Table des matières

Chapitre 1

Introduction

Tant d'ouvrages traitent abondemment et rigoureusement des domaines les plus divers que couvrent les mathématiques ! L'objectif de ce livre n'est pas d'en ajouter un de plus au bas d'une longue liste, mais d'adopter une approche différente.

Je l'ai voulu comme une introduction rapide et intuitive aux notions les plus diverses que l'on rencontre dans ce monde merveilleux et fascinant. A ce titre, les démonstrations que vous rencontrerez au fil de ces pages pourraient faire frémir plus d'un puriste : il s'agit ici, non pas de « prouver » au sens strict du terme, mais de montrer comment les différents concepts s'agencent entre-eux. Il s'agit avant tout de faire acquérir au lecteur une intuition mathématique que j'estime au moins aussi indispensable à l'apprentissage que la rigueur logique. Car enfin, de quoi sert la rigueur à quelqu'un qui ne « voit » pas où on veut en venir ?

Cet ouvrage est également conçu pour qu'un débutant puisse le lire aussi agréablement qu'un mathématicien confirmé. Il existe donc deux méthodes possibles pour parcourir cet ouvrage :

— Les premiers chapitres définissent les notions fondamentales. L'édifice s'éleve ensuite patiemment vers des notions plus avancées. Le lecteur pourra donc entamer ce livre où il le souhaite, en fonction de ses connaissances. Qu'il n'hésite tout de même pas à jeter un oeil sur les notions qui lui semblent élémentaires : il sera parfois surpris par la complexité ou l'élégance de leur construction.

— Afin de guider le lecteur qui voudrait accéder directement au domaine de son choix, une liste des « dépendances » est placée au début de chaque chapitre. Il s'agit en fait d'une liste d'autres chapitres, prérequis conseillés pour une compréhension optimale du chapitre choisi. Ces prérequis ne sont bien entendu pas à prendre au pied de la lettre. Il est même plutôt recommandé de commencer à lire, et de n'utiliser les dépendances que si la compréhension pose problème. N'oubliez pas non plus que ces dépendances sont récursives, à vous de reconstruire l'arbre jusqu'à la racine si besoin est.

1.1 Portée des notations

Les notations sont valables jusqu'à leur redéfinition au sein d'un autre concept. Le plus souvent, elles s'étendent sur une section, parfois pendant un chapitre entier. Certaines notations, les plus importantes, sont valables tout au long de l'ouvrage.

1.2 Abstraction

La première étape de l'abstraction consiste à repérer les propriétés fondamentales communes à plusieurs types différents d'objets mathématiques. Ensuite, on crée une nouvelle structure en imposant qu'elle vérifie ces propriétés. Tous les résultats qui s'en déduiront auront alors un champ d'application plus large que dans les applications de départ.

1.3 Récurrence

Supposons que nous ayons à prouver que A_n est vraie pour tout entier $n \in \mathbb{N}$ tel que $n \geq p$. Si nous prouvons que A_p est vrai et que la véracité de A_n implique celle de A_{n+1}, la démonstration est finie. En effet, on a alors :

$$A_p \Rightarrow A_{p+1} \Rightarrow A_{p+2} \Rightarrow \ldots \Rightarrow A_n$$

pour tout $n \geq p$ arbitraire.

Chapitre 2

Notations

$=$	égalité entre deux objets
$\neq$	différence
$\in$	appartenance à un ensemble
$\notin$	non appartenance à un ensemble
$\equiv$	équivalence
$\Rightarrow$	la condition de gauche implique celle de droite
$\Leftarrow$	la condition de droite implique celle de gauche
$\Leftrightarrow$	les deux conditions s'impliquent mutuellement
$:$	tel que
$\exists$	il existe
$\forall$	pour tout
$\subseteq$	inclusion
$\subset$	inclusion stricte
$\cup$	union d'ensembles
$\cap$	intersection d'ensembles
$\mathfrak{P}$	ensemble des sous-ensembles
Partition	ensemble des partitions
Tribu	ensemble des tribus
Rel	relation
$\le$	ordre large
$<$	ordre strict
$\preccurlyeq$	comparaison entre ensembles
major	ensemble des majorants
minor	ensemble des minorants
max	maximum
min	minimum
sup	supremum
inf	infimum
arg	argument
$\mathfrak{dist}$	distance
adh	adhérence
int	intérieur
∂	frontière
$\mathfrak{B}$	boule
$\mathbb{F}$	fonction
$\mathbb{B}$	ensemble de Boole
$\mathbb{N}$	ensemble des naturels
$\mathbb{Z}$	ensemble des entiers

$\mathbb{Q}$	ensemble des rationnels
$\mathbb{R}$	ensemble des réels
$\mathbb{C}$	ensemble des complexes
$\mathbb{K}$	corps quelconque

Première partie

Ensembles

Chapitre 3

Ensembles

Dépendances

Vous êtes à la racine.

3.1 Définition explicite

Les ensembles sont des regroupements d'objets appelés éléments.

Il existe deux méthodes permettant de définir un ensemble. Lorsqu'il n'existe qu'un nombre fini d'éléments distincts, on peut les énumérer :

$$A = \{a, b, c, ..., z\}$$

On dit alors que x appartient à A, et on le note :

$$x \in A$$

si x fait partie de la liste $a, b, c, ..., z$. Dans le cas contraire, x n'appartient pas à A, ce que l'on note par :

$$x \notin A$$

3.2 Définition implicite

On peut aussi définir un ensemble en demandant que ses éléments respectent certaines conditions. On dit alors que $x \in A$ si x vérifie toutes les conditions nécessaires pour appartenir à l'ensemble A ou que $x \notin A$ si au moins une des conditions n'est pas remplie. Le schéma de ce type de définition s'écrit :

$$A = \{x : \text{ une ou plusieurs conditions sur } x\}$$

Dans ce cas, le nombre d'éléments de l'ensemble peut être fini ou infini.

3.2.1 Variante

On ajoute souvent une condition sur les éléments de l'ensemble :

$$A = \{x \in \Omega : \text{ conditions sur } x\}$$

Dans ce cas, tout candidat x doit en plus appartenir à l'ensemble Ω s'il veut appartenir à l'ensemble A. Cette définition est donc équivalente à :

$$A = \{x : x \in \Omega, \text{ conditions sur } x\}$$

3.3 Notations

Le symbole $\exists$ signifie « il existe » et le symbole $\forall$ signifie « pour tout »

3.4 Ensemble vide

L'ensemble vide $\emptyset = \{\}$ est un cas particulier ne contenant aucun élément :

$$x \notin \emptyset$$

quelle que soit la nature de x.

3.5 Naturels

L'ensemble des nombres naturels se construit à partir d'un élément « racine » 0, auquel on ajoute indéfiniment des successeurs. Le successeur de 0 est noté 0^+ ou 1. On dit aussi que 0 est le prédécesseur de 1 et on le note $1^- = 0$. Arrivé à l'élément i, on ajoute le successeur de i, noté :

$$j = i^+$$

On dit aussi que i est le prédécesseur de j et on le note :

$$i = j^-$$

On a donc :

$$i^{+-} = j^- = i$$

et :

$$j^{-+} = i^+ = j$$

L'ensemble des objets ainsi crée est appelé l'ensemble des nombres naturels et noté $\mathbb{N}$. En l'exprimant au moyen des symboles usuels, on a dans l'ordre de succession :

$$\mathbb{N} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ...\}$$

3.5.1 Notation

On note aussi :

$$i + 1 = i^+$$

et :

$$i - 1 = i^-$$

3.5.2 Element racine

L'élément 0 est le seul naturel à de pas posséder de prédécesseur. On l'appelle pour cette raison l'élément racine de $\mathbb{N}$.

On dit aussi qu'un élément n est nul pour signifier que $n = 0$.

3.5.3 Ensemble discret ou dénombrable

Tout ensemble de la forme :

$$A = \{a_k : k \in \mathbb{N}\}$$

est dit discret ou dénombrable.

3.6 Inclusion

On dit que A est inclus dans B et on note :

$$A \subseteq B$$

si tous les éléments de A appartiennent aussi à B :

$$x \in A \ \Rightarrow \ x \in B$$

On dit alors que A est un sous-ensemble, ou une partie de B.

3.6.1 Stricte

Il y a inclusion stricte :

$$A \subset B$$

lorsque $A \subseteq B$ et que les deux ensembles ne sont pas égaux, ce que l'on note par :

$$A \neq B$$

3.7 Egalité

Deux ensembles A et B sont dit égaux et on le note :

$$A = B$$

si tout élément de A appartient aussi à B et si tout élément de B appartient aussi à A :

$$x \in A \ \Leftrightarrow \ x \in B$$

ce qui revient à dire que l'on a inclusion mutuelle de A et B :

$$A = B \ \Leftrightarrow \ A \subseteq B \text{ et } B \subseteq A$$

3.7.1 Remarque

Dans le cadre des ensembles, on ne se soucie pas de l'ordre :

$$\{a, b\} = \{b, a\}$$

ni du nombre d'apparitions d'un élément :

$$\{a, a, b\} = \{a, b\}$$

3.8 Union

L'union de deux ensembles $A \cup B$ est l'ensemble contenant les éléments de A et les éléments de B. Un élément quelconque de $A \cup B$ peut donc appartenir à A, à B ou aux deux ensembles simultanément :

$$A \cup B = \{x : x \in A \text{ ou/et } x \in B\}$$

3.8.1 Jargon

Le « ou » mathématique est non exclusif. La proposition a ou b signifie que soit a, soit b, soit (a et b) est vérifié.

3.9 Intersection

L'intersection $A \cap B$ est l'ensemble des éléments qui appartiennent à la fois à A et à B :

$$A \cap B = \{x : x \in A \text{ et } x \in B\}$$

3.9.1 Nomenclature

Lorsque l'intersection de deux ensembles est vide, on dit qu'ils sont disjoints.

3.10 Association

Soit les ensembles A, B, C. On définit :

$$A \cup B \cup C = A \cup (B \cup C)$$

Comme les éléments de $A \cup B \cup C$ sont les éléments qui appartiennent à au moins un des ensembles A, B, C, on voit que :

$$A \cup B \cup C = (A \cup B) \cup C$$

On en conclut que :

$$A \cup B \cup C = A \cup (B \cup C) = (A \cup B) \cup C$$

On définit aussi :

$$A \cap B \cap C = A \cap (B \cap C)$$

Comme les éléments de $A \cap B \cap C$ sont les éléments qui appartiennent simultanément à A, B, C, on voit que :

$$A \cap B \cap C = (A \cap B) \cap C$$

On en conclut que :

$$A \cap B \cap C = A \cap (B \cap C) = (A \cap B) \cap C$$

3.11 Commutation

On a clairement :

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

3.12 Distribution

Soit les ensembles A, B, C. Lorsque x appartient à la fois à A et à au moins un des deux ensembles B et C, on sait que x appartient à A et B ou qu'il appartient à A et C, et inversément. On a donc :

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

On dit que l'intersection se distribue sur l'union. On a également la relation :

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

On dit que l'union se distribue sur l'intersection.

3.13 Différence

La différence $A \setminus B$ est l'ensemble des éléments qui appartiennent à A mais pas à B :

$$A \setminus B = \{x : x \in A \text{ et } x \notin B\}$$

3.14 Décomposition

Soit les ensembles A, B. Les éléments de A sont de deux types :
— ceux qui appartiennent également à B
— ceux qui n'appartiennent pas à B
On en conclut que :

$$A = (A \cap B) \cup (A \setminus B)$$

On voit que les deux sous-ensembles de A sont disjoints :

$$(A \cap B) \cap (A \setminus B) = \emptyset$$

3.14.1 Union

Les éléments de $A \cup B$ sont de deux types :
— ceux qui appartiennent seulement à A
— ceux qui appartiennent à B
On en conclut que :

$$A \cup B = (A \setminus B) \cup B$$

On voit que les deux sous-ensembles de $A \cup B$ sont disjoints :

$$(A \setminus B) \cap B = \emptyset$$

3.15 Complémentaire

Si $A \subseteq \Omega$, on dit que $C = \Omega \setminus A$ est le complémentaire de A dans Ω, ou simplement que C est le complémentaire de A lorsque l'ensemble Ω est évident d'après le contexte.

3.15.1 Complémentaire du complémentaire

Soit $A \subseteq \Omega$. Un élément de Ω qui n'appartient pas à $\Omega \setminus A$ appartient à A, et réciproquement. On a donc :

$$\Omega \setminus (\Omega \setminus A) = A$$

3.15.2 Réciprocité

Soit $A \subseteq \Omega$ et son complémentaire :

$$B = \Omega \setminus A$$

En prenant le complémentaire de cette équation, on obtient :

$$\Omega \setminus B = \Omega \setminus (\Omega \setminus A) = A$$

Soit à présent $B \subseteq \Omega$ et son complémentaire :

$$A = \Omega \setminus B$$

En prenant le complémentaire de cette équation, on obtient :

$$\Omega \setminus A = \Omega \setminus (\Omega \setminus B) = B$$

On en conclut l'équivalence :

$$A = \Omega \setminus B \;\Leftrightarrow\; B = \Omega \setminus A$$

3.15.3 Complémentaire d'une union

Soit un ensemble Ω et les sous-ensembles $A, B \subseteq \Omega$. Un élément de Ω qui n'appartient pas à $A \cup B$ n'appartient ni à A ni à B. Il appartient donc à $\Omega \setminus A$ et à $\Omega \setminus B$. Inversément, un élément qui appartient à $\Omega \setminus A$ et à $\Omega \setminus B$ n'appartient ni à A ni à B. On a donc :

$$\Omega \setminus (A \cup B) = (\Omega \setminus A) \cap (\Omega \setminus B)$$

Le complémentaire d'une union est l'intersection des complémentaires.

3.15.4 Complémentaire d'une intersection

Soit $A, B \in \Omega$. Posons :

$$C = \Omega \setminus A \subseteq \Omega$$
$$D = \Omega \setminus B \subseteq \Omega$$

L'expression du complémentaire de $A \cup B$ devient :

$$\Omega \setminus \big[(\Omega \setminus C) \cup (\Omega \setminus D)\big] = C \cap D$$

En prenant le complémentaire des deux membres par rapport à Ω, on obtient :

$$(\Omega \setminus C) \cup (\Omega \setminus D) = \Omega \setminus (C \cap D)$$

Le complémentaire d'une intersection est l'union des complémentaires.

Chapitre 4

Collections

Dépendances

— Chapitre 3 : Les ensembles

4.1 Définition

Une collection est un ensemble d'ensembles, c'est-à-dire un ensemble dont les éléments sont également des ensembles.

4.2 L'ensemble des sous-ensembles

Une collection particulièrement importante est l'ensemble des sous-ensembles d'un ensemble donné A. On le note :

$$\mathfrak{P}(A) = \{S : S \subseteq A\}$$

Remarque : la notation $\mathfrak{P}$ est une notation globale, vous la retrouver dans d'autres chapitres.

4.3 Union

Soit un ensemble Ω et une collection d'ensembles $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$. On définit l'union des ensembles-éléments de $\mathcal{C}$ par :

$$\bigcup \mathcal{C} = \{a \in \Omega : \text{ il existe } A \in \mathcal{C} \text{ tel que } a \in A\}$$

Il s'agit donc de l'ensemble dont chaque élément appartient à au moins un des ensembles-éléments de $\mathcal{C}$.

4.4 Intersection

Soit un ensemble Ω et une collection d'ensembles $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$. On définit l'intersection des ensembles-éléments de $\mathcal{C}$ par :

$$\bigcap \mathcal{C} = \{a \in \Omega : a \in A \text{ pour tout } A \in \mathcal{C}\}$$

Il s'agit donc de l'ensemble dont chaque élément appartient à tous les ensembles-éléments de $\mathcal{C}$.

4.5 Collection paramétrée

Soit un ensemble X et la collection :

$$\mathcal{C} = \{A(x) \in \mathfrak{P}(\Omega) : x \in X\}$$

On appelle ce type de collection une collection paramétrée. L'ensemble X est appelé ensemble de paramètres.

4.5.1 Union

L'union de tous ces ensembles est l'ensemble dont les éléments appartiennent à au moins un des $A(x)$, pour un certain $x \in X$:

$$\bigcup_{x \in X} A(x) = \{a \in \Omega : \text{ il existe } x \in X \text{ tel que } a \in A\}$$

L'intersection est l'ensemble des éléments appartenant à tous les $A(x)$, pour tout les $x \in X$:

$$\bigcap_{x \in X} A(x) = \{a \in \Omega : a \in A(x) \text{ pour tout } x \in X\}$$

4.6 Collections discrètes

Soit $A_1, A_2, A_3, ...$ une collection finie ou infinie d'ensembles. L'union de tous ces ensembles se note :

$$\bigcup_i A_i = A_1 \cup A_2 \cup A_3 \cup ...$$

L'ensemble des éléments qui sont communs à tous ces ensembles se note :

$$\bigcap_i A_i = A_1 \cap A_2 \cap A_3 \cap ...$$

Finie

Dans le cas où la collection est finie, par exemple $A_1, A_2, ..., A_n$, on note simplement :

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup ... \cup A_n$$

ainsi que :

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap ... \cap A_n$$

Infinie

Dans le cas où la collection est infinie, on note simplement :

$$\bigcup_{i=1}^{+\infty} A_i = A_1 \cup A_2 \cup ...$$

ainsi que :

$$\bigcap_{i=1}^{+\infty} A_i = A_1 \cap A_2 \cap ...$$

4.7 Distributivité

On a :

$$A \cap \bigcup_{x \in X} B(x) = \bigcup_{x \in X} \left[A \cap B(x)\right]$$

et :

$$A \cup \bigcap_{x \in X} B(x) = \bigcap_{x \in X} \left[A \cup B(x)\right]$$

Les collections discrètes en sont un cas particulier :

$$A \cap \bigcup_{i} B_i = \bigcup_{i} \left[A \cap B_i\right]$$

$$A \cup \bigcap_{i} B_i = \bigcap_{i} \left[A \cup B_i\right]$$

4.8 Complémentaire

Soit une collection $\{A(x) : x \in X\} \subseteq \mathfrak{P}(\Omega)$ de sous-ensembles de Ω. On a :

$$\Omega \setminus \bigcup_{x \in X} A(x) = \bigcap_{x \in X} \left[\Omega \setminus A(x)\right]$$

et :

$$\Omega \setminus \bigcap_{x \in X} A(x) = \bigcup_{x \in X} \left[\Omega \setminus A(x)\right]$$

Les collections discrètes en sont un cas particulier :

$$\Omega \setminus \bigcup_{i} A_i = \bigcap_{i} \left[\Omega \setminus A_i\right]$$

$$\Omega \setminus \bigcap_{i} A_i = \bigcup_{i} \left[\Omega \setminus A_i\right]$$

Chapitre 5

Partitions

Dépendances

— Chapitre 4 : Les collections

5.1 Définition

Une partition, ou découpage, d'un ensemble A est une collection de sous-ensembles de A :

$$\mathcal{P} = \{P(x) \in \mathfrak{P}(A) : x \in X\}$$

vérifiant certaines propriétés : les ensembles de $\mathcal{P}$ doivent permettre de reconstituer A par leur union :

$$A = \bigcup_{x \in X} P(x)$$

et ne doivent pas se chevaucher. Leur intersection est donc vide dès que $x, y \in X$ vérifient $x \neq y$:

$$P(x) \cap P(y) = \emptyset$$

On note $\text{Partition}(A)$ l'ensemble des collections formant une partition de A.

5.2 Discrètes

Pour qu'une collection $\{A_1, A_2, ...\}$ de sous-ensembles de A forme une partition de A, il faut que :

$$A = \bigcup_i A_i$$

et que :

$$A_i \cap A_j = \emptyset$$

pour tout $i \neq j$.

Chapitre 6

Produit cartésien

Dépendances

— Chapitre 3 : Les ensembles

6.1 Définition

Le produit cartésien $\times$ de deux ensembles A et B permet de construire des ensembles contenant des couples d'éléments (x, y) :

$$A \times B = \{(x, y) : x \in A \quad \text{et} \quad y \in B\}$$

Les éléments x et y sont appelée composantes du couple (x, y).

6.2 Tuple

On généralise aux tuples $(x_1, x_2, ..., x_n)$ formés d'éléments des ensembles $A_1, A_2, ..., A_n$ par :

$$A_1 \times A_2 \times ... \times A_n = \{(x_1, x_2, ..., x_n) : x_1 \in A_1, x_2 \in A_2, ..., x_n \in A_n\}$$

Les éléments $x_1, x_2, ..., x_n$ sont appelés composantes du tuple $(x_1, x_2, ..., x_n)$

6.3 Égalité

On dit que les deux tuples :

$$x = (x_1, x_2, ..., x_n) \in A_1 \times A_2 \times ... \times A_n$$

et :

$$y = (y_1, y_2, ..., y_n) \in A_1 \times A_2 \times ... \times A_n$$

sont égaux et on le note :

$$(x_1, x_2, ..., x_n) = (y_1, y_2, ..., y_n)$$

ou plus simplement :

$$x = y$$

si et seulement si toutes leurs composantes sont identiques :

$$x_i = y_i$$

pour tout $i \in \{1, 2, ..., n\}$.

6.3.1 Remarque

Dans le cadre des tuples, l'ordre compte :

$$(a, b) \neq (b, a)$$

ainsi que le nombre d'apparitions d'un élément :

$$(a, a, b) \neq (a, b)$$

6.4 Puissance

On note A^n l'ensemble des tuples composés de n éléments de A :

$$A^n = \{(x_1, x_2, ..., x_n) : x_1, x_2, ..., x_n \in A\}$$

Un exemple courant :

$$A^2 = A \times A$$

Chapitre 7

Relations

Dépendances

— Chapitre 3 : Les ensembles
— Chapitre 6 : Produit cartésien

7.1 Définition

Une relation R est un ensemble de couples $(x, y) \in A \times B$ reliant des éléments de A à des élements de B. On note $\mathrm{Rel}(A, B)$ l'ensemble des relations sur A, B :

$$\mathrm{Rel}(A, B) = \{R : R \subseteq A \times B\}$$

On dit que $x \in A$ est en relation R avec $y \in B$ si $(x, y) \in R$. On le note aussi :

$$x \ R \ y$$

7.2 Relation inverse

A toute relation R, on associe une relation inverse R^{-1} en intervertissant x et y :

$$R^{-1} = \{(y, x) \in B \times A : (x, y) \in R\}$$

7.3 Relation identité

La relation identité $\mathrm{Id} \subseteq X \times X$ est définie par :

$$\mathrm{Id} = \{(x, x) : x \in X\}$$

Elle vérifie la propriété :

$$\mathrm{Id}^{-1} = \mathrm{Id}$$

7.3.1 Egalité

Si $x = y$, on a $(x, y) \in \mathrm{Id}$ et inversément. L'égalité correspond donc à la relation identité.

7.4 Image

L'image de $x \in A$ par R est l'ensemble des éléments de B en relation avec x :

$$R(x) = \{y \in B : (x, y) \in R\}$$

On généralise la notion d'image aux sous-ensembles $X \subseteq A$:

$$R(X) = \{y \in B : x \in X,\ (x, y) \in R\}$$

7.5 Image inverse

L'image inverse de $y \in B$ est l'ensemble des éléments de A en relation avec y :

$$R^{-1}(y) = \{x \in A : (x, y) \in R\}$$

On généralise la notion d'image inverse aux sous-ensembles $Y \subseteq B$:

$$R^{-1}(Y) = \{x \in A : y \in Y,\ (x, y) \in R\}$$

7.6 Image

Lorsqu'on ne précise pas l'ensemble, l'image de R est simplement l'image de A :

$$\operatorname{im} R = R(A)$$

7.7 Domaine

Le domaine de R est un cas particulier d'image inverse :

$$\operatorname{dom} R = R^{-1}(B) = \{x \in A : y \in B,\ (x, y) \in R\}$$

7.8 Composée

A partir d'une relation R reliant A à B et d'une relation S reliant B à C, on peut construire une relation composée $S \circ R$ (lire S après R) reliant A a C. Pour que le couple (x, z) appartienne à $S \circ R$, on doit pouvoir trouver un élément intermédiaire $y \in B$ tel que $(x, y) \in R$ et $(y, z) \in S$. Ce y doit donc appartenir simultanément à $R(x)$ et à $S^{-1}(z)$. On a par conséquent :

$$S \circ R = \{(x, z) \in A \times C : R(x) \cap S^{-1}(z) \neq \emptyset\}$$

ou encore :

$$S \circ R = \{(x, z) \in A \times C : \exists y \in B : (x, y) \in R,\ (y, z) \in S\}$$

7.9 Puissance

Soit une relation $R \in \mathrm{Rel}(A, A)$. On définit la puissance par :

$$\begin{aligned} R^0 &= \mathrm{Id} \\ R^n &= R \circ R^{n-1} \end{aligned}$$

On a donc en particulier $R^1 = R$ et :

$$R^n = R \circ ... \circ R$$

Chapitre 8

Fonctions

Dépendances

— Chapitre 3 : Les ensembles
— Chapitre 7 : Les relations

8.1 Définitions

Une fonction f de A vers B associe à chaque $x \in A$ un unique élément $f(x) \in B$. On note $\mathbb{F}(A, B)$ l'ensemble des fonctions f de A vers B. On utilise aussi la notation :

$$f : A \mapsto B$$

pour préciser que $f \in \mathbb{F}(A, B)$ et :

$$f : x \mapsto f(x)$$

pour préciser que f associe $x \in A$ à $f(x) \in B$. On dit que $f(x)$ est la valeur de f en x.

Remarque

La notation $f : A \mapsto B$ signifie que :

— $f(x)$ est défini pour tout $x \in A$
— $f(x) \in B$

Par conséquent, si $C \subseteq A$ et si $B \subseteq D$, la condition $f : A \mapsto B$ implique $f : C \mapsto D$.

Synonymes

On parle indifféremment de fonction ou d'application.

8.2 Fonctions discrètes

Dans le cas particulier où $f \in \mathbb{F}(\{1, 2, ..., N\}, B)$, on peut associer à f un nombre $(f_1, f_2, ..., f_N) \in B^N$ par :

$$f_i = f(i)$$

pour tout $i \in \{1, 2, ..., N\}$. Inversément, à tout $(f_1, f_2, ..., f_N) \in B^N$, on associe une fonction $f : \{1, 2, ..., N\} \mapsto B$ par :

$$f(i) = f_i$$

On voit donc l'équivalence :

$$\mathbb{F}(\{1,2,...,N\},B) \equiv B^N$$

Notation

Dans le cas de fonctions quelconques, on pose par analogie :

$$B^A = \mathbb{F}(A,B)$$

8.3 Relation associée

On peut associer à toute fonction $f : A \mapsto B$ une relation $R \in \mathrm{Rel}(A,B)$ définie par :

$$R = \{(x, f(x)) : x \in A\}$$

On a clairement :

$$R(x) = \{f(x)\}$$

8.4 Relation inverse

Soit $f : A \mapsto B$ associée à la relation $R \in \mathrm{Rel}(A,B)$. La relation inverse $R^{-1} \in \mathrm{Rel}(B,A)$ est définie par :

$$R^{-1} = \{(f(x), x) \in B \times A : x \in A\}$$

8.5 Fonction identité

La fonction identité $\mathrm{Id} : A \mapsto A$ est définie par :

$$\mathrm{Id} : x \mapsto \mathrm{Id}(x) = x$$

8.5.1 Relation

La relation associée à la fonction identité s'écrit :

$$\{(x, \mathrm{Id}(x)) \in A^2 : x \in A\} = \{(x,x) \in A^2 : x \in A\}$$

La fonction identité est donc associée à la relation identité.

8.6 Image d'un ensemble

Soit $f : A \mapsto B$. L'image d'un sous-ensemble $X \subseteq A$ par f est l'ensemble des valeurs que prend f en tous les éléments de X :

$$f(X) = \{f(x) : x \in X\}$$

8.7 Image d'une fonction

Soit $f : A \mapsto B$. L'image de f est l'ensemble des valeurs que prend f en tous les éléments de A :

$$\operatorname{im} f = f(A) = \{f(x) : x \in A\}$$

8.8 Image inverse

Soit $f : A \mapsto B$. Pour tout $y \in B$, l'image inverse est l'ensemble défini par :

$$f^{-1}(y) = \{x \in A : f(x) = y\}$$

L'image inverse d'un ensemble $Y \subseteq B$ est définie par :

$$f^{-1}(Y) = \{x \in A : f(x) \in Y\}$$

8.9 Domaine

Soit un ensemble $A \subseteq \Omega$ et une fonction $f : A \mapsto B$. Le domaine de f est l'ensemble des éléments de $x \in \Omega$ tels que $f(x) \in B$ existe. Autrement dit :

$$\operatorname{dom} f = A$$

8.10 Composée

Soit les fonctions $f : A \mapsto B$ et $g : B \mapsto C$.

Supposons que les grandeurs $x \in A$, $y \in B$ et $z \in C$ soient reliées par les égalités $y = f(x)$ et $z = g(y)$. On a alors $z = g(f(x))$. On définit une nouvelle fonction $g \circ f : A \mapsto C$ associée à ce résultat par :

$$g \circ f : x \mapsto (g \circ f)(x) = g(f(x))$$

On nomme $g \circ f$ la composée de f et g. On note aussi :

$$g \circ f(x) = (g \circ f)(x)$$

8.10.1 Association

Soit aussi $h : C \mapsto D$. On remarque que :

$$(h \circ (g \circ f))(x) = h(g(f(x))) = ((h \circ g) \circ f)(x)$$

On note :

$$h \circ g \circ f = (h \circ (g \circ f)) = ((h \circ g) \circ f)$$

8.10.2 Neutre

On constate que

$$\operatorname{Id} \circ f = f \circ \operatorname{Id} = f$$

On dit que la fonction identité est neutre pour la composition.

8.11 Puissance

Soit une fonction $f : A \mapsto A$. La « puissance » d'une fonction est définie au moyen de la composée $\circ$ par :

$$\begin{aligned} f^0 &= \mathrm{Id} \\ f^n &= f \circ f^{n-1} \end{aligned}$$

pour tout $n \in \mathbb{N}$. On a donc en particulier $f^1 = f$ et :

$$f^n = f \circ ... \circ f$$

8.12 Fonction constante

On associe souvent à tout élément $c \in B$ une fonction constante $\hat{c} : A \mapsto B$ définie par :

$$\hat{c}(x) = c$$

pour tout $x \in A$. On note abusivement :

$$\hat{c} = c$$

8.13 Egalité

Deux fonctions $f, g : A \mapsto B$ sont égales si et seulement si leurs valeurs sont égales en tout point $x \in A$:

$$f = g \quad \Leftrightarrow \quad f(x) = g(x)$$

Chapitre 9

Opérations

Dépendances

— Chapitre 8 : Les fonctions

9.1 Introduction

Soit un ensemble Ω. Une opération $\divideontimes$ sur Ω, ou loi de composition interne sur Ω, est une fonction qui, à deux éléments de Ω, associe un troisième élément de Ω appelé résultat. On a donc formellement :

$$\divideontimes : \Omega \times \Omega \mapsto \Omega$$

Si $z \in \Omega$ est le résultat de l'opération $\divideontimes$ sur $x, y \in \Omega$, on le note :

$$z = x \divideontimes y$$

9.2 Neutre

9.2.1 À gauche

On dit qu'un élément $n \in \Omega$ est neutre à gauche pour la loi $\divideontimes$ si :

$$n \divideontimes x = x$$

pour tout $x \in \Omega$.

9.2.2 À droite

On dit qu'un élément $n \in \Omega$ est neutre à droite pour la loi $\divideontimes$ si :

$$x \divideontimes n = x$$

pour tout $x \in \Omega$.

9.2.3 Simultané

On dit qu'un élément $n \in \Omega$ est neutre pour la loi $\divideontimes$ si :

$$x \divideontimes n = n \divideontimes x = x$$

pour tout $x \in \Omega$.

9.2.4 Unicité

Si $m, n \in \Omega$ sont des éléments neutres pour $\divideontimes$, on a :

$$m = m \divideontimes n = n$$

Si un neutre existe, il est unique.

9.3 Inverse

Supposons que $n \in \Omega$ soit l'unique neutre pour la loi $\divideontimes$.

9.3.1 À gauche

On dit qu'un élément $y \in \Omega$ est un inverse à gauche de $x \in \Omega$ pour la loi $\divideontimes$ si :

$$y \divideontimes x = n$$

9.3.2 À droite

On dit qu'un élément $y \in \Omega$ est un inverse à droite de $x \in \Omega$ pour la loi $\divideontimes$ si :

$$x \divideontimes y = n$$

9.3.3 Simultané

On dit qu'un élément $x^\ddagger \in \Omega$ est l'inverse de $x \in \Omega$ pour la loi $\divideontimes$ si :

$$x \divideontimes x^\ddagger = x^\ddagger \divideontimes x = n$$

9.4 Associativité

On dit que $\divideontimes$ est associative si :

$$(x \divideontimes y) \divideontimes z = x \divideontimes (y \divideontimes z)$$

pour tout $x, y, z \in \Omega$. On définit alors :

$$x \divideontimes y \divideontimes z = x \divideontimes (y \divideontimes z) = (x \divideontimes y) \divideontimes z$$

9.4.1 Extension

Soit $x_1, x_2, ..., x_n \in \Omega$. On définit

9.4.2 Unicité de l'inverse

On suppose que $\divideontimes$ admet n comme neutre. Soit $x \in \Omega$ et $y, z \in \Omega$ deux inverses de x. On a :

$$y \divideontimes x \divideontimes z = y \divideontimes (x \divideontimes z) = y \divideontimes n = y$$

et :

$$y \divideontimes x \divideontimes z = (y \divideontimes x) \divideontimes z = n \divideontimes z = z$$

On a donc :

$$y = z$$

Dans le cadre d'une opération associative, l'inverse d'un élément donné est unique.

9.5 Commutativité

On dit que $\divideontimes$ est commutative si :

$$x \divideontimes y = y \divideontimes x$$

pour tout $x, y \in \Omega$.

9.6 Distributivité

Soit $\boxplus, \circledast$ deux lois de composition interne sur Ω. On dit que $\circledast$ se distribue sur $\boxplus$ si :

$$z \circledast (x \boxplus y) = (z \circledast x) \boxplus (z \circledast y)$$

et :

$$(x \boxplus y) \circledast z = (x \circledast z) \boxplus (y \circledast z)$$

pour tout $x, y, z \in \Omega$.

9.7 Opération induite

Soit une opération $\divideontimes$ définie sur l'ensemble Ω. L'opération induite par $\divideontimes$ sur Ω^n est définie par :

$$(x_1, x_2, ..., x_n) \divideontimes (y_1, y_2, ..., y_n) = (x_1 \divideontimes y_1, x_2 \divideontimes y_2, ..., x_n \divideontimes y_n)$$

pour tout :

$$(x_1, x_2, ..., x_n), (y_1, y_2, ..., y_n) \in \Omega^n$$

On note aussi :

$$z = x \divideontimes y$$

pour :

$$x = (x_1, x_2, ..., x_n) \in \Omega^n$$

$$y = (y_1, y_2, ..., y_n) \in \Omega^n$$

et :

$$z = (z_1, z_2, ..., z_n) \in \Omega^n$$

avec :

$$z_i = x_i \divideontimes y_i$$

pour tout $i \in \{1, 2, ..., n\}$.

Chapitre 10

Algèbre

Dépendances

— Chapitre 9 : Les opérations

10.1 Monoïde

Soit un ensemble M sur lequel est défini une opération $\divideontimes$. On dit que le couple $(M, \divideontimes)$ est un monoïde si $\divideontimes$ est associative.

10.1.1 Nomenclature

Lorsque l'opération $\divideontimes$ est évidente d'après le contexte, on dit simplement que M est un monoïde.

10.2 Groupes

Soit un ensemble G sur lequel est défini une opération $\divideontimes$. On dit que le couple $(G, \divideontimes)$ est un groupe si :

— $\divideontimes$ est associative
— Il existe un neutre pour $\divideontimes$
— Chaque élément de G possède un inverse pour $\divideontimes$

10.2.1 Nomenclature

Lorsque l'opération $\divideontimes$ est évidente d'après le contexte, on dit simplement que G est un groupe.

10.2.2 Commutatif

Si $\divideontimes$ est également commutative, on dit que $(G, \divideontimes)$ est un groupe abélien ou commutatif.

10.3 Monoïde et groupe

Soit un monoïde $(G, \divideontimes)$ tel qu'il existe un neutre à droite $n \in G$ pour $\divideontimes$ et que chaque élément $x \in G$ admet un inverse à droite $x^\ddagger$ pour $\divideontimes$:

$$x \divideontimes x^\ddagger = n$$

Soit :

$$y = x^\ddagger \divideontimes x$$

En utilisant l'associativité de $\divideontimes$, on se rend compte que :

$$y = x^\ddagger \divideontimes n \divideontimes x = x^\ddagger \divideontimes x \divideontimes x^\ddagger \divideontimes x = y \divideontimes y$$

En utilisant ce résultat, on obtient :

$$y = y \divideontimes n = y \divideontimes y \divideontimes y^\ddagger = y \divideontimes y^\ddagger = n$$

Donc :

$$x^\ddagger \divideontimes x = n$$

et $x^\ddagger$ est un inverse de x. On a aussi :

$$n \divideontimes x = x \divideontimes x^\ddagger \divideontimes x = x \divideontimes n = x$$

L'élément n est donc l'unique neutre de G et tout élément $x \in G$ admet un inverse. On en conclut que le couple $(G, \divideontimes)$ est un groupe.

10.4 Anneaux

Soit un ensemble A sur lequel sont définies les opérations $\boxplus$, appelée addition, et l'opération $\circledast$, appelée multiplication. On dit que le tuple $(A, \boxplus, \circledast)$ est un anneau si :
- L'addition est commutative et associative
- Il existe un neutre pour l'addition
- Chaque élément de A possède un inverse pour l'addition appelé opposé
- La multiplication est associative
- La multiplication se distribue sur l'addition

10.4.1 Nomenclature

Lorsque les opérations sont évidentes d'après le contexte, on dit simplement que A est un anneau.

10.4.2 Unitaire

Si la multiplication admet également un neutre, on parle d'anneau unitaire.

10.4.3 Notations

Soit un anneau A. L'addition de $x, y \in A$ est généralement notée :

$$x + y$$

La multiplication est généralement notée :

$$x \cdot y$$

On désigne généralement par 0 le neutre pour l'addition :

$$x + 0 = 0 + x = x$$

Si l'anneau A est unitaire, on désigne généralement par 1 le neutre pour la multiplication :

$$x \cdot 1 = 1 \cdot x = x$$

L'inverse de x pour l'addition, aussi nommé opposé, est noté $-x$:

$$x + (-x) = (-x) + x = 0$$

10.4.4 Intègre

Soit un anneau $(A, +, \cdot)$. Si la relation :

$$a \cdot b = 0$$

implique :

$$a = 0 \quad \text{ou} \quad b = 0$$

on dit que $(A, +, \cdot)$ est un anneau intègre.

10.5 Corps

Soit un ensemble K sur lequel sont définies les opérations $\boxplus$, appelée addition, et l'opération $\circledast$, appelée multiplication. On dit que le tuple $(K, \boxplus, \circledast)$ est un corps si :
- L'addition est commutative et associative
- Il existe un neutre pour l'addition
- Chaque élément de K possède un inverse pour l'addition appelé opposé
- La multiplication est associative
- Il existe un neutre pour la multiplication
- Chaque élément de K, à l'exception du neutre pour l'addition, possède un inverse pour la multiplication appelé inverse
- La multiplication se distribue sur l'addition

10.5.1 Nomenclature

Lorsque les opérations sont évidentes d'après le contexte, on dit simplement que K est un corps.

10.5.2 Symbole

On utilise souvent le symbole $\mathbb{K}$ pour désigner un corps générique.

10.5.3 Commutatif

Si la multiplication est également commutative, on parle de corps commutatif.

10.5.4 Notations

Soit un corps K. L'addition de $x, y \in K$ est notée :

$$x + y$$

La multiplication est notée :

$$x \cdot y$$

On désigne généralement par 0 le neutre pour l'addition :

$$x + 0 = 0 + x = x$$

et par 1 le neutre pour la multiplication :

$$x \cdot 1 = 1 \cdot x = x$$

L'inverse de x pour l'addition, aussi nommé opposé, est noté $-x$:

$$x + (-x) = (-x) + x = 0$$

et l'inverse de x pour la multiplication est noté x^{-1} ou $1/x$:

$$x \cdot x^{-1} = x^{-1} \cdot x = 1$$

10.6 Soustraction et division

Lorsque l'addition est définie et que b possède un opposé $-b$, on définit généralement l'opération de soustraction par :

$$a - b = a + (-b)$$

Lorsque la multiplication est définie et que b dispose d'un inverse b^{-1}, on définit généralement l'opération de division par :

$$\frac{a}{b} = a \cdot b^{-1}$$

On note aussi :

$$a/b = \frac{a}{b}$$

10.7 Les puissances

On définit généralement les puissances positives par :

$$\begin{aligned} x^0 &= 1 \\ x^k &= x \cdot x^{k-1} \end{aligned}$$

pour tout $k \in \mathbb{N}$. Si l'inverse de x existe, on définit généralement les puissances négatives par :

$$x^{-k} = (x^{-1})^k$$

pour tout $k \in \mathbb{N}$.

10.8 Somme d'ensembles

Soit l'ensemble Ω sur lequel est définie une opération d'addition. On définit la somme de deux sous-ensembles $A, B \subseteq \Omega$ par :

$$A + B = \{x + y : (x, y) \in A \times B\}$$

Somme directe

Si, pour tout $s \in S$, il existe un et un seul couple $(x, y) \in A \times B$ tel que $s = x + y$, on dit que S est la somme directe de A et B et on le note :

$$S = A \bigoplus B$$

10.9 Multiplication mixte

Soit l'ensemble Ω sur lequel est définie une opération de multiplication. On définit la multiplication de $\lambda \in \Omega$ et de $A \subseteq \Omega$ par :

$$\lambda \cdot A = \{\lambda \cdot x : x \in A\}$$

Deuxième partie

Relations

Chapitre 11

Equivalences

Dépendances

— Chapitre 7 : Les relations

11.1 Définition

Une équivalence $\equiv$ sur un ensemble A est une relation permettant de regrouper les éléments ayant une caractéristique similaire. Choisissons $x, y, z \in A$. Tout élément x étant égal à lui-même, il doit bien entendu être équivalent à lui-même. Notre équivalence doit donc respecter la propriété :

$$x \equiv x$$

Par ailleurs, si x est équivalent à y, l'inverse doit aussi être vrai :

$$x \equiv y \quad \Rightarrow \quad y \equiv x$$

Il est également clair que si x est équivalent y et que y est équivalent à z, notre x doit être équivalent à z. Donc :

$$x \equiv y, \quad y \equiv z \quad \Rightarrow \quad x \equiv z$$

Relation

On peut associer une relation R à toute équivalence en posant :

$$R = \{(x, y) \in A^2 : x \equiv y\}$$

On a alors $x \equiv y$ si et seulement si $(x, y) \in R$.

Multiple

La notation $x \equiv y \equiv z$ signifie que $x \equiv y$ et que $y \equiv z$.

11.2 Classe d'équivalence

Soit $x \in A$. La classe d'équivalence associée à x est l'ensemble des éléments de A qui lui sont équivalents :

$$\mathcal{E}(x) = \{y \in A : y \equiv x\}$$

11.3 Ensemble quotient

Soit la relation $R \subseteq A^2$ associée à l'équivalence $\equiv$ définie sur A. L'ensemble quotient de A par R est la collection des classes d'équivalence :

$$A/R = \{\mathcal{E}(x) \in \mathfrak{P}(A) : x \in A\}$$

11.3.1 Intersection

Soit $x, y \in A$.

— Supposons que x n'est pas équivalent à y et choisissons un z dans l'intersection :

$$z \in \mathcal{E}(x) \cap \mathcal{E}(y)$$

On doit donc avoir $x \equiv z$ et $z \equiv y$, alors que x n'est pas équivalent à y, ce qui contredit la définition des équivalences. Par conséquent, un tel z ne peut pas exister et l'intersection est vide :

$$\mathcal{E}(x) \cap \mathcal{E}(y) = \emptyset$$

— A présent, supposons que x est équivalent à y et choisissons un $z \in \mathcal{E}(x)$. On a donc $z \equiv x$ et $x \equiv y$. On en déduit que $z \equiv y$, d'où $z \in \mathcal{E}(y)$ et $\mathcal{E}(x) \subseteq \mathcal{E}(y)$. Symétriquement, si $w \in \mathcal{E}(y)$, on a $w \equiv y$ et $y \equiv x$, d'où $w \in \mathcal{E}(x)$ et $\mathcal{E}(y) \subseteq \mathcal{E}(x)$. On a donc :

$$\mathcal{E}(x) = \mathcal{E}(y)$$

11.3.2 Union

Quel que soit $x \in A$, tous les éléments de $\mathcal{E}(x)$ sont dans A par définition. L'union :

$$E = \bigcup_{x \in A} \mathcal{E}(x)$$

est donc également inclue dans A. D'un autre coté, tout $x \in A$ étant équivalent à lui-même, chaque $\mathcal{E}(x)$ contient au moins x. On en conclut que A est inclus dans E. La double inclusion nous montre que :

$$A = \bigcup_{x \in A} \mathcal{E}(x)$$

11.3.3 Partition

On déduit de ce qui précède que la collection des classes d'équivalences :

$$\mathcal{P} = A/R$$

forme une partition de A. Cette constation explique la terminologie de quotient, liée à celle de division et de partition.

Chapitre 12

Ordres

Dépendances

— Chapitre 7 : Les relations

12.1 Ordre large

Un ordre, ou ordre large, est une relation permettant de déterminer si un élément d'un ensemble est « plus petit ou égal » qu'un autre (on dit aussi « inférieur » à un autre).

Soit $\le$ un ordre sur l'ensemble Ω. Choisissons $x, y, z \in \Omega$. Comme tout élément x est égal à lui-même, il est forcément « plus petit ou égal » à lui-même, et notre ordre doit respecter la propriété :

$$x \le x$$

Par ailleurs, si x est plus petit ou égal à y et que l'inverse est vrai aussi, on doit avoir l'égalité entre les deux éléments :

$$x \le y, \quad y \le x \quad \Rightarrow \quad x = y$$

Il est également clair que si x est plus petit que y et que si y est plus petit que z, x doit être plus petit que z. Donc :

$$x \le y, \quad y \le z \quad \Rightarrow \quad x \le z$$

12.1.1 Multiple

La notation $x \le y \le z$ signifie que $x \le y$ et que $y \le z$.

12.2 Plus grand ou égal

Soit $x, y \in \Omega$. On dit que x est « plus grand ou égal », ou « supérieur » à y, et on le note :

$$x \ge y$$

si et seulement si :

$$y \le x$$

12.3 Relation associée

On peut associer une relation R à tout ordre en posant :

$$R = \{(x, y) \in \Omega^2 : x \le y\}$$

On a alors $x \le y$ si et seulement si $(x, y) \in R$.

12.4 Ordonné

Soit un ensemble Ω muni d'un ordre $\le$. On dit alors que Ω est un ensemble ordonné ou que le couple $(\Omega, \le)$ est un espace ordonné.

12.5 Ordre total et partiel

Un ordre $\le$ sur Ω est dit total si pour tout $x, y \in \Omega$, on a $x \le y$ ou $y \le x$. Dans le cas contraire, on dit que l'ordre est partiel.

12.6 Ordre sur un produit cartésien

Soit la collection d'ensembles ordonnés $A_1, A_2, ..., A_n$ et les éléments $x_i, y_i \in A_i$. On définit un ordre partiel sur les n-tuples :

$$x = (x_1, x_2, ..., x_n)$$
$$y = (y_1, y_2, ..., y_n)$$

en disant que :

$$x \le y$$

si et seulement si :

$$x_i \le y_i$$

pour tout $i \in \{1, 2, ..., n\}$.

Chapitre 13

Ordre strict

Dépendances

— Chapitre 12 : Les ordres

13.1 Définition

Un ordre strict sur un ensemble Ω est une relation permettant de déterminer si un élément de Ω est « strictement plus petit » qu'un autre. La différence fondamentale avec l'ordre large étant que les deux éléments doivent être distincts.

Soit $<$ un ordre strict sur l'ensemble Ω et $x, y, z \in \Omega$. Comme x ne peut pas être distinct de lui-même, *on ne peut pas avoir* $x < x$. La relation $x < y$ implique donc que $x \neq y$.

La seule solution pour avoir simultanément $x < y$ et $y < x$ serait que $x = y$ comme dans le cas de l'ordre large. Ce n'est pas possible par définition. Les relations $x < y$ et $z < x$ impliquent donc que $y \neq z$. Par contre la propriété :

$$x < y, \quad y < z \quad \Rightarrow \quad x < z$$

est analogue à celle de l'ordre large.

13.1.1 Multiple

La notation $x < y < z$ signifie que $x < y$ et que $y < z$.

13.2 Plus grand

Soit $x, y \in \Omega$. On dit que x est strictement plus grand que y, et on le note :

$$x > y$$

si et seulement si :

$$y < x$$

13.3 Dérivé du large

On peut définir un ordre strict à partir d'un ordre large en disant que $x < y$ si et seulement si $x \leq y$ et $x \neq y$.

13.4 Complémentarité

Supposons que l'ordre $\le$ soit total. Si x, y ne vérifient pas $x \le y$, on doit forcément avoir $y \le x$. Si on avait $x = y$, on aurait aussi $x \le y$ ce qui contredit l'hypothèse. Par conséquent, on doit avoir $x \ne y$. On en déduit que $y < x$.

Réciproquement, si la condition $y < x$ n'est pas vérifiée, on a soit $x = y$, soit $x \le y$. Mais comme $x \le y$ inclut la possibilité que $x = y$, on a simplement $x \le y$.

On a donc soit $x \le y$, soit $x > y$. Ou encore, soit $x < y$, soit $x = y$, soit $x > y$.

Chapitre 14

Extrema

Dépendances

— Chapitre 7 : Les relations
— Chapitre 12 : Les ordres

14.1 Comparaison élément - ensemble

Soit $\le$ un ordre sur l'ensemble Ω et $A \subseteq \Omega$. Nous dirons qu'un objet $m \in \Omega$ est inférieur à l'ensemble A, ou que m est un minorant de A, et nous le noterons :

$$m \le A$$

si $m \le a$ pour tout élément $a \in A$. Inversément, nous dirons que m est supérieur à A, ou que m est un majorant de A, et nous le noterons :

$$m \ge A$$

si $m \ge a$ pour tout élément $a \in A$:

14.2 Comparaison ensemble - ensemble

14.2.1 En-dessous

Soit $\le$ un ordre sur l'ensemble Ω et $A, B \subseteq \Omega$. Nous dirons que A est en-dessous de l'ensemble B, et nous le noterons :

$$A \preccurlyeq B$$

si tout élément $a \in A$ vérifie $a \le B$. Cela revient à imposer l'inégalité :

$$a \le b$$

pour tout couple $(a, b) \in A \times B$.

14.2.2 Au-dessus

On dit que A est au-dessus de l'ensemble B, et on le note :

$$A \succcurlyeq B$$

si et seulement si :

$$B \leqslant A$$

Cela revient à imposer l'inégalité :

$$a \ge b$$

pour tout couple $(a, b) \in A \times B$.

14.2.3 Remarque

Attention, ces comparaisons ne constituent pas un ordre. En particulier, un ensemble quelconque n'est en général ni au-dessus ni en-dessous de lui-même.

14.3 Majorants et minorants

Soit $\le$ un ordre sur l'ensemble de référence Ω et $A \subseteq \Omega$. L'ensemble des majorants de A est l'ensemble des éléments de Ω supérieurs à A :

$$\operatorname{major} A = \{m \in \Omega : m \ge A\}$$

Si $\operatorname{major} A \ne \emptyset$, on dit que A est majoré ou encore que A est borné supérieurement. L'ensemble des minorants de A est l'ensemble des éléments de Ω inférieurs à A :

$$\operatorname{minor} A = \{m \in \Omega : m \le A\}$$

Si $\operatorname{minor} A \ne \emptyset$, on dit que A est minoré ou encore que A est borné inférieurement.

14.3.1 Préciser l'ordre

Lorsqu'on veut préciser l'ordre $\le$ utilisé, on note :

$$\operatorname*{major}_{\le} A = \{m \in \Omega : m \ge A\}$$

$$\operatorname*{minor}_{\le} A = \{m \in \Omega : m \le A\}$$

14.3.2 Comparaisons d'ensembles

On se rend compte que si $B \subseteq \Omega$ vérifie $B \geqslant A$, tous les éléments de B sont dans l'ensemble des majorants :

$$B \geqslant A \ \Rightarrow \ B \subseteq \operatorname{major} A$$

Inversément, si B est en-dessous de A, tous ses éléments sont dans l'ensemble des minorants :

$$B \leqslant A \ \Rightarrow \ B \subseteq \operatorname{minor} A$$

14.4 Éléments maximaux et minimaux

On dit d'un élément $M \in A$ qu'il est maximal dans A si pour tout $a \in A$ vérifiant $a \ge M$, on a $M = a$. Autrement dit, aucun élément de A distinct de M n'est supérieur à ce dernier. M est donc le seul élément de A à être supérieur (au sens large) à lui-même. On a donc $\operatorname{major}\{M\} \cap A = \{M\}$. On note $\operatorname{maxim} A$ l'ensemble des éléments maximaux de A :

$$\operatorname{maxim} A = \Big\{ M \in A : \operatorname{major}\{M\} \cap A = \{M\} \Big\}$$

Symétriquement, on dit d'un élément $m \in A$ qu'il est minimal dans A si pour tout $a \in A$ vérifiant $a \le m$, on a $m = a$. Autrement dit, aucun élément de A distinct de m n'est inférieur à ce dernier. m est donc le seul élément de A à être inférieur (au sens large) à lui-même. On a donc $\operatorname{minor}\{m\} \cap A = \{m\}$. On note $\operatorname{minim} A$ l'ensemble des éléments minimaux de A :

$$\operatorname{minim} A = \Big\{ m \in A : \operatorname{minor}\{m\} \cap A = \{m\} \Big\}$$

14.4.1 Préciser l'ordre

Lorsqu'on veut préciser l'ordre $\le$ utilisé, on note :

$$\operatorname*{maxim}_{\le} A$$

$$\operatorname*{minim}_{\le} A$$

14.5 Maximum et minimum

Considérons le cas où l'on peut trouver des majorants de A appartenant à A. Si $x, y \in A \cap \operatorname{major} A$, on a $x \ge y$ puisque x est un majorant de A et que $y \in A$. Mais on a aussi $y \ge x$ puisque y est également un majorant de A et que $x \in A$. On en conclut que $x = y$ et que l'intersection ne contient qu'un seul élément M :

$$A \cap \operatorname{major} A = \{M\}$$

On dit alors que $M \in A$ est le maximum de l'ensemble A et on le note :

$$M = \max A = \max_{a \in A} a$$

A présent, supposons que l'on puisse trouver des minorants de A appartenant à A. Supposons que $x, y \in A \cap \operatorname{minor} A$. On a alors $x \le y$ puisque x est un minorant de A et que $y \in A$. Mais on a aussi $y \le x$ puisque y est également un minorant de A et que $x \in A$. On en conclut que $x = y$. L'intersection ne contient donc qu'un seul élément m :

$$A \cap \operatorname{minor} A = \{m\}$$

On dit alors que m est le minimum de l'ensemble A et on le note :

$$m = \min A = \min_{a \in A} a$$

14.5.1 Préciser l'ordre

Lorsqu'on veut préciser l'ordre $\le$ utilisé, on note plutôt :

$$\max_{\le} A = \max A$$

$$\min_{\le} A = \min A$$

14.5.2 Maximal et minimal

Supposons que $M = \max A$ existe. On a $M \in A$. Soit $a \in A$ vérifiant $M \le a$. Par définition du maximum, on a aussi $M \ge a$, d'où $M = a$. On en conclut que le maximum est un élément maximal. A présent, soit $x \in \operatorname{maxim} A$. Comme $x \le M$, on en conclut que $x = M$. On a donc :

$$\operatorname{maxim} A = \{\max A\}$$

Supposons que $m = \min A$ existe. On a $m \in A$. Soit $a \in A$ vérifiant $m \ge a$. Par définition du minimum, on a aussi $m \le a$, d'où $m = a$. On en conclut que le minimum est un élément minimal. A présent, soit $x \in \operatorname{minim} A$. Comme $x \ge m$, on en conclut que $x = m$. On a donc :

$$\operatorname{minim} A = \{\min A\}$$

14.6 Supremum et infimum

Soit un ensemble $A \subseteq \Omega$ dont le maximum $M = \max A$ existe et l'élément $b \in \Omega$ tel que $b \ge A$. Comme M est un élément de A, on a forcément $b \ge M$. Mais d'un autre coté, $M \ge A$ par définition. Donc :

$$b \ge M \ge A$$

Comme ces relations sont valables pour tout les b supérieurs à A, on en conclut que :

$$\operatorname{major} A \ge M \ge A$$

Relations qui nous disent que M est le plus petit des objets supérieurs à A. Considérons à présent un cas plus général où nous ne supposons pas que le maximum de A existe. Supposons seulement que l'on puisse trouver un $S \in \Omega$ tel que :

$$\operatorname{major} A \ge S \ge A$$

La première inégalité nous dit que S est un minorant de $\operatorname{major} A$. La seconde nous dit que S est un majorant de A, autrement dit $S \in \operatorname{major} A$. Les deux conditions nous permettent d'affirmer que S est le minimum de l'ensemble des majorants de A :

$$S = \min \operatorname{major} A$$

Cet élément S est donc unique. On l'appelle supremum et on le note :

$$S = \sup A = \sup_{a \in A} a = \min \operatorname{major} A$$

Il est important de remarquer que S n'appartient pas forcément à A. On procéde de la même façon pour étendre la notion de minimum. Si on peut trouver un $I \in \Omega$ tel que :

$$\operatorname{minor} A \le I \le A$$

on en conclut que I est le maximum de l'ensemble des minorants de A et qu'il est donc unique. On l'appelle infimum et on le note :

$$I = \inf A = \inf_{a \in A} a = \max \operatorname{minor} A$$

14.6.1 Préciser l'ordre

Lorsqu'on veut préciser l'ordre $\le$ utilisé, on note :

$$\sup_{\le} A = \min \operatorname{major} A$$

$$\inf_{\le} A = \max \operatorname{minor} A$$

14.6.2 Comparaison des bornes

Si A est non vide, il suffit de choisir $x \in A$ pour se rendre compte que :

$$\inf A \le x \le \sup A$$

d'où :

$$\inf A \le \sup A$$

14.6.3 Cas particulier

Dans le cas où le maximum de A existe, on a par unicité $\sup A = \max A$. Dans le cas où le minimum de A existe, on a par unicité $\inf A = \min A$.

14.7 Ensembles finis

Nous allons montrer que, sous l'hypothèse d'un ordre total, tout ensemble comportant un nombre fini d'éléments admet un maximum et un minimum. Si l'ensemble ne contient qu'un élément, soit $A_1 = \{a\}$, on a $a \le a$ et donc :

$$a = \max\{a\} = \min\{a\}$$

Supposons à présent que l'ensemble comporte deux éléments, soit $A_2 = \{a_1, a_2\}$. Si $a_1 \ge a_2$, on a $a_1 \ge \{a_1, a_2\}$. Dans le cas contraire, on a $a_2 \ge \{a_1, a_2\}$. On en déduit que :

$$\max\{a_1, a_2\} = \begin{cases} a_1 & \text{si } a_1 \ge a_2 \\ a_2 & \text{sinon} \end{cases}$$

Le minimum est donné par l'expression duale :

$$\min\{a_1, a_2\} = \begin{cases} a_1 & \text{si } a_1 \le a_2 \\ a_2 & \text{sinon} \end{cases}$$

Supposons à présent avoir démontré que tout ensemble de $n-1$ éléments admettait un maximum et un minimum. Soit l'ensemble $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ comportant n éléments. L'ensemble $A_{n-1} = \{a_1, a_2, ..., a_{n-1}\}$ contient $n-1$ éléments et admet donc :

$$\sigma = \max\{a_1, a_2, ..., a_{n-1}\}$$
$$\lambda = \min\{a_1, a_2, ..., a_{n-1}\}$$

Posons $\alpha = \max\{\sigma, a_n\}$. On voit que $\alpha \in A$, que $\alpha \ge \sigma \ge A_{n-1}$ et que $\alpha \ge a_n$. On en déduit que $\alpha \ge A$. Donc :

$$\max\{a_1, a_2, ..., a_{n-1}, a_n\} = \max\Big\{ \max\{a_1, a_2, ..., a_{n-1}\}, a_n \Big\}$$

Posons $\beta = \min\{\lambda, a_n\}$. On voit que $\beta \in A$, que $\beta \le \lambda \le A_{n-1}$ et que $\beta \le a_n$. On en déduit que $\beta \le A$. Donc :

$$\min\{a_1, a_2, ..., a_{n-1}, a_n\} = \min\Big\{ \min\{a_1, a_2, ..., a_{n-1}\}, a_n \Big\}$$

Tout ensemble comportant un nombre fini d'éléments et sur lequel est défini un ordre total admet donc un maximum et un minimum. De plus, la démonstration nous donne un moyen d'évaluer récursivement ces extrema.

14.7.1 Notations

On note aussi :

$$\max_{i=1}^{n} a_i = \max\{a_1, a_2, ..., a_n\}$$

$$\min_{i=1}^{n} a_i = \min\{a_1, a_2, ..., a_n\}$$

14.8 Infini

Il arrive qu'aucun élément de Ω ne soit un majorant de Ω. On est alors amené à ajouter à l'ensemble une borne supérieure, notée $+\infty$ (ou ∞), telle que tout élément $a \in \Omega$ vérifie $a < +\infty$. Le concept de l'infini négatif, noté $-\infty$, est similaire. Il s'agit d'une borne inférieure de Ω telle que tout élément $a \in \Omega$ vérifie $a > -\infty$.

14.8.1 Supremum et infimum

Lorsque l'ensemble des majorants de A est vide, le supremum n'existe pas. On dit alors qu'il est infini et on le note :

$$\sup A = +\infty$$

Lorsque l'ensemble des minorants de A est vide, l'infimum n'existe pas. On dit alors qu'il est infini et on le note :

$$\inf A = -\infty$$

14.9 Ordre et inclusion

Soit $A, B \subseteq \Omega$ avec $A \subseteq B$. Soit l'ordre $\le$ défini sur Ω.

14.9.1 Max - Min

Supposons que les maxima existent. Le maximum de A étant dans A, il est aussi dans B. Par conséquent, le maximum de B est supérieur au maximum de A :

$$\max A \le \max B$$

En suivant le même raisonnement avec les minima, on obtient :

$$\min A \ge \min B$$

14.9.2 Sup - Inf

A présent, ne supposons plus l'existence des maxima ou des minima, mais seulement des supremums et infimums. Tout majorant de B sera aussi un majorant de A. On a donc :

$$\operatorname{major} B \subseteq \operatorname{major} A$$

L'ensemble $\operatorname{major} B$ étant inclus dans $\operatorname{major} A$, son minimum doit être supérieur, et on a :

$$\sup A \le \sup B$$

En répétant le même raisonnement avec les minorants, on obtient :

$$\inf A \ge \inf B$$

14.10 Ordre, union et intersection

14.10.1 Max - Min de l'union

Supposons que les maxima de A et B existent. Si l'ordre est total, on a soit $\max A \le \max B$, soit $\max B \le \max A$. Le maximum :

$$M = \max\{\max A, \max B\}$$

existe donc. On voit que M appartient à A ou à B suivant les cas, c'est-à-dire à $A \cup B$ et que :

$$x \le \max A \le M$$
$$y \le \max B \le M$$

pour tout $x \in A$ et $y \in B$. On en conclut que M majore $A \cup B$. On a donc :

$$\max\{\max A, \max B\} = \max(A \cup B)$$

En suivant le même raisonnement avec les minima, on obtient :

$$\min\{\min A, \min B\} = \min(A \cup B)$$

14.10.2 Max - Min de l'intersection

Supposons que les maxima de A et B existent et que leur intersection soit non vide :

$$A \cap B \neq \emptyset$$

Si l'ordre est total, on a soit $\max A \le \max B$, soit $\max B \le \max A$. Le minimum :

$$m = \min\{\max A, \max B\}$$

existe donc. On voit que pour tout $x \in A \cap B$:

$$x \in A \;\Rightarrow\; x \le \max A$$
$$x \in B \;\Rightarrow\; x \le \max B$$

On en conclut que $x \leq m$. L'élément m est donc un majorant de $A \cap B$. Si le maximum de l'intersection existe, on a donc :

$$\max(A \cap B) \leq \min\{\max A, \max B\}$$

En suivant le même raisonnement avec les minima, on obtient :

$$\min(A \cap B) \geq \max\{\min A, \min B\}$$

14.10.3 Sup - Inf de l'union

A présent, ne supposons plus l'existence des maxima ou des minima, mais seulement des supremums et infimums. Comme les majorant de $A \cup B$ sont ceux qui majorent tous les éléments de A et tous les éléments de B, on a :

$$\operatorname{major}(A \cup B) = (\operatorname{major} A) \cap (\operatorname{major} B)$$

Il ne nous reste plus qu'à minimiser les deux membres de cette égalité. Le minimum de l'intersection étant supérieure au maximum des minima, on a :

$$\min\Big[(\operatorname{major} A) \cap (\operatorname{major} B)\Big] \geq \max\{\min \operatorname{major} A, \min \operatorname{major} B\}$$

et :

$$\min \operatorname{major}(A \cup B) \geq \max\{\min \operatorname{major} A, \min \operatorname{major} B\}$$

c'est-à-dire :

$$\sup(A \cup B) \geq \max\{\sup A, \sup B\}$$

En répétant le même raisonnement avec les minorants, on obtient :

$$\operatorname{minor}(A \cup B) = (\operatorname{minor} A) \cap (\operatorname{minor} B)$$

et :

$$\inf(A \cup B) \leq \min\{\inf A, \inf B\}$$

14.10.4 Sup - Inf de l'intersection

Soit $x \in A \cap B$ et $M \in (\operatorname{major} A) \cup (\operatorname{major} B)$. On a soit :
— $M \in \operatorname{major} A$ et alors $M \geq x$ car $x \in A$
— $M \in \operatorname{major} B$ et alors $M \geq x$ car $x \in B$
On en déduit que :

$$(\operatorname{major} A) \cup (\operatorname{major} B) \subseteq \operatorname{major}(A \cap B)$$

Le minimum d'un ensemble inclus dans un autre devant être plus grand, on obtient :

$$\min\Big[(\operatorname{major} A) \cup (\operatorname{major} B)\Big] \geq \min \operatorname{major}(A \cap B)$$

Comme le minimum de l'union est égal au minimum des minima, on a :

$$\min\Big[(\operatorname{major} A) \cup (\operatorname{major} B)\Big] = \min\{\min(\operatorname{major} A), \min(\operatorname{major} B)\}$$

et :

$$\min \operatorname{major}(A \cap B) \le \min\{\min(\operatorname{major} A), \min(\operatorname{major} B)\}$$

On a donc par définition :

$$\sup(A \cap B) \le \min\{\sup A, \sup B\}$$

En répétant le même raisonnement avec les minorants, on obtient :

$$\inf(A \cap B) \ge \max\{\inf A, \inf B\}$$

Chapitre 15

Dualité

Dépendances

— Chapitre 12 : Les ordres

15.1 Ordre dual

Soit un ensemble Ω sur lequel est défini un ordre $\le$. Soit $x, y \in \Omega$. L'ordre dual de $\le$, noté $\le^\star$, est défini par :

$$x \le^\star y$$

si et seulement si :

$$y \le x$$

15.1.1 Ordre primal

Par opposition à l'ordre dual $\le^\star$, l'ordre $\le$ est appelé ordre primal.

15.2 Comparaison élément - ensemble

Soit $A \subseteq \Omega$ et un $m \in \Omega$ vérifiant $m \le A$. On a $m \le a$ pour tout $a \in A$, autrement dit $m \ge^\star a$. On en conclut que :

$$m \ge^\star A$$

Soit $m \in \Omega$ vérifiant $m \ge A$. On a $m \ge a$ pour tout $a \in A$, autrement dit $m \le^\star a$. On en conclut que :

$$m \le^\star A$$

15.3 Majorants et minorants

Soit $A \subseteq \Omega$. Si :

$$m \in \operatorname*{minor}_{\le} A$$

on a $m \le A$. Donc $m \ge^\star A$ et :

$$m \in \operatorname*{major}_{\le^\star} A$$

Si :

$$m \in \operatorname*{major}_{\le^\star} A$$

on a $m \ge A$. Donc $m \le^\star A$ et :

$$m \in \operatorname*{minor}_{\le} A$$

On en conclut que :

$$\operatorname*{major}_{\le^\star} A = \operatorname*{minor}_{\le} A$$

On montre avec des arguments similaires que :

$$\operatorname*{minor}_{\le^\star} A = \operatorname*{major}_{\le} A$$

15.4 Éléments maximaux et minimaux

Soit $A \subseteq \Omega$. Si :

$$m \in \operatorname*{minim}_{\le} A$$

on a $m = a$ pour tout $a \in A$ vérifiant $a \le m$. Donc, si $a \ge^\star m$ on a $a \le m$ et $m = a$. Autrement dit :

$$m \in \operatorname*{maxim}_{\le^\star} A$$

Si :

$$m \in \operatorname*{maxim}_{\le^\star} A$$

on a $m = a$ pour tout $a \in A$ vérifiant $a \ge^\star m$. Donc, si $a \le m$ on a $a \ge^\star m$ et $m = a$. Autrement dit :

$$m \in \operatorname*{minim}_{\le} A$$

On en conclut que :

$$\operatorname*{maxim}_{\le^\star} A = \operatorname*{minim}_{\le} A$$

On montre avec des arguments similaires que :

$$\operatorname*{minim}_{\le^\star} A = \operatorname*{maxim}_{\le} A$$

15.5 Maximum et minimum

Soit $A \subseteq \Omega$. Si l'ensemble des éléments minimaux pour $\le$ se limite au minimum, on a :

$$\underset{\le^\star}{\operatorname{maxim}} A = \underset{\le}{\operatorname{minim}} A = \left\{ \underset{\le}{\min} A \right\}$$

L'ensemble des éléments maximaux pour $\le^\star$ se résumant à un singleton, on en déduit que le maximum pour $\le^\star$ existe :

$$\underset{\le^\star}{\operatorname{maxim}} A = \left\{ \underset{\le^\star}{\max} A \right\}$$

et que :

$$\underset{\le^\star}{\max} A = \underset{\le}{\min} A$$

On montre avec des arguments similaires que :

$$\underset{\le^\star}{\min} A = \underset{\le}{\max} A$$

15.6 Supremum et Infimum

Soit $A \subseteq \Omega$ admettant le supremum :

$$m = \underset{\le^\star}{\sup} A$$

On a :

$$A \le^\star m \le^\star \underset{\le^\star}{\operatorname{major}} A$$

ce qui implique :

$$\underset{\le}{\operatorname{minor}} A = \underset{\le^\star}{\operatorname{major}} A \le m \le A$$

et vice versa. On en déduit que :

$$\underset{\le^\star}{\sup} A = \underset{\le}{\inf} A$$

On montre avec des arguments similaires que :

$$\underset{\le^\star}{\inf} A = \underset{\le}{\sup} A$$

Chapitre 16

Fonctions et ordre

Dépendances

— Chapitre 12 : Les ordres
— Chapitre 14 : Les extrema

16.1 Monotonie

16.1.1 Croissance

On dit qu'une fonction $f : A \to F$ est croissante si :

$$f(x) \ge f(y)$$

pour tout $x, y \in A$ tels que $x \ge y$. Autrement dit, une fonction croissante conserve l'ordre.

16.1.2 Décroissance

On dit qu'une fonction $f : A \to F$ est décroissante si :

$$f(x) \le f(y)$$

pour tout $x, y \in A$ tels que $x \ge y$. Autrement dit, une fonction décroissante inverse l'ordre.

16.1.3 Stricte

On dit qu'une fonction $f : A \to F$ est strictement croissante si :

$$f(x) \ < \ f(y)$$

pour tout $x, y \in A$ tels que $x \ < \ y$.
On dit qu'une fonction $f : A \to F$ est strictement décroissante si :

$$f(x) \ > \ f(y)$$

pour tout $x, y \in A$ tels que $x \ < \ y$.

16.2 Ordre entre fonctions

Soit les fonctions $f, g \in B^A$. Supposons qu'il existe un ordre défini sur B. On dit que f est inférieure à g et on le note :

$$f \le g$$

si et seulement si les valeurs de f sont inférieures aux valeurs de g :

$$f(x) \le g(x)$$

en tout point $x \in A$. Symétriquement, on dit que f est supérieure à g et on le note :

$$f \ge g$$

si :

$$f(x) \ge g(x)$$

pour tout $x \in A$.

16.2.1 Strict

On note également :

- $f < g$ si $f(x) < g(x)$ pour tout $x \in A$
- $f > g$ si $f(x) > g(x)$ pour tout $x \in A$

16.3 Fonctions extrema

Soit un ensemble de paramètres Z et la collection paramétrée de fonctions :

$$\{f_z \in B^A : z \in Z\}$$

Sous réserve d'existence des extrema, on définit la fonction supremum par :

$$\left[\sup_{z \in Z} f_z\right](x) = \sup_{z \in Z} f_z(x)$$

pour tout $x \in A$. On définit la fonction infimum par :

$$\left[\inf_{z \in Z} f_z\right](x) = \inf_{z \in Z} f_z(x)$$

pour tout $x \in A$.

16.3.1 Maximum et minimum

Lorsque les maximum et minimum existent, on définit :

$$\left[\max_{z \in Z} f_z\right](x) = \max_{z \in Z} f_z(x)$$

$$\left[\min_{z \in Z} f_z\right](x) = \min_{z \in Z} f_z(x)$$

16.3.2 Notation

On note aussi :

$$\sup\{f_z : z \in Z\} = \sup_{z \in Z} f_z$$

$$\inf\{f_z : z \in Z\} = \inf_{z \in Z} f_z$$

$$\max\{f_z : z \in Z\} = \max_{z \in Z} f_z$$

$$\min\{f_z : z \in Z\} = \min_{z \in Z} f_z$$

16.3.3 Couples

On définit les fonctions $M = \max\{f, g\}$ et $m = \min\{f, g\}$ par :

$$M(x) = \max\{f, g\}(x) = \max\{f(x), g(x)\}$$
$$m(x) = \min\{f, g\}(x) = \min\{f(x), g(x)\}$$

pour tout $x \in A$.

16.4 Ordre mixte

Soit la fonction $f \in B^A$ et un $c \in B$. Supposons qu'il existe un ordre défini sur B. On dit que f est inférieure à c et on le note :

$$f \le c$$

si les valeurs de f sont inférieures à c :

$$f(x) \le c$$

en tout point $x \in A$. Symétriquement, on dit que f est supérieure à c et on le note :

$$f \ge c$$

si :

$$f(x) \ge c$$

pour tout $x \in A$.

16.4.1 Strict

On note également :
— $f < c$ si $f(x) < c$ pour tout $x \in A$.
— $f > c$ si $f(x) > c$ pour tout $x \in A$.

16.4.2 Fonctions max et min

On définit les fonctions $\max\{f, c\}$ et $\min\{f, c\}$ par :

$$\max\{f, c\}(x) = \max\{f(x), c\}$$
$$\min\{f, c\}(x) = \min\{f(x), c\}$$

pour tout $x \in A$.

16.5 Extrema d'une fonction

Etant donné la fonction $f : \Omega \mapsto B$ et le sous-ensemble $A \subseteq \Omega$, on définit les extrema de f (s'ils existent) par :

$$\max_{x \in A} f(x) = \max\{f(x) : x \in A\}$$

$$\min_{x \in A} f(x) = \min\{f(x) : x \in A\}$$

$$\sup_{x \in A} f(x) = \sup\{f(x) : x \in A\}$$

$$\inf_{x \in A} f(x) = \inf\{f(x) : x \in A\}$$

16.6 Arguments d'extrema

16.6.1 Maximum et minimum

L'ensemble des éléments de A qui maximisent f sur A est noté :

$$\arg\max_{x \in A} f(x) = \left\{ \alpha \in A : f(\alpha) = \max_{x \in A} f(x) \right\}$$

Dans le cas où cet ensemble contient un unique élément, on le note :

$$\alpha = \arg\max_{x \in A} f(x)$$

L'ensemble des éléments de A qui minimisent f sur A est noté :

$$\arg\min_{x \in A} f(x) = \left\{ \beta \in A : f(\beta) = \min_{x \in A} f(x) \right\}$$

Dans le cas où cet ensemble contient un unique élément, on le note :

$$\beta = \arg\min_{x \in A} f(x)$$

16.6.2 Supremum et infimum

L'ensemble des éléments de Ω qui produisent une valeur égale au supremum des valeurs de f sur $A \subseteq \Omega$ est noté :

$$\arg_{\Omega}\sup_{x \in A} f(x) = \left\{ \alpha \in \Omega : f(\alpha) = \sup_{x \in A} f(x) \right\}$$

Dans le cas où cet ensemble contient un unique élément, on le note :

$$\alpha = \arg_{\Omega} \sup_{x \in A} f(x)$$

L'ensemble des éléments de Ω qui produisent une valeur égale à l'infimum des valeurs de f sur $A \subseteq \Omega$ est noté :

$$\arg_{\Omega} \inf_{x \in A} f(x) = \left\{ \beta \in \Omega : f(\beta) = \inf_{x \in A} f(x) \right\}$$

Dans le cas où cet ensemble contient un unique élément, on le note :

$$\beta = \arg_{\Omega} \inf_{x \in A} f(x)$$

16.7 Extrema locaux

On dit que f atteint un minimum local en $a \in A$ si il existe un voisinage U de x tel que :

$$f(a) \le f(x)$$

pour tout $x \in U$. A l'inverse, on dit que f atteint un maximum local en $a \in A$ si il existe un voisinage U de x tel que :

$$f(a) \ge f(x)$$

pour tout $x \in U$.

16.8 Ordre entre fonctions et extrema

Soit les fonctions $f, g : A \mapsto B$ telles que $f \le g$. Supposons que $\sigma \in \operatorname{major} g(A)$. On a $\sigma \ge g(x) \ge f(x)$ pour tout $x \in A$, d'où $\sigma \ge f(x)$ et $\sigma \in \operatorname{major} f(A)$. On en conclut que $\operatorname{major} g(A) \subseteq \operatorname{major} f(A)$. Si les minima existent, on a donc :

$$\min \operatorname{major} f(A) \le \min \operatorname{major} g(A)$$

c'est-à-dire :

$$\sup_{x \in A} f(x) = \sup f(A) \le \sup g(A) = \sup_{x \in A} g(x)$$

Supposons que $\lambda \in \operatorname{minor} f(A)$. On a $\lambda \le f(x) \le g(x)$ pour tout $x \in A$, d'où $\lambda \le g(x)$ et $\lambda \in \operatorname{minor} g(A)$. On en conclut que $\operatorname{minor} f(A) \subseteq \operatorname{minor} g(A)$. Si les maxima existent, on a donc :

$$\max \operatorname{minor} f(A) \le \max \operatorname{minor} g(A)$$

c'est-à-dire :

$$\inf_{x \in A} f(x) = \inf f(A) \le \inf g(A) = \inf_{x \in A} g(x)$$

Chapitre 17

Bijections

Dépendances

— Chapitre 16 : Fonctions et ordre

17.1 Inverse à gauche

Soit $f : A \mapsto B$. Si $l : B \mapsto A$ est une fonction telle que :

$$l \circ f = \mathrm{Id}$$

on dit que l est un inverse à gauche de f.

Unicité

Soit $y \in B$. Si f admet un inverse à gauche, on peut trouver au plus un seul $x \in A$ tel que $f(x) = y$. En effet, supposons que $x_1, x_2 \in A$ avec $f(x_1) = f(x_2) = y$. On a :

$$l(y) = (l \circ f)(x_1) = (l \circ f)(x_2)$$

mais comme $l \circ f = \mathrm{Id}$, et que $\mathrm{Id}(x) = x$ pour tout $x \in A$, on a :

$$l(y) = x_1 = x_2$$

ce qui prouve l'unicité de la solution.

17.2 Inverse à droite

Soit $f : A \mapsto B$. Si $r : B \mapsto A$ est une fonction telle que :

$$f \circ r = \mathrm{Id}$$

on dit que r est un inverse à droite de f.

Existence

Soit $y \in B$. Si f admet un inverse à droite r, on peut trouver au moins un $x \in A$ tel que $f(x) = y$. En effet, si l'on choisit $x = r(y)$, on a :

$$f(x) = (f \circ r)(y) = \mathrm{Id}(y) = y$$

ce qui prouve l'existence de la solution.

17.3 Fonction inverse

Soit $f : A \mapsto B$. Supposons que f admette à la fois un inverse à gauche l et un inverse à droite r. Ces deux inverses sont dès lors identiques :

$$l = l \circ \mathrm{Id} = l \circ f \circ r = \mathrm{Id} \circ r = r$$

Existence et unicité

Soit $y \in B$. On déduit des résultats précédents que l'on peut trouver un unique $x \in A$, donné par $x = r(y)$, tel que $f(x) = y$. Nous pouvons dès lors définir la fonction inverse, notée $f^{-1} : B \mapsto A$, par :

$$f^{-1}(y) = r(y)$$

pour tout $y \in B$. On a donc :

$$f^{-1} = r = l$$

et :

$$f^{-1} \circ f = f \circ f^{-1} = \mathrm{Id}$$

17.3.1 Bijection

Lorsque f admet un inverse f^{-1}, on dit que f est inversible ou que f est une bijection. On note :

$$\mathrm{Bij}(A, B) = \{f \in B^A : \exists f^{-1} \in A^B\}$$

l'ensemble des bijections de A vers B.

17.4 Relation

Lorsque f est inversible, l'image inverse de $y \in B$ se réduit à un ensemble contenant un unique $x \in A$. Soit R la relation associée à f. On a :

$$R^{-1}(y) = \{f^{-1}(y)\}$$

17.5 Inverse d'une composée

Soit $f \in \mathrm{Bij}(A, B)$ et $g \in \mathrm{Bij}(B, C)$. Nous allons essayer de trouver une expression de l'inverse h de la composée de ces deux fonctions. On part de la relation $h \circ g \circ f = \mathrm{Id}$ et on compose à droite par l'inverse de f puis l'inverse de g, ce qui nous donne :

$$h \circ g \circ f \circ f^{-1} \circ g^{-1} = \mathrm{Id} \circ f^{-1} \circ g^{-1} = f^{-1} \circ g^{-1}$$

Mais comme :

$$h \circ g \circ f \circ f^{-1} \circ g^{-1} = h \circ g \circ \mathrm{Id} \circ g^{-1} = h \circ g \circ g^{-1} = h \circ \mathrm{Id} = h$$

on a :

$$h = f^{-1} \circ g^{-1}$$

Partant de la relation $g \circ f \circ h = \mathrm{Id}$ et composant à gauche par l'inverse de g puis par l'inverse de f, on arrive au même résultat $h = f^{-1} \circ g^{-1}$. Donc :

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}$$

On généralise par récurrence à n fonctions :

$$(f_n \circ ... \circ f_2 \circ f_1)^{-1} = f_1^{-1} \circ f_2^{-1} \circ ... \circ f_n^{-1}$$

17.6 Puissances négatives

Si f est inversible, on peut également définir les puissances négatives par :

$$f^{-n} = \left(f^{-1}\right)^n = f^{-1} \circ ... \circ f^{-1}$$

pour tout $n \in \mathbb{N}$.

17.7 Inverse d'une puissance

En considérant le cas particulier de fonctions identiques ($f_1 = ... = f_n = f$), l'expression de l'inverse d'une composition de fonctions nous montre que :

$$\left(f^{-1}\right)^n = f^{-n} = (f^n)^{-1}$$

L'inverse de la puissance est identique à la puissance de l'inverse.

17.8 Idempotence

On dit que deux ensembles A et B sont idempotents si il existe une bijection inversible de A vers B :

$$\mathrm{Bij}(A, B) \neq \emptyset$$

Equivalence

Soit les ensembles A et B. Supposons qu'il existe une bijection $f \in \mathrm{Bij}(A, B)$. On peut alors définir une équivalence sur $A \cup B$ en disant que $x \equiv y$ si $x \in A$ avec $y = f(x) \in B$, ou si $x \in B$ avec $y = f^{-1}(x) \in A$. Dans les autres cas, x ne sera pas équivalent à y. Les classes d'quivalences seront donc de la forme :

$$\mathcal{E}(a) = \{a, f(a)\}$$

pour tout $a \in A$ et de la forme :

$$\mathcal{E}(b) = \{b, f^{-1}(b)\}$$

pour tout $b \in B$. Il y a donc une « équivalence » entre les éléments de A et les éléments de B. Dans ce cas, on confond souvent les éléments x de A avec les éléments de $\hat{x}$ de B, et on note abusivement $x = \hat{x}$.

Ensembles finis

Dans le cas d'ensembles finis, l'idempotence revient à dire que A et B ont le même nombre d'éléments.

17.9 Inverse des fonctions monotones

Théorème 17.9.1. *Soit la fonction $f : A \mapsto B$ avec $B = f(A)$. Si f est strictement croissante (ou décroissante), alors la fonction f est inversible.*

Démonstration 17.9.1. Choisissons $y \in B$ et considérons l'ensemble de solutions :

$$S(y) = \{x : f(x) = y\}$$

Comme $B = f(A)$, cet ensemble est non vide. Supposons $x_1, x_2 \in S(y)$ avec $x_1 < x_2$. On a alors, soit $f(x_1) < f(x_2)$ (si f est strictement croissante), soit $f(x_1) > f(x_2)$ (si f est strictement décroissante). Or, par définition de $S(y)$, on doit avoir $f(x_1) = f(x_2) = y$. Il ne peut donc y avoir d'éléments distincts dans $S(y)$, et cet ensemble est de la forme :

$$S(y) = \{g(y)\}$$

relation qui définit implicitement la fonction $g : B \mapsto A$. On peut trouver un unique $g(y)$ tel que $f(g(y)) = y$. La fonction f est donc inversible et :

$$f^{-1} = g$$

Chapitre 18

Ordre inclusif

Dépendances

- Chapitre 4 : Les collections
- Chapitre 12 : Les ordres
- Chapitre 14 : Les extrema

18.1 Ordre sur les ensembles

Soit l'ensemble Ω et l'ensemble de ses sous-ensembles :

$$\mathfrak{P}(\Omega) = \{A : A \subseteq \Omega\}$$

On dit que $A \in \mathfrak{P}(\Omega)$ est plus petit ou égal à $B \in \mathfrak{P}(\Omega)$ au sens de l'ordre inclusif si et seulement si :

$$A \subseteq B$$

On voit qu'il s'agit d'un ordre partiel.

18.1.1 Attention

Ne pas confondre l'ordre inclusif $\subseteq$ avec la comparaison d'ensembles $\leqslant$ qui est elle basée sur l'ordre entre les éléments des ensembles concernés.

18.2 Extrema inclusifs d'une collection

Soit l'ensemble de paramètres X et la collection paramétrée :

$$\mathcal{C} = \{A(x) \in \mathfrak{P}(\Omega) : x \in X\}$$

18.2.1 Majorant

Si on veut qu'un ensemble quelconque $M \subseteq \Omega$ soit un majorant inclusif de $\mathcal{C}$, il faut que $A(x) \subseteq M$ pour tout $x \in X$. Pour cela, il faut et il suffit que M contienne tous les $A(x)$. Autrement dit, il faut et il suffit que M contienne l'union des ensembles-éléments de la collection. On a donc :

$$\operatorname*{major}_{\subseteq} \mathcal{C} = \left\{ M \in \mathfrak{P}(\Omega) : \bigcup_{x \in X} A(x) \subseteq M \right\}$$

18.2.2 Supremum

Etudions l'ensemble :

$$S = \bigcup_{x \in X} A(x) \in \mathfrak{P}(\Omega)$$

Comme :

$$\bigcup_{x \in X} A(x) \subseteq S$$

on a :

$$S \in \operatorname*{major}_{\subseteq} \mathcal{C}$$

On voit aussi que $S \subseteq M$ pour tout $M \in \operatorname{major}_{\subseteq} \mathcal{C}$. On en déduit que :

$$S = \min_{\subseteq} \operatorname*{major}_{\subseteq} \mathcal{C} = \sup_{\subseteq} \mathcal{C}$$

18.2.3 Minorant

Si on veut qu'un ensemble quelconque $L \subseteq \Omega$ soit un minorant inclusif de $\mathcal{C}$, il faut que $L \subseteq A(x)$ pour tout $x \in X$. Pour cela, il faut et il suffit que L soit contenu dans tous les $A(x)$. Autrement dit, il faut et il suffit que L soit contenu dans l'intersection des ensembles-éléments de la collection. On a donc :

$$\operatorname*{minor}_{\subseteq} \mathcal{C} = \left\{ L \in \mathfrak{P}(\Omega) : L \subseteq \bigcap_{x \in X} A(x) \right\}$$

18.2.4 Infimum

Etudions l'ensemble :

$$I = \bigcap_{x \in X} A(x) \in \mathfrak{P}(\Omega)$$

Comme :

$$I \subseteq \bigcap_{x \in X} A(x)$$

on a :

$$I \in \operatorname*{minor}_{\subseteq} \mathcal{C}$$

On voit aussi que $L \subseteq I$ pour tout $L \in \operatorname{minor}_{\subseteq} \mathcal{C}$. On en déduit que :

$$I = \max_{\subseteq} \operatorname*{minor}_{\subseteq} \mathcal{C} = \inf_{\subseteq} \mathcal{C}$$

18.2.5 Conclusion

Le supremum inclusif d'une collection de sous-ensembles de Ω existe toujours dans $\mathfrak{P}(\Omega)$ et est égal à l'union de tous ces ensembles :

$$\sup_{\subseteq}\{A(x) \in \mathfrak{P}(\Omega) : x \in X\} = \bigcup_{x \in X} A(x)$$

L'infimum inclusif d'une collection de sous-ensembles de Ω existe toujours dans $\mathfrak{P}(\Omega)$ et est égal à l'intersection de tous ces ensembles :

$$\inf_{\subseteq}\{A(x) \in \mathfrak{P}(\Omega) : x \in X\} = \bigcap_{x \in X} A(x)$$

Chapitre 19

Treillis

Dépendances

— Chapitre 18 : L'ordre inclusif

19.1 Définition

Soit l'ensemble Ω et $A, B \subseteq \Omega$. Au sens inclusif, les supremum et infimum de l'ensemble $\{A, B\}$ existent toujours et :

$$\sup_{\subseteq}\{A, B\} = A \cup B$$

$$\inf_{\subseteq}\{A, B\} = A \cap B$$

Le concept de treillis est une généralisation de cette propriété. Soit un ensemble $\mathcal{T}$ sur lequel est défini l'ordre $\le$. On dit que le couple $(\mathcal{T}, \le)$ est un treillis si, pour tout éléments $a, b \in \mathcal{T}$, le supremum et l'infimum de l'ensemble :

$$\{a, b\}$$

existent :

$$\operatorname{minor}\{a, b\} \le \inf\{a, b\} \le \{a, b\} \le \sup\{a, b\} \le \operatorname{major}\{a, b\}$$

Dans la suite, nous considérons un treillis $(\mathcal{T}, \le)$.

19.2 Union généralisée

Soit $a, b \in \mathcal{T}$. Par analogie avec l'union ensembliste, on définit l'opération d'union généralisée $\sqcup$ par :

$$a \sqcup b = \sup\{a, b\}$$

19.3 Intersection généralisée

Soit $a, b \in \mathcal{T}$. Par analogie avec l'intersection ensembliste, on définit l'opération d'intersection généralisée $\sqcap$ par :

$$a \sqcap b = \inf\{a, b\}$$

19.4 Treillis dual

Le treillis dual de $(\mathcal{T}, \le)$ est noté $(\mathcal{T}, \le)^\star$ et défini par :

$$(\mathcal{T}, \le)^\star = (\mathcal{T}, \le^\star)$$

où $\le^\star$ est l'ordre dual de $\le$. On a bien entendu :

$$\sup_{\le^\star}\{a, b\} = \inf_{\le}\{a, b\}$$

et :

$$\inf_{\le^\star}\{a, b\} = \sup_{\le}\{a, b\}$$

Le treillis dual est donc également un treillis.

19.4.1 Union

Soit $a, b \in \mathcal{T}$. On définit l'opération $\sqcup^\star$ par :

$$a \sqcup^\star b = \sup_{\le^\star}\{a, b\} = \inf_{\le}\{a, b\} = a \sqcap b$$

19.4.2 Intersection

Soit $a, b \in \mathcal{T}$. On définit l'opération $\sqcap^\star$ par :

$$a \sqcap^\star b = \inf_{\le^\star}\{a, b\} = \sup_{\le}\{a, b\} = a \sqcup b$$

19.4.3 Primal

Par opposition au treillis dual $(\mathcal{T}, \le)^\star$, le treillis $(\mathcal{T}, \le)$ est appelé treillis primal.

19.5 Éléments nul et universel

19.5.1 Élément nul

Si $\mathcal{T}$ admet un minimum, on l'appelle élément nul de $\mathcal{T}$ et on le note :

$$0 = \min \mathcal{T}$$

Soit $a \in \mathcal{T}$. Si $x \in \operatorname{minor}\{0, a\}$, on a $x \le 0$. Comme on a également $0 \le x$, on en conclut que $x = 0$ et que :

$$\operatorname{minor}\{0, a\} = \{0\}$$

Donc :

$$\inf\{0, a\} = \max\{0\} = 0$$

On a aussi :

$$\{0, a\} \le a \le \operatorname{major}\{0, a\}$$

d'où l'on déduit :

$$\sup\{0, a\} = a$$

En termes d'opérations, ces résultats s'écrivent :

$$0 \sqcap a = 0$$

$$0 \sqcup a = a$$

19.5.2 Élément universel

Si $\mathcal{T}$ admet un maximum, on l'appelle élément universel de $\mathcal{T}$ et on le note :

$$1 = \max \mathcal{T}$$

Si $x \in \operatorname{major}\{1, a\}$, on a $x \ge 1$. Comme on a également $1 \ge x$, on en conclut que $x = 1$ et que :

$$\operatorname{major}\{1, a\} = \{1\}$$

Donc :

$$\sup\{1, a\} = \min\{1\} = 1$$

On a aussi :

$$\{1, a\} \ge a \ge \operatorname{minor}\{1, a\}$$

d'où l'on déduit :

$$\inf\{1, a\} = a$$

En termes d'opérations, ces résultats s'écrivent :

$$1 \sqcup a = 1$$

$$1 \sqcap a = a$$

19.5.3 Dualité

Sous réserve d'existence, on a :

$$0 = \min_{\le} \mathcal{T} = \max_{\le^\star} \mathcal{T}$$

L'élément universel du treillis dual est donc égal à l'élément nul du treillis primal :

$$1^\star = 0$$

Sous réserve d'existence, on a :

$$1 = \max_{\le} \mathcal{T} = \min_{\le^\star} \mathcal{T}$$

L'élément nul du treillis dual est donc égal à l'élément universel du treillis primal :

$$0^\star = 1$$

19.6 Complémentaire

On dit que $x^\ddagger \in \mathcal{T}$ est un complémentaire de $x \in \mathcal{T}$ si :

$$x \sqcup x^\ddagger = 1$$

$$x \sqcap x^\ddagger = 0$$

19.6.1 Dualité

En termes d'opérations du treillis dual, les conditions de complémentarité s'écrivent :

$$x \sqcap^\star x^\ddagger = 0^\star$$

$$x \sqcup^\star x^\ddagger = 1^\star$$

On en conclut que $x^\ddagger$ est également le complémentaire de x au sens du treillis dual.

19.7 Distributivité

On dit que $(\mathcal{T}, \le)$ est un treillis distributif si :

$$a \sqcup (b \sqcap c) = (a \sqcup b) \sqcap (a \sqcup c)$$

$$a \sqcap (b \sqcup c) = (a \sqcap b) \sqcup (a \sqcap c)$$

pour tout $a, b, c \in \mathcal{T}$.

19.8 Booléen

Un treillis $(\mathcal{T}, \le)$ est dit booléen si et seulement si :
- il comprend un élément nul et un élément universel
- il est distributif
- chaque élément de $\mathcal{T}$ admet un complémentaire

19.9 Idempotence

Soit $a \in \mathcal{T}$. On a :

$$\{a\} \le a \le \operatorname{major}\{a\}$$

donc :

$$a = \sup\{a\}$$

On en conclut que :

$$a \sqcup a = \sup\{a, a\} = \sup\{a\} = a$$

On a :

$$\operatorname{minor}\{a\} \le a \le \{a\}$$

donc :

$$a = \inf\{a\}$$

On en conclut que :

$$a \sqcap a = \inf\{a, a\} = \inf\{a\} = a$$

19.10 Sous-ensemble fini

19.10.1 Supremum

Nous allons voir que tout sous-ensemble fini d'un treillis admet un supremum. On sait déjà que c'est vrai pour des ensembles comportant un ou deux éléments. Supposons à présent que ce soit vrai pour les sous-ensembles de maximum $n-1$ éléments, où n est un naturel vérifiant $n \ge 2$. Soit :

$$A = \{a_1, a_2, ..., a_n\} \subseteq \mathcal{T}$$

Choisissons $i \in \{1, 2, ..., n\}$ et posons :

$$x = a_i$$

$$Z = A \setminus \{x\} = \{..., a_{i-1}, a_{i+1}, ...\}$$

L'ensemble Z comportant $n-1$ éléments, il admet un supremum :

$$\mu = \sup Z$$

Posons :

$$\sigma = \sup\{\mu, x\}$$

On a :

$$\sigma \ge \mu \ge Z$$

$$\sigma \ge \{x\}$$

donc :

$$\sigma \ge Z \cup \{x\} = A$$

et :

$$\sigma \in \operatorname{major} A$$

Choisissons :

$$\varkappa \in \operatorname{major} A$$

On a :

$$\varkappa \ge Z$$

$$\varkappa \ge \{x\}$$

La première inégalité nous dit que :

$$\varkappa \in \operatorname{major} Z$$

Le supremum étant le plus petit des majorants, on doit donc avoir :

$$\varkappa \ge \mu$$

Comme on a également :

$$\varkappa \ge x$$

on en conclut que :

$$\varkappa \in \operatorname{major}\{\mu, x\}$$

Le supremum étant le plus petit des majorants, on doit donc avoir :

$$\varkappa \ge \sigma$$

Nous venons de prouver que :

$$\sigma = \min \operatorname{major} A = \sup A$$

Tout sous-ensemble fini non vide $A \subseteq \mathcal{T}$ possède un supremum. Si A compte au moins deux éléments, on a :

$$\sup A = \sup \Big\{ \sup \big(A \setminus \{x\}\big), x \Big\}$$

19.10.2 Infimum

Nous allons voir que tout sous-ensemble fini d'un treillis admet un infimum. On sait déjà que c'est vrai pour des ensembles comportant un ou deux éléments. Supposons à présent que ce soit vrai pour les sous-ensembles de maximum $n - 1$ éléments, où n est un naturel vérifiant $n \ge 2$. Soit :

$$A = \{a_1, a_2, ..., a_n\} \subseteq \mathcal{T}$$

Choisissons $i \in \{1, 2, ..., n\}$ et posons :

$$x = a_i$$

$$Z = A \setminus \{x\} = \{..., a_{i-1}, a_{i+1}, ...\}$$

L'ensemble Z comportant $n - 1$ éléments, il admet un infimum :

$$\gamma = \inf Z$$

Posons :

$$\lambda = \inf\{\gamma, x\}$$

On a :

$$\lambda \le \gamma \le Z$$

$$\lambda \le \{x\}$$

donc :

$$\lambda \le Z \cup \{x\} = A$$

et :

$$\lambda \in \operatorname{minor} A$$

Choisissons :

$$\vartheta \in \operatorname{minor} A$$

On a :

$$\vartheta \le Z$$

$$\vartheta \le \{x\}$$

La première inégalité nous dit que :

$$\vartheta \in \operatorname{minor} Z$$

L'infimum étant le plus grand des minorants, on doit donc avoir :

$$\vartheta \le \gamma$$

Comme on a également :

$$\vartheta \le x$$

on en conclut que :

$$\vartheta \in \operatorname{minor}\{\gamma, x\}$$

L'infimum étant le plus grand des minorants, on doit donc avoir :

$$\vartheta \le \lambda$$

Nous venons de prouver que :

$$\lambda = \max \operatorname{minor} A = \inf A$$

Tout sous-ensemble fini non vide $A \subseteq \mathcal{T}$ possède un infimum. Si A compte au moins deux éléments, on a :

$$\inf A = \inf \Big\{ \inf \big(A \setminus \{x\}\big), x \Big\}$$

19.11 Commutativité

Soit $a, b \in \mathcal{T}$. Comme $\{a, b\} = \{b, a\}$, on a clairement :

$$a \sqcup b = b \sqcup a$$

et :

$$a \sqcap b = b \sqcap a$$

19.12 Associativité

19.12.1 Union

On a :

$$\sup \big\{a, \sup\{b, c\}\big\} = \sup\{a, b, c\} = \sup \big\{ \sup\{a, b\}, c\big\}$$

En terme d'opération, ce résultat implique l'associativité de l'union généralisée :

$$a \sqcup (b \sqcup c) = (a \sqcup b) \sqcup c$$

On définit donc :

$$a \sqcup b \sqcup c = a \sqcup (b \sqcup c) = (a \sqcup b) \sqcup c$$

Plus généralement, on a :

$$a_1 \sqcup a_2 \sqcup ... \sqcup a_n = \sup\{a_1, a_2, ..., a_n\}$$

pour tout $a_1, a_2, ..., a_n \in \mathcal{T}$.

19.12.2 Intersection

Le treillis dual $(\mathcal{T}, \leq^\star)$ étant également un treillis, on a la propriété :

$$a \sqcup^\star (b \sqcup^\star c) = (a \sqcup^\star b) \sqcup^\star c$$

pour tout $a, b, c \in \mathcal{T}$. Cette relation traduite en terme d'opérations du treillis primal nous donne l'associativité de l'intersection généralisée :

$$a \sqcap (b \sqcap c) = (a \sqcap b) \sqcap c$$

On définit donc :

$$a \sqcap b \sqcap c = a \sqcap (b \sqcap c) = (a \sqcap b) \sqcap c$$

Plus généralement, on a :

$$a_1 \sqcap a_2 \sqcap ... \sqcap a_n = \inf\{a_1, a_2, ..., a_n\}$$

pour tout $a_1, a_2, ..., a_n \in \mathcal{T}$.

19.13 Absorption

Soit $a, b \in \mathcal{T}$. Posons :

$$\sigma = \sup\{a, b\}$$

Comme $\sigma \geq a$, on a :

$$a \leq \{a, \sigma\}$$

Choisissons $\lambda \in \mathcal{T}$ tel que :

$$\lambda \leq \{a, \sigma\}$$

on a alors forcément :

$$\lambda \le a$$

d'où l'on conclut que :

$$\mathrm{minor}\{a, \sigma\} \le a \le \{a, \sigma\}$$

L'élément a est donc l'infimum de $\{a, \sigma\}$:

$$a = \inf\{a, \sigma\}$$

Par définition de σ, on a donc :

$$a = \inf\{a, \sup\{a, b\}\}$$

Exprimée en terme d'opérations, cette relation devient :

$$a = a \sqcap (a \sqcup b)$$

Cette même propriété étant valable pour le treillis dual, on a :

$$a = a \sqcap^\star (a \sqcup^\star b)$$

Équation qui peut être réexprimée en termes d'opérations du treillis primal, ce qui nous donne :

$$a = a \sqcup (a \sqcap b)$$

19.14 Treillis d'ensemble

Une collection $\mathcal{C}$ de sous-ensembles d'un ensemble de référence Ω :

$$\mathcal{C} \subseteq \mathfrak{P}(\Omega)$$

est un treillis d'ensembles si et seulement si :

$$\emptyset, \Omega \in \mathcal{C}$$

et :

$$A \cup B \in \mathcal{C}$$

$$A \cap B \in \mathcal{C}$$

pour tout $A, B \in \mathcal{C}$. Le couple $(\mathcal{C}, \subseteq)$ est un cas particulier de treillis.

19.14.1 Élément nul

Un treillis d'ensembles comporte toujours un élément nul :

$$\emptyset = \min_{\subseteq} \mathcal{C}$$

19.14.2 Élément universel

Un treillis d'ensembles comporte toujours un élément universel :

$$\Omega = \max_{\subseteq} \mathcal{C}$$

Troisième partie

Collections

Chapitre 20

Tribus

Dépendances

— Chapitre 4 : Les collections
— Chapitre 18 : L'ordre inclusif

20.1 Définition

Soit un ensemble Ω et une collection de sous-ensembles $\mathcal{T} \subseteq \mathfrak{P}(\Omega)$. On dit que $\mathcal{T}$ forme une tribu sur Ω si et seulement si les conditions suivantes sont remplies :

— $\emptyset \in \mathcal{T}$
— si $A \in \mathcal{T}$, on a aussi $\Omega \setminus A \in \mathcal{T}$
— l'union de toute suite discrète $A_1, A_2, ... \in \mathcal{T}$ appartient aussi à la collection :

$$\bigcup_i A_i \in \mathcal{T}$$

On note $\mathrm{Tribu}(\Omega)$ l'ensemble des tribus sur Ω.

20.2 Corollaire

On conclut directement de la définition que :

$$\Omega = \Omega \setminus \emptyset \in \mathcal{T}$$

20.3 Intersection

Soit les $A_i \in \mathcal{T}$ et leurs complémentaires :

$$B_i = \Omega \setminus A_i \in \mathcal{T}$$

On a aussi :

$$A_i = \Omega \setminus B_i$$

En prenant l'intersection de tous les A_i, on obtient :

$$\bigcap_i A_i = \Omega \setminus \left[\bigcup_i B_i \right] \in \mathcal{T}$$

On en déduit que toute intersection dénombrable d'éléments de $\mathcal{T}$ est également dans $\mathcal{T}$.

20.4 Différence

Soit $A, B \in \mathcal{T}$, on a

$$A \setminus B = A \cap (\Omega \setminus B) \in \mathcal{T}$$

20.5 Tribu engendrée

Les collections d'ensemble étant des ensembles d'ensembles, on peut également les comparer au moyen de l'inclusion. En ce sens, la tribu engendrée par la collection $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$ est la plus petite tribu contenant les éléments de $\mathcal{C}$. On la note :

$$\mathrm{Tribu}(\mathcal{C}, \Omega) = \inf_{\subseteq} \{\mathcal{T} \in \mathrm{Tribu}(\Omega) : \mathcal{C} \subseteq \mathcal{T}\}$$

Chapitre 21

Topologies

Dépendances

- Chapitre 4 : Les collections
- Chapitre 18 : L'ordre inclusif

21.1 Définition

Soit un ensemble Ω et une collection de sous-ensembles $\mathfrak{T} \subseteq \mathfrak{P}(\Omega)$. On dit que $\mathfrak{T}$ forme une topologie sur Ω si et seulement si les conditions suivantes sont remplies :

- $\emptyset, \Omega \in \mathfrak{T}$
- pour toute sous-collection $\mathcal{C} \subseteq \mathfrak{T}$, l'union des ensembles-éléments de $\mathcal{C}$ appartient également à la topologie :

$$\bigcup \mathcal{C} \in \mathfrak{T}$$

- si $A, B \in \mathfrak{T}$, leur intersection appartient également à la topologie :

$$A \cap B \in \mathfrak{T}$$

On note $\operatorname{Topo}(\Omega)$ l'ensemble des topologies sur Ω.

21.2 Espace topologique

Si $\mathfrak{T}$ est une topologie sur Ω, on dit que le couple $(\Omega, \mathfrak{T})$ forme un espace topologique.

21.3 Ouvert

On appelle « ouvert » tout ensemble $U \in \mathfrak{T}$.

21.4 Fermé

On appelle « fermé » tout complémentaire d'un ouvert. Si $F \subseteq \Omega$ est un fermé, on peut donc écrire :

$$F = \Omega \setminus U$$

pour un certain $U \in \mathfrak{T}$. On note la collection des ensembles fermés par :

$$\mathfrak{F} = \{\Omega \setminus U : U \in \mathfrak{T}\}$$

21.4.1 Intersection

Soit une collection d'ouverts $\mathcal{C} \subseteq \mathfrak{T}$ et la collection de fermés correspondant :

$$\mathcal{F} = \{\Omega \setminus U : U \in \mathcal{C}\}$$

L'intersection de tous ces fermés :

$$\bigcap \mathcal{F} = \Omega \setminus \bigcup \mathcal{C}$$

est le complémentaire de l'ensemble :

$$U = \bigcup \mathcal{C}$$

qui est un ouvert. On en conclut qu'une intersection de fermés est un fermé.

21.4.2 Union

Soit les ouverts $U, V \in \mathfrak{T}$ et les fermés correspondant :

$$F = \Omega \setminus U$$
$$G = \Omega \setminus V$$

On a :

$$F \cup G = \Omega \setminus (U \cap V)$$

L'ensemble $U \cap V$ étant un ouvert, on en conclut que l'union de deux fermés est un fermé.

21.5 Ouverts fermés

On a $\{\emptyset, \Omega\} \subseteq \mathfrak{T} \cap \mathfrak{F}$.

21.6 Voisinage

On dit que $V \subseteq \Omega$ est un voisinage de $x \in \Omega$ si il existe un ouvert $U \in \mathfrak{T}$ tel que $x \in U \subseteq V$.

21.7 Adhérence

L'adhérence, ou fermeture, d'un ensemble $A \subseteq \Omega$ est le plus petit (au sens inclusif) ensemble fermé F contenant A. On la note :

$$\operatorname{adh} A = \inf_{\subseteq} \{F \in \mathfrak{F} : A \subseteq F\} = \bigcap \{F \in \mathfrak{F} : A \subseteq F\}$$

21.8 Intérieur

L'intérieur d'un ensemble $A \subseteq \Omega$ est le plus grand ensemble ouvert contenu dans A. On le note :

$$\operatorname{int} A = \sup_{\subseteq} \{U \in \mathfrak{T} : U \subseteq A\} = \bigcup \{U \in \mathfrak{T} : U \subseteq A\}$$

21.9 Frontière

La frontière de A est la différence entre l'adhérence et l'intérieur :

$$\partial A = (\operatorname{adh} A) \setminus (\operatorname{int} A)$$

21.10 Propriétés de l'adhérence et de l'intérieur

21.10.1 Inclusions

On a bien entendu :

$$\operatorname{int} A \subseteq A \subseteq \operatorname{adh} A$$

21.10.2 Nature

Une union quelconque d'ouverts étant un ouvert, on a $\operatorname{int} A \in \mathfrak{T}$. Une intersection quelconque de fermés étant un fermé, on a $\operatorname{adh} A \in \mathfrak{F}$.

21.10.3 Complémentaire

Les fermés F contenant A correspondent aux complémentaires des ouverts $U = \Omega \setminus F$ inclus dans $\Omega \setminus A$. Comme $\bigcap F = \Omega \setminus \bigcup U$, on a :

$$\operatorname{adh} A = \Omega \setminus \operatorname{int}(\Omega \setminus A)$$

Les ouverts U inclus dans A correspondent aux complémentaires des fermés $F = \Omega \setminus U$ contenant $\Omega \setminus A$. La relation $\bigcup U = \Omega \setminus \bigcap F$ nous dit que :

$$\operatorname{int} A = \Omega \setminus \operatorname{adh}(\Omega \setminus A)$$

21.11 Hausdorff

On dit que le couple $(\Omega, \mathfrak{T})$ est un espace de Haussdorff si, pour tout $x, y \in \Omega$ vérifiant $x \neq y$, on peut trouver des ouverts $U, V \in \mathfrak{T}$ tels que :

$$x \in U, \ \ y \in V$$

et :

$$U \cap V = \emptyset$$

21.12 Topologie engendrée

Les collections d'ensemble étant des ensembles d'ensembles, on peut également les comparer au moyen de l'inclusion. En ce sens, la topologie engendrée par la collection $\mathcal{C} \subseteq \mathfrak{P}(\Omega)$ est la plus petite topologie contenant les ensembles-éléments de $\mathcal{C}$. On la note :

$$\operatorname{Topo}(\mathcal{C}, \Omega) = \inf_{\subseteq} \{\mathfrak{T} \in \operatorname{Topo}(\Omega) : \mathcal{C} \subseteq \mathfrak{T}\}$$

Quatrième partie

Nombres

Chapitre 22

Booléens

Dépendances

— Chapitre 9 : Les opérations
— Chapitre 10 : L'algèbre

22.1 Définition

L'ensemble de Boole se définit par :

$$\mathbb{B} = \{0, 1\}$$

On associe souvent à 0 le sens de « faux » et à 1 le sens de « vrai ».

22.2 La loi « et »

Soit $C_1, C_2 \in \mathbb{B}$. La condition « C_1 et C_2 » n'est vraie ($= 1$) que si $C_1 = 1$ et $C_2 = 1$. On définit donc l'opération « ET », notée $\cdot : \mathbb{B} \times \mathbb{B} \mapsto \mathbb{B}$, par :

$$\begin{aligned} 1 \cdot 1 &= 1 \\ 1 \cdot 0 &= 0 \\ 0 \cdot 1 &= 0 \\ 0 \cdot 0 &= 0 \end{aligned}$$

22.3 La loi « ou »

Soit $C_1, C_2 \in \mathbb{B}$. La condition « C_1 ou C_2 » est vraie ($= 1$) dès qu'au moins un des deux $C_i = 1$. On définit donc l'opération « OU », notée $\oplus : \mathbb{B} \times \mathbb{B} \mapsto \mathbb{B}$, par :

$$\begin{aligned} 1 \oplus 1 &= 1 \\ 1 \oplus 0 &= 1 \\ 0 \oplus 1 &= 1 \\ 0 \oplus 0 &= 0 \end{aligned}$$

22.4 Le contraire

Le contraire de vrai est simplement faux et inversément. On définit donc le contraire de $C \in \mathbb{B}$ noté $\neg C$, par :

$$\neg 1 = 0$$
$$\neg 0 = 1$$

Chapitre 23

Naturels

Dépendances

- Chapitre 3 : Les ensembles
- Chapitre 9 : Les opérations
- Chapitre 10 : L'algèbre
- Chapitre 12 : Les ordres
- Chapitre 14 : Les extrema

23.1 Introduction

Nous allons étudier l'ensemble des naturels :

$$\mathbb{N} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ...\}$$

23.2 Addition

L'addition usuelle sur $\mathbb{N}$, notée $+ : \mathbb{N} \times \mathbb{N} \mapsto \mathbb{N}$, est définie par :

$$\begin{aligned} m + 0 &= 0 \\ m + n^+ &= m^+ + n \end{aligned}$$

pour tout $m, n \in \mathbb{N}$.
En utilisant récursivement la définition, on arrive à :

$$\begin{aligned} m + 1 &= m + 0^+ = m^+ + 0 = m^+ \\ m + 2 &= m + 0^{++} = m^+ + 0^+ = m^{++} + 0 = m^{++} \\ &\vdots \\ m + n &= ... = m^{+...+} \end{aligned}$$

où $+...+$ désigne l'opération de succession appliquée n fois. L'addition de n revient donc à effectuer n opérations de succession.

23.2.1 Dualité

Posant $i = m$ et $j = n^+$, on voit clairement que :

$$i + j = i^+ + j^-$$

23.2.2 Neutre additif

Pour $m \in \mathbb{N}$ quelconque, on a :

$$0 + m = 0^{+...+}$$

où +...+ désigne l'opération de succession appliquée m fois. Mais comme $0^{+...+} = m$, on a :

$$0 + m = m$$

On a donc $0 + m = m + 0 = m$. On dit que 0 est neutre pour l'addition. On voit clairement d'après la définition que 0 est le seul neutre pour l'addition.

23.2.3 Commutativité

Soit $m, n \in \mathbb{N}$. Par définition de l'addition, on a :

$$m + n = m^{+...+} = 0^{+.....+}$$

où +.....+ réprésente m suivies de n opérations de successions. On a aussi :

$$n + m = n^{+...+} = 0^{+.....+}$$

où +.....+ réprésente n suivies de m opérations de successions. Les deux résultats étant identiques,on a :

$$m + n = n + m$$

On dit que l'addition sur $\mathbb{N}$ est commutative.

23.2.4 Associativité

Soit $m, n, p \in \mathbb{N}$. Développons l'addition $m + (p + n)$. On a :

$$m + (p + n) = m + p^{+...+}$$

où +...+ désigne l'opération de succession appliquée n fois. Il vient ensuite :

$$m + (p + n) = m^{+} + p^{+...+-}$$
$$\vdots$$
$$m + (p + n) = m^{+...+} + p^{+...+-...-}$$

où $+... + -...-$ désigne l'opération de succession appliquée n fois, suivie de l'opération de prédécession appliquée n fois également. Mais comme le prédécesseur du sucesseur est égal à lui-même par définition, on a $p^{+-} = p$. On constate en développant que la même propriété doit être vérifiée lorsque les opérations de succession et de prédécession sont appliquées un nombre quelconque mais identique de fois. On a donc :

$$p^{+...+-...-} = p$$

En tenant compte de ces résultats dans le développement de la somme ci-dessus, on arrive à :

$$m + (p + n) = m^{+...+} + p$$

Mais $m^{+...+}$ n'est rien d'autre que la somme $m + n$, et :

$$m + (p + n) = (m + n) + p$$

L'addition étant commutative, on peut inverser p et n pour obtenir :

$$m + (n + p) = (m + n) + p$$

On peut donc associer les termes d'une somme comme on le désire, le résultat restera identique. On dit que l'addition sur $\mathbb{N}$ est associative. On note aussi :

$$m + n + p = m + (n + p) = (m + n) + p$$

23.3 Ordre

Soit $m, n \in \mathbb{N}$. On dit que n est plus petit ou égal à m, et on le note :

$$n \le m$$

si on peut trouver un $p \in \mathbb{N}$ tel que :

$$m = n + p$$

En évaluant les additions $n + 1$ et $n^- + 1$, on obtient :

$$\begin{aligned} n^- + 1 &= n \\ n + 1 &= n^+ \end{aligned}$$

Le choix $p = 1$ nous montre alors que :

$$n^- \le n \le n^+$$

23.3.1 Plus grand ou égal

On note aussi :

$$m \ge n$$

pour signifier que $n \le m$.

23.3.2 Total

Soir $m, n \in \mathbb{N}$. Comme m est, directement ou indirectement, un successeur ou un prédécesseur de n, on a soit :

$$m \le n$$

soit :

$$n \le m$$

L'ordre usuel sur $\mathbb{N}$ est total.

23.3.3 Ordre strict

On note :

$$n < m$$

ou :

$$m > n$$

lorsque $n \le m$ et $m \ne n$.

23.3.4 Extrema

Comme tous les éléments de $\mathbb{N}$ sont supérieurs au égaux à 0, on a clairement :

$$\operatorname{minor} \mathbb{N} = \{0\}$$

et :

$$\inf \mathbb{N} = \min \mathbb{N} = 0$$

Par contre, si on suppose avoir trouvé $s \in \operatorname{major} \mathbb{N}$, on a $s \le s^+ \in \mathbb{N}$ ce qui contredit l'hypothèse de majorant. L'ensemble des majorants est donc vide, et le supremum n'existe pas dans $\mathbb{N}$. On note donc :

$$\sup \mathbb{N} = +\infty$$

23.4 Ordre et addition

Soit $a, b, c, d \in \mathbb{N}$. Supposons que :

$$a \le b$$

$$c \le d$$

On peut donc trouver $p, q \in \mathbb{N}$ tels que :

$$\begin{aligned} b &= a + p \\ d &= c + q \end{aligned}$$

En additionnant ces deux équations, on obtient :

$$b + d = a + p + c + q = (a + c) + (p + q)$$

Comme $p + q$ est également un naturel, on en conclut que :

$$a + c \le b + d$$

L'ordre est donc conservé lorsqu'on ajoute au moins autant au grand nombre qu'au petit. On parle d'invariance sous l'addition.

23.5 Positivité

Soit $n \in \mathbb{N}$. La neutralité de 0 pour l'addition nous permet d'écrire :

$$n = 0 + n$$

On en déduit que :

$$n \geq 0$$

pour tout $n \in \mathbb{N}$ et :

$$n \ > \ 0$$

lorsque $n \neq 0$.

23.6 Soustraction

Soit l'ensemble :

$$\Delta = \{(m, n) \in \mathbb{N}^2 : n \leq m\}$$

Choisissons $(m, n) \in \Delta$. On peut trouver $p \in \mathbb{N}$ tel que :

$$m = n + p$$

Comme :

$$n + p = p + n = p^{+...+} = m$$

on voit que :

$$p = m^{-...-}$$

où $-...-$ désigne n opérations de prédécessions. Le p ainsi défini est donc unique. On dit que p est la soustraction de m et n, et on le note :

$$p = m - n$$

23.6.1 $m \leq n$

Si $m \ < \ n$, calculer la soustraction :

$$p = m - n = m^{-...-} = 0^{+...+-...-}$$

reviendrait à effectuer plus d'opérations de prédécessions que de sucessions. Nous serions donc amenés à devoir évaluer le prédécesseur de l'élément racine 0, qui n'existe pas dans $\mathbb{N}$. Cette opération n'est par conséquent pas définie.

23.6.2 Neutralisation

La définition implique directement l'équivalence :

$$m = n = n + 0$$
$$\Longleftrightarrow$$
$$m - n = 0$$

23.6.3 Associativité

Soit $m, n, p \in \mathbb{N}$ avec $n + p \leq m$. On a trivialement :

$$(m - n) - p = m - (n + p)$$

On note aussi :

$$m - n - p = (m - n) - p = m - (n + p)$$

23.6.4 Commutativité

Soit $m, n, p \in \mathbb{N}$ avec $n + p \leq m$. On voit que :

$$m - n - p = m - (n + p) = m - (p + n) = m - p - n$$

23.6.5 Associativité mixte

Soit $m, n, p \in \mathbb{N}$ avec $p \leq n$. On a trivialement :

$$(m + n) - p = m + (n - p)$$

On note aussi :

$$m + n - p = (m + n) - p = m + (n - p)$$

23.6.6 Commutativité mixte

Soit $m, n, p \in \mathbb{N}$ avec $m, p \leq n$. On a trivialement :

$$\begin{aligned} m + n - p = (m + n) - p &= (n + m) - p \\ &= n + (m - p) \\ &= (m - p) + n \\ &= m - p + n \end{aligned}$$

23.7 Ordre et soustraction

Soit $a, b, c, d \in \mathbb{N}$. Supposons que :

$$a \leq b$$

$$c \geq d$$

On peut donc trouver $p, q \in \mathbb{N}$ tel que :

$$b = a + p$$

$$d + q = c$$

En soustrayant ces deux équations, on obtient :

$$b - d - q = a + p - c$$

$$b - d = a - c + p + q$$

Comme $p + q$ est aussi un naturel, on en conclut que :

$$a - c \le b - d$$

L'ordre est donc conservé lorsqu'on soustrait au moins autant au petit nombre qu'au grand. On parle d'invariance pour la soustraction.

23.8 Multiplication

Soit $m, n \in \mathbb{N}$. La multiplication usuelle, notée $\cdot : \mathbb{N} \times \mathbb{N} \mapsto \mathbb{N}$, est définie par :

$$m \cdot 0 = 0$$

$$m \cdot n^+ = (m \cdot n) + m$$

Posant $i = m$ et $j = n^+$, on en déduit que :

$$i \cdot 0 = 0$$

$$i \cdot j = (i \cdot j^-) + i$$

23.8.1 Priorité

Afin d'alléger les notations, nous convenons que la multiplication est toujours prioritaire sur l'addition. On pose donc :

$$i \cdot j + k = (i \cdot j) + k$$

pour tout $i, j, k \in \mathbb{N}$.

La définition peut alors se réécrire :

$$m \cdot 0 = 0$$

$$m \cdot n^+ = m \cdot n + m$$

23.8.2 Notation

Lorsque cela ne pose pas de problème d'ambiguité, on note aussi :

$$m\ n = m \cdot n$$

23.8.3 Absorption

Soit $m \in \mathbb{N}$. On déduit directement de la définition que :

$$\begin{aligned} 0 \cdot m &= 0 \cdot m^- + 0 \\ &\vdots \\ 0 \cdot m &= 0 + ... + 0 = 0 \end{aligned}$$

On a donc $0 \cdot m = m \cdot 0 = 0$. On dit que 0 est absorbant pour la multiplication.

23.8.4 Définition alternative

Soit $m \ge 1$. En utilisant récursivement la définition, on obtient :

$$\begin{aligned} m \cdot n &= m \cdot n^- + m \\ &\vdots \\ &= m \cdot 0 + m + ... + m \\ &= 0 + m + ... + m \\ &= m + ... + m \end{aligned}$$

où le membre de droite compte n terme « m ». Mais comme $m = m^- + 1$, on a :

$$m \cdot n = m^- + 1 + ... + m^- + 1$$

La commutativité nous permet de regrouper les n nombres 1 à la fin :

$$m \cdot n = m^- + ... + m^- + 1 + ... + 1$$

Il est clair que :

$$1 + ... + 1 = 0^{+...+} = n$$

On a dès lors :

$$m \cdot n = m^- + ... + m^- + n$$

Comme le membre de droite contient n termes m^-, on en conclut que :

$$m \cdot n = m^- \cdot n + n$$

qui est une version alternative de la définition.

23.8.5 Neutre multiplicatif

La définition implique que :

$$m \cdot 1 = m \cdot 0 + m = 0 + m = m$$

Mais en appliquant la définition alternative à $1 \cdot m$, il vient :

$$1 \cdot m = 0 \cdot m + m = m$$

On a donc $1 \cdot m = m \cdot 1 = m$. On dit que 1 est neutre pour la multiplication.

23.8.6 Commutativité

Soit $m \in \mathbb{N}$. On a vu que $m \cdot 0 = 0 \cdot m$. L'équation :

$$m \cdot n = n \cdot m$$

Est donc vérifiée pour $n = 0$.
Supposons à présent que :

$$n^- \cdot m = m \cdot n^-$$

En appliquant les deux alternatives de la définition, on a :

$$m \cdot n = m \cdot n^- + m$$

$$n \cdot m = n^- \cdot m + m$$

et donc :

$$m \cdot n = n \cdot m$$

Puisque cette égalité est vraie pour $n = 0 = 1^-$, on en déduit qu'elle est également vraie pour $n = 1 = 2^-$, pour $n = 2 = 3^-$, etc. Elle est donc vérifiée pour tout $n \in \mathbb{N}$: la multiplication est commutative.

23.8.7 Distributivité

Soit $m, n, p \in \mathbb{N}$. On a :

$$m \cdot (n + p) = m + ... + m$$

où le membre de droite compte $n + p$ termes « m ». Par associativité de l'addition, on peut regrouper les termes comme suit :

$$m \cdot (n + p) = (m + ... + m) + (m + ... + m)$$

où la première parenthèse compte n termes et la seconde p termes. On a donc :

$$m \cdot (n + p) = m \cdot n + m \cdot p$$

Par commutativité de la multiplication, on en déduit :

$$(n + p) \cdot m = n \cdot m + p \cdot m$$

On dit que la multiplication se distribue sur l'addition.

Sur la soustraction

Soit $m, n, p \in \mathbb{N}$ avec $p \le n$ et :

$$q = n - p$$

On a :

$$m \cdot (p + q) = m \cdot p + m \cdot q$$

Mais comme $p + q = n$, il vient :

$$m \cdot n = m \cdot p + m \cdot (n - p)$$

On en déduit que :

$$m \cdot (n - p) = m \cdot n - m \cdot p$$

Par commutativité de la multiplication, on a également :

$$(n - p) \cdot m = n \cdot m - p \cdot m$$

On dit que la multiplication se distribue sur la soustraction.

23.8.8 Unicité de l'absorbant

Nous allons voir que l'absorbant est unique. Soit $a_1, a_2 \in \mathbb{N}$ tels que :

$$a_1 \cdot n = a_2 \cdot n = 0$$

pour tout $n \in \mathbb{N}$. On en déduit que :

$$a_1 \cdot n - a_2 \cdot n = 0$$

En utilisant la distributivité, on obtient :

$$(a_1 - a_2) \cdot n = 0$$

Mais comme ce résultat doit être valable pour tout $n \in \mathbb{Z}$, il est valable pour $n = 1$ et on a :

$$(a_1 - a_2) \cdot 1 = a_1 - a_2 = 0$$

Autrement dit, $a_1 = a_2$. L'absorbant est donc unique.

23.8.9 Associativité

Soit $m, n, p \in \mathbb{N}$. On a :

$$m \cdot (n \cdot p) = m + ... + m$$

où le membre de droite compte $n \cdot p$ termes « m ». Par associativité de l'addition, on peut les regrouper en p parenthèses contenant chacune n termes :

$$m \cdot (n \cdot p) = (m + ... + m) + ... + (m + ... + m)$$

ce qui revient à :

$$m \cdot (n \cdot p) = (m \cdot n) + ... + (m \cdot n)$$

Mais comme il y a p parenthèses, on a :

$$m \cdot (n \cdot p) = (m \cdot n) \cdot p$$

La multiplication est donc associative. On note aussi :

$$m \cdot n \cdot p = m \cdot (n \cdot p) = (m \cdot n) \cdot p$$

23.8.10 Produit nul

Soit $m, n \in \mathbb{N}$ tels que :

$$m \cdot n = 0$$

Nous allons prouver qu'au moins un des deux facteurs doit être nul. Supposons que $m, n \neq 0$. On a alors :

$$m \cdot n = n + ... + n = 0$$

Mais comme $n \neq 0$, on a :

$$n > 0$$

et :

$$0 = n + ... + n \geq n > 0$$

ce qui est impossible. On a donc l'implication :

$$m \cdot n = 0 \quad \Rightarrow \quad m = 0 \quad \text{ou} \quad n = 0$$

23.9 Notation décimale

Soit le tuple :

$$(i_0, i_1, i_2, ...i_{n-1}, i_n) \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}^{n+1}$$

La notation décimale associée est définie par :

$$i_n i_{n-1} ... i_2 i_1 i_0 = i_0 + i_1 \cdot 10 + i_2 \cdot 10^2 + ... + i_{n-1} \cdot 10^{n-1} + i_n \cdot 10^n$$

Exemple :

$$7512 = 2 + 1 \cdot 10 + 5 \cdot 10^2 + 7 \cdot 10^3$$

23.10 Division entière

Soit $m, n \in \mathbb{N}$. On définit la division entière $\div : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ par :

$$m \div n = \sup\{k \in \mathbb{N} : k \cdot n \leq m\}$$

On dit que m est le numérateur, n le dénominateur et $m \div n$ le quotient de m par n.

23.10.1 Existence

Soit $m, n \in \mathbb{N}$ avec $n \neq 0$ et :

$$A_{mn} = \{k \in \mathbb{N} : k \cdot n \leq m\}$$

Pour tout naturel p vérifiant $p > m$, on a :

$$p \cdot n = p + ... + p > p > m$$

On en conclut que :

$$A_{mn} \subseteq \{0, 1, ..., m - 1, m\}$$

L'ensemble A_{mn} contient donc un nombre fini d'éléments. Comme l'ordre usuel sur $\mathbb{N}$ est total, on en conclut que A_{mn} admet un maximum identique au suprémum. La division entière de m par n est donc bien définie :

$$m \div n = \max A_{mn} = \sup A_{mn}$$

L'inclusion nous donne l'inégalité des maxima :

$$\max A_{mn} \leq \max\{0, 1, ..., m - 1, m\} = m$$

On en déduit que :

$$m \div n \leq m$$

23.10.2 Division par zéro

Soit $m \in \mathbb{N}$. Essayons d'évaluer la division $m \div 0$. Par absorption de 0, on voit que :

$$k \cdot 0 = 0 \le m$$

pour tout $k \in \mathbb{N}$. Par conséquent :

$$\{k \in \mathbb{N} : k \cdot 0 \le m\} = \mathbb{N}$$

Le supremum n'existe pas dans $\mathbb{N}$ et la division par zéro n'est par conséquent pas définie. On le note symboliquement :

$$m \div 0 = \sup \mathbb{N} = +\infty$$

23.10.3 Division par un

Soit $m \in \mathbb{N}$. On a :

$$m = m \cdot 1 \le m$$

et :

$$(m + 1) \cdot 1 = m + 1 \ > \ m$$

On en déduit que :

$$m \div 1 = m$$

23.10.4 Zéro divisé

Soit $n \in \mathbb{N}$ avec $n \ne 0$. On a :

$$\{k \in \mathbb{N} : k \cdot n \le 0\} = \{0\}$$

La division entière de 0 par n est donc nulle :

$$0 \div n = \sup\{0\} = 0$$

23.10.5 Modulo

Soit $m, n \in \mathbb{N}$. Par définition, on voit que :

$$(m \div n) \cdot n \le m$$

Il est donc licite de définir le naturel :

$$r = m - (m \div n) \cdot n$$

On dit que r est le modulo de m par rapport à n et on le note :

$$m \bmod n = m - (m \div n) \cdot n$$

23.10.6 Décomposition

Soit $m, n \in \mathbb{N}$. Par définition du modulo, a la décomposition :

$$m = (m \div n) \cdot n + m \bmod n$$

23.10.7 Reste

Soit $m, n \in \mathbb{N}$. On dit aussi que $r = m \mod n$ est le reste de la division entière de m par n.

23.10.8 Exacte

Soit $m, n \in \mathbb{N}$. Si $m \mod n = 0$, on dit que la division entière est exacte. Dans ce cas, la décomposition s'écrit simplement :

$$m = (m \div n) \cdot n$$

23.10.9 Bornes

Soit $m, n \in \mathbb{N}$ et $k = m \div n$. On sait déjà que :

$$m \mod n \ge 0$$

par la positivité des naturels. Supposons de plus que :

$$m \mod n = m - k \cdot n \ge n$$

on a alors :

$$m \ge n + k \cdot n = (k + 1) \cdot n$$

ce qui est contraire au caractère de supremum de k. Par conséquent :

$$0 \le m \mod n \le n - 1$$

23.10.10 Solution

Soit $m, n, k, r \in \mathbb{N}$ avec $n \neq 0$. Supposons que :

$$m = k \cdot n + r$$

Par positivité des naturels, on a $r \ge 0$ et :

$$k \cdot n \le m$$

d'où l'on conclut par définition du suprémum que $k \le m \div n$. Si on a également :

$$r \le n - 1$$

on a alors :

$$(k + 1) \cdot n = k \cdot n + n \ > \ k \cdot n + r = m$$

On en conclut que $k + 1 \ > \ m \div n$. Donc :

$$m \div n = k$$

et :

$$m \mod n = r$$

23.10.11 Division de deux produits

Soit $m, n, q, i, j, k \in \mathbb{N}$ avec $n, j \neq 0$ tels que :

$$m = q \cdot n$$

$$i = k \cdot j$$

On a alors les divisions exactes :

$$m \div n = q$$

$$i \div j = k$$

De plus :

$$m \cdot i = q \cdot k \cdot n \cdot j$$

ce qui signifie que :

$$(m \cdot i) \div (n \cdot j) = q \cdot k$$

On a donc :

$$(m \cdot i) \div (n \cdot j) = (m \div n) \cdot (i \div j)$$

23.10.12 Plus grand commun diviseur

Soit $m, n \in \mathbb{N}$ avec $n \neq 0$. On définit le plus grand commun diviseur de m et n par :

$$\mathrm{pgcd}(m, n) = \sup\{k \in \mathbb{N} : m \bmod k = n \bmod k = 0\}$$

Comme les divisions sont exactes, on a :

$$m = \big[m \div \mathrm{pgcd}(m, n)\big] \cdot \mathrm{pgcd}(m, n)$$

et :

$$n = \big[n \div \mathrm{pgcd}(m, n)\big] \cdot \mathrm{pgcd}(m, n)$$

23.10.13 Plus petit commun multiple

Soit $m, n \in \mathbb{N}$. On définit le plus petit commun multiple de m et n par :

$$\mathrm{ppcm}(m, n) = \inf\{k \in \mathbb{N} : k \bmod m = k \bmod n = 0\}$$

Comme les divisions sont exactes, on a :

$$\mathrm{ppcm}(m, n) = \big[\,\mathrm{ppcm}(m, n) \div m\big] \cdot m$$

et :

$$\mathrm{ppcm}(m, n) = \big[\,\mathrm{ppcm}(m, n) \div n\big] \cdot n$$

23.11 Puissance

Soit $m, n \in \mathbb{N}$ avec $n \neq 0$. Les puissances naturelles sont définies par :

$$\begin{aligned} m^0 &= 1 \\ m^n &= m \cdot m^{n-1} \end{aligned}$$

En appliquant récursivement la définition, on obtient :

$$\begin{aligned} m^n &= m \cdot m^{n-1} \\ &\vdots \\ &= m \cdot ... \cdot m \end{aligned}$$

où le membre de droite compte n facteurs « m ».

23.11.1 Priorité

Nous convenons des règles de priorité :

$$m + n^p = m + (n^p)$$

$$m \cdot n^p = m \cdot (n^p)$$

valables pour tout $m, n, p \in \mathbb{N}$.

23.11.2 Somme en exposant

Soit $m, n, p \in \mathbb{N}$. On a :

$$m^{n+p} = m \cdot ... \cdot m$$

où le membre de droite compte $n + p$ facteur m. En regroupant les n premiers facteurs dans une première parenthèse et les p facteurs restant dans la seconde, on a :

$$m^{n+p} = (m \cdot ... \cdot m) \cdot (m \cdot ... \cdot m)$$

qui n'est rien d'autre que :

$$m^{n+p} = m^n \cdot m^p$$

23.11.3 Puissance d'un produit

$$(m \cdot n)^p = m \cdot n \cdot ... \cdot m \cdot n$$

où le membre de droite compte p facteurs $m \cdot n$. La commutativité de la multiplication nous permet de les regrouper les m d'un coté et les n de l'autre :

$$(m \cdot n)^p = m \cdot ... \cdot m \cdot n \cdot ... \cdot n$$

c'est-à-dire :

$$(m \cdot n)^p = m^p \cdot n^p$$

23.11.4 Puissance d'une puissance

Soit $m, n, p \in \mathbb{N}$. On a :

$$\begin{aligned} (m^n)^p &= (m \cdot ... \cdot m)^p \\ &= m \cdot ... \cdot m \end{aligned}$$

où le membre de droite compte $n \cdot p$ facteurs. On en conclut que :

$$(m^n)^p = m^{n \cdot p}$$

23.12 Factorielle

Soit $n \in \mathbb{N}$, $n \neq 0$. On définit la factorielle de n par :

$$\begin{aligned} 0! &= 1 \\ n! &= n \cdot (n-1)! \end{aligned}$$

On a donc :

$$n! = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 1 = 1 \cdot 2 \cdot 3 \cdot ... \cdot n$$

23.12.1 Priorité

On convient que :

$$x \cdot y! = x \cdot (y!)$$

23.13 Monoïde

$(\mathbb{N}, +)$ et $(\mathbb{N}, \cdot)$ sont des monoïdes.

Chapitre 24

Entiers

Dépendances

— Chapitre 23 : Les nombres naturels

24.1 Principe général de l'extension

Nous allons à présent étendre progressivement les opérations vues sur les naturels. Le principe est de partir d'un ensemble X et de construire un ensemble dérivé Y (souvent Y sera X^2). Ensuite, nous définissons l'opération étendue sur Y de telle sorte qu'elle vérifie les nouvelles propriétés demandées en plus de celles déjà acquises sur X.

24.2 Soustraction de naturels

On aimerait bien étendre la soustraction à deux naturels $i, j \in \mathbb{N}$ quelconques. Malheureusement, nous avons vu que si $i \le j$, cette opération n'est pas définie. Pour contourner ce problème nous introduisons la notation différentielle :

$$i - j = (i, j)$$

où $(i, j) \in \mathbb{N}^2$ est appelé nombre entier. Il ne nous reste plus ensuite qu'à définir convenablement les autres opérations pour conserver les propriétés intéressantes qu'elles possèdent sur $\mathbb{N}$.

24.2.1 Notations

Soit $i \in \mathbb{N}$. On note aussi :

$$\begin{aligned} i &= i - 0 = (i, 0) \\ -i &= 0 - i = (0, i) \end{aligned}$$

En particulier :

$$\begin{aligned} 1 &= (1, 0) \\ 0 &= (0, 0) \\ -1 &= (0, 1) \end{aligned}$$

24.3 Addition

Soit $i, j, k, l \in \mathbb{N}$. Pour conserver l'associativité et la commutativité, on doit avoir :

$$(i - j) + (k - l) = (i + j) - k - l = (i + j) - (k + l)$$

Ceci nous incite à définir l'addition des entiers par :

$$(i, j) + (k, l) = (i + j, k + l)$$

24.3.1 Neutre additif

Soit $i, j \in \mathbb{N}$. On a :

$$(i - j) + (0 - 0) = (i + 0) - (j + 0) = i - j$$

Pour tout $n \in \mathbb{N}$, la soustraction native des naturels nous dit que :

$$n - n = 0$$

Afin de rester consistant, on impose que :

$$i - j = (i - j) + 0 = (i - j) + (n - n) = (i + n) - (j + n)$$

ou, en terme de couples :

$$(i, j) = (i + n, j + n)$$

Le neutre pour l'addition s'écrit donc :

$$0 = 0 - 0 = n - n = (0, 0) = (n, n)$$

24.3.2 Équivalence

On voit apparaître les familles :

$$D(i, j) = \{(i + n, j + n) : n \in \mathbb{N}\}$$

où chaque élément de $D(i, j)$ est équivalent à un autre. Nous noterons donc également :

$$i - j = D(i, j)$$

24.4 Définition

On définit l'ensemble des nombres entiers par :

$$\mathbb{Z} = \{i - j : i, j \in \mathbb{N}\}$$

où $i - j$ représente l'ensemble d'équivalence $D(i, j)$. En introduisant les symboles usuels, on a donc :

$$\mathbb{Z} = \{..., -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, ...\}$$

24.4.1 Forme canonique

Soit $i, j \in \mathbb{N}$.

— Si $i \geq j$, le naturel $i - j$ existe et nous pouvons toujours ramener (i, j) par équivalence à :

$$(i - j, 0) = (i - j, 0) + (j, j) = (i - j + j, j) = (i, j)$$

On dit alors que $(i, j) = i - j$ est un entier positif.

— Si $i \leq j$, le naturel $j - i$ existe et nous pouvons toujours ramener (i, j) par équivalence à :

$$(0, j - i) = (0, j - i) + (i, i) = (i, j - i + i) = (i, j)$$

On dit alors que $(i, j) = i - j$ est un entier négatif.

24.4.2 Signe

On définit l'ensemble des entiers positifs par :

$$\mathbb{Z}^+ = \{i = D(i, 0) : i \in \mathbb{N}\}$$

ainsi que l'ensemble des entiers négatifs :

$$\mathbb{Z}^- = \{-i = D(0, i) : i \in \mathbb{N}\}$$

L'ensemble des entiers est bien entendu l'union des deux : $\mathbb{Z} = \mathbb{Z}^+ \cup \mathbb{Z}^-$.

24.4.3 Inclusion

Nous pouvons associer à tout naturel i un entier équivalent $i - 0 = (i, 0)$. Pour cette raison, nous dirons que tout naturel est également un entier, et nous noterons : $\mathbb{N} \subseteq \mathbb{Z}$.

24.4.4 Entiers positifs et naturels

Il existe une bijection $f : \mathbb{N} \mapsto \mathbb{Z}^+$ définie par :

$$f(n) = (n, 0) = n - 0 = n$$

pour tout $n \in \mathbb{N}$. On assimile donc les deux ensembles en écrivant $\mathbb{N} = \mathbb{Z}^+$.

24.5 Opposé

Soit $i, j \in \mathbb{N}$. On déduit de la définition de l'addition que :

$$(i - j) + (j - i) = (i + j) - (j + i) = (i + j) - (i + j) = 0 - 0 = 0$$

ou, en terme de couples :

$$(i, j) + (j, i) = (i + j, i + j) = (0, 0) = 0$$

On dit que $j - i = (j, i)$ est l'opposé de $i - j = (i, j)$, et inversément. On le note :

$$-(i - j) = j - i$$

ou :

$$-(i, j) = (j, i)$$

24.5.1 De l'opposé

On a clairement :

$$-(-(i-j)) = -(j-i) = i-j$$

ou :

$$-(-(i,j)) = -(j,i) = (i,j)$$

L'opposé de l'opposé est l'entier lui-même.

24.5.2 D'une somme

Soit $u, v \in \mathbb{Z}$ et $i, j, k, l \in \mathbb{N}$ tels que :

$$u = i - j$$

$$v = k - l$$

On a :

$$(u+v) + ((-u) + (-v)) = ((i-j) + (k-l)) + ((j-i) + (l-k))$$

ce qui nous donne :

$$(u+v) + ((-u) + (-v)) = (i+k+j+l) - (j+l+i+k) = 0$$

On en conclut que la somme des opposés est égale à l'opposé de la somme :

$$(-u) + (-v) = -(u+v)$$

24.5.3 Notation

On note aussi :

$$-u + v = (-u) + v$$

pour tout $u, v \in \mathbb{Z}$.

24.5.4 Moins un

Un cas particulier important :

$$-(1-0) = 0 - 1 = -1$$

et :

$$-(-1) = 1$$

24.6 Soustraction d'entiers

Soit $i, j, k, l \in \mathbb{N}$. Comme :

$$i + (-j) = (i, 0) + (0, j) = (i, j) = i - j$$

On étend la soustraction à l'ensemble des entiers par :

$$(i, j) - (k, l) = (i, j) + (-(k, l)) = (i, j) + (l, k)$$

ce qu'on peut réécrire par :

$$u - v = u + (-v)$$

pour tout $u, v \in \mathbb{Z}$.

24.6.1 Propriétés

On a clairement :

$$(-i) - j = (-i) + (-j) = -(i + j)$$

$$(-i) - (-j) = (-i) + j = j + (-i) = j - i$$

24.6.2 Notation

On note aussi :

$$-u - v = (-u) - v$$

pour tout $u, v \in \mathbb{Z}$.

24.7 Ordre

Soit $u, v \in \mathbb{Z}$ et $i, j, k, l \in \mathbb{N}$ tels que :

$$u = i - j$$
$$v = k - l$$

Si nous voulons conserver la propriété de conservation de l'ordre sous l'addition, l'inégalité :

$$u = i - j \le k - l = v$$

doit être équivalente à :

$$(i - j) + j + l \le (k - l) + j + l$$

qui nous donne :

$$i + l \le k + j$$

Nous définissons l'ordre sur les entiers en affirmant que :

$$i - j \le k - l$$

si et seulement si :

$$i + l \le j + k$$

24.7.1 Plus grand ou égal

Soit $x, y \in \mathbb{Z}$. On note aussi :

$$y \ge x$$

pour signifier que $x \le y$.

24.7.2 Ordre strict

Soit $x, y \in \mathbb{Z}$. On note :

$$x < y$$

$$y > x$$

lorsque $x \le y$ et $x \ne y$.

24.7.3 Conservation

Soit $u, v, w, z \in \mathbb{Z}$ et $i, j, k, l, m, n, r, s \in \mathbb{N}$ tels que :

$$u = i - j$$
$$v = k - l$$
$$w = m - n$$
$$z = r - s$$

On a :

$$u + w = (i + m) - (j + n)$$
$$v + z = (k + r) - (l + s)$$

Supposons que :

$$u \le v$$
$$w \le z$$

On a :

$$i - j \le k - l$$
$$m - n \le r - s$$

ou :

$$i + l \le j + k$$
$$m + s \le n + r$$

L'ordre des naturels étant conservé sous l'addition, on a :

$$i + l + m + s \le j + k + n + r$$

et :

$$(i + m) - (j + n) \le (k + r) - (l + s)$$

c'est-à-dire :

$$u + w \le v + z$$

L'ordre des entiers est conservé par l'addition.

24.7.4 Opposé

Soit $u, v \in \mathbb{Z}$ tels que :

$$u \le v$$

Comme :

$$-(u + v) = -(u + v)$$

on a :

$$-(u + v) \le -(u + v)$$

et :

$$u + (-(u + v)) \le v + (-(u + v))$$

En développant, on a :

$$-v = u - u - v \le v - u - v = -u$$

c'est-à-dire :

$$-u \ge -v$$

L'ordre sur les opposés est l'inverse de l'ordre original.

Signe

Soit $z \in \mathbb{Z}$. Si $z \ge 0$, on a :

$$-z \le 0$$

Inversément, si $z \le 0$, on a :

$$-z \ge 0$$

24.7.5 Positifs et négatifs

Soit $u \in \mathbb{Z}^+$ et $i \in \mathbb{N}$ tel que :

$$u = i - 0$$

On a :

$$i + 0 \geq 0 + 0$$

et :

$$i - 0 \geq 0 - 0 = 0$$

On en déduit que $u \geq 0$. Réciproquement, soit $u \in \mathbb{Z}$ et $i, j \in \mathbb{N}$ tels que :

$$u = i - j$$

Si $u \geq 0$, on a :

$$i - j \geq 0$$

et :

$$i \geq j$$

On peut donc mettre u sous la forme canonique :

$$u = (i - j, 0) \in \mathbb{Z}^+$$

On a donc :

$$\mathbb{Z}^+ = \{u \in \mathbb{Z} : u \geq 0\}$$

On montre aussi que :

$$\mathbb{Z}^- = \{u \in \mathbb{Z} : u \leq 0\}$$

Par analogie avec les naturels, on dit que les entiers positifs sont les successeurs de 0 et les entiers négatifs les prédécesseurs de 0.

24.8 Signe

La fonction signe est définie par :

$$\operatorname{sign}(z) = \begin{cases} 1 & \text{si } z \geq 0 \\ -1 & \text{si } z < 0 \end{cases}$$

pour tout $z \in \mathbb{Z}$.

24.9 Valeur absolue

Soit $z \in \mathbb{Z}$. On définit la valeur absolue de z par :

$$|z| = \max\{z, -z\}$$

On a donc :

$$|z| = \begin{cases} z & \text{si } z \geq 0 \\ -z & \text{si } z < 0 \end{cases}$$

Comme le nombre positif l'emporte toujours sur le négatif dans le maximum, on a :

$$|z| \geq 0$$

On vérifie que :

$$|x + y| \leq |x| + |y|$$

pour tout $x, y \in \mathbb{Z}$.

24.10 Multiplication

Soit $u \in \mathbb{Z}$. Le naturel 0 est absorbant pour la multiplication. On souhaite conserver cette propriété pour l'entier $0 = 0 - 0$:

$$u \cdot 0 = 0 \cdot u = 0$$

$$u \cdot 1 = 1 \cdot u = u$$

Soit $u, v \in \mathbb{Z}$. Si nous voulons conserver la distributivité, il faut que :

$$u \cdot v + (-u) \cdot v = (u + (-u)) \cdot v = 0 \cdot v = 0$$

On en conclut que :

$$(-u) \cdot v = -(u \cdot v)$$

On a aussi :

$$u \cdot v + u \cdot (-v) = u \cdot (v + (-v)) = u \cdot 0 = 0$$

et :

$$u \cdot (-v) = -(u \cdot v)$$

Enfin :

$$(-u) \cdot (-v) = -(u \cdot (-v)) = -(-(u \cdot v)) = u \cdot v$$

24.10.1 Définition

Soit $u, v \in \mathbb{Z}$ et $i, j, k, l \in \mathbb{N}$ tels que :

$$u = i - j$$

$$v = k - l$$

On a :

$$u \cdot v = (i - j) \cdot (k - l) = (i + (-j)) \cdot (k + (-l))$$

La distributivité nous donne :

$$u \cdot v = i \cdot (k + (-l)) + (-j) \cdot (k + (-l))$$

En l'utilisant une nouvelle fois, on arrive à :

$$u \cdot v = i \cdot k + i \cdot (-l) + (-j) \cdot k + (-j) \cdot (-l)$$

ou :

$$u \cdot v = i \cdot k - i \cdot l - j \cdot k + j \cdot l = (i \cdot k + j \cdot l) - (i \cdot l + j \cdot k)$$

On définit donc la multiplication d'entiers par :

$$(i, j) \cdot (k, l) = (i \cdot k + j \cdot l, i \cdot l + j \cdot k)$$

24.10.2 Entiers positifs

Soit $u, v \in \mathbb{Z}$ tels que $u, v \geq 0$. On peut trouver des naturels i, j tels que :

$$u = i - 0$$

$$v = j - 0$$

On a :

$$u \cdot v = (i - 0) \cdot (j - 0) = (i \cdot j + 0 \cdot 0) - (i \cdot 0 + 0 \cdot j) = i \cdot j - 0 = i \cdot j$$

La multiplication d'entiers positifs correspond à celle des naturels.

24.10.3 Lien avec l'addition

Soit $u, v \in \mathbb{Z}$ avec $v \geq 0$ et $i, j, k \in \mathbb{N}$ tels que :

$$u = i - j$$

$$v = k - 0$$

On a :

$$u \cdot v = (i - j) \cdot (k - 0) = (i \cdot k + j \cdot 0) - (j \cdot k + i \cdot 0) = i \cdot k - j \cdot k$$

En additionnant k fois le même terme u, on obtient :

$$u + ... + u = i \cdot k - j \cdot k = u \cdot v$$

24.10.4 Commutativité

La commutativité de la multiplication sur les entiers découle de celle sur les naturels :

$$\begin{aligned}(k-l)\cdot(i-j) &= (k\cdot i + l\cdot j) - (k\cdot j + l\cdot i) \\ &= (i\cdot k + j\cdot l) - (i\cdot l + j\cdot k) \\ &= (i-j)\cdot(k-l)\end{aligned}$$

24.10.5 Associativité

Soit $u, v, w \in \mathbb{Z}$. Si $u, v, w \geq 0$, on peut les associer aux naturels $i, j, k \in \mathbb{N}$ par :

$$u = i - 0$$

$$v = j - 0$$

$$w = k - 0$$

On a alors :

$$u \cdot v = i \cdot j$$

et :

$$v \cdot w = j \cdot k$$

On a donc :

$$(u \cdot v) \cdot w = (i \cdot j) \cdot (k - 0) = i \cdot j \cdot k - 0 = i \cdot j \cdot k$$

et :

$$u \cdot (v \cdot w) = (i - 0) \cdot (j \cdot k) = i \cdot j \cdot k - 0 = i \cdot j \cdot k$$

On en conclut que :

$$(u \cdot v) \cdot w = u \cdot (v \cdot w)$$

Si un ou plusieurs entiers sont négatifs, soit $-u, -v, -w$, on a :

$$((-u) \cdot v) \cdot w = (-(u \cdot v)) \cdot w = -((u \cdot v) \cdot w)$$

et :

$$((-u) \cdot v) \cdot w = -(u \cdot (v \cdot w)) = (-u) \cdot (v \cdot w)$$

ou :

$$(u \cdot v) \cdot (-w) = -((u \cdot v) \cdot w)$$

et :

$$(u \cdot v) \cdot w = -(u \cdot (v \cdot w)) = u \cdot (-(v \cdot w)) = u \cdot (v \cdot (-w))$$

Les autres cas sont semblables, on a donc :

$$u \cdot v \cdot w = (u \cdot v) \cdot w = u \cdot (v \cdot w)$$

pour tout $u, v, w \in \mathbb{Z}$.

24.10.6 Neutre

On a :

$$(i-j)\cdot 1 = (i-j)\cdot(1-0) = (i\cdot 1 + j\cdot 0) - (i\cdot 0 + j\cdot 1) = i-j$$

et :

$$1\cdot(i-j) = (i-j)\cdot 1 = i-j$$

L'entier $1 = 1-0$ est le neutre pour la multiplication.

24.10.7 Notation

On note aussi :

$$-u\cdot v = -(u\cdot v)$$

pour tout $u, v \in \mathbb{Z}$.

24.10.8 Moins un

On a :

$$(-1)\cdot(i,j) = -(1\cdot(i,j)) = -(i,j)$$

Un cas particulier important :

$$(-1)\cdot(-1) = -(-1) = 1$$

24.11 Ordre et multiplication

24.11.1 Conservation simple

Soit $x, y, z \in \mathbb{Z}$ avec $z \ge 0$. Supposons que :

$$x \le y$$

On a alors :

$$x\cdot z = x + ... + x \le y + ... + y = y\cdot z$$

L'ordre est conservé lorsqu'on multiplie les deux membres de l'inégalité par un entier positif. Par contre :

$$x\cdot(-z) = -x\cdot z \ge -y\cdot z = y\cdot(-z)$$

L'ordre est inversé lorsqu'on multiplie les deux membres de l'inégalité par un entier négatif.

24.11.2 Conservation double

Soit $x, y, u, v \in \mathbb{Z}$ vérifiant :

$$x \le y$$
$$u \le v$$

Si :

$$x, y, u, v \ge 0$$

on a :

$$x \cdot u \le x \cdot v \le y \cdot v$$

24.12 Notation décimale

Soit le tuple :

$$(i_0, i_1, i_2, ...i_{n-1}, i_n) \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}^{n+1}$$

La notation décimale associée est définie par :

$$i_n i_{n-1} ... i_2 i_1 i_0 = i_0 + i_1 \cdot 10 + i_2 \cdot 10^2 + ... + i_{n-1} \cdot 10^{n-1} + i_n \cdot 10^n$$

pour les entiers positifs et :

$$-i_n i_{n-1} ... i_2 i_1 i_0 = -\left(i_0 + i_1 \cdot 10 + i_2 \cdot 10^2 + ... + i_{n-1} \cdot 10^{n-1} + i_n \cdot 10^n\right)$$

pour les entiers négatifs. Exemple :

$$-7512 = -\left(2 + 1 \cdot 10 + 5 \cdot 10^2 + 7 \cdot 10^3\right)$$

24.13 Division entière et modulo

Soit $m, n \in \mathbb{N}$, avec $n \neq 0$. Soit :

$$\begin{aligned} k &= m \div n \\ r &= m \bmod n \end{aligned}$$

On veut étendre la division entière à $\mathbb{Z}$ en conservant la décomposition :

$$m = k \cdot n + r$$

On note que :

$$m = (-k) \cdot (-n) + r$$

On en déduit l'extension :

$$m \div (-n) = -k = -(m \div n)$$

et :

$$m \bmod (-n) = r = m \bmod n$$

On note que :

$$-m = k \cdot (-n) - r$$

On en déduit l'extension :

$$(-m) \div (-n) = k = m \div n$$

et :

$$(-m) \bmod (-n) = -r = -(m \bmod n)$$

On note que :

$$-m = (-k) \cdot n - r$$

On en déduit l'extension :

$$(-m) \div n = -k = -(m \div n)$$

et :

$$(-m) \bmod n = -r = -(m \bmod n)$$

24.14 Puissance

Soit $z \in \mathbb{Z}$ et $n \in \mathbb{N}$. On définit les puissances par :

$$\begin{aligned} z^0 &= 1 \\ z^n &= z \cdot z^{n-1} \end{aligned}$$

On a donc :

$$z^n = z \cdot ... \cdot z$$

z^n est égal au produit de n facteurs z.

24.15 Anneau

$(\mathbb{Z}, +, \cdot)$ est un anneau.

24.16 Propriétés

La majeure partie des résultats démontrés sur $\mathbb{N}$ reste valable sur $\mathbb{Z}$. On le vérifie aisément en utilisant les définitions étendues. Exceptions :

— La positivité des naturels (voir section 23.5), qui est remplacée par la positivité de la valeur absolue

Chapitre 25

Rationnels

Dépendances

— Chapitre 23 : Les nombres naturels
— Chapitre 24 : Les nombres entiers

25.1 Fractions

Soit $a, b \in \mathbb{Z}$. On aimerait bien généraliser la division entière de sorte qu'elle soit tout le temps exacte. Comme ce ne sera en général pas le cas, on contourne le problème en introduisant la notation fractionnaire :

$$\frac{a}{b} = (a, b)$$

où $(a, b) \in \mathbb{Z}^2$ est appelé nombre rationnel. Le nombre a est appelé le numérateur et b le dénominateur. La division entière :

$$a \div b$$

n'étant pas définie lorsque $b = 0$, nous imposerons comme seule restriction que b soit non nul.

Notations

On note aussi les fractions par :

$$a/b = \frac{a}{b}$$

Lorsque $b = 1$, on note plus simplement :

$$a = \frac{a}{1} = (a, 1)$$

et donc en particulier :

$$\begin{array}{rcl} 1 & = & 1/1 \\ 0 & = & 0/1 \\ -1 & = & -1/1 \end{array}$$

Attention

Ne pas confondre les couples $(., .)$ fractionnaires de ce chapitre avec les couples différentiels du chapitre 24 traitant des entiers.

25.2 Produit

La notation fractionnaire étant destinée à représenter une division exacte, nous devons conserver les propriétés de la division entière dans le cas où elle est également exacte. Comme on a :

$$(m \div n) \cdot (i \div j) = (m \cdot i) \div (n \cdot j)$$

pour tout $m, n, i, j \in \mathbb{Z}$ tels que :

$$m \bmod n = i \bmod j = 0$$

on définit le produit entre rationnels par :

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d}$$

ou :

$$(a, b) \cdot (c, d) = (a \cdot c, b \cdot d)$$

pour tout $a, b, c, d \in \mathbb{Z}$ vérifiant $b, d \neq 0$.

25.2.1 Commutativité

Soit $a, b, c, d \in \mathbb{Z}$ vérifiant $b, d \neq 0$. On a :

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d} = \frac{c \cdot a}{d \cdot b} = \frac{c}{d} \cdot \frac{a}{b}$$

La multiplication entre rationnels est commutative.

25.2.2 Associativité

Soit $a, b, c, d, e, f \in \mathbb{Z}$ vérifiant $b, d, f \neq 0$. On a :

$$\frac{a}{b} \cdot \left[\frac{c}{d} \cdot \frac{e}{f}\right] = \frac{a}{b} \cdot \frac{c \cdot e}{d \cdot f} = \frac{a \cdot c \cdot e}{b \cdot d \cdot f}$$

$$\left[\frac{a}{b} \cdot \frac{c}{d}\right] \cdot \frac{e}{f} = \frac{a \cdot c}{b \cdot d} \cdot \frac{e}{f} = \frac{a \cdot c \cdot e}{b \cdot d \cdot f}$$

On en conclut que :

$$\frac{a}{b} \cdot \left[\frac{c}{d} \cdot \frac{e}{f}\right] = \left[\frac{a}{b} \cdot \frac{c}{d}\right] \cdot \frac{e}{f}$$

La multiplication entre rationnels est associative :

$$\frac{a}{b} \cdot \frac{c}{d} \cdot \frac{e}{f} = \frac{a}{b} \cdot \left[\frac{c}{d} \cdot \frac{e}{f}\right] = \left[\frac{a}{b} \cdot \frac{c}{d}\right] \cdot \frac{e}{f}$$

25.2.3 Neutre multiplicatif

Soit $n \in \mathbb{Z}$ avec $n \neq 0$. On a :

$$\frac{a}{b} \cdot 1 = \frac{a}{b} \cdot \frac{1}{1} = \frac{a \cdot 1}{b \cdot 1} = \frac{a}{b}$$

Le rationnel $1 = 1/1$ est le neutre pour la multiplication. Comme :

$$n \div n = 1$$

on impose que :

$$\frac{n}{n} = \frac{1}{1} = 1$$

Cette contrainte nous donne :

$$\frac{a \cdot n}{b \cdot n} = \frac{a}{b} \cdot \frac{n}{n} = \frac{a}{b} \cdot 1 = \frac{a}{b}$$

pour tout $a, b \in \mathbb{Z}$ avec $b \neq 0$.

25.2.4 Équivalence

On voit apparaître les familles :

$$F(a, b) = \left\{ \frac{a \cdot n}{b \cdot n} : n \in \mathbb{Z},\ n \neq 0 \right\}$$

où chaque élément de $F(a, b)$ est équivalent à un autre. Nous noterons donc également :

$$\frac{a}{b} = F(a, b)$$

25.3 Définition

On définit l'ensemble des nombres rationnels par :

$$\mathbb{Q} = \left\{ \frac{a}{b} : a, b \in \mathbb{Z},\ b \neq 0 \right\}$$

25.3.1 Naturels

Se rappelant que tout $a \in \mathbb{Z}$ peut s'exprimer comme la soustraction de deux naturels i, j :

$$a = i - j$$

on arrive à la formulation équivalente :

$$\mathbb{Q} = \left\{ \frac{i - j}{k - l} : i, j, k, l \in \mathbb{N},\ k \neq l \right\}$$

25.3.2 Inclusion

Nous pouvons associer à tout entier a un rationnel équivalent $(a, 1) = a/1$. Pour cette raison, nous dirons que tout entier est également un rationnel, et nous noterons : $\mathbb{Z} \subseteq \mathbb{Q}$.

25.4 Inverse

Soit $a, b \in \mathbb{Z}$ avec $a, b \neq 0$. On a :

$$\frac{a}{b} \cdot \frac{b}{a} = \frac{a \cdot b}{b \cdot a} = 1$$

On dit que b/a est l'inverse de a/b et réciproquement. On le note :

$$\left(\frac{a}{b} \right)^{-1} = \frac{b}{a}$$

25.4.1 De l'inverse

On a clairement :

$$\left[\left(\frac{a}{b}\right)^{-1}\right]^{-1} = \left[\frac{b}{a}\right]^{-1} = \frac{a}{b}$$

L'inverse de l'inverse d'une fraction est égal à elle-même.

25.4.2 D'un produit

Soit $a, b, c, d \in \mathbb{Z} \setminus \{0\}$. On a :

$$\frac{a}{b} \cdot \frac{c}{d} \cdot \frac{d}{c} \cdot \frac{b}{a} = \frac{a}{b} \cdot \frac{c \cdot d}{d \cdot c} \cdot \frac{b}{a} = \frac{a}{b} \cdot 1 \cdot \frac{b}{a} = \frac{a}{b} \cdot \frac{b}{a} = \frac{a \cdot b}{b \cdot a} = 1$$

On en conclut que :

$$\frac{d}{c} \cdot \frac{b}{a} = \left(\frac{a}{b} \cdot \frac{c}{d}\right)^{-1}$$

25.5 Addition

25.5.1 Dénominateur commun

Soit $a, b, n \in \mathbb{Z}$ avec $n \neq 0$. On voit que :

$$\frac{a}{n} + \frac{b}{n} = \frac{a}{1} \cdot \frac{1}{n} + \frac{b}{1} \cdot \frac{1}{n} = a \cdot \frac{1}{n} + b \cdot \frac{1}{n}$$

Si on veut conserver la distributivité de la multiplication sur l'addition, on doit donc avoir :

$$\frac{a}{n} + \frac{b}{n} = (a + b) \cdot \frac{1}{n} = \frac{a + b}{n}$$

Pour additionner deux fractions possédant un même dénominateur, il suffit d'additionner les numérateurs.

