espace vectoriel

Nous disons que E est de "dimension finie" s'il est engendré par une famille finie de vecteurs. Dans le cas contraire, nous disons que E est de "dimension infinie" (nous aborderons ce type d'espaces dans un autre chapitre). Tout espace vectoriel de dimension finie et non réduit au vecteur

nul admet une base. En fait, de tout famille génératrice d'un tel espace nous pouvons extraire une base.

La dimension d'un espace vectoriel est notée :

$$\dim(E) \tag{15.28}$$

Tout espace vectoriel E de dimension finie non nulle n peut être mis en correspondance biunivoque (c'est-à-dire en bijection) avec $\mathbb{R}^n$. Il suffit de choisir une base de E et de faire correspondre à tout vecteur $\vec{x}$ de E le vecteur-colonne dont les termes sont les composantes de $\vec{x}$ dans la base choisie (c'est du bla bla de mathématicien mais ce sera utile quand nous aborderons des espaces plus complexes) :

$$\begin{array}{ccc} E & & \mathbb{R}^n \\ \vec{x}(x_1, x_2, \cdots, x_3) & \leftrightarrow & \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} \end{array}$$

Cette correspondance conserve les opérations d'addition et de multiplication par un scalaire que nous avons déjà vues ; en d'autres termes, elle permet d'effectuer les opérations sur les vecteurs par des opérations sur les nombres.

Remarque Nous disons alors que E et $\mathbb{R}^n$ sont "isomorphes" ou que la correspondance est un isomorphisme (cf. chapitre de Théorie Des Ensembles).

15.3.7 prolongement d'une faille libre

Soit $(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k)$ une famille libre et $(\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m)$ une famille génératrice de E. Si $(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k)$ n'est pas une base de E, nous pouvons extraire une sous-famille $(\vec{v}_{i1}, \vec{v}_{i2}, \cdots, \overrightarrow{v}_{il})$ de $(\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m)$ de telle manière que la famille $(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k, \overrightarrow{v}_{i1}, \vec{v}_{i2}, \cdots, \vec{v}_{il})$ soit une base de E.

Remarque Une telle étude a son utilité lors de passage d'espace mathématique ayant des propriétés données à un autre espace ayant des propriétés mathématiques différentes.

Preuve H1. Nous supposons qu'au moins un des vecteurs $\vec{v}_i$ n'est pas combinaison linéaire des vecteurs $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k$, sinon $(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k)$ engendrerait E et serait donc une base possible de E. Notons ce vecteur $\vec{v}_{i1}$. La famille $(\vec{x}_1, \overrightarrow{x}_2, \cdots, \vec{x}_k, \vec{v}_{i1})$ est alors une famille libre. En effet, la relation :

$$\alpha_1\vec{x}_1 + \cdots + \alpha_k\vec{x}_k + \beta_1\vec{v}_{i1} = \vec{0} \tag{15.29}$$

implique alors tout d'abord que $\beta_1 = 0$, autrement $\vec{v}_{i1}$ serait combinaison linéaire des vecteurs $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k$, et ensuite $\alpha_1 = \cdots = \alpha_k = 0$, puisque les vecteurs $\vec{x}_1, \vec{x}_2, \cdots, \overrightarrow{x}_k$ sont linéairement indépendants.

Si la famille $(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k, \overrightarrow{v}_{i1})$ engendre E, elle est une base possible de E et le théorème est démontré. Dans le cas contraire, le même raisonnement nous assure l'existence d'un autre vecteur $\vec{v}_{i2}$... Si la nouvelle famille en découlant n'est pas un base de E, alors le procédé d'extraction de vecteurs $\vec{v}_i$ de $(\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m)$ se poursuit. Lorsqu'il s'arrête, nous aurons obtenu un "prolongement" de $(\vec{x}_1, \overrightarrow{x}_2, \cdots, \vec{x}_k)$ en une famille libre engendrant E, c'est-à-dire une base de E. •

Il en retourne un corollaire : Tout espace vectoriel de dimension finie et non réduit au vecteur nul admet une base. En fait, de toute famille génératrice d'un tel espace, nous pouvons donc extraire une base.

15.3.8 Rang d'une famille finie

Définition 15.9 *Nous appelons "rang d'une famille" de vecteurs la dimension du sous-espace vectoriel de E qu'elle engendre.*

Montrons que le rang d'une famille de vecteurs $(\vec{x}_1, \vec{x}_2, \cdots, \overrightarrow{x}_k)$ est inférieur ou égal à k et qu'il est égal à k si et seulement si cette famille est libre.

Preuve Ecartons d'abord le cas trivial où le rang de la famille $(\vec{x}_1, \vec{x}_2, \cdots, \overrightarrow{x}_k)$ est nul. D'après le corollaire vu précédemment, nous pouvons alors extraire de cette famille une base du sous-espace vectoriel qu'elle engendre. Le rang est donc inférieur ou égal à k suivant que $(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k)$ est une famille liée ou non.

15.3.9 Sommes directes

Définition 15.10 *Nous disons que la somme $S+T$ de deux sous-espaces vectoriels S et T de E (cas particulier appliqué à un espace de dimensions 2 !) est une "somme directe" si :*

$$S \cap T = \{\vec{0}\} \tag{15.30}$$

Dans ce cas, nous la notons :

$$S \oplus T \tag{15.31}$$

En d'autres termes, la somme de deux sous-espaces vectoriels S et T de E est directe si la décomposition de tout élément de $S+T$ en somme d'un élément de S et d'un élément de T est unique.

Ce concept de décomposition trivial va nous être très utile dans certains théorèmes dont le plus important sur ce site est certainement le théorème spectral (cf. chapitre d'Algèbre Linéaire).

De la somme directe nous pouvons introduire la notion de "sous-espace complémentaire" appelé encore "sous-espace supplémentaire" (selon les pays...) :

Supposons que E soit de dimension finie. Pour tout sous-espace vectoriel S de E, il existe un sous-espace vectoriel T (non unique) de E tel que E soit somme directe de S et T. Nous disons alors que T est un "sous-espace complémentaire" de S dans E.

Preuve Ecartons d'abord les cas triviaux où $S = \{\vec{0}\}$et $S = E$. Le sous-espace vectoriel S admet une base $(\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_k)$, où k est inférieur à la dimension n de E. Par le théorème du prolongement d'une famille libre, cette base peut se prolonger en une base $(\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_k, \overrightarrow{e}_{k+1}, \cdots, \vec{e}_n)$ de E. Soit T le sous-espace vectoriel engendré par la famille $(\vec{e}_{k+1}, \cdots, \vec{e}_n)$. Si $\vec{x}$ est un vecteur quelconque de E, alors $\vec{x} = \vec{s} + \vec{t}$, où $\vec{s}$ est un vecteur de S et $\vec{t}$ un vecteur de T. En outre, $S \cap T = \{\vec{0}\}$ car aucun vecteur, excepté le vecteur nul, ne peut être combinaison linéaire des vecteurs $\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_k$ et des vecteurs $\vec{e}_{k+1}, \cdots, \vec{e}_n$. Nous en concluons donc que :

$$E = S \oplus T \tag{15.32}$$

•

15.4 Espace Affine

L'espace G de la géométrie élémentaire est à la fois usuel et la source de la notion "d'espace affine" que nous allons introduire.

Cet espace G est associé à "l'espace vectoriel" géométrique V par la correspondance entre flèches et vecteurs étudiés jusqu'ici ! La définition suivante ne fait que mettre en évidence les traits dominants de cette correspondance :

Définition 15.11 *Soit U un ensemble non vide d'éléments que nous appellerons "points" et désignerons par les lettres $P, Q, \cdots$; soit en outre E un espace vectoriel. Supposons qu'à tout couple de points (P, Q) corresponde un vecteur noté $\vec{PQ}$. Nous disons alors que que U est un "espace affine" d'espace directeur (ou dit simplement abusivement de "direction") E si les conditions suivantes sont satisfaites :*

C1 *Pour tout point P fixé, la correspondance entre couples (P, Q) et vecteurs $\vec{x}$ est biunivoque, autrement dit, pour tout vecteur $\vec{x}$ il existe un point Q et un seul tel que $\vec{x} = \vec{PQ}$.*

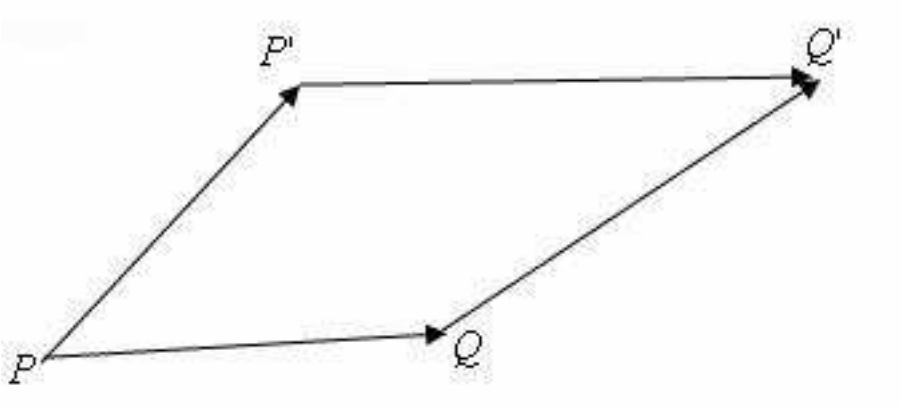


FIGURE 15.5 – Addition vectorielle

C2 *Pour tout triplet de points* (P, Q, R) *:*

$$\vec{PQ} + \vec{QR} = \vec{PR} \tag{15.33}$$

C'est la fameuse "relation de Chasles"

C3 *Si* P *est un point et* $\vec{x}$ *un vecteur, pour exprimer que* Q *est l'unique point tel que* $\vec{x} = \vec{PQ}$*, nous écrivons :*

$$Q = P + \vec{x} \tag{15.34}$$

Bien qu'un peu abusive, cette écriture est comme à l'usage et suggère bien le sens de l'opération qu'elle désigne.

Les propriétés suivantes découlent directement de la définition d'espace affine :

P1 $P + (\vec{x} + \vec{y}) = (P + \vec{x}) + \vec{y}$

P2 Pour tout point P, $\vec{PP} = \vec{0}$. Cela résulte de la condition $\vec{PQ} + \vec{QR} = \vec{PR}$ dans le cas où nous avons $P = Q = R$.

P3 $\vec{PQ} = -\vec{QP}$ Il suffit de poser $R = P$ dans la relation de Chasles $\vec{PQ} + \vec{QR} = \vec{PR}$.

P4 Règle du parallélogramme :

Soit un parallélogramme de la figure 15.5 de sommets P, P', Q', Q (dans le sens des aiguilles d'un montre) et arêtes $\vec{PP'}, \vec{P'Q'}, \vec{QQ'}, \vec{PQ}$.

Nous avons :

$$\vec{PP'} = \vec{QQ'} \tag{15.35}$$

si et seulement si :

$$\vec{PQ'} = \vec{PQ} \tag{15.36}$$

En effet, en remplaçant R par Q' dans la relation de Chasles il vient :

$$\vec{PQ} + \vec{QQ'} = \vec{PQ'} \tag{15.37}$$

et en faisant de même mais en remplaçant R par Q' et Q par P' nous avons :

$$\vec{PP'} + \vec{P'Q'} = \vec{PQ'} \tag{15.38}$$

Nous avons alors en égalisant ces deux dernières relations :

$$\vec{PQ} + \vec{QQ'} = \vec{PP'} + \vec{P'Q'} \tag{15.39}$$

ce qui force l'égalité susmentionnée que nous voulions démontrer.

Précédemment, nous avons vu ce qui faisait qu'un espace G pouvait être muni d'une structure d'espace vectoriel (nous avons vu que nous disons que ce dernier était dès lors "vectorialisé"). Dans le cas général d'un espace affine U, le procédé est le même :

Nous choisissons un point quelconque O de U. La correspondance entre couples (O, P) et vecteurs de l'espace directeur étant alors biunivoque nous définissons l'addition de points et la multiplication d'un point par un scalaire par les opérations correspondantes sur les vecteurs de E. Muni de ces deux opérations, U devient un espace vectoriel, appelé "vectorialisé de U relativement à O". Nous désignerons cet espace par U_0 et appellerons O "l'origine".

Vu la manière dont les opérations ont été définies, il résulte que U_0 est isomorphe à l'espace directeur E :

$$P = \overbrace{0}^{U_0} + \vec{x} \leftrightarrow \overbrace{\vec{x}}^{E} \tag{15.40}$$

Toutefois, cet isomorphisme dépend du choix de l'origine O et en pratique cette origine est choisie sur la base de données inhérentes aux problèmes posés. Par exemple, si une transformation affine admet un point invariant (qui ne bouge pas), il y a avantage a choisir ce point comme origine.

REMARQUE Lorsque nous parlons de dimension d'un espace affine, nous parlons de la dimension de son espace directeur.

REMARQUE L'espace G de la géométrie élémentaire est un espace affine. En effet, sa direction est l'espace géométrique V et les conditions de définition d'un espace affine sont satisfaites. Il faut bien noter qu'au couple de points (P, Q) est associé le vecteur $\vec{PQ}$ et non pas la flèche PQ. En fait, la flèche pouvant être identifiée au couple de points, nous voyons que ce que postule la définition d'un espace affine n'est rien d'autre qu'une forme abstraite de correspondance entre flèches et vecteurs.

REMARQUE Tout espace vectoriel E peut être considéré comme un espace affine de direction E lui-même si au couple de vecteurs $(\vec{x}, \vec{y})$ est associé le vecteur $\vec{y} - \vec{x}$. En effet, les conditions de définition d'un espace affine sont dès lors satisfaites.

15.5 Espaces vectoriels euclidiens

Avant de définir ce qu'est un espace vectoriel euclidien, nous allons au préalable définir quelques outils mathématiques et quelques concepts. Nous pouvons, en choisissant une unité de longueur, mesurer l'intensité de chaque flèche, autrement dit, déterminer sa longueur. Nous pouvons aussi mesurer l'écart angulaire de deux flèches (ou vecteurs) quelconques d'origine commune (non nécessairement distinctes) en prenant comme unité de mesure d'angle par exemple le radian. La mesure de cet écart est alors un nombre compris entre 0 et π, appelé "angle" des deux flèches. Si les deux flèches ont même direction et même sens, leur angle est nul et si elles ont même direction et sens opposé, ce même angle est π.

15.5.1 Norme d'un vecteur

Les flèches représentatives d'un même vecteur $\vec{x}$ ont toutes la même longueur. Nous désignerons cette longueur par la notation :

$$||\vec{x}|| \tag{15.41}$$

et l'appellerons "norme" de $\vec{x}$. Il est clair que la longueur d'un vecteur est nulle si et seulement si sa norme est nulle. Nous dirons qu'un vecteur est unitaire si sa norme est 1.

Si $\vec{x}$ est un vecteur non nul :

$$\frac{1}{||\vec{x}||}\vec{x} \tag{15.42}$$

est un vecteur unitaire colinéaire (nécessairement...) à $\vec{x}$ dont la norme est égale à l'unité et que nous notons $\vec{u}$.

Nous appellerons "angle des vecteurs non nuls" $\vec{x}$ et $\vec{y}$ l'angle de deux flèches d'origine commune représentant l'une $\vec{x}$ et l'autre $\vec{y}$.

Plus rigoureusement cependant une "norme" sur un espace vectoriel réel (ou complexe) E est une application $||\ ||: E \to \mathbb{R}^+$ vérifiant les propriétés :

P1. Positivité

$$\forall x \in E ||x|| \geq 0 \tag{15.43}$$

P2. Linéarité

$$\forall \lambda \in \mathbb{R}, \forall x \in E ||\lambda x|| = |\lambda| \cdot ||x|| \tag{15.44}$$

P3. Nullité

$$\forall x \in E ||x|| = 0 \Leftrightarrow x = 0 \tag{15.45}$$

P4. Inégalité de Minkowski (inégalité triangulaire)

$$\forall x, y \in E ||x + y|| \leq ||x|| + ||y|| \tag{15.46}$$

Remarque Ces propriétés sont principalement imposées par notre approche intuitive de l'espace euclidien et de son interprétation géométrique.

Remarque Nous démontrerons un peu plus loin la propriété P4 sous la dénomination "d'inégalité triangulaire" et nous ferons une étude un peu plus générale de cette inégalité sous la dénomination "d'inégalité de Minkowski" dans le chapitre de topologie.

15.5.2 Produit scalaire vectoriel

Définition 15.12 *Un "espace vectoriel euclidien", est un espace vectoriel (réel et de dimension finie pour les puristes) possédant une opération particulière, le "produit scalaire" que nous noterons (notation spécifique à ce site Internet) en ce qui concerne le cas particulier des vecteurs :*

$$\boxed{||\vec{x}|| \cdot ||\vec{y}|| \cos\theta = \vec{x} \circ \overrightarrow{y}} \tag{15.47}$$

Remarque Nous trouvons dans certains ouvrages (pour information) la notation $(\vec{x}|\overrightarrow{y})$ ou encore $\langle \vec{x}|\vec{y} \rangle$ lors de la généralisation de cette définition comme nous le verrons un peu plus loin.

Remarque Le produit scalaire à une importance énorme dans l'ensemble du domaine des mathématiques et de la physique, ainsi nous le retrouvons dans le calcul différentiel et intégral (de par le produit scalaire fonctionnel), en topologie, en physique quantique en analyse du signal etc... il convient donc de bien comprendre ce qui va suivre.

Remarque Le produit scalaire peut être vu comme une projection de la longueur d'un vecteur sur la longueur d'un autre.

Ce produit scalaire possède les propriétés suivantes (dont la plupart découlent de la définition même du produit scalaire) dans un espace euclidien :

P1. Commutativité $\vec{x} \circ \vec{y} = \vec{y} \circ \overrightarrow{x}$

P2. Associativité $\alpha(\vec{x} \circ \vec{y} = (\alpha \overrightarrow{x}) \circ \vec{y}) = \vec{x} \circ (\alpha \vec{y})$

P3. Distributivité $\vec{x} \circ (\vec{y} + \vec{z}) = \overrightarrow{x} \circ \vec{y} + \vec{x} \circ \vec{z}$

P4 Si $\vec{x} \circ \vec{y} = 0 : \forall \vec{x} \neq \overrightarrow{0} \Rightarrow \vec{y} = \vec{0}$

P5 $||\vec{x}|| = \vec{x} \circ \vec{x}$ et $\vec{x} \circ \vec{x} > 0$ si $\vec{x} = \vec{0}$

P6 $(\alpha\vec{x} + \beta\vec{y}) \circ \vec{z} = \alpha(\vec{x} \circ \vec{z}) + \beta(\vec{y} \circ \overrightarrow{z})$

Seule cette dernière propriété nécessite peut-être une démonstration (de plus un des résultats de la démonstration nous servira à démontrer une autre propriété très importante du produit scalaire) :

Preuve Soit :

$$proj_v\vec{x} = ||\vec{x}|| \cos\theta \left(\frac{1}{||\vec{v}||}\vec{v}\right) \tag{15.48}$$

qui constitue la "projection orthogonale vectorielle" du vecteur $\vec{x}$ sur la normalisaiton à l'unité du vecteur $\vec{v}$.

A l'aide du produit scalaire, le vecteur $proj_v\vec{x}$ peut être exprimé autrement il suffit de de prendre :

$$||\vec{x}||||\vec{y}|| \cos\theta = \vec{x} \circ \vec{y} \Leftrightarrow \cos\theta = \frac{\vec{x} \circ \vec{y}}{||\vec{x}||||\vec{y}||} \tag{15.49}$$

et de l'introduire dans $proj_v\vec{x}$ avec les vecteur adéquats pour obtenir :

$$proj_v\vec{x} = ||\vec{x}|| \cos\theta \left(\frac{1}{||\vec{v}||}\vec{v}\right) = ||\vec{x}|| \left(\frac{\vec{x} \circ \vec{y}}{||\vec{x}||||\vec{y}||}\right)\left(\frac{1}{||\vec{v}||}\vec{v}\right) = \frac{\vec{x} \circ \vec{v}}{||\vec{v}||^2}\vec{v} = \frac{\vec{x} \circ \vec{v}}{\vec{v} \circ \vec{v}}\vec{v} \tag{15.50}$$

Cette expressions vaut également dans le cas où $\vec{x}$ est nul, à condition d'admettre que la projection orthogonale de vecteur nul est nul. La norme de $proj_v\vec{x}$ s'écrit :

$$||proj_v\vec{x}|| = \frac{|\vec{x} \circ \vec{v}|}{||\vec{v}||^2}||\vec{v}|| = \frac{|\vec{x} \circ \overrightarrow{v}|}{||\vec{v}||} \tag{15.51}$$

Si $\vec{v}$ est unitaire, les relations de projections précédentes se simplifient et deviennent :

$$proj_v\vec{x} = (\vec{x} \circ \vec{v})\vec{v} \text{ et } ||proj_v\vec{x}|| = |\vec{x} \circ \vec{v}| \tag{15.52}$$

Par des considérations géométriques élémentaires, il est facile de se rendre compte que :

$$proj_v(\vec{x} + \vec{y}) = proj_v\vec{x} + proj_v\vec{y} \text{ et } proj_v\alpha\vec{x} = \alpha proj_v\vec{x} \tag{15.53}$$

Si nous revenons maintenant à la démonstration de :

$$(\alpha\vec{x} + \beta\vec{y}) \circ \vec{z} = \alpha(\vec{x} \circ \vec{z}) + \beta(\vec{y} \circ \vec{z}) \tag{15.54}$$

Nous avons si $\vec{z} \neq \vec{0}$:

$$proj_v(\alpha\vec{x} + \beta\vec{y}) = proj_v\alpha\vec{x} + proj_v\beta\vec{y} = \alpha proj_v\vec{x} + \beta proj_v\vec{y} \tag{15.55}$$

d'après la définition la propriété de la projection orthogonale :

$$\begin{aligned} &proj_v(\alpha\vec{x} + \beta\vec{y}) = \alpha proj_v\vec{x} + \beta proj_v\vec{y} \\ &\frac{(\alpha\vec{x} + \beta\vec{y}) \circ \vec{z}}{\vec{z} \circ \overrightarrow{z}} = \alpha\frac{\vec{x} \circ \vec{z}}{\vec{z} \circ \overrightarrow{z}}\vec{z} + \beta\frac{\vec{y} \circ \overrightarrow{z}}{\vec{z} \circ \vec{z}}\vec{z} \end{aligned} \tag{15.56}$$

•

d'où la propriété P6 qui s'ensuit par multiplication des deux membres de l'égalité par $\vec{z} \circ \vec{z}$.

Définition 15.13 *L'espace vectoriel E est dit "espace vectoriel proprement euclidien" si $\forall\vec{x} \in E \rightarrow ||\vec{x}|| \geq 0$*

Définition 15.14 *Nous disons que les vecteurs $\vec{x}$ et $\vec{y}$ sont des "vecteurs orthogonaux" s'ils sont non nuls et que leur produit scalaire est nul (leur angle est égal à $\frac{\pi}{2}$).*

Une base $(\vec{e_1}, \vec{e_2}, \vec{e_3})$ de V est dite "base orthonomormale" si les vecteurs $\vec{e_1}, \vec{e_2}, \vec{e_3}$ sont orthogonaux deux à deux et unitaires (donc constituant une famille libre).

REMARQUE Nous verrons en calcul tensoriel (nous aurions pu le faire ici aussi mais bon...) comment à partir d'un ensemble de vecteurs indépendants construire une base orthogonale. C'est ce que le lecteur pourra trouver sous le nom de "méthode d'orthogonalisation de (Grahm-)Schmidt".

Par le raisonnement géométrique, nous voyons que tout vecteur est la somme de ses projections orthogonales sur les vecteurs d'une base orthonormale, autrement dit, si $(\vec{e_1}, \vec{e_2}, \vec{e_3})$ est une base orthonormale :

$$\vec{x} = (\vec{x} \circ \vec{e_1})\vec{e_1} + (\vec{x} \circ \vec{e_2})\vec{e_2} + (\vec{x} \circ \vec{e_3})\vec{e_3} \tag{15.57}$$

Cette décomposition s'obtient également par la propriété de P6 du produit scalaire. En effet, x_1, x_2, x_3 étant les composantes de $\vec{x}$:

$$\begin{aligned} \vec{x} \circ \vec{e_1} &= (x_1 \circ \vec{e_1} + x_2 \circ \vec{e_2} + x_3 \circ \vec{e_3}) \circ \vec{e_1} \\ &= x_1(\vec{e_1} \circ \vec{e_1}) + x_2(\vec{e_2} \circ \vec{e_1}) + x_3(\vec{e_3} \circ \vec{e_1}) \\ &= x_1 \end{aligned} \tag{15.58}$$

puisque $\vec{e_1} \circ \vec{e_1} = 1$ et $\vec{e_2} \circ \vec{e_1} = \overrightarrow{e_3} \circ \vec{e_1} = 0$ et de même :

$$\vec{x} \circ \vec{e_2} = x_2 \text{ et } \vec{x} \circ \vec{e_3} = x_3 \tag{15.59}$$

d'où la décomposition.

Soit x_1, x_2, x_3 et y_1, y_2, y_3 les composantes respectives des vecteurs $\vec{x}$ et $\vec{y}$ dans une base orthonormale canonique nous pouvons écrire le produit scalaire sous la forme :

$$\vec{x} \circ \vec{y} = (x_1\vec{e_1} + x_2\vec{e_2} + x_3\overrightarrow{e_3}) \circ (y_1\vec{e_1} + y_2\vec{e_2} + y_3\vec{e_3}) \tag{15.60}$$

de la propriété P6 du produit scalaire :

$$\begin{aligned} \vec{x} \circ \vec{y} =& x_1\vec{e_1} \circ (y_1\vec{e_1} + y_2\overrightarrow{e_2} + y_3\vec{e_3}) \\ &+ x_2\vec{e_2} \circ (y_1\vec{e_1} + y_2\vec{e_2} + y_3\overrightarrow{e_3}) \\ &+ x_3\vec{e_3} \circ (y_1\vec{e_1} + y_2\vec{e_2} + y_3\overrightarrow{e_3}) \end{aligned} \tag{15.61}$$

en utilisant la propriété P1 et à nouveau P6 :

$$\begin{aligned} \vec{x} \circ \vec{y} =& x_1(y_1\vec{e_1} + y_2\vec{e_2} + y_3\overrightarrow{e_3}) \circ \vec{e_1} \\ &+ x_2(y_1\vec{e_1} + y_2\vec{e_2} + y_3\vec{e_3}) \circ \vec{e_2} \\ &+ x_3(y_1\vec{e_1} + y_2\vec{e_2} + y_3\vec{e_3}) \circ \vec{e_3} \\ \Rightarrow& \\ (y_1\vec{e_1} + y_2\vec{e_2} + y_3\vec{e_3}) \circ \vec{e_1} =& y_1\vec{e_1} \circ \vec{e_1} + y_2\overrightarrow{e_2} \circ \vec{e_1} + y_3\vec{e_3} \circ \vec{e_1} = y_1||\vec{e_1}|| = y_1 \\ (y_1\vec{e_1} + y_2\vec{e_2} + y_3\vec{e_3}) \circ \vec{e_2} =& y_1\overrightarrow{e_1} \circ \vec{e_2} + y_2\vec{e_2} \circ \vec{e_2} + y_3\overrightarrow{e_3} \circ \vec{e_2} = y_2||\vec{e_2}|| = y_2 \\ (y_1\vec{e_1} + y_2\vec{e_2} + y_3\vec{e_3}) \circ \vec{e_3} =& y_1\overrightarrow{e_1} \circ \vec{e_3} + y_2\vec{e_2} \circ \vec{e_3} + y_3\overrightarrow{e_3} \circ \vec{e_3} = y_3||\vec{e_3}|| = y_3 \end{aligned} \tag{15.62}$$

Ce qui nous donne finalement la décomposition :

$$\boxed{\vec{x} \circ \vec{y} = x_1y_1 + x_2y_2 + x_3y_3} \tag{15.63}$$

qui constitue l'une des relations les plus importante dans le domaine du calcul vectoriel et que nous appelons "produit scalaire canonique".

Inégalité de Cauchy-Schwarz

La relation :

$$(\vec{x}\circ\vec{y})\circ(\vec{x}\circ\vec{y}) = \vec{x}\circ\vec{x}+\vec{x}\circ\vec{y}+\vec{y}\circ\vec{x}+\vec{y}\circ\vec{y} = \overrightarrow{x}\circ\vec{x}+\vec{y}\circ\vec{y}+2\vec{x}\circ\vec{y} \tag{15.64}$$

s'écrit également trivialement sous la forme suivante si nous utilisons la notion de norme et la définition du produit scalaire :

$$(\vec{x}\circ\vec{y})\circ(\vec{x}\circ\vec{y}) = ||\overrightarrow{x}+\vec{y}||^2 = ||\vec{x}||^2+||\vec{y}||^2+2\vec{x}\circ\vec{y} \tag{15.65}$$

Si les deux vecteurs $\vec{x}$ et $\vec{y}$ sont orthogonaux, nous retrouvons le résultat d'un théorème fameux : le théorème de Phytagore.

Effectivement, dès lors nous avons $\vec{x}\circ\vec{y}=0$. Ce qui nous donne :

$$\boxed{||\vec{x}+\vec{y}||^2 = ||\vec{x}||^2+||\overrightarrow{y}||^2} \tag{15.66}$$

Cette relation est très importante en physique-mathématique. Il faut s'en souvenir !

Nous appelons également "inégalité de Cauchy-Schwarz" l'inégalité, valable pour tout choix des vecteurs $\vec{x}$ et $\vec{y}$ la relation :

$$\boxed{|\vec{x}\circ\vec{y}| \leq ||\vec{x}||||\vec{y}||} \tag{15.67}$$

Ce qui s'écrit aussi :

$$|\vec{x}\circ\vec{y}|^2 \leq (\vec{x}\circ\vec{x})\cdot(\vec{y}\circ\vec{y}) \tag{15.68}$$

D'abord nous considérerons comme évident que l'égalité n'a lieu qu'en cas de colinéarité des deux vecteurs.

Preuve Nous nous plaçons dans le cas où $\vec{x},\vec{y}\neq\vec{0}$. Alors, pour $\forall\lambda\in\mathbb{R}$ nous avons trivialement selon les propriétés du produit scalaire :

$$(\vec{x}+\lambda\vec{y})\circ(\vec{x}+\lambda\vec{y}) = \vec{x}\circ\vec{x}+2\lambda\vec{x}\circ\vec{y}+\lambda^2\vec{y}\circ\vec{y}\geq 0 \tag{15.69}$$

Il s'agit donc d'une simple équation du deuxième degré où la variable est λ. En se rappelant de ce que nous avons vu lors de notre étude des polynômes du deuxième degré (cf. chapitre de Calcul Algébrique) la relation précédente (le fait qu'elle soit toujours supérieure ou égale à zéro) est satisfaite que si le discriminant b^2-4ac est négatif ou nul. En d'autres termes, si :

$$4(\vec{x}\circ\vec{y})^2-4(\vec{x}\circ\vec{x})(\overrightarrow{y}\circ\vec{y})\leq 0 \tag{15.70}$$

Soit après simplification :

$$(\vec{x}\circ\vec{y})^2-(\vec{x}\circ\vec{x})(\overrightarrow{y}\circ\vec{y})\leq 0\Rightarrow(\vec{x}\circ\vec{y})^2\leq(\vec{x}\circ\vec{x})(\vec{y}\circ\vec{y}) \tag{15.71}$$

•

Lorsque E est $\mathbb{R}^n$, l'inégalité de Cauchy-Schwarz s'écrit avec les composantes des vecteurs :

$$\sum_{i=1}^{n}a_ib_i \leq \sqrt{\sum_{i=1}^{n}a_i^2\sum_{i=1}^{n}b_i^2} \tag{15.72}$$

Dans le cas particulier où $\forall i, b_i = 1$ elle devient :

$$\left(\sum_{i=1}^{n} a_i\right)^2 \leq n \sum_{i=1}^{n} a_i^2 \tag{15.73}$$

ou encore :

$$\left(\frac{1}{n}\sum_{i=1}^{n} a_i\right)^2 \leq \frac{1}{n}\sum_{i=1}^{n} a_i^2 \tag{15.74}$$

ce qui montre que le carré de la moyenne arithmétique est inférieur ou égal à la moyenne arithmétique des carrés. Ce résultat est très important pour l'étude des statistiques.

Inégalité triangulaire

En majorant $2\vec{x} \circ \vec{y}$ par $||\vec{x}||||\vec{y}||$ (à l'aide de l'inégalité de Cauchy-Schwartz) et en mettant ceci dans la relation établie déjà précédement :

$$||\vec{x} + \vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2 + 2\vec{x} \circ \vec{y} \tag{15.75}$$

Nous obtenons

$$||\vec{x} + \vec{y}||^2 \leq ||\vec{x}||^2 + ||\vec{y}||^2 + 2\vec{x} \circ \vec{y} = (||\vec{x}|| + ||\vec{y}||)^2 \tag{15.76}$$

ce qui entraîne la fameuse "inégalité triangulaire" (très utile dans l'étude des suites et séries pour l'étude des convergences ainsi qu'en topologie) :

$$\boxed{||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||} \tag{15.77}$$

Remarque La généralisation de cette inégalité relativement aux propriétés des normes telles que nous le verrons en topologie, donne ce que nous appelons "l'inégalité de Minkowski".

En appliquant une fois l'inégalité triangulaire aux vecteurs et $\vec{x}$ et $(\vec{y} - \vec{x})$ une autre fois aux vecteurs $\vec{y}$ et $(\vec{x} - \vec{y})$ nous obtenons la variante :

$$\boxed{||\vec{x}|| - ||\vec{y}|| \leq ||\vec{x} - \vec{y}||} \tag{15.78}$$

En vertu de la propriété du cosinus et de l'inégalité de Cauchy-Schwartz $\vec{x} \circ \vec{y} \leq ||\vec{x}||||\vec{y}||$, nous pouvons écrire :

$$-1 \leq \frac{\vec{x} \circ \vec{y}}{||\vec{x}||||\vec{y}||} \leq 1 \tag{15.79}$$

relation que nous retrouvons également dans le cadre de l'étude des statistiques (voir chapitre de probabilité et statistiques).