25.5.2 Générique

Soit $a, b, c, d \in \mathbb{Z}$ avec $b, d \neq 0$. Tentons à présent d'obtenir une expression de la somme $(a, b) + (c, d)$. Multipliant la première par $d/d = 1$ et la seconde par $b/b = 1$, il vient :

$$\frac{a}{b} + \frac{c}{d} = \frac{a \cdot d}{b \cdot d} + \frac{c \cdot b}{d \cdot b}$$

Comme les deux fractions du membre de droite ont un même dénominateur, on a finalement :

$$\frac{a}{b} + \frac{c}{d} = \frac{a \cdot d + c \cdot b}{b \cdot d}$$

25.5.3 Neutre additif

Soit $a, b \in \mathbb{Z}$ avec $b \neq 0$. On a :

$$\frac{a}{b} + 0 = \frac{a}{b} + \frac{0}{1} = \frac{a \cdot 1 + 0 \cdot b}{b \cdot 1} = \frac{a}{b}$$

Le rationnel :

$$0 = \frac{0}{1}$$

est le neutre pour l'addition. On a aussi :

$$\frac{0}{n} = \frac{0 \cdot n}{1 \cdot n} = \frac{0}{1} = 0$$

25.6 Opposé

Soit $m, n \in \mathbb{Z}$. On sait que :

$$(-m) \div n = -(m \div n)$$
$$m \div (-n) = -(m \div n)$$
$$(-m) \div (-n) = m \div n$$

A-t-on les même propriétés pour les rationnels ? Soit $a, b \in \mathbb{Z}$ avec $b \neq 0$. On a :

$$\frac{a}{b} + \frac{-a}{b} = \frac{a - a}{b} = \frac{0}{b} = 0$$

L'opposé de a/b est donc $-a/b$:

$$-\frac{a}{b} = \frac{-a}{b}$$

On a également :

$$\frac{a}{-b} = \frac{(-1) \cdot a}{(-1) \cdot (-b)} = \frac{-a}{b} = -\frac{a}{b}$$

et :

$$\frac{-a}{-b} = \frac{(-1) \cdot (-a)}{(-1) \cdot (-b)} = \frac{a}{b}$$

25.7 Représentation canonique

Soit un rationnel $x \in \mathbb{Q}$ et des entiers $a, b \in \mathbb{Z}$ vérifiant $b \neq 0$ et :

$$x = \frac{a}{b}$$

Si $b > 0$, on pose :

$$m = a$$
$$n = b$$

et on a :

$$x = \frac{m}{n}$$

avec $n > 0$. Si $b < 0$, on pose :

$$m = -a$$
$$n = -b$$

et on a :

$$x = \frac{a}{b} = \frac{-a}{-b} = \frac{m}{n}$$

avec $n > 0$. On peut donc toujours se ramener à la représentation canonique :

$$x = \frac{m}{n}$$

où $m, n \in \mathbb{Z}$ et $n > 0$.

25.8 Soustraction

Soit $a, b, c, d \in \mathbb{Z}$ avec $b, d \neq 0$. On définit :

$$\frac{a}{b} - \frac{c}{d} = \frac{a}{b} + \frac{-c}{d} = \frac{a \cdot d - c \cdot b}{b \cdot d}$$

25.9 Ordre

Soit $a, b, c, d \in \mathbb{Z}$ avec $b, d \neq 0$. Nous allons étendre l'ordre sur l'ensemble des rationnels en imposant que la multiplication par un nombre strictement positif conserve l'ordre et que la multiplication par un nombre strictement négatif l'inverse. Soit l'inégalité :

$$\frac{a}{b} \le \frac{c}{d}$$

— Supposons que $b, d \ge 0$. En multipliant par $b \cdot d > 0$, on voit que l'inégalité est équivalente à :

$$a \cdot d \le c \cdot b$$

Nous dirons donc que :

$$\frac{a}{b} \le \frac{c}{d}$$

si et seulement si :

$$a \cdot d \le c \cdot b$$

— Supposons que $b \le 0$ et $d \ge 0$. On a $b \cdot d < 0$ et l'inégalité est équivalente à :

$$a \cdot d \ge c \cdot b$$

Nous dirons donc que :

$$\frac{a}{b} \le \frac{c}{d}$$

si et seulement si :

$$a \cdot d \ge c \cdot b$$

— Supposons que $b \ge 0$ et $d \le 0$. On a $b \cdot d < 0$ et l'inégalité est équivalente à :

$$a \cdot d \ge c \cdot b$$

Nous dirons donc que :

$$\frac{a}{b} \le \frac{c}{d}$$

si et seulement si :

$$a \cdot d \geq c \cdot b$$

— Supposons que $b, d \leq 0$, on a $b \cdot d = (-b) \cdot (-d) > 0$ et l'inégalité est équivalente à :

$$a \cdot d \leq c \cdot b$$

Nous dirons donc que :

$$\frac{a}{b} \leq \frac{c}{d}$$

si et seulement si :

$$a \cdot d \leq c \cdot b$$

25.9.1 Positifs et négatifs

On définit les ensembles des rationnels positifs et négatifs par :

$$\mathbb{Q}^+ = \{x \in \mathbb{Q} : x \geq 0\}$$

$$\mathbb{Q}^- = \{x \in \mathbb{Q} : x \leq 0\}$$

Soit un rationnel x sous la forme canonique :

$$x = \frac{a}{b}$$

Si x est négatif :

$$x = \frac{a}{b} \leq 0 = \frac{0}{1}$$

on a par définition :

$$a \cdot 1 \leq 0 \cdot b = 0$$

autrement dit :

$$a \leq 0$$

Invérsément, si x est positif :

$$x = \frac{a}{b} \geq 0 = \frac{0}{1}$$

on a par définition :

$$a \cdot 1 \geq 0 \cdot b = 0$$

autrement dit :

$$a \geq 0$$

25.9.2 Invariance sous multiplication

Soit les rationnels x, y, z sous forme canonique :

$$x = \frac{a}{b}$$

$$y = \frac{c}{d}$$

$$z = \frac{e}{f}$$

avec $a, b, c, d, e, f \in \mathbb{Z}$ et $b, d, f > 0$. On suppose que $z \geq 0$, c'est-à-dire :

$$e \geq 0$$

et que :

$$x \leq y$$

c'est-à-dire :

$$a \cdot d \leq c \cdot b$$

L'ordre entre entiers est conservé par la multiplication de l'entier positif $e \cdot f$:

$$a \cdot e \cdot d \cdot f \leq c \cdot e \cdot b \cdot f$$

Par définition de l'ordre entre rationnels, on en déduit que :

$$\frac{a \cdot e}{b \cdot f} \leq \frac{c \cdot e}{d \cdot f}$$

c'est-à-dire :

$$x \cdot z \leq y \cdot z$$

L'ordre entre rationnels est conservé sous la multiplication d'un rationnel positif.

25.9.3 Invariance sous addition

Soit les rationnels x, z, u, v sous forme canonique :

$$x = \frac{a}{b}$$

$$y = \frac{c}{d}$$

$$u = \frac{e}{f}$$

$$v = \frac{g}{h}$$

On suppose que :

$$x \leq y$$

$$u \leq v$$

c'est-à-dire :

$$\frac{a}{b} \le \frac{c}{d}$$

$$\frac{e}{f} \le \frac{g}{h}$$

ou :

$$a \cdot d \le c \cdot b$$

$$e \cdot h \le g \cdot f$$

L'ordre entre entiers étant conservé sous la multiplication de $f \cdot h > 0$, on a :

$$a \cdot d \cdot f \cdot h \le c \cdot b \cdot f \cdot h$$

L'ordre entre entiers étant conservé sous la multiplication de $b \cdot d > 0$, on a :

$$e \cdot h \cdot b \cdot d \le g \cdot f \cdot b \cdot d$$

L'ordre entre entiers étant conservé sous l'addition, on a :

$$a \cdot d \cdot f \cdot h + e \cdot h \cdot b \cdot d \le c \cdot b \cdot f \cdot h + g \cdot f \cdot b \cdot d$$

En utilisant la distributivité, on en déduit que :

$$(a \cdot f + e \cdot b) \cdot d \cdot h \le (c \cdot h + g \cdot d) \cdot b \cdot f$$

ou, par définition de l'ordre entre rationnels :

$$\frac{a \cdot f + e \cdot b}{b \cdot f} \le \frac{c \cdot h + g \cdot d}{d \cdot h}$$

qui n'est rien d'autre que :

$$\frac{a}{b} + \frac{e}{f} \le \frac{c}{d} + \frac{g}{h}$$

c'est-à-dire :

$$x + u \le y + v$$

L'ordre entre rationnels est conservé sous l'addition.

25.9.4 Invariance sous soustraction

Soit les rationnels x, z, u, v tels que :

$$x \le y$$

$$u \ge v$$

En ajoutant :

$$(-u) + (-v) \ge (-u) + (-v)$$

à la seconde inégalité, on obtient :

$$u + (-u) + (-v) \ge v + (-u) + (-v)$$

qui se simplifie en :

$$-v \geq -u$$

ou :

$$-u \leq -v$$

On a donc :

$$x + (-u) \leq y + (-v)$$

c'est-à-dire :

$$x - u \leq y - v$$

25.10 Signe

La fonction signe est définie par :

$$\operatorname{sign}(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ -1 & \text{si } x \ < \ 0 \end{cases}$$

pour tout $x \in \mathbb{Q}$

25.11 Valeur absolue

Soit $x \in \mathbb{Q}$. On définit la valeur absolue de x par :

$$|x| = \max\{x, -x\}$$

25.12 Notation décimale

Soit les tuples :

$$(i_0, ..., i_m) \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}^{m+1}$$

et :

$$(j_1, ..., j_n) \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}^n$$

La notation décimale associée est définie par :

$$i_m ... i_0, j_1 ... j_n = i_0 + ... + i_m \cdot 10^m + j_1 \cdot \frac{1}{10} + ... + j_n \cdot \frac{1}{10^n}$$

pour les rationnels positifs et :

$$i_m ... i_0, j_1 ... j_n = -\left(i_0 + ... + i_m \cdot 10^m + j_1 \cdot \frac{1}{10} + ... + j_n \cdot \frac{1}{10^n} \right)$$

pour les rationnels négatifs. Exemple :

$$-7512, 34 = -\left(2 + 1 \cdot 10 + 5 \cdot 10^2 + 7 \cdot 10^3 + 3 \cdot \frac{1}{10} + 4 \cdot \frac{1}{10^2} \right)$$

25.13 Division

Soit $x, y \in \mathbb{Q}$ et $a, b, c, d \in \mathbb{Z}$ avec $b, c, d \neq 0$ et :

$$x = \frac{a}{b}$$

$$y = \frac{c}{d}$$

On dit que $q \in \mathbb{Q}$ est la division de x par y et on le note :

$$\frac{x}{y} = q$$

si :

$$x = q \cdot y$$

Multipliant par y^{-1}, on a bien sur :

$$q = x \cdot y^{-1} = \frac{a}{b} \cdot \frac{d}{c}$$

La division de deux fractions s'écrit donc simplement :

$$\frac{x}{y} = \frac{a/b}{c/d} = \frac{a}{b} \cdot \frac{d}{c}$$

25.14 Puissance

Soit $m, n \in \mathbb{N}$ et $a, b \in \mathbb{Z}$. Les puissances positives s'étendent facilement aux fractions :

$$\left(\frac{a}{b}\right)^n = \frac{a}{b} \cdot ... \cdot \frac{a}{b} = \frac{a^n}{b^n}$$

On définit de plus les puissances négatives par :

$$\left(\frac{a}{b}\right)^{-n} = \left(\frac{b}{a}\right)^n$$

Mais comme un produit d'inverses est égal à l'inverse des produits, on a également :

$$\left(\frac{a}{b}\right)^{-n} = \left(\frac{a^n}{b^n}\right)^{-1}$$

25.14.1 Somme en exposant

Soit $x \in \mathbb{Q}$. On vérifie que :

$$x^{m+n} = x^m \cdot x^n$$

$$x^{m-n} = x^m \cdot x^{-n} = \frac{x^m}{x^n}$$

$$x^{-m+n} = x^{-m} \cdot x^n = \frac{x^n}{x^m}$$

$$x^{-m-n} = x^{-m} \cdot x^{-n} = \frac{1}{x^m \cdot x^n}$$

On a donc bien :

$$x^{i+j} = x^i \cdot x^j$$

pour tout $i, j \in \mathbb{Z}$.

25.15 Carré

Le carré est la deuxième puissance d'un nombre :

$$x^2 = x \cdot x$$

25.15.1 Positivité

Soit $x \in \mathbb{Q}$. Si $x \geq 0$, on pose $p = x$. On a donc $p \geq 0$. La multiplication par un rationnel positif ne modifiant pas l'ordre, on a :

$$p^2 = p \cdot p \geq 0 \cdot p = 0$$

Le carré est donc également positif :

$$x^2 = p^2 \geq 0$$

Si $x \leq 0$, on pose $p = -x$. On a donc $p \geq 0$. On voit que $x = -p$ et que :

$$x^2 = (-p)^2 = (-p) \cdot (-p) = p \cdot p = p^2 \geq 0$$

On a donc :

$$x^2 \geq 0$$

pour tout $x \in \mathbb{Q}$.

25.15.2 Ordre

Soit $x, y \in \mathbb{Q}$ vérifiant :

$$x, y \geq 0$$

et :

$$x \leq y$$

Comme $x \geq 0$, l'inégalité $x \leq y$ est conservée sous la multiplication par x :

$$x^2 = x \cdot x \leq y \cdot x$$

Comme $y \geq 0$, l'inégalité $x \leq y$ est conservée sous la multiplication par y :

$$x \cdot y \leq y \cdot y = y^2$$

En rassemblant ces résultats, il vient :

$$x^2 \leq y \cdot x = x \cdot y \leq y^2$$

on a donc :

$$x^2 \leq y^2$$

La fonction $f : x \mapsto x^2$ est croissante sur l'ensemble des rationnels positifs.

25.15.3 Binôme

Soit $x, y \in \mathbb{Z}$. En utilisant la distributivité, on obtient :

$$(x+y)^2 = (x+y) \cdot (x+y) = x \cdot (x+y) + y \cdot (x+y)$$

En l'utilisant une seconde fois, on arrive à :

$$(x+y)^2 = x \cdot x + x \cdot y + y \cdot x + y \cdot y$$

En utilisant la commutativité, on arrive au développement :

$$(x+y)^2 = x^2 + x \cdot y + x \cdot y + y^2$$

que l'on peut réexprimer comme :

$$(x+y)^2 = x^2 + 2 \cdot x \cdot y + y^2$$

Différence

On a :

$$(x-y)^2 = (x+(-y))^2 = x^2 + 2 \cdot x \cdot (-y) + y^2$$

et donc :

$$(x-y)^2 = x^2 - 2 \cdot x \cdot y + y^2$$

25.16 Corps

$(\mathbb{Q}, +, \cdot)$ est un corps commutatif.

25.17 Densité

Soit le rationnel $\epsilon \in \mathbb{Q}$ vérifiant $\epsilon > 0$ exprimé sous la forme canonique :

$$\epsilon = \frac{m}{n}$$

avec $m, n \in \mathbb{Z}$ et $n > 0$. Comme ϵ est strictement positif, on a :

$$m > 0$$

Existe-t-il un rationnel intermédiaire δ vérifiant :

$$0 < \delta < \epsilon$$

Posons :

$$\delta = \frac{m}{2\ n}$$

Il est clair que $\delta > 0$. On a :

$$n < 2\ n$$

En multipliant par $m > 0$, on obtient :

$$m \cdot n < m \cdot 2 \cdot n$$

Par définition de l'ordre sur les rationnels, cette inégalité est équivalente à :

$$\delta = \frac{m}{2\ n} \ < \ \frac{m}{n} = \epsilon$$

On en conclut que :

$$0 \ < \ \delta \ < \ \epsilon$$

25.17.1 Rationnels distincts

Soit les rationnels x, y vérifiant :

$$x \ < \ y$$

On pose :

$$\epsilon = y - x \ > \ 0$$

On peut trouver un rationnel δ tel que :

$$0 \ < \ \delta \ < \ \epsilon$$

En additionnant cette inégalité avec :

$$x \le x \le x$$

on arrive à :

$$x \ < \ x + \delta \ < \ x + \epsilon$$

Soit :

$$z = x + \delta$$

Comme :

$$x + \epsilon = x + y - x = y$$

on a :

$$x \ < \ z \ < \ y$$

On peut toujours trouver un rationnel strictement compris entre deux rationnels distincts.

Chapitre 26

Fonctions et opérations

Dépendances

— Chapitre 10 : L'algèbre
— Chapitre 23 : Les naturels
— Chapitre 24 : Les entiers
— Chapitre 25 : Les rationnels

26.1 Opérations induites

Soit les fonctions $f, g \in B^A$ et $\divideontimes$ une opération sur A. On définit alors la fonction $f \divideontimes g$ par :

$$(f \divideontimes g)(x) = f(x) \divideontimes g(x)$$

pour tout $x \in A$. Nous avons ainsi défini une opération sur B^A :

$$\divideontimes : B^A \times B^A \mapsto B^A$$

induite par la loi équivalente sur B.

26.1.1 Usuelles

Sur les anneaux et les corps, ou sur tout ensemble où sont définies les opérations usuelles $+, -, \cdot, /$, on aura ainsi :

— les sommes : $(f + g)(x) = f(x) + g(x)$
— les produits : $(f \cdot g)(x) = f(x) \cdot g(x)$
— les soustractions : $(f - g)(x) = f(x) - g(x)$
— les divisions : $(f/g)(x) = f(x)/g(x)$

pour tout $x \in A$.

26.2 Opposé et inverse

Soit la fonction $f \in B^A$. Si l'inverse pour l'addition $-f(x)$ existe pour tout $x \in A$, on définit la fonction opposée $-f$ par :

$$(-f)(x) = -f(x)$$

pour tout $x \in A$. On a alors :

$$f - f = 0$$

Si l'inverse pour la multiplication $1/f(x)$ existe pour tout $x \in A$, on définit la fonction $1/f$ par :

$$(1/f)(x) = 1/f(x)$$

pour tout $x \in A$. On a alors :

$$f \cdot 1/f = 1$$

26.3 Opérations mixte

Soit la fonction $f \in B^A$ et $c \in B$. Étant donnée une opération $\divideontimes$ définie sur B, on définit les opérations mixtes $f \divideontimes c$ et $c \divideontimes f$ par :

$$(f \divideontimes c)(x) = f(x) \divideontimes c$$
$$(c \divideontimes f)(x) = c \divideontimes f$$

pour tout $x \in A$. Dans le cas où l'opération $\divideontimes$ définie sur B est commutative, on a bien entendu $f \divideontimes c = c \divideontimes f$.

Sur les ensembles où sont définies les lois usuelles $+, -, \cdot, /$, on aura ainsi :

— les sommes :

$$(f + c)(x) = f(x) + c$$
$$(c + f)(x) = c + f(x)$$

— les produits :

$$(f \cdot c)(x) = f(x) \cdot c$$
$$(c \cdot f)(x) = c \cdot f(x)$$

— les soustractions :

$$(f - c)(x) = f(x) - c$$
$$(c - f)(x) = c - f(x)$$

— les divisions :

$$(f/c)(x) = f(x)/c$$
$$(c/f)(x) = c/f(x)$$

pour tout $x \in A$.

26.4 Commutateur

Notons qu'en général la composée n'est pas commutative. On peut en effet trouver des fonctions $f, g : A \mapsto A$ telles que $f \circ g \neq g \circ f$.

Cette constatation nous amène à la notion de commutateur. Il s'agit d'une fonction $[f, g] : A \mapsto A$ définie par :

$$[f, g] = f \circ g - g \circ f$$

Cet opérateur est clairement antisymétrique :

$$[f, g] = -[g, f]$$

Dans le cas particulier où $[f, g] = 0$, on dit que les fonctions f et g commutent.

26.5 Puissance fonctionnelle et puissance

Attention à ne pas confondre exposant de la fonction et exposant de la valeur de la fonction en un point :

$$f(x)^n = f(x) \cdot f(x)^{n-1} \neq f^n(x) = (f \circ f^{n-1})(x)$$

26.6 Noyau

Le noyau d'une fonction $f : A \mapsto B$ est l'ensemble des $x \in A$ où les valeurs de f sont égales au neutre pour l'addition :

$$\ker f = \{x \in A : f(x) = 0\}$$

26.7 Support d'une fonction

Le support d'une fonction $f : A \mapsto \Omega$ est l'adhérence de l'ensemble des points où f prend une valeur non nulle :

$$\operatorname{supp} f = \operatorname{adh}\{x \in A : f(x) \neq 0\}$$

Remarque

Attention à ne pas confondre le suprémum d'un ensemble, noté « sup » avec un seul « p », avec le support, noté « supp », qui prend deux « p ».

Chapitre 27

Fonctions indicatrices

Dépendances

- Chapitre 10 : Les structures algébriques
- Chapitre 12 : Les ordres
- Chapitre 23 : Les naturels
- Chapitre 24 : Les entiers
- Chapitre 25 : Les rationnels

27.1 Définition

Soit un ensemble de référence Ω et $A \subseteq \Omega$. La fonction indicatrice

$$\boldsymbol{\delta}_A : \Omega \mapsto \{0, 1\}$$

associée à l'ensemble A permet de déterminer si un élément quelconque de Ω appartient où non à A. Elle est définie par :

$$\boldsymbol{\delta}_A(x) = \begin{cases} 1 & \text{si} \quad x \in A \\ 0 & \text{si} \quad x \in \Omega \setminus A \end{cases}$$

27.1.1 Crochets

On note aussi :

$$\boldsymbol{\delta}[A] = \boldsymbol{\delta}_A$$

et :

$$\boldsymbol{\delta}[A](x) = \boldsymbol{\delta}_A(x)$$

27.2 Ensemble vide

Notons en particulier que :

$$\boldsymbol{\delta}[\emptyset] = 0$$

27.3 Ordre

Soit $A, B \subseteq \Omega$ avec $A \subseteq B$. On voit que :
— pour tout $x \in A$, on a aussi $x \in B$ et $\boldsymbol{\delta}_A(x) = \boldsymbol{\delta}_B(x) = 1$
— pour tout $x \in \Omega \setminus A$, on a $\boldsymbol{\delta}_A(x) = 0 \le \boldsymbol{\delta}_B(x)$
On en conclut que :

$$\boldsymbol{\delta}_A(x) \le \boldsymbol{\delta}_B(x)$$

pour tout $x \in \Omega$, c'est-à-dire :

$$\boldsymbol{\delta}_A \le \boldsymbol{\delta}_B$$

L'ordre des fonctions indicatrices correspond à l'ordre inclusif des ensembles.

27.4 Intersection

Soit $A, B \subseteq \Omega$. On voit que :
— si $x \in A \cap B$, on a $\boldsymbol{\delta}_A(x) \cdot \boldsymbol{\delta}_B(x) = 1 \cdot 1 = 1 = \boldsymbol{\delta}[A \cap B](x)$
— si $x \in A \cap (\Omega \setminus B)$, on a $\boldsymbol{\delta}_A(x) \cdot \boldsymbol{\delta}_B(x) = 1 \cdot 0 = 0 = \boldsymbol{\delta}[A \cap B](x)$
— si $x \in (\Omega \setminus A) \cap B$, on a $\boldsymbol{\delta}_A(x) \cdot \boldsymbol{\delta}_B(x) = 0 \cdot 1 = 0 = \boldsymbol{\delta}[A \cap B](x)$
— si $x \in \Omega \setminus (A \cup B)$, on a $\boldsymbol{\delta}_A(x) \cdot \boldsymbol{\delta}_B(x) = 0 \cdot 0 = 0 = \boldsymbol{\delta}[A \cap B](x)$
Donc :

$$\boldsymbol{\delta}[A \cap B] = \boldsymbol{\delta}_A \cdot \boldsymbol{\delta}_B$$

27.5 Union d'ensembles disjoints

Si $A \cap B = \emptyset$, on a clairement :

$$\boldsymbol{\delta}[A \cup B] = \boldsymbol{\delta}_A + \boldsymbol{\delta}_B$$

27.6 Différence

Soit les ensembles A et B. On a la décomposition :

$$A = (A \setminus B) \cup (A \cap B)$$

On sait que les deux sous-ensembles sont disjoints :

$$(A \setminus B) \cap (A \cap B) = \emptyset$$

On a donc :

$$\boldsymbol{\delta}_A = \boldsymbol{\delta}[A \setminus B] + \boldsymbol{\delta}[A \cap B]$$

On en conclut que :

$$\begin{aligned} \boldsymbol{\delta}[A \setminus B] &= \boldsymbol{\delta}_A - \boldsymbol{\delta}[A \cap B] \\ &= \boldsymbol{\delta}_A - \boldsymbol{\delta}_A \cdot \boldsymbol{\delta}_B \end{aligned}$$

27.7 Union d'ensembles quelconques

Soit les ensembles A et B. On a la décomposition :

$$A \cup B = (A \setminus B) \cup B$$

On voit que les deux sous-ensembles sont disjoints :

$$(A \setminus B) \cap B = \emptyset$$

On en conclut que :

$$\boldsymbol{\delta}[A \cup B] = \boldsymbol{\delta}[A \setminus B] + \boldsymbol{\delta}_B$$

Reprenant l'expression de la différence, on a finalement :

$$\begin{aligned} \boldsymbol{\delta}[A \cup B] &= \boldsymbol{\delta}_A + \boldsymbol{\delta}_B - \boldsymbol{\delta}[A \cap B] \\ &= \boldsymbol{\delta}_A + \boldsymbol{\delta}_B - \boldsymbol{\delta}_A \cdot \boldsymbol{\delta}_B \end{aligned}$$

27.8 Produit cartésien

Pour tout $(x, y) \in A \times B$, on a :

$$\boldsymbol{\delta}[A \times B](x, y) = \boldsymbol{\delta}_A(x) \cdot \boldsymbol{\delta}_B(y)$$

27.9 Delta de Kronecker

Considérons l'ensemble $I \subset \Omega^2$ défini par :

$$I = \{(i, j) \in \Omega^2 : i = j\}$$

Le delta de Kronecker est simplement :

$$\boldsymbol{\delta}_{ij} = \boldsymbol{\delta}_I(i, j)$$

On a donc :

$$\boldsymbol{\delta}_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Notation

On note aussi :

$$\boldsymbol{\delta}^{ij} = \boldsymbol{\delta}_i^j = \boldsymbol{\delta}_{ij}$$

Chapitre 28

Parité

28.1 Définition

On dit qu'une fonction $\sigma : \mathbb{R} \mapsto \mathbb{R}$ est paire si :

$$\sigma(-x) = \sigma(x)$$

pour tout réel x. Inversément, on dit qu'une fonction $\alpha : \mathbb{R} \mapsto \mathbb{R}$ est impaire si :

$$\alpha(-x) = -\alpha(x)$$

pour tout réel x.

28.2 Décomposition

Soit une fonction $\varphi : \mathbb{R} \mapsto \mathbb{R}$. La fonction σ définie par :

$$\sigma(x) = \frac{1}{2} \left[\varphi(x) + \varphi(-x)\right]$$

pour tout réel x vérifie :

$$\sigma(-x) = \frac{1}{2} \left[\varphi(-x) + \varphi(x)\right] = \frac{1}{2} \left[\varphi(x) + \varphi(-x)\right] = \sigma(x)$$

Cette fonction est donc paire. La fonction α définie par :

$$\alpha(x) = \frac{1}{2} \left[\varphi(x) - \varphi(-x)\right]$$

pour tout réel x vérifie :

$$\alpha(-x) = \frac{1}{2} \left[\varphi(-x) - \varphi(x)\right] = -\frac{1}{2} \left[\varphi(x) - \varphi(-x)\right] = -\alpha(x)$$

Cette fonction est donc impaire. On voit que :

$$\sigma(x) + \alpha(x) = \frac{1}{2} \left[\varphi(x) + \varphi(-x) + \varphi(x) - \varphi(-x)\right] = \frac{2}{2}\ \varphi(x) = \varphi(x)$$

On a donc la décomposition :

$$\varphi = \sigma + \alpha$$

où σ est paire et α impaire. On dit que σ est la composante paire de φ et que α est la composante impaire de φ·

Chapitre 29

Distances

Dépendances

— Chapitre 10 : Les structures algébriques
— Chapitre 12 : Les ordres
— Chapitre 21 : Les topologies

29.1 Définition

Afin de généraliser le plus possible la notion de distance au sens usuel, nous allons nous poser la question : quelles sont les caractéristiques fondamentales d'une distance ? Nous en déduirons les propriétés que doit respecter une distance générique.

Soit l'ensemble Ω, un corps $\mathbb{K}$ et une fonction $\mathfrak{dist} : \Omega \times \Omega \to \mathbb{K}$ représentant une distance entre deux éléments de l'ensemble Ω. Choisissons des éléments quelconques $x, y, z \in \Omega$. Au sens usuel, une distance entre deux objets est clairement positive :

$$\mathfrak{dist}(x, y) \ge 0$$

Elle est également symétrique puisqu'on obtient la même distance lorsqu'on intervertit les objets :

$$\mathfrak{dist}(x, y) = \mathfrak{dist}(y, x)$$

Par ailleurs , la distance entre deux objets identiques doit évidemment être nulle :

$$\mathfrak{dist}(x, x) = 0$$

On impose également que le seul élément y qui puisse être à distance nulle de x soit l'élément x lui-même :

$$\mathfrak{dist}(x, y) = 0 \ \Rightarrow \ x = y$$

Enfin, il est toujours plus court d'aller directement de x à z plutôt que de passer par une étape y. On a donc l'inégalité triangulaire :

$$\mathfrak{dist}(x, z) \le \mathfrak{dist}(x, y) + \mathfrak{dist}(y, z)$$

Remarque

Parfois, au lieu d'imposer l'égalité de deux éléments situés à distance nulle l'un de l'autre, on impose juste l'équivalence suivant un critère prédéfini :

$$\mathfrak{dist}(x, y) = 0 \ \Rightarrow \ x \equiv y$$

29.2 Distance à un ensemble

Quelle est la distance à parcourir d'une ville donnée lorsqu'on désire se rendre dans un certain pays ? On a envie de dire que la distance est parcourue dès que l'on a atteint la frontière du pays en question. Comme on choisit généralement le chemin le plus court pour arriver à destination, on se rend compte que la distance ville - pays est le minimum des distances entre la ville et tous les points appartenant au pays.

Maintenant, remplaçons la ville par un élément $x \in \Omega$ et le pays par un ensemble $A \subseteq \Omega$. On définit simplement la distance de x à A comme étant l'infimum des distances de x à un point quelconque de A :

$$\mathfrak{dist}(x, A) = \inf_{a \in A} \mathfrak{dist}(x, a) = \inf\{\mathfrak{dist}(x, a) : a \in A\}$$

Inclusion

Si $B \subseteq A$, on en déduit directement que :

$$\mathfrak{dist}(a, A) \le \mathfrak{dist}(a, B)$$

Notation

On note aussi :

$$\mathfrak{dist}(A, a) = \mathfrak{dist}(a, A)$$

Self-distance

Soit $a \in A$. Comme $\mathfrak{dist}(a, a) = 0$ et que la distance $\mathfrak{dist}(a, b) \ge 0$ pour tout $b \in A$, on voit que le choix $b = a$ minimise $\mathfrak{dist}(a, b)$ sur A. Donc :

$$\mathfrak{dist}(a, A) = 0$$

pour tout élément $a \in A$.

29.3 Distance inter-ensembles

La distance entre deux ensembles A et B est l'infimum des distances possibles entre les couples $(a, b) \in A \times B$:

$$\mathfrak{dist}(A, B) = \inf\{\mathfrak{dist}(a, b) : (a, b) \in A \times B\}$$

On a bien évidemment :

$$\mathfrak{dist}(A, B) = \inf\{\mathfrak{dist}(a, B) : a \in A\} = \inf\{\mathfrak{dist}(A, b) : b \in B\}$$

29.4 Boules

Les boules sont la généralisation des disques et des sphères. Or, ce qui caractérise ces entités, c'est qu'elle incluent des points x vérifiant $\mathfrak{dist}(x, c) \le r$, où c est le centre et r le rayon. On définit par conséquent la boule fermée $\mathfrak{B}[c, r]$ par :

$$\mathfrak{B}[c, r] = \{x \in \Omega : \mathfrak{dist}(x, c) \le r\}$$

où r est un réel positif.

Si l'on veut que la distance soit strictement inférieure à r, on considérera plutôt la définition de la boule ouverte :

$$\mathfrak{B}(c,r) = \{x \in \Omega : \mathfrak{dist}(x,c) \ < \ r\}$$

29.5 Topologie métrique

La topologie usuelle définie sur les ensembles munis d'une distance est celle générée par les boules ouvertes, soit les éléments de la collection :

$$\mathcal{B} = \{\mathfrak{B}(c,r) : c \in \Omega, \ r \in \mathbb{K}, \ r \ > \ 0\}$$

La topologie métrique $\mathfrak{T}$ s'écrit donc :

$$\mathfrak{T} = \mathrm{Topo}(\mathcal{B}, \Omega)$$

Toute union de boules ouvertes est donc un ouvert.

29.6 Intérieur

Soit $A \subseteq \Omega$.

— Soit $x \in \operatorname{int} A$. L'élément x appartient donc à un ouvert U contenu dans A. Comme $x \in U$, on a $U \neq \emptyset$. On en conclut que U doit être une union de boules ouvertes. On peut donc trouver un $a \in U$ et un $\delta \ > \ 0$ tel que $x \in \mathfrak{B}(a,\delta) \subseteq U$. Comme $d = \mathfrak{dist}(a,x) \ < \ \delta$, on a $\delta - d \ > \ 0$. Soit $\epsilon = \delta - d$ et $y \in \mathfrak{B}(x,\epsilon)$. On a :

$$\mathfrak{dist}(a,y) \le \mathfrak{dist}(a,x) + \mathfrak{dist}(x,y) \ < \ d + \epsilon = \delta$$

Donc $\mathfrak{B}(x,\epsilon) \subseteq \mathfrak{B}(a,\delta) \subseteq U \subseteq A$ et $\mathfrak{dist}(x, \Omega \setminus A) \ > \ 0$.

— Réciproquement, si $z \in \Omega$ vérifie $d = \mathfrak{dist}(x, \Omega \setminus A) \ > \ 0$, on a $z \in \mathfrak{B}(z, d/2) \subseteq A$. L'élément z appartient donc à un ouvert contenu dans A. Il appartient donc à l'union des ouverts inclus dans A, c'est-à-dire $z \in \operatorname{int} A$.

On conclut de ce qui précède que :

$$\operatorname{int} A = \{x \in \Omega : \mathfrak{dist}(x, \Omega \setminus A) \ > \ 0\}$$

29.7 Adhérence

On voit que :

$$\operatorname{int}(\Omega \setminus A) = \{x \in \Omega : \mathfrak{dist}(x, A) \ > \ 0\}$$

Le complémentaire de cet ensemble est bien sur constitué des $x \in \Omega$ vérifiant $\mathfrak{dist}(x,A) = 0$. Or, nous avons vu que ce complémentaire n'est rien d'autre que l'adhérence de A :

$$\operatorname{adh} A = \{x \in \Omega : \mathfrak{dist}(x, A) = 0\}$$

29.8 Adhérence carrée

Soit $x \in \Omega$, $A \subseteq \Omega$ et :

$$B = \operatorname{adh} A$$
$$C = \operatorname{adh} B = \operatorname{adh} \operatorname{adh} A$$

Soit $c \in C$. Pour tout $\epsilon \in \mathbb{K}$ avec $\epsilon > 0$, on peut trouver $b \in B$ tel que :

$$\mathfrak{dist}(c, b) \le \epsilon$$

et ensuite $a \in A$ tel que :

$$\mathfrak{dist}(b, a) \le \epsilon$$

On a donc :

$$\mathfrak{dist}(c, a) \le \mathfrak{dist}(c, b) + \mathfrak{dist}(b, a) \le \epsilon + \epsilon$$

L'infimum est par conséquent nul :

$$\mathfrak{dist}(c, A) = \inf_{a \in A} \mathfrak{dist}(c, a) = 0$$

On en déduit que $c \in \operatorname{adh} A$. Nous venons de montrer que :

$$\operatorname{adh} \operatorname{adh} A \subseteq \operatorname{adh} A$$

Mais comme l'inverse est également vrai, on a :

$$\operatorname{adh} \operatorname{adh} A = \operatorname{adh} A$$

Chapitre 30

Limites

Dépendances

— Chapitre 29 : Les distances

30.1 Définition

Soit les ensembles E, F munis respectivement des distances $\mathfrak{dist}_E$ et $\mathfrak{dist}_F$, un sous-ensemble $D \subseteq E$ et la fonction $f : D \mapsto F$. Comme la distance utilisée est sans ambiguité d'après la nature des objets dont elle mesure l'éloignement :

$$\mathfrak{dist}(x, y) = \mathfrak{dist}_E(x, y) \Leftrightarrow x, y \in E$$
$$\mathfrak{dist}(x, y) = \mathfrak{dist}_F(x, y) \Leftrightarrow x, y \in F$$

on note dans la suite de ce chapitre $\mathfrak{dist}$ à la place de $\mathfrak{dist}_E$ et de $\mathfrak{dist}_F$.

Plaçons nous dans $A \subseteq D$. Nous nous intéressons au cas où $f(x)$ se rapproche de plus en plus d'un certain $L \in F$ lorsque $x \in A$ se rapproche suffisamment d'un certain $a \in E$. Pour toute précision $\epsilon \in \mathbb{K}$, $\epsilon > 0$ aussi petite que l'on veut, on doit alors pouvoir trouver un niveau de proximité $\delta(\epsilon) \in \mathbb{K}$, $\delta(\epsilon) > 0$ tel que :

$$\mathfrak{dist}(f(x), L) \le \epsilon$$

pour tout les $x \in A$ vérifiant :

$$\mathfrak{dist}(x, a) \le \delta(\epsilon)$$

Si cette condition est remplie, on dit que L est la limite de f en a, et on note :

$$\lim_{\substack{x \to a \\ x \in A}} f(x) = L$$

30.1.1 Notations

Lorsque l'ensemble A est évident d'après le contexte, on note simplement :

$$\lim_{x \to a} f(x) = \lim_{\substack{x \to a \\ x \in A}} f(x)$$

Au lieu de noter l'ensemble, on peut citer les conditions qui le définissent. Ainsi par exemple :

$$\lim_{\substack{x \to a \\ x \neq a}} f(x) = \lim_{\substack{x \to a \\ x \in X(a)}} f(x)$$

où :

$$X(a) = D \setminus \{a\}$$

Autre application :

$$\lim_{\substack{x \to a \\ x \le a}} f(x) = \lim_{\substack{x \to a \\ x \in I(a)}} f(x)$$

où :

$$I(a) = \{x \in D : x \le a\}$$

30.1.2 Remarque

Rien n'impose que a appartienne à A. L'existence de la limite de f en a n'implique donc pas que la fonction f soit définie en a.

30.1.3 Unicité

Supposons que b et c soient deux limites de f en a. Soit $\epsilon > 0$. On peut alors trouver un $\alpha > 0$ tel que :

$$\mathfrak{dist}(f(x), b) \le \frac{\epsilon}{2}$$

pour tout x vérifiant $\mathfrak{dist}(x, a) \le \alpha$ et un $\beta > 0$ tel que :

$$\mathfrak{dist}(f(x), c) \le \frac{\epsilon}{2}$$

pour tout x vérifiant $\mathfrak{dist}(x, a) \le \beta$. Posons $\delta = \min\{\alpha, \beta\}$. Si x vérifie $\mathfrak{dist}(x, a) \le \delta$, on a :

$$\{ \ \mathfrak{dist}(f(x), b), \ \mathfrak{dist}(f(x), c) \ \} \le \frac{\epsilon}{2}$$

et :

$$\mathfrak{dist}(b, c) \le \mathfrak{dist}(b, f(x)) + \mathfrak{dist}(f(x), c) \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Comme ce doit être vrai pour tout ϵ strictement positif, on en conclut que $\mathfrak{dist}(b, c) = 0$, c'est-à-dire $b = c$. La limite est donc unique.

30.1.4 Inclusion des boules

Soit :

$$B_1(\delta) = \mathfrak{B}(a, \delta) \cap A$$
$$B_2(\epsilon) = \mathfrak{B}(f(a), \epsilon)$$

La définition de la limite revient à exiger que, pour tout $\epsilon > 0$, on puisse trouver un $\delta > 0$ tel que tout élément de $B_1(\delta)$ auquel on applique la fonction f se retrouve dans $B_2(\epsilon)$. Autrement dit, $f(B_1(\delta)) \subseteq B_2(\epsilon)$.

30.1.5 Limite de la distance

Si L est la limite de f en a, la distance doit converger vers zéro par définition :

$$\lim_{\substack{x \to a \\ x \in A}} \mathfrak{dist}(f(x), L) = 0$$

30.2 Chemin

Supposons qu'il existe des sous-ensembles $A, B \subseteq D$ tels que :

$$\lim_{\substack{x \to a \\ x \in A}} f(x) \neq \lim_{\substack{x \to a \\ x \in B}} f(x)$$

On dit alors que la limite dépend du chemin parcouru.

30.3 Limites à l'infini

Si E est un ensemble ordonné, on peut définir la notion de limite à l'infini. Soit une fonction $f : E \mapsto F$ et $L \in F$. Si, pour toute précision $\epsilon > 0$, on peut trouver une borne inférieure $I(\epsilon) \in E$ telle que :

$$\mathfrak{dist}(f(x), L) \leq \epsilon$$

pour tout $x \in E$ vérifiant $x \geq I(\epsilon)$, on dit alors que f tend vers L à l'infini positif et on écrit :

$$\lim_{x \to +\infty} f(x) = L$$

Symétriquement, si pour toute précision $\epsilon > 0$, on peut trouver une borne supérieure $S(\epsilon) \in E$ telle que :

$$\mathfrak{dist}(f(x), L) \leq \epsilon$$

pour tout $x \in E$ vérifiant $x \leq S(\epsilon)$, on dit alors que f tend vers L à l'infini négatif et on écrit :

$$\lim_{x \to -\infty} f(x) = L$$

Notation

On a également la notation alternative :

$$\lim_{x \to \infty} f(x) = \lim_{x \to +\infty} f(x)$$

ainsi que :

$$\begin{aligned} f(+\infty) &= \lim_{x \to +\infty} f(x) \\ f(-\infty) &= \lim_{x \to -\infty} f(x) \end{aligned}$$

30.4 Limite supremum et infimum

Infini positif

Il arrive que la limite à l'infini positif d'une fonction $f : E \mapsto F$ n'existe pas mais que la limite de la fonction $g : E \mapsto F$, que l'on suppose correctement définie pour tout $x \in E$ par :

$$g(x) = \inf\{f(y) : y \in E,\ y \geq x\}$$

existe. On note alors :

$$\liminf_{x\to+\infty} f(x) = \lim_{x\to+\infty} \inf\{f(y) : y \in E,\ y \ge x\}$$

On définit pareillement la limite du supremum :

$$\limsup_{x\to+\infty} f(x) = \lim_{x\to+\infty} \sup\{f(y) : y \in E,\ y \ge x\}$$

Infini négatif

On définit également :

$$\limsup_{x\to-\infty} f(x) = \lim_{x\to-\infty} \sup\{f(y) : y \in E,\ y \le x\}$$

$$\liminf_{x\to-\infty} f(x) = \lim_{x\to-\infty} \inf\{f(y) : y \in E,\ y \le x\}$$

30.5 Limites infinies

Si F est un ensemble ordonné, on peut définir la notion de limite infinie.

Positive

Si, pour tout $G \in F$ aussi grand que l'on veut, on peut trouver une proximité $\delta(G) > 0$ vérifiant :

$$f(x) \ge G$$

pour tout les $x \in A$ assez proche de a :

$$\mathfrak{dist}(x, a) \le \delta(G)$$

on dit alors que f tend vers l'infini positif en a et on écrit :

$$\lim_{x\to a} f(x) = +\infty$$

Négative

Si, pour tout $P \in F$ aussi petit que l'on veut, on peut trouver une proximité $\delta(P) > 0$ vérifiant :

$$f(x) \le P$$

pour tout les $x \in A$ assez proche de a :

$$\mathfrak{dist}(x, a) \le \delta(P)$$

on dit alors que f tend vers l'infini négatif en a et on écrit :

$$\lim_{x\to a} f(x) = -\infty$$

Notation

On a les notations alternatives :

$$\begin{aligned} f(a) = +\infty \quad &\Leftrightarrow \quad \lim_{x\to a} f(x) = +\infty \\ f(a) = -\infty \quad &\Leftrightarrow \quad \lim_{x\to a} f(x) = -\infty \end{aligned}$$

30.6 Limite infinie à l'infini

On suppose que E et F sont ordonnés. On dit que la limite de f à l'infini positif est l'infini positif et on le note :

$$\lim_{x \to +\infty} f(x) = +\infty$$

si, pour tout $G \in F$ aussi grand que l'on veut, on peut trouver une borne inférieure $I(G) \in E$ telle que :

$$f(x) \ge G$$

pour tout $x \in E$ vérifiant $x \ge I(G)$. On dit que la limite de f à l'infini positif est l'infini négatif et on le note :

$$\lim_{x \to +\infty} f(x) = -\infty$$

si, pour tout $P \in F$ aussi petit que l'on veut, on peut trouver une borne inférieure $I(G) \in E$ telle que :

$$f(x) \le P$$

pour tout $x \in E$ vérifiant $x \ge I(G)$. On dit que la limite de f à l'infini négatif est l'infini positif et on le note :

$$\lim_{x \to -\infty} f(x) = +\infty$$

si, pour tout $G \in F$ aussi grand que l'on veut, on peut trouver une borne supérieure $S(G) \in E$ telle que :

$$f(x) \ge G$$

pour tout $x \in E$ vérifiant $x \le S(G)$. On dit que la limite de f à l'infini négatif est l'infini négatif et on le note :

$$\lim_{x \to -\infty} f(x) = -\infty$$

si, pour tout $P \in F$ aussi petit que l'on veut, on peut trouver une borne supérieure $S(G) \in E$ telle que :

$$f(x) \le P$$

pour tout $x \in E$ vérifiant $x \le S(G)$.

Chapitre 31

Limites doubles

Dépendances

— Chapitre 30 : Les limites

31.1 Introduction

Si les ensemble A, B sont munis d'une distance, on définit la distance :

$$\mathfrak{dist}^2 : (A \times B) \times (A \times B) \mapsto \mathbb{K}$$

associée par :

$$\mathfrak{dist}^2\big[(x,y),(a,b)\big] = \max\big\{\mathfrak{dist}(x,a), \mathfrak{dist}(y,b)\big\}$$

pour tout $(x,y),(a,b) \in A \times B$. La fonction définie est-elle une distance ? On a clairement $\mathfrak{dist}^2 \geq 0$ et :

$$\mathfrak{dist}^2\big[(x,y),(a,b)\big] = \mathfrak{dist}^2\big[(a,b),(x,y)\big]$$

On a aussi :

$$\mathfrak{dist}^2\big[(x,y),(x,y)\big] = \max\big\{\mathfrak{dist}(x,x), \mathfrak{dist}(y,y)\big\} = \max\{0,0\} = 0$$

Si :

$$\mathfrak{dist}^2\big[(x,y),(a,b)\big] = 0$$

on a forcément :

$$\mathfrak{dist}(x,a) = \mathfrak{dist}(y,b) = 0$$

Donc $x = a$, $y = b$ et :

$$(x,y) = (a,b)$$

Pour l'inégalité triangulaire, soit $(x,y),(a,b),(c,d) \in A \times B$ et :

$$d = \mathfrak{dist}^2\big[(a,b),(c,d)\big] = \max\big\{\mathfrak{dist}(a,c), \mathfrak{dist}(b,d)\big\}$$

La distance sur A vérifie :

$$\mathfrak{dist}(a,c) \leq \mathfrak{dist}(a,x) + \mathfrak{dist}(x,c)$$

La distance sur B vérifie :

$$\mathfrak{dist}(b, d) \le \mathfrak{dist}(b, y) + \mathfrak{dist}(y, d)$$

On en déduit que :

$$\begin{aligned} d &\le \max\{\mathfrak{dist}(a, x) + \mathfrak{dist}(x, c), \mathfrak{dist}(b, y) + \mathfrak{dist}(y, d)\} \\ &\le \max\{\mathfrak{dist}(a, x), \mathfrak{dist}(b, y)\} + \max\{\mathfrak{dist}(x, c), \mathfrak{dist}(y, d)\} \\ &\le \mathfrak{dist}^2\big[(a, b), (x, y)\big] + \mathfrak{dist}^2\big[(x, y), (c, d)\big] \end{aligned}$$

31.2 Limite en un point

Soit un ensemble F et une fonction $f : A \times B \mapsto F$. On choisit un ensemble $U \subseteq A \times B$. On dit que L est la limite de f en $(a, b) \in A \times B$ et on le note :

$$\lim_{\substack{(x,y)\to(a,b) \\ (x,y)\in U}} f(x, y) = L$$

si, pour toute précision $\epsilon > 0$, on peut trouver un $\delta > 0$ tel que :

$$\mathfrak{dist}(f(x, y), L) \le \epsilon$$

pour tout $(x, y) \in U$ vérifiant :

$$\mathfrak{dist}^2\big[(x, y), (a, b)\big] = \max\{\mathfrak{dist}(x, a), \mathfrak{dist}(y, b)\} \le \delta$$

31.2.1 Formulation équivalente

La condition :

$$\max\{\mathfrak{dist}(x, a), \mathfrak{dist}(y, b)\} \le \delta$$

est équivalente à :

$$\begin{aligned} \mathfrak{dist}(x, a) &\le \delta \\ \mathfrak{dist}(y, b) &\le \delta \end{aligned}$$

On en conclut que :

$$\lim_{\substack{(x,y)\to(a,b) \\ (x,y)\in U}} f(x, y) = L$$

si et seulement si, pour toute précision $\epsilon > 0$, on peut trouver un $\delta > 0$ tel que :

$$\mathfrak{dist}(f(x, y), L) \le \epsilon$$

pour tout $(x, y) \in U$ vérifiant :

$$\begin{aligned} \mathfrak{dist}(x, a) &\le \delta \\ \mathfrak{dist}(y, b) &\le \delta \end{aligned}$$

31.2.2 Extension

Supposons que pour toute précision $\epsilon > 0$ on puisse trouver des $\delta_1, \delta_2 > 0$ tels que :

$$\mathfrak{dist}(f(x,y), L) \le \epsilon$$

pour tout $(x,y) \in U$ vérifiant :

$$\begin{gathered}\mathfrak{dist}(x,a) \le \delta_1 \\ \mathfrak{dist}(y,b) \le \delta_2\end{gathered}$$

En posant :

$$\delta = \min\{\delta_1, \delta_2\} > 0$$

on voit que :

$$\mathfrak{dist}(f(x,y), L) \le \epsilon$$

pour tout $(x,y) \in U$ vérifiant :

$$\begin{gathered}\mathfrak{dist}(x,a) \le \delta \le \delta_1 \\ \mathfrak{dist}(y,b) \le \delta \le \delta_2\end{gathered}$$

On en conclut que :

$$\lim_{\substack{(x,y)\to(a,b) \\ (x,y)\in U}} f(x,y) = L$$

31.3 Limite à l'infini

On suppose que les ensembles A, B sont ordonnés. On dit que $L \in F$ est la limite de $f : A \times B \mapsto F$ à l'infini positif et on le note :

$$\lim_{(x,y)\to+\infty} f(x,y) = L$$

si, pour toute précision $\epsilon > 0$, on peut trouver des bornes inférieures $(I, J) \in A \times B$ telles que :

$$\mathfrak{dist}\big[f(x,y), L\big] \le \epsilon$$

pour tout $(x,y) \in A \times B$ vérifiant :

$$\begin{gathered}x \ge I \\ y \ge J\end{gathered}$$

On dit que $L \in F$ est la limite de $f : A \times B \mapsto F$ à l'infini négatif et on le note :

$$\lim_{(x,y)\to-\infty} f(x,y) = L$$

si, pour toute précision $\epsilon > 0$, on peut trouver des bornes supérieures $(I, J) \in A \times B$ telles que :

$$\mathfrak{dist}\big[f(x,y), L\big] \le \epsilon$$

pour tout $(x,y) \in A \times B$ vérifiant :

$$\begin{gathered}x \le I \\ y \le J\end{gathered}$$

31.4 Dualité

31.4.1 x puis y

Supposons que la limite $\lambda(y)$ définie par :

$$\lambda(y) = \lim_{x \to a} f(x, y)$$

existe pour tout $y \in B$ et que :

$$L = \lim_{y \to b} \lambda(y) = \lim_{y \to b} \lim_{x \to a} f(x, y)$$

est bien définie. Nous supposons également que, quelque soit $\epsilon > 0$, on peut trouver un $\delta > 0$ tel que :

$$\mathfrak{dist}(f(x, y), \lambda(y)) \le \epsilon$$

pour tout $(x, y) \in A \times B$ vérifiant :

$$\mathfrak{dist}(x, a) \le \delta$$

Choissons $\epsilon > 0$. On peut trouver $\delta_1 > 0$ tel que :

$$\mathfrak{dist}(\lambda(y), L) \le \frac{\epsilon}{2}$$

pour tout $(x, y) \in A \times B$ vérifiant :

$$\mathfrak{dist}(y, b) \le \delta_1$$

On peut aussi trouver δ_2 tel que :

$$\mathfrak{dist}(f(x, y), \lambda(y)) \le \frac{\epsilon}{2}$$

pour tout $(x, y) \in A \times B$ vérifiant :

$$\mathfrak{dist}(x, a) \le \delta_2$$

Soit $\delta = \min\{\delta_1, \delta_2\}$ et $(x, y) \in A \times B$ vérifiant :

$$\max\{\mathfrak{dist}(x, a), \mathfrak{dist}(y, b)\} \le \delta$$

On a alors :

$$\begin{aligned} \mathfrak{dist}(f(x, y), L) &\le \mathfrak{dist}(f(x, y), \lambda(y)) + \mathfrak{dist}(\lambda(y), L) \\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

On en conclut que :

$$\lim_{(x,y) \to (a,b)} f(x, y) = L$$

autrement dit :

$$\lim_{(x,y) \to (a,b)} f(x, y) = \lim_{y \to b} \lim_{x \to a} f(x, y)$$

31.4.2 y puis x

Supposons que la limite $\mu(x)$ définie par :

$$\mu(x) = \lim_{y \to b} f(x, y)$$

existe pour tout $x \in A$ et que :

$$M = \lim_{x \to a} \mu(y) = \lim_{x \to a} \lim_{y \to b} f(x, y)$$

est bien définie. Nous supposons également que, quelque soit $\epsilon > 0$, on peut trouver un $\delta > 0$ tel que :

$$\mathfrak{dist}(f(x, y), \mu(x)) \le \epsilon$$

pour tout $(x, y) \in A \times B$ vérifiant :

$$\mathfrak{dist}(y, b) \le \delta$$

Choissons $\epsilon > 0$. On peut trouver $\delta_1 > 0$ tel que :

$$\mathfrak{dist}(\mu(x), M) \le \frac{\epsilon}{2}$$

pour tout $(x, y) \in A \times B$ vérifiant :

$$\mathfrak{dist}(x, a) \le \delta_1$$

On peut aussi trouver δ_2 tel que :

$$\mathfrak{dist}(f(x, y), \mu(x)) \le \frac{\epsilon}{2}$$

pour tout $(x, y) \in A \times B$ vérifiant :

$$\mathfrak{dist}(y, b) \le \delta_2$$

Soit $\delta = \min\{\delta_1, \delta_2\}$ et $(x, y) \in A \times B$ vérifiant :

$$\max\{\mathfrak{dist}(x, a), \mathfrak{dist}(y, b)\} \le \delta$$

On a alors :

$$\begin{aligned} \mathfrak{dist}(f(x, y), M) &\le \mathfrak{dist}(f(x, y), \mu(x)) + \mathfrak{dist}(\mu(x), M) \\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

On en conclut que :

$$\lim_{(x,y) \to (a,b)} f(x, y) = M$$

autrement dit :

$$\lim_{(x,y) \to (a,b)} f(x, y) = \lim_{x \to a} \lim_{y \to b} f(x, y)$$

Chapitre 32

Suites

Dépendances

— Chapitre 30 : Les limites

32.1 Définition

Une suite est une fonction $s : \mathbb{N} \mapsto \Omega$ définie par :

$$s : n \mapsto s_n = s(n)$$

pour tout $n \in \mathbb{N}$. On note $\mathrm{Suite}(\Omega)$ l'ensemble des suites $s \subseteq \Omega$.

32.2 Limite

On dit qu'une suite $s : n \mapsto s_n \in \Omega$ converge vers $L \in \Omega$ au sens de la distance $\mathfrak{dist}$, ou que L est sa limite à l'infini :

$$L = \lim_{n \to \infty} s_n$$

si, pour toute précision $\epsilon > 0$ aussi petite que l'on veut, on peut trouver un naturel $K(\epsilon)$ à partir duquel l'erreur sera au moins aussi faible que demandée. On a donc :

$$\mathfrak{dist}(L, s_n) \le \epsilon$$

pour tout $n \ge K(\epsilon)$.

32.2.1 Notation

Comme la limite d'une suite est toujours sous-entendue vers l'infini, on note :

$$\lim_n s_n = \lim_{n \to \infty} s_n$$

32.3 Limites extrémales

On définit :

$$\limsup_{n \to \infty} s_n = \lim_{n \to \infty} \sup\{s_m : m \in \mathbb{N},\ m \ge n\}$$

$$\liminf_{n \to \infty} s_n = \lim_{n \to \infty} \inf\{s_m : m \in \mathbb{N},\ m \ge n\}$$

32.3.1 Notation

On note aussi :

$$\limsup_n s_n = \limsup_{n\to\infty} s_n$$

$$\liminf_n s_n = \liminf_{n\to\infty} s_n$$

32.4 Équivalence

Soit les suites :

$$s : n \mapsto s_n \in \Omega$$

et :

$$t : n \mapsto t_n \in \Omega$$

On dit que s est équivalente à t, et on le note :

$$s \equiv t$$

si et seulement si la limite de la distance entre les deux suites converge vers zéro :

$$\lim_{n\to\infty} \mathfrak{dist}(s_n, t_n) = 0$$

Quelque soit $\epsilon > 0$, on peut donc trouver un naturel $K(\epsilon)$ tel que :

$$\mathfrak{dist}(s_n, t_n) \le \epsilon$$

pour tout $n \ge K(\epsilon)$.

32.4.1 Remarque

L'équivalence entre s et t n'implique nullement que la limite de s ou de t existe.

32.4.2 Existence des limites

Suppons que $s \equiv t$ et que la limite :

$$\sigma = \lim_{n\to\infty} s_n$$

existe. On a :

$$\mathfrak{dist}(\sigma, t_n) \le \mathfrak{dist}(\sigma, s_n) + \mathfrak{dist}(s_n, t_n)$$

Soit $\epsilon > 0$. En choisissant K_1 tel que :

$$\mathfrak{dist}(\sigma, s_n) \le \frac{\epsilon}{2}$$

pour tout $n \ge K_1$ et K_2 tel que :

$$\mathfrak{dist}(s_n, t_n) \le \frac{\epsilon}{2}$$

pour tout $n \ge K_2$, on voit que :

$$\mathfrak{dist}(\sigma, t_n) \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Cette relation étant valable quel que soit $\epsilon > 0$, on en déduit que la limite des t_n existe et que :

$$\lim_{n \to \infty} t_n = \sigma = \lim_{n \to \infty} s_n$$

Les limites de suites équivalentes sont identiques.

32.5 Ordre

Soit les suites :

$$s : n \mapsto s_n \in \Omega$$

et :

$$t : n \mapsto t_n \in \Omega$$

Si un ordre est défini sur Ω, on dit que s est inférieure à t :

$$s \le t$$

si et seulement si les éléments de s sont inférieurs aux éléments de t :

$$s_n \le t_n$$

pour tout $n \in \mathbb{N}$.

32.6 Monotonie

32.6.1 Croissance

On dit qu'une suite :

$$s : n \mapsto s_n \in \Omega$$

est croissante si :

$$s_i \ge s_j$$

pour tout $i, j \in \mathbb{N}$ tels que $i \ge j$.

32.6.2 Décroissance

On dit qu'une suite :

$$s : n \mapsto s_n \in \Omega$$

est décroissante si :

$$s_i \le s_j$$

pour tout $i, j \in \mathbb{N}$ tels que $i \ge j$.

32.7 Opérations

Soit les suites :

$$s : n \mapsto s_n \in \Omega$$

et :

$$t : n \mapsto t_n \in \Omega$$

Pour toute opération $\divideontimes$ définie sur Ω, on définit l'opération induite $\divideontimes$: $\mathrm{Suite}(\Omega) \times \mathrm{Suite}(\Omega) \mapsto \mathrm{Suite}(\Omega)$ par :

$$(s \divideontimes t)(n) = s_n \divideontimes t_n$$

pour tout $n \in \mathbb{N}$. On note aussi :

$$(s \divideontimes t)_n = (s \divideontimes t)(n)$$

32.7.1 Usuelles

Sur les ensembles où sont définies les opérations usuelles, on aura l'addition :

$$(s + t)_n = s_n + t_n$$

la multiplication :

$$(s - t)_n = s_n - t_n$$

la soustraction :

$$(s \cdot t)_n = s_n \cdot t_n$$

la division :

$$\left[\frac{s}{t}\right]_n = \frac{s_n}{t_n}$$

pour tout $n \in \mathbb{N}$.

32.8 Cauchy

On dit qu'une suite :

$$s : n \mapsto s_n \in \Omega$$

est de Cauchy si, pour toute précision $\epsilon > 0$ aussi petite que l'on veut, on peut trouver un naturel $K(\epsilon)$ à partir duquel la distance entre deux éléments de la suite s_m, s_n sera aussi petite que demandée. On a donc :

$$\mathfrak{dist}(s_m, s_n) \le \epsilon$$

pour tout $m, n \ge K(\epsilon)$.

Suite convergente

Toute suite convergente vers une limite :

$$L = \lim_{n \to \infty} s_n \in \Omega$$

est de Cauchy. En effet, soit $\epsilon > 0$. Si on choisit K tel que :

$$\mathfrak{dist}(L, s_n) \le \epsilon/2$$

pour tout $n \ge K$, on a :

$$\mathfrak{dist}(s_m, s_n) \le \mathfrak{dist}(s_m, L) + \mathfrak{dist}(L, s_n) = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

pour tout $m, n \ge K$.

32.9 Ensemble complet

On dit qu'un ensemble X est complet si toute suite de Cauchy inclue dans X converge vers une limite $L \in X$.

Complétion

On peut compléter tout ensemble A incomplet en créant un ensemble X tel que tout $x \in X$ soit associé à une suite de Cauchy :

$$s : n \mapsto s_n \in A$$

On note alors symboliquement :

$$x = \lim_{n \to \infty} s_n$$

Chapitre 33

Sommes abstraites

Dépendances

— Chapitre 10 : Algèbre

33.1 Introduction

Soit le corps commutatif $\mathbb{K}$, un ensemble Ω et la fonction $f : \Omega \mapsto \mathbb{K}$. On note la somme de f sur X par :

$$\sum_{x \in X} f(x)$$

pour tout sous-ensemble $X \subseteq \Omega$. Il s'agit intuitivement de la somme des $f(x) \in \mathbb{K}$ lorsque x parcourt X. Nous allons voir comment la formaliser.

33.1.1 Notation

Lorsque l'ensemble X est évident d'après le contexte, on convient que :

$$\sum_{x} f(x) = \sum_{x \in X} f(x)$$

33.2 Additivité

Si deux ensembles $X, Y \subseteq \Omega$ ne se chevauchent pas :

$$X \cap Y = \emptyset$$

la somme sur l'union des deux est intuitivement l'addition des sommes sur chacun d'entre-eux :

$$\sum_{z \in X \cup Y} f(z) = \sum_{z \in X} f(z) + \sum_{z \in Y} f(z)$$

33.3 Ensemble vide

Comme $X = X \cup \emptyset$ et $X \cap \emptyset = \emptyset$, on en déduit que :

$$\sum_{x \in X} f(x) = \sum_{x \in X \cup \emptyset} f(x) = \sum_{x \in X} f(x) + \sum_{x \in \emptyset} f(x)$$

La somme sur l'ensemble vide doit donc être le neutre pour l'addition :

$$\sum_{x \in \emptyset} f(x) = 0$$

33.4 Singleton

Il semble également logique d'imposer que la somme sur un ensemble contenant un seul élément $a \in X$ soit égal à $f(a)$:

$$\sum_{x \in \{a\}} f(x) = f(a)$$

Voilà qui complète les caractéristiques génériques des sommes.

33.5 Somme des éléments d'un ensemble

Pour tout $A \subseteq \mathbb{K}$, on note :

$$\sum_{x \in A} x = \sum_{x \in A} \mathrm{Id}(x)$$

la somme des éléments de A.

33.6 Algorithme

Soit l'ensemble X non vide, ensemble dont nous voulons évaluer la somme. Choisissons un élément $a \in A$ et considérons la décomposition :

$$X = \{a\} \cup (X \setminus \{a\})$$

Comme l'intersection des deux ensembles du membre de droite est vide :

$$\{a\} \cap (X \setminus \{a\}) = \emptyset$$

on peut écrire :

$$\sum_{x \in X} f(x) = \sum_{x \in \{a\}} f(x) + \sum_{x \in X \setminus \{a\}} f(x)$$

c'est-à-dire :

$$\sum_{x \in X} f(x) = f(a) + \sum_{x \in X \setminus \{a\}} f(x)$$

On en déduit un algorithme itératif permettant d'estimer la somme :

$$S \approx \sum_{x \in X} f(x)$$

Nous partons de :

$$S_0 = 0$$
$$X_0 = X$$

A chaque étape $k \in \mathbb{N}$, nous choisissons $a_k \in X_k$ et nous adaptons notre estimation par :

$$S_{k+1} = f(a_k) + S_k$$

On retire ensuite a_k de X_k pour éviter de le compter plus d'une fois, ce qui nous donne l'ensemble suivant :

$$X_{k+1} = X_k \setminus \{a_k\}$$

Cet algorithme va nous permettre de formaliser la définition des sommes.

33.7 Ensemble fini

Soit l'ensemble X contenant un nombre fini $N \in \mathbb{N}$ d'éléments :

$$X = \{a_1, a_2, ..., a_N\}$$

En appliquant l'algorithme d'évaluation d'une somme, on finit par arriver à l'itération N avec :

$$X_N = \emptyset$$

On a simplement :

$$\sum_{x \in X} f(x) = S_N + \sum_{x \in \emptyset} f(x) = S_N + 0 = S_N$$

Comme :

$$S_N = f(a_1) + f(a_2) + ... + f(a_N)$$

on a simplement :

$$\sum_{x \in X} f(x) = f(a_1) + f(a_2) + ... + f(a_N)$$

On note :

$$\sum_{k=1}^{N} f(a_k) = f(a_1) + f(a_2) + ... + f(a_N)$$

33.7.1 Numérotation

Les éléments de X peuvent être numérotés différemment. Soit $m, n \in \mathbb{Z}$ et :

$$X = \{a_m, a_{m+1}, ..., a_{n-1}, a_n\}$$

On a alors :

$$\sum_{x \in X} f(x) = f(a_m) + f(a_{m+1}) + ... + f(a_{n-1}) + f(a_n)$$

On note :

$$\sum_{k=m}^{n} f(a_k) = f(a_m) + f(a_{m+1}) + ... + f(a_{n-1}) + f(a_n)$$

33.7.2 Extension

Soit un ensemble X dont on peut extraire un sous-ensemble fini de la forme :

$$F = \{a_m, a_{m+1}, ..., a_{n-1}, a_n\} \subseteq X$$

tel que :

$$f(x) = 0$$

pour tout $x \in X \setminus F$. La somme s'écrit alors :

$$\sum_{x \in X} f(x) = \sum_{x \in F} f(x) = \sum_{k=m}^{n} f(a_k)$$

33.8 Ensemble dénombrable

33.8.1 Naturels

Soit :

$$X = \{a_0, a_1, a_2, ...\} = \{a_k : k \in \mathbb{N}\}$$

Si la suite $n \mapsto S_n$ définie par :

$$S_n = \sum_{k=0}^{n} f(a_k) = f(a_0) + f(a_1) + ... + f(a_n)$$

pour tout $n \in \mathbb{N}$ converge vers un certain $S \in \mathbb{K}$, on définit :

$$\sum_{x \in X} f(x) = \lim_{n \to \infty} S_n = S$$

c'est-à-dire :

$$\sum_{x \in X} f(x) = \lim_{n \to \infty} \sum_{k=0}^{n} f(a_k)$$

On introduit la notation :

$$\sum_{k=0}^{+\infty} f(a_k) = \lim_{n \to \infty} \sum_{k=0}^{n} f(a_k)$$

33.8.2 Entiers

Soit :

$$X = \{..., a_{-2}, a_{-1}, a_0, a_1, a_2, ...\} = \{a_k : k \in \mathbb{Z}\}$$

Si la suite des $n \mapsto S_n$ définie par :

$$S_n = \sum_{k=-n}^{n} f(a_k) = f(a_{-n}) + f(a_{-n+1}) + ... + f(a_{n-1}) + f(a_n)$$

pour tout $n \in \mathbb{N}$ converge vers un certain $S \in \mathbb{K}$, on définit :

$$\sum_{x \in X} f(x) = \lim_{n \to \infty} S_n = S$$

On introduit la notation :

$$\sum_{k=-\infty}^{+\infty} f(a_k) = \lim_{n \to \infty} \sum_{k=-n}^{n} f(a_k)$$

33.8.3 Extension

Soit un ensemble X dont on peut extraire un sous-ensemble D de la forme :

$$D = \{a_k : k \in \mathbb{N}\} \subseteq X$$

ou :

$$D = \{a_k : k \in \mathbb{Z}\} \subseteq X$$

tel que :

$$f(x) = 0$$

pour tout $x \in X \setminus D$. La somme s'écrit alors :

$$\sum_{x \in X} f(x) = \sum_{x \in D} f(x)$$

33.9 Linéarité

Soit les fonctions $f, g : \Omega \mapsto \mathbb{K}$ et l'ensemble fini $X \subseteq \Omega$. On a clairement :

$$\sum_{x \in X} \Big[f(x) + g(x) \Big] = \sum_{x \in X} f(x) + \sum_{x \in X} g(x)$$

Comme le produit se distribue sur l'addition, on a également :

$$\sum_{x \in X} \Big[c \cdot f(x) \Big] = c \cdot \sum_{x \in X} f(x)$$

pour tout $c \in \mathbb{K}$. La somme sur un ensemble fini est linéaire.

33.10 Additivité

Si les ensembles finis X et Y vérifient $X \cap Y = \emptyset$, la commutativité et l'associativité de l'addition nous donnent la propriété d'additivité :

$$\sum_{x \in X \cup Y} f(x) = \sum_{x \in X} f(x) + \sum_{x \in Y} f(x)$$

33.11 Additivité généralisée

Soit les ensembles Ω et Λ comportant un nombre fini d'éléments, la fonction $f : \Omega \mapsto \mathbb{K}$ et la collection d'ensembles :

$$\Theta = \{X(\lambda) \subseteq \Omega : \lambda \in \Lambda\}$$

où $X : \Lambda \mapsto \mathfrak{P}(\Omega)$. On définit la fonction $S : \Lambda \mapsto \mathbb{K}$ représentant les sommes associées par :

$$S(\lambda) = \sum_{x \in X(\lambda)} f(x)$$

On suppose que Θ forme une partition de Ω. On a alors :

$$X(\lambda) \cap X(\mu) = \emptyset$$

pour tout $\lambda, \mu \in \Lambda$ tels que $\lambda \neq \mu$ et :

$$\Omega = \bigcup_{\lambda \in \Lambda} X(\lambda)$$

Alors, l'associativité et la commutativité de l'addition nous permettent de regrouper les termes de Ω par sous-ensembles $X(\lambda)$, et on a :

$$\sum_{x \in \Omega} f(x) = \sum_{\lambda \in \Lambda} S(\lambda) = \sum_{\lambda \in \Lambda} \left[\sum_{x \in X(\lambda)} f(x) \right]$$

33.11.1 Notation

On note aussi :

$$\sum_{\lambda \in \Lambda} \sum_{x \in X(\lambda)} f(x) = \sum_{\lambda \in \Lambda} \left[\sum_{x \in X(\lambda)} f(x) \right]$$

33.12 Produit cartésien

Soit les sous-ensembles $X, Y \subseteq \Omega$ comportant un nombre fini d'éléments et la fonction $f : X \times Y \mapsto \mathbb{K}$. On définit la fonction $A : X \mapsto \mathfrak{P}(X \times Y)$ par :

$$A(x) = \{(x, y) : y \in Y\}$$

pour tout $x \in X$. Les sommes associées s'écrivent :

$$S(x) = \sum_{(\lambda, y) \in A(x)} f(\lambda, y)$$

Comme les éléments de $A(x)$ sont de la forme (x, y), on a forcément $\lambda = x$. Par conséquent, parcourir $A(x)$ revient à parcourir les $y \in Y$ en gardant $\lambda = x$ fixé et on a :

$$S(x) = \sum_{y \in Y} f(x, y)$$

On se rend compte que les ensembles de cette collection ne se chevauchent pas :

$$A(\lambda) \cap A(\mu) = \emptyset$$

pour tout $\lambda, \mu \in X$ tels que $\lambda \neq \mu$. On a aussi :

$$X \times Y = \bigcup_{\lambda \in X} A(\lambda)$$

Nous pouvons par conséquent utiliser l'additivité généralisée, ce qui nous donne :

$$\sum_{(x,y) \in X \times Y} f(x,y) = \sum_{x \in X} S(x)$$

soit, en tenant compte de l'expression de $S(x)$:

$$\sum_{(x,y) \in X \times Y} f(x,y) = \sum_{x \in X} \sum_{y \in Y} f(x,y)$$

On montre également que :

$$\sum_{(x,y) \in X \times Y} f(x,y) = \sum_{y \in Y} \sum_{x \in X} f(x,y)$$

33.13 Somme d'un produit

Soit les fonctions $f, g : \Omega \mapsto \mathbb{K}$ et les sous-ensembles $X, Y \subseteq \Omega$ comportant un nombre fini d'éléments. On a :

$$\sum_{(x,y) \in X \times Y} f(x) \cdot g(y) = \sum_{x \in X} \sum_{y \in Y} f(x) \cdot g(y)$$

Mais comme la valeur de $f(x)$ ne dépend pas de y, on peut appliquer la distributivité du produit sur l'addition pour faire sortir les valeurs de f de la somme sur y et :

$$\sum_{(x,y) \in X \times Y} f(x) \cdot g(y) = \sum_{x \in X} \left[f(x) \cdot \sum_{y \in Y} g(y) \right]$$

A nouveau, comme la somme de g sur Y ne dépend pas de x, on peut la faire sortir et :

$$\sum_{(x,y) \in X \times Y} f(x) \cdot g(y) = \left[\sum_{y \in Y} g(y) \right] \cdot \left[\sum_{x \in X} f(x) \right]$$

La multiplication étant commutative, on a également :

$$\sum_{(x,y) \in X \times Y} f(x) \cdot g(y) = \left[\sum_{x \in X} f(x) \right] \cdot \left[\sum_{y \in Y} g(y) \right]$$

Chapitre 34

Sommes indicées

34.1 Définition

Soit un corps commutatif $\mathbb{K}$ et l'ensemble d'indices $\mathcal{Z}$. Supposons que l'on puisse écrire l'ensemble $A \subseteq \mathbb{K}$ sous la forme :

$$A = \{a_k : k \in \mathcal{Z}\}$$

On associe à A la fonction $\varphi : \mathcal{Z} \mapsto A$ définie par :

$$\varphi(k) = a_k$$

pour tout $k \in \mathcal{Z}$. Pour tout sous-ensemble $I \subseteq \mathcal{Z}$, on définit alors :

$$\sum_{k \in I} a_k = \sum_{k \in I} \varphi(k)$$

sous réserve d'existence de la somme.

34.1.1 Fonction

Soit $f : A \mapsto \mathbb{K}$. On définit :

$$\sum_{k \in I} f(a_k) = \sum_{k \in I} (f \circ \varphi)(k)$$

34.2 Intervalles discrets

Soit $m, n \in \mathbb{Z}$. Nous définissons l'intervalle discret :

$$\mathbb{Z}[m, n] = \{z \in \mathbb{Z} : m \le z \le n\}$$

Si $\mathbb{Z}[m, n] \subseteq \mathcal{Z}$, on a simplement :

$$\sum_{k \in \mathbb{Z}[m,n]} a_k = a_m + a_{m+1} + ... + a_{n-1} + a_n$$

On note aussi :

$$\sum_{k=m}^{n} a_k = \sum_{k=n}^{m} a_k = \sum_{k \in \mathbb{Z}[m,n]} a_k$$

34.3 Linéarité

Soit les ensembles :

$$A = \{a_k : k \in \mathcal{Z}\} \subseteq \mathbb{K}$$
$$B = \{b_k : k \in \mathcal{Z}\} \subseteq \mathbb{K}$$

et $\alpha, \beta \in \mathbb{K}$. Si $I \subseteq \mathcal{Z}$ compte un nombre fini d'éléments, on a clairement :

$$\sum_{k \in I} (\alpha \cdot a_k + \beta \cdot b_k) = \alpha \cdot \sum_{k \in I} a_k + \beta \cdot \sum_{k \in I} b_k$$

34.4 Produit cartésien

Soit les sous-ensembles $I, J \subseteq \mathbb{Z}$ comportant un nombre fini d'éléments, et :

$$A = \{a_{ij} \in E : (i,j) \in I \times J\}$$

Il découle directement de la formule des sommes sur les produits cartésiens que :

$$\sum_{(i,j) \in I \times J} a_{ij} = \sum_{i \in I} \sum_{j \in J} a_{ij} = \sum_{j \in J} \sum_{i \in I} a_{ij}$$

34.4.1 Notation

On note :

$$\sum_{i,j=m}^{n} a_{ij} = \sum_{i=m}^{n} \sum_{j=m}^{n} a_{ij}$$

34.5 Somme d'un produit

Soit les ensembles :

$$A = \{a_k : k \in \mathcal{Z}\} \subseteq \mathbb{K}$$
$$B = \{b_k : k \in \mathcal{Z}\} \subseteq \mathbb{K}$$

et les ensembles $I, J \subseteq \mathcal{Z}$ comportant un nombre fini d'éléments. La somme du produit se déduit du résultat analogue des sommes génériques :

$$\sum_{(i,j) \in I \times J} a_i \cdot b_j = \left[\sum_{i \in I} a_i\right] \cdot \left[\sum_{j \in J} b_j\right]$$

34.6 Lemme du triangle

Soit le triangle discret :

$$\Delta = \{(i,j) \in \mathbb{Z}[0,N] \times \mathbb{Z}[0,N] : \ j \le i\}$$

et un ensemble :

$$A = \{a_{ij} : (i,j) \in \Delta\}$$

Le triangle Δ peut être aussi défini par :

$$\Delta = \{(i,j) \in \mathbb{Z}^2 : 0 \le i \le N, \quad 0 \le j \le i\}$$

La somme sur Δ peut donc se réécrire :

$$\sum_{(i,j)\in\Delta} a_{ij} = \sum_{i=0}^{N} \sum_{j=0}^{i} a_{ij}$$

Une autre définition alternative de Δ nous donne :

$$\Delta = \{(i,j) \in \mathbb{Z}^2 : 0 \le j \le N, \quad j \le i \le N\}$$

On a donc également :

$$\sum_{(i,j)\in\Delta} a_{ij} = \sum_{j=0}^{N} \sum_{i=j}^{N} a_{ij}$$

On en déduit une relation permettant d'inverser les sommes :

$$\sum_{i=0}^{N} \sum_{j=0}^{i} a_{ij} = \sum_{j=0}^{N} \sum_{i=j}^{N} a_{ij} = \sum_{(i,j)\in\Delta} a_{ij}$$

Chapitre 35

Progressions

35.1 Arithmétique

Soit $n \in \mathbb{N}$. Nous allons tenter d'évaluer la somme :

$$S_n = \sum_{i=0}^{n} i = 0 + 1 + 2 + ... + n$$

L'idée est d'exprimer que cette somme est équivalente à :

$$(n - 0) + (n - 1) + ... + (n - n)$$

Si nous posons j tel que $i = n - j$, on voit que l'on a $j = n - i$ et $0 \le j \le n$. Donc :

$$S_n = \sum_{j=0}^{n} (n - j)$$

Développons :

$$\sum_{j=0}^{n} (n - j) = \sum_{j=0}^{n} n - \sum_{j=0}^{n} j$$

Le premier terme du membre de droite peut se réécrire :

$$\sum_{j=0}^{n} n = n \cdot \sum_{j=0}^{n} 1 = n \cdot (n + 1)$$

Pour le second, on a clairement :

$$\sum_{j=0}^{n} j = S_n$$

On en conclut que :

$$S_n = n \cdot (n + 1) - S_n$$

c'est-à-dire :

$$2S_n = n \cdot (n + 1)$$

Divisant par 2, on obtient le résultat final :

$$\sum_{i=0}^{n} i = \frac{n \cdot (n + 1)}{2}$$

35.2 Géométrique

Soit $n \in \mathbb{N}$ et $a \in \mathbb{K}$. Nous allons rechercher une expression de la somme :

$$G_n = \sum_{i=0}^{n} a^i = 1 + a + a^2 + ... + a^n$$

Si $a = 1$, on a simplement :

$$G_n = \sum_{i=0}^{n} a^i = 1 + ... + 1 = n$$

Intéressons-nous à présent au cas où $a \neq 1$. On part du constat que la somme G_n est équivalente à :

$$1 + a \cdot (1 + a + a^2 + ... + a^n) - a^{n+1}$$

On développe en ce sens :

$$\begin{aligned} G_n &= 1 + \sum_{i=1}^{n} a^i \\ &= 1 + a \cdot \sum_{i=0}^{n-1} a^i \\ &= 1 + a \cdot \sum_{i=0}^{n} a^i - a^{n+1} \end{aligned}$$

et finalement, on arrive à l'équation implicite :

$$G_n = 1 + a \cdot G_n - a^{n+1}$$

En soustrayant $a \cdot G_n$ des deux membres, on obtient :

$$(1 - a) \cdot G_n = 1 - a^{n+1}$$

Comme $a \neq 1$, on en déduit que :

$$\sum_{i=0}^{n} a^i = \frac{1 - a^{n+1}}{1 - a}$$

35.2.1 Autre forme

En multipliant numérateur et dénominateur par -1, on obtient la forme équivalente :

$$\sum_{i=0}^{n} a^i = \frac{a^{n+1} - 1}{a - 1}$$

35.3 Factorisation

Les progressions géométriques permettent d'obtenir une importante formule de factorisation. Soit $a, b \in \mathbb{K}$ avec $a \neq 0$. Posons :

$$r = \frac{b}{a}$$

On a alors :

$$(1 - r) \cdot \sum_{i=0}^{n} r^i = 1 - r^{n+1}$$

Multipliant par a^{n+1}, on obtient :

$$(a - b) \cdot a^n \cdot \sum_{i=0}^{n} \frac{b^i}{a^i} = a^{n+1} - b^{n+1}$$

En faisant rentrer le facteur a^n dans la somme, on a en définitive :

$$a^{n+1} - b^{n+1} = (a - b) \cdot \sum_{i=0}^{n} a^{n-i} \cdot b^i$$

35.3.1 Extension

Si $a = 0$, on a :

$$a^{n+1} - b^{n+1} = -b^{n+1}$$

et :

$$\begin{aligned}
(a - b) \cdot \sum_{i=0}^{n} a^{n-i} \cdot b^i &= -b \cdot (0^n + 0^{n-1} \cdot b + ... + 0 \cdot b^{n-1} + b^n) \\
&= -b \cdot b^n = -b^{n+1}
\end{aligned}$$

La formule de factorisation est donc valable pour tout $a, b \in \mathbb{K}$.

35.3.2 Exemples

Voici quelques exemples d'applications :

$$\begin{aligned}
a^2 - b^2 &= (a - b) \cdot (a + b) \\
a^3 - b^3 &= (a - b) \cdot (a^2 + a \cdot b + b^2)
\end{aligned}$$

35.3.3 Symétrie

On a clairement :

$$\sum_{i=0}^{n} a^{n-i} \cdot b^i = a^n + a^{n-1} \cdot b + ... + a \cdot b^{n-1} + b^n = \sum_{i=0}^{n} a^i \cdot b^{n-i}$$

et donc :

$$a^{n+1} - b^{n+1} = (a - b) \cdot \sum_{i=0}^{n} a^i \cdot b^{n-i}$$

35.4 Somme des carrés

Considérons la somme :

$$C_n = \sum_{i=0}^{n} i^2 = 1 + 4 + 16 + ... + n^2$$

La progression arythmétique nous dit que :

$$\frac{1}{2} \cdot i \cdot (i+1) = \sum_{j=0}^{i} j$$

On a par conséquent :

$$\sum_{i=0}^{n} \frac{1}{2} \cdot i \cdot (i+1) = \sum_{i=0}^{n} \sum_{j=0}^{i} j$$

On peut utiliser le lemme du triangle pour inverser les deux sommes du membre de droite :

$$\sum_{i=0}^{n} \sum_{j=0}^{i} j = \sum_{j=0}^{n} \sum_{i=j}^{n} j$$

Comme j ne dépend pas de i, on peut le faire sortir de la somme sur i et le membre de droite devient :

$$\sum_{j=0}^{n} \sum_{i=j}^{n} j = \sum_{j=0}^{n} j \sum_{i=j}^{n} 1 = \sum_{j=0}^{n} j \cdot (n - j + 1)$$

En tenant compte de la linéarité des sommes, on a alors :

$$\begin{aligned} \sum_{j=0}^{n} j \cdot (n - j + 1) &= (n+1) \cdot \sum_{j=0}^{n} j - \sum_{j=0}^{n} j^2 \\ &= \frac{1}{2} \cdot n \cdot (n+1)^2 - C_n \end{aligned}$$

D'un autre côté, on a :

$$\sum_{i=0}^{n} \frac{1}{2} \cdot i \cdot (i+1) = \frac{1}{2} \sum_{i=0}^{n} (i^2 + i) = \frac{1}{2} \cdot C_n + \frac{1}{4} \cdot n \cdot (n+1)$$

En égalisant ces deux expressions, on obtient :

$$\frac{3}{2} \cdot C_n = \frac{1}{4} \cdot \Big[2 \cdot n \cdot (n+1)^2 - n \cdot (n+1) \Big]$$

Ou encore :

$$\begin{aligned} 6 \cdot C_n &= (2 \cdot n \cdot (n+1) - n) \cdot (n+1) \\ &= (2 \cdot n^2 + n) \cdot (n+1) \end{aligned}$$

On a donc la formule permettant d'évaluer la somme des carrés :

$$\sum_{i=0}^{n} i^2 = \frac{(2 \cdot n^2 + n) \cdot (n+1)}{6}$$

Chapitre 36

Différences

36.1 Définition

Etant donné une suite :

$$A = \{a_k : k \in \mathbb{N}\} \subseteq \mathbb{K}$$

on définit l'opérateur des différences $\boldsymbol{\Delta}$ par :

$$\boldsymbol{\Delta} a_k = a_{k+1} - a_k$$

36.2 Addition

Soit les suites :

$$A = \{a_k : k \in \mathbb{N}\} \subseteq \mathbb{K}$$

et :

$$B = \{b_k : k \in \mathbb{N}\} \subseteq \mathbb{K}$$

La différence de l'addition vérifie :

$$\boldsymbol{\Delta}(a_k + b_k) = a_{k+1} + b_{k+1} - a_k - b_k = \boldsymbol{\Delta} a_k + \boldsymbol{\Delta} b_k$$

36.3 Multiplication

Soit les suites :

$$A = \{a_k : k \in \mathbb{N}\} \subseteq \mathbb{K}$$

et :

$$B = \{b_k : k \in \mathbb{N}\} \subseteq \mathbb{K}$$

La différence de la multiplication vérifie :

$$\boldsymbol{\Delta}(a_k \cdot b_k) = a_{k+1} \cdot b_{k+1} - a_k \cdot b_k$$

Ajoutons et soustrayons le terme hybride $a_{k+1} \cdot b_k$. On a :

$$\boldsymbol{\Delta}(a_k \cdot b_k) = a_{k+1} \cdot b_{k+1} - a_{k+1} \cdot b_k + a_{k+1} \cdot b_k - a_k \cdot b_k$$

et :

$$\boldsymbol{\Delta}(a_k \cdot b_k) = a_{k+1} \cdot \boldsymbol{\Delta} b_k + \boldsymbol{\Delta} a_k \cdot b_k$$

Ajoutons et soustrayons le terme hybride $a_k \cdot b_{k+1}$. On a :

$$\boldsymbol{\Delta}(a_k \cdot b_k) = a_{k+1} \cdot b_{k+1} - a_k \cdot b_{k+1} + a_k \cdot b_{k+1} - a_k \cdot b_k$$

et :

$$\boldsymbol{\Delta}(a_k \cdot b_k) = \boldsymbol{\Delta} a_k \cdot b_{k+1} + a_k \cdot \boldsymbol{\Delta} b_k$$

36.4 Somme et différence

On définit également :

$$\boldsymbol{\Delta} \sum_{k=0}^{n} a_k = \sum_{k=1}^{n+1} a_k - \sum_{k=0}^{n} a_k$$

La définition nous donne directement :

$$\boldsymbol{\Delta} \sum_{k=0}^{n} a_k = (a_1 + ... + a_{n+1}) - (a_0 + ... + a_n)$$

Tous les termes se neutralisant sauf a_0 et a_{n+1}, on a :

$$\boldsymbol{\Delta} \sum_{k=0}^{n} a_k = a_{n+1} - a_0$$

On voit aussi que :

$$\sum_{k=0}^{n} \boldsymbol{\Delta} a_k = (a_{n+1} - a_n) + ... + (a_2 - a_1) + (a_1 - a_0)$$

Tous les termes se neutralisant mutuellement sauf le permier et le dernier, on a :

$$\sum_{k=0}^{n} \boldsymbol{\Delta} a_k = a_{n+1} - a_0$$

Ce résultat étant identique au précédent, on a :

$$\boldsymbol{\Delta} \sum_{k=0}^{n} a_k = \sum_{k=0}^{n} \boldsymbol{\Delta} a_k = a_{n+1} - a_0$$

36.5 Sommation par parties

Soit les suites :

$$A = \{a_k : k \in \mathbb{N}\} \subseteq \mathbb{K}$$

et :

$$B = \{b_k : k \in \mathbb{N}\} \subseteq \mathbb{K}$$

On a :

$$\sum_{k=0}^{n} \boldsymbol{\Delta}(a_k \cdot b_k) = a_{n+1} \cdot b_{n+1} - a_0 \cdot b_0$$

En utilisant la loi de différence d'une multiplication, on obtient parallèlement :

$$\sum_{k=0}^{n} \boldsymbol{\Delta}(a_k \cdot b_k) = \sum_{k=0}^{n} \boldsymbol{\Delta} a_k \cdot b_{k+1} + \sum_{k=0}^{n} a_k \cdot \boldsymbol{\Delta} b_k$$

On a donc :

$$\sum_{k=0}^{n} a_k \cdot \boldsymbol{\Delta} b_k = \sum_{k=0}^{n} \boldsymbol{\Delta}(a_k \cdot b_k) - \sum_{k=0}^{n} \boldsymbol{\Delta} a_k \cdot b_{k+1}$$

c'est-à-dire :

$$\sum_{k=0}^{n} a_k \cdot \boldsymbol{\Delta} b_k = a_{n+1} \cdot b_{n+1} - a_0 \cdot b_0 - \sum_{k=0}^{n} \boldsymbol{\Delta} a_k \cdot b_{k+1}$$

Chapitre 37

Suites de rationnels

Dépendances

— Chapitre 25 : Les rationnels
— Chapitre 29 : Les distances
— Chapitre 30 : Les limites

37.1 Définition

Une suite de rationnels, ou suite rationnelle est une suite $s : \mathbb{N} \mapsto \mathbb{Q}$:

$$s : n \mapsto s_n$$

37.2 Distance

On définit une distance sur l'ensemble des rationnels $\mathbb{Q}$ par :

$$\mathfrak{dist}(x, y) = |x - y|$$

pour tout $x, y \in \mathbb{Q}$. On a bien $\mathfrak{dist}(x, y) \geq 0$. La condition $\mathfrak{dist}(x, y) = 0$ implique que $|x - y|$, donc $x - y = 0$ et $x = y$. Enfin :

$$\begin{aligned} \mathfrak{dist}(x, z) &= |x - z| \\ &\leq |x - y| + |y - z| \\ &\leq \mathfrak{dist}(x, y) + \mathfrak{dist}(y, z) \end{aligned}$$

37.3 Équivalence

Soit les suites rationnelles $s, t : \mathbb{N} \mapsto \mathbb{Q}$ vérifiant $s \equiv t$. On a :

$$\lim_{n \to \infty} \mathfrak{dist}(s_n, t_n) = \lim_{n \to \infty} |s_n - t_n| = 0$$

On en déduit que :

$$\lim_{n \to \infty} (s_n - t_n) = 0$$

37.3.1 Réciproque

Supposons que :

$$\lim_{n\to\infty} (s_n - t_n) = 0$$

Choisissons $\epsilon > 0$ et le naturel $K(\epsilon)$ tel que :

$$\mathfrak{dist}(s_n - t_n, 0) \le \epsilon$$

pour tout $k \ge K(\epsilon)$. Par définition de la distance entre rationnels, on a :

$$\mathfrak{dist}(s_n - t_n, 0) = |(s_n - t_n) - 0| = |s_n - t_n|$$

on en conclut que :

$$\lim_{n\to\infty} \mathfrak{dist}(s_n, t_n) = \lim_{n\to\infty} |s_n - t_n| = 0$$

c'est-à-dire $s \equiv t$ par définition.

37.4 Cauchy

37.4.1 Croissante

Soit une suite de Cauchy $u : \mathbb{N} \mapsto \mathbb{Q}$:

$$u : n \mapsto u_n$$

croissante :

$$u_0 \le u_1 \le ... \le u_k \le ...$$

Comme :

$$u_n \ge u_0$$

pour tout $n \in \mathbb{N}$, on voit que la suite est minorée par u_0. On peut trouver un naturel K tel que :

$$\mathfrak{dist}(u_m, u_n) = |u_m - u_n| \le 1$$

pour tout naturels $m, n \ge K$. Soit $n \in \mathbb{N}$ vérifiant $n \ge K$. On a :

$$|u_n - u_K| = u_n - u_K \le 1$$

d'où :

$$u_n \le u_K + 1$$

Pour tout naturel $k \le K$, on a :

$$u_k \le u_n \le u_K + 1$$

En posant :

$$S = u_K + 1$$

on voit que :

$$u_n \le S$$

pour tout $n \in \mathbb{N}$. Une suite de Cauchy croissante est également majorée.

37.4.2 Décroissante

Soit une suite de Cauchy $u : \mathbb{N} \mapsto \mathbb{Q}$:

$$u : n \mapsto u_n$$

décroissante :

$$u_0 \ge u_1 \ge ... \ge u_k \ge ...$$

Comme :

$$u_n \le u_0$$

pour tout $n \in \mathbb{N}$, on voit que la suite est majorée par u_0. On peut trouver un naturel K tel que :

$$\mathfrak{dist}(u_m, u_n) = |u_m - u_n| \le 1$$

pour tout naturels $m, n \ge K$. Soit $n \in \mathbb{N}$ vérifiant $n \ge K$. On a :

$$|u_n - u_K| = u_K - u_n \le 1$$

d'où :

$$u_n \ge u_K - 1$$

Pour tout naturel $k \le K$, on a :

$$u_k \ge u_n \ge u_K - 1$$

En posant :

$$I = u_K - 1$$

on voit que :

$$u_n \ge I$$

pour tout $n \in \mathbb{N}$. Une suite de Cauchy décroissante est également minorée.

37.5 Suite inverse

Soit un rationnel I_0 vérifiant $I_0 > 0$ et la suite $I : \mathbb{N} \setminus \{0\} \mapsto \mathbb{Q}$ définie par :

$$I : n \mapsto I_n = \frac{I_0}{n}$$

pour tout naturel n vérifiant $n \neq 0$. Soit $\epsilon > 0$. Comme I_0 est strictement positif, on a :

$$\epsilon \cdot \frac{1}{I_0} = \frac{\epsilon}{I_0} > 0$$

Soit $\epsilon > 0$. On peut trouver un naturel n tel que :

$$\frac{1}{n} < \frac{\epsilon}{I_0}$$

On a alors :

$$\frac{I_0}{n} < \epsilon$$

et :

$$\mathfrak{dist}\left(0, \frac{I_0}{n}\right) = \left|\frac{I_0}{n} - 0\right| = \left|\frac{I_0}{n}\right| \ < \ \epsilon$$

On en déduit que :

$$\lim_{n \to \infty} \frac{I_0}{n} = 0$$

Chapitre 38

Problème de la racine

38.1 Introduction

Nous allons tenter de déterminer la racine de deux, c'est-à-dire chercher un nombre noté :

$$x = \sqrt{2}$$

tel que :

$$x^2 = \left(\sqrt{2}\right)^2 = 2$$

38.2 Solution rationnelle ?

Nous allons chercher une solution à ce problème sous la forme d'un rationnel $x \in \mathbb{Q}$. Soit :

$$x = \frac{i}{j}$$

avec $i, j \in \mathbb{Z}$ et $j \neq 0$. On a :

$$x^2 = \frac{i^2}{j^2} = \frac{(-i)^2}{(-j)^2} = \frac{(-i)^2}{j^2} = \frac{i^2}{(-j)^2}$$

On peut donc se restreindre aux entiers positifs, autrement dit aux naturels. Si $i = 0$, on a forcément :

$$\frac{i^2}{j^2} = 0 \neq 2$$

ce qui ne résout pas notre problème. Supposons à présent que $i \neq 0$ et posons :

$$k = \mathrm{pgcd}(i, j)$$

On a alors $k \neq 0$ et les quotients :

$$a = i \div k$$

$$b = j \div k$$

vérifiant les divisions exactes :

$$i = a \cdot k$$

$$j = b \cdot k$$

Si nous voulons résoudre le problème, il faut donc avoir :

$$\left(\frac{a}{b}\right)^2 = \frac{a^2}{b^2} = \frac{a^2 \cdot k^2}{b^2 \cdot k^2} = \frac{i^2}{j^2} = 2$$

c'est-à-dire :

$$a^2 = 2\ b^2$$

On a donc :

$$a^2 \div 2 = b^2$$

et :

$$a^2 \bmod 2 = 0$$

38.2.1 Lemme

Soit un naturel $m \in \mathbb{N}$ vérifiant :

$$m^2 \bmod 2 = 0$$

Posons :

$$n = m \div 2$$

$$r = m \bmod 2$$

Le modulo $r \in \mathbb{N}$ vérifie :

$$0 \le r \le 2 - 1 = 1$$

autrement dit $r \in \{0, 1\}$. Si on suppose que $r = 1$, on a :

$$m = 2\ n + 1$$

Développons le carré :

$$m^2 = (2\ n + 1)^2 = 4\ n^2 + 4\ n + 1 = 4\ (n^2 + n) + 1$$

On aurait alors :

$$m^2 \div 2 = 2\ (n^2 + n)$$

et :

$$m^2 \bmod 2 = 1$$

contrairement à l'hypothèse. On en conclut que $r = 0$, la division entière de m par 2 est exacte :

$$m \bmod 2 = 0$$

38.2.2 Modulo nul

Le naturel a vérifiant :

$$a^2 \bmod 2 = 0$$

on a également :

$$a \bmod 2 = 0$$

Si on pose :

$$n = a \div 2$$

on a l'expression de la division exacte :

$$a = 2\ n$$

L'équation à résoudre devient alors :

$$a^2 = 4\ n^2 = 2\ b^2$$

On en déduit que :

$$b^2 = 2\ n^2$$

vérifie :

$$b^2 \div 2 = n^2$$

et :

$$b^2 \bmod 2 = 0$$

On a donc aussi :

$$b \bmod 2 = 0$$

Posons $p = b \div 2$. On a :

$$a = 2\ n$$

$$b = 2\ p$$

et :

$$i = 2\ n \cdot k = n \cdot (2\ k)$$

$$j = 2\ p \cdot k = p \cdot (2\ k)$$

Le naturel $t = 2\ k\ >\ k$ vérifie donc :

$$i \bmod t = j \bmod t = 0$$

ce qui est contraire à l'hypothèse de supremum de $k = \operatorname{pgcd}(i, j)$. On ne peut donc pas trouver de rationnel x vérifiant :

$$x^2 = 2$$

38.3 Suite inférieure

On ne peut pas trouver de solution exacte à l'équation $x^2 = 2$ dans l'ensemble des rationnels, mais on peut l'approcher autant qu'on le souhaite. Soit les suites de naturels $n : k \mapsto n_k$ et $M : k \mapsto M_k$ définies par :

$$n_k = 10^k$$

$$M_k = \sup \left\{ p \in \mathbb{N} : \frac{p^2}{n_k^2} \le 2 \right\}$$

pour tout $k \in \mathbb{N}$. On définit aussi les rationnels associés :

$$x_k = \frac{M_k}{n_k}$$

et les erreurs :

$$E_k = 2 - x_k^2$$

On a :

$$n_0 = 1 \qquad M_0 = 1 \qquad x_0 = 1 \qquad E_0 = 1$$

$$n_1 = 10 \qquad M_1 = 14 \qquad x_1 = 1,4 \qquad E_1 = 0,04$$

$$n_2 = 100 \qquad M_2 = 141 \qquad x_2 = 1,41 \qquad E_2 = 0,0119$$

et ainsi de suite.

38.3.1 Maximum

Soit :

$$A_k = \left\{ p \in \mathbb{N} : \frac{p^2}{n_k^2} \le 2 \right\}$$

Pour tout $p \in \mathbb{N}$ vérifiant $p \ > \ 2 \ n_k$, on a :

$$\frac{p^2}{n_k^2} \ > \ \frac{2^2 \ n_k^2}{n_k^2} = \frac{4 \ n_k^2}{n_k^2} = 4 \ > \ 2$$

On en conclut que :

$$A_k \subseteq \{0, 1, ..., 2 \ n_k - 1, 2 \ n_k\}$$

L'ensemble A_k contient donc un nombre fini d'éléments. Comme l'ordre usuel sur $\mathbb{N}$ est total, on en conclut que A_k admet un maximum identique au suprémum. La suite des M_k est donc bien définie :

$$M_k = \max A_k = \sup A_k$$

38.3.2 Majoration

L'inclusion nous donne l'inégalité des maxima :

$$\max A_k \le \max\{0, 1, ..., 2\ n_k\} = 2\ n_k$$

On en déduit que :

$$M_k \le 2\ n_k$$

En divisant cette inégalité par $n_k > 0$, on obtient une borne constante pour les rationnels associés :

$$x_k = \frac{M_k}{n_k} \le 2$$

La suite des x_k est majorée.

38.3.3 Suite croissante

Le maximum appartenant à l'ensemble, on a :

$$\frac{M_k^2}{n_k^2} \le 2$$

et :

$$\frac{(10\ M_k)^2}{n_{k+1}^2} = \frac{(10\ M_k)^2}{(10\ n_k)^2} = \frac{100\ M_k^2}{100\ n_k^2} = \frac{M_k^2}{n_k^2} \le 2$$

On en déduit que :

$$10\ M_k \in \left\{p \in \mathbb{N} : \frac{p^2}{n_{k+1}^2} \le 2\right\}$$

On conclut de ce résultat et du caractère de maximum de M_{k+1} que :

$$M_{k+1} \ge 10\ M_k$$

et :

$$x_{k+1} = \frac{M_{k+1}}{n_{k+1}} \ge \frac{10\ M_k}{n_{k+1}} = \frac{10\ M_k}{10\ n_k} = \frac{M_k}{n_k} = x_k$$

On en conclut que :

$$x_0 \le x_1 \le x_2 \le ...$$

On a donc $x_i \ge x_j$ pour tout $i, j \in \mathbb{N}$ vérifiant $i \ge j$, la suite des x_k est croissante.

38.3.4 Cadre

Comme M_k est le maximum de A_k, on a :

$$x_k^2 = \frac{M_k^2}{n_k^2} \le 2$$

et :

$$\left(\frac{M_k + 1}{n_k}\right)^2 > 2$$

On a :

$$\left(\frac{M_k + 1}{n_k} \right)^2 = \left(x_k + \frac{1}{n_k} \right)^2$$

En développant le binôme, on arrive à :

$$\left(\frac{M_k + 1}{n_k} \right)^2 = x_k^2 + \frac{2\ x_k}{n_k} + \frac{1}{n_k^2}$$

On a donc l'inégalité :

$$x_k^2 + \frac{2\ x_k}{n_k} + \frac{1}{n_k^2} \ > \ 2$$

qui nous donne la borne inférieure :

$$x_k^2 \ > \ 2 - \left(\frac{2\ x_k}{n_k} + \frac{1}{n_k^2} \right)$$

Posons :

$$\Delta_k = \frac{2\ x_k}{n_k} + \frac{1}{n_k^2}$$

En divisant l'inégalité :

$$1 \le n_k$$

par $n_k^2 \ > \ 0$, on obtient :

$$\frac{1}{n_k^2} \le \frac{1}{n_k}$$

On a donc :

$$\Delta_k \le \frac{2\ x_k}{n_k} + \frac{1}{n_k} = \frac{2\ x_k + 1}{n_k}$$

Comme $x_k \le 2$, on a :

$$\Delta_k \le \frac{2 \cdot 2 + 1}{n_k} = \frac{5}{n_k}$$

On a finalement la borne inférieure :

$$x_k^2 \ > \ 2 - \Delta_k \ge 2 - \frac{5}{n_k}$$

Posons :

$$\alpha_k = \frac{5}{n_k}$$

On a l'encadrement :

$$2 - \alpha_k \ < \ x_k^2 \le 2$$

38.3.5 Bornes d'une différence

Soit $i, j \in \mathbb{N}$ tels que $i \geq j$. On dispose des bornes :

$$2 - \alpha_i \ < \ x_i^2 \leq 2$$
$$2 - \alpha_j \ < \ x_j^2 \leq 2$$

En évaluant la différence de :

$$x_i^2 \leq 2$$

$$x_j^2 \ > \ 2 - \alpha_j$$

on arrive à la majoration :

$$x_i^2 - x_j^2 \leq 2 - (2 - \alpha_j) = \alpha_j$$

Comme la suite est croissante, on a $x_i \geq x_j$. Comme la fonction $f : x \mapsto x^2$ est croissante sur l'ensemble des rationnels positifs, on a également :

$$x_i^2 \geq x_j^2$$

ou :

$$x_i^2 - x_j^2 \geq 0$$

On a donc finalement :

$$\left|x_i^2 - x_j^2\right| = x_i^2 - x_j^2 \leq \alpha_j$$

38.3.6 Factorisation

On a la factorisation :

$$x_i^2 - x_j^2 = (x_i - x_j) \cdot (x_i + x_j)$$

Comme la suite est croissante, on sait que :

$$x_i, x_j \geq x_0 = 1$$

Leur somme vérifie l'inégalité :

$$x_i + x_j \geq 1 + 1 = 2$$

La différence des carrés est donc minorée par :

$$x_i^2 - x_j^2 = (x_i - x_j) \cdot (x_i + x_j) \geq (x_i - x_j) \cdot 2$$

En divisant par 2, on obtient :

$$x_i - x_j \leq \frac{1}{2} \ \left(x_i^2 - x_j^2\right)$$

Comme la suite est croissante, on a $x_i \geq x_j$ et :

$$x_i - x_j \geq 0$$

Donc :

$$|x_i - x_j| = x_i - x_j$$

Comme la fonction $f : x \mapsto x^2$ est croissante sur l'ensemble des rationnels positifs, on a également :

$$x_i^2 \geq x_j^2$$

ou :

$$x_i^2 - x_j^2 \geq 0$$

Donc :

$$\left|x_i^2 - x_j^2\right| = x_i^2 - x_j^2$$

On a donc :

$$|x_i - x_j| \leq \frac{1}{2} \ \left|x_i^2 - x_j^2\right| \leq \frac{\alpha_j}{2}$$

38.3.7 Suite de Cauchy

Soit $\epsilon > 0$. Comme :

$$\lim_{k\to\infty} \alpha_k = \lim_{k\to\infty} \frac{5}{n_k} = \lim_{k\to\infty} \frac{5}{10^k} = 0$$

on peut trouver un naturel K tel que :

$$|\alpha_k| = |\alpha_k - 0| \leq \epsilon$$

pour tout naturel k vérifiant $k \geq K$. On note que :

$$\alpha_k = \frac{5}{10^k} \leq \frac{5}{10^K} = \alpha_K$$

Si $i, j \in \mathbb{N}$ vérifient $i, j \geq K$, on a donc :

$$|x_i - x_j| \leq \frac{1}{2} \ \max\{\alpha_i, \alpha_j\} \leq \alpha_K \leq \epsilon$$

La suite des x_k est de Cauchy.

38.3.8 Erreur

On déduit du cadre de x_k que :

$$-\alpha_k = (2 - \alpha_k) - 2 \ < \ x_k^2 - 2$$

ou :

$$E_k = 2 - x_k^2 \ < \ \alpha_k$$

Comme on a également $x_k^2 \leq 2$ par définition des M_k, on a aussi $E_k \geq 0$ et :

$$|E_k - 0| = |E_k| = E_k \leq \alpha_k$$

La suite des α_k convergeant vers zéro, on en déduit que :

$$\lim_{k\to\infty} E_k = 0$$

38.3.9 Convergence

Soit $\epsilon > 0$. L'erreur convergeant vers zéro, on peut trouver un $K \in \mathbb{N}$ tel que :

$$\mathfrak{dist}(x_k^2, 2) = \left| x_k^2 - 2 \right| = |E_k| = |E_k - 0| \le \epsilon$$

pour tout naturel k vérifiant $k \ge K$. On en déduit que :

$$\lim_{k \to \infty} x_k^2 = 2$$

38.3.10 Supremum

Soit l'ensemble :

$$X = \left\{ x_k^2 : k \in \mathbb{N} \right\}$$

On sait déjà que :

$$x_k^2 = \frac{M_k^2}{n_k^2} \le 2$$

pour tout $k \in \mathbb{N}$. On en conclut que :

$$X \le 2$$

Pour tout rationnel $u \ge 2 \ge X$, on a :

$$u \in \operatorname{major} X$$

On en déduit que :

$$\{u \in \mathbb{Q} : u \ge 2\} \subseteq \operatorname{major} X$$

Soit un rationnel $v < 2$. Posons :

$$\delta = 2 - v > 0$$

On choisit un rationnel ϵ vérifiant :

$$0 < \epsilon < \delta$$

Comme la suite $k \mapsto x_k^2$ converge vers 2, on peut trouver un naturel K tel que :

$$\left| x_k^2 - 2 \right| = 2 - x_k^2 \le \epsilon$$

pour tout naturel k vérifiant $k \ge K$. On a alors :

$$x_k^2 \ge 2 - \epsilon > 2 - \delta = v$$

Donc :

$$v \notin \operatorname{major} X$$

On a donc :

$$\operatorname{major} X = \{u \in \mathbb{Q} : u \ge 2\}$$

et :

$$\sup \left\{ x_k^2 : k \in \mathbb{N} \right\} = \min\{u \in \mathbb{Q} : u \ge 2\} = 2$$

38.4 Suite supérieure

Soit les suites de naturels $n : k \mapsto n_k$ et $m : k \mapsto m_k$ définies par :

$$n_k = 10^k$$

$$m_k = \inf \left\{ p \in \mathbb{N} : \frac{p^2}{n_k^2} \ge 2 \right\}$$

pour tout $k \in \mathbb{N}$. On définit aussi les rationnels associés :

$$y_k = \frac{m_k}{n_k}$$

et les erreurs :

$$e_k = y_k^2 - 2$$

On a :

$$\begin{array}{cccc} n_0 = 1 & m_0 = 2 & y_0 = 2 & e_0 = 2 \\ n_1 = 10 & m_1 = 15 & y_1 = 1,5 & e_1 = 0,25 \\ n_2 = 100 & m_2 = 142 & y_2 = 1,42 & e_2 = 0,0164 \end{array}$$

et ainsi de suite.

38.4.1 Minimum

Soit les ensembles :

$$A_k = \left\{ p \in \mathbb{N} : \frac{p^2}{n_k^2} \le 2 \right\}$$

$$B_k = \left\{ p \in \mathbb{N} : \frac{p^2}{n_k^2} \ge 2 \right\}$$

Le problème de la racine de deux n'admettant pas de solution dans $\mathbb{Q}$, on ne peut trouver de naturel p tel que :

$$\frac{p^2}{n_k^2} = 2$$

Pour tout $p \in \mathbb{N}$, on a donc soit :

$$\frac{p^2}{n_k^2} \ < \ 2$$

et $p \in A_k$, soit :

$$\frac{p^2}{n_k^2} \ > \ 2$$

et $p \in B_k$. On en conclut que :

$$A_k \cup B_k = \mathbb{N}$$

Si on pouvait trouver un $p \in A_k \cap B_k$, on aurait :

$$2 \le \frac{p^2}{n_k^2} \le 2$$

et donc :

$$\frac{p^2}{n_k^2} = 2$$

ce qui est impossible. On en conclut que :

$$A_k \cap B_k = \emptyset$$

Si $p \in \mathbb{N} \setminus A_k$, on a $p \in B_k$ et vice versa. Donc :

$$B_k = \mathbb{N} \setminus A_k$$

Soit $p \in A_k$. Pour tout $u \in \mathbb{N}$ vérifiant $u \le p$, on a :

$$\frac{u^2}{n_k^2} \le \frac{p^2}{n_k^2} \le 2$$

On en conclut que $u \in A_k$. L'ensemble A_k possédant un maximum :

$$M_k = \max A_k$$

il est donc de la forme :

$$A_k = \{0, 1, ..., M_k - 1, M_k\}$$

On en conclut que B_k est de la forme :

$$B_k = \mathbb{N} \setminus A_k = \{M_k + 1, M_k + 2, ...\} = \{p \in \mathbb{N} : p \ge M_k + 1\}$$

On en conclut que le minimum de B_k existe et qu'il s'identifie à l'infimum. Les nombres m_k sont donc bien définis et :

$$m_k = \inf B_k = \min B_k = M_k + 1$$

38.4.2 Minoration

Comme $m_k = M_k + 1$, on a :

$$1 = x_0 \le x_k = \frac{M_k}{n_k} \le \frac{M_k + 1}{n_k} = \frac{m_k}{n_k} = y_k$$

La suite des y_k est minorée.

38.4.3 Suite décroissante

Le minimum appartenant à l'ensemble, on a :

$$\frac{m_k^2}{n_k^2} \ge 2$$

et :

$$\frac{(10\ m_k)^2}{n_{k+1}^2} = \frac{(10\ m_k)^2}{(10\ n_k)^2} = \frac{100\ m_k^2}{100\ n_k^2} = \frac{m_k^2}{n_k^2} \ge 2$$

On en déduit que :

$$10\ m_k \in \left\{ p \in \mathbb{N} : \frac{p^2}{n_{k+1}^2} \ge 2 \right\}$$

On conclut de ce résultat et du caractère de minimum de m_{k+1} que :

$$m_{k+1} \le 10\ m_k$$

et :

$$y_{k+1} = \frac{m_{k+1}}{n_{k+1}} \le \frac{10\ m_k}{n_{k+1}} = \frac{10\ m_k}{10\ n_k} = \frac{m_k}{n_k} = y_k$$

On en conclut que :

$$y_0 \ge y_1 \ge y_2 \ge \ldots$$

On a donc $y_i \le y_j$ pour tout $i, j \in \mathbb{N}$ vérifiant $i \ge j$, la suite des y_k est décroissante.

38.4.4 Cadre

Comme m_k est le minimum de B_k, on a :

$$y_k^2 = \frac{m_k^2}{n_k^2} \ge 2$$

et :

$$\left(\frac{m_k - 1}{n_k} \right)^2 < 2$$

On a :

$$\left(\frac{m_k - 1}{n_k} \right)^2 = \left(y_k - \frac{1}{n_k} \right)^2$$

En développant le binôme, on arrive à :

$$\left(\frac{m_k - 1}{n_k} \right)^2 = y_k^2 - \frac{2\ y_k}{n_k} + \frac{1}{n_k^2}$$

On a donc l'inégalité :

$$y_k^2 - \frac{2\ y_k}{n_k} + \frac{1}{n_k^2} < 2$$

qui nous donne la borne inférieure :

$$y_k^2 < 2 + \left(\frac{2\ y_k}{n_k} - \frac{1}{n_k^2} \right)$$

Posons :

$$\delta_k = \frac{2\ y_k}{n_k} - \frac{1}{n_k^2}$$

Comme :

$$\frac{1}{n_k^2} > 0$$

on a :

$$\delta_k \le \frac{2\ y_k}{n_k}$$

Comme $y_k \le y_0 = 2$, on a :

$$\delta_k \le \frac{2 \cdot 2}{n_k} = \frac{4}{n_k}$$

On a finalement la borne supérieure :

$$y_k^2 \ < \ 2 + \delta_k \le 2 + \frac{4}{n_k}$$

Posons :

$$\gamma_k = \frac{4}{n_k}$$

On a l'encadrement :

$$2 \le y_k^2 \ < \ 2 + \gamma_k$$

38.4.5 Bornes d'une différence

Soit $i, j \in \mathbb{N}$ tels que $i \le j$. On dispose des bornes :

$$2 \le y_i^2 \ < \ 2 + \gamma_i$$

$$2 \le y_j^2 \ < \ 2 + \gamma_j$$

En évaluant la différence de :

$$y_i^2 \ < \ 2 + \gamma_i$$

$$y_j^2 \ge 2$$

on arrive à la majoration :

$$y_i^2 - y_j^2 \le 2 + \gamma_i - 2 = \gamma_i$$

Comme la suite est croissante, on a $y_i \ge y_j$. Comme la fonction $f : x \mapsto x^2$ est croissante sur l'ensemble des rationnels positifs, on a également :

$$y_i^2 \ge y_j^2$$

ou :

$$y_i^2 - y_j^2 \ge 0$$

On a donc finalement :

$$\left|y_i^2 - y_j^2\right| = y_i^2 - y_j^2 \le \gamma_i$$

38.4.6 Factorisation

On a la factorisation :

$$y_i^2 - y_j^2 = (y_i - y_j) \cdot (y_i + y_j)$$

Comme :

$$y_i, y_j \geq 1$$

leur somme vérifie l'inégalité :

$$y_i + y_j \geq 1 + 1 = 2$$

La différence des carrés est donc minorée par :

$$y_i^2 - y_j^2 = (y_i - y_j) \cdot (y_i + y_j) \geq (y_i - y_j) \cdot 2$$

En divisant par 2, on obtient :

$$y_i - y_j \leq \frac{1}{2} \ \left(y_i^2 - y_j^2\right)$$

Comme la suite est croissante, on a $y_i \geq y_j$ et :

$$y_i - y_j \geq 0$$

Donc :

$$|y_i - y_j| = y_i - y_j$$

Comme la fonction $f : x \mapsto x^2$ est croissante sur l'ensemble des rationnels positifs, on a également :

$$y_i^2 \geq y_j^2$$

ou :

$$y_i^2 - y_j^2 \geq 0$$

Donc :

$$\left|y_i^2 - y_j^2\right| = y_i^2 - y_j^2$$

On a donc :

$$|y_i - y_j| \leq \frac{1}{2} \ \left|y_i^2 - y_j^2\right| \leq \frac{\gamma_i}{2}$$

38.4.7 Suite de Cauchy

Soit $\epsilon > 0$. Comme :

$$\lim_{k \to \infty} \gamma_k = \lim_{k \to \infty} \frac{4}{n_k} = \lim_{k \to \infty} \frac{4}{10^k} = 0$$

on peut trouver un naturel K tel que :

$$|\gamma_k| = |\gamma_k - 0| \leq \epsilon$$

pour tout naturel k vérifiant $k \geq K$. On note que :

$$\gamma_k = \frac{4}{10^k} \le \frac{4}{10^K} = \gamma_K$$

Si $i, j \in \mathbb{N}$ vérifient $i, j \ge K$, on a donc :

$$|y_i - y_j| \le \frac{1}{2} \ \max\{\gamma_i, \gamma_j\} \le \gamma_K \le \epsilon$$

La suite des y_k est de Cauchy.

38.4.8 Erreur

On déduit du cadre de y_k que :

$$e_k = y_k^2 - 2 \ < \ \gamma_k$$

Comme on a également $y_k^2 \ge 2$ par définition des m_k, on a aussi $e_k \ge 0$ et :

$$|e_k - 0| = |e_k| = e_k \le \gamma_k$$

La suite des γ_k convergeant vers zéro, on en déduit que :

$$\lim_{k \to \infty} e_k = 0$$

38.4.9 Convergence

Soit $\epsilon \ > \ 0$. L'erreur convergeant vers zéro, on peut trouver un $K \in \mathbb{N}$ tel que :

$$\mathfrak{dist}(y_k^2, 2) = \left|y_k^2 - 2\right| = |e_k| = |e_k - 0| \le \epsilon$$

pour tout naturel k vérifiant $k \ge K$. On en déduit que :

$$\lim_{k \to \infty} y_k^2 = 2$$

38.4.10 Infimum

Soit l'ensemble :

$$Y = \left\{Y_k^2 : k \in \mathbb{N}\right\}$$

On sait déjà que :

$$y_k^2 = \frac{m_k^2}{n_k^2} \ge 2$$

pour tout $k \in \mathbb{N}$. On en conclut que :

$$Y \ge 2$$

Pour tout rationnel $u \le 2 \le Y$, on a :

$$u \in \operatorname{minor} Y$$

On en déduit que :

$$\{u \in \mathbb{Q} : u \le 2\} \subseteq \operatorname{minor} Y$$

Soit un rationnel $v \ > \ 2$. Posons :

$$\delta = v - 2 \; > \; 0$$

On choisit un rationnel ϵ vérifiant :

$$0 \; < \; \epsilon \; < \; \delta$$

Comme la suite $k \mapsto y_k^2$ converge vers 2, on peut trouver un naturel K tel que :

$$\left| y_k^2 - 2 \right| = y_k^2 - 2 \le \epsilon$$

pour tout naturel k vérifiant $k \ge K$. On a alors :

$$y_k^2 \le 2 + \epsilon \; < \; 2 + \delta = v$$

Donc :

$$v \notin \operatorname{minor} Y$$

On a donc :

$$\operatorname{minor} Y = \{u \in \mathbb{Q} : u \le 2\}$$

et :

$$\inf \left\{ y_k^2 : k \in \mathbb{N} \right\} = \max\{u \in \mathbb{Q} : u \le 2\} = 2$$

38.5 Équivalence

On sait que :

$$x_k = \frac{M_k}{n_k} \le \frac{M_k + 1}{n_k} = \frac{m_k}{n_k} = y_k$$

pour tout $k \in \mathbb{N}$. Donc :

$$y_k - x_k \ge 0$$

et :

$$|y_k - x_k| = y_k - x_k$$

On a aussi $x_k^2 \le y_k^2$ et :

$$\left| y_k^2 - x_k^2 \right| = y_k^2 - x_k^2$$

En soustrayant les inégalités :

$$y_k^2 \le 2 + \gamma_k$$
$$x_k^2 \ge 2 - \alpha_k$$

on obtient :

$$y_k^2 - x_k^2 \le (2 + \gamma_k) - (2 - \alpha_k) = \gamma_k + \alpha_k$$

Posons :

$$\varpi_k = \gamma_k + \alpha_k = \frac{4}{n_k} + \frac{5}{n_k} = \frac{9}{n_k}$$

On a :

$$\left|y_k^2 - x_k^2\right| \le \varpi_k$$

Comme $y_k \ge x_k \ge x_0 = 1$, on a :

$$y_k + x_k \ge 1 + 1 = 2$$

La factorisation :

$$y_k^2 - x_k^2 = (y_k - x_k) \cdot (y_k + x_k) \ge (y_k - x_k) \cdot 2$$

nous donne la borne :

$$|y_k - x_k| \le \frac{|y_k^2 - x_k^2|}{2} \le \frac{\varpi_k}{2}$$

Comme :

$$\lim_{k\to\infty} \varpi_k = \lim_{k\to\infty} \frac{9}{n_k} = 0$$

on en déduit que :

$$\lim_{k\to\infty} |y_k - x_k| = 0$$

et :

$$\lim_{k\to\infty} (x_k - y_k) = 0$$

38.6 Conclusion

La racine de deux n'existe pas dans l'ensemble des rationnels, mais on peut trouver des suites de rationnels de Cauchy, croissante et majorée :

$$1 = x_0 \le x_1 \le ... \le x_k \le ... \le 2$$

ou décroissante et minorée :

$$2 = y_0 \ge y_1 \ge ... \ge y_k \ge ... \ge 1$$

dont les carrés convergent vers 2 :

$$\lim_{k\to\infty} x_k^2 = \lim_{k\to\infty} y_k^2 = 2$$

On a également les propriétés extrémales :

$$\sup\left\{x_k^2 : k \in \mathbb{N}\right\} = \inf\left\{y_k^2 : k \in \mathbb{N}\right\} = 2$$

et l'équivalence :

$$\lim_{k\to\infty} (x_k - y_k) = 0$$

On aimerait en déduire que les suites convergent et que :

$$\sqrt{2} = \lim_{k\to\infty} x_k = \lim_{k\to\infty} y_k = \sup\{x_k : k \in \mathbb{N}\} = \inf\{y_k : k \in \mathbb{N}\}$$

Malheureusement, ces limites et extrema n'existent pas dans l'ensemble des rationnels. Nous sommes donc amenés à associer au nombre $\sqrt{2}$ des ensembles associés à ces suites. La généralisation de ces propriétés nous mène à la construction des nombres réels.

Chapitre 39

Réels

Dépendances

— Chapitre 23 : Les nombres naturels
— Chapitre 24 : Les nombres entiers
— Chapitre 25 : Les nombres rationnels

39.1 Suites

On note $\mathfrak{C}^\top$ l'ensemble des suites de Cauchy rationnelles croissantes, $\mathfrak{C}^\perp$ l'ensemble des suites de Cauchy rationnelles décroissantes et :

$$\mathfrak{C} = \mathfrak{C}^\top \cup \mathfrak{C}^\perp$$

l'ensemble des suites de Cauchy rationnelles monotones.

39.1.1 Équivalence

$$\mathcal{E}^\top(s) = \left\{t \in \mathfrak{C}^\top : s \equiv t\right\}$$

$$\mathcal{E}^\top(s) = \left\{t \in \mathfrak{C}^\top : \lim_n (s_n - t_n) = 0\right\}$$

$$\mathcal{E}^\perp(s) = \left\{t \in \mathfrak{C}^\perp : s \equiv t\right\}$$

$$\mathcal{E}^\perp(s) = \left\{t \in \mathfrak{C}^\perp : \lim_n (s_n - t_n) = 0\right\}$$

$$\mathcal{E}(s) = \{t \in \mathfrak{C} : s \equiv t\}$$

$$\mathcal{E}(s) = \left\{t \in \mathfrak{C} : \lim_n (s_n - t_n) = 0\right\}$$

39.2 Ensembles

$$\lambda(u) = \{v \in \mathbb{Q} : v \le u\}$$

$$\theta(u) = \{v \in \mathbb{Q} : v \ge u\}$$

$$s : \mathbb{N} \mapsto \mathbb{Q}$$

$$s : n \mapsto s_n$$

$$\Lambda(s) = \sup_{\subseteq} \{\lambda(s_n) : n \in \mathbb{N}\}$$

$$\Lambda(s) = \bigcup_{n \in \mathbb{N}} \lambda(s_n) = \bigcup_{n \in \mathbb{N}} \{v \in \mathbb{Q} : v \le s_n\}$$

$$\Theta(s) = \inf_{\subseteq} \{\theta(s_n) : n \in \mathbb{N}\}$$

$$\Theta(s) = \bigcap_{n \in \mathbb{N}} \theta(s_n) = \bigcap_{n \in \mathbb{N}} \{v \in \mathbb{Q} : v \ge s_n\}$$

39.3 Séparation

On aimerait bien étendre $\mathbb{Q}$ en construisant un ensemble qui contienne la solution r d'équations telles que $r^2 = 2$. Notons que si $r^2 = 2$ et que l'on veut garder dans cet ensemble étendu les propriétés des rationnels, on doit également avoir $(-r)^2 = r^2 = 2$. Il y aurait donc deux solutions. Nous allons étudier séparément la solution liée aux rationnels positifs. Soit la fonction $f : \mathbb{Q} \mapsto \mathbb{Q}$ définie par :

$$f(x) = \begin{cases} 0 & \text{si } x \le 0 \\ x^2 & \text{si } x > 0 \end{cases}$$

pour tout $x \in \mathbb{Q}$. On part de la constatation que, si la solution de $f(x) = 2$ n'existe pas dans $\mathbb{Q}$, l'équation « sépare » les rationnels en deux catégories : les x tels que $f(x) < 2$ et les y tels que $f(y) > 2$. Considérons l'ensemble :

$$R = \{x \in \mathbb{Q} : f(x) < 2\}$$

Plus on augmente la valeur de $x \in R$, plus l'erreur $e = 2 - x^2 > 0$ diminue. On a par conséquent envie de dire que la solution r est le plus grand des éléments de R. Mais comme ni le maximum ni le supremum n'existent au sens de l'ordre $\le$, nous le considérons plutôt au sens de l'inclusion ensembliste $\subseteq$ sur les sous-ensembles de R :

$$r \equiv \sup_{\subseteq} \mathfrak{P}(R) = R$$

Nous sommes donc amenés à associer un sous-ensemble R des rationnels à chaque élément r de l'ensemble que nous désirons construire. Mais nous n'allons pas prendre n'importe quel sous-ensemble de $\mathbb{Q}$: on désire que les sous-ensembles acceptés vérifie des propriétés analogues à notre R particulier. Or, pour tout $x \in R$, l'ensemble :

$$\Lambda(x) = \{y \in \mathbb{Q} : y < x\}$$

est inclus dans R. Cette propriété est à la base de la construction des réels.

39.4 Définition

On nomme réel tout nombre r associé à un sous-ensemble $R \subseteq \mathbb{Q}$ tel que :
— $R \neq \emptyset$
— Il existe un $\mu \in \mathbb{Q}$ tel que $R \leq \mu$
— Pour tout $x \in R$ et $y \in \mathbb{Q}$ tel que $y < x$, on a $y \in R$:

$$\Lambda(x) = \{y \in \mathbb{Q} : y < x\} \subseteq R$$

— Le maximum de R n'existe pas.
On note $\mathbb{R}$ l'ensemble des réels.

Corollaire

Si le rationnel $x \in \mathbb{Q}$ n'appartient pas à R, on a $y \notin R$ pour tout $y \in \mathbb{Q}$ vérifiant $y > x$ (dans le cas contraire, on aurait $x \notin R$ avec $x < y$ et $y \in R$, ce qui contredit la définition des réels).

Notation

Pour tout réel r associé au sous-ensemble de rationnels R, on note bien entendu :

$$r = \sup R$$

39.5 Inclusions

On peut associer à tout rationnel $x \in \mathbb{Q}$ un ensemble $\Lambda(x)$. Or, on a clairement $\Lambda(x) \neq \emptyset$ et $\Lambda(x) \leq x$. Soit $s \in \Lambda(x)$. On a $\Lambda(s) \subseteq \Lambda(x)$. On peut aussi trouver un $t \in \Lambda(x)$ tel que $s < t < x$. Donc, s ne peut pas être le maximum de $\Lambda(x)$. On en conclut que tout rationnel x correspond au réel associé à $\Lambda(x)$. Les entiers pouvant être considérés comme des cas particuliers de rationnels et les naturels comme des cas particuliers d'entiers, on a donc finalement : $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$.

39.6 Ordre

L'ordre sur les réels découle directement de l'ordre $\subseteq$ sur les ensembles. Soit $r \in \mathbb{R}$ associé au sous-ensemble $R \subseteq \mathbb{Q}$ et $s \in \mathbb{R}$ associé au sous-ensemble $S \subseteq \mathbb{Q}$. On dit que r est plus petit que s et on le note :

$$r \leq s$$

si et seulement si R est inclus dans S :

$$R \subseteq S$$

Totalité

Supposons que pour tout $x \in R$, on ait $x \in S$. On a alors $R \subseteq S$ et $r \leq s$ par définition. Inversément, supposons que l'on puisse trouver un $x \in R$ tel que $x \notin S$. Tous les rationnels $z \in \mathbb{Q}$ tels que $z = x$ ou $z > x$ vérifient $z \notin S$. Par conséquent, si $y \in S$, on doit avoir $y < x$. Mais, par définition des réels et comme $x \in R$, on doit aussi avoir $y \in R$. On en conclut que $S \subseteq R$, et donc $s \leq r$.

Pour tout couple de réels $(r, s) \in \mathbb{R}^2$, on a donc soit $r \leq s$, soit $s \leq r$. L'ordre ainsi défini est donc total.

Ordre strict

On a également l'analogue pour l'ordre et l'inclusion stricts :

$$r < s \quad \Leftrightarrow \quad R \subset S$$

Ce qui revient à dire que $r < s$ si et seulement si $r \le s$ et $r \ne s$.

39.7 Addition

L'addition de $r \in \mathbb{R}$ associé à R et de $s \in \mathbb{R}$ associé à S est définie par :

$$r + s \equiv \{x + y \in \mathbb{Q} : x \in R, \ y \in S\}$$

39.8 Neutre additif

Soit $0 \in \mathbb{Q}$ et le sous-ensemble de rationnel correspondant :

$$Z = \{x \in \mathbb{Q} : x < 0\}$$

Soit $r \in \mathbb{R}$ associé à l'ensemble R et l'ensemble $S = R + Z$ associé à l'addition $r + 0$.

Tout $s \in S$ peut s'écrire sous la forme $s = x + z$, pour un certain $x \in R$ et un certain $z \in Z$. Si $z = 0$, on a bien évidemment $s = x \in R$. Sinon, $z < 0$ et $s < x$, d'où $s \in R$ par définition des réels. On a donc $S \subseteq R$.

Réciproquement, soit un rationnel $x \in R$. Si on avait $y \le x$ pour tout $y \in R$, notre x serait le maximum de R, ce qui n'est pas possible par définition des réels. On peut donc trouver un $y \in R$ tel que $x < y$. Le rationnel $d = x - y$ est strictement négatif et appartient donc à Z. On en déduit que :

$$x = x - y + y = d + y$$

où $d \in Z$ et $y \in R$. Donc, $x \in S$ et $R \subseteq S$.

La double inclusion nous montre alors que $R = S$, autrement dit que $r + 0 = r$. L'élement neutre pour l'addition, noté $0 \in \mathbb{R}$, est donc associé à l'ensemble des rationnels strictement négatifs :

$$0 \equiv \{x \in \mathbb{Q} : x < 0\}$$

39.9 Positifs et négatifs

On définit les ensembles des réels positifs et négatifs par :

$$\mathbb{R}^+ = \{x \in \mathbb{R} : x \ge 0\}$$
$$\mathbb{R}^- = \{x \in \mathbb{R} : x \le 0\}$$

39.10 Signe

La fonction signe est définie par :

$$\operatorname{sign}(x) = \begin{cases} 1 & \text{si } x \ge 0 \\ -1 & \text{si } x < 0 \end{cases}$$

pour tout $x \in \mathbb{R}$

39.11 Opposé

Soit $r \in \mathbb{R}$. On aimerait bien trouver l'opposé $-r \in \mathbb{R}$ tel que :

$$r + (-r) = (-r) + r = 0$$

Si $r \in \mathbb{Q}$, on a simplement :

$$-r \equiv \Lambda(-r) = \{x \in \mathbb{Q} : x < -r\}$$

Si $r \in \mathbb{R} \setminus \mathbb{Q}$, on définit l'association :

$$-r \equiv \{-x : x \in \mathbb{Q} \setminus R\}$$

ou $R \subseteq \mathbb{Q}$ est l'ensemble de rationnels associé à r.

39.12 Soustraction

On définit la soustraction par :

$$r - s = r + (-s)$$

pour tout $r, s \in \mathbb{R}$.

39.13 Valeur absolue

La valeur absolue d'un réel $r \in \mathbb{R}$ est définie par :

$$|r| = \sup\{-r, r\}$$

Propriétés

On a clairement $|-r| = |r|$. Si $r \ge 0$, on a $|r| = r \ge 0$. Si $r < 0$, on a $|r| = -r > 0$. On en conclut que $|r| \ge 0$ pour tout $r \in \mathbb{R}$. Si $|r| = 0$, on a soit $r = 0$ ou $-r = 0$. On en conclut que $r = 0$. Enfin, choisissons $r, s \in \mathbb{R}$:

— Si $r, s \ge 0$, on a $|r + s| = r + s$.
— Si $r \ge 0$ et $s < 0$, on a $|r + s| \le \max\{|r|, |s|\} \le |r| + |s|$.
— Si $r < 0$ et $s \ge 0$, on a $|r + s| \le \max\{|r|, |s|\} \le |r| + |s|$.
— si $r, s < 0$, on a $|r + s| = (-r) + (-s) = |r| + |s|$.

On en conclut que $|r + s| \le |r| + |s|$ pour tout $r, s \in \mathbb{R}$.

39.14 Distance

On définit une distance sur $\mathbb{R}$ par :

$$\mathfrak{dist}(x, y) = |x - y|$$

On a bien $\mathfrak{dist}(x, y) \ge 0$. La condition $\mathfrak{dist}(x, y) = 0$ implique que $|x - y|$, donc $x - y = 0$ et $x = y$. Enfin :

$$\begin{aligned}
\mathfrak{dist}(x, z) &= |x - z| \\
&\le |x - y| + |y - z| \\
&\le \mathfrak{dist}(x, y) + \mathfrak{dist}(y, z)
\end{aligned}$$

39.15 Arrondis

L'arrondi inférieur d'un réel $x \in \mathbb{R}$ associé à X est le plus grand entier dont l'ensemble associé est inclus dans X :

$$\lfloor x \rfloor = \sup\{n \in \mathbb{Z},\ \Lambda(n) \subseteq X\}$$

L'arrondi supérieur est le plus petit entier dont l'ensemble associé inclut X :

$$\lceil x \rceil = \inf\{n \in \mathbb{Z},\ X \subseteq \Lambda(n)\}$$

Eloignement

Supposons que :

$$|\lfloor x \rfloor - x| \ge 1$$

on aurait :

$$\begin{aligned} x - \lfloor x \rfloor &\ge 1 \\ x &\ge \lfloor x \rfloor + 1 \end{aligned}$$

avec $\lfloor x \rfloor + 1$ entier, ce qui contredit l'hypothèse de supremum de l'arrondi inférieur. On déduit l'analogue pour l'arrondi supérieur. On a donc :

$$\max\{|\lceil x \rceil - x|\,, |\lfloor x \rfloor - x|\}\ <\ 1$$

39.16 Suites convergentes

Soit $r \in \mathbb{R}$, l'entier $m \in \mathbb{Z}$ et le naturel $n \in \mathbb{N}$ tel que $n \neq 0$. On considère la suite des rationnels $x_n \in \mathbb{Q}$ définis par :

$$x_n = \frac{m}{2^n}$$

On va tenter de choisir m pour que :

$$\left|\frac{m}{2^n} - r\right| \le \frac{1}{2^n}$$

Cette inégalité est équivalente à :

$$\begin{aligned} \frac{m}{2^n} - r &\le \frac{1}{2^n} \\ r - \frac{m}{2^n} &\le \frac{1}{2^n} \end{aligned}$$

En multipliant par 2^n, on en déduit que :

$$r \cdot 2^n - 1 \le m \le r \cdot 2^n + 1$$

Il suffit donc de prendre :

$$m \in \{\lceil r \cdot 2^n \rceil, \lfloor r \cdot 2^n \rfloor\}$$

pour satisfaire la contrainte demandée. La suite des x_n converge vers r puisque :

$$\lim_{n \to \infty} |x_n - r| = 0$$

On le note :

$$\lim_{n\to\infty} x_n = r$$

On peut donc également identifier tout réel à une suite de rationnels qui converge vers lui.

Opération

On peut se servir de ce résultat pour étendre une opération $\circledast$ définie sur $\mathbb{Q}$. Soit $r, s \in \mathbb{R}$ et les suites de rationnels $x_i, y_i \in \mathbb{Q}$ convergent respectivement vers r et s :

$$\lim_{n\to\infty} x_n = r$$
$$\lim_{n\to\infty} y_n = s$$

On définit alors :

$$r \circledast s = \lim_{n\to\infty} [x_n \circledast y_n]$$

39.17 Multiplication

Soit $r, s \in \mathbb{R}$. et les suites de rationnels x_n et y_n vérifiant :

$$\lim_{n\to\infty} x_n = r$$
$$\lim_{n\to\infty} y_n = s$$

On définit la multiplication par :

$$r \cdot s = \lim_{n\to\infty} x_n \cdot y_n$$

39.18 Inverse

L'inverse d'un réel $r \in \mathbb{R}$ est le réel r^{-1} tel que :

$$r \cdot r^{-1} = r^{-1} \cdot r = 1$$

39.19 Division

Soit $r, s \in \mathbb{R}$ avec $s \neq 0$. On définit la division de r par s par :

$$\frac{r}{s} = r \cdot s^{-1}$$

39.20 Puissance

Soit $x \in \mathbb{R}$ et $n \in \mathbb{N}$. La puissance de x est comme d'habitude :

$$x^0 = 1$$
$$x^n = x \cdot x^{n-1}$$

Les puissances négatives sont données par :

$$x^{-n} = \left(x^{-1}\right)^n$$

Comme l'inverse d'une puissance est la puissance de l'inverse, on a aussi :

$$x^{-n} = (x^n)^{-1}$$

39.21 Racines

Soit $x \in \mathbb{R}$ et $n \in \mathbb{N}$. On dit que z est la n^{ime} racine de x si :

$$z^n = x$$

On note alors indifféremment :

$$z = x^{1/n} = \sqrt[n]{x}$$

Un cas particulier important est celui de la racine carrée :

$$\sqrt{x} = \sqrt[2]{x} = x^{1/2}$$

39.22 Unicité

Amplitude

Soit $x, y, z \in \mathbb{R}$ avec $x, y \geq 0$ et $n \in \mathbb{N}$. Il est clair que si $x \neq y$, on a forcément $x^n \neq y^n$. On en conclut que si x et y sont tels que $x^n = y^n$, on a forcément $x = y$. Les racines sont uniques sur $\mathbb{R}^+$.

Signe

Soit $x \in \mathbb{R}$. On a :

$$x^2 = (-x)^2$$

Par conséquent, x et $-x$ sont des racines carrées de $z = x^2$. Pour conserver l'unicité on impose que :

$$\sqrt{z} = |x| \geq 0$$

Il en va de même pour les puissances paires $2n$, où $n \in \mathbb{N}$ car :

$$x^{2n} = \left(x^2\right)^n = \left((-x)^2\right)^n = (-x)^{2n}$$

Par contre, pour les puissances impaires :

$$x^{2n+1} = \left(x^2\right)^n \cdot x = -\left((-x)^2\right)^n \cdot (-x) = -(-x)^{2n+1}$$

Le signe n'est pas ambigu. En résumé, on a :

$$\sqrt[2n]{x^{2n}} = |x|$$
$$\sqrt[2n+1]{x^{2n+1}} = x$$

39.23 Racine d'un produit

Soit $n \in \mathbb{N}$ et $a, b, x, y \in \mathbb{R}$ avec :

$$a^n = x$$
$$b^n = y$$

On a :

$$a^n \cdot b^n = a \cdot ... \cdot a \cdot b \cdot ... \cdot b = a \cdot b \cdot ... \cdot a \cdot b = (a \cdot b)^n$$

On en déduit que :

$$\sqrt[n]{a^n \cdot b^n} = a \cdot b$$

c'est-à-dire :

$$\sqrt[n]{x \cdot y} = \sqrt[n]{x} \cdot \sqrt[n]{y} = (x \cdot y)^{1/n} = x^{1/n} \cdot y^{1/n}$$

La racine d'un produit est égale au produit des racines.

39.24 Racine d'une racine

Soit $m, n \in \mathbb{N}$ et $a, b, c \in \mathbb{R}$ avec :

$$a^m = b$$
$$b^n = c$$

On voit que :

$$a = b^{1/m} = \left(c^{1/n}\right)^{1/m}$$

Mais on aussi :

$$a^{m \cdot n} = (a^m)^n = b^n = c$$

donc :

$$a = c^{1/(m \cdot n)}$$

On en déduit finalement que :

$$\left(c^{1/n}\right)^{1/m} = c^{1/(m \cdot n)}$$

39.25 Puissances fractionnaires

Soit $x \in \mathbb{R}$ et $m, n \in \mathbb{N}$. On définit les puissances fractionnaires par :

$$x^{m/n} = \left(x^{1/n}\right)^m$$
$$x^{-m/n} = \left(x^{1/n}\right)^{-m} = \frac{1}{x^{m/n}}$$

Comme la racine d'un produit est égale au produit des racines, on a aussi :

$$x^{m/n} = (x^m)^{1/n}$$
$$x^{-m/n} = \left(x^{-m}\right)^{1/n} = \frac{1}{x^{m/n}}$$

39.26 Somme en exposant

Soit $a, b \in \mathbb{Z}$ et $x \in \mathbb{R}$. Posons :

$$z = x^{1/(b \cdot d)}$$

On a alors :

$$x^{1/b} = z^d$$
$$x^{1/d} = z^b$$

On voit que :

$$x^{a/b} \cdot x^{c/d} = \left(z^d\right)^a \cdot \left(z^b\right)^c = z^{a \cdot d} \cdot z^{b \cdot c} = z^{a \cdot d + b \cdot c}$$

c'est-à-dire :

$$x^{a/b} \cdot x^{c/d} = x^{\frac{a \cdot d + b \cdot c}{b \cdot d}}$$

qui n'est rien d'autre que :

$$x^{a/b} \cdot x^{c/d} = x^{\frac{a}{b} + \frac{c}{d}}$$

39.27 Puissance d'une puissance

Soit $a, b \in \mathbb{Z}$ et $x \in \mathbb{R}$. Posons :

$$z = x^{1/(b \cdot d)}$$

On constate que :

$$\left(x^{a/b}\right)^{c/d} = \left(z^{a \cdot d}\right)^{c/d} = z^{a \cdot c}$$

qui n'est rien d'autre que :

$$\left(x^{a/b}\right)^{c/d} = x^{\frac{a \cdot c}{b \cdot d}}$$

39.28 Puissance réelle

Soit $x, s \in \mathbb{R}$ et la suite de rationnels $\{r_1, r_2, ...\}$ convergeant vers s :

$$\lim_{i \to \infty} r_i = s$$

On définit la puissance réelle par :

$$x^s = \lim_{i \to +\infty} x^{r_i}$$

Additivité

Les propriétés des puissances fractionnaires nous montrent que :

$$x^{r+s} = x^r \cdot x^s$$
$$(x^r)^s = x^{r \cdot s}$$

pour tout $x, r, s \in \mathbb{R}$.

Chapitre 40

Extrema réels

Dépendances

— Chapitre **??** : Les nombres réels
— Chapitre **??** : Les intervalles

40.1 Existence

Soit un sous-ensemble $A \subseteq \mathbb{R}$ avec $A \neq \emptyset$. Pour tout réel $x \in A$, on note $Q(x)$ le sous-ensemble de rationnels associé.

— Supposons que A soit majoré ($\operatorname{major} A \neq \emptyset$). Choisissons $\mu \in \operatorname{major} A$ et considérons le sous-ensemble de rationnels $Q(\mu)$ associé à μ. Comme $\mu \geq A$ et comme l'ordre $\leq$ est dérivé de l'ordre inclusif sur les sous-ensembles de rationnels associés, on a $Q(x) \subseteq Q(\mu)$ pour tout $x \in A$. On en conclut que l'union S des $Q(x)$ est inclue dans $Q(\mu)$:

$$S = \bigcup_{x \in A} Q(x) \subseteq Q(\mu)$$

Comme $Q(\mu)$ est majoré, on peut trouver un rationnel σ tel que :

$$S \subseteq Q(\mu) \leq \sigma$$

Donc $S \leq \sigma$, ce qui prouve que S est majoré. D'un autre coté, comme les $Q(x)$ ne sont pas vides, il est clair que leur union S n'est pas vide.
Soit $\alpha \in S$ et le rationnel $\beta \in \mathbb{Q}$ vérifiant $\beta < \alpha$. On peut trouver un $x \in A$ tel que $\alpha \in Q(x)$. Par définition des sous-ensembles de rationnels associés aux réels, on a $\beta \in Q(x)$, donc β appartient à l'union S. Enfin, si S admettait un maximum M, on on pourrait trouver un $x \in A$ tel que $M \in Q(x)$. On aurait aussi $Q(x) \subseteq S \leq M$, et donc $Q(x) \leq M$, ce qui contredit la définition des réels. On en conclut que l'ensemble S correspond à un réel s. Mais on sait que le supremum inclusif est égal à l'union :

$$S = \sup_{\subseteq} \{Q(x) \in \mathfrak{P}(\mathbb{Q}) : x \in A\}$$

L'ordre $\leq$ des réels étant dérivé de $\subseteq$, on en conclut que le supremum de A existe et que :

$$s = \sup A$$

Nous avonc donc prouvé que tout sous-ensemble A non vide majoré de $\mathbb{R}$ admet un supremum.

— Supposons que A soit minoré ($\operatorname{minor} A \neq \emptyset$). Choisissons $\lambda \in \operatorname{minor} A$ et posons :

$$-A = \{-x \in \mathbb{R} : x \in A\}$$

Comme $\lambda \le x$ pour tout $x \in A$, on a $-\lambda \ge -x$ et $-\lambda \in \mathrm{major}(-A) \ne \emptyset$. L'ensemble non vide $-A$ est donc majoré et admet un supremum $S = \sup(-A)$. On a $S \ge -x$ pour tout $x \in A$, donc $I = -S \le x$ et $I \in \mathrm{minor}\, A$. Choisissons $\alpha \in \mathrm{minor}\, A$. On a $\alpha \le x$ pour tout $x \in A$, d'où $-\alpha \ge -x$ et $-\alpha \in \mathrm{major}(-A)$. On en déduit que $-A \le S \le -\alpha$, c'est-à-dire $\alpha \le I \le A$. Le réel $I = -S$ est donc l'infimum de A :

$$\inf A = -\sup(-A)$$

Nous avonc donc prouvé que tout sous-ensemble A non vide minoré de $\mathbb{R}$ admet un infimum.

40.2 Adhérence et distance

Soit $A \subseteq \mathbb{R}$ et $r \in \mathbb{R}$. Soit l'ensemble :

$$D = \{\mathfrak{dist}(r, x) : x \in A\}$$

— Supposons que $r \in \mathrm{adh}\, A$. On a alors :

$$\mathfrak{dist}(r, A) = \inf D = 0$$

Choisissons un réel $\delta > 0$. Si on avait $\mathfrak{dist}(x, r) > \delta$ pour tout $x \in A$, on aurait $0 < \delta \le D$, ce qui contredit l'hypothèse d'infimum nul. Donc, pour tout $\delta > 0$, on peut trouver un $x \in A$ tel que $\mathfrak{dist}(x, r) \le \delta$.

— Réciproquement, supposons que pour tout réel $\delta > 0$, on puisse trouver un $x \in A$ tel que $d = \mathfrak{dist}(x, r) \le \delta$. Dans ce cas, on a $\delta \le d \in D$. On en conclut que $\delta \notin \mathrm{minor}\, D$. Par contre, si $\delta \le 0$, on a $\delta \le 0 \le D$ par positivité de la distance. On en conclut que $\mathrm{minor}\, D =]-\infty, 0]$, d'où :

$$\mathfrak{dist}(r, A) = \inf D = \max \mathrm{minor}\, D = \max\,]-\infty, 0] = 0$$

et $r \in \mathrm{adh}\, A$.

40.3 Eloignement

— Supposons à présent que le supremum $S = \sup A$ existe et que l'on puisse trouver un $\delta > 0$ tel que :

$$\mathfrak{dist}(S, x) = |S - x| = S - x \ge \delta$$

pour tout $x \in A$. Soit alors :

$$y = S - \frac{\delta}{2}$$

On voit que $S > y$ et que :

$$y - x = (y - S) + (S - x) \ge -\frac{\delta}{2} + \delta = \frac{\delta}{2} > 0$$

pour tout $x \in A$, c'est-à-dire $y \ge A$. On a donc $S > y \ge A$, ce qui contredit la définition du supremum. Pour tout $\delta > 0$, on peut donc trouver un $x \in A$ tel que $\mathfrak{dist}(S, x) \le \delta$. On en conclut que $\mathfrak{dist}(S, A) = 0$, c'est-à-dire :

$$\sup A \in \mathrm{adh}\, A$$

— Soit l'ensemble A admettant un infimum $I = \inf A$. Supposons que l'on puisse trouver un $\delta > 0$ tel que :

$$\mathfrak{dist}(I, x) = |I - x| = x - I \geq \delta$$

pour tout $x \in A$. Soit alors :

$$y = I + \frac{\delta}{2}$$

On voit que $I < y$ et que :

$$x - y = (x - I) + (I - y) \geq \delta - \frac{\delta}{2} = \frac{\delta}{2} > 0$$

pour tout $x \in A$, c'est-à-dire $y \leq A$. On a donc $I < y \leq A$, ce qui contredit la définition de l'infimum. Pour tout $\delta > 0$, on peut donc trouver un $x \in A$ tel que $\mathfrak{dist}(I, x) \leq \delta$. On en conclut que $\mathfrak{dist}(I, A) = 0$, c'est-à-dire :

$$\inf A \in \operatorname{adh} A$$

Chapitre 41

Opérations sur les limites

Soit les fonction $f, g : \mathbb{R} \mapsto \mathbb{R}$ vérifiant :

$$F = \lim_{x \to a} f(x)$$

$$G = \lim_{x \to a} g(x)$$

où $F, G \in \mathbb{R}$. Choisissons un réel $\epsilon > 0$. Pour tout $\gamma > 0$, on peut trouver un $\delta(\gamma) > 0$ tel que :

$$\mathfrak{dist}(f(x), F) = |f(x) - F| \le \gamma$$
$$\mathfrak{dist}(g(x), G) = |g(x) - G| \le \gamma$$

pour tout $x \in \mathbb{R}$ tel que $\mathfrak{dist}(x, a) = |x - a| \le \delta(\gamma)$.

41.0.1 Addition

$$|f(x) + g(x) - (F + G)| \le |f(x) - F| + |g(x) - G| = 2\gamma$$

Il suffit donc de choisir $\gamma = \epsilon/2$ pour avoir $|f(x) + g(x) - (F + G)| \le \epsilon$. On en déduit que la limite de $f + g$ est $F + G$, c'est-à-dire :

$$\lim_{x \to a} \big[f(x) + g(x) \big] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$$

Fonction constante

Dans le cas particulier où une des deux fonctions est une constante $G \in \mathbb{R}$, soit :

$$g : x \mapsto G$$

pour tout $x \in \mathbb{R}$, on a :

$$\lim_{x \to a} g(x) = \lim_{x \to a} G = G$$

et :

$$\lim_{x \to a} \big[f(x) + G \big] = \lim_{x \to a} f(x) + G$$

41.0.2 Opposé

On a simplement :

$$\big|\big[-g(x)\big]-(-G)\big| = |G-g(x)| = |g(x)-G| \le \gamma$$

Il suffit donc de choisir $\gamma = \epsilon$ pour avoir $|(-g(x))-(-G)| \le \epsilon$. On en déduit que la limite de $-g$ est $-G$, c'est-à-dire :

$$\lim_{x\to a}\big[-g(x)\big] = -\lim_{x\to a} g(x)$$

41.0.3 Soustraction

On a :

$$\begin{aligned}
\lim_{x\to a}\big[f(x)-g(x)\big] &= \lim_{x\to a}\big[f(x)+\big(-g(x)\big)\big] \\
&= \lim_{x\to a} f(x) + \lim_{x\to a}\big(-g(x)\big) \\
&= \lim_{x\to a} f(x) - \lim_{x\to a} g(x)
\end{aligned}$$

41.0.4 Multiplication

On voit que :

$$\begin{aligned}
|f(x)\cdot g(x) - F\cdot G| &= |f(x)\cdot g(x) - f(x)\cdot G + f(x)\cdot G - F\cdot G| \\
&= |f(x)\cdot(g(x)-G) + (f(x)-F)\cdot G| \\
&\le |f(x)\cdot(g(x)-G)| + |(f(x)-F)\cdot G| \\
&\le |f(x)|\cdot|g(x)-G| + |(f(x)-F)|\cdot|G| \\
&\le |f(x)|\cdot\gamma + \gamma\cdot|G|
\end{aligned}$$

Comme :

$$|f(x)| = |f(x)-F+F| \le |f(x)-F| + F \le \gamma + |F|$$

notre borne peut se réécrire :

$$\begin{aligned}
|f(x)\cdot g(x) - F\cdot G| &\le (\gamma+|F|)\cdot\gamma + \gamma\cdot|G| \\
&\le (\gamma+|F|+|G|)\cdot\gamma
\end{aligned}$$

Si on choisit $\gamma = \min\{1, \epsilon/(1+|F|+|G|)\}$, on a $\gamma \le 1$ et :

$$|f(x)\cdot g(x) - F\cdot G| \le (1+|F|+|G|)\cdot\gamma \le \epsilon$$

On en déduit que la limite de $f\cdot g$ est $F\cdot G$, c'est-à-dire :

$$\lim_{x\to a}\big[f(x)\cdot g(x)\big] = \Big[\lim_{x\to a} f(x)\Big]\cdot\Big[\lim_{x\to a} g(x)\Big]$$

41.0.5 Inverse multiplicatif

Supposons que $G \neq 0$. On a :

$$\begin{aligned}\left|\frac{1}{g(x)} - \frac{1}{G}\right| &= \left|\frac{G - g(x)}{g(x) \cdot |G|}\right| \\ &= \frac{|G - g(x)|}{|g(x)| \cdot |G|} \\ &= \frac{\gamma}{|g(x)| \cdot |G|}\end{aligned}$$

On voit aussi que :

$$\begin{aligned}|g(x)| = |g(x) - G + G| &= |G - (G - g(x))| \\ &\geq |G| - |G - g(x)| \\ &\geq |G| - \gamma\end{aligned}$$

Donc, si $\gamma < |G|$, on a :

$$\left|\frac{1}{g(x)} - \frac{1}{G}\right| \leq \frac{\gamma}{(|G| - \gamma) \cdot |G|}$$

Nous allons voir qu'il est possible de majorer cette expression par ϵ. En effet, la condition :

$$\frac{\gamma}{(|G| - \gamma) \cdot |G|} \leq \epsilon$$

est équivalente à :

$$\gamma \leq \epsilon \cdot (|G| - \gamma) \cdot |G| = G^2 \cdot \epsilon - \epsilon \cdot |G| \cdot \gamma$$

ce qui revient à dire que :

$$(1 + \epsilon \cdot |G|) \cdot \gamma \leq G^2 \cdot \epsilon$$

et enfin :

$$0 < \gamma \leq \frac{G^2 \cdot \epsilon}{1 + \epsilon \cdot |G|}$$

En imposant également $\gamma < |G|$, on obtient la condition suffisante :

$$0 < \gamma < \min\left\{|G|, \frac{G^2 \cdot \epsilon}{1 + \epsilon \cdot |G|}\right\}$$

On a alors $\left|\frac{1}{g(x)} - \frac{1}{G}\right| \leq \epsilon$. On en conclut que la limite de $1/g$ est $1/G$, c'est-à-dire :

$$\lim_{x \to a} \frac{1}{g(x)} = \frac{1}{\lim_{x \to a} g(x)}$$

41.0.6 Fraction

Toujours sous l'hypothèse que $G \neq 0$, on a :

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \left[\lim_{x \to a} f(x)\right] \cdot \left[\lim_{x \to a} \frac{1}{g(x)}\right] = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}$$

Chapitre 42

Limites réelles

42.1 Fonction croissante

Soit une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ croissante et majorée par un certain $M \in \mathbb{R}$:

$$F = \{f(x) \in \mathbb{R} : x \in \mathbb{R}\} \le M$$

Comme l'ensemble de réels F est non vide et majoré, on en conclut qu'il admet un supremum inclus dans l'adhérence :

$$S = \sup F \in \operatorname{adh} F$$

Choisissons un réel $\epsilon > 0$. Comme $\mathfrak{dist}(S, F) = 0$, on peut trouver un $\alpha \in \mathbb{R}$ tel que :

$$\mathfrak{dist}(S, f(\alpha)) = |S - f(\alpha)| \le \epsilon$$

Comme $S \ge F$, on a $|S - f(\alpha)| = S - f(\alpha) \le \epsilon$. Soit un $\beta \in \mathbb{R}$ tel que $\beta \ge \alpha$. La fonction f étant croissante, on a $f(\beta) \ge f(\alpha)$, et donc :

$$\mathfrak{dist}(S, f(\beta)) = S - f(\beta) \le S - f(\alpha) \le \epsilon$$

On en déduit que $f(x)$ converge vers le supremum S lorsque $x \to \infty$. Par unicité de la limite, on a :

$$\lim_{x \to +\infty} f(x) = \sup\{f(x) \in \mathbb{R} : x \in \mathbb{R}\}$$

42.2 Fonction décroissante

Symétriquement, si $g : \mathbb{R} \mapsto \mathbb{R}$ est une fonction décroissante et minorée par un certain $L \in \mathbb{R}$, l'ensemble non vide :

$$G = \{g(x) \in \mathbb{R} : x \in \mathbb{R}\} \ge L$$

admet un infimum inclut dans l'adhérence :

$$I = \inf G \in \operatorname{adh} G$$

Choisissons un réel $\epsilon > 0$. Comme $\mathfrak{dist}(I, G) = 0$, on peut trouver un $\alpha \in \mathbb{R}$ tel que :

$$\mathfrak{dist}(I, g(\alpha)) = |I - g(\alpha)| \le \epsilon$$

Comme $I \le F$, on a $|I - g(\alpha)| = g(\alpha) - I \le \epsilon$. Soit un $\beta \in \mathbb{R}$ tel que $\beta \ge \alpha$. La fonction g étant décroissante, on a $g(\beta) \le g(\alpha)$, et donc :

$$\mathfrak{dist}(I, g(\beta)) = g(\beta) - I \le g(\alpha) - I \le \epsilon$$

On en déduit que $g(x)$ converge vers le supremum I lorsque $x \to \infty$. Par unicité de la limite, on a :

$$\lim_{x \to +\infty} g(x) = \inf\{g(x) \in \mathbb{R} : x \in \mathbb{R}\}$$

42.3 Limites supremum et infimum

Soit une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ majorée et minorée. On définit la famille $\{F(x) : x \in \mathbb{R}\}$ d'ensembles non vides, majorés et minorés par :

$$F(x) = \{f(z) : z \in \mathbb{R},\ z \ge x\}$$

42.3.1 Limite supremum

Si $x \ge y$, on a $F(x) \subseteq F(y)$, et donc $\sup F(x) \le \sup F(y)$. On en conclut que la fonction dérivée $d : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$d(x) = \sup\{f(z) : z \in \mathbb{R},\ z \ge x\}$$

pour tout $x \in \mathbb{R}$ est décroissante et minorée. Elle converge donc vers son infimum :

$$\lim_{x \to +\infty} d(x) = \inf_{x \in \mathbb{R}} d(x) = \inf_{x \in \mathbb{R}} \sup\{f(z) : z \in \mathbb{R},\ z \ge x\}$$

D'un autre coté, la définition de d implique que :

$$\lim_{x \to +\infty} d(x) = \lim_{x \to +\infty} \sup\{f(z) : z \in \mathbb{R},\ z \ge x\} = \limsup_{x \to +\infty} f(x)$$

On en conclut que :

$$\limsup_{x \to +\infty} f(x) = \inf_{x \in \mathbb{R}} \sup\{f(z) : z \in \mathbb{R},\ z \ge x\}$$

42.3.2 Limite infimum

Si $x \ge y$, on a $F(x) \subseteq F(y)$, et donc $\inf F(x) \ge \inf F(y)$. On en conclut que la fonction dérivée $c : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$c(x) = \inf\{f(z) : z \in \mathbb{R},\ z \ge x\}$$

pour tout $x \in \mathbb{R}$ est croissante et majorée. Elle converge donc vers son supremum :

$$\lim_{x \to +\infty} c(x) = \sup_{x \in \mathbb{R}} c(x) = \sup_{x \in \mathbb{R}} \inf\{f(z) : z \in \mathbb{R},\ z \ge x\}$$

D'un autre coté, la définition de c implique que :

$$\lim_{x \to +\infty} c(x) = \lim_{x \to +\infty} \inf\{f(z) : z \in \mathbb{R},\ z \ge x\} = \liminf_{x \to +\infty} f(x)$$

On en conclut que :

$$\liminf_{x \to +\infty} f(x) = \sup_{x \in \mathbb{R}} \inf\{f(z) : z \in \mathbb{R},\ z \ge x\}$$

42.4 Egalité des limites sup et inf

Soit une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ majorée et minorée. Posons :

$$S(x) = \sup\{f(z) : z \in \mathbb{R},\ z \ge x\}$$
$$I(x) = \inf\{f(z) : z \in \mathbb{R},\ z \ge x\}$$

et choisissons un réel $\epsilon > 0$.

— Considérons le cas particulier où les limites sup et inf sont identiques :

$$L = \limsup_{x \to +\infty} f(x) = \liminf_{x \to +\infty} f(x)$$

pour tout $x \in \mathbb{R}$. Choisissons un σ tel que $|S(x) - L| \le \epsilon$ pour tout x vérifiant $x \ge \sigma$ et un τ tel que $|I(x) - L| \le \epsilon$ pour tout x vérifiant $x \ge \tau$. Posant $M = \max\{\sigma, \tau\}$, il vient :

$$f(z) \le S(M) \le L + \epsilon$$
$$f(z) \ge I(M) \ge L - \epsilon$$

pour tout réel z vérifiant $z \ge M$. On a alors :

$$|f(z) - L| \le \epsilon$$

On en conclut que la limite de f à l'infini existe et que :

$$\lim_{x \to +\infty} f(x) = \limsup_{x \to +\infty} f(x) = \liminf_{x \to +\infty} f(x)$$

— Inversément, supposons la limite de f à l'infini existe :

$$L = \lim_{x \to +\infty} f(x)$$

Soit un réel M tel que :

$$|f(x) - L| \le \epsilon$$

pour tout réel x vérifiant $x \ge M$. On a alors :

$$L - \epsilon \le f(x) \le L + \epsilon$$

On en déduit que :

$$S(x) \le L + \epsilon$$
$$I(x) \ge L - \epsilon$$

Le supremum étant supérieur à l'infimum, on a finalement :

$$L - \epsilon \le I(x) \le S(x) \le L + \epsilon$$

et donc :

$$\{|S(x) - L|\,, |I(x) - L|\} \le \epsilon$$

On en conclut que S et I convergent vers L, c'est-à-dire :

$$\limsup_{x \to +\infty} f(x) = \liminf_{x \to +\infty} f(x) = \lim_{x \to +\infty} f(x)$$

42.5 Ordre et limite

Soit les fonctions $f, g : \mathbb{R} \mapsto \mathbb{R}$ vérifiant $f \le g$.

42.5.1 A l'infini

Supposons que les limites à l'infini existent :

$$\lim_{x \to +\infty} f(x) = b$$

$$\lim_{x \to +\infty} g(x) = c$$

Supposons que $b > c$ et posons $\epsilon = (b - c)/4 > 0$. On a alors :

$$b = c + 4\epsilon$$

On peut trouver un réel F tel que :

$$|f(x) - b| \le \epsilon$$

pour tout réel x vérifiant $x \ge F$. De même, on peut trouver un réel G tel que :

$$|g(x) - c| \le \epsilon$$

pour tout réel x vérifiant $x \ge G$. Donc, pour tout réel x vérifiant $x \ge M = \max\{F, G\}$, on a :

$$\{ \ |f(x) - b| \ , \ |g(x) - c| \ \} \le \epsilon$$

On voit que :

$$b - \epsilon = c + 3\epsilon > c + \epsilon$$

On en déduit que :

$$f(x) \ge b - \epsilon > c + \epsilon \ge g(x)$$

ce qui contredit $f \le g$. Notre hypothèse est donc fausse et $b \le c$, c'est-à-dire :

$$\lim_{x \to +\infty} f(x) \le \lim_{x \to +\infty} g(x)$$

42.5.2 Vers un réel

On montre par un raisonnement analogue que, si les limites de f, g en a existent, on a :

$$\lim_{x \to a} f(x) \le \lim_{x \to a} g(x)$$

42.5.3 Supremum et infimum

Pour tout $x \in \mathbb{R}$, on a :

$$\lambda(x) = \inf\{f(z) : z \in \mathbb{R},\ z \ge x\} \le \sigma(x) = \sup\{f(z) : z \in \mathbb{R},\ z \ge x\}$$

On en conclut que $\lambda \le \sigma$. Leurs limites à l'infini respectent donc le même ordre. Mais comme ces limites correspondent aux limites infimum et supremum de f, on a :

$$\liminf_{x \to +\infty} f(x) \le \limsup_{x \to +\infty} f(x)$$

42.6 Ordre et supremum-infimum

Soit les fonctions $f, g : \mathbb{R} \mapsto \mathbb{R}$ majorées, minorées et vérifiant $f \le g$. Posons :

$$\Theta(x) = \{z \in \mathbb{R} : z \ge x\}$$

pour tout $x \in \mathbb{R}$.

42.6.1 Supremum

Soit les fonctions φ, γ définies par :

$$\varphi(x) = \sup\{f(z) : z \in \mathbb{R},\ z \ge x\}$$
$$\gamma(x) = \sup\{g(z) : z \in \mathbb{R},\ z \ge x\}$$

pour tout $x \in \mathbb{R}$. Quel que soit le réel x, on a $f \le g$ sur $\Theta(x)$. On en conclut que :

$$\varphi(x) = \sup f(\Theta(x)) \le \sup g(\Theta(x)) = \gamma(x)$$

ce qui revient à dire que $\varphi \le \gamma$. On doit donc avoir :

$$\lim_{x \to +\infty} \varphi(x) \le \lim_{x \to +\infty} \gamma(x)$$

c'est-à-dire :

$$\limsup_{x \to +\infty} f(x) \le \limsup_{x \to +\infty} g(x)$$

42.6.2 Infimum

Soit les fonctions φ, γ définies par :

$$\varphi(x) = \inf\{f(z) : z \in \mathbb{R},\ z \ge x\}$$
$$\gamma(x) = \inf\{g(z) : z \in \mathbb{R},\ z \ge x\}$$

pour tout $x \in \mathbb{R}$. Quel que soit le réel x, on a $f \le g$ sur $\Theta(x)$. On en conclut que :

$$\varphi(x) = \inf f(\Theta(x)) \le \inf g(\Theta(x)) = \gamma(x)$$

ce qui revient à dire que $\varphi \le \gamma$. On doit donc avoir :

$$\lim_{x \to +\infty} \varphi(x) \le \lim_{x \to +\infty} \gamma(x)$$

c'est-à-dire :

$$\liminf_{x \to +\infty} f(x) \le \liminf_{x \to +\infty} g(x)$$

42.6.3 Egalité

Supposons que :

$$L = \liminf_{x \to +\infty} f(x) = \limsup_{x \to +\infty} g(x)$$

On a alors :

$$L = \liminf_{x \to +\infty} f(x) \le \limsup_{x \to +\infty} f(x) \le \limsup_{x \to +\infty} g(x) = L$$

Les limites sup et inf de f sont donc toutes deux égales à L et :

$$\lim_{x\to+\infty} f(x) = \liminf_{x\to+\infty} f(x) = \limsup_{x\to+\infty} f(x) = L$$

On a aussi :

$$L = \liminf_{x\to+\infty} f(x) \le \liminf_{x\to+\infty} g(x) \le \limsup_{x\to+\infty} g(x) = L$$

Les limites sup et inf de g sont donc toutes deux égales à L et :

$$\lim_{x\to+\infty} g(x) = \liminf_{x\to+\infty} g(x) = \limsup_{x\to+\infty} g(x) = L$$

On en conclut que les limites de f et g existent et que :

$$\lim_{x\to+\infty} f(x) = \lim_{x\to+\infty} g(x)$$

42.7 Cadre

Soit les fonctions $f, S, I : \mathbb{R} \mapsto \mathbb{R}$ majorées, minorées et vérifiant $I \le f \le S$. Supposons que :

$$L = \liminf_{x\to+\infty} I(x) = \limsup_{x\to+\infty} S(x)$$

On a alors :

$$L = \liminf_{x\to+\infty} I(x) \le \liminf_{x\to+\infty} f(x) \le \limsup_{x\to+\infty} f(x) \le \limsup_{x\to+\infty} S(x) = L$$

On en déduit que :

$$\liminf_{x\to+\infty} f(x) = \limsup_{x\to+\infty} f(x) = L$$

La limite de f existe donc et :

$$\lim_{x\to+\infty} f(x) = \liminf_{x\to+\infty} I(x) = \limsup_{x\to+\infty} S(x)$$

Chapitre 43

Suites de réels

43.1 Monotones

Soit une suite de réels $x_1 \le x_2 \le x_3 \le ...$ croissante et majorée. On a alors :

$$\lim_{n \to \infty} x_n = \sup\{x_n \in \mathbb{R} : n \in \mathbb{N}\}$$

Soit une suite de réels $y_1 \ge y_2 \ge y_3 \ge ...$ décroissante et minorée. On a alors :

$$\lim_{n \to \infty} y_n = \inf\{y_n \in \mathbb{R} : n \in \mathbb{N}\}$$

43.2 Limites extrémales

Soit une suite de réels $\{u_n \in \mathbb{R} : n \in \mathbb{N}\}$ majorée et minorée. On a :

$$\limsup_{n \to \infty} u_n = \inf_{n \in \mathbb{N}} \sup\{u_m : m \in \mathbb{N},\ m \ge n\}$$

$$\liminf_{n \to \infty} u_n = \sup_{n \in \mathbb{N}} \inf\{u_m : m \in \mathbb{N},\ m \ge n\}$$

Chapitre 44

Sommes réelles

44.1 Introduction

Nous nous intéressons à des suites de réels $A \subseteq \mathbb{R}$ de la forme :

$$A = \{a_k \in \mathbb{R} : k \in \mathcal{Z}\}$$

44.1.1 Dénombrable

Supposons que $\mathcal{Z} \subseteq \mathbb{Z}$. Si les sommes partielles convergent, on définit :

$$\sum_{k \in \mathbb{N}} a_k = \sum_{k=0}^{+\infty} a_k = \lim_{n \to \infty} \sum_{k=0}^{n} a_k$$

dans le cas où $\mathcal{Z} = \mathbb{N}$ et :

$$\sum_{k \in \mathbb{Z}} a_k = \sum_{k=-\infty}^{+\infty} a_k = \lim_{n \to \infty} \sum_{k=-n}^{n} a_k$$

dans le cas où $\mathcal{Z} = \mathbb{Z}$. On définit aussi :

$$\sum_{k=m}^{+\infty} a_k = \lim_{n \to \infty} \sum_{k=m}^{n} a_k$$

44.1.2 Quelconque

Pour un ensemble $\mathcal{Z}$ quelconque, on pose :

$$Z^+ = \{k \in \mathcal{Z} : a_k \ge 0\}$$

et :

$$Z^- = \{k \in \mathcal{Z} : a_k \ < \ 0\}$$

Sous réserve d'existence du suprémum, on définit :

$$\sum_{k \in Z^+} a_k = \sup \left\{ \sum_{k \in I} a_k : I \subseteq Z^+, \ I \text{ fini} \right\}$$

ainsi que :

$$\sum_{k \in Z^-} a_k = - \sup \left\{ - \sum_{k \in I} a_k : I \subseteq Z^-, \ I \text{ fini} \right\}$$

On définit alors la somme par :

$$\sum_{k \in \mathcal{Z}} a_k = \sum_{k \in Z^+} a_k + \sum_{k \in Z^-} a_k$$

44.2 Additivité

Soit la suite :

$$A = \{a_k \in \mathbb{R} : k \in \mathbb{N}\}$$

On définit la suite $\{S_n : n \in \mathbb{N}\}$ des sommes partielles par :

$$S_n = \sum_{k=0}^{n} a_k$$

pour tout $n \in \mathbb{N}$. Choisissons un $m \in \mathbb{N}$ vérifiant $m \ge 1$. On définit la suite $\{D_n : n \in \mathbb{N}\}$ des sommes partielles commençant en m par :

$$D_n = \sum_{k=m}^{n} a_k$$

pour tout $n \in \mathbb{N}$ vérifiant $n \ge m$. Choisissons un naturel n vérifiant $n \ge m$. On a :

$$S_n = \sum_{k=0}^{m-1} a_k + \sum_{k=m}^{n} a_k = S_{m-1} + D_n$$

Si on pose :

$$E = S_{m-1}$$

on a :

$$S_n = E + D_n$$

pour tout $n \in \mathbb{N}$ vérifiant $n \ge m$. Si la limite de la suite des D_n existe, celle des S_n aussi et :

$$\lim_{n \to \infty} S_n = E + \lim_{n \to \infty} D_n$$

Inversément, si la limite des S_n existe, celle des D_n aussi et :

$$\lim_{n \to \infty} D_n = \lim_{n \to \infty} S_n - E$$

On a par définition :

$$\lim_{n \to \infty} S_n = \sum_{k=0}^{+\infty} a_k$$

et :

$$\lim_{n \to \infty} D_n = \sum_{k=m}^{+\infty} a_k$$

On en conclut que :

$$\sum_{k=0}^{+\infty} a_k = \sum_{k=0}^{m-1} a_k + \sum_{k=m}^{+\infty} a_k$$

44.3 Somme résiduelle

Si la somme des a_k converge, l'additivité nous dit que :

$$\sum_{k=m}^{+\infty} a_k = \sum_{k=0}^{+\infty} a_k - \sum_{k=0}^{m-1} a_k$$

En passant à la limite $m \to \infty$, on a :

$$\begin{aligned}
\lim_{m\to\infty} \sum_{k=m}^{+\infty} a_k &= \lim_{m\to\infty} \left[\sum_{k=0}^{+\infty} a_k - \sum_{k=0}^{m-1} a_k \right] \\
&= \sum_{k=0}^{+\infty} a_k - \lim_{m\to\infty} \sum_{k=0}^{m-1} a_k \\
&= \sum_{k=0}^{+\infty} a_k - \sum_{k=0}^{+\infty} a_k \\
&= 0
\end{aligned}$$

La suite des sommes résiduelles R_m définie par :

$$R_m = \sum_{k=m}^{+\infty} a_k$$

pour tout $m \in \mathbb{N}$ converge donc vers zéro lorsque m tend vers l'infini.

44.4 Termes positifs

Soit la suite positive :

$$P = \{p_k \in \mathbb{R} : p_k \ge 0,\ k \in \mathbb{N}\}$$

et la suite des sommes partielles :

$$S_n = \sum_{k=0}^{n} p_k$$

Choisissons $m, n \in \mathbb{N}$ tels que $m \le n$. On a :

$$S_n = \sum_{k=0}^{m} p_k + \sum_{k=m+1}^{n} p_k = S_m + \sum_{k=m+1}^{n} p_k$$

Par positivité, des p_k, on a :

$$\sum_{k=m+1}^{n} p_k \ge 0$$

et :

$$S_n \geq S_m$$

La suite des S_n est croissante. Si on peut trouver un $M \in \mathbb{R}$ tel que :

$$S_n \leq M$$

pour tout $n \in \mathbb{N}$, la suite est majorée et converge vers son suprémum :

$$\sum_{k=0}^{+\infty} p_k = \lim_{n \to \infty} \sum_{k=0}^{n} p_k = \sup_{n \in \mathbb{N}} \sum_{k=0}^{n} p_k$$

44.5 Convergence absolue

Soit la suite :

$$A = \{a_k \in \mathbb{R} : k \in \mathbb{N}\}$$

La suite dérivée :

$$P = \{|a_k| \in \mathbb{R} : k \in \mathbb{N}\}$$

est positive. Si on peut trouver un $K \in \mathbb{R}$ tel que :

$$\sum_{k=0}^{n} |a_k| \leq K$$

pour tout $n \in \mathbb{N}$, la suite des valeurs absolues converge et :

$$\sum_{k=0}^{+\infty} |a_k| = \sup_{n \in \mathbb{N}} \sum_{k=0}^{n} |a_k|$$

La suite des sommes résiduelles associées converge donc vers zéro :

$$\lim_{m \to \infty} \sum_{k=m}^{+\infty} |a_k| = 0$$

Comme :

$$-|a_k| \leq a_k \leq |a_k|$$

on a :

$$-K \leq -\sum_{k=0}^{n} |a_k| \leq \sum_{k=0}^{n} a_k \leq \sum_{k=0}^{n} |a_k| \leq K$$

et :

$$-K \leq \sum_{k=0}^{n} a_k \leq K$$

La suite des sommes partielles S_n définies par :

$$S_n = \sum_{k=0}^{n} a_k$$

est donc majorée et minorée. Elle admet par conséquent des limites suprémum :

$$\sigma = \limsup_{n\to\infty} S_n = \inf_{n\in\mathbb{N}} \sup\{S_k : k \in \mathbb{N},\ k \ge n\}$$

et infimum :

$$\lambda = \liminf_{n\to\infty} S_n = \sup_{n\in\mathbb{N}} \inf\{S_k : k \in \mathbb{N},\ k \ge n\}$$

La suite des S_n converge-t-elle ? Autrement dit, les limites suprémum et infimum des S_n sont-elles identiques ? On sait déjà que :

$$\lambda \le \sigma$$

Soit la famille de suprémums décroissants définie par :

$$H_m = \sup\{S_n : n \in \mathbb{N}, n \ge m\}$$

pour tout $m \in \mathbb{N}$. On a :

$$\sigma = \inf_{m\in\mathbb{N}} H_m$$

Soit la famille d'infimums croissants définie par :

$$B_m = \inf\{S_n : n \in \mathbb{N}, n \ge m\}$$

pour tout $m \in \mathbb{N}$. On a :

$$\lambda = \sup_{m\in\mathbb{N}} B_m$$

Soit $\epsilon > 0$. Comme la somme résiduelle converge vers zéro, on peut trouver un naturel M tel que :

$$\sum_{k=m}^{+\infty} |a_k| \le \frac{\epsilon}{2}$$

pour tout naturel m vérifiant $m \ge M$. Choisissons des naturels m, n tels que $n \ge m+1$ et $m \ge M$. L'additivité finie nous dit que :

$$\sum_{k=0}^{n} a_k = \sum_{k=0}^{m} a_k + \sum_{k=m+1}^{n} a_k$$

c'est-à-dire :

$$S_n = S_m + \sum_{k=m+1}^{n} a_k$$

ou :

$$S_n - S_m = \sum_{k=m+1}^{n} a_k$$

En prenant la valeur absolue, il vient :

$$|S_n - S_m| = \left| \sum_{k=m+1}^{n} a_k \right| \le \sum_{k=m+1}^{n} |a_k|$$

Les termes $|a_k|$ étant positifs, la suite des D_i définie par :

$$D_i = \sum_{k=m+1}^{i} |a_k|$$

pour tout naturel i vérifiant $i \ge m+1$ est croissante, majorée et converge vers son suprémum. On a donc :

$$\sum_{k=m+1}^{n} |a_k| \le \lim_{i \to \infty} D_i = \sum_{k=m+1}^{+\infty} |a_k|$$

et :

$$|S_m - S_n| \le \sum_{k=m+1}^{+\infty} |a_k|$$

Comme $m+1 > m \ge M$, le terme de droite est majoré par $\epsilon/2$ et :

$$|S_m - S_n| \le \frac{\epsilon}{2}$$

Choisissons $m \ge M$. Pour tout naturel n vérifiant $n \ge m$, on a soit $n = m$ et $S_n = S_m$, soit $n \ge m+1$. On en déduit les inégalités :

$$S_n \le \max\left\{S_m, S_m + \frac{\epsilon}{2}\right\} = S_m + \frac{\epsilon}{2}$$

et :

$$S_n \ge \min\left\{S_m, S_m - \frac{\epsilon}{2}\right\} = S_m - \frac{\epsilon}{2}$$

En passant au suprémum pour les entiers n vérifiant $n \ge m$, on en déduit que :

$$H_m \le S_m + \frac{\epsilon}{2}$$

En passant à l'infimum pour les entiers n vérifiant $n \ge m$, on en déduit que :

$$B_m \ge S_m - \frac{\epsilon}{2}$$

En combinant ces deux inégalités, on obtient :

$$H_m \le S_m + \frac{\epsilon}{2} = \left(S_m - \frac{\epsilon}{2}\right) + \epsilon \le B_m + \epsilon$$

et a fortiori :

$$\inf_{\substack{m \in \mathbb{N} \\ m \ge M}} H_m \le \sup_{\substack{m \in \mathbb{N} \\ m \ge M}} B_m + \epsilon$$

c'est-à-dire :

$$\sigma \le \lambda + \epsilon$$

Cette relation étant valable pour tout $\epsilon > 0$, on en déduit que :

$$\sigma \le \lambda$$

Comme on a également $\lambda \le \sigma$, on en conclut que $\lambda = \sigma$. La somme converge donc et :

$$\sum_{k=0}^{+\infty} a_k = \limsup_{n \to \infty} S_n = \liminf_{n \to \infty} S_n$$

44.6 Progression géométrique infinie

Si le réel a vérifie $|a| > 1$, on voit que a^{n+1} converge vers 0 lorsque n tend vers l'infini. On a alors :

$$\sum_{i=0}^{+\infty} a^i = \lim_{n\to\infty} \frac{a^{n+1} - 1}{a - 1} = \frac{1}{1 - a}$$

Chapitre 45

Complexes

Dépendances

— Chapitre 23 : Les nombres naturels
— Chapitre 24 : Les nombres entiers
— Chapitre 25 : Les nombres rationnels
— Chapitre 39 : Les nombres réels

45.1 Nombre imaginaire

Nous avons vu que le carré d'un nombre positif est forcément positif. Par conséquent, on ne peut pas trouver de $x \in \mathbb{R}$ tel que :

$$x^2 = -1$$

Il nous faut donc inventer un nouvel objet, que l'on nomme nombre imaginaire, et que l'on note $\mathbf{i}$, tel que :

$$\mathbf{i}^2 = -1$$

On peut dès lors étendre la définition de la racine carrée par :

$$\sqrt{-1} = \mathbf{i}$$

Soit $x \in \mathbb{R}$. Comme l'on veut conserver les propriétés des produits et puissances, on note que :

$$(x \cdot \mathbf{i})^2 = x^2 \cdot \mathbf{i}^2 = x^2 \cdot (-1) = -x^2$$

On a par conséquent :

$$\sqrt{-x^2} = x \cdot \mathbf{i}$$

Notation

On note aussi :

$$x\,\mathbf{i} = \mathbf{i}\,x = x \cdot \mathbf{i}$$

Remarque

Attention à ne pas confondre les variables « i » avec le nombre imaginaire « $\mathbf{i}$ ».

45.2 Définition

Nous allons à présent nous intéresser aux propriétés algébriques des couples nombre réel - nombre imaginaire. Pour tout $(a, b) \in \mathbb{R}^2$, nous introduisons la notation complexe :

$$a + \mathbf{i}\, b$$

L'ensemble des nombres complexes est simplement l'ensemble des couples réel - imaginaire :

$$\mathbb{C} = \{a + \mathbf{i}\, b : \ a, b \in \mathbb{R}\}$$

45.3 Parties réelles et imaginaires

Soit $z \in \mathbb{C}$ et $a, b \in \mathbb{R}$ tels que :

$$z = a + \mathbf{i}\, b$$

On nomme a la partie réelle de z, et on la note :

$$\Re(z) = a$$

On nomme b la partie imaginaire de z, et on la note :

$$\Im(z) = b$$

On a donc :

$$z = \Re(z) + \mathbf{i}\, \Im(z)$$

45.4 Complexe conjugué

Soit $z \in \mathbb{C}$ et $a, b \in \mathbb{R}$ tels que :

$$z = a + \mathbf{i}\, b$$

Le complexe conjugué de z est le nombre donné par :

$$\overline{z} = \operatorname{conj}(z) = a + \mathbf{i}(-b) = a - \mathbf{i}\, b$$

On a donc :

$$\Re(\overline{z}) = \Re(z)$$
$$\Im(\overline{z}) = -\Im(z)$$

45.4.1 Conjugué carré

Il est clair d'après la définition que :

$$\operatorname{conj} \operatorname{conj} z = z$$

45.5 Addition

Soient $x, y \in \mathbb{C}$ et $a, b, c, d \in \mathbb{R}$ tels que :

$$x = a + \mathbf{i}\, b$$
$$y = c + \mathbf{i}\, d$$

Comme nous désirons conserver les propriétés de commutativité et d'associativité de l'addition, on a :

$$x + y = a + \mathbf{i}\, b + c + \mathbf{i}\, d = (a + c) + \mathbf{i}(b + d)$$

45.6 Soustraction

Soient $x, y \in \mathbb{C}$ et $a, b, c, d \in \mathbb{R}$ tels que :

$$x = a + \mathbf{i}\, b$$
$$y = c + \mathbf{i}\, d$$

Comme nous souhaitons conserver les propriétés de la soustraction, on a simplement :

$$x + y = a + \mathbf{i}\, b - (c + \mathbf{i}\, d) = a + \mathbf{i}\, b - c - \mathbf{i}\, d = (a - c) + \mathbf{i}(b - d)$$

45.7 Multiplication

Soient $x, y \in \mathbb{C}$ et $a, b, c, d \in \mathbb{R}$ tels que :

$$x = a + \mathbf{i}\, b$$
$$y = c + \mathbf{i}\, d$$

Comme nous voulons conserver les propriétés de la multiplication, nous écrivons :

$$x \cdot y = (a + \mathbf{i}\, b) \cdot (c + \mathbf{i}\, d) = a \cdot c + \mathbf{i}\, a \cdot d + \mathbf{i}\, b \cdot c + \mathbf{i}^2\, b \cdot d$$

En tenant compte de la définition du nombre imaginaire, on a :

$$x \cdot y = a \cdot c + \mathbf{i}\, a \cdot d + \mathbf{i}\, b \cdot c - b \cdot d$$

et finalement :

$$x \cdot y = (a \cdot c - b \cdot d) + \mathbf{i}(a \cdot d + b \cdot c)$$

On a donc :

$$\Re(x \cdot y) = \Re(x) \cdot \Re(y) - \Im(x) \cdot \Im(y)$$
$$\Im(x \cdot y) = \Re(x) \cdot \Im(y) + \Im(x) \cdot \Re(y)$$

45.7.1 Mixte

On en déduit le cas particulier suivant :

$$(a + \mathbf{i}\, b) \cdot c = a \cdot c + \mathbf{i}\, b \cdot c$$

45.7.2 Multiplication par i

Soit $z \in \mathbb{C}$ et $a, b \in \mathbb{R}$ tels que :

$$z = a + \mathbf{i}\, b$$

Comme :

$$\mathbf{i}\, z = \mathbf{i}\, a + \mathbf{i}^2\, b = \mathbf{i}\, a - b$$

on a :

$$\Re(\mathbf{i}\, z) = -b = -\Im(z)$$
$$\Im(\mathbf{i}\, z) = a = \Re(z)$$

45.7.3 Conjugué

On a :

$$\mathrm{conj}(x) \cdot \mathrm{conj}(y) = (a \cdot c - b \cdot d) - \mathbf{i}(a \cdot d + b \cdot c) = \mathrm{conj}(x \cdot y)$$

45.8 Module

Soit $z \in \mathbb{C}$ et $a, b \in \mathbb{R}$ tels que :

$$z = a + \mathbf{i}\, b$$

On définit le module de z par :

$$|z| = \sqrt{a^2 + b^2} \geq 0$$

On voit que le module est un réel positif. Dans le cas d'un réel, le module s'identifie à la valeur absolue, ce qui justifie la notation identique.

45.8.1 Conjugué

On voit que :

$$z \cdot \overline{z} = (a + \mathbf{i}\, b) \cdot (a - \mathbf{i}\, b) = a^2 + b^2 + \mathbf{i}(a \cdot b - a \cdot b) = a^2 + b^2$$

c'est-à-dire :

$$z \cdot \overline{z} = |z|^2$$

45.9 Inverse

Soit $z \in \mathbb{C}$ et $a, b \in \mathbb{R}$ tels que :

$$z = a + \mathbf{i}\, b$$

Multiplions la relation $z \cdot \overline{z} = |z|^2$ par :

$$\left(z \cdot |z|^2\right)^{-1}$$

Il vient :

$$\frac{1}{z} = \frac{\overline{z}}{|z|^2}$$

Nous pouvons donc évaluer l'inverse d'un nombre complexe :

$$\frac{1}{z} = \frac{a - \mathbf{i}\, b}{a^2 + b^2} = \frac{a}{a^2 + b^2} - \mathbf{i}\, \frac{b}{a^2 + b^2}$$

Inverse de i

On a en particulier :

$$\frac{1}{\mathbf{i}} = -\,\mathbf{i}$$

45.10 Division

Soient $x, y \in \mathbb{C}$ et $a, b, c, d \in \mathbb{R}$ tels que :

$$x = a + \mathbf{i}\, b$$
$$y = c + \mathbf{i}\, d$$

On définit la division de deux nombres complexes par :

$$\frac{x}{y} = x \cdot \frac{1}{y} = \frac{x \cdot \overline{y}}{|y|^2}$$

On a donc :

$$\frac{x}{y} = \frac{(a + \mathbf{i}\, b) \cdot (c - \mathbf{i}\, d)}{c^2 + d^2}$$

ou encore :

$$\frac{x}{y} = \frac{(a \cdot c + b \cdot d) + \mathbf{i}(b \cdot c - a \cdot d)}{c^2 + d^2}$$

45.11 Obtention des parties réelles et imaginaires

Soit $z \in \mathbb{C}$ et $a, b \in \mathbb{R}$ tels que :

$$z = a + \mathbf{i}\, b$$

En additionnant z et son conjugué, on obtient :

$$z + \overline{z} = (a + \mathbf{i}\, b) + (a - \mathbf{i}\, b) = 2a = 2\Re(z)$$

ce qui permet d'obtenir une expression de la partie réelle :

$$\Re(z) = \frac{1}{2}(z + \overline{z})$$

En soustrayant z et $\overline{z}$, on obtient :

$$z - \overline{z} = (a + \mathbf{i}\, b) - (a - \mathbf{i}\, b) = 2\,\mathbf{i}\, b = 2\,\mathbf{i}\,\Im(z)$$

ce qui permet d'obtenir une expression de la partie imaginaire :

$$\Im(z) = \frac{1}{2\,\mathbf{i}}(z - \overline{z})$$

Cas particuliers

On en déduit directement que si $z = \overline{z}$, on a $\Im(z) = 0$ et $z \in \mathbb{R}$. Par contre, si $z = -\overline{z}$, on a $\Re(z) = 0$ et z est purement imaginaire.

45.12 Puissance

Soit $n \in \mathbb{N}$ avec $n \neq 0$, $z \in \mathbb{C}$ et $a, b \in \mathbb{R}$ tels que :

$$z = a + \mathbf{i}\, b$$

On définit la n^{ime} puissance d'un nombre complexe par :

$$z^0 = 1$$
$$z^n = z \cdot z^{n-1}$$

Négatives

Les puissances négatives sont données par :

$$z^{-n} = \frac{1}{z^n}$$

Racines

La n^{ime} racine $x \in \mathbb{C}$ de z est définie par :

$$x^n = z$$

On a alors :

$$z = x^{1/n}$$

Fractionnaires

Soit $m \in \mathbb{N}$. on a simplement :

$$z^{m/n} = \left(z^{1/n}\right)^m$$

Réelles

Soit $s \in \mathbb{R}$ et une suite $\{r_i \in \mathbb{Q} : i \in \mathbb{N}\}$ qui converge vers s. On définit :

$$z^s = \lim_{i \to \infty} z^{r_i}$$

45.13 Ordre partiel

Soit les complexes $x, y \in \mathbb{C}$ et les réels $a, b, c, d \in \mathbb{R}$ tels que :

$$x = a + \mathbf{i}\, b$$
$$y = c + \mathbf{i}\, d$$

On dit que x est plus petit que y au sens des composantes, et on le note $x \leq y$ si les deux conditions suivantes sont vérifiées :

$$a \le c$$
$$b \le d$$

45.14 Ordre total

Soit les complexes $x, y \in \mathbb{C}$ et les réels $a, b, c, d \in \mathbb{R}$ tels que :

$$x = a + \mathbf{i}\, b$$
$$y = c + \mathbf{i}\, d$$

On dit que x est plus petit que y et on le note $x \preceq y$ si :

$$a \ < \ c$$

ou si les conditions suivantes sont vérifiées :

$$a = c$$
$$b \le d$$

45.15 Inclusion

Pour tout $a \in \mathbb{R}$, on a $a = a + \mathbf{i}\, 0 \in \mathbb{C}$. On considère donc que $\mathbb{R} \subseteq \mathbb{C}$.

Chapitre 46

Intervalles de réels

Dépendances

— Chapitre **??** : Les nombres réels

46.1 Intervalles bornés

Soit $a, b \in \mathbb{R}$ avec $a \le b$. L'intervalle fermé $[a, b]$ est défini par :

$$[a, b] = \{x \in \mathbb{R} : a \le x \le b\}$$

L'intervalle ouvert $]a, b[$ est défini par :

$$]a, b[= \{x \in \mathbb{R} : a < x < b\}$$

On recontre aussi les intervalles semi-ouverts à gauche :

$$]a, b] = \{x \in \mathbb{R} : a < x \le b\}$$

ou à droite :

$$[a, b[= \{x \in \mathbb{R} : a \le x < b\}$$

46.2 Intervalles non bornés

Soit $a \in \mathbb{R}$. Les intervalles non bornés sont des intervalles partant de l'infini négatif ou/et allant jusqu'à l'infini positif :

$$\begin{aligned}
[a, +\infty[\quad &= \quad \{x \in \mathbb{R} : x \ge a\} \\
]a, +\infty[\quad &= \quad \{x \in \mathbb{R} : x > a\} \\
]-\infty, a] \quad &= \quad \{x \in \mathbb{R} : x \le a\} \\
]-\infty, a[\quad &= \quad \{x \in \mathbb{R} : x < a\} \\
]-\infty, +\infty[\quad &= \quad \mathbb{R}
\end{aligned}$$

46.3 Boules

Soit $c, r, x \in \mathbb{R}$ avec $r \ge 0$. La condition $|x - c| \le r$ est équivalente à $x - c \le r$ et $-(x - c) = c - x \le r$. On en déduit que :

$$x \le c + r$$
$$x \ge c - r$$

ce qui revient à dire que $x \in [c - r, c + r]$. Nous obtenons un résultat analogue avec les inégalités strictes. On peut donc exprimer les boules en terme d'intervalles :

$$\begin{aligned} \mathfrak{B}[c, r] &= [c - r, c + r] \\ \mathfrak{B}(c, r) &= \,]c - r, c + r[\end{aligned}$$

On peut inverser ces relations et exprimer certains intervalles en termes de boules. Soit $a = c - r$ et $b = c + r$. On a alors $a \le b$ et :

$$a + b = c - r + c + r = 2c$$
$$b - a = c + r - c + r = 2r$$

On en déduit que :

$$\begin{aligned} [a, b] &= \mathfrak{B}\left[\frac{a + b}{2}, \frac{b - a}{2}\right] \\ ]a, b[&= \mathfrak{B}\left(\frac{a + b}{2}, \frac{b - a}{2}\right) \end{aligned}$$

46.4 Majorants et minorants

Soit $a, b \in \mathbb{R}$ avec $a \le b$ et $x \in \mathbb{R}$.

Pour que $x \ge [a, b]$, il faut que $x \ge b$. Si $y \in [a, b]$, on aura alors $x \ge b \ge y$ d'où $x \ge y$ par transitivité de l'ordre. On en déduit que :

$$\operatorname{major}[a, b] =]b, +\infty]$$

On obtient le même résultat pour l'intervalle ouvert :

$$\operatorname{major}\ ]a, b[=]b, +\infty]$$

ainsi que pour les intervalles semi-ouverts.

Pour que $x \le [a, b]$, il faut que $x \le a$. Si $y \in [a, b]$, on aura alors $x \le a \le y$ d'où $x \le y$ par transitivité de l'ordre. On en déduit que :

$$\operatorname{minor}[a, b] = [-\infty, a[$$

On obtient le même résultat pour l'intervalle ouvert :

$$\operatorname{minor}\ ]a, b[= [-\infty, a[$$

ainsi que pour les intervalles semi-ouverts.

46.5 Maximum et minimum

Soit $a, b \in \mathbb{R}$ avec $a \le b$. On a clairement :

$$\begin{aligned} \max[a, b] &= \max\ ]a, b] = \max\ ]-\infty, b] = b \\ \min[a, b] &= \min[a, b[= \min[a, +\infty[\ = a \end{aligned}$$

Les autre types d'intervalles n'admettent ni maximum ni minimum.

46.6 Supremum et infimum

Dans le cas où le maximum de l'intervalle I existe, on a $\sup I = \max I$. Dans les autres cas, on a par exemple :

$$\sup\ ]a,b[= \min \operatorname{major}\ ]a,b[= \min [b,+\infty[= b$$

et le même résultat pour les intervalles semi-ouverts.

Dans le cas où le minimum de l'intervalle I existe, on a $\inf I = \min I$. Dans les autres cas, on a par exemple :

$$\inf\ ]a,b[= \max \operatorname{minor}\ ]a,b[= \max\ ]-\infty,a] = a$$

et le même résultat pour les intervalles semi-ouverts.

46.7 Adhérence, intérieur et frontière

On peut vérifier que :

$$\begin{aligned} \operatorname{adh}[a,b] &= \operatorname{adh}\ ]a,b[= [a,b] \\ \operatorname{int}[a,b] &= \operatorname{int}\ ]a,b[=]a,b[\end{aligned}$$

et ainsi de suite. La frontière est donc simplement :

$$\partial [a,b] = \partial\ ]a,b[= \{a,b\}$$

Chapitre 47

Dimension n

Dépendances

— Chapitre ?? : Les naturels
— Chapitre ?? : Les entiers
— Chapitre ?? : Les rationnels
— Chapitre ?? : Les réels
— Chapitre ?? : Les complexes

47.1 Définition

Conformément à la définition de la « puissance » d'un ensemble, $\mathbb{N}^n$ est l'ensemble des n-tuples :

$$\mathbb{N}^n = \{(m_1, m_2, ..., m_n) : m_1, m_2, ..., m_n \in \mathbb{N}\}$$

où chaque composante m_i est un naturel. On a pareillement :

$$\begin{aligned}
\mathbb{Z}^n &= \{(z_1, z_2, ..., z_n) : z_1, z_2, ..., z_n \in \mathbb{N}\} \\
\mathbb{Q}^n &= \{(q_1, q_2, ..., q_n) : q_1, q_2, ..., q_n \in \mathbb{Q}\} \\
\mathbb{R}^n &= \{(r_1, r_2, ..., r_n) : r_1, r_2, ..., r_n \in \mathbb{R}\} \\
\mathbb{C}^n &= \{(c_1, c_2, ..., c_n) : c_1, c_2, ..., c_n \in \mathbb{C}\}
\end{aligned}$$

Dans la suite, nous notons :

$$\mathcal{D} = \{\mathbb{N}^n, \mathbb{Z}^n, \mathbb{Q}^n, \mathbb{R}^n, \mathbb{C}^n\}$$

On dit que les éléments de $\mathcal{D}$ sont des ensembles de dimension n.

47.2 Egalité

Soit l'ensemble $A^n \in \mathcal{D}$ et $x, y \in A^n$. On a :

$$\begin{aligned}
x &= (x_1, x_2, ..., x_n) \\
y &= (y_1, y_2, ..., y_n)
\end{aligned}$$

On dit que $x = y$ si et seulement si $x_i = y_i$ pour tout $i \in \{1, 2, ..., n\}$.
De même, on dit que $x = 0$ si et seulement si $x_i = 0$ pour tout $i \in \{1, 2, ..., n\}$.

47.3 Ordre partiel

Soit l'ensemble $A^n \in \mathcal{D}$ et $x, y \in A^n$. On a :

$$x = (x_1, x_2, ..., x_n)$$
$$y = (y_1, y_2, ..., y_n)$$

On dit que $x \le y$ si et seulement si $x_i \le y_i$ pour tout $i \in \{1, 2, ..., n\}$. On le note aussi $y \ge x$.
On dit que $x < y$ si et seulement si $x_i < y_i$ pour tout $i \in \{1, 2, ..., n\}$. On le note aussi $y > x$.

47.4 Opérations internes

Soit l'ensemble $A^n \in \mathcal{D}$ et $x, y \in A^n$. On a :

$$x = (x_1, x_2, ..., x_n)$$
$$y = (y_1, y_2, ..., y_n)$$

pour certains $x_i, y_i \in A$.

Les opérations sur A^n sont définies par la même opération appliquée aux composantes. L'addition s'écrit donc :

$$x + y = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$$

tandis que la soustraction est donnée par :

$$x - y = (x_1 - y_1, x_2 - y_2, ..., x_n - y_n)$$

47.5 Multiplication mixte

La multiplication mixte $\cdot : A \times A^n \to A^n$ est définie par :

$$\alpha \cdot x = (\alpha \cdot x_1, \alpha \cdot x_2, ..., \alpha \cdot x_n)$$

pour tout $\alpha \in A$.

Notations

On note bien entendu :

$$x \cdot \alpha = \alpha \cdot x$$

Choisissant un $\beta \in A$, on pose également :

$$\alpha \cdot \beta \cdot x = (\alpha \cdot \beta) \cdot x$$

Enfin, le signe de multiplication est souvent omis :

$$\alpha\ x = \alpha \cdot x$$

47.6 Neutre

Soit l'ensemble $A^n \in \mathcal{D}$. On note :

$$0 = (0, 0, ..., 0)$$

l'élément $0 \in A^n$

47.7 Lien avec les fonctions

On voit clairement que ces relations et opérations sont équivalentes aux mêmes relations et opérations définies sur les fonctions :

$$\{1, 2, ..., n\} \mapsto \{f(1), f(2), ..., f(n)\} \equiv (f_1, f_2, ..., f_n)$$

associées.

47.8 Puissance

Soit $x = (x_1, ...x_n), y = (y_1, ..., y_n) \in A^n$. La puissance s'écrit :

$$y^x = y_1^{x_1} \cdot y_2^{x_2} \cdot ... \cdot y_n^{x_n}$$

47.9 Factorielle

Soit $m = (m_1, m_2, ..., m_n) \in \mathbb{N}^n$. On définit :

$$m! = m_1! \cdot m_2! \cdot ... \cdot m_n!$$

Chapitre 48

Matrices

Dépendances

- Chapitre **??** : Les réels
- Chapitre **??** : Les complexes

48.1 Définition

Les matrices sont des tableaux de réels, de complexes, ou de composantes d'une autre nature appartenant à un corps quelconque. Voici un exemple d'une matrice à m lignes et n colonnes :

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

où les a_{ij} appartiennent à un corps $\mathbb{K}$. On a également des versions simplifiées de cette notation :

$$A = (a_{ij})_{i,j} = [a_{ij}]_{i,j}$$

Par convention, le numéro de ligne passe avant le numéro de colonne à l'extérieur de la parenthèse (ou du crochet).

On note $\mathbb{M}(\mathbb{K}, m, n)$ l'ensemble des matrices à m lignes et n colonnes. On dit d'une matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$ qu'elle est de taille (m, n).

48.2 Composantes

Soit $A = (a_{ij})_{i,j}$. La fonction comp_{ij} donne la composante située à la i^{me} ligne et à la j^{me} colonne :

$$a_{ij} = \operatorname*{comp}_{ij} A$$

48.3 Blocs

Soit $A = (a_{ij})_{i,j}$. La fonction $\operatorname{bloc}_{ijkl}$ donne la sous-matrice située entre la i^{me} et la j^{me} ligne, ainsi qu'entre la k^{ime} et la l^{ime} colonne :

$$\operatorname*{bloc}_{ijkl} A = \begin{bmatrix} a_{ik} & a_{i,k+1} & \dots & a_{il} \\ a_{i+1,k} & a_{i+1,k+1} & \dots & a_{i+1,l} \\ \vdots & & \ddots & \vdots \\ a_{jk} & a_{j,k+1} & \dots & a_{jl} \end{bmatrix}$$

On appelle bloc une telle sous-matrice.

48.4 Partitionnement en blocs

Il est parfois très utile de partitionner une matrice en blocs, ou de former une matrice plus grande à partir de matrices de tailles plus petites. On note alors :

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix} = \begin{bmatrix} b_{11} & \dots & b_{1n} & c_{11} & \dots & c_{1p} \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} & c_{m1} & \dots & c_{mp} \\ d_{11} & \dots & d_{1n} & e_{11} & \dots & e_{1p} \\ \vdots & & \vdots & \vdots & & \vdots \\ d_{r1} & \dots & d_{rn} & e_{r1} & \dots & e_{rp} \end{bmatrix}$$

où les $b_{ij}, c_{ij}, d_{ij}, e_{ij}$ sont respectivement les composantes des matrices B, C, D, E. Ces matrices doivent évidemment être de tailles compatibles (nombres de lignes identiques pour B et C, ainsi que pour D et E ; nombres de colonnes identiques pour B et D, ainsi que pour C et E).

48.5 Formes lignes et colonnes

On peut exprimer une matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$ sous la forme de lignes superposées. Soit :

$$A = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{bmatrix}$$

où les $l_i \in \mathbb{M}(\mathbb{K}, 1, n)$ sont les lignes de A. On note :

$$\operatorname*{ligne}_i A = l_i$$

On peut aussi l'exprimer sous la forme de colonnes juxtaposées. Soit :

$$A = [c_1 \ c_2 \dots \ c_n]$$

où les $c_i \in \mathbb{M}(\mathbb{K}, m, 1)$ sont les colonnes de A. On note :

$$\operatorname*{colonne}_i A = c_i$$

48.6 Transposée

Transposer une matrice $A = (a_{ij})_{i,j} \in \mathbb{M}(\mathbb{K}, m, n)$ consiste à agencer ses lignes en colonnes et vice-versa :

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} = (a_{ji})_{i,j}$$

On voit que la transposée $A^T \in \mathbb{M}(\mathbb{K}, n, m)$.
On a clairement :

$$\left(A^T\right)^T = A$$

48.7 Vecteurs lignes et colonnes

Les vecteurs ligne sont des matrices ne possédant qu'une seule ligne. Ils sont donc de taille générique $(1, n)$. Les vecteurs colonne sont des matrices ne possédant qu'une seule colonne. Ils sont donc de taille générique $(m, 1)$.

Les vecteurs *colonne* sont notés :

$$x = (x_i)_i$$

On a aussi :

$$x_i = \operatorname*{comp}_i x$$

Equivalence avec $\mathbb{K}^n$

A partir d'un n-tuple $(x_1, x_2, ..., x_n) \in \mathbb{K}^n$, on peut former un vecteur ligne $x \in \mathbb{M}(\mathbb{K}, 1, n)$ par :

$$x = [x_1 \ x_2 \ \dots \ x_n]$$

et un vecteur colonne $x' \in \mathbb{M}(\mathbb{K}, n, 1)$ par :

$$x' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_1 \ x_2 \ \dots \ x_n]^T$$

On a donc $\mathbb{M}(\mathbb{K}, 1, n), \mathbb{M}(\mathbb{K}, n, 1) \equiv \mathbb{K}^n$.

Notation

Sauf mention contraire, nous convenons dans la suite que tout vecteur $u \in \mathbb{K}^n$ sera associé à un vecteur *colonne* de même nom $u \in \mathbb{M}(\mathbb{K}, n, 1)$. Si nous voulons utiliser un vecteur *ligne*, nous le noterons alors u^T (ou $u^\star$ pour des raisons qui apparaîtront plus loin).

48.8 Carrées et rectangulaires

Les matrices $A \in \mathbb{M}(\mathbb{K}, n, n)$ ayant même nombre de colonnes et de lignes sont dites « carrées ». Les matrices $A \in \mathbb{M}(\mathbb{K}, m, n)$ avec $m \neq n$ sont dites « rectangulaires ». Lorsque $m < n$, on dit que la matrice est « longue ». Si $m > n$, on dit que la matrice est « haute ».

48.9 Matrices diagonales

Une matrice diagonale contient des composantes nulles partout, à l'exception de la diagonale où aucune contrainte n'est fixée. On peut représenter une matrice diagonale $D \in \mathbb{M}(\mathbb{K}, m, n)$ par :

$$D = (d_i \cdot \boldsymbol{\delta}_{ij})_{i,j}$$

On peut aussi former une matrice diagonale de taille (m, n) à partir d'un vecteur $d = (d_i)_i$ de taille $(p, 1)$, où $p \le \min\{m, n\}$. On le note :

$$D = \operatorname*{diag}_{m,n}(d) = \operatorname*{diag}_{m,n}(d_1, ..., d_p)$$

où :

$$\operatorname*{comp}_{ij} \operatorname*{diag}_{m,n}(d_1, ..., d_p) = \begin{cases} d_i & \text{si } i = j \in \{1, 2, ..., p\} \\ 0 & \text{sinon} \end{cases}$$

Carrée

Dans le cas particulier où $m = n$, on note aussi :

$$\operatorname*{diag}_{n}(d) = \operatorname*{diag}_{n,n}(d)$$

48.10 Matrices triangulaires

Les composantes des matrices triangulaires supérieures ne peuvent être non nulles qu'au-dessus de la diagonale, c'est-à-dire lorsque $i \le j$. Elles peuvent se représenter par :

$$T = (t_{ij} \cdot \tau^+_{ij})_{i,j}$$

où :

$$\tau^+_{ij} = \begin{cases} 1 & \text{si } i \le j \\ 0 & \text{si } i > j \end{cases}$$

Les composantes des matrices triangulaires inférieures ne peuvent être non nulles qu'au-dessous de la diagonale, c'est-à-dire lorsque $i \ge j$. Elles peuvent se représenter par :

$$T = (t_{ij} \cdot \tau^-_{ij})_{i,j}$$

où :

$$\tau^-_{ij} = \begin{cases} 1 & \text{si } i \ge j \\ 0 & \text{si } i < j \end{cases}$$

48.11 Egalité

Il est clair que deux matrices $A, B \in \mathbb{M}(\mathbb{K}, m, n)$:

$$A = (a_{ij})_{i,j}$$
$$B = (b_{ij})_{i,j}$$

sont égales ($A = B$) si tous leurs éléments le sont :

$$a_{ij} = b_{ij}$$

pour tout $(i, j) \in \{1, 2, ..., m\} \times \{1, 2, ..., n\}$.

48.12 Ordre partiel

Soit les matrices $A, B \in \mathbb{M}(\mathbb{K}, m, n)$:

$$A = (a_{ij})_{i,j}$$
$$B = (b_{ij})_{i,j}$$

On dit que A est inférieure à B, et on le note $A \le B$, si :

$$a_{ij} \le b_{ij}$$

pour tout $(i, j) \in \{1, 2, ..., m\} \times \{1, 2, ..., n\}$.

48.13 Addition et soustraction

Soit les matrices de tailles identiques $A, B \in \mathbb{M}(\mathbb{K}, m, n)$:

$$A = (a_{ij})_{i,j}$$
$$B = (b_{ij})_{i,j}$$

L'addition et la soustraction des matrices sont simplement définies par l'addition et la soustraction des composantes :

$$A + B = (a_{ij} + b_{ij})_{i,j}$$
$$A - B = (a_{ij} - b_{ij})_{i,j}$$

On a donc clairement :

$$A + B = B + A$$

48.13.1 Neutre

La matrice nulle 0 est le neutre pour cette opération :

$$A + 0 = A$$

On en déduit que tous ses éléments doivent être nuls :

$$0 = (0)_{i,j}$$

On note $0_{m,n}$ au lieu de 0 lorsqu'on veut préciser que 0 est de taille (m, n).

48.13.2 Opposé

L'opposé de A, noté $-A$, est tel que :

$$A + (-A) = 0$$

On a donc clairement :

$$-A = (-a_{ij})_{i,j}$$

48.14 Multiplication mixte

On définit la multiplication mixte $\cdot : \mathbb{K} \times \mathbb{M}(\mathbb{K}, m, n) \to \mathbb{M}(\mathbb{K}, m, n)$ par :

$$\beta \cdot A = (\beta \cdot a_{ij})_{i,j}$$

où $\beta \in \mathbb{K}$ et $A \in \mathbb{M}(\mathbb{K}, m, n)$.
Choissons également $\gamma \in \mathbb{K}$. On note comme d'habitude :

$$A \cdot \beta = \beta \cdot A$$
$$\gamma \cdot \beta \cdot A = (\gamma \cdot \beta) \cdot A$$
$$\beta A = \beta \cdot A$$

48.15 Conjuguée

La conjuguée d'une matrice $A = (a_{ij})_{i,j} \in \mathbb{M}(\mathbb{C}, m, n)$ est définie par la conjuguée des composantes :

$$\operatorname{conj} A = \bar{A} = (\bar{a}_{ij})_{i,j} = (\operatorname{conj} a_{ij})_{i,j}$$

Chapitre 49

Progressions matricielles

49.1 Géométrique

Soit une matrice carréé $A \in \mathbb{M}(\mathbb{K}, n, n)$.

$$G_N = \sum_{k=0}^{N} A^k = I + A + A^2 + ... + A^N$$

49.1.1 Forme gauche

$$G_N = I + A \cdot (I + A + ... + A^N) - A^{N+1}$$

$$G_N = I + A \cdot G_N - A^{N+1}$$

$$(I - A) \cdot G_N = I - A^{N+1}$$

$$(I - A) \cdot \sum_{k=0}^{N} A^k = I - A^{N+1}$$

49.1.2 Forme droite

$$G_N = I + (I + A + ... + A^N) \cdot A - A^{N+1}$$

$$G_N = I + G_N \cdot A - A^{N+1}$$

$$G_N \cdot (I - A) = I - A^{N+1}$$

$$\sum_{k=0}^{N} A^k \cdot (I - A) = I - A^{N+1}$$

49.1.3 Inversible

$$\sum_{k=0}^{N} A^k = (I - A)^{-1} \cdot (I - A^{N+1})$$

$$\sum_{k=0}^{N} A^k = (I - A^{N+1}) \cdot (I - A)^{-1}$$

Chapitre 50

Produits

Dépendances

— Chapitre ?? : Les réels
— Chapitre ?? : Les complexes
— Chapitre ?? : Les sommes

50.1 Introduction

Soit le corps $\mathbb{K}$ sur lequel est défini une relation d'ordre total ainsi des opérations d'addition, de multiplication, de soustraction, de division et de puissance similaires à celles de $\mathbb{R}$.

Soit un ensemble Ω et la fonction $f : \Omega \mapsto \mathbb{K}$. On note le produit de f sur X par :

$$\prod_{x \in X} f(x)$$

pour tout sous-ensemble $X \subseteq \Omega$. Il s'agit intuitivement du produit des $f(x) \in \mathbb{K}$ lorsque x parcourt X. Nous allons voir comment le formaliser.

50.2 Multiplicativité finie

Si deux ensembles $X, Y \subseteq \Omega$ ne se chevauchent pas :

$$X \cap Y = \emptyset$$

le produit sur l'union des deux est intuitivement le produit des produits sur chacun d'entre-eux :

$$\prod_{z \in X \cup Y} f(z) = \left[\prod_{z \in X} f(z) \right] \cdot \left[\prod_{z \in Y} f(z) \right]$$

50.3 Ensemble vide

Comme $X = X \cup \emptyset$ et $X \cap \emptyset = \emptyset$, on en déduit que :

$$\prod_{x \in X} f(x) = \prod_{x \in X \cup \emptyset} f(x) = \prod_{x \in X} f(x) \cdot \prod_{x \in \emptyset} f(x)$$

Le produit sur l'ensemble vide doit donc être le neutre pour la multiplication :

$$\prod_{x \in \emptyset} f(x) = 1$$

50.4 Singleton

Il semble également logique d'imposer que le produit sur un ensemble contenant un seul élément $a \in X$ soit égal à $f(a)$:

$$\prod_{x \in \{a\}} f(x) = f(a)$$

Voilà qui complète les caractéristiques génériques des produits.

50.5 Algorithme

Soit l'ensemble X non vide, ensemble dont nous voulons évaluer le produit. Choisissons un élément $a \in A$ et considérons la décomposition :

$$X = \{a\} \cup (X \setminus \{a\})$$

Comme l'intersection des deux ensembles du membre de droite est vide :

$$\{a\} \cap (X \setminus \{a\}) = \emptyset$$

on peut écrire :

$$\prod_{x \in X} f(x) = \prod_{x \in \{a\}} f(x) \cdot \prod_{x \in X \setminus \{a\}} f(x)$$

c'est-à-dire :

$$\prod_{x \in X} f(x) = f(a) \cdot \prod_{x \in X \setminus \{a\}} f(x)$$

On en déduit un algorithme itératif permettant d'estimer le produit :

$$S \approx \prod_{x \in X} f(x)$$

Nous partons de :

$$S_0 = 0$$
$$X_0 = X$$

A chaque étape k, nous choisissons $a_k \in X_k$ et nous adaptons notre estimation par :

$$S_{k+1} = f(a_k) \cdot S_k$$

On retire ensuite a_k de X_k pour éviter de le compter plus d'une fois, ce qui nous donne l'ensemble suivant :

$$X_{k+1} = X_k \setminus \{a_k\}$$

50.6 Ensemble fini

Soit l'ensemble X contenant un nombre fini $N \in \mathbb{N}$ d'éléments :

$$X = \{a_1, a_2, ..., a_N\}$$

En appliquant l'algorithme d'évaluation d'une somme, on finit par arriver à l'itération N avec :

$$X_N = \emptyset$$

On a simplement :

$$\prod_{x \in X} f(x) = S_N \cdot \prod_{x \in \emptyset} f(x) = S_N \cdot 1 = S_N$$

et :

$$\prod_{x \in X} f(x) = S_N = f(a_1) \cdot f(a_2) \cdot ... \cdot f(a_N)$$

On note :

$$\prod_{k=1}^{N} f(a_k) = f(a_1) \cdot f(a_2) \cdot ... \cdot f(a_N)$$

50.6.1 Numérotation

Les éléments de X peuvent être numérotés différemment. Soit $m, n \in \mathbb{Z}$ et :

$$X = \{a_m, a_{m+1}, ..., a_{n-1}, a_n\}$$

On a alors :

$$\prod_{x \in X} f(x) = f(a_m) \cdot f(a_{m+1}) \cdot ... \cdot f(a_{n-1}) \cdot f(a_n)$$

On note :

$$\prod_{k=m}^{n} f(a_k) = f(a_m) \cdot f(a_{m+1}) \cdot ... \cdot f(a_{n-1}) \cdot f(a_n)$$

50.6.2 Extension

Soit un ensemble X dont on peut extraire un sous-ensemble fini de la forme :

$$F = \{a_m, a_{m+1}, ..., a_{n-1}, a_n\} \subseteq X$$

tel que :

$$f(x) = 1$$

pour tout $x \in X \setminus F$. La somme s'écrit alors :

$$\prod_{x \in X} f(x) = \prod_{x \in F} f(x) = \prod_{k=m}^{n} f(a_k)$$

50.7 Ensemble dénombrable

50.7.1 Naturels

Soit :

$$X = \{a_0, a_1, a_2, ...\} = \{a_k : k \in \mathbb{N}\}$$

Si la suite des $\{S_n : n \in \mathbb{N}\}$ définie par :

$$S_n = \prod_{k=0}^{n} f(a_k) = f(a_0) \cdot f(a_1) \cdot ... \cdot f(a_n)$$

pour tout $n \in \mathbb{N}$ converge vers un certain $S \in \mathbb{K}$, on définit :

$$\prod_{x \in X} f(x) = \lim_{n \to \infty} S_n = S$$

c'est-à-dire :

$$\prod_{x \in X} f(x) = \lim_{n \to \infty} \prod_{k=0}^{n} f(a_k)$$

On introduit la notation :

$$\prod_{k=0}^{+\infty} f(a_k) = \lim_{n \to \infty} \prod_{k=0}^{n} f(a_k)$$

50.7.2 Entiers

Soit :

$$X = \{..., a_{-2}, a_{-1}, a_0, a_1, a_2, ...\} = \{a_k : k \in \mathbb{Z}\}$$

Si la suite des $\{S_n : n \in \mathbb{N}\}$ définie par :

$$S_n = \prod_{k=-n}^{n} f(a_k) = f(a_{-n}) \cdot f(a_{-n+1}) \cdot ... \cdot f(a_{n-1}) \cdot f(a_n)$$

pour tout $n \in \mathbb{N}$ converge vers un certain $S \in \mathbb{K}$, on définit :

$$\prod_{x \in X} f(x) = \lim_{n \to \infty} S_n = S$$

On introduit la notation :

$$\prod_{k=-\infty}^{+\infty} f(a_k) = \lim_{n \to \infty} \prod_{k=-n}^{n} f(a_k)$$

50.7.3 Extension

Soit un ensemble X dont on peut extraire un sous-ensemble D de la forme :

$$D = \{a_k : k \in \mathbb{N}\} \subseteq X$$

ou :

$$D = \{a_k : k \in \mathbb{Z}\} \subseteq X$$

tel que :

$$f(x) = 1$$

pour tout $x \in X \setminus D$. La somme s'écrit alors :

$$\prod_{x \in X} f(x) = \prod_{x \in D} f(x)$$

Chapitre 51

Polynômes

Dépendances

— Chapitre ?? : Les réels
— Chapitre ?? : Les complexes
— Chapitre ?? : Les sommes
— Chapitre ?? : Les produits

51.1 Définition

Soit un corps $\mathbb{K}$. On dit qu'une fonction $p : \mathbb{K} \mapsto \mathbb{K}$ est un polynôme de degré n si on peut trouver des éléments $a_0, a_1, ...a_n \in \mathbb{K}$ tels que :

$$p(x) = \sum_{i=0}^{n} a_i \cdot x^i$$

pour tout $x \in \mathbb{K}$. On note $\mathrm{Poly}(\mathbb{K}, n)$ l'espace des polynômes de degré n définis sur $\mathbb{K}$.

Un cas particulier important est celui des polynômes réels :

$$\mathcal{P}_n = \mathrm{Poly}(\mathbb{R}, n)$$

51.2 Monôme

Un monôme $\mu_i : \mathbb{K} \mapsto \mathbb{K}$ est une fonction de la forme :

$$\mu_i(x) = x^i$$

pour tout $x \in \mathbb{K}$.

51.3 Racines

On dit que r est une racine du polynôme p si :

$$p(r) = 0$$

Nous allons montrer que tout polynôme de degré n non nul admet au plus n racines distinctes.

Ce qui revient à dire que si un polynôme de degré n admet $m + 1$ racines distinctes avec $m \ge n$, alors ce polynôme est forcément nul.

Soit $n = 0$. Considérons un polynôme p_0 défini par $p_0(x) = a_0$ pour un certain $a_0 \in \mathbb{K}$. Comme $n + 1 = 1$, on peut trouver au moins un $x \in \mathbb{K}$ tel que $p_0(x_0) = a_0 = 0$. On en conclut que $a_0 = 0$

et que $p_0(x) = 0$ pour tout $x \in \mathbb{K}$, ce qui revient à dire que $p_0 = 0$. La thèse est donc vérifiée pour $n = 0$.

A présent, supposons que l'affirmation soit vraie pour $n-1$. Choisissons un polynôme p_n de degré n :

$$p_n(x) = \sum_{i=0}^{n} a_i \cdot x^i$$

Supposons que p possède $m+1$ racines

$$p_n(x_0) = p_n(x_1) = ... = p_n(x_m) = 0$$

avec $m \geq n$ et :

$$x_0 \ < \ x_1 \ < \ ... \ < \ x_m$$

On a alors :

$$p_n(x) = p_n(x) - p_n(x_0) = \sum_{i=1}^{n} a_i \cdot (x^i - x_0^i)$$

Mais les propriétés de factorisation des progressions géométriques (voir la section 35.3) nous permettent d'affirmer que :

$$x^i - x_0^i = (x - x_0) \sum_{j=0}^{i-1} x^j \cdot x_0^{i-1-j}$$

on en déduit que :

$$p_n(x) = (x - x_0) \cdot q_{n-1}(x)$$

où le polynôme q_{n-1} est défini par :

$$q_{n-1}(x) = \sum_{i=1}^{n} a_i \sum_{j=0}^{i-1} x^j \cdot x_0^{i-1-j}$$

Cette expression ne faisant intervenir que les puissances $1, x, ..., x^{n-1}$, on voit que $q_{n-1} \in \mathcal{P}_n$ est de degré $n-1$. Considérons les cas des racines $x = x_1, x_2, ..., x_m$ de p_n. On a :

$$0 = p_n(x_k) = (x_k - x_0) \cdot q_{n-1}(x_k)$$

Comme $x_k - x_0 \neq 0$, on peut multiplier par $(x_k - x_0)^{-1}$ pour obtenir :

$$q_{n-1}(x_k) = \frac{p_{n-1}(x_k)}{x_k - x_0} = 0$$

Le polynôme q_{n-1} possède par conséquent au moins $m - 1 \geq n - 1$ racines distinctes. Il est dès lors nul par l'hypothèse de récurrence. La factorisation de p_n devient alors :

$$p_n(x) = (x - x_0) \cdot q_{n-1}(x) = 0$$

pour tout $x \in \mathbb{K}$. Le polynôme p_n est également nul et la thèse est démontrée.

Egalité

Si deux polynômes $p_1, p_2 \in \mathcal{P}_n$ sont égaux en $m+1 \geq n+1$ points distincts, alors le polynôme de degré n :

$$h = p_1 - p_2$$

admet $m+1$ racines. Il est donc nul et $p_1 = p_2$.

51.4 Factorisation

Choisissons un polynôme $p \in \mathcal{P}_n$ qui admet n racines distinctes $x_1, x_2, ..., x_n$. Soit alors $x_0 \notin \{x_1, ..., x_n\}$ et :

$$A = \frac{p(x_0)}{\prod_{i=1}^n (x_0 - x_i)}$$

On voit que le polynôme :

$$q(x) = A \cdot \prod_{i=1}^n (x - x_i)$$

est égal à p en chacune des racines :

$$p(x_i) = q(x_i) = 0 \qquad i = 1, 2, ..., n$$

Mais par la définition de A, on a aussi :

$$p(x_0) = q(x_0)$$

Les deux polynômes sont donc égaux en $n+1$ points distincts. Comme ils sont tous deux de degré n, on en déduit que $p = q$. Les coefficients doivent donc tous être égaux, et comme :

$$p(x) = a_n \cdot x^n + \sum_{i=0}^{n-1} a_i \cdot x^i$$

$$q(x) = A \cdot x^n + \sum_{i=0}^{n-1} b_i(x_1, ..., x_n) \cdot x^i$$

on a $A = a_n$ et :

$$p(x) = a_n \cdot \prod_{i=1}^n (x - x_i)$$

51.5 Binômes canoniques

Le binôme canonique de degré $n \in \mathbb{N}$ est une fonction $b_n : \mathbb{K} \to \mathbb{K}$ définie par :

$$b_n(x) = (1+x)^n$$

pour tout $x \in \mathbb{K}$. On a par exemple :

$$\begin{aligned} b_0(x) &= (1+x)^0 = 1 \\ b_1(x) &= (1+x)^1 = 1 + x \\ b_2(x) &= (1+x)^2 = (1+x) \cdot (1+x) = 1 + 2\ x + x^2 \end{aligned}$$

On voit donc que le binôme canonique de degré n peut s'écrire comme :

$$b_n(x) = (1+x)^n = \sum_{k=0}^{n} a_{nk} \cdot x^k$$

pour certains coefficients $a_{nk} \in \mathbb{K}$. On nomme ces coefficients les « nombres binômiaux », et on les note :

$$\left\{ \begin{array}{c} n \\ k \end{array} \right\} = a_{nk}$$

L'expression de b_0 nous donne :

$$\left\{ \begin{array}{c} 0 \\ 0 \end{array} \right\} = 1$$

On peut évaluer récursivement les nombres binômiaux d'ordres plus élevés en utilisant la définition de la puissance :

$$b_n(x) = (1+x) \cdot b_{n-1}(x)$$

Il vient alors :

$$\begin{aligned}
\sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ k \end{array} \right\} \cdot x^k &= (1+x) \sum_{k=0}^{n-1} \left\{ \begin{array}{c} n-1 \\ k \end{array} \right\} \cdot x^k \\
&= \sum_{k=0}^{n-1} \left\{ \begin{array}{c} n-1 \\ k \end{array} \right\} \cdot x^k + \sum_{k=0}^{n-1} \left\{ \begin{array}{c} n-1 \\ k \end{array} \right\} \cdot x^{k+1} \\
&= \sum_{k=0}^{n-1} \left\{ \begin{array}{c} n-1 \\ k \end{array} \right\} \cdot x^k + \sum_{i=1}^{n} \left\{ \begin{array}{c} n-1 \\ i-1 \end{array} \right\} \cdot x^i
\end{aligned}$$

et finalement :

$$\begin{aligned}
\sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ k \end{array} \right\} \cdot x^k = \left\{ \begin{array}{c} n-1 \\ 0 \end{array} \right\} &+ \sum_{k=1}^{n-1} \left[\left\{ \begin{array}{c} n-1 \\ k \end{array} \right\} + \left\{ \begin{array}{c} n-1 \\ k-1 \end{array} \right\} \right] \cdot x^k \\
&+ \left\{ \begin{array}{c} n-1 \\ n-1 \end{array} \right\} \cdot x^n
\end{aligned}$$

En égalisant les coefficients des $x^0 = 1$, nous avons :

$$\left\{ \begin{array}{c} n \\ 0 \end{array} \right\} = \left\{ \begin{array}{c} n-1 \\ 0 \end{array} \right\}$$

On en déduit par récurrence que :

$$\left\{ \begin{array}{c} n \\ 0 \end{array} \right\} = 1$$

En égalisant les coefficients des x^n, nous avons :

$$\left\{ \begin{array}{c} n \\ n \end{array} \right\} = \left\{ \begin{array}{c} n-1 \\ n-1 \end{array} \right\}$$

On en déduit par récurrence que :

$$\left\{ \begin{array}{c} n \\ n \end{array} \right\} = 1$$

En égalisant les coefficients de x^k pour $k \in \{1, ..., n-1\}$, il vient :

$$\left\{ \begin{array}{c} n \\ k \end{array} \right\} = \left\{ \begin{array}{c} n-1 \\ k \end{array} \right\} + \left\{ \begin{array}{c} n-1 \\ k-1 \end{array} \right\}$$

Il est donc facile d'évaluer les coefficients de b_n à partir des coefficients de b_{n-1}.

51.6 Binômes génériques

Le binôme générique de degré $n \in \mathbb{N}$ est une fonction $B_n : \mathbb{K} \times \mathbb{K} \mapsto \mathbb{K}$ définie par :

$$B_n(x, y) = (x + y)^n$$

pour tout $(x, y) \in \mathbb{K}^2$. Nous avons la forme équivalente :

$$B_n(x, y) = y^n \cdot \left(1 + \frac{x}{y}\right)^n = y^n \cdot b_n\left(\frac{x}{y}\right)$$

c'est-à-dire :

$$B_n(x, y) = \sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ k \end{array} \right\} \cdot x^k \cdot y^{n-k}$$

51.7 Second degré

Le binôme du second degré est couramment utilisé :

$$(x + y)^2 = x\ (x + y) + y\ (x + y) = x^2 + 2\ x\ y + y^2$$

51.8 Symétrie

Par commutativité de l'addition, on a $B_n(x, y) = B_n(y, x)$ et :

$$\sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ k \end{array} \right\} \cdot x^k \cdot y^{n-k} = \sum_{i=0}^{n} \left\{ \begin{array}{c} n \\ i \end{array} \right\} \cdot y^i \cdot x^{n-i}$$

Procédant au changement d'indice $n - i = k$, il vient :

$$\sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ k \end{array} \right\} \cdot x^k \cdot y^{n-k} = \sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ n-k \end{array} \right\} \cdot x^k \cdot y^{n-k}$$

On en déduit en égalisant les coefficients de x^k que :

$$\left\{ \begin{array}{c} n \\ k \end{array} \right\} = \left\{ \begin{array}{c} n \\ n-k \end{array} \right\}$$

51.9 Cas particuliers

En considérant le cas particuliers $x = y = 1$, on constate que :

$$\sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ k \end{array} \right\} = (1+1)^n = 2^n$$

Pour $x = -1$, $y = 1$, on a :

$$\sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ k \end{array} \right\} \cdot (-1)^k = (-1+1)^n = 0^n = 0$$

Lorsque $y = 1 - x$ on arrive à :

$$\sum_{k=0}^{n} \left\{ \begin{array}{c} n \\ k \end{array} \right\} \cdot x^k \cdot (1-x)^{n-k} = (x+1-x)^n = 1^n = 1$$

51.10 Bernstein

Soit $i, n \in \mathbb{N}$ avec $i \le n$. Les polynômes de Bernstein B_i^n sont définis par :

$$B_i^n(t) = \left\{ \begin{array}{c} n \\ i \end{array} \right\} \cdot t^i \cdot (1-t)^{n-i}$$

pour tout $t \in [0, 1]$.

Considérons l'espace fonctionnel $\mathcal{F} = \mathbb{F}([0, 1], \mathbb{K})$. L'opérateur de Bernstein $\mathcal{B}_n : \mathcal{F} \mapsto \mathcal{F}$ est défini par :

$$\mathcal{B}_n(f)(t) = \sum_{i=0}^{n} f(i/n) \cdot B_i^n(t)$$

pour tout $f \in \mathcal{F}$ et pour tout $t \in [0, 1]$.

Soit la fonction constante $c \in \mathcal{F}$ associée à un certain $c \in \mathbb{K}$ et définie par :

$$c(t) = c$$

pour tout $t \in [0, 1]$.

L'opérateur de Bernstein possède l'importante propriété de conserver ces fonctions constantes :

$$\mathcal{B}_n(c)(t) = \sum_{i=0}^{n} c \cdot B_i^n(t) = c \sum_{i=0}^{n} \left\{ \begin{array}{c} n \\ i \end{array} \right\} t^i \cdot (1-t)^{n-i} = c$$

quelle que soit la valeur de $t \in [0, 1]$.

51.11 Division Euclidienne

Soit deux polynomes $a \in \mathcal{P}_n$ et $b \in \mathcal{P}_m$ avec $m \le n$ et :

$$a(x) = \sum_{i=0}^{n} a_i \cdot x^i$$

$$b(x) = \sum_{i=0}^{m} b_i \cdot x^i$$

On dit que q et r sont respectivement le quotient et le reste de la division euclidienne de a et b, et on note :

$$(q, r) = \mathrm{division}(a, b)$$

si :

$$a(x) = b(x) \cdot q(x) + r(x)$$

$$q(x) = \sum_{i=0}^{n-m} q_i x^i$$

$$r(x) = \sum_{i=0}^{p} r_i x^i$$

avec $p < m \le n$.

Cinquième partie

Vecteurs

Chapitre 52

Espaces vectoriels

Dépendances

- Chapitre 10 : Les structures algébriques
- Chapitre **??** : Les sommes

52.1 Introduction

Soit un corps $\mathbb{K}$, un ensemble quelconque A et $n \in \mathbb{N}$. Le but des espaces vectoriels est de fournir un cadre général aux n-tuples de $\mathbb{K}^n$ et aux fonctions de $\mathbb{K}^A$. Nous avons vu la correspondance $\mathbb{K}^n \leftrightarrow \mathbb{K}^A$ dans le cas particulier où A possède un nombre fini d'éléments. Mais le lien entre les deux types d'objets ne s'arrête pas là : la comparaison de deux fonctions se base sur le même principe (étendu) que la comparaison de deux n-tuples. Nous avons également défini des produits mixtes $\cdot : \mathbb{K} \times \mathbb{K}^n \to \mathbb{K}^n$ et $\cdot : \mathbb{K} \times \mathbb{K}^A \to \mathbb{K}^A$ semblables. Les matrices représentant des applications linéaires, nous pouvons également les ajouter dans la liste. Si $u, v \in E$, avec $E \in \{\mathbb{K}^n, \mathbb{K}^A, \mathbb{M}(\mathbb{K}, m, n)\}$, on a :

$$\begin{aligned}(\alpha \cdot \beta) \cdot u &= \alpha \cdot (\beta \cdot u)\\(\alpha + \beta) \cdot x &= \alpha \cdot x + \beta \cdot x\\\alpha \cdot (u + v) &= \alpha \cdot u + \alpha \cdot v\\1 \cdot u &= 1\end{aligned}$$

pour tout $\alpha, \beta \in \mathbb{K}$. De plus l'addition induite sur E par l'addition de $\mathbb{K}$ transforme E en groupe commutatif. On ne peut toutefois pas parler de corps pour E, car la multiplication matricielle n'est pas une multiplication induite, et est non commutative.

Attention

Ne pas confondre les additions définies sur E et $\mathbb{K}$, ni la multiplication de $\mathbb{K}$ avec la multiplication mixte, ni le neutre de E avec celui de $\mathbb{K}$. Lorsqu'il y a un risque d'ambiguité, on parle du vecteur nul $0 \in E$ et du scalaire nul $0 \in \mathbb{K}$.

52.2 Définition

Soit un groupe commutatif pour l'addition E, ainsi qu'un corps $\mathbb{K}$. Si il existe une opération de multiplication mixte $\cdot : \mathbb{K} \times E \mapsto E$ vérifiant les propriétés 52.1 ci-dessus, on dit que E est un espace vectoriel sur $\mathbb{K}$. On nomme alors « vecteurs » les éléments de E et « scalaires » les éléments de $\mathbb{K}$.

Notation

On note aussi :

$$\begin{gathered}
x - y = x + (-1) \cdot y \\
x \cdot \alpha = \alpha \cdot x \\
\alpha \cdot \beta \cdot x = (\alpha \cdot \beta) \cdot x \\
\alpha x = \alpha \cdot x \\
\frac{x}{\alpha} = \alpha^{-1} \cdot x
\end{gathered}$$

Lorsque α a un inverse dans $\mathbb{K}$, on a même les « fractions » :

$$\frac{x}{\alpha} = \frac{1}{\alpha} \cdot x$$

Corollaires

Les propriétés de la multiplication mixte nous montrent directement que :

$$\begin{gathered}
0 \cdot u = (1 - 1) \cdot u = u - u = 0 \\
\alpha \cdot 0 = \alpha \cdot (u - u) = \alpha - \alpha = 0
\end{gathered}$$

Remarque

Le corps $\mathbb{K}$ est souvent $\mathbb{R}$ ou $\mathbb{C}$.

52.3 Sous-espace

On dit que $F \subseteq E$ est un sous-espace vectoriel de E si $0 \in F$ et si :

$$z = \alpha \cdot x + \beta \cdot y$$

appartient à F quels que soient les vecteurs $x, y \in F$ et les scalaires $\alpha, \beta \in \mathbb{K}$.
On vérifie par exemple que E est un sous-espace vectoriel de lui-même.

52.4 Espace engendré

L'espace engendré par les vecteurs $e_1, e_2, ..., e_n \in E$ est l'ensemble des combinaisons linéaires formées à partir des e_i :

$$\operatorname{span}\{e_1, ..., e_n\} = \left\{ \sum_{i=1}^{n} \alpha_i \cdot e_i : \alpha_i \in \mathbb{K} \right\}$$

On vérifie que $\operatorname{span}\{e_1, ..., e_n\}$ est un sous-espace vectoriel de E.

Remarque

Les espaces vectoriels ne pouvant pas s'exprimer comme ci-dessus sont dit de dimension infinie.

52.5 Indépendance linéaire

On dit qu'une série de vecteurs $e_1, ..., e_n$ est linéairement indépendante si pour toute suite de scalaires α_i, la condition :

$$\sum_{i=1}^{n} \alpha_i \cdot e_i = 0$$

implique que tous les scalaires soient nuls :

$$\alpha_i = 0$$

pour tout $i \in \{1, 2, ..., n\}$.

52.6 Coordonnées

Soit les vecteurs linéairement indépendants $(e_1, ..., e_n)$ et $x \in \operatorname{span}\{e_1, ..., e_n\}$. On peut trouver une suite de scalaire α_i tels que :

$$x = \sum_{i=1}^{n} \alpha_i \cdot e_i$$

Supposons que l'on ait également :

$$x = \sum_{i=1}^{n} \beta_i \cdot e_i$$

pour une autre suite de scalaires β_i. En soustrayant les deux équations, on obtient :

$$\sum_{i=1}^{n} (\alpha_i - \beta_i) \cdot e_i = 0$$

L'indépendance linéaire des e_i implique alors que $\alpha_i - \beta_i = 0$, c'est-à-dire :

$$\alpha_i = \beta_i$$

pour tout $i \in \{1, 2, ..., n\}$.

On a donc unicité des coefficients scalaire de la combinaison linéaire. On dit que les α_i sont les coordonnées de x par rapport aux $(e_1, ..., e_n)$.

Base

Par contre, l'existence de telles coordonnées n'est pas garantie pour tout $x \in E$. Ce ne sera le cas que si :

$$E \subseteq \operatorname{span}\{e_1, ..., e_n\}$$

On dit alors que $(e_1, ..., e_n)$ forme une base de E.

Dimension finie

On dit qu'un espace vectoriel E est de dimension finie s'il posséde au moins une base de la forme $(e_1, ..., e_n)$, où $n \in \mathbb{N}$ est fini. Dans le cas où E *ne possède pas* une telle base, il est dit de dimension infinie.

Equivalence

On voit qu'étant donné une base de E, il y a équivalence entre un vecteur $x \in E$ et un élément $(x_1, x_2, ..., x_n) \in \mathbb{K}^n$ formé par ses coordonnées.

Nous noterons donc également (et abusivement) $x = (x_1, x_2, ..., x_n)$, mais attention : il ne faut jamais perdre de vue que les x_i dépendent de la base utilisée. Le vecteur x est lui invariant sous changement de base.

52.7 Absence de redondance

Soit $e_1, ..., e_n \in E$ une suite de vecteurs linéairement indépendants. Soit $i \in \{1, 2, ..., n\}$ et :

$$J(i) = \mathbb{Z}(0, n) \setminus \{i\}$$

Supposons que le vecteur e_i soit une combinaison des autres vecteurs :

$$e_i = \sum_{j \in J(i)} \alpha_j \cdot e_j$$

On a donc :

$$e_i - \sum_{j \in J(i)} \alpha_j \cdot e_j = 0$$

L'hypothèse d'indépendance linéaire voudrait que tous les α_j et le α_i soient nuls. Ce qui n'est manifestement pas le cas puisque $\alpha_i = 1 \neq 0$!

Aucun des vecteurs de la suite n'est donc combinaison des autres. On dit qu'aucun vecteur n'est redondant dans la suite.

52.8 Base canonique sur $\mathbb{K}^n$

Soit $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. On note $\mathfrak{c}_i$ l'élément de $\mathbb{K}^n$ ayant un 1 en i^{me} position et des 0 partout ailleurs. On a donc :

$$\begin{aligned} \mathfrak{c}_1 &= (1, 0, ..., 0) \\ \mathfrak{c}_2 &= (0, 1, 0, ..., 0) \\ &\vdots \\ \mathfrak{c}_n &= (0, ..., 0, 1) \end{aligned}$$

On a alors, pour tout $x = (x_1, ..., x_n) \in \mathbb{K}^n$:

$$x = \sum_{i=1}^{n} x_i \cdot \mathfrak{c}_i$$

52.9 Représentation matricielle

On représente généralement les vecteurs de $\mathbb{K}^n$ par des vecteurs lignes ou colonnes. On parle alors de « vecteurs matriciels ». Le i^{me} vecteur de la base canonique est défini par le vecteur colonne :

$$\mathfrak{c}_i = (\boldsymbol{\delta}_{ij})_j = \begin{bmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}$$

soit :

$$\begin{gathered} \mathfrak{c}_1 = [1 \ 0 \ \ldots \ 0]^T \\ \mathfrak{c}_2 = [0 \ 1 \ \ldots \ 0]^T \\ \vdots \\ \mathfrak{c}_n = [0 \ \ldots \ 0 \ 1]^T \end{gathered}$$

Chapitre 53

Norme

Dépendances

— Chapitre ?? : Les distances
— Chapitre 52 : Les espaces vectoriels

53.1 Norme

Soit S un corps muni d'un ordre $\le$ et E un espace vectoriel sur $\mathbb{K}$. Une norme $\|.\| : E \to \mathbb{R}$ est intuitivement la « grandeur » d'un vecteur $x \in E$. Cette grandeur correspond à la distance séparant le vecteur nul de x :

$$\|x\| \equiv \mathfrak{dist}(x, 0)$$

Soit $x, y, z \in E$.

Comme $\mathfrak{dist}(x, 0) \ge 0$, on impose par analogie que :

$$\|x\| \ge 0$$

De plus, le seul vecteur x de E vérifiant :

$$\|x\| = 0$$

implique que notre distance équivalente $\mathfrak{dist}(x, 0) = 0$ soit nulle, ce qui n'est possible que si $x = 0$. Lorsqu'on fait référence à ces conditions, on dit que la norme est strictement définie positive.

Considérons à présent l'inégalité triangulaire :

$$\mathfrak{dist}(0, x + y) \le \mathfrak{dist}(0, x) + \mathfrak{dist}(x, x + y)$$

Comment faire correspondre $\mathfrak{dist}(x, x + y)$ à une norme ? En demandant simplement que notre distance particulière soit invariante sous translation de x :

$$\mathfrak{dist}(x, x + y) = \mathfrak{dist}(x - x, x + y - x) = \mathfrak{dist}(0, y) \equiv \|y\|$$

On impose donc l'inégalité triangulaire :

$$\|x + y\| \le \|x\| + \|y\|$$

Par ailleurs, lorsqu'on allonge ou réduit un vecteur x d'un facteur $\alpha \in \mathbb{K}$, la distance parcourue sur x devra être allongée ou réduite par la valeur absolue de α :

$$\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$$

53.2 Borne inférieure

On déduit de la définition que :

$$\|-x\| = \|(-1)\cdot x\| = |-1|\cdot\|x\| = \|x\|$$

En posant $z = x + y$, on a :

$$\|z\| \le \|x\| + \|y\| = \|x\| + \|z - x\|$$

c'est-à-dire :

$$\|z - x\| \ge \|z\| - \|x\|$$

Mais comme $\|z - x\| = \|(-1)\cdot(z - x)\| = \|x - z\|$, la propriété vaut également en interchangeant z et x, et on obtient :

$$\|z - x\| \ge \max\{\|z\| - \|x\|, \|x\| - \|z\|\}$$

53.3 Distance associée

On peut associer une distance d à une norme $\|.\|$ en posant :

$$\mathfrak{dist}(x, y) = \|x - y\|$$

En effet :
- $\mathfrak{dist}(x, x) = \|x - x\| = \|0\| = 0$
- si $\mathfrak{dist}(x, y) = 0 = \|x - y\|$, on a forcément $x - y = 0$ et donc $x = y$
- $\mathfrak{dist}(x, y) = \|x - y\| = \|y - x\| = \mathfrak{dist}(y, x)$
- $\mathfrak{dist}(x, y) + \mathfrak{dist}(y, z) = \|x - y\| + \|y - z\| \ge \|x - y + y - z\| = \|x - z\| = \mathfrak{dist}(x, z)$

53.4 Normalisation

On peut toujours normaliser un vecteur $w \ne 0$ pour obtenir un vecteur u de norme 1. Comme $\|w\| \ne 0$, on peut écrire :

$$u = \frac{w}{\|w\|}$$

On a alors :

$$\|u\| = \left\|\frac{w}{\|w\|}\right\| = \frac{1}{\|w\|}\cdot\|w\| = 1$$

Chapitre 54

Espaces de Banach

54.1 Définition

On dit qu'un espace vectoriel X est un espace de Banach si il est complet pour la distance issue de la norme $\mathfrak{dist}(x,y) = \|x-y\|$. Dans la suite, nous considérons un espace de Banach X sur $\mathbb{R}$ ou $\mathbb{C}$.

54.2 Application contractante

On dit qu'une application $A : X \mapsto X$ est contractante s'il existe un $c \in [0,1[\subseteq \mathbb{R}$ tel que :

$$\mathfrak{dist}\big(A(u), A(v)\big) \le c \cdot \mathfrak{dist}(u,v)$$

pour tout $u, v \in X$.

54.3 Suite de Cauchy

Soit une application contractante $A : X \mapsto X$ et $u_0 \in X$. On définit la suite $u_0, u_1, u_2, ...$ par :

$$u_n = A(u_{n-1}) = ... = A^n(u_0)$$

pour tout $n \in \mathbb{N}$. On a alors :

$$\mathfrak{dist}(u_{n+1}, u_n) \le c \cdot \mathfrak{dist}(u_n, u_{n-1}) \le ... \le c^n \cdot \mathfrak{dist}(u_1, u_0)$$

Soit $m, n \in \mathbb{N}$. Les propriétés des distances nous permettent d'écrire :

$$\mathfrak{dist}(u_{n+m}, u_n) \le \sum_{i=0}^{m-1} \mathfrak{dist}(u_{n+i+1}, u_{n+i})$$

Mais comme $\mathfrak{dist}(u_{n+i+1}, u_{n+i}) \le c^{n+i} \cdot \mathfrak{dist}(u_1, u_0)$, on a :

$$\begin{aligned}
\mathfrak{dist}(u_{n+m}, u_n) &\le \sum_{i=0}^{m-1} c^{n+i} \cdot \mathfrak{dist}(u_1, u_0) \\
&\le c^n \cdot \mathfrak{dist}(u_1, u_0) \cdot \sum_{i=0}^{m-1} c^i \\
&\le c^n \cdot \frac{1-c^m}{1-c}
\end{aligned}$$

Finalement, comme $1 - c^m \le 1$ quel que soit $m \in \mathbb{N}$ on obtient une expression qui ne dépend pas de m :

$$\mathfrak{dist}(u_{n+m}, u_n) \le \frac{c^n}{1-c} \cdot \mathfrak{dist}(u_1, u_0)$$

Les éléments de la suite sont donc de plus en plus proche l'un de l'autre lorsque n augmente. Soit $\epsilon > 0$. Comme la suite c^n converge vers 0 lorsque $n \to \infty$, on peut toujours trouver N tel que :

$$c^N \le \frac{\epsilon \cdot (1-c)}{\mathfrak{dist}(u_1, u_0)}$$

Il suffit donc de choisir $i, j \in \mathbb{N}$ tels que $i, j \ge N$ pour avoir :

$$\mathfrak{dist}(u_i, u_j) = \mathfrak{dist}(u_j, u_i) \le \epsilon$$

On en conclut que la suite des u_n est de Cauchy.

54.4 Point fixe

Comme X est complet, notre suite u_n étant de Cauchy converge vers une certaine limite :

$$p = \lim_{n \to \infty} u_n$$

appartenant à X. Analysons le comportement de $A(p)$. On a la borne supérieure :

$$\mathfrak{dist}\big(A(p), p\big) \le \mathfrak{dist}\big(A(p), A(u_n)\big) + \mathfrak{dist}\big(A(u_n), u_n\big) + \mathfrak{dist}\big(u_n, p\big)$$

On sait déjà que $\mathfrak{dist}(u_n, p)$ converge vers 0 par définition de p. On sait aussi que $\mathfrak{dist}\big(A(u_n), u_n\big) = \mathfrak{dist}(u_{n+1}, u_n) \to 0$. On a également :

$$\mathfrak{dist}\big(A(p), A(u_n)\big) \le c \cdot \mathfrak{dist}\big(p, u_n\big)$$

On en conclut que la suite $A(u_n)$ converge vers $A(p)$:

$$A(p) = \lim_{n \to \infty} A(u_n)$$

Les trois termes de la borne supérieure convergeant chacun vers 0, cette borne est aussi petite que l'on veut lorsque n est assez grand. On a donc $\mathfrak{dist}\big(A(p), p\big) = 0$ et :

$$A(p) = p$$

L'élément $p \in X$ est un point fixe de A.

54.5 Unicité

Si p_1 et p_2 sont deux points fixes, on a :

$$A(p_1) = p_1$$
$$A(p_2) = p_2$$

et :

$$\mathfrak{dist}(p_1, p_2) \le c \cdot \mathfrak{dist}\big(A(p_1), A(p_2)\big) \le c \cdot \mathfrak{dist}\,(p_1, p_2)$$

Comme $c < 1$, ce n'est possible que si $\mathfrak{dist}(p_1, p_2) = 0$, c'est-à-dire :

$$p_1 = p_2$$

Le point fixe de A est unique.

54.6 Vitesse de convergence

Nous avons donc montré que la suite des u_n converge vers l'unique point fixe p de A, et ce quel que soit u_0. On a même la propriété suivante nous donnant une borne supérieure pour le taux de convergence :

$$\mathfrak{dist}(u_n, p) \le \mathfrak{dist}\big(A(u_{n-1}), A(p)\big) \le c \cdot \mathfrak{dist}(u_{n-1}, p) \le ... \le c^n \cdot \mathfrak{dist}(u_0, p)$$

Chapitre 55

Continuité

55.1 Dépendances

— Chapitre 55 : Les limites
— Chapitre ?? : Les réels

55.2 Fonctions continues

Une fonction $f : D \to F$ est dite continue en a si :

$$\lim_{\substack{x \to a \\ x \in D}} f(x) = f(a)$$

Soit $A \subseteq D$. On dit qu'une fonction est continue sur A si :

$$\lim_{\substack{x \to a \\ x \in A}} f(x) = f(a)$$

pour tout $a \in A$. On note $\mathrm{Cont}(A, F)$ l'ensemble des fonctions $f : A \mapsto F$ continues sur A.

Remarque

Si F est muni d'une norme, les limites s'évaluent au sens de la distance découlant de la norme.

55.3 Espace vectoriel

Si $\mathbb{K}$ est un corps, on vérifie que $\mathrm{Cont}(A, \mathbb{K})$ est un espace vectoriel sur $\mathbb{K}$. En effet, la fonction nulle 0 est continue. Si $\alpha, \beta \in \mathbb{K}$ et si $f, g \in \mathrm{Cont}(A, \mathbb{K})$, on a :

$$\begin{aligned}\lim_{x \to a}(\alpha \cdot f(x) + \beta \cdot g(x)) &= \alpha \cdot \lim_{x \to a} f(x) + \beta \cdot \lim_{x \to a} g(x) \\ &= \alpha \cdot f(a) + \beta \cdot g(a)\end{aligned}$$

pour tout $a \in A$. On en conclut que $\alpha \cdot f + \beta \cdot g$ est également continue.

55.4 Norme des fonctions continues

Si l'ensemble F est muni d'une norme, on peut définir la norme $\|.\|_{\mathrm{Cont}}$ d'une fonction continue u par :

$$\|u\|_{\mathrm{Cont}} = \sup \left\{ \|u(x)\| : x \in A \right\}$$

Notation

On note aussi :

$$\|u\|_\infty = \|u\|_{\mathrm{Cont}}$$

Convergence uniforme

Cette norme est surtout utilisée lorsqu'il s'agit de mesurer l'écart entre deux fonctions $f, g : A \to B$, en particulier lorsque g représente une approximation de f. Dans ce cas, l'écart $e = f - g$ représente l'erreur la plus élevée de l'estimation :

$$\|e\|_\infty = \|f - g\|_\infty = \sup\{\|f(x) - g(x)\| : x \in A\}$$

Lorsque cette norme particulière de l'erreur tend vers zéro, on parle de convergence uniforme.

55.5 Théorème des valeurs intermédiaires

Les fonctions continues possèdent l'importante propriété suivante.

Théorème 55.5.1.

— *Soit* $f \in \mathrm{Cont}(I, \mathbb{R})$ *où* $I = [a, b]$ *est un intervalle inclus dans* $\mathbb{R}$. *On suppose que :*

$$f(a) < f(b)$$

Soit le réel φ *vérifiant* $f(a) < \varphi < f(b)$. *On peut alors trouver un* $c \in]a, b[$ *tel que* $f(c) = \varphi$.

— *Soit* $g \in \mathrm{Cont}(I, \mathbb{R})$. *On suppose que :*

$$g(a) > g(b)$$

Soit le réel φ *vérifiant* $g(a) > \varphi > g(b)$. *On peut alors trouver un* $c \in]a, b[$ *tel que* $g(c) = \varphi$.

Démonstration 55.5.1. Nous allons démontrer ce résultat par l'absurde.

— Considérons le cas où $f(a) < \varphi < f(b)$. On définit les ensembles :

$$A^+ = \{x \in I : f(x) > \varphi\}$$
$$A^- = \{x \in I : f(x) < \varphi\}$$

Si aucun $c \in I$ ne vérifie $f(c) = \varphi$, on doit avoir clairement $A^+ \cup A^- = I$.
Nous définissons $\alpha = \sup A^-$. Comme $A^- \subseteq I$, on à clairement $\alpha \in I$. Si $\alpha \in A^-$, alors par continuité de f en α, on peut trouver $\delta > 0$ tel que :

$$|f(\alpha + \delta) - f(\alpha)| \le \epsilon = \frac{1}{2}(\varphi - f(\alpha))$$

On a alors clairement $f(\alpha + \delta) < \varphi$ et $\alpha < \alpha + \delta \in A^-$ ce qui contredit l'hypothèse de suprémum pour α.
On doit donc avoir $\alpha \notin A^-$. Mais alors $\alpha \in I \setminus A^- = A^+$. Donc $f(\alpha) > \varphi$. Par continuité de f en α, on peut trouver $\delta > 0$ tel que :

$$|f(\alpha) - f(x)| \le \epsilon = \frac{1}{2}(f(\alpha) - \varphi)$$

pour tout $x \in]\alpha - \delta, \alpha[$. On a alors clairement $f(x) > \varphi$ pour tout $x \in]\alpha - \delta, \alpha[$.
Soit $\beta \in]\alpha - \delta, \alpha[$. Par définition de α, on ne peut pas avoir $\beta \ge A^-$, donc il existe $\gamma \in]\beta, \alpha[\subseteq]\alpha - \delta, \alpha[$ tel que $\gamma \in A^-$. On a donc $f(\gamma) < \varphi$ ce qui contredit la propriété ci-dessus en $x = \gamma$.

— Considérons à présent le cas $g(a) > \varphi > g(b)$. On pose :

$$f = -g$$

On a :

$$f(a) < -\varphi < f(b)$$

On peut donc trouver un $c \in]a, b[$ tel que :

$$f(c) = -\varphi$$

On en conclut que :

$$g(c) = -f(c) = \varphi$$

Généralisation

Soit le réel φ tel que $f(a) \le \varphi \le f(b)$. Si $\varphi \in \{f(a), f(b)\}$, il suffit de prendre $c \in \{a, b\}$ pour avoir $f(a) \le f(c) \le f(b)$. Sinon, on applique le théorème des valeurs intermédiaires et on trouve un c vérifiant $f(a) < f(c) < f(b)$. Mais dans tous les cas, on pourra trouver un $c \in [a, b]$ tel que $f(a) \le f(c) \le f(b)$. De même si $f(a) > f(b)$.

55.6 Théorème de la bijection

Théorème 55.6.1. *Soit $f \in \mathrm{Cont}(I, J)$ où $I = [a, b]$ est un intervalle inclus dans $\mathbb{R}$ et où $J = f(I)$. Si f est strictement croissante (ou décroissante), alors f est inversible et :*

$$f(I) = [\alpha, \beta]$$

avec :

$$\alpha = \min\{f(a), f(b)\}$$
$$\beta = \max\{f(a), f(b)\}$$

Démonstration 55.6.1. Comme $f : I \mapsto J = f(I)$ est strictement croissante ou décroissante, on a vu dans le chapitre traitant des bijections que f est inversible. Choisissons un réel $\varphi \in [\alpha, \beta]$. Le théorème des valeurs intermédiaires nous dit qu'on peut trouver un réel $c \in [a, b]$ tel que $f(c) = \varphi$. On en conclut que $\varphi \in f(I)$. Cette relation étant vérifiée pour tout $\varphi \in [\alpha, \beta]$, on a :

$$[\alpha, \beta] \subseteq f(I)$$

Soit $x \in I$. Comme f est croissante ou décroissante, on doit avoir :

$$f(a) \le f(x) \le f(b)$$

ou :

$$f(a) \ge f(x) \ge f(b)$$

On a donc également :

$$f(I) \subseteq [\alpha, \beta]$$

L'inclusion étant réciproque, on a :

$$f(I) = [\alpha, \beta]$$

55.7 Continuité uniforme

On dit qu'une fonction f est uniformément continue sur A, si pour toute précision $\epsilon > 0$, on peut trouver un $\delta > 0$ tel que :

$$|f(s) - f(t)| \le \epsilon$$

pour tout $s, t \in A$ vérifiant $|s - t| \le \delta$.

Continuité simple

Si f est uniformément continue sur A, on a clairement :

$$\lim_{s \to t} |f(s) - f(t)| = 0$$

et donc :

$$\lim_{s \to t} f(s) = f(t)$$

pour tout $t \in A$. Toute fonction uniformément continue est donc continue.

55.8 Polynômes

Soit $n \in \mathbb{N}$ et $\alpha, \beta \in \mathbb{R}$ avec $\alpha \le \beta$. Nous allons analyser la continuité du monôme $\mu : [\alpha, \beta] \mapsto \mathbb{R}$ défini par :

$$\mu : x \mapsto x^n$$

pour tout $x \in [\alpha, \beta]$. Reprenant les résultats de la section 35.3, nous avons :

$$s^n - t^n = (s - t) \sum_{i=0}^{n-1} s^{n-1-i} \cdot t^i$$

Quelque soient $s, t \in [\alpha, \beta]$, il est clair que

$$|s|, |t| \le M = \max\{|\alpha|, |\beta|\}$$

On a donc :

$$|s^n - t^n| \le |s - t| \cdot n \cdot M^{n-1}$$

Fixons à présent $\epsilon > 0$. Il suffit de prendre :

$$|s - t| \le \delta \le \frac{\epsilon}{n \cdot M^{n-1}}$$

pour avoir :

$$|s^n - t^n| \le \delta \cdot n \cdot M^n \le \epsilon$$

Comme le choix de δ ne dépend ni de s ni de t, le monôme μ est uniformément continu sur $[\alpha, \beta]$. Pour généraliser aux polynômes de la forme :

$$p(x) = \sum_{i=0}^{n} a_i \cdot x^i$$

on part de :

$$p(s) - p(t) = \sum_{i=0}^{n} a_i \cdot (s^i - t^i)$$

On a donc :

$$|p(s) - p(t)| \le \sum_{i=0}^{n} |a_i| \cdot \left|s^i - t^i\right|$$

Mais comme on peut trouver des δ_k tels que :

$$\left|s^k - t^k\right| \le \frac{\epsilon}{\sum_j |a_j|}$$

il suffit de choisir $\delta = \min\{\delta_0, \delta_1, ..., \delta_n\}$ pour avoir :

$$|p(s) - p(t)| \le \epsilon \cdot \frac{\sum_i |a_i|}{\sum_j |a_j|} = \epsilon$$

Cette généralisation montre aussi que toute combinaison linéaire de fonctions uniformément continues est uniformément continue.

Continuité simple

Qu'en est-il de la continuité sur $\mathbb{R}$? Choisissons $a \in \mathbb{R}$ et considérons l'intervalle $I = [a-1, a+1]$. Le polynôme est uniformément continu sur cette intervalle, et $a \in \operatorname{int} I$. Plus précisément, $\mathfrak{dist}(a, \mathbb{R} \setminus I) = 1 > 0$. Donc, si $|x - a| \le 1$, on a forcément $x \in I$. On peut donc se servir de la continuité uniforme sur I pour trouver un $\delta \in]0, 1[$ tel que $|p(x) - p(a)| \le \epsilon$ lorsque $|x - a| \le \delta$. Les polynômes sont donc continus en tout point de $\mathbb{R}$, et par conséquent continus sur $\mathbb{R}$.

55.9 Uniformité

Nous allons à présent montrer que toute fonction continue sur un intervalle de la forme $[\alpha, \beta]$ y est uniformément continue.

Théorème 55.9.1. *Soit la fonction $f \in \operatorname{Cont}([\alpha, \beta], \mathbb{R})$. Etant donné un $\epsilon > 0$ et un $a \in [\alpha, \beta]$, on note $\Delta(a, \epsilon)$ l'ensemble des écarts strictement positifs offrant la précision demandée. Pour tout $\delta \in \Delta(a, \epsilon)$, on aura donc $\delta > 0$ et :*

$$|f(a + h) - f(a)| \le \epsilon$$

pourvu que $h \in \mathbb{R}$ vérifie :

$$|h| \le \delta$$
$$a + h \in [\alpha, \beta]$$

On note les supremums de cette famille d'ensemble par :

$$\sigma(a, \epsilon) = \sup \Delta(a, \epsilon)$$

Nous allons voir que l'intersection de ces ensembles est non vide, même après avoir parcouru tout l'intervalle :

$$\Gamma(\epsilon) = \bigcap_{a \in [\alpha, \beta]} \Delta(a, \epsilon) \neq \emptyset$$

et que l'infimum des supremums est strictement positif :

$$I(\epsilon) = \inf\{\sigma(a, \epsilon) : a \in [\alpha, \beta]\} \ > \ 0$$

Etant donné un $\epsilon \ > \ 0$, on peut donc trouver un $\delta \in \Gamma(\epsilon)$ tel que :

$$|f(x + h) - f(x)| \le \epsilon$$

pour tout $x \in [\alpha, \beta]$ et pour tout h vérifiant :

$$|h| \le \delta$$
$$x + h \in [\alpha, \beta]$$

Posant $s = x + h$ et $t = x$, cela revient à dire que :

$$|f(s) - f(t)| \le \epsilon$$

pour tout $s, t \in [\alpha, \beta]$ vérifiant $|s - t| \le \delta$. La fonction f est donc uniformément continue sur $[\alpha, \beta]$.

Remarques

— La continuité de f nous garantit que ces écarts strictement positifs existent bien, c'est à dire que :

$$\Delta(a, \epsilon) \ne \emptyset$$

quelles que soient les valeurs de $\epsilon \ > \ 0$ et de $a \in [\alpha, \beta]$.

— Par ailleurs, si :

$$0 \ < \ \gamma \le \delta \in \Delta(a, \epsilon)$$

tous les réels présentant un écart inférieur à γ (par rapport à a) auront a fortiori un écart inférieur à δ et satisferont donc la précision ϵ :

$$f\big([a - \gamma, a + \gamma]\big) \subseteq [f(a) - \epsilon, f(a) + \epsilon]$$

Par conséquent, γ appartient à $\Delta(a, \epsilon)$ et :

$$]0, \delta] \subseteq \Delta(a, \epsilon)$$

pour tout $\delta \in \Delta(a, \epsilon)$.

— Si $x \in \,]0, \sigma(a, \epsilon)[$, on a :

$$\psi = \sigma(a, \epsilon) - x \ > \ 0$$

Comme le supremum est dans l'adhérence, la distance à son ensemble est nulle et on peut trouver un $\delta \in \Delta(a, \epsilon)$ tel que :

$$|\sigma(a, \epsilon) - \delta| \le \psi$$

On a donc :

$$\sigma(a, \epsilon) - \delta \le \sigma(a, \epsilon) - x$$

et :

$$x \le \delta$$

On a alors :

$$x \in \left]0, \delta\right] \subseteq \Delta(a, \epsilon)$$

et donc $x \in \Delta(a, \epsilon)$. Cette relation étant vérifiée pour tout $x \in \left]0, \sigma(a, \epsilon)\right[$, on a :

$$\left]0, \sigma(a, \epsilon)\right[\subseteq \Delta(a, \epsilon)$$

— Par définition du supremum, il ne peut avoir d'élément de $\Delta(a, \epsilon)$ supérieur à $\sigma(a, \epsilon)$, et on a également :

$$\Delta(a, \epsilon) \subseteq \left]0, \sigma(a, \epsilon)\right]$$

— Les propositions sur l'intersection non vide et l'infimum strictement positif sont équivalentes. En effet, si $I(\epsilon) > 0$, on a :

$$\sigma(a, \epsilon) \ge I(\epsilon) > 0$$

pour tout a. On a donc :

$$\emptyset \ne \left]0, I(\epsilon)\right[\subseteq \bigcap_{a \in [\alpha, \beta]} (0, \sigma(a, \epsilon)) \subseteq \bigcap_{a \in [\alpha, \beta]} \Delta(a, \epsilon)$$

D'un autre coté, si l'intersection est non nulle, soit :

$$\delta \in \bigcap_{a \in [\alpha, \beta]} \Delta(a, \epsilon)$$

Comme δ appartient à $\Delta(a, \epsilon)$ pour tout $a \in [\alpha, \beta]$, on a :

$$\sigma(a, \epsilon) \ge \delta > 0$$

par définition du supremum localisé en a. Il suffit alors de passer à l'infimum sur a pour obtenir :

$$I(\epsilon) = \inf_{a \in [\alpha, \beta]} \sigma(a, \epsilon) \ge \delta > 0$$

Nous nous attelerons ici à démontrer que $I(\epsilon) > 0$.

Démonstration 55.9.1. Soit $\epsilon > 0$. Considérons la suite d'infimums intermédiaires :

$$D(x) = \inf\{\sigma(\xi, \epsilon) : \xi \in [\alpha, x]\}$$

où $x \in [\alpha, \beta]$. Nous considérons l'ensemble Ψ des éléments tels que cet infimum soit non nul :

$$\Psi = \{x \in [\alpha, \beta] : D(x) > 0\}$$

On a $D(\alpha) = \inf\{\sigma(\alpha, \epsilon)\} = \sigma(\alpha, \epsilon) > 0$. Donc $\alpha \in \Psi$. On a aussi $\Psi \subseteq [\alpha, \beta] \le \beta$. L'ensemble Ψ est non vide et majoré. Il admet donc un supremum :

$$S = \sup \Psi \le \beta$$

— Si $x \in \Psi$ et $a \in [\alpha, x]$, on a :

$$[\alpha, a] \subseteq [\alpha, x]$$

Les propriétés de l'infimum pour l'inclusion nous donnent alors :

$$D(a) \ge D(x) > 0$$

On en conclut que :

$$[\alpha, x] \subseteq \Psi$$

pour tout $x \in \Psi$.

— Soit $a \in [\alpha, \beta]$. Nous allons construire une zone autour de a où l'infimum est strictement positif. Choisissons $\delta(a) \in \Delta(a, \epsilon/2)$ et posons :

$$\gamma(a) = \frac{1}{2}\delta(a) \ > \ 0$$

Considérons à présent l'ensemble :

$$U(a) = [a - \gamma(a), a + \gamma(a)] \cap [\alpha, \beta]$$

Pour tout $b \in U(a)$ et $x \in [b - \gamma(a), b + \gamma(a)] \cap [\alpha, \beta]$, on a alors :

$$\begin{aligned} |b - a| &\le \gamma(a) \ < \ \delta(a) \\ |x - b| &\le \gamma(a) \\ |x - a| &\le |x - b| + |b - a| \le 2\gamma(a) \le \delta(a) \end{aligned}$$

et :

$$|f(b) - f(a)| \le \frac{\epsilon}{2}$$

$$|f(x) - f(a)| \le \frac{\epsilon}{2}$$

On en déduit la borne supérieure :

$$\begin{aligned} |f(x) - f(b)| &\le |f(x) - f(a)| + |f(a) - f(b)| \\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

On a donc :

$$\sigma(b, \epsilon) \ge \gamma(a) \ > \ 0$$

pour tout $b \in U(a)$. En prenant l'infimum, il vient :

$$\inf_{b \in U(a)} \sigma(b, \epsilon) \ge \gamma(a) \ > \ 0$$

— Supposons que $S = \alpha$ et posons :

$$\theta = \min\{\gamma(\alpha), \beta - \alpha\} \ > \ 0$$

On a alors $U(S) = U(\alpha) = [\alpha, \alpha + \theta]$ et :

$$D(\alpha + \theta) = \inf_{b \in U(\alpha)} \sigma(b, \epsilon) \ge \gamma(\alpha) \ > \ 0$$

On en conclut que $\alpha + \theta \in \Psi$ avec $S \ < \ \alpha + \theta$, ce qui contredit $S = \sup \Psi$.

— Supposons à présent que $\alpha \ < \ S \ < \ \beta$ et posons :

$$\begin{aligned} \eta &= \min\{\gamma(S), S - \alpha\} \ > \ 0 \\ \theta &= \min\{\gamma(S), \beta - S\} \ > \ 0 \end{aligned}$$

On a alors $U(S) = [S - \eta, S + \theta]$. Voyons comment se comporte $\sigma(b, \epsilon)$ lorsquer b voyage dans $[\alpha, S + \theta]$. Le supremum étant dans l'adhérence, on peut trouver $\psi \in \Psi$ tel que :

$$|S - \psi| = S - \psi \le \eta$$

Si $b \in [\alpha, \psi]$, on a :

$$\sigma(b, \epsilon) \ge D(\psi) \ > \ 0$$

par définition de Ψ. Mais si $b \in [\psi, S + \theta]$, on a :

$$S - \eta \le \psi \le b \le S + \theta$$

Donc, $b \in [S - \eta, S + \theta] = U(S)$ et :

$$\sigma(b, \epsilon) \ge \gamma(S) \ > \ 0$$

Il suffit donc de poser :

$$\varpi = \min\{\gamma(S), D(\psi)\} \ > \ 0$$

pour avoir $\sigma(b, \epsilon) \ge \varpi$ sur $[\alpha, S + \theta]$. On en déduit que l'infimum est strictement positif :

$$D(S + \theta) \ge \varpi \ > \ 0$$

C'est à dire $S + \theta \in \Psi$ avec $S \ < \ S + \theta$, ce qui contredit la définition du supremum. On doit donc avoir $S = \sup \Psi = \beta$.

— Posons :

$$\theta = \min\{\gamma(\beta), \beta - \alpha\} \ > \ 0$$

On a alors $U(S) = U(\beta) = [S - \theta, S] = [\beta - \theta, \beta]$. Comme le supremum est dans l'adhérence, on peut trouver un $\psi \in \Psi$ tel que :

$$|S - \psi| = S - \psi \le \theta$$

Si $b \in [\alpha, \psi]$, on a :

$$\sigma(b, \epsilon) \ge D(\psi) \ > \ 0$$

par définition de Ψ. Mais si $b \in [\psi, S] = [\psi, \beta]$, on a :

$$S - \theta \le \psi \le b \le \beta$$

Donc, $b \in [S - \theta, S] = U(S)$ et :

$$\sigma(b, \epsilon) \ge \gamma(S) = \gamma(\beta) \ > \ 0$$

Il suffit donc de poser :

$$\varpi = \min\{\gamma(\beta), D(\psi)\} \ > \ 0$$

pour avoir $\sigma(b, \epsilon) \ge \varpi$ sur $[\alpha, \beta]$. On en déduit que l'infimum est strictement positif :

$$D(\beta) \ge \varpi \ > \ 0$$

C'est à dire $\beta \in \Psi$ et :

$$\Psi = [\alpha, \beta]$$

L'infimum $D(x)$ est donc strictement positif sur tout l'intervalle $[\alpha, \beta]$ et on a :

$$I(\epsilon) = D(\beta) \ > \ 0$$
$$]0, I(\epsilon)[\ \subseteq \Gamma(\epsilon)$$

Remarque

Le théorème *n'est pas* applicable aux intervalles ouverts ou semi-ouverts.

55.10 Variations bornées

On déduit de l'uniforme continuité des fonctions continues sur les intervalles que les fonctions continues y sont bornées. En effet, soit $\epsilon >$ et $\delta > 0$ tel que :

$$|f(b) - f(a)| \le \epsilon$$

pour tout $a, b \in [\alpha, \beta]$ tels que $|a - b| \le \delta$. Choisissons $N \in \mathbb{N}$ tel que :

$$\frac{\beta - \alpha}{N} \le \delta$$

Choisissons à présent $x, y \in [\alpha, \beta]$ et posons :

$$h = \frac{y - x}{N}$$

On a alors :

$$|h| = \left| \frac{y - x}{N} \right| \le \frac{\beta - \alpha}{N} \le \delta$$

Posons $x_i = x + i \cdot h$ pour $i = 0, 1, 2, ..., N$. On a alors $x_0 = x$ et $x_N = y$. La variation est bornée par :

$$|f(x) - f(y)| \le \sum_{i=0}^{N} |f(x_i) - f(x_{i-1})| \le (N + 1) \cdot \epsilon$$

55.10.1 Norme

Comme N ne dépend ni de x ni de y, on en conclut que les variations de f sont bornées sur l'intervalle. On a aussi :

$$|f(x)| \le |f(x) - f(\alpha)| + |f(\alpha)| \le M$$

où $M = (N + 1) \cdot \epsilon + |f(\alpha)|$ ne dépend pas du choix de x, ce qui prouve que f est bornée sur l'intervalle. En passant au supremum, on en déduit que la norme est finie et que :

$$\|f\|_\infty \le M$$

55.11 Extrema

Soit $f \in \mathrm{Cont}([a, b], \mathbb{R})$. Comme f est bornée, on peut poser :

$$I = \inf\{f(x) : x \in [a, b]\} = \inf f([a, b])$$

Comme la distance de l'infimum à l'ensemble est nulle, on peut construire une suite $\{x_1, x_2, ...\}$ convergente vers $\lambda \in [a, b]$:

$$\lambda = \lim_{n \to \infty} x_n$$

et telle que :

$$f(x_n) - I \le \frac{1}{2^n}$$

On voit que :

$$\lim_{n \to \infty} f(x_n) = I$$

Mais par continuité de f, on a aussi :

$$\lim_{n \to \infty} f(x_n) = f(\lambda)$$

On en conclut que :

$$f(\lambda) = I$$

On peut donc trouver un réel dans l'intervalle qui minimise la fonction. L'infimum appartient à l'image $f([a, b])$. Il est donc également un minimum et on a :

$$f(\lambda) = \inf f([a, b]) = \min f([a, b])$$

On construit de même un $\sigma \in [a, b]$ qui atteint le supremum :

$$f(\sigma) = \sup f([a, b]) = \max f([a, b])$$

Chapitre 56

Applications linéaires

Dépendances

— Chapitre **??** : Les fonctions

56.1 Définition

Soit les espaces vectoriels E et F sur $\mathbb{K}$ et la fonction $f : E \mapsto F$. On dit que f est linéaire si, pour tout $x, y \in E$ et $\alpha, \beta \in \mathbb{K}$, on a :

$$f(\alpha \cdot x + \beta \cdot y) = \alpha \cdot f(x) + \beta \cdot f(y)$$

On note $\mathrm{Lin}(E, F)$ l'ensemble des fonctions linéaires de E vers F.

56.2 Identité

L'application identité est clairement linéaire.

56.3 Inverse

Soit $u = f(x)$ et $v = f(y)$. Si l'application inverse existe, on a $x = f^{-1}(u)$ et $y = f^{-1}(v)$. En composant à gauche par f^{-1} la définition de la linéarité, on obtient :

$$\alpha \cdot f^{-1}(u) + \beta \cdot f^{-1}(v) = f^{-1}(\alpha \cdot u + \beta \cdot v)$$

ce qui montre que l'inverse est également linéaire.

56.4 Valeur au vecteur nul

Choisissons un $x \in E$. on voit que :

$$f(0) = f(0 \cdot x) = 0 \cdot f(x) = 0$$

La valeur d'une application linéaire s'annule au vecteur nul.

56.5 Norme des applications linéaires

La norme d'une application linéaire est définie comme étant l'extension maximale qu'elle produit :

$$\|f\| = \sup \left\{ \frac{\|f(x)\|}{\|x\|} : x \in E,\ x \neq 0 \right\}$$

On a donc :

$$\|f(x)\| \le \|f\| \cdot \|x\|$$

pour tout $x \in E \setminus \{0\}$.

Vérification

Nous allons vérifier qu'il s'agit bien d'une norme. On a $\|f\| \ge 0$ par positivité de la norme sur E et F. La condition $\|f\| = 0$ implique $\|f(x)\| = 0$ et donc $f(x) = 0$ pour tout $x \neq 0$. Comme f est linéaire, on a aussi $f(0) = 0$ et $f = 0$.

Si f, g sont linéaires, on a :

$$\|f(x) + g(x)\| \le \|f(x)\| + \|g(x)\| \le \|f\| \cdot \|x\| + \|g\| \cdot \|x\| = (\|f\| + \|g\|) \cdot \|x\|$$

pour tout $x \neq 0$. En divisant par $\|x\|$ et en passant au supremum, on obtient :

$$\|f + g\| \le \|f\| + \|g\|$$

Enfin, si $\alpha \in \mathbb{K}$, on a :

$$\frac{\|\alpha \cdot f(x)\|}{\|x\|} = |\alpha| \cdot \frac{\|f(x)\|}{\|x\|}$$

En passant au supremum, on obtient :

$$\|\alpha \cdot f\| = |\alpha| \cdot \|f\|$$

Notation

Lorsqu'il est nécessaire de différentier la norme au sens des applications linéaires d'autres types de normes utilisées, on note :

$$\|f\|_{\mathrm{Lin}} = \|f\|$$

Définition alternative

Soit $N \in \mathbb{K}$, avec $N > 0$ et :

$$B = \{u \in E : \|u\| = N\}$$

Soit $x \in E$ avec $x \neq 0$ et :

$$\lambda = \frac{\|x\|}{N}$$

Définissons :

$$u = \frac{x}{\lambda}$$

On voit que :

$$\|u\| = \left\| \frac{1}{\lambda} \cdot x \right\| = \frac{1}{\lambda} \cdot \|x\| = \frac{N}{\|x\|} \cdot \|x\| = N$$

On a donc $u \in B$. Le rapport des normes s'écrit :

$$\frac{\|f(x)\|}{\|x\|} = \frac{\|f(x)\|}{N \cdot \lambda} = \frac{1}{N} \left\| \frac{f(x)}{\lambda} \right\| = \frac{1}{\|u\|} \cdot \left\| f\left(\frac{x}{\lambda}\right) \right\| = \frac{\|f(u)\|}{\|u\|}$$

On en conclut que :

$$\frac{\|f(x)\|}{\|x\|} = \frac{\|f(u)\|}{\|u\|} \le \sup \left\{ \frac{\|f(v)\|}{\|v\|} : v \in B \right\}$$

Comme ce doit être valable quelque soit $x \ne 0$, on obtient :

$$\|f\| \le \sup \left\{ \frac{\|f(v)\|}{\|v\|} : v \in B \right\}$$

en passant au supremum sur x.
Choisissons à présent $u \in B$. On a alors :

$$\frac{\|f(u)\|}{\|u\|} \le \|f\|$$

En passant au supremum sur u, on obtient :

$$\sup \left\{ \frac{\|f(v)\|}{\|v\|} : v \in B \right\} \le \|f\|$$

On en conclut que les deux supremums sont égaux :

$$\sup \left\{ \frac{\|f(v)\|}{\|v\|} : v \in B \right\} = \|f\|$$

Norme unitaire

Une conséquence importante du résultat ci-dessus est le cas particulier $N = 1$. On a alors :

$$\|f\| = \sup \{ \|f(v)\| : v \in E,\ \|v\| = 1 \}$$

56.6 Norme d'une composée

Soit $f : E \mapsto F$ et $g : F \mapsto G$ deux applications linéaires de normes finies. Si $x \in E$ avec $x \ne 0$ on a $f(x) \in F$ et :

$$\|g \circ f(x)\| \le \|g\| \cdot \|f(x)\| \le \|g\| \cdot \|f\| \cdot \|x\|$$

En divisant par $\|x\| \ne 0$:

$$\frac{\|g \circ f(x)\|}{\|x\|} \le \|g\| \cdot \|f\|$$

et en passant au supremum sur $x \ne 0$, on en conclut que :

$$\|g \circ f\| \le \|g\| \cdot \|f\|$$

56.7 Norme d'une puissance

On a clairement :

$$\|f^n\| = \|f \circ ... \circ f\| \le \|f\|^n$$

56.8 Continuité

Nous allons montrer que, pour tout $f \in \mathrm{Lin}(A, B)$, on a l'équivalence entre l'hypothèse d'une norme de f finie et l'hypothèse de f continue.

Si la norme est finie, on a :

$$\|f(x) - f(a)\| = \|f(x - a)\| \le \|f\| \cdot \|x - a\|$$

qui tend bien vers 0 lorsque x tend vers a. Inversément, si f est continue, on peut trouver un $\delta > 0$ tel que :

$$\|f(x) - f(0)\| \le 1$$

pour tout x vérifiant $\mathfrak{dist}(x, 0) = \|x\| \le \delta$. Posons $B = \{x \in A : \|x\| = \delta\}$. On a alors :

$$\sup_{x \in B} \frac{\|f(x)\|}{\|x\|} = \frac{1}{\delta} \sup_{x \in B} \|f(x) - f(0)\| \le \frac{1}{\delta}$$

La norme est donc finie :

$$\|f\| = \sup_{x \in B} \frac{\|f(x)\|}{\|x\|} \le \frac{1}{\delta} < +\infty$$

56.9 n-linéarité

On dit que la fonction $f : E_1 \times ... \times E_n \mapsto F$ est n-linéaire si elle est linéaire par rapport à chacune des composantes de son argument, les autres composantes restant inchangées :

$$f(..., \alpha x + \beta y, ...) = \alpha \cdot f(..., x, ...) + \beta \cdot f(..., y, ...)$$

pour tout $\alpha, \beta \in \mathbb{K}$ et $x, y \in E$. On note $\mathrm{Lin}_n(E_1, ..., E_n, F)$ l'ensemble des fonctions n-linéaires de $E_1 \times ... \times E_n$ vers F.

Norme

La norme est définie dans ce cas par :

$$\|f\| = \sup \left\{ \frac{\|f(x_1, ..., x_n)\|}{\prod_{i=1}^n \|x_i\|} : (x_1, ..., x_n) \in E_1 \times ... \times E_n \right\}$$

Si cette norme est finie, on a :

$$\|f(x_1, ..., x_n)\| \le \|f\| \cdot \prod_{i=1}^n \|x_i\|$$

pour tout $(x_1, ..., x_n) \in E_1 \times ... \times E_n$.

Bilinéarité

On dit aussi des fonctions 2-linéaires qu'elles sont bilinéaires. La norme d'une fonction $f : E_1 \times E_2 \mapsto F$ bilinéaire est définie par :

$$\|f\| = \sup \left\{ \frac{\|f(u,v)\|}{\|u\| \cdot \|v\|} : (u,v) \in E_1 \times E_2 \right\}$$

Si cette norme est finie, on a :

$$\|f(u,v)\| \le \|f\| \cdot \|u\| \cdot \|v\|$$

pour tout $(u,v) \in E_1 \times E_2$.

56.10 Représentation matricielle

Soit une application linéaire $\mathcal{A} : E \to F$. Choisissons $x \in E$ et posons :

$$y = \mathcal{A}(x)$$

Si on dispose d'une base $(e_1, ..., e_n)$ de E et d'une base $(f_1, ..., f_m)$ de F, on a :

$$x = \sum_{i=1}^{n} x_i \cdot e_i$$

$$y = \sum_{i=1}^{m} y_i \cdot f_i$$

pour certains $x_i, y_i \in \mathbb{K}$. La linéarité de $\mathcal{A}$ implique que :

$$y = \sum_{j=1}^{n} \mathcal{A}(e_j) \cdot x_j$$

Si les $a_{ij} \in \mathbb{K}$ sont les coordonnées de $\mathcal{A}(e_j)$ dans la base des f_i, on a :

$$\mathcal{A}(e_j) = \sum_{i=1}^{m} a_{ij} \cdot f_i$$

En substituant cette expression, on obtient :

$$y = \sum_{i=1}^{m} f_i \sum_{j=1}^{n} a_{ij} \cdot x_j$$

La i^{me} coordonnée de y est donc donnée par :

$$y_i = \sum_{j=1}^{n} a_{ij} \cdot x_j$$

On définit la matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$ associée à $\mathcal{A}$ en posant :

$$A = (a_{ij})_{i,j}$$

Nous définissons ensuite le produit d'une matrice avec le « vecteur » x équivalent de $\mathbb{K}^n$:

$$A \cdot x = \left(\sum_j a_{ij} \cdot x_j \right)_i$$

de telle sorte que :

$$A \cdot x = \mathcal{A}(x)$$

Norme

La norme d'une matrice est la norme de l'application linéaire associée, c'est-à-dire :

$$\|A\|_2 = \sup \left\{ \frac{\|A \cdot x\|}{\|x\|} : x \in \mathbb{K}^n, \ x \neq 0 \right\}$$

Soit :

$$M = \max_{i,j} \left| \operatorname*{comp}_{ij} A \right|$$

On a alors :

$$\|A \cdot x\| \le M \cdot m \cdot n \cdot \max_i x_i \le M \cdot m \cdot n \cdot \|x\|$$

ce qui montre que :

$$\|A\| \le M \cdot m \cdot n \ < \ \infty$$

La norme d'une matrice finie $(m, n \ < \ \infty)$ existe toujours.

Image

L'image d'une matrice est l'image de l'application linéaire associée, c'est-à-dire :

$$\operatorname{im} A = \{A \cdot x : x \in \mathbb{K}^n\}$$

Si $c_i = \text{colonne}_i\, A$, on a :

$$A = [c_1 \ c_2 \ ... \ c_n]$$

On voit que :

$$A \cdot x = \sum_i c_i \cdot x_i$$

autrement dit l'image de A est l'espace vectoriel engendré par ses colonnes :

$$\operatorname{im} A = \operatorname{span}\{c_1, c_2, ..., c_n\}$$

Noyau

Le noyau d'une matrice est le noyau de l'application linéaire associée, c'est-à-dire :

$$\ker A = \{x \in \mathbb{K}^n : A \cdot x = 0\}$$

56.11 Produit matriciel

Soit à présent les matrices $A \in \mathbb{M}(\mathbb{K}, m, n)$ et $B \in \mathbb{M}(\mathbb{K}, n, p)$ données par :

$$A = (a_{ij})_{i,j}$$
$$B = (b_{ij})_{i,j}$$

où les $a_{ij}, b_{ij} \in K$. Soit les applications linéaires $\mathcal{B} : \mathbb{K}^p \to \mathbb{K}^n$ et $\mathcal{A} : \mathbb{K}^n \to \mathbb{K}^m$ définies par :

$$\mathcal{B}(x) = B \cdot x$$
$$\mathcal{A}(y) = A \cdot y$$

pour tout $x \in \mathbb{K}^p$, $y \in \mathbb{K}^n$. Choisissons $z \in \mathbb{K}^m$ et relions x, y, z par :

$$y = \mathcal{B}(x)$$
$$z = \mathcal{A}(y) = \big(\mathcal{A} \circ \mathcal{B}\big)(x)$$

Examinons les composantes de z en fonction de celles de x :

$$z_i = \sum_k a_{ik} \cdot y_k = \sum_k a_{ik} \sum_j b_{kj} \cdot x_j = \sum_{k,j} a_{ik} \cdot b_{kj} \cdot x_j$$

On en déduit que la composée $\mathcal{C} = \mathcal{A} \circ \mathcal{B}$ est représentée par la matrice $C \in \mathbb{M}(\mathbb{K}, m, p)$ de composantes :

$$\operatorname*{comp}_{ij} C = c_{ij} = \sum_k a_{ik} \cdot b_{kj}$$

Il suffit donc de définir le produit matriciel $A \cdot B$ par :

$$A \cdot B = \left(\sum_{k=1}^n a_{ik} \cdot b_{kj} \right)_{i,j}$$

pour avoir :

$$(A \cdot B) \cdot x = \big(\mathcal{A} \circ \mathcal{B}\big)(x)$$

Le produit matriciel représente donc une composée d'applications linéaires. Pour que ce produit soit bien défini, il est nécessaire que le nombre de colonnes n de A et le nombre de lignes de B soient identiques.

On voit également que le produit matrice - vecteur défini précédemment en est un cas particulier lorsque $p = 1$.

Notation

En pratique, on laisse souvent tomber le "$\cdot$" et on note AB au lieu de $A \cdot B$ lorsqu'il est évident que A et B sont deux matrices différentes.

Taille

Le produit d'une matrice de taille (m, n) par une matrice de taille (n, p) est une matrice de taille (m, p).

Lignes et colonnes

Si $x_i^T = \text{ligne}_i(A)$ et $y_j = \text{colonne}_j(B)$, on voit que :

$$\operatorname*{comp}_{ij}(A \cdot B) = x_i^T \cdot y_j$$

Associativité

Soit les matrices $A \in \mathbb{M}(\mathbb{K}, m, n)$, $B \in \mathbb{M}(\mathbb{K}, n, p)$ et $C \in \mathbb{M}(\mathbb{K}, p, q)$ données par :

$$\begin{aligned} A &= (a_{ij})_{i,j} \\ B &= (b_{ij})_{i,j} \\ C &= (c_{ij})_{i,j} \end{aligned}$$

où les $a_{ij}, b_{ij}, c_{ij} \in K$. La relation :

$$A \cdot (B \cdot C) = \left(\sum_{k,l} a_{ik} \cdot b_{kl} \cdot c_{lj} \right)_{i,j} = (A \cdot B) \cdot C$$

nous montre que la multiplication entre matrices est associative. On définit :

$$A \cdot B \cdot C = A \cdot (B \cdot C) = (A \cdot B) \cdot C$$

Distributivité

On a aussi les propriétés de distribution :

$$\begin{aligned} A \cdot (B + C) &= A \cdot B + A \cdot C \\ (B + C) \cdot D &= B \cdot D + C \cdot D \end{aligned}$$

où $A \in \mathbb{M}(\mathbb{K}, m, n)$, $B, C \in \mathbb{M}(\mathbb{K}, n, p)$ et $D \in \mathbb{M}(\mathbb{K}, p, q)$.

Non commutativité

Par contre, on peut trouver des matrices A et B telles que :

$$A \cdot B \neq B \cdot A$$

La multiplication matricielle n'est donc en général pas commutative. D'ailleurs, pour que ces deux produits existent simultanément, il faut que A et B soient toutes deux carrées, ce qui n'est pas forcément le cas.

Commutateur

La matrice associée au commutateur :

$$[\mathcal{A}, \mathcal{B}] = \mathcal{A} \circ \mathcal{B} - \mathcal{B} \circ \mathcal{A}$$

est donnée par le commutateur équivalent :

$$[A, B] = A \cdot B - B \cdot A$$

Transposée

On vérifie que :

$$(A \cdot B)^T = B^T \cdot A^T$$

56.12 Blocs

En utilisant l'associativité de l'addition, on peut facilement vérifier que la formule de multiplication reste valable lorsqu'on considère des blocs de matrices au lieu des éléments, à condition de respecter l'ordre de multiplication. Un exemple fréquemment utilisé :

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11} \cdot B_{11} + A_{12} \cdot B_{21} & A_{11} \cdot B_{12} + A_{12} \cdot B_{22} \\ A_{21} \cdot B_{11} + A_{22} \cdot B_{21} & A_{21} \cdot B_{12} + A_{22} \cdot B_{22} \end{bmatrix}$$

Bloc-diagonale

Un cas particulier important :

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \cdot \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} = \begin{bmatrix} A_1 \cdot B_1 & 0 \\ 0 & A_2 \cdot B_2 \end{bmatrix}$$

56.13 Matrice identité

La matrice identité $I \in \mathbb{M}(\mathbb{K}, n, n)$ correspond à la fonction Id. On a donc :

$$I \cdot x = x$$

pour tout $x \in \mathbb{K}^n$. Si $(\mathfrak{c}_1, ...\mathfrak{c}_n)$ est la base canonique de $\mathbb{K}^n$, on a donc :

$$I \cdot \mathfrak{c}_i = \mathfrak{c}_i$$

ce qui entraîne directement :

$$I = (\boldsymbol{\delta}_{ij})_{i,j}$$

On remarque que :

$$I = [\mathfrak{c}_1 \ \ldots \ \mathfrak{c}_n]$$

Neutre

Comme la fonction identité est neutre pour la composition, la matrice unité correspondante $I \in \mathbb{M}(\mathbb{K}, m, n)$ doit être neutre pour la multiplication avec toutes les matrices de dimensions compatibles. Soit $A \in \mathbb{M}(\mathbb{K}, m, n)$ et $B \in \mathbb{M}(\mathbb{K}, n, p)$. On vérifie que l'on a bien :

$$A \cdot I = A$$
$$I \cdot B = B$$

Notation

On note aussi I_n pour préciser que I est de taille (n, n).

56.14 Inverse

Lorsqu'elle existe, la matrice inverse de A, notée A^{-1}, reflète l'application linéaire inverse sous-jacente. Elle est donc l'unique matrice telle que :

$$A^{-1} \cdot A = A \cdot A^{-1} = I$$

Inverse d'un produit

Soit A et B deux matrices inversibles. Les relations $C \cdot (A \cdot B) = I$ et $(A \cdot B) \cdot D = I$ nous donnent :

$$C = D = B^{-1} \cdot A^{-1}$$

et donc :

$$(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$$

Inverse à gauche et à droite

On dit que L est un inverse à gauche de A si $L \cdot A = I$. On dit que R est un inverse à droite de A si $A \cdot R = I$.

56.15 Puissance

Il est possible de multiplier une matrice carrée A avec elle-même. On peut donc définir l'exposant par :

$$A^0 = I$$
$$A^k = A \cdot A^{k-1}$$

Négative

Si l'inverse A^{-1} existe, on définit également :

$$A^{-k} = (A^{-1})^k$$

56.16 Polynômes matriciels

Ici, $\mathbb{K}$ n'est plus un corps mais l'anneau des matrices X de taille (N, N). On dit que $p : \mathbb{M}(\mathbb{K}, N, N) \mapsto \mathbb{M}(\mathbb{K}, N, N)$ est un polynôme matriciel si il existe $a_0, ..., a_n \in \mathbb{K}$ tels que :

$$p(X) = \sum_{i=0}^{n} a_i \cdot X^i$$

pour tout $X \in \mathbb{M}(\mathbb{K}, N, N)$.

Chapitre 57

Géométrie

57.1 Courbe

Une courbe sur un espace vectoriel E (par exemple $\mathbb{R}^n$) est de la forme :

$$\Lambda = \{\lambda(s) : s \in [\alpha, \beta]\}$$

où $\lambda : [\alpha, \beta] \mapsto E$ est une fonction continue et où $\alpha, \beta \in \mathbb{R}$ vérifient $\alpha \le \beta$.

57.2 Segment

Les segments sont une généralisation des intervalles. Un segment de $u \in E$ vers $v \in E$ est un cas particulier de courbe où $\lambda : [0, 1] \subseteq \mathbb{R} \mapsto E$ est une fonction linéaire définie par :

$$\lambda(s) = u + s \cdot (v - u)$$

pour tout $s \in [0, 1] \subseteq \mathbb{R}$. On voit que $\lambda(0) = u$ et que $\lambda(1) = v$. On note aussi :

$$[u, v] = \lambda([0, 1]) = \{u + s \cdot (v - u) : s \in [0, 1]\} \subseteq E$$

57.2.1 Alternative

On dispose aussi d'une définition alternative. On utilise :

$$L = \{(s, t) \in \mathbb{R}^2 : (s, t) \ge 0 \text{ et } s + t = 1\}$$

et la fonction $\sigma : L \mapsto E$ définie par :

$$\sigma(s, t) = s \cdot u + t \cdot v$$

On a alors $[u, v] = \sigma(L)$. On voit aussi que $\sigma(1, 0) = u$ et $\sigma(0, 1) = v$.

57.3 Enveloppe convexe

Soit $A \subseteq E$ et la collection des segments reliant deux points quelconques de A :

$$\mathcal{S} = \{[u, v] : u, v \in A\}$$

L'enveloppe convexe de A est l'union de tous ces segments :

$$\text{Convexe}(A) = \bigcup \mathcal{S}$$

Pour tout $u, v \in A$ et $(s, t) \in \mathbb{R}^2$ tels que $s, t \ge 0$ et $s + t = 1$, on a donc :

$$s \cdot u + t \cdot v \in \mathrm{Convexe}(A)$$

57.3.1 Inclusion

Il suffit de considérer le choix $(s, t) = (1, 0)$ pour voir que tout $u \in A$ appartient à $\mathrm{Convexe}(A)$. On a donc $A \subseteq \mathrm{Convexe}(A)$.

Ensemble convexe

On dit qu'un ensemble $C \subseteq E$ est convexe si $\mathrm{Convexe}(C) = C$.

57.4 Surface

Une surface de E est de la forme :

$$\Phi = \{\varphi(s, t) : (s, t) \in [a, b] \times [c, d]\}$$

où $\varphi : [a, b] \times [c, d] \mapsto E$ est une fonction continue et où $a, b, c, d \in \mathbb{R}$ vérifient $a \le b$ et $c \le d$.

Chapitre 58

Formes linéaires

Dépendances

— Chapitre **??** : Les fonctions
— Chapitre 56 : Les fonctions linéaires

58.1 Définition

Soit un espace vectoriel E sur $\mathbb{K}$. Une forme linéaire est une fonction linéaire continue $\varphi : E \mapsto \mathbb{K}$.

58.2 Espace dual

L'espace dual $E^\star$ de E est l'ensemble des formes linéaires sur E , autrement dit l'ensemble des fonctions linéaires continues de E vers $\mathbb{K}$:

$$E^\star = \{\varphi \in \mathrm{Lin}(E, \mathbb{K}) : \|\varphi\|_{\mathrm{Lin}} < +\infty\}$$

Il s'agit d'un espace vectoriel pour les opérations d'addition et de multiplication mixte définies sur les fonctions.

58.3 Notation

Pour toute forme $\varphi \in E^\star$ et tout vecteur $v \in E$, on note :

$$\langle \varphi, v \rangle = \varphi(v)$$

ce qui définit implicitement la fonction $\langle , \rangle : E^\star \times E \mapsto \mathbb{K}$.

58.4 Linéarité

Soit $\varphi, \psi \in E^\star$, $u, v \in E$ et $\alpha, \beta \in S$. Comme φ est linéaire, on a :

$$\langle \varphi, \alpha \cdot u + \beta \cdot v \rangle = \alpha \cdot \langle \varphi, u \rangle + \beta \cdot \langle \varphi, v \rangle$$

Symétriquement, la définition des opérations sur les fonctions nous donne également :

$$\langle \alpha \cdot \varphi + \beta \cdot \psi, u \rangle = \alpha \cdot \langle \varphi, u \rangle + \beta \cdot \langle \psi, u \rangle$$

L'application $\langle , \rangle$ est donc bilinéaire.

58.5 Biorthonormalité

On dit que les suites $(\Phi_1, ..., \Phi_m)$ de $E^\star$ et $(e_1, ..., e_n)$ de E sont biorthonormées si :

$$\langle \Phi_i, e_j \rangle = \boldsymbol{\delta}_{ij}$$

pour tout $(i, j) \in \mathbb{Z}(0, m) \times \mathbb{Z}(0, n)$. De telles suites permettent d'évaluer facilement les coefficients des développements en série du type :

$$\varphi = \sum_{i=1}^{m} \alpha_i \cdot \Phi_i$$

où $\alpha_1, ..., \alpha_m \in S$. En effet, il suffit d'évaluer :

$$\varphi(e_j) = \langle \varphi, e_j \rangle = \sum_{i=1}^{m} \alpha_i \cdot \langle \Phi_i, e_j \rangle = \sum_{i=1}^{m} \alpha_i \cdot \boldsymbol{\delta}_{ij} = \alpha_j$$

pour obtenir les valeurs des α_j.

Réciproquement, si :

$$u = \sum_{i=1}^{n} \beta_i \cdot e_i$$

avec $\beta_1, ..., \beta_n \in S$, on a :

$$\Phi_j(u) = \langle \Phi_j, u \rangle = \sum_{i=1}^{n} \beta_i \cdot \langle \Phi_j, e_i \rangle = \sum_{i=1}^{n} \beta_i \cdot \boldsymbol{\delta}_{ij} = \beta_j$$

ce qui nous donne les valeurs des β_j.

Forts de ces résultats, il est aisé d'évaluer :

$$\langle \varphi, u \rangle = \sum_{i,j} \alpha_i \cdot \langle \Phi_i, u_j \rangle \cdot \beta_j = \sum_{i,j} \alpha_i \cdot \boldsymbol{\delta}_{ij} \cdot \beta_j = \sum_{i} \alpha_i \cdot \beta_i$$

On a donc en définitive :

$$\langle \varphi, u \rangle = \sum_{i} \langle \varphi, e_i \rangle \cdot \langle \Phi_i, u \rangle$$

58.6 Similitude

On dit que deux fonctions $u, v \in E$ sont identique au sens des distributions si :

$$\langle \varphi, u \rangle = \langle \varphi, v \rangle$$

pour tout $\varphi \in E^\star$.

Symétriquement, les deux formes $\varphi, \psi \in E^\star$ sont identiques par définition si et seulement si :

$$\langle \varphi, u \rangle = \langle \psi, u \rangle$$

pour tout $u \in E$.

58.7 Espace bidual

On définit l'espace bidual de E, noté $E^{\star\star}$, par :

$$E^{\star\star} = (E^\star)^\star$$

On associe à chaque élément $u \in E$ un élément $\hat{u} \in E^{\star\star}$ par la condition :

$$\hat{u}(\varphi) = \varphi(u)$$

qui doit être vérifiée pour tout $\varphi \in E^\star$. On a donc :

$$\langle \hat{u}, \varphi \rangle = \langle \varphi, u \rangle$$

58.8 Application duale

Soit les espaces vectoriels E et F sur $\mathbb{K}$ et une fonction $A : E \mapsto F$. Le dual de A au sens des formes, s'il existe, est l'unique fonction $A^\star : F^\star \mapsto E^\star$ telle que :

$$\langle A^\star(\varphi), u \rangle = \langle \varphi, A(u) \rangle$$

pour tout $u \in E$ et $\varphi \in F^\star$.

58.9 Formes bilinéaires

Soit les espaces vectoriels E et F sur $\mathbb{K}$. Une forme bilinéaire est une fonction bilinéaire continue $\vartheta : F \times E \mapsto \mathbb{K}$. On utilise une notation analogue à celle des formes :

$$\langle x, \vartheta, u \rangle = \vartheta(x, u)$$

pour tout $x \in F$ et $u \in E$. On voit que :

$$\begin{aligned} \langle \alpha \cdot x + \beta \cdot y, \vartheta, u \rangle &= \alpha \cdot \langle x, \vartheta, u \rangle + \beta \cdot \langle y, \vartheta, u \rangle \\ \langle x, \vartheta, \alpha \cdot u + \beta \cdot v \rangle &= \alpha \cdot \langle x, \vartheta, u \rangle + \beta \cdot \langle x, \vartheta, v \rangle \end{aligned}$$

pour tout $\alpha, \beta \in \mathbb{K}$, $u, v \in E$ et $x, y \in F$.

58.10 Formes quadratiques

Soit la forme bilinéaire $\vartheta : E \times E \mapsto \mathbb{K}$. Une forme quadratique $\mathcal{Q} : E \mapsto \mathbb{K}$ est une fonction de la forme :

$$\mathcal{Q}(x) = \langle x, \vartheta, x \rangle$$

58.11 Représentation matricielle

On peut représenter toute forme linéaire $\varphi \in \mathrm{Lin}(\mathbb{K}^n, \mathbb{K})$ par un vecteur matriciel $\hat{\varphi} \in \mathbb{K}^n$. Etant donné la base canonique $(e_1, ..., e_n)$ de $\mathbb{K}^n$, il suffit de poser :

$$\hat{\varphi}_i = \langle \varphi, e_i \rangle$$

pour avoir :

$$\langle \varphi, u \rangle = \hat{\varphi}^T \cdot u$$

pour tout $u \in \mathbb{K}^n$.

Formes bilinéaires

On peut représenter toute forme bilinéaire $\vartheta \in \mathrm{Lin}(\mathbb{K}^m \times \mathbb{K}^n, \mathbb{K})$ par une matrice $\Theta \in \mathbb{M}(K, m, n)$. Etant donné les bases canoniques $(f_1, ..., f_m)$ de $\mathbb{K}^m$ et $(e_1, ..., e_n)$ de $\mathbb{K}^n$, il suffit de poser :

$$\operatorname*{comp}_{ij} \Theta = \langle f_i, \vartheta, e_j \rangle$$

pour avoir :

$$\langle v, \vartheta, u \rangle = v^T \cdot \Theta \cdot u$$

pour tout $u \in \mathbb{K}^n$ et tout $v \in \mathbb{K}^m$.

Chapitre 59

Produit scalaire

Dépendances

— Chapitre 52 : Les espaces vectoriels
— Chapitre 53 : Les normes

59.1 Introduction

Soit E un espace vectoriel sur $\mathbb{K}$ et une famille de fonctions linéaires $\phi_u \in E^\star$ où $u \in E$ est un paramètre vectoriel. Nous pouvons écrire :

$$\phi_u(v) = \langle \phi_u, v \rangle$$

pour tout $u, v \in E$. Cette expression introduit implicitement le produit dérivé $\langle | \rangle : E \times E \mapsto \mathbb{K}$ défini par :

$$\langle u \mid v \rangle = \langle \phi_u, v \rangle$$

pour tout $u, v \in E$. Ce produit hérite bien entendu la linéarité à droite de la forme associée :

$$\langle u \mid \alpha \cdot v + \beta \cdot w \rangle = \alpha \cdot \langle u \mid v \rangle + \beta \cdot \langle u \mid w \rangle$$

pour tout $u, v, w \in E$ et pour tout $\alpha, \beta \in \mathbb{K}$.
Les produits scalaires sont des cas particuliers de ce type de produit.

Notation

On note aussi $u \cdot v = \langle u \mid v \rangle$.

59.2 Produit scalaire réel

Considérons le cas particulier où $\mathbb{K} = \mathbb{R}$. et un produit linéaire à droite $\langle | \rangle : E \times E \mapsto \mathbb{R}$. Nous voudrions en plus que la valeur de $\langle u \mid u \rangle$ en chaque $u \in E$ puisse représenter la norme de u. Nous imposons donc la positivité :

$$\langle u \mid u \rangle \geq 0$$

Pour compléter le caractère strictement défini positif, on impose également que le seul élément $u \in E$ vérifiant :

$$\langle u \mid u \rangle = 0$$

soit le vecteur nul $u = 0$. Ce qui revient à dire que :

$$\langle u \mid u \rangle > 0$$

pour tout $u \in E \setminus \{0\}$.
Si on peut également interchanger n'importe quels $u, v \in E$ sans changer le résultat :

$$\langle u \mid v \rangle = \langle v \mid u \rangle$$

on dit que $\langle | \rangle$ est un produit scalaire réel sur E.
Nous déduisons directement de la linéarité à droite et de la symétrie que :

$$\langle \alpha \cdot u + \beta \cdot v \mid w \rangle = \alpha \cdot \langle u \mid w \rangle + \beta \cdot \langle v \mid w \rangle$$

pour tout $\alpha, \beta \in \mathbb{R}$ et $u, v, w \in E$. Le produit scalaire réel est bilinéaire.

59.3 Produit scalaire complexe

Examinons à présent le cas $\mathbb{K} = \mathbb{C}$. On demande qu'un produit scalaire $\langle | \rangle : E \times E \mapsto \mathbb{C}$ soit strictement défini positif. Pour cela, les valeurs de $\langle u \mid u \rangle$ doivent être réelles et positives :

$$\langle u \mid u \rangle \in \mathbb{R}$$
$$\langle u \mid u \rangle \geq 0$$

pour tout $u \in E$. Ensuite, il faut également que le seul élément $u \in E$ vérifiant :

$$\langle u \mid u \rangle = 0$$

soit le vecteur nul $u = 0$.
Le caractère réel de $\langle u \mid u \rangle$ implique que :

$$\langle u \mid u \rangle = \overline{\langle u \mid u \rangle}$$

où la barre supérieure désigne comme d'habitude le complexe conjugué. Cette constatation nous mène à une variante de la symétrie. On impose :

$$\langle u \mid v \rangle = \overline{\langle v \mid u \rangle}$$

pour tout $u, v \in E$. On dit que le produit scalaire complexe est hermitien.
La linéarité à droite s'exprime simplement par :

$$\langle u \mid \alpha \cdot v + \beta \cdot w \rangle = \alpha \cdot \langle u \mid v \rangle + \beta \cdot \langle u \mid w \rangle$$

pour tout $u, v, w \in E$ et pour tout $\alpha, \beta \in \mathbb{C}$. On déduit de la linéarité et du caractère hermitien du produit scalaire complexe que :

$$\begin{aligned} \langle \alpha \cdot u + \beta \cdot v \mid w \rangle &= \overline{\langle w \mid \alpha \cdot u + \beta \cdot v \rangle} \\ &= \overline{\alpha} \cdot \overline{\langle w \mid u \rangle} + \overline{\beta} \cdot \overline{\langle w \mid v \rangle} \end{aligned}$$

et finalement :

$$\langle \alpha \cdot u + \beta \cdot v \mid w \rangle = \overline{\alpha} \cdot \langle u \mid w \rangle + \overline{\beta} \cdot \langle v \mid w \rangle$$

On dit que le produit scalaire est antilinéaire à gauche.

Corollaire

En particulier, si u, v, w, x sont des vecteurs de E et si $\alpha = \langle u \mid v \rangle \in \mathbb{C}$, on a :

$$\langle w \mid \alpha \cdot x \rangle = \langle w \mid \langle u \mid v \rangle \cdot x \rangle = \langle u \mid v \rangle \cdot \langle w \mid x \rangle$$
$$\langle \alpha \cdot w \mid x \rangle = \langle \langle u \mid v \rangle \cdot w \mid x \rangle = \langle v \mid u \rangle \cdot \langle w \mid x \rangle$$

Cas particulier

Comme $\overline{x} = x$ pour tout $x \in \mathbb{R} \subseteq \mathbb{C}$, on peut considérer le produit scalaire réel comme un cas particulier de produit scalaire complexe.

59.4 Espace orthogonal

A un vecteur

Soit $x \in H$. On définit l'ensemble $x^\perp$ par :

$$x^\perp = \{z \in E : \langle x \mid z \rangle = 0\}$$

On dit des vecteurs de $x^\perp$ qu'ils sont orthogonaux à x.

A un ensemble

Pour tout sous-ensemble $V \subseteq E$, l'ensemble orthogonal à V est l'ensemble des vecteurs qui sont orthogonaux à tous les éléments de V :

$$V^\perp = \bigcap_{x \in V} x^\perp$$

Pour tout $z \in V^\perp$, on a donc $\langle x \mid z \rangle = 0$ quel que soit $x \in V$.

Nous allons vérifier que $V^\perp$ est un sous-espace vectoriel. Soit $z \in V$. Comme $\langle z \mid 0 \rangle = 0$, on a $0 \in V^\perp$. Soit $x, y \in V^\perp$, $\alpha, \beta \in \mathbb{K}$. On a :

$$\langle z \mid \alpha \cdot x + \beta \cdot y \rangle = \alpha \cdot \langle z \mid x \rangle + \beta \cdot \langle z \mid y \rangle = 0 + 0 = 0$$

ce qui montre que $\alpha \cdot x + \beta \cdot y \in V^\perp$.

59.5 Egalité

Si $u, v \in E$ sont tels que :

$$\langle u \mid w \rangle = \langle v \mid w \rangle$$

pour tout $w \in E$, on a :

$$\langle u - v \mid w \rangle = 0$$

Le choix $w = u - v \in E$ nous donne alors :

$$\langle u - v \mid u - v \rangle = 0$$

ce qui implique $u - v = 0$ et donc $u = v$.

59.6 Base orthonormée

Une base $(e_1, ..., e_n)$ de E est dite orthonormée si le produit scalaire de deux vecteurs $e_i \ne e_j$ s'annule, tandis que le produit scalaire d'un e_i avec lui-même donne l'unité :

$$\langle e_i \mid e_j \rangle = \boldsymbol{\delta}_{ij}$$

Coordonnées

Soit $u \in E$ de coordonnée $u_i \in \mathbb{K}$:

$$u = \sum_{i=1}^n u_i \cdot e_i$$

En effectuant le produit scalaire de u avec e_k, on arrive à :

$$\langle e_k \mid u \rangle = \sum_{i=1}^n u_i \cdot \langle e_k \mid e_i \rangle$$
$$\langle e_k \mid u \rangle = \sum_{i=1}^n u_i \cdot \boldsymbol{\delta}_{ik}$$

Tous les termes de cette dernière somme s'annulent sauf lorsque $i = k$, et on a :

$$\langle e_k \mid u \rangle = u_k$$

On peut donc écrire :

$$y = \sum_{i=1}^n \langle e_i \mid u \rangle \cdot e_i$$

Indépendance linéaire

On peut voir que si une suite de vecteurs e_i est orthonormée, (ils ne forment pas forcément une base) ils sont toujours linéairement indépendant. En effet si les scalaires a_i, sont tels que :

$$\sum_{i=1}^n a_i \cdot e_i = 0$$

on a alors :

$$a_i = \langle e_i \mid 0 \rangle = 0$$

59.7 Produit scalaire et coordonnées

Soit $(e_1, ..., e_n)$ une base de E et $u, v \in E$. On a :

$$u = \sum_i u_i \cdot e_i$$
$$v = \sum_i v_i \cdot e_i$$

pour certains $u_i, v_i \in \mathbb{K}$. Posons :

$$g_{ij} = \langle e_i \mid e_j \rangle$$

où $\langle | \rangle : E \times E \mapsto \mathbb{C}$ est un produit scalaire complexe. Nous pouvons faire sortir les sommes en utilisant les propriétés du produit scalaire, ce qui nous donne :

$$\langle u \mid v \rangle = \sum_{i,j} \overline{u}_i \cdot g_{ij} \cdot v_j$$

Réel

Dans les cas d'un produit scalaire réel, on a $\overline{u}_i = u_i$ et l'expression devient :

$$\langle u \mid v \rangle = \sum_{i,j} u_i \cdot g_{ij} \cdot v_j$$

Base orthonormée

Si la base $(e_1, ..., e_n)$ est orthonormée, l'expression du produit scalaire se simplifie en :

$$\langle u \mid v \rangle = \sum_{i} \overline{u}_i \cdot v_i$$

59.8 Application définie positive

Soit une application linéaire $A : E \mapsto E$. Si le produit scalaire de u avec $A(u)$ est un réel positif :

$$\langle u \mid A(u) \rangle = \langle A(u) \mid u \rangle \geq 0$$

pour tout $u \in E$, on dit que A est définie positive.

59.9 Produit scalaire sur $\mathbb{R}^n$

Soit $x, y \in \mathbb{R}^n$ tels que :

$$x = (x_1, x_2, ..., x_n)$$
$$y = (y_1, y_2, ..., y_n)$$

pour certains $x_i, y_i \in \mathbb{R}$.
Le produit scalaire usuel sur $\mathbb{R}^n$ est défini par :

$$\langle x \mid y \rangle = \sum_{i=1}^{n} x_i y_i$$

59.10 Produit scalaire sur $\mathbb{C}^n$

Soit $x, y \in \mathbb{C}^n$ tels que :

$$x = (x_1, x_2, ..., x_n)$$
$$y = (y_1, y_2, ..., y_n)$$

pour certains $x_i, y_i \in \mathbb{C}$.
Le produit scalaire usuel est défini par :

$$\langle x \mid y \rangle = \sum_{i=1}^{n} \overline{x}_i y_i$$

59.11 Base orthonormée sur $\mathbb{K}^n$

Soit $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Il est clair que la base canonique de $\mathbb{K}^n$:

$$e_i = (\boldsymbol{\delta}_{ij})_{i,j}$$

vérifie :

$$\langle e_i \mid e_j \rangle = \boldsymbol{\delta}_{ij}$$

pour le produit scalaire usuel sur $\mathbb{K}^n$. La suite $(e_1, ..., e_n)$ forme une base orthonormée.

59.12 Représentation matricielle

Soit $x = (x_1, .., x_n), y = (y_1, ..., y_n) \in \mathbb{C}^n$. On définit les vecteurs colonne associé :

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

L'équivalence entre $\mathbb{C}^n$ et $\mathbb{M}(\mathbb{C}, n, 1)$ nous amène à :

$$\langle x \mid y \rangle = \sum_i \overline{x}_i \cdot y_i$$

Le membre de droite n'est rien d'autre que le produit « matriciel » $\overline{x^T} \cdot y$ et on a donc :

$$\langle x \mid y \rangle = \overline{x^T} \cdot y$$

On vérifie que la base $(e_1, ..., e_n)$ est orthonormée pour ce produit scalaire :

$$e_i^T \cdot e_j = \boldsymbol{\delta}_{ij}$$

59.13 Application linéaire

Soit les espaces vectoriels E, F sur $\mathbb{K}$ et une application linéaire $\mathcal{A} : E \mapsto F$. On prend une base $(e_1, .., e_n)$ de E et une base orthonormée $(f_1, ..., f_m)$ de F. Comme les composantes de la matrice associée A sont les coordonnées de $\mathcal{A}(e_j)$ dans la base des f_i, on a :

$$\operatorname*{comp}_{ij} A = \langle f_i \mid \mathcal{A}(e_j) \rangle$$

59.14 Matrice de produit scalaire

Soit un espace vectoriel E sur $\mathbb{K}$ et un produit scalaire $\langle | \rangle$ quelconque défini sur E. Soit $(e_1, ..., e_n)$ une base quelconque de E (nous ne supposons pas qu'elle soit orthonormée). Si $\hat{x}, \hat{y} \in E$, on a :

$$\hat{x} = \sum_i x_i \cdot e_i$$
$$\hat{y} = \sum_i y_i \cdot e_i$$

pour certains $x_i, y_i \in S$. Or, nous avons vu que :

$$\langle \hat{x} \mid \hat{y} \rangle = \sum_{i,j} \overline{x}_i \, \langle e_i \mid e_j \rangle \, y_j$$

Si nous définissons la matrice des produits scalaires $A \in \mathbb{M}(\mathbb{K}, m, n)$ par :

$$\operatorname*{comp}_{ij} A = \langle e_i \mid e_j \rangle$$

nous pouvons réécrire le produit scalaire sous la forme :

$$\langle \hat{x} \mid \hat{y} \rangle = \overline{x^T} \cdot A \cdot y$$

où x, y sont les vecteurs colonne associés à $\hat{x}, \hat{y}$:

$$x = [x_1 \ x_2 \ ... \ x_n]^T$$
$$y = [y_1 \ y_2 \ ... \ y_n]^T$$

Cette matrice possède d'importantes propriétés issues du produit scalaire. On a $\overline{x^T} \cdot A \cdot x > 0$ pour tout $x \neq 0$. On dit que A est une matrice définie positive. Le caractère hermitien du produit scalaire nous donne aussi $\overline{A^T} = A$. On dit que A est une matrice hermitienne, ou auto-adjointe.

Réciproque

Si A est une matrice carrée définie positive et hermitienne, l'application définie par :

$$\langle x \mid y \rangle = \overline{x^T} \cdot A \cdot y$$

est bien un produit scalaire. En effet, la définie positivité de la matrice est équivalente à celle du produit ainsi défini. Pour le caractére hermitien, on a :

$$\begin{aligned}
\langle x \mid y \rangle &= \sum_{i,j} \overline{x}_i \cdot A_{ij} \cdot y_j = \sum_{i,j} \overline{x}_i \cdot \overline{A}_{ji} \cdot y_j \\
&= \sum_{i,j} y_j \cdot \overline{A}_{ji} \cdot \overline{x}_i = \operatorname{conj} \sum_{i,j} \overline{y}_j \cdot A_{ji} \cdot x_i \\
&= \operatorname{conj} \langle y \mid x \rangle
\end{aligned}$$

59.15 Bases de vecteurs matriciels

Un cas particulier important survient lorsque les vecteurs sont des vecteurs matriciels. Soit une suite de vecteurs linéairement indépendants $u_1, u_2, ..., u_m \in \mathbb{M}(\mathbb{K}, n, 1)$. Soit des $x_i, y_i \in \mathbb{K}$ et les vecteurs :

$$X = \sum_{i=1}^{m} x_i \cdot u_i$$
$$Y = \sum_{i=1}^{m} y_i \cdot u_i$$

Si nous considérons les vecteurs $x, y \in \mathbb{M}(\mathbb{K}, m, 1)$ associés :

$$x = [x_1 \ x_2 \ ... \ x_m]^T$$
$$y = [y_1 \ y_2 \ ... \ y_m]^T$$

ainsi que la matrice $U \in \mathbb{M}(\mathbb{K}, n, m)$ rassemblant les u_i :

$$U = [u_1 \ u_2 \ ... \ u_m]$$

on peut réécrire la définition de x, y sous la forme :

$$X = U \cdot x$$
$$Y = U \cdot y$$

On a alors :

$$\langle X \mid Y \rangle = \overline{X^T} \cdot Y = \overline{x^T} \cdot \overline{U^T} \cdot U \cdot y$$

On en conclut que la matrice $A = \overline{U^T} \cdot U \in \mathbb{M}(\mathbb{K}, m, m)$ est une matrice de produit scalaire

Réciproque

Soit $U \in \mathbb{M}(\mathbb{K}, n, m)$ telle que $\ker U = \{0\}$. La matrice $A = \overline{U^T} \cdot U \in \mathbb{M}(\mathbb{K}, m, m)$ est une matrice de produit scalaire. En effet :

$$\overline{x^T} \cdot A \cdot x = \overline{x^T} \cdot \overline{U^T} \cdot U \cdot x = \mathrm{conj}(U \cdot x)^T \cdot (U \cdot x) \geq 0$$

Si $x \neq 0$, on a de plus $U \cdot x \neq 0$ et $\overline{x}^T \cdot A \cdot x \ > \ 0$. Par ailleurs, on a évidemment :

$$\overline{A^T} = \overline{(\overline{U^T} \cdot U)^T} = \overline{U^T} \cdot U = A$$

59.16 Noyau

Soit $l_i = \mathrm{ligne}_i \, A$ et $x \in \ker A$. On a alors :

$$0 = \mathrm{comp}_i(A \cdot x) = l_i \cdot x$$

On en conclut que les lignes de A sont orthogonales aux $\bar{l}_i$. Il en va de même pour toute combinaison linéaire de ces lignes, et :

$$\ker A = \mathrm{span}\{\bar{l}_1, ..., \bar{l}_m\}^\perp$$

Chapitre 60

Norme dérivée du produit scalaire

Dépendances

— Chapitre 52 : Les espaces vectoriels
— Chapitre 53 : Les normes

60.1 Introduction

Soit un espace vectoriel E muni du produit scalaire $\langle | \rangle$. Nous allons analyser les propriétés de l'application $\|.\| : E \mapsto \mathbb{K}$ associée au produit scalaire et définie par :

$$\|x\| = \sqrt{\langle x \mid x \rangle}$$

pour tout $x \in E$.

60.2 Addition

Soit $x, y \in E$ et $\alpha, \beta \in \mathbb{C}$. On a :

$$\begin{aligned}
\|\alpha \cdot x + \beta \cdot y\|^2 &= \langle \alpha \cdot x + \beta \cdot y \mid \alpha \cdot x + \beta \cdot y \rangle \\
&= \overline{\alpha} \cdot \alpha \cdot \langle x \mid x \rangle + \overline{\alpha} \cdot \beta \cdot \langle x \mid y \rangle + \overline{\beta} \cdot \alpha \cdot \langle y \mid x \rangle + \overline{\beta} \cdot \beta \cdot \langle y \mid y \rangle \\
&= |\alpha|^2 \cdot \|x\|^2 + \overline{\alpha} \cdot \beta \cdot \langle x \mid y \rangle + \overline{\beta} \cdot \alpha \cdot \langle y \mid x \rangle + |\beta|^2 \cdot \|y\|^2
\end{aligned}$$

Dans le cas particulier où $\beta = 1$, on a :

$$\begin{aligned}
\|y + \alpha \cdot x\| &= \|y\|^2 + \alpha \cdot \langle y \mid x \rangle + \overline{\alpha} \cdot \langle x \mid y \rangle + |\alpha|^2 \cdot \|x\|^2 \\
&= \|y\|^2 + 2\Re(\alpha \cdot \langle y \mid x \rangle) + |\alpha|^2 \cdot \|x\|^2
\end{aligned}$$

60.3 Théorème de Pythagore

Si $x, y \in E$ sont orthogonaux :

$$\langle x \mid y \rangle = 0$$

on a également $\langle y \mid x \rangle = \operatorname{conj} \langle x \mid y \rangle = 0$ et :

$$\begin{aligned}
\langle x + y \mid x + y \rangle &= \langle x \mid x \rangle + \langle x \mid y \rangle + \langle y \mid x \rangle + \langle y \mid y \rangle \\
&= \langle x \mid x \rangle + \langle y \mid y \rangle
\end{aligned}$$

En exprimant cette relation en terme de $\|.\|$, on obtient :

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2$$

résultat connu sous le nom de théorème de Pythagore.

60.4 Egalité du parallélogramme

En additionnant les équations :

$$\|x+y\|^2 = \langle x \mid x\rangle + \langle x \mid y\rangle + \langle y \mid x\rangle + \langle y \mid y\rangle$$
$$\|x-y\|^2 = \langle x \mid x\rangle - \langle x \mid y\rangle - \langle y \mid x\rangle + \langle y \mid y\rangle$$

on obtient :

$$\|x+y\|^2 + \|x-y\|^2 = 2(\langle x \mid x\rangle + \langle y \mid y\rangle) = 2(\|x\|^2 + \|y\|^2)$$

60.5 Inégalité de Cauchy-Schwartz

Soit $x, y \in E$ et $\lambda \in \mathbb{C}$. On a :

$$\|y - \lambda \cdot x\|^2 = \langle y \mid y\rangle - \lambda \cdot \langle y \mid x\rangle - \overline{\lambda} \cdot \langle x \mid y\rangle + \overline{\lambda} \cdot \lambda \cdot \langle x \mid x\rangle \geq 0$$

Le choix magique de λ (nous verrons d'où il vient en étudiant les projections) est :

$$\lambda = \frac{\langle x \mid y\rangle}{\langle x \mid x\rangle}$$

On a alors :

$$\langle y \mid y\rangle - \frac{\langle x \mid y\rangle \cdot \langle y \mid x\rangle}{\langle x \mid x\rangle} - \frac{\langle y \mid x\rangle \cdot \langle x \mid y\rangle}{\langle x \mid x\rangle} + \frac{\langle y \mid x\rangle \cdot \langle x \mid y\rangle}{\langle x \mid x\rangle^2} \cdot \langle x \mid x\rangle \geq 0$$

En simplifiant les termes, on arrive à :

$$\langle y \mid y\rangle - \frac{\langle x \mid y\rangle \cdot \langle y \mid x\rangle}{\langle x \mid x\rangle} = \langle y \mid y\rangle - \frac{|\langle x \mid y\rangle|^2}{\langle x \mid x\rangle} \geq 0$$

En faisant passer le second terme dans le second membre et en multipliant par $\langle x \mid x\rangle$, on arrive finalement à :

$$|\langle x \mid y\rangle|^2 \leq \langle x \mid x\rangle \cdot \langle y \mid y\rangle$$

En prenant la racine carrée, on obtient une relation connue sous le nom d'inégalite de Cauchy-Schwartz :

$$|\langle x \mid y\rangle| \leq \sqrt{\langle x \mid x\rangle \cdot \langle y \mid y\rangle} = \|x\| \cdot \|y\|$$

60.6 Norme et inégalité de Minkowski

Nous allons à présent vérifier que l'application $\|.\|$ est bien une norme.

Définie positivité

On voit que notre application est strictement définie positive car $\|x\| \geq 0$ pour tout $x \in E$ et :

$$\|x\| = 0 \Rightarrow \langle x \mid x \rangle = 0 \Rightarrow x = 0$$

Produit mixte

La multiplication par un scalaire $\alpha \in \mathbb{C}$ nous donne :

$$\|\alpha \cdot x\| = \sqrt{|\alpha|^2 \cdot \langle x \mid x \rangle} = |\alpha| \cdot \sqrt{\langle x \mid x \rangle} = |\alpha| \cdot \|x\|$$

Inégalité de Minkowski

On a :

$$\begin{aligned} \|x+y\|^2 &= \langle x \mid x \rangle + \langle x \mid y \rangle + \langle y \mid x \rangle + \langle y \mid y \rangle \\ &= \langle x \mid x \rangle + 2\Re(\langle x \mid y \rangle) + \langle y \mid y \rangle \end{aligned}$$

Mais comme $|\Re(\langle x \mid y \rangle)| \leq |\langle x \mid y \rangle| \leq \|x\| \cdot \|y\|$, on a finalement :

$$\|x+y\|^2 \leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2$$

d'où :

$$\|x+y\| \leq \|x\| + \|y\|$$

Cette troisième et dernière propriété étant vérifiée, l'application $\|.\| = \sqrt{\langle | \rangle}$ est bien une norme.

60.7 Distance

On associe une distance à la norme et au produit scalaire par :

$$\mathfrak{dist}(x,y) = \|x-y\| = \sqrt{\langle x-y \mid x-y \rangle}$$

pour tout $x, y \in E$.