Produit scalaire (général)

Voyons maintenant une autre manière un peu plus générale (s'appliquant à des vecteurs ou fonctions), formelle et abstraite pour définir le produit scalaire tout en tentant de rester le plus simple possible (attention dans le cas général la notation du produit scalaire change !) :

Définition 15.15 *Soit E un espace vectoriel réel (!). Une "forme bilinéaire symétrique positive" sur E, est une application :*

$$\phi : \begin{cases} E \times E \to \mathbb{R} \\ (x,y) \mapsto \langle x|y\rangle \end{cases} \tag{15.80}$$

qui vérifie (par définition !) :

P1 La positivité :

$$\forall x \in E \qquad \phi(x,x) = \langle x|x\rangle \geq 0 \in \mathbb{R} \tag{15.81}$$

P2 La nullité (définie)

$$\forall x \in E \qquad \phi(x,x) = \langle x|x\rangle = 0 \Rightarrow x = 0 \tag{15.82}$$

P3 La symétrie

$$\forall (x,y) \in E^2 \qquad \phi(x,y) = \phi(y,x) \Leftrightarrow \langle x|y\rangle = \langle y|x\rangle \tag{15.83}$$

P4 La bilinéarité *(forme bilinéaire) avec, dans l'ordre, la "linéarité à gauche" et la "linéarité à droite" :*

$$\forall (x,y,z) \in E^3, \quad \forall (\lambda,\mu) \in \mathbb{R}^2 \quad \begin{cases} \langle \lambda x + \mu y|z\rangle = \lambda\langle x|z\rangle + \mu\langle y|z\rangle \\ \langle x|\lambda y + \mu z\rangle = \lambda\langle x|y\rangle + \mu\langle x|z\rangle \end{cases} \tag{15.84}$$

REMARQUE A nouveau, ces propriétés sont principalement imposées par notre approche intuitive de l'espace euclidien et de son interprétation géométrique.

Définition 15.16 *Un espace E muni d'un produit scalaire est appelé un "espace préhilbertien". Si E est de dimension finie, nous parlons alors "d'espace euclidien".*

Nous avons vu en topologie (cf. chapitre de Topologie) que les propriétés du produit scalaire sont les briques de bases pour définir une norme et donc une distance dans un espace métrique. Cette distance sera alors donnée selon ce que nous avons obtenu en topologie :

$$\forall x, y \in E \quad d(x,y) := ||x - y|| \tag{15.85}$$

Définition 15.17 *Nous disons qu'un espace E muni d'un produit scalaire $\langle \ | \ \rangle$ est un "espace Hilbertien" ou "espace de Hilbert" si cet espace est complet pour la métrique définie ci-dessus.*

En d'autres termes, avoir un espace métrique muni donc d'une distance générée par un produit scalaire est une chose. Ensuite, avoir une distance mesurable en est une autre. Un espace de Hilbert a donc des distances mesurables au sens topologique du terme car l'ensemble sur lequel on travaille est continu et n'importe quel point peut-être approché indéfiniment (imaginez avoir une règle et que vous ne pouvez pas avec cette règle approcher les points qui définissent les dimensions de votre objet... ce serait gênant...). Donc sans espace complet une grande partie des théorèmes de l'analyse fonctionnelle ne pourraient pas être utilisés dans l'étude des espace vectoriels et cela serait très gênant en physique quantique ondulatoire par exemple...

Formellement, rappelons qu'un espace métrique est complet si toutes les suites de Cauchy (cf. chapitre des Suites Et Séries) de cet espace sont convergentes (cf. chapitre sur les Fractals) dans un espace métrique (cf. chapitre de Topologie).

15.5.3 Produit vectoriel

Le produit vectoriel de deux vecteurs est une opération propre à la dimension 3. Pour l'introduire, il faut préalablement orienter l'espace destiné à le recevoir. L'orientation étant définie au moyen de la notion de "déterminant", nous commencerons par une brève introduction à l'étude de cette notion. Cette étude sera reprise plus tard dans le détail lors de l'analyse des systèmes linéaires dans le chapitre d'algèbre linéaire.

Définition 15.18 *Nous appelons "déterminant" des vecteurs-colonnes de $\mathbb{R}^2$ (pour la forme générale du déterminant se reporter au chapitre d'algèbre linéaire) :*

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{15.86}$$

et nous notons :

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \tag{15.87}$$

le nombre (produit soustrait en croix) :

$$a_1 b_2 - a_2 b_1 \tag{15.88}$$

Nous appelons déterminant des vecteurs-colonnes de (cf. chapitre d'Algèbre Linéaire) :

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \tag{15.89}$$

et nous notons :

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \tag{15.90}$$

le nombre :

$$a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} - a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} = a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_1 b_3 c_2 - a_2 b_1 c_3 - a_3 b_2 c_1 \tag{15.91}$$

Ainsi, la fonction qui associe à tout couple de vecteurs-colonnes de $\mathbb{R}^2$ (à tout triplet de vecteurs-colonnes de $\mathbb{R}^3$) son déterminant est appelé "déterminant d'ordre 2" (respectivement d'ordre 3).

Le déterminant a comme propriété d'être multiplié par -1 si l'un de ses vecteurs colonnes est remplacé par son opposé ou si deux de ses vecteurs-colonnes sont échangés (la vérification étant simple nous nous abstiendrons de la démonstration, sauf sur demande). En plus, le déterminant est non nul si et seulement si ses vecteurs-colonnes sont linéairement indépendants (la démonstration se trouve quelques lignes plus bas et est d'une grande importance en mathématique).

Définition 15.19 *Soit x_1, x_2, x_3 et y_1, y_2, y_3 les composantes respectives des vecteurs $\vec{x}$ et $\vec{y}$ dans la base orthonormale $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$. Nous appelons "produit vectoriel" de $\vec{x}$ et $\vec{y}$, et nous notons indisctement :*

$$\vec{x} \times \vec{y} = \vec{x} \wedge \vec{y} \tag{15.92}$$

le vecteur :

$$(x_2 y_3 - x_3 y_2)\vec{e}_1 + (x_3 y_1 - x_1 y_3)\vec{e}_2 + (x_1 y_2 - x_2 y_1)\vec{e}_3 \tag{15.93}$$

ou sous forme de composantes :

$$\boxed{\vec{x} \times \vec{y} = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}} \tag{15.94}$$

REMARQUE La première notation est la notation internationale due à Gibbs (que nous utiliserons tout au long de ce site), la deuxième est la notation français due à Burali-Forti (assez embêtant car se confond avec l'opérateur ET en logique).

REMARQUE Il est assez embêtant de retenir par coeur les relations qui forment le produit vectoriel habituellement. Mais heureusement il existe au moins trois bons moyens mnémotechniques :

1. Le plus rapide consiste à retrouver l'une des expressions des composantes du produit vectoriel et ensuite par décrémentation des indices (en recommencent à 3 lorsque qu'on arrive à 0) de connaître toutes les autres composantes. Encore faut-il trouver un moyen simple de se souvenir d'une des composantes. Un bon moyen est la propriété mathématique suivante de deux vecteur colinéaires permettant facilement de retrouver la troisème composante (celle selon l'axe Z) :
   Soit deux vecteurs colinéaires dans un même plan, alors :

$$\vec{x} = \lambda\vec{y} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \Leftrightarrow \frac{x_1}{y_1} = \frac{x_2}{y_2} \Leftrightarrow x_1y_2 - x_2y_1 = 0 \tag{15.95}$$

   Nous retrouvons donc bien l'expression de la troisième composante du produit vectoriel de deux vecteurs (non nécessairement colinéaires... eux !).
2. La seconde mais que nous verrons lors de notre étude du calcul tensoriel consiste à utiliser le symbole d'antisymétrie (également appelé "tenseur de Levi-Civita"). Cette méthode est certainement la plus esthétique d'entre toutes mais pas nécessairement la plus rapide à développer. Nous donnons ici juste l'expression sans plus d'explications pour l'instant (elle est également utile pour l'expression du déterminant par extension) :

$$(\vec{x} \times \vec{y})_i = \varepsilon_{ijk} x_j y_k \tag{15.96}$$

3. Cette dernière méthode est assez simple et triviale aussi mais elle utilise implicitement la première méthode : la i-ème composante $\vec{x} \times \vec{y}$ est le déterminant des deux colonnes privées de leur i-ème terme, le deuxième déterminant étant cependant pris avec le signe "-" tel que :

$$\vec{x} \times \vec{y} = \begin{vmatrix} x2 & y2 \\ x3 & y3 \end{vmatrix} \vec{e}_1 - \begin{vmatrix} x1 & y1 \\ x3 & y3 \end{vmatrix} \vec{e}_2 + \begin{vmatrix} x1 & y1 \\ x2 & y2 \end{vmatrix} \vec{e}_3 \tag{15.97}$$

Il est important, même si c'est relativement simple, de se rappeler que les différents produits vectoriels pour les vecteurs d'une base orthogonale sont :

$$\begin{gathered} \vec{e}_1 \times \vec{e}_1 = \vec{e}_2 \times \vec{e}_2 = \vec{e}_3 \times \vec{e}_3 = 0 \\ \vec{e}_1 \times \vec{e}_2 = \vec{e}_3, \quad \vec{e}_2 \times \overrightarrow{e}_3 = \vec{e}_1, \quad \vec{e}_3 \times \vec{e}_1 = \overrightarrow{e}_2 \end{gathered} \tag{15.98}$$

Le produit vectoriel jouit aussi propriétés suivantes que nous allons démontrer :

P1 Antisymétrie

$$\vec{x} \times \vec{y} = -(\vec{y} \times \vec{x}) \tag{15.99}$$

P2 Linéarité

$$(\alpha\vec{x} \times \beta\vec{y}) \times \vec{z} = \alpha(\overrightarrow{x} \times \vec{z}) + \beta(\vec{y} \times \vec{z}) \tag{15.100}$$

P3 Si et seulement si et sont linéairement indépendants (très important !)

$$\vec{x} \times \vec{y} \neq \vec{0} \tag{15.101}$$

P4 Non associativité

$$\vec{x} \times (\vec{y} \times \vec{z}) \neq (\vec{x} \times \overrightarrow{y}) \times \vec{z} \tag{15.102}$$

Les deux premières propriétés découlent directement de la définition et la propriété P4 se vérifié aisément en développant les composantes et en comparant les résultats obtenus.

Démontrons alors la troisième propriété qui est très importante en algèbre linéaire.

Preuve Soient deux vecteurs $\vec{x}(x_1, x_2, x_3)$ et $\vec{y}(y_1, y_2, y_3)$. Si les deux vecteurs sont linéairement dépendants alors il existe α tel que nous puissions écrire :

$$\vec{y} = \alpha \cdot \vec{x} \tag{15.103}$$

Si nous développons le produit vectoriel des deux vecteurs dépendants à un facteur α près, nous obtenons :

$$\begin{aligned}\vec{x} \times \vec{y} &= \vec{x} \times \alpha\vec{x} \\ &= (x_2\alpha x_3 - x_3\alpha x_2)\vec{e}_1 + (x_3\alpha x_1 - x_1\alpha x_3)\vec{e}_2 + (x_1\alpha x_2 - x_2\alpha x_1)\vec{e}_3\end{aligned} \tag{15.104}$$

Il va sans dire que le résultat ci-dessus est égal au vecteur nul si effectivement les deux vecteurs sont linéairement dépendants. •

Si nous supposons maintenant que les deux vecteurs et sont linéairement indépendants et non nuls, nous devons démontrer que le produit vectoriel est :
P3.1. Orthogonal (perpendiculaire) à $\vec{x}$ et $\vec{y}$
P3.2. De norme $||\vec{x}||||\vec{y}|| \sin(\theta)$, où θ est l'angle entre $\vec{x}$ et $\vec{y}$

Preuve Commencons par la première propriété P3.1 (première importance en physique !) :

$$\begin{aligned}\vec{x} \circ (\vec{x} \times \vec{y}) &= x_1 \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - x_2 \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + x_3 \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \\ &= \begin{vmatrix} x_1 & x_1 & y_2 \\ x_2 & x_2 & y_2 \\ x_3 & x_3 & y_3 \end{vmatrix} \\ &= 0\end{aligned} \tag{15.105}$$

ce qui montre bien que le vecteur $\vec{x}$ est perpendiculaire au vecteur résultant du produit vectoriel entre $\vec{x}$ et $\vec{y}$! •

Terminons avec la deuxième propriété P3.2 (aussi de première importance en physique !) :

Preuve Soit le carré de la norme du produit vectoriel $||\vec{x} \times \vec{y}||^2$. D'après la définition du produit vectoriel nous avons :

$$\begin{aligned}||\vec{x} \times \vec{y}||^2 &= (x_2y_3 - x_3y_2)^2 + (x_3y_1 - x_1y_3)^2 + (x_1y_2 - x_2y_1)^2 \\ &= (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) - (x_1y_1 + x_2y_2 + x_3y_3)^2 \\ &= ||\vec{x}||^2||\vec{y}||^2 - ||\vec{x}||^2||\vec{y}||^2 \cos^2\theta \\ &= ||\vec{x}||^2||\vec{y}||^2(1 - \cos^2\theta) \\ &= ||\vec{x}||^2||\vec{y}||^2 \sin^2\theta\end{aligned} \tag{15.106}$$

Donc finalement :

$$\boxed{||\vec{x} \times \vec{y}|| = ||\vec{x}||||\vec{y}|| \sin\theta} \tag{15.107}$$

•

Nous remarquerons que dans le cas où E est l'espace vectoriel géométrique, la norme du produit vectoriel représente l'aire du parallélogramme construit sur des représentants $\vec{x}$ et $\vec{y}$ d'origine commune (cfr figure 15.6).

Si $\vec{x}$ et $\vec{y}$ sont linéairement indépendants, le triplet $(\vec{x}\times\vec{y}, \vec{x}, \vec{y})$ et donc aussi le triplet $(\vec{x}, \vec{y}, \vec{x}\times\vec{y})$ sont directs.

En effet, (z_1, z_2, z_3) étant les composantes de $\vec{x} \times \vec{y}$ (dans la base $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$), le déterminant de passage de $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$ à $(\vec{x} \times \overrightarrow{y}, \vec{x}, \vec{y})$ (par exemple) s'écrit :

$$\begin{vmatrix} z_1 & x_1 & y_1 \\ z_2 & x_2 & y_2 \\ z_3 & x_3 & y_3 \end{vmatrix} = z_1^2 + z_2^2 + z_3^2 \tag{15.108}$$

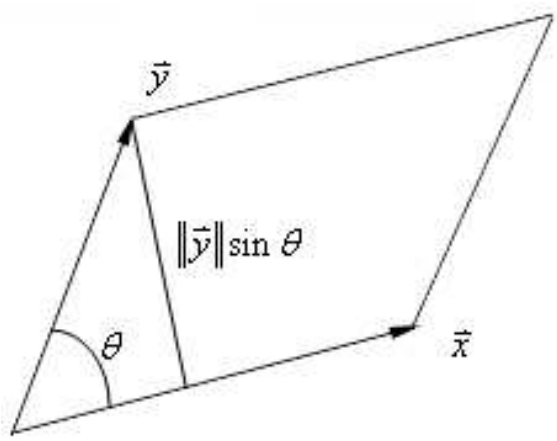


FIGURE 15.6 – Norme du produit vectoriel dans un espace géométrique

Ce déterminant est donc positif, puisqu'au moins un des z_i n'est pas nul, d'après la troisième propriété d'indépendance linéaire du produit vectoriel.

Voici encore quelques propriétés très importantes d'utilité pratique du produit vectoriel (en physique particulièrement) qui sont triviales à vérifier si les développements sont effectués (nous pouvons les faire sur demande si jamais !) :

P1

$$\vec{x} \times (\vec{y} \times \vec{z}) = (\vec{x} \circ \vec{z})\vec{y} - (\vec{x} \circ \vec{y})\vec{z} \tag{15.109}$$

REMARQUE Cette relation est appelée la "règle de Grassmann" et il est important de noter que sans les paranthèses le résultat n'est pas unique.

P2

$$(\vec{x} \times \vec{y}) \circ (\vec{z} \times \vec{v}) = (\vec{x} \circ \vec{z})(\vec{y} \circ \vec{v}) - (\vec{x} \circ \vec{v})(\vec{y} \circ \vec{z}) \tag{15.110}$$

P3

$$(\vec{x} \times \vec{y}) \circ \vec{z} = -(\vec{x} \times \vec{z}) \circ \vec{y} \tag{15.111}$$

P4

$$(\vec{x} \times \vec{y}) \circ \vec{z} = \vec{x} \circ (\vec{y} \times \vec{z}) \tag{15.112}$$

P5

$$||\vec{x} \times \vec{y}||^2 = (\vec{x} \circ \vec{x})^2(\vec{y} \circ \vec{y})^2 - (\vec{x} \circ \vec{y})^2 \tag{15.113}$$

15.5.4 Produit mixte

Nous pouvons étendre la définition du produit vectoriel à une autre type d'outil mathématique que nous appelons le "produit mixte" :

Définition 15.20 *Nous appelons "produit mixte" des vecteurs x, y, z le double produit :*

$$\vec{x} \circ (\vec{y} \times \vec{z}) \tag{15.114}$$

souvent condensé sous la notation suivante :

$$[\vec{x}, \vec{y}, \vec{z}] \tag{15.115}$$

D'après ce que nous avons vu lors de la définition du produit scalaire et vectoriel , le produit mixte peut également s'écrire :

$$||\vec{x}|| \circ ||\vec{y} \times \vec{z}|| \cos\theta \tag{15.116}$$

Nous remarquerons que dans le cas où E est l'espace vectoriel eucliden $\mathbb{R}^3$, la valeur absolue du produit mixte symbole le volume (orienté) du parallélépipède, construit sur des représentants x, y, z d'origine commune.

Remarque Il est assez trivial que le produit mixte est un extension à 3 dimension du produit vectoriel. Effectivement, dans l'espression du produire mixte, le produit vectoriel représente la surface de base du parallélépipède et le produit scalaire projette un des vecteurs sur le vecteur résultant du produit vectoriel ce qui donner la hauteur h du parallélépipède.

De par les propriétés de commutativité du produit scalaire, nous avons :

$$(\vec{x} \times \vec{y}) \circ \vec{z} = \vec{z} \circ (\vec{x} \times \vec{y}) \tag{15.117}$$

et le lecteur peut vérifier sans aucune peine (nous le ferons s'il y a demande) en développant les composantes que :

$$(\vec{x} \times \vec{y}) \circ \vec{z} = \vec{x} \circ (\vec{y} \times \vec{z}) \tag{15.118}$$

Le produit mixte jouit également des propriétés que le lecteur ne devrait avoir aucun mal à vérifier en développant les composantes mis à part peut-être P3 qui découle des propriétés du produit scalaire et vectoriel (nous pouvons développer sur demande si jamais !) :

P1

$$[\vec{x}, \vec{y}, \vec{z}] = [\vec{z}, \vec{x}, \vec{y}] = [\vec{y}, \vec{z}, \vec{x}] = -[\vec{y}, \vec{x}, \vec{z}] = -[\vec{z}, \vec{y}, \vec{x}] = -[\vec{x}, \vec{z}, \vec{y}] \tag{15.119}$$

P2

$$[\alpha\vec{x} + \beta\vec{y}, \vec{z}, \vec{v}] = \alpha[\vec{x}, \vec{z}, \vec{v}] + \beta[\vec{y}, \vec{z}, \vec{v}] \tag{15.120}$$

P3

$$[\vec{x}, \vec{y}, \vec{z}] \neq 0 \tag{15.121}$$

si et seulement si x, y, z sont linéairement indépendants

P4

$$(\vec{x} \times \vec{y}) \times (\vec{z} \times \vec{v}) = [\vec{x}, \vec{y}, \vec{v}] \cdot \vec{z} - [\vec{x}, \vec{y}, \vec{z}] \cdot \vec{v} \tag{15.122}$$

Remarque Nous reviendrons sur le produit mixte lors de notre étude du calcul tensoriel car il permet d'arriver à un résultat très intéressant en particulier en ce qui concerne la relativité générale !

15.6 Espaces vectoriels fonctionnels

Soit $C^k_{[a,b]}$ l'ensemble des fonctions réelles k-fois dérivables dans l'intervalle fermé $[a, b]$. Nous désignerons les éléments de cet ensemble par les lettres $\vec{f}, \vec{g}, \cdots$

La valeur de $\vec{f}$ au point t sera bien évidemment notée $\vec{f}(t)$. Dire que $\vec{f} = \vec{g}$ équivaudra donc à dire que :

$$\vec{f} = \vec{g}, \quad \forall t \in [a, b] \tag{15.123}$$

De manière abrégée, nous écrirons $\vec{f} \equiv \vec{g}$, le signe $\equiv$ indiquant ainsi que les deux membres sont égaux pour tout t de l'intervalle $[a, b]$.

Considérons les deux opérations suivantes :

1. $\vec{f}(t) + \vec{g}(t)$ définie par la formule $(\vec{f} + \vec{g})(t) \equiv \vec{f}(t) + \vec{g}(t)$
2. $\alpha\vec{f}$ définie par la formule $(\alpha\vec{f})(t) \equiv \alpha \overrightarrow{f}(t)$

Ces deux opérations satisfont à toutes les conditions des vecteurs d'un espace vectoriel que nous avons déjà définies au début de ce chapitre (associativité, commutativité, vecteur nul, vecteur opposé, distributivité, élément neutre) et munissent donc $C^k_{[a,b]}$ d'une structure d'espace vectoriel. Le vecteur nul de cet espace étant bien évidemment la fonction nulle et l'opposé de $\vec{f}$ étant la fonction $-\vec{f}$.

Il est intéressant de constater que $C^k_{[a,b]}$ en tant qu'espace vectoriel est une généralisation de $\mathbb{R}^n$ au cas continu. Nous pouvons en effet concevoir tout vecteur (a_i) de $\mathbb{R}^n$ sous la forme d'une fonction réelle définie dans l'ensemble $\{1, 2, \cdots, n\}$: la valeur de cette fonction au point i est tout simplement a_i.

REMARQUE Les polynômes de degré n et à une inconnue forment aussi un exemple d'espace vectoriel fonctionnel de dimension $n+1$ (à chaque coefficient du polynôme correspond une composante du vecteur).

Le champ d'application privilégié de la théorie abstraite du produit scalaire est constitué par les espaces vectoriels fonctionnels. Nous appelons ainsi aussi produit scalaire canonique dans $C_{[a,b]}$ l'opération définie par la relation :

$$\vec{f} \circ \vec{g} = \int_a^b \vec{f}(t)\vec{g}(t)dt \tag{15.124}$$

Cette opération définit bien un produit scalaire, les propriétés de ce dernier sont vérifiées et, en outre, l'intégrale :

$$\vec{f} \circ \vec{f} = ||\vec{f}||^2 = \int_a^b \vec{f}^2(t)dt \tag{15.125}$$

est positive si la fonction continue $\vec{f}$ n'est pas identiquement nulle.

15.7 Espace vectoriels hermitiens

L'objectif de ce qui va suivre n'est pas de faire une étude détaillée du sujet des espaces vectoriels complexes mais juste de donner la bagage et le vocabulaire minimum nécessaire à l'étude de certaines théories physiques comme la physique quantique par exemple.

Lorsque les scalaires qui apparaissent dans la définition de la notion d'espace vectoriel sont des nombres complexes (cf. chapitre sur les Nombres), et non plus des nombres réels, nous parlons alors "d'espaces vectoriels complexes".

REMARQUE Rigoureusement dans la communication courante, il devrait systématiquement être fait mention si nous parlons d'espace vectoriel réel ou d'espace vectoriel complexe...

Citons quelques exemples d'espaces vectoriels complexes :

E1 L'espace vectoriel $\mathbb{C}^n$ des vecteurs-colonnes à n termes complexes ($\mathbb{C}$ étant identifié à $\mathbb{C}$). Nous rencontrerons de tels vecteurs dans le chapitre de physique quantique relativiste entre autres.

E2 L'espace vectoriel des polynômes à coefficients complexes en une indéterminée. Nous rencontrerons ce genre d'espaces dans les chapitres de physique quantique ondulatoire ou encore en chimie quantique.

E3 L'espace vectoriel des fonctions complexes d'une variable réelle ou complexe dérivable ou non. Nous rencontrerons ce genre d'espace très fréquemment dans la section de mécanique globalement et surtout dans les chapitres de l'électrodynamique ou encore de mécanique ondulatoire.

Il s'agit ici d'adapter ce que nous avons fait précédemment au espace vectoriels complexes. L'exemple suivant nous montre que nous ne pouvons pas transposer tel quel les définitions précédentes. En effet, considérons l'espace vectoriel $\mathbb{C}^n$. Comme pour $\mathbb{R}^n$, nous pourrions être tenté de définir un produit scalaire sur $\mathbb{C}^n$ par :

$$\forall x, y \in \mathbb{C}^n, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \Rightarrow \langle x|y\rangle = x_1 y_1 + \ldots + x_n y_n \tag{15.126}$$

avec $x_i, y_i \in \mathbb{C}$

Malheureusement, nous nous apercevons que cette définition n'est pas satisfaisante car nous aurions alors :

$$\langle x|x\rangle = x_1^2 + \ldots + x_n^2 \tag{15.127}$$

et cette quantité n'est pas en général un nombre réel dans l'espace des complexes ce qui viole la propriété de positivité du produit scalaire et donc empêche d'introduire tout concept de distance.

Nous ne pourrions donc plus définir une norme en posant $||x|| := \sqrt{\langle x|x\rangle}$. Pour que $\langle x|x\rangle$ soit un nombre réel positif nous voyons qu'il faudrait plutôt définir le produit scalaire comme ceci :

$$\forall x, y \in \mathbb{C}^n, \quad \langle x|y\rangle = x_1\overline{y_1} + \ldots + x_n\overline{y_n} \tag{15.128}$$

Dans ce cas nous avons :

$$\langle x|x\rangle := \overline{x_1} \cdot x_1 + \ldots + \overline{x_n} \cdot x_n = |x_1|^2 + \ldots + |x_n|^2 \tag{15.129}$$

qui est bien un nombre réel positif. A partir de là, nous pouvons à nouveau définir une norme sur l'espace vectoriel complexe $\mathbb{C}^n$ en posant :

$$||\vec{x}|| = \sqrt{\langle x|x\rangle} = \sqrt{|x_1|^2 + \ldots + |x_n|^2} \tag{15.130}$$

Nous allons à présent montrer comment définir un produit scalaire sur en espace vectoriel complexe dans le cas général.

15.7.1 Produit hermitien

Définition 15.21 *Soit H un espace vectoriel complexe (!). Nous appelons "produit scalaire" ou plus exactement "produit hermitien" sur H, une application :*

$$\phi : \begin{cases} H \times H \to \mathbb{C} \\ (x, y) \quad \mapsto \langle x|y\rangle \end{cases} \tag{15.131}$$

qui vérifie :

P1. La positivité

$$\forall x \in H \quad \phi(x, x) = \langle x|x\rangle \geq 0 \in \mathbb{R}^+ \tag{15.132}$$

P2. Nullité (définie)

$$\forall x \in H \quad \phi(x,x) = \langle x|x\rangle = 0 \Leftrightarrow x = 0 \tag{15.133}$$

P3. Symétrie hermitienne

$$\forall (x,y) \in H^2 \quad \phi(x,y) = \overline{\phi(y,x)} \Leftrightarrow \langle x|y\rangle = \overline{\langle y|x\rangle} \tag{15.134}$$

P4. La bilinéarité (forme bilinéaire) change un peu aussi... ce qui fait que nous parlons alors de "sesquilinéarité". Nous parlons alors, dans l'ordre, d'anti-linéarité à gauche et de linéarité à droite tel que :

$$\forall (x,y,z) \in H^3, \forall (\lambda,\mu) \in \mathbb{C}^2 \begin{cases} \langle \lambda x + \mu y|z\rangle = \overline{\lambda}\langle x|z\rangle + \overline{\mu}\langle y|z\rangle \\ \langle x|\lambda y + \mu z\rangle = \lambda\langle x|y\rangle + \mu\langle x|z\rangle \end{cases} \tag{15.135}$$

REMARQUE Certains mathématiciens mettent l'antilinéarité à droite. C'est simplement une question de convention qui n'a aucune importance.

REMARQUE Le lecteur remarquera peut-être sans peine que si les éléments des définitions précédentes sont tous dans $\mathbb{R}$ alors la sesquilinéarité se réduit à la bilinéarité et le caractère hermitien à la symétrie. Donc le produit hermitien se réduit au produit scalaire.

REMARQUE Nous souhaitons donner pour l'instant uniquement le minimum sur le vaste sujet que sont les espaces vectoriels complexes afin que le lecteur puisse lire sans trop de peine le début du chapitre de physique quantique ondulatoire.

Lorsque nous munissons un espace vectoriel complexe d'un produit scalaire alors au même titre qu'un espace vectoriel réel devient un espace vectoriel euclidien ou préhilbertien, l'espace vectoriel complexe devient donc ce que nous appelons un "espace vectoriel hermitien" (terme assez souvent utilisé dans le chapitre de physique quantique ondulatoire).

Définition 15.22 *Encore une fois, nous disons qu'un espace H muni d'un produit hermitien $\langle \ | \ \rangle$ est un "espace de Hilbert" si cet espace est complet pour la métrique définie ci-dessus.*

Ainsi, les espaces de Hilbert sont une généralisation comprenant les produits scalaires et produits hermitiens des espaces euclidiens, préhilbertiens et hermitiens.

15.7.2 Types d'espaces vectoriels

Pour résumer tout cela :
- Nous appelons espace préhilbertien (réel ou complexe) tout espace vectoriel, de dimension finie ou non, muni d'un produit scalaire.
- Nous appelons espace de Hilbert (réel ou complexe) tout espace préhilbertien complet (en tant qu'espace normé).
- Nous appelons espace euclidien tout espace vectoriel réel de dimension finie muni d'un produit scalaire.
- Nous appelons espace hermitien tout espace vectoriel complexe de dimension finie muni d'un produit scalaire.

Nous savons que tout espace vectoriel (réel ou complexe) normé de dimension finie est complet. Ainsi, les espaces euclidiens et les espaces hermitiens sont des espaces de Hilbert (respectivement réels ou complexes).

15.8 Systèmes de coordonnées

Nous allons aborder ici, l'aspect des changements de coordonnées des composantes de vecteurs non pas d'une base à une autre (pour cela il faut aller voir le chapitre d'algèbre linéaire) mais d'un système de coordonnées à un autre. Ce type de transformation à un fort potentiel en physique (ainsi qu'en mathématique) lorsque nous souhaitons simplifier l'étude de systèmes physiques dont les équations deviennent plus facilement manipulables dans d'autres système de coordonnées.

Définition 15.23 *En mathématiques, un "système de coordonnées" permet de faire correspondre à chaque point d'un espace à n dimensions, un n-uplet de scalaires.*

Remarque Dans beaucoup de cas, les scalaires considérés sont des nombres réels, mais il est possible d'utiliser des nombres complexes ou des éléments d'un quelconque corps (cf. chapitre de Théorie Des Ensembles). Plus généralement, les coordonnées peuvent provenir d'un anneau ou d'une autre structure algébrique apparentée.

Bien que nous soyons dans un chapitre et une section de mathématiques du site, nous nous permettrons dans ce qui va suivre de faire une liaison directe avec la physique pour ce qui concerne les expressions de la vitesse et de l'accélération dans différentes systèmes de coordonnées (désolé pour les matheux...)

15.8.1 Système de coordonnées cartésiennes

Nous ne souhaitons pas trop nous attarder sur ce système car il est bien connu de tout le monde habituellement. Rappelons cependant que la plupart du temps, en physique, les systèmes cartésiens dans lequels nous travaillons sont $\mathbb{R}^2$ (deux dimensions spatiales), $\mathbb{R}^3$ (trois dimensions spatiales) voir $\mathbb{R}^4$ ou $\mathbb{C}^4$ (trois dimensions spatiales et une temporelle) lorsque nous travaillons en relativité.

Les systèmes cartésiens sont représentés par des vecteurs de base orthonormés (dans le sens qu'ils sont linéairement indépendant et de norme unité) qui forment une "base euclidienne" d'un espace vectoriel euclidien...

Dans $\mathbb{R}^3$ (cas le plus fréquent), il y a trois vecteurs de base notés traditionnellement :

$$\vec{e}_1, \vec{e}_2, \vec{e}_3 \tag{15.136}$$

Dans ce système, la position d'un point P (repérable par un vecteur $\vec{r}$) est défini par le triplet de nombres noté (en calcul tensoriel) :

$$(x_1, x_2, x_3) \tag{15.137}$$

et en physique plus conventionellement :

$$(x, y, z) \tag{15.138}$$

où habituellement la coordonnée (z) représente la hauteur (la verticale), la coordonnée (x) la largeur et la coordonnée (y) la longueur (évidemment ces choix sont complétement arbitraires).

Ce point P peut être repéré par un vecteur noté arbitrairement $\vec{r}$ dans la base $(\vec{e}_i)$ par la relation (utilisant la notation tensorielle) :

$$\vec{x} = x^i \vec{e}_i \tag{15.139}$$

En physique, si nous travaillons avec des coordonnées, c'est toujours pour pouvoir déterminer l'emplacement d'un élément. Or, comme nous le verrons plus rigoureusement en mécanique analytique, le physicien travaille avec les notions suivantes (chaque élément dépendant souvent du temps) :

Position

$$\vec{r} = \Big(x(t), y(t), z(t), t\Big) \tag{15.140}$$

Vitesse

$$\frac{d\vec{r}}{dt} = \dot{\vec{r}} = \vec{v} = \Big(\dot{x}(t), \dot{y}(t), \dot{z}(t), t\Big) \tag{15.141}$$

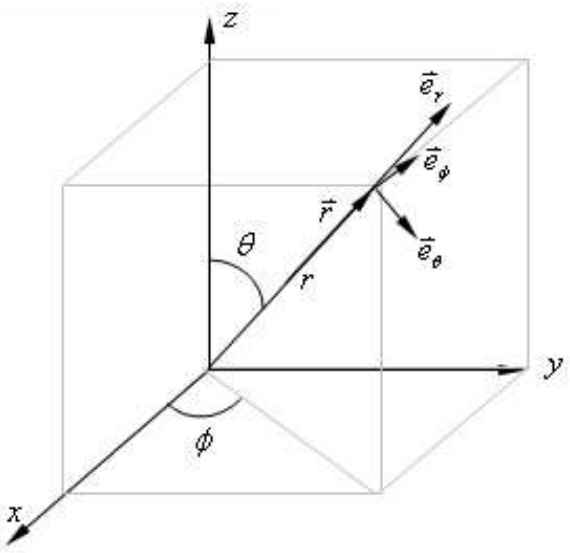


FIGURE 15.7 – Système de coordonnées sphérique

Accélération

$$\frac{d\vec{v}}{dt} = \dot{\vec{v}} = \vec{a} = \left(\ddot{x}(t), \ddot{y}(t), \ddot{z}(t), t\right) \tag{15.142}$$

Maintenant voyons comment s'expriment ces différentes notions dans des systèmes telles que les coordonnnées sphériques, cylindriques et polaires.

15.8.2 Système de coordonnées sphériques

Le choix de commencer avec ce système de coordonnées n'est pas un hasard. Il a pour avantage d'être une généralisation des systèmes cylindrique et polaire dont nous en retrouverons par la suite plus facilement les expressions de la position, de la vitesse et de l'accélération .

Nous représentons traditionnellement un système à coordonnées sphérique comme la figure 15.7.

Nous voyons très bien si nous connaissons bien les relations trigonométriques élémentaires (voir le chapitre du même nom dans la section de géométrie) que nous avons les transformations :

$$\begin{aligned}
r &= \sqrt{x^2+y^2+z^2}, \quad 0 \leq r < \infty \\
\theta &= \arccos\left(\frac{z}{r}\right) = \arccos\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right), \quad 0 \leq \theta < \pi \\
\phi &= \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right) = \arccos\left(\frac{x}{r\sin\theta}\right) = \arcsin\left(\frac{y}{r\sin\theta}\right), \quad 0 \leq \phi < 2\pi
\end{aligned} \tag{15.143}$$

et inversement :

$$x = r\sin\theta\cos\phi \quad y = r\sin\theta\sin\phi \quad z = r\cos\theta \tag{15.144}$$

Maintenant il nous faut trouver les expressions qui relient les vecteurs de la base sphérique que nous choisissons de noter $\vec{e}_r, \vec{e}_\theta, \vec{e}_\phi$ avec les vecteurs de la base cartésienne $\vec{e}_x, \vec{e}_y, \vec{e}_z$. Nous avons,

comme nous pouvons le voir sur la figure 15.7 :

$$\begin{aligned}
\vec{e}_r =& \frac{\vec{r}}{r} = \frac{x\vec{e}_x + y\vec{e}_y + z\vec{e}_z}{r} = \frac{r\sin\theta\cos\phi\vec{e}_x + r\sin\theta\sin\phi\vec{e}_y + r\cos\theta\vec{e}_z}{r} \\
=& \sin\theta\cos\phi\vec{e}_x + \sin\theta\sin\phi\vec{e}_y + \cos\theta\vec{e}_z \\
\vec{e}_\theta =& \frac{\vec{e}_z \times \overrightarrow{e}_r}{\sin\theta} = \frac{1}{\sin\theta}\begin{pmatrix}0\\0\\1\end{pmatrix} \times \begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\phi\end{pmatrix} = \frac{1}{\sin\theta}\begin{pmatrix}-\sin\theta\sin\phi\\ \sin\theta\cos\phi\\ 0\end{pmatrix} \\
=& \frac{1}{\sin\theta}\begin{pmatrix}-\sin\theta\sin\phi\\ 0\\ 0\end{pmatrix} + \frac{1}{\sin\theta}\begin{pmatrix}0\\ sin\theta\cos\phi\\ 0\end{pmatrix} = -\vec{e}_x\sin\phi + \vec{e}_y\cos\phi \\
\vec{e}_\phi =& \vec{e}_\theta \times \vec{e}_r = \begin{pmatrix}-\sin\phi\\ \cos\phi\\ 0\end{pmatrix} \times \begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{pmatrix} = \begin{pmatrix}\cos\phi\cos\theta - 0\cdot\sin\theta\sin\phi\\ 0\cdot\sin\theta\cos\phi - (-\sin\phi)\cos\theta\\ -\sin\phi\cdot\sin\theta\sin\phi - \cos\phi\sin\theta\cos\phi\end{pmatrix} \\
=& \begin{pmatrix}\cos\phi\cos\theta\\ \sin\phi\cos\theta\\ -\sin\phi\sin\theta\sin\phi - \cos\phi\sin\theta\cos\phi\end{pmatrix} = \vec{e}_x\cos\theta\cos\phi + \vec{e}_y\cos\theta\sin\phi - \overrightarrow{e}_z\sin\theta
\end{aligned} \tag{15.145}$$

Explications pour la seconde ligne : En divisant par $\sin\theta$, nous nous assurons de par les propriétés de la norme du produit vectoriel que $||\vec{e}_\phi|| = ||\vec{e}_z||||\vec{e}_r||\sin\theta = 1$ sera bien normalisé à l'unité.

Remarque Il est important de remarquer que le produit vectoriel de deux vecteurs de base donne toujours le troisième vecteur de base perpendiculaire (comme pour les coordonnées cartésiennes donc !).

Pour des besoins ultérieurs, déterminons les différentielles partielles de chacune des ces coordonnées :

$$\begin{aligned}
&\frac{\partial\vec{e}_r}{\partial r} = 0 \\
&\frac{\partial\vec{e}_r}{\partial\theta} = \overrightarrow{e}_x\cos\theta\cos\phi + \vec{e}_y\cos\theta\sin\phi - \vec{e}_z\sin\theta = \vec{e}_\theta \\
&\frac{\partial\vec{e}_r}{\partial\phi} = -\vec{e}_x\sin\theta\sin\phi + \overrightarrow{e}_y\sin\theta\cos\phi = (-\vec{e}_x\sin\theta + \vec{e}_y\cos\phi)\sin\theta = \vec{e}_\phi\sin\theta \\
&\frac{\partial\vec{e}_\phi}{\partial r} = 0 \\
&\frac{\partial\vec{e}_\phi}{\partial\theta} = 0 \\
&\frac{\partial\vec{e}_\phi}{\partial\phi} = -\vec{e}_x\cos\phi - \vec{e}_y\sin\phi = -(\vec{e}_r\sin\theta + \vec{e}_\theta\cos\theta) \\
&\frac{\partial\vec{e}_\theta}{\partial r} = 0 \\
&\frac{\partial\vec{e}_\theta}{\partial\theta} = -\vec{e}_x\sin\cos\phi - \vec{e}_y\sin\theta\sin\phi - \vec{e}_z\cos\theta = -\vec{e}_r \\
&\frac{\partial\vec{e}_\theta}{\partial\phi} = -\vec{e}_x\cos\theta\sin\phi + \vec{e}_y\cos\theta\cos\phi = \vec{e}_\phi\cos\theta
\end{aligned} \tag{15.146}$$

Nous allons également utiliser plus tard (pour l'étude des opérateurs vectoriels) la variation exprimée en coordonnées sphériques :

$$\begin{aligned}
d\vec{r} = d(r\vec{e}_r) = \vec{e}_r dr + r d\vec{e}_r &= \overrightarrow{e}_r dr + r\left(\frac{\partial\vec{e}_r}{\partial r}dr + \frac{\partial\vec{e}_r}{\partial\theta}d\theta + \frac{\partial\vec{e}_r}{\partial\phi}d\phi\right) \\
&= \vec{e}_r dr + \vec{e}_\theta r d\theta + \vec{e}_\phi r\sin\theta d\phi
\end{aligned} \tag{15.147}$$

Pour exprimer la vitesse et l'accélération en coordonnées sphériques, nous aurons également besoin des dérivées par rapport au temps :

$$\begin{aligned}
\dot{\vec{e}}_r &= \frac{\partial \vec{e}_r}{\partial r}\dot{r} + \frac{\partial \overrightarrow{e}_r}{\partial \theta}\dot{\theta} + \frac{\partial \vec{e}_r}{\partial \phi}\dot{\phi} = \vec{e}_\theta\dot{\theta} + \vec{e}_\phi\dot{\phi}\sin\theta \\
\dot{\vec{e}}_\theta &= \frac{\partial \vec{e}_\theta}{\partial r}\dot{r} + \frac{\partial \vec{e}_\theta}{\partial \theta}\dot{\theta} + \frac{\partial \overrightarrow{e}_\theta}{\partial \phi}\dot{\phi} = -\vec{e}_r\dot{\theta} + \overrightarrow{e}_\phi\dot{\phi}\cos\theta \\
\dot{\vec{e}}_\phi &= \frac{\partial \vec{e}_\phi}{\partial r}\dot{r} + \frac{\partial \vec{e}_\phi}{\partial \theta}\dot{\theta} + \frac{\partial \overrightarrow{e}_\phi}{\partial \phi}\dot{\phi} = -(\vec{e}_r\sin\dot{\theta} + \overrightarrow{e}_\theta\cos\dot{\theta})\dot{\phi}
\end{aligned} \tag{15.148}$$

Donc si nous faisons maintenant un peu de physique, nous avons :

$$\vec{v} = \dot{\vec{r}} = \dot{\vec{e}}_r r + \dot{r}\vec{e}_r \tag{15.149}$$

ce qui nous amène à (nous aurons besoin de cette relation en astrophysique) :

$$\boxed{\vec{v} = \dot{r}\vec{e}_r + \vec{e}_\theta r\dot{\theta} + \overrightarrow{e}_\phi r\dot{\phi}\sin\theta} \tag{15.150}$$

Pour l'accélération nous obtenons (exactement dans la même démarche que pour déterminer l'expression de la vitesse) :

$$\boxed{\begin{aligned}\vec{a} =& \vec{e}_r(\ddot{r} - r\dot{\theta}^2 - r\dot{\phi}^2\sin\theta) + \overrightarrow{e}_\theta(r\ddot{\theta} + 2\dot{r}\dot{\theta} - r\dot{\phi}^2\sin\theta\cos\theta) \\ &+ \vec{e}_\phi(r\ddot{\phi}\sin\theta + 2r\dot{\theta}\dot{\phi}\cos\theta + 2\dot{r}\dot{\phi}\sin\theta)\end{aligned}} \tag{15.151}$$

15.8.3 Système de coordonnées cylindriques

Le système de coordonnées cylindriques (très utile dans l'étude des système à mouvements hélicoïdaux) est assez semblable à celui des coordonnées sphériques puisqu'il peut être vu comme une tranche de la sphère. Soit la figure 15.8, il vient sans peine qu'en coordonnées polaires (à un changement de notation près pour l'angle par rapport au schéma ci-dessus) :

$$r = \sqrt{x^2 + y^2} \quad \phi = \arccos\left(\frac{x}{r}\right) \quad \phi = \arcsin\left(\frac{y}{r}\right) \text{ et } \phi = \arctan\left(\frac{x}{y}\right) \tag{15.152}$$

et inversement :

$$x = r\cos\phi \quad y = r\sin\phi \quad z = z \tag{15.153}$$

Maintenant il nous faut trouver les expressions qui relient les vecteurs de la base polaire que nous choisisson de noter $\vec{e}_r, \vec{e}_\phi, \vec{e}_z$ (au lieu de $\hat{r}, \hat{\phi}, \hat{z}$ comme cela se fait traditionnellement) avec les vecteurs de la base cartésienne $\vec{e}_x, \vec{e}_y, \vec{e}_z$. Nous avons, identiquement à ce que nous avons fait pour les coordonnées sphériques :

$$\begin{aligned}
\vec{e}_r &= \frac{\vec{r}}{r} = \frac{x\vec{e}_x + y\overrightarrow{e}_y}{r} = \vec{e}_x\cos\phi + \vec{e}_y\sin\phi \\
\vec{e}_\phi &= \frac{\vec{e}_z \times \vec{e}_r}{\sin\theta} = -\vec{e}_x\sin\phi + \vec{e}_y\cos\phi \\
\vec{e}_z &= \vec{e}_z
\end{aligned} \tag{15.154}$$

Explications pour la seconde ligne : en divisant par $\sin\theta$, nous nous assurons de par les propriétés de la norme du produit vectoriel que $||\vec{e}_\phi|| = ||\vec{e}_z||||\vec{e}_r||\sin\theta = 1$ sera bien normalisé à l'unité. Dans le cas des coordonnées cylindriques l'angle étant de toute façon droit, nous ne serions pas obligé d'indiquer cette division, mais nous avons fait ce choix par souci d'homogénéité avec les développements précédents.

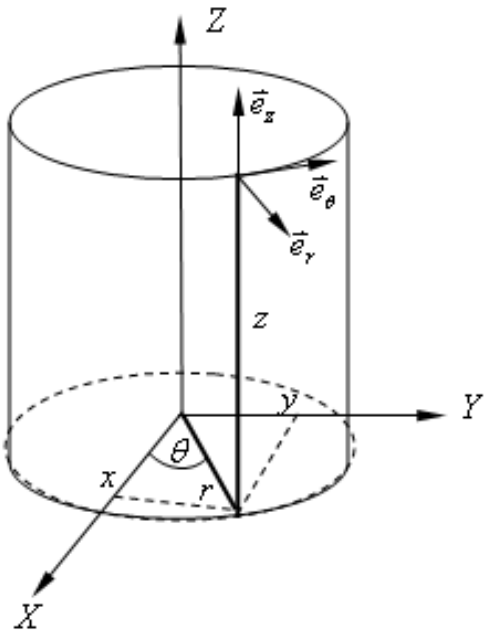


FIGURE 15.8 – Système de coordonnées cylindriques

REMARQUE Il est important de remarquer que le produit vectoriel de deux vecteurs de base donne toujours le troisième vecteur de base perpendiculaire (comme pour les coordonnées cartésiennes et sphériques donc !).

Pour des besoins ultérieurs, déterminons les différentielles partielles de chacune des ces coordonnées :

$$\begin{aligned}
&\frac{\partial \vec{e}_r}{\partial r} = 0 \quad ; \quad \frac{\partial \vec{e}_r}{\partial \phi} = -\vec{e}_x \sin\phi + \vec{e}_y \cos\phi = \vec{e}_\phi \quad ; \quad \frac{\partial \vec{e}_r}{\partial z} = 0 \\
&\frac{\partial \vec{e}_\phi}{\partial r} = 0 \quad ; \quad \frac{\partial \vec{e}_\phi}{\partial \phi} = -\vec{e}_x \cos\phi - \vec{e}_y \sin\phi = -\vec{e}_r \quad ; \quad \frac{\partial \vec{e}_\phi}{\partial z} = 0 \\
&\frac{\partial \vec{e}_z}{\partial r} = 0 \quad ; \quad \frac{\partial \vec{e}_z}{\partial \phi} = 0 \quad ; \quad \frac{\partial \vec{e}_z}{\partial z} = 0
\end{aligned} \tag{15.155}$$

Nous allons également utiliser plus tard (pour l'étude des opérateurs vectoriels) la variation $d\vec{r}$ exprimée en coordonnées cylindriques :

$$d\vec{r} = d(r\vec{e}_r) = \vec{e}_r dr + r d\vec{e}_r = \vec{e}_r dr + r\left(\frac{\partial \vec{e}_r}{\partial r}dr + \frac{\partial \vec{e}_r}{\partial \phi}d\phi + \frac{\partial \vec{e}_r}{\partial z}dz\right) = \vec{e}_r dr + r\vec{e}_\phi d\phi \tag{15.156}$$

Pour exprimer la vitesse et l'accélération en coordonnées sphériques, nous aurons également besoin des dérivées par rapport au temps :

$$\begin{aligned}
\dot{\vec{e}}_r &= \frac{\partial \vec{e}_r}{\partial r}\dot{r} + \frac{\partial \vec{e}_r}{\partial \phi}\dot{\phi} + \frac{\partial \vec{e}_r}{\partial z}\dot{z} = \vec{e}_\phi \dot{\phi} s \\
\dot{\vec{e}}_\theta &= \frac{\partial \vec{e}_r}{\partial r}\dot{r} + \frac{\partial \vec{e}_r}{\partial \phi}\dot{\phi} + \frac{\partial \vec{e}_r}{\partial z}\dot{z} = -\vec{e}_r \dot{\phi} \\
\dot{\vec{e}}_\phi &= \frac{\partial \vec{e}_r}{\partial r}\dot{r} + \frac{\partial \vec{e}_r}{\partial \phi}\dot{\phi} + \frac{\partial \vec{e}_r}{\partial z}\dot{z} = \vec{0}
\end{aligned} \tag{15.157}$$

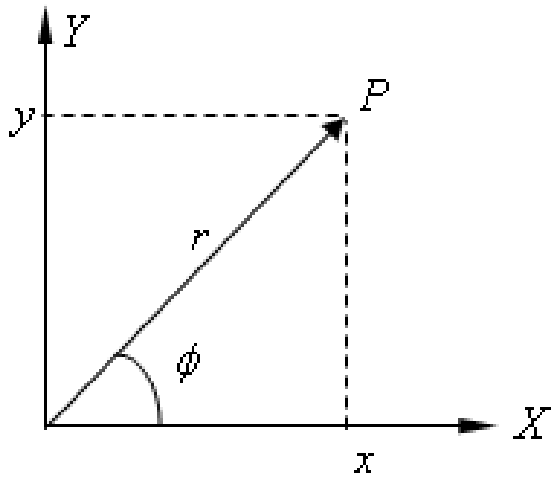


FIGURE 15.9 – Système de coordonnées polaires

Donc si nous faisons maintenant un peu de physique, nous avons (rappellons que la composante z est indépendante des autres composantes cylindriques) :

$$\vec{v} = \dot{\vec{r}} + \dot{\vec{k}} = \dot{\vec{e}}_r r + \dot{r}\dot{\vec{e}}_r + \vec{e}_z \dot{z} + \dot{\vec{e}}_z z \tag{15.158}$$

ce qui nous amène à :

$$\boxed{\vec{v} = \dot{r}\vec{e}_r + \vec{e}_\phi r\dot{\phi} + \vec{e}_z \dot{z}} \tag{15.159}$$

Pour l'accélération nous obtenons (exactement dans la même démarche que pour déterminer l'expression de la vitesse) :

$$\boxed{\vec{a} = \vec{e}_r \left(\ddot{r} - r\dot{\phi}^2\right) + \vec{e}_\phi \left(r\ddot{\phi} + 2\dot{r}\dot{\phi}\right) + \vec{e}_z \ddot{z}} \tag{15.160}$$

15.8.4 Système de coordonnées polaires

Le système de coordonnées polaires est très semblable à celui des coordonnées cylindriques puisqu'il peut être vu comme un retranchement d'une dimension (la hauteur) du système cylindrique (figure **??**.

Ainsi, il vient sans peine qu'en coordonnées polaires :

$$r = \sqrt{x^2 + y^2} \qquad \phi = \arccos\left(\frac{x}{r}\right) \qquad \phi = \arcsin\left(\frac{y}{r}\right) \text{ et } \phi = \arctan\left(\frac{y}{x}\right) \tag{15.161}$$

et inversément :

$$x = r\cos\phi \quad , \quad y = r\sin\phi \tag{15.162}$$

Maintenant il nous faut trouver les expressions qui relient les vecteurs de la base polaire que je choisis de noter $\vec{e}_r, \vec{e}_\phi$ (au lieu de $\hat{r}, \hat{\phi}$ comme cela ce fait traditionnellement) avec les vecteurs de la

base cartésienne $\vec{e}_x, \vec{e}_y, \vec{e}_z$. Nous avons identiquement à ce que nous avions fait pour les coordonnées sphériques :

$$\begin{aligned}\vec{e}_r &= \frac{\vec{r}}{r} = \frac{x\vec{e}_x + y\vec{e}_y}{r} = \vec{e}_x \cos\phi + \vec{e}_y \sin\phi \\ \vec{e}_\phi &= \frac{\vec{e}_z \times \vec{e}_r}{\sin\theta} = -\vec{e}_x \sin\phi + \vec{e}_y \cos\phi\end{aligned} \tag{15.163}$$

Explications pour la seconde ligne : en divisant par $\sin\theta$, nous nous assurons de par les propriétés de la norme du produit vectoriel que $||\vec{e}_\phi|| = ||\vec{e}_z||||\vec{e}_r|| \sin\theta = 1$ sera bien normalisé à l'unité. Dans le cas des coordonnées polaires l'angle étant de toute façon droit, nous ne serions pas obligé d'indiquer cette division, mais nous avons fait ce choix par souci d'homogénéité avec les développements précédents.

Pour des besoins ultérieurs, déterminons les différentielles partielles de chacune des ces coordonnées :

$$\begin{aligned}\frac{\partial \vec{e}_r}{\partial r} = 0 \quad ; \quad \frac{\partial \vec{e}_r}{\partial \phi} &= -\vec{e}_x \sin\phi + \vec{e}_y \cos\phi = \vec{e}_\phi \\ \frac{\partial \vec{e}_\phi}{\partial r} = 0 \quad ; \quad \frac{\partial \vec{e}_\phi}{\partial \phi} &= -\vec{e}_x \cos\phi - \vec{e}_y \sin\phi = -\vec{e}_r\end{aligned} \tag{15.164}$$

Nous allons également utiliser plus tard (pour l'étude des opérateurs vectoriels) la variation $d\vec{r}$ exprimée en coordonnées cylindriques :

$$\begin{aligned}d\vec{r} = d(r\vec{e}_r) = \vec{e}_r dr + r d\vec{e}_r = \vec{e}_r dr + r\left(\frac{\partial \vec{e}_r}{\partial r}dr + \frac{\partial \vec{e}_r}{\partial \phi}d\phi\right) \\ = \vec{e}_r dr + \vec{e}_\phi r d\phi\end{aligned} \tag{15.165}$$

Pour exprimer la vitesse et l'accélération en coordonnées sphériques, nous aurons également besoin des dérivées par rapport au temps :

$$\begin{aligned}\dot{\vec{e}}_r &= \frac{\partial \vec{e}_r}{\partial r}\dot{r} + \frac{\partial \vec{e}_r}{\partial \phi}\dot{\phi} = \vec{e}_\phi \dot{\phi} \\ \dot{\vec{e}}_\phi &= \frac{\partial \vec{e}_\phi}{\partial r}\dot{r} + \frac{\partial \vec{e}_\phi}{\partial \phi}\dot{\phi} = -\vec{e}_r \dot{\phi}\end{aligned} \tag{15.166}$$

Donc si nous faisons maintenant un peu de physique, nous avons :

$$\vec{v} = \dot{\vec{r}} = \dot{\vec{e}}_r r + \dot{r}\vec{e}_r \tag{15.167}$$

ce qui nous amène à :

$$\boxed{\vec{v} = \dot{r}\vec{e}_r + \vec{e}_\phi r\dot{\phi}} \tag{15.168}$$

où le premier terme est la composante radiale de la vitesse et le second la composante tangentielle de la vitesse (angulaire).

Pour l'accélération nous obtenons (exactement dans la même démarche que pour déterminer l'expression de la vitesse) :

$$\boxed{\vec{a} = \vec{e}_r(\ddot{r} - \dot{\phi}^2 r) + \vec{e}_\phi(2\dot{r}\dot{\phi} + r\ddot{\phi})} \tag{15.169}$$

où le premier terme est l'accélération radiale, le second l'accélération centripète, le troisième l'accélération de Coriolis et le quatrième l'accélération tangentielle.

15.9 Opérateurs différentiels

Définition 15.24 *Un champ scalaire, vectoriel ou tensoriel, dans un volume V, est une application qui, à tout point $\vec{x}$ de ce volume V, associe respectivement une grandeur scalaire, vectorielle ou tensorielle.*

Ainsi, l'application f qui, à tout point $\vec{x}$ de V, de coordonnées spatiales x, y, z associe la valeur scalaire $f(\vec{x}) = f(x, y, z)$ est un champs scalaire dans V.

En chaque point d'un volume traversé par un fluide en mouvement, le vecteur qui coïncide à chaque instant avec la vitesse de la particule changeante qui passe en ce point à ce même instant définit un champ vectoriel 3D, éventuellement variable dans le temps. Les champs ainsi définis constituent un outil mathématique de base dans l'ensemble de la physique.

REMARQUE Lorsque nous représentons graphiquement un champ scalaire, l'ensemble des points continus de valeur égale constituent ce que l'on appelle des "isolignes" ou plus couramment "courbes de niveau".

Le gradient, la divergence et le rotationnel sont les trois principaux opérateurs différentiels linéaires du premier ordre que nous allons présenter ici. Cela signifie qu'ils ne font intervenir que des dérivées partielles (ou différentielles) premières des champs, à la différence, par exemple, du laplacien qui fait intervenir des dérivées partielles d'ordre deux.

Nous les rencontrerons en particulier en mécanique des fluides et en électromagnétisme ainsi que physique quantique ondulatoire où ils permettent d'exprimer facilement certaines propriétés.

15.9.1 Gradient d'un champ scalaire

Le gradient est un opérateur qui s'applique à un champ de scalaires et le transforme en champ de vecteurs. Intuitivement, le gradient indique la direction de la plus grande variation du champ scalaire, et l'intensité de cette variation. Par exemple, le gradient de l'altitude est dirigé selon la ligne de plus grande pente et sa norme augmente avec la pente.

Soit un champ scalaire tridimensionnel $f(x, y, z)$, où x et y et z sont les coordonnées cartésiennes d'un point M de l'espace. Lorsque M se déplace dans l'espace selon le vecteur dM de composantes dx, dy et dz, le champ scalaire f varie de df selon la différentielle totale :

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \tag{15.170}$$

A partir de cette relation, nous pouvons définir "l'opérateur gradient" d'un champ scalaire tel que :

$$df = \overrightarrow{grad}(f) \circ d\vec{r} = \vec{\nabla}(f) \circ d\vec{r} \tag{15.171}$$

où :

$$\overrightarrow{grad}(f) \tag{15.172}$$

est un terme vectoriel appelé le "gradient du champ scalaire f". Pour condenser l'écriture, nous utilisons parfois le symbole $\overrightarrow{grad}$ nommé le "nabla du champ scalaire f".

Le vecteur obtenu par le calcul de du gradient sur les quatre propriétés suivante :

P1 Ses composantes représentent la variation (pente) de la fonction f selon les différentes directions de l'espace.

P2 Sa norme est la variation maximale de f en fonction de la distance.

P3 Sa direction est selon la variation maximale de f en fonction de la distance.

P4 Le sens indique les valeurs où f augmente.

A partir de la définition et de la différentielle totale, nous obtenons

$$\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = \vec{\nabla}(f)_x dx + \vec{\nabla}(f)_y dy + \vec{\nabla}(f)_z dz \tag{15.173}$$

Ce qui nous amène à poser que :

$$\vec{\nabla}(f)_x = \frac{\partial f}{\partial x}dx \quad ; \quad \vec{\nabla}(f)_y = \frac{\partial f}{\partial y}dy \quad ; \quad \vec{\nabla}(f)_z = \frac{\partial f}{\partial z}dz \tag{15.174}$$

et donc que finalement l'opérateur "gradient en coordonnées cartésiennes" est donné par :

$$\boxed{\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)} \tag{15.175}$$

Finalement nous voyons que le gradient d'un champ scalaire $f(x, y, z)$ est le champ vectoriel dont les composantes en chaque point sont les trois dérivées du champ scalaire f par rapport aux trois coordonnées spatiales, notées ici x, y, z.

La variation de f pour un déplacement $d\vec{r}$ est donc le produit scalaire de $d\vec{r}$ par le gradient du champ f. Or, un déplacement infinitésimal effectué le long d'une isoligne (décrivant une isosurface), du champ scalaire tridimensionnel $f(x, y, z)$ n'engendre aucune variation df de f. Le produit scalaire évoqué est donc nul dans ce cas, ce qui implique que $d\vec{r}$ et $\vec{\nabla}(f)$ sont perpendiculaires.

En considérant cette fois un déplacement perpendiculaire aux isolignes, nous montrons facilement que le vecteur gradient de f est dirigé depuis les faibles valeurs de f vers les fortes valeurs de f. Son module étant d'autant plus grand que f varie rapidement au voisinage du point considéré.

Par sa direction, son sens et son module, le vecteur gradient d'un champ en un point comporte donc des indications sur la manière dont varie le champ autour de ce point.

Remarque La condition nécessaire et suffisante pour qu'un champ de vecteurs soit le gradient d'un champ scalaire f est que ce champ vectoriel soit irrotationnel (voir plus loin l'opérateur "rotationnel d'un champ vectoriel").

Après avoir défini le gradient en coordonnées cartésiennes x, y, z nous devons nous intéresser à l'expression de cet opérateur dans d'autres systèmes de coordonnées. Il est fréquent en physique d'avoir à utiliser les coordonnées cylindriques, polaires et sphériques pour simplifier l'étude formelle de systèmes physiques. Ainsi, si nous faisons référence à notre étude des systèmes de coordonnées, nous avons (rappel) d'abord en coordonnées polaires :

$$d\vec{r} = \vec{e}dr + \vec{e}_\phi d\phi \tag{15.176}$$

Or, avec la définition du gradient en coordonnées cartésiennes, nous avons en coordonnées polaires la définition suivante :

$$df = \vec{\nabla}(f) \circ d\vec{r} \tag{15.177}$$

Si nous exprimons la différentielle totale exacte (cf. chapitre de Calcul Différentiel Et Intégral) de df nous obtenons les relations suivantes :

$$df = \frac{\partial f}{\partial r}dr + \frac{\partial f}{\partial \phi}d\phi = \vec{\nabla}(f) \circ d\vec{r} = \vec{\nabla}(f) \circ (\vec{e}_r dr + \vec{e}rd\phi) \tag{15.178}$$

Ce qui nous permet d'obtenir la relation :

$$df = \frac{\partial f}{\partial r}dr + \frac{\partial f}{\partial \phi}d\phi = \vec{\nabla}(f) \circ (\vec{e}_r dr + \vec{e}_\phi rd\phi) = \vec{\nabla}(f)_r dr + \vec{\nabla}(f)_\phi rd\phi \tag{15.179}$$

donc :

$$df = \frac{\partial f}{\partial r}dr + \frac{\partial f}{\partial \phi}d\phi = \vec{\nabla}(f)_r dr + \vec{\nabla}(f)_\phi rd\phi \tag{15.180}$$

ce qui nous amène à :

$$\vec{\nabla}(f)_r = \frac{\partial f}{\partial r} \quad ; \quad \vec{\nabla}(f)_\phi = \frac{1}{r}\frac{\partial f}{\partial \phi} \tag{15.181}$$

Ainsi le "gradient en coordonnées polaires" s'exprime comme :

$$\boxed{\vec{\nabla} = \left(\frac{\partial}{\partial r}, \frac{1}{r}\frac{\partial f}{\partial \phi}\right)} \tag{15.182}$$

Occupons nous maintenant de l'expression du gradient en coordonnées cylindriques. Rappelons que lors de l'étude des différents systèmes de coordonnées nous avons obtenu pour les coordonnées cylindriques :

$$d\vec{r} = \vec{e}dr + \vec{e}_\phi r d\phi + \vec{e}_z dz \tag{15.183}$$

Donc nous savons déjà que l'expression du gradient en coordonnées cylindriques sera identique à celle en coordonnées polaires à l'exception de l'ajout de la composante verticale z indépendante des autres coordonnées. Ainsi, nous obtenons l'opérateur "gradient en coordonnées cylindriques" :

$$\boxed{\vec{\nabla} = \left(\frac{\partial}{\partial r}, \frac{1}{r}\frac{\partial f}{\partial \phi}, \frac{\partial}{\partial z}\right)} \tag{15.184}$$

Occupons nous maintenant de l'expression du gradient en coordonnées sphériques. Rappelons que lors de l'étude des différents systèmes de coordonnées nous avons obtenu pour les coordonnées sphériques :

$$d\vec{r} = \vec{e}_r dr + \vec{e}_\theta r d\theta + \vec{e}_\phi r \sin\theta d\phi \tag{15.185}$$

Or, avec la définition du gradient en coordonnées cartésiennes, nous avons en coordonnées sphériques la définition suivante :

$$df = \vec{\nabla}(f) \circ d\vec{r} \tag{15.186}$$

Si nous exprimons la différentielle totale de df nous obtenons les relations suivantes :

$$df = \frac{\partial f}{\partial r}dr + \frac{\partial f}{\partial \phi}d\phi + \frac{\partial f}{\partial \theta}d\theta = \vec{\nabla}(f) \circ d\vec{r} = \vec{\nabla}(f) \circ d(\vec{e}_r dr + \vec{e}_\theta r d\theta + \vec{e}_\phi r \sin\theta d\phi) \tag{15.187}$$

Ce qui nous permet d'obtenir la relation (nous utilisons maintenant la notation qui use de l'opérateur "nabla") :

$$\begin{aligned} \frac{\partial f}{\partial r}dr + \frac{\partial f}{\partial \phi}d\phi + \frac{\partial f}{\partial \theta}d\theta &= \vec{\nabla}(f) \circ d(\vec{e}_r dr + \vec{e}_\theta r d\theta + \vec{e}_\phi r \sin\theta d\phi) \\ &= \vec{\nabla}(f)_r dr + \vec{\nabla}(f)_\theta r d\theta + \vec{\nabla}(f)_\phi r \sin\theta d\phi \end{aligned} \tag{15.188}$$

La relation :

$$\frac{\partial f}{\partial r}dr + \frac{\partial f}{\partial \phi}d\phi + \frac{\partial f}{\partial \theta}d\theta = \vec{\nabla}(f)_r dr + \vec{\nabla}(f)_\theta r d\theta + \vec{\nabla}(f)_\phi r \sin\theta d\phi \tag{15.189}$$

Nous impose :

$$\vec{\nabla}(f)_r = \frac{\partial f}{\partial r} \quad ; \quad \vec{\nabla}(f)_\theta = \frac{1}{r}\frac{\partial f}{\partial \phi} \quad ; \quad \vec{\nabla}(f)_\phi = \frac{1}{r\sin\theta}\frac{\partial f}{\partial \theta} \tag{15.190}$$

Ainsi l'opérateur "gradient en coordonnées sphériques" s'exprime comme :

$$\boxed{\vec{\nabla} = \left(\frac{\partial}{\partial r}, \frac{1}{r}\frac{\partial f}{\partial \phi}, \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}\right)} \tag{15.191}$$

Nous avons donc finalement vu toutes les expressions du gradient dans les systèmes cartésiens, polaires, cylindriques et sphériques.

15.9.2 Gradient d'un champ de vecteurs

Le gradient d'un champ vectoriel $\vec{f}(x,y,z)$ est le champ dit "champ tensoriel" défini par les neuf relations suivantes en coordonnées cartésiennes :

$$\vec{\nabla}(\vec{f}) = \frac{\partial f_i}{\partial x_j} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \end{bmatrix} \tag{15.192}$$

Nous utiliserons un tel gradient lors de notre étude dans le chapitre de Génie Météo de l'effet Papillon dont l'origine vient de la détermination des équations de Navier-Stokes en Mécanique des Milieux Continus.