60.8 Produit scalaire à partir de la norme

En soutrayant les équations :

$$\begin{aligned} \|x+y\|^2 &= \langle x \mid x \rangle + 2\Re(\langle x \mid y \rangle) + \langle y \mid y \rangle \\ \|x-y\|^2 &= \langle x \mid x \rangle - 2\Re(\langle x \mid y \rangle) + \langle y \mid y \rangle \end{aligned}$$

on obtient :

$$\|x+y\|^2 - \|x-y\|^2 = 4\Re(\langle x \mid y \rangle)$$

Comme $\Re(\mathbf{i}\, z) = -\Im(z)$, on a aussi :

$$\begin{aligned} \|x + \mathbf{i}\, y\|^2 &= \langle x \mid x \rangle + 2\Re(\mathbf{i}\, \langle x \mid y \rangle) + \langle y \mid y \rangle \\ &= \langle x \mid x \rangle - 2\Im(\langle x \mid y \rangle) + \langle y \mid y \rangle \\ \|x - \mathbf{i}\, y\|^2 &= \langle x \mid x \rangle - 2\Re(\mathbf{i}\, \langle x \mid y \rangle) + \langle y \mid y \rangle \\ &= \langle x \mid x \rangle + 2\Im(\langle x \mid y \rangle) + \langle y \mid y \rangle \end{aligned}$$

En soustrayant ces deux résultats, on a donc :

$$\|x + \mathbf{i}\, y\|^2 - \|x - \mathbf{i}\, y\|^2 = -4\Im(\langle x \mid y \rangle)$$

On en conclut que :

$$\begin{aligned} \langle x \mid y \rangle &= \Re(\langle x \mid y \rangle) + \mathbf{i}\,\Im(\langle x \mid y \rangle) \\ &= \frac{1}{4}(\|x+y\|^2 + \|x-y\|^2) + \frac{\mathbf{i}}{4}(\|x - \mathbf{i}\, y\|^2 - \|x + \mathbf{i}\, y\|^2) \end{aligned}$$

60.9 Norme et coordonnées

Soit $(e_1, ..., e_n)$ une base de E et $u \in E$. On a :

$$u = \sum_i u_i \cdot e_i$$

pour certains $u_i, v_i \in \mathbb{K}$. La norme s'écrit alors :

$$\|u\| = \sqrt{\sum_{i,j} \overline{u}_i \cdot \langle e_i \mid e_j \rangle \cdot u_j}$$

Base orthonormée

Si la base est orthonormée, les seuls termes ne s'annulant pas sont ceux où $i = j$, et on a :

$$\|u\| = \sqrt{\sum_i |u_i|^2}$$

60.10 Norme sur $\mathbb{K}^n$

Soit $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. On définit une norme sur $\mathbb{K}^n$, dite norme euclidienne, à partir du produit scalaire :

$$\|x\| = \sqrt{\langle x \mid x \rangle} = \sqrt{\sum_{i=1}^n |x_i|^2}$$

Normes k

Par extension, on définit une série de normes k par :

$$\|x\|_k = \left(\sum_i |x_i|^k \right)^{1/k}$$

Lorsque k devient très grand, il est clair que la contribution du $|x_i|^k$ le plus grand en valeur absolue devient énorme par rapport aux autres contributions de la norme. On peut vérifier que :

$$\lim_{k \mapsto +\infty} \|x\|_k = \max_{i=1}^n |x_i|$$

On s'inspire de ce résultat pour définir :

$$\|x\|_\infty = \max_{i=1}^n |x_i|$$

On nomme $\|.\|_\infty$ la norme « max ».

Attention, une norme k quelconque ne dérive en général pas d'un produit scalaire et ne possède donc pas les propriétés que nous avons vu pour la norme $\|.\| = \|.\|_2 = \sqrt{\langle | \rangle}$.

60.11 Norme sur $\mathbb{C}$

Soit $(a, b) \in \mathbb{R}^2$ et $z = a + \mathbf{i}\, b$. Il est clair que le module :

$$|z| = |a + \mathbf{i}\, b| = \sqrt{a^2 + b^2} = \|(a, b)\|$$

définit une norme sur $\mathbb{C}$.

60.12 Représentation matricielle

Soit le vecteur matriciel $x = [x_1 \; ... \; x_n]^T$. Sa norme s'écrit :

$$\|x\| = \sqrt{\langle x \mid x \rangle} = \sqrt{\overline{x}^T \cdot x}$$

Chapitre 61

Applications adjointes

Dépendances

- Chapitre 58 : Les formes linéaires
- Chapitre 60 : Les produits scalaires

61.1 Adjoint au sens des formes linéaires

Soit les espaces vectoriels E et F sur $\mathbb{K}$ et l'application linéaire $A : E \mapsto F$. Si $A^\star : F^\star \mapsto E^\star$ est l'unique fonction de E^F vérifiant :

$$\langle \varphi, A(u) \rangle = \langle A^\star(\varphi), u \rangle$$

pour tout $\varphi \in F^\star$ et $u \in E$, on dit que $A^\star$ est l'application duale (ou adjointe) de A au sens des formes linéaires.

61.2 Adjoint au sens du produit scalaire

Soient E et F deux espaces vectoriels munis de produits scalaires et l'application linéaire $A : E \mapsto F$. Si $A^\star : F \mapsto E$ est l'unique fonction de E^F vérifiant :

$$\langle v \mid A(u) \rangle = \langle A^\star(v) \mid u \rangle$$

pour tout $(u, v) \in E \times F$, on dit que $A^\star$ est l'application duale (ou adjointe) de A au sens du produit scalaire. Nous supposons dans la suite que les applications rencontrées possèdent une application adjointe.

Applications auto-adjointes

Si $E = F$ et $A = A^\star$, on dit que A est hermitienne ou auto-adjointe.

61.3 Identité

Comme :

$$\langle v \mid \mathrm{Id}(u) \rangle = \langle \mathrm{Id}(v) \mid u \rangle = \langle v \mid u \rangle$$

on a bien évidemment $\mathrm{Id}^\star = \mathrm{Id}$.

61.4 Adjoint d'une combinaison linéaire

Soit deux applications linéaires $A, B : E \mapsto F$ et $\alpha, \beta \in \mathbb{K}$. Si $(u, v) \in E \times F$, on a :

$$\begin{aligned}
\big\langle (\overline{\alpha} \cdot A^\star + \overline{\beta} \cdot B^\star)(v) \mid u \big\rangle &= \alpha \cdot \langle A^\star(v) \mid u \rangle + \beta \cdot \langle B^\star(v) \mid u \rangle \\
&= \alpha \cdot \langle v \mid A(u) \rangle + \beta \cdot \langle v \mid A(u) \rangle \\
&= \langle v \mid (\alpha \cdot A + \beta \cdot B)(u) \rangle
\end{aligned}$$

On en conclut que :

$$(\alpha \cdot A + \beta \cdot B)^\star = \overline{\alpha} \cdot A^\star + \overline{\beta} \cdot B^\star$$

L'opérateur $^\star : A \mapsto A^\star$ est antilinéaire.

61.5 Bidual

On remarque que :

$$\langle u \mid A^\star(v) \rangle = \overline{\langle A^\star(v) \mid u \rangle} = \overline{\langle v \mid A(u) \rangle} = \langle A(u) \mid v \rangle$$

pour tout $(u, v) \in E \times F$. On a donc :

$$(A^\star)^\star = A$$

61.6 Adjoint d'une composée

Soit un troisième espace vectoriel G. Si les applications adjointes des applications linéaires $A : E \mapsto F$ et $B : F \mapsto G$ existent, on a :

$$\langle v \mid (A \circ B)(u) \rangle = \langle A^\star(v) \mid B(u) \rangle = \langle (B^\star \circ A^\star)(v) \mid u \rangle$$

pour tout $(u, v) \in E \times G$. On en conclut que :

$$(A \circ B)^\star = B^\star \circ A^\star$$

61.7 Construction d'applications auto-adjointes

Nous allons voir que nous pouvons construire deux applications auto-adjointes à partir de n'importe quelle application linéaire $A : E \mapsto F$ admettant une application duale $A^\star : F \mapsto E$.

— L'application $A^\star \circ A : E \mapsto E$ vérifie :

$$\langle A^\star \circ A(v) \mid u \rangle = \langle A(v) \mid A(u) \rangle = \langle v \mid A^\star \circ A(u) \rangle$$

pour tout $u, v \in E$. On en déduit que :

$$(A^\star \circ A)^\star = A^\star \circ A$$

— L'application $A \circ A^\star : F \mapsto F$ vérifie :

$$\langle A \circ A^\star(x) \mid y \rangle = \langle A^\star(x) \mid A^\star(y) \rangle = \langle x \mid A \circ A^\star(y) \rangle$$

pour tout $x, y \in F$. On en déduit que :

$$(A \circ A^\star)^\star = A \circ A^\star$$

61.8 Linéarité de l'adjoint

Soit $u, v \in E$, $x, y \in F$ et $\alpha, \beta \in \mathbb{K}$. On a :

$$\begin{aligned}
\langle A^\star(\alpha \cdot x + \beta \cdot y) \mid u \rangle &= \langle \alpha \cdot x + \beta \cdot y \mid A(u) \rangle \\
&= \overline{\alpha} \cdot \langle x \mid A(u) \rangle + \overline{\beta} \cdot \langle y \mid A(u) \rangle \\
&= \overline{\alpha} \cdot \langle A^\star(x) \mid u \rangle + \overline{\beta} \cdot \langle A^\star(y) \mid u \rangle \\
&= \langle \alpha \cdot A^\star(x) + \beta \cdot A^\star(y) \mid u \rangle
\end{aligned}$$

Comme ce doit être valable pour tout $u \in E$, on en conclut que l'application adjointe est linéaire :

$$A^\star(\alpha \cdot x + \beta \cdot y) = \alpha \cdot A^\star(x) + \beta \cdot A^\star(y)$$

61.9 Norme de l'adjoint

Soit une application linéaire $A : E \mapsto F$ de norme finie. Si $x \in F$ est un vecteur non nul, on a :

$$\|A^\star(x)\|^2 = \langle A^\star(x) \mid A^\star(x) \rangle = \langle A \circ A^\star(x) \mid x \rangle \le \|A \circ A^\star(x)\| \cdot \|x\|$$

et donc :

$$\|A^\star(x)\|^2 \le \|A\| \cdot \|A^\star(x)\| \cdot \|x\|$$

Si $\|A^\star(x)\| \neq 0$, on peut diviser par $\|A^\star(x)\|$. On obtient alors :

$$\|A^\star(x)\| \le \|A\| \cdot \|x\|$$

Par positivité des normes, on remarque que cette relation est également valable lorsque $\|A^\star(x)\| = 0 \le \|A\| \cdot \|x\|$. En divisant par la norme de x, on obtient :

$$\frac{\|A^\star(x)\|}{\|x\|} \le \|A\|$$

Il ne nous reste plus qu'à passer au supremum sur x pour en conclure que :

$$\|A^\star\| \le \|A\|$$

Mais comme $(A^\star)^\star = A$, on a aussi :

$$\|A\| = \|(A^\star)^\star\| \le \|A^\star\|$$

Ces deux inégalités nous montrent que :

$$\|A^\star\| = \|A\|$$

61.10 Inverse

Supposons que A soit inversible. On a :

$$\begin{aligned}
\mathrm{Id} = \mathrm{Id}^\star = (A^{-1} \circ A)^\star = A^\star \circ (A^{-1})^\star \\
\mathrm{Id} = \mathrm{Id}^\star = (A \circ A^{-1})^\star = (A^{-1})^\star \circ A^\star
\end{aligned}$$

On en conclut que $A^\star$ est également inversible et que :

$$(A^\star)^{-1} = (A^{-1})^\star$$

61.11 Noyau et image

Soit l'application linéaire $A : E \mapsto F$. Soit $u \in \ker A$ et $v \in \operatorname{im} A^\star$. On peut donc trouver un $x \in F$ tel que $v = A^\star(x)$. On a :

$$\langle u \mid v \rangle = \langle u \mid A^\star(x) \rangle = \langle A(u) \mid x \rangle = \langle 0 \mid x \rangle = 0$$

d'où $u \in (\operatorname{im} A^\star)^\perp$ et $\ker A \subseteq (\operatorname{im} A^\star)^\perp$. Inversément, si $u \in (\operatorname{im} A^\star)^\perp$, on a :

$$\langle A(u) \mid x \rangle = \langle u \mid A^\star(x) \rangle = 0$$

pour tout $x \in E$. On en conclut que $A(u) = 0$, c'est-à-dire $u \in \ker A$. On a donc aussi $(\operatorname{im} A^\star)^\perp \subseteq \ker A$. Ces deux inclusions nous montrent finalement que :

$$\ker A = (\operatorname{im} A^\star)^\perp$$

Comme le bidual revient à l'application d'origine, on a aussi :

$$\ker A^\star = (\operatorname{im} A)^\perp$$

61.12 Représentation matricielle

Soit $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ et la matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$ représentant l'application linéaire $\mathcal{A} : \mathbb{K}^n \mapsto \mathbb{K}^m$. Soit $A^\star \in \mathbb{M}(\mathbb{K}, n, m)$ représentant $\mathcal{A}^\star$. On a :

$$\langle y \mid \mathcal{A}(x) \rangle = \overline{y^T} \cdot A \cdot x$$

ainsi que :

$$\langle \mathcal{A}^\star(y) \mid x \rangle = \left(\overline{A^\star} \cdot \overline{y} \right)^T \cdot x = \overline{y^T} \cdot \left(\overline{A^\star} \right)^T \cdot x$$

Les deux produits scalaires devant être égaux par définition de la dualité, on doit clairement avoir :

$$\left(\overline{A^\star} \right)^T = A$$

c'est-à-dire :

$$A^\star = \overline{A^T} = \operatorname{conj} A^T$$

Ce résultat prouve l'existence et l'unicité de l'adjoint dans le cas d'espaces de dimension finie.

Cas particulier

Dans le cas d'une matrice réelle, on a $\overline{A} = A$ et :

$$A^\star = A^T$$

61.13 Adjoint d'un produit

On peut vérifier que :

$$(A \cdot B)^\star = B^\star \cdot A^\star$$

pour toutes matrices $A \in \mathbb{M}(\mathbb{K}, m, n)$ et $B \in \mathbb{M}(\mathbb{K}, n, p)$ de dimensions compatibles pour la multiplication.

Chapitre 62

Tenseurs

Dépendances

— Chapitre 52 : Les vecteurs
— Chapitre 56 : Les fonctions linéaires
— Chapitre 58 : Les formes
— Chapitre 60 : Les produits scalaires

62.1 Introduction

Nous présentons deux variantes de la définition des tenseurs. La première, classique, est basée sur les formes linéaires. La seconde, basée sur les vecteurs et la généralisation des produits scalaires, a l'avantage de mettre en relief la structure particulière des tenseurs, ainsi que la multitude de fonctions que l'on peut leur associer.

62.2 Produit tensoriel de formes linéaires

Soit deux espaces vectoriels E et F sur $\mathbb{K}$, ainsi que les fonctions $\varphi \in \operatorname{Lin}(E, S)$ et $\psi \in \operatorname{Lin}(F, S)$. On définit le produit tensoriel de ces deux fonctions, noté $\varphi \otimes \psi$, par :

$$(\varphi \otimes \psi)(a, b) = \varphi(a) \cdot \psi(b)$$

pour tout $a \in E$ et $b \in F$. La fonction $\varphi \otimes \psi$ est donc une forme bilinéaire vérifiant :

$$\langle a, \varphi \otimes \psi, b \rangle = \langle \varphi, a \rangle \cdot \langle \psi, b \rangle$$

Forme associée

On peut réécrire la définition comme :

$$(\varphi \otimes \psi)(a, b) = \langle \psi(b) \cdot \varphi, a \rangle = (\psi(b) \cdot \varphi)(a)$$

Nous pouvons donc associer à chaque $b \in F$ une forme linéaire $\psi(b) \cdot \varphi$ que nous notons :

$$\langle \varphi \otimes \psi, b \rangle = \varphi \cdot \psi(b)$$

Dualité

On voit en échangeant φ et ψ que :

$$(\psi \otimes \varphi)(b, a) = \psi(b) \cdot \varphi(a) = (\varphi \otimes \psi)(a, b)$$

Associativité

Le produit tensoriel étant associatif, on note :

$$\varphi \otimes \psi \otimes \omega = \varphi \otimes (\psi \otimes \omega) = (\varphi \otimes \psi) \otimes \omega$$

pour toutes formes linéaires φ, ψ, ω définies sur les espaces vectoriels E, F, G.

Notation

On convient également de la notation :

$$\big(\langle \psi \otimes \varphi, u \rangle \big)(x) = \psi(x) \cdot \langle \varphi, u \rangle$$

pour tout $x \in F$. On a donc :

$$\langle \psi \otimes \varphi, u \rangle = \psi \cdot \langle \varphi, u \rangle$$

62.3 Tenseurs de formes linéaires

Soit les espaces vectoriels $E_1, E_2, ..., E_n$ sur $\mathbb{K}$. On nomme tenseur d'ordre n les formes n-linéaires de $E_1 \times E_2 \times ... \times E_n$ vers $\mathbb{K}$.

62.4 Produit tensoriel de deux vecteurs

Soit les espaces vectoriels E et F sur $\mathbb{K}$. Choisissons $a \in F$ et $b \in E$. Le produit tensoriel $a \otimes b$ est l'application linéaire de E vers F définie par :

$$(a \otimes b)(c) = a \cdot \langle b \mid c \rangle$$

pour tout $c \in E$. On vérifie aisément que ce produit est bilinéaire.

Associativité mixte

Il est clair d'après les définitions des opérations que :

$$(\alpha \cdot a) \otimes b = \alpha \cdot (a \otimes b)$$
$$a \otimes (b \cdot \alpha) = (a \otimes b) \cdot \alpha$$

pour tout $\alpha \in \mathbb{K}$.

Distributivité

On a :

$$(a + b) \otimes c = a \otimes c + b \otimes c$$
$$a \otimes (c + d) = a \otimes c + a \otimes d$$

pour tout $a, b \in F$ et $c, d \in E$.

62.5 Contractions

Soit les espaces vectoriels E, F, G sur $\mathbb{K}$. On étend la notion de « produit scalaire » par les relations :

$$\langle a \otimes b \mid c \rangle = a \cdot \langle b \mid c \rangle$$
$$\langle b \mid c \otimes d \rangle = \langle b \mid c \rangle \cdot d$$
$$\langle a \otimes b \mid c \otimes d \rangle = \langle b \mid c \rangle \cdot (a \otimes d)$$

valables pour tout $(a, b, c, d) \in G \times F \times F \times E$.

62.6 Contractions doubles

Soit $b \in F$ et $c \in E$. Afin de rester consistant avec le produit tensoriel des formes, nous définissons la forme associée à $b \otimes c$ par :

$$\varphi(a, d) = \langle a \,|(\, b \otimes c)(d) \rangle = \langle a \mid b \rangle \cdot \langle c \mid d \rangle$$

pour tout $d \in E$ et $a \in F$. On note en général cette forme au moyen de la double contraction :

$$\langle a | \, b \otimes c \, | d \rangle = \langle a \mid b \rangle \cdot \langle c \mid d \rangle$$

62.7 Dualité

Soit le scalaire $\alpha \in \mathbb{C}$ et les vecteurs $a, u \in E_1$ et $b, v \in E_2$. On a :

$$\langle u \mid \alpha \cdot (a \otimes b)(v) \rangle = \alpha \cdot \langle u \mid a \rangle \cdot \langle b \mid v \rangle$$

et :

$$\begin{aligned} \langle \overline{\alpha} \cdot (b \otimes a)(u) \mid v \rangle &= \alpha \cdot \mathrm{conj}(\langle a \mid u \rangle) \cdot \langle b \mid v \rangle \\ &= \alpha \cdot \langle u \mid a \rangle \cdot \langle b \mid v \rangle \end{aligned}$$

On en déduit que :

$$(\alpha \cdot a \otimes b)^\star = \overline{\alpha} \cdot b \otimes a$$

62.8 Combinaison linéaire

Soit des vecteurs a_i, b_i, c_i, d_i et des scalaires $\theta_{ij}, \upsilon_{ij}$. On étend la définition des contractions à des combinaisons linéaires de la forme :

$$T = \sum_{i,j} \theta_{ij} \cdot a_i \otimes b_j$$
$$U = \sum_{i,j} \upsilon_{ij} \cdot c_i \otimes d_j$$

en imposant simplement la linéarité. On a donc par exemple :

$$\langle T \mid U \rangle = \sum_{i,j,k,l} \theta_{ij} \cdot \upsilon_{kl} \cdot \langle b_j \mid c_k \rangle \cdot (a_i \otimes d_l)$$

62.9 Tenseurs d'ordre deux

Soit $a \in F$ et $b \in E$. Un objet de la forme :

$$T = a \otimes b$$

est un cas particulier de « tenseur d'ordre deux ». Il est formellement défini comme une application linéaire de E vers F, mais il est en fait bien plus riche puisqu'on peut associer différents types de fonctions et de formes à chaque contraction possible impliquant ce tenseur. Les notions de tenseur et de contraction sont en fait étroitement liées.

On note $\mathbb{T}_2(F, E)$ l'espace vectoriel généré par ce type de tenseurs d'ordre deux.

62.10 Tenseurs d'ordre un et zéro

La possibilité d'associer une forme linéaire à chaque vecteur $a \in E$ par la contraction avec un autre vecteur, qui est dans ce cas un simple produit scalaire :

$$\varphi_a(b) = \langle a \mid b \rangle$$

nous incite à considérer les vecteurs comme des « tenseurs d'ordre un ». Nommant $\mathbb{T}_1(E)$ l'ensemble des tenseurs d'ordre un, on a simplement $\mathbb{T}_1(E) = E$. Quant aux scalaires, ils seront considérés comme des « tenseurs d'ordre zéro ». On note $\mathbb{T}_0 = \mathbb{K}$ l'ensemble des tenseurs d'ordre zéro.

62.11 Associativité

Soit $a \in G$, $b \in F$ et $c \in E$. Comme $\mathrm{Lin}(E, F)$ est également un espace vectoriel, on a :

$$\Big[a \otimes \big[(b \otimes c)(u)\big]\Big](v) = \big[a \otimes b \, \langle c \mid u \rangle \,\big](v) = a \, \langle c \mid u \rangle \, \langle b \mid v \rangle$$

valable pour tout $u \in E$ et tout $v \in F$. Mais $\mathrm{Lin}(F, G)$ est aussi un espace vectoriel, et l'on a aussi :

$$\Big[\big[(a \otimes b)(v)\big] \otimes c\Big](u) = \Big[\, \langle b \mid v \rangle \, a \otimes c\Big](u) = \langle b \mid v \rangle \, \langle c \mid u \rangle \, a$$

Le résultat étant le même, le produit tensoriel est associatif, et nous notons :

$$a \otimes b \otimes c = a \otimes (b \otimes c) = (a \otimes b) \otimes c$$

l'application linéaire définie par :

$$(a \otimes b \otimes c)(u, v) = \langle c \mid u \rangle \, \langle b \mid v \rangle \, a$$

62.12 Tenseur d'ordre n

Considérons les espaces vectoriels $E_1, ..., E_n$, les séries de vecteurs $a_k^1, a_k^2, ...a_k^{N_k} \in E_k$ et les scalaires $\theta_{ij...r} \in \mathbb{K}$. Un tenseur d'ordre n est une combinaison linéaire de la forme :

$$T = \sum_{i,j,...,r} \theta_{ij...r} \cdot a_1^i \otimes a_2^j \otimes ... \otimes a_n^r$$

On note $\mathbb{T}_n(E_1, E_2, ..., E_n)$ l'espace des tenseurs d'ordre n.

Lorsque tous les espaces vectoriels sont égaux, soit $E = E_1 = E_2 = ... = E_n$, on note $\mathbb{T}_n(E) = \mathbb{T}_n(E, E, ..., E)$.

Attention

Ne pas confondre les *indices supérieurs* (le i des vecteurs a_j^i par exemple), très utilisés en calcul tensoriel, avec les puissances ! Dans le contexte des tenseurs, une éventuelle puissance d'un scalaire θ_i^j serait notée au besoin par :

$$\left(\theta_j^i\right)^m = \theta_j^i \cdot ... \cdot \theta_j^i$$

62.13 Dualité

Soit les séries de vecteurs $a_k^i \in E_k$, les scalaires $\theta_{ij...rs} \in \mathbb{K}$, et le tenseur associé :

$$A = \sum_{i,j,...,r,s} \theta_{ij...rs} \cdot a_1^i \otimes a_2^j \otimes ... \otimes a_{n-1}^r \otimes a_n^s$$

On vérifie que le dual s'obtient en inversant l'ordre des vecteurs et en conjuguant les coordonnées :

$$A^\star = \sum_{i,j,...,r,s} \overline{\theta}_{ij...rs} \cdot a_n^s \otimes a_{n-1}^r \otimes ... \otimes a_2^j \otimes a_1^i$$

62.14 Réduction

Soit les séries de vecteurs $a_k^i \in E_k$, les scalaires $\theta_{ij...rs} \in \mathbb{K}$, et le tenseur associé :

$$A = \sum_{i,j,...,r,s} \theta_{ij...rs} \cdot a_1^i \otimes a_2^j \otimes ... \otimes a_{n-1}^r \otimes a_n^s$$

Les opérations de réduction consistent à construire des tenseurs d'ordre $n-1$, notés $A_-(s)$ et $A^-(i)$, en retirant les vecteurs a_n^s (réduction à droite) ou les vecteurs a_1^i (réduction à gauche) :

$$A_-(s) = \sum_{i,j,...,r} \theta_{ij...rs} \cdot a_1^i \otimes a_2^j \otimes ... \otimes a_{n-1}^r$$

$$A^-(i) = \sum_{j,...,r,s} \theta_{ij...rs} \cdot a_2^j \otimes ... \otimes a_{n-1}^r \otimes a_n^s$$

62.15 Contraction généralisée

Soit les séries de vecteurs $a_k^i \in E_k$ et $b_k^j \in F_k$, les scalaires $\eta_{i...r}, \theta_{j...s} \in \mathbb{K}$, et les tenseurs :

$$A = \sum_{i,...,r} \eta_{i...r} \cdot a_1^i \otimes ... \otimes a_m^r$$

$$B = \sum_{j,...,s} \theta_{j...s} \cdot b_1^j \otimes ... \otimes b_n^s$$

La contraction d'ordre 0 consiste simplement à juxtaposer les deux tenseurs au moyen du produit tensoriel :

$$\langle A \odot B \rangle_0 = \sum_{i,...,r,j,...,s} \eta_{i...r} \cdot \theta_{j...s} \cdot a_1^i \otimes ... \otimes a_m^r \otimes b_1^j \otimes ... \otimes b_n^s$$

Soit $p \in \mathbb{N}$ tel que $0 \le p \le \min\{m, n\}$. Si :

$$E_m = F_1$$
$$E_{m-1} = F_2$$
$$\vdots$$
$$E_{m-p+1} = F_p$$

on peut définir la contraction $\langle\odot\rangle_p$ d'ordre p de deux tenseurs par récurrence :

$$\langle A \odot B\rangle_p = \sum_{r,j} \left\langle a_m^r \mid b_1^j \right\rangle \cdot \left\langle A_-(r) \odot B^-(j) \right\rangle_{p-1}$$

Notation

Soit A un tenseur d'ordre m et B un tenseur d'ordre n. Le produit tensoriel est bien évidemment identique à la contraction d'ordre 0. On le note :

$$A \otimes B = \langle A \odot B\rangle_0$$

La contraction d'ordre 1 est notée comme un produit scalaire :

$$\langle A \mid B\rangle = \langle A \odot B\rangle_1$$

ou comme un produit lorsqu'il n'y a pas de confusion possible :

$$A \cdot B = \langle A \mid B\rangle$$

Enfin, la contraction maximale d'ordre $M = \min\{m, n\}$ est notée par :

$$A : B = \langle A \odot B\rangle_M$$

Exemple

Considérons le cas :

$$A = a_1 \otimes ... \otimes a_m$$
$$B = b_1 \otimes ... \otimes b_n$$

En utilisant p fois la récurrence, on obtient :

$$\langle A \odot B\rangle_p = \langle a_m \mid b_1\rangle \cdot ... \cdot \langle a_{m-p+1} \mid b_p\rangle \cdot a_1 \otimes ... \otimes a_{m-p} \otimes b_{p+1} \otimes ... \otimes b_n$$

soit un tenseur d'ordre $m - p + n - p$.

62.16 Contraction double généralisée

On définit également la contraction double utilisant trois facteurs par :

$$\langle Y | \, T \, | X\rangle_{m,n} = \langle Y \odot \langle T \odot X\rangle_n\rangle_m = \langle\langle Y \odot T\rangle_m \odot X\rangle_n$$

Notation

On note :

$$\langle y | \, A \, | x\rangle = \langle y | \, A \, | x\rangle_{1,1}$$
$$y \cdot A \cdot x = \langle y | \, A \, | x\rangle$$

62.17 Coordonnées

Si chaque E_k dispose d'une base de vecteurs $(e_k^1, e_k^2, ...e_k^{N_k})$, tout tenseur $T \in \mathbb{T}_n(E_1, E_2, ..., E_n)$ peut être écrit sous la forme :

$$T = \sum_{i,...,r} \theta_{i...r} \cdot e_1^i \otimes ... \otimes e_n^r$$

On dit que les $\theta_{i...r}$ sont les coordonnées de T. L'indépendance linéaire des e_k^i nous garantit que ces coordonnées sont uniques.

Il y a donc :

$$N = \prod_i N_i$$

coordonnées déterminant un tenseur.

Bases orthonormées

Si les bases sont orthonormées :

$$\left\langle e_k^i \mid e_k^j \right\rangle = \boldsymbol{\delta}_{ij}$$

les coordonnées s'obtiennent aisément au moyen des contractions. On évalue :

$$\left\langle e_n^s \otimes ... \otimes e_1^j \odot T \right\rangle_n = \sum_{i,...,r} \theta_{i...r} \cdot \left\langle e_1^j \mid e_1^i \right\rangle \cdot ... \cdot \left\langle e_n^s \mid e_n^r \right\rangle$$

$$\left\langle e_n^s \otimes ... \otimes e_1^j \odot T \right\rangle_n = \sum_{i,...,s} \theta_{i...r} \cdot \boldsymbol{\delta}_{ji} \cdot ... \cdot \boldsymbol{\delta}_{sr}$$

Le produit des deltas de Kronecker s'annulant partout sauf aux indices $(i, ..., r) = (j, ..., s)$, on a finalement :

$$\left\langle e_n^s \otimes ... \otimes e_1^j \odot T \right\rangle_n = \theta_{j...s}$$

ce qui nous donne les valeurs des $\theta_{j...s}$.

62.18 Norme

La norme d'un tenseur A est analogue aux normes dérivées d'un produit scalaire. On utilise ici la contraction maximale et le tenseur duel de A, afin que les espaces des vecteurs soient compatibles :

$$\|A\| = \sqrt{A^\star : A}$$

Bases orthonormées

Supposons que A soit représenté par rapport aux bases orthonormées $(e_k^1, e_k^2, ...e_k^{N_k})$:

$$A = \sum_{i,...,r} \theta_{i...r} \cdot e_1^i \otimes ... \otimes e_n^r$$

pour certains $\theta_{i...r} \in \mathbb{K}$. La norme nous donne dans ce cas :

$$\|A\| = \sqrt{A^\star : A} = \sqrt{\sum_{i,...,r} |\theta_{i...r}|^2}$$

62.19 Tenseur identité

Le tenseur identité $\boldsymbol{\mathcal{I}} \in \mathbb{T}_2(E)$ est défini par :

$$\boldsymbol{\mathcal{I}} \cdot u = \langle \boldsymbol{\mathcal{I}} \odot u \rangle_1 = u$$

pour tout $u \in E$.

Base orthonormée

Soit $(e_1, ..., e_n)$ une base orthonormée de E. On a alors :

$$\boldsymbol{\mathcal{I}} \cdot e_i = e_i$$

et :

$$\langle \boldsymbol{\mathcal{I}} \odot e_j \otimes e_i \rangle_2 = \langle \boldsymbol{\mathcal{I}} \cdot e_i \mid e_j \rangle = \boldsymbol{\delta}_{ij}$$

On en conclut que :

$$\boldsymbol{\mathcal{I}} = \sum_{i,j} \boldsymbol{\delta}_{ij} \cdot e_i \otimes e_j = \sum_i e_i \otimes e_i$$

62.20 Inverse

Soit un tenseur $T \in \mathbb{T}_2(E)$. Si il existe un tenseur $T^{-1} \in \mathbb{T}_2(E)$ dont les contractions d'ordre 1 avec T donnent le tenseur identité :

$$T \cdot T^{-1} = \left\langle T \odot T^{-1} \right\rangle_1 = T^{-1} \cdot T = \left\langle T^{-1} \odot T \right\rangle_1 = \boldsymbol{\mathcal{I}}$$

on dit que T^{-1} est l'inverse de T.

62.21 Trace

Soit $(e_1, ..., e_n)$ est une base orthonormée de E et le tenseur $T \in \mathbb{T}_2(E)$:

$$T = \sum_{i,j} \theta_{ij} \cdot e_i \otimes e_j$$

On définit sa trace par :

$$\operatorname{tr} T = T : \boldsymbol{\mathcal{I}} = \sum_{i,j} \theta_{ij} \cdot \boldsymbol{\delta}_{ij} = \sum_i \theta_{ii}$$

62.22 Cadre

On dit que les vecteurs $e_i \in E$ forment un cadre de E si :

$$\sum_i e_i \otimes e_i = \boldsymbol{\mathcal{I}}$$

On a alors, pour tout vecteur u de E :

$$\sum_i \langle e_i \otimes e_i \mid u \rangle = u$$

c'est-à-dire :

$$u = \sum_i \langle e_i \mid u \rangle \, e_i$$

par définition de la contraction. Prenant le produit scalaire avec un autre vecteur $v \in F$, on obtient :

$$\langle v \mid u \rangle = \sum_i \langle v \mid e_i \rangle \langle e_i \mid u \rangle$$

62.23 Représentation matricielle

Soit les vecteurs $u \in F = \mathbb{M}(K, m, 1) \equiv \mathbb{K}^m$ et $v, w \in E = \mathbb{M}(K, n, 1) \equiv \mathbb{K}^n$. La matrice :

$$A = u \cdot v^\star = \begin{bmatrix} u_1 \cdot \bar{v}_1 & u_1 \cdot \bar{v}_2 & \dots & u_1 \cdot \bar{v}_n \\ u_2 1 \cdot \bar{v}_1 & u_2 \cdot \bar{v}_2 & \dots & u_2 \cdot \bar{v}_n \\ \vdots & & \ddots & \vdots \\ u_m \cdot \bar{v}_1 & u_m \cdot \bar{v}_2 & \dots & u_m \cdot \bar{v}_n \end{bmatrix}$$

représente une application linéaire de $\mathbb{K}^n$ vers $\mathbb{K}^m$. Comme le produit matriciel est associatif, cette fonction possède la propriété :

$$A \cdot w = (u \cdot v^\star) \cdot w = u \cdot (v^\star \cdot w) = u \cdot \langle v \mid w \rangle$$

Ce résultat étant identique à la définition d'un tenseur d'ordre deux, on voit que les matrices sont équivalentes à des tenseurs d'ordre deux : $\mathbb{M}(K, m, n) \equiv \mathbb{T}_2(\mathbb{K}^m, \mathbb{K}^n)$. On parle donc de tenseurs matriciels pour désigner cette représentation. On définit par équivalence le produit tensoriel de deux vecteurs matriciels par :

$$u \otimes v = u \cdot v^\star$$

62.24 Contraction d'ordre 1

Soit les espaces vectoriels E_1, E_2, E_3 sur $\mathbb{K}$ et les suites de vecteurs $(e_k^1, e_k^2, ...e_k^{N_k})$ formant des bases orthonormées des E_k. Soit les tenseurs :

$$\mathcal{A} = \sum_{i,k} a_{ik} \cdot e_1^i \otimes e_2^k$$

$$\mathcal{B} = \sum_{l,j} b_{lj} \cdot e_2^l \otimes e_3^j$$

où $a_{ij}, b_{ij} \in \mathbb{K}$. Calculons leur contraction d'ordre 1 :

$$\mathcal{C} = \langle \mathcal{A} \odot \mathcal{B} \rangle_1 = \sum_{i,j,k,l} a_{ik} \cdot b_{lj} \cdot \left\langle e_2^k \mid e_2^l \right\rangle \cdot e_1^i \otimes e_3^j$$

$$\mathcal{C} = \sum_{i,j,k,l} a_{ik} \cdot b_{lj} \cdot \boldsymbol{\delta}_{kl} \cdot e_1^i \otimes e_3^j$$

$$\mathcal{C} = \sum_{i,j,k} a_{ik} \cdot b_{kj} \cdot e_1^i \otimes e_3^j$$

On voit que les coordonnées obtenues :

$$c_{ij} = \sum_k a_{ik} \cdot b_{kj}$$

correspondent à une matrice $C = (c_{ij})_{i,j}$ telle que :

$$C = A \cdot B$$

où les matrices A et B sont données par :

$$A = (a_{ij})_{i,j}$$
$$B = (b_{ij})_{i,j}$$

La contraction d'ordre 1 correspond donc au produit matriciel, qui correspond lui-même à la composition d'application linéaires.

62.25 Contraction maximale

Soit les espaces vectoriels E_1, E_2, E_3 sur $\mathbb{K}$ et les suites de vecteurs $(e_k^1, e_k^2, ...e_k^{N_k})$ formant des bases orthonormées des E_k. Soit les tenseurs :

$$\mathcal{A} = \sum_{i,k} a_{ik} \cdot e_1^i \otimes e_2^k$$
$$\mathcal{B} = \sum_{l,j} b_{lj} \cdot e_2^l \otimes e_3^j$$

où $a_{ij}, b_{ij} \in \mathbb{K}$. Calculons leur contraction maximale :

$$\alpha = \mathcal{A} : \mathcal{B} = \langle \mathcal{A} \odot \mathcal{B} \rangle_2 = \sum_{i,j,k,l} a_{ik} \cdot b_{lj} \cdot \boldsymbol{\delta}_{kl} \cdot \boldsymbol{\delta}_{ij}$$
$$\alpha = \sum_{i,k} a_{ik} \cdot b_{ki}$$

Ce résultat nous incite à définir l'opération équivalente sur les matrices associées $A = (a_{ij})_{i,j}$ et $B = (b_{ij})_{i,j}$:

$$A : B = \sum_{i,j} a_{ij} \cdot b_{ji}$$

62.26 Base canonique

Soit $\mathfrak{c}_{m,i}$ les vecteurs de la base canonique de $\mathbb{K}^m$ et $\mathfrak{c}_{n,i}$ les vecteurs de la base canonique de $\mathbb{K}^n$. Si $A = (a_{ij})_{i,j} \in \mathbb{M}(\mathbb{K}, m, n)$, on a clairement :

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \cdot \mathfrak{c}_{m,i} \otimes \mathfrak{c}_{n,j} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \cdot \mathfrak{c}_{m,i} \cdot \mathfrak{c}_{n,j}^\star$$

Lignes

On a :

$$\mathfrak{c}_{m,k}^\star \cdot A = \sum_{k,i,j} a_{ij} \cdot \boldsymbol{\delta}_{ik} \cdot \mathfrak{c}_{n,j}^\star = \sum_{j} a_{kj} \cdot \mathfrak{c}_{n,j}^\star$$

La matrice de taille $(1, n)$ obtenue est donc la k^{eme} ligne de A :

$$\mathfrak{c}_{m,k}^\star \cdot A = \underset{k}{\text{ligne}}\, A$$

Colonnes

On a :

$$A \cdot \mathfrak{c}_{n,k} = \sum_{k,i,j} a_{ij} \cdot \mathfrak{c}_{m,i} \cdot \boldsymbol{\delta}_{jk} = \sum_i a_{ik} \cdot \mathfrak{c}_{m,i}$$

La matrice de taille $(m, 1)$ obtenue est donc la k^{eme} colonne de A :

$$A \cdot \mathfrak{c}_{n,k} = \operatorname*{colonne}_k A$$

Identité

Dans le cas où $m = n$, on a en particulier :

$$\sum_{i=1}^n \mathfrak{c}_{n,i} \otimes \mathfrak{c}_{n,i} = \sum_{i=1}^n \mathfrak{c}_{n,i} \cdot \mathfrak{c}_{n,i}^\star = I$$

62.27 Normes matricielles

La norme de Frobenius d'une matrice $A = (a_{ij})_{i,j}$ est la norme du tenseur associé. On a donc :

$$\|A\|_F = \sqrt{A^\star : A} = \sqrt{\sum_{i,j} \bar{a}_{ij} \cdot a_{ij}} = \sqrt{\sum_{i,j} |a_{ij}|^2}$$

Trace

La trace d'une matrice $A = (a_{ij})_{i,j} \in \mathbb{M}(K, m, n)$ est la trace du tenseur sous-jacent, et donc :

$$\operatorname{tr}(A) = \sum_{i \in I} a_{ii}$$

où $I = \{1, 2, ..., \min\{m, n\}\}$.

Chapitre 63

Produit extérieur

Dépendances

- Chapitre 52 : Les vecteurs
- Chapitre 62 : Les tenseurs

63.1 Dimension 2

Soit l'espace vectoriel $E = \mathrm{span}\{e_1, e_2\}$ sur S, où (e_1, e_2) forme est une base orthonormée. Soit $u, v \in E$. On a :

$$u = \sum_{i=1}^{2} u_i e_i$$
$$v = \sum_{i=1}^{2} v_i e_i$$

pour certains $u_i, v_i \in S$.

Les symboles $\boldsymbol{\epsilon}_{ij} \quad (i, j = 1, 2)$ sont définis de telle sorte que pour tout u, v la double somme :

$$u \wedge v = \sum_{i,j=1}^{2} \boldsymbol{\epsilon}_{ij} u_i v_j$$

représente (au signe près) la surface du parallélogramme dont les sommets sont $(0, 0)$, (u_1, u_2), (v_1, v_2) et $(u_1 + v_1, u_2 + v_2)$. On impose de plus l'antisymétrie :

$$u \wedge v = -v \wedge u$$

afin de donner une orientation à ce parallélogramme. Ces contraintes nous donnent :

$$\begin{aligned}
u = e_1, \quad v = e_1 \quad &\Rightarrow \quad u \wedge v = 0 \\
u = e_1, \quad v = e_2 \quad &\Rightarrow \quad u \wedge v = 1 \\
u = e_2, \quad v = e_2 \quad &\Rightarrow \quad u \wedge v = 0 \\
u = e_2, \quad v = e_1 \quad &\Rightarrow \quad u \wedge v = -1
\end{aligned}$$

ce qui nous amène à :

$$\boldsymbol{\epsilon}_{11} = \boldsymbol{\epsilon}_{22} = 0$$
$$\boldsymbol{\epsilon}_{12} = 1 \qquad \boldsymbol{\epsilon}_{21} = -1$$

On voit qu'il y a un certain arbitraire dans notre choix, on aurait pu également choisir :

$$\boldsymbol{\epsilon}_{11} = \boldsymbol{\epsilon}_{22} = 0$$
$$\boldsymbol{\epsilon}_{12} = -1 \qquad \boldsymbol{\epsilon}_{21} = 1$$

Le choix du signe détermine ce que l'on appelle l'orientation de l'espace E.

63.2 Dimension 3

Soit l'espace vectoriel $E = \mathrm{span}\{e_1, e_2, e_3\}$ sur S, où (e_1, e_2, e_3) forme est une base orthonormée. Soit $u, v, w \in E$. On a :

$$u = \sum_{i=1}^{3} u_i e_i$$
$$v = \sum_{i=1}^{3} v_i e_i$$
$$w = \sum_{i=1}^{3} w_i e_i$$

pour certains $u_i, v_i, w_i \in S$.

Les symboles $\boldsymbol{\epsilon}_{ijk} \quad (i, j = 1, 2, 3)$ sont définis de telle sorte que pour tout u, v, w le scalaire :

$$u \wedge v \wedge w = \sum_{i,j,k=1}^{3} \boldsymbol{\epsilon}_{ijk} u_i v_j w_k$$

représente (au signe près) le volume du parallélipipède dont les côtés sont définis par u, v et w. On impose également l'antisymétrie

$$u \wedge v \wedge w = -u \wedge w \wedge v$$
$$u \wedge v \wedge w = -v \wedge u \wedge w$$

Par un procédé analogue au cas bidimensionnel, on montre qu'un choix possible est donné par :

$$\boldsymbol{\epsilon}_{123} = 1$$
$$\boldsymbol{\epsilon}_{ijj} = \boldsymbol{\epsilon}_{iij} = \boldsymbol{\epsilon}_{iji} = 0$$
$$\boldsymbol{\epsilon}_{ijk} = \boldsymbol{\epsilon}_{jki} = \boldsymbol{\epsilon}_{kij}$$
$$\boldsymbol{\epsilon}_{ijk} = -\boldsymbol{\epsilon}_{jik} \qquad \boldsymbol{\epsilon}_{ijk} = -\boldsymbol{\epsilon}_{ikj}$$

Produit vectoriel

On peut également former un vecteur avec l'opérateur $\wedge$. On pose simplement :

$$u \wedge v = \sum_{i,j,k=1}^{3} \left(\boldsymbol{\epsilon}_{ijk} u_j v_k\right) e_i$$

Les composantes sont donc données par :

$$(u \wedge v)_1 = u_2 v_3 - u_3 v_2$$
$$(u \wedge v)_2 = u_3 v_1 - u_1 v_3$$
$$(u \wedge v)_3 = u_1 v_2 - u_2 v_1$$

On peut relier le produit extérieur au produit scalaire en constatant que :

$$\langle u \mid v \wedge w \rangle = \sum_{i,j,k} \boldsymbol{\epsilon}_{ijk} u_i v_j w_k = u \wedge v \wedge w$$

63.3 Permutations en dimension N

Voyons quelles sont les propriétés communes aux $\boldsymbol{\epsilon}_*$ présentés ci-dessus. On remarque que $\boldsymbol{\epsilon}_{12} = \boldsymbol{\epsilon}_{123} = 1$. On note aussi que $\boldsymbol{\epsilon}_*$ est antysimétrique puisque l'inversion de deux indices change le signe. Ces propriétés nous permettent de généraliser la définition du produit extérieur à un espace de dimension N. On impose que le $\boldsymbol{\epsilon}$ à N indices vérifie les conditions suivantes :

$$\boldsymbol{\epsilon}_{1,2,...,N} = 1$$
$$\boldsymbol{\epsilon}_{i...j...k...l} = -\boldsymbol{\epsilon}_{i...k...j...l}$$

Ces conditions nous permettent de retrouver la valeur de n'importe quel $\boldsymbol{\epsilon}_{ijk...l}$. Si deux indices sont égaux, l'antisymétrie nous permet d'affirmer que :

$$\boldsymbol{\epsilon}_{i...j...j...k} = -\boldsymbol{\epsilon}_{i...j...j...k}$$

et donc :

$$\boldsymbol{\epsilon}_{i...j...j...k} = 0$$

Les seuls $\boldsymbol{\epsilon}$ non nuls sont donc ceux dont tous les indices $(i, j, k, ..., s)$ sont distincts, c'est-à-dire les permutations de $(1, 2, 3, ..., N)$. On se rend compte que si $p \in \mathbb{N}$ est le nombre de permutations de couples d'indices nécessaires pour obtenir $(i, j, k, ..., s)$ à partir de $(1, 2, 3, ..., N)$, on a :

$$\boldsymbol{\epsilon}_{ijk...l} = (-1)^p$$

63.4 Tenseur de permutation

Soit $(e_1, e_2, ..., e_N)$ une base orthonormée d'un espace vectoriel E sur S. Nous définissons le tenseur de permutation $\mathcal{E} \in \mathbb{T}_N(E)$ par :

$$\mathcal{E} = \sum_{i_1,i_2,...,i_N} \boldsymbol{\epsilon}_{i_1,i_2,...,i_N} \cdot e_{i_1} \otimes e_{i_2} \otimes ... \otimes e_{i_N}$$

63.5 Produit extérieur généralisé

Soit $(e_1, e_2, ..., e_N)$ une base orthonormée d'un espace vectoriel E sur $\mathbb{K}$, $M \leq N$ et les vecteurs $u_1, u_2, ...u_M \in E$ de coordonnées u_k^i :

$$u_k = \sum_{i=1}^{N} u_k^i e_i$$

Nous définissons leur produit extérieur par une contraction d'ordre M :

$$u_1 \wedge u_2 \wedge ... \wedge u_M = \langle \mathcal{E} \odot u_M \otimes ... \otimes u_1 \rangle_M$$

Par orthonormalité de la base, on a :

$$\langle e_j \mid u_k \rangle = u_k^j$$

Le produit extérieur s'écrit donc :

$$\mathcal{U} = u_1 \wedge u_2 \wedge ... \wedge u_M = \sum_{i_1,i_2,...i_N} \boldsymbol{\epsilon}_{i_1,i_2,...i_N} \cdot e_{i_1} \otimes ... \otimes e_{i_{N-M}} \cdot u_M^{i_N} \cdot ... \cdot u_1^{i_{N-M+1}}$$

ce qui nous donne les coordonnées du tenseur $\mathcal{U} \in \mathbb{T}_{N-M}(E)$ par rapport à la base $(e_1, ..., e_N)$:

$$U_{i_1,...,i_{N-M}} = \sum_{i_{N-M+1},...,i_N} \boldsymbol{\epsilon}_{i_1,i_2,...i_N} \cdot u_1^{i_{N-M+1}} \cdot ... \cdot u_M^{i_N}$$

On vérifie les propriétés suivantes :

$$u \wedge v = -v \wedge u$$
$$u \wedge u = 0$$
$$(\alpha u + \beta v) \wedge w = \alpha u \wedge w + \beta v \wedge w$$
$$w \wedge (\alpha u + \beta v) = \alpha w \wedge u + \beta w \wedge v$$

pour tout $u, v, w, ... \in E$ et $\alpha, \beta \in S$.

N vecteurs

Dans le cas où $M = N$, le produit extérieur est le scalaire :

$$\Delta = u_1 \wedge u_2 \wedge ... \wedge u_N = \sum_{i,j,...,k=1}^{N} \boldsymbol{\epsilon}_{ij...k} \cdot u_1^i \cdot u_2^j \cdot ... \cdot u_N^k$$

On appelle le Δ ainsi obtenu le déterminant des N vecteurs u_i, et on le note :

$$\det(u_1, ..., u_N) = u_1 \wedge u_2 \wedge ... \wedge u_N$$

63.6 Déterminant d'une matrice

Soit une matrice $A \in \mathbb{M}(K, N, N)$:

$$A = (a_{ij})_{i,j}$$

et les vecteurs correspondants :

$$a_i = \sum_j a_{ij} e_j$$

On définit alors simplement :

$$\det(A) = a_1 \wedge ... \wedge a_N = \sum_{i_1,i_2,...i_N} \boldsymbol{\epsilon}_{i_1,i_2,...i_N} a_{1,i_1} ... a_{N,i_N}$$

Chapitre 64

Matrices élémentaires

Dépendances

— Chapitre **??** : Les matrices
— Chapitre 62 : Les tenseurs

64.1 Introduction

Les matrices élémentaires constituent une classe importante de matrice. Elles permettent d'obtenir d'importants résultats utiles tant sur le plan théorique que pour les applications numériques. Une matrice élémentaire de taille (n, n) est déterminée par un scalaire $\alpha \in \mathbb{K}$ et le produit tensoriel de deux vecteurs $u, v \in \mathbb{K}^n$:

$$\mathrm{Elem}(\alpha, u, v) = I + \alpha \cdot u \otimes v = I + \alpha \cdot u \cdot v^\star$$

64.2 Inverse

Considérons le produit :

$$\begin{aligned}
\mathrm{Elem}(\alpha, u, v) \cdot \mathrm{Elem}(\beta, u, v) &= I + (\alpha + \beta) \cdot u \cdot v^\star + \alpha \cdot \beta \cdot u \cdot v^\star \cdot u \cdot v^\star \\
&= I + [\alpha + \beta + \alpha \cdot \beta \cdot (v^\star \cdot u)] \cdot u \cdot v^\star
\end{aligned}$$

On voit que si on peut trouver un scalaire β tel que :

$$\alpha + \beta + \alpha \cdot \beta \cdot (v^\star \cdot u) = 0$$

le produit des deux matrices sera égal à la matrice identité. Ce sera possible si $1 + \alpha \cdot v^\star \cdot u \neq 0$. On a alors :

$$\beta = -\frac{\alpha}{1 + \alpha \cdot (v^\star \cdot u)}$$

et :

$$\mathrm{Elem}(\alpha, u, v) \cdot \mathrm{Elem}(\beta, u, v) = I$$

Par symétrie, il est clair que le produit des deux matrices ne change pas lorsqu'on intervertit α et β. On a donc aussi :

$$\mathrm{Elem}(\beta, u, v) \cdot \mathrm{Elem}(\alpha, u, v) = I$$

Ces deux conditions étant remplies, on a :

$$\mathrm{Elem}(\beta, u, v) = \mathrm{Elem}(\alpha, u, v)^{-1}$$

Les matrices élémentaires sont donc très faciles à inverser.

64.3 Matrices élémentaires de transformation

Nous allons construire une matrice élémentaire qui transforme un vecteur donné $x \in \mathbb{M}(\mathbb{K}, n, 1)$ non nul en un autre vecteur donné $y \in \mathbb{M}(\mathbb{K}, n, 1)$ de même taille.

Colonne

On cherche une matrice élémentaire E_{yx} telle que $E_{yx} \cdot x = y$. L'équation :

$$(I + \alpha \cdot u \cdot v^\star) \cdot x = x + \alpha \cdot u \cdot (v^\star \cdot x) = y$$

nous donne la condition :

$$\alpha \cdot (v^\star \cdot x) \cdot u = y - x$$

On peut donc choisir par exemple $u = y - x$ et $v = x$. On a alors $\alpha = 1/(x^\star \cdot x) \neq 0$ et on se retrouve avec la matrice élémentaire :

$$E_{yx} = I + \frac{1}{x^\star \cdot x} \cdot (y - x) \cdot x^\star$$

Inverse

Si l'inverse existe, il s'agit d'une matrice élémentaire de paramètre scalaire :

$$\beta = -\frac{1}{x^\star \cdot x + x^\star \cdot (y - x)} = -\frac{1}{x^\star \cdot y}$$

Sous réserve que $x^\star \cdot y \neq 0$, on a donc :

$$E_{yx}^{-1} = I - \frac{1}{x^\star \cdot y} \cdot (y - x) \cdot x^\star$$

Ligne

On cherche une matrice élémentaire E_{yx} telle que $x^\star \cdot E_{yx} = y^\star$. L'équation :

$$x^\star \cdot (I + \alpha \cdot u \cdot v^\star) = x^\star + \alpha \cdot (x^\star \cdot u) \cdot v^\star = y^\star$$

nous donne la condition :

$$\alpha \cdot (x^\star \cdot u) \cdot v^\star = y^\star - x^\star$$

ou :

$$\overline{\alpha} \cdot (u^\star \cdot x) \cdot v = y - x$$

On peut donc choisir par exemple $v = y - x$ et $u = x$. On a alors $\alpha = \overline{\alpha} = 1/(x^\star \cdot x) \neq 0$ et on se retrouve avec la matrice élémentaire :

$$E_{yx} = I + \frac{1}{x^\star \cdot x} \cdot x \cdot (y - x)^\star$$

Inverse

Si l'inverse existe, il s'agit d'une matrice élémentaire de paramètre scalaire :

$$\beta = -\frac{1}{x^\star \cdot x + (y - x)^\star \cdot x} = -\frac{1}{y^\star \cdot x}$$

Sous réserve que $y^\star \cdot x \neq 0$, on a donc :

$$E_{yx}^{-1} = I - \frac{1}{y^\star \cdot x} \cdot x \cdot (y - x)^\star$$

64.4 Matrices élémentaires de permutation

Les matrices élémentaires de permutations permettent de permuter deux lignes ou deux colonnes d'une matrice. Soit $\mathfrak{c}_1, ..., \mathfrak{c}_n$ les vecteurs de la base canonique de $\mathbb{K}^n$. La matrice de permutation élémentaire de taille (n, n) et de paramètres i, j est définie par :

$$\underset{n,i,j}{\text{permut}} = I - (\mathfrak{c}_i - \mathfrak{c}_j) \cdot (\mathfrak{c}_i - \mathfrak{c}_j)^\star$$

Dans la suite, nous considérons $A \in \mathbb{M}(\mathbb{K}, m, n)$ et $P = \text{permut}_{n,i,j}$.

Permutation des colonnes

Soit les colonnes $C_i = A \cdot \mathfrak{c}_i$. On a :

$$A \cdot P = A - (C_i \cdot \mathfrak{c}_i^\star + C_j \cdot \mathfrak{c}_j^\star) + (C_j \cdot \mathfrak{c}_i^\star + C_i \cdot \mathfrak{c}_j^\star)$$

Les colonnes i et j de A sont donc permutées par multiplication à droite d'une matrice de permutation.

Permutation des lignes

Soit les lignes $L_i = \mathfrak{c}_i^\star \cdot A$. On a :

$$P \cdot A = A - (\mathfrak{c}_i \cdot L_i + \mathfrak{c}_j \cdot L_j) + (\mathfrak{c}_j \cdot L_i + \mathfrak{c}_i \cdot L_j)$$

Les lignes i et j de A sont donc permutées par multiplication à gauche d'une matrice de permutation.

Symétrie

On constate que la transposée et la duale sont égales à la matrice elle-même :

$$\underset{n,i,j}{\text{permut}} = \underset{n,i,j}{\overset{T}{\text{permut}}} = \underset{n,i,j}{\overset{\star}{\text{permut}}}$$

Inverse

Soit $\Delta_{ij} = \mathfrak{c}_i - \mathfrak{c}_j$ et :

$$P = \underset{n,i,j}{\text{permut}} = I - \Delta_{ij} \cdot \Delta_{ij}^\star$$

Le produit de cette matrice avec elle-même s'écrit :

$$P \cdot P = I - 2\Delta_{ij} \cdot \Delta_{ij}^\star + \Delta_{ij} \cdot \Delta_{ij}^\star \cdot \Delta_{ij} \cdot \Delta_{ij}^\star$$

Mais comme $\Delta_{ij}^\star \cdot \Delta_{ij} = 2$, on a :

$$P \cdot P = I - 2\Delta_{ij} \cdot \Delta_{ij}^\star + 2\Delta_{ij} \cdot \Delta_{ij}^\star$$

Les deux derniers termes s'annihilent et :

$$P \cdot P = I$$

Les matrices de permutations élémentaires sont donc égales à leur propre inverse :

$$\operatorname*{permut}_{n,i,j} = \operatorname*{permut}_{n,i,j}^{-1}$$

64.5 Matrices de permutations

Une matrice de permutation P est une matrice de la forme :

$$P = P_1 \cdot ... \cdot P_n$$

où les P_i sont des matrices élémentaires de permutation.

Inverse

$$P^\star \cdot P = P_n \cdot ... \cdot P_1 \cdot P_1 \cdot ... \cdot P_n = I$$
$$P \cdot P^\star = P_1 \cdot ... \cdot P_n \cdot P_n \cdot ... \cdot P_1 = I$$

Donc $P^\star = P^{-1}$. Comme $P^\star = P^T$, la transposée d'une matrice de permutation est identique à son inverse. On dit que ces matrices sont orthogonales.

Chapitre 65

Systèmes linéaires et inverses

65.1 Gauss-Jordan

Nous allons tenter de diagonaliser une matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$ en une matrice diagonale aussi proche que possible de la matrice identité. Soit les $\mathfrak{c}_i$, vecteurs de la base canonique de $\mathbb{K}^m$ ou $\mathbb{K}^n$ suivant le contexte. Si A est non nulle, on peut trouver un i et un j tels que :

$$a = \operatorname*{comp}_{ij} A \neq 0$$

On considère les matrices élémentaires de permutation $P_0 = \mathrm{permut}_{n,i,1}$ et $Q_0 = \mathrm{permut}_{n,j,1}$, et on évalue $P_0 \cdot A \cdot Q_0$ pour inverser les lignes 1 et i et les colonnes 1 et j de A. On se retrouve alors avec une matrice de la forme :

$$Q_0 \cdot A \cdot P_0 = [x \ c_2 \ ... \ c_n]$$

où x est la première colonne et où $x^\star \cdot \mathfrak{c}_1 = a \neq 0$. On utilise ensuite une première matrice élémentaire :

$$E_0 = I + \frac{1}{x^\star \cdot x} \cdot (\mathfrak{c}_1 - x) \cdot x^\star$$

pour transformer la première colonne x en $y = \mathfrak{c}_1$. Partitionnant à part la première ligne et la première colonne du résultat, on a :

$$E_0 \cdot P_0 \cdot A \cdot Q_0 = \begin{bmatrix} 1 & z^\star \\ 0 & \ddots \end{bmatrix}$$

Posons $u^\star = [1 \ z^\star]$ pour la première ligne. On a $\mathfrak{c}_1^\star \cdot u = 1 \neq 0$. On utilise une seconde matrice élémentaire :

$$F_0 = I + \frac{1}{u^\star \cdot u} \cdot u \cdot (\mathfrak{c}_1 - u)^\star$$

pour transformer la première ligne $u^\star$ en $v^\star = \mathfrak{c}_1^\star$. Mais comme $(\mathfrak{c}_1 - u)^\star = [0 \ -z^\star]$, on a :

$$u \cdot (\mathfrak{c}_1 - u)^\star = \begin{bmatrix} 1 \\ z \end{bmatrix} \cdot \begin{bmatrix} 0 & -z^\star \end{bmatrix} = \begin{bmatrix} 0 & -z^\star \\ 0 & -z \cdot z^\star \end{bmatrix}$$

On voit aussi que $u^\star \cdot u = 1 + z^\star \cdot z$. La matrice élémentaire F_0 est donc de la forme :

$$F_0 = I + \frac{1}{1 + z^\star \cdot z} \cdot \begin{bmatrix} 0 & -z^\star \\ 0 & -z \cdot z^\star \end{bmatrix}$$

On obtient alors une matrice modifiée de la forme :

$$E_0 \cdot P_0 \cdot A \cdot Q_0 \cdot F_0 = \begin{bmatrix} 1 & z^\star \\ 0 & \ddots \end{bmatrix} + \frac{1}{1 + z^\star \cdot z} \cdot \begin{bmatrix} 1 & z^\star \\ 0 & \ddots \end{bmatrix} \cdot \begin{bmatrix} 0 & -z^\star \\ 0 & -z \cdot z^\star \end{bmatrix}$$

et finalement :

$$E_0 \cdot P_0 \cdot A \cdot Q_0 \cdot F_0 = \begin{bmatrix} 1 & 0 \\ 0 & A^{(n-1)} \end{bmatrix}$$

Recommençons le même procédé pour transformer, au moyen des matrices de permutation $P^{(n-1)}, Q^{(n-1)}$ et élémentaires $E^{(n-1)}, F^{(n-1)}$, la première colonne et la première ligne de $A^{(n-1)}$ en e_1 et $e_1^\star$. Si on pose :

$$\begin{aligned} P_1 &= \begin{bmatrix} 1 & 0 \\ 0 & P^{(n-1)} \end{bmatrix} = \begin{bmatrix} I_1 & 0 \\ 0 & P^{(n-1)} \end{bmatrix} \\ Q_1 &= \begin{bmatrix} 1 & 0 \\ 0 & Q^{(n-1)} \end{bmatrix} = \begin{bmatrix} I_1 & 0 \\ 0 & Q^{(n-1)} \end{bmatrix} \\ E_1 &= \begin{bmatrix} 1 & 0 \\ 0 & E^{(n-1)} \end{bmatrix} = \begin{bmatrix} I_1 & 0 \\ 0 & E^{(n-1)} \end{bmatrix} \\ F_1 &= \begin{bmatrix} 1 & 0 \\ 0 & F^{(n-1)} \end{bmatrix} = \begin{bmatrix} I_1 & 0 \\ 0 & F^{(n-1)} \end{bmatrix} \end{aligned}$$

il vient :

$$E_1 \cdot P_1 \cdot E_0 \cdot P_0 \cdot A \cdot Q_0 \cdot F_0 \cdot Q_1 \cdot F_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & A^{(n-2)} \end{bmatrix}$$

Soit $p = \min\{m, n\}$. On peut répéter le même processus r fois en utilisant à l'étape k :

$$\begin{aligned} P_k &= \begin{bmatrix} I_k & 0 \\ 0 & P^{(n-k)} \end{bmatrix} \\ Q_k &= \begin{bmatrix} I_k & 0 \\ 0 & Q^{(n-k)} \end{bmatrix} \\ E_k &= \begin{bmatrix} I_k & 0 \\ 0 & E^{(n-k)} \end{bmatrix} \\ F_k &= \begin{bmatrix} I_k & 0 \\ 0 & F^{(n-k)} \end{bmatrix} \end{aligned}$$

jusqu'à ce que la matrice $A^{(n-r)}$ soit nulle, ou jusqu'à ce qu'on ait atteint $r = p$. Posons :

$$\begin{aligned} E &= E_r \cdot P_r \cdot \ldots \cdot E_0 \cdot P_0 \\ F &= Q_0 \cdot F_0 \cdot \ldots \cdot Q_r \cdot F_r \end{aligned}$$

On a alors schématiquement :

$$E \cdot A \cdot F = C$$

où :

$$C \in \left\{ I_p, \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \begin{bmatrix} I_m & 0 \end{bmatrix}, \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \right\}$$

suivant que $r = p = m = n$, $r = p = n$, $r = p = m$ ou $r < p$.

Inverse

Les matrices élémentaires de permutation sont inversibles, d'inverse identique à elles-mêmes. On a donc $P_i^{-1} = P_i$ et $Q_i^{-1} = Q_i$. Les matrices élémentaires de transformation sont également inversibles. En effet, $x^\star \cdot y = x^\star \cdot \mathfrak{c}_1 = a \ne 0$. L'inverse de E_0 existe donc et s'écrit :

$$E_0^{-1} = I - \frac{1}{a} \cdot (\mathfrak{c}_1 - x) \cdot x^\star$$

D'un autre coté $v^\star \cdot u = \mathfrak{c}_1^\star \cdot u = 1 \ne 0$. L'inverse de F_0 existe aussi et s'écrit :

$$F_0^{-1} = I - u \cdot (\mathfrak{c}_1 - u)^\star$$

Comme on procède de même à chaque étape, l'inverse de chaque matrice élémentaire existe. En appliquant la formule d'inversion d'un produit, on obtient :

$$\begin{aligned} E^{-1} &= P_0 \cdot E_0^{-1} \cdot ... \cdot P_r \cdot E_r^{-1} \\ F^{-1} &= F_r^{-1} \cdot Q_r \cdot ... \cdot F_0^{-1} \cdot Q_0 \end{aligned}$$

Posons $L = E^{-1}$ et $R = F^{-1}$. En multipliant l'équation $E \cdot A \cdot F = C$ à gauche par L et à droite par R, on obtient :

$$A = L \cdot C \cdot R$$

Une telle décomposition est appelée décomposition de Gauss-Jordan et notée :

$$(L, C, R) = \mathrm{GaussJordan}(A)$$

Rang

Le r ainsi obtenu est appelé rang de la matrice A. On le note :

$$r = \operatorname{rang} A$$

On a par construction $r \le p$.

65.2 Systèmes linéaires

Nous allons à présent utiliser la décomposition de Gauss-Jordan pour analyser les espaces de solutions :

$$S(y) = \{x \in \mathbb{K}^n : A \cdot x = y\}$$

pour tout $y \in \mathbb{K}^m$. On a :

$$A \cdot x = L \cdot C \cdot R \cdot x = y$$

Multiplions cette équation par L^{-1}. Il vient :

$$C \cdot R \cdot x = L^{-1} \cdot y$$

On pose :

$$\begin{aligned} z &= R \cdot x \\ b &= L^{-1} \cdot y \end{aligned}$$

Le système linéaire s'écrit alors :

$$C \cdot z = b$$

Plein rang, matrice carrée

Si $r = m = n$, on a :

$$A = L \cdot I \cdot R = L \cdot R$$

La matrice A est un produit de matrices inversibles. Elle est donc inversible et :

$$A^{-1} = R^{-1} \cdot L^{-1}$$

L'équation $A \cdot x = y$ admet donc pour unique solution $x \in S(y)$ le vecteur :

$$x = A^{-1} \cdot y = R^{-1} \cdot L^{-1} \cdot y$$

Plein rang, matrice haute

Si $r = n < m$, on a :

$$C = \begin{bmatrix} I_n \\ 0 \end{bmatrix}$$

Si on partitionne $b = L^{-1} \cdot y$ en deux vecteurs b_1, b_2 de tailles $(n, 1)$ et $(m - n, 1)$, on a :

$$\begin{bmatrix} I_n \\ 0 \end{bmatrix} \cdot z = \begin{bmatrix} z \\ 0 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Il y a donc deux conditions pour que $x = R^{-1} \cdot z$ soit dans $S(y)$:

$$z = b_1$$
$$0 = b_2$$

Si $b_2 \neq 0$, il n'existe pas de solution. Si $b_2 = 0$, il existe une unique solution $x \in S(y)$, qui s'écrit :

$$x = R^{-1} \cdot z = R^{-1} \cdot b_1$$

Remarquons que l'on peut toujours trouver un y tel qu'il existe au moins une solution. En effet, il suffit de choisir :

$$y = L \cdot \begin{bmatrix} b_1 \\ 0 \end{bmatrix}$$

D'un autre coté, il existe toujours un y tel qu'il n'existe pas de solution. En effet, il suffit de choisir :

$$y = L \cdot \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

où $b_2 \neq 0$.

Plein rang, matrice longue

Si $r = m < n$, on se retrouve alors avec :

$$C = \begin{bmatrix} I_m & 0 \end{bmatrix}$$

Si on partitionne z en deux vecteurs z_1, z_2 de tailles $(m, 1)$ et $(n - m, 1)$, on a :

$$C \cdot z = \begin{bmatrix} I_m & 0 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = I_m \cdot z_1 + 0 \cdot z_2 = z_1$$

La condition pour que $x = R^{-1} \cdot z$ soit dans $S(y)$ se résume à :

$$z_1 = b = L^{-1} \cdot y$$

Nous n'avons par contre aucune condition sur $z_2 \in \mathbb{M}(\mathbb{K}, n - m)$. Il y a donc une infinité de solutions $x \in S(y)$, de la forme :

$$x = R^{-1} \cdot z = R^{-1} \cdot \begin{bmatrix} L^{-1} \cdot y \\ z_2 \end{bmatrix}$$

Rang incomplet

Supposons que $r < p$. On partitionne alors z en deux vecteurs z_1 et z_2 de tailles $(r, 1)$ et $(n-r, 1)$ et b en deux vecteurs b_1 et b_2 de tailles $(r, 1)$ et $(m-r, 1)$. Avec ces notations, le produit $C \cdot z$ s'écrit :

$$C \cdot z = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ 0 \end{bmatrix}$$

L'équation $C \cdot z = b$ prend donc la forme :

$$\begin{bmatrix} z_1 \\ 0 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Les deux conditions pour que $x = R^{-1} \cdot z$ soit dans $S(y)$ sont donc que $z_1 = b_1$ et que $b_2 = 0$. Il n'y a aucune condition sur z_2. Si $b_2 \neq 0$ il n'y a pas de solution. Si $b_2 = 0$, il existe une infinité de solutions $x \in S(y)$ de la forme :

$$x = R^{-1} \cdot \begin{bmatrix} b_1 \\ z_2 \end{bmatrix}$$

Remarquons que l'on peut toujours trouver un y tel qu'il existe au moins une solution. En effet, il suffit de choisir :

$$y = L \cdot \begin{bmatrix} b_1 \\ 0 \end{bmatrix}$$

Un choix particulier est par exemple $y = b = 0$.

D'un autre coté, il existe toujours un y tel qu'il n'existe pas de solution. En effet, il suffit de choisir :

$$y = L \cdot \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

où $b_2 \neq 0$.

Rang et existence

On conclut de ce qui précède que si $r < m$ ou $r < p$, on peut toujours trouver un y tel qu'il n'existe pas de solution. Pour qu'il existe au moins une solution du système quel que soit y, il faut donc avoir $r = m = p$.

Rang et unicité

On conclut de ce qui précède que si $r < n$ ou $r < p$, on peut toujours trouver un y tel qu'il existe une infinité de solution, et même une infinité de solutions non nulles puisque $z_2 \neq 0$ peut être quelconque.

Pour qu'il existe au maximum une solution du système quel que soit y, il faut donc avoir $r = n = p$.

65.3 Matrices hautes

Soit une matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$ strictement haute ($m > n$). Un inverse à droite R vérifiant $A \cdot B = I$ ne peut pas exister, car sinon il suffirait de prendre $x = B \cdot y$ pour avoir :

$$A \cdot x = A \cdot B \cdot y = I \cdot y = y$$

Le système $A \cdot x = y$ admettrait toujours au moins une solution, ce qui contredit les résultats précédents.

65.4 Matrices longues

Soit une matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$ strictement longue ($n > m$). Un inverse à gauche B vérifiant $B \cdot A = I$ ne peut pas exister, car sinon on aurait :

$$(B \cdot A)^\star = A^\star \cdot B^\star = I^\star = I$$

La matrice $B^\star$ serait donc un inverse à droite de la matrice $A^\star$. Or, $A^\star$ est de taille (n, m), donc strictement haute, et ne peut pas admettre d'inverse à droite. On en conclut qu'une matrice strictement longue ne peut pas admettre d'inverse à gauche.

65.5 Matrices carrées

Supposons que $m = n$ et considérons deux matrices carrées $A, B \in \mathbb{M}(\mathbb{K}, n, n)$ telles que :

$$A \cdot B = I$$

Il existe au moins une solution du système $A \cdot x = y$ quel que soit y, car il suffit de prendre $x = B \cdot y$ pour avoir $A \cdot x = A \cdot B \cdot y = y$. On a donc forcément $r = m \le n$. Mais comme $m = n$, cette solution est unique. On en déduit que l'application linéaire associée à A est inversible, donc A aussi et :

$$A \cdot (B - A^{-1}) = I - I = 0$$

En multipliant à gauche par A^{-1}, il vient simplement $B - A^{-1} = 0$, c'est-à-dire $B = A^{-1}$. On a donc également :

$$B \cdot A = A^{-1} \cdot A = I$$

65.6 Complément de Schur

Soit le système suivant :

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} F \\ G \end{bmatrix}$$

où A, B, C, D, F, G et x, y sont respectivement des matrices et des vecteurs matriciels de tailles compatibles. On a :

$$A \cdot x + B \cdot y = F$$
$$C \cdot x + D \cdot y = G$$

Si A est carrée et inversible, la première relation nous permet d'éliminer x :

$$x = A^{-1} \cdot F - A^{-1} \cdot B \cdot y$$

On substitue alors dans la seconde relation, et on obtient :

$$C \cdot A^{-1} \cdot F - C \cdot A^{-1} \cdot B \cdot y + D \cdot y = G$$

En plaçant y en évidence, on obtient :

$$(D - C \cdot A^{-1} \cdot B) \cdot y = G - C \cdot A^{-1} \cdot F$$

Sous réserve d'inversibilité, il ne nous reste alors plus qu'à résoudre par rapport à y :

$$y = (D - C \cdot A^{-1} \cdot B)^{-1} \cdot (G - C \cdot A^{-1} \cdot F)$$

qui nous donne ensuite :

$$x = A^{-1} \cdot [F - B \cdot (D - C \cdot A^{-1} \cdot B)^{-1} \cdot (G - C \cdot A^{-1} \cdot F)]$$

Si A est facilement inversible, et si $D - C \cdot A^{-1} \cdot B$ est de taille réduite comparativement à la taille de A, il peut être avantageux de résoudre le système global en (x, y) de cette façon.

65.7 Dimension

Soit un espace vectoriel E sur $\mathbb{K}$ possédant deux bases $(e_1, e_2, ..., e_n)$ et $(f_1, f_2, ..., f_m)$. Comme les $f_i \in E$, on a :

$$f_i = \sum_{k=1}^{n} a_{ik} \cdot e_k$$

pour certains $a_{ij} \in \mathbb{K}$. Comme les $e_k \in E$, on a également :

$$e_k = \sum_{j=1}^{m} b_{kj} \cdot f_j$$

pour certains $b_{kj} \in \mathbb{K}$. On en conclut que :

$$f_i = \sum_{j=1}^{m} \sum_{k=1}^{n} a_{ik} \cdot b_{kj} \cdot f_j$$

Par indépendance linéaire des f_i, on doit donc avoir :

$$\sum_{k=1}^{m} a_{ik} \cdot b_{kj} = \boldsymbol{\delta}_{ij}$$

Si nous introduisons les matrices $A \in \mathbb{M}(\mathbb{K}, m, n)$ et $B \in \mathbb{M}(\mathbb{K}, n, m)$ de composantes a_{ij} et b_{ij} respectivement, on a donc $A \cdot B = I_m$. La matrice A admettant un inverse à droite, elle ne peut pas être strictement haute et on a $m \le n$. Mais on a aussi :

$$e_k = \sum_{k=1}^{n} \sum_{j=1}^{m} b_{kj} \cdot a_{ji} \cdot e_i$$

d'où l'on conclut que $B \cdot A = I_n$. La matrice A admettant un inverse à gauche, elle ne peut pas être strictement longue et on a également $n \le m$. On conclut de ces deux inégalités que $m = n$. Si on possède une base de E comptant n vecteurs, on peut-être sûr que toute autre base de E comptera également n vecteurs. On dit que $m = n$ est la dimension de E et on le note :

$$\dim E = n$$

Corollaire

La base canonique de $\mathbb{K}^n$ comptant n éléments, on en déduit que $\mathbb{K}^n$ est de dimension n.

65.8 Indépendance linéaire

Soit un espace vectoriel E de dimension m et une suite de vecteurs linéairement indépendants $(u_1, ..., u_n)$ de E. Si $(e_1, ..., e_m)$ est une base de E, on peut trouver des coordonnées $a_{ki} \in \mathbb{K}$ telles que :

$$u_i = \sum_{k=1}^{m} a_{ki} \cdot e_k$$

Pour tout $w \in E$, on sait aussi que l'on peut trouver des coordonnées $y_k \in \mathbb{K}$ telles que :

$$w = \sum_{k=1}^{m} y_k \cdot e_k$$

Nous allons à présent examiner si on peut également trouver des coordonnées $x_i \in \mathbb{K}$ de w par rapport aux u_i :

$$w = \sum_{i=1}^{n} x_i \cdot u_i$$

Si cette hypothèse est vérifiée, on doit avoir :

$$\sum_{k=1}^{m} y_k \cdot e_k = w = \sum_{k=1}^{m} \sum_{i=1}^{n} a_{ki} \cdot x_i \cdot e_k$$

L'indépendance linéaire des e_k implique alors que :

$$y_k = \sum_{i=1}^{n} a_{ki} \cdot x_i$$

Utilisant les vecteurs matriciels associés $x = [x_1 \ ... \ x_n]^\star$ et $y = [y_1 \ ... \ y_m]^\star$ ainsi que la matrice $A = (a_{ki})_{k,i}$ de taille (m, n), on se retrouve avec le système :

$$y = A \cdot x$$

Soit r le rang de A. On sait déjà que $r \leq \min\{m, n\}$. Examinons les différents cas :

— Si on avait $r < n$, on pourrait trouver un y tel qu'il existe une infinité de solution en x, ce qui contredit l'unicité des coordonnées de w par rapport à la suite de vecteurs linéairement indépendant $(u_1, ..., u_n)$. On doit donc avoir $r = n \leq m$.

— Dans le cas où $r = n < m$, on peut trouver un y tel que la solution n'existe pas : la suite de u_i ne forme donc pas une base de E.

— Dans le cas où $r = n = m$, il existe toujours une unique solution, la suite des u_i forme alors une base de E. Il suffit donc de trouver une suite de m vecteurs indépendants dans un espace de dimension m pour former une base de cet espace.

Chapitre 66

Matrices unitaires

66.1 Conservation du produit scalaire

Les matrices unitaires sont une généralisation des rotations. Or, la propriété essentielle d'une rotation est qu'elle conserve les distances. Comme les distances usuelles découlent des normes usuelles, elles-mêmes dérivées du produit scalaire, on impose que ce soit le produit scalaire qui soit conservé. On veut donc que :

$$\langle U \cdot x \mid U \cdot y \rangle = x^\star \cdot U^\star \cdot U \cdot y = \langle x \mid y \rangle = x^\star \cdot y$$

pour toute matrice unitaire U de taille (m, n) et tout $x, y \in \mathbb{K}^n$. Comme cette relation doit être valable pour tout x, y, elle doit l'être pour les vecteurs de la base canonique :

$$\operatorname*{comp}_{ij} U = \mathfrak{c}_i^\star \cdot U^\star \cdot U \cdot \mathfrak{c}_j = \mathfrak{c}_i^\star \cdot \mathfrak{c}_j = \boldsymbol{\delta}_{ij}$$

On en conclut que :

$$U^\star \cdot U = I$$

Si cette condition est vérifiée, on dit que U est une matrice unitaire.

Norme

La conservation de la norme usuelle $\|x\| = \sqrt{\langle x \mid x \rangle}$ découle de celle du produit scalaire. Si U est unitaire, on a donc $\|U \cdot x\| = \|x\|$.

66.2 Matrices élémentaires unitaires

Soit la matrice élémentaire :

$$U = \operatorname{Elem}(\alpha, u, u) = I + \alpha \cdot u \cdot u^\star$$

où $\alpha \in \mathbb{R}$. Si on veut que U soit unitaire, il faut que :

$$\begin{aligned} I = U^\star \cdot U &= (I + \alpha \cdot u \cdot u^\star) \cdot (I + \alpha \cdot u \cdot u^\star) \\ &= I + [2 \cdot \alpha + \alpha^2 \cdot (u^\star \cdot u)] \cdot u \cdot u^\star \end{aligned}$$

Il suffit donc que α vérifie la condition :

$$2 \cdot \alpha + \alpha^2 \cdot (u^\star \cdot u) = \alpha \cdot [2 + \alpha \cdot (u^\star \cdot u)] = 0$$

Si on choisit $\alpha = 0$, on a $U = I$ qui est bien une matrice unitaire. Dans le cas où $\alpha \neq 0$, on doit avoir :

$$2 + \alpha \cdot (u^\star \cdot u) = 0$$

Si u est un vecteur non nul, on aura bien sûr $u^\star \cdot u > 0$. Il suffit alors de prendre :

$$\alpha = -\frac{2}{u^\star \cdot u}$$

On se retrouve donc avec des matrices élémentaires unitaires du type :

$$U = I - \frac{2}{u^\star \cdot u} \cdot u \cdot u^\star$$

On les note :

$$\text{Unitaire}(u) = I - \frac{2}{u^\star \cdot u} \cdot u \cdot u^\star$$

66.2.1 Propriétés

On a donc $U^\star \cdot U = I$. Comme U est carrée, on a également $U^{-1} = U^\star$. On voit également que :

$$U^\star = I - \frac{2}{u^\star \cdot u} \cdot u \cdot u^\star = U$$

On en conclut que $U^{-1} = U$.

66.3 Matrices de Householder

Soit deux vecteurs $x, y \in \mathbb{K}^n$. On aimerait bien trouver une matrice élémentaire unitaire U telle que $U \cdot x \approx y$. Par analogie avec les matrices de transformation élémentaire, on pose $u = x - y$ et $U = \text{Unitaire}(u)$. On a alors :

$$\begin{aligned}
U \cdot x &= x - \frac{2u^\star \cdot x}{u^\star \cdot u} \cdot u \\
&= \frac{(u^\star \cdot u^\star) \cdot x - 2(u^\star \cdot x) \cdot x + 2(u^\star \cdot x) \cdot y}{u^\star \cdot u^\star} \\
&= \frac{(u^\star \cdot u^\star - 2u^\star \cdot x) \cdot x + 2(u^\star \cdot x) \cdot y}{u^\star \cdot u^\star}
\end{aligned}$$

Développons le coefficient de x :

$$u^\star \cdot u^\star - 2u^\star \cdot x = (x^\star \cdot x - x^\star \cdot y - y^\star \cdot x + y^\star \cdot y) - (2x^\star \cdot x - 2y^\star \cdot x)$$

Comme il s'agit d'une transformation unitaire, il est logique de demander que le produit scalaire soit conservé. On a donc $x^\star \cdot x = y^\star \cdot y$. Si on impose de plus que $x^\star \cdot y$ soit réel, on a :

$$x^\star \cdot y = \text{conj}(x^\star \cdot y) = y^\star \cdot x$$

et :

$$u^\star \cdot u^\star - 2u^\star \cdot x = (2x^\star \cdot x - 2y^\star \cdot x) - (2x^\star \cdot x - 2y^\star \cdot x) = 0$$

On a alors :

$$U \cdot x = \gamma \cdot y$$

où :

$$\gamma = \frac{2u^\star \cdot x}{u^\star \cdot u^\star} \in \mathbb{C}$$

Le vecteur $U \cdot x$ est donc identique à y à un scalaire près.

Base canonique

Un cas particulier intéressant est celui où y est proportionnel au i^{eme} vecteur de la base canonique :

$$y = \alpha \cdot \mathfrak{c}_i$$

avec $\alpha \in \mathbb{C}$. Si $x_i = \operatorname{comp}_i x$, on a alors :

$$y^\star \cdot x = \overline{\alpha} \cdot x_i$$

— Si $x_i = 0$, on a $y^\star \cdot x = 0 \in \mathbb{R}$. Il suffit alors de choisir α pour que le produit scalaire soit conservé :

$$y^\star \cdot y = |\alpha|^2 = x^\star \cdot x$$

On peut donc prendre :

$$\alpha = \sqrt{x^\star \cdot x} \in \mathbb{R}$$

— Considérons à présent le cas où $x_i \neq 0$. Si on prend $\alpha = \lambda \cdot x_i$ où $\lambda \in \mathbb{R}$ est quelconque, on a :

$$y^\star \cdot x = \lambda \cdot \overline{x}_i \cdot x_i = \lambda \cdot |x_i|^2 \in \mathbb{R}$$

Comme on exige que la norme soit conservée, il faut de plus que :

$$y^\star \cdot y = \lambda^2 \cdot |x_i|^2 = x^\star \cdot x$$

On doit donc avoir $\lambda^2 = x^\star \cdot x / |x_i|^2$ et :

$$\alpha = \lambda \cdot x_i = \sqrt{\frac{x^\star \cdot x}{|x_i|^2}} \cdot x_i$$

Il ne nous reste alors plus qu'à poser $u = x - \alpha \cdot \mathfrak{c}_i$ et $U = \text{Unitaire}(u)$ pour obtenir :

$$U \cdot x = \gamma \cdot \mathfrak{c}_i$$

pour un certain $\gamma \in \mathbb{C}$.

66.4 Décomposition de Householder

Soit une matrice A de taille (m, n) et le partitionnement en colonnes :

$$A = [x \ x_2 \ ... \ x_n]$$

Soit H_1 la matrice de Householder qui permet de transformer x en $\gamma \cdot \mathfrak{c}_1$, pour un certain $\gamma_1 \in \mathbb{C}$. On a alors :

$$H_1 \cdot A = \begin{bmatrix} \gamma_1 & \cdots \\ 0 & A^{(n-1)} \end{bmatrix}$$

On peut répéter le même processus avec $A^{(n-1)}$ et la matrice de Householder $H^{(n-1)}$ correspondante. Si on pose alors :

$$H_2 = \begin{bmatrix} 1 & 0 \\ 0 & H^{(n-1)} \end{bmatrix} = \begin{bmatrix} I_1 & 0 \\ 0 & H^{(n-1)} \end{bmatrix}$$

il vient :

$$H_2 \cdot H_1 \cdot A = \begin{bmatrix} \gamma_1 & \dots & \dots \\ 0 & \gamma_2 & \dots \\ 0 & 0 & H^{(n-2)} \end{bmatrix}$$

On peut répéter le processus $p = \min\{m, n\}$ fois en utilisant à l'étape $k+1$:

$$H_{k+1} = \begin{bmatrix} I_k & 0 \\ 0 & H^{(n-k)} \end{bmatrix}$$

Posons :

$$H = H_p \cdot ... \cdot H_1$$

Si $m < n$, on obtient à la fin du processus :

$$H \cdot A = R = \begin{bmatrix} \gamma_1 & \dots & \dots & \dots \\ 0 & \ddots & \dots & \dots \\ 0 & 0 & \gamma_m & \dots \end{bmatrix}$$

Si $m = n$, on obtient à la fin du processus :

$$H \cdot A = R = \begin{bmatrix} \gamma_1 & \dots & \dots \\ 0 & \ddots & \dots \\ 0 & 0 & \gamma_m \end{bmatrix}$$

Si $m > n$, on obtient à la fin du processus :

$$H \cdot A = R = \begin{bmatrix} \gamma_1 & \dots & \dots \\ 0 & \ddots & \dots \\ 0 & 0 & \gamma_m \\ 0 & 0 & 0 \end{bmatrix}$$

On voit que dans tous les cas la matrice R est triangulaire supérieure. Posons :

$$Q = H^{-1} = H_1 \cdot ... \cdot H_p = H^\star$$

On a alors la décomposition :

$$A = Q \cdot R$$

où Q est une matrice unitaire et R une matrice triangulaire supérieure. On note cette décomposition :

$$(Q, R) = \text{Householder}(A)$$

Sixième partie

Intégrales

Chapitre 67

Mesures

67.1 Dépendances

- Chapitre **??** : Les ensembles
- Chapitre **??** : Les ordres et extréma
- Chapitre **??** : Les fonctions

67.2 Introduction

L'objectif des mesures est de « mesurer » des ensembles, ou plutôt des sous-ensembles d'un ensemble donné. Soit l'ensemble Ω et une tribu de sous-ensembles $\mathcal{T} \subseteq \mathfrak{P}(\Omega)$. Une mesure sur $\mathcal{T}$ est une fonction $\mu : \mathcal{T} \mapsto \mathbb{R}$ associant une valeur réelle à chaque ensemble de la tribu. On demande que cette mesure soit positive :

$$\mu(A) \ge 0$$

pour tout $A \in \mathcal{T}$. Il semble également logique que la mesure d'un ensemble vide soit nulle :

$$\mu(\emptyset) = 0$$

Pour toute suite discrète (finie ou infinie) $\{A_1, A_2, ...\} \subseteq \mathcal{T}$ d'ensembles disjoints deux à deux, on a :

$$A_i \cap A_j = \emptyset$$

pour tout (i, j) tels que $i \neq j$. On exige dans ce cas que la mesure vérifie la propriété d'additivité :

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i)$$

Inclusion

Soit $A, B \in \mathcal{T}$ avec $A \subseteq B$. Comme $C = B \setminus A$ et A vérifient $C \cup A = B$ et $C \cap A = \emptyset$, on a :

$$\mu(B) = \mu(C) + \mu(A) \ge \mu(A)$$

La mesure d'un ensemble « plus petit » au sens de l'inclusion est donc plus petite :

$$\mu(A) \le \mu(B)$$

Union

Soit $A, B \in \mathcal{T}$. Comme $A \cup B = (A \setminus B) \cup B$ et $(A \setminus B) \cap B = \emptyset$, on a :

$$\mu(A \cup B) = \mu(A \setminus B) + \mu(B)$$

Comme $A \setminus B \subseteq A$, on a aussi $\mu(A \setminus B) \le \mu(A)$. On en déduit que :

$$\mu(A \cup B) \le \mu(A) + \mu(B)$$

On peut en conclure par récurrence que :

$$\mu\left(\bigcup_{i=0}^{n} A_i\right) \le \sum_{i=0}^{n} \mu(A_i)$$

Puis, par passage à la limite :

$$\mu\left(\bigcup_{i=0}^{+\infty} A_i\right) \le \sum_{i=0}^{+\infty} \mu(A_i)$$

Appellation

On dit qu'un ensemble A est mesurable (pour μ) si $A \in \mathcal{T}$. Dans la suite, nous considérons une mesure $\mu : \mathcal{T} \mapsto \mathbb{R}$ et un ensemble mesurable $A \in \mathcal{T}$.

67.3 Lebesgue

La mesure de Lebesgue μ_L est définie sur la tribu $\mathcal{T}$ engendrée par les ensembles ouverts de $\mathbb{R}$. Elle exprime simplement la longueur d'un intervalle. Pour tout :

$$I \in \left\{ \ [a,b], \ \ ]a,b[\ , \ \ ]a,b], [a,b[\ \right\}$$

on a simplement :

$$\mu_L(I) = b - a$$

Soit $\mathfrak{I}$ l'ensemble des collections au plus dénombrables d'intervalles ouverts disjoints. Pour tout $A \in \mathcal{T}$, on définit :

$$\mu_I(A) = \inf \left\{ \sum_{n \in N} \mu_L(I_n) : \{I_n : n \in N \subseteq \mathbb{N}\} \in \mathfrak{I}, \ A \subseteq \bigcup_{n \in N} I_n \right\}$$

et :

$$\mu^S(A) = \sup \left\{ \sum_{n \in N} \mu_L(I_n) : \{I_n : n \in N \subseteq \mathbb{N}\} \in \mathfrak{I}, \ \bigcup_{n \in N} I_n \subseteq A \right\}$$

Si :

$$\mu_I(A) = \mu^S(A)$$

on dit que l'ensemble A est mesurable au sens de Lebesgue et on définit :

$$\mu_L(A) = \mu_I(A) = \mu^S(A)$$

67.3.1 Mesure nulle

Pour tout ensemble N inclus dans un ensemble $A \in \mathcal{T}$ de mesure nulle :

$$\mu_L(A) = 0$$

on définit :

$$\mu_L(N) = 0$$

67.3.2 Singleton

On voit que les ensembles de la forme $\{a\} = [a, a]$ sont de mesure nulle :

$$\mu_L(\{a\}) = a - a = 0$$

On en conclut que pour toute suite discrète de réels $a_1, a_2, ...$, on a :

$$\mu_L(\{a_1, a_2, ...\}) = \sum_i \mu_L(\{a_i\}) = 0$$

On a aussi :

$$\begin{aligned} b - a = \mu_L([a, b]) &= \mu_L([a, b] \setminus \{a_1, a_2, ...\}) + \mu_L(\{a_1, a_2, ...\}) \\ &= \mu_L([a, b] \setminus \{a_1, a_2, ...\}) + 0 \end{aligned}$$

et donc :

$$\mu_L([a, b] \setminus \{a_1, a_2, ...\}) = b - a$$

67.4 Mesure de Stieltjes

On associe à toute fonction croissante $g : \mathbb{R} \mapsto \mathbb{R}$ une mesure de Stieltjes μ_g. Pour tout :

$$I \in \{\ [a, b],\ ]a, b[\ ,\ ]a, b], [a, b[\ \}$$

on définit :

$$\mu_g(I) = g(b) - g(a)$$

Soit $\mathfrak{J}$ l'ensemble des collections au plus dénombrables d'intervalles ouverts disjoints. Pour tout $A \in \mathcal{T}$, on définit :

$$\mu_g(A) = \inf \left\{ \sum_{n \in N} \mu_L(I_n) : \{I_n : n \in N \subseteq \mathbb{N}\} \in \mathfrak{J},\ A \subseteq \bigcup_{n \in N} I_n \right\}$$

67.4.1 Mesure nulle

Pour tout ensemble N inclus dans un ensemble $A \in \mathcal{T}$ de mesure nulle :

$$\mu_g(A) = 0$$

on définit :

$$\mu_g(N) = 0$$

67.5 Dirac

La mesure de Dirac μ_D^a en a est définie sur $\mathfrak{P}(\Omega)$. Il s'agit d'une mesure permettant de détecter si un $A \subseteq \Omega$ donné contient a. Elle est donc basée sur les fonctions indicatrices :

$$\mu_D^a(A) = \boldsymbol{\delta}_A(a) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{si } a \notin A \end{cases}$$

67.6 Mesure produit

Soit les tribus $\mathcal{T}_1$ et $\mathcal{T}_2$ et la tribu produit :

$$\mathcal{P} = \{A \times B : A \in \mathcal{T}_1,\ B \in \mathcal{T}_2\}$$

A partir de mesures $\mu : \mathcal{T}_1 \mapsto \mathbb{R}$ et $\nu : \mathcal{T}_2 \mapsto \mathbb{R}$, on peut construire une mesure produit $\mu \otimes \nu : \mathcal{P} \mapsto \mathbb{R}$ par :

$$(\mu \otimes \nu)(A \times B) = \mu(A) \cdot \nu(B)$$

Dimension n

On généralise la mesure de Lebesgue sur $\mathbb{R}^n$ par :

$$\mu_L(\ ]a_1, b_1[\ \times\ ]a_2, b_2[\ ...\times\ ]a_n, b_n[\ ) = \prod_{i=1}^{n} (b_i - a_i)$$

et l'extension à la tribu engendrée par les ouverts de $\mathbb{R}^n$ au moyen des supremum et infimum.

67.7 Fonction mesurable

On dit qu'une fonction $f : A \mapsto \mathbb{R}$ est mesurable (au sens de la tribu $\mathcal{T}$) si la relation f^{-1} vérifie $f^{-1}(]a, +\infty[) \in \mathcal{T}$ et $f^{-1}(] - \infty, a[) \in \mathcal{T}$ pour tout $a \in \mathbb{R}$. On a donc :

$$\{x \in A : f(x)\ >\ a\} \in \mathcal{T}$$
$$\{x \in A : f(x)\ <\ a\} \in \mathcal{T}$$

Corollaires

On a :

$$\{x \in A : f(x) \ge a\} = A \setminus \{x \in A : f(x)\ <\ a\} \in \mathcal{T}$$
$$\{x \in A : f(x) \le a\} = A \setminus \{x \in A : f(x)\ >\ a\} \in \mathcal{T}$$

et :

$$\{x \in A : f(x) = a\} = \{x \in A : f(x) \ge a\} \cap \{x \in A : f(x) \le a\} \in \mathcal{T}$$

Les mesures de tous ces ensembles sont donc bien définies pour tout $a \in \mathbb{R}$.

67.8 Opposé d'une fonction mesurable

Soit une fonction mesurable $f : A \mapsto \mathbb{R}$. On a :

$$\{x \in A : -f(x) > a\} = \{x \in A : f(x) < -a\} \in \mathcal{T}$$
$$\{x \in A : -f(x) < a\} = \{x \in A : f(x) > -a\} \in \mathcal{T}$$

On en déduit que la fonction opposée $-f$ est mesurable.

67.9 Fonctions extrema

Soit la suite $\{f_n : n \in \mathbb{N}\}$ de fonctions mesurables. Posons :

$$S = \sup\{f_n : n \in \mathbb{N}\}$$
$$I = \inf\{f_n : n \in \mathbb{N}\}$$

On a :

$$\{x \in A : S(x) > a\} = \bigcup_n \{x \in A : f_n(x) > a\} \in \mathcal{T}$$
$$\{x \in A : S(x) < a\} = \bigcap_n \{x \in A : f_n(x) < a\} \in \mathcal{T}$$

On en conclut que $\sup_n f_n$ est mesurable. Symétriquement, on a :

$$\{x \in A : I(x) > a\} = \bigcap_n \{x \in A : f_n(x) > a\} \in \mathcal{T}$$
$$\{x \in A : I(x) < a\} = \bigcup_n \{x \in A : f_n(x) < a\} \in \mathcal{T}$$

On en conclut que $\inf_n f_n$ est mesurable.

Chapitre 68

Essentialité

Dépendances

— Chapitre 71 : Les mesures

68.1 Introduction

Soit la mesure $\mu : \mathcal{T} \mapsto \mathbb{R}$ et l'ensemble $A \in \mathcal{T}$. On a envie de dire que tout sous-ensemble de mesure nulle de A est « négligeable ». L'essentiel de l'information demeurera donc si on s'abstrait d'un quelconque ensemble $N \subseteq A$ vérifiant $N \in \mathcal{T}$ et $\mu(N) = 0$. Le résultat de cette abstraction, soit $A \setminus N$, est appelé sous-ensemble essentiel de A. On note :

$$\mathfrak{E}(A) = \{A \setminus N : N \subseteq A,\ N \in \mathcal{T},\ \mu(N) = 0\}$$

l'ensemble des sous-ensembles essentiels de A.

68.2 Ensembles de mesure nulle

Inclusion

Soit $A, B \in \mathcal{T}$ avec $A \subseteq B$ et $\mu(B) = 0$. On a alors :

$$0 \le \mu(A) \le \mu(B) = 0$$

On en déduit que :

$$\mu(A) = 0$$

Un ensemble inclus dans un ensemble de mesure nulle est également de mesure nulle.

Union

Soit $A, B \in \mathcal{T}$ avec $\mu(A) = \mu(B) = 0$. On a alors :

$$0 \le \mu(A \cup B) \le \mu(A) + \mu(B) = 0 + 0 = 0$$

On en déduit que :

$$\mu(A \cup B) = 0$$

De même, la mesure d'une union d'une suite finie d'ensembles de mesure nulle est de mesure nulle :

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = 0$$

On a le même résultat pour les suites infinies :

$$\mu\left(\bigcup_{i=1}^{+\infty} A_i\right) = 0$$

68.3 Mesure d'un sous-ensemble essentiel

Soit $E \in \mathfrak{E}(A)$. On peut alors trouver $Z \in \mathcal{T}$ tel que $Z \subseteq A$, $E = A \setminus Z$ et vérifiant $\mu(Z) = 0$. Comme les ensembles E et Z sont disjoints et d'union égale à A, on a :

$$\mu(A) = \mu(E) + \mu(Z) = \mu(E) + 0 = \mu(E)$$

On a donc :

$$\mu(E) = \mu(A)$$

pour tout sous-ensemble essentiel $E \in \mathfrak{E}(A)$.

68.4 Convergence de la mesure

Soit une suite d'ensembles $\{A_n \in \mathcal{T} : n \in \mathbb{N}\}$ telle que $A_m \subseteq A_n$ pour tout naturels m, n vérifiant $m \le n$. Posons :

$$A = \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{T}$$

Considérons la suite d'ensemble $\{D_n \subseteq A : n \in \mathbb{N}\}$ définie par :

$$\begin{aligned} D_0 &= A_0 \\ D_n &= A_n \setminus A_{n-1} \end{aligned}$$

pour tout $n \ge 1$. Comme $A_{n-1}, D_n \subseteq A_n$, on a $A_n = A_{n-1} \cup D_n$. Nous allons montrer par récurrence que :

$$A_n = \bigcup_{k=0}^{n} D_k$$

On sait que c'est vrai pour $n = 0$. Supposons que ce soit vrai pour $n - 1$. On a :

$$A_n = D_n \cup A_{n-1} = D_n \cup \bigcup_{k=0}^{n-1} D_k = \bigcup_{k=0}^{n} D_k$$

Si x appartient à au moins un des A_n, il appartient à au moins un des D_k pour $0 \le k \le n$, et inversément. On en déduit que :

$$A = \bigcup_{n\in\mathbb{N}} A_n = \bigcup_{k\in\mathbb{N}} D_k$$

Soit les naturels i, j avec $i \ne j$. Soit $m = \max\{i, j\}$ et $n = \min\{i, j\}$. Si $D_i = \emptyset$ ou si $D_j = \emptyset$, on a forcément $D_i \cap D_j = \emptyset$. Sinon, soit $x \in D_m$. On a alors $x \in A_m$ et $x \notin A_{m-1}$. Comme $i \ne j$, on a $m - 1 \ge n$ et $A_n \subseteq A_{m-1}$. On en déduit que :

$$x \notin A_n = \bigcup_{k=0}^{n} D_k \subseteq A_{m-1}$$

En particulier, x ne peut pas appartenir à D_n. On en conclut que $D_i \cap D_j = D_m \cap D_n = \emptyset$. Les propriétés de la mesure nous disent alors que :

$$\mu(A_n) = \sum_{k=0}^{n} \mu(D_k)$$

et que :

$$\mu(A) = \sum_{n \in \mathbb{N}} \mu(D_n)$$

Par définition de la somme infinie sur $\mathbb{N}$, on a :

$$\sum_{n \in \mathbb{N}} \mu(D_n) = \lim_{n \to \infty} \sum_{k=0}^{n} \mu(D_k)$$

et donc :

$$\lim_{n \to \infty} \mu(A_n) = \mu(A)$$

68.5 Convergence étendue

Soit une suite d'ensembles $\{A_n \in \mathcal{T} : n \in \mathbb{N}\}$ telle que, pour tout $n \in \mathbb{N}$, on puisse trouver un sous-ensemble essentiel $E_n \subseteq A_n$ vérifiant $E_n \subseteq A_{n+1}$. Posons :

$$A = \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{T}$$

Pour tout $n \in \mathbb{N}$, on peut donc trouver un $Z_n \subseteq A_n$ vérifiant $\mu(Z_n) = 0$ et $E_n = A_n \setminus Z_n \subseteq A_{n+1}$. Notons Z l'union de ces ensembles :

$$Z = \bigcup_{n \in \mathbb{N}} Z_n$$

et analysons le comportement des $C_n = A_n \setminus Z$. Notons C l'union de ces ensembles. On a :

$$C = \bigcup_{n \in \mathbb{N}} C_n = \bigcup_{n \in \mathbb{N}} A_n \setminus Z = A \setminus Z$$

Supposons que $x \in C_n$. On a alors $x \in A_n$ et $x \notin Z_n$, d'où $x \in A_{n+1}$. Comme $x \notin Z$, on a aussi $x \in A_{n+1} \setminus Z = C_{n+1}$. On en conclut que $C_n \subseteq C_{n+1}$. La récurrence :

$$C_0 \subseteq C_1 \subseteq C_2 \subseteq C_3 \subseteq ...$$

nous montre alors que $C_m \subseteq C_n$ pour tout $m, n \in \mathbb{N}$ vérifiant $m \le n$. La mesure converge par conséquent vers la mesure de l'union :

$$\lim_{n \to \infty} \mu(C_n) = \mu(C) = \mu(A \setminus Z)$$

Mais comme C est un sous-ensemble essentiel de A, on a $\mu(C) = \mu(A)$. Pour la même raison, on a $\mu(C_n) = \mu(A_n)$. On a donc finalement :

$$\lim_{n \to \infty} \mu(A_n) = \mu(A)$$

68.6 Ordre faible

Infériorité essentielle

Soit deux fonctions $f, g : A \mapsto \mathbb{R}$ telles que $f - g$ soit mesurable. On dit que f est *essentiellement* inférieure à g, et on le note :

$$f \lesssim g$$

si on peut trouver un sous-ensemble essentiel S de A tel que $f(x) \le g(x)$ en tout point $x \in S$. On a donc :

$$\mu(\{x \in A : f(x) \ > \ g(x)\}) = 0$$

Supériorité essentielle

Soit deux fonctions $f, g : A \mapsto \mathbb{R}$ telles que $f - g$ soit mesurable. On dit que f est *essentiellement* supérieure à g, et on le note :

$$f \gtrsim g$$

si on peut trouver un sous-ensemble essentiel S de A tel que $f(x) \ge g(x)$ en tout point $x \in S$. On a donc :

$$\mu(\{x \in A : f(x) \ < \ g(x)\}) = 0$$

Validité

La définition de $f \lesssim g$ revient à imposer que :

$$\mu(\{x \in A : f(x) - g(x) \ > \ 0\}) = 0$$

La définition de $f \gtrsim g$ revient à imposer que :

$$\mu(\{x \in A : f(x) - g(x) \ < \ 0\}) = 0$$

Comme $f - g$ est mesurable, ces notions sont correctement définies.

68.7 Transitivité

Soit les fonctions $f, g, h : A \mapsto \mathbb{R}$ telles que :

$$f \lesssim g$$

$$g \lesssim h$$

Si on pose :

$$Z = \{x \in A : f(x) \ > \ g(x)\}$$
$$N = \{x \in A : g(x) \ > \ h(x)\}$$

on a $\mu(Z) = \mu(N) = 0$. Comme une union finie d'ensembles de mesure nulle est de mesure nulle, on a $\mu(N \cup Z) = 0$. Pour tout $A \setminus (N \cup Z)$, on a $f(x) \le g(x)$ et $g(x) \le h(x)$. On en déduit que $f(x) \le h(x)$. Notre f est donc inférieure à h sur A sauf sur l'ensemble de mesure nulle $N \cup Z$. On en conclut que la fonction étagée f est essentiellement inférieure à h :

$$f \lesssim h$$

68.8 Conservation sous l'addition

Supposons que $f \lesssim g$ et que $u \lesssim v$. Les ensembles :

$$\begin{aligned} Z &= \{x \in A : f(x) \ > \ g(x)\} \\ N &= \{x \in A : u(x) \ > \ v(x)\} \end{aligned}$$

sont de mesure nulle. Leur union est donc également de mesure nulle et on a bien entendu $f(x) + u(x) \le g(x) + v(x)$ sur $A \setminus (Z \cup N)$ c'est-à-dire sur une sous-ensemble essentiel de A. On en conclut que :

$$f + u \lesssim g + v$$

68.9 Conservation sous la soustraction

Supposons que $f \lesssim g$ et que $u \gtrsim v$. Les ensembles :

$$\begin{aligned} Z &= \{x \in A : f(x) \ > \ g(x)\} \\ N &= \{x \in A : u(x) \ < \ v(x)\} \end{aligned}$$

sont de mesure nulle. Leur union est donc également de mesure nulle et on a bien entendu $f(x) - u(x) \le g(x) - v(x)$ sur $A \setminus (Z \cup N)$ c'est-à-dire sur une sous-ensemble essentiel de A. On en conclut que :

$$f - u \lesssim g - v$$

68.10 Fonctions max et min

Max

Supposons que $f, g \lesssim h$. Les ensembles :

$$\begin{aligned} Z &= \{x \in A : f(x) \ > \ h(x)\} \\ N &= \{x \in A : g(x) \ > \ h(x)\} \end{aligned}$$

sont de mesure nulle. Leur union est donc également de mesure nulle et on a bien entendu $\max\{f(x), g(x)\} \le h(x)$ sur $A \setminus (Z \cup N)$ c'est-à-dire sur une sous-ensemble essentiel de A. On en conclut que :

$$\max\{f, g\} \lesssim h$$

Min

Supposons que $f, g \gtrsim h$. Les ensembles :

$$\begin{aligned} Z &= \{x \in A : f(x) \ < \ h(x)\} \\ N &= \{x \in A : g(x) \ < \ h(x)\} \end{aligned}$$

sont de mesure nulle. Leur union est donc également de mesure nulle et on a bien entendu $\min\{f(x), g(x)\} \ge h(x)$ sur $A \setminus (Z \cup N)$ c'est-à-dire sur une sous-ensemble essentiel de A. On en conclut que :

$$\min\{f, g\} \gtrsim h$$

68.11 Egalité

On dit que f est essentiellement égale à g, et on le note :

$$f \approx g$$

si et seulement si $f \lesssim g$ et $f \gtrsim g$. Les ensembles :

$$\begin{aligned} Z &= \{x \in A : f(x) \ < \ g(x)\} \\ N &= \{x \in A : f(x) \ > \ g(x)\} \end{aligned}$$

sont alors de mesure nulle. Donc, l'ensemble :

$$D = \{x \in A : f(x) \ne g(x)\} = Z \cup N$$

est également de mesure nulle.

Chapitre 69

Extrema essentiels

Dépendances

— Chapitre 71 : Les mesures

69.1 Bornes

Soit la fonction mesurable $f : A \mapsto \mathbb{R}$.

Supérieure

On dit que f est *essentiellement* inférieure au réel $\sigma \in \mathbb{R}$, et on le note :

$$f \lesssim \sigma$$

si on peut trouver un sous-ensemble essentiel S de A tel que $f(x) \le \sigma$ en tout point $x \in S$. On a donc :

$$\mu(\{x \in A : f(x) \ > \ \sigma\}) = 0$$

On dit aussi que f est essentiellement majorée par σ.

Inférieure

On dit que f est *essentiellement* supérieure au réel $\lambda \in \mathbb{R}$, et on le note :

$$f \gtrsim \lambda$$

si on peut trouver un sous-ensemble essentiel S de A tel que $f(x) \ge \lambda$ en tout point $x \in S$. On a donc :

$$\mu(\{x \in A : f(x) \ < \ \lambda\}) = 0$$

On dit aussi que f est essentiellement minorée par λ.

69.2 Extrema essentiels

Soit une fonction mesurable $f : A \mapsto \mathbb{R}$. Pour tout $\sigma \in \mathbb{R}$, notons :

$$\Psi(\sigma) = \{x \in A : f(x) \ > \ \sigma\}$$

Comme on ne se soucie pas des ensembles de mesure nulle, on peut choisir (si elle existe) une borne supérieure $\sigma \in \mathbb{R}$ telle que $f \lesssim \sigma$, ce qui revient à imposer que $\mu(\Psi(\sigma)) = 0$. Notons :

$$\Theta = \{\sigma \in \mathbb{R} : f \lesssim \sigma\} = \{\sigma \in \mathbb{R} : \mu(\Psi(\sigma)) = 0\}$$

La finalité des bornes étant d'encadrer au plus près un ensemble, nous considérons l'infimum (s'il existe) des réels σ possédant cette propriété, et nous l'appelons supremum essentiel :

$$\operatorname*{ess\,sup}_{x \in A} f(x) = \inf \Theta$$

Symétriquement, on note :

$$\Gamma(\lambda) = \{x \in A : f(x) \ < \ \lambda\}$$

pour tout $\lambda \in \mathbb{R}$ et :

$$\Lambda = \{\lambda \in \mathbb{R} : f \gtrsim \lambda\} = \{\lambda \in \mathbb{R} : \mu(\Gamma(\lambda)) = 0\}$$

Si le supremum de Λ existe, on définit alors l'infimum essentiel par :

$$\operatorname*{ess\,inf}_{x \in A} f(x) = \sup \Lambda$$

Nous disposons donc des bornes :

$$\operatorname*{ess\,sup}_{x \in A} f(x) = \inf\{\sigma \in \mathbb{R} : f \lesssim \sigma\}$$

$$\operatorname*{ess\,inf}_{x \in A} f(x) = \sup\{\lambda \in \mathbb{R} : f \gtrsim \lambda\}$$

Notation

On note aussi :

$$\operatorname{ess\,sup}\{f(x) : x \in A\} = \operatorname*{ess\,sup}_{x \in A} f(x)$$
$$\operatorname{ess\,inf}\{f(x) : x \in A\} = \operatorname*{ess\,inf}_{x \in A} f(x)$$

Au besoin, la mesure μ utilisée est indiquée par :

$$\operatorname*{ess\,\overset{\mu}{sup}}_{x \in A} f(x)$$

$$\operatorname*{ess\,\overset{\mu}{inf}}_{x \in A} f(x)$$

69.3 Existence et estimation

Supremum essentiel

Supposons que f soit essentiellement inférieure (sur A) à un certain $\sigma \in \mathbb{R}$ et que f soit essentiellement supérieure à un certain $\lambda \in \mathbb{R}$ sur un ensemble $L \subseteq A$ de mesure strictement positive. On a $\sigma \in \Theta \neq \emptyset$. Pour tout $\alpha \in \mathbb{R}$ tel que $\alpha \geq \sigma$, on a $\Psi(\alpha) \subseteq \Psi(\sigma)$. Comme $\mu(\Psi(\sigma)) = 0$, l'ensemble $\Psi(\alpha)$ est inclus dans un ensemble de mesure nulle. Il est donc lui-même de mesure nulle et $\alpha \in \Theta$. On en conclut que :

$$[\sigma, +\infty[\subseteq \Theta$$

pour tout $\sigma \in \Theta$. Comme f est essentiellement supérieure à λ sur L, on peut trouver un ensemble de mesure nulle $Z \subseteq L$ tel que :

$$L \setminus Z = \{x \in L : f(x) \ge \lambda\}$$

Comme $L \setminus Z$ est un sous-ensemble essentiel de L, on a $\mu(L \setminus Z) = \mu(L) > 0$. Posons :

$$C = \{x \in A : f(x) \ge \lambda\}$$

Comme $L \subseteq A$, on a $L \setminus Z \subseteq C$ et $\mu(C) \ge \mu(L \setminus Z) > 0$. Soit $\beta \in \mathbb{R}$ vérifiant $\beta < \lambda$. La condition $f(x) \ge \lambda$ implique que $f(x) > \beta$. On en déduit que :

$$C \subseteq \{x \in A : f(x) > \beta\} = \Psi(\beta)$$

On a alors $\mu(\Psi(\beta)) \ge \mu(C) > 0$, ce qui implique que β ne peut pas appartenir à Θ. Tous les éléments $\theta \in \Theta$ vérifient donc $\theta \ge \lambda$ et :

$$\Theta \subseteq [\lambda, +\infty[$$

Comme $\lambda \le \Theta$, l'ensemble de réels Θ est non vide et minoré. Il admet donc un infimum :

$$\operatorname*{ess\,sup}_{x \in A} f(x) = \inf \Theta$$

On se rappelle que l'inclusion $X \subseteq Y$ implique que $\inf X \ge \inf Y$. Comme $\lambda = \inf[\lambda, +\infty[$ et $\sigma = \inf[\sigma, +\infty[$, les inclusions :

$$[\sigma, +\infty[\subseteq \Theta \subseteq [\lambda, +\infty[$$

nous montrent que :

$$\lambda \le \operatorname*{ess\,sup}_{x \in A} f(x) \le \sigma$$

Infimum essentiel

Supposons que f soit essentiellement supérieure (sur A) à un certain $\lambda \in \mathbb{R}$ et que f soit essentiellement inférieure à un certain $\sigma \in \mathbb{R}$ sur un ensemble $S \subseteq A$ de mesure strictement positive. On a $\lambda \in \Lambda \neq \emptyset$. Pour tout $\alpha \in \mathbb{R}$ tel que $\alpha \le \lambda$, on a $\Gamma(\alpha) \subseteq \Gamma(\lambda)$. Comme $\mu(\Gamma(\lambda)) = 0$, l'ensemble $\Gamma(\alpha)$ est inclus dans un ensemble de mesure nulle. Il est donc lui-même de mesure nulle et $\alpha \in \Lambda$. On en conclut que :

$$]-\infty, \lambda] \subseteq \Lambda$$

pour tout $\lambda \in \Lambda$. Comme f est essentiellement inférieure à σ sur S, on peut trouver un ensemble de mesure nulle $Z \subseteq S$ tel que :

$$S \setminus Z = \{x \in S : f(x) \le \sigma\}$$

Comme $S \setminus Z$ est un sous-ensemble essentiel de S, on a $\mu(S \setminus Z) = \mu(S) > 0$. Posons :

$$C = \{x \in A : f(x) \le \sigma\}$$

Comme $S \subseteq A$, on a $S \setminus Z \subseteq C$ et $\mu(C) \ge \mu(S \setminus Z) > 0$. Soit $\beta \in \mathbb{R}$ vérifiant $\beta > \sigma$. La condition $f(x) \le \sigma$ implique que $f(x) < \beta$. On en déduit que :

$$C \subseteq \{x \in A : f(x) < \beta\} = \Gamma(\beta)$$

On a alors $\mu(\Gamma(\beta)) \ge \mu(C) > 0$, ce qui implique que β ne peut pas appartenir à Λ. Tous les éléments $\gamma \in \Lambda$ vérifient donc $\gamma \le \sigma$ et :

$$\Theta \subseteq]-\infty, \sigma]$$

Comme $\sigma \ge \Lambda$, l'ensemble de réels non vide Λ est majoré. Il admet donc un supremum :

$$\operatorname*{ess\,inf}_{x \in A} f(x) = \sup \Lambda$$

On se rappelle que l'inclusion $X \subseteq Y$ implique que $\sup X \le \sup Y$. Comme $\lambda = \sup\ ]-\infty, \lambda]$ et $\sigma = \sup\ ]-\infty, \sigma]$, les inclusions :

$$]-\infty, \lambda] \ \subseteq \Theta \subseteq \ ]-\infty, \sigma]$$

nous montrent que :

$$\lambda \le \operatorname*{ess\,inf}_{x \in A} f(x) \le \sigma$$

69.4 Intersection

Supposons que A soit de mesure strictement positive et que $\Theta \cap \Lambda \neq \emptyset$. Pour tout $\alpha \in \mathbb{R}$, posons :

$$\Xi(\alpha) = \{x \in A : f(x) = \alpha\}$$

et :

$$\Upsilon(\alpha) = \{x \in A : f(x) \neq \alpha\}$$

Il est clair que :

$$A = \Xi(\alpha) \cup \Upsilon(\alpha)$$

Comme imposer la différence revient à imposer soit la supériorité soit l'infériorité stricte, on a :

$$\Upsilon(\alpha) = \Psi(\alpha) \cup \Gamma(\alpha)$$

Soit à présent $\alpha \in \Theta \cap \Lambda$. Comme $\mu(\Psi(\alpha)) = \mu(\Gamma(\alpha)) = 0$, leur union est également de mesure nulle :

$$\mu(\Upsilon(\alpha)) = 0$$

Si l'intersection $\Theta \cap \Lambda$ n'est pas vide, la fonction f est donc essentiellement constante sur A et $\Xi(\alpha) = A \setminus \Upsilon(\alpha)$ est un sous-ensemble essentiel de A. On a donc $\mu(\Xi(\alpha)) = \mu(A) > 0$. Soit à présent $\beta \in \mathbb{R}$ et supposons que $\beta \neq \alpha$. Si $f(x) = \beta$, on a forcément $f(x) \neq \alpha$, donc :

$$\Xi(\beta) = \{x \in A : f(x) = \beta\} \subseteq \{x \in A : f(x) \neq \alpha\}$$

L'ensemble $\Xi(\beta)$ étant inclus dans un ensemble de mesure nulle, on doit avoir $\mu(\Xi(\beta)) = 0$. Comme $\mu(\Xi(\alpha)) > 0$ pour tout $\alpha \in \Theta \cap \Lambda$, β ne peut pas appartenir à $\Theta \cap \Lambda$. On en conclut que l'intersection vérifie soit :

$$\Theta \cap \Lambda = \emptyset$$

soit :

$$\Theta \cap \Lambda = \{\alpha\}$$

pour un certain $\alpha \in \mathbb{R}$.

Extrema

Supposons que f soit essentiellement inférieure à σ et essentiellement supérieure à λ sur A qui est de mesure strictement positive. On en déduit que les supremum et infimum essentiels existent :

$$S = \text{ess sup}\{f(x) : x \in A\}$$
$$I = \text{ess inf}\{f(x) : x \in A\}$$

et que $\lambda \le \{S, I\} \le \sigma$. Supposons que $S < I$ et posons $\delta = I - S > 0$. Soit le réel strictement positif $\epsilon = \delta/4$. Comme l'infimum est dans l'adhérence, on peut trouver un $\alpha \in \Theta$ tel que $|\alpha - S| \le \epsilon$, d'où $\alpha \le S + \epsilon$. On a alors :

$$[\alpha, +\infty[\subseteq \Theta$$

Comme le supremum est dans l'adhérence, on peut trouver un $\beta \in \Lambda$ tel que $|I - \beta| \le \epsilon$, d'où $\beta \ge I - \epsilon$. On a alors :

$$] - \infty, \beta] \subseteq \Lambda$$

On voit que :

$$\begin{aligned} S + \epsilon &= S + \frac{\delta}{4} \\ I - \epsilon &= S + \delta - \frac{\delta}{4} = S + \frac{3\delta}{4} \end{aligned}$$

On a donc :

$$\alpha \le S + \epsilon < I - \epsilon \le \beta$$

et $\alpha < \beta$. On a aussi :

$$[\alpha, \beta] =] - \infty, \beta] \cap [\alpha, +\infty[\subseteq \Theta \cap \Lambda$$

Donc $\alpha, \beta \in \Theta \cap \Lambda$ avec $\alpha \neq \beta$ ce qui impossible. Notre hypothèse est donc fausse et $I \le S$, c'est-à-dire :

$$\operatorname*{ess\,inf}_{x \in A} f(x) \le \operatorname*{ess\,sup}_{x \in A} f(x)$$

69.5 Ordre

Soit les fonctions $f, g : A \mapsto \mathbb{R}$ telles que $f \lesssim g$.

Supremum essentiel

On a bien entendu :

$$f \lesssim g \lesssim \operatorname*{ess\,sup}_{x \in A} g(x)$$

On en déduit que :

$$\operatorname*{ess\,sup}_{x \in A} f(x) \le \operatorname*{ess\,sup}_{x \in A} g(x)$$

Infimum essentiel

On a bien entendu :

$$\operatorname*{ess\,inf}_{x \in A} f(x) \lesssim f \lesssim g$$

On en déduit que :

$$\operatorname*{ess\,inf}_{x \in A} f(x) \le \operatorname*{ess\,inf}_{x \in A} g(x)$$

69.6 Addition

Soit les fonctions $f, g : A \mapsto \mathbb{R}$.

Supremum essentiel

On sait que :

$$f \lesssim \sigma = \operatorname{ess\,sup}\{f(x) : x \in A\}$$
$$g \lesssim \tau = \operatorname{ess\,sup}\{g(x) : x \in A\}$$

On a donc :

$$f + g \lesssim \sigma + \tau$$

On en conclut que :

$$\operatorname*{ess\,sup}_{x \in A} [f(x) + g(x)] \le \operatorname*{ess\,sup}_{x \in A} f(x) + \operatorname*{ess\,sup}_{x \in A} g(x)$$

Infimum essentiel

On sait que :

$$f \gtrsim \lambda = \operatorname{ess\,inf}\{f(x) : x \in A\}$$
$$g \gtrsim \gamma = \operatorname{ess\,inf}\{g(x) : x \in A\}$$

On a donc :

$$f + g \gtrsim \lambda + \gamma$$

On en conclut que :

$$\operatorname*{ess\,inf}_{x \in A} [f(x) + g(x)] \ge \operatorname*{ess\,inf}_{x \in A} f(x) + \operatorname*{ess\,inf}_{x \in A} g(x)$$

Chapitre 70

Fonctions étagées

Dépendances

- Chapitre ?? : Les fonctions
- Chapitre 71 : Les mesures

70.1 Définition

Soit la mesure $\mu : \mathcal{T} \mapsto \mathbb{R}$, l'ensemble $\Phi \in \mathcal{T}$ et $A \in \mathcal{T}$ vérifiant $A \subseteq \Phi$. On dit que la fonction $w : \Phi \to \mathbb{R}$ est étagée sur A si on peut trouver une collection finie $\{A_1, ..., A_n\} \subseteq \mathcal{T}$ formant une partition de A et des réels $w_i \in \mathbb{R}$ tels que :

$$w(x) = \sum_{i=1}^{n} w_i \cdot \boldsymbol{\delta}_{A_i}(x)$$

pour tout $x \in A$. On note $\text{Etagee}(A)$ l'ensemble des fonctions étagées sur A.

70.2 Mesurabilité

Soit $a \in \mathbb{R}$. Comme :

$$\{x \in A : w(x) > a\} = \bigcup \{A_i : i \in \{1, ..., n\},\ w_i > a\}$$

et que l'union s'opère sur un nombre fini d'ensembles, on a :

$$\{x \in A : w(x) > a\} \in \mathcal{T}$$

On a aussi :

$$\{x \in A : w(x) < a\} = \bigcup \{A_i : i \in \{1, ..., n\},\ w_i < a\} \in \mathcal{T}$$

Les fonctions étagées sont mesurables.

70.3 Multiplication par une fonction indicatrice

Soit $B \subseteq A$. On a :

$$\begin{aligned} w(x) \cdot \boldsymbol{\delta}_B(x) &= \sum_{i=1}^n w_i \cdot \boldsymbol{\delta}_{A_i}(x) \cdot \boldsymbol{\delta}_B(x) \\ &= \sum_{i=1}^n w_i \cdot \boldsymbol{\delta}[A_i \cap B](x) \end{aligned}$$

Soit $B_i = A_i \cap B$. La fonction $u = w \cdot \boldsymbol{\delta}_B$ peut se définir par :

$$u(x) = w(x) \cdot \boldsymbol{\delta}_B(x) = \sum_{i=1}^n w_i \cdot \boldsymbol{\delta}_{B_i}(x)$$

pour tout $x \in A$. Si $i \ne j$, on a :

$$B_i \cap B_j = (A_i \cap B) \cap (A_j \cap B) = (A_i \cap A_j) \cap B = \emptyset$$

L'union des B_i nous donne :

$$\bigcup_i B_i = \bigcup_i (A_i \cap B) = B \cap \bigcup_i A_i = B \cap A = B$$

On en conclut que les ensembles B_i forment une partition de B. La fonction u est donc une fonction étagée sur B. Comme u est nulle sur $A \setminus B$, on a même :

$$u = 0 \cdot \boldsymbol{\delta}[A \setminus B] + \sum_{i=1}^n w_i \cdot \boldsymbol{\delta}_{B_i}$$

ce qui montre que u est une fonction étagée sur $(A \setminus B) \cup B = A$.

70.4 Essentialité

Bornes

Soit l'ensemble M rassemblant les indices correspondant aux ensembles de mesure non nulle :

$$M = \{i \in \{1, ..., n\} : \mu(A_i) \ > \ 0\}$$

et l'ensemble N rassemblant les indices correspondant aux ensembles de mesure nulle :

$$N = \{i \in \{1, ..., n\} : \mu(A_i) = 0\}$$

Par positivité de la mesure, on a clairement $M \cup N = \{1, ..., n\}$. Comme l'union d'ensembles de mesure nulle est également de mesure nulle, l'ensemble :

$$Z = \bigcup_{i \in N} A_i$$

vérifie $\mu(Z) = 0$. L'ensemble :

$$A \setminus Z = A \setminus \bigcup_{i \in N} A_i = \bigcup_{i \in M} A_i$$

est donc un sous-ensemble essentiel de A. Comme :

$$\min_{i \in M} w_i \le w(x) \in \{w_i : i \in M\} \le \max_{i \in M} w_i$$

pour tout $x \in A \setminus Z$, on a :

$$\min\{w_i : i \in M\} \lesssim w \lesssim \max\{w_i : i \in M\}$$

Extrema

Soit :

$$\alpha = \arg\max_{i \in M} w_i \in M$$
$$\beta = \arg\min_{i \in M} w_i \in M$$

On a alors :

$$\sigma = w_\alpha = \max\{w_i : i \in M\}$$
$$\lambda = w_\beta = \min\{w_i : i \in M\}$$

Comme $w \lesssim \sigma$, on doit avoir :

$$\operatorname*{ess\,sup}_{x \in A} w(x) \le \sigma$$

Mais on a aussi $w = w_\alpha = \sigma$ sur A_α, et a fortiori $w \ge \sigma$ sur A_α, avec $\mu(A_\alpha) > 0$, ce qui implique :

$$\operatorname*{ess\,sup}_{x \in A} w(x) \ge \sigma$$

Ces deux inégalités nous imposent que :

$$\operatorname*{ess\,sup}_{x \in A} w(x) = \sigma$$

Comme $w \gtrsim \lambda$, on a :

$$\operatorname*{ess\,inf}_{x \in A} w(x) \ge \lambda$$

Mais on a aussi $w = w_\beta = \lambda$ sur A_β, et a fortiori $w \le \lambda$ sur A_β, avec $\mu(A_\beta) > 0$, ce qui implique :

$$\operatorname*{ess\,inf}_{x \in A} w(x) \le \lambda$$

Ces deux inégalités nous imposent que :

$$\operatorname*{ess\,inf}_{x \in A} w(x) = \lambda$$

Les extrema essentiels d'une fonction étagée existent toujours et sont très simples à évaluer :

$$\operatorname*{ess\,sup}_{x \in A} w(x) = \max\{w_i : i \in \{1, ..., n\},\ \mu(A_i) > 0\}$$

$$\operatorname*{ess\,inf}_{x \in A} w(x) = \min\{w_i : i \in \{1, ..., n\},\ \mu(A_i) > 0\}$$

70.5 Intégrale d'une mesure

L'intégrale d'une mesure $\mu : \mathcal{T} \mapsto \mathbb{R}$ sur $A \in \mathcal{T}$ est simplement la mesure de A :

$$\int_A d\mu(x) = \mu(A)$$

70.6 Intégrale d'une constante

On étend la définition aux constantes $c \in \mathbb{R}$ par linéarité :

$$\int_A c\ d\mu(x) = c \cdot \int_A d\mu(x) = c \cdot \mu(A)$$

Fonction nulle

La fonction nulle peut être considérée comme la constante 0. On a donc :

$$\int_A 0 \ d\mu(x) = 0 \cdot \int_A d\mu(x) = 0$$

70.7 Intégrale d'une fonction étagée

Soit un ensemble A et une fonction $w \in \mathrm{Etagee}(A)$. On peut trouver des $w_i \in \mathbb{R}$ et des $A_i \in \mathcal{T}$ formant une partition de A et tels que :

$$w(x) = \sum_{i=1}^{n} w_i \cdot \boldsymbol{\delta}_{A_i}(x)$$

pour tout $x \in A$. On étend la définition de l'intégrale en imposant la linéarité :

$$\int_A w(x) \ d\mu(x) = \sum_{i=1}^{n} \int_{A_i} w(x) \ d\mu(x)$$

La fonction étagée w étant constante et valant w_i sur chaque A_i, on a :

$$\int_{A_i} w(x) \ d\mu(x) = \int_{A_i} w_i \ d\mu(x) = w_i \cdot \mu(A_i)$$

et finalement :

$$\int_A w(x) \ d\mu(x) = \sum_{i=1}^{n} w_i \cdot \mu(A_i)$$

70.8 Unicité

Nous allons vérifier que, pour une fonction étagée w donnée, l'intégrale ainsi définie ne dépend pas de la partition utilisée. Soit les collections finies d'ensembles $A_i \in \mathcal{T}$ et $B_j \in \mathcal{T}$ telles que $\{A_1, A_2, ..., A_m\}$ et $\{B_1, B_2, ..., B_n\}$ forment deux partitions de A et telles que l'on puisse trouver des $a_i, b_j \in \mathbb{R}$ vérifiant :

$$w(x) = \sum_{i=1}^{m} a_i \cdot \boldsymbol{\delta}[A_i](x) = \sum_{j=1}^{n} b_j \cdot \boldsymbol{\delta}[B_j](x)$$

pour tout $x \in A$.

Intersection des partitions

Posons $P = \{1, ..., m\} \times \{1, ..., n\}$ et $C_{ij} = A_i \cup B_j \in \mathcal{T}$ pour tout $(i, j) \in P$. Si $i \neq k$ ou si $j \neq l$, on a :

$$C_{ij} \cap C_{kl} = A_i \cap B_j \cap A_k \cap B_l = (A_i \cap A_k) \cap (B_j \cap B_l) = \emptyset$$

On voit également que :

$$\bigcup_j C_{ij} = \bigcup_j (A_i \cap B_j) = A_i \cap \bigcup_j B_j = A_i \cap A = A_i$$
$$\bigcup_i C_{ij} = \bigcup_i (A_i \cap B_j) = B_j \cap \bigcup_i A_i = A \cap B_j = B_j$$

Leur union globale sur les (i, j) s'écrit donc :

$$\bigcup_{i,j} C_{ij} = \bigcup_i A_i = A$$

Les C_{ij} forment également une partition de A.

Nous savons que les C_{ij} ne se chevauchent pas et que leur union sur j permet de reformer les A_i. De même, leur union sur i permet de reformer les B_j. On en conclut que :

$$\mu(A_i) = \sum_j \mu(C_{ij})$$
$$\mu(B_j) = \sum_i \mu(C_{ij})$$

et :

$$\boldsymbol{\delta}[A_i] = \sum_j \boldsymbol{\delta}[C_{ij}]$$
$$\boldsymbol{\delta}[B_j] = \sum_i \boldsymbol{\delta}[C_{ij}]$$

Expression alternative

Utilisant les propriétés des C_{ij}, la fonction w peut se réécrire comme suit :

$$w(x) = \sum_{(i,j)\in P} a_i \cdot \boldsymbol{\delta}[C_{ij}](x) = \sum_{(i,j)\in P} b_j \cdot \boldsymbol{\delta}[C_{ij}](x)$$

On en conclut que :

$$\sum_{(i,j)\in P} (a_i - b_j) \cdot \boldsymbol{\delta}[C_{ij}](x) = 0$$

Implication de l'égalité

Soit un couple $(i, j) \in P$ tel que $\boldsymbol{\delta}[C_{ij}] \neq 0$, c'est-à-dire $C_{ij} \neq \emptyset$. Soit $x \in C_{ij}$. Pour tout $(k, l) \in P \setminus (i, j)$, on a $C_{kl} \cap C_{ij} = \emptyset$ et donc $\boldsymbol{\delta}[C_{kl}](x) = 0$. La somme :

$$\sum_{(k,l)\in P} (a_k - b_l) \cdot \boldsymbol{\delta}[C_{kl}](x) = 0$$

se réduit donc à :

$$(a_i - b_j) \cdot \boldsymbol{\delta}[C_{ij}](x) = 0$$

On doit donc avoir $a_i - b_j = 0$, c'est-à-dire $a_i = b_j$. On pose :

$$F = \{(i, j) \in P : C_{ij} \neq \emptyset\}$$

Intégrales

L'intégrale évaluée sur le découpage des A_i s'écrit :

$$I = \sum_i a_i \cdot \mu(A_i) = \sum_{(i,j)\in P} a_i \cdot \mu(C_{ij})$$

Mais comme $\mu(C_{ij}) = \mu(\emptyset) = 0$ pour tout $(i, j) \in P \setminus F$, on a finalement :

$$I = \sum_{(i,j) \in F} a_i \cdot \mu(C_{ij})$$

De même, l'intégrale évaluée sur le découpage des B_j s'écrit :

$$J = \sum_j b_j \cdot \mu(B_j) = \sum_{(i,j) \in P} b_j \cdot \mu(C_{ij}) = \sum_{(i,j) \in F} b_j \cdot \mu(C_{ij})$$

Comme $a_i = b_j$ pour tous les couples $(i, j) \in F$, on en déduit que :

$$J = \sum_{(i,j) \in F} a_i \cdot \mu(C_{ij}) = I$$

L'intégrale d'une fonction étagée ne dépend donc pas du choix de la partition utilisée.

70.9 Ordre

Soit les fonctions étagées $u, v \in \text{Etagee}(A)$ vérifiant $u \lesssim v$. Soit les partitions $\{U_1, ..., U_m\}$ et $\{V_1, ..., V_n\}$ et les réels u_i, v_j tels que :

$$u(x) = \sum_{i=1}^{m} u_i \cdot \boldsymbol{\delta}_{U_i}(x)$$
$$v(x) = \sum_{j=1}^{n} v_j \cdot \boldsymbol{\delta}_{V_j}(x)$$

pour tout $x \in A$. Si $W_{ij} = U_i \cap V_j$, on sait que :

$$u(x) = \sum_{i,j} u_i \cdot \boldsymbol{\delta}[W_{ij}](x)$$
$$v(x) = \sum_{i,j} v_j \cdot \boldsymbol{\delta}[W_{ij}](x)$$

Posons $P = \{1, ..., m\} \times \{1, ..., n\}$ et :

$$M = \{(i, j) \in P : \mu(W_{ij}) > 0\}$$

Considérons le choix $(i, j) \in P$ et $x \in W_{ij}$. Comme u est essentiellement inférieure à v, on doit avoir $u_i \le v_j$ dès que $\mu(W_{ij}) > 0$, c'est-à-dire lorsque $(i, j) \in M$. Examinons à présent les intégrales. Les mesures des W_{ij} s'annulant pour tout $(i, j) \in P \setminus M$, on a :

$$\int_A u(x)\, d\mu(x) = \sum_{i,j} u_i \cdot \mu(W_{ij}) = \sum_{(i,j) \in M} u_i \cdot \mu(W_{ij})$$

et :

$$\int_A v(x)\, d\mu(x) = \sum_{i,j} v_j \cdot \mu(W_{ij}) = \sum_{(i,j) \in M} v_j \cdot \mu(W_{ij})$$

La propriété $u_i \le v_j$ étant vérifiée pour tout $(i, j) \in M$, on en conclut que :

$$\int_A u(x)\, d\mu(x) \le \int_A v(x)\, d\mu(x)$$

L'intégrale conserve l'ordre essentiel entre fonctions étagées.

Signe

Comme la fonction constante 0 est un cas particulier de fonction étagée, l'intégrale d'une fonction étagée u essentiellement supérieure à zéro est positive. De même, l'intégrale d'un fonction étagée v essentiellement inférieure à zéro est négative.

Egalité

Soit deux fonctions étagées u, v essentiellement égales ($u \approx v$). On a $u \lesssim v$ et l'intégrale de u est inférieure à l'intégrale de v. Mais on a aussi $u \gtrsim v$ et l'intégrale de u est supérieure à l'intégrale de v. On en conclut que :

$$\int_A u(x)\ d\mu(x) = \int_A v(x)\ d\mu(x)$$

70.10 Encadrement

Soit $w \in \mathrm{Etagee}(A)$. Comme les constantes sont des cas particuliers de fonctions étagées et que l'on a :

$$\operatorname*{ess\ inf}_{x \in A} w(x) \lesssim w \lesssim \operatorname*{ess\ sup}_{x \in A} w(x)$$

on en conclut en intégrant que :

$$\mu(A) \cdot \operatorname*{ess\ inf}_{x \in A} w(x) \le \int_A w(x)\ d\mu(x) \le \mu(A) \cdot \operatorname*{ess\ sup}_{x \in A} w(x)$$

70.11 Propriétés extrémales

Soit $w \in \mathrm{Etagee}(A)$. Si u est une fonction étagée vérifiant $u \lesssim w$, l'intégrale de u est inférieure à l'intégrale de w. Par ailleurs, le choix $u = w \lesssim w$ reproduit bien entendu l'intégrale de w. On en conclut que :

$$\int_A w(x)\ d\mu(x) = \sup \left\{ \int_A u(x)\ d\mu(x) : u \in \mathrm{Etagee}(A),\ u \lesssim w \right\}$$

On aboutit avec un raisonnement analogue à :

$$\int_A w(x)\ d\mu(x) = \inf \left\{ \int_A v(x)\ d\mu(x) : v \in \mathrm{Etagee}(A),\ v \gtrsim w \right\}$$

70.12 Additivité

Soit les ensembles A et B disjoints :

$$A \cap B = \emptyset$$

et une fonction $w \in \mathrm{Etagee}(A \cup B)$. On peut trouver des $w_i \in \mathbb{R}$ et des $C_i \in \mathcal{T}$ formant une partition de $A \cup B$ et tels que :

$$w(x) = \sum_i w_i \cdot \boldsymbol{\delta}_{C_i}(x)$$

pour tout $x \in A$. On sait que les $A_i = C_i \cap A$ forment une partition de A et que les $B_i = C_i \cap B$ forment une partition de B. On a aussi :

$$A_i \cap B_i = C_i \cap A \cap C_i \cap B = C_i \cap (A \cap B) = \emptyset$$

Comme :

$$C_i = C_i \cap (A \cup B) = (C_i \cap A) \cup (C_i \cap B) = A_i \cup B_i$$

on a $\boldsymbol{\delta}[C_i] = \boldsymbol{\delta}[A_i] + \boldsymbol{\delta}[B_i]$ et :

$$\begin{aligned} w &= \sum_i w_i \cdot (\boldsymbol{\delta}[A_i] + \boldsymbol{\delta}[B_i]) \\ &= \sum_i w_i \cdot \boldsymbol{\delta}[A_i] + \sum_i w_i \cdot \boldsymbol{\delta}[B_i] \end{aligned}$$

Evaluons l'intégrale de w sur $A \cup B$. Il vient :

$$\begin{aligned} \int_{A \cup B} w(x)\ d\mu(x) &= \sum_i w_i \cdot \mu(A_i) + \sum_i w_i \cdot \mu(B_i) \\ &= \int_A w(x)\ d\mu(x) + \int_B w(x)\ d\mu(x) \end{aligned}$$

L'intégrale des fonctions étagées est donc additive sur les ensembles concernés.

70.13 Propriétés diverses

Soit $u, v \in \mathrm{Etagee}(A)$. On peut trouver des $u_i, v_j \in \mathbb{R}$ et des $U_i, V_i \in \mathcal{T}$ tels que $\{U_1, ..., U_n\}$ et $\{V_1, ..., V_m\}$ forment des partitions de A et vérifiant :

$$u(x) = \sum_{i=1}^{n} u_i \cdot \boldsymbol{\delta}[U_i](x)$$

$$v(x) = \sum_{j=1}^{m} v_j \cdot \boldsymbol{\delta}[V_j](x)$$

pour tout $x \in A$. Nous avons vu que les intersections $W_{ij} = U_i \cup V_j \in \mathcal{T}$ forment aussi une partition de A et qu'elles possèdent les propriétés :

$$\mu(U_i) = \sum_j \mu(W_{ij})$$

$$\mu(V_j) = \sum_i \mu(W_{ij})$$

et :

$$\boldsymbol{\delta}[U_i] = \sum_j \boldsymbol{\delta}[W_{ij}]$$

$$\boldsymbol{\delta}[V_j] = \sum_i \boldsymbol{\delta}[W_{ij}]$$

Les fonctions u, v peuvent donc également s'exprimer comme :

$$u(x) = \sum_{i,j} u_i \cdot \boldsymbol{\delta}[W_{ij}](x)$$

$$v(x) = \sum_{i,j} v_j \cdot \boldsymbol{\delta}[W_{ij}](x)$$

Espace vectoriel

Soit $\alpha, \beta \in \mathbb{R}$. On voit que la fonction $w : A \mapsto \mathbb{R}$ définie par :

$$w(x) = \alpha \cdot u(x) + \beta \cdot v(x)$$

est également une fonction étagée puisque :

$$w(x) = \sum_{i,j} (\alpha \cdot u_i + \beta \cdot v_j) \cdot \boldsymbol{\delta}[W_{ij}](x)$$

Nous venons de montrer que toute combinaison linéaire de fonctions étagées est une fonction étagée. L'ensemble Etagee(A) forme donc un espace vectoriel sur $\mathbb{R}$.

Intégrale

Evaluons l'intégrale de w. On obtient :

$$\begin{aligned}
\int_A w(x)\ d\mu(x) &= \sum_{i,j} (\alpha \cdot u_i + \beta \cdot v_j) \cdot \mu(W_{ij}) \\
&= \alpha \sum_i u_i \sum_j \mu(W_{ij}) + \beta \sum_j v_j \sum_i \mu(W_{ij}) \\
&= \alpha \sum_i u_i \cdot \mu(U_i) + \beta \sum_j v_j \cdot \mu(V_j) \\
&= \alpha \int_A u(x)\ d\mu(x) + \beta \int_A v(x)\ d\mu(x)
\end{aligned}$$

On en conclut que :

$$\int_A (\alpha \cdot u(x) + \beta \cdot v(x))\ d\mu(x) = \alpha \cdot \int_A u(x)\ d\mu(x) + \beta \cdot \int_A v(x)\ d\mu(x)$$

Nous venons de prouver la linéarité de l'intégrale des fonctions étagées.

Fonctions max et min

On voit que la fonction :

$$\max\{u, v\}(x) = \sum_{i,j} \max\{u_i, v_j\} \cdot \boldsymbol{\delta}[W_{ij}](x)$$

est également une fonction étagée. Il en va de même pour :

$$\min\{u, v\}(x) = \sum_{i,j} \min\{u_i, v_j\} \cdot \boldsymbol{\delta}[W_{ij}](x)$$

70.14 Mesure nulle

Soit un ensemble A vérifiant $\mu(A) = 0$ et une fonction w étagée sur A définie par :

$$w(x) = \sum_{i=1}^{n} w_i \cdot \boldsymbol{\delta}_{A_i}(x)$$

pour tout $x \in A$. Les w_i sont bien entendu des réels et les A_i forment une partition de A. Comme les A_i sont inclus dans l'ensemble A qui est de mesure nulle, on a $\mu(A_i) = 0$. L'intégrale s'écrit donc :

$$\int_A w(x)\ d\mu(x) = \sum_{i=1}^{n} w_i \cdot \mu(A_i) = \sum_{i=1}^{n} w_i \cdot 0 = 0$$

L'intégrale d'une fonction étagée sur un ensemble de mesure nulle est nulle.

70.15 Mesurabilité et opérations

Soit la mesure $\mu : \mathcal{T} \mapsto \mathbb{R}$, l'ensemble $A \in \mathcal{T}$, une fonction mesurable $f : A \mapsto \mathbb{R}$ et une fonction étagée $w \in \mathrm{Etagee}(A)$. Considérons un choix de $w_i \in \mathbb{R}$ et de $A_i \in \mathcal{T}$ formant une partition de A et tels que :

$$w(x) = \sum_{i=1}^{n} w_i \cdot \boldsymbol{\delta}_{A_i}(x)$$

pour tout $x \in A$.

Addition

On a clairement :

$$\begin{aligned}
\{x \in A : f(x) + w(x)\ >\ a\} &= \bigcup_{i=1}^{n} \{x \in A_i : f(x) + w(x)\ >\ a\} \\
&= \bigcup_{i=1}^{n} \{x \in A_i : f(x) + w_i\ >\ a\} \\
&= \bigcup_{i=1}^{n} \{x \in A_i : f(x)\ >\ a - w_i\}
\end{aligned}$$

Comme f est mesurable, les ensembles :

$$F_i = \{x \in A_i : f(x)\ >\ a - w_i\}$$

appartiennent à $\mathcal{T}$. Comme l'union opère sur un nombre fini d'ensembles, on a :

$$\{x \in A : f(x) + w(x)\ >\ a\} = \bigcup_{i=1}^{n} F_i \in \mathcal{T}$$

On conclut de même que :

$$\{x \in A : f(x) + w(x)\ <\ a\} = \bigcup_{i=1}^{n} \{x \in A_i : f(x)\ <\ a - w_i\} \in \mathcal{T}$$

La fonction $f + w$ est donc mesurable. Il en va de même pour $-(f + w) = -f - w$.

Soustraction

Comme l'opposé d'une fonction étagée est une fonction étagée :

$$-w(x) = -\sum_{i=1}^{n} w_i \cdot \boldsymbol{\delta}_{A_i}(x) = \sum_{i=1}^{n} (-w_i) \cdot \boldsymbol{\delta}_{A_i}(x)$$

la fonction $f - w = f + (-w)$ est également mesurable. Il en va de même pour $-(f - w) = w - f$.

Chapitre 71

Décomposition en fonctions positives

71.1 Dépendances

- Chapitre ?? : Les ensembles
- Chapitre ?? : Les ordres et extréma
- Chapitre ?? : Les fonctions

71.2 Introduction

On considère une fonction mesurable $f : A \mapsto \mathbb{R}$ et les fonctions positives associées $f^+, f^- : A \mapsto \mathbb{R}$ définies par :

$$\begin{aligned} f^+(x) &= \max\{f(x), 0\} \ge 0 \\ f^-(x) &= \max\{-f(x), 0\} \ge 0 \end{aligned}$$

71.3 Mesurabilité

On constate que les ensembles suivant sont dans la tribu :

$$\begin{aligned} A^+ &= \{x \in A : f(x) > 0\} \in \mathcal{T} \\ A^0 &= \{x \in A : f(x) = 0\} \in \mathcal{T} \\ A^- &= \{x \in A : f(x) < 0\} \in \mathcal{T} \end{aligned}$$

On a clairement $A = A^+ \cup A^0 \cup A^-$. On voit que :

- Sur A^+, on a $f^+ = f > 0$ et $f^- = 0$
- Sur A^0, on a $f^+ = f = 0$ et $f^- = -f = 0$
- Sur A^- on a $f^+ = 0$ et $f^- = -f > 0$

Conditions $> a$

Soit un réel $a < 0$. La condition $f^+(x) \ge 0 > a$ implique que $f^+(x) > a$. Comme f^+ est positive, cette condition est satisfaite pour tout $x \in A$. Il en va de même pour f^-. Donc :

$$\{x \in A : f^+(x) > a\} = \{x \in A : f^+(x) \ge 0\} = A \in \mathcal{T}$$
$$\{x \in A : f^-(x) > a\} = \{x \in A : f^-(x) \ge 0\} = A \in \mathcal{T}$$

Soit un réel $a \ge 0$. La condition $f^+(x) > a$ implique que $f^+(x) > 0$. Un x vérifiant cette condition est donc forcément dans A^+, où $f^+ = f$. On peut alors faire sortir la condition $x \in A^+$ en utilisant l'intersection. Comme l'intersection de deux ensembles d'une tribu est dans la tribu, on a finalement :

$$\begin{aligned}\{x \in A : f^+(x) > a\} &= \{x \in A^+ : f^+(x) > a\} \\ &= \{x \in A^+ : f(x) > a\} \\ &= \{x \in A : f(x) > a\} \cap A^+ \in \mathcal{T}\end{aligned}$$

La condition $f^-(x) > a$ implique que $f^-(x) > 0$. Un x vérifiant cette condition est donc forcément dans A^-, où $f^- = -f$. On peut alors faire sortir la condition $x \in A^-$ en utilisant l'intersection. La fonction $-f$ étant mesurable, on a finalement :

$$\begin{aligned}\{x \in A : f^-(x) > a\} &= \{x \in A^- : f^-(x) > a\} \\ &= \{x \in A^- : -f(x) > a\} \\ &= \{x \in A : -f(x) > a\} \cap A^- \in \mathcal{T}\end{aligned}$$

Conditions $< a$

Soit un réel $a \le 0$. Comme $f^+ \ge 0$, la condition $f^+(x) < a \le 0$ n'est satisfaite pour aucun $x \in A$. Il en va de même pour f^-. Donc :

$$\begin{gathered}\{x \in A : f^+(x) < a\} = \emptyset \in \mathcal{T} \\ \{x \in A : f^-(x) < a\} = \emptyset \in \mathcal{T}\end{gathered}$$

Soit un réel $a > 0$. Les $x \in A$ vérifiant La condition $f^+(x) < a$ sont de deux types :
- Si $x \in A^- \cup A^0$, on a $f^+(x) = 0 < a$. La condition est alors a fortiori satisfaite.
- Sinon, $x \in A \setminus (A^- \cup A^0) = A^+$ et il faut toujours imposer $f^+(x) < a$. Par contre, on a alors $f = f^+$.

On a donc :

$$\begin{aligned}\{x \in A : f^+(x) < a\} &= (A^- \cup A^0) \cup \{x \in A^+ : f^+(x) > a\} \\ &= A^- \cup A^0 \cup \{x \in A^+ : f(x) > a\} \in \mathcal{T}\end{aligned}$$

Les $x \in A$ vérifiant La condition $f^-(x) < a$ sont de deux types :
- Si $x \in A^+ \cup A^0$, on a $f^-(x) = 0 < a$. La condition est alors a fortiori satisfaite.
- Sinon, $x \in A \setminus (A^+ \cup A^0) = A^-$ et il faut toujours imposer $f^-(x) < a$. Par contre, on a alors $f = -f^-$.

On a donc :

$$\begin{aligned}\{x \in A : f^-(x) < a\} &= (A^+ \cup A^0) \cup \{x \in A^- : f^-(x) > a\} \\ &= A^- \cup A^0 \cup \{x \in A^- : -f(x) > a\} \in \mathcal{T}\end{aligned}$$

On conclut de ce qui précède que les fonctions f^+ et f^- sont mesurables.

Chapitre 72

Intégrales

Dépendances

— Chapitre **??** : Les fonctions
— Chapitre 71 : Les mesures

72.1 Introduction

Soit une fonction $f : A \mapsto \mathbb{R}$ mesurable et essentiellement positive ($f \gtrsim 0$) :

$$\mu(\{x : f(x) \ < \ 0\}) = 0$$

Pour toute fonction étagée $w : A \mapsto \mathbb{R}$, on sait que la fonction $w - f$ est mesurable. Le réel :

$$\mu(\{x \in A : w(x) \ > \ f(x)\}) = \mu(\{x \in A : w(x) - f(x) \ > \ 0\})$$

est donc bien défini. On peut dès lors introduire l'ensemble des fonctions étagées w essentiellement inférieures à f :

$$\mathcal{E}_A(f) = \{w \in \mathrm{Etagee}(A) : w \lesssim f\}$$

Evaluant les intégrales des fonctions étagées appartenant à $\mathcal{E}_A(f)$, on obtient l'ensemble de réels associé :

$$\mathcal{I}_A(f) = \left\{ \int_A w(x) \ d\mu(x) \ : w \in \mathcal{E}_A(f) \right\}$$

Cet ensemble est non vide car $0 \in \mathcal{E}_A(f)$ produit une intégrale $0 \in \mathcal{I}_A(f)$. Deux cas peuvent se présenter :

— L'ensemble de réels $\mathcal{I}_A(f)$ est majoré. Il admet donc un supremum. Nous pouvons alors définir l'intégrale de f sur A par :

$$\int_A f(x) \ d\mu(x) = \sup \left\{ \int_A w(x) \ d\mu(x) \ : w \in \mathcal{E}_A(f) \right\}$$

On dit alors que f est intégrable sur A. On le note parfois :

$$\int_A f(x) \ d\mu(x) \ < \ +\infty$$

— Dans le cas contraire, on a $\sup \mathcal{I}_A(f) = +\infty$. On définit par analogie :

$$\int_A f(x) \ d\mu(x) = +\infty$$

Consistance

Les propriétés extrémales des intégrales des fonctions étagées nous montrent que, dans le cas particulier où $f \in \mathrm{Etagee}(A)$, l'intégrale obtenue par le supremum est identique à l'intégrale au sens des fonctions étagées.

72.2 Signe quelconque

Soit une fonction mesurable $f : A \mapsto \mathbb{R}$. On peut décomposer f en fonctions positives $f^+, f^- : A \mapsto \mathbb{R}$ définies par :

$$\begin{aligned} f^+(x) &= \max\{f(x), 0\} \ge 0 \\ f^-(x) &= \max\{-f(x), 0\} \ge 0 \end{aligned}$$

pour tout $x \in A$. On a alors :

$$f = f^+ - f^-$$

Les fonctions f^+ et f^- étant également mesurables, leurs intégrales sont bien définies (finies ou infinies). Plusieurs cas peuvent se présenter :

— Les fonctions f^+ et f^- sont intégrables sur A. On étend alors la définition de l'intégrale par :

$$\int_A f(x)\ d\mu(x) = \int_A f^+(x)\ d\mu(x) - \int_A f^-(x)\ d\mu(x)$$

On dit alors que f est intégrable.

— La fonction f^+ produit une intégrale infinie, mais f^- est intégrable sur A. On définit alors :

$$\int_A f(x)\ d\mu(x) = +\infty$$

— La fonction f^- produit une intégrale infinie, mais f^+ est intégrable sur A. On définit alors :

$$\int_A f(x)\ d\mu(x) = -\infty$$

— Si aucune des fonction f^+, f^- n'est intégrable, l'intégrale de f n'est pas définie.

72.3 Intégrale de Riemann

Soit une fonction f mesurable et essentiellement positive. Posons :

$$\mathcal{F}_A(f) = \{w \in \mathrm{Etagee}(A) : w \gtrsim f\}$$

et :

$$\mathcal{J}_A(f) = \left\{ \int_A w(x)\ d\mu(x)\ \ : w \in \mathcal{F}_A(f) \right\}$$

Si f est intégrable et que son intégrale vérifie la dualité :

$$\int_A f(x)\ d\mu(x) = \sup \mathcal{I}_A(f) = \inf \mathcal{J}_A(f)$$

on dit que f est intégrable au sens de Riemann.

72.4 Fonctions essentiellement positives

Soit une fonction f mesurable et essentiellement positive.

Signe de l'intégrale

Comme la fonction étagée nulle $0 \in \mathcal{E}_A(f)$, on a :

$$\int_A f(x)\ d\mu(x) \ge \int_A 0\ d\mu(x) = 0$$

par définition du supremum. L'intégrale d'une fonction essentiellement positive est positive.

Intégrale nulle

Supposons que :

$$\int_A f(x)\ d\mu(x) = 0$$

Soit le réel $\epsilon > 0$. Si il existait un ensemble $C \subseteq A$ de mesure non nulle tel que $f > \epsilon$ sur C, la fonction étagée $w = \epsilon \cdot \boldsymbol{\delta}_C$ vérifierait $w \lesssim f$ et appartiendrait donc à $\mathcal{E}_A(f)$. On aurait alors par définition du supremum :

$$\int_A f(x)\ d\mu(x) \ge \int_A \epsilon \cdot \boldsymbol{\delta}_C(x)\ d\mu(x) = \epsilon \cdot \mu(C)\ >\ 0$$

ce qui contredit l'hypothèse. On doit donc avoir $f \lesssim \epsilon$ quel que soit $\epsilon > 0$. On en conclut que $f \lesssim 0$. Mais comme on a également $f \gtrsim 0$, on en déduit que $f \approx 0$. Une fonction essentiellement positive présentant une intégrale nulle est essentiellement nulle.

72.5 Ordre

Fonctions positives

Soit les fonctions essentiellement positives et intégrables $f, g : A \to \mathbb{R}$ telles que $f \lesssim g$. Soit $w \in \mathcal{E}_A(f)$. Comme on a $w \lesssim f$ et $f \lesssim g$, on en conclut que $w \lesssim g$, c'est-à-dire que $w \in \mathcal{E}_A(g)$. On a donc $\mathcal{E}_A(f) \subseteq \mathcal{E}_A(g)$. On en déduit que $\mathcal{I}_A(f) \subseteq \mathcal{I}_A(g)$, d'où :

$$\sup \mathcal{I}_A(f) \le \sup \mathcal{I}_A(g)$$

Comme ces supremums sont par définition égaux aux intégrales correspondantes, on a :

$$\int_A f(x)\ d\mu(x) \le \int_A g(x)\ d\mu(x)$$

Généralisation

On considère à présent le cas de fonctions f, g de signe quelconque, toujours avec $f \lesssim g$. On définit les fonctions positives :

$$\begin{aligned} f^+(x) &= \max\{f(x), 0\} \\ f^-(x) &= \max\{-f(x), 0\} \\ g^+(x) &= \max\{g(x), 0\} \\ g^-(x) &= \max\{-g(x), 0\} \end{aligned}$$

On a donc $f^+ \lessapprox g^+$ et $g^- \lessapprox f^-$. On en conclut que :

$$\int_A f^+(x)\ d\mu(x) - \int_A f^-(x)\ d\mu(x) \le \int_A g^+(x)\ d\mu(x) - \int_A g^-(x)\ d\mu(x)$$

d'où, par définition :

$$\int_A f(x)\ d\mu(x) \le \int_A g(x)\ d\mu(x)$$

L'ordre des intégrales correspond à l'ordre essentiel sur les fonctions.

Egalité

Soit deux fonctions intégrables f, g vérifiant $f \approx g$. On a $f \lessapprox g$ et l'intégrale de f est inférieure à l'intégrale de g. Mais on a aussi $f \gtrapprox g$ et l'intégrale de f est supérieure à l'intégrale de g. On en conclut que :

$$\int_A f(x)\ d\mu(x) = \int_A g(x)\ d\mu(x)$$

72.6 Valeur moyenne

La valeur moyenne d'une fonction f sur A est définie par :

$$M = \frac{1}{\mu(A)} \int_A f(x)\ d\mu(x)$$

72.7 Bornes

Soit une fonction intégrable $f : A \mapsto \mathbb{R}$.

— Si f est essentiellement majorée, on a :

$$f \lessapprox \text{ess sup}\{f(x) : x \in A\}$$

En intégrant sur A, il vient :

$$\int_A f(x)\ d\mu(x) \le \mu(A) \cdot \operatorname*{ess\,sup}_{x \in A} f(x)$$

— Si f est essentiellement minorée, on a :

$$f \gtrapprox \text{ess inf}\{f(x) : x \in A\}$$

En intégrant sur A, il vient :

$$\int_A f(x)\ d\mu(x) \ge \mu(A) \cdot \operatorname*{ess\,inf}_{x \in A} f(x)$$

— Si f est majorée et minorée, et si A est de mesure non nulle, on peut réexprimer ces bornes en terme de valeur moyenne de f :

$$\operatorname*{ess\,inf}_{x \in A} f(x) \le \frac{1}{\mu(A)} \int_A f(x)\ d\mu(x) \le \operatorname*{ess\,sup}_{x \in A} f(x)$$

Corollaires

Si f est essentiellement nulle, les supremum et infimum essentiels s'annulent, et :

$$0 \le \int_A f(x)\ d\mu(x) \le 0$$

On en conclut que :

$$\int_A f(x)\ d\mu(x) = 0$$

Il en va de même si $\mu(A) = 0$.

72.8 Opposé

Soit une fonction intégrable f et $g = -f$. On a :

$$\begin{aligned} g^+(x) &= \max\{g(x), 0\} = \max\{-f(x), 0\} = f^-(x) \\ g^-(x) &= \max\{-g(x), 0\} = \max\{f(x), 0\} = f^+(x) \end{aligned}$$

et :

$$\int_A g(x)\ d\mu(x) = \int_A f^-(x)\ d\mu(x) - \int_A f^+(x)\ d\mu(x) = -\int_A f(x)\ d\mu(x)$$

L'intégrale de la fonction opposée est donc l'opposé de l'intégrale :

$$\int_A (-f(x))\ d\mu(x) = -\int_A f(x)\ d\mu(x)$$

72.9 Multiplication par une fonction indicatrice

Fonctions positives

Soit une fonction f essentiellement positive et intégrable sur A et l'ensemble $C \in \mathcal{T}$ vérifiant $C \subseteq A$. Soit le réel $\epsilon > 0$.

— Comme le supremum est dans l'adhérence, on peut trouver une fonction $u \in \mathcal{E}_A(f \cdot \boldsymbol{\delta}_C)$ telle que :

$$\left| \int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x) - \int_A u(x)\ d\mu(x) \right| \le \epsilon$$

On a donc :

$$\int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x) - \epsilon \le \int_A u(x)\ d\mu(x)$$

Les ensembles $A \setminus C$ et C étant disjoints et d'union égale à A, on a aussi :

$$\int_A u(x)\ d\mu(x) = \int_{A \setminus C} u(x)\ d\mu(x) + \int_C u(x)\ d\mu(x)$$

Sur $A \setminus C$, on a $u \lesssim f \cdot \boldsymbol{\delta}_C = 0$. La fonction u y est donc essentiellement négative. Son intégrale y est donc négative et :

$$\int_A u(x)\ d\mu(x) \le \int_C u(x)\ d\mu(x)$$

La fonction étagée u étant essentiellement inférieure à f sur C, on a aussi :

$$\int_C u(x)\ d\mu(x) \le \int_C f(x)\ d\mu(x)$$

Rassemblant ces résultats, on obtient :

$$\int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x) - \epsilon \le \int_C f(x)\ d\mu(x)$$

Cette inégalité étant satisfaite quel que soit $\epsilon > 0$, on a :

$$\int_C f(x)\ d\mu(x) \ge \int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x)$$

— Comme le supremum est dans l'adhérence, on peut trouver un $w \in \mathcal{E}_C(f)$ tel que :

$$\int_C f(x)\ d\mu(x) - \epsilon \le \int_C w(x)\ d\mu(x)$$

Considérons l'extension de w sur A définie par :

$$u = \begin{cases} w & \text{sur } C \\ 0 & \text{sur } A \setminus C \end{cases}$$

On voit que u est une fonction étagée. On a :

$$\int_A u(x)\ d\mu(x) = \int_{A\setminus C} u(x)\ d\mu(x) + \int_C u(x)\ d\mu(x)$$

Sur $A \setminus C$, u est nulle et l'intégrale correspondante l'est aussi. Sur C, on a $u = w$. Donc :

$$\int_A u(x)\ d\mu(x) = 0 + \int_C u(x)\ d\mu(x) = \int_C w(x)\ d\mu(x)$$

Comme :

$$u \lesssim \begin{cases} f & \text{sur } C \\ 0 & \text{sur } A \setminus C \end{cases}$$

on voit que $u \in \mathcal{E}_A(f \cdot \boldsymbol{\delta}_C)$ et que :

$$\int_A u(x)\ d\mu(x) \le \int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x)$$

par définition du supremum. En rassemblant ces résultats, on obtient :

$$\int_C f(x)\ d\mu(x) - \epsilon \le \int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x)$$

Ce résultat étant vérifié pour tout $\epsilon > 0$, on a finalement :

$$\int_C f(x)\ d\mu(x) \le \int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x)$$

L'intégrale de f sur C étant à la fois inférieure et supérieure à l'intégrale de $f \cdot \boldsymbol{\delta}_C$ sur A, elle doit lui être égale :

$$\int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x) = \int_C f(x)\ d\mu(x)$$

Généralisation

Soit une fonction f mesurable et sa décomposition en fonctions positives $f = f^+ - f^-$. On voit que :

$$(f^+ - f^-) \cdot \boldsymbol{\delta}_C = f^+ \cdot \boldsymbol{\delta}_C - f^- \cdot \boldsymbol{\delta}_C$$

Comme on a aussi :

$$\begin{aligned}(f \cdot \boldsymbol{\delta}_C)^+ &= f^+ \cdot \boldsymbol{\delta}_C \\ (f \cdot \boldsymbol{\delta}_C)^- &= f^- \cdot \boldsymbol{\delta}_C\end{aligned}$$

on en conclut que :

$$\begin{aligned}\int_A f(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x) &= \int_A f^+(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x) - \int_A f^-(x) \cdot \boldsymbol{\delta}_C(x)\ d\mu(x) \\ &= \int_C f^+(x)\ d\mu(x) - \int_C f^-(x)\ d\mu(x) \\ &= \int_C f(x)\ d\mu(x)\end{aligned}$$

72.10 Additivité

Fonctions positives

Soit deux ensembles A et B disjoints ($A \cap B = \emptyset$) et une fonction f essentiellement positive et mesurable sur $A \cup B$. Comme A et B sont disjoints, on a $\boldsymbol{\delta}[A \cap B] = 0$ et :

$$1 = \boldsymbol{\delta}[A \cup B](x) = \boldsymbol{\delta}_A(x) + \boldsymbol{\delta}_B(x)$$

pour tout $x \in A \cup B$. Soit le réel $\epsilon > 0$.

— Comme le supremum est dans l'adhérence, on peut trouver des fonctions $u \in \mathcal{E}_{A\cup B}(f \cdot \boldsymbol{\delta}_A)$ et $v \in \mathcal{E}_{A\cup B}(f \cdot \boldsymbol{\delta}_B)$ telles que leurs intégrales soit situées à une distance inférieure à ϵ des intégrales de $f \cdot \boldsymbol{\delta}_A$ et $f \cdot \boldsymbol{\delta}_B$. Comme $A, B \subseteq A \cup B$, on a :

$$\begin{aligned}\int_{A\cup B} u(x)\ d\mu(x) &\geq \int_{A\cup B} f(x) \cdot \boldsymbol{\delta}_A(x)\ d\mu(x) - \epsilon \\ &\geq \int_A f(x)\ d\mu(x) - \epsilon \\ \int_{A\cup B} v(x)\ d\mu(x) &\geq \int_{A\cup B} f(x) \cdot \boldsymbol{\delta}_B(x)\ d\mu(x) - \epsilon \\ &\geq \int_B f(x)\ d\mu(x) - \epsilon\end{aligned}$$

Par linéarité des intégrales des fonctions étagées, on a :

$$\begin{aligned}\int_{A\cup B} [u(x) + v(x)]\ d\mu(x) &= \int_{A\cup B} u(x)\ d\mu(x) + \int_{A\cup B} v(x)\ d\mu(x) \\ &\geq \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x) - 2\epsilon\end{aligned}$$

On voit que :

$$u + v \lesssim f \cdot \boldsymbol{\delta}_A + f \cdot \boldsymbol{\delta}_B = f \cdot (\boldsymbol{\delta}_A + \boldsymbol{\delta}_B) = f$$

La fonction $u+v$ est donc essentiellement inférieure à f et on a $u+v \in \mathcal{E}_{A\cup B}(f)$. La définition du supremum nous montre alors que :

$$\int_{A\cup B} f(x)\ d\mu(x) \ge \int_{A\cup B} [u(x) + v(x)]\ d\mu(x)$$

En rassemblant ces résultats, on arrive à :

$$\int_{A\cup B} f(x)\ d\mu(x) \ge \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x) - 2\epsilon$$

Ce résultat étant valable pour tout $\epsilon > 0$ et aucune grandeur ne dépendant de ϵ, on a finalement :

$$\int_{A\cup B} f(x)\ d\mu(x) \ge \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x)$$

— Le supremum étant dans l'adhérence, on peut trouver une fonction $w \in \mathcal{E}_{A\cup B}(f)$ telle que :

$$\int_{A\cup B} f(x)\ d\mu(x) - \epsilon \le \int_{A\cup B} w(x)\ d\mu(x)$$

Posons $u = w \cdot \boldsymbol{\delta}_A$ et $v = w \cdot \boldsymbol{\delta}_B$. On a :

$$w = w \cdot (\boldsymbol{\delta}_A + \boldsymbol{\delta}_B) = w \cdot \boldsymbol{\delta}_A + w \cdot \boldsymbol{\delta}_B = u + v$$

La multiplication par une fonction indicatrice conservant le caractère étagé d'une fonction, u et v sont étagées. On a :

$$u = \begin{cases} w & \text{sur } A \\ 0 & \text{sur } B \end{cases} \lesssim \begin{cases} f & \text{sur } A \\ 0 & \text{sur } B \end{cases}$$

et :

$$v = \begin{cases} 0 & \text{sur } A \\ w & \text{sur } B \end{cases} \lesssim \begin{cases} 0 & \text{sur } A \\ f & \text{sur } B \end{cases}$$

On en conclut que u est essentiellement inférieure à $f \cdot \boldsymbol{\delta}_A$ et que v est essentiellement inférieure à $f \cdot \boldsymbol{\delta}_B$. On a $u \in \mathcal{E}_{A\cup B}(f \cdot \boldsymbol{\delta}_A)$ et $v \in \mathcal{E}_{A\cup B}(f \cdot \boldsymbol{\delta}_B)$. On peut donc écrire :

$$\int_{A\cup B} u(x)\ d\mu(x) \le \int_{A\cup B} f(x) \cdot \boldsymbol{\delta}_A(x)\ d\mu(x)$$
$$\int_{A\cup B} v(x)\ d\mu(x) \le \int_{A\cup B} f(x) \cdot \boldsymbol{\delta}_B(x)\ d\mu(x)$$

inégalités que l'on peut réécrire sous la forme :

$$\int_{A\cup B} u(x)\ d\mu(x) \le \int_A f(x)\ d\mu(x)$$
$$\int_{A\cup B} v(x)\ d\mu(x) \le \int_B f(x)\ d\mu(x)$$

En utilisant la linéarité des intégrales des fonctions étagées, on a :

$$\begin{aligned}\int_{A\cup B} f(x)\ d\mu(x) - \epsilon &\le \int_{A\cup B} w(x)\ d\mu(x)\\ &\le \int_{A\cup B} [u(x)+v(x)]\ d\mu(x)\\ &\le \int_{A\cup B} u(x)\ d\mu(x) + \int_{A\cup B} v(x)\ d\mu(x)\\ &\le \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x)\end{aligned}$$

Comme cette inégalité doit être valable pour tout $\epsilon > 0$, on a finalement :

$$\int_{A\cup B} f(x)\ d\mu(x) \le \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x)$$

L'intégrale sur $A \cup B$ devant être à la fois supérieure et inférieure à la somme de l'intégrale sur A et de l'intégrale sur B, on en conclut que :

$$\int_{A\cup B} f(x)\ d\mu(x) = \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x)$$

Généralisation

Nous allons étendre le résultat précédent aux fonctions mesurables quelconques. Soit la fonction mesurable f. En utilisant la décomposition en fonctions positives $f = f^+ - f^-$, on obtient :

$$\begin{aligned}\int_{A\cup B} f(x)\ d\mu(x) &= \int_{A\cup B} f^+(x) - \int_{A\cup B} f^-(x)\ d\mu(x)\\ &= \int_A f^+ + \int_B f^+ - \int_A f^- - \int_B f^-\\ &= \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x)\end{aligned}$$

Corollaire

Si $C \subseteq A$, les ensembles $A \setminus C$ et C sont disjoints et d'union égale à A. On a donc :

$$\int_A f(x)\ d\mu(x) = \int_{A\setminus C} f(x)\ d\mu(x) + \int_C f(x)\ d\mu(x) \ge \int_C f(x)\ d\mu(x)$$

Invariance

Soit un sous-ensemble essentiel C de A. On a alors $C = A \setminus Z$ avec $Z \subseteq A$ et $\mu(Z) = 0$. Les ensembles C et $Z = A \setminus C$ sont disjoints et d'union égale à A. Comme l'intégrale de f sur l'ensemble de mesure nulle Z est nulle, on a :

$$\begin{aligned}\int_A f(x)\ d\mu(x) &= \int_C f(x)\ d\mu(x) + \int_Z f(x)\ d\mu(x)\\ &= \int_C f(x)\ d\mu(x) + 0\\ &= \int_C f(x)\ d\mu(x)\end{aligned}$$

L'intégrale est invariante lorsqu'on ajoute ou s'abstrait d'un ensemble de mesure nulle.

72.11 Union

Soit à présent $A, B \in \mathcal{T}$ quelconques et une fonction intégrable $f : A \cup B \mapsto \mathbb{R}$. Comme $A \setminus B$ et B sont disjoints et d'union égale à $A \cup B$, on a :

$$\int_{A\cup B} f(x)\ d\mu(x) = \int_{A\setminus B} f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x)$$

On peut également décomposer A en les ensembles disjoints $A \setminus B$ et $A \cap B$. Donc :

$$\int_A f(x)\ d\mu(x) = \int_{A\setminus B} f(x)\ d\mu(x) + \int_{A\cap B} f(x)\ d\mu(x)$$

et :

$$\int_{A\setminus B} f(x)\ d\mu(x) = \int_A f(x)\ d\mu(x) - \int_{A\cap B} f(x)\ d\mu(x)$$

En injectant cette relation dans l'expression de l'intégrale sur $A \cup B$, on obtient :

$$\int_{A\cup B} f(x)\ d\mu(x) = \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x) - \int_{A\cap B} f(x)\ d\mu(x)$$

Corollaire

Si l'ensemble $A \cap B$ vérifie $\mu(A \cap B) = 0$, l'intégrale correspondante est nulle et :

$$\int_{A\cup B} f(x)\ d\mu(x) = \int_A f(x)\ d\mu(x) + \int_B f(x)\ d\mu(x)$$

Il suffit donc d'avoir une intersection de mesure nulle (et pas forcément vide) pour vérifier l'additivité.

72.12 Convergence des mesures

Soit une fonction $f : A \mapsto \mathbb{R}$ essentiellement positive et intégrable. On a :

$$I = \int_A f(x)\ d\mu(x)\ <\ +\infty$$

Posons :

$$C(\alpha) = \{x \in A : f(x) \le \alpha\}$$
$$Z(\alpha) = \{x \in A : f(x)\ >\ \alpha\}$$

pour tout $\alpha \in \mathbb{R}$. Il est clair que $A = C(\alpha) \cup Z(\alpha)$. On a :

$$\int_{Z(\alpha)} f(x)\ d\mu(x) \le \int_{C(\alpha)} f(x)\ d\mu(x) + \int_{Z(\alpha)} f(x)\ d\mu(x) = I$$

Mais comme $f \gtrsim \alpha$ sur $Z(\alpha)$, on a :

$$\operatorname*{ess\,inf}_{x\in Z(\alpha)} f(x) \ge \alpha$$

et :

$$\mu(Z(\alpha)) \cdot \alpha \le \mu(Z(\alpha)) \cdot \operatorname*{ess\,inf}_{x\in Z(\alpha)} f(x) \le \int_{Z(\alpha)} f(x)\ d\mu(x)$$

Rassemblant ces résultats, il vient :

$$\mu(Z(\alpha)) \cdot \alpha \le I$$

Soit $\epsilon > 0$. Si on choisit $\alpha \ge I/\epsilon$, on a :

$$\mu(Z(\alpha)) \le \frac{I}{\alpha} \le \epsilon$$

On en déduit que :

$$\lim_{\alpha \to +\infty} \mu(Z(\alpha)) = \lim_{\alpha \to +\infty} \mu(\{x \in A : f(x) > \alpha\}) = 0$$

72.13 Convergence des intégrales

Soit une fonction $f : A \mapsto \mathbb{R}$ essentiellement positive et intégrable. On a :

$$I = \int_A f(x) \, d\mu(x) < +\infty$$

Soit la collection $\{C(\alpha) \in \mathfrak{P}(A) : \alpha \in \mathbb{R}\}$ et la collection associée $\{Z(\alpha) \in \mathfrak{P}(A) : \alpha \in \mathbb{R}\}$ définie par :

$$Z(\alpha) = A \setminus C(\alpha)$$

Nous supposons que :

$$\lim_{\alpha \to \infty} \mu(Z(\alpha)) = 0$$

et que $Z(\alpha) \subseteq Z(\beta)$ pour tout $\alpha, \beta \in \mathbb{R}$ vérifiant $\alpha \ge \beta$.

Soit le réel $\epsilon > 0$. Comme le supremum est dans l'adhérence, on peut trouver un $w \in \mathcal{E}_A(f)$ tel que :

$$\int_A f(x) \, d\mu(x) \le \int_A w(x) \, d\mu(x) + \epsilon$$

Comme $A = C(\alpha) \cup Z(\alpha)$, on a :

$$\int_A f(x) \, d\mu(x) = \int_{C(\alpha)} f(x) \, d\mu(x) + \int_{Z(\alpha)} f(x) \, d\mu(x)$$

$$\int_A w(x) \, d\mu(x) = \int_{C(\alpha)} w(x) \, d\mu(x) + \int_{Z(\alpha)} w(x) \, d\mu(x)$$

pour tout $\alpha \in \mathbb{R}$. On en déduit que :

$$\int_{Z(\alpha)} f(x) \, d\mu(x) \le \int_{C(\alpha)} w(x) \, d\mu(x) - \int_{C(\alpha)} f(x) \, d\mu(x) + \int_{Z(\alpha)} w(x) \, d\mu(x) + \epsilon$$

Soit la fonction $I : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$I(\alpha) = \int_{Z(\alpha)} f(x) \, d\mu(x) \ge 0$$

On sait que $w \lesssim f$. Par conséquent, $\int w \le \int f$ sur $C(\alpha)$ et :

$$I(\alpha) \le \int_{Z(\alpha)} w(x) \, d\mu(x) + \epsilon$$

Comme w est une fonction étagée, son supremum existe :

$$M = \operatorname*{ess\ sup}_{x \in A} w(x)$$

De plus, on a $C(\alpha) \subseteq A$ pour tout $\alpha \in \mathbb{R}$, d'où :

$$S(\alpha) = \operatorname*{ess\ sup}_{x \in C(\alpha)} w(x) \le M$$

On a donc :

$$\int_{Z(\alpha)} w(x)\ d\mu(x) \le \mu(Z(\alpha)) \cdot S(\alpha) \le \mu(Z(\alpha)) \cdot M$$

et :

$$I(\alpha) \le \mu(Z(\alpha)) \cdot M + \epsilon$$

pour tout $\alpha \in \mathbb{R}$. Si $\alpha \ge \beta$, on a $Z(\alpha) \subseteq Z(\beta)$, d'où $I(\alpha) \le I(\beta)$. La fonction I est décroissante et minorée par zéro. Elle admet donc une limite à l'infini. Comme M ne dépend pas de α, on a :

$$\lim_{\alpha \to +\infty} [\mu(Z(\alpha)) \cdot M + \epsilon] = 0 \cdot M + \epsilon = \epsilon$$

La limite de cette expression existe donc, et $\limsup = \lim$. On en conclut que :

$$0 \le \lim_{\alpha \to +\infty} I(\alpha) \le \epsilon$$

Ce résultat devant être satisfait pour tout $\epsilon > 0$, on a :

$$\lim_{\alpha \to +\infty} I(\alpha) = 0$$

c'est-à-dire :

$$\lim_{\alpha \to +\infty} \int_{Z(\alpha)} f(x)\ d\mu(x) = 0$$

Corollaire

On en déduit que :

$$\begin{aligned} \int_A f(x)\ d\mu(x) &= \lim_{\alpha \to +\infty} \int_{C(\alpha)} f(x)\ d\mu(x) + \lim_{\alpha \to +\infty} \int_{Z(\alpha)} f(x)\ d\mu(x) \\ &= \lim_{\alpha \to +\infty} \int_{C(\alpha)} f(x)\ d\mu(x) + 0 \end{aligned}$$

d'où :

$$\lim_{\alpha \to +\infty} \int_{C(\alpha)} f(x)\ d\mu(x) = \int_A f(x)\ d\mu(x)$$

Généralisation

Soit f est une fonction intégrable de signe quelconque et sa décomposition en fonctions positives $f = f^+ - f^-$. On a :

$$\lim_{\alpha \to +\infty} \int_{C(\alpha)} f^+(x)\ d\mu(x) = \int_A f^+(x)\ d\mu(x)$$

$$\lim_{\alpha \to +\infty} \int_{C(\alpha)} f^-(x)\ d\mu(x) = \int_A f^-(x)\ d\mu(x)$$

On en déduit que :

$$\begin{aligned}
\lim_{\alpha \to +\infty} \int_{C(\alpha)} f(x)\ d\mu(x) &= \lim_{\alpha \to +\infty} \int_{C(\alpha)} f^+ - \lim_{\alpha \to +\infty} \int_{C(\alpha)} f^- \\
&= \int_A f^+ - \int_A f^- \\
&= \int_A f(x)\ d\mu(x)
\end{aligned}$$

72.14 Adhérence

Soit un ensemble A vérifiant $\mu(A) > 0$ et une fonction $f : A \mapsto \mathbb{R}$ mesurable, essentiellement positive et majorée :

$$M = \operatorname*{ess\ sup}_{x \in A} f(x) < +\infty$$

Comme $\mu(A) > 0$ et $f \gtrsim 0$, on a :

$$M \ge \operatorname*{ess\ inf}_{x \in A} f(x) \ge 0$$

Soit :

$$\Theta = \left\{ \operatorname*{ess\ sup}_{x \in A} [f(x) - w(x)] : w \in \mathcal{E}_A(f) \right\}$$

et :

$$\lambda = \inf_{w \in \mathcal{E}_A(f)} \operatorname*{ess\ sup}_{x \in A} [f(x) - w(x)]$$

Comme la fonction nulle $0 \in \mathcal{E}_A(f)$, on a :

$$\lambda \le \operatorname*{ess\ sup}_{x \in A} [f(x) - 0] = M$$

Choisissons $w \in \mathcal{E}_A(f)$. On a $w \lesssim f$, d'où $f - w \gtrsim 0$ et :

$$\operatorname*{ess\ sup}_{x \in A} [f(x) - w(x)] \ge \operatorname*{ess\ inf}_{x \in A} [f(x) - w(x)] \ge 0$$

Donc $\Theta \ge 0$ et $\lambda = \inf \Theta \ge 0$.

Si $M = 0$, on a également $\lambda \le 0$, d'où $\lambda = 0$. Supposons à présent que $M > 0$. Nous allons construire une suite $\{w_n : n \in \mathbb{N}\} \subseteq \mathcal{E}_A(f)$ convergeant vers f au sens du supremum essentiel. Posons $w_0 = 0 \in \mathcal{E}_A(f)$, le réel $\Delta_0 = M/2$ et l'ensemble :

$$A_0 = \{x \in A : f(x) - w_0(x) > \Delta_0\}$$

On construit ensuite la fonction w_1 par :

$$w_1 = w_0 + \Delta_0 \cdot \boldsymbol{\delta}[A_0] = \begin{cases} w_0 + \Delta_0 & \text{sur } A_0 \\ w_0 & \text{sur } A \setminus A_0 \end{cases}$$

La fonction $\Delta_0 \cdot \boldsymbol{\delta}[A_0]$ étant étagée, w_1 l'est aussi. En utilisant $w_0 \lesssim f$ sur A et $w_0 + \Delta_0 < f$ sur A_0, on obtient :

$$w_1 = \begin{cases} w_0 + \Delta_0 & \text{sur } A_0 \\ w_0 & \text{sur } A \setminus A_0 \end{cases} \lesssim \begin{cases} f & \text{sur } A_0 \\ f & \text{sur } A \setminus A_0 \end{cases}$$

On en conclut que $w_1 \lesssim f$ et que $w_1 \in \mathcal{E}_A(f)$. Sur $A \setminus A_0$, on a par définition $f - w_1 = f - w_0 \le \Delta_0$. Sur A_0, on a :

$$f - w_1 = f - w_0 - \Delta_0 \le M - \frac{M}{2} = \frac{M}{2}$$

On en conclut que :

$$\operatorname*{ess\,sup}_{x \in A} [f(x) - w_1(x)] \le \frac{M}{2} = \Delta_0$$

Supposons à présent être arrivé à l'étape $n - 1$ avec la fonction étagée w_{n-1} vérifiant $w_{n-1} \lesssim f$ et :

$$\operatorname*{ess\,sup}_{x \in A} [f(x) - w_{n-1}(x)] \le \Delta_{n-1}$$

On construit l'étape n par :

$$\begin{aligned} \Delta_n &= \Delta_{n-1}/2 \\ A_n &= \{x \in A : f(x) - w_{n-1}(x) > \Delta_n\} \\ w_n &= w_{n-1} + \Delta_n \cdot \boldsymbol{\delta}[A_n] \end{aligned}$$

La fonction $\Delta_n \cdot \boldsymbol{\delta}[A_n]$ étant étagée, w_n l'est aussi. En utilisant $w_{n-1} \lesssim f$ sur A et $w_{n-1} + \Delta_n < f$ sur A_n, on obtient :

$$w_n = \begin{cases} w_{n-1} + \Delta_n & \text{sur } A_n \\ w_{n-1} & \text{sur } A \setminus A_n \end{cases} \lesssim \begin{cases} f & \text{sur } A_n \\ f & \text{sur } A \setminus A_n \end{cases}$$

On en conclut que $w_n \lesssim f$ et que $w_n \in \mathcal{E}_A(f)$. Sur $A \setminus A_n$, on a $f - w_n = f - w_{n-1} \le \Delta_n$. Sur A_n, on a :

$$f - w_n = f - w_{n-1} - \Delta_n \le \Delta_{n-1} - \Delta_n = 2\Delta_n - \Delta_n = \Delta_n$$

On a donc :

$$\operatorname*{ess\,sup}_{x \in A} [f(x) - w_n(x)] \le \Delta_n$$

Comme :

$$\Delta_n = \frac{\Delta_{n-1}}{2} = ... = \frac{\Delta_0}{2^{n-1}} = \frac{M}{2^n}$$

On voit que :

$$0 \le \operatorname*{ess\,sup}_{x \in A} [f(x) - w_n(x)] \le \frac{M}{2^n}$$

Comme :

$$\limsup_{n\to\infty} \frac{M}{2^n} = \liminf_{n\to\infty} 0 = 0$$

la limite de $\operatorname{ess\,sup}_{x\in A}[f(x) - w_n(x)]$ existe et :

$$\lim_{n\to\infty} \operatorname*{ess\,sup}_{x\in A}[f(x) - w_n(x)] = 0$$

Pour tout réel $\epsilon > 0$, on peut donc trouver un n tel que :

$$\operatorname*{ess\,sup}_{x\in A}[f(x) - w_m(x)] \le \epsilon$$

pour tout $m \in \mathbb{N}$ vérifiant $m \ge n$. Par conséquent, λ ne peut pas être strictement positif et :

$$\inf_{w\in\mathcal{E}_A(f)} \operatorname*{ess\,sup}_{x\in A}[f(x) - w(x)] = 0$$

72.15 Multiplication par un réel

Soit une fonction intégrable f essentiellement positive et le réel $\lambda > 0$. Si w est une fonction étagée, on voit que $w \le f \Leftrightarrow \lambda \cdot w \le \lambda \cdot f$. On peut donc associer une fonction $\lambda \cdot w \in \mathcal{E}_A(\lambda \cdot f)$ à toute fonction $w \in \mathcal{E}_A(f)$, et réciproquement. Par linéarité de l'intégrale des fonctions étagées, on en conclut que :

$$\int_A \lambda \cdot w(x)\ d\mu(x) = \lambda \cdot \int_A w(x)\ d\mu(x)$$

En passant au supremum sur $w \in \mathcal{E}_A(f) \Leftrightarrow \lambda \cdot w \in \mathcal{E}_A(\lambda \cdot f)$, on obtient donc :

$$\int_A \lambda \cdot f(x)\ d\mu(x) = \lambda \cdot \int_A f(x)\ d\mu(x)$$

Le même résultat est bien évidemment vérifié lorsque $\lambda = 0$. Lorsque $\lambda < 0$, on pose $\alpha = -\lambda > 0$. On a alors :

$$\begin{aligned}
\int_A \lambda \cdot f(x)\ d\mu(x) &= \int_A -\alpha \cdot f(x)\ d\mu(x) \\
&= -\int_A \alpha \cdot f(x)\ d\mu(x) \\
&= -\alpha \int_A f(x)\ d\mu(x) \\
&= \lambda \int_A f(x)\ d\mu(x)
\end{aligned}$$

Nous avons donc prouvé la relation pour tout $\lambda \in \mathbb{R}$.

72.16 Addition

Positives et majorées

Soit deux fonctions $f, g : A \mapsto \mathbb{R}$ intégrables, essentiellement positives et majorées :

$$\begin{aligned}
\operatorname{ess\,sup}\{f(x) : x \in A\} &< +\infty \\
\operatorname{ess\,sup}\{g(x) : x \in A\} &< +\infty
\end{aligned}$$

Soit le réel $\epsilon > 0$.

— Le supremum étant dans l'adhérence, on peut trouver un $u \in \mathcal{E}_A(f)$ tel que :

$$\int_A u(x)\ d\mu(x) \ge \int_A f(x)\ d\mu(x) - \epsilon$$

et un $v \in \mathcal{E}_A(g)$ tel que :

$$\int_A v(x)\ d\mu(x) \ge \int_A g(x)\ d\mu(x) - \epsilon$$

On sait que $u + v$ est une fonction étagée. Les conditions $u \lesssim f$ et $v \lesssim g$ impliquent que $u+v \lesssim f+g$. On a donc $u+v \in \mathcal{E}_A(f+g)$. En utilisant la linéarité des intégrales de fonctions étagées et la définition du supremum, il vient :

$$\begin{aligned}\int_A u(x)\ d\mu(x) + \int_A v(x)\ d\mu(x) &= \int_A [u(x) + v(x)]\ d\mu(x) \\ &\le \int_A [f(x) + g(x)]\ d\mu(x)\end{aligned}$$

En rassemblant ces résultats, on obtient :

$$\int_A f(x)\ d\mu(x) + \int_A g(x)\ d\mu(x) - 2\epsilon \le \int_A [f(x) + g(x)]\ d\mu(x)$$

Cette relation étant valable quel que soit $\epsilon > 0$, on a :

$$\int_A f(x)\ d\mu(x) + \int_A g(x)\ d\mu(x) \le \int_A [f(x) + g(x)]\ d\mu(x)$$

— Comme f et g sont essentiellement majorées, on peut trouver un $u \in \mathcal{E}_A(f)$ et un $v \in \mathcal{E}_A(g)$ vérifiant :

$$\text{ess sup}\{f(x) - u(x) : x \in A\} \le \epsilon$$
$$\text{ess sup}\{g(x) - v(x) : x \in A)\} \le \epsilon$$

Donc $f \lesssim u + \epsilon$ et $g \lesssim v + \epsilon$. On en conclut que $f + g \lesssim u + v + 2\epsilon$, d'où :

$$\begin{aligned}\int_A [f(x) + g(x)]\ d\mu(x) &\le \int_A [u(x) + v(x) + 2\epsilon]\ d\mu(x) \\ &\le \int_A u(x)\ d\mu(x) + \int_A v(x)\ d\mu(x) + 2\epsilon \cdot \mu(A)\end{aligned}$$

par linéarité des intégrales de fonctions étagées. On a aussi, par définition du supremum :

$$\int_A u(x)\ d\mu(x) \le \int_A f(x)\ d\mu(x)$$

$$\int_A v(x)\ d\mu(x) \le \int_A g(x)\ d\mu(x)$$

On a finalement la borne :

$$\int_A [f(x) + g(x)]\ d\mu(x) \le \int_A f(x)\ d\mu(x) + \int_A g(x)\ d\mu(x) + 2\epsilon \cdot \mu(A)$$

Cette inégalité devant être satisfaite pour tout $\epsilon > 0$, on en déduit que :

$$\int_A [f(x) + g(x)]\ d\mu(x) \le \int_A f(x)\ d\mu(x) + \int_A g(x)\ d\mu(x)$$

L'intégrale de $f+g$ devant être simultanément supérieure et inférieure à la somme de l'intégrale de f et de l'intégrale de g, on en conclut que :

$$\int_A [f(x)+g(x)]\ d\mu(x) = \int_A f(x)\ d\mu(x) + \int_A g(x)\ d\mu(x)$$

Positives

Soit deux fonctions f, g intégrables et essentiellement positives. Posons :

$$F(\alpha) = \{x \in A : f(x) \le \alpha\}$$
$$G(\alpha) = \{x \in A : g(x) \le \alpha\}$$

et :

$$N(\alpha) = \{x \in A : f(x)\ >\ \alpha\} = A \setminus F(\alpha)$$
$$T(\alpha) = \{x \in A : g(x)\ >\ \alpha\} = A \setminus G(\alpha)$$

pour tout $\alpha \in \mathbb{R}$. Pour tous réels α, β vérifiant $\alpha \ge \beta$, il est clair que $N(\alpha) \subseteq N(\beta)$ et $T(\alpha) \subseteq T(\beta)$. Soit :

$$C(\alpha) = F(\alpha) \cap G(\alpha) = A \setminus (N(\alpha) \cup T(\alpha))$$

et :

$$Z(\alpha) = A \setminus C(\alpha) = N(\alpha) \cup T(\alpha)$$

Choisissons des réels α, β tels que $\alpha \ge \beta$ et un $x \in Z(\alpha)$. On a soit $x \in N(\alpha)$, d'où $x \in N(\beta)$ et $x \in Z(\beta)$, soit $x \in T(\alpha)$, d'où $x \in T(\beta)$ et $x \in Z(\beta)$. On en conclut que $Z(\alpha) \subseteq Z(\beta)$. On voit que :

$$0 \le \mu(Z(\alpha)) \le \mu(N(\alpha)) + \mu(T(\alpha))$$

Comme :

$$\limsup_{\alpha \to +\infty} [\mu(N(\alpha)) + \mu(T(\alpha))] = \liminf_{\alpha \to +\infty} 0 = 0$$

la fonction $\alpha \mapsto \mu(Z(\alpha))$ converge et :

$$\lim_{\alpha \to +\infty} \mu(Z(\alpha)) = 0$$

Si φ est intégrable, par exemple $\varphi \in \{f, g, f+g\}$, on a donc :

$$\lim_{\alpha \to +\infty} \int_{C(\alpha)} \varphi(x)\ d\mu(x) = \int_A \varphi(x)\ d\mu(x)$$

Sur $C(\alpha) \subseteq F(\alpha)$, on a $0 \lesssim f \lesssim \alpha$. Le supremum essentiel existe par conséquent sur $C(\alpha)$ et :

$$0 \le \operatorname{ess\ sup}\{f(x) : x \in C(\alpha)\} \le \alpha$$

On conclut de même que :

$$0 \le \operatorname{ess\ sup}\{g(x) : x \in C(\alpha)\} \le \alpha$$

Les fonctions f et g sont donc essentiellement majorées sur $C(\alpha)$ et on a :

$$\int_{C(\alpha)} [f(x)+g(x)]\ d\mu(x) = \int_{C(\alpha)} f(x)\ d\mu(x) + \int_{C(\alpha)} g(x)\ d\mu(x)$$

Passant à la limite pour $\alpha \to +\infty$, on obtient alors :

$$\int_A [f(x) + g(x)] \ d\mu(x) = \int_A f(x) \ d\mu(x) + \int_A g(x) \ d\mu(x)$$

Soustraction

Soit f, g deux fonctions intégrables essentiellement positives vérifiant $h = f - g \gtrsim 0$. Comme $f = g + h$, on a :

$$\int_A f(x) \ d\mu(x) = \int_A g(x) \ d\mu(x) + \int_A h(x) \ d\mu(x)$$

On en déduit que :

$$\int_A h(x) \ d\mu(x) = \int_A f(x) \ d\mu(x) - \int_A g(x) \ d\mu(x)$$

c'est-à-dire :

$$\int_A [f(x) - g(x)] \ d\mu(x) = \int_A f(x) \ d\mu(x) - \int_A g(x) \ d\mu(x)$$

Généralisation

Soit deux fonctions intégrables f, g et leurs décompositions en fonctions positives $f = f^+ - f^-$ et $g = g^+ - g^-$. Posons :

$$\begin{aligned} s^+ &= f^+ + g^+ \\ s^- &= f^- + g^- \end{aligned}$$

On a :

$$f + g = f^+ - f^- + g^+ - g^- = (f^+ + g^+) - (f^- + g^-) = s^+ - s^-$$

Décomposons A en les ensembles :

$$\begin{aligned} P &= \{x \in A : s^+(x) - s^-(x) \ge 0\} \\ M &= \{x \in A : s^+(x) - s^-(x) \ < \ 0\} \end{aligned}$$

Comme $s^+ - s^- \gtrsim 0$ sur P, on a :

$$\int_P [s^+(x) - s^-(x)] \ d\mu(x) = \int_P s^+(x) \ d\mu(x) - \int_P s^-(x) \ d\mu(x)$$

Comme $s^+ - s^- \lesssim 0$ sur M, on y a $s^- - s^+ \gtrsim 0$ et :

$$\begin{aligned} \int_M [s^+(x) - s^-(x)] \ d\mu(x) &= -\int_M [s^-(x) - s^+(x)] \ d\mu(x) \\ &= -\left[\int_M s^-(x) \ d\mu(x) - \int_M s^+(x) \ d\mu(x)\right] \\ &= \int_M s^+(x) \ d\mu(x) - \int_M s^-(x) \ d\mu(x) \end{aligned}$$

Rassemblant ces résultats, il vient :

$$\begin{aligned}
\int_A [f(x) + g(x)] \ d\mu(x) &= \int_A [s^+(x) - s^-(x)] \ d\mu(x) \\
&= \int_P [s^+(x) - s^-(x)] \ d\mu(x) + \int_M [s^+(x) - s^-(x)] \ d\mu(x) \\
&= \int_P s^+ - \int_P s^- + \int_M s^+ - \int_M s^- \\
&= \int_A s^+ - \int_A s^- \\
&= \int_A [f^+(x) + g^+(x)] \ d\mu(x) - \int_A [f^-(x) + g^-(x)] \ d\mu(x) \\
&= \int_A f^+ + \int_A g^+ - \int_A f^- - \int_A g^- \\
&= \int_A f(x) \ d\mu(x) + \int_A g(x) \ d\mu(x)
\end{aligned}$$

L'intégrale d'une somme est la somme des intégrales.

72.17 Linéarité

Soit les réels α, β et les fonctions intégrables $f, g : A \mapsto \mathbb{R}$. On a :

$$\begin{aligned}
\int_A [\alpha \cdot f(x) + \beta \cdot g(x)] \ d\mu(x) &= \int_A \alpha \cdot f(x) \ d\mu(x) + \int_A \beta \cdot g(x) \ d\mu(x) \\
&= \alpha \cdot \int_A f(x) \ d\mu(x) + \beta \cdot \int_A g(x) \ d\mu(x)
\end{aligned}$$

L'intégrale est linéaire.

Chapitre 73

Intégrales et mesures

73.1 Addition de mesure

Soit les mesures γ, λ et la mesure μ définie par :

$$\mu = \gamma + \lambda$$

Soit un ensemble A et une fonction $w \in \mathrm{Etagee}(A)$. On peut trouver des $w_i \in \mathbb{R}$ et des $A_i \in \mathcal{T}$ formant une partition de A et tels que :

$$w(x) = \sum_{i=1}^{n} w_i \cdot \boldsymbol{\delta}_{A_i}(x)$$

On a alors :

$$\begin{aligned}
\int_A w(x)\ d\mu(x) &= \sum_{i=1}^{n} w_i \cdot \mu(A_i) \\
&= \sum_{i=1}^{n} w_i \cdot (\gamma(A_i) + \lambda(A_i)) \\
&= \sum_{i=1}^{n} w_i \cdot \gamma(A_i) + \sum_{i=1}^{n} w_i \cdot \lambda(A_i) \\
&= \int_A w(x)\ d\gamma(x) + \int_A w(x)\ d\lambda(x)
\end{aligned}$$

En passant au suprémum sur toutes les fonctions étagées w essentiellement inférieures à une fonction positive intégrable, puis en utilisant la définition d'une fonction intégrable f on en déduit que :

$$\int_A f(x)\ d(\gamma + \lambda)(x) = \int_A f(x)\ d\gamma(x) + \int_A f(x)\ d\lambda(x)$$

73.2 Mesure de Lebesgue

Soit $A \subseteq \mathbb{R}^n$ et μ_L la mesure de Lebesgue. On note alors $d\mu_L(x) = dx$ ou :

$$d\mu_L(x) = dx = dx_1\ ...\ dx_n$$

et :

$$\int_A f(x)\ dx = \int_A f(x)\ d\mu_L(x)$$

73.3 Intégrale de Stieltjes

Soit la fonction croissante $\gamma : A \mapsto \mathbb{R}$. L'intégrale de Stieltjes associée à γ se note :

$$\int_A f(x)\ d\gamma(x)$$

Elle se définit d'après la mesure de Stieltjes μ_γ associée à γ :

$$\int_A f(x)\ d\gamma(x) = \int_A f(x)\ d\mu_\gamma(x)$$

73.3.1 Fonction à variation bornée

Nous allons étendre la définition à des fonctions non nécessairement croissantes. Soit une fonction $g : [\alpha, \beta] \mapsto \mathbb{R}$ et $S[a, b]$ l'ensemble des suites réelles croissantes inclues dans $[a, b]$:

$$S[a, b] = \Big\{ \{x_0, x_1, x_2, ..., x_n\} : a \le x_0 \le x_1 \le x_2 \le ... \le b,\ n \in \mathbb{N} \Big\}$$

Nous supposons que la fonction g nous permet de définir les fonctions σ, λ associées par :

$$\sigma(x) = \sup \left\{ \sum_{i=0}^{n} \max\{g(x_{i+1}) - g(x_i), 0\} : \{x_0, x_1, ..., x_n\} \in S[\alpha, x] \right\}$$

et :

$$\lambda(x) = -\inf \left\{ \sum_{i=0}^{n} \min\{g(x_{i+1}) - g(x_i), 0\} : \{x_0, x_1, ..., x_n\} \in S[\alpha, x] \right\}$$

pour tout $x \in [\alpha, \beta]$. Il est clair que les fonctions σ, λ sont croissantes. On peut donc définir l'intégrale de Stieltjes associée à g par :

$$\int_A f(x)\ dg(x) = \int_A f(x)\ d\sigma(x) - \int_A f(x)\ d\lambda(x)$$

73.3.2 Positivité

L'intégrale de Stieltjes d'une fonction essentiellement positive n'est généralement pas positive.

73.4 Mesure pondérée

Soit une fonction essentiellement positive φ telle que la fonction μ associée définie par :

$$\mu(A) = \int_A \varphi(x)\ dx$$

soit une mesure. Considérons une fonction $w \in \mathcal{E}_A(f)$, avec :

$$w = \sum_i w_i \cdot \boldsymbol{\delta}_{A_i}$$

où les A_i forment une partition de A et où les $w_i \in \mathbb{R}$. On a alors :

$$\int_A w(x)\ d\mu(x) = \sum_i w_i \cdot \int_{A_i} \varphi(x)\ dx$$

Par linéarité de l'intégrale, on voit aussi que :

$$\begin{aligned}\int_A w(x) \cdot \varphi(x) \ dx &= \sum_i w_i \cdot \int_A \varphi(x) \cdot \boldsymbol{\delta}_{A_i}(x) \ dx \\ &= \sum_i w_i \cdot \int_{A_i} \varphi(x) \ dx\end{aligned}$$

On en conclut que :

$$\int_A w(x) \ d\mu(x) = \int_A w(x) \cdot \varphi(x) \ dx$$

Comme φ est positive, on a l'équivalence entre $w \lesssim f$ et $w \cdot \varphi \lesssim f \cdot \varphi$. En passant au supremum, on obtient donc :

$$\int_A f(x) \ d\mu(x) = \int_A f(x) \cdot \varphi(x) \ dx$$

pour toute fonction intégrable $f : A \to \mathbb{R}$. On note symboliquement :

$$d\mu(x) = \varphi(x) \ dx$$

Problème inverse

Remarquons qu'on ne sait généralement pas faire correspondre une fonction φ à une mesure μ donnée.

73.5 Mesure de Dirac

L'intégrale d'une fonction $f : \mathbb{R} \to \mathbb{R}$ par rapport à une mesure de Dirac μ_D^a est fort simple à calculer. En effet :

$$\int_A d\mu_D^a(x) = \mu_D^a(A) = \boldsymbol{\delta}_A(a)$$

par définition. Si $a \notin A$, l'ensemble A est de mesure nulle et :

$$\int_A f(x) \ d\mu_D^a(x) = 0$$

Considérons à présent le cas où $a \in A$. L'ensemble $A \setminus \{a\}$ étant disjoint de $\{a\}$ et de mesure nulle :

$$\int_{A \setminus \{a\}} d\mu_D^a(x) = \boldsymbol{\delta}_{A \setminus \{a\}}(a) = 0$$

on a :

$$\int_A f(x) \ d\mu_D^a(x) = \int_{\{a\}} f(x) \ d\mu_D^a(x)$$

La fonction f étant constante sur $\{a\}$ et valant $f(a)$, on a finalement :

$$\int_A f(x) \ d\mu_D^a(x) = f(a) \cdot \boldsymbol{\delta}_{\{a\}}(a) = f(a)$$

Dans le cas général, on a donc :

$$\int_A f(x)\ d\mu_D^a(x) = f(a) \cdot \boldsymbol{\delta}_A(a)$$

On note également :

$$\int_A f(x)\ \boldsymbol{\delta}(x-a)\ dx = f(a) \cdot \boldsymbol{\delta}_A(a)$$

73.6 Pondérée - Dirac

Soit une mesure μ définie par :

$$d\mu(x) = \left[\varphi(x) + \sum_i \alpha_i \cdot \boldsymbol{\delta}(x - x_i)\right]\ dx$$

où les $\alpha_i \in \mathbb{R}$. Si les x_i appartiennent tous à A, l'intégrale d'une fonction f par rapport à cette mesure s'écrit :

$$\int_A f(x)\ d\mu(x) = \int_A f(x) \cdot \varphi(x)\ dx + \sum_i \alpha_i \cdot f(x_i)$$

Chapitre 74

Intégrales unidimensionnelles

74.1 Introduction

Soit une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ et $a, b \in \mathbb{R}$ tels que $a \le b$.

Intervalle

On définit la notation particulière :

$$\int_a^b f(x) \ d\mu(x) = \int_{[a,b]} f(x) \ d\mu(x)$$

Intervalle inversé

On étend la notation à des « intervalles » inversés par :

$$\int_b^a f(x) \ d\mu(x) = -\int_a^b f(x) \ d\mu(x)$$

Intervalle ouvert

Pour les intervalles ouverts, on considère la limite :

$$\int_{]a,b[} f(x) \ d\mu(x) = \lim_{(s,t)\to(a,b)} \int_s^t f(x) \ d\mu(x)$$

Infini

Quand une des bornes de l'intervalle tend vers l'infini, on définit l'intégrale par :

$$\int_a^{+\infty} f(x) \ d\mu(x) = \lim_{b\to+\infty} \int_a^b f(x) \ d\mu(x)$$
$$\int_{-\infty}^b f(x) \ d\mu(x) = \lim_{a\to-\infty} \int_a^b f(x) \ d\mu(x)$$

Sur l'ensemble des réels

Enfin, l'intégrale sur $\mathbb{R}$ entier est définie par :

$$\int_{-\infty}^{+\infty} f(x) \ d\mu(x) = \int_{\mathbb{R}} f(x) \ d\mu(x) = \lim_{a\to-\infty} \int_{-a}^a f(x) \ d\mu(x)$$

74.2 Additivité d'intervalles

Soit $a, b, c, \alpha, \beta \in \mathbb{R}$ avec $\alpha \le a, b, c \le \beta$. Soit une fonction intégrable $f : [\alpha, \beta] \mapsto \mathbb{R}$. Posons :

$$I(x, y) = \int_x^y f(\xi)\ d\xi$$

où $d\xi = d\mu_L(\xi)$ est la mesure de Lebesgue. Par définition, on a :

$$I(y, x) = -I(x, y)$$

Si $a \le b \le c$, on a :

$$\mu_L([a, b] \cap [b, c]) = \mu_L(\{b\}) = 0$$

L'additivité nous donne alors :

$$I(a, c) = I(a, b) + I(b, c)$$

Si $a \le c \le b$, on a :

$$\mu_L([a, c] \cap [c, b]) = \mu_L(\{c\}) = 0$$

et :

$$I(a, b) = I(a, c) + I(c, b)$$

On en déduit que :

$$I(a, c) = I(a, b) - I(c, b) = I(a, b) + I(b, c)$$

On vérifie pareillement, en considérant tous les cas, que $I(a, b) = I(a, c) + I(c, b)$ quel que soit l'ordre des réels a, b, c. On a donc :

$$\int_a^c f(x)\ dx = \int_a^b f(x)\ dx + \int_b^c f(x)\ dx$$

74.3 Valeur moyenne d'une intégrale

Soit la fonction intégrable et continue $f : [a, b] \mapsto \mathbb{R}$. Pour la mesure de Lebesgue, on a :

$$\begin{aligned} \text{ess sup}\{f(x) : x \in [a, b]\} &= \max f([a, b]) \\ \text{ess inf}\{f(x) : x \in [a, b]\} &= \min f([a, b]) \end{aligned}$$

Les bornes d'une fonction continue sur un intervalle fermé étant atteintes, on peut trouver des réels $\sigma, \lambda \in [a, b]$ tels que :

$$\begin{aligned} f(\sigma) &= \max f([a, b]) \\ f(\lambda) &= \min f([a, b]) \end{aligned}$$

Les bornes de l'intégrales nous disent que :

$$f(\lambda) \le \frac{1}{b-a} \int_a^b f(s) ds \le f(\sigma)$$

Le théorème des valeurs intermédiaires nous permet alors de trouver un c compris entre λ et σ (et donc dans $[a, b]$) tel que :

$$f(c) = \frac{1}{b-a} \int_a^b f(s) ds$$

Chapitre 75

Les espaces fonctionnels

Dépendances

- Chapitre 71 : Les normes
- Chapitre 60 : Les produits scalaires
- Chapitre 72 : Les intégrales

75.1 Fonctions intégrables

Soit une fonction $f : A \mapsto \mathbb{R}$ et sa décomposition en fonctions positives $f = f^+ - f^-$. Soit $|f| : A \mapsto \mathbb{R}$ définie par :

$$|f|(x) = |f(x)|$$

pour tout $x \in A$. Si $f(x) \geq 0$, on a $f^-(x) = 0$ et :

$$|f(x)| = f(x) = f^+(x) = f^+(x) + f^-(x)$$

Si $f(x) < 0$, on a $f^+(x) = 0$ et :

$$|f(x)| = -f(x) = f^-(x) = f^+(x) + f^-(x)$$

On en conclut que $|f| = f^+ + f^-$. Si f est intégrable, f^+ et f^- le sont aussi et :

$$\int_A |f(x)|\ d\mu(x) = \int_A f^+(x)\ d\mu(x) + \int_A f^-(x)\ d\mu(x) < +\infty$$

La fonction $|f|$ est donc également intégrable. Inversément, si $|f|$ est intégrable, on a $f^+ = |f| - f^- \leq |f|$ et $f^- = |f| - f^+ \leq |f|$, d'où :

$$\int_A f^+(x)\ d\mu(x) \leq \int_A |f(x)|\ d\mu(x) < +\infty$$

$$\int_A f^-(x)\ d\mu(x) \leq \int_A |f(x)|\ d\mu(x) < +\infty$$

La fonction f est donc également intégrable. On en conclut que l'on peut représenter l'ensemble des fonctions intégrables par :

$$\mathrm{Leb}(A) = \left\{ f \in \mathbb{R}^A : \int_A |f(x)|\ d\mu(x) < +\infty \right\}$$

75.2 Intégrale complexe

Soit une fonction à valeurs complexes $u : A \mapsto \mathbb{C}$. On définit les fonctions à valeurs réelles associées $v, w : A \mapsto \mathbb{R}$ par :

$$\phi(x) = \Re(u(x))$$
$$\psi(x) = \Im(u(x))$$

L'intégrale de la fonction u est alors définie par :

$$\int_A u(x)\ d\mu(x) = \int_A \phi(x)\ d\mu(x) + \mathbf{i} \int_A \psi(x)\ d\mu(x)$$

75.3 Produit scalaire

Par analogie avec le produit scalaire sur $\mathbb{K}^n$:

$$\langle x \mid y \rangle = \sum_i \overline{x}_i \cdot y_i \equiv \sum_i \overline{x}(i) \cdot y(i)$$

on définit le produit scalaire entre deux fonctions $u, v : A \mapsto \mathbb{C}$ par :

$$\langle u \mid v \rangle = \int_A \overline{u(x)} \cdot v(x)\ d\mu(x)$$

Cette application est clairement hermitienne et linéaire à droite. Comme l'intégrale d'une fonction positive est positive, on a aussi :

$$\langle u \mid u \rangle = \int_A |u(x)|^2\ d\mu(x) \ge 0$$

Il ne s'agit cependant pas tout à fait d'un produit scalaire, car la condition :

$$\langle u \mid u \rangle = \int_A |u(x)|^2\ d\mu(x) = 0$$

n'implique pas que $u = 0$ partout sur A. Par contre, l'annulation de cette intégrale implique que la fonction positive $x \mapsto |u(x)|^2$ soit essentiellement nulle. On aura donc également $u \approx 0$. L'application $\langle | \rangle$ est donc essentiellement un produit scalaire. On parlera aussi de produit scalaire au sens faible. La norme essentielle (ou norme faible) associée est :

$$\|u\| = \sqrt{\langle u \mid u \rangle} = \sqrt{\int_A |u(x)|^2\ d\mu(x)}$$

Sur quel espace ce « produit scalaire » est-il correctement défini ? Il faut que la norme associée soit finie :

$$\|u\|^2 = \langle u \mid u \rangle = \int_A |u(x)|^2\ d\mu(x)\ <\ +\infty$$

Si les normes de u et v sont finies, Cauchy-Schwartz nous garantit que :

$$\left| \int_A \overline{u(x)} \cdot v(x)\ d\mu(x) \right| \le \left[\int_A |u(x)|^2\ d\mu(x) \right] \cdot \left[\int_A |v(x)|^2\ d\mu(x) \right]\ <\ +\infty$$

Nous définissons donc notre produit scalaire fonctionnel sur l'espace :

$$\mathrm{Leb}^2(A) = \left\{ u \in \mathbb{C}^A : \int_A |u(x)|^2\ d\mu(x)\ <\ +\infty \right\}$$

nommé espace de Lebesgue de degré 2.

75.4 Espaces de Lebesgue

L'espace de Lebesgue de degré k est l'ensemble des fonctions telles que l'intégrale de la puissance k existe (et ne soit donc pas infinie) :

$$\mathrm{Leb}^k(A, B) = \left\{ u \in \mathbb{C}^A : \int_A |u(x)|^k \ d\mu(x) \ < \ +\infty \right\}$$

Norme

La norme usuelle sur cet espace est définie par :

$$\|u\|_k = \left[\int_A |u(x)|^k \ dx \right]^{1/k}$$

Il s'agit d'une norme faible.

75.5 Espace de fonctions essentiellement bornées

Par analogie avec la norme « max », on définit l'espace des fonctions essentiellement bornées par :

$$\mathrm{Leb}^\infty(A, B) = \left\{ u \in \mathbb{R}^A : \text{ess sup}\{|u(x)| : x \in A\} \ < \ +\infty \right\}$$

75.6 Noyau

On peut généraliser le produit scalaire usuel de $\mathrm{Leb}^2(A, B)$ en généralisant la notion de « matrice de produit scalaire ». On choisit une fonction $K : A^2 \to B$ appelée « noyau » et on définit le produit scalaire associé :

$$\langle u | \, K \, | v \rangle = \langle u \mid v \rangle_K = \int_{A^2} u(x) \cdot K(x, y) \cdot v(y) \ d\mu(x) \ d\mu(y)$$

Ce noyau doit bien entendu respecter certaines propriétés afin de s'assurer que $\langle | \rangle_K$ soit bien un produit scalaire.

Chapitre 76

Convergence et intégration

76.1 Convergence monotone

Soit une suite de fonctions intégrables :

$$\{u_n \in \mathbb{R}^A : n \in \mathbb{R}\}$$

Nous supposons que $u_n \gtrsim 0$ pour tout $n \in \mathbb{N}$ et que la suite soit essentiellement croissante :

$$u_0 \lesssim u_1 \lesssim u_2 \lesssim u_3 \lesssim ...$$

Nous supposons également que la fonction :

$$s = \sup\{u_n : n \in \mathbb{N}\}$$

est bien définie et intégrable :

$$S = \int_A s(x) \ d\mu(x) \ < \ +\infty$$

Convergence des fonctions

Pour tout $n \in \mathbb{N}$, on a $u_n \lesssim u_{n+1}$ et l'ensemble :

$$Z_n = \{x \in A : u_n(x) \ > \ u_{n+1}(x)\}$$

est de mesure nulle. Il en va donc de même pour leur union :

$$Z = \bigcup_n Z_n$$

L'ensemble $\Phi = A \setminus Z$ est donc un sous-ensemble essentiel de A. Soit $x \in \Phi$. On a alors :

$$u_0(x) \le u_1(x) \le u_2(x) \le ...$$

La suite $\{u_n(x) : n \in \mathbb{N}\}$ est donc croissante et majorée par $s(x) = \sup_n u_n(x)$. Elle converge dès lors vers son supremum :

$$\lim_{n \to \infty} u_n(x) = s(x)$$

Suite d'intégrales

On définit la suite $\{I_n : n \in \mathbb{N}\}$ par :

$$I_n = \int_A u_n(x) \ d\mu(x)$$

pour tout $n \in \mathbb{N}$. Comme u_n croît essentiellement avec n, on a $I_m \le I_n$ pour tout $m, n \in \mathbb{N}$ vérifiant $m \le n$. Comme $u_n \le s$, on a aussi $I_n \le S$. On en conclut que la suite des I_n converge vers son supremum et que :

$$L = \lim_{n\to\infty} I_n \le S = \int_A s(x) \ d\mu(x)$$

Fonction étagée

Soit le réel $\epsilon > 0$. Choisissons $w \in \mathcal{E}_A(s)$ telle que :

$$\int_A s(x) \ d\mu(x) \le \int_A w(x) \ d\mu(x) + \epsilon$$

Comme $w \lesssim s$ l'ensemble :

$$N = \{x \in A : w(x) > s(x)\}$$

est de mesure nulle. Posons $\Psi = \Phi \setminus N$. La fonction w étant étagée, on dispose d'une partition $\{A_1, ..., A_N\}$ de A et de réels w_i tels que :

$$w = \sum_i w_i \cdot \boldsymbol{\delta}[A_i]$$

Posons $\Psi_i = A_i \cap \Psi$. Comme $\Psi \subseteq A$, les Ψ_i forment une partition de Ψ et on a :

$$w \cdot \boldsymbol{\delta}[\Psi] = \sum_i w_i \cdot \boldsymbol{\delta}[A_i] \cdot \boldsymbol{\delta}[\Psi] = \sum_i w_i \cdot \boldsymbol{\delta}[\Psi_i]$$

On en déduit que :

$$\int_\Psi w(x) \ d\mu(x) = \int_A w(x) \cdot \boldsymbol{\delta}_\Psi(x) \ d\mu(x) = \sum_i w_i \cdot \mu(\Psi_i)$$

L'intégrale étant invariante par abstraction d'un ensemble de mesure nulle, on a :

$$\int_\Psi w(x) \ d\mu(x) = \int_\Phi w(x) \ d\mu(x) = \int_A w(x) \ d\mu(x)$$

et finalement :

$$\int_A w(x) \ d\mu(x) = \sum_i w_i \cdot \mu(\Psi_i)$$

Mesures convergentes

Choisissons un réel α vérifiant $0 < \alpha < 1$ et posons :

$$C_n = \{x \in \Psi : \alpha \cdot w(x) \le u_n(x)\}$$

Soit :

$$\begin{aligned} W_n^i &= C_n \cap \Psi_i = \{x \in \Psi_i : \alpha \cdot w(x) \le u_n(x)\} \\ X_n^i &= Z_n \cap \Psi_i = \{x \in \Psi_i : u_n(x) > u_{n+1}(x)\} \end{aligned}$$

Comme $X_n^i \subseteq Z_n$, on a $\mu(X_n^i) = 0$. Si $x \in W_n^i \setminus X_n^i$, on a :

$$\alpha \cdot w(x) \le u_n(x) \le u_{n+1}(x)$$

On en déduit que $x \in W_{n+1}^i$. Donc $W_n^i \setminus X_n^i \subseteq W_{n+1}^i$ et :

$$\lim_{n\to\infty} \mu(W_n^i) = \mu\left(\bigcup_n W_n^i\right)$$

Comme les $W_n^i \subseteq \Psi_i$, il est clair que leur union sur n est inclue dans Ψ_i. Soit $x \in \Psi_i$.

- Si $s(x) = 0$, on a $0 \le u_n(x) \le s(x) = 0$ pour tout $n \in \mathbb{N}$ et $u_n(x) = 0$. On a aussi $w(x) \le s(x) = 0$, d'où $\alpha \cdot w(x) \le 0 = u_n(x)$ et $x \in W_n^i$.
- Considérons à présent le cas où $s(x) > 0$. On se rappelle que $\Psi_i \subseteq \Psi = \Phi \setminus N$. Donc $x \notin N$ et on a $w(x) \le s(x)$. Multipliant cette relation par $\alpha > 0$, on obtient $\alpha \cdot w(x) \le \alpha \cdot s(x)$. Soit le réel :

  $$\delta = (1 - \alpha) \cdot s(x) > 0$$

  Sur $\Psi_i \subseteq \Psi \subseteq \Phi$, la suite des $\{u_n(x) : n \in \mathbb{N}\}$ converge vers $s(x)$. On peut donc trouver un $K \in \mathbb{N}$ tel que $|u_n(x) - s(x)| \le \delta$ pour tout naturel n vérifiant $n \ge K$. On a alors :

  $$u_n(x) \ge s(x) - \delta = s(x) - (1 - \alpha) \cdot s(x) = \alpha \cdot s(x) \ge \alpha \cdot w(x)$$

  Il existe donc un naturel n tel que $x \in C_n$, d'où $x \in C_n \cap \Psi_i = W_n^i$.

Notre x appartient donc à l'union sur n des W_n^i. On en conclut que Ψ_i est inclu dans l'union sur n des W_n^i. La réciproque étant également vraie, on a :

$$\bigcup_n W_n^i = \Psi_i$$

et :

$$\lim_{n\to\infty} \mu(W_n^i) = \mu(\Psi_i)$$

Convergence des intégrales

Fixons $n \in \mathbb{N}$. Comme $C_n \subseteq A$, les $W_n^i = C_n \cap \Psi_i$ forment une partition de C_n et on a :

$$w \cdot \boldsymbol{\delta}[C_n] = \sum_i w_i \cdot \boldsymbol{\delta}[C_n] \cdot \boldsymbol{\delta}[\Psi_i] = \sum_i w_i \cdot \boldsymbol{\delta}[W_n^i]$$

On en déduit que :

$$\int_{C_n} w(x)\ d\mu(x) = \int_A w(x) \cdot \boldsymbol{\delta}_{C_n}(x)\ d\mu(x) = \sum_i w_i \cdot \mu(W_n^i)$$

En passant à la limite sur n, on obtient donc :

$$\begin{aligned}
\lim_{n\to\infty} \int_{C_n} w(x)\ d\mu(x) &= \sum_i w_i \cdot \lim_{n\to\infty} \mu(W_n^i) \\
&= \sum_i w_i \cdot \mu(\Psi_i) \\
&= \int_A w(x)\ d\mu(x)
\end{aligned}$$

Comme $\alpha \cdot w \le u_n$ sur C_n, on a a fortiori l'infériorité essentielle et :

$$\int_{C_n} \alpha \cdot w(x)\ d\mu(x) = \alpha \int_{C_n} w(x)\ d\mu(x) \le \int_A u_n(x)\ d\mu(x) = I_n$$

En passant à la limite sur n et en multipliant par $1/\alpha$, on en déduit que :

$$\int_A w(x)\ d\mu(x) \le \frac{1}{\alpha} \lim_{n\to\infty} I_n = \frac{L}{\alpha}$$

Se rappelant la propriété de w par rapport à s, on a :

$$\int_A s(x)\ d\mu(x) - \epsilon \le \int_A w(x)\ d\mu(x) \le \frac{L}{\alpha}$$

Donc :

$$S - \epsilon = \int_A s(x)\ d\mu(x) - \epsilon \le \frac{L}{\alpha}$$

Ce résultat devant être satisfait pour tout $\epsilon > 0$ et pour tout $0 > \alpha > 1$, on a finalement $S \le L$. Mais nous avons vu précédemment que $L \le S$. On en conclut que $S = L$, c'est-à-dire :

$$\int_A \lim_{n\to\infty} u_n(x)\ d\mu(x) = \lim_{n\to\infty} \int_A u_n(x)\ d\mu(x)$$

76.2 Convergence des sommes

Soit une suite $\{f_n \in \mathbb{R}^A : n \in \mathbb{N}\}$ de fonctions intégrables essentiellement positives. Posons :

$$u_n = \sum_{i=0}^{n} f_i$$

On voit que la suite des u_n est croissante et essentiellement positive. Si la somme converge, on a :

$$\lim_{n\to\infty} u_n(x) = \sum_{i=0}^{+\infty} f_i(x) = s(x)$$

pour tout $x \in A$. Si la fonction s ainsi définie est intégrable, on a donc :

$$\int_A \sum_{i=0}^{+\infty} f_i(x)\ d\mu(x) = \lim_{n\to\infty} \int_A \sum_{i=0}^{n} f_i(x)\ d\mu(x)$$

76.3 Lemme de Fatou

Soit une suite $\{f_n \in \mathbb{R}^A : n \in \mathbb{N}\}$ de fonctions intégrables essentiellement positives. Posons :

$$u_n = \inf\{f_m : m \in \mathbb{N},\ m \ge n\}$$

et supposons que la fonction $s = \sup_n u_n$ soit bien définie et intégrable. Par construction, la suite des u_n est croissante et essentiellement positive. On a donc $\int_A \lim_n u_n = \lim_n \int_A u_n$, c'est-à-dire :

$$\int_A \liminf_{n\to\infty} f_n(x)\ d\mu(x) = \lim_{n\to\infty} \int_A u_n(x)\ d\mu(x)$$

Comme la limite des $U_n = \int_A u_n$ existe, on a $\liminf_n U_n = \lim_n U_n$ et :

$$\int_A \liminf_{n\to\infty} f_n(x)\ d\mu(x) = \liminf_{n\to\infty} \int_A u_n(x)\ d\mu(x)$$

Comme $u_n \le f_n$, on a $U_n \le F_n = \int_A f_n$. On en conclut que $\liminf_n U_n \le \liminf_n F_n$:

$$\liminf_{n\to\infty} \int_A u_n(x)\ d\mu(x) \le \liminf_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

Rassemblant ces résultats, il vient :

$$\int_A \liminf_{n\to\infty} f_n(x)\ d\mu(x) \le \liminf_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

76.4 Convergence dominée

Soit une suite de fonctions $\{f_n : n \in \mathbb{N}\}$ intégrables sur Ω. Nous supposons qu'il existe un sous-ensemble essentiel A de Ω et une fonction $f : A \mapsto \mathbb{R}$ telle que :

$$\lim_{n\to\infty} f_n(x) = f(x)$$

pour tout $x \in A$. Nous supposons également qu'il existe une fonction intégrable $\varphi : A \mapsto \mathbb{R}$ telle que $|f_n| \le \varphi$ pour tout $n \in \mathbb{N}$.

Soit $x \in A$. Puisque la suite $\{|f_n(x)| : n \in \mathbb{N}\}$ est inférieure à $\varphi(x)$, sa limite $|f(x)|$ vérifie $|f(x)| \le \varphi(x)$. On a donc $|f| \le \varphi$ et $\int_A |f| \le \int_A \varphi < +\infty$, ce qui montre que f est intégrable. La majoration par φ nous montre également que $\varphi - \max\{f_n, -f_n\} \ge 0$, d'où :

$$\varphi - f_n \ge 0$$
$$\varphi + f_n \ge 0$$

On peut donc appliquer le lemme de Fatou à la suite de fonctions $\psi_n = \varphi - f_n$. On obtient :

$$\int_A \liminf_{n\to\infty} [\varphi(x) - f_n(x)]\ d\mu(x) \le \liminf_{n\to\infty} \int_A [\varphi(x) - f_n(x)]\ d\mu(x)$$

On sait que la fonction φ ne dépend pas de n et que la limite de f_n existe. On a donc $\liminf_n f_n = \limsup_n f_n = \lim f_n$. Comme $\inf(-X) = -\sup(X)$, on a aussi :

$$\liminf_n (-f_n) = -\limsup_n f_n = -\lim_n f_n = -f$$

On en déduit que :

$$\int_A \liminf_n [\varphi - f_n] = \int_A \varphi - \int_A f$$

Pour le second membre, on a :

$$\liminf_n \int_A [\varphi - f_n] = \int_A \varphi + \liminf_n \left[- \int_A f_n \right] = \int_A \varphi - \limsup_n \int_A f_n$$

On se retrouve donc avec l'inégalité :

$$\int_A \varphi - \int_A f \le \int_A \varphi - \limsup_n \int_A f_n$$

Eliminant l'intégrale de φ, il vient :

$$\int_A f(x)\ d\mu(x) \ge \limsup_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

Appliquons à présent le lemme de Fatou à la suite de fonctions $\omega_n = \varphi + f_n$. On obtient :

$$\int_A \liminf_{n\to\infty} [\varphi(x) + f_n(x)]\ d\mu(x) \le \liminf_{n\to\infty} \int_A [\varphi(x) + f_n(x)]\ d\mu(x)$$

Utilisant les mêmes remarques que précédemment et éliminant l'intégrale de φ, il vient :

$$\int_A f(x)\ d\mu(x) \le \liminf_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

Rassemblant ces résultats, on en déduit que :

$$\limsup_{n\to\infty} \int_A f_n(x)\ d\mu(x) \le \liminf_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

Mais on sait que $\liminf(X) \le \limsup(X)$. On conclut de ces deux inégalités que :

$$\limsup_{n\to\infty} \int_A f_n(x)\ d\mu(x) = \liminf_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

La limite de la suite d'intégrales existe donc et :

$$\lim_{n\to\infty} \int_A f_n(x)\ d\mu(x) = \limsup_{n\to\infty} \int_A f_n(x)\ d\mu(x) = \liminf_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

Les bornes faisant intervenir l'intégrale de f deviennent :

$$\lim_{n\to\infty} \int_A f_n(x)\ d\mu(x) \le \int_A f(x)\ d\mu(x) \le \lim_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

d'où :

$$\int_A f(x)\ d\mu(x) = \lim_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

Nous venons de prouver que :

$$\int_A \lim_{n\to\infty} f_n(x)\ d\mu(x) = \lim_{n\to\infty} \int_A f_n(x)\ d\mu(x)$$

L'intégrale étant invariante sous abstraction d'un ensemble de mesure nulle, on a même :

$$\int_\Omega f(x)\ d\mu(x) = \lim_{n\to\infty} \int_\Omega f_n(x)\ d\mu(x)$$

pour autant que f soit définie sur Ω.

Chapitre 77

Additivité généralisée

Dépendances

— Chapitre ?? : Les fonctions
— Chapitre 71 : Les mesures

77.1 Introduction

Mesure induite

Soit les ensembles Ω et Ψ, les tribus $\mathcal{T} \subseteq \mathfrak{P}(\Omega)$ et $\mathcal{U} \subseteq \mathfrak{P}(\Psi)$ et les mesures $\mu : \mathcal{T} \mapsto \mathbb{R}$ et $\nu : \mathcal{U} \mapsto \mathbb{R}$.

On considère un ensemble $X \subseteq \Psi$ mesurable pour ν et paramétrant la collection $\mathcal{C} = \{P(x) \in \mathfrak{P}(\Omega) : x \in X\}$ formant une partition de Ω. Nous supposons également que chaque ensemble-élément de $\mathcal{C}$ est mesurable pour μ. Choisissons un sous-ensemble quelconque $A \subseteq \Omega$. Posons $A(x) = P(x) \cap A$ et :

$$\mathcal{P}(A) = \{A(x) : x \in X\} = \{P(x) \cap A : x \in X\}$$

On a :

$$\bigcup_{x \in X} A(x) = A \cap \bigcup_{x \in X} P(x) = A \cap \Omega = A$$

Si $x, y \in X$ vérifient $x \neq y$, on a aussi :

$$A(x) \cap A(y) = P(x) \cap P(y) \cap A = \emptyset$$

On en déduit que $\mathcal{P}(A)$ forme une partition de A. Soit la collection $\mathcal{M}$ des sous-ensembles A de Ω tels que la fonction $x \mapsto \mu(A(x))$ soit mesurable. Si $\mathcal{M}$ forme une tribu, on peut définir une mesure $\sigma : \mathcal{M} \mapsto \mathbb{R}$ par la relation :

$$\sigma(A) = \int_X \mu(A(x)) \, d\nu(x)$$

Validité

Par positivité de μ et de l'intégrale, on a clairement $\sigma \geq 0$. L'ensemble vide étant de mesure nulle au sens de μ, on a :

$$\begin{aligned}
\sigma(\emptyset) &= \int_X \mu(\emptyset \cap P(x)) \, d\nu(x) \\
&= \int_X \mu(\emptyset) \, d\nu(x) \\
&= \int_X 0 \, d\nu(x) \\
&= 0
\end{aligned}$$

Soit $A, B \subseteq \Omega$ avec $A \cap B = \emptyset$. On a :

$$\begin{aligned}
A(x) \cup B(x) &= (A \cap P(x)) \cup (B \cap P(x)) \\
&= (A \cup B) \cap P(x) \\
&= (A \cup B)(x)
\end{aligned}$$

Par additivité de μ et linéarité de l'intégrale, on a :

$$\begin{aligned}
\sigma(A \cup B) &= \int_X \mu((A \cup B)(x)) \, d\nu(x) \\
&= \int_X \mu(A(x) \cup B(x)) \, d\nu(x) \\
&= \int_X [\mu(A(x)) + \mu(B(x))] \, d\nu(x) \\
&= \int_X \mu(A(x)) \, d\nu(x) + \int_X \mu(B(x)) \, d\nu(x) \\
&= \sigma(A) + \sigma(B)
\end{aligned}$$

La fonction σ est donc également additive et représente bien une mesure.

Notation

Soit une fonction $f : A \mapsto \mathbb{R}$ et la fonction $I : X \mapsto \mathbb{R}$ définie par :

$$I(x) = \int_{A(x)} f(y) \, d\mu(y)$$

pour tout $x \in X$. On note dans la suite :

$$\int_X \int_{A(x)} f(y) \, d\mu(y) \, d\nu(x) = \int_X I(x) \, d\nu(x)$$

ou encore :

$$\int_X d\nu(x) \int_{A(x)} f(y) \, d\mu(y) = \int_X I(x) \, d\nu(x)$$

Fonctions étagées

Soit une fonction étagée $w : A \mapsto \mathbb{R}$. On dispose d'une partition $\{A_1, ..., A_n\}$ de A et de réels w_i tels que :

$$w = \sum_i w_i \cdot \boldsymbol{\delta}_{A_i}$$

Evaluons son intégrale :

$$\begin{aligned}
\int_A w(z)\ d\sigma(z) &= \sum_i w_i \cdot \sigma(A_i) \\
&= \sum_i w_i \int_X \mu(A_i(x))\ d\nu(x) \\
&= \int_X \sum_i w_i \cdot \mu(A_i(x))\ d\nu(x) \\
&= \int_X \left[\int_{A(x)} w(y)\ d\mu(y)\right]\ d\nu(x)
\end{aligned}$$

On a donc :

$$\int_A w(z)\ d\sigma(z) = \int_X\ d\nu(x) \int_{A(x)} w(y)\ d\mu(y)$$

pour toute fonction étagée.

Ordre

Soit les fonctions mesurables f, g vérifiant $f \lesssim g$ au sens de la mesure σ. Soit :

$$N = \{z \in A : f(z)\ >\ g(z)\}$$

Comme $N \subseteq A$, on a $N = N \cap A$ et :

$$\begin{aligned}
N(x) = N \cap P(x) &= N \cap A \cap P(x) \\
&= N \cap A(x) \\
&= \{z \in A(x) : f(z)\ >\ g(z)\}
\end{aligned}$$

La mesure de N étant σ-nulle, on a :

$$\sigma(N) = \int_X \mu(N(x))\ d\nu(x) = 0$$

Comme μ est positive, elle est essentiellement positive. On en conclut que la fonction $x \mapsto \mu(N(x))$ est essentiellement nulle sur X (au sens de la mesure ν). L'ensemble :

$$Z = \{x \in X : \mu(N(x))\ >\ 0\}$$

vérifie $\nu(Z) = 0$. Le sous-ensemble :

$$E = X \setminus Z = \{x \in X : \mu(N(x)) = 0\}$$

est donc ν-essentiel dans X. Soit les fonctions $F, G : X \mapsto \mathbb{R}$ définies par :

$$F(x) = \int_{A(x)} f(y)\ d\mu(y)$$

$$G(x) = \int_{A(x)} g(y)\ d\mu(y)$$

Pour tout $x \in E$, on a $\mu(N(x)) = 0$. Le sous-ensemble :

$$C(x) = A(x) \setminus N(x) = \{y \in A(x) : f(y) \le g(y)\}$$

est donc μ-essentiel dans $A(x)$. On a donc $w \lesssim f$ au sens de μ sur $A(x)$ et :

$$F(x) = \int_{A(x)} f(y)\ d\mu(y) \le \int_{A(x)} g(y)\ d\mu(y) = G(x)$$

On a donc $F \le G$ sur E qui est un sous-ensemble essentiel de X. Donc, $F \lesssim G$ au sens de ν sur X et :

$$\int_X F(x)\ d\nu(x) \le \int_X G(x)\ d\nu(x)$$

Autrement dit :

$$\int_X\ d\nu(x) \int_{A(x)} f(y)\ d\mu(y) \le \int_X\ d\nu(x) \int_{A(x)} g(y)\ d\mu(y)$$

Positives majorées

Soit une fonction intégrable $f : A \mapsto \mathbb{R}$ essentiellement positive et majorée au sens de σ :

$$\operatorname*{ess\overset{\sigma}{sup}}_{x \in A} f(x)\ <\ +\infty$$

Soit un réel $\epsilon\ >\ 0$.

— Le supremum étant dans l'adhérence, on peut trouver une fonction étagée w essentiellement inférieure à f au sens de σ et telle que :

$$\int_A f(z)\ d\sigma(z) - \epsilon \le \int_A w(z)\ d\sigma(z)$$

On a aussi :

$$\begin{aligned} \int_A w(x)\ d\sigma(x) &= \int_X\ d\nu(x) \int_{A(x)} w(y)\ d\mu(y) \\ &\le \int_X\ d\nu(x) \int_{A(x)} f(y)\ d\mu(y) \end{aligned}$$

d'où finalement :

$$\int_A f(x)\ d\sigma(x) - \epsilon \le \int_X\ d\nu(x) \int_{A(x)} f(y)\ d\mu(y)$$

Comme cette relation est valable pour tout $\epsilon\ >\ 0$, on doit avoir :

$$\int_A f(x)\ d\sigma(x) \le \int_X\ d\nu(x) \int_{A(x)} f(y)\ d\mu(y)$$

— Comme f est essentiellement majorée, on peut trouver une fonction étagée w essentiellement inférieure à f au sens de σ et telle que :

$$\operatorname*{ess\overset{\sigma}{sup}}_{x \in A} [f(x) - w(x)] \le \epsilon$$

On a donc $f \lesssim w + \epsilon$ au sens de σ et :

$$\int_X\ d\nu(x) \int_{A(x)} f(y)\ d\mu(y) \le \int_X\ d\nu(x) \int_{A(x)} [w(y) + \epsilon]\ d\mu(y)$$

On voit que :

$$\begin{aligned}
\int_X d\nu(x) \int_{A(x)} [w(y) + \epsilon] \ d\mu(y) &= \int_X d\nu(x) \left[\int_{A(x)} w(y) \ d\mu(y) + \epsilon \cdot \mu(A(x)) \right] \\
&= \int_X d\nu(x) \int_{A(x)} w(y) \ d\mu(y) + \epsilon \cdot \int_X \mu(A(x)) \ d\nu(x) \\
&= \int_X d\nu(x) \int_{A(x)} w(y) \ d\mu(y) + \epsilon \cdot \sigma(A)
\end{aligned}$$

Or :

$$\int_X d\nu(x) \int_{A(x)} w(y) \ d\mu(y) = \int_A w(z) \ d\sigma(z) \le \int_A f(z) \ d\sigma(z)$$

par définition du supremum. Rassemblant tous ces résultats, il vient :

$$\int_X d\nu(x) \int_{A(x)} f(y) \ d\mu(y) \le \int_A f(z) \ d\sigma(z) + \epsilon \cdot \sigma(A)$$

Cette inégalité devant être satisfaite pour tout $\epsilon > 0$, on en conclut que :

$$\int_X d\nu(x) \int_{A(x)} f(y) \ d\mu(y) \le \int_A f(z) \ d\sigma(z)$$

L'intégrale double $\int_X \int_{A(x)}$ devant être simultanément supérieure et inférieure à l'intégrale $\int_A$, on a :

$$\int_A f(z) \ d\sigma(z) = \int_X d\nu(x) \int_{A(x)} f(y) \ d\mu(y)$$

Positives

Posons :

$$C(\alpha) = \{x \in A : f(x) \le \alpha\}$$
$$Z(\alpha) = \{x \in A : f(x) > \alpha\}$$

et :

$$C(\alpha, x) = C(\alpha) \cap A(x) = \{x \in A(x) : f(x) \le \alpha\}$$
$$Z(\alpha, x) = A(x) \setminus C(\alpha) = \{x \in A(x) : f(x) > \alpha\}$$

Comme f est essentiellement majorée sur $C(\alpha)$, on a :

$$\int_{C(\alpha)} f(z) \ d\sigma(z) = \int_X d\nu(x) \int_{C(\alpha,x)} f(y) \ d\mu(y)$$

Les propriétés de $Z(\alpha) = A \setminus C(\alpha)$ nous montrent que :

$$\lim_{\alpha \to +\infty} \int_{C(\alpha)} f(z) \ d\sigma(z) = \int_A f(z) \ d\sigma(z)$$

On obtient bien entendu le même résultat en se restreignant aux entiers $n \in \mathbb{N}$:

$$\lim_{n \to \infty} \int_{C(n)} f(z) \ d\sigma(z) = \int_A f(z) \ d\sigma(z)$$

Comme $Z(\alpha, x)$ vérifie les mêmes propriétés, on a :

$$\lim_{n\to\infty} \int_{C(n,x)} f(y)\ d\mu(y) = \int_{A(x)} f(y)\ d\mu(y)$$

Pour tout $n \in \mathbb{N}$, posons :

$$u_n(x) = \int_{C(n,x)} f(y)\ d\mu(y)$$

Il s'agit d'une suite de fonctions positives. Comme l'inégalité $m \le n$ implique $C(m,x) \subseteq C(n,x)$, on a $u_m \le u_n$. La suite est donc croissante et converge vers :

$$\lim_{n\to\infty} u_n(x) = \int_{A(x)} f(y)\ d\mu(y)$$

La convergence monotone nous montre alors que :

$$\begin{aligned}
\lim_{n\to\infty} \int_X u_n(x)\ d\nu(x) &= \int_X \lim_{n\to\infty} u_n(x)\ d\nu(x) \\
&= \int_X d\nu(x) \int_{A(x)} f(y)\ d\mu(y)
\end{aligned}$$

D'un autre coté, on a :

$$\begin{aligned}
\lim_{n\to\infty} \int_X u_n(x)\ d\nu(x) &= \lim_{n\to\infty} \int_X d\nu(x) \int_{C(n,x)} f(y)\ d\mu(y) \\
&= \lim_{n\to\infty} \int_{C(n)} f(z)\ d\sigma(z) \\
&= \int_A f(z)\ d\sigma(z)
\end{aligned}$$

On en conclut finalement que :

$$\int_A f(z)\ d\sigma(z) = \int_X d\nu(x) \int_{A(x)} f(y)\ d\mu(y)$$

Signe quelconque

Soit une fonction intégrable $f : A \mapsto \mathbb{R}$ et sa décomposition en fonctions positives $f = f^+ - f^-$. On a :

$$\begin{aligned}
\int_A f(z)\ d\sigma(z) &= \int_A f^+(z)\ d\sigma(z) - \int_A f^-(z)\ d\sigma(z) \\
&= \int_X d\nu(x) \int_{A(x)} f^+(y)\ d\mu(y) - \int_X d\nu(x) \int_{A(x)} f^-(y)\ d\mu(y) \\
&= \int_X \left[\int_{A(x)} f^+(y)\ d\mu(y) - \int_{A(x)} f^-(y)\ d\mu(y) \right] d\nu(x) \\
&= \int_X d\nu(x) \int_{A(x)} f(y)\ d\mu(y)
\end{aligned}$$

77.2 Fubini

Sur $\mathbb{R}^2$, la mesure de Lebesgue, que nous notons ici σ, est basée sur des ensembles de la forme :

$$P = [a, b] \times [c, d]$$

où $a, b, c, d \in \mathbb{R}$ et $a \leq b$, $c \leq d$. Si μ est la mesure de Lebesgue sur $\mathbb{R}$, on a alors :

$$\sigma(P) = (\mu \otimes \mu)([a, b] \times [c, d]) = \mu([a, b]) \cdot \mu([c, d]) = (b - a) \cdot (d - c)$$

Considérons le partionnement formé des segments $[(x, c), (x, d)]$ pour tous les x compris entre a et b :

$$\mathcal{P} = \{[(x, c), (x, d)] : x \in [a, b]\}$$

Comme :

$$[(x, c), (x, d)] = \{(x, y) : y \in [c, d]\}$$

on définit la mesure μ d'un tel segment par extension de la mesure de Lebesgue :

$$\mu([(x, c), (x, d)]) = d - c$$

La mesure σ_x qui en découle s'évalue :

$$\sigma_x(A) = \int_a^b (d - c) dx = (d - c) \int_a^b dx = (d - c) \cdot (b - a)$$

Considérons le partionnement alternatif :

$$\mathcal{Q} = \{[(a, y), (b, y)] : y \in [c, d]\}$$

Comme :

$$[(a, y), (b, y)] = \{(x, y) : x \in [a, b]\}$$

on définit la mesure μ d'un tel segment par extension de la mesure de Lebesgue :

$$\mu([(a, y), (b, y)]) = b - a$$

La mesure σ_y qui en découle s'évalue :

$$\sigma_y(A) = \int_c^d (b - a) dy = (b - a) \int_c^d dy = (b - a) \cdot (d - c)$$

On en conclut que $\sigma_x = \sigma_y = \sigma$. Si f est une fonction intégrable sur $A = [\alpha, \beta] \times [\gamma, \delta] \subseteq \mathbb{R}^2$, on a alors :

$$\int_A f(x, y)\ d\sigma(x, y) = \int_\alpha^\beta dx \int_\gamma^\delta f(x, y)\ dy = \int_\gamma^\delta dy \int_\alpha^\beta f(x, y)\ dx$$

On note souvent $d\sigma(x, y) = dx\ dy$ et :

$$\int_A f(x, y)\ dx\ dy = \int_\alpha^\beta dx \int_\gamma^\delta f(x, y)\ dy = \int_\gamma^\delta dy \int_\alpha^\beta f(x, y)\ dx$$

Dimension n

Soit $A = [a_1, b_1] \times [a_2, b_2] ... \times [a_n, b_n] \subseteq \mathbb{R}^n$. On a :

$$\int_A f(x)\ dx = \int_{a_1}^{b_1} dx_1\ ... \int_{a_n}^{b_n} dx_n\ f(x_1, ...x_n)$$

où dx correspond à la mesure de Lebesgue $\sigma = \mu \otimes ... \otimes \mu$.

77.3 Produit cartésien

On peut généraliser Fubini dans certaines conditions. On a alors :

$$\int_{A\times B} f(x, y)\ d\mu(x)\ d\nu(y) = \int_B d\nu(y)\ \int_A f(x, y)\ d\mu(x)$$

et symétriquement :

$$\int_{A\times B} f(x, y)\ d\mu(x)\ d\nu(y) = \int_A\ d\mu(x) \int_B f(x, y)\ d\nu(y)$$

77.4 Domaine régulier

Soit les réels a, b et les fonctions $S, I : [a, b] \mapsto \mathbb{R}$ permettant de définir l'ensemble :

$$A = \{(x, y) \in [a, b] \times \mathbb{R} : I(x) \le y \le S(x)\}$$

Posons :

$$A(x) = \{(x, y) : I(x) \le y \le S(x)\}$$

On voit que $A(x) \cap A(z) = \emptyset$ si $x \neq z$ et que :

$$A = \bigcup_{x\in[a,b]} A(x)$$

On a donc :

$$\int_A f(x, y)\ d\sigma(x, y) = \int_a^b d\nu(x)\ \int_{I(x)}^{S(x)} f(x, y)\ d\mu(y)$$

77.5 Lemme du triangle

Un petit lemme intéressant permettant de permuter l'intégration de deux variables. Soit le triangle Δ :

$$\Delta = \{(s, t) \in \mathbb{R}^2 : 0 \le s, t \le T, \quad s \ge t\}$$

On peut redéfinir cet ensemble de deux manières équivalentes :

$$\Delta = \{(s, t) \in \mathbb{R}^2 : 0 \le s \le T, \quad 0 \le t \le s\}$$
$$\Delta = \{(s, t) \in \mathbb{R}^2 : 0 \le t \le T, \quad t \le s \le T\}$$

On a donc :

$$\int_\Delta f(x)\ dx = \int_0^T\ ds \int_0^s f(s, t)\ dt = \int_0^T\ dt \int_t^T f(s, t)\ ds$$

Cas particulier

En particulier, si la fonction à intégrer ne dépend que de t, on a :

$$\int_0^T \, ds \int_0^s u(t) \, dt = \int_0^T u(t) \, dt \int_t^T \, ds = \int_0^T (T - t) \cdot u(t) \, dt$$

Chapitre 78

Sommes et intégrales

78.1 Introduction

Soit une fonction décroissante $f : \mathbb{R} \mapsto \mathbb{R}$. Choisissons $k \in \mathbb{Z}$. Comme $f(k)$ maximise f sur $[k, k+1]$, l'intégrale vérifie :

$$\int_k^{k+1} f(x)\ dx \le \int_k^{k+1} f(k)\ dx = f(k) \cdot 1$$

Comme $f(k)$ minimise f sur $[k-1, k]$, l'intégrale vérifie :

$$f(k) = f(k) \cdot 1 = \int_{k-1}^{k} f(k)\ dx \le \int_{k-1}^{k} f(x)\ dx$$

On en déduit l'encadrement :

$$\int_k^{k+1} f(x)\ dx \le f(k) \le \int_{k-1}^{k} f(x)\ dx$$

Soit $m, n \in \mathbb{Z}$. On a :

$$\sum_{k=m}^{n} \int_k^{k+1} f(x)\ dx = \int_m^{n+1} f(x)\ dx$$

et :

$$\sum_{k=m}^{n} \int_{k-1}^{k} f(x)\ dx = \int_{m-1}^{n} f(x)\ dx$$

En sommant les inégalités sur $k \in \mathbb{Z}[m, n]$, on obtient :

$$\int_m^{n+1} f(x)\ dx \le \sum_{k=m}^{n} f(k) \le \int_{m-1}^{n} f(x)\ dx$$

78.2 Sommes infinies

Sous réserve de convergence des sommes et des intégrales, on a :

$$\int_0^{+\infty} f(x)\ dx \le \sum_{k=0}^{+\infty} f(k) \le \int_{-1}^{+\infty} f(x)\ dx$$

Septième partie

Différentielles

Chapitre 79

Différentielles

Dépendances

— Chapitre 56 : Les fonctions linéaires

79.1 Définition

Soit les espaces vectoriels Ω, F sur $\mathbb{K}$.

L'idée à la base de la notion de différentielles est de linéariser localement une fonction $f : \Omega \mapsto F$ autour d'un point $a \in \Omega$. Pour tout $h \in \Omega$ suffisamment petit, on veut donc avoir :

$$f(a+h) - f(a) \approx \mathfrak{D}^f_a(h)$$

où $\mathfrak{D}^f_a$ est une application linéaire de Ω vers F. On suppose que la norme existe de $\mathfrak{D}^f_a$ existe, de sorte que nous puissions écrire :

$$\left\|\mathfrak{D}^f_a(h)\right\| \le \left\|\mathfrak{D}^f_a\right\| \cdot \|h\|$$

On voit que la norme de la différentielle tend plus vite que h vers 0. On demande que la norme de l'erreur donnée par :

$$E(h) = f(a+h) - f(a) - \mathfrak{D}^f_a(h)$$

devienne négligeable par rapport à :

$$\left\|\mathfrak{D}^f_a\right\| \cdot \|h\|$$

lorsque h tend vers 0. Comme la norme de la différentielle ne varie pas, il nous suffit d'imposer que :

$$\lim_{\substack{h\to 0\\ h\neq 0}} \frac{\|E(h)\|}{\|h\|} = 0$$

Si ces conditions sont vérifiées, on dit que f est différentiable en a et que $\mathfrak{D}^f_a$ est la différentielle de f en a. On a alors :

$$f(a+h) - f(a) = \mathfrak{D}^f_a(h) + E(h)$$

Notation

Tout au long de ce chapitre, nous notons :

$$\lim_{h \to 0} = \lim_{\substack{h \to 0 \\ h \neq 0}}$$

$$\lim_{b \to a} = \lim_{\substack{b \to a \\ b \neq a}}$$

79.2 Continuité

Si a est différentiable en a, elle est forcément continue en a. En effet, la différentielle est bien continue :

$$\left\| \mathfrak{D}_a^f(0) \right\| \le \left\| \mathfrak{D}_a^f \right\| \cdot \|0\| = 0$$

Par définition, pour tout $\epsilon > 0$, on peut trouver $\delta > 0$ tel que :

$$\frac{\|E(h)\|}{\|h\|} \le \epsilon$$

c'est-à-dire :

$$\|E(h)\| \le \epsilon \cdot \|h\|$$

On en déduit que l'erreur est continue en $h = 0$. On a donc :

$$\lim_{h \to 0} \|E(h)\| = 0$$

L'expression de $f(a+h)$ peut donc s'écrire :

$$\lim_{h \to 0} f(a+h) = f(a) + \lim_{h \to 0} \left[\mathfrak{D}_a^f(h) + E(h) \right] = f(a)$$

79.3 Dérivées partielles

Nous allons à présent voir comment obtenir les composantes de la différentielle dans le cas d'espaces de dimensions finies. Nous disposons donc d'une base $(\varpi_1, ..., \varpi_n)$ de Ω et d'une base $(\phi_1, ..., \phi_m)$ de F. Nous introduisons les composantes $f_1, ..., f_m : \Omega \mapsto F$ de f telles que :

$$f(x) = \sum_{i=1}^{m} f_i(x) \cdot \phi_i$$

pour tout $x \in \Omega$. Nous procédons de même pour l'erreur :

$$E(h) = \sum_{i=1}^{m} E_i(h) \cdot \phi_i$$

pour tout $h \in \Omega$. Nous allons également utiliser les coordonnées $h_1, ..., h_n \in \mathbb{K}$ de h :

$$h = \sum_{i=1}^{n} h_i \cdot \varpi_i$$

Par linéarité :

$$f(a+h) - f(a) = \sum_{j=1}^{n} \mathfrak{D}_a^f(\varpi_j) \cdot h_j + E(h)$$

Mais on peut trouver des $\Delta_{ij} \in \mathbb{K}$ tels que :

$$\mathfrak{D}^f_a(\varpi_j) = \sum_{i=1}^m \Delta_{ij} \cdot \phi_i$$

Injectons les expressions des composantes dans F. On obtient :

$$\sum_i \phi_i \cdot \left[f_i(a+h) - f_i(a) - E_i(h) - \sum_{j=1}^n \Delta_{ij} \cdot h_j \right] = 0$$

Par indépendance linéaire des ϕ_i, on a alors :

$$f_i(a+h) - f_i(a) = \sum_{j=1}^n \Delta_{ij} \cdot h_j + E_i(h)$$

pour tout $i \in \mathbb{Z}(1,m)$. On nomme Δ_{ij} la dérivée partielle de f_i par rapport à ϖ_j. On la note :

$$\partial_j f_i(a) = \frac{\partial f_i}{\partial x_j}(a) = \Delta_{ij}$$

Attention

Ne pas confondre la frontière ∂A d'un ensemble A avec la dérivée ∂f d'une fonction f.

79.4 Dérivées et limites

Soit $\lambda \in \mathbb{K}$. Si l'on choisit $h = \lambda\ \varpi_j$, on a $h_k = \lambda\ \boldsymbol{\delta}_{jk}$ et :

$$\begin{aligned} f_i(a + \lambda\ \varpi_j) - f_i(a) &= \sum_{k=1}^n \partial_k f_i(a) \cdot \lambda \cdot \boldsymbol{\delta}_{jk} + E_i(h) \\ &= \lambda \cdot \partial_j f_i(a) + E_i(h) \end{aligned}$$

En divisant l'équation ci-dessus par λ puis en faisant tendre λ vers 0, on obtient :

$$\partial_j f_i(a) = \lim_{\lambda \to 0} \left[\frac{f_i(a + \lambda\ \varpi_j) - f_i(a)}{\lambda} - \frac{E_i(h)}{\lambda} \right]$$

Comme l'erreur doit converger plus vite vers zéro que la norme $\|h\| = \lambda$, la limite du second terme du membre de droite s'annule et on a :

$$\partial_j f_i(a) = \lim_{\lambda \to 0} \frac{f_i(a + \lambda\ \varpi_j) - f_i(a)}{\lambda}$$

Ce qui montre que la dérivée partielle $\partial_j f_i$ est la variation de f_i obtenue lorsqu'on fait varier la j^{me} variable (celle correspondant à ϖ_j).

79.5 Représentation matricielle

On associe à la différentielle $\mathfrak{D}^f_a$ la matrice Jacobienne $\partial f(a)$ de f en a définie par :

$$\partial f(a) = \Big(\partial_j f_i(a) \Big)_{i,j}$$

On peut alors écrire la linéarisation de f sous forme de produit matriciel :

$$f(a+h) - f(a) = \partial f(a) \cdot h + E(h)$$

où f, h, E sont les vecteurs colonnes associés aux grandeurs du même nom.

79.5.1 Vecteurs associés

On dispose de n vecteurs représentant chacun la dérivée des composantes de f par rapport à la j^{me} variable :

$$\partial_j f(a) = \Big(\partial_j f_i(a) \Big)_i$$

et de m vecteurs représentant chacun les dérivées de la i^{me} composante de f :

$$\partial f_i(a) = \Big(\partial_j f_i(a) \Big)_j$$

79.5.2 Notation

Dans le cas où la variable porte un nom par défaut, comme par exemple :

$$f : x \mapsto f(x)$$

on note aussi :

$$\frac{\partial f}{\partial x}(a) = \partial f(a)$$

Plusieurs sous-variables

Lorsque plusieurs sous variables portent un nom par défaut, comme par exemple :

$$f : (x, y) \mapsto f(x, y)$$

on note aussi :

$$\frac{\partial f}{\partial x}(a) = \partial_x f(a)$$

pour la Jacobienne par rapport aux variables $x = (x_1, ..., x_s)$ et :

$$\frac{\partial f}{\partial y}(a) = \partial_y f(a)$$

pour la Jacobienne par rapport aux variables $y = (y_1, ..., y_t)$

Symbolique

On a aussi les notations symboliques :

$$df = \frac{\partial f}{\partial x^T} \ dx$$
$$df_i = \sum_j \frac{\partial f_i}{\partial x_j} \ dx_j$$

où df représente une petite variation de f suite à une petite variation dx de x.
On utilise parfois la transposée de la Jacobienne :

$$\Big[\partial f(x) \Big]^T = \frac{\partial f^T}{\partial x} = \left(\frac{\partial f}{\partial x^T} \right)^T$$

79.5.3 Appellation

Pour des fonctions du type $f : \mathbb{R}^n \mapsto \mathbb{R}$, la Jacobienne se réduit à un vecteur matriciel. On dit alors que ∂f est le gradient de f.

79.6 Dérivées ordinaires

Dans le cas où $m = n = 1$, il n'y a qu'une dérivée partielle, $\partial_1 f_1$, que l'on appelle alors dérivée ordinaire. On a les équivalences :

$$\lambda \ \phi_1 \quad \Leftrightarrow \quad \lambda \quad \Leftrightarrow \quad \lambda \ \epsilon_1$$

On peut considérer Ω et F comme équivalents à $\mathbb{K}$ et se restreindre à des fonctions $f : \mathbb{K} \mapsto \mathbb{K}$ sans perte de généralité. La dérivée ordinaire est alors simplement :

$$\partial f(a) = \lim_{\lambda \to 0} \frac{f(a + \lambda) - f(a)}{\lambda}$$

79.6.1 Notation

On note aussi :

$$\frac{df}{dx}(a) = \lim_{\lambda \to 0} \frac{f(a + \lambda) - f(a)}{\lambda}$$

On a alors :

$$f(a + h) = f(a) + \frac{df}{dx}(a) \cdot h + E(h)$$

et :

$$\lim_{\lambda \to 0} \frac{E(\lambda)}{\lambda} = \lim_{\lambda \to 0} \left[\frac{f(a + \lambda) - f(a)}{\lambda} - \frac{df}{dx}(a) \right] = 0$$

79.6.2 Définition équivalente

Si on pose $b = a + \lambda$, on voit que $\lambda = b - a$ et que la convergence $h \to 0$ est équivalente à $b \to a$. On a donc :

$$\frac{df}{dx}(a) = \lim_{b \to a} \frac{f(b) - f(a)}{b - a}$$

79.7 L'application dérivée

Si f est différentiable en tout vecteur a de $A \subseteq \Omega$, on dit que f est différentiable sur A. On peut alors définir une application dérivée $\partial f : A \mapsto \mathbb{M}(\mathbb{K}, m, n)$ définie par :

$$\partial f : a \mapsto \partial f(a)$$

pour tout $a \in A$. Si cette nouvelle application ∂f est également continue, on dit que f est continûment différentiable. On note $\mathrm{Cont}^1(A, F)$ l'ensemble des fonctions continûment différentiables de A vers F.

79.8 Hessienne

Soit $f : \Omega \mapsto \mathbb{K}$ avec Ω de dimension finie n. Supposons que f est différentiable sur $A \subseteq \Omega$. La dimension de $\mathbb{K}$ sur $\mathbb{K}$ étant 1, la Jacobienne $\partial f(a)$ se réduit à un vecteur matriciel de composantes $\partial_i f$. Si la dérivée $\partial f : A \mapsto \mathbb{K}^n$ est elle-même différentiable, on nomme l'application définie par :

$$\partial^2 f = \partial\left(\partial f\right)$$

la dérivée seconde de f. Il s'agit d'une fonction qui transforme un élément de A en un « vecteur matriciel de vecteurs matriciels » appartenant à $\mathbb{M}(\mathbb{K}^n, n, 1)$. On peut assimiler cet objet à une matrice équivalente de taille (n, n) dont les composantes sont des éléments de $\mathbb{K}$. En définitive, nous avons $\partial^2 f(a) \in \mathbb{M}(\mathbb{K}, n, n)$ pour tout $a \in A$. Cette matrice est appellée hessienne de f en a. Ses composantes sont données par :

$$\partial^2_{ij} f(a) = \partial_i \left(\partial_j f\right)(a)$$

Notation

On note aussi :

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \partial^2_{ij} f(a)$$

Lorsque $i = j$, on note :

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial^2 f}{\partial x_i \partial x_i} = \partial^2_{ii} f(a)$$

En termes matriciels, cela donne :

$$\frac{\partial^2 f}{\partial x \partial x}(a) = \partial^2 f(a)$$

Dérivée ordinaire

Dans le cas où la dimension de Ω est un 1, on peut l'assimiler à $\mathbb{K}$, on a alors une fonction $f : \mathbb{K} \mapsto \mathbb{K}$ possédant une seule dérivée seconde, que l'on note :

$$\frac{d^2 f}{dt^2}(a) = \partial^2 f(a)$$

Continuité

On note $\mathrm{Cont}^2(A, \mathbb{K})$ l'ensemble des fonctions dont la dérivée seconde est continue sur A.

79.9 Dérivée d'ordre k

Soit $f : \mathbb{K} \mapsto \mathbb{K}$ et $k \in \mathbb{N}$. Pour autant que la fonction f soit suffisamment dérivable, on définit par récurrence la fonction $\partial^k f : \mathbb{K} \mapsto \mathbb{K}$:

$$\begin{aligned} \partial^0 f &= f \\ \partial^k f &= \partial\big(\partial^{k-1} f\big) \end{aligned}$$

La dérivée d'ordre k de f est donc la fonction obtenue lorsqu'on applique un nombre k de fois l'opérateur de dérivation ∂ à la fonction f :

$$\partial^k f = (\partial \circ ... \circ \partial)(f)$$

79.9.1 Ensembles

On note $\mathrm{Cont}^k(A, \mathbb{K})$ l'ensemble des fonctions $f : A \mapsto \mathbb{K}$ dont la dérivée d'ordre k :

$$\partial^k f : A \mapsto \mathbb{K}$$

existe et est continue sur A.

Infini

Si la dérivée $\partial^k f$ existe pour tout $k \in \mathbb{N}$, on dit que f est indéfiniment dérivable. On note $\mathrm{Cont}^\infty(A, \mathbb{K})$ l'ensemble des fonctions indéfiniment dérivables.

Ordre 0

On voit que $\mathrm{Cont}^0 = \mathrm{Cont}$.

79.9.2 Homéomorphisme

On note $\mathrm{Hom}^k(A, \mathbb{K})$ l'ensemble des bijections f de Cont^k telles que la fonction f^{-1} soit aussi dans Cont^k.

79.9.3 Notations

Pour $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$, on écrit également :

$$\partial^\alpha f = \partial^{\alpha_1 ... \alpha_n} f = \partial_1^{\alpha_1} ... \partial_n^{\alpha_n} f$$

79.10 Fonctions à intégrale continue

Soit une fonction $u : A \mapsto B$. Si la fonction v définie par :

$$v(x) = \int_a^x u(x) d\mu(x)$$

est continue, on dit que u est à intégrale continue. On note $\mathrm{Cont}_\mu^{-1}(A, B)$ l'ensemble des fontions à intégrale continue.

79.11 Différentiabilité uniforme

On dit qu'une fonction f est uniformément différentiable sur A si pour tout $\epsilon > 0$, on peut trouver un $\delta > 0$ tel que :

$$|f(s) - f(t) - \partial f(t) \cdot (s - t)| \le \epsilon \cdot |s - t|$$

quel que soit $s, t \in A$ vérifiant $|s - t| \le \delta$.

Chapitre 80

Dérivées

Dépendances

— Chapitre **??** : Les différentielles

80.1 Fonction constante

Si f est constante, on peut trouver $c \in F$ tel que :

$$f(x) = c$$

pour tout $x \in \Omega$. On a alors :

$$\partial_j f(a) = \lim_{\lambda \to 0} \frac{c - c}{\lambda} = 0$$

Par conséquent :

$$\partial f(a) = 0$$

80.2 Identité

Considérons le cas particulier où $m = n$ et où $f = \mathrm{Id}$. Lorsque $i = j$, nous avons :

$$\partial_i f_i(a) = \lim_{\lambda \to 0} \frac{x_i + \lambda - x_i}{\lambda} = \lim_{\lambda \to 0} \frac{\lambda}{\lambda} = 1$$

Lorsque $i \neq j$, on a par contre :

$$\partial_j f_i(a) = \lim_{\lambda \to 0} \frac{x_i - x_i}{\lambda} = \lim_{\lambda \to 0} \frac{0}{\lambda} = 0$$

On en conclut que :

$$\partial f(a) = \partial \mathrm{Id}(a) = (\boldsymbol{\delta}_{ij})_{i,j} = I$$

La Jacobienne de la fonction identité est la matrice identité.

80.3 Composition de fonctions

Nous allons nous intéresser à présent au moyen d'obtenir la dérivée d'une composée de fonctions $f : E \mapsto F$ et $g : F \mapsto G$. Supposons que f soit différentiable en a et que g soit différentiable en $b = f(a)$. On a alors :

$$df(a) = f(a + da) - f(a) = \partial f(a) \cdot da + E_f(da)$$
$$dg(b) = g(b + db) - g(b) = \partial g(b) \cdot db + E_g(db)$$

Choisissons en particulier $db = f(a + da) - f(a)$. On a alors :

$$\begin{aligned} dg(b) &= g(f(a + da)) - g(f(a)) \\ &= (g \circ f)(a + da) - (g \circ f)(a) \\ &= d(g \circ f)(a) \end{aligned}$$

Mais d'un autre coté :

$$\begin{aligned} dg(b) &= \partial g(f(a)) \cdot df(a) + E_g(df(a)) \\ &= \partial g(f(a)) \cdot (\partial f(a) \cdot da + E_f(da)) + E_g(df(a)) \\ &= \partial g(f(a)) \cdot \partial f(a) \cdot da + E_f(da) \cdot da + E_g(df(a)) \end{aligned}$$

Il est aisé de vérifier que :

$$\lim_{h \to 0} \frac{\partial g(f(a)) \cdot E_f(da) + E_g(df(a))}{\|h\|} = 0$$

puisque $\partial g(f(a))$ ne dépend pas de da. On a donc montré que $g \circ f$ est différentiable en a et que :

$$\partial(g \circ f)(a) = \partial g(f(a)) \cdot \partial f(a) = (\partial g \circ f)(a) \cdot \partial f(a)$$

La dérivée d'une composée de fonctions est donc tout simplement le produit des Jacobiennes.

Notation

Soit le schéma fonctionnel :

$$(x_1, ..., x_n) \mapsto (y_1, ..., y_m) \mapsto (z_1, ..., z_p)$$

On note aussi :

$$\frac{\partial z_i}{\partial x_j} = \sum_k \frac{\partial z_i}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_j}$$

$$\frac{\partial z}{\partial x^T} = \frac{\partial z}{\partial y^T} \cdot \frac{\partial y}{\partial x^T}$$

Pour $z : t \mapsto (x_1(t), x_2(t), ..., x_n(t))$, on a également :

$$\frac{dz}{dt} = \sum_k \frac{\partial z}{\partial x_k} \cdot \frac{dx_k}{dt}$$

Si $x, y, z \in \mathbb{K}$, on a encore :

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

80.4 Inverse fonctionnel

En considérant le cas particulier $g = f^{-1}$, on a $g \circ f = \mathrm{Id}$. Choisissons un vecteur a où f est différentiable et posons $b = f(a)$. On a :

$$\partial f^{-1}(b) \cdot \partial f(a) = \partial \mathrm{Id}(a) = I$$

La jacobienne de l'inverse d'une fonction est donc l'inverse matriciel de sa jacobienne :

$$\partial f^{-1}(b) = [\partial f(a)]^{-1}$$

Notation

On note aussi :

$$\frac{\partial y}{\partial x} = \left(\frac{\partial x}{\partial y}\right)^{-1}$$

ou, pour des fonctions $y : \mathbb{K} \mapsto \mathbb{K}$:

$$\frac{dy}{dx} = \left(\frac{dx}{dy}\right)^{-1}$$

80.5 Addition

Si f, g sont différentiables en a, on a :

$$f(a + h) - f(a) = \partial f(a) \cdot h + E_f(h)$$
$$g(a + h) - g(a) = \partial g(a) \cdot h + E_g(h)$$

En additionnant les équations ci-dessus, on obtient :

$$[f(a + h) + g(a + h)] - [f(a) + g(a)] = [\partial f(a) + \partial g(a)] \cdot h$$
$$+ E_f(h) + E_g(h)$$

Il est clair que :

$$\lim_{h \to 0} \frac{\|E_f(h) + E_g(h)\|}{\|h\|} = 0$$

La fonction $f + g$ est donc différentiable en a et :

$$\partial(f + g)(a) = \partial f(a) + \partial g(a)$$

On peut montrer, cette fois en soustrayant les deux équations que :

$$\partial(f - g)(a) = \partial f(a) - \partial g(a)$$

80.6 Produit scalaire

Considérons deux fonctions f, g différentiables en a :

$$f(a + h) = f(a) + \partial f(a) \cdot h + E_f(h)$$
$$g(a + h) = g(a) + \partial g(a) \cdot h + E_g(h)$$

Leur produit scalaire s'écrit :

$$\langle f(a+h) \mid g(a+h)\rangle = \langle f(a) \mid g(a)\rangle + \langle f(a) \mid \partial g(a) \cdot h\rangle + \langle \partial f(a) \cdot h \mid g(a)\rangle + E_{f \cdot g}(h)$$

où :

$$E_{f \cdot g}(h) = \langle E_f(h) \mid E_g(h)\rangle + \langle E_f(h) \mid g(a)\rangle + \langle E_f(h) \mid \partial g(a)\rangle + \langle f(a) \mid E_g(h)\rangle + \langle \partial f(a) \mid E_g(h)\rangle$$

On a donc :

$$\lim_{h \to 0} \frac{\|E_{f \cdot g}(h)\|}{\|h\|} = 0$$

ce qui nous montre que la différentielle du produit scalaire s'écrit :

$$\mathfrak{D}_a^{\langle f \mid g\rangle}(h) = \langle f(a) \mid \partial g(a) \cdot h\rangle + \langle \partial f(a) \cdot h \mid g(a)\rangle$$

En terme de composantes, on a :

$$\mathfrak{D}_a^{\langle f \mid g\rangle}(h) = \sum_{j=1}^{n} \varpi_j \cdot h_j \cdot \sum_{i=1}^{m} [f_i(a) \cdot \partial_j g_i(a) + \partial_j f_i(a) \cdot g_i(a)]$$

La représentation matricielle s'écrit donc :

$$\partial \langle f \mid g\rangle (a) = [\partial g(a)]^T \cdot f(a) + [\partial f(a)]^T \cdot g(a)$$

80.6.1 Dérivée ordinaire

Dans le cas où $m = n = 1$, cette expression se simplifie en :

$$\frac{d}{dx}(f \cdot g)(a) = f(a) \cdot \frac{dg}{dx}(a) + \frac{df}{dx}(a) \cdot g(a)$$

80.6.2 Constante

Si une des deux fonctions est constante, soit $g(x) = c$ pour tout x, on a :

$$\partial g(x) = 0$$

et :

$$\frac{d}{dx}(f \cdot c)(a) = f(a) \cdot 0 + \frac{df}{dx}(a) \cdot c = c \cdot \frac{df}{dx}(a)$$

Notation

On note aussi :

$$d(f \cdot g) = df \cdot g + f \cdot dg$$

80.7 Inverse multiplicatif

Soit les fonctions $f, g : \mathbb{K} \mapsto \mathbb{K}$ reliées par l'équation :

$$f \cdot g = 1$$

En dérivant, on obtient :

$$\frac{df}{dx} \cdot g + f \cdot \frac{dg}{dx} = \frac{d1}{dx} = 0$$

Donc, si $g \neq 0$, on a :

$$\frac{df}{dx} = -\frac{f}{g} \cdot \frac{dg}{dx}$$

Mais comme $f(x) = 1/g(x)$, cela nous donne :

$$\frac{d}{dx}\left(\frac{1}{g}\right)(x) = -\frac{1}{g(x)^2} \cdot \frac{dg}{dx}(x)$$

80.8 Fraction

En appliquant les résultats précédents, on obtient :

$$\frac{d}{dx}\left(\frac{f}{g}\right)(x) = \frac{d}{dx}\left(f \cdot \frac{1}{g}\right)(x) = \frac{df}{dx}(x) \cdot g(x) - \frac{f(x)}{g(x)^2} \cdot \frac{dg}{dx}(x)$$

et finalement :

$$\frac{d}{dx}\left(\frac{f}{g}\right)(x) = \frac{\frac{df}{dx}(x) \cdot g(x) - f(x) \cdot \frac{dg}{dx}}{g(x)^2}$$

80.9 Dérivée d'une limite

Soit la suite de fonctions :

$$F = \{f_n \in \mathbb{R}^{\mathbb{R}} : n \in \mathbb{N}\}$$

convergeant en tout point $x \in \mathbb{R}$ vers une fonction $f : \mathbb{R} \mapsto \mathbb{R}$:

$$\lim_{n \to \infty} f_n(x) = f(x)$$

Si la fonction f est différentiable en $a \in \mathbb{R}$, on a :

$$\begin{aligned}
\partial f(a) &= \lim_{h \to 0} \frac{1}{h}\big[f(a+h) - f(a)\big] \\
&= \lim_{h \to 0} \frac{1}{h}\left[\lim_{n \to \infty} f_n(a+h) - \lim_{n \to \infty} f_n(a)\right] \\
&= \lim_{h \to 0} \lim_{n \to \infty} \frac{f_n(a+h) - f_n(a)}{h}
\end{aligned}$$

Chapitre 81

Différentielles et polynômes

Dépendances

— Chapitre **??** : Les différentielles

81.1 Polynômes

Soit $n \in \mathbb{N}$. Nous allons analyser la différentiabilité du monôme $\mu : \mathbb{R} \mapsto \mathbb{R}$ défini par :

$$\mu : t \mapsto t^n$$

pour tout $t \in \mathbb{R}$. La formule de factorisation nous donne :

$$s^n - t^n = (s - t) \sum_{i=0}^{n-1} s^i \cdot t^{n-1-i}$$

On a donc :

$$\frac{s^n - t^n}{s - t} = \sum_{i=0}^{n-1} s^i \cdot t^{n-1-i}$$

Passant à la limite $s \to t$, on obtient :

$$\lim_{s \to t} \frac{s^n - t^n}{s - t} = \sum_{i=0}^{n-1} t^i \cdot t^{n-1-i} = \sum_{i=0}^{n-1} t^{n-1} = n \cdot t^{n-1}$$

On en conclut que la dérivée existe sur $\mathbb{R}$ et que :

$$\frac{d}{dt}(t^n) = \lim_{s \to t} \frac{s^n - t^n}{s - t} = n \cdot t^{n-1}$$

La dérivée d'une combinaison linéaire étant identique à la combinaison linéaire des dérivées (voir dérivée d'une somme et la multiplication par une constante), on en conclut que tous les polynômes sont dérivables sur $\mathbb{R}$.

Uniformité

Choisissons $\alpha, \beta \in \mathbb{R}$ avec $\alpha \leq \beta$ et analysons la différentiabilité sur l'intervalle $[\alpha, \beta]$. Posons :

$$e(s, t) = \frac{s^n - t^n}{s - t} - \frac{d}{dt}(t^n)$$

Si $n = 1$, on a :

$$e(s,t) = \frac{s-t}{s-t} - 1 = 0$$

Le monôme de degré 1 est donc uniformément différentiable. Considérons à présent le cas où $n \ge 2$. Le passage à la limite nous montre que :

$$\frac{d}{dt}(t^n) = \sum_{i=0}^{n-1} t^i \cdot t^{n-1-i}$$

En utilisant les propriétés des sommes, on obtient :

$$\begin{aligned} e(s,t) &= \sum_{i=0}^{n-1} s^i \ t^{n-1-i} - \sum_{i=0}^{n-1} t^i \ t^{n-1-i} \\ &= \sum_{i=0}^{n-1} (s^i - t^i) \ t^{n-1-i} \end{aligned}$$

En factorisant tous les $s^i - t^i$, on a alors :

$$e(s,t) = \sum_{i=0}^{n-1} t^{n-1-i} \ (s-t) \ \sum_{k=0}^{i-1} s^k \ t^{i-1-k}$$

et comme $s - t$ ne dépend pas de i :

$$e(s,t) = (s-t) \sum_{i=0}^{n-1} t^{n-1-i} \ \sum_{k=0}^{i-1} s^k \ t^{i-1-k}$$

Si on pose $M = \max\{|\alpha|, |\beta|\}$, on a clairement $|s|, |t| \le M$. On peut alors trouver la borne supérieure :

$$\begin{aligned} |e(s,t)| &\le |s-t| \sum_{i=0}^{n-1} M^{n-1-i} \ \sum_{k=0}^{i-1} M^{i-1} \\ &\le |s-t| \sum_{i=0}^{n-1} M^{n-1-i} \ i \ M^{i-1} \\ &\le |s-t| \ M^{n-2} \ \sum_{i=0}^{n-1} i \\ &\le \frac{1}{2} \ |s-t| \ M^{n-2} \ (n-1) \ n \end{aligned}$$

Fixons à présent $\epsilon > 0$. Il suffit de prendre :

$$|s-t| \le \delta \le \frac{2\epsilon}{M^{n-2} \cdot (n-1) \cdot n}$$

pour avoir :

$$|e(s,t)| \le \frac{M^{n-2} \cdot (n-1) \cdot n \cdot \delta}{2} \le \epsilon$$

Comme on a :

$$\mu(s) - \mu(t) - \partial\mu(t) = s^n - t^n - n \cdot t^{n-1} = e(s,t) \cdot (s-t)$$

on dispose de la borne supérieure :

$$|\mu(s) - \mu(t) - \partial\mu(t)| \le |e(s,t)| \cdot |s-t| \le \epsilon \cdot |s-t|$$

Comme le choix de δ ne dépend ni de s ni de t, le monôme μ est uniformément différentiable sur $[\alpha, \beta]$.

On généralise aisément à un polynôme quelconque :

$$p(x) = \sum_{i=0}^{n} a_i \cdot x^i$$

en constatant que :

$$|p(s) - p(t) - \partial p(t) \cdot (s-t)| \le |s-t| \sum_{i=0}^{n} |a_i| \cdot |e_i(s,t)|$$

où e_i est l'erreur obtenue avec le monôme de degré i. Mais comme on peut trouver des δ_k tels que :

$$|e_i(s,t)| \le \frac{\epsilon}{\sum_j |a_j|}$$

il suffit de choisir $\delta = \min\{\delta_0, \delta_1, ..., \delta_n\}$ pour avoir :

$$|p(s) - p(t) - \partial p(t) \cdot (s-t)| \le |s-t| \cdot \epsilon \cdot \frac{\sum_i |a_i|}{\sum_j |a_j|} = |s-t| \cdot \epsilon$$

Tout polynôme est uniformément différentiable sur des intervalles de la forme $[\alpha, \beta]$. Cette généralisation montre aussi que toute combinaison linéaire de fonctions uniformément différentiables est uniformément différentiable.

Chapitre 82

Dérivées des puissances

82.1 Introduction

Nous allons évaluer les dérivées des fonctions $f : x \mapsto x^\alpha$ où $x, \alpha \in \mathbb{R}$.

82.2 L'inverse multiplicatif

Commençons par :

$$x \cdot y = 1$$
$$y = \frac{1}{x} = x^{-1}$$

On en déduit que :

$$x \cdot dy + y \cdot dx = d(1) = 0$$

ce qui nous donne :

$$\frac{d}{dx}\left(\frac{1}{x}\right) = \frac{dy}{dx} = -\frac{y}{x} = -\frac{1}{x^2}$$

82.3 Puissances négatives

Considérons la relation :

$$y = x^{-n}$$

où $n \in \mathbb{N}$. On définit la variable intermédiaire z telle que :

$$z = x^{-1}$$
$$y = z^n$$

On en déduit :

$$\begin{aligned}
\frac{dy}{dx} = \frac{dy}{dz} \cdot \frac{dz}{dx} &= n \cdot z^{n-1} \cdot \left(-\frac{1}{x^2}\right) \\
&= -n \cdot x^{1-n} \cdot x^{-2} \\
&= (-n) \cdot x^{-n-1}
\end{aligned}$$

82.4 Racines

Toujours pour $n \in \mathbb{N}$, considérons la relation :

$$y = x^n \qquad \Leftrightarrow \qquad x = y^{1/n}$$

Posons $\alpha = 1/n$. On a :

$$\frac{dy}{dx} = n \cdot x^{n-1} = n \cdot y \cdot y^{-\alpha}$$

Donc :

$$\frac{dx}{dy} = \alpha \cdot y^{\alpha - 1}$$

82.5 Puissances fractionnaires

Choisissons à présent :

$$y = x^{m/n}$$

où $m \in \mathbb{Z}$ et $n \in \mathbb{N}$. Définissons la variable intermédiaire :

$$z = x^m$$

On a alors :

$$y = z^{1/n}$$

Posons $\alpha = m/n$. La dérivée s'écrit :

$$\begin{aligned} \frac{dy}{dx} = \frac{dy}{dz} \cdot \frac{dz}{dx} \quad &= \quad \frac{1}{n} \cdot z^{\frac{1}{n}-1} \cdot m \cdot x^{m-1} \\ &= \quad \alpha \cdot x^{\alpha - m} \cdot x^{m-1} \\ &= \quad \alpha \cdot x^{\alpha - 1} \end{aligned}$$

On a donc :

$$\frac{d}{dx}\left(x^\alpha\right) = \alpha \cdot x^{\alpha - 1}$$

82.6 Puissances réelles

Par passage à la limite, on obtient :

$$\frac{d}{dx}\left(x^\alpha\right) = \alpha \cdot x^{\alpha - 1}$$

pour tout $\alpha \in \mathbb{R}$.

Chapitre 83

Différentielles et matrices

83.1 Applications linéaires

Soit $A \in \mathbb{M}(\mathbb{R}, m, n)$. Considérons la fonction $\mathcal{A} : \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}$ définie par :

$$\mathcal{A}(x) = A \cdot x$$

En terme de composantes, on a :

$$\mathcal{A}_i(x) = \sum_j A_{ij} \cdot x_j$$

Les dérivées s'écrivent :

$$\frac{\partial \mathcal{A}_i}{\partial x_k}(x) = A_{ik}$$

On a donc :

$$\frac{\partial}{\partial x}(A \cdot x) = A$$

83.2 Formes linéaires

Soit $u \in \mathbb{R}^n$. Considérons la forme linéaire $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ définie par :

$$\varphi(x) = x^\star \cdot u$$

En terme de composantes, on a :

$$\varphi(x) = \sum_i x_i \cdot u_i$$

Les dérivées s'écrivent :

$$\frac{\partial \varphi}{\partial x_k}(x) = u_i$$

On a donc :

$$\frac{\partial}{\partial x}(x^\star \cdot u) = u$$

Comme $u^\star \cdot x = x^\star \cdot u$, on a aussi :

$$\frac{\partial}{\partial x}(u^\star \cdot x) = u$$

83.3 Formes bilinéaires

Soit à présent $A \in \mathbb{M}(\mathbb{R}, m, n)$. Considérons la forme bilinéaire $\vartheta : \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}$ définie par :

$$\vartheta(x, y) = x^\star \cdot A \cdot y$$

En terme de composantes, on a :

$$\vartheta(x, y) = \sum_{i,j} x_i \cdot A_{ij} \cdot y_j$$

Les dérivées s'écrivent :

$$\frac{\partial \vartheta}{\partial x_k}(x, y) = \sum_j A_{kj} \cdot y_j$$

$$\frac{\partial \vartheta}{\partial y_k}(x, y) = \sum_i x_i \cdot A_{ik}$$

On a donc :

$$\frac{\partial}{\partial x}(x^\star \cdot A \cdot y) = A \cdot y$$

$$\frac{\partial}{\partial y}(x^\star \cdot A \cdot y) = A^\star \cdot x$$

Les dérivées secondes s'en déduisent alors :

$$\frac{\partial^2}{\partial x \partial x}(x^\star \cdot A \cdot y) = 0$$

$$\frac{\partial^2}{\partial y \partial y}(x^\star \cdot A \cdot y) = 0$$

$$\frac{\partial^2}{\partial y \partial x}(x^\star \cdot A \cdot y) = A$$

83.4 Formes quadratiques

Soit à présent $A \in \mathbb{M}(\mathbb{R}, n, n)$. Considérons la forme quadratique $\mathcal{Q} : \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}$ définie par :

$$\mathcal{Q}(x) = x^\star \cdot A \cdot x$$

En terme de composantes, on a :

$$\mathcal{Q}(x) = \sum_{i,j} x_i \cdot A_{ij} \cdot x_j$$

Les dérivées s'écrivent :

$$\begin{aligned}
\frac{\partial \mathcal{Q}}{\partial x_k}(x) &= \sum_j A_{kj} \cdot x_j + \sum_i x_i \cdot A_{ik} \\
&= \sum_i (A_{ki} + A_{ik}) \cdot x_i
\end{aligned}$$

On a donc :

$$\frac{\partial}{\partial x}(x^\star \cdot A \cdot x) = (A + A^\star) \cdot x$$

La dérivée seconde est alors immédiate :

$$\frac{\partial^2}{\partial x \partial x}(x^\star \cdot A \cdot x) = A + A^\star$$

Un cas fréquent est celui d'une matrice H hermitienne ($H^\star = H$). On a alors :

$$\frac{\partial}{\partial x}(x^\star \cdot H \cdot x) = 2 \cdot H \cdot x$$

et :

$$\frac{\partial^2}{\partial x \partial x}(x^\star \cdot H \cdot x) = 2 \cdot H$$

83.5 Produit matriciel

Soient les fonctions $A : \mathbb{R} \mapsto \mathbb{M}(\mathbb{R}, m, n)$ et $B : \mathbb{R} \mapsto \mathbb{M}(\mathbb{R}, n, p)$ qui, à chaque réel t, associent des matrices de composantes réelles. Nous allons tenter de trouver une expression de la dérivée du produit matriciel $A \cdot B$. Les propriétés des différentielles nous permettent d'écrire :

$$\partial \sum_{k=1}^n A_{ik}(t) \cdot B_{kj}(t) = \sum_{k=1}^n \partial A_{ik}(t) \cdot B_{kj}(t) + \sum_{k=1}^n A_{ik}(t) \cdot \partial B_{kj}(t)$$

On a donc :

$$\partial(A \cdot B) = \partial A \cdot B + A \cdot \partial B$$

ou, symboliquement :

$$d(A \cdot B) = dA \cdot B + A \cdot dB$$

83.6 Matrice inverse

Considérons le cas où A est carrée ($m = n$). En dérivant la relation $A \cdot A^{-1} = I$, on obtient :

$$0 = dI = d(A \cdot A^{-1}) = dA \cdot A^{-1} + A \cdot dA^{-1}$$

et donc :

$$d(A^{-1}) = -A^{-1} \cdot dA \cdot A^{-1}$$

Chapitre 84

Résolution d'équations

84.1 Introduction

Soit une fonction $F : \mathbb{R}^n \mapsto \mathbb{R}^n$ et l'espace des solutions :

$$S = \{s \in \mathbb{R}^n : F(s) = 0\} = \ker F$$

Nous allons tenter d'obtenir itérativement une estimation d'un $s \in S$. 0n part d'un certain $x_0 \in \mathbb{R}^n$ et on essaie à chaque itération :

$$x_{k+1} = I(x_k) = x_k + \Delta_k$$

d'améliorer la qualité de notre estimation, c'est à dire de rapprocher $F(x_{k+1})$ de zéro. On espère que pour K assez grand, on aura :

$$F(x_K) \approx \lim_{k\to\infty} F(x_k) = 0$$

A moins que l'on ait déjà une vague idée d'une région $X \subseteq \mathbb{R}^n$ contenant une solution $s \in S$, on choisit en général $x_0 = 0$.

84.2 Newton-Raphson

On demande à chaque étape que le développement du premier ordre autour de x_k s'annule en $x_{k+1} = x_k + \Delta_k$. On impose donc :

$$F(x_{k+1}) \approx F(x_k) + \partial F(x_k) \cdot \Delta_k \approx 0$$

On est amenés à résoudre le système :

$$\partial F(x_k) \cdot \Delta_k = -F(x_k)$$

Si la matrice $\partial F(x_k)$ est inversible, on a :

$$\Delta_k = -\left[\partial F(x_k)\right]^{-1} \cdot F(x_k)$$

et :

$$x_{k+1} = x_k - \left[\partial F(x_k)\right]^{-1} \cdot F(x_k)$$

84.3 Banach

On voit que la condition $F(s) = 0$ est équivalente à $F(s) + s = s$. On définit donc la fonction $f : \mathbb{R}^n \to \mathbb{R}^n$ par :

$$f(x) = F(x) + x$$

pour tout $x \in \mathbb{R}^n$, et on cherche un s tel que $f(s) = F(s) + s = s$. Si f est contractante, on applique le théorème de Banach en itérant simplement par :

$$x_{k+1} = f(x_k) = f^{k+1}(x_0)$$

Pour K assez grand, on a alors :

$$x_K \approx \lim_{k\to\infty} x_k = s$$

84.4 Méthode hybride d'Aitken

Supposons à présent que $F, f : \mathbb{R} \to \mathbb{R}$ avec $f(x) = F(x)+x$ pour tout $x \in \mathbb{R}$. L'idée est d'utiliser simultanément les deux approches $F(s) = 0$ et $f(s) = F(s) + s = s$. Soit :

$$F'(x) = \frac{dF}{dx}(x)$$

Supposons qu'au bout de k itérations on ait obtenu x_k comme approximation de la solution s. On commence par effectuer deux f-itérations en partant de x_k :

$$\begin{aligned} u_0 &= x_k \\ u_1 &= f(u_0) \\ u_2 &= f(u_1) \end{aligned}$$

Les valeurs de F s'écrivent alors :

$$\begin{aligned} F(u_0) = f(u_0) - u_0 = u_1 - u_0 \\ F(u_1) = f(u_1) - u_1 = u_2 - u_1 \end{aligned}$$

On se sert ensuite du développement :

$$F(u_0) \approx F(u_1) + F'(u_1) \cdot (u_0 - u_1)$$

pour approximer F' :

$$F'(u_1) \approx \frac{F(u_0) - F(u_1)}{u_0 - u_1} = \frac{F(u_1) - F(u_0)}{u_1 - u_0}$$

On a donc :

$$F'(u_1) \approx \frac{(u_2 - u1) - (u_1 - u_0)}{u_1 - u_0} = \frac{(u_2 - 2u_1 + u_0)}{u_1 - u_0}$$

On applique ensuite l'itération de Newton-Raphson :

$$x_{k+1} = x_k - \frac{F(x_k)}{F'(x_k)}$$

en remplaçant la dérivée F' par son approximation :

$$x_{k+1} = x_k - \frac{(u_1 - u_0)^2}{u_2 - 2u_1 + u_0}$$

On espère que la suite des x_k ainsi définie converge plus vite vers la solution s que la suite des $f^k(x_0)$.

Huitième partie

Analyse

Chapitre 85

Théorème de Rolle

Dépendances

— Chapitre **??** : Les différentielles
— Chapitre 72 : Les intégrales

85.1 Extrema locaux

Soit $f : A \mapsto \mathbb{R}$ avec $A \subseteq \mathbb{R}^n$. Supposons que f soit différentiable et atteigne un minimum local en $a \in \operatorname{int} A$. On a :

$$f(a+h) - f(a) = \mathfrak{D}^f_a(h) + E(h)$$
$$f(a-h) - f(a) = -\mathfrak{D}^f_a(h) + E(-h)$$

Fixons $\epsilon > 0$. On peut trouver $\delta_1 > 0$ tel que :

$$|E(h)| \le \epsilon \cdot |h|$$

pour tout $h \in \mathfrak{B}(0, \delta_1)$. On peut aussi trouver $\delta_2 > 0$ tel que :

$$f(a) \le f(a+h)$$

pour tout $h \in \mathfrak{B}(0, \delta_2)$. Posons $\delta = \min\{\delta_1, \delta_2\}$ et choisissons $h \in \mathfrak{B}(0, \delta)$. On a également $-h \in \mathfrak{B}(0, \delta)$. Donc :

$$\mathfrak{D}^f_a(h) = f(a+h) - f(a) - E(h) \ge -E(h) \ge -\epsilon \cdot \|-h\|$$
$$\mathfrak{D}^f_a(h) = f(a) - f(a-h) + E(-h) \le E(-h) \le \epsilon \cdot \|h\|$$

On en conclut que :

$$\left|\mathfrak{D}^f_a(h)\right| \le \epsilon \cdot \|h\|$$

Posons $\gamma = \delta/2$ et remarquons que l'ensemble de norme fixe $N = \{h \in \mathbb{R}^n : \|h\| = \gamma\}$ est inclus dans $\mathfrak{B}(0, \delta)$. Les propriétés des applications linéaires nous disent que :

$$\left\|\mathfrak{D}^f_a\right\| = \sup\left\{\frac{1}{\gamma}\left\|\mathfrak{D}^f_a(h)\right\| : h \in N\right\}$$

Or, la borne nous dit que :

$$\frac{1}{\gamma}\left|\mathfrak{D}^f_a(h)\right| \le \epsilon$$

quel que soit $\epsilon > 0$ et $h \in N$. Donc :

$$\left\|\mathfrak{D}_a^f\right\| = 0$$

ce qui implique que :

$$\mathfrak{D}_a^f = 0$$

La différentielle s'annule donc en un minimum local. On montre de la même manière que la différentielle s'annule en un maximum local.

La Jacobienne

La Jacobienne étant la représentation matricielle de la différentielle, elle s'annule également aux extrema locaux.

85.2 Théorème de Rolle

Soit $f \in \operatorname{Cont}^1([a,b],\mathbb{R})$ avec $f(a) = f(b)$. Comme f est continue, il existe $\lambda, \sigma \in [a,b]$ tels que :

$$f(\lambda) = \min f([a,b])$$
$$f(\sigma) = \max f([a,b])$$

85.2.1 Configurations

Plusieurs cas peuvent se présenter :

— $\lambda < f(a) = f(b) < \sigma$: dans ce cas, la fonction atteint ses deux bornes à l'intérieur de l'intervalle :

$$\{\lambda, \sigma\} \subseteq]a, b[$$

Comme les extrema sont aussi des extrema locaux, on a :

$$\partial f(\lambda) = \partial f(\sigma) = 0$$

— $\lambda < f(a) = f(b) = \sigma$: dans ce cas, la fonction atteint son minimum à l'intérieur de l'intervalle :

$$\lambda \in]a, b[$$

et on a :

$$\partial f(\lambda) = 0$$

— $\lambda = f(a) = f(b) < \sigma$: dans ce cas, la fonction atteint son maximum à l'intérieur de l'intervalle :

$$\sigma \in]a, b[$$

et on a :

$$\partial f(\sigma) = 0$$

— $\lambda = f(a) = f(b) = \sigma$: dans ce cas, on a :

$$\lambda \le f(x) \le \sigma = \lambda$$

pour tout $x \in [a,b]$, et donc :

$$f(x) = \lambda$$

La fonction f est constante et $\partial f = 0$. On peut donc prendre n'importe quel $c \in]a,b[$, on aura :

$$\partial f(c) = 0$$

85.2.2 Conclusion

Dans tous les cas, on a au moins un $c \in]a,b[$ tel que :

$$\partial f(c) = 0$$

85.3 Théorème des accroissements finis

Soit $f \in \mathrm{Cont}^1([a,b],\mathbb{R})$ et la fonction g associée définie par :

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a} \cdot (x - a)$$

pour tout $x \in [a,b]$. Comme $g(a) = g(b) = f(a)$, on peut trouver un $c \in]a,b[$ tel que :

$$0 = \partial g(c) = \partial f(c) - \frac{f(b) - f(a)}{b - a}$$

On a donc :

$$\partial f(c) = \frac{f(b) - f(a)}{b - a}$$

On vient ainsi de démontrer le théorème des accroissements finis.

Dimension n

On peut généraliser ce théorème à une fonction $f \in \mathrm{Cont}^1(\mathbb{R}^m,\mathbb{R}^n)$. Soit $u, v \in \mathbb{R}^m$. On considére le segment $[u,v]$ et la fonction associée $\phi : [0,1] \mapsto \mathbb{R}^m$ définie par :

$$\phi(t) = u + t \cdot (v - u)$$

pour tout $t \in [0,1]$. On pose alors :

$$g(t) = (f \circ \phi)(t) = f(u + t \cdot (v - u))$$

pour tout $t \in [0,1]$. En appliquant le résultat précédent aux composantes g_i sur l'intervalle $[0,1]$, on obtient un $s \in]0,1[$ tel que :

$$\partial g_i(s) = \frac{g_i(1) - g_i(0)}{1 - 0} = g_i(1) - g_i(0) = f_i(v) - f_i(u)$$

En appliquant la formule permettant d'évaluer la dérivée d'une composition de fonctions, on obtient :

$$\partial g_i(s) = \sum_j \partial_j f_i(u + s \cdot (v - u)) \cdot (v_j - u_j)$$

Utilisant la notation matricielle, on a donc :

$$f(v) - f(u) = \partial f(u + s \cdot (v - u)) \cdot (v - u)$$

Ce qui revient à dire qu'il existe un $w \in [u, v] \subseteq \mathbb{R}^n$ tel que :

$$f(v) - f(u) = \partial f(w) \cdot (v - u)$$

85.4 Théorème de Cauchy

Le théorème des accroissements finis nous donne un résultat sous la forme symbolique :

$$\frac{df}{dx} = \frac{\boldsymbol{\Delta} f}{\boldsymbol{\Delta} x}$$

Nous allons maintenant généraliser ce théorème, et obtenir le résultat :

$$\frac{df}{dg} = \frac{\boldsymbol{\Delta} f}{\boldsymbol{\Delta} g}$$

où $f, g \in \mathrm{Cont}^1([a, b], \mathbb{R})$ et $a, b \in \mathbb{R}$. Considérons à cette fin la fonction h définie par :

$$h(x) = [f(b) - f(a)] \cdot g(x) - f(x) \cdot [g(b) - g(a)]$$

On remarque que :

$$\begin{aligned} h(a) &= f(b) \cdot g(a) - f(a) \cdot g(a) - f(a) \cdot g(b) + f(a) \cdot g(a) \\ &= f(b) \cdot g(a) - f(a) \cdot g(b) \end{aligned}$$

$$\begin{aligned} h(b) &= f(b) \cdot g(b) - f(a) \cdot g(b) - f(b) \cdot g(b) + f(b) \cdot g(a) \\ &= f(b) \cdot g(a) - f(a) \cdot g(b) \end{aligned}$$

Donc :

$$h(a) = h(b)$$

Appliquant le théorème de Rolle à h, on peut donc trouver un $c \in]a, b[$ tel que :

$$0 = \partial h(c) = [f(b) - f(a)] \cdot \partial g(c) - \partial f(c) \cdot [g(b) - g(a)]$$

On a donc :

$$[f(b) - f(a)] \cdot \partial g(c) = \partial f(c) \cdot [g(b) - g(a)]$$

Si $\partial f(c) \neq 0$ et $f(b) \neq f(a)$, on peut le mettre sous la forme :

$$\frac{\partial g(c)}{\partial f(c)} = \frac{g(b) - g(a)}{f(b) - f(a)}$$

85.5 Théorème de l'Hospital

Soient F, G deux fonctions continues sur $I = [\alpha, \beta]$ et dérivables $I \setminus \{a\}$, avec $a \in \mathrm{int}\ I$. Supposons que les deux fonctions s'annulent en a :

$$F(a) = G(a) = 0$$

Soit alors $h \neq 0$ tel que $b = a + h \in I \setminus \{a\}$. En appliquant le théorème de Cauchy à F et G, on trouve un $t \in]0, 1[$ tel que :

$$[F(b) - F(a)] \cdot \partial G(a + t \cdot h) = [G(b) - G(a)] \cdot \partial F(a + t \cdot h)$$

Mais comme F et G s'annulent en a, on a :

$$F(b) \cdot \partial G(a + t \cdot h) = G(b) \cdot \partial F(a + t \cdot h)$$

Si de plus ∂G ne s'annule pas sur I, on peut écrire :

$$\frac{\partial F(a + t \cdot h)}{\partial G(a + t \cdot h)} = \frac{F(b)}{F(a)}$$

On voit en faisant tendre h vers 0 que les limites, si elles existent, doivent être identiques. On a donc :

$$\lim_{x \to a} \frac{F(x)}{G(x)} = \lim_{x \to a} \frac{\partial F(a)}{\partial G(a)}$$

85.6 Uniformité

Nous allons à présent montrer que toute fonction continument différentiable sur un intervalle de la forme $[\alpha, \beta]$ y est uniformément différentiable.

Soit une fonction $f \in \mathrm{Cont}^1([\alpha, \beta], \mathbb{R})$. Comme la dérivée ∂f est continue sur $[\alpha, \beta]$, elle y est uniformément continue. Fixons $\epsilon > 0$. On peut donc trouver un $\delta > 0$ tel que :

$$|\partial f(s) - \partial f(t)| \le \epsilon$$

pour tout $s, t \in [\alpha, \beta]$ vérifiant $|s - t| \le \delta$. Si $s = t$, on a bien évidemment :

$$|f(t) - f(t) - \partial f(t) \cdot (t - t)| = 0 \le \epsilon \cdot (t - t) = 0$$

Considérons à présent le cas $s \ne t$. Nous pouvons supposer sans perte de généralité que $s < t$. Le théorème des accroissements finis nous dit qu'on peut trouver un $\gamma \in]s, t[$ tel que :

$$\partial f(\gamma) = \frac{f(t) - f(s)}{t - s}$$

On a donc :

$$\frac{f(t) - f(s)}{t - s} - \partial f(s) = \partial f(\gamma) - \partial f(s)$$

Mais comme $|\gamma - s| \le |t - s| \le \delta$, on a $|\partial f(\gamma) - \partial f(s)| \le \epsilon$ et :

$$\left| \frac{f(t) - f(s)}{t - s} - \partial f(s) \right| \le \epsilon$$

On a donc bien :

$$|f(t) - f(s) - \partial f(s) \cdot (t - s)| \le \epsilon \cdot |t - s|$$

pour tout $s, t \in [\alpha, \beta]$ vérifiant $|s - t| \le \delta$.

Remarque

Le théorème *n'est pas* applicable aux autres types d'intervalles. Cela ne marche pas sur $]\alpha, \beta[$ par exemple.

Chapitre 86

Théorème fondamental

Dépendances

- Chapitre **??** : Les différentielles
- Chapitre 72 : Les intégrales

86.1 Dérivée de l'intégrale

Soit une fonction $f \in \mathrm{Cont}([\alpha, \beta], \mathbb{R})$, un réel $a \in [\alpha, \beta]$ et la fonction intégrale associée $I : [\alpha, \beta] \mapsto \mathbb{R}$ définie par :

$$I(x) = \int_a^x f(s)\ ds$$

pour tout $x \in [\alpha, \beta]$. On constate que :

$$I(a) = \int_a^a f(s)\ ds = \int_{\{a\}} f(s)\ ds$$

La mesure de Lebesgue du singleton $\{a\}$ étant nulle, l'intégrale s'annule et on a :

$$I(a) = 0$$

Nous allons chercher à évaluer la dérivée de la fonction intégrale I en un point quelconque $b \in [\alpha, \beta]$.

Nous utilisons dans la suite la notation abrégée :

$$\int_x^y = \int_x^y f(s)\ ds$$

Par additivité, on a :

$$\int_a^{b+h} = \int_a^b + \int_b^{b+h}$$

c'est-à-dire :

$$I(b+h) = I(b) + \int_b^{b+h}$$

pour tout réel h tel que $b + h \in [\alpha, \beta]$. Quelle est la valeur de l'intégrale sur $[b, b+h]$? Fixons $\epsilon > 0$. Par continuité de f, on peut choisir $\delta > 0$ tel que :

$$|f(b+s) - f(b)| \le \epsilon$$

pour tout s vérifiant $|s| \le \delta$. Cela n'est possible que si :

$$f(b) - \epsilon \le f(b+s) \le f(b) + \epsilon$$

On a donc :

$$\sup \Big\{ f(b+s) : |s| \le \delta \Big\} \le f(b) + \epsilon$$
$$\inf \Big\{ f(b+s) : |s| \le \delta \Big\} \ge f(b) - \epsilon$$

Si nous choisissons $h \in (0, \delta)$, l'intégrale peut donc être majorée et minorée par :

$$\int_b^{b+h} \le (f(b) + \epsilon) \cdot \mu_L([a, a+h]) = (f(b) + \epsilon) \cdot h$$

et :

$$\int_b^{b+h} \ge (f(b) - \epsilon) \cdot \mu_L([a, a+h]) = (f(b) - \epsilon) \cdot h$$

Nous disposons donc des inégalités :

$$(f(b) - \epsilon) \cdot h \le \int_b^{b+h} \le (f(b) + \epsilon) \cdot h$$

Autrement dit :

$$f(b) - \epsilon \le \frac{1}{h} \int_b^{b+h} \le f(b) + \epsilon$$

D'un autre coté, on a :

$$\frac{1}{h} \int_b^{b+h} f(s)\ ds = \frac{I(b+h) - I(b)}{h}$$

Passons à la limite $\delta \to 0$. On a alors $h \to 0$ et :

$$f(b) - \epsilon \le \lim_{h \to 0} \frac{1}{h} \int_b^{b+h} \le f(b) + \epsilon$$

Ces inégalités devant être valables pour tout $\epsilon > 0$, on a forcément :

$$\lim_{h \to 0} \frac{I(b+h) - I(b)}{h} = f(b)$$

On en conclut que I est dérivable et que :

$$\frac{dI}{dx}(b) = \lim_{h \to 0} \frac{I(b+h) - I(b)}{h} = f(b)$$

Autrement dit :

$$\frac{d}{dx} \int_a^x f(s)\ ds = f(x)$$

86.2 Intégrale de la dérivée

Soit $\alpha, \beta \in \mathbb{R}$ avec $\alpha \le \beta$. Soit la fonction $F \in \mathrm{Cont}^1([\alpha, \beta], \mathbb{R})$ et sa dérivée continue :

$$f = \partial F = \frac{dF}{ds}$$

Comme F est continument différentiable sur $[\alpha, \beta]$, elle y est uniformément différentiable. De même, f est continue sur $[\alpha, \beta]$. Elle y est donc uniformément continue. Nous allons tenter d'évaluer l'intégrale :

$$\int_a^b f(s)\ ds = \int_a^b \frac{dF}{dx}(s)\ ds$$

avec $a, b \in [\alpha, \beta]$ et $a \le b$.
Nous utilisons dans la suite la notation abrégée :

$$\int_x^y = \int_x^y f(s)\ ds$$

86.2.1 L'idée

L'idée intuitive est que :

$$\boldsymbol{\Delta} F = \sum_i \boldsymbol{\Delta} F_i = \sum_i \frac{\boldsymbol{\Delta} F_i}{\boldsymbol{\Delta} x_i} \cdot \boldsymbol{\Delta} x_i$$

En passant à la limite $\boldsymbol{\Delta} x_i \to 0$, on soupçonne alors le résultat suivant :

$$\boldsymbol{\Delta} F = \int_a^b \frac{dF}{dx}\ dx$$

86.2.2 La réalisation

Fixons $\epsilon > 0$. Comme f est uniformément continue, nous savons qu'il existe $\vartheta > 0$ tel que :

$$|f(x+h) - f(x)| \le \epsilon$$

pour tout x, h vérifiant $x, x+h \in [\alpha, \beta]$ et $|h| \le \vartheta$. On en déduit que :

$$\sup_{\xi \in [x-\vartheta, x+\vartheta]} f(\xi) \le f(x) + \epsilon$$
$$\inf_{\xi \in [x-\vartheta, x+\vartheta]} f(\xi) \ge f(x) - \epsilon$$

On a donc les bornes pour l'intégrale :

$$(f(x) - \epsilon) \cdot h \le \int_x^{x+h} \le (f(x) + \epsilon) \cdot h$$

Comme F est uniformément différentiable, nous pouvons trouver $\varpi > 0$ tel que :

$$|F(x+h) - F(x) - f(x) \cdot h| \le \epsilon \cdot h$$

pour tout x, h vérifiant $x, x+h \in [\alpha, \beta]$ et $|h| \le \varpi$. On en déduit que :

$$F(x+h) - F(x) - f(x) \cdot h \le \epsilon \cdot h$$
$$f(x) \cdot h - (F(x+h) - F(x)) \le \epsilon \cdot h$$

En considérant ces deux inégalités par rapport au centre $f(x) \cdot h$, on obtient :

$$F(x+h) - F(x) - \epsilon \cdot h \le f(x) \cdot h \le F(x+h) - F(x) + \epsilon \cdot h$$

En soustrayant ou en ajoutant $\epsilon \cdot h$ à ces inégalités, on a :

$$\begin{aligned} F(x+h) - F(x) - 2\epsilon \cdot h &\le f(x) \cdot h - \epsilon \cdot h \le &F(x+h) - F(x) \\ F(x+h) - F(x) &\le f(x) \cdot h + \epsilon \cdot h \le &F(x+h) - F(x) + 2\epsilon \cdot h \end{aligned}$$

Si $|h| \le \min\{\vartheta, \varpi\}$, nous avons de nouvelles bornes pour l'intégrale :

$$F(x+h) - F(x) - 2\epsilon \cdot h \le \int_x^{x+h} \le F(x+h) - F(x) + 2\epsilon \cdot h$$

Choisissons à présent $n \in \mathbb{N}$ tel que :

$$\left| \frac{b-a}{n} \right| \le \min\{\vartheta, \varpi\}$$

Posons $h = (b-a)/n$ et définissons la série :

$$x_i = a + i \cdot h$$

On a alors $a = x_0 \le x_1 \le ... \le x_n = b$. Les propriétés des sommes nous disent que :

$$\sum_{i=1}^n (F(x_i) - F(x_{i-1})) = F(x_n) - F(x_0) = F(b) - F(a)$$

D'un autre coté, on a clairement :

$$\sum_{i=1}^n \int_{x_{i-1}}^{x_i} = \int_a^b$$

Si nous appliquons les bornes précédentes avec $x = x_{i-1}$, nous avons $x + h = x_i$ et :

$$F(x_i) - F(x_{i-1}) - 2\epsilon \cdot h \le \int_{x_{i-1}}^{x_i} \le F(x_i) - F(x_{i-1}) + 2\epsilon \cdot h$$

En sommant sur $i = 1, 2, ..., n$, nous obtenons par conséquent :

$$F(b) - F(a) - 2\epsilon \cdot h \cdot n \le \int_a^b \le F(b) - F(a) + 2\epsilon \cdot h \cdot n$$

Mais comme $h \cdot n = b - a$, cela devient :

$$F(b) - F(a) - 2\epsilon \cdot (b-a) \le \int_a^b \le F(b) - F(a) + 2\epsilon \cdot (b-a)$$

Ces bornes devant être satisfaites pour tout $\epsilon > 0$, on en déduit que :

$$\int_a^b = F(b) - F(a)$$

On a donc finalement :

$$\int_a^b f(s) \, ds = \int_a^b \frac{dF}{ds}(s) \, ds = F(b) - F(a)$$

86.2.3 Primitive

Cette relation permet de calculer l'intégrale d'une fonction continue $f : t \mapsto f(t)$ lorsqu'on connaît une fonction F vérifiant :

$$\frac{dF}{dt} = f$$

On appelle « primitive » de f une telle fonction F.

86.2.4 Notation

On note aussi :

$$\boldsymbol{\Delta} F = \int dF$$

86.3 Polynômes

On sait que :

$$\frac{d}{dt}(t^n) = n \cdot t^{n-1}$$

Comme n est constante, on peut le réécrire :

$$\frac{d}{dt}\left(\frac{t^n}{n}\right) = t^{n-1}$$

ou, en posant $m = n - 1$:

$$\frac{d}{dt}\left(\frac{t^{m+1}}{m+1}\right) = t^m$$

L'intégrale s'écrit donc :

$$\int_a^b t^m \ dt = \frac{b^{m+1} - a^{m+1}}{m+1}$$

On a en particulier :

$$\int_0^x t^m \ dt = \frac{x^{m+1}}{m+1}$$

86.3.1 Exemples

$$\int_0^x t \ dt = \frac{x^2}{2}$$

$$\int_0^x t^2 \ dt = \frac{x^3}{3}$$

86.4 Valeur moyenne

86.4.1 Accroissements finis

Soit des réels distincts a, b vérifiant $a < b$, la fonction $f \in \mathrm{Cont}([a,b], \mathbb{R})$ et la fonction $F : [a,b] \mapsto \mathbb{R}$ définie par :

$$F(x) = \int_a^x f(t)\ dt$$

pour tout $x \in [a,b]$. Comme $F \in \mathrm{Cont}^1([a,b], \mathbb{R})$, le théorème des accroissements finis nous dit qu'on peut trouver un $c \in]a,b[$ tel que :

$$\partial F(c) = \frac{F(b) - F(a)}{b - a}$$

On sait que :

$$\partial F(c) = f(c)$$

On a aussi :

$$F(b) = \int_a^b f(t)\ dt$$

et :

$$F(a) = \int_a^a f(t)\ dt = \int_{\{a\}} f(t)\ dt$$

La mesure de Lebesgue du singleton $\{a\}$ étant nulle, l'intégrale s'annule et on a :

$$F(a) = 0$$

On a donc $F(b) - F(a) = F(b)$ et :

$$f(c) = \frac{1}{b - a} \int_a^b f(t)\ dt$$

86.4.2 Théorème de Cauchy

Soit des réels distincts a, b vérifiant $a < b$, les fonctions $f, g \in \mathrm{Cont}([a,b], \mathbb{R})$ et les fonction $F, G : [a,b] \mapsto \mathbb{R}$ définies par :

$$\begin{aligned} F(x) &= \int_a^x f(t)\ dt \\ G(x) &= \int_a^x g(t)\ dt \end{aligned}$$

pour tout $x \in [a,b]$. Comme $F, G \in \mathrm{Cont}^1([a,b], \mathbb{R})$, le théorème de Cauchy nous dit qu'on peut trouver un $c \in]a,b[$ tel que :

$$\partial F(c) \cdot \big[G(b) - G(a)\big] = \big[F(b) - F(a)\big] \cdot \partial G(c)$$

On sait que :

$$\partial F(c) = f(c)$$
$$\partial G(c) = g(c)$$

On a aussi :

$$\begin{aligned} F(b) &= \int_a^b f(t)\ dt \\ G(b) &= \int_a^b g(t)\ dt \end{aligned}$$

et :

$$\begin{aligned} F(a) &= \int_a^a f(t)\ dt = 0 \\ G(a) &= \int_a^a g(t)\ dt = 0 \end{aligned}$$

On en conclut que :

$$f(c) \int_a^b g(t)\ dt = g(c) \int_a^b f(t)\ dt$$

Si l'intégrale de g et $g(c)$ sont non nuls, on peut mettre cette relation sous la forme :

$$\frac{f(c)}{g(c)} = \frac{\int_a^b f(t)\ dt}{\int_a^b g(t)\ dt}$$

86.5 Intégration par parties

Soient $f, g \in \mathrm{Cont}^1(\mathbb{R}, \mathbb{R})$. On se rappelle que :

$$\partial(f \cdot g) = \partial f \cdot g + f \cdot \partial g$$

et comme on a :

$$\int_a^b \partial(f \cdot g)\ dx = f(b) \cdot g(b) - f(a) \cdot g(a)$$

on obtient la formule d'intégration par parties :

$$\int_a^b f(x) \cdot \partial g(x)\ dx = (f \cdot g)(b) - (f \cdot g)(a) - \int_a^b \partial f(x) \cdot g(x)\ dx$$

86.5.1 Stieltjes

Le résultat est également valable lorsqu'on utilise les mesures de Stieltjes associées à f et g :

$$\int_a^b f(x) \cdot dg(x) = \boldsymbol{\Delta}(f \cdot g) - \int_a^b g(x) \cdot df(x)$$

86.5.2 Dérivée constante

On considère le cas particulier où $\partial g = 1$. Une exemple de fonction g vérifiant cette propriété est simplement $g = \mathrm{Id}$. On a donc $g(x) = x$ et :

$$\int_a^b f(x)\ dx = \int_a^b f(x) \cdot 1\ dx = \int_a^b f(x) \cdot \partial g(x)\ dx$$

L'intégration par parties nous donne :

$$\int_a^b f(x)\ dx = f(b) \cdot b - f(a) \cdot a - \int_a^b \partial f(x) \cdot x\ dx$$

86.6 Changement de variable

Considérons une fonction $f \in \mathrm{Cont}(\mathbb{R}, \mathbb{R})$ et un changement de variable $x = \varphi(s)$ où $\varphi \in \mathrm{Hom}^1(\mathbb{R}, \mathbb{R})$. Soit la mesure de lebesgue $\mu([\alpha, \beta]) = \beta - \alpha$.

86.6.1 L'idée

$$\sum_i f_i \cdot \boldsymbol{\Delta} x_i = \sum_i f_i \cdot \frac{\boldsymbol{\Delta} x_i}{\boldsymbol{\Delta} s_i} \cdot \boldsymbol{\Delta} s_i$$

On devrait donc avoir par passage à la limite :

$$\int f\ dx = \int f\ \frac{dx}{ds}\ ds$$

86.6.2 La réalisation

On applique le même procédé qu'à la section 86.2. Si x est proche de y, on a :

$$\int_x^y \approx f(x) \cdot (y - x)$$

Posant $s = \varphi^{-1}(x)$ et $t = \varphi^{-1}(y)$, on a aussi :

$$y - x = \varphi(t) - \varphi(s) = \frac{d\varphi}{ds}(s) \cdot (t - s) + e(|s - t|)$$

où e converge plus vite que $s - t$ vers 0. On en conclut que :

$$\int_x^y \approx (f \circ \varphi)(s) \cdot \frac{d\varphi}{ds}(s) \cdot (t - s)$$

On remarque que le second membre est une approximation de l'intégrale de la fonction :

$$F(s) = (f \circ \varphi)(s) \cdot \frac{d\varphi}{ds}(s)$$

sur l'intervalle $[s, t] = [\varphi^{-1}(x), \varphi^{-1}(y)]$. Il ne nous reste plus qu'à sommer sur tous les petits intervalles $[x_{i-1}, x_i]$ et à passer à la limite $h = x_i - x_{i-1} \to 0$ pour obtenir :

$$\int_a^b f(x)\ dx = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} (f \circ \varphi)(s) \cdot \frac{d\varphi}{ds}(s)\ ds$$

Chapitre 87

Développements de Taylor

87.1 Polynômes de Taylor

Considérons un polynôme $p : \mathbb{R} \mapsto \mathbb{R}$ de degré n défini par :

$$p(x) = \sum_{i=0}^{n} \gamma_i \cdot x^i$$

pour tout $x \in \mathbb{R}$. Calculons ses dérivées :

$$\begin{aligned}
\partial p(x) &= \sum_{i=1}^{n} \gamma_i \cdot i \cdot x^{i-1} \\
\partial^2 p(x) &= \sum_{i=2}^{n} \gamma_i \cdot i \cdot (i-1) \cdot x^{i-2} \\
&\vdots \\
\partial^k p(x) &= \sum_{i=k}^{n} \gamma_i \cdot \frac{i!}{(i-k)!} \cdot x^{i-k} \\
&\vdots \\
\partial^n p(x) &= n! \cdot \gamma_n
\end{aligned}$$

Lorsqu'on évalue ces dérivées en 0, seuls les termes en $x^{k-k} = 1$ ne s'annulent pas. On obtient donc :

$$\partial^k p(0) = \frac{k!}{0!} \cdot \gamma_k = k! \cdot \gamma_k$$

ce qui nous donne l'expression des coefficients de p en fonction de ses dérivées en 0 :

$$\gamma_k = \frac{1}{k!} \cdot \partial^k p(0)$$

Le polynôme peut donc se réécrire :

$$p(x) = \sum_{i=0}^{n} \frac{1}{i!} \cdot \partial^i p(0) \cdot x^i$$

Cette expression est appelée développement de Taylor de p autour de 0.

Généralisation

Soit $a \in \mathbb{R}$. La fonction r définie par :

$$r(t) = p(t + a) = \sum_{i=0}^{n} \gamma_i \cdot (t + a)^i$$

pour tout $t \in \mathbb{R}$ est clairement un polynôme de degré n. On a $r(0) = p(a)$ et plus généralement :

$$\partial^i r(0) = \partial^i p(a)$$

pout tout $i \ge 0$. Le développement de Taylor de r autour de 0 s'écrit :

$$r(t) = \sum_{i=0}^{n} \frac{1}{i!} \cdot \partial^i r(0) \cdot t^i$$

ou encore :

$$r(t) = \sum_{i=0}^{n} \frac{1}{i!} \cdot \partial^i p(a) \cdot t^i$$

En posant $x = t + a$, on a $t = x - a$ et :

$$p(x) = p(t + a) = r(t)$$

Le développement devient :

$$p(x) = \sum_{i=0}^{n} \frac{1}{i!} \cdot \partial^i p(a) \cdot (x - a)^i$$

Cette expression est nommée développement de Taylor de p autour de a.

87.2 Opérateur de Taylor

Soit $\alpha, \beta \in \mathbb{R}$ avec $\alpha < \beta$, une fonction $f \in \mathrm{Cont}^N([\alpha, \beta], \mathbb{R})$ et $a \in [\alpha, \beta]$. Par analogie avec le développement de Taylor des polynômes, on définit l'opérateur de Taylor T_a^N par :

$$T_a^N(f)(x) = \sum_{k=0}^{N} \frac{1}{k!} \cdot \partial^k f(a) \cdot (x - a)^k$$

pour tout $x \in [\alpha, \beta]$.

87.2.1 Erreur

L'erreur E_a^N de l'opérateur T_a^N est donnée par :

$$E_a^N(f)(x) = f(x) - T_a^N(f)(x)$$

pour tout $x \in [\alpha, \beta]$.

87.2.2 Polynômes

Si p est un polynôme de degré N, on a bien entendu $T_a^N(p) = p$ pour tout $a \in \mathbb{R}$ et $E_a^N(p) = 0$.

87.3 Forme intégrale

87.3.1 Premier ordre

Soit $\alpha, \beta \in \mathbb{R}$ avec $\alpha < \beta$ et la fonction $f \in \mathrm{Cont}^2([\alpha, \beta], \mathbb{R})$. Le théorème fondamental nous dit que :

$$\int_a^x \partial f(t)\ dt = f(x) - f(a)$$

pour tout $a, x \in [\alpha, \beta]$. Appliquant le même théorème à la dérivée ∂f, on a aussi :

$$\int_a^x \partial^2 f(t)\ dt = \partial f(x) - \partial f(a)$$

Intégration par parties

Soit $u = \partial f$ et $v = \mathrm{Id}$. on a :

$$\int_a^x u(x)\ \partial v(x)\ dx = \int_a^x \partial f(t) \cdot 1\ dt = \int_a^x \partial f(t)\ dt$$

L'intégration par parties nous donne :

$$\int_a^x u(x)\ \partial v(x)\ dx = v(x)\ u(x) - v(a)\ u(a) - \int_a^x v(t)\ \partial u(t)\ dt$$

En tenant compte des définitions de u et v, on obtient :

$$\int_a^x \partial f(t)\ dt = x\ \partial f(x) - a\ \partial f(a) - \int_a^x t\ \partial^2 f(t)\ dt$$

Appliquons le théorème fondamental au membre de gauche :

$$f(x) - f(a) = x\ \partial f(x) - a\ \partial f(a) - \int_a^x t\ \partial^2 f(t)\ dt$$

ou encore :

$$f(x) = f(a) + x\ \partial f(x) - a\ \partial f(a) - \int_a^x t\ \partial^2 f(t)\ dt$$

En multipliant la relation :

$$\partial f(x) - \partial f(a) = \int_a^x \partial^2 f(t)\ dt$$

par x, on arrive au résultat :

$$x\ \partial f(x) = x\ \partial f(a) + \int_a^x x\ \partial^2 f(t)\ dt$$

L'expression de $f(x)$ devient alors :

$$f(x) = f(a) + x\ \partial f(a) + \int_a^x x\ \partial^2 f(t)\ dt - a\ \partial f(a) - \int_a^x t\ \partial^2 f(t)\ dt$$

et finalement :

$$f(x) = f(a) + (x - a) \cdot \partial f(a) + \int_a^x (x - t) \cdot \partial^2 f(t)\ dt$$

Le membre de droite est appelé développement de Taylor du premier ordre de f sous forme intégrale.

87.3.2 Second ordre

Soit $f \in \mathrm{Cont}^3([\alpha, \beta], \mathbb{R})$. Comme $\mathrm{Cont}^3 \subseteq \mathrm{Cont}^2$, f admet un développement de Taylor du premier ordre sous forme intégrale. Nous allons intégrer par parties le terme :

$$\int_a^x (x-t) \cdot \partial^2 f(t) \, dt$$

On sait que :

$$\frac{d}{dt} \left[\frac{1}{2} (x-t)^2 \right] = (x-t) \cdot (-1) = -(x-t)$$

Posons $u = \partial^2 f$ et :

$$v : t \mapsto \frac{1}{2} (x-t)^2$$

On a :

$$\int_a^x \partial v(t) \ u(t) \ dt = - \int_a^x (x-t) \ \partial^2 f(t) \ dt$$

et :

$$\int_a^x v(t) \ \partial u(t) \ dt = \frac{1}{2} \int_a^x (x-t)^2 \ \partial^3 f(t) \ dt$$

Enfin :

$$\begin{aligned}
\int_a^x \partial(v \cdot u)(t) \ dt &= \frac{1}{2} \ (x-x)^2 \ \partial^2 f(x) - \frac{1}{2} \ (x-a)^2 \ \partial^2 f(a) \\
&= 0 - \frac{1}{2} \ (x-a)^2 \ \partial^2 f(a) \\
&= -\frac{1}{2} \ (x-a)^2 \ \partial^2 f(a)
\end{aligned}$$

On en conclut que :

$$- \int_a^x (x-t) \ \partial^2 f(t) \ dt = -\frac{1}{2} \ (x-a)^2 \ \partial^2 f(a) - \frac{1}{2} \int_a^x (x-t)^2 \ \partial^3 f(t) \ dt$$

ou encore :

$$\int_a^x (x-t) \ \partial^2 f(t) \ dt = \frac{1}{2} \ (x-a)^2 \ \partial^2 f(a) + \frac{1}{2} \int_a^x (x-t)^2 \ \partial^3 f(t) \ dt$$

Le développement du premier ordre peut dont se réécrire :

$$f(x) = f(a) + (x-a) \ \partial f(a) + \frac{1}{2} \ (x-a)^2 \ \partial^2 f(a) + \frac{1}{2} \int_a^x (x-t)^2 \ \partial^3 f(t) \ dt$$

Le membre de droite est appelé développement du second ordre de f sous forme intégrale.