Par les quatre relations suivantes en coordonnées polaires :

$$\vec{\nabla}(\vec{f}) = \begin{bmatrix} \frac{\partial f_1}{\partial r} & \frac{1}{r}\frac{\partial f_1}{\partial \phi} \\ \frac{\partial f_2}{\partial r} & \frac{1}{r}\frac{\partial f_2}{\partial \phi} \end{bmatrix} \tag{15.193}$$

Par les neuf relations suivantes en coordonnées cylindriques :

$$\vec{\nabla}(\vec{f}) \begin{bmatrix} \frac{\partial f_1}{\partial r} & \frac{1}{r}\frac{\partial f_1}{\partial \phi} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial r} & \frac{1}{r}\frac{\partial f_2}{\partial \phi} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial r} & \frac{1}{r}\frac{\partial f_3}{\partial \phi} & \frac{\partial f_3}{\partial z} \end{bmatrix} \tag{15.194}$$

Par les neuf relations suivantes en coordonnées sphériques :

$$\vec{\nabla}(\vec{f}) \begin{bmatrix} \frac{\partial f_1}{\partial r} & \frac{1}{r}\frac{\partial f_1}{\partial \theta} & \frac{1}{r\sin\theta}\frac{\partial f_1}{\partial \phi} \\ \frac{\partial f_2}{\partial r} & \frac{1}{r}\frac{\partial f_2}{\partial \theta} & \frac{1}{r\sin\theta}\frac{\partial f_2}{\partial \phi} \\ \frac{\partial f_3}{\partial r} & \frac{1}{r}\frac{\partial f_3}{\partial \theta} & \frac{1}{r\sin\theta}\frac{\partial f_3}{\partial \phi} \end{bmatrix} \tag{15.195}$$

Nous avons donc finalement vu toutes les expressions du gradient d'un champ vectoriel dans les systèmes cartésiens, polaires, cylindriques et sphériques.

15.9.3 Divergences d'un champs de vecteurs

La divergence s'applique à un champ de vecteurs et le transforme en un champ de scalaires. Intuitivement, et dans les cas le plus courant, la divergence d'un champ vectoriel exprime sa tendance à provenir ou converger vers certains points.

Cependant, il faut distinguer rigoureusement deux contributions à la divergence que nous définirons rigoureusement un peu loin : l'une due aux variations de direction appelée la "divergence directionnelle" et l'autre due aux variations de modules (norme) appelée la "divergence modulaire". Ainsi, pour des champs simples, nous pouvons imaginer des cas où la divergence ne serait que modulaire et d'autres, où elle ne serait que directionnelle. Nous pourrions aussi construire un champ où les deux type de divergence coexistent, mais d'effets contraintes (convergence modulaire et divergence directionnelle par exemple).

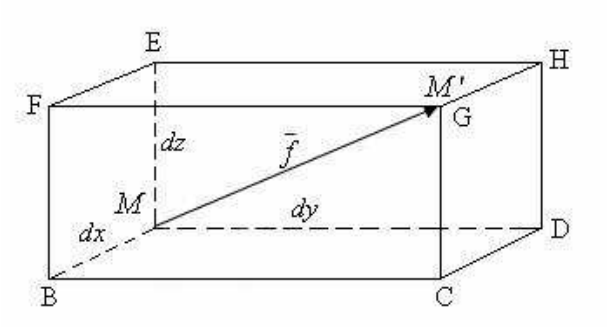


FIGURE 15.10 – Déplacement de $\vec{f}$ à travers un parallélépipède rectangle de côtés dx, dy et dz

Prenons par exemple un vecteur $\vec{f}$ de l'espace et faisons lui traverser une surface S quelconque. Les physiciens assimilent alors la quantité $\vec{f}$ qui se dirige suivant la normale à la surface au flux de $\vec{f}$ à travers S .

Pour se convaincre de cette analogie nous pouvons imaginer un fluide coulant sur une surface plane, le flux à travers la surface est évidemment nul, par contre si le fluide coule verticalement à travers une surface horizontale le flux sera maximal. Il est alors immédiat de vouloir représenter le flux par le produit scalaire de $\vec{f}$ avec la normale $\vec{n}$ de la surface S.

REMARQUE Il faut toujours prendre garde à la direction de $\vec{n}$ car en un point quelconque d'une surface on a en général deux normales.

Si la surface est plane la normale est la même partout mais si elle change suivant les endroits, nous nous intéresserons alors à un petit élément de surface ds.

Si un petit élément de flux est défini par :

$$d\Phi = (\vec{f} \circ \vec{n})ds \tag{15.196}$$

alors le flux total sera donné par :

$$\Phi = \int_S (\vec{f} \circ \vec{n})ds \tag{15.197}$$

ce qui est parfois noté (c'est un peu abusif mais pourquoi pas...) :

$$\Phi = \int_S \vec{f} d\vec{s} \tag{15.198}$$

Supposons maintenant que notre vecteur $\vec{f}$ déplace un point $M(x, y, z)$ de l'espace en $M'(x + dx, y+dy, z+dz)$ à travers un parallélépipède rectangle de côtés dx, dy et dz (cfr figure 15.10). Nous pouvons décomposer le mouvement (flux) à travers chaque face du parallélépipède (décompositions dans la base orthonormée). Par exemple, si nous nous intéressons à l'élément décomposé du flux à travers la face (dy, dz) décrite par les sommets $BCFG$ nous avons bien évidemment $\vec{n} = (1, 0, 0)$.

Il nous faut encore déterminer comment représenter le flux $\vec{f}$ pour cette direction. Comme le flux est une fonction, c'est-à-dire que chacune de ces composantes peut être dépendante des trois composantes de l'espace (si nous prenons le cas particulier d'une fonction dans $\mathbb{R}^3$) nous avons :

$$\vec{f} = \begin{pmatrix} f_x(x, y, z) \\ f_y(x, y, z) \\ f_z(x, y, z) \end{pmatrix} \tag{15.199}$$

REMARQUE Ceux qui ne sont pas convaincus peuvent aller lire le début du chapitre d'électrodynamique où nous prenons le champ électrique comme (excellent) exemple.

Alors la variation du flux selon x sera donnée par :

$$\vec{f}_x = \begin{pmatrix} f_x(x + dx, y, z) \\ f_y(x + dx, y, z) \\ f_z(x + dy, y, z) \end{pmatrix} - \begin{pmatrix} f_x(x, y, z) \\ f_y(x, y, z) \\ f_z(x, y, z) \end{pmatrix} \tag{15.200}$$

ce qui nous donne :

$$d\Phi_x = (\vec{(f)}_x \circ \vec{n}_x)dydz = \Big(f_x(x+dx,y,z) - f_x(x,y,z)\Big)dydz = \frac{\partial f_x}{\partial x}dxdydz \tag{15.201}$$

De même pour les deux autres faces :

$$\begin{aligned} d\Phi_y &= (\vec{(f)}_y \circ \vec{n}_y)dxdz = \Big(f_y(x,y+dy,z) - f_x(x,y,z)\Big)dxdz = \frac{\partial f_y}{\partial y}dxdydz \\ d\Phi_z &= (\vec{(f)}_z \circ \vec{n}_z)dxdy = \Big(f_z(x,y,z+dz) - f_x(x,y,z)\Big)dxdy = \frac{\partial f_z}{\partial z}dxdydz \end{aligned} \tag{15.202}$$

d'où en sommant :

$$d\Phi = \left(\frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} + \frac{\partial f_z}{\partial z}\right)dxdydz = div(\vec{f})dV = \vec{\nabla} \circ \vec{f}dV \tag{15.203}$$

Par rapport à la première expression de Φ, le terme $dxdydz$ est donc un élément de volume et non plus de surface. Nous avons aussi un résultat intéressant :

$$d\Phi = \int_V \left(\vec{\nabla} \circ \vec{f}\right) dV = \int_V \left(\vec{f} \circ \vec{n}\right) dS \tag{15.204}$$

Remarque Voir les exemples pratiques dans le chapitre d'électrodynamique où par exemple pour le champ électrique la divergence est nulle pour un charge sphérique libre car les vecteurs pointent dans des directions différentes (divergence modulaire) et les modules décroissent comme l'inverse du carrée du rayon (convergence modulaire). Les deux contributions sont oppositions et donc la divergence totale est nulle.

Le développement ci-dessus est appelé "théorème d'Ostrogradsky" ou "théorème de Gauss-Ostrogradsky" ou encore "théorème de la divergence" et définit en fait la divergence totale de $\vec{f}$ dans un volume comme le flux de $\vec{f}$ à travers les parois du volume (surface Gauss), ce qu'exprime bien le nom divergence.

Nous définissons "l'opérateur divergence" par la relation suivante (la notation tensorielle a été utilisée afin d'abréger l'écriture) dans un espace à n dimensions :

$$div(\vec{f}) = \vec{\nabla} \circ \vec{f} = \sum_{i=0}^{n} \frac{\partial f_i}{\partial x_i} \tag{15.205}$$

Ainsi, nous avec pour l'opérateur "divergence en coordonnées cartésiennes" :

$$\boxed{\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)} \tag{15.206}$$

Si la divergence d'un champ de vecteurs est identiquement nulle en tous les points d'un repère Eulérien, l'intégrale triple du flux de ce champs à travers un volume V sera :

$$\Phi = \iiint_V (\vec{\nabla} \circ \vec{f})dV = 0 \tag{15.207}$$

Il en résulte que le flux de ce champ de vecteurs à travers les bords du volume est nul, c'est-à-dire que le flux entrant compense le flux sortant. Nous disons qu'un tel champ de vecteurs de divergence nulle présente un flux conservatif.

Pour déterminer l'expression de la divergence en coordonnées polaires rappelons les relations démontrées plus haut :

$$\begin{aligned} \vec{e}_r &= \frac{\vec{r}}{r} = \frac{x\vec{e}_x + y\vec{e}_y}{r} = \vec{e}_x \cos\phi + \vec{e}_y \sin\phi \\ \vec{e}_\phi &= \frac{\vec{e}_z \times \vec{e}_r}{\sin\theta} = -\vec{e}_x \sin\phi + \vec{e}_y \cos\phi \end{aligned} \tag{15.208}$$

Soit à présent $\vec{f}:\mathbb{R}^2\to\mathbb{R}^2$ une fonction vectorielle. Nous avons :

$$\vec{f}=f_x\vec{e}_x+f_y\vec{e}_y=f_r\vec{e}_r+f_\theta\vec{e}_\theta \tag{15.209}$$

Connaissant l'expression de $\vec{e}_r,\vec{e}_\theta$ en fonction de $\vec{e}_x,\vec{e}_y$, à partir de l'expression ci-dessus nous en déduisons :

$$\begin{aligned}f_x&=\cos\phi\cdot f_r-\sin\phi\cdot f_\phi\\ f_y&=\sin\phi\cdot f_r+\cos\phi\cdot f_\phi\end{aligned} \tag{15.210}$$

La divergence de $\vec{f}$ est définie par $\vec{\nabla}\circ\vec{f}=\frac{\partial f_x}{\partial x}+\frac{\partial f_y}{\partial y}$. Nous avons :

$$\vec{\nabla}\circ\vec{f}=\vec{\nabla}(f_x)\circ\vec{e}_x+\vec{\nabla}(f_y)\circ\vec{e}_y \tag{15.211}$$

Le premier terme vaut (application du gradient en coordonnées cylindriques) :

$$\begin{aligned}\vec{\nabla}(f_x)\circ\vec{e}_x&=\left(\frac{\partial f_x}{\partial r}\vec{e}_r+\frac{1}{r}\frac{\partial f_x}{\partial\phi}\vec{e}_\phi\right)\circ\vec{e}_x\\ &=\frac{\partial f_x}{\partial r}\vec{e}_r\circ\vec{e}_x+\frac{1}{r}\frac{\partial f_x}{\partial\phi}\vec{e}_\phi\circ\vec{e}_x\\ &=\frac{\partial f_x}{\partial r}\cos\phi-\frac{1}{r}\frac{\partial f_x}{\partial\phi}\sin\phi\end{aligned} \tag{15.212}$$

de la même façon nous obtenons (nous pouvons détailler sur demande) :

$$\begin{aligned}\vec{\nabla}(f_y)\circ\vec{e}_y&=\frac{\partial f_y}{\partial r}\vec{e}_r\circ\vec{e}_y+\frac{1}{r}\frac{\partial f_y}{\partial\phi}\vec{e}_\phi\circ\vec{e}_y\\ &=\frac{\partial f_y}{\partial r}\sin\phi+\frac{1}{r}\frac{\partial f_y}{\partial\phi}\cos\phi\end{aligned} \tag{15.213}$$

En additionnant les trois termes et en exprimant les dérivées partielles des fonctions en fonction des dérivées partielles des fonctions f_x,f_y à l'aide des relations :

$$\begin{aligned}f_x&=\cos\phi\cdot f_r-\sin\phi\cdot f_\phi\\ f_y&=\sin\phi\cdot f_r+\cos\phi\cdot f_\phi\end{aligned} \tag{15.214}$$

Nous obtenons :

$$\begin{aligned}\vec{\nabla}\circ\vec{f}=&\frac{\partial f_x}{\partial r}\cos\phi-\frac{1}{r}\frac{\partial f_x}{\partial\phi}\sin\phi+\frac{\partial f_y}{\partial r}\sin\phi+\frac{1}{r}\frac{\partial f_y}{\partial\phi}\cos\phi\\ =&\cos^2\phi\frac{\partial f_r}{\partial r}-\cos\phi\sin\phi\frac{\partial f_\phi}{\partial r}-\frac{1}{r}\sin\phi\left(-\sin\phi f_r+\cos\phi\frac{\partial f_r}{\partial\phi}-\cos\phi f_\phi-\sin\phi\frac{\partial f_\phi}{\partial\phi}\right)\\ &+\sin^2\phi\frac{\partial f_r}{\partial r}+\sin\phi\cos\phi\frac{\partial f_\phi}{\partial r}+\frac{1}{r}\cos\phi\left(\cos\phi f_r+\sin\phi\frac{\partial f_r}{\partial\phi}-\sin\phi f_\phi+\cos\phi\frac{\partial f_\phi}{\partial\phi}\right)\end{aligned} \tag{15.215}$$

Après simplification :

$$\vec{\nabla}\circ\vec{f}=\frac{\partial f_r}{\partial r}+\frac{1}{r}f_r+\frac{1}{r}\frac{\partial f_\phi}{\partial\phi}=\frac{1}{r}\frac{\partial(rf_r)}{\partial r}+\frac{1}{r}\frac{\partial f_\phi}{\partial\phi} \tag{15.216}$$

L'expression de l'opérateur "divergence en coordonnées polaires" est alors :

$$\boxed{\vec{\nabla}=\left(\frac{1}{r}\frac{\partial(r}{\partial r},\frac{1}{r}\frac{\partial}{\partial\phi}\right)} \tag{15.217}$$

Pour déterminer l'expression de l'opérateur divergence en coordonnées cylindriques rappelons les relations :

$$\begin{aligned}\vec{e}_r&=\frac{\vec{r}}{r}=\frac{x\vec{e}_x+y\vec{e}_y}{r}=\vec{e}_x\cos\phi+\vec{e}_y\sin\phi\\ \vec{e}_\phi&=\frac{\vec{e}_z\times\vec{e}_r}{\sin\theta}=-\vec{e}_x\sin\phi+\vec{e}_y\cos\phi\\ \vec{e}_z&=\vec{e}_z\end{aligned} \tag{15.218}$$

Soit à présent $\vec{f}:\mathbb{R}^3\to\mathbb{R}^3$ une fonction vectorielle. Nous avons :

$$\vec{f}=f_x\vec{e}_x+f_y\vec{e}_y+f_z\vec{e}_z=f_r\vec{e}_r+f_\phi\vec{e}_\phi+f_z\vec{e}_z \tag{15.219}$$

Connaissant l'expression de $\vec{e}_r,\vec{e}_\phi,\vec{e}_z$ en fonction de $\vec{e}_x,\vec{e}_y,\vec{e}_z$, à partir de l'expression ci-dessus nous en déduisons :

$$\begin{aligned}f_x&=\cos\phi\cdot f_r-\sin\phi\cdot f_\phi\\f_y&=\sin\phi\cdot f_r+\cos\phi\cdot f_\phi\\f_z&=f_z\end{aligned} \tag{15.220}$$

La divergence de $\vec{f}$ est définie par $\vec{\nabla}\circ\vec{f}=\frac{\partial f_x}{\partial x}+\frac{\partial f_y}{\partial y}+\frac{\partial f_z}{\partial z}$. Nous avons :

$$\vec{\nabla}\circ\vec{f}=\vec{\nabla}(f_x)\circ\vec{e}_x+\vec{\nabla}(f_y)\circ\vec{e}_y+\vec{\nabla}(f_z)\circ\vec{e}_z \tag{15.221}$$

Le premier terme vaut (application du gradient en coordonnées cylindriques) :

$$\begin{aligned}\vec{\nabla}(f_x)\circ\vec{e}_x&=\left(\frac{\partial f_x}{\partial r}\vec{e}_r+\frac{1}{r}\frac{\partial f_x}{\partial\phi}\vec{e}_\phi+\frac{\partial f_x}{\partial z}\vec{e}_z\right)\circ\vec{e}_x\\&=\frac{\partial f_x}{\partial r}\vec{e}_r\circ\vec{e}_x+\frac{1}{r}\frac{\partial f_x}{\partial\phi}\vec{e}_\phi\circ\vec{e}_x+\frac{\partial f_x}{\partial z}\vec{e}_z\circ\vec{e}_x\\&=\frac{\partial f_x}{\partial r}\cos\phi-\frac{1}{r}\frac{\partial f_x}{\partial\phi}\sin\phi\end{aligned} \tag{15.222}$$

de la même façon nous obtenons (nous pouvons détailler sur demande) :

$$\begin{aligned}\vec{\nabla}(f_y)\circ\vec{e}_y&=\frac{\partial f_y}{\partial r}\vec{e}_r\circ\vec{e}_y+\frac{1}{r}\frac{\partial f_y}{\partial\phi}\vec{e}_\phi\circ\vec{e}_y+\frac{\partial f_y}{\partial z}\vec{e}_z\circ\vec{e}_y\\&=\frac{\partial f_y}{\partial r}\sin\phi+\frac{1}{r}\frac{\partial f_y}{\partial\phi}\cos\phi\end{aligned} \tag{15.223}$$

et :

$$\begin{aligned}\vec{\nabla}(f_z)\circ\vec{e}_z&=\frac{\partial f_z}{\partial r}\vec{e}_r\circ\vec{e}_z+\frac{1}{r}\frac{\partial f_z}{\partial\phi}\vec{e}_\phi\circ\vec{e}_z+\frac{\partial f_z}{\partial z}\vec{e}_z\circ\vec{e}_z\\&=\frac{\partial f_z}{\partial z}\end{aligned} \tag{15.224}$$

En additionnant les trois termes et en exprimant les dérivées partielles des fonctions $\int_x,\int_y,\int_z$ en fonction des dérivées partielles des fonctions $\int_r,\int_\phi,\int_z$ à l'aide des relations :

$$\begin{aligned}f_x&=\cos\phi\cdot f_r-\sin\phi\cdot f_\phi\\f_y&=\sin\phi\cdot f_r+\cos\phi\cdot f_\phi\\f_z&=f_z\end{aligned} \tag{15.225}$$

Nous obtenons :

$$\begin{aligned}\vec{\nabla}\circ\vec{f}=&\frac{f_x}{\partial r}\cos\phi-\frac{1}{r}\frac{f_x}{\partial\phi}\sin\phi+\frac{f_y}{\partial r}\sin\phi+\frac{1}{r}\frac{f_y}{\partial\phi}\cos\phi+\frac{f_z}{\partial z}\\=&\cos^2\phi\frac{f_r}{\partial r}-\cos\phi\sin\phi\frac{f_\phi}{\partial r}-\frac{1}{r}\sin\phi\left(-\sin\phi f_r+\cos\phi\frac{f_r}{\partial\phi}-\cos\phi f_\phi-\sin\phi\frac{f_\phi}{\partial\phi}\right)\\&+\sin^2\phi\frac{f_r}{\partial r}+\sin\phi\cos\phi\frac{f_\phi}{\partial r}+\frac{1}{r}\cos\phi\left(\cos\phi f_r+\sin\phi\frac{f_r}{\partial\phi}-\sin\phi f_\phi+\cos\phi\frac{f_\phi}{\partial\phi}\right)+\frac{f_z}{\partial z}\end{aligned} \tag{15.226}$$

Après simplification :

$$\vec{\nabla}\circ\vec{f}=\frac{f_r}{\partial r}+\frac{1}{r}f_r+\frac{1}{r}\frac{f_\phi}{\partial\phi}+\frac{f_z}{\partial z}\qquad=\frac{1}{r}\frac{\partial(rf_r)}{\partial r}+\frac{1}{r}\frac{f_\phi}{\partial\phi}+\frac{f_z}{\partial z} \tag{15.227}$$

L'expression de l'opérateur "divergence en coordonnées cylindriques" est alors :

$$\vec{\nabla} = \left(\frac{1}{r} \frac{\partial (r}{\partial r}, \frac{1}{r} \frac{\partial}{\partial \phi}, \frac{\partial}{\partial z} \right) \tag{15.228}$$

Pour obtenir l'expression de la divergence en coordonnes sphériques, rappelons les relations :

$$\begin{aligned} \vec{e}_r &= \sin\theta \cos\phi \vec{e}_x + \sin\theta \sin\phi \vec{e}_y + \cos\theta \vec{e}_z \\ \vec{e}_\phi &= -\vec{e}_x \sin\phi + \vec{e}_y \cos\phi \\ \vec{e}_\theta &= -\vec{e}_x \cos\theta \cos\phi - \vec{e}_y \cos\theta \sin\phi - \vec{e}_z \sin\theta \end{aligned} \tag{15.229}$$

Soit à présent $\vec{f} : \mathbb{R}^3 \to \mathbb{R}^3$ une fonction vectorielle. Nous avons :

$$\vec{f} = \vec{f}_x \vec{e}_x + \vec{f}_y \vec{e}_y + \vec{f}_z \vec{e}_z = \vec{f}_r \vec{e}_r + \vec{f}_\theta \vec{e}_\theta + \vec{f}_\phi \vec{e}_\phi \tag{15.230}$$

Connaissant l'expression de $\vec{e}_r, \vec{e}_\theta, \vec{e}_\phi$ en fonction de $\vec{e}_x, \vec{e}_y, \vec{e}_z$, à partir de l'expression ci-dessus nous en déduisons :

$$\begin{aligned} f_x &= \sin\theta \cos\phi \cdot f_r + \cos\theta \sin\phi \cdot f_\theta - \sin\phi \cdot f_\phi \\ f_y &= \sin\theta \sin\phi \cdot f_r + \cos\theta \sin\phi \cdot f_\theta + \cos\phi \cdot f_\phi \\ f_z &= \cos\theta \cdot f_r - \sin\theta \cdot f_\theta \end{aligned} \tag{15.231}$$

La divergence de est définie par . Nous avons :

(229)

Le premier terme vaut (application du gradient en coordonnées sphériques) :

(230)

de la même façon nous obtenons (nous pouvons détailler sur demande) :

(231)

et :

(232)

En additionnant les trois termes et en exprimant les dérivées partielles des fonctions en fonction des dérivées partielles des fonctions à l'aide des relations :

(233)

nous obtenons (nous pouvons développer les détails intermédiaires sur demande) :

(234)

Ainsi, l'expression de la divergence en coordonnées sphériques devient :

(235)

et donc l'opérateur de "divergence en coordonnées sphériques" est alors :

(236)

Nous avons donc finalement vu toutes les expressions de la divergence d'un champ vectoriel dans les systèmes cartésiens, polaires, cylindriques et sphériques. RotationnelS d'un champ de vecteurs

Le rotationnel d'un champ de vecteurs peut être vu (c'est une simplification !) comme le champ de vecteurs dont les lignes de champs sont perpendiculaires à celles dont nous avons calculé le rotationnel comme le montre l'exemple particulier ci-dessous :

(237)

Le rotationnel transforme ainsi un champ de vecteurs en un autre champ de vecteurs. Plus difficile à se représenter précisément que le gradient et la divergence, il exprime intuitivement la tendance qu'a un champ à tourner autour d'un point (la manière dont il est tordu).

Exemples :

E1. Dans une tornade, le vent tourne autour de l'oeil du cyclone et le champ vectoriel vitesse du vent a un rotationnel non nul autour de l'oeil.

E2. Le rotationnel du champ des vitesses d'un disque qui tourne à vitesse constante est constant, dirigé selon l'axe de rotation et orienté de telle sorte que la rotation ait lieu, par rapport à lui, dans le sens direct.

Un champ de vecteurs est dit "irrotationnel" lorsque le rotationnel de ce champ est identiquement nul en tous les points de l'espace.

a divergence donne certaines indications sur le comportement d'un vecteur ou d'un champ de vecteurs : comment il se dirige par rapport à la normale et comment il traverse les surfaces, mais

c'est insuffisant. Prenons un champ qui aurait la forme d'un cylindre et un autre champ qui aurait la forme d'une hélice de même diamètre que le cylindre. S'ils se dirigent dans la même direction leur divergence sera identique alors que les mouvements sont bien différents. Il faut donc que nous déterminions la manière dont le champ est courbé quand il traverse une surface : ceci va être déterminé par la circulation (comme le travail d'une force par exemple) du vecteur le long d'une courbe fermée, obtenue avec la somme des produits scalaires sur le contour :

(238)

en fait ça revient au même de regarder comment est tordu le vecteur par rapport à la normale à la surface ce qui nous amène à définir le "rotationnel" ou "vecteur tourbillon" en écrivant :

(239)

qui établit donc une relation entre l'intégrale curviligne et l'intégrale de surface (on transforme donc une intégrale curviligne sur un parcours fermé en une intégrale de surface délimitée par ledit parcours).

En d'autres termes, le rotationnel se calcule en utilisant le fait que la circulation autour d'un circuit élémentaire fermé d'un champ de vecteurs est égal au flux de son rotationnel à travers la surface élémentaire immédiate engendrée par ce circuit.

Ceci est le "théorème de Stokes" (qui est plus rigoureusement démontrable avec un formalisme mathématique assez lourd) qui est donc en fait une définition de l'opérateur rotationnel dont nous allons chercher l'expression mathématique explicite :

Soit un champ vectoriel défini dans un espace donné. Nous voulons donc calculer la circulation du autour d'un contour fermé C :

(240)

Nous choisissons comme contour C le contour d'un rectangle infinitésimal de côté plong dans et parallèle au plan xOy (remarquez que nous parcourons le contour de façon a toujours avoir la surface à notre gauche) :

(241)

Pour les deux côtés horizontaux, la contribution à la circulation est :

(242)

Ce qui nous autorise à écrire :

(243)

De même, pour les faces verticales :

(244)

Ainsi, nous avons la circulation selon z :

(245)

Ce qui s'écrit aussi sous la forme générale traditionnelle suivante :

(246)

et constitue le non moins fameux "théorème de Green" ou appelé encore "théorème de Green-Riemann".

Et que nous écrirons dans le cas qui nous intéresse :

(247)

Par permutation circulaire nous obtenons alors :

(248)

Soit sous forme vectorielle condensée :

(249)

Ce qui permet de mieux comprendre la notation, ou la définition non intuitive du rotationnel dans beaucoup d'ouvrages et qui est :

(250)

soit le produit vectoriel de l'opérateur gradient avec le champ vectoriel.

Donc nous avons finalement démontré le théorème de Stokes qui donne bien :

(251)

et en même temps le rotationnel en coordonnées cartésiennes.

Cherchons maintenant à déterminer l'expression du rotationnel en coordonnées cylindriques ((le rotationnel en coordonnées polaires n'étant pas définissable).

En réutilisant la même technique que pour le rotationnel en coordonnées cylindriques, nous écrivons la circulation de le long d'un contour correspondant à un petit morceau de cylindre orthogonal à (Oz) : .

(252)

Nous avons alors en fixant z (attention le n'a rien à voir avec le r du rayon du cylindre... la notation peut être confuse nous en sommes désolé !) :

(253)

la circulation totale donne donc après regroupement des termes :

(254)

Nous ne pouvons pas à cette étape directement comparer avec le rotationnel car il nous est difficile de faire apparaître la différentielle de la surface si nous regardons les différentielles qui apparaissent actuellement dans la circulation. Le mieux est alors tout diviser par :

(255)

Donc :

(256)

Maintenant nous déterminons le rotationnel en fixant . Le problème revient à avoir donc un rectangle dans l'espace que nous parcourons pour déterminer la circulation. Or, nous savons déjà que est le résultat du rotationnel pour un rectangle en coordonnées cartésiennes :

(257)

à la différence que dans les coordonnées cylindrique il faut substituer z par , x par z, y par r, par et finalement par (ce choix s'impose toujours simplement parce que le circulation se fait de telle manière que la surface soit toujours à notre gauche) . Ce qui nous donne :

(258)

Il ne nous reste plus qu'à trouver la composante du rotationnel en r (soit quand r est fixé). Le calcul est alors plus délicat puisqu'il s'agit de parcourir (positivement toujours !) une surface courbée par la variation de l'angle .

Nous avons alors en fixant r :

(259)

la circulation totale donne donc après regroupement des termes :

(260)

Nous ne pouvons pas à cette étape directement comparer avec le rotationnel car il nous est difficile de faire apparaître la différentielle de la surface si nous regardons les différentielles qui apparaissent actuellement dans la circulation. Le mieux est alors tout diviser par :

(261)

Donc finalement :

(262)

Et finalement le rotationnel en coordonnées cylindriques dans sa globalité est donné par :

(263)

Le lecteur pourrai aisément vérifier que ce résultat est simplement le gradient en coordonnées cylindriques appliqué au champ vectoriel .

Pour s'en persuader, montrons maintenant directement l'expression du rotationnel en coordonnées sphériques directement en montrant ceci via le produit vectoriel du gradient en coordonnées sphériques avec le champ vectoriel .

D'abord rappelons que nous avons obtenu pour le gradient en coordonnées sphériques :

(264)

Donc il vient :

(265)

ce que nous pouvons aussi écrire en décomposant à l'aide des vecteurs de base :

(266)

A l'aide des dérivées partielles que nous avions démontrées lors de notre introduction plus haut du système de coordonnées sphériques il vient :

(267)

Les produits vectoriels avec le vecteur colinéaires s'annulent. Il reste donc :

(268)

Comme le produit vectoriel de deux vecteurs de base donne le vecteur orthogonal correspondant (positivement ou négativement) nous avons alors :

(269)

En regroupant les termes il vient :

(270)

Soit en simplifiant :

(271)

Soit finalement :

(272) LAPLACIENS D'UN CHAMP SCALAIRE

Le laplacien d'un champ scalaire est le champ scalaire qui mesure la différence entre la valeur de la fonction en un point et sa moyenne autour de ce point. En d'autres termes, la dérivée partielle deuxième mesure les variations de la pente au point étudiée dans un entourage immédiat et selon une dimension à la fois. Si la dérivée partielle deuxième est nulle selon x, alors la pente est constante dans un entourage immédiat et selon cette dimension, cela implique que la valeur de la fonction au point étudié est la moyenne de son entourage (selon une dimension)

Cet opérateur s'obtient à partir de la divergence du gradient et nous la notons (écriture tensorielle) :

(273)

Le laplacien est nul, ou assez petit, lorsque la fonction varie sans à-coup. Les fonctions vérifiant l'équation de Laplace sont appelées "fonctions harmoniques".

Donc l'opérateur "laplacien en coordonnées cartésiennes" est :

(274)

Le laplacien d'un champ scalaire et dans d'autres systèmes de coordonnées est un peu plus long à développer. Il existe plusieurs méthodes et parmi celles existantes j'ai choisi celles dont le type de raisonnement et les outils utilisés semblaient pertinents. Il est intéressant d'aborder différentes stratégies mais bien sûr il existe des méthodes plus simples que celle présentée ci-dessous.

Soit le laplacien en coordonnées cartésiennes dans d'un champ scalaire f :

(275)

Pour déterminer cette expression en coordonnées polaires, nous allons utiliser la différentielle totale et la règle de chaîne en coordonnées polaires :

(276)

donc pour une dérivée seconde :

(277)

or, nous avons pour les coordonnées polaires :

et (278)

d'où :

et

et (279)

d'où :

(280)

et compte tenu que les dérivées partielles secondes sont continues, alors les dérivées croisées sont égales selonle théorème de Schwarz (cf. chapitre de Dalcul Différentiel Et Intégral) :

(281)

Donc :

(282)

De façon similaire, nous aurons :

(283)

d'où l'expression du laplacien en coordonnées polaires en sommant les deux dernères expressions :

(284)

Donc l'opérateur "laplacien en coordonnées polaires" est finalement donné par :

(285)

Pour trouver l'expression du laplacien en coordonnées sphériques, nous allons utiliser l'intuition du physicien et les notions de similitude.

Nous allons tout d'abord nous aider de la figure ci-dessus pour savoir de quoi l'on parle :

(286)

Rappelons que les relations entre coordonnées cartésiennes et sphériques sont données par les relations :

(287)

Nous allons considérer maintenant les similitudes suivantes :

Coordonnées cylindriques : et

Coordonnées sphériques : et

Construisons un tableau de correspondance :

(288)

L'objectif est de jouer avec cette correspondance avec d'abord le laplacien en coordonnées cylindriques où l'on a soustrait des deux côtés de l'égalité le terme . Ainsi :

(289)

utilisons le tableau de correspondance et nous obtenons :

(290)

Le deuxième terme de l'égalité de cette dernière relation est l'équivalent sphérique du terme n°1 du laplacien en coordonnées cylindriques :

(291)

Maintenant examinons le terme :

Identiquement lorsque nous avons déterminé la relation :

(292)

nous obtenons :

(293)

avec :

et (294)

ce qui nous permet d'écrire :

(295)

si nous jouons encore avec le tableau de correspondance, nous avons :

(296)

nous divisons cette relation des deux côtés par et ainsi nous obtenons :

(297)

Nous avons donc ci-dessus l'équivalent sphérique du terme n°2 du laplacien en coordonnées cylindriques :

(298)

Le troisième et dernier terme est très simple à déterminer. Nous remplacons par afin d'obtenir :

(299)

En rassemblant tous les termes obtenus précédemment, nous obtenons enfin la forme étendue du laplacien en coordonnées sphériques si utilisé en physique :

(300)

Nous pouvons raccourcir cette expression et factorisant les termes :

(301)

Si nous condensons encore un peu, nous obtenons l'expression finale de l'opérateur "laplacien en coordonnées sphériques" appelé aussi "laplacien sphérique" :

(302) LaplacienS d'un champ vectoriel

Le laplacien d'un champ vectoriel est le champ vectoriel défini par (notation tensorielle) :

(303)

dont les composantes sont les laplacien des composantes.

Ainsi, en coordonnées cartésiennes :

(304)

Le laplacien d'un champ de vecteurs, appelé fréquemment "laplacien vectoriel", en d'autres systèmes de coordonnées est assez simple à obtenir à partir de la connaissance du laplacien d'un champ scalaire dans ces mêmes coordonnées. Ainsi, en coordonnées polaires, nous avons pour le laplacien d'un champ vectoriel la relation suivante :

(305)

et en coordonnées cylindriques :

(306)

et finalement en coordonnées sphériques :

(307) IDENTITÉS

Les opérateurs différentiels scalaires et vectoriels ont des identités remarquables très simples que nous retrouverons très souvent en physique.

Voyons d'abord les relations qui n'ont aucun sens (au cas où vous tomberiez dessus sans faire exprès...) :

ou (308)

Le rotationnel d'une divergence n'existe pas puisque l'opérateur rotationnel s'applique à un champ vectoriel alors que la divergence est un scalaire.

ou (309)

Le rotationnel d'un laplacien scalaire n'existe pas puisque l'opérateur rotationnel s'applique à un champ vectoriel alors que par construction, le laplacien est un scalaire.

Voyons maintenant quelques propriétés remarquables sans démonstrations (cependant si vous en avez besoin car vous n'y arrivez pas seul, n'hésitez pas à nous contacter, nous compléterons) :

I1. Par construction le laplacien scalaire est la divergence du gradient du champ :

(310)

I2. Le rotationnel du gradient est nul :

(311)

I3. La divergence du rotationnel d'un champ vectoriel est toujours nulle :

ou (312)

Démonstration :

(313)

C.Q.F.D.

I4. Le rotationnel du rotationnel d'un champ vectoriel est égal au gradient de la divergence de ce champ moins son laplacien vectoriel :

ou (314)

Démonstration :

(315)

Il est ensuite facile de vérifier que cette dernière égalité est égale à :

(316)

C.Q.F.D.

I5. La multiplication de l'opérateur nabla par le produit scalaire de deux vecteurs est égal à... (voir ci-dessous), qui donne une relation très utile en mécanique des fluides :

(317)

I6. Le produit scalaire du rotationnel d'un vecteur est la différence des opérateurs commutés tel que :

(318) Nous réutiliserons cette dernière relation lors de notre étude en électromagnétisme de la pression de radiation (entre autres). RÉSUMÉ

Dans le cadre de ce site Internet, nous faisons usage des différentes notations présentées et résumées dans le tableau ci-dessous. Leur usage permet dans le cadre de différentes théories d'éviter des confusions avec d'autres êtres mathématiques. C'est embêtant certes mais il faudra faire avec.

ÊTRE MATHEMATIQUE

NOTATIONS

Gradient d'un champ scalaire

Gradient d'un champ vectoriel

Divergence d'un champ de vecteurs

Rotationnel d'un champ de vecteurs

rot()

Laplacien d'un champ scalaire

Laplacien d'un champ vectoriel

Et pour les pragmatique voici un résumé des explications des opérateurs les plus importants en physique :

- Le gradient signifie "la pente" (exemple : le champ électrique est la pente du potentiel électrostatique).

Les différentes expressions du gradient (mises sous la forme de l'opérateur nabla) en coordonnées cartésiennes, polaires, cylindriques et sphériques sont les suivantes :

(319)

(320)

(321)

(322)

- La divergence caractérise un flux de quelque chose qui vient de quelque part, d'une source, ou qui y va. Si la divergence n'est pas nulle, c'est qu'il y a concentration autour d'un point, donc la densité augmente (ou diminue, c'est selon le signe). Ca peut être la densité de charges électriques ou bien la masse volumique. D'où le fameux théorème qui dit que le flux (ce qui passe dans une surface) est égal à l'intégrale de la divergence (ce qui reste).

Les différentes expressions de la divergence (mises sous la forme de l'opérateur nabla) en coordonnées cartésiennes, polaires, cylindriques et sphériques sont les suivantes :

(323)
(324)
(325)
(326)

- Le rotationnel caractérise l'existence d'un tourbillon (très utilisé en mécanique des fluides). S'il y a un tourbillon, on peut suivre une ligne de courant sur une courbe fermée sans qu'elle change de sens : la circulation ne sera pas nulle (elle vaut l'intégrale du rotationnel).

Les différentes expressions du rotationnel en coordonnées cartésiennes, cylindriques et sphériques sont les suivantes :

- Le laplacien d'un champ scalaire est le champ scalaire qui mesure la différence entre la valeur de la fonction en un point et sa moyenne autour de ce point. En d'autres termes, la dérivée partielle deuxième mesure les variations de la pente au point étudiée dans un entourage immédiat et selon une dimension à la fois. Si la dérivée partielle deuxième est nulle selon une direction, alors la pente est constante dans un entourage immédiat et selon cette dimension, cela implique que la valeur de la fonction au point étudié est la moyenne de son entourage (selon une dimension).

Les différentes expressions du laplacien (mises sous la forme de l'opérateur nabla) en coordonnées cartésiennes, polaires et sphériques sont les suivantes :

(327)
(328)
(329)

CHIMIE

16 Chimie quantique 483
Puits de potentiel rectangulaire tridimensionnel infini 483 ● Vibrations moléculaires 485 ● Atome hydrogénoïde 487

17 Chimie moléculaire 497
Approximation orbitalaire 498 ● Méthode C.L.O.A. 500

18 Chimie analytique 505
Mélanges simples 505 ● Réactions 506

19 Chimie thermique 509
Transformations chimiques 509 ● Grandeurs molaires 510 ● Enthalpie standard de réaction 513

16

Chimie quantique

Avant que le lecteur aille plus loin dans la lecture de cette section du site, nous souhaitons rappeler que celui-ci traite de mathématique appliquée et de physique théorique. Ainsi, nous ne traiterons dans cette section que de chimie théorique (chimique quantique, thermochimie, chimie cinétique, ...).

Ce choix fait suite à l'évolution du visage de la chimie ces vingt dernières décennies. De science en grande partie descriptive, elle tend à devenir déductive : à côté de l'expérience, la place du calcul grandit constamment. La chimie théorique, appelée également "chimie physique", application des méthodes de la physique à la chimie, trop souvent encore considérée comme une discipline en soi, imprègne toute la chimie moderne : l'abord d'un problème quelconque exige son concours et le chimiste doit en posséder une connaissance approfondie. Or, au niveau de l'enseignement de la chimie comme branche secondaire, ce rôle de la chimie-physique se manifeste déjà : il en résulte un relèvement de ce niveau, une augmentation du degré d'abstraction, et un risque de rebuter l'étudiante moyen. Enfin, il ne s'agit pas de charger les connaissances en y incorporant plus d'éléments nouveaux, mais bien de convertir le mode d'approche de cette discipline en substituant le plus souvent l'énoncé de connaissances encyclopédiques avec un exposé raisonné des questions fondamentales.

Une bonne compréhension de la chimie nécessite selon nous obligatoirement un crochet obligatoire par la physique quantique (cf. section de physique atomique) afin d'avoir au moins une approche de ce qu'est un atome et des différentes orbites électroniques avant de parler de liaisons, des différentes méthodes de remplissage des orbites électroniques, d'oxydo-réduction, de remplissage des couches, et autres...

Dans ce sens, nous allons commencer par l'étude du cas particulier de l'atome d'hydrogène qui a une importance capitale pour toute la suite (étude des atomes polyélectroniques). Il convient donc de parcourir les prochaines lignes avec toute l'attention possible et d'en saisir le maximum de subtilités.

16.1 Puits de potentiel rectangulaire tridimensionnel infini

Nous avons étudié en physique quantique corpusculaire l'atome d'hydrogène en utilisant la mécanique relativiste de Bohr-Sommerfeld. De ce modèle il en est ressorti une quantification simpliste (mais pas trop fausse comme nous le verrons plus loin) de certaines propriétés de la matière.

En physique quantique ondulatoire, nous avons étudié le puits de potentiel rectiligne infini et l'oscillateur harmonique sans donner beaucoup plus d'exemple. Maintenant nous allons nous orienter dans la résolution de problèmes plus proches de ceux utiles en chimie dans l'objectif d'étudier l'atome hydrogénoïde.

Nous allons considérer maintenant une particule se mouvant librement dans la boîte tridimensionnelle ci-dessous.

equation (53.1)

L'énergie potentielle de ce système est donnée par :

$$(53.2)\quad \begin{array}{ll} E_{\text{pot}}(x,y,z)=0 & 0<x<a,\quad 0<y<b,\quad 0<z<c,\\ E_{\text{pot}}(x,y,z)=\infty & \text{ailleurs}\end{array} \tag{16.1}$$

Comme dans le cas unidimensionnel (voir section de physique atomique), les murs de potentiel infini empêchent la particule de quitter la boîte, et la fonction d'onde n'est non nulle que pour equation se trouvant à l'intérieur de la boîte. Elle s'annule nécessairement dès que l'un des murs est atteint. L'équation de SCHRÖDINGER que nous devons résoudre est donc :

$$\Delta\Psi + \frac{2m}{\hbar} - (E_{\text{tot}} - E_{\text{pot}})\Psi = 0 \tag{53.3}$$ (16.2)

et les conditions aux bornes se lisent :

$$\begin{aligned} \Psi(0,y,z) &= \Psi(a,y,z) = 0 \\ \Psi(x,0,z) &= \Psi(x,b,z) = 0 \\ \Psi(x,y,0) &= \Psi(x,y,c) = 0 \end{aligned} \tag{53.4}$$ (16.3)

Notons que l'Hamiltonien peut s'écrire comme la somme de l'Hamiltonien selon chaque axe (nous parlons des opérateurs bien évidemment !). Nous avons donc :

$$H = H_x + H_y + H_z \tag{53.5}$$ (16.4)

où :

$$H_x = \left(\frac{-\hbar^2}{2m}\right)\frac{\partial^2}{\partial^2 x}, \quad H_y = \left(\frac{-\hbar^2}{2m}\right)\frac{\partial^2}{\partial^2 y}, \quad H_z = \left(\frac{-\hbar^2}{2m}\right)\frac{\partial^2}{\partial^2 z} \tag{53.6}$$ (16.5)

relations dont nous avons démontré la provenance dans la section de physique quantique ondulatoire.

Une telle forme est dite "forme séparable" : l'Hamiltonien est la somme d'opérateurs individuels H_i chacun ne dépendant que d'une seule variable ou degré de liberté q_i. Cette forme traduit le caractère indépendant des mouvements décrits par les variables q_i.

Rappelons-nous que la probabilité conjointe de deux événements indépendants est le produit des probabilités individuelles des deux événements, pris séparément (cf. chapitre de Probabilités). Nous nous attendons donc à ce que la densité de probabilité de présence dans l'espace (cf. chapitre de Physique Quantique Ondulatoire) de configuration multidimensionnel soit, dans le cas où l'Hamiltonien est de forme séparable, un simple produit de densités de probabilités individuelles. En fait, la forme séparable de l'Hamiltonien permet une séparation de variables sur la fonction d'onde elle-même.

Écrivons donc les solutions de l'équation de SCHRÖDINGER sous la forme :

$$\Psi(x,y,z) = \xi(x)\theta(y)\zeta(z) \tag{53.7}$$ (16.6)

d'une produit de trois facteurs chacun dépendant d'une seule coordonnée.

Substituant cette écriture dans l'équation de SCHRÖDINGER on obtient sans développements :

$$\theta(y)\zeta(z)(H_x\xi(x)) + \xi(x)\zeta(z)(H_y\theta(y)) + \xi(x)\theta(y)(H_z\zeta(z)) = (E_{\text{tot}} - E_{\text{pot}})\xi(x)\theta(y)\zeta(z) \tag{53.8}$$ (16.7)

ou encore, en divisant les deux membres de ceci par $\xi(x)\theta(y)\zeta(z)$:

$$\frac{1}{\xi(x)}(H_x\xi(x)) + \frac{1}{\theta(y)}(H_y\theta(y)) + \frac{1}{\zeta(z)}(H_z\zeta(z)) = (E_{\text{tot}} - E_{\text{pot}}) \tag{53.9}$$ (16.8)

ce qui est une forme beaucoup plus esthétique et facile à mémoriser.

Cette équation demande que la somme des trois termes dans le membre de gauche soit égale à une constante dans le cadre d'un système conservatif (c'est ce qui intéresse souvent les chimistes) ! Chacun de ces trois termes ne dépendant que d'une et une seule variable, pour que leur somme soit égale à une constante, il faut que chaque terme soit lui-même constant. En effet, en prenant la dérivée des deux membres de la relation précédente par rapport à x, par exemple, nous avons :

$$\frac{\mathrm{d}\left(\frac{1}{\xi(x)}(H_x\xi(x))\right)}{\mathrm{d}x} = 0 \tag{53.10}$$ (16.9)

ce qui signifie que $\frac{1}{\xi(x)}(H_x\xi(x))$ doit bien être une constante et que nous noterons E_x puisque ce terme exprime une énergie. Nous avons alors :

$$H_x\xi(x) = E_x\xi(x) \tag{16.10}$$

De même, nous obtenons :

$$H_y\theta(y) = E_y\theta(y) \quad \text{et} \quad H_z\zeta(z) = E_z\zeta(z) \tag{16.11}$$

Notons que chacune des équations séparées que nous venons d'obtenir, pour le mouvement de la particule dans les trois directions spatiales est l'équation de SCHRÖDINGER dans une boîte unidimensionnelle. Ainsi, les trois relations obtenues précédemment décrivent chacune indépendamment le mouvement dans la direction respective x, y, z, limité aux intervalles respectifs :

$$[0,a], \quad [0,b], quad[0,c] \tag{16.12}$$

elle doit être résolue avec conditions aux bornes :

$$\begin{aligned} \xi(0) = \xi(a) = 0 \\ \theta(0) = \theta(b) = 0 \\ \zeta(0) = \zeta(c) = 0 \end{aligned} \tag{16.13}$$

Les résultats obtenus dans le chapitre de Physique Quantique Ondulatoire lors de la résolution de l'équation de SCHRÖDINGER dans le cas de du puits rectiligne nous donne directement :

$$\begin{aligned} \xi_\lambda(x) &= \sqrt{\frac{2}{a}}\sin\left(\frac{\lambda\pi x}{a}\right) \quad E_{x,\lambda} = \frac{\lambda^2h^2}{8ma^2} \quad \text{avec} \quad \lambda \in \mathbb{N}^* \\ \theta_\mu(y) &= \sqrt{\frac{2}{b}}\sin\left(\frac{\mu\pi y}{b}\right) \quad E_{y,\mu} = \frac{\mu^2h^2}{8mb^2} \quad \text{avec} \quad \mu \in \mathbb{N}^* \\ \zeta_\nu(z) &= \sqrt{\frac{2}{c}}\sin\left(\frac{\nu\pi z}{c}\right) \quad E_{z,\nu} = \frac{\nu^2h^2}{8mc^2} \quad \text{avec} \quad \nu \in \mathbb{N}^* \end{aligned} \tag{16.14}$$

En résumé les états stationnaires de la particule dans la boîte tridimensionnelle sont spécifiés par trois nombres quantiques entiers strictement positifs equation. Les fonctions d'onde sont finalement :

$$\Psi_{\lambda\mu\nu} = \sqrt{\frac{8}{abc}}\sin\left(\frac{\lambda\pi x}{a}\right)\sin\left(\frac{\mu\pi y}{b}\right)\sin\left(\frac{\nu\pi z}{c}\right) \tag{16.15}$$

et leurs énergies respectives (valeurs propres) :

$$E_{\lambda\mu\nu} = \frac{h^2}{8m}\left(\frac{\lambda^2}{a^2} + \frac{\mu^2}{b^2} + \frac{\nu^2}{c^2}\right) \tag{16.16}$$

La technique de séparation de variables détaillée ci-haut, n'est applicable que parce que l'Hamiltonien est de forme séparable. Il vient automatiquement dès lors que la densité de probabilité tridimensionnelle equation est le produit des densités de probabilités equation, comme nous l'avions anticipé. Nous notons aussi que l'énergie de mouvement dans l'espace tridimensionnel est la somme des énergies de mouvements dans les trois directions spatiales : l'indépendance de ces trois directions ou degrés de liberté, implique donc l'additivité de leur énergie.

16.2 Vibrations moléculaires

Nous avons étudié en physique quantique ondulatoire l'oscillateur harmonique. C'est maintenant en chimie que l'on va utiliser toute la puissance des résultats obtenus de l'étude de ce système.

L'oscillateur harmonique est un modèle des vibrations moléculaires, et est représenté par un potentiel parabolique de type :

$$E_p = \frac{1}{2}\mathrm{c^{te}}x^2 \tag{16.17}$$

pour une molécule diatomique. Mais nous avons vu dans le chapitre de Physique Nucléaire que equation ce qui fait que nous avons finalement pour une molécule diatomique :

$$(53.19) E_p = \frac{m\omega_0^2}{2}x^2 \tag{16.18}$$

Pour une molécule polyatomique, nous aurons in extenso (par l'additivité de l'énergie) :

$$(53.20) E_{\text{pot}}(q_i) = \frac{1}{2}\sum_i \omega_i^2 q_i^2 \tag{16.19}$$

Les quantités equation sont des fréquences (ou plutôt, plus correctement des pulsations) vibrationnelles d'une molécule, diatomique dans le premier cas, et polyatomique dans le second cas. Dans la première équation, la variable x représente l'élongation de la liaison entre les deux atomes A et B (comme avec un ressort) dans une molécule diatomique, c'est-à-dire equation, où R est la longueur instantanée de cette liaison, et equation est sa valeur d'équilibre. Dans le cas d'une molécule polyatomique, le potentiel décrivant les vibrations moléculaires ne prend la forme séparable et termes de sommation ci-dessus qu'en terme de variables spéciales equation qui dénotent des mouvements collectifs des noyaux, et qui sont appelées "modes normaux de vibrations".

Nous avons aussi vu en physique quantique ondulatoire que l'Hamiltonien d'une molécule diatomique (problème de l'oscillateur harmonique) peut s'écrire sous la forme :

$$(53.21) H = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{m\omega^2}{2}x^2 \tag{16.20}$$

Pour une molécule polyatomique cette relation devient logiquement :

$$(53.22) H = \sum_i \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial q_i^2} + \frac{m\omega^2}{2}q_i^2\right) \tag{16.21}$$

L'Hamiltonien est clairement de forme séparable ci-dessus : c'est une somme d'Hamiltoniens unidimensionnels, chacun ne dépendant que d'un seul mode equation comme variable, et décrivant ce mode comme étant un ressort unique, ou oscillateur harmonique de masse unitaire equation et de pulsation d'oscillation equation. Par conséquent, une séparation des variables equation est possible, réduisant l'équation de SCHRÖDINGER indépendante du temps en un certain nombre d'équations du même type que celle d'un oscillateur harmonique unidimensionnel. Il suffit donc de connaître l'expression de la fonction d'onde pour un oscillateur harmonique unidimensionnel, ce que nous avons déjà fait en physique quantique ondulatoire où nous avions obtenu :

$$(53.23) E_n = \hbar\omega\left(n + \frac{1}{2}\right) \quad \text{avec} \quad n \in \mathbb{N} \tag{16.22}$$

et

$$(53.24) \Psi_n = \left(A\sqrt{\pi}n!2^n\right)^{-1/2} \mathrm{e}^{-Q^2/2} H_n(Q) \tag{16.23}$$

La figure ci-dessous montre le graphique des premières fonctions d'onde de la relation précédente ainsi que celui de leurs densités de probabilité de présence respectives. On note les mêmes structures nodales que celles des fonctions propres d'une particule dans une boîte unidimensionnelle.

equation (53.25)

Ci-dessus les premiers niveaux d'énergie de l'oscillateur unidimensionnel avec (a) leur fonction propre associée, (b) la distribution de probabilité de présence associée.

Dans la limite des très grandes valeurs de n, la distribution de probabilité se rapproche de plus en plus de celle prédite par la mécanique classique, l'oscillateur réside pour la majeure partie du temps au voisinage des points de rebroussement définis par l'intersection du potentiel equation avec le niveau n. Cette tendance est illustrée ci-dessous :

equation (53.26)

Pour une molécule polyatomique l'expression de la quantification d'énergie devient :

$$(53.27) E_n = \sum_i \hbar\omega_i\left(n_i + \frac{1}{2}\right) \quad \text{avec} \quad n \in \mathbb{N}, \forall i = 1, 2, 3 \ldots, n \tag{16.24}$$

et les fonctions/états propres deviennent :

$$(53.28)\Psi_n(q_i) = (A\sqrt{\pi}n!2^n)^{-1/2}\prod_i H_{n_i}(Q_i)\mathrm{e}^{-Q_i^2/2} \tag{16.25}$$

avec :

$$(53.29)Q_i = \sqrt{\frac{\omega_i}{\hbar}}q_i \tag{16.26}$$

Ces deux dernières relations sont très importantes parce qu'elles permettent parmi tant d'autres de :

- prédire le spectre de la molécule (spectroscopie) ;
- d'étudier les bandes d'énergie (d'où viennent les notions de bande de conduction et de valence) ;
- de localiser les liaisons entre atomes et donc les propriétés chimiques.

16.3 Atome hydrogénoïde

Nous considérons ici la quantification d'un système générique comportant deux corps (particules) en interaction mutuelle et se mouvant dans l'espace tridimensionnel. On démontrera dans un premier temps que, si la séparation des variables dynamiques décrivant individuellement chacun des deux corps est impossible, par contre, le mouvement d'ensemble du système (celui du centre de masse) et le mouvement interne, dit encore mouvement relatif, sont séparables. En outre, si le potentiel est centro-symétrique, le mouvement interne peut encore se décomposer en un mouvement de rotation et un mouvement radial. La quantification du mouvement rotationnel est intimement reliée à celle du moment cinétique.

Nous nous intéresserons ici à la mécanique d'un système atomique ne comportant qu'un seul électron. C'est un système à deux particules : un noyau, de masse M et de charge equation, et un électron de masse equation et de charge equation. Il est décrit par l'hamiltonien suivant :

$$\begin{aligned}(53.30)\quad H &= E_c^{\text{noyau}} + E_c^{\text{électron}} + E_{\text{pos}}^{système} = \frac{p_N^2}{2M} + \frac{p_e^2}{2m_e} + E_{\text{pos}}(\vec{r}_e - \vec{R}_N) \\ &- \frac{\hbar^2}{2M}\nabla_N^2 - \frac{\hbar^2}{2m_e}\nabla_e^2 + E_{\text{pos}}(\vec{r}_e - \vec{R}_N)\end{aligned} \tag{16.27}$$

Rappelez-vous qu'en physique atomique nous avions démontré lors de l'étude des opérateurs fonctionnels que :

$$(53.31)p = \frac{1}{\mathrm{i}}\hbar\nabla = \frac{1}{\mathrm{i}}\frac{h}{2\pi}\hbar\nabla \tag{16.28}$$

Donc $\vec{r}_e$ et $\vec{R}_N$ sont les vecteurs de position de l'électron et du noyau, respectivement.

L'énergie potentielle étant donnée par (cf. chapitre d'Électrostatique) :

$$(53.32)E_{\text{pos}}(\vec{r}_e - \vec{R}_N) = -\frac{Zq_e^2}{4\pi\varepsilon||\vec{r}_e - \vec{R}_N||} \tag{16.29}$$

Les mouvements des deux particules sont corrélés car les deux charges interagissent à travers leur champ électrique mutuel. On ne peut donc pas effectuer une séparation de variables entre equation et equation. Par contre, une séparation de variables est possible entre la coordonnée de centre de masse (voir la définition du centre de masse en mécanique classique) :

$$(53.33)\vec{R}_{CM} = \frac{m_e\vec{r}_e + M\vec{R}_N}{m_e + M} \tag{16.30}$$

et la coordonnée relative de l'électron par rapport au noyau :

$$(53.34)\vec{r}_{\text{rel}} = \vec{r}_e - \vec{R}_N \tag{16.31}$$

Nous obtenons dès lors :

$$(53.35)\vec{p}_{CM} = \frac{\hbar}{\mathrm{i}}\vec{\nabla}_{CM} = \frac{\hbar}{\mathrm{i}}\left(\vec{\nabla}_e + \vec{\nabla}_N\right) = \vec{p}_e + \vec{p}_N \tag{16.32}$$

et :

$$(53.36)\vec{p}_{\mathrm{rel}} = \frac{M\vec{p}_e - m_e\vec{p}_N}{m_e + M} \tag{16.33}$$

L'hamiltonien dans le référentiel de centre de masse s'écrit donc :

$$(53.37)H = \frac{p_{CM}^2}{2M_{\mathrm{tot}}} + \frac{p_{\mathrm{rel}}^2}{2\mu} + E_{\mathrm{tot}}(\vec{r}_{\mathrm{rel}}) \tag{16.34}$$

où equation est la masse totale du système, et :

$$(53.38)\mu = \frac{m_e M}{M_{\mathrm{tot}}} \tag{16.35}$$

est sa masse réduite.

Nous voyons clairement que, l'hamiltonien H est mis sous une forme séparable et qu'on peut l'écrire :

$$(53.39)H = H_{CM} + H_{\mathrm{rel}} \tag{16.36}$$

avec :

$$(53.40)H_{CM} = \frac{p_{CM}^2}{2M_{\mathrm{tot}}} \quad \text{et} \quad H_{\mathrm{rel}} = \frac{p_{\mathrm{rel}}^2}{2\mu} + E_{\mathrm{tot}}(\vec{r}_{\mathrm{rel}}) \tag{16.37}$$

En termes des coordonnées equation et equation, la fonction décrivant un état stationnaire du système à deux corps est donc un produit de fonctions d'onde individuelles (rappelons que la probabilité conjointe de deux événements est le produit de leur probabilité), l'une pour le mouvement du centre de masse, l'autre pour le mouvement relatif :
equation (53.41)
et l'énergie de cet état est la somme des énergies de mouvement respectives :
equation (53.42)
avec
equation et equation (53.43)

REMARQUE Cette approche de séparer la fonction d'onde en la composition d'une fonction d'onde du centre de masse et du mouvement relatif est utilisée également dans le cadre de l'étude des atomes polyélectroniques mais à une différence près : comme le noyau est dès lors beaucoup plus massif que le cortège des électrons (en approximation...), le centre de masse est assimilé au noyau de l'atome et le mouvement relatif à l'ensemble du cortège électronique. Cette approche approximative est très connue sous la dénomination "d'approximation de Born-Oppenheimer".

L'hamiltonien apparaissant dans la première de ces deux relations a été défini plus haut comme valant :
equation (53.44)
Ce mouvement est celui d'une particule de masse equation dans une boite tridimensionnelle de volume infini. Les fonctions propres et valeurs propres pour ce mouvement étant déjà obtenues dans notre étude précédente, on se limitera à l'étude de l'équation séparée pour le mouvement relatif, ou mouvement interne. Comme aucune confusion ne sera dès lors plus possible entre les différentes Hamiltoniens, nous laisserons tomber, pour simplifier les notations, la mention rel en indice inférieur.
Avec equation donné par la relation que nous avons démontrée précédemment :
equation (53.45)
et la relation (aussi démontrée précédemment) :
equation (53.46)
nous obtenons alors l'équation de SCHRÖDINGER pour le mouvement relatif :
equation (53.47)

ou écrit autrement :
equation (53.48)
Dans le cas où l'énergie potentielle equation est de source centrosymétrique, c'est-à-dire qu'il ne dépend que de la longueur du vecteur position equation, et non de son orientation, l'équation précédente, telle qu'écrit, en coordonnées cartésiennes, n'est pas séparable : en effet, en coordonnées cartésiennes, la longueur de equation est données par :
equation (53.49)
et l'énergie potentielle n'est pas séparable en trois composantes chacune dépendant que d'une seule des trois variables x, y, z. L'hamiltonien n'est donc pas de forme séparable. Cependant, l'équation précédente est séparable en coordonnées sphériques. Car, dans ce système de coordonnées, le potentiel ne dépend que d'une des trois variables sphériques, le rayon r. Il est indépendant des deux angles equation et equation.
Si nous nous référons au résultat obtenu lors de l'étude des expressions du laplacien dans différents systèmes de coordonnées, dans le chapitre de Calcul Vectoriel, nous avions obtenu pour le laplacien en coordonnées sphériques, l'expression suivante :
equation (53.50)
L'hamiltonien :
equation (53.51)
devient dès lors (simple distribution et nouvelle manière de noter) :
equation (53.52)
où :
equation (53.53)
est l'opérateur énergie cinétique pour le mouvement radial de l'électron par rapport au noyau, et equation est l'opérateur "associé" au carré du vecteur moment cinétique :
equation (53.54)

16.3.1 Rotateur rigide

Si nous considérons maintenant le cas d'un système appelé "rotateur rigide" où nous négligeons (restreignons serait un terme plus adapté) les degrés de liberté d'oscillation (c'est à ce système que l'on a affaire dans le cas des molécules diatomiques ou polyatomique linéaires), les seules coordonnées mises en jeu sont les angles equation et equation qui fixent l'orientation du rotateur.
Ainsi, dans ce cas r est fixé et nous avons :
equation, equation (53.55)
Vu les contraintes sur le potentiel, il est facile de comprendre pourquoi le rotateur est dit "rigide".
L'Hamiltonien se réduit alors à :
equation (53.56)
où :
equation (53.57) est le moment d'inertie (cf. chapitre de Mécanique Classique) de la masse réduite du système.

Remarque Nous associons l'opérateur (conséquent...) equation à un moment cinétique, pour la simple raison qu'il en a les unités. Effectivement, rappelons que nous avons démontré en physique quantique ondulatoire que lorsque le spin est nul (donc dans le cadre de notre étude de l'atome hydrogénoïde ici présent, le spin ne sera pas pris en compte dans un premier temps) et que nous avons affaire à une seule particule alors le moment cinétique (que nous noterons donc L dans ce chapitre au lieu de b) est donné par :
equation et que equation avec equation (53.58)

Finalement, nous pouvons écrire l'équation de Schrödinger sous la forme :
equation (53.59)
Rappelons aussi que nous avions obtenu (cf. chapitre de Physique Quantique Ondulatoire) :
equation (53.60)
Passons maintenant des coordonnées rectangulaires x, y, z aux coordonnées sphériques equation. Rappelons (cf. chapitre de Calcul Vectoriel) que :
equation et equation (53.61)
exprimons les différentielles totales :
equation (53.62)

Ces relations peuvent être écrites comme une transformation orthogonale des différentielles totales : equation par :

equation (53.63)

ou encore par la transformation inverse (au besoin... il suffit de vérifier que les deux matrices de transformation multipliées entre elles donnent la matrice unité) :

equation (53.64)

Il en résulte par exemple :

equation (53.65)

et finalement :

equation (53.66)

Ainsi, en tenant compte de ces relations, nous obtenons par exemple, dans le cas de l'opérateur :

equation (53.67)

Le résultat suivant :

equation (53.68)

De même :

equation (53.69)

Finalement, nous avons donc peu de libertés de mouvement pour notre rotateur rigide (car il est très rigide...) et nous pouvons écrire pour l'équation de SCHRÖDINGER :

equation (53.70)

où equation est rappelons-le, vu comme un opérateur linéaire fonctionnel, et l'énergie totale E comme sa valeur propres correspondante.

Dès lors, nous pouvons écrire que l'opérateur moment cinétique (nous changeons la notation afin de ne pas confondre par la suite opérateur et valeur propre conformément aux remarques que nous avions faites lors des énoncées des postulats de la physique quantique ondulatoire) :

equation (53.71)

Ainsi, les fonctions propres equation de equation sont solutions de l'équation aux valeurs et fonctions propres :

equation (53.72)

c'est-à-dire à l'équation différentielle :

equation (53.73)

où equation est bien évidemment la valeur propre de equation. Une solution simple de cette équation différentielle serait :

equation (53.74)

avec comme condition d'uniformité selon les propriétés des formes complexes (cf. chapitre sur les Nombres) :

equation (53.75)

Cette condition mathématique, impose la quantification évidente et remarquable suivante :

equation avec equation (53.76)

où (rappel) equation est le nombre quantique magnétique.

Sachant que (cf. chapitre de Physique Quantique Corpusculaire) :

equation (53.77)

Nous pouvons écrire :

equation (53.78)

Dès lors, nous retrouvons le(s) résultat(s) que nous avions en physique quantique corpusculaire et ondulatoire :

equation (53.79)

Ce qui est assez satisfaisant, même remarquable et jouissif (pour ne pas le dire...).

Ainsi, la mesure d'une composante du moment cinétique fournit toujours un multiple entier relatif de equation qui apparaît comme l'unité naturelle du moment cinétique.

Les fonctions propres communes (!!!) aux opérateurs equation et equation sont dans une cadre plus général nécessairement de la forme (méthode de séparation des variables) :

equation (53.80)

Comme le rotateur est rigide, nous avons equation. Ce facteur s'éliminera de lui-même dans l'équation aux valeurs propres et fonctions propre que nous déterminerons de suite. Donc nous pouvons ne pas le prendre en compte. Finalement, nous pouvons écrire :

equation (53.81)

Ce qui nous amène à l'équation aux valeurs et fonctions propres :

equation (53.82)
C'est-à-dire :
equation (53.83)
d'où :
equation (53.84)
En posant :
equation (53.85)
et donc :
equation (53.86)
nous obtenons une équation différentielle du type "Fuchs" donnée par :
equation (53.87)
D'où finalement :
equation (53.88)
Dont les coefficients présentent des pôles (singularités) en equation. Or, rappelons que nous avons :
equation (53.89)
Ainsi l'équation différentielle s'écrit :
equation (53.90)
Dont les coefficients présentent des pôles (singularités) en equation. Or, rappelons que nous avons :
equation (53.91)
Ainsi l'équation différentielle s'écrit :
equation (53.92)
Une solution non triviale étant, connaissant les équations différentielles de type Fuchs..., les polynômes de Legendre (cf. chapitre de Calcul Algébrique). Ce que vous pouvez vérifier en injectant cette solution dans l'équation différentielle :
equation (53.93)
Donc finalement, nous avons des fonctions propres communes (car rappelons que les polynômes de Legendre sont orthogonaux entre eux) qui seront :
equation (53.94)
Pour normaliser la partie dépendante de equation, nous savons que l'exponentielle doit être multipliée par equation (voir chapitre de Statistiques, la partie concernant la fonction de distribution de Gauss-Laplace) ce qui nous permet déjà d'écrire :
equation (53.95)

Remarque Il n'y pas besoin de faire de calculs compliqués pour calculer le facteur de normalisation de l'exponentielle, car dans le cadre d'une intégration sur tout l'espace, les trois facteurs de equation sont indépendants les uns des autres. Ainsi l'intégrale sera le produit des intégrales (cf. chapitre de Calcul Différentiel Et Intégral).