87.3.3 Ordre N

Soit $f \in \mathrm{Cont}^{N+1}([\alpha, \beta], \mathbb{R})$. On montre en intégrant par parties que :

$$\int_a^x (x-t)^{k-1} \ \partial^k f(t) \ dt = \frac{1}{k} \ (x-a)^k \ \partial^k f(a) + \frac{1}{k} \int_a^x (x-t)^k \ \partial^{k+1} f(t) \ dt$$

pour tout $k \in \mathbb{Z}[2, N]$. On en déduit par récurrence le développement de Taylor d'ordre N de f sous forme intégrale :

$$f(x) = \sum_{k=0}^{N} \frac{1}{k!} \cdot \partial^k f(a) \cdot (x-a)^k + \frac{1}{N!} \int_a^x (x-t)^N \ \partial^{N+1} f(t) \ dt$$

87.4 Erreur

On a :

$$E_a^N(f)(x) = f(x) - T_a^N(f)(x) = \frac{1}{N!} \int_a^x (x-t)^N \ \partial^{N+1} f(t) \ dt$$

En appliquant le théorème de Cauchy entre a et x aux fonctions F, G définies par :

$$\begin{aligned} F(z) &= \int_a^z (x-t)^N \ \partial^{N+1} f(t) \ dt \\ G(z) &= \int_a^z (x-t)^N \ dt \end{aligned}$$

pour tout $z \in [\alpha, \beta]$, on voit que l'on peut trouver un $c \in]a, x[$ si $a < x$ ou un $c \in]x, a[$ si $x < a$ tel que :

$$(x-c)^N \ F(x) = (x-c)^N \ \partial^{N+1} f(c) \ G(x)$$

ou encore :

$$\partial^{N+1} f(c) \ G(x) = F(x)$$

Comme :

$$\begin{aligned} G(x) = \int_a^x (x-t)^N \ dt &= -\big[(x-x)^{N+1} - (x-a)^{N+1}\big]/(N+1) \\ &= -\big[0 - (x-a)^{N+1}\big]/(N+1) \\ &= (x-a)^{N+1}/(N+1) \end{aligned}$$

on a :

$$\partial^{N+1} f(c) \ \frac{(x-a)^{N+1}}{N+1} = F(x) = \int_a^x (x-t)^N \ \partial^{N+1} f(t) \ dt$$

On en déduit que :

$$E_a^N(f)(x) = \partial^{N+1} f(c) \ \frac{(x-a)^{N+1}}{(N+1)!}$$

87.5 Forme différentielle

Soit une fonction $f \in \mathrm{Cont}^{N+1}([\alpha, \beta], \mathbb{R})$ et $a, x \in [\alpha, \beta]$. On définit la fonction $F : [\alpha, \beta] \mapsto \mathbb{R}$ par :

$$\begin{aligned} F(t) &= \sum_{k=0}^{N} \frac{1}{k!} \cdot \partial^k f(t) \cdot (x-t)^k \\ &= f(t) + \partial f(t)\ (x-t) + \partial^2 f(t)\ \frac{(x-t)^2}{2} + ... \end{aligned}$$

pour tout $t \in [\alpha, \beta]$. On a :

$$F(x) = f(x) + \partial f(x)\ (x-x) + \partial^2 f(x)\frac{(x-x)^2}{2} + ... = f(x) + 0 = f(x)$$

et :

$$F(a) = f(a) + \partial f(a)\ (x-a) + \partial^2 f(a)\frac{(x-a)^2}{2} + ... = T_a^N(f)(x)$$

La dérivée de F s'écrit :

$$\begin{aligned} \partial F(t) = \partial f(t) &+ \big[\partial f(t)\ (-1) + \partial^2 f(t)\ (x-t)\big] \\ &+ \left[-\partial^2 f(t)\ (x-t) + \partial^3 f(t)\ \frac{(x-t)^2}{2}\right] \\ &\quad ... \\ &+ \left[-\partial^{N-1} f(t)\ \frac{(x-t)^{N-2}}{(N-2)!} + \partial^N f(t)\ \frac{(x-t)^{N-1}}{(N-1)!}\right] \\ &+ \left[-\partial^{N} f(t)\ \frac{(x-t)^{N-1}}{(N-1)!} + \partial^{N+1} f(t)\ \frac{(x-t)^{N}}{N!}\right] \end{aligned}$$

On voit que tous les termes s'annulent sauf le dernier, et :

$$\partial F(t) = \partial^{N+1} f(t)\ \frac{(x-t)^N}{N!}$$

Soit $G \in \mathrm{Cont}^1([\alpha, \beta], \mathbb{R})$. On peut appliquer le théorème de Cauchy à F et G entre a et x. On dispose alors d'un $c \in]a, x[$ si $a < x$ ou d'un $c \in]x, a[$ si $x < a$ tel que :

$$\partial F(c)\ \big[G(x) - G(a)\big] = \big[F(x) - F(a)\big]\ \partial G(c)$$

On a :

$$F(x) - F(a) = f(x) - T_a^N(f)(x) = E_a^N(f)(x)$$

On en conclut que :

$$E_a^N(f)(x)\ \partial G(c) = \partial^{N+1} f(c)\ \frac{(x-c)^N}{N!}\ \big[G(x) - G(a)\big]$$

87.5.1 Forme de Lagrange

Soit le choix :

$$G : t \mapsto (x-t)^{N+1}$$

on a :

$$G(x) = (x-x)^{N+1} = 0$$

et :

$$G(a) = (x - a)^{N+1}$$

La dérivée s'écrit :

$$\partial G(t) = -(N + 1)\ (x - t)^N$$

La relation de Cauchy devient :

$$-E_a^N(f)(x)\ (N + 1)\ (x - c)^N = -\partial^{N+1} f(c)\ \frac{(x - c)^N}{N!}\ (x - a)^{N+1}$$

On a donc l'expression de l'erreur :

$$E_a^N(f)(x) = \partial^{N+1} f(c)\ \frac{(x - a)^{N+1}}{(N + 1)!}$$

87.5.2 Forme de Cauchy

Soit le choix :

$$G : t \mapsto t - a$$

on a :

$$G(x) = x - a$$

et :

$$G(a) = a - a = 0$$

La dérivée s'écrit :

$$\partial G(t) = 1$$

La relation de Cauchy devient :

$$E_a^N(f)(x) = \partial^{N+1} f(c)\ \frac{(x - c)^N}{N!}\ (x - a)$$

87.6 Borne

Soit $f \in \mathrm{Cont}^{N+1}([\alpha, \beta], \mathbb{R})$. Comme $\partial^{N+1} f$ est continue, sa norme $\|.\|_\infty$ sur $[\alpha, \beta]$ est finie et on a :

$$\left|E_a^N(f)(x)\right| \le \left\|\partial^{N+1} f\right\|_\infty\ \frac{|x - a|^{N+1}}{(N + 1)!}$$

On peut majorer cette expression en constatant que :

$$|x - a| \le |\beta - \alpha|$$

La borne de l'erreur devient alors :

$$\left|E_a^N(f)(x)\right| \le \left\|\partial^{N+1} f\right\|_\infty\ \frac{|\beta - \alpha|^{N+1}}{(N + 1)!}$$

Le membre de droite ne dépendant pas de x, on a :

$$\left\|E_a^N(f)\right\|_\infty \le \left\|\partial^{N+1} f\right\|_\infty\ \frac{|\beta - \alpha|^{N+1}}{(N + 1)!}$$

87.7 Convergence

Soit $f \in \mathrm{Cont}^\infty([\alpha, \beta], \mathbb{R})$. Si on peut trouver un $\sigma \in \mathbb{R}$ tel que :

$$\|\partial^n f\|_\infty \le \sigma$$

pour tout $n \in \mathbb{N}$, on a :

$$\left\|E_a^N(f)\right\|_\infty \le \sigma \ \frac{|\beta - \alpha|^{N+1}}{(N+1)!}$$

On en conclut que :

$$0 \le \lim_{N\to\infty} \left\|E_a^N(f)\right\|_\infty \le \sigma \ \lim_{N\to\infty} \frac{|\beta - \alpha|^{N+1}}{(N+1)!} = 0$$

L'erreur converge vers zéro quand N tend vers l'infini :

$$\lim_{N\to\infty} \left\|E_a^N(f)\right\|_\infty = 0$$

87.8 Dimension n

87.8.1 Premier ordre

Soit $\Omega \subseteq \mathbb{R}^n$, la fonction $f \in \mathrm{Cont}^1(\Omega, \mathbb{R})$ et les vecteurs $u, v \in \mathbb{R}^n$ tels que le segment $[u, v]$ est inclus dans Ω. On définit la fonction $\lambda : [0, 1] \mapsto \mathbb{R}^n$ associée au segment $[u, v]$ par :

$$\lambda(s) = u + s \cdot (v - u)$$

pour tout $s \in [0, 1]$, ainsi que la fonction $\varphi = f \circ \lambda$ qui vérifie :

$$\varphi(s) = (f \circ \lambda)(s) = f(u + s \cdot (v - u))$$

pour tout $s \in [0, 1]$. On pose :

$$h = v - u$$

On a :

$$\varphi(0) = f(u)$$

La dérivée s'écrit :

$$\partial\varphi(s) = \sum_i \partial_i f(u + s \cdot h) \cdot h_i$$

ou, en utilisant la notation matricielle :

$$\partial\varphi(s) = \partial f(u + s \cdot h) \cdot h$$

On a la valeur particulière :

$$\partial\varphi(0) = \partial f(u) \cdot h$$

La dérivée seconde s'écrit :

$$\partial^2\varphi(s) = \sum_{i,j} h_j \cdot \partial^2_{ji} f(u + s \cdot h) \cdot h_i$$

ou, en utilisant la notation matricielle :

$$\partial^2\varphi(s) = h^\star \cdot \partial^2 f(u + s \cdot h) \cdot h$$

Le développement du premier ordre de φ autour de 0 s'écrit donc :

$$\varphi(s) = f(u) + s \cdot \partial f(u) \cdot h + E^1_u(s, h)$$

avec :

$$E^1_u(s, h) = h^\star \cdot \partial^2 f(u + c \cdot h) \cdot h \cdot \frac{(c-0)^2}{2} = h^\star \cdot \partial^2 f(u + c \cdot h) \cdot h \cdot \frac{c^2}{2}$$

pour un certain $c \in]0, s[$. Mais comme :

$$\varphi(1) = f(u + h) = f(v)$$

on en déduit le développement de f :

$$f(v) = f(u) + \partial f(u) \cdot (v - u) + \mathcal{E}^1_u(h)$$

avec :

$$\mathcal{E}^1_u(h) = h^\star \cdot \partial^2 f(u + c \cdot h) \cdot h \cdot \frac{c^2}{2}$$

pour un certain $c \in]0, 1[$.

Borne

Soit :

$$M^2 = \max_{i,j} \left\| \partial^2_{ij} f \right\|_\infty$$

On a :

$$\left| \mathcal{E}^1_u(h) \right| \le \frac{1}{2} \cdot n^2 \cdot M^2 \cdot \|h\|^2$$

87.8.2 Second ordre

Soit $f \in \mathrm{Cont}^3(\Omega, \mathbb{R})$. Avec les mêmes notations que précédemment, on a :

$$\partial^2\varphi(0) = h^\star \cdot \partial^2 f(u) \cdot h$$

La dérivée tierce de φ s'écrit :

$$\partial^3\varphi(s) = \sum_{i,j,k} \partial^3_{kji} f(u + s \cdot h) \cdot h_i \cdot h_j \cdot h_k$$

ou, en utilisant la notation tensorielle :

$$\partial^3\varphi(s) = \partial^3 f(u + s \cdot h) : h \otimes h \otimes h$$

Le développement du second ordre de φ autour de 0 s'écrit :

$$\varphi(s) = f(u) + s\ \partial f(u) \cdot h + \frac{s^2}{2}\ h^\star \cdot \partial^2 f(u) \cdot h + E^2_u(s, h)$$

avec :

$$E_u^2(s,h) = \partial^3 f(u + c \cdot h) : h \otimes h \otimes h \cdot \frac{c^3}{6}$$

pour un certain $c \in \left]0, s\right[$. Mais comme :

$$\varphi(1) = f(u + h) = f(v)$$

on en déduit le développement de f :

$$f(v) = f(u) + \partial f(u) \cdot h + h^\star \cdot \partial^2 f(u) \cdot h + \mathcal{E}_u^2(h)$$

avec :

$$\mathcal{E}_u^2(h) = \partial^3 f(u + c \cdot h) : h \otimes h \otimes h \cdot \frac{c^3}{6}$$

pour un certain $c \in \left]0, 1\right[$.

Borne

Soit :

$$M^3 = \max_{i,j,k} \left\| \partial^3_{ijk} f \right\|_\infty$$

On a :

$$\left| \mathcal{E}_u^2(h) \right| \le \frac{1}{6} \cdot n^3 \cdot M^3 \cdot \|h\|^3$$

87.9 Notation

Soit la fonction $E : \Omega \subseteq \mathbb{R}^m \mapsto \mathbb{R}^n$, la fonction $b : \mathbb{R} \mapsto \mathbb{R}$ et le vecteur $h \in \Omega$. On note $E \sim o\left(b(h)\right)$, ou on dit que E est en $o\left(b(h)\right)$, pour signifier que :

$$\lim_{h \to 0} \frac{\|E(h)\|}{b(\|h\|)} = 0$$

On note $E \sim O\left(b(h)\right)$, ou on dit que E est en $O\left(b(h)\right)$, pour signifier qu'il existe $M \in \mathbb{R}$ tel que :

$$\|E(h)\| \le M \cdot b(\|h\|)$$

pour tout $h \in \Omega$.

87.9.1 Puissance

Une famille de fonction souvent employée est la puissance :

$$b_k : x \mapsto x^k$$

pour un certain $k \in \mathbb{N}$. On a alors $o\left(h^k\right)$ si :

$$\lim_{h \to 0} \frac{\|E(h)\|}{\|h\|^k} = 0$$

et $O\left(h^k\right)$ si :

$$\|E(h)\| \le M \cdot \|h\|^k$$

87.9.2 Relation

Si $E \sim O\left(h^k\right)$, on a :

$$0 \le \lim_{h \to 0} \frac{\|E(h)\|}{\|h\|^{k-1}} \le \lim_{h \to 0} \frac{M \ \|h\|^k}{\|h\|^{k-1}} = 0$$

d'où :

$$\lim_{h \to 0} \frac{\|E(h)\|}{\|h\|^{k-1}} = 0$$

et $E \sim o\left(h^{k-1}\right)$.

87.9.3 Cas particulier

Le $O(1)$ implique une erreur bornée en valeur absolue, le $o(1)$ implique la continuité et le $o(h)$ la différentiabilité.

87.9.4 Développement de Taylor

Pour toute fonction $f \in \mathrm{Cont}^{N+1}(\Omega, \mathbb{R}^n)$, l'erreur $E_a^N(f)$ du développement de Taylor d'ordre N est en $O\left(h^{N+1}\right)$.

87.10 Extrapolation de Richardson

Supposons qu'une fonction v nous donne une approximation de V respectant :

$$v(h) \approx V + C \cdot h^m + O(h^{m+1})$$

pour un certain $C \in \mathbb{R}$ et pour tout $h \in [0, R] \subseteq \mathbb{R}$. L'entier m est appelé l'ordre de l'approximation. Supposons que l'on dispose de deux estimations de $V_1 = v(h)$ et $V_2 = v(h/k)$. On a alors :

$$V_1 = v(h) = V + C \cdot h^m + O(h^{m+1})$$
$$V_2 = v\left(h/k\right) = V + C \cdot \left(\frac{h}{k}\right)^m + O(h^{m+1})$$

On se sert de la première équation pour obtenir une expression de $C \cdot h^m$:

$$C \cdot h^m = V_1 - V + O(h^{m+1})$$

Posons :

$$r = \frac{1}{k^m}$$

On a alors :

$$V_2 = V + r \cdot C \cdot h^m + O(h^{m+1}) = V + r \cdot (V_1 - V) + O(h^{m+1})$$

On en conclut que :

$$(1 - r) \cdot V = V_2 - r \cdot V_1 + O(h^{m+1})$$

Ce qui nous donne l'approximation :

$$V = \frac{V_2 - r \cdot V_1}{1 - r} + O(h^{m+1})$$

Cette approximation est plus précise, car l'erreur n'est plus en $O(h^m)$ mais en $O(h^{m+1})$. On appelle cette technique l'extrapolation de Richardson.

Cas particulier

Un cas particulier intéressant est celui où l'approximation est d'ordre 1 et où $k = 2$. On a alors :

$$V = 2V_2 - V_1 + O(h^2) = V_2 + (V_2 - V_1) + O(h^2)$$

ce qui revient à faire l'approximation $V - V_2 \approx V_2 - V_1$.

Chapitre 88

Développements d'Hadamard

Dépendances

— Chapitre ?? : Les différentielles
— Chapitre 72 : Les intégrales

88.1 Lemme de Hadamard

Soit la fonction $f : \mathbb{R}^m \to \mathbb{R}^n$ et les vecteurs $u, v \in \mathbb{R}^m$. On définit la fonction $\lambda : [0,1] \mapsto \mathbb{R}^m$ associée au segment $[u,v] \subseteq \mathbb{R}^m$ par :

$$\lambda(s) = u + s \cdot (v - u)$$

pour tout $s \in [0,1]$. On a bien entendu $\lambda(0) = u$ et $\lambda(1) = v$. On définit également la fonction $\varphi = f \circ \lambda$ qui vérifie :

$$\varphi(s) = (f \circ \lambda)(s) = f(u + s \cdot (v - u))$$

pour tout $t \in [0,1]$. On voit que $\varphi(0) = f(u)$ et $\varphi(1) = f(v)$. Donc, en termes de composantes dans $\mathbb{R}^n$, on a :

$$f_i(v) - f_i(u) = \varphi_i(1) - \varphi_i(0) = \int_0^1 \frac{d\varphi_i}{ds}(s)\ ds$$

où $i \in \{1, 2, ..., n\}$.

Voyons quelle est la forme de la dérivée :

$$\begin{aligned}
\frac{d\varphi_i}{ds}(s) &= \sum_j \partial_j f_i(u + s \cdot (v - u)) \cdot \partial \lambda_j(s) \\
&= \sum_j \partial_j f_i(u + s \cdot (v - u)) \cdot (v_j - u_j)
\end{aligned}$$

où $j \in \{1, 2, ..., m\}$. Si nous définissons :

$$G_{ij}(u, v) = \int_0^1 \partial_j f_i(u + s \cdot (v - u))\ ds$$

nous obtenons alors l'expression de la variation :

$$f_i(v) - f_i(u) = \sum_j G_{ij}(u, v) \cdot (v_j - u_j)$$

En termes matriciels :

$$G(u,v) = \big[G_{ij}(u,v)\big]_{i,j} = \left[\int_0^1 \partial_j f_i(u + s \cdot (v-u)) \ ds\right]_{i,j}$$

est donc l'intégrale de la Jacobienne :

$$G(u,v) = \int_0^1 \partial f(u + s \cdot (v-u)) \ ds$$

et :

$$f(v) - f(u) = G(u,v) \cdot (v-u)$$

88.2 Développement du second ordre

Soit la fonction $f \in \mathrm{Cont}^2(\mathbb{R}^n, \mathbb{R})$ et les vecteurs $a, h \in \mathbb{R}^n$. On définit la fonction $\lambda : [0,1] \mapsto \mathbb{R}^n$ associée au segment $[a, a+h]$:

$$\lambda(s) = a + s \cdot h$$

pour tout $s \in [0,1]$. Le lemme de Hadamard nous dit que :

$$f(a+h) - f(a) = \int_0^1 \partial f(a + s \cdot h) \cdot h \ ds$$

Par définition de la dérivée seconde, on a :

$$\partial_i f(a + s \cdot h) = \partial_i f(a) + \sum_j \partial^2_{ji} f(a) \cdot h_j \cdot s + e_i(s \cdot h)$$

où l'erreur e vérifie :

$$\lim_{h \to 0} \frac{\|e(h)\|}{\|h\|} = 0$$

L'intégrale s'écrit alors :

$$f(a+h) - f(a) = \sum_i \int_0^1 \left[\partial_i f(a) + \sum_j \partial^2_{ji} f(a) \cdot h_j \cdot s + e_i(s \cdot h)\right] \cdot h_i \ ds$$

La grandeur $\partial_i f(a) \cdot h_i$ ne dépendant pas de s, on a :

$$\int_0^1 \partial_i f(a) \cdot h_i \ ds = \partial_i f(a) \cdot h_i \cdot (1 - 0) = \partial_i f(a) \cdot h_i$$

D'un autre coté, comme $s^2/2$ est une primitive de s, on a :

$$\int_0^1 s \ ds = \frac{1}{2} \cdot (1^2 - 0^2) = \frac{1}{2}$$

et donc :

$$\int_0^1 \partial^2_{ji} f(a) \cdot h_j \cdot h_i \cdot s \ ds = \frac{1}{2} \partial^2_{ji} f(a) \cdot h_j \cdot h_i$$

Posons :

$$\mathcal{E}_2(h) = \sum_i \int_0^1 e_i(s \cdot h) \cdot h_i \ ds$$

On a alors :

$$f(a+h) - f(a) = \sum_i \partial_i f(a) \cdot h_i + \frac{1}{2} \sum_{i,j} h_j \cdot \partial^2_{ji} f(a) \cdot h_i + \mathcal{E}_2(h)$$

En termes matriciels, cette expression fait intervenir la Jacobienne et la Hessienne :

$$f(a+h) - f(a) = \partial f(a) \cdot h + \frac{1}{2} \ h^\star \cdot \partial^2 f(a) \cdot h + \mathcal{E}_2(h)$$

88.2.1 Comportement de l'erreur

Nous savons que, pour toute précision $\epsilon > 0$, nous pouvons trouver un $\delta > 0$ tel que :

$$\frac{\|e(h)\|}{\|h\|} \le \epsilon$$

pour tout h vérifiant $\|h\| \le \delta$. Comme $|e_i| \le \|e\|$ et $|h_i| \le \|h\|$, on a alors :

$$\begin{aligned} |\mathcal{E}_2(h)| &\le \sum_i \left| \int_0^1 e_i(s \cdot h) \cdot h_i \ ds \right| \\ &\le n \cdot \epsilon \cdot \|h\|^2 \end{aligned}$$

L'erreur décroît donc plus vite que $\|h\|^2$:

$$\lim_{h \to 0} \frac{|\mathcal{E}_2(h)|}{\|h\|^2} = 0$$

88.2.2 Dérivées ordinaires

Lorsque $n = 1$, le développement est simplement :

$$f(a+h) = f(a) + \frac{df}{dx}(a) \cdot h + \frac{d^2 f}{dx^2}(a) \cdot \frac{h^2}{2} + \mathcal{E}_2(h)$$

On constate qu'il est analogue au développement de Taylor d'ordre deux autour de a.

88.3 Développement du troisième ordre

Soit la fonction $f \in \mathrm{Cont}^3(\mathbb{R}^n, \mathbb{R})$ et les vecteurs $a, h \in \mathbb{R}^n$. En évaluant le développement du second ordre de chaque $\partial_i f$, on a :

$$\partial_i f(a + s \cdot h) = \partial_i f(a) + \sum_j \partial_{ji} f(a) \cdot h_j \cdot s + \sum_{j,k} h_k \cdot \partial^3_{kji} f(a) \cdot h_j \cdot \frac{s^2}{2} + e_i(h)$$

où $e \sim o(h^2)$. En intégrant, nous obtenons une estimation de la variation de f :

$$f(a+h) - f(a) = \sum_i \int_0^1 \partial_i f(a + s \cdot h) \cdot h_i \ ds$$

Posons :

$$\begin{aligned}
I_1(h) &= \sum_i \int_0^1 \partial_i f(a) \cdot h_i \; ds \\
I_2(h) &= \sum_{i,j} \int_0^1 h_j \cdot \partial_{ji} f(a) \cdot h_i \cdot s \; ds \\
I_3(h) &= \frac{1}{2} \sum_{i,j,k} \int_0^1 h_k \cdot \partial^3_{kji} f(a) \cdot h_j \cdot h_i \cdot s^2 \; ds \\
\mathcal{E}_3(h) &= \sum_i \int_0^1 e_i(h) \cdot h_i \; ds
\end{aligned}$$

Comme $s^3/3$ est une primitive de s^2, on a :

$$\int_0^1 s^2 \; ds = \frac{1}{3} \cdot (1^3 - 0^3) = \frac{1}{3}$$

Les intégrales s'écrivent donc :

$$\begin{aligned}
I_1(h) &= \sum_i \partial_i f(a) \cdot h_i \\
I_2(h) &= \frac{1}{2} \sum_{i,j} h_i \cdot \partial_{ji} f(a) \cdot h_j \\
I_3(h) &= \frac{1}{6} \sum_{i,j,k} \partial^3_{kji} f(a) \cdot h_i \cdot h_j \cdot h_k
\end{aligned}$$

et la variation de f est donnée par :

$$f(a+h) - f(a) = I_1(h) + I_2(h) + I_3(h) + \mathcal{E}_3(h)$$

En terme de notations tensorielles, on peut l'écrire symboliquement :

$$f(a+h) - f(a) = \partial f(a) \cdot h + \frac{1}{2} h^\star \cdot \partial^2 f(a) \cdot h + \frac{1}{6} \left\langle \partial^3 f(a) \odot h \otimes h \otimes h \right\rangle_3$$

88.3.1 Comportement de l'erreur

Nous savons que, pour toute précision $\epsilon > 0$, nous pouvons trouver un $\delta > 0$ tel que :

$$\frac{\|e(h)\|}{\|h\|^2} \le \epsilon$$

pour tout h vérifiant $\|h\| \le \delta$. Comme $|e_i(h)| \le \|e(h)\|$ et $|h_i| \le \|h\|$, on a :

$$\begin{aligned}
|\mathcal{E}_3(h)| &\le \sum_i \left| \int_0^1 e_i(h) \cdot h_i \; ds \right| \\
&\le n \cdot \epsilon \cdot \|h\|^3
\end{aligned}$$

L'erreur $|\mathcal{E}_3(h)|$ est donc en $o\left(h^3\right)$.

88.3.2 Dérivées ordinaires

Lorsque $n = 1$, le développement est simplement :

$$f(a+h) = f(a) + \frac{df}{dx}(a) \cdot h + \frac{d^2 f}{dx^2}(a) \cdot \frac{h^2}{2} + \frac{d^3 f}{dx^3} \cdot \frac{h^3}{6} + \mathcal{E}_3(h)$$

On constate qu'il est analogue au développement de Taylor d'ordre trois autour de a.

Chapitre 89

Symétrie

Soit la fonction $f \in \mathrm{Cont}^2(\mathbb{R}^2, \mathbb{R})$. Posons $\partial_x = \partial_1$ et $\partial_y = \partial_2$. Nous allons tenter d'évaluer les dérivées secondes $\partial_{xy} = \partial_{12}$ et $\partial_{yx} = \partial_{21}$. On note :

$$\begin{aligned}
\varphi_{11} &= \varphi(x, y) \\
\varphi_{21} &= \varphi(x + h, y) \\
\varphi_{12} &= \varphi(x, y + h) \\
\varphi_{22} &= \varphi(x + h, y + h)
\end{aligned}$$

où $\varphi = f$ ou une de ses dérivées. Comme $\partial_{xy} = \partial_x \partial_y$ et $\partial_{yx} = \partial_y \partial_x$, on a par définition :

$$\Delta_{xy} = \partial_y f_{21} - \partial_y f_{11} = \partial_{xy} f_{11} \cdot h + o(h)$$
$$\Delta_{yx} = \partial_x f_{12} - \partial_x f_{11} = \partial_{yx} f_{11} \cdot h + o(h)$$

Multiplié par h, cela devient :

$$\Delta_{xy} \cdot h = \partial_{xy} f_{11} \cdot h^2 + o(h^2)$$
$$\Delta_{yx} \cdot h = \partial_{yx} f_{11} \cdot h^2 + o(h^2)$$

On dispose également des développements d'ordre deux :

$$f_{22} - f_{21} = \partial_y f_{21} \cdot h + \partial_{yy} f_{21} \cdot \frac{h^2}{2} + o(h^2)$$
$$f_{12} - f_{11} = \partial_y f_{11} \cdot h + \partial_{yy} f_{11} \cdot \frac{h^2}{2} + o(h^2)$$
$$f_{22} - f_{12} = \partial_x f_{12} \cdot h + \partial_{xx} f_{12} \cdot \frac{h^2}{2} + o(h^2)$$
$$f_{12} - f_{11} = \partial_x f_{11} \cdot h + \partial_{xx} f_{11} \cdot \frac{h^2}{2} + o(h^2)$$

On en conclut que :

$$\Delta_{xy} \cdot h = D + \Delta_{yy} + o(h^2)$$
$$\Delta_{yx} \cdot h = D + \Delta_{xx} + o(h^2)$$

où l'on a posé :

$$D = f_{22} - f_{21} - f_{12} + f_{11}$$
$$\Delta_{yy} = (\partial_{yy} f_{11} - \partial_{yy} f_{21}) \cdot h^2$$
$$\Delta_{xx} = (\partial_{xx} f_{11} - \partial_{xx} f_{12}) \cdot h^2$$

Par continuité de $\partial_{xx} f$ et de $\partial_{yy} f$, on a :

$$\lim_{h\to 0} \frac{\Delta_{yy}}{h^2} = \lim_{h\to 0} (\partial_{yy} f_{11} - \partial_{yy} f_{21}) = 0$$
$$\lim_{h\to 0} \frac{\Delta_{xx}}{h^2} = \lim_{h\to 0} (\partial_{xx} f_{11} - \partial_{xx} f_{12}) = 0$$

On en conclut que $\Delta_{xx}, \Delta_{yy} \sim o(h^2)$. Comme la somme de deux erreurs en $o(h^2)$ donne également une erreur en $o(h^2)$, on a :

$$\Delta_{xy} \cdot h = D + o(h^2) + o(h^2) = D + o(h^2)$$
$$\Delta_{yx} \cdot h = D + o(h^2) + o(h^2) = D + o(h^2)$$

et :

$$\partial_{xy} f_{11} \cdot h^2 = D + o(h^2)$$
$$\partial_{yx} f_{11} \cdot h^2 = D + o(h^2)$$

On en conclut que la différence $\partial_{xy} f_{11} - \partial_{yx} f_{11} \sim o(1)$ tend vers 0 avec h, ce qui n'est possible que si :

$$\partial_{xy} f_{11} = \partial_{yx} f_{11}$$

Nous avons donc prouvé que :

$$\partial_{xy} f(x, y) = \partial_{yx} f(x, y)$$

Généralisation

On peut bien entendu généraliser à une fonction $f \in \mathrm{Cont}^2(\mathbb{R}^n, \mathbb{R})$. On a alors :

$$\partial_{ij} f = \partial_{ji} f$$

où $i, j \in \{1, 2, ..., n\}$. Si $H = \partial^2 f$, on écrit aussi ce résultat sous la forme :

$$H^\star = H$$

89.1 Dérivation par rapport à un paramètre

Nous allons a présent examiner ce qu'il se passe lorsque les bornes de l'intervalle d'intégration ($a, b : \mathbb{R} \mapsto \mathbb{R}$) et la fonction à intégrer ($f : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$) varient par rapport à un paramètre. Soit la fonction $I : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$I(t) = \int_{a(t)}^{b(t)} f(s, t)\ ds$$

Pour une valeur donnée de t, posons :

$$\phi_t(s) = f(s, t)$$

L'intégrale de ϕ_t peut s'évaluer si nous connaissons une primitive ψ_t telle que :

$$\frac{d\psi_t}{ds}(s) = \phi_t(s)$$

Mais cette expression consiste à évaluer la variation de ψ lorsque s varie, t étant fixé. Cela revient donc à une dérivée partielle par rapport à s. Donc, si nous connaissons une fonction F telle que :

$$\frac{\partial F}{\partial s}(s,t) = f(s,t)$$

nous pouvons réécrire l'intégrale :

$$\int_{a(t)}^{b(t)} f(s,t)\ ds = F(b(t),t) - F(a(t),t)$$

Il ne nous reste plus alors qu'à évaluer la dérivée de I par rapport à t en utilisant la règle des compositions de fonctions :

$$\frac{dI}{dt}(t) = \frac{\partial F}{\partial s}(b(t),t) \cdot \frac{db}{dt}(t) + \frac{\partial F}{\partial t}(b(t),t) - \frac{\partial F}{\partial s}(a(t),t) \cdot \frac{db}{dt}(t) - \frac{\partial F}{\partial t}(a(t),t)$$

Si $F \in \mathrm{Cont}^2(\mathbb{R}^2, \mathbb{R})$, la symétrie des dérivées secondes nous permet d'écrire :

$$\frac{\partial F}{\partial t} = \frac{\partial}{\partial t}\left[\frac{\partial f}{\partial s}\right] = \frac{\partial}{\partial s}\left[\frac{\partial f}{\partial t}\right]$$

La dérivée partielle de F par rapport à t est donc une primitive de la dérivée partielle de f par rapport à t. On en déduit que :

$$\int_{\alpha}^{\beta} \frac{\partial f}{\partial t}(s,t)\ ds = \frac{\partial F}{\partial t}(\beta,t) - \frac{\partial F}{\partial t}(\alpha,t)$$

pour tout $\alpha, \beta \in \mathbb{R}$. Pour un t fixé quelconque, on peut poser $\alpha = a(t)$ et $\beta = b(t)$. Il vient alors :

$$\frac{dI}{dt}(t) = \frac{\partial F}{\partial s}(b(t),t) \cdot \frac{db}{dt}(t) - \frac{\partial F}{\partial s}(a(t),t) \cdot \frac{db}{dt}(t) + \int_{a(t)}^{b(t)} \frac{\partial f}{\partial t}(s,t)\ ds$$

89.2 Différences finies

Soit une fonction $f : \mathbb{R}^2 \mapsto \mathbb{R}^m$ deux fois continument dérivable. Nous allons voir comment évaluer des approximations des dérivées premières et secondes de f. On note $\partial_x = \partial_1$ et $\partial_y = \partial_2$. On choisit les réels x, y et la variation non nulle $h \in \mathbb{R}$.

Dérivées premières

Soustrayons les développements d'ordre deux :

$$f(x+h,y) \approx f(x,y) + h \cdot \partial_x f(x,y) + \frac{h^2}{2} \cdot \partial^2 f(x,y)$$
$$f(x-h,y) \approx f(x,y) - h \cdot \partial_x f(x,y) + \frac{h^2}{2} \cdot \partial^2 f(x,y)$$

On obtient :

$$f(x+h,y) - f(x-h,y) \approx 2h \cdot \partial_x f(x,y)$$

et donc :

$$\partial_x f(x,y) \approx \frac{f(x+h,y) - f(x-h,y)}{2h}$$

L'erreur est en $o(h) = o(h^2)/h$. En procédant de même avec y, on obtient :

$$\partial_y f(x,y) \approx \frac{f(x,y+h) - f(x,y-h)}{2h}$$

Dérivées secondes

On additionne cette fois les mêmes développements et on obtient :

$$f(x+h,y) + f(x-h,y) \approx 2f(x,y) + \frac{h^2}{2} \cdot \partial^2 f(x,y) + o(h^2)$$

et donc :

$$\partial^2_{xx} f(x,y) \approx \frac{f(x+h,y) - 2f(x,y) + f(x-h,y)}{h^2}$$

L'erreur est en $o(1) = o(h^2)/h^2$, et donc aussi petite que l'on veut pourvu que $h \neq 0$ soit suffisamment petit. En procédant de même avec y, on obtient :

$$\partial^2_{yy} f(x,y) \approx \frac{f(x,y+h) - 2f(x,y) + f(x,y-h)}{h^2}$$

On vérifie également en évaluant les développements en $(x \pm h, y \pm h)$ que :

$$\partial^2_{xy} f(x,y) \approx \frac{f(x+h,y+h) - f(x+h,y-h) - f(x-h,y+h) + f(x-h,y-h)}{h^2}$$

La dernière dérivée seconde s'évalue approximativement par $\partial^2_{yx} f(x,y) = \partial^2_{xy} f(x,y)$.

Généralisation

Nous allons voir comment généraliser ces résultats aux dérivées ∂_{ij} d'une fonction $F : \mathbb{R}^n \mapsto \mathbb{R}^m$. Soit $u \in \mathbb{R}^n$ et les vecteurs de la base canonique $\mathfrak{c}_i \in \mathbb{R}^n$. On définit les fonctions $f_{ij} : \mathbb{R}^n \mapsto \mathbb{R}^m$ par :

$$f_{ij}(x,y) = F(u + x \cdot \mathfrak{c}_i + y \cdot \mathfrak{c}_j)$$

On a clairement :

$$\begin{aligned} \partial_i F(u) &= \partial_x f_{ij}(0,0) \\ \partial_{ii} F(u) &= \partial_{xx} f_{ij}(0,0) \\ \partial_{ij} F(u) &= \partial_{xy} f_{ij}(0,0) \\ \partial_{jj} F(u) &= \partial_{yy} f_{ij}(0,0) \end{aligned}$$

Il suffit donc d'utiliser les méthodes d'approximations des dérivées de f_{ij} pour approximer les dérivées de F.

Chapitre 90

Distributions

90.1 Dépendances

- Chapitre ?? : Les fonctions
- Chapitre 56 : Les fonctions linéaires

90.2 Formes et fonctions

On peut toujours associer une forme linéaire φ à une fonction intégrable quelconque $\hat{\varphi}$ en définissant :

$$\langle \varphi, u \rangle = \int_A u(x) \cdot \hat{\varphi}(x) \, dx$$

Inversément, on ne pourra pas toujours trouver une fonction $\hat{\varphi}$ correspondant à une forme linéaire φ donnée. On définira malgré tout l'intégrale généralisée en notant :

$$\int_A u(x) \cdot \varphi(x) \, dx = \langle \varphi, u \rangle$$

où il ne faut pas perdre de vue que φ n'est pas nécessairement une fonction.

90.3 Formes et mesures

Soit $u : A \mapsto \mathbb{R}$. A toute mesure μ, on peut associer une forme linéaire $\hat{\mu}$ par :

$$\langle \hat{\mu}, u \rangle = \int_A u(x) \, d\mu(x)$$

Inversément, à toute forme linéaire $\hat{\mu}$, on peut associer une fonction $\mu : \mathfrak{P}(\mathbb{R}) \mapsto \mathbb{R}$ par :

$$\mu(A) = \langle \hat{\mu}, \boldsymbol{\delta}_A \rangle$$

Toutefois, rien ne garantit que la fonction μ ainsi définie est une mesure. En particulier, rien ne garantit qu'elle soit positive.

90.4 Fonction et forme bilinéaire

A toute fonction $\hat{K} : A \times B \mapsto F$, on peut associer une forme bilinéaire K par :

$$\langle u, K, v \rangle = \int_{A \times B} u(x) \cdot K(x, y) \cdot v(y) \, d\mu(x) \, d\nu(y)$$

pour toutes fonctions $u, v : A \mapsto B$. Inversément, à toute forme bilinéaire K, on peut associer une intégrale généralisée en notant :

$$\int_{A\times B} u(x) \cdot K(x, y) \cdot v(y) \ d\mu(x) \ d\nu(y) = \langle u, K, v\rangle$$

90.5 Définition

Nous nous intéressons ici au cas où l'espace vectoriel E est un ensemble de fonctions intégrables : $E = \mathcal{F} \subseteq \mathrm{Leb}^2(\mathbb{R}, \mathbb{R})$. Les limites à l'infini doivent alors forcément s'annuler

$$\lim_{x\to+\infty} u(x) = \lim_{x\to-\infty} u(x) = 0$$

pour tout $u \in \mathcal{F}$.

90.6 Delta de Dirac

La distribution $\delta \in F^\star$ de Dirac est définie par :

$$\langle\delta, u\rangle = u(0)$$

pour tout $u \in F$. Elle correspond bien sûr à l'intégrale :

$$\int_{\mathbb{R}} \delta(x) \cdot u(x) \ dx = u(0)$$

On remarque que :

$$\int_{A^2} \delta(\xi - x) \cdot K(\xi, \eta) \cdot \delta(\eta - y) \ d\mu(\xi) \ d\nu(\eta) = K(x, y)$$

90.7 Dérivée

En intégrant par parties, on a :

$$\int_{\mathbb{R}} \frac{du}{dx}(x) \cdot v(x) \ dx = \lim_{a\to+\infty} [u(a) \cdot v(a) - u(-a) \cdot v(-a)] - \int_{\mathbb{R}} u(x) \cdot \frac{dv}{dx}(x) \ dx$$

mais comme les limites à l'infini s'annulent, cette expression se réduit à :

$$\int_{\mathbb{R}} \frac{du}{dx}(x) \cdot v(x) \ dx = -\int_{\mathbb{R}} u(x) \cdot \frac{dv}{dx}(x) \ dx$$

Par extension, on définit la dérivée $\frac{du}{dx}$ d'une distribution u par :

$$\left\langle \frac{du}{dx}, v \right\rangle = -\left\langle u, \frac{dv}{dx} \right\rangle$$

pour tout $v \in F$.

90.7.1 Échelon

Comme application, considérons la fonction échelon e_+ :

$$e_+(x) = \boldsymbol{\delta}_{[0,+\infty)} = \begin{cases} 1 & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$$

Pour tout $v \in F$, on a :

$$\begin{aligned} \left\langle \frac{de_+}{dx}, v \right\rangle &= -\left\langle e_+, \frac{dv}{dx} \right\rangle \\ &= -\int_0^{+\infty} \frac{dv}{dx}(x) dx \end{aligned}$$

Appliquons à présent le théorème fondamental. Il vient :

$$\left\langle \frac{de_+}{dx}, v \right\rangle = -\left[\lim_{x \to +\infty} v(x) - v(0) \right] = v(0)$$

On en déduit que :

$$\frac{de_+}{dx} = \delta$$

au sens des distributions.

90.8 Dilatation

Soit d_a l'opérateur de dilatation :

$$d_a(u)(x) = u(a \cdot x)$$

où $a > 0$ est un réel strictement positif.
Le changement de variable $\xi = a \cdot x$ nous donne $d\xi = a\ dx$ et donc :

$$\int_{\mathbb{R}} \hat{u}(a\ x)\ v(x)\ dx = \frac{1}{a} \int_{\mathbb{R}} \hat{u}(\xi)\ v\left(\xi/a\right)\ d\xi$$

On définit donc l'extension de cet opérateur aux distributions par :

$$\langle d_a(u), v \rangle = \frac{1}{a} \left\langle u, d_{1/a}(v) \right\rangle$$

90.9 Réflexion

L'opérateur de réflexion r se définit par :

$$r(u)(x) = u(-x)$$

Le changement de variable $\xi = -x$ nous donne $d\xi = -dx$ et donc :

$$\begin{aligned} \int_{\mathbb{R}} \hat{u}(-x)\ v(x)\ dx &= \lim_{a \to +\infty} \int_{-a}^{a} \hat{u}(-x)\ v(x)\ dx \\ &= \lim_{a \to +\infty} -\int_{a}^{-a} \hat{u}(\xi)\ v(-\xi)\ d\xi \\ &= \lim_{a \to +\infty} \int_{a}^{-a} \hat{u}(\xi)\ v(-\xi)\ d\xi \end{aligned}$$

On définit donc l'extension de cet opérateur aux distributions par :

$$\langle r(u), v \rangle = \langle u, r(v) \rangle$$

90.10 Translation

L'opérateur de translation t_a est défini par :

$$t_a(u)(x) = u(x-a)$$

Le changement de variable $\xi = x - a$ nous donne $d\xi = dx$ et donc :

$$\int_{\mathbb{R}} \hat{u}(x-a) v(x) dx = \int_{\mathbb{R}} \hat{u}(\xi) v(\xi + a) d\xi$$

On définit donc les extensions de cet opérateur aux distributions par :

$$\langle t_a(u), v \rangle = \langle u, t_{-a}(v) \rangle$$

90.11 Convolution

Les intégrales unidimensionnelles permettent de définir l'opérateur de convolution $\circledcirc$. Soit deux fonctions $u, v : \mathbb{R} \mapsto \mathbb{R}$, leur convolution est une fonction $u \circledcirc v : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$(u \circledcirc v)(t) = \int_{-\infty}^{+\infty} u(t-s)\ v(s)\ ds$$

pour tout $t \in \mathbb{R}$.

90.11.1 Dirac

En utilisant les résultats ci-dessus, on arrive facilement à :

$$\int_{\mathbb{R}} u(x)\ \delta(x-a)\ dx = u(a)$$

Comme :

$$\int_{\mathbb{R}} u(x)\ \delta(-x)\ dx = \int_{\mathbb{R}} u(-x)\ \delta(x)\ dx = u(0)$$

on en déduit que $\delta(-x) = \delta(x)$ et :

$$\int_{\mathbb{R}} \delta(x-y)\ u(y)\ dy = u(x)$$

c'est-à-dire :

$$\delta \circledcirc u = u$$

La distribution de Dirac est neutre pour le produit de convolution. On peut montrer que ce neutre est unique.

90.12 Corrélation

Les intégrales unidimensionnelles permettent de définir l'opérateur de corrélation $\natural$. Soit deux fonctions $u, v : \mathbb{R} \mapsto \mathbb{R}$, leur corrélation est une fonction $u \natural v : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$(u \natural v)(t) = \int_{-\infty}^{+\infty} u(s+t)\ v(s)\ ds$$

pour tout $t \in \mathbb{R}$.

Chapitre 91

Formes différentielles

Dépendances

— Chapitre ?? : Les différentielles
— Chapitre 72 : Les intégrales

91.1 Intégrale d'un tenseur

Soit $(\mathfrak{c}_1, \mathfrak{c}_2, ..., \mathfrak{c}_n)$ la bace canonique de $\mathbb{R}^n$ et la fonction tensorielle $T : A \mapsto \mathbb{T}_m(\mathbb{R}^n)$ qui, à chaque $x \in A$ associe un tenseur $T(x)$ de la forme :

$$T(x) = \sum_{i,j,...,p} t_{ij...p}(x) \cdot \mathfrak{c}_i \otimes \mathfrak{c}_j \otimes ... \otimes \mathfrak{c}_p$$

L'intégrale de cette fonction est définie par :

$$\int_A T(x)\ d\mu(x) = \sum_{i,j,...,p} I_{ij...p} \cdot \mathfrak{c}_i \otimes \mathfrak{c}_j \otimes ... \otimes \mathfrak{c}_p$$

où chaque coordonnée $I_{ij...p}$ est l'intégrale de la coordonnée correspondante de T :

$$I_{ij...p} = \int_A t_{ij...p}(x)\ d\mu(x)$$

91.2 Produit extérieur

Soit $d\mu = dx = dx_1\ ...\ dx_n$ la mesure de Lebesgue sur $\mathbb{R}^n$. On sait que dx représente la mesure de l'élément de volume $[x_1, x_1 + dx_1] \times ... \times [x_n, x_n + dx_N]$. Etant donné que nous avons construit le produit extérieur pour représenter (au signe près) des mesures de surfaces et de volumes, il est tout à fait naturel de le faire intervenir dans une mesure de Lebesgue. Soit la base canonique $(e_1, e_2, ..., e_n)$ de $\mathbb{R}^n$ et les vecteurs :

$$\delta_i = dx_i \cdot e_i$$

où les dx_i sont bien évidemment des scalaires. Si nous évaluons le produit extérieur des ces vecteurs, nous obtenons :

$$\delta_1 \wedge \delta_2 \wedge ... \wedge \delta_n = \sum_{i,j,...,k} \boldsymbol{\epsilon}_{ij...k} \cdot dx_1 \cdot \delta_{1i} \cdot dx_2 \cdot \delta_{2j} \cdot ... \cdot dx_n \cdot \delta_{nk}$$

Le seul terme ne s'annulant pas étant $\boldsymbol{\epsilon}_{1,2,...,n} = 1$, on a finalement :

$$\delta_1 \wedge \delta_2 \wedge ... \wedge \delta_n = dx_1 \cdot dx_2 \cdot ... \cdot dx_n = dx$$

Cette constatation nous amène à définir une mesure plus générale. Considérons à présent des vecteurs infinitésimaux $\upsilon_1, \upsilon_2, ..., \upsilon_n \in \mathbb{R}^n$, c'est à dire des vecteurs dont la norme tendra vers zéro dans l'intégrale. Afin de garantir la positivité de la mesure, nous définissons :

$$du = |\upsilon_1 \wedge \upsilon_2 \wedge ... \wedge \upsilon_n|$$

91.3 Tenseur différentiel

Il est même possible de définir des tenseurs différentiels dU en choisissant $m \le n$ et en posant :

$$dU = \upsilon_1 \wedge \upsilon_2 \wedge ... \wedge \upsilon_m$$

Il est clair que $dU \in \mathbb{T}_{n-m}(\mathbb{R}^n)$. On nomme ce type de tenseur une forme différentielle.

91.4 Paramétrisation

Soit $m, n \in \mathbb{N}$ avec $m \le n$ et la fonction $\phi : U \subseteq \mathbb{R}^m \mapsto \mathbb{R}^n$, dérivable et inversible. Le but est de paramétrer x sur $\phi(U)$ par la relation $x = \phi(u)$ pour tout $u \in U$. Nous utilisons la base canonique $(e_1, e_2, ..., e_m)$ de $\mathbb{R}^m$ et nous posons :

$$\delta_i = \phi(u + du_i\ e_i) - \phi(u) = \partial_i \phi(u)\ du_i$$

Sur $\phi(U)$, on utilise le tenseur différentiel :

$$dX = \delta_1 \wedge \delta_2 \wedge ... \wedge \delta_m$$

La linéarité du produit extérieur nous permet d'ecrire :

$$\begin{aligned} dX &= \partial_1\phi(u) \wedge \partial_2\phi(u) \wedge ... \wedge \partial_m\phi(u)\ du_1\ du_2\ ...\ du_m \\ &= \partial_1\phi(u) \wedge \partial_2\phi(u) \wedge ... \wedge \partial_m\phi(u)\ du \end{aligned}$$

On définit le tenseur $W \in \mathbb{T}_{n-m}(\mathbb{R}^n)$ associé à dX par :

$$W(u) = \partial_1\phi(u) \wedge \partial_2\phi(u) \wedge ... \wedge \partial_m\phi(u)$$

Deux cas peuvent alors se présenter.

Fonction tensorielle

On peut évaluer l'intégrale d'une fonction tensorielle $f : \mathbb{R}^n \mapsto \mathbb{T}_p(\mathbb{R}^n)$ en utilisant la contraction maximale avec dX. Comme on a l'équivalence $x \in \phi(U) \leftrightarrow u \in U$, on a alors :

$$\int_{\phi(U)} f(x) : dX = \int_U (f \circ \phi)(u) : W(u)\ du$$

Dans le cas particulier où $p = n - m$, on obtiendra un scalaire.

Fonction scalaire

On peut évaluer l'intégrale d'une fonction scalaire $f : \mathbb{R}^n \mapsto \mathbb{R}$ en utilisant la norme de dX. On a alors $dx = \|dX\|$ et :

$$\int_{\phi(U)} f(x) dx = \int_U (f \circ \phi)(u) \cdot \|W(u)\|\ du$$

Pavé

Un cas particulier important est celui où $U = [\alpha_1, \beta_1] \times .. \times [\alpha_m, \beta_m]$ pour certains $\alpha_i, \beta_i \in \mathbb{R}$. On a alors :

$$\int_{\phi(U)} \sim \int_{\alpha_1}^{\beta_1} du_1 \int_{\alpha_2}^{\beta_2} du_2 \ ... \int_{\alpha_m}^{\beta_m} du_m$$

91.5 Changement de variable

Nous considérons à présent le cas où $m = n$. Nous utilisons la base canonique $(e_1, e_2, ..., e_n)$ de $\mathbb{R}^n$ et nous posons de nouveau :

$$\delta_i = \phi(u + du_i \ e_i) - \phi(u) = \partial_i \phi(u) \ du_i$$

On utilise la mesure :

$$dx = |\delta_1 \wedge \delta_2 \wedge ... \wedge \delta_n|$$

On a alors :

$$\begin{aligned} dx &= |\partial_1 \phi(u) \wedge \partial_2 \phi(u) \wedge ... \wedge \partial_n \phi(u) \ du| \\ &= \left| \sum_{i,j,...,k} \boldsymbol{\epsilon}_{ij...k} \cdot \partial_1 \phi_i(u) \cdot \partial_2 \phi_j(u) \cdot \ \ldots \ \cdot \partial_n \phi_k(u) \right| \ du \\ &= |\det \partial \phi(u)| \ du \end{aligned}$$

On voit donc apparaître le déterminant de la Jacobienne de ϕ. Comme on a l'équivalence $x \in A \leftrightarrow u \in \phi^{-1}(A)$, le changement de variable peut s'écrire :

$$\int_A f(x) \ dx = \int_{\phi^{-1}(A)} (f \circ \phi)(u) \cdot |\det \partial \phi(u)| \ du$$

Pavé

Un cas particulier important est celui où $\phi^{-1}(A) = [\alpha_1, \beta_1] \times .. \times [\alpha_n, \beta_n]$ pour certains $\alpha_i, \beta_i \in \mathbb{R}$. On a alors :

$$\int_A \sim \int_{\alpha_1}^{\beta_1} du_1 \int_{\alpha_2}^{\beta_2} du_2 \ ... \int_{\alpha_n}^{\beta_n} du_n$$

91.6 Intégrales de ligne vectorielles

Soit une fonction continue $\gamma : [a, b] \to \mathbb{R}^n$ définissant la courbe $\Lambda = \gamma([a, b])$. L'intégrale de ligne d'une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ sur cette courbe est l'intégrale de la contraction d'ordre 1 de f avec $d\gamma$, qui revient ici au produit scalaire du vecteur $f(x) \in \Lambda$ par le vecteur $\partial \gamma(t)$. On a donc :

$$\int_\Lambda f \cdot d\Lambda = \int_a^b \langle (f \circ \gamma)(t) \mid \partial \gamma(t) \rangle \, dt$$

91.7 Intégrales de ligne scalaires

Dans le cas d'une fonction $g : \mathbb{R}^n \mapsto \mathbb{R}$, on utilise comme mesure la longueur $\|\partial\gamma(t)\|$ de chaque petit segment $d\Lambda$. On a alors :

$$\int_\Lambda g \ d\Lambda = \int_a^b (g \circ \gamma)(t) \cdot \|\partial\gamma(t)\| \ dt$$

91.8 Contour fermé

Si $\gamma(a) = \gamma(b)$, on dit que le contour fermé, et on note en général :

$$\oint_\Lambda = \int_\Lambda$$

91.9 Intégrales de surface vectorielles

Soit $f : \mathbb{R}^n \mapsto \mathbb{R}^n$, $\sigma : A \subseteq \mathbb{R}^{n-1} \mapsto \mathbb{R}^n$ et la surface $\Theta = \sigma(A)$. On définit les vecteurs :

$$\delta_i = \frac{\partial\sigma}{\partial u_i} du_i$$

pour $i = 1, ..., n-1$. L'intégrale de surface est simplement la contraction d'ordre 1 :

$$\int_\Theta f \cdot d\Theta = \int_A \langle (f \circ \sigma)(u) \mid \delta_1 \wedge ... \wedge \delta_{n-1} \rangle$$

qui nous donne un scalaire. Dans le cas particulier où $n = 3$ et où $A = [U_1, U_2] \times [V_1, V_2]$, on a :

$$\int_\Theta f \cdot d\Theta = \int_{U_1}^{U_2} du \int_{V_1}^{V_2} (f \circ \sigma)(u, v) \cdot \left(\frac{\partial\sigma}{\partial u}(u, v) \wedge \frac{\partial\sigma}{\partial v}(u, v) \right) \ dv$$

91.10 Intégrales de surface scalaires

Soit $f : \mathbb{R}^n \mapsto \mathbb{R}^n$, $\sigma : A \subseteq \mathbb{R}^{n-1} \mapsto \mathbb{R}^n$ et la surface $\Theta = \sigma(A)$. On définit les vecteurs :

$$\delta_i = \frac{\partial\sigma}{\partial u_i} du_i$$

pour $i = 1, ..., n-1$. Utilisant comme mesure la norme du produit extérieur des δ_i, on obtient :

$$\int_\Theta f \ d\Theta = \int_A (f \circ \sigma)(u) \cdot \|\delta_1 \wedge ... \wedge \delta_{n-1}\|$$

Dans le cas particulier où $n = 3$ et où $A = [U_1, U_2] \times [V_1, V_2]$, on a :

$$\int_\Theta f \ d\Theta = \int_{U_1}^{U_2} du \int_{V_1}^{V_2} (f \circ \sigma)(u, v) \cdot \left\| \frac{\partial\sigma}{\partial u}(u, v) \wedge \frac{\partial\sigma}{\partial v}(u, v) \right\| \ dv$$

91.11 Intégrale de flux

Soit $A \subseteq \mathbb{R}^n$ et la fonction $a : \mathbb{R}^n \to \mathbb{R}$ telle que :

$$A = \{x \in \mathbb{R}^n : a(x) \le 0\}$$

On s'arrange de plus pour avoir a constante sur la frontière :

$$\partial A = \{x \in \mathbb{R}^n : a(x) = 0\}$$

On introduit le vecteur normal :

$$n = \frac{1}{\left\| \frac{\partial a}{\partial x} \right\|} \cdot \frac{\partial a}{\partial x}$$

L'intégrale du flux sortant de la fonction $f : \mathbb{R}^n \to \mathbb{R}^n$ est alors donnée par :

$$\int_{\partial A} \langle f \mid n \rangle \ d\mu$$

91.12 Différentielle

Soit une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$, les vecteurs infinitésimaux $\delta_1, ..., \delta_{n-1} \in \mathbb{R}^n$ et la forme différentielle :

$$\omega = f \cdot \delta_1 \wedge \delta_2 \wedge ... \wedge \delta_{n-1}$$

Si f est différentiable, on définit la différentielle de ω par :

$$d\omega = \sum_i \frac{\partial f}{\partial x_i} \cdot \kappa_i \wedge \delta_1 \wedge \delta_2 \wedge ... \wedge \delta_{n-1}$$

où :

$$\kappa_i = dx_i \cdot e_i$$

On note aussi symboliquement :

$$d\omega = df \wedge dx_1 \wedge ... \wedge dx_{n-1}$$

On peut montrer sous certaines conditions que l'intégrale sur la frontière de A est alors donnée par :

$$\int_{\partial A} \omega = \int_A d\omega$$

91.13 Théorème de Stokes

Soit $f, g : \mathbb{R}^2 \mapsto \mathbb{R}$ et les vecteurs infinitésimaux :

$$\delta x = e_1 \ dx$$
$$\delta y = e_2 \ dy$$

Considérons la forme différentielle :

$$\omega = f\delta x + g\delta y$$

Si les fonctions sont différentiables, on a alors :

$$d\omega = \frac{\partial f}{\partial x}\delta x \wedge \delta x + \frac{\partial f}{\partial y}\delta y \wedge \delta x + \frac{\partial g}{\partial x}\delta x \wedge \delta y + \frac{\partial g}{\partial y}\delta y \wedge \delta y$$

Mais comme :

$$\delta x \wedge \delta x = \delta y \wedge \delta y = 0$$
$$\delta y \wedge \delta x = -\delta x \wedge \delta y$$

il vient :

$$d\omega = \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) \delta x \wedge \delta y$$

En intégrant, on obtient alors :

$$\int_{\partial A} (f \ \delta x + g \ \delta y) = \int_A \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx \wedge dy$$

Mais comme nous somme dans la base canonique, on a $\delta x \wedge \delta y = dx \ dy$ et :

$$\int_{\partial A} (f \ \delta x + g \ \delta y) = \int_A \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) \ dx \ dy$$

Chapitre 92

Géométrie différentielle

92.1 Dépendances

— Chapitre 52 : Les vecteurs
— Chapitre 60 : Les produits scalaires
— Chapitre 62 : Les tenseurs

92.2 Coordonnées curvilignes

Soit l'espace vectoriel $E = \mathbb{R}^n$ sur $\mathbb{R}$. Les coordonnées curvilignes sont basées sur la notion de position r, exprimée comme une fonction de certains paramètres $x \in \mathbb{R}^n$ que nous appelons « coordonnées » de r :

$$r = \rho(x)$$

où $\rho : \mathbb{R}^n \to \mathbb{R}^n$. Nous envisageons également le cas du changement de variable. La position dépend alors d'un autre jeu de coordonnées $y \in \mathbb{R}^n$:

$$r = \sigma(y)$$

où $\sigma : \mathbb{R}^n \to \mathbb{R}^n$. Nous définissons les vecteurs fondamentaux e_i et e^i au moyen de ces fonctions :

$$e_i(x) = \frac{\partial \rho}{\partial x^i}(x)$$
$$e^i(y) = \frac{\partial \sigma}{\partial y_i}(y)$$

de telle sorte que :

$$dr = \sum_i e_i \, dx^i = \sum_i e^i \, dy_i$$

Nous supposons que $(e_1, ..., e_n)$ et $(e^1, ..., e^n)$ sont des bases de E.

Courbe

Dans le cas où x et y ne dépendent que d'un paramètre $t \in \mathbb{R}$, on a :

$$\frac{dr}{dt} = \sum_i e_i \, \frac{dx^i}{dt} = \sum_i e^i \, \frac{dy_i}{dt}$$

On dit alors que la position r décrit une courbe.

92.3 Changement de variable

Si les fonctions ρ et σ sont inversibles, on a :

$$x = \rho^{-1}(r) = (\rho^{-1} \circ \sigma)(y) = \phi(y)$$
$$y = \sigma^{-1}(r) = (\sigma^{-1} \circ \rho)(x) = \psi(x)$$

où nous avons implicitement défini $\phi = \rho^{-1} \circ \sigma$ et $\psi = \sigma^{-1} \circ \rho$. Nous notons $\frac{\partial x^i}{\partial y_j}$ et $\frac{\partial y_i}{\partial x^j}$ les coordonnées des dérivées de ϕ et ψ suivant les bases formées par les e_i et les e^i :

$$\frac{\partial \phi}{\partial y_j} = \sum_i \frac{\partial x^i}{\partial y_j} \ e_i$$
$$\frac{\partial \psi}{\partial x^j} = \sum_i \frac{\partial y_i}{\partial x^j} \ e^i$$

La composition des dérivées nous donne les relations :

$$e_i = \sum_j \frac{\partial r}{\partial y_j} \ \frac{\partial y_j}{\partial x^i} = \sum_j \frac{\partial y_j}{\partial x^i} \ e^j$$
$$e^i = \sum_j \frac{\partial r}{\partial x^j} \ \frac{\partial x^j}{\partial y_i} = \sum_j \frac{\partial x^j}{\partial y_i} \ e_j$$

qui nous permettent de relier les e_i aux e^j et inversément.

92.4 Produit scalaire

Les produits intérieurs entre vecteurs de base se notent habituellement :

$$g_{ij} = \langle e_i \mid e_j \rangle$$
$$g_i^j = \left\langle e_i \mid e^j \right\rangle$$
$$g^{ij} = \left\langle e^i \mid e^j \right\rangle$$

Il est clair d'après les propriétés de symétrie de ce produit que :

$$g_{ij} = g_{ji}$$
$$g^{ij} = g^{ji}$$
$$g_i^j = g_j^i$$

Le produit scalaire de deux vecteurs $a, b \in E$ définis par :

$$a = \sum_i a^i \ e_i = \sum_i a_i \ e^i$$
$$b = \sum_i b^i \ e_i = \sum_i b_i \ e^i$$

peut s'écrire indifféremment comme :

$$\langle a \mid b \rangle = \sum_{i,j} g_{ij} \ a^i \ b^j$$
$$\langle a \mid b \rangle = \sum_{i,j} g^{ij} \ a_i \ b_j$$
$$\langle a \mid b \rangle = \sum_{i,j} g^i_j \ a_i \ b^j$$
$$\langle a \mid b \rangle = \sum_{i,j} g^j_i \ a^i \ b_j$$

Et en particulier, la longeur ds d'un changement de position dr vérifie :

$$(ds)^2 = \langle dr \mid dr \rangle = \sum_{i,j} g_{ij} \ dx^i \ dx^j = \sum_{i,j} g^{ij} \ dy_i \ dy_j$$

De plus, les relations entre les vecteurs e_i et les vecteurs e^i permettent de déduire, en utilisant la linéarité du produit scalaire :

$$g_{ij} = \sum_k \frac{\partial y_k}{\partial x^i} \ g^k_j = \sum_{k,l} \frac{\partial y_k}{\partial x^i} \ \frac{\partial y_l}{\partial x^j} \ g^{kl}$$
$$g^{ij} = \sum_k \frac{\partial x_k}{\partial y^i} \ g^j_k = \sum_{k,l} \frac{\partial x^k}{\partial y_i} \ \frac{\partial x^l}{\partial y_j} \ g_{kl}$$

92.5 Dérivées primales d'un vecteur

Nous allons à présent voir comment évolue un vecteur $a \in E$, que l'on note sous la forme :

$$a = \sum_i a^i \ e_i$$

où les coordonnées $a^i \in \mathbb{R}$ tout comme les vecteurs de base e_i dépendent des coordonnées x^i. La règle de dérivation d'un produit nous donne :

$$da = \sum_i da^i \ e_i + \sum_k a^k \ de_k$$

La différentielle da^i s'obtient directement :

$$da^i = \sum_j \frac{\partial a^i}{\partial x^j} \ dx^j$$

On peut suivre la même règle avec de_i :

$$de_k = \sum_j \frac{\partial e_k}{\partial x^j} \ dx^j$$

Les symboles de Christoffel $\left\{ \begin{array}{c} i \\ kj \end{array} \right\}$ sont définis comme les coordonnées de $\frac{\partial e_k}{\partial x^j}$ suivant la base $(e_1, ..., e_n)$:

$$\frac{\partial e_k}{\partial x^j} = \sum_i \left\{ \begin{array}{c} i \\ kj \end{array} \right\} \ e_i$$

Notons que comme :

$$\frac{\partial e_k}{\partial x^j} = \frac{\partial^2 r}{\partial x^j \partial x^k} = \frac{\partial^2 r}{\partial x^k \partial x^j} = \frac{\partial e_j}{\partial x^k}$$

on a la symétrie :

$$\left\{ \begin{array}{c} i \\ kj \end{array} \right\} = \left\{ \begin{array}{c} i \\ jk \end{array} \right\}$$

On peut évaluer ces symboles si on connait par exemple les valeurs des :

$$\begin{aligned} \left\langle e^i \mid \frac{\partial e_k}{\partial x^j} \right\rangle &= \sum_m \left\{ \begin{array}{c} m \\ kj \end{array} \right\} \left\langle e^i \mid e_m \right\rangle \\ &= \sum_m \left\{ \begin{array}{c} m \\ kj \end{array} \right\} g^i_m \end{aligned}$$

On a alors, pour chaque choix de k, j un système linéaire à résoudre. Il suffit d'inverser la matrice $G = (g^i_m)_{i,m}$ pour obtenir les valeurs des symboles.

La dérivation d'un vecteur $a \in E$ s'écrit alors :

$$da = \sum_{i,j} e_i \, dx^j \left[\frac{\partial a^i}{\partial x^j} + \sum_k \left\{ \begin{array}{c} i \\ kj \end{array} \right\} a^k \right]$$

On définit les coordonnées :

$$\nabla_j a^i = \frac{\partial a^i}{\partial x^j} + \sum_k \left\{ \begin{array}{c} i \\ kj \end{array} \right\} a^k$$

Dans le cas où les coordonnées dépendent d'un paramètre $t \in \mathbb{R}$, on a :

$$\begin{aligned} \frac{da}{dt} &= \sum_{i,j} e_i \, \frac{dx^j}{dt} \left[\frac{\partial a^i}{\partial x^j} + \sum_k \left\{ \begin{array}{c} i \\ kj \end{array} \right\} a^k \right] \\ &= \sum_i e_i \left[\frac{da^i}{dt} + \sum_{j,k} \left\{ \begin{array}{c} i \\ kj \end{array} \right\} a^k \, \frac{dx^j}{dt} \right] \end{aligned}$$

92.5.1 Dérivée seconde et géodésique

Considérons le cas :

$$a = \frac{dr}{dt} = \sum_i e_i \, \frac{dx^i}{dt}$$

Les coordonnées de a sont clairement $a^i = \frac{dx^i}{dt}$ et la dérivée seconde :

$$\frac{d^2 r}{dt^2} = \frac{d}{dt} \frac{dr}{dt} = \frac{da}{dt}$$

s'écrit :

$$\frac{d^2 r}{dt^2} = \sum_i e_i \left[\frac{d^2 x^i}{dt^2} + \sum_{j,k} \left\{ \begin{array}{c} i \\ kj \end{array} \right\} \frac{dx^k}{dt} \, \frac{dx^j}{dt} \right]$$

Les courbes $x^i = x^i(t)$ vérifiant $\frac{d^2 r}{dt^2} = 0$ sont appelées des géodésiques.

92.6 Dérivées duales d'un vecteur

Nous allons recommencer le même processus, écrivant cette fois $a \in E$ sous la forme :

$$a = \sum_i a_i \ e^i$$

Les coordonnées $a_i \in \mathbb{R}$ tout comme les vecteurs de base e^i dépendent des coordonnées y_i. En suivant la même méthode que ci-dessus, on obtient :

$$da = \sum_{i,j} e^i \ dy_j \left[\frac{\partial a_i}{\partial y_j} + \sum_k \left\{ \begin{array}{c} kj \\ i \end{array} \right\} \ a_k \right]$$

où l'on a introduit de nouveaux symboles de Christoffel, définis par :

$$\frac{\partial e^k}{\partial y_j} = \sum_i \left\{ \begin{array}{c} kj \\ i \end{array} \right\} \ e^i$$

Ces nouveaux symboles présentent la symétrie :

$$\left\{ \begin{array}{c} kj \\ i \end{array} \right\} = \left\{ \begin{array}{c} jk \\ i \end{array} \right\}$$

92.7 Dérivées d'un tenseur

On étend simplement la notion de dérivée aux tenseurs en appliquant la formule :

$$d(a \otimes b) = da \otimes b + a \otimes db$$

où a et b sont deux tenseurs d'ordre quelconque. Par exemple, pour le tenseur :

$$T = \sum_{i,j} T^i_j \ e_i \otimes e^j$$

on a :

$$dT = \sum_{i,j} \left[dT^i_j \ e_i \otimes e^j + T^i_j \ de_i \otimes e^j + T^i_j \ e_i \otimes de^j \right]$$

qui devient, en introduisant les symboles de Christoffel :

$$dT = \sum_{i,j} e_i \otimes e^j \left[dT^i_j + \sum_{k,m} \left\{ \begin{array}{c} i \\ mk \end{array} \right\} \ T^m_j \ dx^k + \sum_{k,m} \left\{ \begin{array}{c} mk \\ j \end{array} \right\} \ T^i_m \ dy_k \right]$$

92.8 Produit scalaire et symboles de Christoffel

Lorsqu'on différentie les g_{ij}, on obtient :

$$\begin{aligned} dg_{ij} &= \langle de_i \mid e_j \rangle + \langle e_i \mid de_j \rangle \\ &= \sum_{k,l} \left\{ \begin{array}{c} k \\ il \end{array} \right\} \ g_{kj} \ dx^l + \sum_{k,l} \left\{ \begin{array}{c} k \\ jl \end{array} \right\} \ g_{ik} \ dx^l \end{aligned}$$

On en déduit que :

$$\frac{\partial g_{ij}}{\partial x^l} = \sum_k \left\{ \begin{array}{c} k \\ il \end{array} \right\} g_{kj} + \sum_k \left\{ \begin{array}{c} k \\ jl \end{array} \right\} g_{ik}$$

Définissons :

$$\gamma_{ijl} = \sum_k \left\{ \begin{array}{c} k \\ il \end{array} \right\} g_{kj}$$

Les propriétés de symétrie des symboles de Christoffel nous montrent que :

$$\gamma_{ijl} = \gamma_{ilj}$$

Et comme (changement de l'indice l en k) :

$$\frac{\partial g_{ij}}{\partial x^k} = \gamma_{ijk} + \gamma_{jik}$$

On en déduit :

$$\begin{aligned} \frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ki}}{\partial x^j} &= 2\ \gamma_{jik} \\ &= 2 \sum_l \left\{ \begin{array}{c} l \\ jk \end{array} \right\} g_{li} \end{aligned}$$

AFAIRE : LA FIN DU CHAPITRE EST A DÉBROUILLONNER

92.9 Bases biorthonormées

92.9.1 produit scalaire

Nous considérons tout au long de cette section le cas particulier où les bases sont biorthonormées, c'est-à-dire :

$$g_i^j = \boldsymbol{\delta}_i^j$$

On déduit des relations liant les g^{ij}, g_{ij} aux g_i^j que :

$$g_{ik} g^{kj} = \sum_{k,l,m} \frac{\partial y_l}{\partial x^i} \frac{\partial x^m}{\partial y_k} g_k^l g_m^j$$

$$g_{ik} g^{kj} = \sum_{k,l,m} \frac{\partial y_l}{\partial x^i} \frac{\partial x^m}{\partial y_k} \boldsymbol{\delta}_k^l \boldsymbol{\delta}_m^j$$

$$g_{ik} g^{kj} = \sum_k \frac{\partial y_k}{\partial x^i} \frac{\partial x^j}{\partial y_k}$$

$$g_{ik} g^{kj} = \frac{\partial x^i}{\partial x^j} = \boldsymbol{\delta}_i^j$$

On aurait de même :

$$g^{ik} g_{kj} = \boldsymbol{\delta}_j^i$$

92.9.2 Coordonnées

Les coordonnées d'un tenseur de la forme :

$$T = \sum_{i,j,k,l} T^{k...l}_{i...j} e^i \otimes ... \otimes e^j \otimes e_k \otimes ... \otimes e_l$$

où il y a m indices $i...j$ et n indices $k...l$ s'obtiennent facilement en utilisant la contraction double :

$$T^{k...l}_{i...j} = \langle e_j \otimes ... \otimes e_i | \, T \, \left| e^l \otimes ... \otimes e^k \right\rangle_{m,n}$$

92.10 Dérivées des changements de variable

$$\frac{\partial x^i}{\partial y_j} = \left\langle e_i \mid \frac{\partial \phi}{\partial y_j} \right\rangle$$
$$\frac{\partial y_i}{\partial x^j} = \left\langle e^i \mid \frac{\partial \psi}{\partial x^j} \right\rangle$$

92.10.1 Christoffel

Tenant compte de cette identité, l'équation reliant les symboles de Christoffel aux produits scalaires devient :

$$\left\{ \begin{array}{c} m \\ jk \end{array} \right\} = \frac{1}{2} \sum_i g^{im} \left[\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ki}}{\partial x^j} \right]$$

92.10.2 Dérivée d'un vecteur

La relation :

$$d \left\langle e^i \mid e_j \right\rangle = d\boldsymbol{\delta}^j_i = 0$$

nous conduit à :

$$\left\langle de^i \mid e_j \right\rangle + \left\langle e^i \mid de_j \right\rangle = 0$$
$$\sum_{k,l} \left\{ \begin{array}{c} ik \\ l \end{array} \right\} \left\langle e^l \mid e_j \right\rangle dy_k + \sum_{k,l} \left\{ \begin{array}{c} l \\ jk \end{array} \right\} \left\langle e^i \mid e_l \right\rangle dx^k = 0$$
$$\sum_k \left\{ \begin{array}{c} ik \\ l \end{array} \right\} dy_k = - \sum_k \left\{ \begin{array}{c} l \\ jk \end{array} \right\} dx^m$$

Par ailleurs :

$$da_i = \frac{\partial a_i}{\partial y_j} dy_j = \frac{\partial a_i}{\partial x^j} dx^j$$

On peut donc réexprimer la dérivée duale comme :

$$da = \sum_{i,j} e^i dx^j \left[\frac{\partial a_i}{\partial x^j} - \sum_k \left\{ \begin{array}{c} k \\ ij \end{array} \right\} a_k \right]$$

92.10.3 Gradient

On peut également définir le gradient d'un vecteur par :

$$\nabla a = \sum_{i,j} \nabla_j a^i e_i \otimes e^j$$

de telle sorte que l'on ait :

$$da = \langle \nabla a \mid dr \rangle = \nabla a \cdot dr$$

92.10.4 Dérivée d'un tenseur

$$dT = \sum_{i,j,k} e_i \otimes e^j dx^k \left[\frac{\partial T^i_j}{\partial x^k} + \sum_m \left\{ \begin{array}{c} i \\ mk \end{array} \right\} T^m_j - \sum_m \left\{ \begin{array}{c} m \\ jk \end{array} \right\} T^i_m \right]$$

On définit alors les coordonnées :

$$\nabla_k T^i_j = \frac{\partial T^i_j}{\partial x^k} + \sum_m \left\{ \begin{array}{c} i \\ mk \end{array} \right\} T^m_j - \sum_m \left\{ \begin{array}{c} m \\ jk \end{array} \right\} T^i_m$$

92.10.5 Tenseur de courbure

Appliquons la formule de dérivation des coordonnées d'un tenseur dans le cas particulier où :

$$T^i_j = \nabla_j a^i$$

On a :

$$\frac{\partial T^i_j}{\partial x^k} = \frac{\partial^2 a^i}{\partial x^j \partial x^k} + \sum_m \left\{ \begin{array}{c} i \\ jm \end{array} \right\} \frac{\partial a^m}{\partial x^k} + \sum_m \frac{\partial}{\partial x^k} \left\{ \begin{array}{c} i \\ jm \end{array} \right\} a^m$$

$$\textstyle\sum_l \left\{ \begin{array}{c} i \\ kl \end{array} \right\} T^l_j = \sum_l \left\{ \begin{array}{c} i \\ kl \end{array} \right\} \frac{\partial a^i}{\partial x^j} + \sum_{l,m} \left\{ \begin{array}{c} i \\ kl \end{array} \right\} \left\{ \begin{array}{c} l \\ jm \end{array} \right\} a^m$$

$$\textstyle -\sum_l \left\{ \begin{array}{c} l \\ jk \end{array} \right\} T^i_l = -\sum_l \left\{ \begin{array}{c} l \\ jk \end{array} \right\} \frac{\partial a^i}{\partial x^l} - \sum_{l,m} \left\{ \begin{array}{c} l \\ jk \end{array} \right\} \left\{ \begin{array}{c} i \\ lm \end{array} \right\} a^m$$

La somme de tous ces termes vaut $\nabla_k T^i_j = \nabla_k \nabla_j a^i$. En interchangeant les indices j et k, on obtient $\nabla_j \nabla_k a^i$. On en déduit, en utilisant les propriétés de symétrie que :

$$\nabla_k \nabla_j a^i - \nabla_j \nabla_k a^i = \sum_m R^i_{m,kj} a^m$$

où les $R^{\cdots}_{\cdots}$ sont définis par :

$$R^i_{m,kj} = \frac{\partial}{\partial x^k} \left\{ \begin{array}{c} i \\ jm \end{array} \right\} - \frac{\partial}{\partial x^j} \left\{ \begin{array}{c} i \\ km \end{array} \right\} + \sum_l \left\{ \begin{array}{c} i \\ kl \end{array} \right\} \left\{ \begin{array}{c} l \\ jm \end{array} \right\} - \sum_l \left\{ \begin{array}{c} i \\ jl \end{array} \right\} \left\{ \begin{array}{c} l \\ km \end{array} \right\}$$

Ce sont les coordonnées du tenseur de courbure de Riemann-Christoffel.

Chapitre 93

L'espace vectoriel des polynômes

93.1 Introduction

AFAIRE : ARRANGER LE CHAPITRE

Il est clair d'après la définition des polynômes que les espaces $\mathcal{P}_n$ sont des espaces vectoriels pour l'ensemble des scalaires $S = \mathbb{R}$ et que :

$$\mathcal{P}_n = \text{span}\{\mu_0, \mu_1, ..., \mu_n\}$$

Nous allons montrer que $(\mu_0, \mu_1, ..., \mu_n)$ forme une base de $\mathcal{P}_n$. Pour cela, il nous reste à prouver l'indépendance linéaire des μ_i :

$$\sum_{i=0}^{n} a_i \mu_i = 0 \quad \Rightarrow \quad a_0 = a_1 = ... = a_n = 0$$

c'est-à-dire :

$$\sum_{i=0}^{n} a_i x^i = 0 \quad \forall x \in \mathbb{R} \quad \Rightarrow \quad a_0 = a_1 = ... = a_n = 0$$

Nous allons le montrer par récurrence.

Comme $\mu_0 = 1$ on a évidemment :

$$a_0 1 = 0 \quad \Rightarrow \quad a_0 = 0$$

et la thèse est vraie pour $n = 0$. Supposons à présent qu'elle soit vraie pour $n - 1$. Choisissons $p \in \mathcal{P}_n$:

$$p(x) = \sum_{i=0}^{n} a_i x^i$$

et supposons que $p(x) = 0$ pour tout $x \in \mathbb{R}$. Comme $p(0) = 0$, on a :

$$a_0 = 0$$

donc :

$$p(x) = \sum_{i=1}^{n} a_i x^i = x q(x) = 0$$

où l'on à définit $q \in \mathcal{P}_{n-1}$ par :

$$q(x) = \sum_{i=1}^{n} a_i x^{i-1}$$

Il est clair que, pour tout $x \neq 0$, $q(x) = 0$. Mais comme les polynômes sont des fonctions continues, on a :

$$q(0) = \lim_{\substack{x \to 0 \\ x \neq 0}} q(x) = 0$$

Donc q s'annule également en 0. On en conclut que $q(x)$ est nul pour tout $x \in \mathbb{R}$. Par l'hypothèse de récurrence, les coefficients de ce polynôme sont tous nuls :

$$a_1 = a_2 = ... = a_n = 0$$

Rassemblant les résultats, il vient :

$$a_0 = a_1 = ... = a_n = 0$$

et $(\mu_0, \mu_1, ..., \mu_n)$ forme bien une base de $\mathcal{P}_n$.

93.2 Polynômes orthogonaux

Nous allons à présent voir comment construire des suites de polynômes orthogonaux pour le produit scalaire :

$$\langle p \mid q \rangle = \int_a^b p(x) q(x) d\mu(x)$$

ou, lorsque c'est possible :

$$\langle p \mid q \rangle = \int_a^b p(x) q(x) w(x) dx$$

93.2.1 Récurrence

On pourrait bien entendu partir de la suite de la base canonique de monômes $(1, x, x^2, ..., x^n)$ et l'orthogonaliser en utilisant le procédé de Gram-Schmidt, mais on peut arriver à un algorithme plus rapide en utilisant les propriétés des polynômes. Soit $(\phi_n)_n$ une suite de polynômes orthonormés, où ϕ_i est de degré i. On a donc :

$$\langle \phi_m \mid \phi_n \rangle = \int_A \phi_m(x) \phi_n(x) d\mu(x) = \delta_{mn}$$

Supposons que $(\phi_0, ..., \phi_n)$ forme une base de $\mathcal{P}_n$. On peut vérifier que $(\phi_0, ..., \phi_n, x\phi_n)$ forme une base de $\mathcal{P}_{n+1}$. On peut donc représenter ϕ_{n+1} comme :

$$\phi_{n+1}(x) = a_n x \phi_n(x) + b_n \phi_n(x) + c_n \phi_{n-1}(x) + \sum_{i=0}^{n-2} d_i \phi_i(x)$$

Soit $i \in \{0, ..., n-2\}$. La condition d'orthogonalité de ϕ_{n+1} avec ϕ_i s'écrit :

$$\langle \phi_i \mid \phi_{n+1} \rangle = a_n \langle \phi_i \mid x\phi_n \rangle + b_n \langle \phi_i \mid \phi_n \rangle + c_n \langle \phi_i \mid \phi_{n-1} \rangle + \sum_{j=0}^{n-2} d_j \langle \phi_i \mid \phi_j \rangle = 0$$

L'orthogonalité implique que :

$$\langle \phi_i \mid \phi_n \rangle = \langle \phi_i \mid \phi_{n-1} \rangle = 0$$
$$\langle \phi_i \mid \phi_j \rangle = \delta_{ij}$$

On a aussi :

$$\langle \phi_i \mid x\phi_n \rangle = \int_A x\phi_i(x)\phi_n(x)d\mu(x) = \langle x\phi_i \mid \phi_n \rangle$$

Mais comme ϕ_i est de degré i, $x\phi_i$ est de degré $i+1$ et on peut l'exprimer comme :

$$x\phi_i = \sum_{j=0}^{i+1} \alpha_i \phi_j$$

Le produit scalaire devient alors :

$$\langle \phi_i \mid x\phi_n \rangle = \sum_{j=0}^{i+1} \alpha_i \langle \phi_j \mid \phi_n \rangle = 0$$

puisque $j \leq i+1 < n$. On en conclut que :

$$\langle \phi_i \mid \phi_{n+1} \rangle = \sum_{j=0}^{n-2} d_j \delta_{ij} = d_i = 0$$

Les conditions :

$$\langle \phi_{n+1} \mid \phi_n \rangle = 0$$
$$\langle \phi_{n+1} \mid \phi_{n-1} \rangle = 0$$

impliquent respectivement que :

$$b_n = -a_n \langle \phi_n \mid x\phi_n \rangle$$
$$c_n = -a_n \langle \phi_{n-1} \mid x\phi_n \rangle$$

La condition de normalisation :

$$\langle \phi_{n+1} \mid \phi_{n+1} \rangle = a_n \langle x\phi_n \mid \phi_{n+1} \rangle = 1$$

nous donne alors la valeur de a_n :

$$a_n^2 \langle x\phi_n \mid x\phi_n \rangle - a_n^2 \langle x\phi_n \mid \phi_n \rangle^2 - a_n^2 \langle x\phi_n \mid \phi_{n-1} \rangle^2 = 1$$
$$a_n = \left[\langle x\phi_n \mid x\phi_n \rangle - \langle x\phi_n \mid \phi_n \rangle^2 - \langle x\phi_n \mid \phi_{n-1} \rangle^2 \right]^{-1/2}$$

On voit donc que le choix du produit scalaire détermine :

$$\phi_0 = \frac{1}{\sqrt{\langle 1 \mid 1 \rangle}}$$
$$\phi_1 = a_1(x - \langle \phi_0 \mid x \rangle \phi_0)$$

ainsi que toute la suite de polynômes.

93.2.2 Approximation

Soit une suite de polynômes orthonormaux $(\phi_0, ...\phi_n)$ pour le produit scalaire :

$$\langle u \mid v \rangle = \int_A u(x)v(x)d\mu(x)$$

Nous cherchons l'approximation de u :

$$w(x) = \sum_{i=0}^{n} w_i \phi_i(x)$$

qui minimise l'erreur au sens intégral :

$$\langle u - w \mid u - w \rangle = \int_A [u(x) - w(x)]^2 d\mu(x)$$

sur $\mathcal{P}_n$. Imposant que la dérivée par rapport aux w_i soit nulle, on obtient :

$$2 \int_A \phi_i(x)[u(x) - w(x)] d\mu(x) = 0$$

Mais comme :

$$w_i = \int_A \phi_i(x) w(x) d\mu(x)$$

on obtient :

$$w_i = \int_A \phi_i(x) u(x) d\mu(x) = \langle \phi_i \mid u \rangle$$

Ce qui n'a rien d'étonnant au vu des résultats du chapitre **??**. On peut vérifier facilement que la hessienne de l'erreur par rapport aux w_i est bien positive. L'approximation ainsi définie :

$$w(x) = \sum_{i=0}^{n} \phi_i(x) \int_A \phi_i(y) u(y) d\mu(y)$$
$$w(x) = \sum_{i=0}^{n} \int_A \phi_i(x) \phi_i(y) u(y) d\mu(y)$$

minimise donc bien l'erreur sur l'ensemble des polynômes de degré n.

93.2.3 Intégration de Gauss

Soit une suite de polynômes orthonormaux $(\phi_0, ...\phi_n)$ pour le produit scalaire :

$$\langle u \mid v \rangle = \int_A u(x) v(x) d\mu(x)$$

Considérons la formule d'intégration :

$$I(f) = \sum_{i=0}^{n} w_i f(x_i)$$

supposée approximer l'intégrale :

$$\langle f \rangle = \langle f \mid 1 \rangle = \int_A f(x) d\mu(x)$$

Fixons les points $x_0 < x_1 < ... < x_n$ et imposons que la formule soit exacte pour $\phi_0, ..., \phi_n$. On a :

$$\langle \phi_k \rangle = \sum_{i=0}^{n} w_i \phi_k(x_i)$$

où $k = 0, 1, ..., n$. Définissant les matrices et vecteurs :

$$\varphi = (\langle \phi_k \rangle)_k$$
$$W = (w_i)_i$$
$$\Phi = (\phi_i(x_j))_{i,j}$$

ces conditions se ramènent à :

$$\Phi W = \varphi$$

Si la matrice $\Phi(n+1, n+1)$ est inversible, on a alors :

$$W = \Phi^{-1}\varphi$$

La formule est alors valable pour tout polynôme de $\mathcal{P}_n$. Notons que

$$\langle \phi_k \rangle = \frac{1}{\phi_0} \langle \phi_k \mid \phi_0 \rangle$$

s'annule pour tout $k \neq 0$. Si les racines de ϕ_{n+1} sont toutes distinctes, on peut choisir les x_i tels que :

$$\phi_{n+1}(x_i) = 0$$

On a alors :

$$\langle \phi_{n+1} \rangle = I(\phi_{n+1}) = 0$$

et la formule devient valable sur $\mathcal{P}_{n+1}$. Mieux, considérons un polynôme p de degré $n+m+1$ où $m \geq 0$ et sa division euclidienne par ϕ_{n+1}. On a :

$$p(x) = q(x)\phi_{n+1}(x) + r(x)$$

Comme q est de degré m, on a :

$$q = \sum_{i=0}^{m} q_i \phi_i$$

Si $m \leq n$, on a donc :

$$\langle q\phi_{n+1} \rangle = \sum_{i=0}^{n} q_i \langle \phi_i \mid \phi_{n+1} \rangle = 0$$

et :

$$\langle p \rangle = \langle r \rangle = I(r)$$

puisque r est de degré n au plus. Comme ϕ_{n+1} s'annule en les x_i, on a aussi ;

$$I(p) = I(r)$$

Rassemblant tout ces résultats, on obtient :

$$\int_A f(x) d\mu(x) = \sum_{i=0}^{n} w_i f(x_i)$$

pour tout polynôme $f \in \mathcal{P}_{2n+1}$. En pratique, on utilise ces formules d'intégration pour des fonctions qui ne sont pas forcément des polynômes.

93.3 Legendre

Les polynômes de Legendre sont orthogonaux pour le produit scalaire :

$$\int_{-1}^{1} P_n(x)P_m(x)dx = \frac{2}{2n+1}\delta_{mn}$$

Ils obéissent à la récurrence :

$$P_0(x) = 1$$
$$P_1(x) = x$$
$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$$

93.4 Interpolation

Un problème d'interpolation consiste à trouver les coefficients : $a_i \in \mathbb{R}$ tels que la fonction :

$$u = \sum_{i=1}^{n} a_i u_i$$

où les u_i sont des polynômes de degré n, vérifie :

$$\langle \phi_i, u \rangle = y_i$$

pour tout $i = 1, 2, ..., n$, où les ϕ_i sont des formes linéaires de $\mathcal{P}_N^D$ et les y_i des réels donnés. On utilise couramment des bases biorthogonales :

$$\langle \phi_i, u_j \rangle = \delta_{ij}$$

et on a alors simplement :

$$a_i = \langle \phi_i, u \rangle$$

L'exemple le plus courant est :

$$\langle \phi_i, u \rangle = u(x_i)$$
$$y_i = f(x_i)$$

pour une certaine fonction f à interpoler. Les conditions ci-dessus se résument alors à l'égalité de f et de u en un nombre fini de points :

$$u(x_i) = f(x_i)$$

On rencontre parfois aussi le cas :

$$\langle \phi_i, u \rangle = \frac{du}{dx}(x_i)$$
$$y_i = \frac{df}{dx}(x_i)$$

93.4.1 Lagrange

Les polynômes de Lagrange Λ_i sont biorthogonaux aux formes :

$$\langle \phi_i, u \rangle = u(x_i)$$

On a donc :

$$\langle \phi_j, \Lambda_i \rangle = \Lambda_i(x_j) = \delta_{ij}$$

Le polynôme Λ_i doit donc s'annuler en tout les points x_j, où $j \neq i$. On peut donc le factoriser comme :

$$\Lambda_i(x) = A_i \prod_{j \in E_i} (x - x_j) = A_i P_i(x)$$

où $E_i = \{1, 2, ..., n\} \setminus \{i\}$. Mais comme $\Lambda_i(x_i) = 1$, on a :

$$A_i = \frac{1}{P_i(x_i)}$$

et :

$$\Lambda_i(x) = \prod_{j \in E_i} \frac{(x - x_j)}{(x_i - x_j)}$$

Donc si on souhaite construire un polynôme :

$$w(x) = \sum_{i=1}^{n} u_i \Lambda_i(x)$$

qui interpole u en les x_i :

$$u(x_i) = w(x_i)$$

pour tout $i = 1, 2, ..., n$, il faut et il suffit de prendre :

$$u_i = \langle \phi_i, u \rangle = u(x_i)$$

93.4.2 Newton

L'interpolation de Newton utilise des polynômes construit récursivement à partir des polynômes de degré inférieur. Soit f la fonction à interpoler, $p_{i,j}$ le polynôme de degré $j - i$:

$$p_{ij}(x) = \sum_{j=0}^{j-i} a_k x^k$$

tels que :

$$p_{ij}(x_k) = f(x_k)$$

pour tous $k \in \{i, i+1, ..., j\}$. On voit que si $i = j$, on a :

$$p_{ii} = f(x_i)$$

Pour $i < j$, on peut construire les $p_{i,j}$ par récurrence. On vérifie que :

$$p_{ij}(x) = \frac{(x - x_i)p_{i+1,j}(x) - (x - x_j)p_{i,j-1}(x)}{x_j - x_i}$$

satisfait bien aux conditions d'interpolation ci-dessus.

Neuvième partie

Optimisation

Chapitre 94

Optimisation libre

94.1 Minimum

Théorème 94.1.1. *Soit $\varphi \in \mathrm{Cont}^2(\mathbb{R}^n, \mathbb{R})$. Supposons que que $a \in \mathbb{R}^n$ annule la Jacobienne (on parlera ici plutôt de gradient) :*

$$\partial\varphi(a) = 0$$

Supposons également que la Hessienne en a soit définie positive, c'est-à-dire que :

$$\Delta^\star \cdot \partial^2\varphi(a) \cdot \Delta \geq 0$$

pour tout $\Delta \in \mathbb{R}^n$ qui vérifie $\Delta \neq 0$. Si ces conditions sont remplies, nous nous proposons de montrer que φ atteint un minimum local en a. On peut donc trouver $\delta > 0$ tel que :

$$\varphi(a) \leq \varphi(a + \Delta)$$

pour tout $\Delta \in \mathbb{R}^n$ vérifiant $\|\Delta\| \leq \delta$. A l'inverse, si φ atteint un minimum local en a, les conditions sur le gradient et la Hessienne seront remplies.

— Supposons que les conditions sur le gradient et la hessienne soit remplies. Le développement d'ordre deux :

$$\varphi(a + \Delta) = \varphi(a) + \partial\varphi(a) \cdot \Delta + \frac{1}{2}\Delta^\star \cdot \partial^2\varphi(a) \cdot \Delta + E(\Delta)$$

où $E \sim o\left(\Delta^2\right)$ devient alors simplement :

$$\varphi(a + \Delta) = \varphi(a) + \frac{1}{2}\Delta^\star \cdot \partial^2\varphi(a) \cdot \Delta + E(\Delta)$$

Choisissons $h \in \mathbb{R}^n$. Pour un $\lambda \in \mathbb{R}$ quelconque, posons $\Delta = \lambda \cdot h$. On a alors :

$$\varphi(a + \Delta) = \varphi(a) + \frac{\lambda^2}{2} \cdot h^\star \cdot \partial^2\varphi(a) \cdot h + E(\lambda \cdot h)$$

Mais comme la Hessienne est définie positive et que E converge plus vite que $\|\Delta\|^2 = \lambda^2 \cdot \|h\|^2$ vers 0, il suffit de choisir $\lambda > 0$ assez petit pour avoir :

$$|E(\lambda \cdot h)| \leq \frac{\lambda^2}{2} \cdot h^\star \cdot \partial^2\varphi(a) \cdot h$$

on a alors :

$$\frac{1}{2}\Delta^\star \cdot \partial^2\varphi(a) \cdot \Delta + E(\Delta) \geq \frac{1}{2}\Delta^\star \cdot \partial^2\varphi(a) \cdot \Delta - |E(\Delta)| \geq 0$$

et :

$$\varphi(a+\Delta) = \varphi(a) + \frac{1}{2} \cdot \Delta^\star \cdot \partial^2 \varphi(a) \cdot \Delta + E(\Delta) \ge \varphi(a)$$

Nous avons donc bien un minimum local de φ en a.

— Inversément, si φ atteint un minimum local en a, nous avons vu que la différentielle s'annulait. La jacobienne s'annule donc aussi et le développement d'ordre deux devient :

$$\varphi(a+\Delta) = \varphi(a) + \frac{1}{2} \Delta^\star \cdot \partial^2 \varphi(a) \cdot \Delta + E(\Delta) \ge \varphi(a)$$

La condition de minimum local nous dit donc que :

$$\Delta^\star \cdot \partial^2 f(a) \cdot \Delta + E(\Delta) \ge 0$$

Choisissons à nouveau $h \in \mathbb{R}^n$ et posons $\Delta = \lambda \cdot h$ pour un $\lambda \in \mathbb{R}$ quelconque. On a alors :

$$\frac{\lambda^2}{2} \cdot h^\star \cdot \partial^2 f(a) \cdot h + E(\lambda \cdot h) \ge 0$$

Divisant par λ^2, on obtient :

$$\frac{1}{2} \cdot h^\star \cdot \partial^2 f(a) \cdot h + \frac{E(\lambda \cdot h)}{\lambda^2 \cdot \|h\|^2} \cdot \|h\|^2 \ge 0$$

Si on fait tendre $\lambda > 0$ vers 0, on arrive à la relation :

$$\frac{1}{2} \cdot h^\star \cdot \partial^2 f(a) \cdot h \ge 0$$

Comme ce résultat est valable quel que soit $h \in \mathbb{R}^n$, on en conclut que la Hessienne est définie positive.

94.2 Maximum

Un raisonnement analogue nous montre que :

Théorème 94.2.1. *Soit $\varphi \in \mathrm{Cont}^2(\mathbb{R}^n, \mathbb{R})$. Supposons que que $a \in \mathbb{R}^n$ annule la Jacobienne (on parlera ici plutôt de gradient) :*

$$\partial \varphi(a) = 0$$

Supposons également que la Hessienne en a soit définie négative, c'est-à-dire que :

$$\Delta^\star \cdot \partial^2 \varphi(a) \cdot \Delta \le 0$$

pour tout $\Delta \in \mathbb{R}^n$ qui vérifie $\Delta \neq 0$. Si ces conditions sont remplies, φ atteint un maximum local en a. On peut donc trouver $\delta > 0$ tel que :

$$\varphi(a) \ge \varphi(a+\Delta)$$

pour tout $\Delta \in \mathbb{R}^n$ vérifiant $\|\Delta\| \le \delta$. A l'inverse, si φ atteint un maximum local en a, les conditions sur le gradient et la Hessienne seront remplies.

94.3 Equivalence

En pratique, on peut toujours se ramener à un problème de minimisation. En effet, maximiser une fonction revient à minimiser son opposé :

$$\arg\max_{x \in A} \varphi(x) = \arg\min_{x \in A} (-\varphi(x))$$

Nous nous restreindrons donc dans la suite aux problèmes de minimisation.

94.4 Dérivées ordinaires

Pour des fonctions $f : \mathbb{R} \mapsto \mathbb{R}$, les conditions se simplifient en :

$$\frac{df}{dt}(a) = 0$$
$$\frac{d^2 f}{dt^2}(a) \geq 0$$

pour un minimum et en :

$$\frac{df}{dt}(a) = 0$$
$$\frac{d^2 f}{dt^2}(a) \leq 0$$

pour un maximum.

94.5 Point de selle

Soit une fonction $\mathfrak{L} : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$. Un point de selle $(\gamma, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ est un couple d'élément qui minimise $\mathfrak{L}(x, y)$ par rapport à x et qui la maximise par rapport à y. On aura alors :

$$\mathfrak{L}(\gamma, y) \leq \mathfrak{L}(\gamma, \lambda) \leq \mathfrak{L}(x, \lambda)$$

94.6 Convexité

Soit l'ensemble :

$$L = \{(s, t) \in \mathbb{R}^2 : (s, t) \geq 0 \text{ et } s + t = 1\}$$

Une fonction $\varphi : \mathbb{R}^n \mapsto \mathbb{R})$ est dite convexe si pour tout $u, v \in \mathbb{R}^n$ et $(s, t) \in L$, on a :

$$\varphi(s \cdot u + t \cdot v) \leq s \cdot \varphi(u) + t \cdot \varphi(v)$$

Formulation équivalente

En substituant $s = 1 - t$, on obtient :

$$\varphi(u + t \cdot (v - u)) \leq \varphi(u) + t \cdot (\varphi(v) - \varphi(u))$$

ou :

$$\varphi(u + t \cdot (v - u)) - \varphi(u) \leq t \cdot (\varphi(v) - \varphi(u))$$

Différentielle

Si φ est différentiable, on peut écrire par définition :

$$\mathfrak{D}^{\varphi}_u(t \cdot (v - u)) + o\left(t \cdot (v - u)\right) \leq t \cdot (\varphi(v) - \varphi(u))$$

On peut bien entendu faire sortir le t par linéarité :

$$t \cdot \mathfrak{D}^{\varphi}_u(v - u) + o\left(t \cdot (v - u)\right) \leq t \cdot (\varphi(v) - \varphi(u))$$

Il ne nous reste plus qu'à diviser par t et à faire tendre $t \to 0$ pour annuler le terme d'erreur. On a alors la borne supérieure :

$$\mathfrak{D}_u^\varphi(v-u) \le \varphi(v) - \varphi(u)$$

En termes matriciels, cela se réécrit :

$$\partial\varphi(u) \cdot (v-u) \le \varphi(v) - \varphi(u)$$

Hessienne

Supposons à présent que la fonction convexe $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ soit deux fois continûment différentiable. Soit $u, v \in \mathbb{R}^n$. Posons $\Delta = v - u$. On considère le développement d'ordre deux :

$$\varphi(v) = \varphi(u) + \partial\varphi(u) \cdot \Delta + \frac{1}{2} \cdot \Delta^\star \cdot \partial^2\varphi(u) \cdot \Delta + o\left(\Delta^2\right)$$

La borne supérieure de la différentielle nous dit que :

$$\partial\varphi(u) \cdot \Delta \le \varphi(v) - \varphi(u)$$

On a donc :

$$\frac{1}{2} \cdot \Delta^\star \cdot \partial^2\varphi(u) \cdot \Delta + o\left(\Delta^2\right) = \varphi(v) - \varphi(u) - \partial\varphi(u) \cdot \Delta \ge 0$$

c'est-à-dire :

$$\Delta^\star \cdot \partial^2\varphi(u) \cdot \Delta + o\left(\Delta^2\right) \ge 0$$

Choisissons $\delta \in \mathbb{R}^n$ et $t \in \mathbb{R}$. Considérons le cas où $v = u + t \cdot \delta$. On a alors $\Delta = t \cdot \delta$ et :

$$t^2 \cdot \delta^\star \cdot \partial^2\varphi(u) \cdot \delta + o\left(t^2 \cdot \delta^2\right) \ge 0$$

Il suffit alors de diviser par t^2 et de faire tendre $t \to 0$ pour obtenir :

$$\delta^\star \cdot \partial^2\varphi(u) \cdot \delta \ge 0$$

Comme ce doit être valable pour tout $\delta \in \mathbb{R}^n$, la Hessienne est définie positive en tout $u \in \mathbb{R}^n$.

Globalité

Considérons le minimum global :

$$\lambda \in \arg\min_{x \in \mathbb{R}^n} \varphi(x)$$

Supposons à présent que φ atteigne un minimum local en γ. On sait que $\varphi(\lambda) \le \varphi(\gamma)$ par définition de λ. Mais comme $\partial\varphi(\gamma) = 0$, on a aussi :

$$0 = \partial\varphi(\gamma) \cdot (\lambda - \gamma) \le \varphi(\lambda) - \varphi(\gamma)$$

On en déduit que :

$$\varphi(\gamma) \le \varphi(\lambda)$$

donc $\varphi(\gamma) = \varphi(\lambda)$ est également un minimum global.
Attention toutefois, cela ne prouve aucunement que le point minimisant φ est unique.

Corollaire

Un corollaire important de ces résultats est que si φ est convexe et deux fois différentiable et que l'on trouve un point x tel que $\partial\varphi(x) = 0$, ce point minimise localement φ car la Hessienne est définie positive. Le réel $\varphi(x)$ est donc également le minimum global de φ.

94.7 Convexité stricte

Une fonction $\varphi : \mathbb{R}^n \mapsto \mathbb{R})$ est dite strictement convexe si pour tout $u, v \in \mathbb{R}^n$ tels que $u \neq v$ et tout $(s, t) \in L \setminus \{(1, 0), (0, 1)\}$, on a :

$$\varphi(s \cdot u + t \cdot v) \ < \ s \cdot \varphi(u) + t \cdot \varphi(v)$$

Minima

Supposons qu'il existe deux $u, v \in \mathbb{R}^n$ distincts $(u \neq v)$ tels que $\varphi(u) = \varphi(v)$ soit le minimum global de φ. On a alors :

$$\begin{aligned} \varphi(s \cdot u + t \cdot v) &< s \cdot \varphi(u) + t \cdot \varphi(v) \\ &< (s + t) \cdot \varphi(u) = \varphi(u) \end{aligned}$$

ce qui n'est pas possible puisque $\varphi(u)$ est minimum global. Il existe donc au plus un seul $u \in \mathbb{R}^n$ minimisant globalement φ :

$$u = \arg \min_{x \in \mathbb{R}^n} \varphi(x)$$

94.8 Concavité

Une fonction $\psi : \mathbb{R}^n \mapsto \mathbb{R}$ est dite concave si pour tout $u, v \in \mathbb{R}^n$ et $(s, t) \in L$, on a :

$$\psi(s \cdot u + t \cdot v) \geq s \cdot \psi(u) + t \cdot \psi(v)$$

On voit que si ψ est concave, $\varphi = -\psi$ est convexe. L'équivalence max-min nous montre alors que tout maximum local de ψ est également un maximum global.

94.9 Concavité stricte

Une fonction $\psi : \mathbb{R}^n \mapsto \mathbb{R}$ est dite strictement concave si pour tout $u, v \in \mathbb{R}^n$ tels que $u \neq v$ et tout $(s, t) \in L \setminus \{(1, 0), (0, 1)\}$, on a :

$$\psi(s \cdot u + t \cdot v) \ > \ s \cdot \psi(u) + t \cdot \psi(v)$$

On voit que si ψ est strictement concave, $\varphi = -\psi$ est strictement convexe. L'équivalence max-min nous montre alors qu'il existe au plus un seul élément de $\mathbb{R}^n$ maximisant globalement ψ.

94.10 Equation du second degré

Soit $a, b, c \in \mathbb{R}$ avec $a \neq 0$. Soit le polynôme du second degré, ou polynôme de degré 2 défini par :

$$p(x) = a \cdot x^2 + b \cdot x + c$$

pour tout $x \in \mathbb{R}$. Nous allons chercher un éventuel extrema γ de ce polynôme. La dérivée première s'écrit :

$$\frac{dp}{dx}(x) = 2 \cdot a \cdot x + b$$

La condition :

$$\frac{dp}{dx}(\gamma) = 2 \cdot a \cdot \gamma + b = 0$$

nous donne :

$$\gamma = -\frac{b}{2a}$$

La dérivée seconde est donnée par :

$$\frac{d^2p}{dx^2}(x) = 2 \cdot a$$

Si $a > 0$, la dérivée seconde est toujours positive et γ minimise localement p. On vérifie que p est strictement convexe, et on en déduit que γ est l'unique réel qui minimise globalement p.

Par contre, si $a < 0$, la dérivée seconde est toujours négative et γ maximise localement p. On vérifie que p est strictement concave, et on en déduit que γ est l'unique réel qui maximise globalement p.

94.10.1 Racines

Intéressons-nous à présent à l'écart par rapport à γ :

$$\delta = x - \gamma = x + \frac{b}{2a}$$

On a donc :

$$x = \gamma + \delta = -\frac{b}{2a} + \delta$$

et :

$$x^2 = (\gamma + \delta)^2 = \delta^2 - \frac{\delta \cdot b}{a} + \frac{b^2}{4 \cdot a^2}$$

En injectant ces relations dans la définition du polynôme, on obtient :

$$\begin{aligned} p(\gamma + \delta) &= a \cdot \delta^2 - b \cdot \delta + \frac{b^2}{4 \cdot a} + b \cdot \delta - \frac{b^2}{2 \cdot a} + c \\ &= a \cdot \delta^2 - \frac{b^2}{4 \cdot a} + c \end{aligned}$$

Cette expression nous permet d'obtenir les racines d'un polynôme du second degré. La condition :

$$p(\delta + \gamma) = 0$$

nous donne :

$$\delta^2 = \frac{b^2 - 4 \cdot a \cdot c}{4 \cdot a^2}$$

équation qui admet deux solution δ_+, δ_- :

$$\delta_+ = \frac{\sqrt{b^2 - 4 \cdot a \cdot c}}{2 \cdot a}$$
$$\delta_- = -\frac{\sqrt{b^2 - 4 \cdot a \cdot c}}{2 \cdot a}$$

Nous avons donc deux racines x_+, x_- :

$$x_+ = \frac{-b + \sqrt{b^2 - 4 \cdot a \cdot c}}{2 \cdot a}$$

$$x_- = \frac{-b - \sqrt{b^2 - 4 \cdot a \cdot c}}{2 \cdot a}$$

telles que $p(x_+) = p(x_-) = 0$.

94.11 Moindres-carrés

Soit la matrice $A \in \mathbb{M}(\mathbb{R}, m, n)$ et le vecteur matriciel $b \in \mathbb{M}(\mathbb{R}, m, 1)$. On aimerait trouver le $\xi \in \mathbb{M}(\mathbb{R}^n, n, 1)$ qui minimise la norme de l'erreur e définie par :

$$e(x) = A \cdot x - b$$

pour tout $x \in \mathbb{M}(\mathbb{R}^n, n, 1)$. Comme la fonction $\|x\| \mapsto \|x\|^2$ est strictement croissante sur $\|x\| \in \mathbb{R}^+$, cela revient à minimiser :

$$\mathcal{E}(x) = \|x\|^2 = e(x)^\star \cdot e(x) = (A \cdot x - b)^\star \cdot (A \cdot x - b)$$

En développant, on obtient :

$$\mathcal{E}(x) = x^\star \cdot A^\star \cdot A \cdot x - x^\star \cdot A^\star \cdot b - b^\star \cdot A \cdot x + b^\star \cdot b$$

Comme $(A^\star \cdot A)^\star = A^\star \cdot A$ et $x^\star \cdot A^\star \cdot b = b^\star \cdot A^\star \cdot x$, l'annulation de la dérivée nous donne :

$$\partial \mathcal{E}(\xi) = 2A^\star \cdot A \cdot \xi - 2A^\star \cdot b = 0$$

d'où l'on tire directement :

$$A^\star \cdot A \cdot \xi = A^\star \cdot b$$

Si $A^\star \cdot A$ est inversible, on en déduit que :

$$\xi = (A^\star \cdot A)^{-1} \cdot A^\star \cdot b$$

Orthogonalité

Considérons la partition en colonne $A = [c_1 \ ... \ c_n]$. On a alors :

$$A^\star = \begin{bmatrix} c_1^\star \\ \vdots \\ c_n^\star \end{bmatrix}$$

La propriété :

$$A^\star \cdot (A \cdot \xi - b) = A^\star \cdot A \cdot \xi - A^\star \cdot b = 0$$

nous dit donc que les colonnes de A sont orthogonales au vecteur $r = A \cdot \xi - b$:

$$\langle c_i \mid r \rangle = c_i^\star \cdot r = \operatorname*{ligne}_i [A^\star \cdot (A \cdot \xi - b)] = 0$$

Approximation de fonctions

On désire obtenir une approximation w d'une fonction u dont on connaît les valeurs aux points $x_1, ..., x_m$ en minimisant l'erreur :

$$\sum_{i=1}^{m} (u(x_i) - w(x_i))^2$$

On choisit alors les fonctions $\varphi_1(x), ..., \varphi_n(x)$, où $n \le m$ et on pose :

$$w(x) = \sum_{i=1}^{m} a_i \cdot \varphi_i(x)$$

En utilisant les matrices :

$$\begin{aligned} A &= [\varphi_j(x_i)]_{i,j} \\ a &= [a_1 \ a_2 \ ... \ a_m]^T \\ b &= [u(x_1) \ u(x_2) \ ... \ u(x_n)]^T \end{aligned}$$

on peut réécrire le problème de minimisation comme suit :

$$a = \arg\min_z (A \cdot z - b)^\star \cdot (A \cdot z - b)$$

La solution est donc :

$$a = (A^\star \cdot A)^{-1} \cdot A^\star \cdot b$$

Chapitre 95

Projections

95.1 Minimisation de l'écart

Soit un espace vectoriel E sur $\mathbb{R}$ et des vecteurs $x, y \in E$, où $x \neq 0$. La projection de y sur x est le vecteur :

$$p \in \operatorname{span}\{x\} = \{\gamma \cdot x : \gamma \in \mathbb{R}\}$$

qui minimise sur $\operatorname{span}\{x\}$ la norme usuelle $\|e\| = \sqrt{\langle e \mid e \rangle}$ de l'écart e séparant y de p. Afin de trouver le $\lambda \in \mathbb{R}$ minimisant cet écart, on utilise la fonction $\mathcal{E} : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$\mathcal{E}(\gamma) = \langle y - \gamma \cdot x \mid y - \gamma \cdot x \rangle = \langle y \mid y \rangle - 2 \cdot \gamma \cdot \langle x \mid y \rangle + \gamma^2 \cdot \langle x \mid x \rangle$$

pour tout $\gamma \in \mathbb{R}$. Nous imposons l'annulation de la dérivée en un certain $\gamma = \lambda$ qui minimise potentiellement l'écart :

$$\partial \mathcal{E}(\lambda) = -2 \cdot \langle x \mid y \rangle + 2 \cdot \lambda \cdot \langle x \mid x \rangle = 0$$

ce qui nous donne :

$$\lambda = \frac{\langle x \mid y \rangle}{\langle x \mid x \rangle}$$

Produit scalaire complexe

Nous nous plaçons à présent dans le cas plus général où E est un espace vectoriel sur $\mathbb{C}$. Le λ défini plus haut devient donc un complexe. Nous allons déterminer s'il minimise la fonction $\mathcal{E} : \mathbb{C} \mapsto \mathbb{R}$ définie par :

$$\mathcal{E}(\gamma) = \|y - \gamma \cdot x\|^2$$

pour tout $\gamma \in \mathbb{C}$. Considérons le vecteur d'écart :

$$e = y - \lambda \cdot x$$

On voit qu'il vérifie la propriété :

$$\langle x \mid e \rangle = \langle x \mid y \rangle - \lambda \cdot \langle x \mid x \rangle = 0$$

On a donc aussi :

$$\langle e \mid x \rangle = \operatorname{conj} \langle x \mid e \rangle = 0$$

On dit que le vecteur d'écart est orthogonal à x. Soit l'écart $\delta = \gamma - \lambda$. On a alors $\gamma = \lambda + \delta$ et :

$$y - \gamma \cdot x = y - \lambda \cdot x - \delta \cdot x = e - \delta \cdot x$$

En utilisant les propriétés du produit scalaire, on en déduit :

$$\begin{aligned} \mathcal{E}(\gamma) &= \langle e - \delta \cdot x \mid e - \delta \cdot x \rangle \\ &= \langle e \mid e \rangle - \delta \cdot \langle e \mid x \rangle - \overline{\delta} \cdot \langle x \mid e \rangle + |\delta|^2 \cdot \langle x \mid x \rangle \\ &= \langle e \mid e \rangle + |\delta|^2 \cdot \langle x \mid x \rangle \end{aligned}$$

Comme $|\delta|^2 \cdot \langle x \mid x \rangle$ est un réel ≥ 0, on a finalement :

$$\mathcal{E}(\gamma) \geq \langle e \mid e \rangle = \mathcal{E}(\lambda)$$

Notre paramètre λ ainsi défini minimise donc bien $\mathcal{E}$ sur $\mathbb{C}$. De plus, si $\delta \neq 0$ on a $|\delta|^2 > 0$ et $\mathcal{E}(\gamma) > \langle e \mid e \rangle$, ce qui prouve que :

$$\lambda = \arg \min_{\gamma \in \mathbb{C}} \mathcal{E}(\gamma)$$

est l'unique complexe à minimiser $\mathcal{E}$. La racine carrée étant une fonction monotone strictement croissante sur $\mathbb{R}$, la norme $\|e\| = \sqrt{\mathcal{E}(\lambda)}$ est donc également minimisée et la projection s'écrit :

$$p = \lambda \cdot x = \frac{\langle x \mid y \rangle}{\langle x \mid x \rangle} \cdot x$$

95.2 Théorème de Pythagore

Calculons à présent la norme de y en partant de la relation :

$$y = p + e$$

On déduit de la propriété d'orthogonalité que :

$$\langle p \mid e \rangle = \lambda \cdot \langle x \mid e \rangle = 0$$

On peut donc appliquer le théorème de pythagore, qui nous dit que :

$$\|y\|^2 = \|p + e\|^2 = \|p\|^2 + \|e\|^2$$

95.3 Carré de la distance

Si :

$$D = \min_{\gamma \in \mathbb{C}} \|y - \gamma \cdot x\|$$

on a donc :

$$D^2 = \|y - \lambda \cdot x\|^2 = \|e\|^2 = \|y\|^2 - \|p\|^2$$

et finalement :

$$D^2 = \|y\|^2 - |\lambda|^2 \cdot \|x\|^2$$

95.4 Cauchy-Schwartz

Par positivité du produit scalaire, on a bien évidemment :

$$\|e\|^2 = \langle e \mid e \rangle \geq 0$$

Le théorème de Pythagore implique donc que :

$$\langle p \mid p \rangle = \|p\|^2 \leq \|y\|^2 = \langle y \mid y \rangle$$

En substituant $p = \lambda \cdot x$, on obtient :

$$\overline{\lambda} \cdot \lambda \cdot \langle x \mid x \rangle \leq \langle y \mid y \rangle$$

Mais comme :

$$\overline{\lambda} = \operatorname{conj} \frac{\langle x \mid y \rangle}{\langle x \mid x \rangle} = \frac{\langle y \mid x \rangle}{\langle x \mid x \rangle}$$

on a :

$$\frac{\langle y \mid x \rangle}{\langle x \mid x \rangle} \cdot \frac{\langle x \mid y \rangle}{\langle x \mid x \rangle} \cdot \langle x \mid x \rangle \leq \langle y \mid y \rangle$$

En simplifiant et en multipliant par la norme carrée de x, on a finalement :

$$\langle y \mid x \rangle \cdot \langle x \mid y \rangle \leq \langle x \mid x \rangle \cdot \langle y \mid y \rangle$$

relation dont la racine nous donne l'inégalité de Cauchy-Schwartz :

$$|\langle x \mid y \rangle| \leq \|x\| \cdot \|y\|$$

95.5 Propriétés extrémales

L'égalité :

$$\|p\|^2 = \|y\|^2$$

n'est atteinte que lorsque $\|e\|^2 = 0$, c'est-à-dire $e = 0$ et $y = \lambda \cdot x$ pour un certain $\lambda \in \mathbb{C}$. Ce constat nous amène aux problèmes d'optimisations suivants. Soit un vecteur $c \neq 0$ fixé et :

$$d = \|c\| = \sqrt{\langle c \mid c \rangle}$$

On cherche à maximiser ou à minimiser :

$$\varphi(x) = \langle c \mid x \rangle$$

sur l'ensemble $B = \mathfrak{B}(0, r) = \{x : \|x\| \leq r\}$. L'inégalité de Cauchy-Schwartz nous dit que :

$$-d \cdot r \leq -\|c\| \cdot \|x\| \leq \langle c \mid x \rangle \leq \|c\| \cdot \|x\| \leq d \cdot r$$

Nous allons chercher nos solutions sous la forme $x = \alpha \cdot c$, pour un certain $\alpha \in \mathbb{C}$. On a alors :

$$\|\alpha \cdot c\| = |\alpha| \cdot \|c\| = |\alpha| \cdot d \leq r$$

et donc $|\alpha| \leq r/d$. Si on choisit $\alpha = r/d$, on a :

$$\langle c \mid x \rangle = \frac{r}{d} \cdot \langle c \mid c \rangle = \frac{r}{d} \cdot d^2 = d \cdot r$$

La borne supérieure est atteinte. La fonction φ est donc maximisée :

$$\eta = \frac{r}{d} \cdot c \in \arg\max_{x \in B} \langle c \mid x \rangle$$

Si on choisit $\alpha = -r/d$, on a :

$$\langle c \mid x \rangle = -\frac{r}{d} \cdot \langle c \mid c \rangle = -d \cdot r$$

La borne inférieure est atteinte. La fonction φ est donc minimisée :

$$\theta = -\frac{r}{d} \cdot c \in \arg\min_{x \in B} \langle c \mid x \rangle$$

95.6 Sous-espace vectoriel

Soit l'espace vectoriel E sur $\mathbb{R}$ et un sous-espace vectoriel $U \subseteq E$ possédant une base orthonormée $(u_1, ..., u_n)$. La projection d'un vecteur quelconque $z \in E$ sur U est le vecteur $p \in U$ qui minimise la norme de l'écart entre z et p. On cherche donc le $\lambda = (\lambda_1, ..., \lambda_n) \in \mathbb{R}^n$ qui minimise la fonction $\mathcal{E} : \mathbb{R}^n \mapsto \mathbb{R}$ définie par :

$$\mathcal{E}(\gamma) = \left\| z - \sum_{i=1}^{n} \gamma_i \cdot u_i \right\|^2$$

pour tout $\gamma = (\gamma_1, ..., \gamma_n) \in \mathbb{R}^n$. En utilisant $\langle u_i \mid u_j \rangle = \boldsymbol{\delta}_{ij}$, on obtient :

$$\begin{aligned} \mathcal{E}(\gamma) &= \langle z \mid z \rangle - 2 \sum_i \gamma_i \cdot \langle u_i \mid z \rangle + \sum_{i,j} \gamma_i \cdot \gamma_j \cdot \langle u_i \mid u_j \rangle \\ &= \langle z \mid z \rangle - 2 \sum_i \gamma_i \cdot \langle u_i \mid z \rangle + \sum_i \gamma_i^2 \end{aligned}$$

Imposant $\partial_k \mathcal{E}(\lambda_1, ..., \lambda_n) = 0$, on obtient les n équations :

$$-2\lambda_k \cdot \langle u_k \mid z \rangle + 2\lambda_k = 0$$

On en déduit que le choix :

$$\lambda = (\lambda_1, ..., \lambda_n) = (\langle u_1 \mid z \rangle, ..., \langle u_n \mid z \rangle)$$

minimise potentiellement $\mathcal{E}$. Notre projection potentielle de z sur U est donc donnée par :

$$p = \sum_i \langle u_i \mid z \rangle \cdot u_i$$

Produit scalaire complexe

Nous nous plaçons à présent dans le cas plus général où E est un espace vectoriel sur $\mathbb{C}$. Le λ défini plus haut devient donc un élément de $\mathbb{C}^n$. Nous allons déterminer s'il minimise la fonction $\mathcal{E} : \mathbb{C}^n \mapsto \mathbb{R}$ définie par :

$$\mathcal{E}(\gamma) = \left\| z - \sum_{i=1}^{n} \gamma_i \cdot u_i \right\|^2$$

pour tout $\gamma = (\gamma_1, ..., \gamma_n) \in \mathbb{C}^n$. Choisissons $x \in U$. On a alors :

$$x = \sum_i x_i \cdot u_i$$

où $x_i = \langle u_i \mid z \rangle$. Considérons le vecteur d'écart :

$$e = z - p$$

On a alors :

$$\begin{aligned}
\langle x \mid e \rangle &= \sum_i \overline{x_i} \cdot \langle u_i \mid z - p \rangle \\
&= \sum_i \overline{x_i} \cdot \langle u_i \mid z \rangle - \sum_{i,j} \overline{x_i} \cdot \langle u_i \mid u_j \rangle \cdot \lambda_j \\
&= \sum_i \langle x \mid u_i \rangle \cdot \langle u_i \mid z \rangle - \sum_i \langle x \mid u_i \rangle \cdot \langle u_i \mid z \rangle = 0
\end{aligned}$$

Le vecteur d'écart est donc orthogonal à tous les $x \in U$. On a également :

$$\langle e \mid x \rangle = \operatorname{conj} \langle x \mid e \rangle = 0$$

Soit l'écart $\delta = \gamma - \lambda$. On a alors $\gamma = \lambda + \delta$. Posons :

$$\begin{aligned}
g &= \sum_i \gamma_i \cdot u_i \\
p &= \sum_i \lambda_i \cdot u_i \\
\Delta &= \sum_i \delta_i \cdot u_i
\end{aligned}$$

On a alors $z - g = z - p - \Delta = e - \Delta$ et :

$$\begin{aligned}
\mathcal{E}(\gamma) &= \langle e - \Delta \mid e - \Delta \rangle \\
&= \langle e \mid e \rangle - \langle e \mid \Delta \rangle - \langle \Delta \mid e \rangle + \langle \Delta \mid \Delta \rangle
\end{aligned}$$

Comme $\Delta \in U$, ses produits scalaires avec e s'annulent par orthogonalité et on a finalement :

$$\mathcal{E}(\gamma) = \langle e \mid e \rangle + \langle \Delta \mid \Delta \rangle \ge \langle e \mid e \rangle = \mathcal{E}(\lambda)$$

Les scalaires λ_k minimisent donc bien la norme de l'écart sur U.

Théorème de Pythagore

On a $z = p+e$. Comme p est un vecteur de U, on a $\langle p \mid e \rangle = \langle e \mid p \rangle = 0$. Le théorème de Pythagore est donc applicable :

$$\|z\|^2 = \|p\|^2 + \|e\|^2$$

Carré de la distance

Soit :

$$D = \min\{\|z - v\| : v \in U\}$$

Par orthonormalité, on a :

$$\|p\|^2 = \sum_{i,j} \mathrm{conj}(\langle u_i \mid z \rangle) \cdot \langle u_j \mid z \rangle \cdot \langle u_i \mid u_j \rangle = \sum_i |\langle u_i \mid z \rangle|^2$$

et donc :

$$D^2 = \|z\|^2 - \|p\|^2 = \|z\|^2 - \sum_i |\langle u_i \mid z \rangle|^2$$

95.7 Tenseur de projection

On peut réécrire la projection de z sur U sous la forme :

$$p = \sum_i u_i \cdot \langle u_i \mid z \rangle$$

Cette expression ressemble à la contraction d'ordre 1 d'un tenseur d'ordre 2. Effectivement, si on pose :

$$\mathcal{P} = \sum_i u_i \otimes u_i$$

on a alors :

$$p = \mathcal{P} \cdot z = \langle \mathcal{P} \odot z \rangle_1$$

Identité locale

Pour tout $y \in U$, on a :

$$y = \sum_i \langle u_i \mid y \rangle \cdot u_i$$

et :

$$\mathcal{P} \cdot y = \sum_i u_i \cdot \langle u_i \mid y \rangle$$

Ces deux expressions étant identiques, on a $\mathcal{P} \cdot y = y$. Le tenseur de projection correspond donc localement (sur U) au tenseur identité.

Invariance

Pour tout $z \in E$, on a $y = \mathcal{P} \cdot z \in U$. On en conclut que :

$$\mathcal{P} \cdot z = y = \mathcal{P} \cdot y = \mathcal{P} \cdot (\mathcal{P} \cdot z)$$

Donc $\mathcal{P} = \mathcal{P} \cdot \mathcal{P}$.

Complémentaire

Soit $(u_1, ..., u_n)$ une base orthonormée de E. On voit que le tenseur identité :

$$\boldsymbol{\mathcal{I}} = \sum_{i=1}^{n} u_i \otimes u_i$$

est le tenseur de projection de E dans lui-même. On considère le tenseur de projection sur $\mathrm{span}\{u_1, ..., u_m\}$, où $m \le n$:

$$\mathcal{P} = \sum_{i=1}^{m} u_i \otimes u_i$$

Le tenseur complémentaire est défini par :

$$\mathcal{Q} = \boldsymbol{\mathcal{I}} - \mathcal{P} = \sum_{i=1}^{n} u_i \otimes u_i - \sum_{i=1}^{m} u_i \otimes u_i = \sum_{i=m+1}^{n} u_i \otimes u_i$$

Il s'agit donc d'un tenseur de projection sur l'espace complémentaire $\mathrm{span}\{u_{m+1}, ..., u_n\}$. Il est lié à l'écart de projection car $e = z - p = \mathcal{Q} \cdot z$. On remarque l'orthogonalité :

$$\mathcal{Q} \cdot \mathcal{P} = (\boldsymbol{\mathcal{I}} - \mathcal{P}) \cdot \mathcal{P} = \mathcal{P} - \mathcal{P} \cdot \mathcal{P} = 0$$

De même :

$$\mathcal{P} \cdot \mathcal{Q} = \mathcal{P} \cdot \mathcal{P} - \mathcal{P} = 0$$

On a aussi sans surprise :

$$\mathcal{Q} \cdot \mathcal{Q} = \mathcal{Q} - \mathcal{Q} \cdot \mathcal{P} = \mathcal{Q}$$

95.8 Gram-Schmidt

Nous allons construire une suite de vecteurs orthonormaux $(u_1, u_2, ...u_n)$ à partir d'une suite $(a_1, a_2, ..., a_n)$ de vecteurs linéairement indépendants de E. L'indépendance linéaire nous garantit que $a_1 \neq 0$. On peut donc normaliser pour obtenir le premier vecteur de la série :

$$u_1 = \frac{a_1}{\|a_1\|}$$

On a alors :

$$\langle u_1 \mid u_1 \rangle = \frac{\langle a_1 \mid a_1 \rangle}{\|a_1\|^2} = \frac{\langle a_1 \mid a_1 \rangle}{\langle a_1 \mid a_1 \rangle} = 1$$

Nous projetons a_2 sur u_1 en utilisant le tenseur $\mathcal{P}_1 = u_1 \otimes u_1$ et nous évaluons l'écart :

$$e_2 = (\boldsymbol{\mathcal{I}} - \mathcal{P}_1) \cdot a_2 = a_2 - u_1 \cdot \langle u_1 \mid a_2 \rangle$$

Les propriétés d'orthogonalité de l'écart nous assurent alors que $\langle u_1 \mid e_2 \rangle = 0$. L'indépendance linéaire nous garantit que $e_2 \neq 0$. On peut donc normaliser en utilisant $u_2 = e_2 / \|e_2\|$. On a alors clairement $\langle u_2 \mid u_2 \rangle = 1$ et $\langle u_1 \mid u_2 \rangle = 0$.

Supposons à présent avoir trouvé la suite orthonormée $(u_1, u_2, ..., u_k)$, où $k \leq n-1$. Nous projetons a_{k+1} sur l'espace vectoriel $U_k = \mathrm{span}\{u_1, u_2, ..., u_k\}$ en utilisant le tenseur :

$$\mathcal{P}_k = \sum_{i=1}^{k} u_i \otimes u_i$$

et nous évaluons l'écart :

$$e_{k+1} = (\boldsymbol{\mathcal{I}} - \mathcal{P}_k) \cdot a_{k+1} = a_{k+1} - \sum_{i=1}^{k} u_i \cdot \langle u_i \mid a_{k+1} \rangle$$

Les propriétés d'orthogonalité de l'écart nous assurent alors que $\langle u_j \mid e_{k+1} \rangle = 0$ pour tout $j \in \{1, 2, ..., k\}$. L'indépendance linéaire nous garantit que $e_{k+1} \neq 0$. On peut donc normaliser en utilisant :

$$u_{k+1} = \frac{e_{k+1}}{\|e_{k+1}\|}$$

On a alors clairement $\langle u_j \mid u_{k+1} \rangle = \boldsymbol{\delta}_{j,k+1}$.
Nous disposons donc à la fin du processus d'une suite de vecteurs $(u_1, u_2, ..., u_n)$ orthonormée :

$$\langle u_i \mid u_j \rangle = \boldsymbol{\delta}_{ij}$$

Ce procédé est appelé processus d'orthogonalisation de Gram-Schmidt.

Remarque

Dans le cas où l'indépendance linéaire n'est pas garantie, il est toujours possible d'adapter l'algorithme en enlevant dynamiquement de la liste des a_k les vecteurs donnant un écart de projection nul. On se retrouve alors à la fin du processus avec une suite orthonormée $(u_1, ..., u_m)$, où $m \le n$.

95.9 Représentation matricielle

Soient $u_1, u_2, ..., u_m \in \mathbb{M}(\mathbb{K}, n, 1)$ des vecteurs matriciels orthonormés pour le produit scalaire usuel :

$$u_i^\star \cdot u_j = \boldsymbol{\delta}_{ij}$$

La matrice de projection associée au tenseur de projection sur $U = \operatorname{span}\{u_1, ..., u_m\}$ s'écrit :

$$P = \sum_{k=1}^m u_k \otimes u_k = \sum_{k=1}^m u_k \cdot u_k^\star$$

Il s'agit donc d'une matrice de taille (n, n). Elle permet de projeter tout vecteur matriciel $z \in \mathbb{M}(\mathbb{K}, n, 1)$ sur U :

$$P \cdot z = \sum_{k=1}^m u_k \cdot u_k^\star \cdot z = \sum_{k=1}^m u_k \cdot \langle u_k \mid z \rangle$$

95.10 Factorisation

En terme de composantes, si $u_k = (u_{kj})_j$ et si $P = (p_{ij})_{i,j}$, on a :

$$p_{ij} = \sum_{k=1}^m u_{ki} \cdot \bar{u}_{kj}$$

expression qui ressemble furieusement à un produit matriciel. Soit la matrice U de taille (n, m) rassemblant les m vecteurs u_k :

$$U = [u_1 \ u_2 \ \dots \ u_m]$$

On a alors :

$$\begin{aligned}
\operatorname*{comp}_{ik} U &= u_{ki} \\
\operatorname*{comp}_{kj} U^\star &= \operatorname{conj} \operatorname*{comp}_{jk} U = \bar{u}_{kj}
\end{aligned}$$

Le produit ci-dessus peut donc s'écrire :

$$p_{ij} = \sum_{k=1}^{m} b_{ik} \cdot c_{kj}$$

où $b_{ik} = \mathrm{comp}_{ik} U$ et $c_{kj} = \mathrm{comp}_{kj} U^\star$. On a donc finalement :

$$P = U \cdot U^\star$$

95.10.1 Propriétés

Comme les u_k sont orthonormaux, on a :

$$\underset{ij}{\mathrm{comp}}(U^\star \cdot U) = u_i^\star \cdot u_j = \langle u_i \mid u_j \rangle = \boldsymbol{\delta}_{ij}$$

On a donc $U^\star \cdot U = I$, la matrice identité de taille (m, m). On a également, comme attendu :

$$P^2 = U \cdot U^\star \cdot U \cdot U^\star = U \cdot I \cdot U^\star = U \cdot U^\star = P$$

Si on pose $Q = I - P$, on a aussi :

$$P \cdot Q = Q \cdot P = P - P^2 = 0$$

Cas particulier

Les m vecteurs u_i étant linéairement indépendants dans $\mathbb{K}^n$ qui est de dimension n, on doit forcément avoir $m \le n$. Dans le cas où $m = n$, on a de plus $U^\star = U^{-1}$ et :

$$\sum_{i=1}^{n} u_i \otimes u_i = P = U \cdot U^{-1} = I$$

95.11 Vecteurs A-orthogonaux

Soit une matrice de produit scalaire $A \in \mathbb{M}(\mathbb{K}, n, n)$. On peut utiliser le procédé de Gram-Schmidt pour construire une base de vecteurs orthogonaux pour le produit scalaire :

$$\langle x \mid y \rangle = x^\star \cdot A \cdot y$$

On part d'une suite $(a_1, a_2, ..., a_n)$ de vecteurs matriciels linéairement indépendants de $\mathbb{K}^n$ (typiquement la base canonique). On commence par normaliser a_1 :

$$u_1 = \frac{a_1}{\sqrt{a_1^\star \cdot A \cdot a_1}}$$

et on construit étape après étape la suite des u_i. Supposons être arrivé à la suite $(u_1, u_2, ..., u_k)$ vérifiant $\langle u_i \mid u_j \rangle = u_i^\star \cdot A \cdot u_j = \boldsymbol{\delta}_{ij}$, où $k \le n - 1$. Nous projetons a_{k+1} sur l'espace vectoriel $\mathrm{span}\{u_1, u_2, ..., u_k\}$ et nous évaluons l'écart :

$$\begin{aligned} e_{k+1} &= a_{k+1} - \sum_{i=1}^{k} u_i \cdot \langle u_i \mid a_{k+1} \rangle \\ &= a_{k+1} - \sum_{i=1}^{k} u_i \cdot u_i^\star \cdot A \cdot a_{k+1} \end{aligned}$$

Ensuite, nous normalisons le résultat :

$$u_{k+1} = \frac{e_{k+1}}{\sqrt{e_{k+1}^\star \cdot A \cdot e_{k+1}}}$$

Nous disposons donc à la fin du processus d'une suite de vecteurs $(u_1, u_2, ..., u_n)$ orthonormée :

$$\langle u_i \mid u_j \rangle = u_i^\star \cdot A \cdot u_j = \boldsymbol{\delta}_{ij}$$

On dit que la suite des u_i est A-orthonormée. Si on définit la famille de matrices :

$$U_m = [u_1 \ u_2 \ ... \ u_m]$$

pour tout $m \le n$, on peut réecrire l'orthogonalité comme suit :

$$U_m^\star \cdot A \cdot U_m = I_m$$

On note aussi $U = U_n$.

Complément orthogonal

Si les vecteurs orthonormaux $(u_1, ..., u_p)$, où $p \le n$, sont donnés, on peut simplement commencer le processus à $k = p$ pour compléter la suite de vecteurs jusqu'à $k = n$. On obtient alors la suite orthonormée $(u_1, ..., u_p, u_{p+1}, ..., u_n)$. On dit que $(u_{p+1}, ..., u_n)$ est le complément orthogonal de $(u_1, ..., u_p)$.

Orthogonalité usuelle

Dans le cas où l'on choisit $A = I$, cette méthode offre un moyen d'obtenir (ou de compléter) une suite de vecteurs u_i tels que $u_i^\star \cdot u_j = \boldsymbol{\delta}_{ij}$ et des matrices U_m correspondantes telles que $U_m^\star \cdot U_m = I_m$. Comme $U = U_n$ est carrée, on a de plus :

$$U^{-1} = U^\star$$

Systèmes linéaires

Lorsqu'on dispose de n vecteurs A-orthonormés, il est facile de résoudre le système linéaire $A \cdot x = y$. Les u_i forment alors une base de $\mathbb{K}^n$ et on peut trouver des scalaires $x_i \in \mathbb{K}$ tels que :

$$x = \sum_{i=1}^n x_i \cdot u_i$$

En prenant le produit scalaire de x avec u_k, on obtient :

$$u_k^\star \cdot A \cdot x = \sum_{i=1}^n (u_k^\star \cdot A \cdot u_i) \cdot x_i = x_k$$

On voit donc apparaître une expression analogue à celle d'une projection usuelle :

$$x = \sum_{i=1}^n u_i \cdot (u_i^\star \cdot A \cdot x) = \sum_{i=1}^n u_i \cdot u_i^\star \cdot y$$

Attention, analogue n'est pas identique, les u_i ne sont en général pas orthonormés pour le produit scalaire usuel.

Chapitre 96

Algorithmes d'optimisation libre

96.1 Introduction

Nous allons présenter des algorithmes permettant de résoudre approximativement des problèmes de minimisation d'une fonction φ sur $\mathbb{R}^n$. Ces algorithmes partent d'un point initial $x_0 \in \Omega$ et itèrent schématiquement comme suit :

$$x_{k+1} = I(x_k) = x_k + p_k$$

pour un certain $p_k \in \mathbb{R}^n$. On espère bien entendu que la suite converge et que :

$$x_N \approx \lim_{k \to \infty} x_k = \arg \min_{x \in \Omega} \varphi(x)$$

pour N assez grand. Nous adoptons les notations :

$$J = (\partial \varphi)^\star$$

pour le gradient, de taille $(n, 1)$, et :

$$H = \partial^2 \varphi$$

pour la hessienne, de taille (n, n). On note également :

$$\Phi_k = \Phi(x_k)$$

pour toute fonction Φ (par exemple, $\Phi \in \{\varphi, J, H\}$).

96.2 Minimum dans une direction

Soit l'itération :

$$x_{k+1} = x_k - \alpha_k \cdot p_k$$

où $\alpha_k \in \mathbb{R}$ et $p_k \in \mathbb{R}^n$. On choisit généralement le paramètre α_k de façon à minimiser le développement d'ordre deux :

$$\varphi_{k+1} \approx \varphi_k - \alpha_k \cdot J_k^\star \cdot p_k + \frac{\alpha_k^2}{2} \cdot p_k^\star \cdot H_k \cdot p_k$$

En imposant l'annulation de la dérivée de ce développement par rapport à α_k, on en déduit que :

$$-J_k^\star \cdot p_k + \alpha_k \cdot p_k^\star \cdot H_k \cdot p_k = 0$$

La valeur optimale de α_k s'écrit :

$$\alpha_k = \frac{J_k^\star \cdot p_k}{p_k^\star \cdot H_k \cdot p_k}$$

96.3 Méthode de la plus grande descente

La méthode de la plus grande pente consiste à partir à chaque itération du point $x^{(k)}$ et à suivre la direction δ_k où φ descend le plus rapidement dans le voisinage immédiat. En première approximation, si :

$$x_{k+1} = x_k + \alpha_k \cdot \delta_k$$

pour un certain $\alpha_k \in \mathbb{R}$, on a :

$$\varphi_{k+1} \approx \varphi_k + \alpha_k \cdot J_k^\star \cdot \delta_k$$

Nous choisissons le vecteur δ_k qui minimise $J_k^\star \cdot \delta_k = \langle J_k \mid \delta_k \rangle$ sur $\mathfrak{B}(0, \|J_k\|)$, c'est-à-dire :

$$\delta_k = -J_k$$

On a alors :

$$x_{k+1} = x_k - \alpha_k \cdot J_k$$

La valeur optimale de α_k s'écrit donc :

$$\alpha_k = \frac{J_k^\star \cdot J_k}{J_k^\star \cdot H_k \cdot J_k}$$

96.4 Newton

Il s'agit ici d'optimiser le pas s_k :

$$x_{k+1} = x_k + s_k$$

pour minimiser le développement :

$$\varphi_{k+1} \approx \varphi_k + J_k^\star \cdot s_k + \frac{1}{2} \cdot s_k^\star \cdot H_k \cdot s_k$$

Mais comme $H = H^\star$, l'annulation du gradient par rapport à s_k nous donne :

$$J_k + H_k \cdot s_k \approx 0$$

On en déduit la valeur optimale :

$$s_k = -H_k^{-1} \cdot J_k$$

96.5 Gradients conjugués

Soit l'itération :

$$x_{k+1} = x_k - \alpha_k \cdot p_k$$

où $\alpha_k \in \mathbb{R}$ et $p_k \in \mathbb{R}^n$. On a comme d'habitude :

$$\alpha_k = \frac{J_k^\star \cdot p_k}{p_k^\star \cdot H_k \cdot p_k}$$

Lorsque $k = 0$, on prend le gradient comme direction :

$$p_0 = J_0$$

Pour $k \ge 1$, on choisit p_k comme combinaison linéaire du gradient J_k et du pas précédent p_{k-1} :

$$p_k = J_k - \beta_k \cdot p_{k-1}$$

où $\beta_k \in \mathbb{R}$. L'idée des gradients conjugué est de construire une suite de p_k orthogonaux entre-eux afin de minimiser la fonction dans toutes les directions. Au bout de n itérations, on espère avoir construit une base de $\mathbb{R}^n$ et être très proche du minimum global. En fait, nous n'allons pas vérifier que toutes les directions sont orthogonales, mais seulement que deux directions successives le sont. On demande donc l'orthogonalité au sens du produit scalaire défini par H :

$$p_k^\star \cdot H_k \cdot p_{k-1} = J_k^\star \cdot H_k \cdot p_{k-1} - \beta_k \cdot p_{k-1}^\star \cdot H_k \cdot p_{k-1} = 0$$

On en déduit la valeur de β_k :

$$\beta_k = \frac{J_k^\star \cdot H_k \cdot p_{k-1}}{p_{k-1}^\star \cdot H_k \cdot p_{k-1}}$$

Nous allons à présent obtenir une valeur approximative de β_k en fonction des variations du gradient. Le développement d'ordre un de J s'écrit :

$$J_k - J_{k-1} \approx H_k \cdot (x_k - x_{k-1}) = -\alpha_k \cdot H_k \cdot p_{k-1}$$

On en déduit que :

$$\beta_k \approx \frac{J_k^\star \cdot (J_k - J_{k-1})}{p_{k-1}^\star \cdot (J_k - J_{k-1})}$$

96.6 Moindres carrés

Soit la fonction $f : \mathbb{R}^n \mapsto \mathbb{R}^m$. On cherce à minimiser la fonction $\varphi = f^\star \cdot f/2$. Le problème de minimisation s'écrit alors :

$$\arg\min_x \frac{1}{2} f(x)^\star \cdot f(x)$$

Dans ce cas, si on définit :

$$D = \partial f$$

le gradient s'écrit :

$$J = D^\star \cdot f$$

Si on suppose que les dérivées secondes de f sont négligeables par rapport aux dérivées premières, on a :

$$H \approx D^\star \cdot D$$

La méthode de Newton devient dans ce cas :

$$x_{k+1} = x_k + s_k$$

avec :

$$s_k = -[D_k^\star \cdot D_k]^{-1} \cdot D_k^\star \cdot f_k$$

Zéros

Si $S = \ker f \neq \emptyset$, soit $\gamma \in S$. On a alors :

$$0 \le \min \frac{1}{2} f(x)^\star \cdot f(x) \le \frac{1}{2} f(\gamma)^\star \cdot f(\gamma) = 0$$

On en déduit que tout $x \in \mathbb{R}^n$ minimisant φ vérifiera $f(x) = 0$. La méthode de minimisation nous fournit donc une approximation d'un tel x.

96.7 Levenberg-Marquardt

C'est une variante de la méthode des moindres carrés, utile dans les cas où $D_k^\star \cdot D_k$ est numériquement proche d'une matrice non inversible. On ajoute alors la matrice identité multipliée par un scalaire λ sur la diagonale :

$$s_k = -[D_k^\star \cdot D_k + \lambda_k \cdot I]^{-1} \cdot D_k^\star \cdot f_k$$

Chapitre 97

Solveurs itératifs

97.1 Dépendances

— Chapitre ?? : Les matrices

97.2 Méthodes itératives

Il s'agit de résoudre itérativement (et approximativement) en x un système linéaire $A \cdot x = b$, où A est une matrice et b un vecteur. L'itération générique s'écrit :

$$x_{k+1} = I(x_k)$$

et on espère que la suite des $x_k \in \mathbb{R}^n$ converge vers la solution. On initialise en général les algorithmes avec $x_0 = 0$.

97.3 Résolution par minimisation

Soit une matrice H de taille (n, n) hermitienne ($H = H^\star$) et définie positive :

$$x^\star \cdot H \cdot x \ge 0$$

pour tout $x \in \mathbb{R}^n$. Considérons la fonction définie par :

$$\varphi(x) = \frac{1}{2} x^\star \cdot H \cdot x - x^\star \cdot b$$

Si un certain x minimise φ sur $\mathbb{R}^n$, on doit avoir :

$$\partial\varphi(x) = H \cdot x - b = 0$$

ce qui revient à résoudre le système linéaire :

$$H \cdot x = b$$

Comme $\partial^2\varphi = H$ est définie positive, résoudre le système $H \cdot x = b$ en x revient donc à minimiser φ sur $\mathbb{R}^n$. On peut obtenir une approximation de la solution en utilisant la méthode de la plus grande descente, le gradient étant donné par :

$$J = (\partial\varphi)^\star = H \cdot x - b$$

On a donc des itérations k de la forme :

$$x_{k+1} = x_k + \alpha_k \cdot (b - H \cdot x)$$

pour un $\alpha_k \in \mathbb{R}$ bien choisi. On peut aussi utiliser un autre algorithme, comme les gradients conjugués par exemple.

Résidu

On remarque que la direction de la descente est donné par le résidu :

$$-J = b - H \cdot x$$

Newton

On pourrait bien entendu résoudre par la méthode de Newton, mais cela reviendrait alors à inverser la matrice H directement, ce que l'on cherche précisément à éviter ici.

Généralisation

Soit la matrice A de taille (N, n), où $N \geq n$, et $y \in \mathbb{R}^N$. Considérons le système général :

$$A \cdot x = y$$

à résoudre en $x \in \mathbb{R}^n$. Si on multiplie à gauche par $A^\star$, on obtient :

$$A^\star \cdot A \cdot x = A^\star \cdot y$$

Posons :

$$H = A^\star \cdot A$$
$$b = A^\star \cdot y$$

On est donc amenés à résoudre le système :

$$H \cdot x = b$$

où H est clairement hermitienne et définie positive puisque :

$$x^\star \cdot (A^\star \cdot A) \cdot x = (A \cdot x)^\star \cdot (A \cdot x) \geq 0$$

Cette technique présente deux avantages. Premièrement, si H est inversible, la solution :

$$x = H^{-1} \cdot b = (A^\star \cdot A)^{-1} \cdot A^\star \cdot y$$

minimise la norme $\|A \cdot x - y\|$ même si la solution de $A \cdot x = y$ n'existe pas. Il est même possible d'utiliser cette méthode avec des matrices A strictement hautes. Ensuite, les propriétés de H nous disent qu'une solution du système minimise la fonction φ associée. Il est donc possible de se ramener au problème de minimisation :

$$x \in \arg\min_{z \in \mathbb{R}^n} \left[\frac{1}{2} z^\star \cdot A^\star \cdot A \cdot z - z^\star \cdot A^\star \cdot y \right]$$

que l'on peut de nouveau résoudre itérativement en utilisant par exemple la méthode de la plus grande descente ou les gradients conjugués.

97.4 Espaces de Krylov

Les espaces de Krylov sont des espaces de vecteurs engendrés par les puissances de la matrice A appliquées à un vecteur initial u :

$$\mathrm{Krylov}_m(A, u) = \mathrm{span}\{u, A \cdot u, A^2 \cdot u, ..., A^{m-1} \cdot u\}$$

On a donc des éléments $x \in \mathrm{Krylov}_m(A, u)$ de la forme :

$$x = \left[\sum_{i=0}^{m-1} \alpha_i \cdot A^i\right] \cdot u$$

pour certains $\alpha_i \in \mathbb{K}$. Il s'agit donc de polynômes matriciels appliquées à u.

97.5 Méthode de projection

Soit la matrice A de taille (n, n), le vecteur $b \in \mathbb{R}^n$ et le système $A \cdot x = b$ à résoudre en $x \in \mathbb{R}^n$. On choisit $m \leq n$ beaucoup plus petit que n et les vecteurs $(v_1, ..., v_m)$ de $\mathbb{R}^n$. Nous allons tenter de minimiser l'erreur en x_{k+1} produite par l'itération générique suivante :

$$x_{k+1} = x_k + \sum_{i=1}^{m} v_i \cdot y_i$$

où les $y_i \in \mathbb{R}$. Posons :

$$\begin{aligned} V &= [v_1 \; ... \; v_m] \\ y &= [y_1 \; ... \; y_m]^\star \end{aligned}$$

On peut réécrire l'itération sous la forme :

$$x_{k+1} = x_k + V \cdot y$$

Nous avons vu en résolvant les moindres carrés qu'une condition nécessaire de minimisation était d'imposer l'orthogonalité entre $(A \cdot x - b)$ et les colonnes de A. Nous imposant ici une variante :

$$w_i^\star \cdot (b - A \cdot x_{k+1}) = 0$$

pour certains $w_1, ..., w_m \in \mathbb{R}^n$. En posant :

$$W = [w_1 \; ... \; w_m]$$

cette condition peut s'écrire :

$$W^\star \cdot (b - A \cdot x_{k+1}) = 0$$

Posons :

$$r_k = b - A \cdot x_k$$

On a alors :

$$b - A \cdot x_{k+1} = b - A \cdot x_k - A \cdot V \cdot y = r_k - A \cdot V \cdot y$$

La condition d'orthogonalité devient :

$$W^\star \cdot (r_k - A \cdot V \cdot y) = W^\star \cdot r_k - W^\star \cdot A \cdot V \cdot y = 0$$

Si la matrice $W^\star \cdot A \cdot V$ est inversible, on a :

$$y = (W^\star \cdot A \cdot V)^{-1} \cdot W^\star \cdot r_k$$

Comme V et W sont de taille (n, m), la matrice $W^\star \cdot A \cdot V$ est de taille (m, m), donc beaucoup plus petite que A. Le système correspondant est donc beaucoup plus facile à résoudre.

Orthonormés

Si On choisit $V = W$ et si les vecteurs $v_i = w_i$ sont A-orthonormés, on a simplement :

$$\operatorname*{comp}_{ij} W^\star \cdot A \cdot V = w_i^\star \cdot A \cdot v_j = v_i^\star \cdot A \cdot v_j = \boldsymbol{\delta}_{ij}$$

et :

$$(W^\star \cdot A \cdot V)^{-1} = W^\star \cdot A \cdot V = I$$

Krylov

On peut par exemple choisir des vecteurs de base de $\mathrm{Krylov}_m(A, r_k)$ pour former les colonnes des matrices V et W. On peut aussi utiliser aussi des vecteurs de base de $\mathrm{Krylov}_m(A^\star, r_k)$.

97.6 Minimisation du résidu

On tente de minimiser la norme quadratique du résidu :

$$\mathcal{E}(y) = (r_k - A \cdot V \cdot y)^\star \cdot (r_k - A \cdot V \cdot y)$$

La condition d'annulation de la dérivée par rapport à y nous donne la valeur optimale :

$$y = (V^\star \cdot A^\star \cdot A \cdot V)^{-1} \cdot (V^\star \cdot A^\star \cdot r_k)$$

97.7 Gradients biconjugués

Soit la matrice A de taille (n, n) et les vecteurs $b, c \in \mathbb{K}^n$. Les gradients biconjugués permettent d'obtenir simultanément les deux solutions $x, y \in \mathbb{K}^n$ telles que :

$$\begin{aligned} A \cdot x &= b \quad \text{(système primal)} \\ A^\star \cdot y &= c \quad \text{(système dual)} \end{aligned}$$

Soit (x_k, y_k) l'estimation de (x, y) obtenues à l'itération k et les résidus :

$$\begin{aligned} r_k &= b - A \cdot x_k \\ s_k &= c - A^\star \cdot y_k \end{aligned}$$

L'algorithme est initialisé par $x_0 = y_0 = 0$. On a alors $r_0 = b$ et $s_0 = c$. L'itération est donnée par :

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k \cdot p_k \\ y_{k+1} &= y_k + \gamma_k \cdot q_k \end{aligned}$$

où $p_k, q_k \in \mathbb{K}^n$ et $\alpha_k, \gamma_k \in \mathbb{K}$. Par analogie avec la plus grande descente, les premiers pas s'écrivent $p_0 = r_0$ et $q_0 = s_0$. On adapte ensuite à chaque itération les p_k, q_k par :

$$p_{k+1} = r_{k+1} + \beta_k \cdot p_k$$
$$q_{k+1} = s_{k+1} + \delta_k \cdot q_k$$

où $\beta_k, \delta_k \in \mathbb{K}$. On impose l'orthogonalité des résidus ainsi que l'A-orthogonalité des pas successifs :

$$s_k^\star \cdot r_{k+1} = r_k^\star \cdot s_{k+1} = q_k^\star \cdot A \cdot p_{k+1} = p_k^\star \cdot A^\star \cdot q_{k+1} = 0$$

On voit que :

$$\begin{aligned}
r_{k+1} &= b - A \cdot x_{k+1} \\
&= b - A \cdot x_k - \alpha_k \cdot A \cdot p_k \\
&= r_k - \alpha_k \cdot A \cdot p_k \\
s_{k+1} &= c - A^\star \cdot y_{k+1} \\
&= c - A^\star \cdot y_k - \gamma_k \cdot A^\star \cdot q_k \\
&= s_k - \gamma_k \cdot A^\star \cdot q_k
\end{aligned}$$

Les conditions d'orthogonalité nous donnent donc les valeurs des coefficients

$$\alpha_k = \frac{s_k^\star \cdot r_k}{s_k^\star \cdot A \cdot p_k} \qquad \gamma_k = \frac{r_k^\star \cdot s_k}{r_k^\star \cdot A^\star \cdot q_k}$$
$$\beta_k = -\frac{q_k^\star \cdot A \cdot r_{k+1}}{q_k^\star \cdot A \cdot p_k} \qquad \delta_k = -\frac{p_k^\star \cdot A^\star \cdot s_{k+1}}{p_k^\star \cdot A^\star \cdot q_k}$$

Simplification des coefficients

Il existe un procédé plus rapide pour évaluer les coefficients. En prenant le dual des relations entre les résidus successifs, on obtient $(s_k - s_{k+1})^\star = \overline{\gamma_k} \cdot q_k^\star \cdot A$. Donc :

$$q_k^\star \cdot A \cdot r_{k+1} = \frac{1}{\overline{\gamma_k}} \cdot (s_k - s_{k+1})^\star \cdot r_{k+1} = -\frac{1}{\overline{\gamma_k}} \cdot s_{k+1}^\star \cdot r_{k+1}$$

et :

$$\begin{aligned}
q_k^\star \cdot A \cdot p_k &= q_k^\star \cdot A \cdot r_k - \beta_k \cdot q_k^\star \cdot A \cdot p_{k-1} = q_k^\star \cdot A \cdot r_k \\
&= \frac{1}{\overline{\gamma_k}} \cdot (s_k - s_{k+1})^\star \cdot r_k = \frac{1}{\overline{\gamma_k}} \cdot s_k^\star \cdot r_k
\end{aligned}$$

On en conclut que :

$$\beta_k = \frac{s_{k+1}^\star \cdot r_{k+1}}{s_k^\star \cdot r_k}$$

D'un autre coté, $(r_k - r_{k+1})^\star = \overline{\alpha_k} \cdot p_k^\star \cdot A^\star$. On obtient en suivant un raisonnement analogue :

$$\delta_k = \frac{r_{k+1}^\star \cdot s_{k+1}}{r_k^\star \cdot s_k}$$

Chapitre 98

Optimisation sous contrainte

98.1 Introduction

Ce chapitre traite de la résolution numérique de problèmes d'optimisation sous contraintes, ainsi que de leurs fondements théoriques. Les applications de ces méthodes sont nombreuses : maximisation du profit sous contrainte de ne pas dépasser un certain budget, maximiser la solidité d'une structure sans dépasser un certain poids, ...

98.2 Problème de minimisation sous contrainte

Soit les fonctions convexes et et deux fois continument différentiables $\omega_1, \omega_2, ..., \omega_m : \mathbb{R}^n \mapsto \mathbb{R}$ et la fonction $\omega : \mathbb{R}^n \mapsto \mathbb{R}^m$ définie par :

$$\omega(x) = (\omega_1(x), \omega_2(x), ..., \omega_m(x))$$

pour tout $x \in \mathbb{R}^n$. On considère l'ensemble associé :

$$\Omega = \{x \in \mathbb{R}^n : \omega(x) \le 0\}$$

Par définition de l'ordre partiel sur $\mathbb{R}^m$, tout $x \in \Omega$ doit donc vérifier les m contraintes :

$$\begin{array}{rcl} \omega_1(x) & \le & 0 \\ \omega_2(x) & \le & 0 \\ \dots & & \\ \omega_m(x) & \le & 0 \end{array}$$

Nous considérons le problème général suivant : trouver un $\gamma \in \Omega$ qui minimise la fonction convexe et et deux fois continument différentiable $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ sur Ω :

$$\gamma \in \arg\min_{x \in \Omega} \varphi(x)$$

Nous avons donc un problème de minimisation à n paramètres :

$$x = (x_1, ..., x_n)$$

et m contraintes correspondant aux ω_i. On dit que φ est la fonction objectif et ω la fonction contraignante.

Lien avec la minimisation libre

Il est important de se rendre compte que si ξ est la solution de :

$$\partial\varphi(\xi) = 0$$

on sait que :

$$\xi \in \arg\min_{x\in\mathbb{R}^n} \varphi(x)$$

mais rien ne nous dit que ξ respecte les contraintes exigées ! On ne peut donc pas dans le cas général appliquer la technique classique de la minimisation libre pour résoudre des problèmes de minimisation contraints.

98.3 Convexité

Soit $u, v \in \Omega$ et $s, t \in \mathbb{R}$ tels que $s, t \ge 0$ et $s + t = 1$. On a alors par convexité des ω_i :

$$\omega(s \cdot u + t \cdot v) \le s \cdot \omega(u) + t \cdot \omega(v) \le 0$$

On en conclut que $w = s \cdot u + t \cdot v \in \Omega$. On a donc Convexe$(\Omega) = \Omega$. L'ensemble Ω est convexe.

98.4 Frontière

Soit $x \in \partial\Omega$. Comme $x \in \operatorname{adh}\Omega$, on a $\mathfrak{dist}(x, \Omega) = 0$. On peut donc construire une suite $u_n \in \Omega$ convergeant vers x. La continuité des ω_i et la définition de Ω nous disent que :

$$\omega_i(x) = \lim_{n\to\infty} \omega_i(u_n) \le \limsup_{n\to\infty} \omega_i(u_n) \le 0$$

pour tout $i \in \{1, ..., m\}$. Comme $x \notin \operatorname{int}\Omega$, on a aussi $\mathfrak{dist}(x, \mathbb{R}^n \setminus \Omega) = 0$ et on peut construire une suite $v_n \in \mathbb{R}^n \setminus \Omega$ convergeant vers x. On peut alors trouver un $i \in \{1, ..., m\}$ tel que :

$$\omega_i(x) = \lim_{n\to\infty} \omega_i(v_n) \ge \liminf_{n\to\infty} \omega_i(v_n) \ge 0$$

Ces deux inégalités nous disent que, pour tout $x \in \partial\Omega$, il existe un i tel que $\omega_i(x) = 0$.

98.5 Pénalité

L'idée est de laisser l'objectif inchangé sur Ω mais de l'augmenter suffisamment en dehors pour que le minimum sur Ω soit également le minimum global. On utilise une fonction p convexe et différentiable vérifiant :

$$\begin{aligned} p(x) &= 0 \text{ pour tout } x \in \Omega \\ p(x) &> 0 \text{ pour tout } x \in \mathbb{R}^n \setminus \Omega \end{aligned}$$

et on construit donc un nouvel objectif ψ par :

$$\psi(x) = \varphi(x) + p(x)$$

On dit alors que $p(x)$ est un terme de pénalité. On a :

$$\min_{x\in\Omega} \psi(x) = \min_{x\in\Omega} \varphi(x) = \varphi(\gamma)$$

Choisissons :

$$\chi \in \arg \min_{x \in \mathbb{R}^n \setminus \Omega} \psi(x)$$

Si p est assez élevée en dehors de Ω pour avoir $\psi(\chi) \ge \varphi(\gamma)$, on a :

$$\min_{x \in \mathbb{R}^n} \psi(x) = \min\{\psi(\chi), \varphi(\gamma)\} = \varphi(\gamma)$$

Si de plus $\psi(\chi) > \varphi(\gamma)$, aucun minimum global ne pourra être en dehors de Ω (sinon il serait minimisé par $\varphi(\gamma)$). On aura donc :

$$\arg \min_{x \in \mathbb{R}^n} \psi(x) \subseteq \Omega$$

et il suffit de choisir :

$$\gamma \in \arg \min_{x \in \mathbb{R}^n} \psi(x)$$

pour obtenir une solution à notre problème. De par les propriétés des fonctions convexes, γ sera une solution de $\partial\psi(\gamma) = 0$.

98.6 Lagrangien

Par définition de Ω, on peut trouver au moins un $\omega_i(x) > 0$ pour tout $x \notin \Omega$. Il est donc possible de s'en servir pour former un terme de pénalité. Comme chaque ω_i est convexe, $y_i \cdot \omega_i$ le sera aussi pour tout $y_i \in \mathbb{R}$ tel que $y_i \ge 0$ (si y_i était strictement négatif, on aurait $y_i \cdot \omega_i$ concave). Posons :

$$\mathcal{P} = \{y \in \mathbb{R}^m : y \ge 0\}$$

L'objectif modifié s'écrit alors :

$$\mathfrak{L}(x, y) = \varphi(x) + \sum_{i=1}^{m} y_i \cdot \omega_i(x)$$

pour tout $x \in \mathbb{R}^n$ et tout $y = (y_1, ..., y_m) \in \mathcal{P}$. On dit que les y_i sont les multiplicateurs de Lagrange et que $\mathfrak{L}$ est le lagrangien associé au problème de minimisation. Utilisant la notation du produit scalaire entre vecteurs matriciels, on le note sous la forme schématique :

$$\mathfrak{L}(x, y) = \varphi(x) + y^\star \cdot \omega(x)$$

98.7 Conditions de Kuhn-Tucker

Nous allons à présent essayer de caractériser un choix $y = \lambda$ tel que γ minimise globalement $\mathfrak{L}(x, \lambda)$:

$$\mathfrak{L}(\gamma, \lambda) = \min_{x \in \mathbb{R}^n} \mathfrak{L}(x, \lambda)$$

Puisqu'il s'agit d'un problème de minimisation libre, on sait que la dérivée doit s'annuler en γ :

$$\frac{\partial \mathfrak{L}}{\partial x}(\gamma, \lambda) = \partial\varphi(\gamma) + \lambda^\star \cdot \partial\omega(\gamma) = 0$$

En terme de composantes, on a donc :

$$\partial\varphi(\gamma) + \sum_{i=1}^{m} \lambda_i \cdot \partial\omega_i(\gamma) = 0$$

D'un autre coté, γ doit appartenir à Ω, où le terme de pénalité doit également s'annuler :

$$\lambda^\star \cdot \omega(\gamma) = 0$$

On a donc :

$$\mathfrak{L}(\gamma, \lambda) = \varphi(\gamma) + \lambda^\star \cdot \omega(\gamma) = \varphi(\gamma) + 0 = \varphi(\gamma)$$

En terme de composantes, la condition d'annulation du terme de pénalité s'écrit :

$$\sum_i \lambda_i \cdot \omega_i(\gamma) = 0$$

Mais comme $\lambda_i \ge 0$ et $\omega_i(\gamma) \le 0$, on en déduit que $\lambda_i \cdot \omega_i(\gamma) \le 0$. Que se passerait-il si on pouvait trouver un k tel que $\lambda_k \cdot \omega_k(\gamma) \ < \ 0$? On aurait :

$$\sum_i \lambda_i \cdot \omega_i(\gamma) \le \lambda_k \cdot \omega_k(\gamma) \ < \ 0$$

ce qui est incompatible avec la condition d'annulation du terme de pénalité. On a donc :

$$\lambda_i^\star \cdot \omega_i(\gamma) = 0$$

pour tout $i \in \{1, 2, ..., m\}$. Les conditions :

$$\partial\varphi(\gamma) + \sum_{i=1}^m \lambda_i \cdot \partial\omega_i(\gamma) = 0$$

$$\lambda_i^\star \cdot \omega_i(\gamma) = 0$$

$$\omega(\gamma) \le 0$$

sont appellées conditions de Kuhn-Tucker.