Finalement, nous devons trouver equation tel que :
equation (53.96)
et (ce que nous allons démontrer juste plus bas) :
equation (53.97)
En résumé, nous écrivons :
equation (53.98)
Vérification (attention c'est un peu long et il est conseillé de le relire plusieurs fois) :
Nous considérons les fonctions définies par :
equation où equation (53.99)
avec :
equationet equation (53.100)
Le but est de montrer que ces fonctions sont orthogonales (dans equation où equation) et de trouver les constantes equation telles que equation.
D'abord, démontrons pour les besoins ultérieurs que pour equation :
equation (53.101)

Preuve Si et seulement si equation l'égalité est évidente. Supposons equation (donc le cas général en dehors du cas particulier évident précédent) et soit P un polynôme réel de degré equation.
Posons :

equation (53.102)
Montrons que (produit scalaire fonctionnel) :
equation (53.103)
dans equation Effectivement, rappelons que nous avons faire le changement de variable equation. En intégrant par parties nous obtenons :
equation
remarquons que pour tout equation, equation est nul en equation. Par suite (par extension), la relation précédente se simplifie en :
equation (53.104)
Après equationintégrations par parties nous obtenons :
equation (53.105)
Si equation alors l'expression précédente montre trivialement que :
equation (53.106)
Si equation alors en posant :
equation (53.107)
Nous obtenons :
equation (53.108)
remarquons encore une fois que equation s'annule en equation pour tout equation. En intégrant equation fois par parties l'expression précédente nous trouvons :
equation (53.109)
or h est un polynôme de degré equation .
Effectivement, le premier facteur est de degré 2m et la dérivée equation-ème de equation est de degré equation, dès lors :
equation (53.110)
donc equation est un polynôme de degré equation et sachant que equation est à une constante près égal au l-ème polynôme de Legendre (cf. chapitre de Calcul Algébrique) nous avons alors :
equation (53.111)
Nous venons de montrer que equation est orthogonal à tout polynôme de degré equation.
equation

equation est un polynôme de degré l (il suffit de le vérifier pour quelques valeurs) donc cherchons s'il existe une constante equation telle que :
equation (53.112)
avec rappelons-le :
equation (53.113)
Nous pouvons déterminer la constante C en comparant les coefficients dominants des polynômes :
equation et equation (53.114)
Le coefficient dominant de equation est :
equation (53.115)
et le coefficient dominant de equation est :
equation (53.116)
Ainsi :
equation (53.117)
c'est-à-dire :
equation (53.118)
Nous aurions donc pour equation (on intègre par parties autant de fois qu'il le faut à gauche et à droite - nécessairement - pour obtenir ce résultat) :
equation (53.119)
Maintenant, établissons une relation remarquable qui existerait entre equation (et qui nous sera utile par la suite) . Supposons pour cela equation et rappelons la base qu'à la base :
equation (53.120)
Donc cela nous amène à écrire (rien de particulier) :
equation (53.121)
Par les résultats précédents (equation) :
equation (53.122)
cela nous amène à écrire :
equation (53.123)
Ainsi, nous obtenons :
equation (53.124)

Nous allons à présent (enfin !) montrer que les fonctions equation sont orthogonales.
D'abord, démontrons que :
equation> (53.125)
où equation est le l-ème polynôme de Legendre.

Preuve D'abord, nous savons que polynômes de Legendre satisfont la formule de récurrence suivante (cf. chapitre de Calcul Algébrique) :
equation (53.126)
pour equation .
En multipliant l'égalité ci-dessus par equation et en intégrant, nous obtenons :
equation (53.127)
Or :
equation (53.128)
Rappelons que les equation polynômes forment une base orthogonale dont les polynômes qui l'engendrent sont de degré croissant de 0 à n, donc un polynôme d'ordre inférieur - exprimé dans un sous-espace vectoriel - sera toujours perpendiculaire aux vecteurs (polynômes) générant les dimensions supérieures. Ainsi, si nous prenons l'exemple de equation engendré par la base equation, alors un vecteur equation exprimé par la combinaison linéaire de equation, sera toujours perpendiculaire à equation et donc un produit scalaire avec celui-ci.
Et donc par suite :
equation (53.129)
Posons :
equation (53.130)
L'expression précédente devient :
equation avec equation (53.131)
Ainsi par récurrence :
equation (53.132)
De plus comme :
equation (53.133)
et :
equation (53.134)
Donc :
equation (53.135)
et :
equation (53.136)
Nous avons donc bien démontré (c'est juste au cas ou... vous ne suiveriez plus...) que :
equation (53.137)
equation

Enfin, attaquons à ce qui nous intéresse enfin. C'est-à-dire démontrer que :
equation (53.138)

Preuve Si equation :
equation (53.139)
où equation.
remarque Remarque : Rappelons que le jacobien en coordonnées sphériques est equation (cf. chapitre de Calcul Différentiel Et Intégral) et comme la fonction intégrée n'est pas dépendante de r, nous l'avons sortie de l'intégrale (equation intégrera la fonction R(r) présente dans l'équation de Schrödinger). fin remarque
Et avec :
equation (53.140)
Si equation et equationalors d'abord le produit scalaire se simplifie :
equation (53.141)
En faisant le changement de variable equation nous obtenons :
equation (53.142)
Supposons equation :
equation (53.143)
où equation est le l-ème polynôme de Legendre. Ainsi l'expression du produit scalaire devient :
equation (53.144)
Si nous posons :
equation (53.145)
alors la relation devient :
equation. (53.146)

En intégrant m fois par parties l'expression ci-dessus nous obtenons :
equation (53.147)
Or equation est un polynôme de degré k. Sachant que equation, cette dernière intégrale est nulle pour les mêmes raisons que celles évoquées précédemment. Donc :
equation (53.148)
Si equation alors nous avions démontré que :
equation (53.149)
et donc :
equation car equation (53.150)
Il ne reste qu'à traiter le cas equation. Supposons à nouveau equation. Alors comme avant nous avons :
equation (53.151)
et :
equation (53.152)
Posons :
equation (53.153)
La relation devient alors :
equation (53.154)
En intégrant m fois par parties nous trouvons :
equation (53.155)
equation est un polynôme de degré l dont le coefficient dominant vaut :
equation (53.156)
equation étant orthogonal à tout polynôme de degré strictement inférieur à l, l'expression peut s'écrire :
equation (53.157)
Or, nous avons démontré que :
equation (53.158)
donc :
equation (53.159)
Si equation nous savons que nous obtenons le résultat.
equation

Finalement ce dernier résultat nous donne aussi la condition de normalisation :
equation (53.160)
Et donc finalement :
equation (53.161)
est bien une famille orthonormale.

Enfin, après cette interlude fortement mathématique (mais instructif quant à la méthodologie d'approche), nous voyons (ce qui est logique) que à chaque valeur de l correspondent donc 2l+1 fonctions propres equation. Nous disons encore que la valeur equation est 2l+1 fois dégénérée puisque :
equation (53.162)

Voici quelques valeurs de la fonction equation qui génère que nous appelons communément des "harmoniques sphériques" :
equation (53.163)

Il est possible de tracer ces fonctions dans Maple en procédant de la manière suivante (pour les harmoniques sphérique complexes) :

```
> y22 :=sqrt(15/(2*Pi))*(sin(phi))**2*exp(2*I*theta); > plot3d(Re(y22),theta=0..2*Pi,phi=0..Pi,
coords=spherical,axes=normal, labels=[x,y,z], scaling=CONSTRAINED) ;
```

Voyons quelques tracés de ces magnifiques harmoniques sphériques :

- equation (correspondant à au moins à equation !) donne une sphère (valeur constante quelque soient equation) dont la densité de probabilité peut être présentée par la "carte photographique" ou "carte de densité" (la densité dans un état donnée y est représenté par la densité de points clairs sur un fond foncé) :
equation (53.164)
Ce qui représente les orbitales 1s possibles.
- equation donnent (pour equation au moins !) :
equation (53.165)
Ce qui représente les orbitales 2p possibles, dont la densité de probabilité peut être représentée par sa carte de densité et d'isodensité :
equation equation (53.166)
- equation (pour equation au moins !) :

equation (53.167)

Ce qui représente 5 orbitales centrosymétriques 3d possibles, dont la densité de probabilité peut être représentée par les (les deux dernières cartes représentent equation) cartes de densité :

equation (53.168)

- equation (pour equation au moins !) :

equation (53.169)

Ce qui représente 5 orbitales anti-centrosymétriques 4f possibles, dont la densité de probabilité peut être représentée par (dans l'ordre : equation) :

equation equation

equation equation (53.170)

Les résultats précédents nous amènent donc à écrire :

equationequation (53.171)

Remplaçant ceci dans l'équation de Schrödinger :

equation (53.172)

Nous obtenons (equation dans le rotateur rigide mais equation dans le cas de l'atome hydrogénoïde) :

equation (53.173)

Comme il n'y a dans cette relation, aucun opérateur qui agit sur equation, nous pouvons le simplifier de façon à obtenir :

equation (1) (53.174)

où nous voyons que dans ce cas général de l'atome isolé que les niveaux d'énergie ne dépendant plus de equation (en raison de la symétrie sphérique du potentiel). Nous disons alors que les niveaux correspondant aux mêmes valeurs de n et de l sont tous confondus quelque soient les valeurs de equation.

Dans le cas où equation dérive du potentiel de Coulomb en 1/r, cette équation radiale ne donne lieu à une solution R(r) normable (différente de zéro aussi donc...) que pour des valeurs de l'énergie répondant à la loi de quantification suivante (tiens donc... quelle coïncidence, nous retrouvons l'expression démontrée dans les vieux modèles de la physique quantique corpusculaire) :

equation (53.175)

où equation est la constant de Rydberg telle que nous l'avons déterminée en physique quantique corpusculaire. Ainsi, dans ce cas les niveaux d'énergie correspondant aux mêmes valeurs de n sont tous confondus quelque soit l.

Pour une valeur donnée du nombre quantique principal n (rappelons que nous avons vu dans le chapitre de Physique Quantique Corpusculaire que equation), il est possible de vérifier qu'il existe plusieurs solutions pour la fonction R(r) selon la valeur du nombre quantique azimutal l. D'où l'identification des solutions par la paire (n,l). Nous les notons equation. Ce sont des fonctions réelles de la variable r données par (il suffit de vérifier si elles marchent elles satisfont l'équation de Schrödinger, ce qui est facile, nous en ferons un exemple un peu plus loin) :

equation (53.176)

où :

equation (53.177)

est l'équivalent du rayon de Bohr (pour la masse réduite) que nous avions déterminé dans le chapitre de Physique Quantique Corpusculaire avec comme différence que nous avons ici une masse réduite au lieu d'une masse simple.

Vérifions cependant si notre équation de Schrödinger est bien vérifiée (prenons equation) :

equation (53.178)

Ce qui correspond bien au résultat attendu.

Ce qui sous forme graphique nous donne pour la partie radiale equation :

equation (53.179)

Etudions un peu plus en détail la fonction radiale dans le cas de l'atome d'hydrogène ! :

Dans le cas de l'orbital atomique 1s (cas particulier mais nous pourrions faire les mêmes calculs qui suivent avec toutes les autres orbitales !) nous avons donc pour l'atome d'hydrogène :

equation (53.180)

C'est donc bien une fonction exponentielle décroissante comme le montrait les graphiques précédents. Avant de continuer rappelons (cf. chapitre de Physique Quantique Ondulatoire) que :

equation (53.181)

Or, en coordonnées sphériques (voir le début de ce chapitre) :

equation (53.182)

Il vient alors comme nous l'avons vu plus haut :

equation (53.183)

Il s'ensuit que :

equation (53.184)

Avec ce résultat nous allons pouvoir calculer la probabilité radiale de trouver l'électron sur chaque orbitale atomique ! Ainsi il vient immédiatement avec le résultat précédent :

equation (53.185)

Et donc dans le cas de notre orbitale atomique 1s :

equation (53.186)

Il est maintenant super intéressant de calculer le point r où la probabilité de trouver l'électron est maximale sur l'orbitale 1s !

Pour cela, nous remarquons que equation passe par un maximum lorsque nous avons trivialement :

equation (53.187)

soit :

equation (53.188)

Soit :

equation (53.189)

Ce qui est remarquable car nous retrouvons le résultat du modèle de Bohr.

Pour résumér un pout tout cela, les états stationnaires de l'atome d'hydrogène sont spécifiés par trois nombre quantiques equation et la fonction d'onde de SCHRÖDINGER étant donnée au final par :

equation (53.190)

Nous avons alors la nomenclature traditionnelle suivante dans le cas de l'atome d'hydrogène : Nous pouvons inclure le spin de l'électron dans la description de la structure électronique de l'atome.

n	l	m_l	fonction	nomenclature
1	0	0	$\Psi_{1,0,0}$	1s
2	0	0	$\Psi_{2,0,0}$	2s
	1	1	$\Psi_{2,1,1}$	2p1
		0	$\Psi_{2,1,0}$	2p0
		-1	$\Psi_{2,1,-1}$	2p-1
3	0	0	$\Psi_{3,0,0}$	3s
	1	1	$\Psi_{3,1,1}$	3p-1
		0	$\Psi_{3,1,0}$	3p0
		-	$\Psi_{3,1,-1}$	3p-1
	2	2	$\Psi_{3,2,2}$	3d2
		1	$\Psi_{3,2,1}$	3d1
		0	$\Psi_{3,2,0}$	3d0
		-1	$\Psi_{3,2,-1}$	3d-1
		-2	$\Psi_{3,2,-2}$	3d-2

TABLEAU 16.1 – Nomenclature de l'atome d'hydrogène

Si nous traitons le spin comme un degré de liberté additionnel alors, l'absence de terme d'interaction entre les degrés de liberté classiques (positions dans l'espace réelle) et le spin, interaction appelée "couplage spin-orbite", dans l'hamiltonien précédent, implique que nous pouvons écrire la fonction d'onde totale, spin inclus, sous la forme de produit :

equation (53.192)

où nous avons ajouter le nombre quantique de spin equation (cf. chapitre de Physique Quantique Corpusculaire).

La même remarque que nous avions fait dans le chapitre de Physique Quantique Corpusculaire s'applique dès lors : les niveaux restent equation fois dégénérés.

17

Chimie moléculaire

La chimie moléculaire est le domaine central qui interconnecte par l'étude des molécules de nombreuses techniques de pointe prometteuses de ce début 21ème siècle qui sont pour ne citer que les plus connues de la masse : la biologie moléculaire, les matériaux moléculaires, l'électronique moléculaire, les polymères...

Sachant qu'il a été découvert expérimentalement qu'une même molécule peut avoir plusieurs fonctions très différentes, leur étude théorique permet de bien les utiliser (avec parfois un meilleur rendement en termes de R&D) dans leurs domaines d'application.

Nous ne sommes à notre époque qu'aux balbutiements de ce que la nature a fait avec beaucoup de temps et de hasard : c'est-à-dire des molécules complexes fonctionnant comme des nanomachines capables localement (site actif) de filtrer, d'oxyder, de faire de catalyses... et bien d'autres manipulations (y'a qu'à observer votre corps !).

Une molécule sera systématiquement traitée à notre époque avec l'équation de Schrödinger (donc pas de cas relativistes et pas de prise en compte des spins) sous la forme habituelle :

equation (54.1)

soit sous une forme également stationnaire (indépendante du temps) où pour rappel equationest une fonction propre et E une valeur propre de l'application H.

En réalité, les fonctions d'ondes sont impossibles à calculer formellement avec les outils mathématiques contemporains et la seule chose que nous savons faire sont des calculs numériques (méthode des perturbations). C'est la raison pour laquelle certains centres de chimie se transforment avec le temps en des centres de calculs où le caractère prédictif (et peu coûteux) de la chimie quantique s'affirme de plus en plus.

Il reste cependant bien évidemment indispensable, comme toujours, de comment les modèles théoriques sont construits et leur hypothèses sous-jacentes.

Mais nous pouvons quand même avec les calculs prédire la forme de molécules de taille raisonnable, l'énergie de leurs liaisons internes, leur capacité énergétique sous l'effort d'une déformation, la forme des orbitales moléculaires (O.M.), l'énergie de transitions d'états (lorsque des parties de la molécule se déplacent dans celle-ci), leur réactivité vis-à-vis d'un milieu réactionnel...

Nous distinguons couramment deux cas d'études de la chimie moléculaire :

- Mécanique quantique : toutes les interactions entre particules sont prises en compte sous l'hypothèse de quelques simplifications acceptables.

- Mécanique moléculaire : Pour de grosses molécules, nous ne nous occupons plus du problème électronique mais des interactions de certains paramètres sur lesquels nous voulons nous concentrer.

Par exemple l'hémoglobine (protéine transportant l'oxygène dans les muscles) est une énorme structure moléculaire dont on ne va étudier que le site actif avec les outils de la mécanique quantique. Le comportement global de la molécule elle-même sera traité avec les outils de la mécanique moléculaire.

Il s'ensuite que excepté pour les atomes hydrogénoïdes, nous ne pouvons pas décrire analytiquement une molécule d'un point de vue purement quantique ! Toutes les méthodes quantiques actuelles reposent sur une ou plusieurs approximations. Les fonctions d'onde ne sont donc qu'approchées et le niveau de calcul est adapté en fonction de ce que nous voulons montrer et de la précision que nous recherchons (en cherchant à minimiser le temps de calcul pour des problèmes de coûts...). La bonne compréhension des approximations permet de formuler des modèles simples ne nécessitant qu'un minimum de calculs (souvent triviaux).

Nous nous proposons ici de montrer deux modèles courants (et les plus simples) :

17.1 Approximation orbitalaire

Une molécule est bien évidemment un problème extrêmement complexe : N noyaux, n électron et tout est en mouvement !

equation (54.2)

L'hamiltonien (cf. chapitre de Physique Quantique Ondulatoire) :

equation (54.3)

est alors de la forme cauchemardesque mais intuitive (le G en indice de l'hamiltonien signifie "Général") suivante :

equation (54.4)

où :

1. equation est l'énergie cinétique des K noyaux de masse equation de la molécule.
2. equation est l'énergie cinétique des n électrons de masse equation
3. equationest l'énergie potentielle due à l'attraction électron(-)/noyaux(+)
4. equation est l'énergie potentielle de répulsion électron(-)/électron(-)
5. equation est l'énergie potentielle de répulsion noyau(+)/noyau(+)

Souvent nous retrouvons ces termes sous la forme suivante de l'équation de Schrödinger dans la littérature :

equation (54.5)

Une première approximation que nous pourrions tenter c'est de découpler le mouvement des noyaux de celui des électrons. Effectivement, comme le noyau est beaucoup plus massif (environ 2000 fois) que le cortège des électrons, le centre de masse est assimilé au noyau de l'atome et le mouvement relatif à l'ensemble du cortège électronique. Cette approche approximative est très connue sous la dénomination "d'approximation de Born-Oppenheimer" :

equation (54.6)

qui nous permet alors d'étudier les orbitales moléculaires. Mais malheureusement cette approximation n'est pas suffisante à cause du terme de répulsion inter-électronique (la double somme) qui empêche d'utiliser la technique de séparation des variables comme nous l'avons fait dans le chapitre de Chimie Quantique avec l'atome d'hydrogène.

Par ailleurs cette dernière équation s'écrit également sous la forme de la première ligne du couple d'équation ci-dessous (équation de Schrödinger des électrons et des noyaux) :

equation (54.7)

Ce système d'équations constitue ce que certains appellent "l'approximation adiabatique" (? ? ?).

L'idée suivante va être d'utiliser la propriété suivante :

Soit deux opérateur A et B, f(u) et g(v) les fonctions propres associées aux valeurs propres a et b. Alors f(u)g(v) est fonction propre de l'opérateur somme A+B avec la valeur propre associée a+b.

Ce qui s'écrit :

equation (54.8)

Démonstration :

equation (54.9)

C.Q.F.D.

Et c'est ce que nous allons faire pour décomposer l'hamiltonien nélectronique equation en somme d'hamiltonien mono-électroniques indépendants en sachant de ce qui précède que si nous trouvons la fonction propre pour chacun (ce qui est relativement plus facile) il suffira des les multiplier pour avoir la fonction propre globale.

Ainsi, nous écrivons :

equation (54.10)

et donc nous devrons trouver pour chaque i :

equation (54.11)

Pour ensuite avoir :

equation (54.12)

avec donc :

equation (54.13)

Cette approche par hamiltoniens mono-électronique va nous amener à remplacer :

equation (54.14)

par la somme des hamiltoniens pour un électron appelés "hamiltoniens effectifs" :

equation (54.15)

Cette méthode d'approximation est parfois appelée en chimie théorique "approximation des électrons indépendants" ou encore "approximation orbitalaire" elle consiste donc à globaliser les interactions individuelles électron-électron et à écrire que chaque électron évolue dans un potentiel moyen résultant de la présence de l'ensemble des autres électrons.

La "méthode de Slater" consiste par définition à écrire cette dernière relation sous la forme :

equation (54.16)

où equation est appelé "constante d'écran".

Cette méthode de Slater revient en gros à remplacer les termes purement électroniques par une constante. Elle peut être considérée comme une méthode paramétrique car les constantes ont été déterminées de façon purement expérimentale.

Le principe de calcul empirique de la constante d'écran est relativement simple. Dans un atome polyélectronique, les électrons de coeurs sont sur des orbites beaucoup plus contractés alors que les électrons de valence qui vont être responsable des propriétés chimiques de l'atome en question sont sur des orbites beaucoup plus relâchées.

L'attraction exercée par le noyau sur ces électrons est beaucoup plus faible que celle exercée sur les électrons de coeur et ces électrons ne perçoivent qu'une partie de la charge atomique.

Slater a alors proposé que cette charge effective, qui est le plus souvent notée equation pouvait être calculée en tenant compte de la constant d'écran. Cette constante représente donc l'effet moyen des autres électrons sur l'électron considéré de l'hamiltonien effectif i.

equation (54.17)

Pour un électron périphérique, nous devrons donc considérer que sa constante d'écran est due à tous les électrons placés sur des orbites égales ou inférieures. La tradition (ou plutôt le "truc") veut que le calcul se fasse en regroupant les orbitales atomiques en plusieurs groupes 1s/2s, 2p/3s, 3p/3d/4s, 4p/4d/4f/5s, 5p/etc...

Ensuite le calcul est simple car il se fait sur la base d'un tableau de valeurs prédéfinies et il suffit d'additionner les contributions d'écran de tous les électrons selon le tableau ci-dessous :

n'<n-1
n'=n-1
n'=n
n'>n
1s
0.30
0
ns, np
1
0.85
0.35
0
nd, nf
1
1
0.35
0 (54.18)

Ce tableau mérite quelques explications bien évidemment ! :

L'indice n'indique le nombre du groupe qui contribue à la constante d'écran alors que n est le nombre du groupe de l'électron que nous considérons.

exempleExemple :

Dans le cas du Carbone de configuration 1s22s22p2, la charge nucléaire est equation. Un électron 1s n'est écranté que par l'autre électron 1s, la charge effective qu'il voit est donc de :

equation (54.19)

Un électron 2s ou 2p est écranté par les deux électrons 1s et par les 3 autres électrons 2s et 2p. La charge effective par laquelle il est attiré est alors :

equation (54.20)

Nous voyons donc bien que la charge effective ressentie décroit assez rapidement !

17.2 Méthode C.L.O.A.

Cette méthode, plutôt qualitative, consiste à considérer que la fonction d'onde moléculaire est une Combinaisons Linéaire des Orbitales Atomiques contrairement à la méthode précédente où nous multiplions les hamiltoniens effectifs.

Cette méthode est importante car elle est à la base d'une grande partie du vocabulaire courant des chimistes dès que la chimie effectuée est assez pointue !

Basons nous sur l'exemple de la molécule de dihydrogène equation. L'idée est alors la suivante :

Si nous avons la fonction de l'orbitale atomique equation de equation et respectivement la fonction equation de equation, alors nous posons que l'orbitale moléculaire dicentrique (liée à deux atomes) correspondante est donnée par :

equation (54.21)

qui défini un système quantique à deux états propre.

Mais comme nous le savons bien, en réalité, seul le carre de la fonction d'onde possède un sens physique (probabilité de présence). Ainsi, si nous supposons que la fonction d'onde n'est pas à valeur dans equation, nous avons pour l'unique électron qui nous intéresse (1s) :

equation (54.22)

où nous admettons que :

- equation représente la probabilité de présence d'être près de A
- equation représente la probabilité de présence d'être près de B
- equation représente la probabilité de présence de l'électron qui joue la liaison A-B

Dans le cas particulier de la molécule diatomique symétrique que nous avons choisi comme exemple, les atomes A et B jouent le même rôle et il n'y a donc pas de raison que l'électron soit plus près de A que de B ou inversement.

Ainsi, la probabilité de trouver l'électron près de A est donc égale à la probabilité de le trouver près de B.

equation (54.23)

De plus dans ce cas les orbitales equation et equation sont totalement identiques (orbitales 1s toutes deux du même atome) et il n'y a donc pas lieu de les distinguer. Nous avons donc :

equation (54.24)

Nous avons donc deux solutions pour equation qui sont (ces deux solutions peuvent se trouver sous des notations très variées dans la littérature) :

equation (54.25)

si equation et :

equation (54.26)

si equation.

Remarque : Attention ! Nous ne pouvons pas poser pour les deux dernière relations que equation. Cette dernière égalité a lieu en tout point que si la distance qui sépare les deux noyaux est nulle (ce qui est peu probable) ou, s'ils sont distants d'une certaine valeur D au milieu de celle-ci.

Ces deux expressions seront simultanément solutions de l'équation de Schrödinger. Nous obtenons donc deux orbitales moléculaires à partir des deux orbitales atomiques dans le cas de la molécule diatomique symétrique.

La fonction :

equation (54.27)

est appelée "fonction liante" car elle correspond à un renforcement de la probabilité de présence de l'électron entre les atomes A et B ce qui correspond à la création de la liaison !

Inversement, la fonction :

equation (54.28)

est appelée "fonction anti-liante" car elle correspond à une diminution de la probabilité de présence de l'électron entre les atomes A et B ce qui correspond à la destruction de la liaison !

En définitive, en ce recouvrant, les deux orbitales atomiques de même énergie donnent naissance à deux orbitales moléculaires d'énergies différentes, l'une liante stabilisée et l'autre anti-liante déstabilisée.

Nous avons évidemment au vu de ce qui précède, que dans des cas plus complexe, que le niveau d'énergie de l'orbitale moléculaire liante est plus petite que celle de l'anti-liante (nous allons le démontrer rigoureusement en détails plus bas).

Ainsi, il faut respectivement plus d'énergie pour ioniser l'électron de l'orbitale liante equation qu'il n'en faut pour ioniser l'électron de l'orbital anti-liante equation. Donc il est communément admis de dire que l'énergie de la liaison liante est plus forte que l'anti-liante (mais nous en ferons la démonstration juste un peu plus loin).

Indiquons aussi qu'en chimie, une liaison chimique dans laquelle chacun des atomes liés met en commun un électron d'une de ses couches externes afin de former un doublet d'électrons liant les deux atomes est communément appelée "liaison covalente".

Les chimistes disent alors que la liaison covalente implique le partage équitable d'une seule paire d'électrons, appelé "doublet liant" (mais dont au fait qu'un seul des électrons l'est vraiment). Chaque atome fournissant un électron, la paire d'électrons étant alors délocalisée entre les deux atomes comme nous l'avons démontré.

Ce sont les raisons qui font que nous disons couramment que la liaison equation est une liaison chimique covalente créée entre deux atomes par recouvrement axial d'orbitales.

Creusons maintenant cette approche ! Les orbitales moléculaires doivent être normées. Ce qui signifie que :

equation (54.29)

Ce qui donne, puisque les orbitales atomiques sont normées pour equationet sont des fonctions réelles :

equation (54.30)

Puisque a (nombre réel dans notre cas particuler) est imposé comme étant une constante, il vient immédiatement :

equation (54.31)

D'où pour equation :

equation (54.32)

De manière identique, nous avons pour equation :

equation (54.33)

Si nous avons equation , il vient la forme suivant que nous trouvons dans la littérature :

equation (54.34)

Faisons un petit exemple en utilisant comme l'orbitale atomique la plus basse (1s) de l'atome d'hydrogène dans le cas d'une liaison dihydrogèneequation pour laquelle nous avons démontré à la fin du chapitre de Chimie Quantique que :

equation (54.35)

Il vient alors que :

equation et equation (54.36)

avec pour rappel :

equation (54.37)

Il vient alors pour l'orbitale moléculaire liante de niveau s :

equation (54.38)

et pour l'anti-liante de niveau s :

equation (54.39)

Nous voyons alors immédiatement que equation s'annule au milieu des deux protons car equation. L'orbitale moléculaire anti-liante présente donc un plan nodal et les électrons sont essentiellement localisés sur les protons.

Par contre, pour l'orbitale moléculaire equation la densité ne s'annule pas. Nous comprenons alors aisément qu'un électron de equation assure la stabilité de la molécule et qui est donc responsable de la liaison chimique.

Nous en déduisons donc que la stabilisation électronique due à l'interaction de deux orbitales identiques est proportionnelle à leur recouvrement. Plus le recouvrement est grand, plus la stabilisation est importante.

Il y a une approche encore plus technique utilisant la notation de Dirac et qui a l'avantage de permettre de la détermination des valeurs propres de l'énergie.

D'abord nous écrivons l'expression générale de l'équation indépendante du temps de Schrödinger avec la notation Ket-Bra pour une orbitale moléculaire, superposition de deux orbitales atomiques :

equation (54.40)

Soit sous forme explicite :

equation (54.41)

Si nous multiplions par le bra equation à gauche et en tenant compte que a, b ainsi que les valeurs propres de l'énergie sont des constantes, nous obtenons l'équation suivante :

equation (54.42)

De même, nous obtenons avec :

equation (54.43)

Simplifions encore les écritures :

equation (54.44)

Par symétrie du problème dans le cas du dihydrogène, nous posons :

equation (54.45)

qui sont appelées "intégrales de résonance" car il s'agit d'un terme relatif à la combinaisons (résonance) des deux orbitales atomiques propres aux deux atomes de base de la structure moléculaire.

Nous avons également :

equation (54.46)

qui sont appelées "intégrales coulombiennes" car elles correspondant selon le cinquième postulat de la physique quantique ondulatoire (voir chapitre du même nom) à la valeur moyenne de l'énergie totale de l'électron.

Nous avons évidemment aussi :

equation (54.47)

qui sont appelées "intégrales de recouvrement" car les deux orbitales atomiques de même type de chacun des atomes se recouvrent.

Et enfin, nous avons toujours par symétrie de notre cas particulier :

equation (54.48)

Nous pouvons alors écrire, puisque les intégrales de recouvrement sont unitaires :

equation (54.49)

Ces deux équations sont appelées "équations séculaires". La solution triviale, mais à priori non physique car elle signifierait que l'électron aurait une densité de probabilité nulle en tout point de l'espace correspond à equation.

Il existe une solution non triviale, si et seulement si, le déterminant (cf. chapitre d'Algèbre Linéaire), appelé dans la chimie moléculaire "déterminant séculaire", suivant est nul :

equation (54.50)

Comme nous avons par symétrie de notre cas particulier :

equation (54.51)

Il vient alors :

equation (54.52)

D'où :

equation (54.53)

Ce qui nous donne deux solutions (+) :

equation (54.54)

et (-) :

equation (54.55)

Nous avons donc :

equation et equation (54.56)

Mais pour pouvoir faire le calcul des niveaux d'énergie de manière détaillée, encore faut-il avoir la forme de l'hamiltonien... et celui utilisant les deux électrons de la molécule de dihydrogène est déjà bien tordu... pour simplifier l'étude, nous nous réduisons au cas du cation (ion positif) equation constitué de deux protons et un électron :

equation (54.57)

Nous avons alors en nous basant sur la relation obtenue au début de ce chapitre :

equation (54.58)

La relation suivante :

equation (54.59)

où les deux premier termes dans la parenthèse sont pou rappel associés à l'énergie potentielle de l'électron et le dernier à l'énergie potentielle de répulsion des protons (le premier terme à droite de l'égalité étant l'énergie cinétique de l'électron).

Essayons maintenant d'ordonner l'énergie ces deux orbitales moléculaires. Pour cela, écrivons :

equation et equation (54.60)

Rappelons que pour qu'un système soit stable, les énergies En doivent être négatives, ce qui correspond à des états stables (il faut un apport de l'énergie pour les défaire) et nous impose à cause de la forme de equation :

equation (54.61)

Sachant ceci, il vient :

equation (54.62)

Dès lors, nous voyons que les notations ne sont pas conformes à l'usage de la physique quantique car normalement l'indice 1 est réservé au niveau de plus basse énergie. Nous écrirons donc à l'avenir :

equation et equation (54.63)

avec les fonctions propres associées equation et equation et donc :

equation (54.64)

Nous pouvons également remarquer une chose importante ! C'est que si nous considérons les atomes isolément, les termes d'interactions s'annulent et nous avons :

equation (54.65)

Donc nous avons la différence qualitative entre un atome isolé et un système diatomique simple :

equation (54.66)

Ce qui signifie que l'énergie du plus bas niveau d'une molécule diatomique est inférieure à l'énergie d'un atome isolé qui est proche de equation. Cette observation confirme que le système est stabilisé en énergie par rapport aux deux atomes isolés, ce qui paraît cohérent avec le constat expérimental de l'existence de telles molécules.

L'usage veut que les chimistes représentent les différences d'énergie sous la forme suivante pour notre cas particulier :

equation (54.67)

Nous concluons donc que lorsque deux atomes (contribuant chacun avec un électron) s'unissent, leurs orbitales atomiques vont se combiner pour générer deux orbitales moléculaires, l'une equation d'énergie plus basse et une seconde equation d'énergie plus élevée que celle des atomes isolés. Ainsi, la rupture qui fera partir l'électron avec un des atomes sera exothermique par rapport aux atomes isolés.

18

Chimie analytique

La chimie est une science très complexe à n-corps que les mathématiques ne peuvent expliquer sans l'apport de simulations numériques sur ordinateur ou d'approximations quant à l'utilisation de la théorique quantique (voir section d'atomistique). D'ici que ces outils soient suffisamment puissants et accessibles à tout le monde, la chimie reste une science principalement expérimentale basée sur l'observation de différentes propriétés de la matière dont voici quelques définitions fortes importantes (que nous retrouvons d'ailleurs dans d'autres domaines que la chimie).