98.8 Point de selle

Supposons que $(\gamma, \lambda) \in \Omega \times \mathcal{P}$ vérifie les conditions de Kuhn-Tucker. Choisissons $x \in \mathbb{R}^n$ et $y \in \mathcal{P}$. L'annulation du gradient ($\partial_x \mathfrak{L}(\gamma, \lambda) = 0$) et la convexité du lagrangien par rapport à la variable x nous assurent que :

$$\mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(x, \lambda)$$

D'un autre coté, on a $\lambda^\star \cdot \omega(\gamma) = 0$ et $\mathfrak{L}(\gamma, \lambda) = \varphi(\gamma)$. On sait aussi que $y \ge 0$ et que $\omega(\gamma) \le 0$. Donc $y^\star \cdot \omega(\gamma) \le 0$ et :

$$\mathfrak{L}(\gamma, y) = \varphi(\gamma) + y^\star \cdot \omega(\gamma) \le \varphi(\gamma) = \mathfrak{L}(\gamma, \lambda)$$

On voit que λ maximise le lagrangien par rapport à la variable y. Ces deux propriétés extrémales nous montrent que (γ, λ) est un point de selle :

$$\mathfrak{L}(\gamma, y) \le \mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(x, \lambda)$$

Réciproque

Nous venons de voir que les conditions de Kuhn-Tucker remplies sur $\Omega \times \mathcal{P}$ nous offraient un point de selle sur $\mathbb{R}^n \times \mathcal{P}$. Nous allons voir que la réciproque est également vraie. Supposons que $(\gamma, \lambda) \in \mathbb{R}^n \times \mathcal{P}$ soit un point de selle du lagrangien :

$$\mathfrak{L}(\gamma, y) \le \mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(x, \lambda)$$

pour tout $(x, y) \in \mathbb{R}^n \times \mathcal{P}$. On a :

$$\mathfrak{L}(\gamma, y) - \mathfrak{L}(\gamma, \lambda) = (y - \lambda)^\star \cdot \omega(\gamma) \le 0$$

relation qui doit être valable pour tout $y \ge 0$. Fixons $i \in \{1, 2, ..., m\}$ et choisissons :

$$y_k = \begin{cases} \lambda_k & \text{si } k \neq i \\ 0 & \text{si } k = i \end{cases}$$

on en déduit que $(y - \lambda)^\star \cdot \omega(\gamma) = -\lambda_i \cdot \omega_i(\gamma) \le 0$, c'est-à-dire :

$$\lambda_i \cdot \omega_i(\gamma) \ge 0$$

Considérons à présent le choix :

$$y_k = \begin{cases} \lambda_k & \text{si } k \neq i \\ 2\lambda_i & \text{si } k = i \end{cases}$$

On en déduit alors que :

$$(y - \lambda)^\star \cdot \omega(\gamma) = \lambda_i \cdot \omega_i(\gamma) \le 0$$

On conclut de ces deux inégalités que :

$$\lambda_i \cdot \omega_i(\gamma) = 0$$

Enfin, si nous prenons :

$$y_k = \begin{cases} \lambda_k & \text{si } k \neq i \\ \lambda_i + 1 & \text{si } k = i \end{cases}$$

on voit que :

$$(y - \lambda)^\star \cdot \omega(\gamma) = 1 \cdot \omega_i(\gamma) = \omega_i(\gamma) \le 0$$

ce qui nous montre que $\gamma \in \Omega$.
De plus, comme :

$$\gamma \in \arg\min_{x \in \mathbb{R}^n} \mathfrak{L}(x, \lambda)$$

les propriétés des fonctions dérivables nous imposent que $\partial_x \mathfrak{L}(\gamma, \lambda) = 0$.

98.9 Solution

Supposons que (γ, λ) soit un point de selle du lagrangien. Choisissons $x \in \Omega$. On a $\omega(x) \le 0$ et $\lambda^\star \cdot \omega(x) \le 0$. On en déduit que :

$$\mathfrak{L}(x, \lambda) = \varphi(x) + \lambda^\star \cdot \omega(x) \le \varphi(x)$$

On a donc bien :

$$\varphi(\gamma) = \mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(x, \lambda) \le \varphi(x)$$

ce qui prouve que γ minimise φ sur Ω. La solution du problème de point de selle du lagrangien contient donc la solution de notre problème de minimisation sous contrainte.

Généralisation

Remarquons que nous n'avons pas utilisé la convexité de φ ou de ω pour démontrer que γ est bien la solution de notre problème. On peut donc appliquer la technique du point de selle du lagrangien à une classe de fonctions beaucoup plus générale.

Attention, si φ et/ou ω ne sont plus supposées convexes, le point de selle impliquera toujours les conditions de Kuhn-Tucker mais l'inverse ne sera plus vrai. En pratique, on essaiera malgré tout de trouver une solution aux conditions de Kuhn-Tucker, et on déterminera a posteriori :

— si le minimum trouvé est bien un minimum local (test de la Hessienne définie positive au γ obtenu)

— si ce minimum local est aussi un minimum sur Ω

98.10 Problème de maximisation sous contrainte

Soit les fonctions *concaves* et deux fois continument différentiables $\vartheta_1, \vartheta_2, ..., \vartheta_m : \mathbb{R}^n \mapsto \mathbb{R}$ et la fonction $\vartheta : \mathbb{R}^n \mapsto \mathbb{R}^m$ définie par :

$$\vartheta(x) = (\vartheta_1(x), \vartheta_2(x), ..., \vartheta_m(x))$$

pour tout $x \in \mathbb{R}^n$. On considère l'ensemble associé :

$$\Theta = \{x \in \mathbb{R}^n : \vartheta(x) \ge 0\}$$

Nous considérons le problème général suivant : trouver un $\gamma \in \Theta$ qui maximise la fonction concave et et deux fois continument différentiable $\psi : \mathbb{R}^n \mapsto \mathbb{R}$ sur Θ :

$$\gamma \in \arg\max_{x \in \Theta} \psi(x)$$

On note que ce problème est équivalent au suivant :

$$\gamma \in \arg\min_{x \in \Omega} (-\psi(x))$$

où :

$$\Omega = \{x \in \mathbb{R}^n : -\vartheta(x) \le 0\}$$

98.10.1 Lagrangien

On sait que maximiser ψ revient à minimiser $-\psi$. Comme $-\psi$ et $-\vartheta$ sont convexes, on peut utiliser le lagrangien défini par :

$$\mathfrak{L}_{\min}(x, y) = \Big[-\psi(x) \Big] + y^\star \cdot \Big[-\vartheta(x) \Big] = -\psi(x) - y^\star \cdot \vartheta(x)$$

pour tout $(x, y) \in \mathbb{R}^n \times \mathcal{P}$. La solution (γ, λ) vérifie alors :

$$\mathfrak{L}_{\min}(\gamma, y) \le \mathfrak{L}_{\min}(\gamma, \lambda) \le \mathfrak{L}_{\min}(x, \lambda)$$

On voit que la fonction opposée :

$$\mathfrak{L}_{\max}(x, y) = -\mathfrak{L}_{\min}(x, y) = \psi(x) + y^\star \cdot \vartheta(x)$$

vérifie les conditions du point de selle inversées :

$$\mathfrak{L}_{\max}(x, \lambda) \le \mathfrak{L}_{\max}(\gamma, \lambda) \le \mathfrak{L}_{\max}(\gamma, y)$$

Il s'agit du lagrangien natif du problème de maximisation. Il est maximisé par rapport à x et minimisé par rapport aux multiplicateurs de Lagrange.

98.10.2 Kuhn-Tucker

Les conditions de Kuhn-Tucker associées à $\mathfrak{L}_{\min}$ s'écrivent :

$$-\partial\psi(\gamma) - \sum_{i=1}^m \lambda_i \cdot \partial\vartheta_i(\gamma) = 0$$

$$-\lambda_i^\star \cdot \vartheta_i(\gamma) = 0$$

$$-\vartheta(\gamma) \le 0$$

Elles sont équivalentes aux conditions analogues associées à $\mathfrak{L}_{\max}$:

$$\partial\psi(\gamma) + \sum_{i=1}^m \lambda_i \cdot \partial\vartheta_i(\gamma) = 0$$

$$\lambda_i^\star \cdot \vartheta_i(\gamma) = 0$$

$$\vartheta(\gamma) \ge 0$$

98.11 Contraintes d'égalité

Soit les fonctions deux fois continument différentiables $\varrho_1, \varrho_2, ..., \varrho_m : \mathbb{R}^n \mapsto \mathbb{R}$ et la fonction $\varrho : \mathbb{R}^n \mapsto \mathbb{R}^m$ définie par :

$$\varrho(x) = (\varrho_1(x), \varrho_2(x), ..., \varrho_m(x))$$

pour tout $x \in \mathbb{R}^n$. On considère l'ensemble associé :

$$\Phi = \{x \in \mathbb{R}^n : \varrho(x) = 0\}$$

Nous considérons le problème général suivant : trouver un $\gamma \in \Phi$ qui minimise la fonction deux fois continument différentiable $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ sur Φ :

$$\gamma \in \arg\min_{x \in \Phi} \varphi(x)$$

98.11.1 Lagrangien

On peut réécrire l'ensemble Φ sous la forme :

$$\Phi = \{x \in \mathbb{R}^n : \varrho(x) \le 0, \ \varrho(x) \ge 0\}$$

On introduit donc deux séries de multiplicateurs et le lagrangien défini par :

$$L(x, y, z) = \varphi(x) + y^\star \cdot \varrho(x) + z^\star \cdot (-\varrho(x))$$

pour tout $(x, y, z) \in \mathbb{R}^n \times \mathcal{P} \times \mathcal{P}$. On constate qu'il ne dépend que de x et de $y - z$:

$$L(x, y, z) = \varphi(x) + (y - z)^\star \cdot \varrho(x)$$

On voit que $u = y - z$ est libre d'aller où il veut dans $\mathbb{R}^m$. On pose :

$$\mathfrak{L}(x, u) = \varphi(x) + u^\star \cdot \varrho(x)$$

pour tout $(x, u) \in \mathbb{R}^n \times \mathbb{R}^m$.

98.11.2 Kuhn-Tucker

Comme on veut que γ minimise $\mathfrak{L}$ par rapport à la variable x sur $\mathbb{R}^n$, on impose $\partial_x \mathfrak{L}(\gamma, \lambda) = 0$. On doit avoir aussi $\varrho(x) = 0$. Mais comme $\partial_u \mathfrak{L}(x, u) = \varrho(x)$, on se retrouve finalement avec les conditions nécessaires de Kuhn-Tucker :

$$\partial_x \mathfrak{L}(\gamma, \lambda) = 0$$

$$\partial_u \mathfrak{L}(\gamma, \lambda) = 0$$

98.11.3 Point de selle

Nous allons montrer que tout point de selle est solution du problème de minimisation. Soit (γ, λ) tel que :

$$\mathfrak{L}(\gamma, u) \le \mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(x, \lambda)$$

pour tout $(x, u) \in \mathbb{R}^n \times \mathbb{R}^m$. On en déduit que :

$$\mathfrak{L}(\gamma, u) - \mathfrak{L}(\gamma, \lambda) \le 0$$

pour tout $u \in \mathbb{R}^m$, c'est-à-dire :

$$(u - \lambda)^\star \cdot \varrho(\gamma) \le 0$$

Considérons la base canonique $(\mathfrak{c}_1, ..., \mathfrak{c}_m)$ de $\mathbb{R}^m$. Si on prend $u = \lambda + \mathfrak{c}_i$, on obtient la condition

$$(u - \lambda)^\star \cdot \varrho(\gamma) = \varrho_i(\gamma) \le 0$$

Mais comme u n'est pas forcément positif, on peut aussi prendre $u = \lambda - \mathfrak{c}_i$. On obtient alors la condition :

$$(u-\lambda)^\star \cdot \varrho(\gamma) = -\varrho_i(\gamma) \le 0$$

On déduit de ces deux inégalités que $\varrho(\gamma) = 0$, c'est à dire $\gamma \in \Phi$.
A présent, soit $x \in \Phi$. Comme $\varrho(\gamma) = \varrho(x) = 0$, on a :

$$\varphi(\gamma) = \mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(x, \lambda) = \varphi(x)$$

Le γ issu du point de selle est donc bien la solution de notre problème de minimisation contraint.

98.12 Récapitulation

Considérons l'ensemble :

$$\Omega = \{x \in \mathbb{R}^n : \omega(x) \le 0,\ \vartheta(x) \ge 0,\ \varrho(x) = 0\}$$

98.12.1 Minimisation

Si on veut minimiser φ sur Ω, on utilisera le lagrangien défini par :

$$\mathfrak{L}(x, y, z, u) = \varphi(x) + y^\star \cdot \omega(x) - z^\star \cdot \vartheta(x) + u^\star \cdot \varrho(x)$$

pour tout (x, y, z, u) tels que $y, z \ge 0$. Le point de selle $(\gamma, \lambda, \mu, \nu)$ vérifiera :

$$\mathfrak{L}(\gamma, y, z, u) \le \mathfrak{L}(\gamma, \lambda, \mu, \nu) \le \mathfrak{L}(x, \lambda, \mu, \nu)$$

Les conditions de Kuhn-Tucker s'écriront :

$$\partial\varphi(\gamma) + \sum_i [\lambda_i \cdot \partial\omega_i(x) - \mu_i \cdot \partial\vartheta_i(x) + \nu_i \cdot \partial\varrho_i(x)] = 0$$

$$\lambda_i \cdot \omega_i(x) = 0$$

$$\mu_i \cdot \vartheta_i(x) = 0$$

$$\omega_i(x) \le 0$$

$$\vartheta_i(x) \ge 0$$

$$\varrho_i(x) = 0$$

98.12.2 Maximisation

Si on veut maximiser ψ sur Ω, on utilisera le lagrangien défini par :

$$\mathfrak{L}(x, y, z, u) = \psi(x) - y^\star \cdot \omega(x) + z^\star \cdot \vartheta(x) + u^\star \cdot \varrho(x)$$

pour tout (x, y, z, u) tels que $y, z \ge 0$. Le point de selle $(\gamma, \lambda, \mu, \nu)$ vérifiera :

$$\mathfrak{L}(x, \lambda, \mu, \nu) \le \mathfrak{L}(\gamma, \lambda, \mu, \nu) \le \mathfrak{L}(\gamma, y, z, u)$$

Les conditions de Kuhn-Tucker s'écriront :

$$\partial\psi(\gamma) + \sum_i [-\lambda_i \cdot \partial\omega_i(x) + \mu_i \cdot \partial\vartheta_i(x) + \nu_i \cdot \partial\varrho_i(x)] = 0$$

$$\lambda_i \cdot \omega_i(x) = 0$$

$$\mu_i \cdot \vartheta_i(x) = 0$$

$$\omega_i(x) \le 0$$

$$\vartheta_i(x) \ge 0$$

$$\varrho_i(x) = 0$$

98.12.3 Equivalence

On exprime parfois ces problèmes en utilisant des contraintes équivalentes. Si on pose $s = -\omega(x)$, on a $s \ge 0$. De même, on pose $t = \vartheta(x) \ge 0$. On définit alors Γ comme l'ensemble des (x, s, t) vérifiant les conditions :

$$\begin{aligned} s, t &\ge 0 \\ s + \omega(x) &= 0 \\ t - \vartheta(x) &= 0 \\ \varrho(x) &= 0 \end{aligned}$$

Et on minimise $\varphi(x)$ (ou on maximise $\psi(x)$) sur les $(x, s, t) \in \Gamma$.

98.13 Découplage

Minimisation

Soit le problème de minimisation :

$$\gamma \in \arg\min_{x\in\Omega} \varphi(x)$$

où :

$$\Omega^+ = \{x \in \mathbb{K}^n : \omega(x) \le 0,\ x \ge 0\}$$

La condition $x \ge 0$ est équivalente à $-x \le 0$. On définit le lagrangien :

$$L(x, y, z) = \varphi(x) + y^\star \cdot \omega(x) - z^\star \cdot x$$

La solution de notre problème vérifie les conditions du point de selle :

$$L(\gamma, y, z) \le L(\gamma, \lambda, \mu) \le L(x, \lambda, \mu)$$

pour tout $x \in \mathbb{R}^n$ et $y, z \ge 0$. Si $x \ge 0$, on a :

$$-z^\star \cdot x \le 0$$

par positivité de z. On a alors :

$$L(x, y, z) = \varphi(x) + y^\star \cdot \omega(x) - z^\star \cdot x \le \varphi(x) + y^\star \cdot \omega(x) = L(x, y, 0)$$

On en conclut qu'un choix permettant de maximiser L par rapport à son troisième argument est $\mu = 0$. Introduisons le lagrangien modifié :

$$\mathfrak{L}(x, y) = L(x, y, 0) = \varphi(x) + y^\star \cdot \omega(x)$$

Le problème du point de selle est équivalent à trouver $(\gamma, \lambda) \ge 0$ tel que :

$$\mathfrak{L}(\gamma, y) \le \mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(x, \lambda)$$

pour tout $(x, y) \ge 0$. La différence est qu'on ne fait plus apparaître explicitement la contrainte de positivité de x dans le lagrangien.

Maximisation

Soit le problème de minimisation :

$$\gamma \in \arg\max_{x \in \Theta} \psi(x)$$

où :

$$\Theta = \{x \in \mathbb{K}^n : \vartheta(x) \ge 0,\ x \ge 0\}$$

On pose le lagrangien :

$$L(x, y, z) = \psi(x) + y^\star \cdot \omega(x) + z^\star \cdot x$$

La solution de notre problème vérifie les conditions du point de selle inversé :

$$L(x, \lambda, \mu) \le L(\gamma, \lambda, \mu) \le L(\gamma, y, z)$$

pour tout $x \in \mathbb{R}^n$ et $y, z \ge 0$. Si $x \ge 0$, on a :

$$z^\star \cdot x \ge 0$$

par positivité de z. On a alors :

$$L(x, y, z) \ge L(x, y, 0)$$

On en conclut qu'un choix permettant de minimiser L par rapport à son troisième argument est $\mu = 0$. Introduisons le lagrangien modifié :

$$\mathfrak{L}(x, y) = L(x, y, 0) = \psi(x) + y^\star \cdot \vartheta(x)$$

Le problème du point de selle est équivalent à trouver $(\gamma, \lambda) \ge 0$ tel que :

$$\mathfrak{L}(x, \lambda) \le \mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(\gamma, y)$$

pour tout $(x, y) \ge 0$.

98.14 Dualité en optimisation linéaire

98.14.1 Problème primal

Soit une matrice réelle A de taille (m, n), l'ensemble :

$$\Theta = \{x \in \mathbb{R}^n : A \cdot x \le b,\ x \ge 0\}$$

et le vecteur $c \in \mathbb{R}^n$. Nous nous intéressons au problème de maximisation linéaire sous contrainte :

$$\gamma \in \arg\max_{x \in \Theta} \left[c^\star \cdot x\right]$$

98.14.2 Lagrangien

La contrainte $A \cdot x \le b$ est équivalente à :

$$b - A \cdot x \ge 0$$

On introduit donc le lagrangien :

$$\begin{aligned}\mathfrak{L}(x, y) &= c^\star \cdot x + y^\star \cdot (b - A \cdot x) \\ &= c^\star \cdot x + y^\star \cdot b - y^\star \cdot A \cdot x\end{aligned}$$

défini pour tout $(x, y) \ge 0$.

98.14.3 Dualité

Comme on travaille dans les réels, on a par symétrie du produit scalaire :

$$\begin{aligned}c^\star \cdot x &= x^\star \cdot c \\ y^\star \cdot b &= b^\star \cdot y \\ y^\star \cdot A \cdot x &= (A \cdot x)^\star \cdot y = x^\star \cdot A^\star \cdot y\end{aligned}$$

Le lagrangien du problème primal peut donc se réécrire :

$$\begin{aligned}\mathfrak{L}(x, y) &= x^\star \cdot c + b^\star \cdot y - x^\star \cdot A^\star \cdot y \\ &= b^\star \cdot y + x^\star \cdot (c - A^\star \cdot y)\end{aligned}$$

On voit que la fonction duale définie par :

$$\mathfrak{L}^\star(y, x) = \mathfrak{L}(x, y) = b^\star \cdot y + x^\star \cdot (c - A^\star \cdot y)$$

pour tout $(y, x) \ge 0$ peut également être considérée comme un lagrangien formé à partir de l'objectif $y \mapsto b^\star \cdot y$ et d'une des deux contraintes suivantes :

$$? \begin{cases} c - A^\star \cdot y \le 0 \\ c - A^\star \cdot y \ge 0 \end{cases}$$

Résoudre le problème primal revient à trouver un point $(\gamma, \lambda) \ge 0$ vérifiant :

$$\mathfrak{L}(x, \lambda) \le \mathfrak{L}(\gamma, \lambda) \le \mathfrak{L}(\gamma, y)$$

pour tout $x, y \ge 0$. En utilisant la définition de $\mathfrak{L}^\star$, ces conditions se réécrivent :

$$\mathfrak{L}^\star(\lambda, x) \le \mathfrak{L}^\star(\lambda, \gamma) \le \mathfrak{L}^\star(y, \gamma)$$

Le couple (λ, γ) est donc solution d'un problème de minimisation en y. Dans un problème de minimisation, les contraintes associées aux multiplicateurs de lagrange doivent être négatives. On a donc $c - A^\star \cdot y \le 0$, autrement dit :

$$A^\star \cdot y \ge c$$

Le problème primal est donc équivalent au problème dual suivant.

98.14.4 Problème dual

Trouver :

$$\lambda \in \arg\min_{y \in \Omega} (b^\star \cdot y)$$

où :

$$\Omega = \{y \in \mathbb{R}^m : A^\star \cdot y \ge c,\ y \ge 0\}$$

98.14.5 Valeurs extrémales

Les conditions de Khun-Tucker des problèmes primal et dual impliquent :

$$\lambda^\star \cdot (b - A \cdot \gamma) = 0$$

$$\gamma^\star \cdot (c - A^\star \cdot \lambda) = 0$$

Les valeurs des lagrangiens aux points de selle s'écrivent donc :

$$\mathfrak{L}(\gamma, \lambda) = c^\star \cdot \gamma + \lambda^\star \cdot (b - A \cdot \gamma) = c^\star \cdot \gamma$$

$$\mathfrak{L}^\star(\lambda, \gamma) = b^\star \cdot \lambda + \gamma^\star \cdot (c - A^\star \cdot \lambda) = b^\star \cdot \lambda$$

La définition du lagrangien dual implique que :

$$\mathfrak{L}(\gamma, \lambda) = \mathfrak{L}^\star(\lambda, \gamma)$$

c'est-à-dire :

$$c^\star \cdot \gamma = b^\star \cdot \lambda$$

Le maximum de $x \mapsto c^\star \cdot x$ sur Θ est donc égal au minimum de $y \mapsto b^\star \cdot y$ sur Ω. Dans la suite, on note :

$$V = \max_{x \in \Theta} (c^\star \cdot x) = \min_{y \in \Omega} (b^\star \cdot y)$$

98.14.6 Bornes

Par définition des maxima et minima, on a :

$$c^\star \cdot x \le V \le b^\star \cdot y$$

pour tout $x \in \Theta$ et $y \in \Omega$. Si les deux valeurs $c^\star \cdot x$ et $b^\star \cdot y$ sont proches, les éléments x et y sont presque optimaux.

98.14.7 Attention

Ne pas confondre la notion de dualité max − min avec la dualité au sens des applications linéaires.

98.15 Minimisation de la norme sous contraintes linéaires

Soit la matrice $A \in \mathbb{M}(\mathbb{R}, m, n)$, le vecteur colonne $b \in \mathbb{M}(\mathbb{R}, m, 1)$ et l'ensemble :

$$\Phi = \{x \in \mathbb{M}(\mathbb{R}, n, 1) : A \cdot x = b\}$$

On veut trouver le $\gamma \in \Phi$ qui minimise la norme usuelle $\|x\| = \sqrt{x^\star \cdot x}$ sur Φ. Comme la fonction $\|x\| \mapsto \|x\|^2 /2$ est une fonction monotone strictement croissante sur $\|x\| \in \mathbb{R}^+$, cela revient à minimiser :

$$\gamma = \arg \min_{x \in \Phi} \left(\frac{1}{2} \cdot x^\star \cdot x \right)$$

Afin de résoudre ce problème, on introduit le lagrangien :

$$\mathfrak{L}(x, u) = \frac{1}{2} \cdot x^\star \cdot x + u^\star \cdot (b - A \cdot x)$$

On impose les conditions de Kuhn-Tucker :

$$\partial_x \mathfrak{L}(\gamma, \lambda) = \gamma - A^\star \cdot \lambda = 0$$

$$\partial_u \mathfrak{L}(\gamma, \lambda) = b - A \cdot \gamma = 0$$

On en déduit que $\gamma = A^\star \cdot \lambda$ et que :

$$A \cdot A^\star \cdot \lambda = b$$

Si $A \cdot A^\star$ est inversible, on a donc :

$$\lambda = (A \cdot A^\star)^{-1} \cdot b$$

et :

$$\gamma = A^\star \cdot (A \cdot A^\star)^{-1} \cdot b$$

Chapitre 99

Valeurs propres

99.1 Minimisation d'une forme quadratique avec norme contrainte

Soit un réel $R \ne 0$, la matrice $H \in \mathbb{M}(\mathbb{R}, m, n)$ hermitienne et définie positive, et l'ensemble :

$$\Omega = \{x \in \mathbb{K}^n : \|x\| = \sqrt{x^\star \cdot x} = R\}$$

On peut reformuler Ω de manière équivalente par :

$$\Omega = \{x \in \mathbb{K}^n : x^\star \cdot x = R^2\}$$

Soit l'objectif $\varphi : \mathbb{K}^n \mapsto \mathbb{K}$ défini par :

$$\varphi(x) = x^\star \cdot H \cdot x$$

pour tout $x \in \mathbb{K}^n$. On veut trouver le $x \in \Omega$ qui minimise φ sur Ω. Définissons le lagrangien :

$$\mathfrak{L}(x, \lambda) = x^\star \cdot H \cdot x + \lambda \cdot (R^2 - x^\star \cdot x)$$

où $\lambda \in \mathbb{R}$. On impose les conditions de Kuhn-Tucker :

$$\partial_x \mathfrak{L}(x, \lambda) = 2H \cdot x - 2\lambda \cdot x = 0$$
$$\partial_\lambda \mathfrak{L}(x, \lambda) = R^2 - x^\star \cdot x = 0$$

Toute solution optimale $x \in \Omega$ vérifie donc l'équation :

$$H \cdot x = \lambda \cdot x$$

pour un certain $\lambda \in \mathbb{R}$. Nous allons généraliser cette propriété aux applications linéaires.

Norme unité

Un cas particulier souvent utilisé est celui où $R = 1$.

99.2 Application linéaire

Soit l'espace vectoriel E sur $\mathbb{K}$ et une application linéaire $A : E \mapsto E$. Si le couple $(u, \lambda) \in (E \setminus \{0\}) \times \mathbb{K}$ vérifie :

$$A(u) = \lambda \cdot u$$

On dit que u est vecteur propre de A correspondant à la valeur propre λ.

Rapport de Rayleigh

En prenant le produit scalaire de u avec la relation $A(u) = \lambda \cdot u$, on obtient :

$$\langle u \mid A(u) \rangle = \lambda \cdot \langle u \mid u \rangle$$

d'où l'on tire :

$$\lambda = \frac{\langle u \mid A(u) \rangle}{\langle u \mid u \rangle}$$

Une telle expression est appelée rapport de Rayleigh. Par extension, on définit :

$$R(v) = \frac{\langle v \mid A(v) \rangle}{\langle v \mid v \rangle}$$

pou tout vecteur non nul $v \in E$.

Définie positive

Si A est définie positive, toute valeur propre $\lambda = R(u) \geq 0$ est un réel positif.

Formulation équivalente

On voit qu'il est équivalent de chercher un scalaire λ et un vecteur $u \neq 0$ tel que :

$$(A - \lambda \cdot \mathrm{Id})(u) = A(u) - \lambda \cdot u = 0$$

99.3 Opérateurs auto-adjoints

Si l'application linéaire A est auto-adjointe ($A^\star = A$), ses valeurs et vecteurs propres possèdent d'importantes propriétés. Soit $u \in E$ et $\lambda \in \mathbb{C}$ tels que $A(u) = \lambda \cdot u$ et $\langle u \mid u \rangle = 1$. On a alors :

$$\lambda = \langle u \mid A(u) \rangle = \langle A(u) \mid u \rangle = \operatorname{conj} \langle u \mid A(u) \rangle = \bar{\lambda}$$

La valeur propre doit donc être réelle.

Orthonormalité

Soit la suite de vecteurs propres $u_i \in \Omega$ et de valeurs propres $\lambda_i \in \mathbb{R}$. On a :

$$\langle u_i \mid A(u_j) \rangle = \lambda_j \cdot \langle u_i \mid u_j \rangle$$

ainsi que :

$$\langle u_i \mid A(u_j) \rangle = \langle A(u_i) \mid u_j \rangle = \lambda_i \cdot \langle u_i \mid u_j \rangle$$

En soustrayant ces deux équations, on obtient :

$$(\lambda_j - \lambda_i) \cdot \langle u_i \mid u_j \rangle = 0$$

Le produit scalaire doit s'annuler lorsque les valeurs propres diffèrent. Si toutes les valeurs propres sont distinctes, les vecteurs propres forment une suite orthonormée :

$$\langle u_i \mid u_j \rangle = \boldsymbol{\delta}_{ij}$$

Orthogonalité par rapport à l'application linéaire

Si les vecteurs propres $u_i \in \Omega$ sont orthonormés, on a :

$$\langle u_i \mid A(u_j) \rangle = \lambda_j \cdot \langle u_i \mid u_j \rangle = \lambda_j \cdot \boldsymbol{\delta}_{ij}$$

99.4 Représentation tensorielle

Supposons que la suite de vecteurs propres $(u_1, ..., u_n)$ soit orthonormée. Pour tout $x \in \mathrm{span}\{u_1, ..., u_n\}$, on a alors :

$$x = \sum_i \langle u_i \mid x \rangle \cdot u_i$$

et :

$$\begin{aligned} A(x) &= \sum_i \langle u_i \mid x \rangle \cdot A(u_i) \\ &= \sum_i \lambda_i \cdot \langle u_i \mid x \rangle \cdot u_i \end{aligned}$$

On peut donc représenter A sur $\mathrm{span}\{u_1, ..., u_n\}$ par le tenseur associé :

$$\mathcal{A} = \sum_i \lambda_i \cdot u_i \otimes u_i$$

de sorte que :

$$A(x) = \mathcal{A} \cdot x = \langle \mathcal{A} \odot x \rangle_1$$

99.5 Inverse

Supposons que les vecteurs propres u_i forment une base orthonormée de E, et que les valeurs propres correspondantes λ_i soient non nulles. Si $x, y \in E$ sont tels que $A(x) = \mathcal{A} \cdot x = y$, on a :

$$y = \sum_i \langle u_i \mid y \rangle \cdot u_i = \mathcal{A} \cdot x = \sum_i \lambda_i \cdot u_i \cdot \langle u_i \mid x \rangle$$

On en déduit en comparant les coefficients de chaque vecteur u_i que $\lambda_i \cdot \langle u_i \mid x \rangle = \langle u_i \mid y \rangle$, d'où :

$$\langle u_i \mid x \rangle = \frac{1}{\lambda_i} \cdot \langle u_i \mid y \rangle$$

Mais ces produits scalaires sont les coordonnées de x par rapport aux u_i :

$$x = \sum_i \langle u_i \mid x \rangle \cdot u_i = \sum_i \frac{1}{\lambda_i} \cdot \langle u_i \mid y \rangle \cdot u_i$$

Donc, si on pose :

$$\mathcal{A}^{-1} = \sum_i \frac{1}{\lambda_i} \cdot u_i \otimes u_i$$

on a :

$$x = \mathcal{A}^{-1} \cdot y$$

99.6 Représentation matricielle

On dit que $\lambda \in \mathbb{K}$ est la valeur propre de la matrice $A \in \mathbb{M}(\mathbb{K}, n, n)$ correspondant au vecteur propre non nul $u \in \mathbb{K}^n$ si ils sont valeurs et vecteurs propres de l'application linéaire sous-jacente :

$$A \cdot u = \lambda \cdot u$$

Formulation équivalente

On voit qu'il est équivalent de chercher un scalaire λ et un vecteur $u \neq 0$ tel que :

$$(A - \lambda \cdot I) \cdot u = 0$$

Rapport de Rayleigh

On peut évaluer λ à partir de u en multipliant l'équation ci-dessus à gauche par $u^\star$. On obtient alors :

$$u^\star \cdot A \cdot u = \lambda \cdot u^\star \cdot u$$

et le rapport de Rayleigh en u s'en suit :

$$\lambda = \frac{u^\star \cdot A \cdot u}{u^\star \cdot u} = R(u)$$

99.7 Invariance

Soit la valeur propre λ de A et le vecteur propre correspondant $u \neq 0$. On a :

$$A \cdot u = \lambda \cdot u$$

Multiplions cette équation à gauche par une matrice inversible Q :

$$Q \cdot A \cdot u = \lambda \cdot Q \cdot u$$

Posons à présent $v = Q \cdot u$. On a alors $u = Q^{-1} \cdot v$ et :

$$Q \cdot A \cdot Q^{-1} \cdot v = \lambda \cdot v$$

On constate aussi que $v \neq 0$, car sinon :

$$0 \neq u = Q^{-1} \cdot v = Q^{-1} \cdot 0 = 0$$

ce qui est impossible. On en conclut que λ est aussi une valeur propre de la matrice :

$$B = Q \cdot A \cdot Q^{-1}$$

On dit que les valeurs propres sont des invariants sous transformation linéaire réversible.

99.8 Matrices triangulaires

Nous allons démontrer par récurrence sur la taille n d'une matrice triangulaire supérieure T que les composantes diagonales sont des valeurs propres de T.

— Si $n = 1$, soit les uniques composantes $\lambda = \operatorname{comp}_{11}(T)$ et $\mu = \operatorname{comp}_1(u)$. On a :

$$T \cdot u = [\lambda] \cdot [\mu] = \lambda \cdot [1] \cdot [\mu] = \lambda \cdot u$$

ce qui montre que la composante diagonale λ est une valeur propre de T.

— Supposons que le résultat soit vrai pour $n - 1$. Partitionnons à part la dernière ligne et la dernière colonne d'une matrice $T \in \mathbb{M}(\mathbb{K}, n, n)$:

$$T = \begin{bmatrix} T^{(n-1)} & z \\ 0 & \lambda_n \end{bmatrix}$$

Donc, $T^{(n-1)}$ est une matrice triangulaire de la forme :

$$T^{(n-1)} = \begin{bmatrix} \lambda_1 & \dots & \dots \\ 0 & \ddots & \dots \\ 0 & 0 & \lambda_{n-1} \end{bmatrix}$$

où les composantes diagonales $\lambda_1, ..., \lambda_{n-1}$ vérifient :

$$T^{(n-1)} \cdot u_i^{(n-1)} = \lambda_i \cdot u_i^{(n-1)}$$

pour certains vecteurs propres $u_i^{(n-1)} \neq 0$. Nous allons chercher les vecteurs propres $u \in \mathbb{K}^n$ de T sous la forme :

$$u = \begin{bmatrix} v \\ \mu \end{bmatrix}$$

où $v \in \mathbb{K}^{n-1}$ et $\mu \in \mathbb{C}$. Si on choisit $v = u_i^{(n-1)}$ pour $i \in \{1, 2, ..., n - 1\}$ et $\mu = 0$, on a :

$$\begin{aligned} T \cdot u &= \begin{bmatrix} T^{(n-1)} & z \\ 0 & \lambda_n \end{bmatrix} \cdot \begin{bmatrix} u_i^{(n-1)} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} T^{(n-1)} \cdot u_i^{(n-1)} \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda_i \cdot u_i^{(n-1)} \\ 0 \end{bmatrix} \\ &= \lambda_i \cdot \begin{bmatrix} u_i \\ 0 \end{bmatrix} = \lambda_i \cdot u \end{aligned}$$

Les valeurs propres de $T^{(n-1)}$ sont donc également valeurs propres de T. Il nous reste à examiner le cas de λ_n. Posons :

$$A = \begin{bmatrix} \lambda_1 - \lambda_n & \dots & \dots & \dots \\ 0 & \ddots & \dots & \dots \\ 0 & 0 & \lambda_{n-1} - \lambda_n & \dots \end{bmatrix}$$

On doit donc trouver un vecteur $u \in \mathbb{K}^n$ non nul tel que :

$$(T - \lambda_n \cdot I) \cdot u = \begin{bmatrix} A \\ 0 \end{bmatrix} \cdot u = \begin{bmatrix} A \cdot u \\ 0 \end{bmatrix} = 0$$

Il suffit donc que $u \neq 0$ vérifie $A \cdot u = 0$. Mais comme A est de taille $(n - 1, n)$, le rang r vérifie $r \leq n - 1 < n$. Le système admet donc une infinité de solutions non nulles et λ_n est également une valeur propre de T.

99.9 Forme de Schur

Soit une matrice carrée A de taille (n, n). On choisit la valeur propre λ_1 la plus grande de A et on évaluer un vecteur propre non nul correspondant u_1 en résolvant le système :

$$(A - \lambda_1 \cdot I) \cdot u_1 = 0$$

On construit alors le complément orthogonal $(u_2, ..., u_n)$ tel que $(u_1, ..., u_n)$ forme une suite orthonormée. On a donc $u_j^\star \cdot u_i = \boldsymbol{\delta}_{ij}$. Soit la matrice U de taille (n, n) définie par :

$$U_1 = [u_1 \ u_2 \ ... \ u_n]$$

On a bien sur $U_1^{-1} = U_1^\star$ et :

$$\begin{aligned} A \cdot U_1 &= \begin{bmatrix} A \cdot u_1 & A \cdot u_2 & \dots & A \cdot u_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 \cdot u_1 & A \cdot u_2 & \dots & A \cdot u_n \end{bmatrix} \end{aligned}$$

En multipliant à gauche par la duale, on obtient :

$$\begin{aligned} U_1^\star \cdot A \cdot U_1 &= \begin{bmatrix} (\lambda_1 \cdot u_1^\star \cdot u_1) & (u_1^\star \cdot A \cdot u_2) & \dots & (u_1^\star \cdot A \cdot u_n) \\ (\lambda_1 \cdot u_2^\star \cdot u_1) & (u_2^\star \cdot A \cdot u_2) & \dots & (u_2^\star \cdot A \cdot u_n) \\ \vdots & \vdots & \vdots & \vdots \\ (\lambda_1 \cdot u_n^\star \cdot u_1) & (u_n^\star \cdot A \cdot u_2) & \dots & (u_n^\star \cdot A \cdot u_n) \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & (u_1^\star \cdot A \cdot u_2) & \dots & (u_1^\star \cdot A \cdot u_n) \\ 0 & (u_2^\star \cdot A \cdot u_2) & \dots & (u_2^\star \cdot A \cdot u_n) \\ \vdots & \vdots & \vdots & \vdots \\ 0 & (u_n^\star \cdot A \cdot u_2) & \dots & (u_n^\star \cdot A \cdot u_n) \end{bmatrix} \end{aligned}$$

ou, plus schématiquement :

$$U_1^\star \cdot A \cdot U_1 = \begin{bmatrix} \lambda_1 & \dots \\ 0 & A^{(n-1)} \end{bmatrix}$$

On peut réitérer le même raisonnement en utilisant la plus grande valeur propre de $A^{(n-1)}$ et une matrice unitaire correspondante $U^{(n-1)}$. Posons :

$$U_2 = \begin{bmatrix} 1 & 0 \\ 0 & U^{(n-1)} \end{bmatrix}$$

On obtient :

$$U_2^\star \cdot U_1^\star \cdot A \cdot U_1 \cdot U_2 = \begin{bmatrix} \lambda_1 & \dots & \dots \\ 0 & \lambda_2 & \dots \\ 0 & 0 & A^{(n-2)} \end{bmatrix}$$

On continue ainsi, en utilisant à l'étape $k + 1$ la plus grande valeur propre de $A^{(n-k)}$ et :

$$U_{k+1} = \begin{bmatrix} I_k & 0 \\ 0 & U^{(n-k)} \end{bmatrix}$$

On s'arrête évidemment lorsque $k + 1 = n$. Posons :

$$U = U_1 \cdot ... \cdot U_n$$

On obtient à la fin du processus une matrice triangulaire :

$$U^\star \cdot A \cdot U = T = \begin{bmatrix} \lambda_1 & \dots & \dots \\ 0 & \ddots & \dots \\ 0 & 0 & \lambda_n \end{bmatrix}$$

Comme :

$$U^{-1} = U_n^\star \cdot ... \cdot U_1^\star = U^\star$$

on obtient, en multipliant la relation précédente à gauche par U et à droite par $U^\star$ la décomposition :

$$A = U \cdot T \cdot U^\star$$

On appelle cette décomposition la forme de Schur et on la note :

$$(T, U) = \mathrm{Schur}(A)$$

Valeurs propres

La matrice T étant triangulaire supérieure, ses éléments diagonaux sont des valeurs propres de T. Par invariance sous transformation, on voit aussi que les valeurs propres de :

$$A = U \cdot T \cdot U^\star = U \cdot T \cdot U^{-1}$$

sont égales aux valeurs propres de T. La décomposition de Schur d'une matrice A nous permet donc de connaître les valeurs propres en examinant la matrice triangulaire obtenue.

Ordre

Nous allons montrer par récurrence sur la taille de A que les valeurs propres sont triées par ordre décroissant. Si $n = 1$, il n'y a rien à montrer. Supposons à présent que le résultat soit vrai pour $n-1$. Soit :

$$(R, V) = \mathrm{Schur}(A^{(n-1)})$$

Par l'hypothèse de récurrence, R contient sur sa diagonale les valeurs propres de $A^{(n-1)}$ triées par ordre décroissant :

$$\lambda_2 \ge \lambda_3 \ge ...$$

Mais on a aussi :

$$\begin{bmatrix} 1 & 0 \\ 0 & V^\star \end{bmatrix} \cdot U_1^\star \cdot A \cdot U_1 \cdot \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} \lambda_1 & \dots \\ 0 & R \end{bmatrix}$$

Les valeurs propres de $A^{(n-1)}$ sont donc également les valeurs propres de A. Par construction de l'algorithme, λ_1 est la plus grande valeur propre et on a $\lambda_1 \ge \max\{\lambda_2, ..., \lambda_n\}$. On en conclut finalement que :

$$\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge ...$$

99.10 Algorithme QR

Le but est d'évaluer les valeurs propres d'une matrice A. Pour cela, on tente d'obtenir une approximation de la forme de Schur. Il s'agit d'un algorithme itératif qui part de :

$$A_0 = A$$

A chaque itération k, on décompose la matrice A_k obtenue à l'itération précédente en utilisant l'algorithme de Householder :

$$(Q_k, R_k) = \text{Householder}(A_k)$$

On a alors $A_k = Q_k \cdot R_k$, avec $Q_k^\star = Q_k^{-1}$. On construit ensuite une nouvelle matrice A_{k+1} par :

$$A_{k+1} = Q_k^\star \cdot A_k \cdot Q_k = Q_k^\star \cdot Q_k \cdot R_k \cdot Q_k = R_k \cdot Q_k$$

Au bout de p itérations, la matrice A_p produite par l'algorithme vérifie :

$$A_p = Q_p^\star \cdot ... \cdot Q_0^\star \cdot A \cdot Q_0 \cdot ... \cdot Q_p$$

En inversant cette relation, on obtient :

$$A = Q_0 \cdot ... \cdot Q_p \cdot A_p \cdot Q_p^\star \cdot ... \cdot Q_0^\star$$

En injectant l'identité $A_p \cdot Q_p^\star = R_p$, on arrive à :

$$A = Q_0 \cdot ...Q_p \cdot R_p \cdot Q_{p-1}^\star \cdot ... \cdot Q_0^\star$$

Si les suites :

$$\mathcal{U}_p = Q_0 \cdot ...Q_p$$

et R_p convergent :

$$\lim_{p\to\infty} \mathcal{U}_p = U$$
$$\lim_{p\to\infty} R_p = R$$

on a bien évidemment :

$$\lim_{p\to\infty} \mathcal{U}_{p-1}^\star = U^\star$$

On en déduit la forme approximative :

$$A = U \cdot R \cdot U^\star \approx \mathcal{U}_N \cdot R_N \cdot \mathcal{U}_N^\star$$

pour N suffisamment grand. Par invariance sous transformation réversible, les estimations des valeurs propres de A sont sur la diagonale de la matrice triangulaire supérieure $R_N \approx R$. On peut se servir de ces estimations pour appliquer l'algorithme de Schur et construire une matrice triangulaire où les valeurs propres sont triées par ordre décroissant sur la diagonale.

99.11 Matrices hermitiennes

Si A est hermitienne ($A = A^\star$), la forme de Schur $(T, U) = \mathrm{Schur}(A)$ implique que :

$$A = U \cdot T \cdot U^\star = A^\star = U \cdot T^\star \cdot U^\star$$

On en déduit en multipliant à gauche par $U^\star$ et à droite par U que $T = T^\star$. Mais comme T est triangulaire supérieure, on a $t_{ij} = \mathrm{comp}_{ij}(T) = 0$ si $i > j$. De l'autre coté de la diagonale, on a $t_{ji} = \bar{t}_{ij} = 0$. On en conclut que les seuls éléments potentiellement non nuls doivent se trouver sur la diagonale ($i = j$). La matrice T se réduit donc à une matrice diagonale formée à partir des valeurs propres λ_i (non nécessairement distinctes) de A :

$$T = \Lambda = \operatorname*{diag}_n(\lambda_1, ..., \lambda_n)$$

En multipliant $A = U \cdot \Lambda \cdot U^\star$ à droite par U, on a :

$$A \cdot U = \Lambda \cdot U$$

Donc, si $u_i = \mathrm{colonne}_i\, U$, on a :

$$A \cdot u_i = \lambda_i \cdot u_i$$

Les vecteurs propres sont donc donnés par les colonnes de U. Comme $U^\star \cdot U = I$, on a la propriété d'orthonormalité :

$$u_i^\star \cdot u_j = \boldsymbol{\delta}_{ij}$$

99.12 Théorème de Courant-Fisher

Soit une matrice hermitienne A et sa forme de Schur :

$$A = U \cdot \Lambda \cdot U^\star$$

Nous posons :

$$\begin{aligned} \lambda_i &= \operatorname*{comp}_{ii} \Lambda \\ u_i &= \operatorname*{colonne}_i U \end{aligned}$$

pour les valeurs propres et vecteurs propres correspondants. Les valeurs propres étant triées par ordre décroissant, on a :

$$\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge ...$$

On considère le rapport de Rayleigh défini par :

$$R(x) = \frac{x^\star \cdot A \cdot x}{x^\star \cdot x}$$

pour tout $x \in \mathbb{C}^n \setminus \{0\}$. Choisissons un tel x. Comme les u_i forment une base orthonormée, on a :

$$x = \sum_{i=1}^n a_i \cdot u_i$$

où $a_i = \langle u_i \mid x \rangle$ par orthonormalité. Le rapport s'écrit alors :

$$\begin{aligned} R(x) &= \frac{\sum_{i,j=1}^n \overline{a}_i \cdot a_j \cdot u_i^\star \cdot A \cdot u_j}{\sum_{i,j=1}^n \overline{a}_i \cdot a_j \cdot u_i^\star \cdot u_j} \\ &= \frac{\sum_{i=1}^n |a_i|^2 \cdot \lambda_i}{\sum_{j=1}^n |a_j|^2} \end{aligned}$$

Posons :

$$w_i = \frac{|a_i|^2}{\sum_{j=1}^n |a_j|^2}$$

Il est clair que les w_i sont des réels positifs et que :

$$\sum_{i=1}^n w_i = \frac{\sum_{i=1}^n |a_i|^2}{\sum_{j=1}^n |a_j|^2} = 1$$

On a alors :

$$R(x) = \sum_{i=1}^n w_i \cdot \lambda_i$$

Propriétés extrémales

Nous allons analyser les extrema du rapport de Rayleigh sur les ensembles :

$$\begin{aligned} \mathcal{P}_m &= \mathrm{span}\{u_1, u_2, ..., u_m\} \setminus \{0\} \\ \mathcal{Q}_m &= \mathcal{P}_{m-1}^\perp \setminus \{0\} \end{aligned}$$

— Si $x \in \mathcal{P}_m$, les seules coordonnées non nulles sont $a_1, ..., a_m$. Il en va donc de même des w_i et :

$$R(x) = \sum_{i=1}^m w_i \cdot \lambda_i$$

On voit aussi que :

$$\begin{aligned} 1 = \sum_{i=1}^n w_i &= \sum_{i=1}^m w_i + \sum_{i=m+1}^n w_i \\ &= \sum_{i=1}^m w_i + 0 \\ &= \sum_{i=1}^m w_i \end{aligned}$$

Les valeurs propres étant ordonnées par ordre décroissant, on a donc :

$$R(x) \geq \lambda_m \sum_{i=1}^m w_i = \lambda_m$$

Cette borne inférieure est atteinte en $u_m \in \mathcal{P}_m$:

$$R(u_m) = \lambda_m$$

On en conclut que λ_m est le minimum de R sur l'espace $\mathcal{P}_m$:

$$\lambda_m = \min_{x \in \mathcal{P}_m} \frac{x^\star \cdot A \cdot x}{x^\star \cdot x}$$

et que u_m est solution du problème de minimisation :

$$u_m \in \arg\min_{x \in \mathcal{P}_m} \frac{x^\star \cdot A \cdot x}{x^\star \cdot x}$$

— Si $x \in \mathcal{Q}_m$, on doit avoir $\langle z \mid x \rangle = 0$ pour tout $z \in \mathcal{P}_{m-1}$, et en particulier :

$$a_k = \langle u_k \mid x \rangle = 0$$

pour tout $k \in \{1, ..., m-1\}$. Les $m-1$ premières coordonnées sont nulles. Les $m-1$ premiers w_i le sont donc aussi et :

$$R(x) = \sum_{i=m}^{n} w_i \cdot \lambda_i$$

La somme des w_i se simplifie alors en :

$$\sum_{i=m}^{n} w_i = 1$$

Les valeurs propres étant ordonnées par ordre décroissant, on a donc :

$$R(x) \le \lambda_m \sum_{i=m}^{n} w_i = \lambda_m$$

Cette borne supérieure est atteinte en $u_m \in \mathcal{Q}_m$:

$$R(u_m) = \lambda_m$$

On en conclut que λ_m est le maximum de R sur l'espace $\mathcal{Q}_m$:

$$\lambda_m = \max_{x \in \mathcal{Q}_m} \frac{x^\star \cdot A \cdot x}{x^\star \cdot x}$$

et que u_m est solution du problème de maximisation :

$$u_m \in \arg\max_{x \in \mathcal{Q}_m} \frac{x^\star \cdot A \cdot x}{x^\star \cdot x}$$

Minimax

Soit $m \in \mathbb{N}$ et la collection d'espaces vectoriels de dimension m générés par des bases orthonormées :

$$\mathcal{D}_m = \{\mathrm{span}\{v_1, ..., v_m\} : v_i \in \mathbb{C}^n,\ \langle v_i \mid v_j \rangle = \delta_{ij} \text{ pour tout } i, j \in \{1, ..., m\}\}$$

On définit des collections associées par :

$$\begin{aligned} \mathcal{V}_m &= \{X \setminus \{0\} : X \in \mathcal{D}_m\} \\ \mathcal{W}_m &= \{X^\perp \setminus \{0\} : X \in \mathcal{D}_{m-1}\} \end{aligned}$$

— Soit $X \in \mathcal{V}_m$ et la suite $(v_1, ..., v_m)$ formant une base orthonormée de X. Il est équivalent d'imposer la contrainte $x \in X \cap \mathcal{Q}_m$, ou d'imposer simultanément que $x \neq 0$ s'écrive :

$$x = \sum_{j=1}^{m} a_j \cdot v_j$$

et que x soit orthogonal à $u_1, ..., u_{m-1}$:

$$\langle u_i \mid x \rangle = \sum_{j=1}^{m} \langle u_i \mid v_j \rangle \cdot a_j = 0$$

pour tout $i \in \{1, ..., m-1\}$. Soit la matrice $C \in \mathbb{M}(\mathbb{C}, m-1, m)$ de composantes :

$$\operatorname*{comp}_{ij} C = \langle u_i \mid v_j \rangle$$

et le vecteur $a = [a_1 \ ... \ a_m]^\star$. On doit alors avoir $C \cdot a = 0$. Cette matrice étant strictement longue, il existe une infinité de solution dans l'espace :

$$S = \{a \in \mathbb{C}^m : C \cdot a = 0\}$$

On peut donc choisir $a \neq 0$ dans S correspondant à un $x \neq 0$ appartenant à $X \cap \mathcal{Q}_m$. On a alors :

$$\min_{z \in X} R(z) \le R(x) \le \max_{z \in \mathcal{Q}_m} R(z) = \lambda_m$$

L'espace X produit donc un minimum inférieur ou égal à λ_m. Le cas particulier $X = \mathcal{P}_m \in \mathcal{V}_m$ atteignant la borne, on en déduit que :

$$\lambda_m = \max_{X \in \mathcal{V}_m} \min_{x \in X} \frac{x^\star \cdot A \cdot x}{x^\star \cdot x}$$

— Soit $Y \in \mathcal{W}_m$. On peut alors trouver un $X \in \mathcal{D}_{m-1}$ tel que $Y = X^\perp \setminus \{0\}$. Soit la suite $(v_1, ..., v_{m-1})$ formant une base orthonormée de X. Il est équivalent d'imposer la contrainte $x \in Y \cap \mathcal{P}_m$, ou d'imposer simultanément que $x \neq 0$ s'écrive :

$$x = \sum_{j=1}^{m} a_j \cdot u_j$$

et que x soit orthogonal à $v_1, ..., v_{m-1}$:

$$\langle v_i \mid x \rangle = \sum_{j=1}^{m} \langle v_i \mid u_j \rangle \cdot a_j = 0$$

pour tout $i \in \{1, ..., m-1\}$. Soit la matrice $C \in \mathbb{M}(\mathbb{C}, m-1, m)$ de composantes :

$$\operatorname*{comp}_{ij} C = \langle v_i \mid u_j \rangle$$

et le vecteur $a = [a_1 \ ... \ a_m]^\star$. On doit alors avoir $C \cdot a = 0$. Cette matrice étant strictement longue, il existe une infinité de solution dans l'espace :

$$S = \{a \in \mathbb{C}^m : C \cdot a = 0\}$$

On peut donc choisir $a \neq 0$ dans S correspondant à un $x \neq 0$ appartenant à $Y \cap \mathcal{P}_m$. On a alors :

$$\lambda_m = \min_{z \in \mathcal{P}_m} R(z) \le R(x) \le \max_{z \in Y} R(z)$$

L'espace Y produit donc un maximum supérieur ou égal à λ_m. Le cas particulier $Y = \mathcal{Q}_m \in \mathcal{V}_m$ atteignant la borne, on en déduit que :

$$\lambda_m = \min_{Y \in \mathcal{W}_m} \max_{x \in Y} \frac{x^\star \cdot A \cdot x}{x^\star \cdot x}$$

Remarque

Si les valeurs propres étaient ordonnées par ordre croissant :

$$\lambda_1 \le \lambda_2 \le \lambda_3 \le ...$$

on aurait bien entendu :

$$\lambda_m = \min_{X \in \mathcal{V}_m} \max_{x \in X} \frac{x^\star \cdot A \cdot x}{x^\star \cdot x} = \max_{Y \in \mathcal{W}_m} \min_{x \in Y} \frac{x^\star \cdot A \cdot x}{x^\star \cdot x}$$

Résolution numérique

Les propriétés extrémales nous offrent un moyen d'obtenir numériquement des approximations des valeurs et vecteurs propres d'une matrice hermitienne.

Valeur propre maximale

Nous allons nous servir d'un algorithme d'optimisation pour obtenir le maximum λ_1 de R sur $\mathcal{Q}_1 = \mathbb{R}^n$. A titre d'exemple, nous choisissons « la plus grande montée », qui n'est rien d'autre que l'opposé de la plus grande descente. On part d'un vecteur $x \neq 0$ et on considère l'itération k :

$$x_{k+1} = x_k + \alpha_k \cdot \delta_k$$

où $\alpha_k \in \mathbb{R}$ et $\delta_k \in \mathbb{R}^n$. Soit $J_k = (\partial R(x_k))^\star$ et le développement :

$$R(x_{k+1}) \approx R(x_k) + \alpha_k \cdot J_k^\star \cdot \delta_k$$

On a vu que pour maximiser $J_k^\star \cdot \delta_k = \langle J_k \mid \delta_k \rangle$ sur $\mathfrak{B}(0, \|J_k\|)$, il suffit de choisir $\delta_k = J_k$. On pose donc :

$$x_{k+1} = x_k + \alpha_k \cdot J_k$$

Si $H_k = \partial^2 R(x_k)$, la valeur optimale de α_k s'écrit :

$$\alpha_k = \frac{J_k^\star \cdot J_k}{J_k^\star \cdot H_k \cdot J_k}$$

On peut aussi utiliser un autre algorithme de minimisation libre pour construire une suite x_k dont on espère que $\lim_{k \to \infty} x_k = u_1$ et $\lim_{k \to \infty} R(x_k) = \lambda_1$.

Autres valeurs propres

Supposons que l'on ait déjà obtenu une bonne approximation de $(\lambda_1, ..., \lambda_{m-1})$ et de $(u_1, ..., u_{m-1})$. Si on veut que $x \in \mathcal{Q}_m$, il faut et il suffit d'imposer que :

$$\langle u_1 \mid x \rangle = ... = \langle u_{m-1} \mid x \rangle = 0$$

On est donc amenés à construire le complémentaire orthogonal $(v_m, ..., v_n)$ de $(u_1, ..., u_{m-1})$ et à poser :

$$V = [v_m \ v_{m+1} \ ... \ v_n]$$

On a alors $\mathrm{span}\{v_m, ..., v_n\} = \mathcal{Q}_m$. Pour tout $x \in \mathcal{Q}_m$, on peut donc trouver $z = [z_m \ ... \ z_n]^\star \in \mathbb{R}^{n-m+1}$ tel que :

$$x = \sum_{i=m}^{n} z_i \cdot v_i = V \cdot z$$

On pose donc :

$$\varrho(z) = R(V \cdot z) = \frac{z^\star \cdot V^\star \cdot A \cdot V \cdot z}{z^\star \cdot V^\star \cdot V \cdot z}$$

Par orthonormalité, on a $V^\star \cdot V = I$. On est donc finalement amené à maximiser :

$$\varrho(z) = \frac{z^\star \cdot V^\star \cdot A \cdot V \cdot z}{z^\star \cdot z}$$

sur $\mathbb{R}^{n-m+1}$. Pour cela, on part de $z_0 \neq 0$ et on itère :

$$z_{k+1} = z_k + \alpha_k \cdot \partial\varrho(z_k)$$

où :

$$\alpha_k = \frac{\partial\varrho(z_k)^\star \cdot \partial\varrho(z_k)}{\partial\varrho(z_k)^\star \cdot \partial^2\varrho(z_k) \cdot \partial\varrho(z_k)}$$

Chapitre 100

Valeurs singulières

100.1 Décomposition en valeurs singulières

Soit les espaces vectoriels E et F et une application linéaire $A : E \mapsto F$ admettant un dual $A^\star : F \mapsto E$. Les applications $A^\star \circ A$ et $A \circ A^\star$ étant auto-adjointes, il y a fort à parier que leurs valeurs et vecteurs propres possèdent d'importantes propriétés.

Supposons que $A^\star \circ A$ admette les valeurs propres $\lambda_i \in \mathbb{K}$ triées par ordre décroissant ($\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge ...$) et correspondant aux vecteurs propres $v_i \in E$ formant une suite orthonormée. On a donc :

$$A^\star \circ A(v_i) = \lambda_i \cdot v_i$$

On voit que les vecteurs $z_i = A(v_i) \in F$ possèdent la propriété :

$$A^\star(z_i) = A^\star \circ A(v_i) = \lambda_i \cdot v_i$$

et :

$$A \circ A^\star(z_i) = A(\lambda_i \cdot v_i) = \lambda_i \cdot A(v_i) = \lambda_i \cdot z_i$$

Les z_i sont donc vecteurs propres de $A \circ A^\star$ de valeurs propres λ_i identiques à celles de $A^\star \circ A$. On a l'orthogonalité :

$$\langle z_i \mid z_j \rangle = \langle A(v_i) \mid A(v_j) \rangle = \langle v_i \mid A^\star \circ A(v_j) \rangle = \lambda_j \cdot \langle v_i \mid v_j \rangle = \lambda_i \cdot \boldsymbol{\delta}_{ij}$$

On voit aussi que les valeurs propres sont positives :

$$\lambda_i = \langle A(v_i) \mid A(v_i) \rangle \ge 0$$

Comme elles sont également triées par ordre décroissant, on a $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_r > 0$ pour un certain $r \in \mathbb{N}$, et $\lambda_n = 0$ pour tout $n > r$. Dans la suite, nous nous restreignons aux valeurs propres non nulles. On peut alors poser :

$$\sigma_i = \sqrt{\lambda_i} > 0$$

afin de normaliser les z_i :

$$u_i = \frac{1}{\sigma_i} \cdot z_i = \frac{1}{\sigma_i} \cdot A(v_i)$$

On a alors :

$$\langle u_i \mid u_j \rangle = \langle v_i \mid v_j \rangle = \boldsymbol{\delta}_{ij}$$

ainsi que :

$$A^\star(u_i) = \frac{\lambda_i}{\sigma_i} \cdot v_i = \sigma_i \cdot v_i$$

Nous disposons donc des relations primales et duales :

$$A(v_i) = \sigma_i \cdot u_i$$
$$A^\star(u_i) = \sigma_i \cdot v_i$$

Pour tout $x \in \operatorname{span}\{v_1, ..., v_r\}$, on a :

$$x = \sum_{i=1}^{r} \langle v_i \mid x \rangle \cdot v_i$$

et :

$$\begin{aligned} A(x) &= \sum_{i=1}^{r} \langle v_i \mid x \rangle \cdot A(v_i) \\ &= \sum_{i=1}^{r} \langle v_i \mid x \rangle \cdot \sigma_i \cdot u_i \end{aligned}$$

100.2 Représentation tensorielle

On conclut de ce qui précède que A peut être représentée sur $\operatorname{span}\{v_1, ..., v_n\}$ par le tenseur associé :

$$\mathcal{A} = \sum_{i=1}^{r} \sigma_i \cdot u_i \otimes v_i$$

de sorte que :

$$A(x) = \mathcal{A} \cdot x = \langle \mathcal{A} \odot x \rangle_1 = \sum_{i=1}^{r} \sigma_i \cdot u_i \cdot \langle v_i \mid x \rangle$$

On appelle une telle représentation une décomposition en valeurs singulières.

100.3 Dualité

Le tenseur dual est donc :

$$\mathcal{A}^\star = \sum_{i=1}^{r} \sigma_i \cdot v_i \otimes u_i$$

Propriétés

On retrouve sans surprise les représentation de :

$$\begin{aligned} \mathcal{A}^\star \cdot \mathcal{A} &= \sum_{i,j=1}^{r} \sigma_i \cdot \sigma_j \cdot \langle u_i \mid u_j \rangle \cdot v_i \otimes v_j \\ &= \sum_{i=1}^{r} \sigma_i^2 \cdot v_i \otimes v_i \end{aligned}$$

et de :

$$\begin{aligned} \mathcal{A} \cdot \mathcal{A}^\star &= \sum_{i,j=1}^r \sigma_i \cdot \sigma_j \cdot \langle v_i \mid v_j \rangle \cdot u_i \otimes u_j \\ &= \sum_{i=1}^r \sigma_i^2 \cdot u_i \otimes u_i \end{aligned}$$

en fonction de leurs valeurs et vecteurs propres.

100.4 Inverse

Supposons que $(v_1, ..., v_r)$ forme une base de E et que $(u_1, ...u_r)$ forme une base de F. Soit $x \in E$ et $y \in F$ tels que $y = A(x) = \mathcal{A} \cdot x$. On a :

$$y = \sum_{i=1}^r \langle u_i \mid y \rangle \cdot u_i = \mathcal{A} \cdot x = \sum_{i=1}^r \sigma_i \cdot u_i \cdot \langle v_i \mid x \rangle$$

On en déduit en comparant que $\sigma_i \cdot \langle v_i \mid x \rangle = \langle u_i \mid y \rangle$, ce qui nous donne les produits scalaires correspondant aux coordonnées de x par rapport aux v_i :

$$\langle v_i \mid x \rangle = \frac{1}{\sigma_i} \cdot \langle u_i \mid y \rangle$$

On a donc :

$$x = \sum_i \langle v_i \mid x \rangle \cdot v_i = \sum_i \frac{1}{\sigma_i} \cdot \langle u_i \mid y \rangle \cdot v_i$$

Donc, si on pose :

$$\mathcal{A}^{-1} = \sum_{i=1}^r \frac{1}{\sigma_i} \cdot v_i \otimes u_i$$

on a :

$$x = \mathcal{A}^{-1} \cdot y$$

100.5 Pseudo-inverse

Nous ne supposons à présent plus que les suites de vecteurs $(u_1, ..., u_r)$ et $(v_1, ..., v_r)$ forment des bases de E et F, mais nous définissons malgré tout par analogie le tenseur pseudo-inverse de A par :

$$\mathcal{A}^\dagger = \sum_{i=1}^r \frac{1}{\sigma_i} \cdot v_i \otimes u_i$$

Le pseudo-inverse $A^\dagger$ de l'application linéaire correspondante A est donc défini par :

$$A^\dagger(y) = \mathcal{A}^\dagger \cdot y = \sum_{i=1}^r \frac{1}{\sigma_i} \cdot v_i \cdot \langle u_i \mid y \rangle$$

Tenseurs de projections

On voit que :

$$\begin{aligned}\mathcal{A}^\dagger \cdot \mathcal{A} &= \sum_{i,j=1}^r \frac{1}{\sigma_i} \cdot \sigma_j \cdot \langle u_i \mid u_j \rangle \cdot v_i \otimes v_j \\ &= \sum_{i=1}^r v_i \otimes v_i\end{aligned}$$

correspond au tenseur de projection sur $\mathrm{span}\{v_1, ..., v_r\}$. De même :

$$\begin{aligned}\mathcal{A} \cdot \mathcal{A}^\dagger &= \sum_{i,j=1}^r \sigma_i \cdot \frac{1}{\sigma_j} \cdot \langle v_i \mid v_j \rangle \cdot u_i \otimes u_j \\ &= \sum_{i=1}^r u_i \otimes u_i\end{aligned}$$

correspond au tenseur de projection sur $\mathrm{span}\{u_1, ..., u_r\}$.

Dualité

On a clairement :

$$(\mathcal{A} \cdot \mathcal{A}^\dagger)^\star = \mathcal{A} \cdot \mathcal{A}^\dagger$$
$$(\mathcal{A}^\dagger \cdot \mathcal{A})^\star = \mathcal{A}^\dagger \cdot \mathcal{A}$$

Produits

On déduit des résultats ci-dessus que :

$$\begin{aligned}\mathcal{A} \cdot \mathcal{A}^\dagger \cdot \mathcal{A} &= \sum_{i,j=1}^r \sigma_i \cdot \langle v_i \mid v_j \rangle \cdot u_i \otimes v_j \\ &= \sum_{i=1}^r \sigma_i \cdot u_i \otimes v_i \\ &= \mathcal{A}\end{aligned}$$

et :

$$\begin{aligned}\mathcal{A}^\dagger \cdot \mathcal{A} \cdot \mathcal{A}^\dagger &= \sum_{i,j=1}^r \frac{1}{\sigma_i} \cdot \langle u_i \mid u_j \rangle \cdot v_i \otimes u_j \\ &= \sum_{i=1}^r \frac{1}{\sigma_i} \cdot v_i \otimes u_i \\ &= \mathcal{A}^\dagger\end{aligned}$$

Orthogonalité

Soit le tenseur identité $\boldsymbol{\mathcal{I}}$. On déduit de ce qui précède les propriétés d'orthogonalité :

$$\mathcal{A} \cdot (\boldsymbol{\mathcal{I}} - \mathcal{A}^\dagger \cdot \mathcal{A}) = 0$$
$$\mathcal{A}^\dagger \cdot (\boldsymbol{\mathcal{I}} - \mathcal{A} \cdot \mathcal{A}^\dagger) = 0$$

100.6 Représentation matricielle

Soit une matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$ et $p = \min\{m, n\}$. L'algorithme de décomposition en valeurs singulières est très simple. On évalue :

$$(\Lambda_1, U) = \mathrm{Schur}(A \cdot A^\star)$$
$$(\Lambda_2, V) = \mathrm{Schur}(A^\star \cdot A)$$

On a alors $U, \Lambda_1 \in \mathbb{M}(\mathbb{K}, n, n)$ et $V, \Lambda_2 \in \mathbb{M}(\mathbb{K}, m, m)$. Comme les matrices $A^\star \cdot A$ et $A \cdot A^\star$ sont hermitiennes et que leurs valeurs propres sont identiques, les matrices « triangulaires » obtenues sont en fait diagonales et :

$$\Lambda_1 = \underset{n}{\mathrm{diag}}(\lambda_1, ...\lambda_p)$$
$$\Lambda_2 = \underset{m}{\mathrm{diag}}(\lambda_1, ...\lambda_p)$$

On pose alors $\sigma_i = \sqrt{\lambda_i}$ pour $i \in \{1, 2, ..., p\}$ et on a $\sigma_1 \ge \sigma_2 \ge ... \ge \sigma_r > 0$ et $\sigma_{r+1} = ... = \sigma_p = 0$. Les colonnes de U et de V sont les vecteurs propres correspondant :

$$u_i = \underset{i}{\mathrm{colonne}}\, U$$
$$v_i = \underset{i}{\mathrm{colonne}}\, V$$

On a également $U^{-1} = U^\star$ et $V^{-1} = V^\star$. La décomposition en valeurs singulières de A s'écrit :

$$A = \sum_{i=1}^{r} \sigma_i \cdot u_i \otimes v_i = \sum_{i=1}^{r} \sigma_i \cdot u_i \cdot v_i^\star$$

Si nous posons :

$$S = \underset{m,n}{\mathrm{diag}}(\sigma_1, ..., \sigma_r)$$

on peut réécrire la décomposition de A sous la forme :

$$A = U \cdot S \cdot V^\star$$

On note alors :

$$(U, S, V) = \mathrm{DVS}(A)$$

100.7 Pseudo-inverse

Le pseudo-inverse est donné par :

$$A^\dagger = \sum_{i=1}^{r} \frac{1}{\sigma_i} \cdot v_i \cdot u_i^\star$$

On a donc :

$$S^\dagger = \underset{n,m}{\mathrm{diag}} \left(\frac{1}{\sigma_1}, ..., \frac{1}{\sigma_r} \right)$$

et :

$$A^\dagger = V \cdot S^\dagger \cdot U^\star$$

100.8 Systèmes linéaires

Moindres carrés

Soit la matrice $A \in \mathbb{M}(\mathbb{K}, m, n)$, le vecteur matriciel $b \in \mathbb{K}^m$ et l'erreur produite par $x \in \mathbb{K}^n$:

$$e(x) = b - A \cdot x$$

On dit aussi que $e(x)$ est le résidu du système en x. Nous allons tenter de minimiser :

$$\mathcal{E}(x) = e(x)^\star \cdot e(x) = \|e(x)\|^2$$

en utilisant la décomposition en valeurs singulières $A = \sum_{i=1}^r \sigma_i \cdot u_i \cdot v_i^\star$. Comme $(v_1, ..., v_n)$, suite orthonormée et linéairement indépendante, forme une base de $\mathbb{K}^n$, on peut exprimer x en fonction de ses coordonnées dans cette base :

$$x = \sum_{i=1}^n x_i \cdot v_i = \sum_{i=1}^n \langle v_i \mid x \rangle \cdot v_i$$

Comme $(u_1, ..., u_m)$, suite orthonormée et linéairement indépendante, forme une base de $\mathbb{K}^m$, on peut exprimer b comme :

$$b = \sum_{i=1}^m \langle u_i \mid b \rangle \cdot u_i$$

On a également :

$$A \cdot x = \sum_{i=1}^r \sigma_i \cdot u_i \cdot \langle v_i \mid x \rangle = \sum_{i=1}^r \sigma_i \cdot u_i \cdot x_i$$

On en conclut que l'erreur s'écrit :

$$e(x) = \sum_{i=1}^r (\langle u_i \mid b \rangle - \sigma_i \cdot x_i) \cdot u_i + \sum_{i=r+1}^m \langle u_i \mid b \rangle \cdot u_i$$

Posons :

$$e_i(x) = \begin{cases} \langle u_i \mid b \rangle - \sigma_i \cdot x_i & \text{si } i \in \{1, ..., r\} \\ \langle u_i \mid b \rangle & \text{si } i \in \{r+1, ..., m\} \end{cases}$$

On a alors $e(x) = \sum_{i=1}^m e_i(x) \cdot u_i$ et :

$$\mathcal{E}(x) = \|e(x)\|^2 = \sum_{i,j=1}^m \bar{e}_i(x) \cdot e_j(x) \cdot u_i^\star \cdot u_j = \sum_{i=1}^m |e_i(x)|^2$$

On a donc :

$$\mathcal{E}(x) = \sum_{i=1}^r |\langle u_i \mid b \rangle - \sigma_i \cdot x_i|^2 + \sum_{i=r+1}^m |\langle u_i \mid b \rangle|^2$$

Dans le cas où l'on travaille avec des réels, l'annulation de la dérivée par rapport aux x_i nous donne :

$$2(\langle u_i \mid b \rangle - \sigma_i \cdot x_i) = 0$$

lorsque $i \in \{1, ..., r\}$. Nous n'avons par contre aucune contrainte sur $x_{r+1}, ..., x_n$. Un choix satisfaisant les conditions ci-dessus est donc :

$$x_i = \begin{cases} \langle u_i \mid b \rangle / \sigma_i & \text{si } i \in \{1, ..., r\} \\ 0 & \text{si } i \in \{r+1, ..., n\} \end{cases}$$

Notre x potentiellement optimal s'écrit donc :

$$x = \sum_{i=1}^{r} \frac{1}{\sigma_i} \cdot \langle u_i \mid b \rangle \cdot v_i$$

La somme ressemble à une expression faisant intervenir le pseudo-inverse. En effet, on a :

$$A^\dagger \cdot b = \sum_{i=1}^{r} \frac{1}{\sigma_i} \cdot v_i \cdot \langle u_i \mid b \rangle = x$$

Considérons à présent le cas général complexe. On voit que pour le choix $x = A^\dagger \cdot b$:

$$\mathcal{E}(x) = \sum_{i=r+1}^{m} |\langle u_i \mid b \rangle|^2$$

On en déduit la borne inférieure de l'erreur :

$$\mathcal{E}(z) = \sum_{i=1}^{r} |\langle u_i \mid b \rangle - \sigma_i \cdot \langle v_i \mid z \rangle|^2 + \sum_{i=r+1}^{m} |\langle u_i \mid b \rangle|^2 \ge \sum_{i=r+1}^{n} |\langle u_i \mid b \rangle|^2 = \mathcal{E}(x)$$

pour tout $z \in \mathbb{C}^n$. Le choix $x = A^\dagger \cdot b$ minimise bien la norme de l'erreur sur $\mathbb{C}^n$:

$$x = A^\dagger \cdot b \in \arg \min_{z \in \mathbb{C}^n} \mathcal{E}(z)$$

Les $x_{r+1}, ..., x_n$ étant des complexes arbitraires, nous allons montrer que l'ensemble optimal s'écrit :

$$\arg \min_{z \in \mathbb{C}^n} \mathcal{E}(z) = \Gamma = \left\{ \left(A^\dagger \cdot b + \sum_{i=r+1}^{n} x_i \cdot v_i \right) : \ x_{r+1}, ..., x_n \in \mathbb{C} \right\}$$

En effet, on a $\mathcal{E}(z) = \mathcal{E}(x)$ pour tout $z \in \Gamma$. On voit aussi que tout choix de $z \notin \Gamma$ provoque :

$$\sum_{i=1}^{r} |\langle u_i \mid b \rangle - \sigma_i \cdot \langle v_i \mid z \rangle|^2 \ > \ 0$$

et donc $\mathcal{E}(z) \ > \ \mathcal{E}(x)$.

Projection

On peut réécrire Γ sous la forme :

$$\Gamma = \{A^\dagger \cdot b\} + \text{span}\{v_{r+1}, ..., v_n\}$$

Soit la matrice de projection :

$$P = \sum_{i=r+1}^{n} v_i \otimes v_i$$

On sait que $P \cdot z \in \text{span}\{v_{r+1}, ..., v_n\}$ pour tout $z \in \mathbb{C}^n$ et que $P \cdot y = y$ pour tout $y \in \text{span}\{v_{r+1}, ..., v_n\}$. On en conclut que tout $x \in \Gamma$ peut s'écrire sous la forme :

$$x = A^\dagger \cdot b + P \cdot z$$

pour un certain $z \in \mathbb{K}^n$. La matrice de projection P est également la complémentaire de la projection sur $\mathrm{span}\{v_1, ..., v_r\}$. Or, on a a vu que $A^\dagger \cdot A$ est précisément cette matrice de projection. On retrouve donc fort logiquement :

$$I - A^\dagger \cdot A = \sum_{i=1}^{n} v_i \otimes v_i - \sum_{i=1}^{r} v_i \otimes v_i = \sum_{i=r+1}^{n} v_i \otimes v_i = P$$

On a donc en définitive des vecteurs optimaux de la forme :

$$x = A^\dagger \cdot b + (I - A^\dagger \cdot A) \cdot z$$

Solutions

Soit l'espace des solutions :

$$S = \{x \in \mathbb{C}^n : A \cdot x = b\} = \{x \in \mathbb{C}^n : \mathcal{E}(x) = 0\}$$

Si $\langle u_{r+1} \mid b \rangle = ... = \langle u_m \mid b \rangle = 0$, le minimum de l'erreur est nul et $\mathcal{E}(x) = 0$ pour tout $x \in \Gamma$. On en conclut que $x \in S$, d'où $\Gamma \subseteq S$. D'un autre coté, tout $z \in S$ minimise $\mathcal{E}(z) = 0$. On a donc également $S \subseteq \Gamma$ et finalement $\Gamma = S$.

Inversément, si $S \neq \emptyset$, on conclut que $\langle u_{r+1} \mid b \rangle = ... = \langle u_m \mid b \rangle = 0$.

Norme contrainte

Supposons que $S \neq \emptyset$. Soit $x \in \Gamma = S$, que l'on écrit sous la forme :

$$\begin{aligned} x &= A^\dagger \cdot b + \sum_{i=r+1}^{n} x_i \cdot v_i \\ &= \sum_{i=1}^{r} \frac{1}{\sigma_i} \cdot v_i \cdot \langle u_i \mid b \rangle + \sum_{i=r+1}^{n} x_i \cdot v_i \end{aligned}$$

Par orthonormalité des v_i, on a :

$$\|x\|^2 = x^\star \cdot x = \sum_{i=1}^{r} \frac{1}{\sigma_i^2} \cdot |\langle u_i \mid b \rangle|^2 + \sum_{i=r+1}^{n} |x_i|^2$$

On voit que :

$$\|x\|^2 \geq \sum_{i=1}^{r} \frac{1}{\sigma_i^2} \cdot |\langle u_i \mid b \rangle|^2 = \left\|A^\dagger \cdot b\right\|^2$$

On en conclut que le choix $x = A^\dagger \cdot b$ minimise la norme de x sur S :

$$A^\dagger \cdot b \in \arg\min_{z \in S} \|z\|^2$$

Lien avec les résultats précédents

— On a montré précédemment en dérivant les expressions matricielles que le choix :

$$x = (A^\star \cdot A)^{-1} \cdot A^\star \cdot b$$

minimise également l'erreur $\mathcal{E}$ sur $\mathbb{R}^n$ lorsque l'inverse de $A^\star \cdot A$ existe. Si tel est le cas, on a :

$$(A^\star \cdot A)^{-1} = \sum_{i=1}^{r} \frac{1}{\sigma_i^2} \cdot v_i \otimes v_i$$

et :

$$\begin{aligned} (A^\star \cdot A)^{-1} \cdot A^\star &= \sum_{i,j=1}^{r} \frac{\sigma_j}{\sigma_i^2} \cdot \langle v_i \mid v_j \rangle \cdot v_i \otimes u_j \\ &= \sum_{i=1}^{r} \frac{1}{\sigma_i} \cdot v_i \otimes u_i = A^\dagger \end{aligned}$$

— On a vu aussi en utilisant les multiplicateurs de lagrange que le choix :

$$x = A^\star \cdot (A \cdot A^\star)^{-1} \cdot b$$

minimise également la norme de x sur S lorsque l'inverse de $A \cdot A^\star$ existe. Si tel est le cas, on a :

$$(A \cdot A^\star)^{-1} = \sum_{i=1}^{r} \frac{1}{\sigma_i^2} \cdot u_i \otimes u_i$$

et :

$$\begin{aligned} A^\star \cdot (A \cdot A^\star)^{-1} &= \sum_{i,j=1}^{r} \frac{\sigma_i}{\sigma_j^2} \cdot \langle u_i \mid u_j \rangle \cdot v_i \otimes u_j \\ &= \sum_{i=1}^{r} \frac{1}{\sigma_i} \cdot v_i \otimes u_i = A^\dagger \end{aligned}$$

100.9 Image et noyau

Tout vecteur $b = A \cdot x = \sum_{i=1}^{r} \sigma_i \cdot \langle v_i \mid x \rangle \cdot u_i$ est exprimé comme une combinaison linéaire des $(u_1, ..., u_r)$. On en conclut que $\operatorname{im} A \subseteq \operatorname{span}\{u_1, ..., u_r\}$. Réciproquement, si $b \in \operatorname{span}\{u_1, ..., u_r\}$, on a $\langle u_{r+1} \mid b \rangle = \langle u_m \mid b \rangle = 0$ et l'espace des solutions $\{x \in \mathbb{C}^n : A \cdot x = b\}$ n'est pas vide. On en conclut que $\operatorname{span}\{u_1, ..., u_r\} \subseteq \operatorname{im} A$. D'où finalement :

$$\operatorname{im} A = \operatorname{span}\{u_1, ..., u_r\}$$

Tout vecteur $z \in \operatorname{span}\{v_{r+1}, ..., v_n\}$ vérifie $\langle v_1 \mid z \rangle = ... = \langle v_r \mid z \rangle = 0$. On en déduit que $A \cdot z = \sum_{i=1}^{r} \sigma_i \cdot 0 \cdot u_i = 0$ et que $\operatorname{span}\{v_{r+1}, ..., v_n\} \subseteq \ker A$. Réciproquement, si $z \in \ker A$, on a $\sum_{i=1}^{r} \sigma_i \cdot \langle v_i \mid z \rangle \cdot u_i = 0$ ce qui implique $\langle v_1 \mid z \rangle = ... = \langle v_r \mid z \rangle = 0$. On en conclut que $\ker A \subseteq \operatorname{span}\{v_{r+1}, ..., v_n\}$. D'où finalement :

$$\ker A = \operatorname{span}\{v_{r+1}, ..., v_n\}$$

100.10 Normes

La décomposition en valeurs singulières permet d'évaluer facilement la norme usuelle des applications linéaires :

$$\|A\|_{\text{Lin}} = \sup_{x \neq 0} \frac{\|A \cdot x\|}{\|x\|} = \max\{\sigma_1, ..., \sigma_r\}$$

ainsi que la norme de Frobénius :

$$\|A\|_F = \sqrt{A^\star : A} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

100.11 Fonctions de matrices

La décomposition en valeurs singulières permet d'étendre la définition d'une fonction $f : \mathbb{R} \mapsto \mathbb{R}$. Soit la décomposition de A :

$$A = \sum_{i=1}^{r} \sigma_i \cdot u_i \cdot v_i^\star$$

On définit alors :

$$f(A) = \sum_{i=1}^{r} f(\sigma_i) \cdot u_i \cdot v_i^\star$$

Chapitre 101

Espaces de Hilbert

101.1 Définition

Un espace de Hilbert H est un espace de Banach dont la distance découle d'un produit scalaire $\langle | \rangle$ défini sur H :

$$\mathfrak{dist}(u, v) = \|u - v\| = \sqrt{\langle u - v \mid u - v \rangle}$$

Dans la suite, nous considérons un espace de Hilbert H sur $\mathbb{K}$.

101.2 Utilisation

Les espaces de Hilbert permettent de généraliser en dimension infinie nombre de résultats valables en dimension finie. Parmi eux, les espaces fonctionnels sont très utilisés. Ils permettent entre-autres de relier la solution de certaines équations différentielles à des problèmes de minimisation.

101.3 Complétude du noyau

Soit $\varphi \in H^\star$ et une suite de Cauchy $\{x_1, x_2, ...\} \subseteq \ker \varphi \subseteq H$. L'espace H étant complet, cette suite converge vers $x \in H$. Par continuité de φ, on a :

$$\langle \varphi, x \rangle = \lim_{n \to \infty} \langle \varphi, x_n \rangle = 0$$

On en conclut que $x \in \ker \varphi$. L'espace vectoriel $\ker \varphi$ est donc complet.

101.4 Complétude de l'image

Soit l'application linéaire continue $A : H \mapsto H$ et une suite de Cauchy $\{y_1, y_2, ...\} \subseteq \operatorname{im} A \subseteq H$. L'espace H étant complet, cette suite converge vers $y \in H$. Comme les $y_i \in \operatorname{im} A$, on peut trouver des $x_i \in H$ tels que $A(x_i) = y_i$. Par continuité de A, on a :

$$y = \lim_{n \to \infty} y_n = \lim_{n \to \infty} A(x_n) = A(x) \in \operatorname{im} A$$

On en conclut que $y \in \operatorname{im} A$. L'espace vectoriel $\operatorname{im} A$ est donc complet.

101.5 Complétude de l'espace orthogonal

Soit $V \subseteq H$. Considérons une suite de Cauchy $\{x_1, x_2, ...\} \subseteq V^\perp \subseteq H$. Comme H est complet, cette suite est convergente vers $x \in H$. Choisissons $z \in V$. On a :

$$\langle x - x_n \mid z \rangle = \langle x \mid z \rangle - \langle x_n \mid z \rangle = \langle x \mid z \rangle$$

On en déduit que :

$$|\langle x \mid z \rangle| = |\langle x - x_n \mid z \rangle| \le \|x - x_n\| \cdot \|z\|$$

Comme cette relation doit être valable pour tout n et que $\|x - x_n\|$ converge vers 0, on en déduit que :

$$\langle x \mid z \rangle = 0$$

c'est-à-dire $x \in V^\perp$. On en conclut que $V^\perp$ est un espace complet.

101.6 Théorème de projection

Théorème 101.6.1. *Soit $V \subseteq H$ un sous-espace vectoriel complet. Pour tout $x \in H$, il existe un unique $u \in V$ vérifiant :*

$$\|x - u\| = \inf_{z \in V} \|x - z\| = \mathfrak{dist}(x, V)$$

Démonstration 101.6.1. Soit $x \in H$ et :

$$D = \{\|x - z\| : z \in V\} \subseteq \mathbb{R}$$

L'ensemble D est inclu dans $\mathbb{R}$ et minoré puisque $0 \le D$. On en conclut que l'infimum $\lambda = \inf D$ existe. On a par définition $\mathfrak{dist}(x, V) = \lambda$. Comme $D \subseteq \mathbb{R}$, on a $\lambda \in \operatorname{adh} D$ et la distance au sens des réels est nulle :

$$\mathfrak{dist}(\lambda, D) = \inf_{r \in D} |\lambda - r| = 0$$

On peut donc contruire une suite de $z_n \in V$ tels que :

$$\lim_{n \to \infty} \|x - z_n\| = \lambda$$

Choisissons $\epsilon > 0$ et $m, n \in \mathbb{N}$ tels que :

$$\max\{\|x - z_m\|, \|x - z_n\|\} \le \lambda + \epsilon$$

et appliquons l'égalité du parallèlogramme à :

$$v = x - z_m$$
$$w = x - z_n$$

On a alors :

$$v + w = 2x - (z_m + z_n)$$
$$v - w = z_n - z_m$$

et donc :

$$\|2x - (z_m + z_n)\|^2 + \|z_n - z_m\|^2 = 2\left(\|x - z_m\|^2 + \|x - z_m\|^2\right)$$

Comme :

$$\|2x - (z_m + z_n)\|^2 = \left\|2\left(x - \frac{z_m + z_n}{2}\right)\right\|^2 = 4\left\|x - \frac{z_m + z_n}{2}\right\|^2$$

on peut réécrire ce résultat sous la forme :

$$\|z_n - z_m\|^2 + 4\left\|x - \frac{z_m + z_n}{2}\right\|^2 = 2\left(\|x - z_m\|^2 + \|x - z_m\|^2\right)$$

et donc :

$$\|z_n - z_m\|^2 = 2\left(\|x - z_m\|^2 + \|x - z_m\|^2\right) - 4\left\|x - \frac{z_m + z_n}{2}\right\|^2$$

Comme V est un espace vectoriel, on a $(z_m + z_n)/2 \in V$. Donc, par définition de l'infimum :

$$\left\|x - \frac{z_m + z_n}{2}\right\| \ge \lambda$$

On a aussi :

$$\|x - z_m\|^2 + \|x - z_m\|^2 \le 2(\lambda + \epsilon)^2$$

On en déduit que :

$$\|z_n - z_m\|^2 \le 4(\lambda + \epsilon)^2 - 4\lambda^2 = 8\lambda \cdot \epsilon + 4\epsilon^2$$

ce qui montre que les z_n forment une suite de Cauchy. L'espace V étant complet, cette suite converge vers un certain $u \in V$. On a alors :

$$\|x - u\| = \lim_{n\to\infty} \|x - z_n\| = \lambda = \mathfrak{dist}(x, V)$$

Nous avons donc prouvé l'existence d'un vecteur $u \in V$ minimisant la distance à x. Supposons à présent que $u, v \in V$ vérifient :

$$\|x - u\| = \|x - v\| = \lambda$$

Appliquons le parallélogramme à $u - x$ et $x - v$. On a :

$$(u - x) + (x - v) = u - v$$
$$(u - x) - (x - v) = u + v - 2x$$

et :

$$\begin{aligned} \|u - v\|^2 + \|u + v - 2x\|^2 &= 2(\|u - x\|^2 + \|x - v\|^2) \\ &= 2(\lambda^2 + \lambda^2) = 4\lambda^2 \end{aligned}$$

On en conclut que :

$$\begin{aligned} \|u - v\|^2 &= 4\lambda^2 - \|u + v - 2x\|^2 \\ &= 4\lambda^2 - 4\left\|\frac{u + v}{2} - x\right\|^2 \end{aligned}$$

Mais comme $(u + v)/2 \in V$, on a :

$$\left\|\frac{u + v}{2} - x\right\|^2 \ge \lambda^2$$

et :

$$0 \le \|u - v\|^2 \le 4\lambda^2 - 4\lambda^2 = 0$$

Donc, $\|u - v\| = 0$ et $u = v$, ce qui prouve l'unicité de la solution optimale.

101.7 Application projective

On peut donc définir l'application de projection $P : H \mapsto V$ par :

$$\|x - P(x)\| = \inf_{z \in V} \|x - z\|$$

ou :

$$P(x) = \arg_H \inf_{z \in V} \|x - z\|$$

Comme $P(x) \in V$, cela revient à :

$$P(x) = \arg \min_{z \in V} \|x - z\|$$

101.8 Orthogonalité

Soit $x \in H$ et $u = P(x)$. L'écart $e = x - u$ possède-t-il les mêmes propriétés d'orthogonalité qu'en dimension finie ? Soit $\alpha \in \mathbb{K}$ et $v \in V$. On a $u + \alpha \cdot v \in V$ et donc :

$$\|x - u\|^2 = \|e\|^2 \le \|x - u - \alpha \cdot v\|^2 = \|e - \alpha \cdot v\|^2$$

En développant ce dernier membre, on obtient :

$$\|e\|^2 \le \|e\|^2 - 2\Re(\alpha \cdot \langle e \mid v \rangle) + |\alpha|^2 \cdot \|v\|^2$$

On en déduit que :

$$|\alpha|^2 \cdot \|v\|^2 - 2\Re(\alpha \cdot \langle e \mid v \rangle) \ge 0$$

Utilisant la définition du produit complexe, il vient :

$$|\alpha|^2 \cdot \|v\|^2 - 2\Re(\alpha) \cdot \Re(\langle e \mid v \rangle) + 2\Im(\alpha) \cdot \Im(\langle e \mid v \rangle) \ge 0$$

Choisissons $\alpha = \gamma \in \mathbb{R}$, avec $\gamma > 0$. On a $\Re(\alpha) = \gamma$, $\Im(\alpha) = 0$ et $|\alpha|^2 = \gamma^2$. Donc :

$$\gamma^2 \cdot \|v\|^2 - 2\gamma \cdot \Re(\langle e \mid v \rangle) \ge 0$$

Si nous divisons par γ et que nous faisons passer le second terme dans le second membre, il vient :

$$2\Re(\langle e \mid v \rangle) \le \gamma \cdot \|v\|^2$$

Comme ce doit être valable pour tout γ strictement positif, on en conclut que $\Re(\langle e \mid v \rangle) \le 0$. Recommençons le même procédée avec $\alpha = -\gamma < 0$. On a :

$$\gamma^2 \cdot \|v\|^2 + 2\gamma \cdot \Re(\langle e \mid v \rangle) \ge 0$$

et :

$$2\Re(\langle e \mid v \rangle) \ge -\gamma \cdot \|v\|^2$$

Comme ce doit être valable pour tout $\delta = -\gamma$ strictement négatif, on en conclut que $\Re(\langle e \mid v \rangle) \ge 0$. D'où finalement $\Re(\langle e \mid v \rangle) = 0$.

Choisissons à présent $\alpha = \mathbf{i}\,\gamma$, où le réel $\gamma > 0$ et où $\mathbf{i} = \sqrt{-1}$. On a $\Re(\alpha) = 0$, $\Im(\alpha) = \gamma$ et $|\alpha|^2 = \gamma^2$. Donc :

$$\gamma^2 \cdot \|v\|^2 + 2\gamma \cdot \Im(\langle e \mid v \rangle) \ge 0$$

On en déduit que :

$$2\Im(\langle e \mid v\rangle) \ge -\gamma \cdot \|v\|^2$$

Comme ce doit être valable pour tout $\delta = -\gamma$ strictement négatif, on en conclut que $\Im(\langle e \mid v\rangle) \ge 0$. Recommençons le même procédée avec $\alpha = -\mathbf{i}\,\gamma$, avec $\gamma > 0$. On a :

$$\gamma^2 \cdot \|v\|^2 - 2\gamma \cdot \Im(\langle e \mid v\rangle) \ge 0$$

et :

$$2\Im(\langle e \mid v\rangle) \le \gamma \cdot \|v\|^2$$

Comme ce doit être valable pour tout γ strictement positif, on en conclut que $\Im(\langle e \mid v\rangle) \le 0$. D'où finalement $\Im(\langle e \mid v\rangle) = 0$. Ce produit scalaire ayant des parties réelles et imaginaires nulles, il est nul :

$$\langle e \mid v\rangle = 0$$

Cette relation d'orthogonalité étant valable pour tout $v \in V$, on en conclut que $e \in V^\perp$.

101.9 Identité locale

Soit $v \in V$. Il est clair que le choix $u = v$ minimise la distance $\|v - u\| \ge 0$ sur $u \in V$ puisque $\|v - v\| = 0$. Par unicité de la solution optimale, on en déduit que $P(v) = v$ pour tout $v \in V$.

101.10 Invariance

Comme $v = P(x) \in V$ pour tout $x \in H$, on a :

$$P^2(x) = P \circ P(x) = P(v) = v = P(x)$$

d'où $P^2 = P$.

101.11 Somme directe

On peut donc exprimer tout $x \in H$ comme une somme :

$$x = u + e$$

où $u = P(x) \in V$ et $e = x - P(x) \in V^\perp$. Nous allons voir que cette décomposition est unique. Soit $x \in H$ et les vecteurs $u, v \in V$ et $y, z \in V^\perp$ tels que :

$$x = u + y = v + z$$

On a :

$$0 = \|x - x\|^2 = \|u + y - v - z\|^2 = \|(u - v) + (y - z)\|^2$$

Comme $u - v \in V$ et $y - z \in V^\perp$, on a $\langle u - v \mid y - z\rangle = 0$. On peut donc appliquer le théorème de Pythagore. On obtient :

$$0 = \|(u - v) + (y - z)\|^2 = \|u - v\|^2 + \|y - z\|^2$$

ce qui n'est possible que si $\|u - v\| = \|y - z\| = 0$, c'est-à-dire $u = v$ et $y = z$, ce qui prouve l'unicité de la décomposition. L'espace H est donc la somme directe de V et de $V^\perp$:

$$H = V \bigoplus V^\perp$$

101.12 Biorthogonal

— Soit $x \in V$. Pour tout $z \in V^\perp$, on a :

$$\langle x \mid z \rangle = 0$$

Donc $x \in (V^\perp)^\perp$ et $V \subseteq (V^\perp)^\perp$.

— Soit $x \in (V^\perp)^\perp \subseteq H$. On pose $u = P(x) \in V$ et $v = x - P(x) \in V^\perp$. On a donc $x = u + v$. Par définition de $(V^\perp)^\perp$, on a :

$$\langle x \mid z \rangle = 0$$

pour tout $z \in V^\perp$. Comme $u \in V$, on a aussi $\langle u \mid z \rangle = 0$ et finalement :

$$0 = \langle x \mid z \rangle = \langle u \mid z \rangle + \langle v \mid z \rangle = \langle v \mid z \rangle$$

Si on choisit $z = v$, cela donne $\langle v \mid v \rangle = 0$, d'où $v = 0$ et $x = u \in V$. On en conclut que $(V^\perp)^\perp \subseteq V$.

On conclut de ces deux inclusions que :

$$(V^\perp)^\perp = V$$

101.13 Théorème de représentation de Riesz

Nous allons à présent établir une équivalence entre les formes linéaires de $H^\star$ et le produit scalaire sur H.

Théorème 101.13.1. *Pour toute forme linéaire $\varphi \in H^\star$, il existe un et un seul $u \in H$ tel que :*

$$\langle \varphi, x \rangle = \langle u \mid x \rangle$$

pour tout $x \in H$. On dit que u représente φ sur H.

Démonstration 101.13.1. Soit le noyau :

$$N = \ker \varphi = \{x \in H : \langle \varphi, x \rangle = 0\}$$

— Si $N = H$, il suffit de choisir $u = 0$. On a alors :

$$\langle \varphi, x \rangle = 0 = \langle 0 \mid x \rangle = \langle u \mid x \rangle$$

pour tout $x \in H$. Pour prouver l'unicité, si on a aussi $v \in H$ représente également φ, le choix $x = v$ nous donne :

$$\langle \varphi, v \rangle = 0 = \langle v \mid v \rangle$$

ce qui implique que $v = 0$.

— Dans le cas contraire, on a $H \setminus N \neq \emptyset$. Comme N est complet, on peut lui appliquer les résultats relatifs aux projections, dont la somme directe $H = N \bigoplus N^\perp$. Choisissons $a \in H \setminus N$. On a $\langle \varphi, a \rangle \neq 0$ et une unique décomposition $a = b + c$, où $b \in N$ et $c \in N^\perp$. On a donc par définition $\langle \varphi, b \rangle = 0$ et :

$$0 \neq \langle \varphi, a \rangle = \langle \varphi, b \rangle + \langle \varphi, c \rangle = \langle \varphi, c \rangle$$

En partant de vecteurs de la forme $\gamma \cdot c$, où $\gamma \in \mathbb{K}$, on peut obtenir la valeur de φ en n'importe quel $x \in H$:

$$\langle \varphi, \gamma \cdot c \rangle = \gamma \cdot \langle \varphi, c \rangle = \langle \varphi, x \rangle$$

Il suffit donc de choisir :

$$\gamma = \frac{\langle \varphi, x \rangle}{\langle \varphi, c \rangle}$$

Considérons un $x \in H$ quelconque et sa décomposition unique $x = y + z$, où $y \in N$ et $z \in N^\perp$. Si :

$$w = \frac{\langle \varphi, x \rangle}{\langle \varphi, c \rangle} \cdot c$$

on a $\langle \varphi, w \rangle = \langle \varphi, x \rangle$. Posons $v = x - w$. On a alors :

$$\langle \varphi, v \rangle = \langle \varphi, x \rangle - \langle \varphi, w \rangle = 0$$

On en déduit que $v \in N$. Comme on a aussi $w \in N^\perp$ et $x = v + w$, on conclut par unicité de la décomposition que $v = y$ et $w = z$. Pour tout $u \in H$, on a :

$$\langle u \mid x \rangle = \langle u \mid v \rangle + \langle u \mid w \rangle$$

Par analogie avec $\langle \varphi, v \rangle = 0$, on voudrait bien que $\langle u \mid v \rangle = 0$. Pour cela, il suffit de choisir $u \in N^\perp$, par exemple $u = \lambda \cdot c$ pour un certain $\lambda \in \mathbb{K}$. Si on veut que u représente φ, il faut en particulier que $\langle u \mid c \rangle = \langle \varphi, c \rangle$, c'est-à-dire :

$$\langle \lambda \cdot c \mid c \rangle = \overline{\lambda} \cdot \langle c \mid c \rangle = \langle \varphi, c \rangle$$

d'où :

$$\lambda = \frac{\overline{\langle \varphi, c \rangle}}{\langle c \mid c \rangle}$$

et :

$$u = \frac{\overline{\langle \varphi, c \rangle}}{\langle c \mid c \rangle} \cdot c$$

Soit $x \in H$. On a la décomposition :

$$x = \frac{\langle \varphi, x \rangle}{\langle \varphi, c \rangle} \cdot c + v$$

où $v \in N$. Donc :

$$\begin{aligned} \langle u \mid x \rangle &= \frac{\langle \varphi, x \rangle}{\langle \varphi, c \rangle} \cdot \frac{\langle \varphi, c \rangle}{\langle c \mid c \rangle} \cdot \langle c \mid c \rangle + \langle u \mid v \rangle \\ &= \langle \varphi, x \rangle \end{aligned}$$

Nous avons prouvé l'existence d'un $u \in H$ représentant φ. Pour l'unicité, si u et p représentent φ, on a :

$$\langle u - p \mid x \rangle = \langle u \mid x \rangle - \langle p \mid x \rangle = \langle \varphi, x \rangle - \langle \varphi, x \rangle = 0$$

pour tout $x \in H$, et en particulier pour $x = u - p$, d'où :

$$\|u - p\|^2 = \langle u - p \mid u - p \rangle = 0$$

ce qui implique $u - p = 0$ et donc $u = p$.

101.14 Extension aux formes bilinéaires

Théorème 101.14.1. *Si $\vartheta : H \times H \mapsto \mathbb{K}$ est une forme bilinéaire de norme finie sur H, il existe une unique application linéaire et continue $A : H \mapsto H$ telle que :*

$$\langle u, \vartheta, v \rangle = \langle A(u) \mid v \rangle$$

pour tout $u, v \in H$.

Démonstration 101.14.1. Pour tout $u \in H$, on définit la forme linéaire $\varphi_u : H \mapsto \mathbb{K}$ par :

$$\langle \varphi_u, x \rangle = \langle u, \vartheta, x \rangle$$

pour tout $x \in H$. Pour $u \in H$ fixé, φ_u est continue car :

$$\|\langle \varphi_u, x \rangle\| = \|\langle u, \vartheta, x \rangle\| \le \|\vartheta\|_{\mathrm{Lin}} \cdot \|u\| \cdot \|x\|$$

d'où $\|\varphi_u\| \le \vartheta_{\mathrm{Lin}} \cdot \|u\| < +\infty$. On peut donc trouver un unique représentant $A(u) \in H$ tel que :

$$\langle A(u) \mid x \rangle = \langle \varphi_u, x \rangle = \langle u, \vartheta, x \rangle$$

pour tout $x \in H$, ce qui définit l'application $A : u \in H \mapsto A(u) \in H$. Cette application est linéaire car :

$$\begin{aligned} \langle A(\alpha \cdot u + \beta \cdot v) \mid x \rangle &= \langle \alpha \cdot u + \beta \cdot v, \vartheta, x \rangle \\ &= \alpha \cdot \langle u, \vartheta, x \rangle + \beta \cdot \langle v, \vartheta, x \rangle \\ &= \alpha \cdot \langle A(u) \mid x \rangle + \beta \cdot \langle A(v) \mid x \rangle \end{aligned}$$

pour tout $u, v, x \in H$ et $\alpha, \beta \in \mathbb{K}$. Elle est également continue car :

$$\|A(u)\|^2 = \langle A(u) \mid A(u) \rangle = \langle u, \vartheta, A(u) \rangle \le \|\vartheta\|_{\mathrm{Lin}} \cdot \|u\| \cdot \|A(u)\|$$

On a donc soit $A = 0$ et a fortiori A continue, soit :

$$\|A(u)\| \le \|\vartheta\|_{\mathrm{Lin}} \cdot \|u\| < +\infty$$

101.15 Application adjointe

Théorème 101.15.1. *Soit une application $A : H \mapsto H$ linéaire et continue. Il existe une unique application $A^\star : H \mapsto H$ telle que :*

$$\langle A^\star(v) \mid u \rangle = \langle v \mid A(u) \rangle$$

pour tout $(u, v) \in H^2$.

Démonstration 101.15.1. Choisissons $v \in H$. L'application $\varphi_v : u \mapsto \langle v \mid A(u) \rangle$ est une forme linéaire. Elle est continue car :

$$|\langle \varphi_v, u \rangle| = |\langle v \mid A(u) \rangle| \le \|v\| \cdot \|A(u)\| \le \|v\| \cdot \|A\| \cdot \|u\|$$

On a donc :

$$|\varphi_v| \le \|A\| \cdot \|v\|$$

Le théorème de Riesz nous dit qu'il existe un unique $A^\star(v) \in H$ tel que :

$$\langle A^\star(v) \mid u \rangle = \langle \varphi_v, u \rangle = \langle v \mid A(u) \rangle$$

pour tout $u \in H$. Comme ce résultat est également valable quel que soit $v \in H$, nous avons défini l'application $A^\star : H \mapsto H$ demandée.

101.16 Théorème de Lax-Milgram

Théorème 101.16.1. *Soit une forme bilinéaire $\vartheta : H \times H \mapsto \mathbb{K}$ de norme finie. Nous supposons qu'il existe un réel $\varrho > 0$ tel que :*

$$\langle u, \vartheta, u \rangle \ge \varrho \cdot \|u\|^2$$

pour tout $u \in H$. On dit que ϑ est coercive. Soit une forme linéaire et continue $\varphi : H \mapsto \mathbb{K}$. Nous allons montrer qu'il existe un unique $s \in H$ tel que :

$$\langle s, \vartheta, v \rangle = \langle \varphi, v \rangle$$

pour tout $v \in H$.

Démonstration 101.16.1. En appliquant le théorème de Riesz, on peut trouver un unique $f \in H$ qui représente φ sur H :

$$\langle \varphi, v \rangle = \langle f \mid v \rangle$$

pour tout $v \in H$. On dispose aussi d'une application linéaire continue A telle que :

$$\langle u, \vartheta, v \rangle = \langle A(u) \mid v \rangle$$

pour tout $u, v \in H$. Pour tout s vérifiant la condition du théorème, on doit donc avoir :

$$\langle f \mid v \rangle = \langle \varphi, v \rangle = \langle s, \vartheta, v \rangle = \langle A(s) \mid v \rangle$$

et donc $\langle f \mid v \rangle = \langle A(s) \mid v \rangle$ pour tout $v \in H$. Par linéarité, on en déduit que :

$$\langle f - A(s) \mid v \rangle = 0$$

Le cas particulier $v = f - A(s)$ nous donne :

$$\langle f - A(s) \mid f - A(s) \rangle = \|f - A(s)\|^2 = 0$$

On en déduit que $f - A(s) = 0$, autrement dit $A(s) = f$.

On sait que l'image $V = \operatorname{im} A \subseteq H$ est un espace complet. On a dès lors la somme directe $H = V \bigoplus V^\perp$. Soit $u \in V^\perp$. On a $\langle A(x) \mid u \rangle = 0$ pour tout $x \in H$. Le choix $x = u$ combiné avec la coercivité de ϑ nous donne :

$$0 \le \varrho \cdot \|u\|^2 \le \langle u, \vartheta, u \rangle = \langle A(u) \mid u \rangle = 0$$

Divisant alors par ϱ strictement positif, on obtient $\|u\|^2 = 0$, ce qui n'est possible que si $u = 0$. On en déduit que $V^\perp = \{0\}$. Pour tout $x \in H$, on a donc la décomposition $x = y + 0 = y$ avec $y \in V$. On en déduit que $H = V = \operatorname{im} A$. Donc $f \in \operatorname{im} A$, ce qui prouve l'existence d'au moins un $s \in H$ tel que $A(s) = f$.

D'un autre coté, si $s, t \in H$ vérifient $A(s) = A(t) = f$, on a $A(s - t) = A(s) - A(t) = f - f = 0$. Donc :

$$0 \le \varrho \cdot \|s - t\|^2 \le \langle s - t, \vartheta, s - t \rangle = \langle A(s - t) \mid s - t \rangle = \langle 0 \mid s - t \rangle = 0$$

ce qui implique que $\|s - t\|^2 = 0$, d'où $s = t$. La solution s est donc unique.

Inverse

Nous avons également prouvé que pour tout $v \in H$, il existe un unique $u \in H$ vérifiant $A(u) = v$. L'application A est donc inversible et $A^{-1}(v) = u$.

101.17 Suite orthonormée

Si une suite $\{u_i \in H : i \in \mathbb{N}\}$ vérifie :

$$\langle u_i \mid u_j \rangle = \delta_{ij}$$

pour tout $i, j \in \mathbb{N}$, on dit qu'elle est orthonormée. Dans la suite, nous considérons une suite $\{u_i \in H : i \in \mathbb{N}\}$ orthonormée.

101.18 Inégalité de Bessel

Soit $x \in H$ et une suite orthonormée $\{u_i \in H : i \in \mathbb{N}\}$. Par analogie avec les projections sur des espaces de dimension finie, on pose (sous réserve de convergence) :

$$p = \sum_{i=1}^{+\infty} \langle u_i \mid x \rangle \cdot u_i$$

Soit $e = x - p$. On a :

$$\begin{aligned}
\langle u_k \mid e \rangle &= \langle u_k \mid x \rangle - \sum_i \langle u_i \mid x \rangle \langle u_k \mid u_i \rangle \\
&= \langle u_k \mid x \rangle - \sum_i \langle u_i \mid x \rangle \delta_{ik} \\
&= \langle u_k \mid x \rangle - \langle u_k \mid x \rangle \\
&= 0
\end{aligned}$$

On en conclut que :

$$\langle p \mid e \rangle = \sum_i \langle x \mid u_i \rangle \cdot \langle u_i \mid e \rangle = 0$$

On a donc :

$$\|x\|^2 = \|p + e\|^2 = \|p\|^2 + \|e\|^2$$

d'où :

$$\|x\|^2 - \|p\|^2 = \|e\|^2 \geq 0$$

Par orthonormalité, la norme de p vérifie :

$$\|p\|^2 = \sum_i |\langle u_i \mid x \rangle|^2$$

d'où :

$$\|x\|^2 - \sum_i |\langle u_i \mid x \rangle|^2 \geq 0$$

c'est-à-dire :

$$\sum_i |\langle u_i \mid x \rangle|^2 \leq \|x\|^2$$

101.19 Base hilbertienne

Si, pour tout $x \in H$, la suite des projections finies :

$$p_n = \sum_{i=1}^{n} \langle u_i \mid x \rangle \cdot u_i$$

converge vers x :

$$x = \lim_{n \to \infty} \sum_{i=1}^{n} \langle u_i \mid x \rangle \cdot u_i$$

au sens de la distance dérivant du produit scalaire, on dit que les u_i forment une base de Hilbert (ou une base hilbertienne) de H. On a alors :

$$x = \sum_{i=1}^{+\infty} \langle u_i \mid x \rangle \cdot u_i$$

Les scalaires :

$$x_i = \langle u_i \mid x \rangle$$

sont appelés les coordonnées de x par rapport aux u_i. Dans la suite, nous considérons une suite $\{u_i \in H : i \in \mathbb{N}\}$ formant une base hilbertienne de H.

101.20 Egalité de Parseval

Soit $x \in H$ et $\epsilon > 0$. On peut trouver un $n \in \mathbb{N}$ tel que :

$$D = \left\| x - \sum_{i=1}^{n} \langle u_i \mid x \rangle \cdot u_i \right\| \le \epsilon$$

Mais les propriétés des projections nous disent que :

$$0 \le D^2 = \|x\|^2 - \sum_{i=1}^{n} |\langle u_i \mid x \rangle|^2$$

et donc :

$$0 \le \|x\|^2 - \sum_{i=1}^{n} |\langle u_i \mid x \rangle|^2 \le \epsilon^2$$

En faisant tendre $n \to \infty$, on a $\epsilon \to 0$ et donc :

$$0 \le \|x\|^2 - \sum_{i=1}^{+\infty} |\langle u_i \mid x \rangle|^2 \le 0$$

c'est-à-dire :

$$\|x\|^2 = \sum_{i=1}^{+\infty} |\langle u_i \mid x \rangle|^2$$

101.21 Produit scalaire

Soit $x, y \in E$. On a :

$$x = \sum_{i=1}^{+\infty} x_i \cdot u_i$$

$$y = \sum_{i=1}^{+\infty} y_i \cdot u_i$$

où $x_i = \langle u_i \mid x \rangle$ et $y_i = \langle u_i \mid y \rangle$. Par orthonormalité, leur produit scalaire s'écrit :

$$\langle x \mid y \rangle = \lim_{n \to \infty} \sum_{i=1}^{n} \overline{x}_i \cdot y_i$$

On a donc :

$$\langle x \mid y \rangle = \sum_{i=1}^{+\infty} \overline{x}_i \cdot y_i$$

Chapitre 102

Théorie spectrale

102.1 Progression géométrique

Soit un espace de Hilbert H et une application linéaire $A : H \mapsto H$. On voit que :

$$\begin{aligned} \sum_{k=0}^{n} A^k &= \mathrm{Id} + A + A^2 + ... + A^n \\ A \circ \sum_{k=0}^{n} A^k &= A + A^2 + A^3 + ... + A^{n+1} \end{aligned}$$

En soustrayant ces deux équations, on obtient :

$$(\mathrm{Id} - A) \circ \sum_{k=0}^{n} A^k = \mathrm{Id} - A^{n+1}$$

On montre de même que :

$$\left[\sum_{k=0}^{n} A^k \right] \circ (\mathrm{Id} - A) = \mathrm{Id} - A^{n+1}$$

Infinie

Si $\|A\| < 1$, on a $\|A^n\| \le \|A\|^n \to 0$ lorsque $n \to \infty$ et donc $A^n \to 0$. On en déduit que :

$$(\mathrm{Id} - A) \circ \sum_{k=0}^{+\infty} A^k = \lim_{n\to\infty} (\mathrm{Id} - A^{n+1}) = \mathrm{Id}$$

On a aussi :

$$\left[\sum_{k=0}^{+\infty} A^k \right] \circ (\mathrm{Id} - A) = \mathrm{Id}$$

On en déduit que :

$$(\mathrm{Id} - A)^{-1} = \sum_{k=0}^{+\infty} A^k$$

102.2 Convergence

Soit une application linéaire continue $A : H \mapsto H$. Choisissons $\lambda \in \mathbb{C}$ tel que $\|A\| < |\lambda|$. On a :

$$\left\|A^k\right\| \le \|A\|^k < |\lambda|^k$$

Posons $r = \|A\| / |\lambda| < 1$ et divisons par le module de λ puissance k :

$$\frac{\left\|A^k\right\|}{|\lambda|^k} \le \frac{\|A\|^k}{|\lambda|^k} = r^k < 1$$

On en déduit que :

$$\sum_{k=0}^{n} \frac{\left\|A^k\right\|}{|\lambda|^k} \le \sum_{k=0}^{n} r^k = \frac{1-r^k}{1-r} \le \frac{1}{1-r}$$

La suite des :

$$S_n = \sum_{k=0}^{n} \frac{\left\|A^k\right\|}{|\lambda|^k}$$

étant croissante et bornée, elle converge vers son supremum :

$$S = \sum_{k=0}^{+\infty} \frac{\left\|A^k\right\|}{|\lambda|^k} = \lim_{n\to\infty} S_n = \sup_{n\in\mathbb{N}} S_n$$

Soit $x \in H$ et la suite définie par $u_0 = x$ et :

$$u_n = \sum_{k=0}^{n} \frac{1}{\lambda^k} \cdot A^k(x)$$

Pour tout $m, n \in \mathbb{N}$ avec $m \ge n$, on a :

$$\|u_m - u_n\| = \left\| \sum_{k=n+1}^{m} \frac{1}{\lambda^k} \cdot A^k(x) \right\| \le \sum_{k=n+1}^{m} \frac{\left\|A^k(x)\right\|}{|\lambda|^k}$$

Majorons par la norme de A^k, puis la norme de A puissance k :

$$\|u_m - u_n\| \le \left[\sum_{k=n+1}^{m} \frac{\left\|A^k\right\|}{|\lambda|^k} \right] \cdot \|x\| \le \left[\sum_{k=n+1}^{m} \frac{\|A\|^k}{|\lambda|^k} \right] \cdot \|x\|$$

et effectuons le changement de variable $i = k - (n+1)$. Il vient :

$$\|u_m - u_n\| \le \frac{\|A\|^{n+1}}{|\lambda|^{n+1}} \left[\sum_{i=0}^{m-n-1} \frac{\|A\|^i}{|\lambda|^i} \right] \cdot \|x\| = r^{n+1} \cdot S_{m-n-1} \cdot \|x\|$$

Mais comme $S_{m-n-1} \le S$, on a finalement :

$$\|u_m - u_n\| \le r^{n+1} \cdot S \cdot \|x\|$$

qui converge vers 0 lorsque $n \to \infty$. La suite des u_n est donc de Cauchy et converge vers un certain $L(x) \in H$, ce qui définit l'application $L : H \mapsto H$, que l'on note :

$$L = \sum_{k=0}^{+\infty} \frac{1}{\lambda^k} \cdot A^k$$

Continuité

Cette application est de norme finie et donc continue car pour tout $n \in \mathbb{N}$:

$$\|u_n\| \le \left[\sum_{k=0}^{n} \frac{\|A\|^k}{|\lambda|^k} \right] \cdot \|x\| = \frac{1 - r^{n+1}}{1 - r} \cdot \|x\| \le \frac{\|x\|}{1 - r}$$

On a donc :

$$\|L(x)\| = \lim_{n \to \infty} \|u_n\| \le \frac{\|x\|}{1 - r}$$

d'où :

$$\|L\| \le \frac{1}{1 - r}$$

Inverse

Comme l'application $B = A/\lambda$ vérifie $\|B\| < 1$, on a :

$$(\lambda \cdot \mathrm{Id} - A) \circ (\frac{1}{\lambda} \cdot L) = (\mathrm{Id} - \frac{1}{\lambda} \cdot A) \circ L = \mathrm{Id}$$
$$(\frac{1}{\lambda} \cdot L) \circ (\lambda \cdot \mathrm{Id} - A) = L \circ (\mathrm{Id} - \frac{1}{\lambda} \cdot A) = \mathrm{Id}$$

et donc :

$$(\lambda \cdot \mathrm{Id} - A)^{-1} = \frac{1}{\lambda} \cdot L$$

102.3 Exponentielle

Soit une application linéaire continue $A : H \mapsto H$. Sous réserve de convergence, on définit l'opérateur $\exp(A)$ associé par :

$$\exp(A) = \sum_{k=0}^{+\infty} \frac{1}{k!} \cdot A^k$$

On a donc :

$$\exp(A)(u) = \sum_{k=0}^{+\infty} \frac{1}{k!} \cdot A^k(u)$$

pour tout $u \in H$.

Chapitre 103

Calcul variationnel

103.1 Méthode des variations

Soit l'espace fonctionnel $\mathcal{F}$ et la forme $I : \mathcal{F} \mapsto \mathbb{R}$. Un problème variationnel consiste à chercher une fonction u qui minimise I sur $\mathcal{F}$:

$$I(u) \le I(v)$$

pour tout $v \in \mathcal{F}$. L'astuce consiste à transformer ce problème en considérant la famille de fonctions $\{J_w : w \in \mathcal{F}\}$ définies par :

$$J_w(\epsilon) = I(u + \epsilon \cdot w)$$

pour tout $\epsilon \in \mathbb{R}$. Comme $\mathcal{F}$ est un espace vectoriel, $u + \epsilon \cdot w \in \mathcal{F}$. On en déduit que :

$$J_w(0) = I(u) \le J_w(\epsilon)$$

Si les fonctions J_w sont dérivables, les propriétés des extrema des fonctions $\mathbb{R} \mapsto \mathbb{R}$ nous disent que :

$$\frac{dJ_w}{d\epsilon}(0) = 0$$

pour toute variation $w \in \mathcal{F}$. Si la dérivée seconde existe, on doit également avoir :

$$\frac{d^2 J_w}{d\epsilon^2}(0) \ge 0$$

Ces équations nous permettent de caractériser la solution u de notre problème variationnel.

103.2 Discrétisation

Une technique couramment employée pour résoudre les problèmes variationnels est de choisir une suite de fonctions $\varphi_i \in \mathcal{F}$ linéairement indépendantes et de minimiser sur l'espace vectoriel $\mathcal{F}_n = \operatorname{span}\{\varphi_1, \varphi_2, ..., \varphi_n\} \subseteq \mathcal{F}$. On espère que la solution obtenue sera proche de la solution exacte. Ce sera par exemple le cas si on a schématiquement :

$$\lim_{n \to \infty} \operatorname{adh} \mathcal{F}_n = \mathcal{F}$$

au sens de la distance $\mathfrak{dist}$ définie sur $\mathcal{F}$. Pour toute précision $\epsilon > 0$, on pourra alors trouver un $n \in \mathbb{N}$ assez grand et un $u_n \in \mathcal{F}_n$ tels que :

$$\mathfrak{dist}(u_n, u) \le \epsilon$$

103.3 Moindres carrés

Soit $f : A \mapsto \mathbb{R}$ et le sous-ensemble $\mathcal{F} \subseteq \mathbb{F}(A, \mathbb{R})$. On cherche la fonction $u \in \mathcal{F}$ qui se rapproche le plus possible de f au sens intégral des moindres carrés. On cherche donc le u qui minimise la fonctionnelle :

$$I(u) = \int_A \Big[u(x) - f(x)\Big]^2 dx$$

Afin de résoudre ce problème, on pose :

$$J_v(\epsilon) = I(u + \epsilon \cdot v) = \int_A \Big[u(x) + \epsilon \cdot v(x) - f(x)\Big]^2 dx$$

pour tout $\epsilon \in \mathbb{R}$ et $v \in \mathcal{F}$. La dérivée s'écrit :

$$\frac{dJ_v}{d\epsilon}(\epsilon) = \int_A 2\big[u(x) + \epsilon \cdot v(x) - f(x)\big] \cdot v(x)\ dx$$

Comme elle doit s'annuler en $\epsilon = 0$, on a :

$$\frac{dJ_v}{d\epsilon}(0) = 2 \int_A \big[u(x) - f(x)\big] \cdot v(x)\ dx = 0$$

donc :

$$\int_A \big[u(x) - f(x)\big] \cdot v(x)\ dx = 0$$

pour tout $v \in F$.

Discrétisation

On résout approximativement ce problème en posant :

$$u_n(x) = \sum_{i=1}^n U_i \cdot \varphi_i(x)$$

où les $\varphi_i \in \mathcal{F}$ sont des fonctions connues et où les U_i sont des réels à déterminer. Comme on désire que $u_n \approx u$, on impose :

$$\int_A (u_n(x) - f(x)) \cdot \varphi_i(x)\ dx = 0$$

On a alors :

$$\int_A \varphi_i \sum_j U_j \cdot \varphi_j\ dx = \int_A \varphi_i \cdot f\ dx$$

ou :

$$\sum_j \left[\int_A \varphi_i \cdot \varphi_j\ dx\right] \cdot U_j = \int_A \varphi_i \cdot f\ dx$$

Définissons les matrices $A \in \mathbb{M}(\mathbb{R}, n, n)$, $B \in \mathbb{M}(\mathbb{R}, n, 1)$ et $U \in \mathbb{M}(\mathbb{R}, n, 1)$:

$$\begin{aligned} A &= \left[\int_A \varphi_i \cdot \varphi_j\ dx\right]_{i,j} \\ B &= \left[\int_A \varphi_i \cdot f\ dx\right]_i \\ U &= [U_i]_i \end{aligned}$$

Le problème peut alors s'exprimer sous forme matricielle :

$$A \cdot U = B$$

Si la matrice A est inversible, la solution est donnée par :

$$U = A^{-1} \cdot B$$

et nous en déduisons notre approximation $u_n = \sum_i U_i \cdot \varphi_i$.

103.4 Lax-Milgram

Soit un espace fonctionnel de Hilbert $\mathcal{F}$. Nous considérons une forme bilinéaire $a : \mathcal{F} \times \mathcal{F} \mapsto \mathbb{R}$, de norme finie, coercive et symétrique :

$$\langle u, a, v \rangle = \langle v, a, u \rangle$$

pour tout $u, v \in F$. Nous considérons également une forme linéaire continue $b : \mathcal{F} \mapsto \mathbb{R}$ et nous définissons la fonctionnelle $I : \mathcal{F} \mapsto \mathbb{R}$ par :

$$I(v) = \frac{1}{2} \langle v, a, v \rangle - \langle b, v \rangle$$

— Supposons que I atteigne un minimum global en u. On a alors :

$$I(u) \le I(v)$$

pour tout $v \in F$. Choisissons un quelconque $w \in \mathcal{F}$ et posons :

$$J_w(\epsilon) = I(u + \epsilon \cdot w)$$

pour tout $\epsilon \in \mathbb{R}$. Tenant compte des propriétés de a et b, on obtient :

$$J_w(\epsilon) = \frac{1}{2} \langle u, a, u \rangle + \epsilon \cdot \langle u, a, w \rangle + \frac{\epsilon^2}{2} \cdot \langle w, a, w \rangle - \langle b, u \rangle - \epsilon \cdot \langle b, w \rangle$$

La dérivée s'écrit :

$$\frac{dJ_v}{d\epsilon}(\epsilon) = \langle u, a, w \rangle + \epsilon \cdot \langle w, a, w \rangle - \langle b, w \rangle$$

La condition extrémale sur u nous dit que $J_w(0) \le J_w(\epsilon)$ pour tout $\epsilon \in \mathbb{R}$. On doit donc avoir :

$$\frac{dJ_v}{d\epsilon}(0) = \langle u, a, w \rangle - \langle b, w \rangle = 0$$

c'est-à-dire :

$$\langle u, a, w \rangle = \langle b, w \rangle$$

pour tout $w \in \mathcal{F}$. Le théorème de Lax-Milgram nous dit qu'il existe un unique $u \in \mathcal{F}$ vérifiant cette condition.

— Supposons à présent que $u \in \mathcal{F}$ soit l'unique élément de $\mathcal{F}$ vérifiant :

$$\langle u, a, w \rangle = \langle b, w \rangle$$

pour tout $w \in \mathcal{F}$. Choisissons $v \in \mathcal{F}$ et posons $w = v - u$. Comme $\langle u, a, w \rangle = \langle b, w \rangle$, on a :

$$\begin{aligned} I(v) &= I(u+w) \\ &= \frac{1}{2} \langle u, a, u \rangle + \langle u, a, w \rangle + \frac{1}{2} \cdot \langle w, a, w \rangle - \langle b, u \rangle - \langle b, w \rangle \\ &= \frac{1}{2} \langle u, a, u \rangle + \frac{1}{2} \cdot \langle w, a, w \rangle - \langle b, u \rangle \\ &= I(u) + \frac{1}{2} \langle w, a, w \rangle \ge I(u) \end{aligned}$$

On a donc :

$$u = \arg\min_{v \in \mathcal{F}} I(v)$$

Orthogonalité

Soit le sous-espace vectoriel $\mathcal{G} \subseteq \mathcal{F}$ et :

$$v \in \arg\min_{z \in \mathcal{G}} I(z)$$

On a alors :

$$\langle v, a, w \rangle = \langle b, w \rangle$$

pour tout $w \in \mathcal{G}$.
Supposons que $u \in \mathcal{F}$ minimise I sur $\mathcal{F}$ et posons $e = v - u$. On a :

$$\langle e, a, w \rangle = \langle v, a, w \rangle - \langle u, a, w \rangle = \langle b, w \rangle - \langle b, w \rangle = 0$$

Cette propriété est similaire aux propriétés d'orthogonalité vue dans les projections. On est tenté d'en déduire que le v en question minimise la « norme » de l'écart e au sens de a. Soit $z \in \mathcal{G}$ et $\delta = z - v$. On a $z - u = z - v + v - u = \delta + e$ et :

$$\langle z - u, a, z - u \rangle = \langle \delta, a, \delta \rangle + 2 \langle e, a, \delta \rangle + \langle e, a, e \rangle$$

Mais comme $\delta \in \mathcal{G}$, on a $\langle e, a, \delta \rangle = 0$ et :

$$\langle z - u, a, z - u \rangle = \langle \delta, a, \delta \rangle + \langle e, a, e \rangle \ge \langle e, a, e \rangle$$

ce qui prouve que :

$$v \in \arg\min_{z \in \mathcal{G}} \langle z - u, a, z - u \rangle$$

Borne inférieure

Comme a est coercive, on peut trouver un réel $\varrho > 0$ tel que :

$$a(u, u) \ge \varrho \cdot \|u\|^2$$

pour tout $u \in \mathcal{F}$. On sait aussi que :

$$\|b\| = \sup\{|b(u)| : u \in \mathcal{F}, \ \|u\| = 1\} < +\infty$$

Posons :

$$K(\epsilon) = I(\epsilon \cdot v) = \frac{\epsilon^2}{2} \cdot \langle v, a, v \rangle - \epsilon \cdot \langle b, v \rangle$$

et cherchons la valeur de $\epsilon = \gamma$ qui minimise cette fonction. On trouve :

$$\frac{dK}{d\epsilon}(\gamma) = \gamma \cdot \langle v, a, v\rangle - \langle b, v\rangle = 0$$

Donc :

$$\gamma = \frac{\langle b, v\rangle}{\langle v, a, v\rangle}$$

et :

$$K(\gamma) = \frac{\langle b, v\rangle^2}{2\,\langle v, a, v\rangle} - \frac{\langle b, v\rangle^2}{\langle v, a, v\rangle} = -\frac{\langle b, v\rangle^2}{2\,\langle v, a, v\rangle}$$

On en déduit que :

$$I(v) = K(1) \ge K(\gamma) = -\frac{\langle b, v\rangle^2}{2\,\langle v, a, v\rangle}$$

Mais comme :

$$\langle b, v\rangle^2 \le \|b\|^2 \cdot \|v\|^2$$
$$\langle v, a, v\rangle \ge \varrho \cdot \|v\|^2$$

on a :

$$\frac{\langle b, v\rangle^2}{2\,\langle v, a, v\rangle} \le \frac{\|b\|^2}{2\varrho}$$

et :

$$I(v) \ge -\frac{\|b\|^2}{2\varrho}$$

Ce résultat étant valable pour tout $v \in \mathcal{F}$, on a la borne inférieure :

$$\inf_{v\in\mathcal{F}} I(v) \ge -\frac{\|b\|^2}{2\varrho}$$

Discrétisation

Afin de trouver une approximation $u_n \approx u$, on pose :

$$u_n = \sum_{i=1}^{n} U_i \cdot \varphi_i$$

où les $\varphi_i \in \mathcal{F}$ sont connues et où les U_i sont des réels à déterminer. La linéarité de a et b nous permet de développer :

$$I(u_n) = \frac{1}{2}\sum_{i,j=1}^{n} U_i \cdot \langle \varphi_i, a, \varphi_j\rangle \cdot U_j - \sum_{i=1}^{n} \langle b, \varphi_i\rangle\, U_i$$

Si nous définissons les matrices :

$$\begin{aligned} A &= [\langle \varphi_i, a, \varphi_j\rangle]_{i,j} \\ B &= [\langle b, \varphi_i\rangle]_i \\ U &= [U_i]_i \end{aligned}$$

on peut réécrire $I(u_n)$ sous forme matricielle :

$$I(u_n) = J(U) = \frac{1}{2} U^\star \cdot A \cdot U - U^\star \cdot B$$

La condition $\partial J(U) = 0$ nous amène à :

$$A \cdot U - B = 0$$

Si la matrice A est inversible, on en déduit :

$$U = A^{-1} \cdot B$$

ce qui nous donne U et donc u_n.

103.5 Valeurs propres

Soit un espace vectoriel E et les formes bilinéaires $a, b : E \times E \mapsto \mathbb{R}$. Nous supposons que a, b sont symétriques et définies positives. Nous allons tenter de minimiser la fonctionnelle :

$$I(v) = \langle v, a, v \rangle$$

sur l'ensemble :

$$\Omega = \{v \in E : \langle v, b, v \rangle = 1\}$$

L'espace Ω n'est malheureusement pas un sous-espace vectoriel. Par exemple, si u appartient à Ω, on a :

$$\langle 2u, b, 2u \rangle = 4 \langle u, b, u \rangle = 4 \neq 1$$

Donc $2u \notin \Omega$. Par conséquent, nous ne pouvons pas utiliser les techniques de projection sur un espace vectoriel pour résoudre ce problème. Nous allons donc employer les techniques de minimisation sous contraintes. Définissons le lagrangien :

$$\mathfrak{L}(v, y) = \langle v, a, v \rangle + y \cdot (1 - \langle v, b, v \rangle)$$

pour tout $(v, y) \in E \times \mathbb{R}$. La fonction u minimisant I sur Ω peut dès lors s'obtenir via le point de selle (u, λ) :

$$\mathfrak{L}(u, y) \leq \mathfrak{L}(u, \lambda) \leq \mathfrak{L}(v, \lambda)$$

pour tout $(v, y) \in E \times \mathbb{R}$. Nous allons évaluer le minimum sur v en utilisant la méthode des variations. Soit $w \in E$ et $\epsilon \in \mathbb{R}$. On pose :

$$\begin{aligned} J_v(\epsilon) &= \mathfrak{L}(u + \epsilon \cdot w, \lambda) \\ &= \langle u + \epsilon \cdot w, a, u + \epsilon \cdot w \rangle + \lambda \cdot [1 - \langle u + \epsilon \cdot w, b, u + \epsilon \cdot w \rangle] \end{aligned}$$

Utilisant les propriétés de a, b et la contrainte $\langle u, b, u \rangle = 1$, il vient :

$$\begin{aligned} J_v(\epsilon) = \langle u, a, u \rangle + 2\epsilon \cdot \langle w, a, u \rangle + \epsilon^2 \cdot \langle w, a, w \rangle \\ -\lambda \cdot [2\epsilon \cdot \langle w, b, u \rangle + \epsilon^2 \cdot \langle w, b, w \rangle] \end{aligned}$$

Les propriétés du point de selle en (u, λ) nous garantissent que :

$$J_v(0) \leq J_v(\epsilon)$$

pour tout $\epsilon \in \mathbb{R}$. La dérivée :

$$\frac{dJ_v}{d\epsilon}(\epsilon) = 2\langle w, a, u\rangle + 2\epsilon \cdot \langle w, a, w\rangle - 2\lambda \cdot [\langle w, b, u\rangle + \epsilon \cdot \langle w, b, w\rangle]$$

doit donc s'annuler en $\epsilon = 0$:

$$\frac{dJ_v^\lambda}{d\epsilon}(0) = 2\langle w, a, u\rangle - 2\lambda \cdot \langle w, b, u\rangle = 0$$

c'est-à-dire :

$$\langle w, a, u\rangle = \lambda \cdot \langle w, b, u\rangle$$

pour tout $w \in E$. On dit alors que λ est valeur propre de a, b correspondant à la fonction propre u.

Propriétés extrémales

Si nous choisissons $w = u$, nous obtenons :

$$\langle u, a, u\rangle = \lambda \cdot \langle u, b, u\rangle = \lambda$$

La valeur propre est donc la valeur minimale de I sur Ω :

$$\lambda = \langle u, a, u\rangle = \min_{v\in\Omega} \langle v, a, v\rangle$$

Rapport de Rayleigh

Supposons que b soit strictement définie positive. Posons $x = \alpha \cdot u$, pour un certain $\alpha \in \mathbb{R}$ non nul. On a $\langle x, a, x\rangle = \alpha^2 \cdot \langle u, a, u\rangle$ et $\langle x, b, x\rangle = \alpha^2 \cdot \langle u, b, u\rangle = \alpha^2$. Donc :

$$\lambda = \langle u, a, u\rangle = \frac{\langle x, a, x\rangle}{\langle x, b, x\rangle}$$

Considérons la même expression pour un quelconque $z \in E$ non nul :

$$R(z) = \frac{\langle z, a, z\rangle}{\langle z, b, z\rangle}$$

On dit que $R(z)$ est le rapport de Rayleigh de z. Soit $v = z/\sqrt{\langle z, b, z\rangle}$. On a :

$$\langle v, a, v\rangle = \frac{\langle z, a, z\rangle}{\langle z, b, z\rangle} = R(z)$$

et :

$$\langle v, b, v\rangle = \frac{\langle z, b, z\rangle}{\langle z, b, z\rangle} = 1$$

ce qui prouve que $v \in \Omega$. On a donc :

$$\lambda = R(x) \le \langle v, a, v\rangle = R(z)$$

On en conclut que x minimise le rapport de Rayleigh sur $E_0 = E \setminus \{0\}$:

$$\lambda = R(x) = \arg\min_{z\in E_0} R(z) = \arg\min_{z\neq 0} R(z)$$

Comme $\Omega \subseteq F$, on en conclut que la réciproque est vraie : si x minimise R sur E_0, le vecteur $x/\sqrt{\langle x, b, x\rangle}$ minimise la fonctionnelle I sur Ω. Les deux problèmes de minimisation sont donc équivalents.

Suite de vecteurs propres

On peut en fait définir une suite de vecteurs et de valeurs propres. Soit $\Omega_1 = \Omega$ et $e_1 = u$. On pose récursivement :

$$\Omega_{n+1} = \{v \in \Omega : \langle v, b, e_1 \rangle = ... = \langle v, b, e_n \rangle = 0\}$$

et :

$$e_{n+1} \in \arg \min_{v \in \Omega_{n+1}} \langle v, a, v \rangle$$

On aura alors bien sûr $\Omega_{n+1} \subseteq \Omega_n \subseteq ...\Omega_1$. Les propriétés extrémales des valeurs propres nous montrent que $\lambda_{n+1} \ge \lambda_n \ge ... \ge \lambda_1$.

Discrétisation

On pose :

$$u_n = \sum_{i=1}^{n} U_i \cdot \varphi_i$$

On a alors :

$$\langle u_n, a, u_n \rangle = \sum_{i,j=1}^{n} U_i \cdot \langle \varphi_i, a, \varphi_j \rangle \cdot U_j$$
$$\langle u_n, b, u_n \rangle = \sum_{i,j=1}^{n} U_i \cdot \langle \varphi_i, b, \varphi_j \rangle \cdot U_j$$

Avec les matrices :

$$A = [\langle \varphi_i, a, \varphi_j \rangle]_{i,j}$$
$$B = [\langle \varphi_i, b, \varphi_j \rangle]_{i,j}$$

le lagrangien peut s'écrire :

$$\mathfrak{L}(u_n, y) = U^\star \cdot A \cdot U + y \cdot [1 - U^\star \cdot B \cdot U]$$

Les conditions :

$$\partial_v \mathfrak{L}(u_n, \lambda) = 2A \cdot U - 2\lambda \cdot B \cdot U = 0$$
$$\partial_y \mathfrak{L}(u_n, \lambda) = 1 - U^\star \cdot B \cdot U = 0$$

nous amènent au problème :

$$A \cdot U = \lambda \cdot B \cdot U$$
$$U^\star \cdot B \cdot U = 1$$

à résoudre en U. Notons que si on choisit les φ_i « orthonormées » dans le sens du pseudo-produit scalaire introduit par b, on a :

$$\langle \varphi_i, b, \varphi_j \rangle = \delta_{ij}$$

et $B = I$. On peut par exemple construire une telle suite de φ_i en utilisant la méthode de Gram-Schmidt. Le problème se simplifie alors en :

$$A \cdot U = \lambda \cdot U$$
$$U^\star \cdot U = 1$$

Chapitre 104

Algorithmes d'optimisation contrainte

104.1 Introduction

Nous allons présenter des algorithmes permettant de résoudre approximativement des problèmes de minimisation d'une fonction φ sur un ensemble $\Omega \subseteq \mathbb{R}^n$. Ces algorithmes partent d'un point initial $x_0 \in \Omega$ et itèrent schématiquement comme suit :

$$x_{k+1} = I(x_k) = x_k + p_k$$

pour un certain $p_k \in \mathbb{R}^n$. On espère bien entendu que la suite converge et que :

$$x_N \approx \lim_{k \to \infty} x_k = \arg\min_{x \in \Omega} \varphi(x)$$

pour N assez grand. Nous adoptons les notations :

$$J = \partial\varphi$$

pour le gradient, de taille $(n, 1)$ et :

$$H = \partial^2\varphi$$

pour la hessienne, de taille (n, n). On note également :

$$\Phi_k = \Phi(x_k)$$

pour toute fonction Φ (par exemple, $\Phi \in \{\varphi, J, H\}$).

104.2 Contraintes linéaires

Nous considérons le cas de contraintes linéaires :

$$\Omega = \{x \in \mathbb{R}^n : A \cdot x = b\}$$

où A est une matrice de taille (m, n) et b un vecteur de taille $(m, 1)$. On choisit généralement $x_0 = 0$. L'itération k part d'un x_k satisfaisant les contraintes :

$$A \cdot x_k = b$$

Si on veut que $x_{k+1} = x_k + s_k \in \Omega$, il est nécessaire et suffisant que $A \cdot (x_k + s_k) = b$, ou que :

$$A \cdot s_k = A \cdot (x_k + s_k) - A \cdot x_k = b - b = 0$$

On doit donc avoir $s_k \in \ker A$. Soit les $u_i \in \mathbb{R}^m$, $v_i \in \mathbb{R}^n$ et $\sigma_i \in \mathbb{R}$ constituant la décomposition en valeurs singulières de A. On forme une matrice à partir des vecteurs de base de $\ker A$:

$$V = [v_{r+1} \dots v_n]$$

On a alors :

$$s_k = \sum_{i=r+1}^{n} p_{k,i} \cdot v_i = V \cdot p_k$$

où $p_k = [p_{k,r+1} \dots p_{k,n}]^\star$.

Newton projeté

Il s'agit de l'adaptation de la méthode de Newton :

$$x_{k+1} = x_k + V \cdot p_k$$

On minimise le développement d'ordre deux :

$$\varphi_{k+1} \approx \varphi_k + J_k^\star \cdot V \cdot p_k + \frac{1}{2} \cdot p_k^\star \cdot V^\star \cdot H_k \cdot V \cdot p_k$$

L'annulation du gradient par rapport à p_k nous donne :

$$V^\star \cdot J_k + V^\star \cdot H_k \cdot V \cdot p_k = 0$$

et donc :

$$p_k = -(V^\star \cdot H_k \cdot V)^{-1} \cdot V^\star \cdot J_k$$

104.3 Contraintes d'inégalité

Examinons à présent le cas où :

$$\Omega = \{x \in \mathbb{R}^n : \omega(x) \le 0\}$$

Nous laissons à présent tomber le k des itérations, afin d'alléger les notations.

Méthodes de penalité

La méthode de pénalité consiste à ajouter une fonction positive à φ lorsque x sort de Ω. Si l'on augmente la valeur de la fonction pénalité, on rapproche alors le minimum global de Ω. La solution à notre problème contraint devrait donc être approchée en faisant tendre l'amplitude de la pénalité vers l'infini. Soit la fonction $\varpi : \mathbb{R}^n \mapsto \mathbb{R}^m$ définie par :

$$\varpi_i(x) = \max\{\omega_i(x), 0\}$$

On a par construction $\varpi(x) = 0$ pour tout $x \in \Omega$ et $\varpi(x) > 0$ pour tout $x \notin \Omega$. On ajoute la somme des carrés $\varpi(x)^\star \cdot \varpi(x) = \sum_i \varpi_i(x)^2$ multipliée par un paramètre réel $k \ge 0$ à la fonction objectif pour obtenir l'objectif modifié :

$$\psi_k(x) = \varphi(x) + k \cdot \varpi(x)^\star \cdot \varpi(x)$$

On utilise ensuite un algorithme de minimisation libre pour évaluer le minimum global :

$$\mu(k) = \arg \min_{x \in \mathbb{R}^n} \psi_k(x)$$

On espère que μ converge à l'infini et que :

$$\lim_{k \to \infty} \mu(k) = \arg \min_{x \in \Omega} \varphi(x)$$

Il suffit dans ce cas de choisir k assez grand pour obtenir une estimation de la solution du problème contraint.

Chapitre 105

Réseaux de neurones

105.1 Définition

Commençons par la description d'un neurone i de fonction caractéristique σ. La relation entre les entrées x_j et la sortie y_i s'écrit :

$$y_i = \sigma \left(\sum_j (w_{ij}\ x_j) + b_i \right)$$

Un réseau de neurones est composé de neurones reliés entres eux (la sortie d'un neurone peut servir d'entrée à un autre).

La fonction caractéristique est généralement l'une de celles décrites ci-dessous :

$$\begin{gathered}
\sigma(x) = \operatorname{sign}(x) \\
\sigma(x) = \tanh(x) \\
\sigma(x) = \exp(-x^2) \\
\sigma(x) = x \exp(-x^2)
\end{gathered}$$

105.2 Perceptron à une couche

Le perceptron monocouche est composé d'une rangée de neurones reliant les entrées x_i aux sorties y_i (le perceptron multicouche est composé de monocouches assemblées l'une à la suite de l'autre). La relation entrées-sorties s'écrit :

$$y_i = P_\theta(x) = c + \sum_j v_j\ \sigma \left(\sum_k (w_{jk}\ x_k) + b_j \right)$$

On écrit $y_i = P_\theta(x)$ pour mettre en évidence l'influence des paramètres du résau $\theta = (v, w, b, c)$ sur la sortie y.

105.3 Entrainement

Les réseaux de neurones sont principalement utilisés afin de calquer le comportement d'un système difficille à modéliser par d'autres méthodes. On dispose d'un certain nombre de vecteurs $y^{(n)} \in \mathbb{R}^M$, $x^{(n)} \in \mathbb{R}^N$, où $n = 1, ..., K$. On aimerait bien trouver le vecteur des paramètres, θ, qui correspond le mieux à cette série d'entrées-sorties. On va alors entrainer le réseau de neurones défini par $y = R_\theta(x)$ en utilisant une méthode d'optimisation non linéaire afin d'obtenir la solution de

$$\theta^* = \arg\min_\theta \sum_n \left\| y^{(n)} - R_\theta\left(x^{(n)}\right) \right\|^2$$

Dixième partie

Equations différentielles

Chapitre 106

Equations différentielles ordinaires

106.1 Fonctions Lipschitziennes

Les fonctions Lipschitziennes sont des fonctions à variations bornées :

$$\operatorname{Lip}(A, B) = \{f \in \mathbb{F}(A, B) : \exists L \in \mathbb{R} : \forall x, y \in A : \|f(x) - f(y)\| \leq L \|x - y\|\}$$

106.2 Problème différentiel d'ordre un

Soit $f \in \operatorname{Lip}(\mathbb{R} \times \mathbb{R}^n, \mathbb{R})$, et l'application A définie par :

$$A(u)(t) = u_0 + \int_0^t f(s, u(s)) \, ds$$

pour toute fonction $u : U \subseteq \mathbb{R} \mapsto \mathbb{R}$.
On peut montrer que A est contractante pour la distance définie pour toutes fonctions u, v par :

$$\mathfrak{dist}(u, v) = \sup_{a \leq t \leq b} \|u(t) - v(t)\|$$

où $\|.\|$ est la normee usuelle sur $\mathbb{R}^n$:

$$\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$$

Cette application admet donc un unique point fixe u tel que :

$$u = A(u)$$

En dérivant cette relation, on obtient :

$$\frac{du}{dt}(t) = \frac{d}{dt} A(u)(t) = \frac{d}{dt} \int_0^t f(s, u(s)) \, ds$$

La dérivée de l'intégrale vaut $f(t, u(t)$ et on a :

$$\frac{du}{dt}(t) = f(t, u(t))$$

On a aussi :

$$u(0) = u_0 + \int_0^0 f(s, u(s)) \, ds = u_0$$

Notre point fixe u est donc solution du problème différentiel :

$$\begin{aligned}\frac{du}{dt}(t) &= f(t,u(t)) \\ u(0) &= u_0\end{aligned}$$

Inversément, en intégrant la première équation ci-dessous entre 0 et t, on obtient la relation :

$$u(t) - u_0 = \int_0^t f(s, u(s))\ ds$$

autrement dit :

$$u(t) = u_0 + \int_0^t f(s, u(s))\ ds = A(u)$$

Toute solution du problème différentiel est donc point fixe de A. Comme ce point fixe est unique, on en conclut que la solution du problème différentiel l'est aussi.

106.2.1 Convergence

La série des $u^{(n)}$ définie par :

$$u^{(n)}(t) = A\big(u^{(n-1)}\big) = u_0 + \int_0^t f\big(s, u^{(n-1)}(s)\big)\ ds$$

converge au sens de la distance sup définie ci-dessus vers la solution de ce problème différentiel :

$$\lim_{n\to+\infty} \sup_{a\le t\le b} \left\| u(t) - u^{(n)}(t) \right\| = 0$$

106.2.2 Notation

Pour toute fonction :

$$u : \mathbb{R} \mapsto \mathbb{R},\ t \mapsto u(t)$$

on note aussi :

$$\dot{u} = \frac{du}{dt}$$
$$\ddot{u} = \frac{d^2u}{dt^2}$$

ou :

$$u'(t) = \frac{du}{dt}(t)$$
$$u''(t) = \frac{d^2u}{dt^2}(t)$$

106.3 Problème différentiel d'ordre quelconque

Soit les fonctions :

$$a_0, a_1, ..., a_{n-1} : \mathbb{R} \mapsto \mathbb{R}$$

et une solution $u : \mathbb{R} \mapsto \mathbb{R}$ du problème différentiel d'ordre n :

$$\begin{aligned} a_0(t) \cdot u(t) + \sum_{i=1}^{n-1} a_i(t) \cdot \frac{d^i u}{dt^i}(t) &= 0 \\ u(0) &= U_0 \\ \frac{d^i u}{dt^i}(0) &= U_i \qquad (i = 1, ..., n-1) \end{aligned}$$

Ce problème peut se ramener à un problème différentiel d'ordre un. Pour cela on définit la fonction $v : \mathbb{R} \mapsto \mathbb{R}^n$ par :

$$v(t) = \Big(v_i(t)\Big)_{i=0,...,n-1} = \left(\frac{d^i u}{dt^i}(t)\right)_{i=0,...,n-1}$$

pour tout $t \in \mathbb{R}$. On voit alors que :

$$\frac{dv}{dt} = \begin{bmatrix} 0 & 1 & 0 & \dots & \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & & & & \\ 0 & \dots & \dots & 0 & 1 \\ -a_0/a_n & -a_1/a_n & \dots & \dots & -a_{n-1}/a_n \end{bmatrix} \cdot v$$

La condition initiale s'écrit :

$$v(0) = (U_i)_{i=0,...,n-1}$$

Comme il existe une et une seule solution v au problème d'ordre un associé, le problème différentiel d'ordre n admet également une unique solution :

$$u : t \mapsto v_0(t)$$

106.4 Problème aux limites

Soit la fonctionnelle $I : \mathrm{Cont}^2([a, b], \mathbb{R}) \mapsto \mathbb{R}$. Nous allons tenter de minimiser I sur l'ensemble :

$$\mathcal{F} = \{u \in \mathrm{Cont}^2([a, b], \mathbb{R}) : u(a) = U_1, \ u(b) = U_2\}$$

On voit que les valeurs de toute fonction $u \in \mathcal{F}$ sont contraintes aux extrémités du domaine de u. On parle dans ce cas de problème aux limites.

Soit $u \in \mathcal{F}$ et la fonction w appartenant à l'ensemble :

$$\mathcal{W} = \{w \in \mathrm{Cont}^2([a, b], \mathbb{R}) : w(a) = w(b) = 0\}$$

On a alors $u + \epsilon \cdot w \in \mathcal{F}$ pour tout $w \in \mathcal{W}$ et tout $\epsilon \in \mathbb{R}$, car :

$$u(a) + \epsilon \cdot w(a) = u(a) = U_1$$
$$u(b) + \epsilon \cdot w(b) = u(b) = U_2$$

On considère la famille de fonctions $J_w : \mathbb{R} \mapsto \mathbb{R}$ définies par :

$$J_w(\epsilon) = I(u + \epsilon \cdot w)$$

pour tout $w \in \mathcal{W}$ et tout $\epsilon \in \mathbb{R}$. Si u minimise I sur $\mathcal{F}$, on a évidemment $J_w(\epsilon) \ge J_w(0)$ pour tout $w \in \mathcal{W}$ et pour tout $\epsilon \in \mathbb{R}$. Il est par conséquent nécessaire que la condition de stationarité :

$$\frac{dJ_w}{d\epsilon}(0) = 0$$

soit satisfaite pour tout $w \in \mathcal{W}$.

106.4.1 Équation d'Euler

Un exemple typique de fonctionnelle I est définie pour tout $u \in \text{Cont}^2([a,b],\mathbb{R})$ par :

$$I(u) = \int_a^b f(t,u(t),u'(t))\ dt$$

avec :

$$f \in \text{Cont}^2([a,b]\times\mathbb{R}^2,\mathbb{R}),\ \ (t,u,v) \mapsto f(t,u,v)$$

On a :

$$J_w(\epsilon) = \int_a^b f(t, u+\epsilon\cdot w, u'+\epsilon\cdot w')\ dt$$

Donc :

$$\frac{dJ_w}{d\epsilon}(0) = \int_a^b \left(\frac{\partial f}{\partial u}(t,u,u')\cdot w + \frac{\partial f}{\partial v}(t,u,u')\cdot w' \right) dt = 0$$

Nous allons tenter d'intégrer par parties le deuxième terme du membre de droite. On sait que :

$$\frac{d}{dt}\left[\frac{\partial f}{\partial v}(t,u(t),u'(t))\ w\right] = \left[\frac{\partial^2 f}{\partial t\partial v} + \frac{\partial^2 f}{\partial u\partial v}\ u' + \frac{\partial^2 f}{\partial v\partial v}\ u''\right]\ w + \frac{\partial f}{\partial v}(t,u,u')\ w'$$

En utilisant le théorème fondamental et les conditions sur w, on arrive à :

$$\int_a^b \frac{d}{dt}\left[\frac{\partial f}{\partial v}\ w\right] dt = \frac{\partial f}{\partial v}(b,u(b),u'(b))\ w(b) - \frac{\partial f}{\partial v}(a,u(a),u'(a))\ w(a) = 0$$

On en déduit que :

$$\int_a^b \frac{\partial f}{\partial v}(t,u,u')\ w'\ dt = -\int_a^b \left(\frac{\partial^2 f}{\partial t\partial u} + \frac{\partial^2 f}{\partial u\partial u}\ u' + \frac{\partial^2 f}{\partial u\partial u}\ u''\right)\ w\ dt$$

La condition de stationarité sur J_w devient alors :

$$\int_a^b \left(\frac{\partial f}{\partial u} - \frac{\partial^2 f}{\partial t\partial u} - \frac{\partial^2 f}{\partial u\partial u}\ u' - \frac{\partial^2 f}{\partial u\partial u}\cdot u''\right)\ w\ dt = 0$$

Comme cette équation est valable pour tout les w dans $\mathcal{W}$, on en déduit que u vérifie l'équation différentielle :

$$\frac{\partial f}{\partial u} - \frac{\partial^2 f}{\partial t\partial u} - \frac{\partial^2 f}{\partial u\partial u}\ u' - \frac{\partial^2 f}{\partial u\partial u}\cdot u'' = 0$$

où les dérivées de f sont bien entendu évaluées en $(t,u(t),u'(t))$.

106.5 Sturm-Liouville

Nous considérons à présent une application importante du théorème de Lax-milgram. Soit l'espace :

$$F = \{u \in \mathrm{Cont}^2([a,b],\mathbb{R}) : u(a) = u(b) = 0\}$$

et les fonctions :

$$p, q, f : \mathbb{R} \mapsto \mathbb{R}$$

Considérons la fonctionnelle $\mathcal{L} : F \mapsto \mathrm{Cont}([a,b],\mathbb{R})$ définie par :

$$\mathcal{L}(u) = \frac{d}{dt}\left[p \cdot \frac{du}{dt}\right] - q \cdot u + f$$

et le problème différentiel avec conditions aux limites associé :

$$\begin{aligned} \mathcal{L}(u) &= 0 \\ u(a) = u(b) &= 0 \end{aligned}$$

Choisissons $v \in F$ et intégrons l'équation :

$$\mathcal{L}(u) \cdot v = 0$$

sur $[a,b]$. On obtient :

$$-\int_a^b \frac{d}{dt}\left[p(t)\ \frac{du}{dt}(t)\right]\ v(t)\ dt + \int_a^b q(t)\ u(t)\ v(t)\ dt = \int_a^b f(t)\ v(t)\ dt$$

Nous allons tenter d'intégrer par parties. On a :

$$\frac{d}{dt}\left[p(t)\ \frac{du}{dt}(t)\ v(t)\right] = \frac{d}{dt}\left[p(t)\ \frac{du}{dt}(t)\right]\ v(t) + p(t)\ \frac{du}{dt}(t)\ \frac{dv}{dt}(t)$$

En appliquant le théorème fondamental, on obtient :

$$\int_a^b \frac{d}{dt}\left[p(t)\ \frac{du}{dt}(t)\ v(t)\right]\ dt = p(b)\ \frac{du}{dt}(b)\ v(b) - p(a)\ \frac{du}{dt}(a)\ v(a) = 0$$

On en conclut que :

$$\int_a^b \frac{d}{dt}\left[p(t)\ \frac{du}{dt}(t)\right]\ v(t)\ dt = -\int_a^b p(t)\ \frac{du}{dt}(t)\ \frac{dv}{dt}(t)\ dt$$

L'intégrale de $\mathcal{L}(u) \cdot v = 0$ devient :

$$\int_a^b \left[p(t)\ \frac{du}{dt}(t)\ \frac{dv}{dt}(t) + q(t)\ u(t)\ v(t)\right]\ dt = \int_a^b f(t)\ v(t)\ dt$$

Donc, si on définit

$$\begin{aligned} a(u,v) &= \int_a^b \left[p(t) \cdot \frac{du}{dt}(t) \cdot \frac{dv}{dt}(t) + q(t) \cdot u(t) \cdot v(t)\right]\ dt \\ b(v) &= \int_a^b f(t) \cdot v(t)\ dt \end{aligned}$$

on a :

$$a(u, v) = b(v)$$

pour tout $v \in F$. En appliquant le théorème de Lax-Milgram, on en déduit que la solution du probléme :

$$\begin{aligned} \mathcal{L}(u) &= 0 \\ u(a) = u(b) &= 0 \end{aligned}$$

minimise sur F la fonctionnelle I définie pour toute fonction $v \in \mathrm{Cont}^2([a, b], \mathbb{R})$ par :

$$I(v) = \int_a^b \left[p(x) \ \left(\frac{dv}{dx}(x) \right)^2 + q(x) \ v(x)^2 \right] \ dx - \int_a^b f(x) \ v(x) \ dx$$

106.6 Séparation des variables

Soit les fonctions :

$$a, b : \mathbb{R} \mapsto \mathbb{R}$$

et $f : \mathbb{R}^2 \mapsto \mathbb{R}$ définie par :

$$f(t, u) = a(t) \cdot b(u)$$

Si f est lipschitzienne, le probléme différentiel :

$$\begin{aligned} \frac{du}{dt}(t) &= f(t, u(t)) = a(t) \cdot b\big(u(t)\big) \\ u(0) &= u_0 \end{aligned}$$

admet une unique solution. En faisant passer u dans le premier membre et dt dans le second, la première équation peut se réécrire symboliquement :

$$\frac{du}{a(u)} = b(t) dt$$

En intégrant les deux membres, il vient :

$$\int_{u(0)}^{u(s)} \frac{du}{A(u)} = \int_0^s B(t) dt$$

106.7 Dérivées des fonctions usuelles

106.7.1 Arcsinus

AFAIRE : ARRANGER LA FIN DU CHAPITRE
Soit la relation :

$$\begin{gathered} y = \sin(x) \\ x = \arcsin(y) \end{gathered}$$

Comme

$$\frac{dy}{dx} = \cos(x)$$
$$\sin(x)^2 + \cos(x)^2 = 1$$

on a

$$\frac{dy}{dx} = \sqrt{1-y^2}$$
$$\frac{d}{dy}\arcsin(y) = \frac{1}{\sqrt{1-y^2}}$$

Table

$$\frac{d\tan(x)}{dx} = 1 + \tan(x)^2$$
$$\frac{d\arcsin(x)}{dx} = \frac{1}{\sqrt{1-x^2}}$$
$$\frac{d\arccos(x)}{dx} = -\frac{1}{\sqrt{1-x^2}}$$
$$\frac{d\arctan(x)}{dx} = \frac{1}{1+x^2}$$
$$\cos(x) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} x^{2k}$$
$$\sin(x) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}$$

106.7.2 Fonctions usuelles

Le théorème fondamental appliqué aux dérivées des fonctions usuelles du chapitre **??** nous permet d'obtenir les résultats suivants :

$$\int_0^x \frac{1}{\sqrt{1-\xi^2}} d\xi = \arcsin(x)$$
$$\int_0^x \frac{1}{1+\xi^2} d\xi = \arctan(x)$$

Chapitre 107

Exponentielle

107.1 Dépendances

— Chapitre 106 : Équations différentielles ordinaires

107.2 Introduction

L'exponentielle est définie comme l'unique solution $\exp : \mathbb{R} \mapsto \mathbb{R}$ du problème différentiel :

$$\begin{aligned} \frac{d\exp}{dt}(t) &= \exp(t) \\ \exp(0) &= 1 \end{aligned}$$

107.3 Développement de Taylor

On a :

$$\frac{d\exp}{dt}(0) = \exp(0) = 1$$

On montre par récurrence que :

$$\frac{d^k\exp}{dt^k}(0) = \frac{d^{k-1}\exp}{dt^{k-1}}(0) = 1$$

Le développement de Taylor autour de 0 s'écrit donc :

$$\exp(t) = \sum_{k=0}^{+\infty} \frac{t^k}{k!}$$

107.4 Additivité

Soit $t \in \mathbb{R}$. On remarque que les applications $f, g : \mathbb{R} \mapsto \mathbb{R}$ définies par :

$$\begin{aligned} f &: s \mapsto \exp(s+t) \\ g &: s \mapsto \exp(s) \cdot \exp(t) \end{aligned}$$

pour tout $s \in \mathbb{R}$ vérifient :

$$\begin{aligned}\partial f(s) &= \exp(s+t) = f(s) \\ f(0) &= \exp(0+t) = \exp(t)\end{aligned}$$

et :

$$\begin{aligned}\partial g(s) &= \exp(s) \cdot \exp(t) = g(s) \\ g(0) &= \exp(0) \cdot \exp(t) = 1 \cdot \exp(t) = \exp(t)\end{aligned}$$

Par unicité de la solution en u du problème différentiel :

$$\begin{aligned}\partial u(s) &= u(s) \\ u(0) &= \exp(t)\end{aligned}$$

on en déduit que :

$$\exp(s+t) = \exp(s) \cdot \exp(t)$$

107.5 Miroir

On déduit de l'additivité que :

$$1 = \exp(0) = \exp(t-t) = \exp(t) \cdot \exp(-t)$$

pour tout $t \in \mathbb{R}$. On en conclut que :

$$\exp(-t) = \frac{1}{\exp(t)}$$

107.6 Limites

On a :

$$\lim_{t \to +\infty} \exp(t) = \lim_{t \to +\infty} (1 + t + \frac{t^2}{2} + ...) \ge \lim_{t \to +\infty} t = +\infty$$

La limite à l'infini positif est donc infinie :

$$\lim_{t \to +\infty} \exp(t) = +\infty$$

En utilisant le changement de variable $t = -s$, on obtient la limite à l'infini négatif :

$$\lim_{s \to -\infty} \exp(s) = \lim_{t \to +\infty} \exp(-t) = \lim_{t \to +\infty} \frac{1}{\exp(t)} = 0$$

107.7 Image

Si $t \ge 0$ il est clair que $\exp(t) > 0$ puisqu'il s'agit d'une somme infinie de termes strictement positifs. Si $s \le 0$, on a $t = -s \ge 0$ et :

$$\exp(s) = \exp(-t) = \frac{1}{\exp(t)} \ > \ 0$$

On en conclut que :

$$\exp : \mathbb{R} \mapsto \mathbb{R}^+ \setminus \{0\}$$

Comme la fonction exp est continue et croît avec t sur $\mathbb{R}$ de :

$$\lim_{t \to -\infty} \exp(t) = 0$$

jusqu'à :

$$\lim_{t \to +\infty} \exp(t) = +\infty$$

on a :

$$\exp(\mathbb{R}) =]0, +\infty[= \mathbb{R}^+ \setminus \{0\}$$

107.7.1 Réels positifs

Comme la fonction exp est continue et croît avec t sur $\mathbb{R}^+$ de :

$$\exp(0) = 1$$

jusqu'à :

$$\lim_{t \to +\infty} \exp(t) = +\infty$$

on a :

$$\exp(\mathbb{R}^+) = [1, +\infty[$$

107.7.2 Réels négatifs

Comme la fonction exp est continue et croît avec s sur $\mathbb{R}^-$ de :

$$\lim_{s \to -\infty} \exp(s) = 0$$

jusqu'à :

$$\exp(0) = 1$$

on a :

$$\exp(\mathbb{R}^-) =]0, 1]$$

107.8 Intégrale

Comme la fonction exp est une primitive d'elle-même, on a :

$$\int_a^b \exp(t)\ dt = \exp(b) - \exp(a)$$

En faisant tendre a vers $-\infty$, on voit que :

$$\int_{-\infty}^b \exp(t)\ dt = \lim_{a \to -\infty} \Big(\exp(b) - \exp(a) \Big) = \exp(b)$$

Les autres intégrales à bornes infinies sont infinies :

$$\int_{-\infty}^{+\infty} \exp(t)\ dt = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \Big(\exp(b) - \exp(a) \Big) = +\infty$$

$$\int_{a}^{+\infty} \exp(t)\ dt = \lim_{b \to +\infty} \Big(\exp(b) - \exp(a) \Big) = +\infty$$

Chapitre 108

Logarithme

108.1 Introduction

La fonction $\exp : \mathbb{R} \mapsto \mathbb{R}^+ \setminus \{0\}$ étant strictement croissante et d'image égale à $\mathbb{R}^+ \setminus \{0\}$, elle est inversible. On définit le logarithme :

$$\ln : \mathbb{R}^+ \setminus \{0\} \mapsto \mathbb{R}$$

comme la fonction inverse de exp :

$$\ln = \exp^{-1}$$

Pour tout $x, y \in \mathbb{R}$ tels que $y = \exp(x) > 0$, on a donc :

$$\ln(y) = x$$

108.2 Valeurs particulières

Le cas particulier $x = 0$ et $y = \exp(0) = 1$ nous montre que :

$$\ln(1) = 0$$

108.3 Dérivée

Soit les réels x, y tels que :

$$x = \ln(y)$$

On a alors par définition :

$$y = \exp(x)$$

La dérivée de cette relation s'écrit symboliquement :

$$\frac{dy}{dx} = \exp(x) = y$$

Comme la dérivée d'une fonction inverse est l'inverse de la dérivée, on a :

$$\frac{dx}{dy} = \frac{1}{y}$$

c'est-à-dire :

$$\frac{d \ln}{dy}(y) = \frac{1}{y}$$

108.4 Développement de Taylor

Soit la fonction $u : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$u(t) = \ln(1 + t)$$

pour tout réel t. On a :

$$u(0) = \ln(1 + 0) = \ln(1) = 0$$

La dérivée s'écrit :

$$\partial u(t) = \frac{1}{1 + t}$$

et en particulier :

$$\partial u(0) = \frac{1}{1 + 0} = 1$$

On procède de même pour la dérivée seconde :

$$\partial^2 u(t) = -\frac{1}{(1 + t)^2}$$

et en particulier :

$$\partial^2 u(0) = -\frac{1}{(1 + 0)^2} = -1$$

on procède de même pour la dérivée tierce :

$$\partial^3 u(t) = \frac{2}{(1 + t)^3}$$

et en particulier :

$$\partial^3 u(0) = \frac{2}{(1 + 0)^3} = 2$$

on procède de même pour la dérivée quarte :

$$\partial^4 u(t) = \frac{-6}{(1 + t)^4}$$

et en particulier :

$$\partial^4 u(0) = \frac{-6}{(1 + 0)^4} = -6$$

On voit que :

$$\partial^k u(0) = (-1)^{1+k} \cdot (k - 1)!$$

pour tout $k \in \mathbb{N}$ vérifiant $k \geq 1$. Le développement de Taylor s'écrit donc :

$$\ln(1 + t) = \sum_{k=1}^{+\infty} \frac{(-1)^{1+k} \cdot (k - 1)!}{k!} \ t^k$$

Comme $k! = k \cdot (k-1)!$, on a :

$$\frac{(k-1)!}{k!} = \frac{1}{k}$$

Le développement s'écrit donc finalement :

$$\ln(1+t) = \sum_{k=1}^{+\infty} \frac{(-1)^{1+k}}{k} \ t^k$$

En posant $x = 1 + t$, on obtient la forme équivalente :

$$\ln(x) = \sum_{k=1}^{+\infty} \frac{(-1)^{1+k}}{k} \ (x-1)^k$$

108.5 Additivité

Soit $a, b \in \mathbb{R}$. En utilisant l'additivité de l'exponentielle, on obtient :

$$\exp\Big[\ln(a \cdot b)\Big] = a \cdot b = \exp\big[\ln(a)\big] \cdot \exp\big[\ln(b)\big] = \exp\Big[\ln(a) + \ln(b)\Big]$$

En prenant le logarithme de cette égalité, on en déduit que :

$$\ln(a \cdot b) = \ln(a) + \ln(b)$$

108.6 Miroir

Soit $a \in \mathbb{R}$. On a :

$$\ln(a) + \ln\left(\frac{1}{a}\right) = \ln\left(a \cdot \frac{1}{a}\right) = \ln(1) = 0$$

On en conclut que :

$$\ln\left(\frac{1}{a}\right) = -\ln(a)$$

108.7 Soustraction

Soit les réels a, b. On a :

$$\ln\left(\frac{a}{b}\right) = \ln(a) + \ln\left(\frac{1}{b}\right) = \ln(a) - \ln(b)$$

108.8 Intégrale de $x \mapsto 1/x$

Comme ln est une primitive de la fonction :

$$u : \mathbb{R} \setminus \{0\}, \ x \mapsto 1/x$$

On a :

$$\int_a^b \frac{1}{x} \, dx = \ln(b) - \ln(a) = \ln\left[\frac{b}{a}\right]$$

108.9 Gaussienne

Soit les réels γ et u_0. Nous cherchons la fonction $u : \mathbb{R} \mapsto \mathbb{R}$ solution du problème différentiel :

$$\begin{aligned} \frac{du}{dt} &= \gamma \cdot t \cdot u \\ u(0) &= u_0 \end{aligned}$$

en procédant par séparation de variables :

$$\frac{du}{u} = \gamma \cdot t \ dt$$

En intégrant :

$$\int_{u_0}^{u(s)} \frac{du}{u} = \int_0^s \gamma \cdot t \ dt$$

on obtient :

$$\ln(u(s)) - \ln(u_0) = \gamma \cdot \frac{s^2}{2}$$

ou :

$$\ln\left(\frac{u(s)}{u_0}\right) = \gamma \cdot \frac{s^2}{2}$$

En prenant l'exponentielle, on arrive à la solution :

$$u(s) = u_0 \cdot \exp\left(\gamma \cdot \frac{s^2}{2}\right)$$

Une fonction de cette forme est appelée gaussienne.