Définitions :

D1. Une "propriété subjective" est une propriété basée sur une impression personnelle/individuelle, par exemple, la beauté, la sympathie, la couleur, l'utilité, etc...

D2. Une "propriété objective" est une propriété ressentie (qui ne peut être contredite), par exemple, sa masse, son volume, sa forme, etc...

D3. Une "propriété qualitative" est une propriété qui descriptive donnéeà l'aide de mots. Par exemple : forme ovale, magnétique, conducteur, etc...

D4. Une "propriété quantitative" est une propriété qui se chiffre. Par exemple : sa masse, son volume, sa densité, etc...

D5. Une "propriété caractéristique" est une propriété exclusive qui permet d'identifier une substance pure. Elle ne change pas même si l'on transforme physiquement la matière, par exemple, sa masse volumique, son point d'ébullition, son point de fusion, etc...

D6. Un corps est dit un "corps pur" lorsque tout échantillon de ce corps présente des valeurs de constantes bien définies et identiques à celles de l'ensemble (densité, point de fusion, point d'ébullition, indice de réfraction, etc.).

remarque Remarque : Nous connaissons environ 2'000'000 substances pures différentes en ce début de 21ème sciècle (c'est dire... qu'il y a du travail derrière). fin remarque

D7. Nous nommons "corps composés", les corps purs qui, soumis à des procédés chimiques, restituent leurs composants.

D8. Si nous effectuons la séparation des mélanges et la décomposition des combinaisons, nous obtenons finalement des corps qui sont indécomposables par les méthodes chimiques classiques ; nous les nommons "éléments" ou "corps simples".

Ainsi, la plus petite partie d'une combinaison chimique présentant encore toutes les propriétés de celle-ci est la "molécule" de cette combinaison. La plus petite partie d'un élément ou corps simple est "l'atome" de cet élément.

Quelques rappels d'ordre général au préalable :

R1. Un mélange est dit "hétérogène" en chimie si ces constituants sont immédiatement discernables à l'oeil nu ou au microscope

R2. Une mélange est dit "homogène" en chimie si ces constituant ne sont pas discernables à l'oeil nu ou au microscope.

18.1 Mélanges simples

Avant de partir dans des équations plus ou moins compliquées, le cas le plus simple d'application des mathématiques à la chimie par lequel nous pouvons commencer est la gestion des mélanges pour les opérations d'analyse et de contrôle de réactions chimiques simples avec deux mélanges.

Considérons deux exemples typiques et particuliers :

1. Soit une solution (jaune) de 10 millilitres d'une solution qui contient une concentration d'acide à 30%. Combien de millilitres d'acide pur (bleu) devons-nous rajouter pour augmenter la concentration (verte) à 50% ?

equation (55.1)

Puisque l'inconnue est la quantité d'acide pur à rajouter, nous la noterons x. Nous avons alors :

equation (55.2)

Ce qui donne :

equation (55.3)

Il vient alors trivialement :

equation (55.4)

Donc 4 millilitres d'acide devraient être ajoutés à la solution d'origine.

2. Un jerricane contient 8 litres d'essence et d'huile pour faire fonctionner un agrégat. Si 40% du mélange initial est de l'essence, combien devrions-nous enlever du mélange (reste en rose) pour le remplacer par de l'essence pure (vert clair) pour que le mélange final (vert caci) contienne 60% d'essence ?

equation (55.5)

Nous noterons x l'inconnue qui est le nombre de litre du mélange initial à retirer et à remplacer par l'essence pure qui étant de quantité égale est aussi x. Nous avons alors :

equation (55.6)

Ce qui donne :

equation (55.7)

Il vient alors trivialement :

equation (55.8)

Donc 2.6 litres devraient être enlevés du mélange d'origine et être remplacer par 2.6 d'essence pure.

Bref voilà pour les mélanges. Nous pouvons aller beaucoup plus loin, et faire beaucoup plus compliqué avec plus d'inconnues mais nous nous arrêterons là pour l'instant.

18.2 Réactions

Puisque l'étude principale en chimie consiste à observer les résultats de mélanges de corps purs et/ou composés il convient d'abord de nous attarder sur les règles de bases qui régissent ces mélanges dans des conditions normales de pression et de température (C.N.T.P).

Il convient au préalable de préciser que nous n'allons pas étudier dans ce chapitre ce qui crée les liaisons entre les éléments car ceci est le rôle de la chimie quantique et moléculaire (voir chapitres précédents). De plus, nous insistons sur le fait que chaque élément théorique sera illustré d'un exemple pratique auquel il peut être utile de se reporter parfois pour mieux comprendre.

Considérons maintenant un système chimique fermé (sans transfert de masse donc !), nous traduisons la modification de la composition (s'il y a lieu et si elle existe) du système chimique par une équation de réaction de la forme (le système ne va pas toujours dans les deux sens !) :

equation (55.9)

appelée "équation de bilan" où les coefficients equation sont appelés "coefficients stoechiométriques" dans le sens où ils indiquent les "proportions d'or", rigoureusement appelées "proportions stoechiométriques", nécessaires tel qu'à des conditions normales la réaction puisse avoir lieu et où les equation sont les produits réagissant (purs ou composés) et les equation produits formés.

Attention ! Dans l'écriture de l'équation ci-dessus, nous imposons que tous equation sans exception réagissent à la récation chimique et donc que tous les equation sont dépendants.

Si les proportions d'or sont respectées (tels que les coefficients soient bien stoechiométriques !) et existent lors de l'écriture de l'équation de réaction, alors pour tout equation nous avons :

equation (55.10)

cette proposition n'est démontrable que si les coefficients stoechiométriques d'un côté ou l'autre de la réaction varient de manière proportionnelles. L'expérience montre que dans des conditions normales de température et de pression (C.N.T.P) cela est bien le cas.

Dès lors, la stoechiométrie de la réaction impose que s'il disparaît dans le système equation moles de equation , equation moles de equation avec respectivement une variation de matière des produits equation..., il apparaîtra en conséquence equation moles de equation, equation moles de

equation, ... avec respectivement une variation de matière des produits equation... en respectant les proportionnalités des coefficients stoechiométriques tel que nous puissions écrire "l'équation du bilan de matière" :

equation (55.11)

où equation est appelé "avancement élémentaire de la réaction" (fréquemment on prendra les valeurs absolues des rapports pour ne pas avoir à réfléchir sur le signe des variations).

La division des variations equation et equation par leurs coefficient stoechiométriques se justifie uniquement pour des raison de normalisation ayant pour objectif de rapporter equation à une valeur comprise entre 0 et 1 (soit entre 0

Ces dernière égalités indiquent simplement que si l'un des produits réactif disparaît en une quantité donnée, les autres produits réactifs voient leur quantité diminuée en rapport à leur coefficient stoechiométrique de manière à conserver la proportion d'or de la réaction.

L'écriture du bilanénergétique peut être allégé par l'introduction des coefficients stoechiométriques algébriques equation tels que : equation pour un produit formé, equation pour un produit réagissant.

Finalement nous pouvons écrire :

equation (55.12)

que nous retrouvons également souvent dans la littérature avec la valeur absolue au numérateur !

Dès lors, avec cette convention algébrique, l'équation de réaction comme elle existe, permet d'écrire :

equation (55.13)

ce qui signifie que la somme algébrique du nombre total de composés purs des réactifs et produits formés est toujours nulle.

Il est clair qu'à l'instant initial de la réaction nous choisissons pour l'avancement la valeur equation (sa valeur maximale étant égale à l'unité), instant auquel les quantités de matière sont equation.

L'intégration de l'expression différentielle du bilan de matière donne bien évidemment :

equation (55.14)

relation que nous retrouvons dans les tableaux d'avancements (voir plus bas) en se souvenant bien que equation pour un produit formé, equation pour un produit réagissant.

Se pose alors la question : Quelle est la valeur maximale equation de l'avancement d'une réaction ? Eh bien la réponse a cette question et ma foi fort simple. La valeur maximale d'avancement d'une réaction ayant les proportions stoechiométriques (nous respectons la traditions ainsi plutôt qu'en parlant de proportions d'or...) est telle qu'elle a lieu lorsque les réactifs auront tous disparus et dès lors il est nécessairement donné par :

equation (55.15)

pour ce que nous appelons le "réactif limitant", c'est-à-dire le produit réagissant qui disparaît (a toujours la plus petite valeur de molarité) en premier et qui arrête la réaction attendue !

Il peut être utile de définir le "taux d'avancement" equation donné par la grandeur intensive :

equation (55.16)

ce qui de manière formelle donne :

equation (55.17)

exempleExemple :

Considérons pour illustrer tous ces concepts la réaction (diazote avec hydrogène donnant de l'ammoniac) :

equation (55.18)

où les lettres latines représentent des corps purs (atomes) dont le nom ne nous importe pas. Les indices représentent tout simple la combinaison des atomes pour obtenir une molécule. Nous avons dans cette réaction :

equation et equation (55.19)

Le lecteur remarque que nous avons bien selon notre convention pour le bilan de masse :

equation (55.20)

Si nous considérons qu'il y a une mole de chaque corps composé, cela nous donne pour les proportions stoechiométriques (à un facteur equation près pour toutes les valeurs) :

equation (55.21)

Si à un moment equation donné, nous obtenons par mesure :

equation (55.22)

Quel est l'avancement de la réaction ?
equation (55.23)
soit autrement écrit nous en sommes à 10
Le taux de conversion de equation y relatif est donc :
equation (55.24)
Et quelle est la valeur maximale d'avancement equation du réactif limitant ?
Le ou les réactif(s) limitant sont bien évidemment de par la définition donnée par les chimistes (...) equation et equation puisqu'ils sont en quantité stoechiométriques dans cet exemple particulier.
Donc dans le cadre de l'exemple ci-dessus où nous avons equation pour le equation alors :
equation (55.25)
Les chimistes utilisent également souvent ce qu'ils appellent un "tableau d'avancement".
Voyons de quoi il s'agit avec notre exemple :
Équation
equation
+ equation
=equation
État initial
equation
1
3
0
État intermédi.
equation
equation
equation
equation
État final
equation
equation
equation
equation
Recherchons equation à partir de ce tableau. Le réactif limitant est donc soit equation, soit equation.
Donc pour equation :
equation (55.26)
et pour equation :
equation (55.27)
Le plus petit equation étant égal pour les deux réactifs, tous les réactifs sont ici limitant.

19

Chimie thermique

La thermochimie est la branche qui s'intéresse, historiquement, aux phénomènes thermiques et aux bilans qui accompagnent les réactions chimiques. Elle puise ses fondements principalement dans la thermodynamique.

Nous ne saurions donc que recommander au lecteur d'avoir parcouru le chapitre traitant de la thermodynamique dans la section de mécanique car de nombreuses notions seront supposées connues dans ce chapitre.

Par ailleurs, ce chapitre doit étudié en parallèle de celui de Chimie Analytique (cela peut être embêtant mais il faut faire avec).

19.1 Transformations chimiques

Soit le système fermé siège de la réaction chimique (cf. chapitre de Chimie Analytique) :

equation (56.1)

Nous allons, pour simplifier, considérer que la réaction chimique est totale et que les produits réagissants sont utilisées en quantités stoechiométriques (état 1 : equation) pour donner les produits formés, également en quantités stoechiométriques (état 2 :equation).

Si la transformation s'effectue à volume (quasi-)constant, le travail sur l'atmosphère ambiant est nul car (cf. chapitre de Thermodynamique) :

equation (56.2)

L'application du premier principe se réduit et permet alors d'écrire :

equation (56.3)

où equation est dans le cadre de la chimie thermique appelée "chaleur de réaction à volume constant", échangée bien sûr entre le système et le milieu extérieur (nous n'écrivons pas le delta pour indiquer qu'il s'agit d'une variation par tradition...).

Rappelons que :

- Si equation, la réaction est dite "endothermique" (le système reçoit de la chaleur du milieu extérieur)
- Si equation la réaction est dite "exothermique" (le système cède de la chaleur au milieu extérieur)
- Si equation la réaction est dite "athermique" (pas d'échange de chaleur avec le milieu extérieur)

Remarque Rappelons aussi qu'un système fermé n'est pas un système isolé ! Pour un rappel des différentes définitions nous renvoyons le lecteur au chapitre de Thermodynamique.

Si la réaction se fait à pression constante (cas le plus courant dans la pratique), soit isobare, nous avons alors :

equation (56.4)

Remarque Le choix des indices d'intégration sont différents à précédemment pour différencier le fait qu'une réaction à pression ou à volume constant ne sont pas nécessairement identiques.

L'application du premier principe de la thermodynamique, entre les deux états permet d'écrire :

equation (56.5)

où equation est la quantité de chaleur, appelée "chaleur de réaction à pression constante", échangée entre le système et le milieu extérieur (equation est donc une variation... même si la notation traditionnelle malheureuse des thermodynamiciens ne le met pas en évidence...).

En utilisant la définition de l'enthalpie, nous pouvons écrire cette dernière relation sous la forme :

equation (56.6)

Si nous travaillons avec les volumes molaires, ceux des phases condensées (solides et liquides et donc) étant négligeables devant le volume molaire gazeux, seuls les constituants gazeux auront une enthalpie très différente de leur énergie interne (voir l'exemple dans le chapitre de Thermodynamique). Nous aurions dès lors dans le cadre de l'approximation des gaz parfaits (cf. chapitre de Thermodynamique) :

equation (56.7)

Dans le cadre des gaz parfaits, la relation antéprécédente peut donc s'écrire :

equation (56.8)

Or, comme (cf. chapitre de Mécanique Des Milieux Continus) equation et equation sont tout deux les mêmes états finaux d'une même réaction complète et que pour un gaz mono-atomique :

equation (56.9)

donc l'énergie interne equation et equation ne dépend que du nombre de constituants mais... ils sont égaux puisque même état final d'une même réaction !

Ainsi, nous avons :

equation (56.10) .

En posant equation (différence entre le nombre de moles de gaz des produits formés et ceux des produits réagissant), nous pouvons écrire pour une réaction chimique :

equation (56.11)

ce qui permet de différencier l'énergie mise en jeu entre réaction isobare et isochore et de chercher le meilleur choix à faire en fonction des objectifs industriels.

Il est intéressant de remarque que si le delta du nombre de moles est nul. Les variations de chaleur isobare ou isochore sont égales et qu'il n'y a donc pas de raison de préférer à priorie l'une ou l'autre des transformations.

Évidemment dans la pratique le problème consiste à connaître les valeurs des différentes variables de la relation. Ces valeurs se trouvant dans d'immenses bases de données auxquelles les chimistes ont accès... cette relation n'est que très très rarement utilisée dans la pratique et de toute façon elle est basée sur des hypothèses simplificatrices trop contraignantes pour être d'un vrai intérêt pratique.

19.2 Grandeurs molaires

Définition 19.1 *D1. Par convention, la "mole" est la quantité de matière d'un système contenant autant d'espèces chimiques qu'il y a d'atomes de carbone dans 12 [g] de carbone 12 (cf. chapitre de Physique Nucléaire).*

Le nombre d'atomes de carbone contenus dans 12 [g] de carbone est égal au nombre d'Avogadro donné approximativement par :

equation (56.12)

Cela signifie in extenso et par construction qu'un mole d'eau, de fer, d'électrons, contient respectivement toujours un nombre d'atomes égale au nombre d'Avogadro.

Remarquons qu'avec un système mélangé c'est un non sens mathématique de faire la somme des masses molaiers des constituants pour avoir la masse molaire totale. La masse molaire est donc une grandeur intensive !

Définition 19.2 *D2. La "masse molaire atomique" est la masse d'une mole d'atomes des éléments chimiques concernés.*

REMARQUE Nous trouvons ces masses molaires atomiques dans les classifications périodiques. Mais il faut surtout savoir que celles qui sont indiquées prennent en compte les isotopes naturels (ce qui est bien normal puisque chimiquement ils sont indiscernables à part pour le chimiste nucléaire ou le physicien nucléaire). Donc la valeur indiquée dans les tables se calcule comme la somme des proportions respectives des masses molaires des différents isotopes (la validité de cette manière de calculer est toute relative...).

Définition 19.3 *D3. La "masse molaire moléculaire" est égale à la somme des masses molaires atomiques des éléments chimiques qui la constitue.*

Il vient dès lors immédiatement le constat suivant : la masse m d'un échantillon constitué d'une quantité de n moles d'espèces chimiques tout identiques de masse molaire equation est donnée par la relation :

equation (56.13)

De manière un peu plus formelle et sous un aspect thermodynamique, voici aussi comment nous pouvons définir la masse molaire :

Soit X une grandeur extensive relative à un système monophasique (voir le chapitre de Thermodynamique pour le vocabulaire dont il est fait usage) et soit un élément de volume dV de ce système, autour d'un point courant M et contenant la quantité de matière dn. Nous lui associons la grandeur extensive dX proportionnelle à dn tel que :

equation (56.14)

ce qui fait que equation est une grandeur intensive (quotient de deux grandeurs extensives selon ce qui a été vu dans le chapitre de Thermodynamique) que nous appellerons par définition "grandeur molaire associée" à X.

Nous en déduisons que :

equation (56.15)

l'intégrale portant sur tout le système monophasique.

Dans le cas d'une phase uniforme, equation étant constant en tout point, nous pouvons simplement écrire :

equation (56.16)

Remarque En gros l'idée est de dire que la masse d'un système chimique monophasique est proportionnelle à la masse molaire de celui-ci à un facteur entier près représentant le nombre de ses constituants (où le nombre de moles pour être plus exact).

Définition 19.4 *D4. Dès lors que le système est hétérogène, nous utilisons le concept de "titre molaire", défini par :*

equation (56.17)

equation étant donc le titre molaire d'une espèce equation dont la quantité de matière (le nombre de moles par exemple) est equation avec equation étant la quantité de matière totale de la phase envisagée.

Il en résulte que, pour toutes les espèces de la phase envisagée, equation, ce qui veut dire que, s'il y a n espèces chimiques, il suffit de connaître n-1 titres molaires pour toutes les connaîtres.

Si la phase considérée est gazeuse et en considérant un gaz parfait suivant la loi de Boyle-Mariotte (approximation de l'équation de Van der Waals démontrée dans le chapitre de Mécanique Statistique) nous avons :

equation et equation (56.18)

nous avons dès lors aussi la possibilité dans le cas de phases gazeuses d'exprimer le titre molaire comme :

equation (56.19)

Remarque Nous pourrions faire de même avec le volume.

Définition 19.5 *D5. Nous définissons le "titre massique associée à l'espèce" equation par le rapport :*

equation (56.20)

avec equation étant la masse totale de la phase étudiée. Nous avons bien sûr également equation.

Définition 19.6 *D6. Nous définissons la "concentration molaire volumique" ou "molarité" par le rapport (attention à ne pas confondre la notation avec la chaleurs massiques) :*

equation (56.21)

Remarque Il existe d'autres grandeurs de composition beaucoup moins utilisées que equation ou equation. Nous pouvons citer la "concentration massique volumique" equation, la "molalité" (quotient de la quantité de matière de l'espèce equation par la masse du solvant", etc.

En chimie thermique, nous considérons très souvent les gaz comme parfaits et nous faisons jouer un rôle privilégié à une pression particulière appelée "pression standard" equation et dont la valeur vaut (pression atmosphérique) :

equation (56.22)

Définition 19.7 *D6. Nous disons qu'un gaz (parfait) est dans l'état standard si sa pression est égale à la pression standard equation de 1 [bar].*

Définition 19.8 *D7. Nous appelons "grandeur molaire standard" equationd'un constituant la valeur de la grandeur molaire de ce même constituant pris à l'état standard, c'est-à-dire sous la pression equation.*

REMARQUES

R1. Toute grandeur molaire standard est bien évidemment intensive : la pression étant fixée par l'état standard, celle-ci ne dépend plus que de la température.

R2. Toute grandeur standard est notée grâce à l'exposant "°" : equation est donc le volume molaire standard. Par contre, la grandeur molaire standard n'est pas toujours spécifiée avec le petit m en indice, il faut donc parfois être prudent avec ce que l'on manipule.

Dans le cas du gaz parfait, le volume molaire se calcule grâce à l'équation d'état du gaz parfait. il en résulte :

equation (56.23)

Nous constatons bien évidemment que le volume molaire standard du gaz parfait dépend de la température.

Si nous faisons ce calcul aux "conditions normales de température et de pression" (abrégé es C.N.T.P.), c'est-à-dire à une température de 293.15 [K] (soit 20 [°C]) et à une pression de 1 [atm] (soit 101.325 [kPa]), nous trouvons alors un volume de 22.4 [L/mol] qui est une valeur très connue par les chimistes.

REMARQUES

1. Dans une large gamme de températures et de pressions, le volume molaire des gaz réels n'est généralement pas très différent de celui d'un gaz parfait.
2. Dans le cas d'un état condensé, nous ne disposons pas en général d'une équation d'état mais nous pouvons néanmoins mesurer le volume molaire.

Nous pouvons définir ainsi par extension d'autres grandeurs standards découlant de celles que nous avions définies en thermodynamique :

- "L'énergie interne molaire standard" (grandeur intensive puisque exprimée par unite molaire) equationequation
- "L'enthalpie molaire standard" (grandeur intensive puisque exprimée par unite molaire) equation avec :
  equation (56.24)
  Il est important que le lecteur observe que l'enthalpie ne dépend que de la température (et de l'énergie interne).

REMARQUE Pour les états condensés le volume standard est très faible en unités S.I. ce qui fait que equation. Il est cependant très difficile de parler de pression pour les états condensés donc cette approximation est à prendre avec des pincettes.

Si nous considérons maintenant une fonction extensive X (comme le volume par exemple !) définie sur un système chimique gazeux en évolution. Nous pouvons à priori, exprimer X en fonction des deux variables intensives T, P (car une fonction extensive est soit produit/rapport de deux grandeurs intensives, soit une somme de grandeurs extensives) et des différentes quantités de matières equation de equation tel que :

equation (56.25)

Si tous les produits (réagissants et formés) sont dans leur état standard la fonction extensive, notée dès lors equation, se met sous la forme :

equation (56.26)

où la pression n'intervient plus puisque fixée à sa valeur standard. Le gaz n'est alors décrit par sa température et de la quantité de ses constituants !

Si nous considérons cependant une évolution infiniment petite du système à température et pression constantes (car nous supposons les proportions des composants du gaz comme variant toujours en proportions) les différentes quantités de matières varient alors selon la différentielle total exacte (cf. chapitre de Calcul Différentiel et Intégral) :

equation (56.27)

où n'interviennent bien évidemment par la contrainte précitée (température et pression constantes) que les différentielles par rapport aux quantités de matières qui varient.

Nous pouvons alors être amenés à définir artificiellement (rien ne nous en empêche, ce n'est pas faux !) la grandeur molaire standard intensive qui ne dépend que de la température :

equation (56.28)

Dès lors :

equation (56.29)

mais nous avons aussi défini en dans le chapitre de Chimie Analytique la relation :

equation (56.30)

qui exprimait, rappelons-le, la variation de la quantité de matière d'un des composés de la réaction par rapport à sa proportion stoechiométrique (constante) et l'avancement de la réaction. Il vient dès lors :

equation (56.31)

ainsi que :

equation (56.32)

Par définition, nous appelons dès lors "grandeur standard de réaction associée à la fonction extensive X" la grandeur (notation mal choisie par les chimistes...) :

equation (56.33)

qui est une grandeur intensive qui ne dépend que de la température. Cette relation peut aussi s'écrire :

equation (56.34)

De manière générale, les chimistes appellent "opérateur de Lewis", noté equation, la dérivée d'une grandeur X (standardisée ou non), par rapport à l'état d'avancement de la réaction equationà température et pression constantes.

Remarque Le symbole figure avec la lettre r en indice pour bien montrer qu'il s'agit d'une grandeur de réaction. En d'autres termes, il s'agit de variation de la grandeur molaire standard considérée au cours de la réaction concernée pour un avancement de réaction d'une mole à une pression de 1 bar pour un gaz parfait.

Il ne faut par ailleurs pas oublier que les coefficients stoechiométriques des réactifs sont positifs et ceux des produits négatifs (cf. chapitre de Chimie Analytique).

Il existe deux grandeurs de réaction jouant des rôles importants en chimie :

1. L'énergie interne molaire de réaction, appelée plus souvent "énergie interne de réaction standard", d'un système chimique :
   equation (56.35)
2. L'enthalpie molaire de réaction, appelée plus souvent "enthalpie de réaction standard", d'un système chimique :
   equation (56.36)

19.3 Enthalpie standard de réaction

Dès lors nous pouvons considérer les deux cas suivants suivant en connaissant la relation (cf. chapitre de Thermodynamique) :

equation : (56.37)

1. Si les equation sont dans un état condensé, puisque la pression interne ne s'applique pas nous avons :

equation (56.38)

qui reste quand même à prendre avec des pincettes suivant les scénarios !

2. Si les equation sont à l'état gazeux (supposés parfaits) :

equation (56.39)

Il en résulte que seuls les gaz vont intervenir dans cette relation :

equation (56.40)

que nous écrivons conventionnellement :

equation (56.41)

Il en résulte que dans le cas particulier où equation (qui est au fait une notation... malheureuse... pour la somme algébrique des proportions stoechiométriques est nulle) pour une température donnée alors :

equation (56.42)

où il faut se rappeler que les coefficients stoechiométriques des produits sont comptés comme positifs, alors que ceux des réactifs sont négatifs (cf. chapitre de Chimie Analytique).

Ainsi, la variation de la fonction enthalpie correspond à la variation de la quantité de chaleur mise en jeu dans une transformation isobare à une température T donnée. Raison pour laquelle celle-ci est parfois notée equation.

Une réaction chimique qui a une enthalpie de réaction (qui est pour rappel la variation instantanée de l'enthalpie pendant une réaction) négative est dite exothermique, puisqu'elle libère de la chaleur dans l'environnement (pression constante oblige par définition de cette enthalpie de réaction !), alors qu'une réaction chimique dont l'enthalpie de réaction est positive est dite endothermique puisqu'elle nécessite alors un apport de chaleur pour se produire.

Ainsi, selon les développements précédents, si nous notons avec un indice p les produits et avec un indice i les réactifs, nous trouvons souvent l'enthalpie standard de réaction sous la forme suivante si les coefficients stoechiométriques sont tous comptés comme positifs :

equation (56.43)

sous cette forme, nous voyons alors bien que l'enthalpie standard de réaction correspond donc à la différence des enthalpies molaires partielles entre les produits et les réactifs. Il s'agit ni plus ni moins de la "loi de Hess" définie au 19ème siècle par le chimiste suisse Henri Hess.

Puisque dans un système à l'équilibre, l'énergie initiale est toujours supérieure ou égale à l'énergie finale (les systèmes tendent toujours à aller vers un état plus stable ayant une énergie minimale), alors l'enthalpie standard de réaction equation ne peut être que négative ou nulle.

Si une réaction chimique à pression constante et à une température donnée ne donne qu'un seul et unique composé chimique (produit) alors l'enthalpie standard de réaction est appelée "enthalpie standard de formation" et notée equation.

Au fait, l'intérêt de la relation antéprécédente est que le chimiste peut simplement, sans avoir à connaître les quantités de matières mises en jeu, déterminer juste en connaissant les coefficients stoechiométrique d'une réaction chimique gazeuse ou condensée (s'il accepte qu'il s'agira alors d'une approximation pour ce dernier cas) isobare si la variation instantanée de l'énergie interne molaire lors de l'avancement d'une réaction à une température donnée est égale à la variation instantanée de l'enthalpie molaire.

Deux cas pratiques se présentent alors :

1. La différence entre la variation instantanée de l'énergie interne molaire et l'enthalpie molaire est nulle : Dès lors, la réaction chimique (à la température donnée) n'occupe pas instantanément un volume plus grand et ne perd donc aucune énergie à repousser ("inutilement") la pression du gaz environnant la réaction étudiée (cela peut être vu comme une économie d'argent en termes énergétiques dans une industrie chimique). Dans ce cas, l'enthalpie standard de réaction est égale simplement à la chaleur de réaction à pression constante equation.
2. La différence entre la variation instantanée de l'énergie interne molaire et l'enthalpie molaire est positive : Dès lors, la réaction chimique (à la température donnée) occupe instantanément un volume plus grand et perd donc une part d'énergie à repousser ("inutilement") la pression du gaz environnant la réaction étudiée (cela peut être vu comme une perte d'argent en terme de rendement énergétique dans une industrie chimique).

REMARQUE Évidemment il est tout à fait possible d'imaginer une entreprise qui tire profit de la variation du volume d'une réaction (le cas 2) qui repousse la gaz environnant avec un système de piston pour produire ensuite de l'énergie mécanique... dont il est serait possible dans certaines situations de ne jamais perdre d'argent.

Une petite difficulté se présente cependant...l'enthalpie standard d'un corps pur simple (corps formé que d'un seul type d'atomes) ne peut être calculée de façon absolue car elle dépend de l'énergie interne qui est très difficilement calculable (il faut utiliser les outils de la physique quantique qui soulèvent des problèmes insurmontables aujourd'hui). Cela suppose de définir une échelle

arbitraire d'enthalpies molaires en définissant un zéro arbitraire d'enthalpie et adopté comme à l'échelle internationale (ce qui n'est malheureusement pas le cas !).

Ainsi, pour pouvoir dresser des tables d'enthalpies molaires standards il a été choisi de définir l'échelle d'enthalpie de la manière suivante : l'enthalpie molaire standard d'un corps pur simple stable dans l'état standard est nulle à 298 [K]. Il s'ensuit que l'enthalpie de formation standard d'un corps pur simple est donc toujours nulle.

ANNEXES

A Repères biographiques 519

A

Repères biographiques

Cette annexe présente une poignée d'hommes qui postulent une étrange renommée. Selon les règles de l'histoire qu'on enseigne à l'école primaire, ils n'existent pas, ils n'ont commandé aucune armée, ils n'ont envoyé personne à la mort, ils n'ont dirigé aucun empire et ils n'ont eu qu'une part minime dans les grandes décisions historiques. Certains ont acquis quelque célébrité, mais aucun ne fut jamais un héros national. Pourtant leur oeuvre a davantage influencé le cours de l'histoire que bien des actes accomplis par des hommes d'état auréolés d'une gloire très supérieure. Elle a produit plus de boulversements que le va-et-vient des armées en bataille par-dessus les frontières, elle a fait plus pour le bonheur ou le malheur que les édits des rois et des assemblées. Car leur oeuvre, c'est d'avoir façonné l'esprit de l'homme.

Qui propage ses idées manie un pouvoir bien supérieur à celui de l'épée ou du sceptre : aussi ont-ils façonné et dirigé le monde. Pour la plupart, ils n'ont pas levé le moindre petit doigt pour agir physiquement ; ils ont travaillé essentiellement en intellectuels, dans le silence et l'oubli, sans se soucier outre mesure du monde environnant. Mais, dans leur sillage, des empires se sont écroulés et des continents disloqués, des régimes politiques se sont soit renforcés soit érodés, les classes se sont dressés les unes contre les autres, ainsi que les nations. Non pas sous l'effet d'un noir complot, mais de par la puissance extraordinaire de leurs idées. Qui sont ces hommes (femmes) ? : des savants, économistes, chimistes, biologistes, mathématiciens, physiciens, informaticiens, ingénieurs,...

Les biographies ci-dessous des scientifiques les plus connus à travers le monde sont triées par ordre alphabétique. Si vous souhaitez que nous rajoutions une entrée, il vous suffit de nous envoyer par e-mail les nom et prénom de la personne concernée et la raison pour laquelle vous aimeriez la voir figurer dans la liste ci-dessous. Nous étudierons ensuite la proposition et prendrons la décision qui s'impose quant au choix de l'insérer.

Les tailles des biographies ne sont pas proportionnelles au nombre d'articles publiés ou découvertes effectuées, mais à la quantité d'informations trouvées sur ces personnages sur l'Internet ou dans la littérature.

La liste n'est pas exhaustive mais son objectif est rendre hommage et de se remémorer les grands hommes qui ont fait des sciences pures et exactes ce qu'elles sont aujourd'hui et qui ont consacré une partie ou l'entier de leur vie à la des sciences, l'art le plus contraint.

Attention ! En physique (aussi bien qu'en mathématique) qu'une théorie, une équation, une constante ou autre porte rarement le nom de son vrai inventeur. Ce fait est largement connu dans la chez les scientifiques et est souvent source de moqueries de la communauté.



Alembert Jean le Rond (1717, Dunedin, New Zealand — 1783, Edinburgh, Scotland) — Enfant naturel d'un commissaire d'artillerie, abandonné sur les marches de la chapelle parisienne de Saint-Jean-Le-Rond, le futur grand philosophe, mathématicien et physicien est recueilli par un vitrier qui recevra secrètement une pension pour subvenir à l'éducation du jeune garçon qui étudiera brillamment le droit, la médecine et les mathématiques. Suite à la publication de divers mémoires (sur le calcul intégral, sur la réfraction des corps solides), d'Alembert entre à l'Académie des sciences (1741). On lui doit le célèbre principe de la quantité de mouvement, dit "principe de d'Alembert" dans son Traité de dynamique (1743). En astronomie, il est l'auteur (1749) d'un traité sur la précession des équinoxes qu'il explique au moyen de la théorie de la gravitation universelle de Newton et d'une solution partielle au

problème des trois corps. D'Alembert établit aussi une théorie mathématique des cordes vibrantes en étudiant la nature composite du son (harmoniques).



Ampère, André-Marie (1775-1836) à 18 ans il connaît la majeur partie des oeuvres mathématiques de son temps. Mathématicien de premier ordre, il montre comment l'on doit utiliser cette science, qu'il considérait comme une branche de la philosophie, à l'étude des découvertes des faits physiques pour en donner une relation définitive. En quelques semaines, Ampère établit les bases de toute une science à laquelle il donne le nom d'électromagnétisme. Il cherche à comprendre le magnétisme des aimants et en tire un hypothèse de "courants particulaires" (orbites électroniques et orientation du spin aujourd'hui).Il établit également l'égalité du nombre de molécules dans des volumes égaux de Gaz de natures différentes, mais mesurés dans des conditions identiques de température et de pression (observation expérimentale de Gay-Lussac).



Archimède, De Syracuse (287-212 av. J.-C.), mathématicien et ingénieur grec célèbre à la fois comme mécanicien théoricien et comme constructeur de machines. Archimède de Syracuse eut une production mathématique exceptionnelle, dont une partie nous est parvenue dans des traités comme Sur la sphère et le cylindre ; la Mesure du cercle ; la Quadrature de la parabole ; Des spirales ; Des conoïdes et sphéroïdes ; la Méthode, Des corps flottants... C'est à partir de ses travaux mécaniques que les principales anecdotes le mettant en scène, comme celle du levier ou du bain, vont se constituer. La célèbre maxime : «Donnez-moi une place où me tenir et je mettrai la terre en mouvement» est un écho populaire de la contribution archimédienne à la statique, exposée dans le traité des Équilibres. Archimède démontre la loi du levier, introduit la notion fondamentale de centre de gravité, et détermine ces barycentres pour les principales figures géométriques planes. Il en est de même pour l'anecdote d'Archimède, jaillissant nu de son bain, en criant « Eurêka », parce qu'il venait, dit-on, de trouver le moyen de résoudre le problème que lui avait posé le roi Hiéron. En fait, le récit est une mise en scène spectaculaire de la découverte du principe fondamental de l'hydrostatique (communément appelé depuis "principe d'Archimède"). En géométrie, l'œuvre d'Archimède développe celle d'Eudoxe de Cnide telle que nous la connaissons par le livre XII des Éléments d'Euclide : il s'agit de comparer les mesures des figures planes et solides, en particulier des figures curvilignes. Ainsi Archimède démontre que le volume du cylindre circonscrit à une sphère est égal à une fois et demie le volume de celle-ci et que la surface latérale du cylindre est égale à celle de la sphère ou quatre fois la surface d'un grand cercle. Donc, si l'on sait calculer la surface du cercle, on connaîtra celle de la sphère, du cylindre, son volume et celui de la sphère, etc. En plus des résultats déjà cités, Archimède a réussi la quadrature d'un segment de parabole (il est égal aux quatre tiers du triangle inscrit qui a même base et hauteur) et la cubature de certains conoïdes et sphéroïdes (solides de révolution engendrés par une portion de conique). Son résultat le plus célèbre est cependant le plus simple et concerne le cercle. Archimède ramène sa quadrature à un autre problème : la rectification de sa circonférence, c'est-à-dire «trouver une ligne droite égale qui lui soit égale», problème qu'il résout à l'aide d'une courbe géométrique qu'on appelle désormais "spirale d'Archimède". En outre, il calcule des valeurs approchées du rapport circonférence/diamètre (ce que nous appelons le nombre "Pi").



Avogadro, Amedeo conte di Quaregna e di Ceretto (1776-1856), Fils d'un magistrat de Turin, Amadeo Avogadro commence par suivre la voie paternelle. Il passe une licence de droit en 1795 et s'inscrit au barreau de sa ville natale. Mais son goût pour la physique et les mathématiques, auxquelles il n'a cessé de s'intéresser en solitaire, le pousse à entamer sur le tard des études scientifiques. En 1809, il fait une communication à l'Académie royale de Turin ; le succès qu'il remporte grâce à elle lui permet d'obtenir un poste de professeur au Collège royal de Verceil. En 1820, l'Université de Turin crée pour lui une chaire de physique qu'il gardera jusqu'à la fin de sa vie. C'est en étudiant les lois régissant la compression et la dilatation des gaz qu'Avogadro énonce en 1811 l'hypothèse restée célèbre sous le nom de "loi d'Avogadro". Reposant sur la théorie atomique de Dalton et la loi de Gay-Lussac sur les rapports volumiques, la théorie d'Avogadro indique que deux volumes égaux de gaz différents, dans les mêmes conditions de

température et de pression, contiennent le même nombre de molécules. Sous son apparente simplicité, cette loi comporte des implications importantes ; grâce à elle, il devient possible de déterminer la masse molaire d'un gaz à partir de celle d'un autre. Mais les chimistes de l'époque, plus intéressés par les expériences, boudent quelque peu les études théoriques d'Avogadro qui ne seront d'ailleurs universellement reconnues que cinquante ans plus tard. Le nom d'Avogadro reste également lié à celui du nombre d'Avogadro (6.023.10E23) indiquant le nombre de molécules contenues dans une seule mole.



Bachelier, Louis (1870-1946) est né au Havre dans une famille de négociants. Il apparaît à sa majorité sur les listes électorales du Havre en 1892 comme représentant de commerce à la même adresse professionnelle que son père. Après avoir effectué son service militaire, à l'âge de 22 ans il reprend ses études à la faculté des sciences de Paris. Elles sont couronnées par une licence ès sciences en 1895 (mention passable) et par la soutenance en 1900 de sa non moins fameuse et méconnue thèse de doctorat en mathématiques. Bien que cette thèse soit considérée aujourd'hui comme un travail précurseur en théorie des probabilités et en théorie financière, elle ne vaut à l'époque à son auteur qu'une mention honorable. De 1913 à 1914 Bachelier dispensa un cours libre de théorie des probabilités appliquées à la mécanique, la balistique et la biométrie. Il fut également chargé de conférences supplémentaires sur les mathématiques générales de 1913 à 1914. Ce n'est qu'après la guerre de 1914-1918 qu'il obtient un premier poste de chargé de cours à la faculté des sciences de Besançon. Après divers remplacements à Dijon puis à Rennes, il revient à Besançon en 1927 comme professeur titulaire de la chaire de calcul différentiel et intégral, poste qu'il occupe jusqu'à sa retraite en 1937. Louis Bachelier a parmi ses nombreux travaux été le premier a avoir introduit la continuité dans les problèmes de probabilité en prenant le temps comme variable. En particulier, il a élaboré une théorie mathématique du mouvement brownien cinq ans avant Albert Einstein. Il est également bien avant Norbert Wiener, le premier à avoir défini la fonction du mouvement brownien et donné un grand nombre de ses propriétés.



Banach, Stefan (1892-1945) mathématicien polonais qui a posé les bases de l'analyse fonctionnelle. Né à Cracovie en 1892, en Autriche-Hongrie (actuellement ville polonaise). Banach fit ses études secondaires à Cracovie ; il se révéla particulièrement brillant en mathématiques et en sciences naturelles, mais son désintérêt pour les autres matières l'empêcha d'obtenir les meilleures mentions. La vie (au moins mathématique) de Banach va basculer au printemps 1916, quand il rencontre Steinhaus à Cracovie. Avec Otto Nikodym, ils décident de fonder une société mathématique. La recherche mathématique de Banach commence là. Son premier article est cosigné avec Steinhaus. Steinhaus lui avait parlé d'une propriété qu'il ne parvenait pas à démontrer, et après quelques jours de réflexion, Banach exhiba un contre-exemple. Il est difficile de dire ce qu'il serait advenu de l'activité mathématique de Banach sans la rencontre avec Steinhaus, mais toujours est-il qu'il entama à compter de celle-ci une recherche intense et fructueuse. Banach retourne à Lvov en 1920 où un poste d'assistant lui est proposé. Il soutient sa thèse en 1922, et c'est dans cette thèse qu'apparaissent pour la première fois la notion d'espace de Banach, qu'y sont démontrés les théorèmes fondamentaux sur ces objets, où on y évoque la topologie faible... Bref, cette thèse marque la naissance de l'analyse fonctionnelle. En 1929, il fonde avec Steinhaus la revue mathématique Studia Math, consacrée au développement de l'analyse fonctionnelle, et en 1939 il est élu président de la société mathématique de Pologne. En 1945, peu avant la fin de la Seconde Guerre Mondiale, il décède d'un long cancer. De nombreux théorèmes sont associés au nom de Banach, qu'il les ait démontrés lui-même, ou qu'ils fassent référence à ces idées. Citons entre autres : le théorème de Hahn-Banach de prolongement des formes linéaires continues, le théorème de Banach-Steinhaus, de Banach-Alaoglu, le théorème du point fixe de Banach, ainsi que le paradoxe de Banach-Tarski.



Bell, John (1928-1990) fut dès la plus petite enfance attiré par les livres traitant des sciences. À cause de problèmes financiers familiaux il ne put poursuivre immédiatement des études académiques. Il travailla donc pendant une année en tant que technicien au département de physique de l'université de Queen's à Belfast avant de devenir étudiant en 1945 dans ce même département. Il sortit premier

de sa promotion en mathématiques-physique. Bell trouva dans les années 60 une nouvelle inspiration dans les bases de la théorie quantique, une matière supposée épuisée par les résultats de la discussion de Bohr-Einstein trente ans plus tôt, et ignorée par pratiquement tous ceux qui ont employé la théorie quantique entre-temps. Effectivement, Bell était intrigué par les incertitudes quantique de Heisenberg et voulait creuser le sujet en montrant que la discussion de tels concepts comme le "réalisme", le "déterminisme" et la "localité" pouvaient êtres affiliés dans un rapport mathématique rigoureux : "les inégalités de Bell" vérifiables expérimentalement. Bell poussa très loin les doutes qu'il avait sur les principes d'incertitudes au point qu'il en irrita même son professeur (Sloane) qui lui fit remarquer que maintenant il allait un peu trop loin ! Bell attendit son travail de thèse pour développer ses idées. Malheureusement à nouveau à cause de problèmes financiers, il du repousser ses recherches à plus tard et joindre le centre anglais de recherche atomique à Harwell. Pendant sa carrière il épousa une femme (Mary Bell) qui l'aida dans le développement de ses travaux sur les principes fondamentaux de la théorique quantique. C'est en 1951 avec Rudolf Peierls que Bell développa sa célèbre théorie C.P.T. (Charge, Parity, Time). Malheureusement pour Bell, Gerhard Lüders et Wolfgang Pauli arrivèrent au même résultat dans la même période et c'est à eux que furent attribué les crédits de la découverte. Les développements théoriques de Bell sont à l'origine de la cryptographie et de la théorique de l'information quantique. L'attention à la théorie quantique de l'information a énormément augmentée au cours des dernières années, et le sujet semble sûr d'être l'un des secteurs scientifiques dont la croissance sera la plus importante au 21ème siècle. Un autre travail de première importance de Bell en 1969 fut la participation au développement de "l'anomalie A.B.J." (Adler-Bell-Jackiw) dans la théorie quantique des champs. Ces trois physiciens montrèrent que le modèle algébrique standard contentait une erreur. Effectivement la quantification du modèle des champs brise une symétrie. Bell fut nommé pour le prix Nobel, qu'il aurait certainement obtenu s'il n'était pas décédé d'une attaque cérébrale en 1990.



Bernoulli, Daniel (1700-1782), savant suisse qui découvrit les principes de base du comportement d'un fluide. Il étudia l'écoulement des fluides et formula le principe selon lequel la pression exercée par un fluide est inversement proportionnelle à sa vitesse d'écoulement. Il utilisa des concepts atomistes pour ébaucher la première théorie cinétique des gaz, en exprimant leur comportement en termes de probabilités sous des conditions particulières de pression et de température. On peut le considérer comme l'un des fondateurs de l'hydrodynamique.



Bessel, Friedrich (1784-1846) Né à Minden en Westphalie, Bessel commença à travailler très jeune comme commis. Attiré par la navigation maritime, il s'intéressa aux observations nautiques, construisant lui-même son sextant et étudiant l'astronomie à ses heures de liberté. Il calcula la trajectoire de la comète de Halley, résultat qui fut immédiatement publié et lui permit d'obtenir, en 1806, un emploi d'assistant à l'observatoire de Lilienthal. En 1810, il devint directeur du nouvel observatoire de Königsberg, tout en poursuivant des études mathématiques. Il dut enseigner les mathématiques à ses étudiants en astronomie jusqu'en 1825 (date à laquelle Jacobi vint enseigner cette matière à Königsberg). Toute sa vie fut consacrée à l'astronomie (il écrivit plus de 350 articles) et, peu avant sa mort, il commença l'étude du mouvement d'Uranus, problème qui devait aboutir à la découverte de Neptune. En mathématiques, Bessel est connu pour avoir introduit les fonctions qui portent son nom, les utilisant pour la première fois, en 1817, lors de l'étude d'un problème de Kepler, et les employant plus complètement sept ans plus tard pour étudier les perturbations planétaires.



Bohr, Niels Henrik David (1885-1962), physicien danois, prix Nobel en 1922, pour sa contribution à la physique nucléaire et à la compréhension de la structure atomique. La théorie de Bohr sur la structure atomique, pour laquelle il reçut le prix Nobel de physique en 1922, fut publiée entre 1913 et 1915. Son travail s'inspira du modèle nucléaire de l'atome de Rutherford, dans lequel l'atome est considéré comme un noyau compact entouré d'un essaim d'électrons. Le modèle pose en principe que l'atome n'émet de rayonnement électromagnétique que lorsqu'un

électron se déplace d'un niveau quantique à un autre. Ce modèle contribua énormément aux développements ultérieurs de la physique atomique théorique.



Boltzmann, Ludwig (1844-1906), physicien autrichien qui contribua à établir les bases de la mécanique statistique. Ayant fait ses études à Vienne et à Oxford, il enseigna la physique dans différentes universités allemandes et autrichiennes pendant plus de quarante ans. Développant la théorie cinétique des gaz, notamment à partir des travaux de Maxwell, il établit que la seconde loi de la thermodynamique pouvait être obtenue sur la base de l'analyse statistique. Calculant le nombre de particules dotées d'une énergie donnée, il établit la statistique dite de Maxwell-Boltzmann. Il exprima l'entropie S d'un système en fonction de la probabilité W de son état. Il put aussi établir de manière théorique la loi de Stefan relative au rayonnement d'un corps noir. Mais il lui fallut expliquer comment les principes mécaniques, où les phénomènes sont réversibles, pouvaient engendrer des lois thermodynamiques décrivant des phénomènes marqués par l'irréversibilité. Il avança l'idée que les évolutions irréversibles, quoiqu'elles ne soient que des possibilités parmi d'autres, sont si probables que ce sont pratiquement toujours elles qui se produisent.



Boole, Georges (1815-1864) Mathématicien et logicien anglais, Boole est le créateur de la logique symbolique. Né à Lincoln et fils d'un petit commerçant, il reçut ses premières leçons de mathématiques de son père, qui lui apprit aussi à fabriquer des instruments d'optique. En dehors des conseils de son père et de quelques années passées dans les écoles locales, Boole est un autodidacte. Quand les affaires de son père déclinèrent, il fut obligé de travailler pour aider sa famille et, dès seize ans, il enseigna dans des écoles de village ; à vingt ans, il ouvrit sa propre école à Lincoln. Pendant ses loisirs, il étudiait les mathématiques à l'Institut de mécanique, créé vers cette époque ; c'est là qu'il se familiarisa avec les Principia de Newton, la Mécanique céleste de Laplace et la Mécanique analytique de Lagrange et qu'il commença à résoudre des problèmes d'algèbre supérieure. Boole soumit au nouveau Cambridge Mathematical Journal une série d'articles originaux dont le premier est Recherches sur la théorie des transformations analytiques ; ces articles portaient sur les équations différentielles et sur les invariants par transformation linéaire. En 1844, il étudie les liens entre l'algèbre et le calcul infinitésimal dans un important mémoire publié dans les Transactions de la Royal Society, qui lui décerne une médaille cette même année pour sa contribution à l'analyse (c'est-à-dire l'utilisation de l'algèbre dans l'étude des infiniment petits et grands). Développant de nouvelles idées sur la méthode en logique et confiant dans le symbolisme qu'il avait élaboré à partir de ses recherches mathématiques, il publie, en 1847, un opuscule, Mathematical Analysis of Logic, dans lequel il soutient que la logique doit être rattachée aux mathématiques et non à la philosophie. Bien qu'il n'eût aucun titre universitaire, Boole fut, sur la base de ses publications, nommé en 1849 professeur au Queen's College à Cork, en Irlande. Boole est un des premiers auteurs anglais à écrire sur la logique ; il a mis en évidence l'analogie entre les symboles algébriques et ceux qui représentent les formes logiques et les syllogismes en montrant que les symboles des quantités peuvent être isolés de ceux des opérations. Avec Boole, en 1847 et en 1854, commence l'algèbre de la logique, c'est-à-dire ce qu'on appelle de nos jours l'algèbre de Boole. Dans son ouvrage de 1854, Boole énonce complètement sa nouvelle méthode symbolique d'inférence logique, qui permet, étant donné des propositions contenant un certain nombre de termes, d'en tirer, par traitement symbolique des prémisses, des conclusions qui étaient logiquement contenues dans les prémisses. Il rechercha aussi une méthode générale en calcul des probabilités, qui aurait permis, à partir des probabilités connues d'un système d'événements donnés, de déterminer la probabilité de tout autre événement relié logiquement aux événements donnés.



Borel, Émile (1871-1956), reçu major à l'X et à ULM, il choisit cette dernière et se consacre aux mathématiques. Il fonda l'institut Henri-Poincaré et fut élu député de l'Aveyron et maire de Saint-Afrique. Il étudie les mesures d'ensembles et notamment, définit les ensembles de mesure nulle et l'ensemble des boréliens, sur lequel on peut définir une mesure. Il se tourne ensuite vers les probabilités et la physique mathématique. Borel est également considéré comme un mathéma-

ticien constructiviste. Il fut à l'origine de la théorie des jeux stratégiques et de la cybernétique que développeront von Neumann et Morgenstern. Son élève Henri Lebesgue utilisera ses résultats en topologie et théorie de la mesure pour sa théorie de l'intégration.



Bose, Satyendranath (1894-1974) Mathématicien et physicien indien, connu pour ses contributions à la théorie quantique. Né à Calcutta, Bose a fait ses études au Presidency College de Calcutta. En 1924, il propose une description statistique des systèmes quantiques, reprise par Albert Einstein, et qui n'impose aucune restriction sur la distribution en énergie des particules du système. Cette description est connue sous le nom de "statistique de Bose-Einstein", par opposition à la "statistique de Fermi-Dirac". Appliquée à la théorie du rayonnement du corps noir, cette nouvelle statistique conduit à la formule de distribution de Planck et permet de traiter ce rayonnement comme un gaz de photons. Dans le domaine de la physique des particules élémentaires, la statistique de Bose-Einstein impose à la fonction d'onde des particules (dans l'équation de Schrödinger) d'être parfaitement symétrique pour l'ensemble des variables d'espace et de spin. Les particules obéissant à cette statistique (photons, mésons p, etc.) sont appelés des bosons. Professeur de physique aux universités de Calcutta et de Dacca, Satyendranath Bose a été nommé, en 1958, professeur national des Indes.



Broglie, Louis Victor, prince de (1892-1987), physicien français et lauréat du prix Nobel, qui apporta une contribution essentielle à la théorie quantique avec ses études de la radiation électromagnétique. Né à Dieppe, Louis de Broglie fit ses études à Paris. Il essaya de cerner la nature dualiste de la matière et de l'énergie. Sa découverte de la nature ondulatoire des électrons (1924) lui valut le prix Nobel de physique en 1929. Il fut élu à l'Académie des sciences en 1933 et à l'Académie française en 1943. Il fut nommé professeur de physique théorique à l'université de Paris (1928), secrétaire perpétuel de l'Académie des sciences (1942), et conseiller au Commissariat à l'énergie atomique (1945).

RÉFÉRENCES ET INDEXES

Références 527
Index général 529
Index des noms propres 533

Références

[1] Georges Ifrah. *Histoire universelle des chiffres*. Robert Laffont, Paris, 1994. ISBN 2221057791. 29

Index général

Symbols

écart 300
équation
    caractéristique 377
    de Bessel 428
    différentielle 374
        1er ordre, 374
        linéaire, 375
    diophantienne 247
    homogène 375
    linéaire 281
    ordre d'une 374

A

accroissement fini 322
adhérence 307
analitique
    expression 278
analyse vectorielle 433
angle
    directeur 439
anneau
    Boole 315
asymptote 289
    horizontale 289
    oblique 289
    verticale 289
atlas 309
    maximal 310
automorphisme
    intérieur 269
axe
    abscisse 275
    ordonnée 275

B

base 300
    canonique 439
    dénombrable 313
    euclidienne 459
    orthonormale 446
Bernoulli 396
Bessel 428
    fonction de 422
    séries de 422
borélien 313
borne
    inférieure 339
    intégration 339
    supérieure 339
boule
    fermée 305
    fermée généralisée 307
    généralisée 307
    ouverte 305
    ouverte généralisée 307

C

carte 309
    compatible 310
Cauchy 390
Cauchy-Schwarz 448
champ
    tensoriel 469
champs
    scalaire 466
Chasles 443
classe
    monotone 315
coéfficient
    de Fourrier 405
coefficient
    directeur 319
    dominant 245
combinaison
    convexe 437
    linéaire 437
condition initiale 340
constante
    d'Euler-Mascheroni 373
    Gamma d'Euler 373
contante
    Euler 330
convergence
    absolue 430
coordonnée
    curviligne 344
cosinus
    directeur 440
courbe de niveau 466
couverture
    ouverte 305
cycle 271
    ordre 271

D

décroissance
    exponentielle 291
dépendance
    fonctionelle 282
dérivée
    intérieure 333
déterminant 450
diamètre 308

différomorphisme 309
différentiel 320
différentielle
  df 323
  partielle 324
  totaleexacte 324
dimension
  finie 440
  infinie 440
distance 277, 300
  équivalente 303
  discrète 302
  euclidienne 302
  fonction 300
  hölderienne 302
  induite 301
  semi 300
  ultra-métrique 300
distingué
  groupe 269
divergence
  directionnelle 469
  en coordonnées polaires 472
  en coordonnées cylindriques 474
  modulaire 469
domaine
  d'existence 282
  naturel de définition 282
droite vectorielle 438

E

E.D. 374
ensemble
  mesurable 311
equation
  conditionnelle 239
  identité 239
espace
  affine 442
  de Hilbert 450
  euclidien 450
  Hausdorff 300
  hilbertien 450
  métrique 300
  mesurable 311
  préhilbertien 450
  séparé 300
  topologique 300
  vectoriel 436, 442
espace vectoriel
  complexe 456
  fonctionnel 455
  hermitien 456
Euler
  constante 330
exponentielle
  naturelle 295

F

famille
  liée 438
  libre 438
flêche
  équivalente 433
  colinéaires 433
flèche 433
fonction
  élémentaire 284
  affine 281
  algébrique 285
  bivalente 274
  bornée 306
  composée 284
  constante 283
  continue 288
  contractante 304
  croissante 280, 283
  décroissante 281, 283
  dérivée 320
  de croissance 291
  de Bessel 422
  de Dirac 369
  delta 369
  distance 300
  exponentielle 284
  Gamme d'Euler 370
  impaire 283
  k-lipschitzienne 303
  logarithmique 284
  monotone 281
  périodique 283
  paire 283
  polynomiale 285
  puissance 284
  strictement croissante 281
  strictement décroissante 281
  strictement monotone 281
  transcendante 284
  trigonométrique 284
  univalente 274
  zêta 402
fontion
  d'état 326
formule
  Leibniz 331
Fourier
  transformée de 419
fraction
  rationnelle 285

G

gap 308
Gauss-Ostrogradsky
  théorème 471
Gellman 265
Gibbs 416
Girard (Albert) 392
gradient 466
  coordonnées cylindriques 468
  coordonnées polaires 468
  d'un champs scalaire 466
  du champs scalaire 466
Grahm-Scmidt 447
Gram-Schmidt 297
graphique 275

plan 275
Grassmann 454
groupe
abélien 315
alterné 272
cercle 264
diédral 266
distingué 269
linéaire 262
orthogonal 264
orthogonal réel 264
permutation 268
spécial linéaire 263, 266
spécial orthogonal 264
spécial unitaire 265
unitaire 265, 266

H

Hôpital 322
Hausdorff 308
axiome 300
espace 300
Hilbert 450
homéomorphisme 309

I

inégalité
de Minkowski 449
triangulaire 449
inéquation **239**
intégrale
définie 341
eulérienne 370
indéfinie 339
Riemann 338
intégration
par parties 348
intégrtale
définie 339
isoligne 466
isométrie 303

J

jacobien 347

K

Kronecker 298

L

Lagrange 404
limite 285
à droite 287
à gauche 287
Lodovico Ferrari 251
logarithme 292
népérien 293
naturel 293

M

méthode de réduction à l'unité 435
métrique 300
matrice
jacobienne 347
module 277

N

nabla
du champs scalaire 466
Napier 293
Newton
binôme 243
nombre
dérivé 319
de Bernoulli 396
norme 277, 444

O

opérateur
de divergence 471
gradient 466
orbite 267
Ostrogradsky
théorème 471

P

période 283
Parseval 418
partie
fermée 307
ouverte 307
parties 306
Pascal
triangle 243
Pauli
matrice 265
pente
moyenne 319
permutation
impaire 272
paire 272
Phénomène de Gibbs 416
Phytagore 448
plan vectoriel 438
point 3
(espace vectoriel) 442
adhérent 307
point fixe 431
polynôme
algébrique 244
caractéristique 377
de Bernoulli 396, 398
de MacLaurin 405
de Taylor 405
degré 285
Legendre 254
primitive 339
produit
hermitien 457
mixte 454
vectoriel 451
produit scalaire
canonique 447
progression
arithmétique 387
géométrique 387
harmonique 387

propriété
isométrique 304

Q

quadrant 275

R

règle
d'Alembert 429
de Bernoulli 322
de Cauchy 429
de Grassmann 454
de l'Hospital 322
règle de trois 435
règle des rapports et proportions 435
racine 245
raison 388
rang
d'une famille 441
rapport simple et directe 435
rapport simple et inverse 436
rayon
de convergence 429
relation
de Chasles 443
reste
de Lagrange 404
Riemann 402
Rolle
théorème 320

S

série
alternée 430
de MacLaurin 405
de Taylor 405
divergente 389
MacLaurin 403
numérique 392
Taylor limitée 403
troginométrique 405
séries
de Bessel 422
Schwarz 325
semi-distance 300
somme
de la série 393
directe 442
partielle 392
sous-espace
complémentaire 442
métrique 301
supplémentaire 442
vectoriel 437
spectre de phase 411
sphère 305
généralisée 307
stabilisateur 267
structure
différentiable 310
structure algébrique vectorielle 436
suite 387
bornée 388
constante stationnaire 388
croissante 388
décroissante 388
de Cauchy 390
majorée 387
minorée 387
récurrente 392
système
de coordonnées 309

T

TAF 322
terme 437
test de l'intégrale 428
théorème
de factorisation 246
de Parseval 418
du point fixe 431
Fourier 410
Fourier-Dirichlet 410
Phytagore (de) 448
Schwarz 325
topologie 299, 300
induite 305
métrique 305
transformée de Fourier 419
transformée de Fourier inverse 419
transformation
affine 262
transposition 271
tribu 311
borélienne 313
engendrée 312
Lebesgue 313
trace 313

V

variété 309
différentiable 309
topologique 309
variable
d'entrée 282
dépendante 282
indépendante 282
muette 339
transformée 274
vecteur 434, 436
colonne 437
nul 434
orthogonal 446
pseudo-vecteur 434
Viète 235
voisinage 304

Index des noms propres

Symbols

écart 300
équation
    caractéristique 377
    de Bessel 428
    différentielle 374
        1er ordre, 374
        linéaire, 375
    diophantienne 247
    homogène 375
    linéaire 281
    ordre d'une 374
Cantor 82
    théorème de Cantor-Bernstein 82

A

accroissement fini 322
adhérence 307
analitique
    expression 278
analyse vectorielle 433
angle
    directeur 439
anneau
    Boole 315
asymptote 289
    horizontale 289
    oblique 289
    verticale 289
atlas 309
    maximal 310
automorphisme
    intérieur 269
axe
    abscisse 275
    ordonnée 275

B

base 300
    canonique 439
    dénombrable 313
    euclidienne 459
    orthonormale 446
Bernoulli 396
Bessel 428
    fonction de 422
    séries de 422
borélien 313
borne
    inférieure 339
    intégration 339
    supérieure 339
boule
    fermée 305
    fermée généralisée 307
    généralisée 307
    ouverte 305
    ouverte généralisée 307

C

carte 309
    compatible 310
Cauchy 390
Cauchy-Schwarz 448
champ
    tensoriel 469
champs
    scalaire 466
Chasles 443
classe
    monotone 315
coéfficient
    de Fourrier 405
coefficient
    directeur 319
    dominant 245
combinaison
    convexe 437
    linéaire 437
condition initiale 340
constante
    d'Euler-Mascheroni 373
    Gamma d'Euler 373
contante
    Euler 330
convergence
    absolue 430
coordonnée
    curviligne 344
cosinus
    directeur 440
courbe de niveau 466
couverture
    ouverte 305
cycle 271
    ordre 271

D

décroissance
    exponentielle 291
dépendance
    fonctionelle 282
dérivée
    intérieure 333

déterminant 450
diamètre 308
difféomorphisme 309
différentiel 320
différentielle
  df 323
  partielle 324
  totaleexacte 324
dimension
  finie 440
  infinie 440
distance 277, 300
  équivalente 303
  discrète 302
  euclidienne 302
  fonction 300
  hölderienne 302
  induite 301
  semi 300
  ultra-métrique 300
distingué
  groupe 269
divergence
  directionnelle 469
  en coordonnées polaires 472
  en coordonnées cylindriques 474
  modulaire 469
domaine
  d'existence 282
  naturel de définition 282
droite vectorielle 438

E

E.D. 374
ensemble
  mesurable 311
equation
  conditionnelle 239
  identité 239
espace
  affine 442
  de Hilbert 450
  euclidien 450
  Hausdorff 300
  hilbertien 450
  métrique 300
  mesurable 311
  préhilbertien 450
  séparé 300
  topologique 300
  vectoriel 436, 442
espace vectoriel
  complexe 456
  fonctionnel 455
  hermitien 456
Euler
  constante 330
exponentielle
  naturelle 295

F

famille
  liée 438
  libre 438
flèche
  équivalente 433
  colinéaires 433
flèche 433
fonction
  élémentaire 284
  affine 281
  algébrique 285
  bivalente 274
  bornée 306
  composée 284
  constante 283
  continue 288
  contractante 304
  croissante 280, 283
  décroissante 281, 283
  dérivée 320
  de croissance 291
  de Bessel 422
  de Dirac 369
  delta 369
  distance 300
  exponentielle 284
  Gamme d'Euler 370
  impaire 283
  k-lipschitzienne 303
  logarithmique 284
  monotone 281
  périodique 283
  paire 283
  polynomiale 285
  puissance 284
  strictement croissante 281
  strictement décroissante 281
  strictement monotone 281
  transcendante 284
  trigonométrique 284
  univalente 274
  zêta 402
fontion
  d'état 326
formule
  Leibniz 331
Fourier
  transformée de 419
fraction
  rationnelle 285

G

GÖDEL 3, 6, 82, 83
gap 308
Gauss-Ostrogradsky
  théorème 471
Gellman 265
Gibbs 416
Girard (Albert) 392
gradient 466
  coordonnées cylindriques 468
  coordonnées polaires 468
  d'un champs scalaire 466
  du champs scalaire 466

Grahm-Scmidt 447
Gram-Schmidt 297
graphique 275
    plan 275
Grassmann 454
groupe
    abélien 315
    alterné 272
    cercle 264
    diédral 266
    distingué 269
    linéaire 262
    orthogonal 264
    orthogonal réel 264
    permutation 268
    spécial linéaire 263, 266
    spécial orthogonal 264
    spécial unitaire 265
    unitaire 265, 266

H

Hôpital 322
Hausdorff 308
    axiome 300
    espace 300
Hilbert 450
homéomorphisme 309

I

inégalité
    de Minkowski 449
    triangulaire 449
inéquation **239**
intégrale
    définie 341
    eulérienne 370
    indéfinie 339
    Riemann 338
intégration
    par parties 348
intégrtale
    définie 339
isoligne 466
isométrie 303

J

jacobien 347

K

Kronecker 298

L

Lagrange 404
limite 285
    à droite 287
    à gauche 287
Lodovico Ferrari 251
logarithme 292
    népérien 293
    naturel 293

M

méthode de réduction à l'unité 435
métrique 300
matrice
    jacobienne 347
module 277

N

nabla
    du champs scalaire 466
Napier 293
Newton
    binôme 243
nombre
    dérivé 319
    de Bernoulli 396
norme 277, 444

O

opérateur
    de divergence 471
    gradient 466
orbite 267
Ostrogradsky
    théorème 471

P

période 283
Parseval 418
partie
    fermée 307
    ouverte 307
parties 306
Pascal
    triangle 243
Pauli
    matrice 265
pente
    moyenne 319
permutation
    impaire 272
    paire 272
Phénomène de Gibbs 416
Phytagore 448
plan vectoriel 438
point
    (espace vectoriel) 442
    adhérent 307
point fixe 431
polynôme
    algébrique 244
    caractéristique 377
    de Bernoulli 396, 398
    de MacLaurin 405
    de Taylor 405
    degré 285
    Legendre 254
primitive 339
produit
    hermitien 457
    mixte 454
    vectoriel 451
produit scalaire
    canonique 447
progression

arithmétique 387
géométrique 387
harmonique 387
propriété
isométrique 304

Q

quadrant 275

R

règle
d'Alembert 429
de Bernoulli 322
de Cauchy 429
de Grassmann 454
de l'Hospital 322
règle de trois 435
règle des rapports et proportions 435
racine 245
raison 388
rang
d'une famille 441
rapport simple et directe 435
rapport simple et inverse 436
rayon
de convergence 429
relation
de Chasles 443
reste
de Lagrange 404
Riemann 402
Rolle
théorème 320

S

série
alternée 430
de MacLaurin 405
de Taylor 405
divergente 389
MacLaurin 403
numérique 392
Taylor limitée 403
troginométrique 405
séries
de Bessel 422
Schwarz 325
semi-distance 300
somme
de la série 393
directe 442
partielle 392
sous-espace
complémentaire 442
métrique 301
supplémentaire 442
vectoriel 437
spectre de phase 411
sphère 305
généralisée 307
stabilisateur 267
structure
différentiable 310
structure algébrique vectorielle 436
suite 387
bornée 388
constante stationnaire 388
croissante 388
décroissante 388
de Cauchy 390
majorée 387
minorée 387
récurrente 392
système
de coordonnées 309

T

TAF 322
terme 437
test de l'intégrale 428
théorème
de factorisation 246
de Parseval 418
du point fixe 431
Fourier 410
Fourier-Dirichlet 410
Phytagore (de) 448
Schwarz 325
topologie 299, 300
induite 305
métrique 305
transformée de Fourier 419
transformée de Fourier inverse 419
transformation
affine 262
transposition 271
tribu 311
borélienne 313
engendrée 312
Lebesgue 313
trace 313

V

variété 309
différentiable 309
topologique 309
variable
d'entrée 282
dépendante 282
indépendante 282
muette 339
transformée 274
vecteur 434, 436
colonne 437
nul 434
orthogonal 446
pseudo-vecteur 434
Viète 235
voisinage 304