Chapitre 109

Fonctions hyperboliques

109.1 Introduction

Le cosinus hyperbolique cosh est défini comme étant la composante paire de l'exponentielle. On a donc :

$$\cosh(x) = \frac{1}{2} \Big[\exp(x) + \exp(-x) \Big]$$

pour tout $x \in \mathbb{R}$. Le sinus hyperbolique sinh est défini comme étant la composante impaire de l'exponentielle. On a donc :

$$\sinh(x) = \frac{1}{2} \Big[\exp(x) - \exp(-x) \Big]$$

pour tout $x \in \mathbb{R}$.

109.1.1 Décomposition

On a :

$$\exp = \cosh + \sinh$$

avec :

$$\begin{aligned} \cosh(x) &= \cosh(-x) \\ \sinh(x) &= -\sinh(-x) \end{aligned}$$

pour tout réel x.

109.1.2 Valeurs particulières

On a :

$$\cosh(0) = \frac{1}{2}\Big[\exp(0) + \exp(0) \Big] = \frac{1}{2}(1+1) = 1$$

et :

$$\sinh(0) = \frac{1}{2}\Big[\exp(0) - \exp(0) \Big] = 0$$

109.2 Relation fondamentale

Soit un réel x et :

$$c = \cosh(x)$$
$$s = \sinh(x)$$

Le carré du cosinus hyperbolique se développe en :

$$c^2 = \frac{1}{4} \left[\exp(x)^2 + 2 \ \exp(x) \cdot \exp(-x) + \exp(-x)^2 \right]$$

Comme $\exp(x) \cdot \exp(-x) = 1$, le développement devient :

$$c^2 = \frac{1}{4} \left[\exp(x)^2 + \exp(-x)^2 + 2 \right]$$

Le carré du sinus hyperbolique se développe en :

$$s^2 = \frac{1}{4} \left[\exp(x)^2 - 2 \ \exp(x) \cdot \exp(-x) + \exp(-x)^2 \right]$$

Comme $\exp(x) \cdot \exp(-x) = 1$, le développement devient :

$$s^2 = \frac{1}{4} \left[\exp(x)^2 + \exp(-x)^2 - 2 \right]$$

En soustrayant ces deux équations, on obtient :

$$\begin{aligned} c^2 - s^2 &= \frac{\exp(x)^2 + \exp(-x)^2 + 2 - \exp(x)^2 - \exp(-x)^2 + 2}{4} \\ &= \frac{4}{4} = 1 \end{aligned}$$

On a donc :

$$\cosh(x)^2 - \sinh(x)^2 = 1$$

109.3 Dérivées

Pour tout réel x, on a :

$$\frac{d\cosh}{dx}(x) = \frac{1}{2} \Big[\exp(x) - \exp(-x) \Big] = \sinh(x)$$

et :

$$\frac{d\sinh}{dx}(x) = \frac{1}{2} \Big[\exp(x) + \exp(-x) \Big] = \cosh(x)$$

109.4 Intégrales

Comme sinh est une primitive de cosh, on a :

$$\int_a^b \cosh(x) \ dx = \sinh(b) - \sinh(a)$$

Comme cosh est une primitive de sinh, on a :

$$\int_a^b \sinh(x) \ dx = \cosh(b) - \cosh(a)$$

109.5 Tangente

La tangente hyperbolique tanh est définie par :

$$\tanh(x) = \frac{\sinh x}{\cosh x}$$

pour tout $x \in \mathbb{R}$.

109.5.1 Dérivée

On a :

$$\frac{d \tanh}{dx}(x) = \frac{\cosh(x)}{\cosh(x)} - \frac{\sinh(x) \cdot \sinh(x)}{\cosh(x)^2} = 1 - \tanh(x)^2$$

109.5.2 Problème différentiel

Comme :

$$\tanh(0) = \frac{\sinh(0)}{\cosh(0)} = \frac{0}{1} = 0$$

la tangente hyperbolique est solution $u : \mathbb{R} \mapsto \mathbb{R}$ du problème différentiel :

$$\begin{aligned} \partial u(t) &= 1 - u(t)^2 \\ u(0) &= 0 \end{aligned}$$

vérifié pour tout $t \in \mathbb{R}$.

Chapitre 110

Exponentielle matricielle

110.1 Introduction

Soit une matrice $A \in \mathfrak{M}(\mathbb{R}, n, n)$ et la fonction :

$$E_A : \mathbb{R} \mapsto \in \mathfrak{M}(\mathbb{R}, n, n)$$

définie comme étant l'unique solution de :

$$\frac{dE_A}{dt} = A \cdot E_A$$

$$E_A(0) = I$$

On définit alors l'exponentielle matricielle par :

$$\exp(A) = E_A(1)$$

110.2 Matrice nulle

Dans le cas où $A = 0$, on a :

$$\frac{dE_0}{dt} = 0 \cdot E_0 = 0$$

En intégrant, on voit que :

$$E_0(t) - E_0(0) = \int_0^t 0 \, dt = 0$$

La fonction E_0 est donc constante et vaut $E_0(t) = E_0(0) = I$ pour tout $t \in \mathbb{R}$. On en conclut que :

$$\exp(0) = I$$

110.3 Développement de Taylor

Soit la fonction :

$$u : \mathbb{R} \mapsto \mathbb{M}(\mathbb{R}, n, n), \ t \mapsto \exp(A \cdot t)$$

On a :

$$u(0) = \exp(A \cdot 0) = \exp(0) = I$$

et :

$$\frac{du}{dt}(0) = A \cdot u(0) = A \cdot I = A$$

On montre par récurrence que :

$$\frac{d^k u}{dt^k}(0) = A \cdot \frac{d^{k-1} u}{dt^{k-1}}(0) = A \cdot A^{k-1} = A^k$$

En évaluant le développement de Taylor de u autour de $t = 0$ on obtient :

$$\exp(A \cdot t) = \sum_{k=0}^{+\infty} \frac{1}{k!} \cdot A^k \cdot t^k$$

Évaluons la dérivée de ce développement :

$$\begin{aligned}
\frac{d}{dt} \exp(A \cdot t) &= \sum_{k=1}^{+\infty} \frac{1}{k!} \cdot A^k \cdot k \cdot t^{k-1} \\
&= A \cdot \sum_{k=1}^{+\infty} \frac{1}{(k-1)!} \cdot A^{k-1} \cdot t^{k-1} \\
&= A \cdot \sum_{k=0}^{+\infty} \frac{1}{k!} \cdot A^k \cdot k \cdot t^k \\
&= A \cdot \exp(A \cdot t)
\end{aligned}$$

On a aussi :

$$\exp(A \cdot 0) = I + \sum_{k=1}^{+\infty} \frac{1}{k!} \cdot A^k \cdot 0^k = I$$

La fonction $t \mapsto \exp(A \cdot t)$ est donc identique à la solution E :

$$E(t) = \exp(A \cdot t)$$

110.4 Développement en série de l'exponentielle

Le cas particulier $t = 1$ nous donne le développement de l'exponentielle matricielle :

$$\exp(A) = \sum_{k=0}^{+\infty} \frac{1}{k!} \cdot A^k$$

Sur $\mathbb{R}$

Quand $n = 1$ et $A = 1$, on retrouve le développement de l'exponentielle usuelle :

$$\exp(t) = \sum_{k=0}^{+\infty} \frac{t^k}{k!}$$

110.5 Additivité

Soit les fonctions $f, g : \mathbb{R} \mapsto \mathbb{R}$ définies par :

$$\begin{aligned} f &: s \mapsto \exp\Big(A \cdot (s+t)\Big) \\ g &: s \mapsto \exp(A \cdot s) \cdot \exp(A \cdot t) \end{aligned}$$

pour tout $s \in \mathbb{R}$. On a :

$$\begin{aligned} \partial f(s) &= A \cdot \exp\Big(A \cdot (s+t)\Big) = A \cdot f(s) \\ f(0) &= \exp\Big(A \cdot (0+t)\Big) = \exp(A \cdot t) \end{aligned}$$

et :

$$\begin{aligned} \partial g(s) &= A \cdot \exp(A \cdot s) \cdot \exp(t) = A \cdot g(s) \\ g(0) &= \exp(A \cdot 0) \cdot \exp(A \cdot t) = I \cdot \exp(A \cdot t) = \exp(A \cdot t) \end{aligned}$$

Par unicité de la solution en u du problème différentiel :

$$\begin{aligned} \partial u(s) &= A \cdot u(s) \\ u(0) &= \exp(A \cdot t) \end{aligned}$$

on en déduit que :

$$\exp\Big(A \cdot (s+t)\Big) = \exp(A \cdot s) \cdot \exp(A \cdot t)$$

110.6 Miroir

L'additivité nous dit que :

$$\exp(A \cdot (t-t)) = \exp(A \cdot t) \cdot \exp(A \cdot (-t)) = \exp(A \cdot t) \cdot \exp(-A \cdot t)$$

Mais la condition initiale nous dit que :

$$\exp(A \cdot (t-t)) = \exp(A \cdot 0) = I$$

On a donc :

$$\exp(A \cdot t) \cdot \exp(-A \cdot t) = I$$

Comme les matrices sont carrées, on en déduit que l'inverse matriciel de $\exp(A \cdot t)$ existe et s'écrit :

$$\exp(A \cdot t)^{-1} = \exp(-A \cdot t)$$

Le cas particulier $t = 1$ nous donne :

$$\exp(A)^{-1} = \exp(-A)$$

110.7 Solution vectorielle

Soit un vecteur $x_0 \in \mathbb{R}^n$ et la fonction $x : \mathbb{R} \mapsto \mathbb{R}^n$ définie par :

$$x(t) = \exp(A \cdot t) \cdot x_0$$

pour tout $t \in \mathbb{R}$. On a :

$$\dot{x}(t) = A \cdot \exp(A \cdot t) \cdot x_0 = A \cdot x$$

et :

$$x(0) = \exp(A \cdot 0) \cdot x_0 = I \cdot x_0 = x_0$$

Notre fonction x est donc l'unique solution de l'équation différentielle :

$$\frac{dx}{dt} = A \cdot x$$

$$x(0) = x_0$$

Sur $\mathbb{R}$

Dans le cas où $n = 1$, et $A = 1$, on obtient l'exponentielle usuelle, qui est donc solution de :

$$\begin{aligned} \dot{u} &= u \\ u(0) &= 1 \end{aligned}$$

110.8 Valeurs propres

Il existe un lien entre l'exponentielle d'une matrice hermitienne et ses valeurs propres. Soit la fonction $X : \mathbb{R} \mapsto \mathbb{R}^n$ vérifiant l'équation différentielle :

$$\dot{X}(t) = A \cdot X(t)$$

où A est une matrice carrée hermitienne. Comme $A = A^\star$, on sait que la forme de Schur :

$$A = U \cdot \Lambda \cdot U^\star$$

nous donne une matrice carrée unitaire U qui vérifie par conséquent :

$$U^\star = U^{-1}$$

et une matrice diagonale :

$$\Lambda = (\lambda_i \cdot \delta_{ij})_{i,j}$$

où les λ_i sont les valeurs propres de A. Si on effectue le changement de variable :

$$X = U \cdot Y \quad \Leftrightarrow \quad Y = U^\star \cdot X$$

l'équation différentielle devient :

$$U \cdot \dot{Y} = A \cdot U \cdot Y$$

En multipliant à gauche par $U^\star$, on obtient :

$$\dot{Y} = U^\star \cdot A \cdot U \cdot Y = \Lambda \cdot Y$$

Exprimée en terme de composantes $Y = (y_i)_i$, cette dernière équation devient :

$$\dot{y}_i = \lambda_i \cdot y_i$$

dont la solution est :

$$y_i(t) = y_i(0) \cdot \exp(\lambda_i \cdot t)$$

Comme :

$$\sum_k \frac{1}{k!} \Lambda^k \cdot t^k = \left(\sum_k \frac{1}{k!} \lambda_i^k \cdot t^k \cdot \boldsymbol{\delta}_{ij} \right)_{i,j}$$

on voit que :

$$\exp(\Lambda \cdot t) = \Big(\exp(\lambda_i \cdot t) \cdot \boldsymbol{\delta}_{ij} \Big)_{i,j}$$

On en conclut que :

$$Y(t) = \exp(\Lambda \cdot t) \cdot Y(0)$$

La condition initiale sur Y est liée à celle sur X par :

$$Y(0) = U^\star \cdot X(0)$$

On a donc :

$$Y(t) = \exp(\Lambda \cdot t) \cdot U^\star \cdot X(0)$$

et :

$$X(t) = U \cdot Y(t) = U \cdot \exp(\Lambda \cdot t) \cdot U^\star \cdot X(0)$$

Par définition de l'exponentielle matricielle, on a aussi :

$$X(t) = \exp(A \cdot t) \cdot X(0)$$

On en conclut que :

$$\exp(A \cdot t) \cdot X(0) = U \cdot \exp(\Lambda \cdot t) \cdot U^\star \cdot X(0)$$

Au point $t = 1$, on a :

$$\exp(A) \cdot X(0) = U \cdot \exp(\Lambda) \cdot U^\star \cdot X(0)$$

Cette relation étant vérifiée quelque soit $X(0) \in \mathbb{R}^n$, on en conclut la relation liant l'exponentielle d'une matrice hermitienne à ses valeurs propres :

$$\exp(A) = U \cdot \exp(\Lambda) \cdot U^\star$$

110.9 Dérivée

AFAIRE : dérivée de u(t) = exp(L(t)), arranger la fin du chapitre

$$\frac{du}{dt}(t) = \frac{dL}{dt}(t) \cdot u(t)$$

? ? ? ? ? ? ?

110.10 Intégrale

Soit la fonction R :

$$R : \mathbb{R} \mapsto \mathfrak{M}(\mathbb{R}, n, n)$$

solution de :

$$\begin{aligned} \dot{R}(t) &= L(t) \cdot R(t) \\ R(0) &= I \end{aligned}$$

On vérifie que :

$$R(t) = \exp \int_0^t L(s) ds$$

110.11 Systèmes linéaires

Considérons à présent le problème linéaire suivant :

$$\dot{u}(t) = L(t) \cdot u(t) + b(t)$$

$$u(0) = u_0$$

où on a :

$$\begin{aligned} L &: \mathbb{R} \mapsto \mathfrak{M}(\mathbb{R}, n, n) \\ b &: \mathbb{R} \mapsto \mathfrak{M}(\mathbb{R}, n, 1) \equiv \mathbb{R}^n \\ u &: \mathbb{R} \mapsto \mathfrak{M}(\mathbb{R}, n, 1) \equiv \mathbb{R}^n \end{aligned}$$

Nous allons effectuer un changement de variable afin de résoudre ce problème. Nous supposons par la suite que $R(t)$ est inversible pour tout t. Posons $u = R \cdot x$. On constate tout de suite en utilisant $R(0) = I$ que $x(0) = u_0$. On obtient aussi, en dérivant $u = R \cdot x$:

$$\dot{u} = \dot{R} \cdot x + R \cdot \dot{x} = L \cdot R \cdot x + R \cdot \dot{x}$$

En comparant avec l'équation différentielle dont u est solution :

$$\dot{u} = L \cdot u + b = L \cdot R \cdot x + b$$

on obtient :

$$\begin{gathered} \dot{x} = R^{-1} \cdot b \\ x(0) = u_0 \end{gathered}$$

On en déduit que :

$$x(t) = u_0 + \int_0^t [R(s)]^{-1} \cdot b(s) ds$$

et :

$$u(t) = R(t) \cdot u_0 + \int_0^t R(t) \cdot [R(s)]^{-1} \cdot b(s) ds$$

Dans le cas où $L(t) = L$ ne dépend pas de t, on peut montrer que

$$R(s+t) = R(s) \cdot R(t)$$

en vérifiant que $\varphi(s) = R(s+t)$ et $\psi(s) = R(s) \cdot R(t)$ sont solutions en w de :

$$\frac{dw}{ds}(s) = L \cdot w(s)$$

$$w(0) = R(t)$$

On a alors évidemment $R(-s) = [R(s)]^{-1}$ et

$$u(t) = R(t) \cdot u_0 + \int_0^t R(t-s) \cdot b(s) ds$$

La solution est donc donnée par l'intégrale de convolution de R et b.

110.12 Conditions initiales

Nous allons à présent étudier ce qu'il se passe lorsqu'on dérive la solution par rapport à la condition initiale $u_0 = x$. Posons $u(x,t)$ la solution de :

$$\frac{\partial u}{\partial t}(x,t) = f(t, u(x,t))$$
$$u(x,0) = x$$

Nous allons utiliser la notation

$$u_x(x,t) = \frac{\partial u}{\partial x^T}(x,t)$$

En intervertissant l'ordre de dérivation, on arrive à

$$\frac{\partial u_x}{\partial t}(x,t) = \frac{\partial}{\partial x^T} \frac{\partial u}{\partial t}(x,t)$$
$$\frac{\partial u_x}{\partial t}(x,t) = \frac{\partial f}{\partial u^T}(t, u(x,t)) \cdot u_x(x,t)$$

Par ailleurs, il est clair que :

$$u_x(x,0) = I$$

Utilisant les résultats de la section 110.11 avec :

$$L(t) \mapsto \frac{\partial f}{\partial u^T}(t, u(x,t))$$
$$R(t) \mapsto u_x(x,t)$$

nous obtenons :

$$u_x(x,t) = \exp \int_0^t \frac{\partial f}{\partial u^T}(s, u(x,s)) ds$$

Chapitre 111

Fonctions trigonométriques

111.1 Dépendances

— Chapitre ?? : Les complexes

111.2 Introduction

Les fonctions trigonométriques cosinus (cos), sinus (sin) et associées peuvent se définir à partir des rotations dans le plan $\mathbb{R}^2$. Pour cela, commençons par définir ce qu'est une rotation. Soit la fonction :

$$Q : \mathbb{R} \mapsto \mathbb{M}(\mathbb{R}, 2, 2), \qquad \theta \mapsto Q(\theta)$$

qui associe à chaque valeur de l'angle de rotation $\theta \in \mathbb{R}$ une matrice carrée réelle représentant une rotation dans le plan $\mathbb{R}^2$.

111.3 Angle nul

Il semble logique de demander qu'une rotation d'angle 0 d'un vecteur ne modifie pas ce vecteur, c'est-à-dire :

$$Q(0) = I$$

111.4 Produit scalaire

Une rotation doit conserver les angles entre les vecteurs, ainsi que leur norme. Autrement dit, le produit scalaire sur $\mathbb{R}^2$:

$$\langle x \mid y \rangle = x^\star \cdot y$$

doit être conservé :

$$\langle Q(\theta) \cdot x \mid Q(\theta) \cdot y \rangle = \langle x \mid y \rangle$$

pour tout $x, y \in \mathbb{R}^2$, c'est-à-dire :

$$(Q(\theta) \cdot x)^\star \cdot (Q(\theta) \cdot y) = x^\star \cdot Q(\theta)^\star \cdot Q(\theta) \cdot y = x^\star \cdot y$$

On en déduit que :

$$Q(\theta)^\star \cdot Q(\theta) = I$$

Comme $Q(\theta)$ est carrée, on a :

$$Q(\theta)^\star = Q(\theta)^{-1}$$

où le $^{-1}$ désigne l'inverse matriciel. Quelques calculs suffisent à nous montrer que pour satisfaire cette condition, la forme de la matrice doit être l'une des deux solutions suivantes :

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \qquad \text{ou} \qquad \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}$$

où $\cos, \sin : \mathbb{R} \mapsto \mathbb{R}$ sont des fonctions à déterminer vérifiant la relation fondamentale :

$$\cos(\theta)^2 + \sin(\theta)^2 = 1$$

Mais comme $Q(0) = I$, l'élément $(1, 1)$ de la matrice doit être identique à l'élément $(2, 2)$ et on a :

$$Q(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

avec :

$$\cos(0) = 1$$
$$\sin(0) = 0$$

111.4.1 Relation fondamentale et bornes

Soit $x \in \mathbb{R}$. Comme les fonctions cos et sin sont à valeurs dans $\mathbb{R}$, on a :

$$\cos(x)^2 \geq 0$$
$$\sin(x)^2 \geq 0$$

On déduit de la relation fondamentale :

$$\cos(x)^2 + \sin(x)^2 = 1$$

les inégalités :

$$\cos(x)^2 = 1 - \sin(x)^2 \leq 1$$
$$\sin(x)^2 = 1 - \cos(x)^2 \leq 1$$

On en déduit les bornes :

$$-1 \leq \cos \leq 1$$
$$-1 \leq \sin \leq 1$$

111.5 Différentielle

Posons $x = \cos(\theta)$ et $y = \sin(\theta)$. En différentiant la condition :

$$x^2 + y^2 = 1$$

on obtient :

$$2\ x\ dx + 2\ y\ dy = 0$$

En particulier, si $(x, y) = (1, 0)$, le vecteur (dx, dy) doit être de la forme :

$$dx = 0$$
$$dy = \delta$$

Une rotation infinitésimale doit donc modifier le vecteur $(1, 0)$ en :

$$(1, 0) + (dx, dy) = (1, \delta)$$

On en conclut que :

$$Q(\delta) \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} \approx \begin{bmatrix} 1 \\ \delta \end{bmatrix}$$

Compte tenu des propriétés de symétrie de $Q(\delta)$, on a donc :

$$Q(\delta) \approx \begin{bmatrix} 1 & -\delta \\ \delta & 1 \end{bmatrix}$$

lorsque δ suffisamment proche de 0. Plus précisément, on a :

$$Q(\delta) = \begin{bmatrix} 1 & -\delta \\ \delta & 1 \end{bmatrix} + \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}$$

On impose que les valeurs absolues des composantes de l'erreur convergent plus vite vers zéro que δ :

$$\lim_{\delta \to 0} \frac{|E_{ij}(\delta)|}{\delta} = 0$$

D'après la forme des matrices, il est clair que :

$$E_{11} = E_{22}$$
$$E_{12} = -E_{21}$$

On en déduit qu'il suffit de vérifier la convergence des composantes $(1, 1)$ et $(2, 1)$. La première nous dit que :

$$\lim_{\delta \to 0} \frac{\cos(\delta) - 1}{\delta} = 0$$

et la seconde :

$$\lim_{\delta \to 0} \frac{\sin(\delta) - \delta}{\delta} = 0$$

ou :

$$\lim_{\delta \to 0} \frac{\sin(\delta)}{\delta} = 1$$

111.6 Additivité

Une rotation d'angle θ_1 suivie d'une rotation d'angle θ_2 doit donner le même résultat qu'une rotation directe d'angle $\theta_1 + \theta_2$, ce qui s'écrit :

$$Q(\theta_1 + \theta_2) = Q(\theta_2) \cdot Q(\theta_1)$$

On en déduit directement que :

$$Q(\theta) \cdot Q(-\theta) = Q(\theta - \theta) = Q(0) = I$$

Comme $Q(\theta)$ est carrée, on a :

$$Q(-\theta) = Q(\theta)^{-1} = Q(\theta)^\star$$

Au niveau des composantes, la propriété d'additivité implique que :

$$\begin{bmatrix} \cos(\theta_2) & -\sin(\theta_2) \\ \sin(\theta_2) & \cos(\theta_2) \end{bmatrix} \cdot \begin{bmatrix} \cos(\theta_1) & -\sin(\theta_1) \\ \sin(\theta_1) & \cos(\theta_1) \end{bmatrix} = \begin{bmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{bmatrix}$$

En effectuant le produit matriciel et en comparant composante par composante, on obtient les formules suivantes :

$$\cos(\theta_1+\theta_2) = \cos(\theta_1)\ \cos(\theta_2) - \sin(\theta_1)\ \sin(\theta_2)$$
$$\sin(\theta_1+\theta_2) = \sin(\theta_1)\ \cos(\theta_2) + \cos(\theta_1)\ \sin(\theta_2)$$

111.7 Dérivées

La dérivée de la fonction cos s'écrit :

$$\frac{d\cos}{d\theta}(\theta) = \lim_{\delta\to 0} \frac{\cos(\theta+\delta) - \cos(\theta)}{\delta}$$

En appliquant les formules d'additivité ci-dessus avec $\theta_1 = \theta$ et $\theta_2 = \delta$, et en se rappelant les propriétés des fonctions cos et sin pour des angles $\delta \to 0$, on obtient :

$$\begin{aligned} \frac{d\cos}{d\theta}(\theta) &= \lim_{\delta\to 0} \frac{\cos(\theta)\ \cos(\delta) - \sin(\theta)\ \sin(\delta) - \cos(\theta)}{\delta} \\ &= \lim_{\delta\to 0} \frac{\cos(\theta)\ (\cos(\delta)-1) - \sin(\theta)\ \sin(\delta)}{\delta} \\ &= -\sin(\theta) \end{aligned}$$

La dérivée de la fonction sin s'écrit :

$$\frac{d\sin}{d\theta}(\theta) = \lim_{\delta\to 0} \frac{\sin(\theta+\delta) - \sin(\theta)}{\delta}$$

En procédant comme précédemment, on arrive à :

$$\begin{aligned} \frac{d\sin}{d\theta}(\theta) &= \lim_{\delta\to 0} \frac{\sin(\theta)\ \cos(\delta) + \cos(\theta)\ \sin(\delta) - \sin(\theta)}{\delta} \\ &= \lim_{\delta\to 0} \frac{\sin(\theta)\ (\cos(\delta)-1) + \cos(\theta)\ \sin(\delta)}{\delta} \\ &= \cos(\theta) \end{aligned}$$

On a donc :

$$\frac{d\cos}{d\theta}(\theta) = -\sin(\theta)$$

$$\frac{d\sin}{d\theta}(\theta) = \cos(\theta)$$

111.8 Équations différentielles

La dérivée seconde de la fonction cos s'écrit :

$$\frac{d^2 \cos}{d\theta^2}(\theta) = \frac{d(-\sin)}{d\theta}(\theta) = -\cos(\theta)$$

On a aussi $\cos(0) = 1$ et :

$$\frac{d \cos}{d\theta}(0) = -\sin(0) = 0$$

La fonction cos est donc l'unique solution $u : \mathbb{R} \mapsto \mathbb{R}$ du problème différentiel :

$$\begin{aligned} \partial^2 u(t) &= -u(t) \\ u(0) &= 1 \\ \partial u(0) &= 0 \end{aligned}$$

vérifié pour tout $t \in \mathbb{R}$. La dérivée seconde de la fonction sin s'écrit :

$$\frac{d^2 \sin}{d\theta^2}(\theta) = \frac{d \cos}{d\theta}(\theta) = -\sin(\theta)$$

On a aussi $\sin(0) = 0$ et :

$$\frac{d \sin}{d\theta}(0) = \cos(0) = 1$$

La fonction sin est donc l'unique solution $v : \mathbb{R} \mapsto \mathbb{R}$ du problème différentiel :

$$\begin{aligned} \partial^2 v(t) &= -v(t) \\ v(0) &= 0 \\ \partial v(0) &= 1 \end{aligned}$$

vérifié pour tout $t \in \mathbb{R}$.

111.9 Racines

Nous allons maintenant nous occuper du problème des éventuelles racines des fonctions trigonométriques. Soit :

$$u(t) = \cos(t)$$
$$v(t) = \sin(t)$$

La fonction u vérifie :

$$u(0) = 1$$
$$\partial u = -v$$

La fonction v vérifie :

$$v(0) = 0$$
$$\partial v = u$$

Soit φ l'infimum des réels positifs donnant une valeur négative à la fonction u :

$$\varphi = \inf\{x \in \mathbb{R} : x \geq 0,\ u(x) < 0\}$$

Nous allons montrer que φ est un réel, c'est-à-dire que $\varphi < +\infty$. Le théorème fondamental du calcul différentiel et intégral nous dit que :

$$\begin{aligned} u(t) - u(0) &= \int_0^t \frac{du}{dt}(s)\ ds \\ v(t) - v(0) &= \int_0^t \frac{dv}{dt}(s)\ ds \end{aligned}$$

En tenant compte des propriétés de u, v, ces relations se réécrivent :

$$\begin{aligned} u(t) - 1 &= -\int_0^t v(s) ds \\ v(t) - 0 &= \int_0^t u(s) ds \end{aligned}$$

On a donc :

$$\begin{aligned} u(t) &= 1 - \int_0^t v(s) ds \\ v(t) &= \int_0^t u(s) ds \end{aligned}$$

Comme u est dérivable, elle est continue et on peut trouver un $\epsilon \in \,]0, \varphi[$ tel que, pour tout $t \in \,]0, \epsilon[$:

$$|u(t) - u(0)| = |u(t) - 1| < 1$$

On en déduit que :

$$1 - u(t) < 1$$

c'est-à-dire :

$$u(t) > 0$$

La positivité de u entraîne celle de v :

$$v(t) = \int_0^t u(s) ds > 0$$

pour tout $t \in \,]0, \epsilon[$.
Pour $t \in \,]\epsilon, \varphi[$, la définition de φ nous dit que $u \geq 0$ sur $]0, \varphi[$ et donc :

$$\begin{aligned} v(t) &= \int_0^t u(s) ds = \int_0^\epsilon u(s) ds + \int_\epsilon^t u(s) ds \\ &\geq \int_0^\epsilon u(s) ds > 0 \end{aligned}$$

On pose :

$$\delta = \int_0^\epsilon u(s) ds$$

pour alléger les notations. Toujours avec $t \in \,]\epsilon, \varphi[$, il vient :

$$\begin{aligned} u(t) &= 1 - \int_0^t v(s)ds \leq 1 - \int_\epsilon^t v(s)ds \\ &\leq 1 - (t - \epsilon)\ \delta \end{aligned}$$

On peut donc trouver t tel que $u(t) \leq 0$. En effet, l'égalité :

$$1 - (t - \epsilon)\ \delta = 0$$

est équivalente à :

$$t = \frac{1}{\delta} + \epsilon < +\infty$$

donc :

$$u\left(\frac{1}{\delta} + \epsilon\right) \leq 0$$

Comme u est continue, elle ne peut pas devenir négative sans passer par 0 et il existe au moins un :

$$\psi \leq \frac{1}{\delta} + \epsilon$$

tel que $u(\psi) = 0$. Donc :

$$\varphi \leq \frac{1}{\delta} + \epsilon < +\infty$$

111.10 Périodicité

Nous allons à présent montrer d'importantes propriétés de périodicité des fonctions trigonométriques. Considérons la plus petite racine positive de u :

$$\psi = \min\{x \geq 0 : u(x) = 0\}$$

Vu que $u^2 + v^2 = 1$, on doit avoir $v(\psi)^2 = 1 - u(\psi)^2 = 1$. Donc $v(\psi) = \pm 1$. Mais $v \geq \delta > 0$ sur l'intervalle (ϵ, ψ) et par continuité :

$$v(\psi) = \lim_{x \to \psi} v(x) \geq 0$$

La seule solution acceptable est donc $v(\psi) = 1$. Donc v est solution du problème :

$$\begin{gathered} \frac{d^2 v}{dt^2} = -v \\ v(\psi) = 1 \\ \frac{dv}{dt}(\psi) = 0 \end{gathered}$$

Mais la fonction définie par $f(t) = u(t - \psi)$ vérifie également ce problème. Par unicité, on en déduit :

$$u(t - \psi) = v(t)$$

En particulier, en $t = 2\psi$, on a :

$$v(2\psi) = u(\psi) = 0$$

Donc $u(2\psi)^2 = 1 - v(2\psi)^2 = 1$. Quel est le signe de $u(2\psi)$? Choisissons $s \in [\psi, 2\psi]$ et $t = s - \psi \in [0, \psi]$. On a :

$$v(s) = u(t) \geq 0$$

par définition de ψ et continuité de u (la fonction de peut pas devenir négative avant de passer par 0). Donc :

$$u(t) = -\int_{\psi}^{t} v(s) ds \leq 0$$

et par continuité $u(2\psi) = -1$. Par unicité de la solution de :

$$\begin{aligned} \frac{d^2}{dt^2}(-u) &= -(-u) \\ -u(0) &= -1 \\ \frac{d}{dt}(-u)(0) &= 0 \end{aligned}$$

on a :

$$u(t - 2\psi) = -u(t)$$

Répétant le même procédé, on obtient successivement :

$$\begin{array}{cc} u(0) = 1 & v(0) = 0 \\ u(\psi) = 0 & v(\psi) = 1 \\ u(2\psi) = -1 & v(2\psi) = 0 \\ u(3\psi) = 0 & v(3\psi) = -1 \\ u(4\psi) = 1 & v(4\psi) = 0 \end{array}$$

111.10.1 Extension

Donc $u(4\psi) = u(0)$ et $v(4\psi) = v(0)$. On en déduit que $Q(4\psi) = Q(0) = I$. Mais, par additivité des rotations,

$$Q(t + 4\psi) = Q(t)\ Q(4\psi) = Q(t)$$

Et donc :

$$\begin{aligned} u(t + 4\psi) &= u(t) \\ v(t + 4\psi) &= v(t) \end{aligned}$$

pour tout $t \in \mathbb{R}$. Définissant le nombre π par :

$$\pi = 2\psi$$

on peut écrire la périodicité des fonctions trigonométriques :

$$\begin{aligned} \cos(\theta + 2\pi) &= \cos(\theta) \\ \sin(\theta + 2\pi) &= \sin(\theta) \end{aligned}$$

On en déduit directement que :

$$\begin{aligned} \cos(\theta + 2k\pi) &= ... = \cos(\theta) \\ \sin(\theta + 2k\pi) &= ... = \sin(\theta) \end{aligned}$$

pour tout $k \in \mathbb{Z}$.

111.11 Angle double

L'additivité nous donne :

$$\cos(2\ x) = \cos(x + x) = \cos(x) \cdot \cos(x) - \sin(x) \cdot \sin(x)$$

On a donc :

$$\cos(2\ x) = \cos(x)^2 - \sin(x)^2$$

111.11.1 Alternative

En utilisant :

$$\cos(x)^2 + \sin(x)^2 = 1$$

on dispose des formulations alternatives :

$$\cos(2\ x) = 2\cos(x)^2 - 1$$

et :

$$\cos(2\ x) = 1 - 2\sin(x)^2$$

111.11.2 Relations inverses

Il est aisé d'inverser ces deux relations, on a :

$$\cos(x)^2 = \frac{\cos(2\ x) + 1}{2}$$

et :

$$\sin(x)^2 = \frac{1 - \cos(2\ x)}{2}$$

111.12 Intégrale

Comme sin est une primitive de cos, on a :

$$\int_a^b \cos(t)\ dt = \sin(b) - \sin(a)$$

Comme cos est une primitive de $-\sin$, on a :

$$-\int_a^b \sin(t)\ dt = \cos(b) - \cos(a)$$

ou :

$$\int_a^b \sin(t)\ dt = \cos(a) - \cos(b)$$

111.13 Tangente

La tangente tan est définie par :

$$\tan(x) = \frac{\sin(x)}{\cos(x)}$$

pour tout $x \in \mathbb{R}$.

111.13.1 Dérivée

On a :

$$\frac{d\tan}{dx}(x) = \frac{\cos(x)}{\cos(x)} - \frac{\sin(x) \cdot \big(-\sin(x)\big)}{\cos(x)^2} = 1 + \tan(x)^2$$

111.13.2 Problème différentiel

Comme :

$$\tan(0) = \frac{\sin(0)}{\cos(0)} = \frac{0}{1} = 0$$

la tangente est solution $u : \mathbb{R} \mapsto \mathbb{R}$ du problème différentiel :

$$\begin{aligned} \partial u(t) &= 1 + u(t)^2 \\ u(0) &= 0 \end{aligned}$$

vérifié pour tout $t \in \mathbb{R}$.

111.14 Angle entre vecteurs

Dans le cas où le produit scalaire est réel, on peut trouver un réel $\theta \in [0, \pi]$ tel que :

$$-1 \le \cos(\theta) = \frac{\langle x \mid y \rangle}{\|x\| \ \|y\|} \le 1$$

On dit alors que θ est l'angle formé par les deux vecteurs x et y.

111.15 Angles entre espaces

Autre application des valeurs singulières, les angles entres espaces vectoriels générés par les vecteurs colonnes orthonormés des matrices :

$$X \in \mathbb{M}(\mathbb{R}, k, n)$$
$$Y \in \mathbb{M}(\mathbb{R}, l, n)$$

On a donc :

$$X^\star \cdot X = I_k$$
$$Y^\star \cdot Y = I_l$$

Les valeurs singulières de la matrice des produits scalaires :

$$Y^\star \cdot X$$

nous donne les cosinus de ces angles.

111.16 Coordonnées polaires

AFAIRE : CLARIFIER LA FIN DU CHAPITRE
Soit les vecteurs (c_1, c_2) formant une base orthonormée de $\mathbb{R}^2$:

$$\langle c_i \mid c_j \rangle = \delta_{ij}$$

et ne dépendant pas de la position :

$$\frac{\partial c_i}{\partial x_j} = 0$$

Soit le changement de variable vers $y = (R, \theta)$, exprimé par :

$$x_1 = R \cdot \cos(\theta)$$
$$x_2 = R \cdot \sin(\theta)$$

$$r = x_1 \cdot c_1 + x_2 \cdot c_2$$
$$dr = e_R \; dR + e_\theta \; d\theta$$

$$e_R = \frac{\partial r}{\partial R} = \cos(\theta) \cdot c_1 + \sin(\theta) \cdot c_2$$
$$e_\theta = \frac{\partial r}{\partial \theta} = -R \cdot \sin(\theta) c_1 + R \cdot \cos(\theta) \cdot c_2$$

$$\frac{\partial e_R}{\partial R} = 0$$
$$\frac{\partial e_R}{\partial \theta} = -\sin(\theta) \cdot c_1 + \cos(\theta) \cdot c_2 = \frac{1}{R} \cdot e_\theta$$
$$\frac{\partial e_\theta}{\partial R} = -\sin(\theta) \cdot c_1 + \cos(\theta) \cdot c_2 = \frac{1}{R} \cdot e_\theta$$
$$\frac{\partial e_\theta}{\partial \theta} = -R \cdot \cos(\theta) \cdot c_1 - R \cdot \sin(\theta) \cdot c_2 = -R \cdot e_R$$

$$de_R = \frac{\partial e_R}{\partial R} \; dR + \frac{\partial e_R}{\partial \theta} \; d\theta = \frac{1}{R} \; d\theta \; e_\theta$$
$$de_\theta = \frac{\partial e_\theta}{\partial R} \; dR + \frac{\partial e_\theta}{\partial \theta} \; d\theta = \frac{1}{R} \; dR \; e\theta - R \; d\theta$$

$$a = a^R \cdot e_R + a^\theta \cdot e_\theta$$

$$da = \frac{\partial a^R}{\partial R} \cdot e_R \; dR + \left(\frac{\partial a^R}{\partial \theta} - R \cdot a^\theta \right) \cdot e_R \; d\theta +$$
$$\left(\frac{\partial a^\theta}{\partial R} + \frac{a^\theta}{R} \right) \cdot e_\theta \; dR + \left(\frac{\partial a^\theta}{\partial \theta} + \frac{a^R}{R} \right) \cdot e_\theta \; d\theta$$

Chapitre 112

Fonctions trigonométriques inverses

Les fonctions trigonométriques ne sont pas inversible. Soit $y \in \mathbb{R}$ et $s \in \mathbb{R}$ une solution du problème :

$$\sin(s) = y$$

alors, pour tout $k \in \mathbb{N}$, on a :

$$\sin(x + 2\ k\ \pi) = \sin(x) = y$$

L'ensemble des solutions :

$$S(x) = \{x \in \mathbb{R}$$

De même pour la fonction cos. Par contre, elles sont localement inversible, et on peut définir les fonctions arcsin, arccos, arctan par :

$$\arcsin(y) = x$$
$$\Leftrightarrow$$
$$y = \sin(x)$$
$$x \in [-\pi/2, \pi/2]$$
$$\arccos(y) = x$$
$$\Leftrightarrow$$
$$y = \cos(x)$$
$$x \in [0, \pi]$$
$$\arctan(y) = x$$
$$\Leftrightarrow$$
$$y = \tan(x)$$
$$x \in [-\pi/2, \pi/2]$$

Chapitre 113

Equations aux dérivées partielles

113.1 Courbes caractéristiques

Soit $u \in F = \mathrm{Cont}^1(\mathbb{R}^2, \mathbb{R})$ et l'équation aux dérivées partielles à résoudre sur $\Omega \subseteq \mathbb{R}^2$:

$$a(x, y, u)\ u_x(x, y) + b(x, y, u)\ u_y(x, y) = c(x, y, u)$$

où nous introduisons les notations :

$$u_x(x, y) = \frac{\partial u}{\partial x}(x, y)$$
$$u_y(x, y) = \frac{\partial u}{\partial y}(x, y)$$

Les coefficients a, b, c sont en général des fonctions de x, y, u mais ne peuvent pas dépendre de u_x ni de u_y. Soit à présent la courbe Γ définie par :

$$\Gamma = \{(w_x(t), w_y(t)) : t \in \mathbb{R}\}$$

où w_x et w_y sont des fonctions dérivables. Définissons la restriction de u à Γ :

$$\varphi(t) = u\left(w_x(t), w_y(t)\right)$$

Si on s'arrange pour que :

$$\frac{dw_x}{dt}(t) = a(w_x(t), w_y(t), u(w_x(t), w_y(t)))$$
$$\frac{dw_y}{dt}(t) = b(w_x(t), w_y(t), u(w_x(t), w_y(t)))$$

On a alors :

$$\frac{d\varphi}{dt} = u_x\ a + u_y\ b = c$$

Définissons alors :

$$f : (t, u) \mapsto c\left(w_x(t), w_y(t), u\right)$$

On a :

$$\frac{d\varphi}{dt}(t) = f(t, u(t))$$

qui est une équation différentielle ordinaire en t. On peut donc connaître φ et donc u sur Γ si on ajoute la condition initiale :

$$\varphi(0) = u_0$$

On dit alors que Γ est une courbe caractéristique de l'équation aux dérivées partielles.

113.2 Fonction de Green

Soit un espace fonctionnel $F \subseteq \mathrm{Leb}^2(\mathbb{R}^n, \mathbb{R})$ et une forme $\langle , \rangle : F^D \times F \mapsto \mathbb{R}$ à laquelle on associe par abus de notation :

$$\int_A u(x) \cdot v(x) \, dx = \langle u, v \rangle$$

où $A \subseteq \mathbb{R}^n$.
Soit un opérateur $L : F \mapsto \mathrm{Leb}^2(\mathbb{R}^n, \mathbb{R})$ qui vérifie :

$$\langle u, L(v) \rangle = \langle L(u), v \rangle$$

pour tout $u, v \in F$. On dit d'un tel opérateur qu'il est auto-adjoint.
Nous nous intéressons à l'équation différentielle :

$$L(u) = f$$

où $f \in F$.
Introduisons la distribution δ de Dirac :

$$\int_A \delta(x - a) \; f(x) \; dx = f(a)$$

et définissons la famille de solutions v_x telles que :

$$L(v_x)(y) = \delta(y - x)$$

On peut alors définir la fonction de Green G :

$$G(x, y) = v_x(y)$$

Mais les propriétés de L nous permettent d'écrire :

$$\langle L(v_x), u \rangle = \langle v_x, L(u) \rangle$$

Si u est la solution de $L(u) = f$, l'équation précédente peut se formeuler comme :

$$\int_A \delta(y - x) \; u(y) \; dy = \int_A v_x(y) \; f(y) \; dy$$

et finalement :

$$u(x) = \int_A G(x, y) \; f(y) \; dy$$

113.2.1 Exemple d'opérateur auto-adjoint

Comme exemple d'opérateur auto-adjoint, citons :

$$L : u \mapsto \mathbf{\Delta} u = \sum_i \frac{\partial^2 u}{\partial x_i^2}$$

Chapitre 114

Algorithmes de résolution d'EDO

114.1 Introduction

AFAIRE : ARRANGER LE CHAPITRE
Le but est de calculer une approximation de la solution y de

$$\frac{dy}{dx}(x) = f(x, y(x))$$
$$y(0) = y_0$$

Pour cela, on choisit n points x_i, et on tente de progresser en évaluant les approximations successives $y_{i+1} \approx y(x_{i+1})$ à partir de y_i, pour $i = 0, 1, 2, ...$ On pose :

$$h_i = x_{i+1} - x_i$$

114.1.1 Euler

On se sert de la formulation intégrale correspondant à l'équation différentielle que l'on veut résoudre :

$$y_{i+1} = y_i + \int_{x_i}^{x_{i+1}} f(x, y(x))\ dx$$

Si h_i est suffisamment petit, la fonction y sera plus ou moins constante sur l'intervalle $[x_i, x_{i+1}]$ et on peut approximer la formulation intégrale par :

$$y_{i+1} = y_i + h_i \cdot f(x_i, y_i)$$

Cette méthode se nomme Euler explicite.

114.1.2 Predicteur - Correcteur

On part de nouveau de :

$$y_{i+1} = y_i + \int_{x_i}^{x_{i+1}} f(x, y(x))\ dx$$

On commence par calculer une première estimation de $y(x_{i+1})$ en utilisant la méthode d'Euler explicite :

$$y^*_{i+1} = y_i + h_i \cdot f(x_i, y_i)$$

La valeur y^*_{i+1} ainsi obtenue est nommé prédicteur.

Une fois cette première estimation évaluée, on construit une meilleure approximation de l'intégrale en supposant que, pour $t \in [0,1]$:

$$f\left(x_i + t \cdot h_i, y(x_i + th_i)\right) \approx f_i + t \cdot \left(f^*_{i+1} - f_i\right)$$

avec :

$$\begin{aligned} f_i &= f(x_i, y_i) \\ f^*_{i+1} &= f(x_{i+1}, y^*_{i+1}) \end{aligned}$$

On obtient :

$$\begin{aligned} \int_{x_i}^{x_{i+1}} f(x, y(x))\ dx &\approx h_i \int_0^1 \left[f_i + t \cdot (f^*_{i+1} - f_i)\right] dt \\ &\approx \frac{h_i}{2} \cdot \left(f_i + f^*_{i+1}\right) \end{aligned}$$

L'étape correctrice s'écrit donc :

$$y_{i+1} = y_i + \frac{h_i}{2} \cdot \left[f(x_i, y_i) + f(x_{i+1}, y^*_{i+1})\right]$$

114.1.3 Taylor

Partant des dérivées de la fonction $y(x) = f(x, y(x))$:

$$\begin{aligned} \frac{dy}{dx} &= f \\ \frac{d^2y}{dx^2} &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} f \\ \frac{d^3y}{dx^3} &= ... \end{aligned}$$

on peut écrire le développement en série de Taylor de y autour de x_i :

$$y_{i+1} = y_i + f(x_i, h_i)\ h_i + \frac{df}{dx}(x_i, y_i)\ \frac{h_i^2}{2} + ...$$

Chapitre 115

Algorithmes de résolution d'EDP

115.1 Résolution par les caractéristiques

AFAIRE : ARRANGER LE CHAPITRE

Soit une équation du premier ordre à résoudre :

$$a(x, y, u)\ u_x(x, y) + b(x, y, u)\ u_y(x, y) = c(x, y, u)$$

Supposons que l'on connaisse la valeur de la solution :

$$u_{i0} = u(x_{i0}, y_{i0})$$

pour $i = 1, 2, ..., n$. Les équations :

$$\frac{dx}{dt} = a \qquad \frac{dy}{dt} = b \qquad \frac{du}{dt} = c$$

nous permettent de construire simultanément les courbes caractéristiques et la solution. Par exemple, si on utilise le schéma d'Euler explicite, on a :

$$\begin{aligned} x_{i,k+1} &= x_{ik} + h_k\ a(x_{ik}, y_{ik}, u_{ik}) \\ y_{i,k+1} &= y_{ik} + h_k\ b(x_{ik}, y_{ik}, u_{ik}) \\ u_{i,k+1} &= u_{ik} + h_k\ c(x_{ik}, y_{ik}, u_{ik}) \end{aligned}$$

115.2 Différences finies

On vérifie sur le développement de Taylor de u que :

$$\begin{aligned} u_x(x, y) &\approx \frac{u(x + h, y) - u(x - h, y)}{2h} \\ u_y(x, y) &\approx \frac{u(x, y + h) - u(x, y - h)}{2h} \end{aligned}$$

pour les dérivées premières et :

$$\begin{aligned} u_{xx}(x, y) &\approx \frac{1}{h^2}\ (u(x + h, y) - 2u(x, y) + u(x - h, y)) \\ u_{yy}(x, y) &\approx \frac{1}{h^2}\ (u(x, y + h) - 2u(x, y) + u(x, y - h)) \\ u_{xy}(x, y) &\approx \frac{1}{4h^2}\ (u(x + h, y + h) + u(x - h, y - h) \\ &\qquad -u(x + h, y - h) - u(x - h, y + h)) \end{aligned}$$

pour les dérivées secondes. L'erreur converge vers 0 aussi vite que h^2. Posons :

$$U_{ij} = u(ih, jh)$$

Les dérivées approximatives s'écrivent :

$$u_x \approx \Delta_x U_{ij} = \frac{1}{2h} \left(U_{i+1,j} - U_{i-1,j} \right)$$
$$u_y \approx \Delta_y U_{ij} = \frac{1}{2h} \left(U_{i,j+1} - U_{i,j-1} \right)$$

et :

$$u_{xx} \approx \Delta_x^2 U_{ij} = \frac{1}{h^2} \left(U_{i+1,j} - 2U_{ij} + U_{i-1,j} \right)$$
$$u_{yy} \approx \Delta_y^2 U_{ij} = \frac{1}{h^2} \left(U_{i,j+1} - 2U_{ij} + U_{i,j-1} \right)$$
$$u_{xy} \approx \Delta_x \Delta_y U_{ij} = \frac{1}{4h^2} \left(U_{i+1,j+1} - U_{i+1,j-1} - U_{i-1,j+1} + U_{i-1,j-1} \right)$$

On substitue alors ces expressions dans l'équation :

$$F(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy}) = 0$$

et on obtient un système linéaire à résoudre en les U_{ij}.

115.3 Eléments finis

Soient des espaces fonctionnels $F, H \subset \mathbb{F}(\Omega, \mathbb{R})$ et un opérateur linéaire :

$$L \in \mathrm{Lin}(F, H)$$

Nous cherchons à résoudre de manière approchée l'équation différentielle associée :

$$L(u) = f$$

où f est une fonction de H. La méthode des éléments finis consiste à imposer l'annulation de l'intégrale du résidu $L(u)$, pondéré par des fonctions ψ_i :

$$\int_\Omega [L(u)(x) - f(x)] \ \psi_i(x) \ dx = 0$$

pour $i = 1, 2, ..., n$. Afin de résoudre ce problème, on discrétise la solution approchée u :

$$u(x) = \sum_{i=1}^{n} U_i \ \varphi_i(x)$$

où les U_i sont des réels et les φ_i des fonctions de F. On définit les grandeurs :

$$A_{ij} = \int_\Omega L(\varphi_i)(x) \ \psi_i(x) \ dx$$
$$F_i = \int_\Omega f(x) \ \psi_i(x) \ dx$$

et les matrices :

$$U = (U_i)_i$$
$$A = (A_{ij})_{i,j}$$
$$F = (F_i)_i$$

En utilisant la linéarité de L, l'équation des résidus pondérés :

$$\int_\Omega L(u)(x)\ \psi_i(x)\ dx = \int_\Omega f(x)\ \psi_i(x)\ dx$$

devient :

$$A\ U = F$$

soit un système linéaire à résoudre en U :

$$U = A^{-1}\ F$$

ce qui nous donne une forme approchée u de la solution exacte.

Onzième partie

Analyse dans le plan complexe

Chapitre 116

Exponentielle complexe

116.1 Dépendances

— Chapitre **??** : Les complexes

116.2 Introduction

AFAIRE : ARRANGER LE CHAPITRE
Considérons l'unique solution x de l'équation différentielle :

$$\frac{dx}{dt}(t) = z \cdot x(t)$$
$$x(0) = 1$$

où $z \in \mathbb{C}$. On définit alors l'exponentielle d'un nombre complexe par :

$$\exp(z \cdot t) = x(t)$$

On a donc simplement :

$$\exp(z) = x(1)$$

116.3 Additivité

Comme les fonctions :

$$s(t) = \exp((z_1 + z_2) \cdot t)$$
$$p(t) = \exp(z_1 t) \cdot \exp(z_2 t)$$

vérifient la même équation différentielle :

$$\frac{ds}{dt}(t) = (z_1 + z_2)\ s(t)$$
$$s(0) = 1$$

$$\frac{dp}{dt}(t) = z_1\ p(t) + z_2\ p(t) = (z_1 + z_2)\ p(t)$$
$$p(0) = 1$$

on a $s(t) = p(t)$ pour tout t. En considérant le cas $t = 1$, on arrive à la propriété d'additivité des exponentielles :

$$\exp(z_1 + z_2) = \exp(z_1) \cdot \exp(z_2)$$

116.4 Lien avec les fonctions trigonométriques

Considérons maintenant le cas particulier où $u(t) = \exp(\mathbf{i}\, t)$:

$$\frac{du}{dt}(t) = \mathbf{i}\, u(t)$$
$$u(0) = 1$$

On a alors :

$$\frac{d^2u}{dt^2}(t) = -u(t) \qquad u(0) = 1 \qquad \frac{du}{dt}(0) = \mathbf{i}$$

Soit la fonction $u : \mathbb{R} \mapsto \mathbb{R}$ définie pour tout réel t par :

$$u(t) = \cos(t) + \mathbf{i} \sin(t)$$

On voit que :

$$\frac{du}{dt} = -\sin(t) + \mathbf{i} \cos(t)$$

vérifie la même équation. Par unicité de la solution, on a :

$$\exp(\mathbf{i}\, t) = \cos(t) + \mathbf{i} \sin(t)$$

On en déduit directement que :

$$\exp(\mathbf{i}\, \pi/2) = \mathbf{i}$$
$$\exp(\mathbf{i}\, \pi) = -1$$
$$\exp(3\pi\, \mathbf{i}\, /2) = \mathbf{i}$$
$$\exp(2\pi\, \mathbf{i}) = 1$$

ainsi que la périodicité :

$$\exp(t + 2\pi) = \exp(t)$$

pour tout $t \in \mathbb{R}$.
En utilisant l'additivité, on note que :

$$\exp(z) = \exp(a) \cdot \exp(\mathbf{i}\, b) = \exp(a) \cdot (\cos(b) + \mathbf{i} \sin(b))$$

De la même façon, si nous prenons $u(t) = \exp(-i.t)$, nous en déduisons :

$$\frac{d^2u}{dt^2}(t) = -u(t)$$
$$u(0) = 1 \qquad \frac{du}{dt}(0) = -i$$

Donc :

$$\exp(-\,\mathbf{i}\, t) = \cos(t) - \mathbf{i} \sin(t)$$

En additionnant, puis en soustrayant les relations :

$$\exp(\mathbf{i}\, u) = \cos(u) + \mathbf{i} \sin(u)$$
$$\exp(-\,\mathbf{i}\, u) = \cos(u) - \mathbf{i} \sin(u)$$

on obtient :

$$\cos(u) = \frac{\exp(\mathbf{i}\, u) + \exp(-\mathbf{i}\, u)}{2}$$
$$\sin(u) = \frac{\exp(\mathbf{i}\, u) - \exp(-\mathbf{i}\, u)}{2\,\mathbf{i}}$$

Choisissant un angle $\theta \in \mathbb{R}$ tel que :

$$\cos(\theta) = \frac{\Re(z)}{|z|}$$
$$\sin(\theta) = \frac{\Im(z)}{|z|}$$

on peut réexprimer z comme :

$$z = |z| \left(\cos(\theta) + \mathbf{i} \sin(\theta)\right) = |z| \exp(\mathbf{i}\, \theta)$$

Comme les fonctions cos sont 2π périodiques, il existe une infinité d'angles θ vérifiant cette propriété. On définit l'argument de z comme l'unique θ vérifiant cette propriété et se trouvant dans l'intervalle $[0, 2\pi)$.

$$\theta = \arg(z) \quad \Leftrightarrow \quad \begin{cases} z = |z| \exp(\mathbf{i}\, \theta) \\ \theta \in [0, 2\pi) \end{cases}$$

Inspiré par relation :

$$z = |z| \exp(\mathbf{i} \arg(z))$$

et cherchant à étendre la propriété :

$$\ln(x \cdot y) = \ln(x) + \ln(y)$$

du logarithme sur $\mathbb{R}$, on définit le logarithme d'un complexe par :

$$\ln(z) = \ln(|z|) + \ln(\exp(\mathbf{i} \arg(z)))$$
$$\ln(z) = \ln(|z|) + \mathbf{i} \arg(z)$$

Pour un $z \in \mathbb{C}$ donné, l'ensemble des $y \in \mathbb{C}$ tels que :

$$\exp(y) = z$$

peut s'écrire :

$$\mathcal{Y} = \{y_k = \ln(|z|) + \mathbf{i} \arg(z) + 2\pi\, \mathbf{i}\, k : k \in \mathbb{Z}\}$$

Les complexes permettent de retrouver aisément les propriétés fondamentales des fonctions trigonométriques cos, sin. En se rappelant que :

$$\exp(-\mathbf{i}\, t) \cdot \exp(it) = \exp(it - it) = \exp(0) = 1$$

on retombe sur :

$$(\cos(t) + \mathbf{i} \sin(t))(\cos(t) - \mathbf{i} \sin(t)) = 1$$
$$\cos(t)^2 + \sin(t)^2 = 1$$

En dérivant la relation reliant l'exponentielle aux fonctions trigonométriques, on obtient :

$$\frac{d}{dt}(\cos(t) + \mathbf{i}\sin(t)) = \frac{d}{dt}\exp(\mathbf{i}\,t)$$
$$\frac{d}{dt}\cos(t) + \mathbf{i}\,\frac{d}{dt}\sin(t) = \mathbf{i}\exp(\mathbf{i}\,t)$$
$$\frac{d}{dt}\cos(t) + \mathbf{i}\,\frac{d}{dt}\sin(t) = \mathbf{i}\cos(t) - \sin(t)$$

on a bien :

$$\frac{d}{dt}\cos(t) = -\sin(t)$$
$$\frac{d}{dt}\sin(t) = \cos(t)$$

On a aussi :

$$\cos(u+v) + \mathbf{i}\sin(u+v) = \exp\left[\mathbf{i}(u+v)\right]$$
$$\cos(u+v) + \mathbf{i}\sin(u+v) = \exp\left[\mathbf{i}\,u\right] \cdot \exp\left[\mathbf{i}\,v\right]$$
$$\cos(u+v) + \mathbf{i}\sin(u+v) = (\cos(u) + \mathbf{i}\sin(u))(\cos(v) + \mathbf{i}\sin(v))$$

ce qui donne, tous calculs faits :

$$\cos(u+v) = \cos(u)\cos(v) - \sin(u)\sin(v)$$
$$\sin(u+v) = \sin(u)\cos(v) + \sin(v)\cos(u)$$

Chapitre 117

Dérivation et intégration dans le plan complexe

117.1 Dépendances

— Chapitre **??** : Les différentielles
— Chapitre **??** : Les intégrales

117.2 Dérivées

On définit la dérivée d'une fonction $f : \mathbb{C} \mapsto \mathbb{C}$ par analogie avec la dérivée des fonctions réelles :

$$\frac{df}{dz}(z) = \lim_{\substack{h \to 0 \\ h \in \mathbb{C}}} \frac{f(z+h) - f(z)}{h}$$

Si f est dérivable sur $A \subseteq \mathbb{C}$, on dit qu'elle est analytique sur A.
On définit les fonctions $u, v : \mathbb{R}^2 \mapsto \mathbb{R}$ associées à f par :

$$u(x,y) = \Re\left(f(x + \mathbf{i}\, y)\right)$$
$$v(x,y) = \Im\left(f(x + \mathbf{i}\, y)\right)$$

On a alors $f(z) = u(x,y) + \mathbf{i}\, v(x,y)$. Si f est dérivable, la limite ne peut pas dépendre du chemin suivi pour arriver en (x,y). On peut donc choisir h de la forme Δx ou $\mathbf{i}\,\Delta y$ avec $\Delta x, \Delta y \in \mathbb{R}$. Cela donne :

$$\frac{df}{dz}(z) = \lim_{\Delta x \to 0} \frac{\frac{\partial u}{\partial x}\Delta x + \mathbf{i}\,\frac{\partial v}{\partial x}\Delta x}{\Delta x}$$
$$\frac{df}{dz}(z) = \lim_{\Delta y \to 0} \frac{\frac{\partial u}{\partial y}\Delta y + \mathbf{i}\,\frac{\partial v}{\partial y}\Delta y}{\mathbf{i}\,\Delta y}$$

et par comparaison :

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$
$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

En dérivant de nouveau ces équations, on obtient :

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$
$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0$$

Les puissances de z sont dérivables. On vérifie que :

$$\frac{dz^n}{dz} = nz^{n-1}$$

117.3 Extension de l'intégrale au plan complexe

Soit une fonction $f : A \mapsto \mathbb{C}$. Comme $u = \Re(f)$ et $v = \Im(f)$ sont des fonctions à valeurs réelles :

$$u, v : A \mapsto \mathbb{R}$$

leur intégrale est bien définie. On définit alors l'intégrale de f par :

$$\int_A f d\mu = \int_A u d\mu + \mathbf{i} \int_A v d\mu$$

117.4 Intégrale de ligne

Soit une fonction $\gamma : [a, b] \mapsto \mathbb{C}$ et la courbe :

$$l = \{\gamma(t) : t \in [a, b]\} \subset \mathbb{C}$$

On définit l'intégrale de ligne de f sur l par :

$$\int_l f(z)dz = \int_a^b (f \circ \gamma)(t) \frac{d\gamma}{dt}(t)dt$$

On peut également écrire :

$$\int_l f(z)dz = \int_l (u + \mathbf{i}\, v)(dx + \mathbf{i}\, dy)$$

$$\int_l f(z)dz = \int_l (udx - vdy) + \mathbf{i} \int_l (udy + vdx)$$

où les intégrales des membres de droites sont des intégrales de ligne classiques de fonctions $\mathbb{R}^2 \mapsto \mathbb{R}$.

117.4.1 Théorème fondamental

Soit une fonction analytique $F : \mathbb{C} \mapsto \mathbb{C}$. La dérivée d'une composée de fonction nous donne :

$$\frac{d}{dt}(F \circ \gamma)(t) = (\frac{dF}{dz} \circ \gamma)(t) \frac{d\gamma}{dt}(t)$$

On en déduit, en appliquant le théorème fondamental que :

$$\int_l \frac{dF}{dz}(z)dz = F(\gamma(b)) - F(\gamma(a))$$

117.4.2 Contour fermé

Dans la suite, lorsque l est une courbe fermée et est donc la frontière d'un certain ensemble $l = \partial S$ inclus dans $\mathbb{C}$, nous notons :

$$\oint_l f(z)dz = \int_{\partial S} f(z)dz$$

Soit f une fonction analytique et $l = \partial S$ une courbe fermée. On a :

$$\oint_l f(z)dz = \oint_l (udx - vdy) + \mathbf{i} \oint_l (udy + vdx)$$

En appliquant le théorème de Stokes, il vient :

$$\oint_l f(z)dz = \int_S (-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y})dS + \mathbf{i} \int_S (\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y})dS$$

Mais comme f est analytique :

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$
$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

Les termes du membre de droite s'annulent donc et :

$$\oint_l f(z)dz = 0$$

117.4.3 Cercle

Soit C le cercle de centre a et de rayon R :

$$C = \{\gamma(\theta) = a + R \exp(\mathbf{i}\, \theta) : \theta \in [0, 2\pi)\}$$

On a alors :

$$\frac{d\gamma}{d\theta}(\theta) = R\, \mathbf{i} \exp(\mathbf{i}\, \theta)$$

et :

$$(z - a)^k = (R \exp(\mathbf{i}\, \theta))^k$$

pour tout $z \in C$ et $k \in \mathbb{Z}$. Donc :

$$\oint_C \frac{dz}{(z-a)^k} = \int_0^{2\pi} \frac{R\, \mathbf{i} \exp(\mathbf{i}\, \theta)}{R^k \exp(\mathbf{i}\, k\theta)}$$
$$\oint_C \frac{dz}{(z-a)^k} = \mathbf{i}\, R^{1-k} \int_0^{2\pi} \exp[\mathbf{i}(1-k)\theta] d\theta$$

Mais les propriétés de périodicité des fonctions trigonométriques et de l'exponentielle associée nous permettent de vérifier que :

$$\int_0^{2\pi} \exp[\mathbf{i}\, l\theta] d\theta = 2\pi \delta_{l,0}$$

pour tout $l \in \mathbb{Z}$. On a donc :

$$\oint_C \frac{dz}{(z-a)^k} = 2\pi\, \mathbf{i}\, \delta_{k,1}$$

pour tout $k \in \mathbb{Z}$. L'intégrale s'annule pour tout les $k \neq 1$.

Vu que la fonction $f(z) = (z-a)^{-k}$ est analytique partout sauf en a, on obtient le même résultat pour toute courbe fermée entourant a :

$$\oint_{\partial S} \frac{dz}{(z-a)^k} = 2\pi\, \mathbf{i}\, \delta_{k,1} \delta_S(a)$$

117.4.4 Théorème de Cauchy

Si f est analytique, la fonction

$$g(z) = \begin{cases} \frac{f(z)-f(a)}{z-a} & \text{si } z \neq a \\ \frac{df}{dz}(a) & \text{si } z = a \end{cases}$$

est analytique aussi. On a donc :

$$\oint_l g(z)dz = 0$$

Soit $a \in \mathbb{C}$ et une courbe fermée entourant a telle que $a \notin l$. On a alors :

$$\oint_l \frac{f(z)-f(a)}{z-a}dz = \oint_l g(z)dz = 0$$

On en déduit que :

$$f(a)\oint_{\partial S} \frac{1}{z-a}dz = \oint_{\partial S} \frac{f(z)}{z-a}dz$$

ce qui nous donne la valeur de $f(a)$ en fonction d'une intégrale :

$$f(a) = \frac{1}{2\pi\,\mathbf{i}} \oint_{\partial S} \frac{f(z)}{z-a}dz$$

En dérivant k fois cette dernière relation par rapport à a, on obtient :

$$\frac{df}{dz}(a) = \frac{k!}{2\pi\,\mathbf{i}} \oint_{\partial S} \frac{f(z)}{(z-a)^{k+1}}dz$$

117.4.5 Séries de Laurent

Supposons que f puisse s'écrire comme une combinaison linéaire de puissances entière de $z-a$. On a :

$$f(z) = \sum_{n=-\infty}^{+\infty} \alpha_n (z-a)^n$$

Multipliant le tout par $(z-a)^{-k-1}$ et intégrant sur un contour entourant a, on obtient :

$$\oint_l f(z)(z-a)^{-k-1}dz = \sum_{n=-\infty}^{+\infty} \alpha_n \oint_l (z-a)^{n-k-1}dz$$

Mais nous avons vu que seule l'intégrale de $(z-a)^{-1}$ ne s'annule pas. On en déduit que :

$$\alpha_k = \frac{1}{2\pi\,\mathbf{i}} \oint_l \frac{f(z)}{(z-a)^{k+1}}dz$$

117.4.6 Théorème des résidus

Considérons le cas où :

$$f(z) = \sum_{n=-p}^{+\infty} \alpha_n (z-a)^n$$

pour un certain $p \in \mathbb{Z}$. On dit alors que f à un pôle d'ordre p en a. En multpliant le tout par $(z-a)^p$ et en dérivant $p-1$ fois, on obtient :

$$\frac{d^{p-1}}{dz^{p-1}}[(z-a)^p f(z)] = (p-1)!\left[a_{-1} + a_0(z-a) + a_1(z-a)^2 + ...\right]$$

On a donc :

$$\lim_{z\to a} \frac{d^{p-1}}{dz^{p-1}}[(z-a)^p f(z)] = (p-1)!a_{-1}$$

Mais comme a_{-1} n'est rien d'autre, à un facteur $2\pi\,\mathbf{i}$ près, que l'intégrale de f sur un contour entourant a, on a :

$\oint_l f(z)dz = 2\pi\,\mathbf{i}\lim_{z\to a} \frac{1}{(p-1)!}\frac{d^{p-1}}{dz^{p-1}}[(z-a)^p f(z)]$

Chapitre 118

Polynômes et exponentielles

118.1 Chebyshev

Les polynômes de Chebyshev T_n sont défini par :

$$T_n : x \mapsto \cos(n \arccos(x))$$
$$T_n(\cos(\theta)) = \cos(n\theta)$$

Il est clair que :

$$T_0(x) = 1$$
$$T_1(x) = x$$

Les propriétés des fonctions trigonométriques :

$$\cos[(n+m)\theta] = \cos[n\theta]\cos[m\theta] + \sin[n\theta]\sin[m\theta]$$
$$\cos[(n-m)\theta] = \cos[n\theta]\cos[m\theta] - \sin[n\theta]\sin[m\theta]$$

nous montrent que :

$$T_{n+m}(x) + T_{n-m}(x) = 2T_n(x)T_m(x)$$

On en déduit entre autre que :

$$T_{n+1}(x) + T_{n-1}(x) = 2xT_n(x)$$

Les fonctions T_n sont donc des polynômes. Comme :

$$T_n\left[\cos\left(\frac{2k+1}{2n}\right)\right] = 0$$

les racines sont données par :

$$x_k = \cos\left(\frac{2k+1}{2n}\right)$$

pour $k = 0, 1, ..., n-1$. On déduit aussi la propriété suivante de la définition :

$$(T_n \circ T_m)(x) = T_{m \cdot n}(x) = (T_m \circ T_n)(x)$$

Nous allons voir que les T_n sont orthogonaux moyennant un certain poids. Si $m = n = 0$, il est clair que :

$$\int_0^\pi T_0(\cos(\theta))T_0(\cos(\theta))d\theta = \int_0^\pi d\theta = \pi$$

Si $m, n \neq 0$, on a :

$$\int_0^\pi T_m(\cos(\theta))T_n(\cos(\theta))d\theta = \int_0^\pi \cos(m\theta)\cos(n\theta)d\theta$$
$$\int_0^\pi T_m(\cos(\theta))T_n(\cos(\theta))d\theta = \frac{1}{2}\int_0^\pi \cos[(m+n)\theta]d\theta + \frac{1}{2}\int_0^\pi \cos[(m-n)\theta]d\theta$$

Si $m = n$, on a alors :

$$\int_0^\pi T_m(\cos(\theta))T_n(\cos(\theta))d\theta = \frac{1}{2n}[\sin(2n\pi) - \sin(0)] + \frac{\pi}{2}$$

La première intégrale s'annule donc par périodicité de la fonction sin. A présent, si $m \neq n$, on a :

$$\begin{gathered}\int_0^\pi T_m(\cos(\theta))T_n(\cos(\theta))d\theta = \\ \frac{1}{2(m+n)}[\sin[(m+n)\pi] - \sin(0)] + \\ \frac{1}{2(m-n)}[\sin[(m-n)\pi] - \sin(0)] = 0\end{gathered}$$

Rassemblant ces résultats, on obtient :

$$\int_0^\pi T_m(\cos\theta)T_n(\cos\theta)d\theta = \begin{cases} \pi & \text{si } m = 0 \\ \pi/2 & \text{si } m = n \neq 0 \\ 0 & \text{si } m \neq n \end{cases}$$

Le changement de variable $x = \cos(\theta)$,

$$dx = -\sin(\theta)d\theta$$
$$d\theta = \frac{dx}{\sqrt{1-x^2}}$$

nous donne :

$$\int_0^\pi T_m(\cos\theta)T_n(\cos\theta)d\theta = \int_{-1}^1 \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}dx$$

On en déduit que :

$$\int_{-1}^1 \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}dx = \delta_{mn}\frac{\pi}{2}(1+\delta_{m,0})$$

pour tout $m, n \in \mathbb{N}$.

118.2 Hermite

Les polynômes de Hermite sont orthogonaux pour le produit scalaire :

$$\int_{-\infty}^{+\infty} H_n(x)H_m(x)\exp(-x^2)dx = 2^n n!\sqrt{\pi}\delta_{mn}$$

Ils obéissent à la récurrence :

$$H_0(x) = 1$$
$$H_1(x) = 2x$$
$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$$

118.3 Laguerre

Les polynômes de Laguerre sont orthogonaux pour le produit scalaire :

$$\int_0^{+\infty} L_n(x) L_m(x) \exp(-x) dx = \delta_{mn}$$

Ils obéissent à la récurrence :

$$L_0(x) = 1$$
$$L_1(x) = 1 - x$$
$$(n+1)L_{n+1}(x) = (2n+1-x)L_n(x) - nL_{n-1}(x)$$

Chapitre 119

Analyse de Fourier

119.1 Séries de Fourier

AFAIRE : ARANGER LE CHAPITRE

Constatons tout d'abord que :

$$\int_0^{2\pi} \exp(\mathbf{i}\, kx)dx = \begin{cases} 2\pi & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$$

pour tout $k \in \mathbb{Z}$. En effet, si $k = 0$, l'intégrale devient :

$$\int_0^{2\pi} \exp(0)dx = \int_0^{2\pi} 1dx = 2\pi$$

Par contre, si $k \neq 0$, le changement de variable :

$$s = \mathbf{i}\, kx$$
$$ds = \mathbf{i}\, kdx$$

nous mène à :

$$\int_0^{2\pi} \exp(\mathbf{i}\, kx)dx = \frac{1}{\mathbf{i}\, k} \int_0^{2\pi\, \mathbf{i}\, k} \exp(s)ds$$

le théorème fondamental nous donne alors :

$$\int_0^{2\pi} \exp(\mathbf{i}\, kx)dx = \frac{1}{\mathbf{i}\, k}[\exp(2\pi\, \mathbf{i}\, k) - \exp(0)] = 0$$

pour tout $k \in \mathbb{Z}$, par périodicité des exponentielles imaginaires. On généralise ce résultat en utilisant le changement de variable :

$$t = a + \frac{T}{2\pi}x$$

ce qui nous donne :

$$\int_a^{a+T} \exp\left(\frac{2\pi\, \mathbf{i}\, kx}{T}\right) dx = \begin{cases} T & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$$

Définissons à présent le produit scalaire :

$$\langle u \mid v \rangle = \int_{-T/2}^{T/2} \bar{u}(x)v(x)dx$$

Soient les fonctions e_k :

$$e_k : x \mapsto \exp\left(\frac{2\pi\,\mathbf{i}\,kx}{T}\right)$$

où $k \in \mathbb{Z}$. Notons que :

$$\bar{e}_m e_n = \exp\left(-\frac{2\pi\,\mathbf{i}\,mx}{T}\right)\exp\left(\frac{2\pi\,\mathbf{i}\,nx}{T}\right)$$
$$\bar{e}_m e_n = \exp\left(\frac{2\pi\,\mathbf{i}(n-m)x}{T}\right)$$

La suite des e_k est donc orthonormée :

$$\langle e_m \mid e_n \rangle = \int_{-T/2}^{T/2} \exp\left(\frac{2\pi\,\mathbf{i}(n-m)x}{T}\right) dx = 2\pi\delta_{mn}$$

Soit à présent $u \in \operatorname{span}\{e_k : k \in \mathbb{Z}\}$:

$$u(x) = \sum_{k=-\infty}^{+\infty} \hat{u}_k e_k$$

Nous avons vu au chapitre **??** que les composantes d'un vecteur (ici la fonction u) par rapport à une base orthonormée s'écrivent :

$$\hat{u}_k = \langle e_k \mid u \rangle = \int_{-T/2}^{T/2} u(x)\exp\left(-\frac{2\pi\,\mathbf{i}\,kx}{T}\right) dx$$

Soit à présent v :

$$v(x) = \sum_{k=-\infty}^{+\infty} \hat{v}_k e_k$$

Utilisant à nouveau les propriétés des bases orthonormées, on a :

$$\langle u \mid v \rangle = \sum_{k=-\infty}^{+\infty} \operatorname{conj}(\hat{u}_k)\hat{v}_k$$

119.2 Transformeée de Fourier discrète

Dans le cas où :

$$a = \exp(\mathbf{i}\,x)$$
$$a^k = \exp(\mathbf{i}\,kx)$$

l'équation de la progression géométrique du chapitre **??** :

$$\sum_{i=0}^{n} a^i = \frac{1-a^{n+1}}{1-a}$$

devient :

$$\sum_{k=0}^{n} \exp(\mathbf{i}\,kx) = \frac{1-\exp(\mathbf{i}(n+1)x)}{1-\exp(\mathbf{i}\,x)}$$

On voit alors que si $(n+1)x$ est un multiple entier de 2π, la somme s'annule par périodicité sauf dans le cas où x est un multiple entier de 2π, où elle vaut $n+1$. Posant $N = n-1$ et $x = 2\pi l/N$, on obtient :

$$\sum_{k=0}^{N-1} \exp\left(\frac{2\pi\,\mathbf{i}\,kl}{N}\right) = \begin{cases} N & \text{si } m = l/N \in \mathbb{Z} \\ 0 & \text{si } m = l/N \notin \mathbb{Z} \end{cases}$$

pour tout $k, l, m, N \in \mathbb{Z}$. Dans la suite nous utilisons la notation :

$$e(k,l) = \exp\left(\frac{2\pi\,\mathbf{i}\,kl}{N}\right)$$

Nous allons en déduire la forme de la transformeée de Fourier discrète. Soit la série $u_k : k = 0, ..., N-1$:

$$u_k = \sum_{n=-N/2}^{N/2-1} U_n e(k,n)$$

où nous supposons que $N/2 \in \mathbb{N}$. Multipliant par $e(k,-m)$ où $m \in \{-N/2, ..., N/2-1\}$, et utilisant l'additivité des exponentielles, il vient :

$$u_k e(k,-m) = \sum_{n=-N/2}^{N/2-1} U_n e(k,n) e(k,-m)$$

$$u_k e(k,-m) = \sum_{n=-N/2}^{N/2-1} U_n e(k,n-m)$$

Sommons à présent sur k :

$$\sum_{k=0}^{N-1} u_k e(k,-m) = \sum_{n=-N/2}^{N/2-1} U_n \sum_{k=0}^{N-1} e(k,n-m)$$

Mais la somme sur k du membre de droite s'annule sauf lorsque $n-m$ est multiple entier de N. Comme m, n sont dans $\{-N/2, ..., N/2-1\}$, le seul cas possible est ici $m = n$. On a donc :

$$U_m = \frac{1}{N} \sum_{k=0}^{N-1} u_k e(k,-m)$$

Nous avons donc obtenu une bijection entre les u_k et les U_n :

$$u_k = \sum_{n=-N/2}^{N/2-1} U_n \exp\left(\frac{2\pi\,\mathbf{i}\,kn}{N}\right)$$

$$U_n = \frac{1}{N} \sum_{k=0}^{N-1} u_k \exp\left(-\frac{2\pi\,\mathbf{i}\,kn}{N}\right)$$

119.3 Transformeée de Fourier

On peut présenter le résultat précédent sous une autre forme. On peut poser :

$$x_k = k\Delta x$$
$$y_n = n\Delta y$$

ainsi que :

$$u_k = u(x_k)$$
$$U_n = U(y_n)$$

pour des fonctions $u, v \in X \subset \mathbb{F}(\mathbb{R}, \mathbb{R})$. Soit :

$$u(x_k) = \sum_{n=-N/2}^{N/2-1} U(y_n) \exp(2\pi \mathbf{i}\, x_k y_n) \Delta y$$

où $k = -N/2, ..., N/2-1$. Suivant le même procédé que ci-dessus, on aboutit à la relation inverse :

$$U(y_n) = \sum_{k=-N/2}^{N/2-1} u(x_k) \exp(-2\pi \mathbf{i}\, x_k y_n) \Delta x$$

pour autant que :

$$\Delta x \Delta y = \frac{1}{N}$$

Choisissons :

$$\Delta x = \Delta y = \frac{1}{\sqrt{N}}$$

On voit que Δx et Δy tendent alors vers 0 lorsque N tend vers l'infini. Les sommes ci-dessus se rapprochent donc de plus en plus d'intégrales sur l'intervalle $[-\sqrt{N}/2, \sqrt{N}/2]$, qui tend lui-même vers $(-\infty, +\infty) = \mathbb{R}$. On arrive donc aux relations :

$$u(x) = \int_{-\infty}^{+\infty} U(y) \exp(2\pi \mathbf{i}\, xy) dy$$
$$U(y) = \int_{-\infty}^{+\infty} u(x) \exp(-2\pi \mathbf{i}\, xy) dx$$

On définit alors la transformeée de Fourier $\mathcal{F} : X \mapsto X$ et son inverse par :

$$U(y) = \mathcal{F}(u)(y) = \int_{-\infty}^{+\infty} u(x) \exp(-2\pi \mathbf{i}\, xy) dx$$
$$u(x) = \mathcal{F}^{-1}(U)(x) = \int_{-\infty}^{+\infty} U(y) \exp(2\pi \mathbf{i}\, xy) dy$$

119.3.1 Delta de Dirac

On déduit des relations ci dessus que :

$$u(x) = \int_{-\infty}^{+\infty} \exp(2\pi \mathbf{i}\, xy) dy \int_{-\infty}^{+\infty} u(z) \exp(-2\pi \mathbf{i}\, zy) dz$$
$$u(x) = \int_{-\infty}^{+\infty} u(z) dz \int_{-\infty}^{+\infty} \exp(2\pi \mathbf{i}(x - z)y) dy$$

On en déduit la relation fondamentale :

$$\int_{-\infty}^{+\infty} \exp(2\pi\,\mathbf{i}(x-z)y)dy = \delta(x-z)$$

qui est l'extension de l'orthonormalité des bases discrètes. Ce n'est donc pas une intégrale au sens classique du terme, mais une distribution.

Cette relation nous montre aussi, lorsque $z = 0$, que :

$$\mathcal{F}(1)(x) = \delta(x)$$

Inversément, on a :

$$\mathcal{F}(\delta)(x) = \int_{-\infty}^{+\infty} \delta(x)\exp(-2\pi\,\mathbf{i}\,xy)dx = \exp(0) = 1$$

119.3.2 Produit scalaire

Considérons à présent le produit scalaire :

$$\langle u \mid v \rangle = \int_{-\infty}^{+\infty} \bar{u}(x)v(x)dx$$

et examinons $\langle \mathcal{F}(u) \mid \mathcal{F}(v) \rangle$. En utilisant la propriété d'orthonormalité, on arrive à :

$$\int_{-\infty}^{+\infty} \mathrm{conj}\left[\mathcal{F}(u)(y)\right]\mathcal{F}(v)(y)dy = \int_{\mathbb{R}^2} \bar{u}(x)v(z)\exp\left(2\pi\,\mathbf{i}(x-z)y\right)dxdz$$

$$\int_{-\infty}^{+\infty} \mathrm{conj}\left[\mathcal{F}(u)(y)\right]\mathcal{F}(v)(y)dy = \int_{\mathbb{R}^2} \bar{u}(x)v(z)\delta(x-z)dxdz$$

et donc :

$$\int_{-\infty}^{+\infty} \mathrm{conj}\left[\mathcal{F}(u)(y)\right]\mathcal{F}(v)(y)dy = \int_{-\infty}^{+\infty} \bar{u}(x)v(x)dx$$

La transformeée de Fourier possède donc la propriété de conserver le produit scalaire :

$$\langle \mathcal{F}(u) \mid \mathcal{F}(v) \rangle = \langle u \mid v \rangle$$

119.3.3 Convolution

La transformeée d'un produit de convolution s'écrit :

$$\mathcal{F}(u \star v)(y) = \int_{-\infty}^{+\infty} \exp(-2\pi\,\mathbf{i}\,xy)dx \int_{-\infty}^{+\infty} u(x-z)v(z)dz$$

Considérons le changement de variable :

$$\xi = x - z$$
$$\eta = z$$

On a alors $x = \xi + \eta$ et :

$$\mathcal{F}(u \star v)(y) = \int_{-\infty}^{+\infty} u(\xi) \exp(-2\pi \,\mathbf{i}\, \xi y) d\xi \int_{-\infty}^{+\infty} v(\eta) \exp(-2\pi \,\mathbf{i}\, \eta y) d\eta$$

c'est-à-dire :

$$\mathcal{F}(u \star v) = \mathcal{F}(u)\mathcal{F}(v)$$

119.3.4 Dérivées

Soit u une fonction qui s'annule à l'infini. La transformeée de Fourier de sa dérivée s'écrit :

$$\mathcal{F}\left(\frac{du}{dx}\right)(y) = \int_{-\infty}^{+\infty} \frac{du}{dx}(x) \exp(-2\pi \,\mathbf{i}\, xy) dx$$

En intégrant par partie, et en tenant compte du fait que les limites à l'infini sont nulles, il vient :

$$\mathcal{F}\left(\frac{du}{dx}\right)(y) = -(-2\pi \,\mathbf{i}\, y) \int_{-\infty}^{+\infty} u(x) \exp(-2\pi \,\mathbf{i}\, xy) dx$$

et finalement :

$$\mathcal{F}\left(\frac{du}{dx}\right)(y) = 2\pi \,\mathbf{i}\, y \mathcal{F}(y)$$

119.3.5 Déphasage

Considérons l'opérateur de translation :

$$t_a(u)(x) = u(x-a)$$

La transformeée de Fourier d'une translatée de la fonction u s'écrit :

$$\mathcal{F}\left[t_a(u)\right](y) = \int_{-\infty}^{+\infty} u(x-a) \exp(-2\pi \,\mathbf{i}\, xy) dx$$

$$\mathcal{F}\left[t_a(u)\right](y) = \exp(-2\pi \,\mathbf{i}\, ay) \int_{-\infty}^{+\infty} u(x-a) \exp(-2\pi \,\mathbf{i}(x-a)y) dx$$

Mais comme $dx = d(x-a)$, on en déduit que :

$$\mathcal{F}\left[t_a(u)\right](y) = \exp(-2\pi \,\mathbf{i}\, ay)\mathcal{F}(u)(y)$$

119.3.6 Dilatation

Soit d_a l'opérateur de dilatation :

$$d_a(u)(x) = u(ax)$$

où $a > 0$ est un réel strictement positif.
La transformeée de Fourier de la fonction dilaté $d_a(u)$ s'écrit :

$$\mathcal{F}\left[d_a(u)\right](y) = \int_{-\infty}^{+\infty} u(ax)\exp(-2\pi\,\mathbf{i}\,xy)dx$$

Considérons le changement de variable $z = ax$. On a alors :

$$\mathcal{F}\left[d_a(u)\right](y) = \frac{1}{a}\int_{-\infty}^{+\infty} u(z)\exp\left(-2\pi\,\mathbf{i}\,z\frac{y}{a}\right)dz$$

$$\mathcal{F}\left[d_a(u)\right](y) = \frac{1}{a}\mathcal{F}(u)(y/a)$$

c'est-à-dire :

$$\mathcal{F}\left[d_a(u)\right] = \frac{1}{a}\mathcal{F}\left[d_{1/a}(u)\right]$$

119.3.7 Gaussienne

Définissons la gaussienne G :

$$G(x) = \exp(-\pi x^2)$$

Sa transformeée s'écrit :

$$\mathcal{F}(G)(y) = \int_{-\infty}^{+\infty} \exp(-\pi x^2 - 2\pi\,\mathbf{i}\,xy)dx$$

Il est clair que :

$$\exp\left[-\pi(x+\mathbf{i}\,y)^2\right] = \exp(-\pi x^2 - 2\pi\,\mathbf{i}\,xy)\exp(\pi y^2)$$

Il vient donc, en effectuant le changement de variable $\xi = x + \mathbf{i}\,y$:

$$\mathcal{F}(G)(y) = \exp(-\pi y^2)\int_{-\infty}^{+\infty} \exp(-\pi\xi^2)d\xi$$

Mais nous avons vu au chapitre **??** que cette dernière intégrale vaut 1. Donc :

$$\mathcal{F}(G) = G$$

La transformeée de Fourier laisse G inchangée.

119.4 Transformeée de Fourier

La transformeée de Fourier (voir chapitre 119) est définie par :

$$\mathcal{F}(u)(y) = \int_{\mathbb{R}} u(x)\exp(2\pi\,\mathbf{i}\,xy)dx$$

où $\mathbf{i} = \sqrt{-1}$. On a :

$$\int_{\mathbb{R}} \mathcal{F}(u)(y)v(y)dy = \int_{\mathbb{R}} u(x)v(y)\exp(2\pi\,\mathbf{i}\,xy)dxdy$$

$$\int_{\mathbb{R}} \mathcal{F}(u)(y)v(y)dy = \int_{\mathbb{R}} u(x)\mathcal{F}(v)(x)dx$$

et donc :

$$\langle \mathcal{F}(u), v \rangle = \langle u, \mathcal{F}(v) \rangle$$

Chapitre 120

Ondelettes

Les ondelettes constitue un outil d'analyse de signal puissant. L'idée est de décomposer une fonction en différentes échelles, des échelles grossières jusqu'à des résolution très fines. Contrairement aux séries de Fourier, elles ont une influence locale et des relations simples entres les différentes échelles représentées, ce qui permet de passer rapidement (en $\mathcal{O}(N)$) d'une représentation à une autre.

Elles sont fort utilisées en compression de donnée (son, image, ...) où les coefficients les plus négligeables sont ignorés.

120.1 Introduction

AFAIRE : ARRANGER LE CHAPITRE

Le plus simple est de commencer avec un exemple célèbre : les ondelettes de Haar :

$$\varphi(x) = \boldsymbol{\delta}_{[0,1[}(x)$$
$$\psi(x) = \boldsymbol{\delta}_{[0,1/2[}(x) - \boldsymbol{\delta}_{[1/2,1[}(x)$$

La fonction φ est appelée fonction d'échelle, tandis que ψ est l'ondelette proprement dite. La propriété principale d'un système d'ondelette est qu'il permet facilement de changer d'échelle. Ainsi, on a :

$$\varphi(x) = \varphi(2x) + \varphi(2x-1) = a_0^0\varphi(2x) + a_1^0\varphi(2x-1)$$
$$\psi(x) = \varphi(2x) - \varphi(2x-1) = a_0^1\varphi(2x) + a_1^1\varphi(2x-1)$$

où l'on a posé $a_0^0 = 1$, $a_1^0 = 1$, $a_0^1 = 1$, $a_1^1 = -1$. En inversant les deux relations ci-dessus, on arrive à :

$$\varphi(2x) = \frac{1}{2}(\varphi(x) + \psi(x))$$
$$\varphi(2x-1) = \frac{1}{2}(\varphi(x) - \psi(x))$$

Les deux représentations ci-dessous sont donc équivalentes :

$$f(x) = f_1\varphi(2x) + f_2\varphi(2x-1)$$
$$f(x) = \frac{f_1+f_2}{2}\varphi(x) + \frac{f_1-f_2}{2}\psi(x)$$

La différence est que l'on représente f tantôt comme une somme de fonctions de base de même échelle, tantôt comme une superposition d'échelles différentes. En effet, $\varphi(2x), \varphi(2x-1), \psi(x)$ représentent l'échelle fine 1/2, tandis que $\varphi(x)$ représente l'échelle grossière 1 (un petit dessin peut aider).

Une autre propriété importante est que le support des fonctions φ et ψ est borné, et donc la fonction $\varphi_{jk} : x \mapsto \varphi(2^j x - k)$ à un support S_j de plus en plus fin lorsque j augmente. On dit que les ondelettes ont une influence locale.

120.2 Définition

Nous allons construire un système d'ondelettes générique à partir des coefficients :

$$a_k^i \quad : \quad i = 0, 1, ..., m-1 \qquad k, m \in \mathbb{Z}, m > 1$$

Adoptant la convention $\sum_k = \sum_{k\in\mathbb{Z}}$, nous demandons que ces coefficients vérifient les propriétés de normalisation :

$$\sum_k a_{p+mk}^0 = 1$$
$$\sum_k a_k^i = m\boldsymbol{\delta}_{i,0}$$

et d'orthogonalisation suivantes :

$$\sum_k a_{k+mr}^i a_{k+ms}^j = m\boldsymbol{\delta}_{ij}\boldsymbol{\delta}_{rs}$$
$$\sum_k \sum_{i=0}^{m-1} a_{r+mk}^i a_{s+mk}^i = m\boldsymbol{\delta}_{rs}$$

Par ailleurs, afin d'assurer le caractère compact des fonctions associées, on impose la nullité des coefficients en dehors d'un certain domaine :

$$a_k^i = 0 \qquad \forall k \notin \{0, 1, 2, ...mg - 1\}$$

où $g > 0$ est un entier fixé.

Nous allons voir dans la suite que ces propriétés permettent de montrer d'importants résultats sur le comportement des ondelettes associées. Remarquons déjà les conséquences directes suivantes :

$$\sum_k a_{p-mk}^0 = 1$$
$$\sum_k \sum_{i=0}^{m-1} a_{r-m.k}^i a_{s-m.k}^i = m\boldsymbol{\delta}_{rs}$$

Dans la suite, nous utilisons le produit scalaire :

$$\langle u \mid v \rangle = \int_{\mathbb{R}} u(x)v(x)dx$$

120.3 Construction

La fonction d'échelle s'obtient en partant de :

$$\varphi^{(0)}(x) = \boldsymbol{\delta}_{[0,1[}(x)$$

et en itérant :

$$\varphi^{(n)}(x) = \sum_k a_k^0 \varphi^{(n-1)}(mx - k)$$

Supposant que l'algorithme converge vers un point fixe (voir chapitre **??**), on définit la fonction d'échelle :

$$\varphi = \lim_{n\to\infty} \varphi^{(n)}$$

Qui possède la propriété :

$$\varphi(x) = \sum_k a_k^0 \varphi(mx - k)$$

Le système d'ondelettes est définit finalement par :

$$\varphi(x) = \sum_k a_k^0 \varphi(mx - k)$$
$$\psi^i(x) = \sum_k a_k^i \varphi(mx - k)$$

120.4 Propriétés

Les propriétés des a_k^i permettent de montrer, par récurrence sur n, que :

$$\int_{\mathbb{R}} \varphi(x)dx = 1$$
$$\int_{\mathbb{R}} \psi^i(x)dx = 0$$

En effet, on a $\int_{\mathbb{R}} \varphi^{(0)} = 1$ et si on suppose que $\int_{\mathbb{R}} \varphi^{(n-1)} = 1$ alors :

$$\int_{\mathbb{R}} \varphi^{(n)}(x)dx = \sum_k a_k^0 \int_{\mathbb{R}} \varphi^{(n-1)}(mx - k)dx$$
$$\int_{\mathbb{R}} \varphi^{(n)}(x)dx = \left(\sum_k a_k^0\right) \frac{1}{m} \int_{\mathbb{R}} \varphi^{(n-1)}(\xi)d\xi = 1$$

Par ailleurs on a :

$$\int_{\mathbb{R}} \psi^i(x)dx = \sum_k a_k^i \int_{\mathbb{R}} \varphi(mx - k)dx$$
$$\int_{\mathbb{R}} \psi^i(x)dx = \left(\sum_k a_k^i\right) \frac{1}{m} \int_{\mathbb{R}} \varphi(\xi)d\xi = 0$$

On peut aussi montrer par récurrence :

$$\mathrm{supp}(\varphi) \subseteq \left[0, 1 + (g-1)\sum_{k=0}^{+\infty} \left(\frac{1}{m}\right)^k\right] = \left[0, 1 + (g-1)\frac{m}{m-1}\right]$$

Le support de φ est donc bien borné.

Une autre propriété importante qui va nous permettre de prouver la convergence des développements en ondelettes est la partition de l'unité :

$$\sum_k \varphi(x - k) = 1 \qquad \forall x \in \mathbb{R}$$

En effet, il est clair que $\sum_k \varphi^{(0)}(x - k) = 1$. Partant de l'hypothèse de récurrence $\sum_k \varphi^{(n-1)}(x - k) = 1$, on obtient :

$$\sum_k \varphi^{(n)}(x-k) = \sum_k \sum_l a_l^0 \varphi^{(n-1)}\left(m(x-k)-l\right)$$
$$\sum_k \varphi^{(n)}(x-k) = \sum_l \sum_k a_l^0 \varphi^{(n-1)}(mx-mk-l)$$

Posant $p = mk + l$, cette dernière équation devient :

$$\sum_k \varphi^{(n)}(x-k) = \sum_p \varphi^{(n-1)}(mx-p) \sum_k a_{p-mk}^0$$

Mais comme $\sum_k a_{p-mk}^0 = 1$ quel que soit la valeur de p, on en déduit que :

$$\sum_k \varphi^{(n)}(x-k) = \sum_p \varphi^{(n-1)}(mx-p) = 1$$

et la propriété de partition de l'unité est démontréé en passant à la limite :

$$\sum_k \varphi(x-k) = \lim_{n\to+\infty} \sum_k \varphi^{(n)}(x-k) = 1$$

120.5 Niveaux

On définit la fonction d'échelle et les ondelettes du niveau $j \in \mathbb{N}$ comme étant les dilatations translatées des fonctions φ, ψ^i :

$$\varphi_{jk}(x) = m^{j/2}\varphi(m^j x - k)$$
$$\psi_{jk}^i(x) = m^{j/2}\psi^i(m^j x - k)$$

Le coefficient $m^{j/2}$ permet d'obtenir les normalisations : $\langle \varphi_{jk} \mid \varphi_{jk} \rangle = 1$ et $\left\langle \psi_{jk}^i \mid \psi_{jk}^i \right\rangle = 1$

La définition du système d'ondelettes permet d'obtenir les relations suivantes entres échelles voisines :

$$\varphi_{jk} = \frac{1}{\sqrt{m}} \sum_l a_l^0 \varphi_{j+1,mk+l} = \frac{1}{\sqrt{m}} \sum_p a_{p-mk}^0 \varphi_{j+1,p}$$
$$\psi_{jk}^i = \frac{1}{\sqrt{m}} \sum_l a_l^i \varphi_{j+1,mk+l} = \frac{1}{\sqrt{m}} \sum_p a_{p-mk}^i \varphi_{j+1,p}$$

120.6 Développement en série et algorithme de Mallat

120.6.1 Convergence

Commençons par remarquer que :

$$\sup_{\xi\in\mathbb{R}} \varphi_{jk}(\xi) = m^{j/2} \sup_{\xi\in\mathbb{R}} \varphi(\xi) = m^{j/2} M$$

où M ne dépend pas de j.

Soit la mesure de Lebesgue μ définie par :

$$\mu([a,b]) = |b-a|$$

On a alors :

$$\mu(\operatorname{supp}(\varphi_{jk})) = m^{-j}\mu(\operatorname{supp}(\varphi)) = m^{-j}\boldsymbol{\delta}_0$$

où $\boldsymbol{\delta}_0$ de dépend pas de j. Le support des fonctions φ_{jk} est donc de plus en plus fin lorsque j augmente.

Nous définissons la famille d'ensembles :

$$K_j(x) = \{k \in \mathbb{Z} : \varphi_{jk}(x) \neq 0\}$$

où $x \in \mathbb{R}$. On obtient, à partir des propriétés du support de φ, que :

$$m^j x - k \in \operatorname{supp}(\varphi) \Rightarrow m^j x - 2g \leq k \leq m^j x$$

pour tout $k \in K_j(x)$. On a donc :

$$\# K_j(x) \leq 2g$$

Pour un réel x fixé, le nombre de $\varphi_{jk}(x)$ non nuls est donc borné, et ce quelle que soit la valeur de $j \in \mathbb{N}$.

Choisissons $x, y \in \operatorname{supp}(\varphi_{jk})$. On a alors :

$$|x - y| \leq \boldsymbol{\delta}_0 m^{-j}$$

et par continuité de f :

$$|f(x) - f(y)| \leq \epsilon_j$$

avec :

$$\lim_{j \to +\infty} \epsilon_j = 0$$

Utilisant un simple changement de variable $\xi = m^j y - k$, on obtient :

$$\int_{\mathbb{R}} \varphi_{jk}(y) dy = \int_{\mathbb{R}} m^{j/2} \varphi\left(m^j y - k\right) dy$$
$$\int_{\mathbb{R}} \varphi_{jk}(y) dy = \int_{\mathbb{R}} m^{j/2} \varphi(\xi) m^{-j} d\xi$$
$$\int_{\mathbb{R}} \varphi_{jk}(y) dy = m^{-j/2} \int_{\mathbb{R}} \varphi(\xi) d\xi = m^{-j/2}$$

Comme $m^j x \in \mathbb{R}$, on a aussi :

$$\sum_k \varphi_{jk}(x) = m^{j/2} \sum_k \varphi(m^j x - k) = m^{j/2}$$

Donc :

$$\sum_k \varphi_{jk}(x) \int_{\mathbb{R}} \varphi_{jk}(y) dy = m^{-j/2} \sum_k \varphi_{jk}(x)$$
$$\sum_k \varphi_{jk}(x) \int_{\mathbb{R}} \varphi_{jk}(y) dy = m^{-j/2} m^{j/2} = 1$$

On en déduit en particulier que :

$$f(x) = \sum_k \varphi_{jk}(x) \int_{\mathbb{R}} f(x) \varphi_{jk}(y) dy$$

On aimerait bien montrer que les fonctions échelles de niveau j quelconque forment un cadre (voir chapitre **??**), c'est-à-dire que :

$$f = \sum_k \langle f \mid \varphi_{jk} \rangle \, \varphi_{jk}$$

Pour x fixé, seuls les k appartenant à $K_j(x)$ vont donner une contribution non nulle à la somme :

$$\sum_k \langle f \mid \varphi_{jk} \rangle \, \varphi_{jk}(x) = \sum_{k \in K_j(x)} \langle f \mid \varphi_{jk} \rangle \, \varphi_{jk}(x)$$

Rassemblant tous les résultats ci-dessus, on en déduit que :

$$\left| f(x) - \sum_k \langle f \mid \varphi_{jk} \rangle \, \varphi_{jk}(x) \right| = \left| \sum_{k \in K_j} \varphi_{jk}(x) \int_{\mathbb{R}} (f(x) - f(y)) \, \varphi_{jk}(y) dy \right|$$

$$\left| f(x) - \sum_k \langle f \mid \varphi_{jk} \rangle \, \varphi_{jk}(x) \right| \le 2gMm^{j/2} \epsilon_j \boldsymbol{\delta}_0 m^{-j} M m^{j/2} = 2gM^2 \boldsymbol{\delta}_0 \epsilon_j$$

qui tend bien vers 0 lorsque j tend vers l'infini.

120.6.2 Algorithme de Mallat

Considérons à présent l'approximation suivante :

$$f(x) \approx \tilde{f}(x) = \sum_k c_{Rk} \varphi_{Rk}(x)$$

où $c_{Rk} = \langle f \mid \varphi_{Rk} \rangle$. Nous allons maintenant montrer par récurrence que pour tout $J \le R$:

$$\tilde{f}(x) = \sum_k c_{Jk} \varphi_{Jk}(x) + \sum_k \sum_{i=1}^{m-1} \sum_{j=J}^{R} d^i_{jk} \psi^i_{jk}(x)$$

où :

$$c_{jk} = \langle f \mid \varphi_{jk} \rangle$$
$$d^i_{jk} = \left\langle f \mid \psi^i_{jk} \right\rangle$$

Il suffit pour cela de constater que :

$$\begin{aligned}
\sum_k c_{jk} \varphi_{jk} + \sum_k \sum_{i=1}^{m-1} d^i_{jk} \psi^i_{jk} &= \sum_k \langle f \mid \varphi_{jk} \rangle \, \varphi_{jk} + \sum_k \sum_{i=1}^{m-1} \left\langle f \mid \psi^i_{jk} \right\rangle \psi^i_{jk} \\
&= \frac{1}{m} \sum_{p,q} \sum_k \sum_{i=0}^{m-1} \langle f \mid \varphi_{j+1,p} \rangle \, \varphi_{j+1,q} a^i_{p-mk} a^i_{q-mk} \\
&= \sum_{p,q} \langle f \mid \varphi_{j+1,p} \rangle \, \varphi_{j+1,q} \frac{1}{m} \sum_k \sum_{i=0}^{m-1} a^i_{p-mk} a^i_{q-mk} \\
&= \sum_{p,q} \langle f \mid \varphi_{j+1,p} \rangle \, \varphi_{j+1,q} \boldsymbol{\delta}_{pq} \\
&= \sum_p \langle f \mid \varphi_{j+1,p} \rangle \, \varphi_{j+1,p}
\end{aligned}$$

en utilisant les propriétés des a^i_k. On peut donc construire le développement du niveau $j+1$ à partir des fonctions d'échelles et des ondelettes du niveau j. La définition des c_{jk}, d^i_{jk} combinée aux équations reliant les échelles voisines nous donne :

$$c_{jk} = \langle f \mid \varphi_{jk} \rangle = \frac{1}{\sqrt{m}} \sum_p \langle f \mid \varphi_{j+1,p} \rangle \, a^0_{p-mk}$$

$$c_{jk} = \frac{1}{\sqrt{m}} \sum_p c_{j+1,p} a^0_{p-m.k} = \frac{1}{\sqrt{m}} \sum_l c_{j+1,l+mk} a^0_l$$

ainsi que :

$$d^i_{jk} = \langle f \mid \psi^i_{jk} \rangle = \frac{1}{\sqrt{m}} \sum_p \langle f \mid \varphi_{j+1,p} \rangle \, a^i_{p-mk}$$

$$d^i_{jk} = \frac{1}{\sqrt{m}} \sum_p c_{j+1,p} a^i_{p-mk} = \frac{1}{\sqrt{m}} \sum_l c_{j+1,l+mk} a^i_l$$

La relation inverse s'obtient à partir de :

$$\sum_p \langle f \mid \varphi_{j+1,p} \rangle \, \varphi_{j+1,p} = \sum_k c_{jk} \varphi_{jk} + \sum_k \sum_{i=1}^{m-1} d^i_{jk} \psi^i_{jk}$$

$$\sum_p c_{j+1,p} \varphi_{j+1,p} = \frac{1}{\sqrt{m}} \sum_p \varphi_{j+1,p} \left[\sum_k c_{jk} a^0_{p-mk} + \sum_k \sum_{i=1}^{m-1} d^i_{jk} a^i_{p-mk} \right]$$

En comparant les coefficients des $\varphi_{j+1,p}$, et comme cette équation doit être valable pour tout x, on a :

$$c_{j+1,p} = \frac{1}{\sqrt{m}} \sum_k c_{jk} a^0_{p-m.k} + \sum_k \sum_{i=1}^{m-1} d^i_{jk} a^i_{p-mk}$$

$$c_{j+1,p} = \frac{1}{\sqrt{m}} \sum_k c_{jk} a^0_{p-m.k} + \sum_k \sum_{i=1}^{m-1} d^i_{jk} a^i_{p-mk}$$

On voit donc que pour passer d'une représentation mono-échelle

$$f(x) = \sum_k c_{Rk} \varphi_{Rk}(x)$$

à une représentation multi-échelle

$$f(x) = \sum_k c_{Jk} \varphi_{Jk}(x) + \sum_k \sum_{i=1}^{m-1} \sum_{j=J}^{R} d^i_{jk} \psi^i_{jk}(x)$$

(ou l'inverse) il suffit d'appliquer successivement les équations ci-dessus reliant entre-eux les coefficients $c_{j+1,p}$, c_{jk}, d^i_{jk}, pour $j = R-1, R-2, ..., J$. En pratique, on approxime les coefficients du niveau de résolution maximal R par :

$$c_{Rk} \approx m^{R/2} f\left(\frac{k}{m^R}\right)$$

Lorsque f est assez régulière pour avoir la convergence :

$$f(x) = \lim_{R \to +\infty} \sum_k c_{Rk} \varphi_{Rk}(x)$$

on écrit :

$$f(x) = \sum_k c_{Jk} \varphi_{Jk}(x) + \sum_k \sum_{i=1}^{m-1} \sum_{j=J}^{+\infty} d^i_{jk} \psi^i_{jk}(x)$$

Douzième partie

Probabilités

Chapitre 121

Probabilité

121.1 Probabilité

Une probabilité $\mathbb{P}$ sur un ensemble d'événements Ω est une mesure définie sur $\mathcal{S} = \{A : A \subseteq \Omega\}$ et à valeurs dans $[0, 1]$:

$$\mathbb{P} : \mathcal{S} \mapsto [0, 1]$$

Cette probabilité doit vérifier la normalisation :

$$\mathbb{P}[\Omega] = 1$$

ainsi que l'additivité :

$$\mathbb{P}\left[\bigcup_i \Phi_i\right] = \sum_i \mathbb{P}[\Phi_i]$$

lorsque les ensembles Φ_i sont disjoints deux à deux :

$$\Phi_i \cap \Phi_j = \begin{cases} \Phi_i & i = j \\ \emptyset & i \neq j \end{cases}$$

On en déduit directement que :

$$\mathbb{P}[\Phi] = \mathbb{P}[\Phi \cup \emptyset] = \mathbb{P}[\Phi] + \mathbb{P}[\emptyset]$$

d'où $\mathbb{P}[\emptyset] = 0$.
La grandeur $\mathbb{P}[\Phi]$ peut s'interpréter comme la probabilité que l'un des événements de Φ se réalise.

121.2 Variable aléatoire

Une variable aléatoire X associe une valeur réelle a chaque élément de Ω. On a donc $X : \Omega \mapsto \mathbb{R}$.

121.3 Mesure induite

Etant donné une variable aléatoire X, on peut définir une mesure induite $\mathcal{L}_X : \mathfrak{P}(\mathbb{R}) \mapsto [0, 1]$, qui exprime la probabilité qu'un événement $\omega \in \Omega$ donne une valeur appartenant à un sous-ensemble $U \subseteq \mathbb{R}$:

$$\mathcal{L}_X(U) = \mathbb{P}\left[X^{-1}(U)\right] = \mathbb{P}[\{\omega \in \Omega : X(\omega) \in U\}]$$

121.3.1 Variables conjointes

La mesure induite par deux variables aléatoires X et Y se définit par :

$$\mathcal{L}_{X,Y}(D) = \mathbb{P}\left[\{\omega \in \Omega : (X(\omega), Y(\omega)) \in D\}\right]$$

pour tout $D \subseteq \mathbb{R}^2$.
On voit clairement que :

$$\mathcal{L}_X(U) = \mathcal{L}_{X,Y}(U \times \mathbb{R})$$
$$\mathcal{L}_Y(U) = \mathcal{L}_{X,Y}(\mathbb{R} \times U)$$

121.4 Collection induite

Soit X une variable aléatoire et $U \subseteq \mathbb{R}$. On définit le sous-ensemble de Ω :

$$\Theta(X, U) = \{\omega \in \Omega : X(\omega) \in U\}$$

ou de manière équivalente en utilisant la relation inverse X^{-1} :

$$\Theta(X, U) = X^{-1}(U)$$

La collection $\Lambda(X)$ induite par X est un ensemble regroupant les $\Theta(X, U)$ pour tous les sous-ensembles de $\mathbb{R}$:

$$\Lambda(X) = \{\Theta(X, U) : U \subseteq \mathbb{R}\}$$

Comme :

$$\Theta(X, \emptyset) = \emptyset$$
$$\Theta(X, \mathbb{R}) = \Omega$$

il est clair que l'on a $\emptyset, \Omega \in \Lambda(X)$ quelle que soit la variable aléatoire X.

Fonctions indicatrices

Si $\Phi \subseteq \Omega$ et $X = \boldsymbol{\delta}_\Phi$, on a :

$$\Theta(\boldsymbol{\delta}_\Phi, \{1\}) = \{\omega : \boldsymbol{\delta}_\Phi(\omega) = 1\} = \Phi$$
$$\Theta(\boldsymbol{\delta}_\Phi, \{0\}) = \{\omega : \boldsymbol{\delta}_\Phi(\omega) = 0\} = \Omega \setminus \Phi$$

De même, si un ensemble $U \subseteq \mathbb{R}$:
- ne contient ni 1 ni 0, on a $\Theta(\boldsymbol{\delta}_\Phi, U) = \emptyset$
- contient 1 et 0, on a $\Theta(\boldsymbol{\delta}_\Phi, U) = \Omega$
- contient 1 et pas 0, on a $\Theta(\boldsymbol{\delta}_\Phi, U) = \Phi$
- contient 0 et pas 1, on a $\Theta(\boldsymbol{\delta}_\Phi, U) = \Omega \setminus \Phi$

On a donc :

$$\Lambda(\boldsymbol{\delta}_\Phi) = \{\emptyset, \Omega, \Phi, \Omega \setminus \Phi\}$$

121.5 Espérance

L'espérance d'une variable aléatoire X est simplement une moyenne pondérée par les probablités que X prennent telle ou telle valeur :

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega)$$

121.5.1 Indicatrice

Notons que pour tout $\Phi \subseteq \Omega$, on a :

$$\begin{aligned} \mathbb{E}[\boldsymbol{\delta}_\Phi] &= \int_\Omega \boldsymbol{\delta}_\Phi \ d\mathbb{P} \\ &= \int_\Phi d\mathbb{P} \end{aligned}$$

et donc :

$$\mathbb{E}[\boldsymbol{\delta}_\Phi] = \mathbb{P}[\Phi]$$

121.5.2 Fonction d'une variable aléatoire

Pour toute fonction $G : \mathbb{R} \mapsto \mathbb{R}$, on a bien évidemment $G \circ X : \Omega \mapsto \mathbb{R}$ et on peut définir :

$$\mathbb{E}[G(X)] = \int_\Omega (G \circ X)(\omega) \ d\mathbb{P}(\omega)$$

121.5.3 Fonction de plusieurs variables aléatoires

De même, si X et Y sont deux variables aléatoires, pour toute fonction $G : \mathbb{R}^2 \mapsto \mathbb{R}$, on a évidemment $G(X, Y) \in \mathbb{R}$ et on peut définir :

$$\mathbb{E}[G(X,Y)] = \int_\Omega G\left(X(\omega), Y(\omega)\right) \ d\mathbb{P}(\omega)$$

Le cas particulier $G(X, Y) = a\ X + b\ Y$, où $a, b \in \mathbb{R}$, nous montre la linéarité de l'espérance, qui découle directement de celle de l'intégrale :

$$\mathbb{E}[a\ X + b\ Y] = a\ \mathbb{E}[X] + b\ \mathbb{E}[Y]$$

121.6 Espérance et mesure induite

Soit une variable aléatoire X et la fonction étagée $G : \mathbb{R} \mapsto \mathbb{R}$ définie pour tout $x \in \mathbb{R}$ par :

$$G(x) = \sum_i g_i \ \boldsymbol{\delta}_{A_i}(x)$$

où les A_i forment une partition de $\mathbb{R}$ et où les g_i sont supposés sans perte de généralité être des réels distincts. Soit la partition de Ω constituée des ensembles :

$$\Omega_i = X^{-1}(A_i) = \{\omega \in \Omega : X(\omega) \in A_i\}$$

On voit que $(G \circ X)(\omega) = g_i$ pour tout $\omega \in \Omega_i$. Calculons l'espérance de $G(X)$:

$$\begin{aligned}
\mathbb{E}[G(X)] &= \int_\Omega (G \circ X)(\omega)\ d\mathbb{P}(\omega) \\
&= \sum_i \int_{\Omega_i} (G \circ X)(\omega)\ d\mathbb{P}(\omega) \\
&= \sum_i \int_{\Omega_i} g_i\ d\mathbb{P}(\omega) \\
&= \sum_i g_i \int_{\Omega_i} d\mathbb{P}(\omega) \\
&= \sum_i g_i\ \mathbb{P}[\Omega_i]
\end{aligned}$$

Par définition de la mesure induite, on a :

$$\mathcal{L}_X(A_i) = \mathbb{P}\left[X^{-1}(A_i)\right] = \mathbb{P}[\Omega_i]$$

L'espérance de $G(X)$ peut donc s'exprimer comme :

$$\mathbb{E}[G(X)] = \sum_i g_i\ \mathcal{L}_X(A_i)$$

Mais le membre de droite n'est autre que l'intégrale de G sur $\mathbb{R}$ utilisant la mesure $\mathcal{L}_X$:

$$\mathbb{E}[G(X)] = \int_\mathbb{R} G(x)\ d\mathcal{L}_X(x)$$

Comme cette expression doit être valable pour toute fonction en escalier, on en conclut que :

$$\mathbb{E}[G(X)] = \int_\mathbb{R} G(x)\ d\mathcal{L}_X(x)$$

pour toute fonction intégrable G.

121.6.1 Identité

Le cas particulier $G = \mathrm{Id}$ nous donne :

$$\mathbb{E}[X] = \int_\mathbb{R} x\ d\mathcal{L}_X(x)$$

121.6.2 Densité

Si il existe une fonction $f_X : \mathbb{R} \mapsto \mathbb{R}$ telle que $d\mathcal{L}_X = f_X\ dx$, où dx correspond à la mesure de Lebesgue sur $\mathbb{R}$, on a :

$$\mathbb{E}[G(X)] = \int_\mathbb{R} G(x)\ f_X(x)\ dx$$

ainsi que :

$$\mathbb{E}[X] = \int_\mathbb{R} x\ f_X(x)\ dx$$

On nomme cette fonction f_X la densité de la variable aléatoire X.
Remarquons que f_X est positive par positivité de la mesure. Comme :

$$\mathbb{E}[1] = 1$$

on obtient la propriété de normalité :

$$\int_{\mathbb{R}} f_X(x)\ dx = 1$$

Variable aléatoire gaussienne

Une variable aléatoire est dite normale de paramètres μ, σ si sa fonction densité vérifie :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

121.6.3 Variables conjointes

Soit les variables aléatoires X, Y et la fonction étagée $G : \mathbb{R}^2 \mapsto \mathbb{R}$ définie pour tout $x, y \in \mathbb{R}$ par :

$$G(x,y) = \sum_i g_i\ \boldsymbol{\delta}_{A_i}(x,y)$$

où les A_i forment une partition de $\mathbb{R}^2$ et où les g_i sont supposés sans perte de généralité être des réels distincts. Soit la partition de Ω constituée des ensembles :

$$\Omega_i = \{\omega \in \Omega : (X(\omega), Y(\omega)) \in A_i\}$$

On voit que $G(X(\omega), Y(\omega)) = g_i$ pour tout $\omega \in \Omega_i$. Calculons l'espérance de $G(X,Y)$:

$$\begin{aligned}
\mathbb{E}\left[G(X,Y)\right] &= \int_\Omega G(X(\omega), Y(\omega))\ d\mathbb{P}(\omega) \\
&= \sum_i \int_{\Omega_i} G(X(\omega), Y(\omega))\ d\mathbb{P}(\omega) \\
&= \sum_i \int_{\Omega_i} g_i\ d\mathbb{P}(\omega) \\
&= \sum_i g_i \int_{\Omega_i} d\mathbb{P}(\omega) \\
&= \sum_i g_i\ \mathbb{P}\left[\Omega_i\right]
\end{aligned}$$

Par définition de la mesure induite, on a :

$$\mathcal{L}_{X,Y}(A_i) = \mathbb{P}\left[\Omega_i\right]$$

L'espérance de $G(X)$ peut donc s'exprimer comme :

$$\mathbb{E}\left[G(X,Y)\right] = \sum_i g_i\ \mathcal{L}_{X,Y}(A_i)$$

Mais le membre de droite n'est autre que l'intégrale de G sur $\mathbb{R}^2$ utilisant la mesure $\mathcal{L}_{X,Y}$:

$$\mathbb{E}\left[G(X,Y)\right] = \int_{\mathbb{R}^2} G(x,y)\ d\mathcal{L}_{X,Y}(x,y)$$

Comme cette expression doit être valable pour toute fonction en escalier, on en conclut que :

$$\mathbb{E}\left[G(X,Y)\right] = \int_{\mathbb{R}^2} G(x,y)\ d\mathcal{L}_{X,Y}(x,y)$$

pour toute fonction intégrable G.

121.6.4 Densité conjointe

Si il existe une fonction $f_{X,Y} : \mathbb{R}^2 \mapsto \mathbb{R}$ telle que $d\mathcal{L}_{X,Y} = f_{X,Y}\ dx\ dy$, où $dx\ dy$ correspond à la mesure de Lebesgue sur $\mathbb{R}^2$, on a :

$$\mathbb{E}\left[G(X,Y)\right] = \int_{\mathbb{R}^2} G(x,y)\ f_{X,Y}(x,y)\ dx\ dy$$

En considérant le cas particulier $G(X,Y) = X$, on obtient :

$$\begin{aligned} \mathbb{E}\left[X\right] &= \int_{\mathbb{R}^2} x\ f_{X,Y}(x,y)\ dx\ dy \\ &= \int_{\mathbb{R}} x\ \left[\int_{\mathbb{R}} f_{X,Y}(x,y)\ dy\right]\ dx \end{aligned}$$

En définissant la fonction associée f_X par :

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\ dy$$

on peut dès lors écrire l'espérance de X comme :

$$\mathbb{E}\left[X\right] = \int_{\mathbb{R}} x\ f_X(x)\ dx$$

En suivant le même déroulement pour $\mathbb{E}\left[Y\right]$, et en définissant :

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\ dx$$

on peut écrire l'espérance de Y comme :

$$\mathbb{E}\left[Y\right] = \int_{\mathbb{R}} y\ f_Y(y)\ dy$$

Distribution normale

On dit que les variables aléatoires $X_1, ..., X_N$ présentent une distribution normale multivariée si il existe :

$$\mu = (\mu_i)_i$$
$$\Theta = (\sigma_{ij})_{i,j}$$

tels que la fonction densité associée à $X = (X_1, ..., X_N)^T$ s'écrive :

$$f_X(x) = \frac{1}{2\pi^{n/2} \det A} \exp\left(-\frac{1}{2}(x-\mu)^T \cdot \Theta^{-1} \cdot (x-\mu)\right)$$

pour tout $x \in \mathbb{R}^N$. On a alors :

$$\mathbb{E}\left[X_i\right] = \mu_i$$
$$\operatorname{cov}\left(X_i, X_j\right) = \sigma_{ij}$$

On a aussi la fonction génératrice :

$$\Psi_X(u) = \exp\left(u^T \cdot \mu + \frac{1}{2} u^T \cdot \Theta^{-1} \cdot u\right)$$

pour tout $u \in \mathbb{R}^N$.

121.7 Fonction génératrice des moments

On définit le moment générateur d'une densité par :

$$\Psi_X(u) = \mathbb{E}\left[\exp(X \cdot u)\right]$$

L'intérêt de cette fonction est qu'elle permet de calculer facilement les espérances des puissances naturelles de X. En effet :

$$\frac{d^k\Psi_X}{du^k}(u) = \mathbb{E}\left[X^k \ \exp(X \cdot u)\right]$$

et donc :

$$\frac{d\Psi}{du}(0) = \mathbb{E}\left[X^k \ \exp(0)\right] = \mathbb{E}\left[X^k\right]$$

121.7.1 Variable gaussienne

A titre d'exemple, nous calculons le moment générateur associé à une densité gaussienne :

$$\Psi(u) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp(xu) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

On obtient en développant :

$$\begin{aligned} \Psi(u) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp\left(xu - \frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp(\mu u + \frac{1}{2}u^2\sigma^2) \int_{\mathbb{R}} \exp\left(-\frac{(x-(\mu+u\sigma^2))^2}{2\sigma^2}\right) dx \end{aligned}$$

Comme l'intégrale vaut $\sqrt{2\pi}\sigma$, on obtient finalement :

$$\Psi(u) = \exp(u\mu + \frac{1}{2}u^2\sigma^2)$$

121.8 Variance

La variance de X est la variation carrée moyenne de X autour de son espérance $\mathbb{E}[X]$:

$$\mathrm{var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Comme la variable $Z = (X - \mathbb{E}[X])^2$ est positive, son espérance doit également etre positive et $\mathrm{var}(X) \geq 0$.

En développant la définition et en utilisant la linéarité de l'espérance, on obtient :

$$\begin{aligned} \mathrm{var}(X) &= \mathbb{E}\left[X^2 - 2\ X \cdot \mathbb{E}[X] + \mathbb{E}[X]^2\right] \\ &= \mathbb{E}\left[X^2\right] - 2\ \mathbb{E}[X] \cdot \mathbb{E}[X] + \mathbb{E}[X]^2 \cdot \mathbb{E}[1] \\ &= \mathbb{E}\left[X^2\right] - 2\ \mathbb{E}[X]^2 + \mathbb{E}[X]^2 \end{aligned}$$

soit :

$$\mathrm{var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

Invariance sous translation

Notons que si X, Y sont deux variables aléatoires reliées par :

$$Y = X + a$$

où $a \in \mathbb{R}$, on a :

$$\begin{aligned}
\operatorname{var}(Y) &= \mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right] \\
&= \mathbb{E}\left[(X + a - \mathbb{E}[X + a])^2\right] \\
&= \mathbb{E}\left[(X + a - \mathbb{E}[X] - a)^2\right] \\
&= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \operatorname{var}(X)
\end{aligned}$$

La variance est donc invariante sous translation :

$$\operatorname{var}(X + a) = \operatorname{var}(X)$$

121.9 Covariance

La covariance de deux variables aléatoire X, Y se définit par :

$$\operatorname{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])]$$

En développant et en utilisant la linéarité de l'espérance, on obtient :

$$\begin{aligned}
\operatorname{cov}(X, Y) &= \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[Y] \cdot \mathbb{E}[X] + \mathbb{E}[X] \cdot \mathbb{E}[Y] \\
&= \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]
\end{aligned}$$

On voit également que la variance d'une variable aléatoire X n'est rien d'autre que sa covariance avec elle-même :

$$\operatorname{var}(X) = \operatorname{cov}(X, X)$$

Invariance sous translation

Suivant le même raisonnement que pour la variance, on considère les variables aléatoires W, X, Y, Z reliées par :

$$\begin{aligned}
W &= X + a \\
Z &= Y + b
\end{aligned}$$

où $a, b \in \mathbb{R}$. La covariance entre W et Z s'exprime alors :

$$\begin{aligned}
\operatorname{cov}(W, Z) &= \mathbb{E}[(W - \mathbb{E}[W])(Z - \mathbb{E}[Z])] \\
&= \mathbb{E}[(X + a - \mathbb{E}[X] - a)(Y + b - \mathbb{E}[Y] - b)] \\
&= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \operatorname{cov}(X, Y)
\end{aligned}$$

La covariance est donc invariante sous translation :

$$\operatorname{cov}(X + a, Y + b) = \operatorname{cov}(X, Y)$$

121.10 Variance d'une combinaison linéaire

Nous utilisons la notation :

$$X_0 = X - \mathbb{E}[X]$$

pour toute variable aléatoire X. Cette variables aléatoire X_0 a la propriété d'avoir une espérance nulle car :

$$\mathbb{E}[X_0] = \mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0$$

La variance d'une telle variable peut s'écrire :

$$\mathrm{var}(X_0) = \mathbb{E}\left[X_0^2\right] - \mathbb{E}[X_0]^2 = \mathbb{E}\left[X_0^2\right]$$

Quant à la covariance, elle s'écrit :

$$\mathrm{cov}(X_0, Y_0) = \mathbb{E}[X_0\ Y_0] - \mathbb{E}[X_0]\ \mathbb{E}[Y_0] = \mathbb{E}[X_0\ Y_0]$$

Soit les réels a, b. Par linéarité de l'espérance, on a :

$$\mathbb{E}[a\ X + b\ Y] = a\ \mathbb{E}[X] + b\ \mathbb{E}[Y]$$

La variance de la combinaison linéaire $a\ X + b\ Y$ s'écrit :

$$\begin{aligned}
\mathrm{var}(a\ X + b\ Y) &= \mathbb{E}\left[(a\ X + b\ Y - \mathbb{E}[a\ X + b\ Y])^2\right] \\
&= \mathbb{E}\left[(a\ X + b\ Y - a\ \mathbb{E}[X] - b\ \mathbb{E}[Y])^2\right] \\
&= \mathbb{E}\left[(a\ X_0 + b\ Y_0)^2\right]
\end{aligned}$$

En développant, on arrive à :

$$\begin{aligned}
\mathrm{var}(a\ X + b\ Y) &= \mathbb{E}\left[a^2\ X_0^2 + 2\ a\ b\ X_0\ Y_0 + b^2\ Y_0^2\right] \\
&= a^2\ \mathbb{E}\left[X_0^2\right] + 2\ a\ b\ \mathbb{E}[X_0\ Y_0] + b^2\ \mathbb{E}\left[Y_0^2\right]
\end{aligned}$$

et donc :

$$\mathrm{var}(a\ X + b\ Y) = a^2\ \mathrm{var}(X_0) + 2\ a\ b\ \mathrm{cov}(X_0, Y_0) + b^2\ \mathrm{var}(Y_0)$$

L'invariance sous translation nous permet alors d'écrire :

$$\mathrm{var}(a\ X + b\ Y) = a^2\ \mathrm{var}(X) + 2\ a\ b\ \mathrm{cov}(X, Y) + b^2\ \mathrm{var}(Y)$$

121.11 Produit scalaire

Nous allons voir que la covariance est un produit scalaire. Nous utilisons la notation :

$$X_0 = X - \mathbb{E}[X]$$

pour toute variable aléatoire X. Cette variables aléatoire X_0 a la propriété d'avoir une espérance nulle car :

$$\mathbb{E}[X_0] = \mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0$$

On en déduit que :

$$\mathrm{cov}\left(X_0, Y_0\right) = \mathbb{E}\left[X_0\ Y_0\right] - \mathbb{E}\left[X_0\right]\ \mathbb{E}\left[Y_0\right] = \mathbb{E}\left[X_0\ Y_0\right]$$

La symétrie est vérifiée :

$$\mathrm{cov}\left(Y_0, X_0\right) = \mathbb{E}\left[Y_0 \cdot X_0\right] = \mathbb{E}\left[X_0 \cdot Y_0\right] = \mathrm{cov}\left(X_0, Y_0\right)$$

En ce qui concerne le caractère défini positif, on a :

$$\mathrm{cov}\left(X_0, X_0\right) = \mathbb{E}\left[X_0^2\right] \ge 0$$

De plus, si X_0 est tel que $\mathrm{cov}\left(X_0, X_0\right) = 0$, on a :

$$\int_\Omega X_0^2\ d\mathbb{P}(\omega) = 0$$

ce qui entraîne la nullité essentielle $X_0 \approx 0$ sur Ω.
Soit les réels a, b. On voit que la linéarité est bien respectée :

$$\begin{aligned}
\mathrm{cov}\left(X_0, a\ Y_0 + b\ Z_0\right) &= \mathbb{E}\left[X_0\ \left(a\ Y_0 + b\ Z_0\right)\right] \\
&= a\ \mathbb{E}\left[X_0\ Y_0\right] + b\ \mathbb{E}\left[X_0\ Z_0\right] \\
&= a\ \mathrm{cov}\left(X_0, Y_0\right) + b\ \mathrm{cov}\left(X_0, Z_0\right)
\end{aligned}$$

Nous venons de montrer que la covariance est essentiellement un produit scalaire pour toute variable aléatoires à espérance nulles X_0, Y_0. Comme la covariance est invariante sous translation, on voit que :

$$\mathrm{cov}\left(X, Y\right) = \mathrm{cov}\left(X_0, Y_0\right)$$

est également un produit scalaire pour toutes variables aléatoires X, Y.

121.11.1 Cauchy-Schwartz

En appliquant l'inégalité de Cauchy-Schwartz à ce produit scalaire, on obtient :

$$\mathrm{cov}\left(X, Y\right)^2 \le \mathrm{cov}\left(X, X\right)\ \mathrm{cov}\left(Y, Y\right) = \mathrm{var}\left(X\right)\ \mathrm{var}\left(Y\right)$$

où, en prenant la racine :

$$\mathrm{cov}\left(X, Y\right) \le \sqrt{\mathrm{var}\left(X\right)\ \mathrm{var}\left(Y\right)}$$

121.12 Probabilité conditionnelle

On définit une nouvelle famille de probabilités :

$$\mathbb{P}\left[A|B\right] = \frac{\mathbb{P}\left[A \cap B\right]}{\mathbb{P}\left[B\right]}$$

où A, B sont des sous-ensembles quelconque de Ω, et où B est tel que :

$$\mathbb{P}\left[B\right] > 0$$

Comme $B \cap B = B$, on a :

$$\mathbb{P}\left[B|B\right] = 1$$

On est donc certain qu'un événement de B va se produire. En fait, pour tout ensemble C tel que $B \subseteq C$, on a $C \cap B = B$ et :

$$\mathbb{P}[C|B] = 1$$

On déduit de l'inégalité :

$$\mathbb{P}[A \cap B] \le \mathbb{P}[B]$$

que :

$$\mathbb{P}[A|B] \le 1$$

D'un autre coté, comme $\mathbb{P}[B] \le 1$, on a :

$$\mathbb{P}[A|B] \ge \mathbb{P}[A \cap B] \ge 0$$

L'additivité est également satisfaite :

$$\begin{aligned}
\mathbb{P}[\cup_i A_i|B] &= \frac{\mathbb{P}[(\cup_i A_i) \cap B]}{\mathbb{P}[B]} \\
&= \frac{\mathbb{P}[\cup_i (A_i \cap B)]}{\mathbb{P}[B]} \\
&= \sum_i \frac{\mathbb{P}[A_i \cap B]}{\mathbb{P}[B]} = \sum_i \mathbb{P}[A_i|B]
\end{aligned}$$

pour toute famille de A_i disjoints deux à deux. Les fonctions :

$$\mathbb{P}_B[A] = \mathbb{P}[A|B]$$

forment donc bien une famille de probabilités. On dit que $\mathbb{P}[A|B]$ est la probabilité conditionnelle de A sachant B.

Lorsque $B = \Omega$, on retrouve d'ailleurs :

$$\mathbb{P}[A|\Omega] = \mathbb{P}[A]$$

Indépendance

On dit que deux ensembles d'événements A et B sont indépendants si :

$$\mathbb{P}[A|B] = \mathbb{P}[A]$$

c'est-à-dire si :

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$$

Application

Une technique fréquemment employée pour évaluer $\mathbb{P}[A]$ est d'utiliser une partition $B_1, ..., B_n$ de Ω. Utilisant $A = A \cup \Omega$, on a alors :

$$\mathbb{P}[A] = \sum_i \mathbb{P}[A \cap B_i] = \sum_i \mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]$$

121.13 Espérance conditionnelle à un ensemble

Soit $A \subseteq \Omega$. On a vu que :

$$\mathbb{E}\left[\boldsymbol{\delta}_A\right] = \mathbb{P}\left[A\right]$$

pour toute fonction indicatrice d'un sous-ensemble A de Ω. Par analogie, on aimerait bien obtenir une expression d'une espérance conditionnelle vérifiant :

$$\mathbb{E}\left[\boldsymbol{\delta}_A | B\right] = \mathbb{P}\left[A|B\right]$$

pour un ensemble $B \subseteq \Omega$ donné vérifiant $\mathbb{P}\left[B\right] > 0$.

Soit $\Omega_1, ..., \Omega_N$ une partition de Ω et Z une variable aléatoire en escalier :

$$Z(\omega) = \sum_i Z_i \ \boldsymbol{\delta}_{\Omega_i}(\omega)$$

On voit que :

$$\begin{aligned} \mathbb{E}\left[Z|B\right] &= \sum_i Z_i \ \mathbb{E}\left[\boldsymbol{\delta}_{\Omega_i}|B\right] \\ &= \sum_i Z_i \ \mathbb{P}\left[\Omega_i|B\right] \end{aligned}$$

Or :

$$\mathbb{P}\left[\Omega_i|B\right] = \frac{\mathbb{P}\left[\Omega_i \cap B\right]}{\mathbb{P}\left[B\right]}$$

On a donc :

$$\mathbb{E}\left[Z|B\right] = \frac{1}{\mathbb{P}\left[B\right]} \sum_i Z_i \ \mathbb{P}\left[\Omega_i \cap B\right]$$

Considérons la nouvelle partition :

$$\begin{gathered} \Phi_i^+ = \Omega_i \cap B \\ \Phi_i^- = \Omega_i \cap (\Omega \setminus B) \end{gathered}$$

Comme $\Phi_i^+ \cup \Phi_i^- = \Omega_i$, on a clairement $\boldsymbol{\delta}_{\Phi_i^+} + \boldsymbol{\delta}_{\Phi_i^-} = \boldsymbol{\delta}_{\Omega_i}$ et on peut réexprimer Z comme :

$$Z(\omega) = \sum_i Z_i \ \boldsymbol{\delta}_{\Phi_i^+}(\omega) + \sum_i Z_i \ \boldsymbol{\delta}_{\Phi_i^-}(\omega)$$

L'expression de l'espérance conditionelle devient :

$$\mathbb{E}\left[Z|B\right] = \frac{1}{\mathbb{P}\left[B\right]} \left[\sum_i Z_i \ \mathbb{P}\left[\Phi_i^+ \cap B\right] + \sum_i Z_i \ \mathbb{P}\left[\Phi_i^- \cap B\right]\right]$$

Remarquons que par construction :

$$\begin{gathered} \Phi_i^+ \cap B = \Phi_i^+ \\ \Phi_i^- \cap B = \emptyset \end{gathered}$$

Par conséquent, les termes en $\mathbb{P}\left[\Phi_i^- \cap B\right]$ s'annulent et on a :

$$\mathbb{E}\left[Z|B\right] = \frac{1}{\mathbb{P}\left[B\right]} \sum_i Z_i \ \mathbb{P}\left[\Phi_i^+\right]$$

Mais comme $\bigcup_i \Phi_i^+ = B$, les Φ_i^+ forment une partition de B et on peut écrire cette expression sous la forme intégrale :

$$\mathbb{E}\left[Z|B\right] = \frac{\int_B Z \quad d\mathbb{P}}{\int_B \ d\mathbb{P}}$$

Comme cette relation doit être valable pour toute variable aléatoire en escalier Z, elle l'est également pour une variable aléatoire quelconque X :

$$\mathbb{E}\left[X|B\right] = \frac{\int_B X \quad d\mathbb{P}}{\int_B \ d\mathbb{P}}$$

Densité conditionnelle

Soient X, Y deux variables aléatoires. Un cas particulier important d'espérance conditionnelle est celui où :

$$B_y = \{\omega : Y(\omega) = y\}$$

On note alors :

$$\mathbb{E}\left[X|Y = y\right] = \mathbb{E}\left[X|B_y\right]$$

On remarque que :

$$(X, Y)(B_y) = \{(x, y) \in \mathbb{R}^2 : x \in \mathbb{R}\}$$

Par conséquent, si il existe une fonction densité $f_{X,Y}$ associée à X, Y, on peut écrire :

$$\begin{aligned}
\int_{B_y} X \ d\mathbb{P} &= \int_{(X,Y)(B_y)} x \ f_{X,Y}(x, y) \ dx \ dy \\
&= \int_{\mathbb{R}} x \ f_{X,Y}(x, y) \ dx
\end{aligned}$$

ainsi que :

$$\int_{B_y} d\mathbb{P} = \int_{\mathbb{R}} f_{X,Y}(x, y) \ dx$$

L'espérance conditionnelle s'écrit alors :

$$\mathbb{E}\left[X|Y = y\right] = \frac{\int_{\mathbb{R}} x \ f_{X,Y}(x, y) \ dx}{\int_{\mathbb{R}} f_{X,Y}(x, y) \ dx}$$

Donc, si on définit :

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{\int_{\mathbb{R}} f_{X,Y}(x, y) \ dx}$$

on a tout simplement :

$$\mathbb{E}\left[X|Y = y\right] = \int_{\mathbb{R}} x \ f_{X|Y}(x, y) \ dx$$

121.14 Espérance conditionnelle à une tribu

Tribu et espace fonctionnel

Soit $\Gamma \subseteq \mathfrak{P}(\Omega)$ une collection de sous-ensembles de Ω formant une tribu sur Ω (voir section **??**), et $\mathcal{F}(\Gamma)$ l'ensemble des variables aléatoires W telles que :

$$\Lambda(W) \subseteq \Gamma$$

où $\Lambda(W)$ est la collection induite par W.

Minimisation

L'espérance conditionnelle est construite comme le meilleur estimateur au sens des moindres carrés d'une variable aléatoire X sur $\mathcal{F}(\Gamma)$. Soit la fonctionnelle $I : \mathcal{F}(\Gamma) \mapsto \mathbb{R}$ représentant l'erreur :

$$I(Z) = \int_\Omega [Z(\omega) - X(\omega)]^2 \ d\mathbb{P}(\omega)$$

Nous allons minimiser I sur $\mathcal{F}(\Gamma)$. Pour ce faire, on utilise la technique du calcul variationnel (voir chapitre 103). On commence par définir :

$$J_W(\epsilon) = I(Z^* + \epsilon W) = \int_\Omega (Z^* + \epsilon W - X)^2 \ d\mathbb{P}$$

où la variable aléatoire Z^* est l'optimum recherché, et où $W \in \mathcal{F}(\Gamma)$, $\epsilon \in \mathbb{R}$. La dérivée s'écrit :

$$\frac{dJ_W}{d\epsilon}(\epsilon) = \int_\Omega 2(Z^* + \epsilon W - X)W \ d\mathbb{P} = 0$$

Comme celle-ci doit s'annuler en $\epsilon = 0$, on a :

$$\frac{dJ_W}{d\epsilon}(0) = \int_\Omega 2(Z^* - X)W \ d\mathbb{P} = 0$$

Autrement dit :

$$\int_\Omega W Z^* \ d\mathbb{P} = \int_\Omega W X \ d\mathbb{P}$$

équation qui doit être vérifiée pour tout $W \in \mathcal{F}(\Gamma)$.

Unicité

Nous supposons dorénavant que $\mathcal{F}(\Gamma)$ est un espace vectoriel. Soient $Z_1, Z_2 \in \mathcal{F}(\Gamma)$ des variables aléatoires qui minimisent tous deux la fonctionnelle I. On a :

$$\int_\Omega W Z_1 \ d\mathbb{P} = \int_\Omega W Z_2 \ d\mathbb{P} = \int_\Omega W X \ d\mathbb{P}$$

pour tout $W \in \mathcal{F}(\Gamma)$. Donc :

$$\int_\Omega W(Z_1 - Z_2) \ d\mathbb{P} = 0$$

Mais comme $Z_1 - Z_2 \in \mathcal{F}(\Gamma)$, il suffit de considérer le cas $W = Z_1 - Z_2$ pour avoir :

$$\int_\Omega (Z_1 - Z_2)^2 \ d\mathbb{P} = 0$$

On en conclut que $Z_1 = Z_2$ presque partout sur Ω. L'espérance conditionnelle est donc unique pour X et Γ donnés.

Définition

Forts de ces résultats, on définit l'espérance de X conditionnellement à la tribu Γ comme étant :

$$\mathbb{E}\left[X|\Gamma\right] = \arg\min_{Z\in\mathcal{F}(\Gamma)} \int_\Omega \left[Z - X\right]^2 \ d\mathbb{P}$$

On a donc :

$$\int_\Omega W \ \mathbb{E}\left[X|\Gamma\right] \ d\mathbb{P} = \int_\Omega W \ X \ d\mathbb{P}$$

pour tout $W \in \mathcal{F}(\Gamma)$.

Fonctions indicatrices

Soit un ensemble $\Phi \in \Gamma$. Les propriétés de Γ nous disent que $\Omega \setminus \Phi \in \Gamma$. Donc :

$$\Lambda(\boldsymbol{\delta}_\Phi) = \{\emptyset, \Omega, \Phi, \Omega \setminus \Phi\} \subseteq \Gamma$$

et $\boldsymbol{\delta}_\Phi \in \mathcal{F}(\Gamma)$. On en déduit que :

$$\int_\Omega \boldsymbol{\delta}_\Phi \ \mathbb{E}\left[X|\Gamma\right] \ d\mathbb{P} = \int_\Omega \boldsymbol{\delta}_\Phi \ X \ d\mathbb{P}$$

c'est-à-dire :

$$\int_\Phi \mathbb{E}\left[X|\Gamma\right] \ d\mathbb{P} = \int_\Phi X \ d\mathbb{P}$$

pour tout $\Phi \in \Gamma$.
Comme $\Omega \in \Gamma$, on a en particulier :

$$\int_\Omega \mathbb{E}\left[X|\Gamma\right] \ d\mathbb{P} = \int_\Omega X \ d\mathbb{P}$$

c'est-à-dire :

$$\mathbb{E}\left[\mathbb{E}\left[X|\Gamma\right]\right] = \mathbb{E}\left[X\right]$$

Variable aléatoire dans l'espace fonctionnel

Une conséquence directe de la définition de l'espérance conditionnelle est que si $Z \in \mathcal{F}(\Gamma)$, on a :

$$\int_\Omega (Z - Z)^2 \ d\mathbb{P} = 0$$

Par conséquent, Z minimise la fonctionnelle :

$$I(Y) = \int_\Omega (Y - Z)^2 \ d\mathbb{P} \ge 0$$

sur $\mathcal{F}(\Gamma)$ et :

$$\mathbb{E}\left[Z|\Gamma\right] = Z$$

Tour

Soit la tribu $\Delta \subseteq \Gamma$ et X une variable aléatoire et $W \in \mathcal{F}(\Delta)$. On a :

$$\Lambda(W) \subseteq \Delta \subseteq \Gamma$$

Par conséquent $W \in \mathcal{F}(\Gamma)$ et les équations suivantes sont vérifiées :

$$\int_\Omega W \ \mathbb{E}\left[X|\Delta\right] \ d\mathbb{P} = \int_\Omega W \ X \ d\mathbb{P}$$
$$\int_\Omega W \ \mathbb{E}\left[X|\Gamma\right] \ d\mathbb{P} = \int_\Omega W \ X \ d\mathbb{P}$$

On en déduit que :

$$\int_\Omega W \ \mathbb{E}\left[X|\Delta\right] \ d\mathbb{P} = \int_\Omega W \ \mathbb{E}\left[X|\Gamma\right] \ d\mathbb{P}$$

Comme cette dernière équation est valable pour tout $W \in \mathcal{F}(\Delta)$, on en déduit que $\mathbb{E}\left[X|\Delta\right]$ est le meilleur estimateur de $\mathbb{E}\left[X|\Gamma\right]$ sur $\mathcal{F}(\Delta)$. Ce qui revient à dire que :

$$\mathbb{E}\left[\mathbb{E}\left[X|\Gamma\right]|\Delta\right] = \mathbb{E}\left[X|\Delta\right]$$

Couple de variables aléatoires

Etant donné deux variables aléatoires X, Y, on définit :

$$\mathbb{E}\left[X|Y\right] = \mathbb{E}\left[X|\Lambda(Y)\right]$$

Comme $\Gamma = \Lambda(Y)$, l'espace $\mathcal{F}(\Gamma)$ est l'ensemble des variables aléatoires W telles que :

$$\Lambda(W) \subseteq \Lambda(Y)$$

121.15 Ensemble discret

Nous allons à présent considérer le cas particulier où l'ensemble des événements peut s'écrire comme :

$$\Omega = \{\omega_i : i \in \mathbb{N}\}$$

Nous notons p_i les probabilités associées aux singletons :

$$p_i = \mathbb{P}\left[\{\omega_i\}\right]$$

Étant donnée une variable aléatoire X, on note :

$$x_i = X(\omega_i)$$

L'espérance d'une telle variable s'écrit simplement :

$$\mathbb{E}\left[X\right] = \sum_i x_i \ p_i$$

Chapitre 122

Statistiques

122.0.1 Indépendance

On dit que les variables aléatoires X_1, X_2, ..., X_N sont indépendantes si :

$$\mathbb{E}\left[\prod_i X_i\right] = \prod_i \mathbb{E}\left[X_i\right]$$

On en déduit que :

$$\mathrm{cov}\left(X_i, X_j\right) = \mathrm{var}\left(X_i\right)\ \boldsymbol{\delta}_{ij}$$

et donc :

$$\mathrm{var}\left(\sum_i X_i\right) = \sum_i \mathrm{var}\left(X_i\right)$$

122.1 Echantillons

Nous nous intéressons dans la suite de ce chapitre à des échantillons de N variables aléatoires indépendantes $X_1, ..., X_N$ telles que :

$$\mathbb{E}\left[X_i\right] = \mu$$
$$\mathrm{cov}\left(X_i, X_j\right) = \sigma\ \boldsymbol{\delta}_{ij}$$

122.2 L'inégalité de Markov

Soit une variable aléatoire X. On définit la variable associée :

$$Y = \begin{cases} a^2 & \text{si } |X - b| \ge a \\ 0 & \text{si } |X - b| < a \end{cases}$$

Comme :

$$Y \le (X - b)^2$$

on a $\mathbb{E}\left[Y\right] \le \mathbb{E}\left[(X - b)^2\right]$. D'un autre coté :

$$\mathbb{E}\left[Y\right] = a^2\ \mathbb{P}\left[|X - b| \ge a\right]$$

Rassemblant ces deux résultats, on obtient la propriété :

$$\mathbb{P}\left[|X-b| \geq a\right] \leq \frac{1}{a^2}\ \mathbb{E}\left[(X-b)^2\right]$$

connue sous le nom d'inégalité de Markov.
Le cas particulier $b = \mathbb{E}\left[X\right]$ nous donne :

$$\mathbb{P}\left[|X-\mathbb{E}\left[X\right]| \geq a\right] \leq \frac{1}{a^2}\ \mathrm{var}\left(X\right)$$

122.3 La loi des grands nombres

Soit la moyenne :

$$M_N = \frac{1}{N}\sum_{i=1}^{N} X_i$$

On a :

$$\mathbb{E}\left[M_N\right] = \frac{1}{N}\ N\ \mu = \mu$$

L'indépendance entre les variables nous amène à :

$$\begin{aligned}
\mathrm{var}\left(M_N\right) &= \frac{1}{N^2}\mathrm{var}\left(\sum_i X_i\right) \\
&= \frac{1}{N^2}\sum_i \mathrm{var}\left(X_i\right) \\
&= \frac{1}{N^2}\ N\ \sigma^2
\end{aligned}$$

et donc :

$$\mathrm{var}\left(M_N\right) = \frac{\sigma^2}{N}$$

Soit $a > 0$. L'inégalité de Markov nous dit que :

$$\mathbb{P}\left[|M_N-\mu| \geq a\right] \leq \frac{\sigma^2}{a^2 N}$$

Soit à présent $\epsilon > 0$. Si on veut :

$$\mathbb{P}\left[|M_N-\mu| \geq a\right] \leq \frac{\sigma^2}{a^2 N}\ <\ \epsilon$$

il suffit de choisir :

$$N > \frac{\sigma^2}{a^2\epsilon}$$

On en conclut que :

$$\lim_{N\to+\infty}\mathbb{P}\left[M_N = \mu\right] = 1$$

122.4 Fréquence et probabilité

Appliquons la loi des grands nombres à la fonction indicatrice $\boldsymbol{\delta}_A$. On a alors $X_i = 1$ lorsque $\omega \in A$ et $X_i = 0$ lorsque $\omega \notin A$. La moyenne s'écrit donc :

$$M_N = \frac{n(A)}{N}$$

où $n(A)$ est le nombre de X_i valant 1, autrement dit le nombre d'événements ω appartenant à A. Comme :

$$\mu = \mathbb{E}\left[\boldsymbol{\delta}_A\right] = \mathbb{P}\left[A\right]$$

on en déduit que la fréquence $n(A)/N$ converge vers la probabilité de A :

$$\lim_{N\to+\infty} \mathbb{P}\left[\frac{n(A)}{N} = \mathbb{P}\left[A\right]\right] = 1$$

122.5 Estimateurs non biaisés

Soit une fonction $G : \mathbb{R}^n \mapsto \mathbb{R}$:

$$G : (X_1, ..., X_N) \mapsto G(X_1, ..., X_N)$$

On dit que $\hat{G} : \mathbb{R}^n \mapsto \mathbb{R}$ est un estimateur non biaisé de G si :

$$\mathbb{E}\left[\hat{G}\right] = \mathbb{E}\left[G\right]$$

122.6 Estimation des espérance et des variances

Soit :

$$M_N(X_1, ..., X_N) = \frac{1}{N}\sum_{i=1}^{N} X_i$$

La loi des grands nombres nous dit que :

$$\mathbb{E}\left[M_N\right] = \mu$$

La moyenne M_N est donc un estimateur non biaisé de l'espérance μ.
Soit les variables à espérances nulles :

$$X_i^* = X_i - \mu$$
$$M_N^* = M_N - \mu$$

On obtient directement :

$$M_N^* = \frac{1}{N}\sum_i X_i^*$$

On voit également que :

$$X_i - M_N = X_i - \mu + \mu - M_N = X_i^* - M_N^*$$

Donc :

$$\mathbb{E}\left[\sum_i (X_i - M_N)^2\right] = \mathbb{E}\left[\sum_i (X_i^* - M_N^*)^2\right]$$

En développant, on obtient successivement :

$$\begin{aligned}
\mathbb{E}\left[\sum_i (X_i - M_N)^2\right] &= \mathbb{E}\left[\sum_i (X_i^*)^2\right] - 2\ \mathbb{E}\left[M_N^* \sum_i X_i^*\right] + \mathbb{E}\left[(M_N^*)^2\right] \\
&= \sum_i \mathbb{E}\left[(X_i^*)^2\right] - 2\ N\ \mathbb{E}\left[(M_N^*)^2\right] + \mathbb{E}\left[(M_N^*)^2\right]
\end{aligned}$$

Mais comme :

$$\operatorname{var}(M_N^*) = \mathbb{E}\left[(M_N^*)^2\right] = \frac{\sigma^2}{N}$$
$$\mathbb{E}\left[(X_i^*)^2\right] = \operatorname{var}(X_i) = \sigma^2$$

l'expression devient :

$$\mathbb{E}\left[\sum_i (X_i - M_N)^2\right] = (N - 2 + 1)\ \sigma^2 = (N - 1)\ \sigma^2$$

On en conclut que :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - M_N)^2$$

est un estimateur non biaisé de la variance :

$$\mathbb{E}\left[S^2\right] = \sigma^2$$

122.7 Maximum de vraisemblance

Il s'agit de trouver les paramètres $\hat{\theta}$ (espérance, variance, ...) qui maximisent la vraisemblance :

$$V(\hat{\theta}) = \prod_i \mathbb{P}\left[\{\omega : X_i(\omega) = x_i\} | \theta = \hat{\theta}\right]$$

Notons que cela revient à maximiser :

$$\ln \prod_{i=1}^{N} \mathbb{P}\left[\{\omega : X_i(\omega) = x_i\} | \theta = \hat{\theta}\right] = \sum_{i=1}^{N} \ln \mathbb{P}\left[\{\omega : X_i(\omega) = x_i\} | \theta = \hat{\theta}\right]$$

ce qui est souvent plus facile.
En pratique, lorsque la fonction de densité f_θ est connue, on maximise :

$$\phi(\theta) = \sum_i \ln f_\theta(x_i)$$

en imposant :

$$\frac{\partial \phi}{\partial \theta}(\hat{\theta}) = 0$$

122.8 Echantillon de densité donnée

Il s'agit d'un algorithme permettant de générer N nombres aléatoires :

$$\{x_1, ..., x_N\}$$

suivant la densité f. Soit $\epsilon \geq 0$ une erreur maximale et $[a, b]$ tel que :

$$\int_a^b f(x)\ dx = 1 - \epsilon$$

Soit :

$$M = \sup_{x \in [a,b]} f(x)$$

et la génératrice :

$$\mathrm{rand}(a, b)$$

qui renvoie des variables aléatoires de densité uniforme sur $[a, b]$.
On part de $A_0 = \emptyset$. A chaque itération, on génére deux nombres de densités uniformes :

$$x = \mathrm{rand}(a, b)$$
$$y = \mathrm{rand}(0, M)$$

Afin de modifier cette densité, on n'ajoute x à la liste déjà obtenue :

$$A_i = A_{i-1} \cup \{x\}$$

que si $y < f(x)$. Autrement, on ne fait rien et on passe à l'itération suivante.
La comparaison de y et de $f(x)$ sert donc de filtre à l'algorithme.

Chapitre 123

Calcul stochastique

123.1 Processus stochastique

Un processus stochastique est une fonction :

$$X : [0, +\infty) \times \Omega \mapsto \mathbb{R}, \quad (t, \omega) \mapsto X(t, \omega)$$

On sous-entend souvent l'événement ω, et on note $X(t) = X(t, \omega)$.

123.2 Intégrale d'Ito

Il s'agit d'une intégrale utilisant un processus stochastique X comme mesure :

$$I(t) = \int_0^t f(s)\ dX(s) = \lim_{\delta \to 0} \sum_k f(t_k)(X(t_{k+1}) - X(t_k))$$

123.3 Variation quadratique

Soit $\delta > 0$ et les N temps $t_k = k \cdot \delta$ où $k = 0, ..., \lceil \frac{T}{\delta} \rceil$. On définit la variation quadratique d'une fonction f :

$$\langle f \rangle (T) = \lim_{\delta \to 0} \sum_k (f(t_{k+1}) - f(t_k))^2$$

Si la dérivée de f existe, la variation quadratique s'annule car :

$$(f(t_{k+1}) - f(t_k))^2 \to \delta^2\ \frac{df}{dt}(t_k)^2$$

Comme $\delta^2 \to \delta\ ds$, on a :

$$\langle f \rangle (T) = \lim_{\delta \to 0} \delta \int_0^T \left(\frac{df}{dt}(s) \right)^2 ds = 0$$

123.4 Variation conjointe

Considérons maintenant deux processus stochastiques X, Y. Nous définissons la variation conjointe $\langle X, Y \rangle$:

$$\langle X, Y \rangle (T) = \lim_{\delta \to 0} \sum_k (X(t_{k+1}) - X(t_k))\ (Y(t_{k+1}) - Y(t_k))$$

Dans le cas où les dérivées de X et de Y existent, on a évidemment : $\langle X, Y \rangle = 0$.

Pour une fonction $f : \mathbb{R}^2 \mapsto \mathbb{R}$ quelconque, nous avons :

$$df(X,Y) = f(X + dX, Y + dY) - f(X,Y)$$

Dans le cas particulier où $f(X,Y) = X \cdot Y$, cette expression se réduit à :

$$\begin{aligned} d(X \cdot Y) &= (X + dX) \cdot (Y + dY) - X \cdot Y \\ &= dX \cdot Y + X \cdot dY + dX \cdot dY \end{aligned}$$

Mais comme :

$$\langle X, Y \rangle (t) = \int_0^t dX \cdot dY$$

on a en définitive :

$$\begin{aligned} X(t)Y(t) - X(0)Y(0) &= \int_0^t d(XY)(s) \\ &= \int_0^t X(s)\ dY(s) + \int_0^t Y(s)\ dX(s) + \langle X, Y \rangle (t) \end{aligned}$$

123.5 Relations variations quadratiques - conjointes

La définition nous donne directement :

$$\langle X \rangle = \langle X, X \rangle$$

On peut aussi vérifier que :

$$(X + Y)^2 - (X - Y)^2 = 4\ X\ Y$$

d'où l'on déduit :

$$\langle X, Y \rangle = \frac{1}{4}(\langle X + Y \rangle - \langle X - Y \rangle)$$

123.6 Variation d'ordre quelconque

Soit $\delta > 0$ et les temps $t_k = k\delta$ où $k = 0, ..., \lceil \frac{T}{\delta} \rceil$. On définit la variation d'ordre n d'une fonction f :

$$\langle f \rangle^n (T) = \lim_{\delta \to 0} \sum_k (f(t_{k+1}) - f(t_k))^n$$

123.7 Calcul d'Ito

Soit une fonction $F : \mathbb{R}^n \mapsto \mathbb{R}$ et N processus stochastiques X_i dont les variations d'ordre $n \geq 3$ s'annulent. Soit $X = (X_1, ..., X_N)$. On peut écrire le développement en série de Taylor d'ordre 2 :

$$F(X + \Delta) - F(X) \approx \frac{\partial F}{\partial X}(X)\Delta + \frac{1}{2}\Delta^T \frac{\partial^2 F}{\partial X \partial X}(X)\Delta$$

En faisant tendre $\Delta \to 0$, on obtient :

$$dF = \frac{\partial F}{\partial X} \, dX + \frac{1}{2} \, dX^T \frac{\partial^2 F}{\partial X \partial X} \, dX$$

On a donc la formule de Ito pour une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$:

$$dF = \sum_i \frac{\partial F}{\partial X_i} \, dX_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 F}{\partial X_i \partial X_j} \, dX_i \, dX_j$$

Ce qui nous permet d'évaluer une variation de F :

$$F(X(t)) - F(X(0)) = \sum_i \int_0^t \frac{\partial F}{\partial X_i}(X(s)) \, dX_i(s) +$$
$$\frac{1}{2} \sum_{i,j} \int_0^t \frac{\partial^2 F}{\partial X_i \partial X_j}(X(s)) \, d\left\langle X_i, X_j \right\rangle(s)$$

123.7.1 Dérivées ordinaires

Dans le cas d'une seule variable, on a :

$$dF = \sum_i \frac{dF}{dX} \, dX + \frac{1}{2} \frac{d^2 F}{dX^2} \, dX \, dX$$

Ce qui nous permet d'évaluer une variation de F :

$$F(X(t)) - F(X(0)) = \sum_i \int_0^t \frac{dF}{dX}(X(s)) \, dX(s) +$$
$$\frac{1}{2} \int_0^t \frac{d^2 F}{dX^2}(X(s)) d\left\langle X \right\rangle(s)$$

123.8 Mouvement Brownien

Un mouvement brownien est un processus stochastique :

$$B : [0, +\infty) \times \Omega \mapsto \mathbb{R}, \quad (t, \omega) \mapsto B(t, \omega)$$

continu par rapport à t :

$$B_\omega : t \mapsto B(t, \omega) \in \mathrm{Cont}([0, +\infty))$$

De plus, si on définit :

$$\mathcal{B}_t : \omega \mapsto B(t, \omega)$$

on a la propriété d'indépendance des variations temporelles :

$$\mathrm{cov}\left(\mathcal{B}_u - \mathcal{B}_t, \mathcal{B}_t - \mathcal{B}_s\right) = 0$$

pour tout $s < t < u$ positifs. On demande aussi qu'une variation $\mathcal{B}_t - \mathcal{B}_s$ suive une loi normale d'espérance nulle et de variance $t - s$:

$$\mathbb{E}\left[\mathcal{B}_t - \mathcal{B}_s\right] = 0$$
$$\mathrm{var}\left(\mathcal{B}_t - \mathcal{B}_s\right) = t - s$$

Variation quadratique

Si les mouvements browniens sont continus, ils ne sont pas dérivables. Comme les variations sont normalement distribuées avec une moyenne nulle et une variance $t-s$, on en déduit (en utilisant par exemple le moment générateur des densités normales) :

$$\mathbb{E}\left[(\mathcal{B}_t-\mathcal{B}_s)^2\right]=t-s$$
$$\operatorname{var}\left((\mathcal{B}_t-\mathcal{B}_s)^2\right)=2\ (t-s)^2$$

La variation quadratique des mouvement browniens peut s'écrire :

$$\langle B_\omega\rangle(T)=\lim_{\delta\to 0}\sum_k(B_\omega(t_{k+1})-B_\omega(t_k))^2$$

Lorsque $\delta\to 0$, on a $N\to+\infty$ et la loi des grands nombres nous dit que chaque terme de la somme de droite converge vers la variance δ. Comme on a N termes, on obtient :

$$\sum_k B_\omega(t_{k+1})-B_\omega(t_k)\to N\delta=T$$

On a donc :

$$\langle\mathcal{B}_\omega\rangle(T)=T$$

Ce que l'on note symboliquement sous forme différentielle par :

$$dB(t)\cdot dB(t)=dt$$

Variations d'ordre quelconque

Les variations $\langle B\rangle^n$ d'un mouvement brownien s'annulent pour $n\geq 3$.

Multidimensionnel

Nous définissons un mouvement Brownien de dimension n comme une collection de n mouvements Browniens B_i indépendants et vérifiant :

$$\langle B_i,B_j\rangle(t)=\boldsymbol{\delta}_{ij}\cdot t$$

Calul d'Ito

Dans le cas de N mouvement browniens B_i, les équations d'Ito deviennent :

$$dF=\sum_i\frac{\partial F}{\partial X_i}dB_i+\frac{1}{2}\sum_{i,j}\frac{\partial^2 F}{\partial X_i\partial X_j}\ dB_i\ dB_j$$

Mais comme $dB_idB_j=d\langle B_i,B_j\rangle=\boldsymbol{\delta}_{ij}\ dt$, on a :

$$dF=\sum_i\frac{\partial F}{\partial X_i}\ dB_i+\frac{1}{2}\sum_i\frac{\partial^2 F}{\partial X_i\partial X_i}\ dt$$

Ce qui nous permet d'évaluer une variation de F :

$$F(X(t))-F(X(0))=\sum_i\int_0^t\frac{\partial F}{\partial X_i}(X(s))\ dX_i(s)+$$
$$\frac{1}{2}\sum_i\int_0^t\frac{\partial^2 F}{\partial X_i\partial X_i}(X(s))\ ds$$

Dérivée ordinaire

Le cas particulier unidimensionnel nous donne :

$$dF(B) = \frac{dF}{dX}(B)\ dB + \frac{1}{2}\frac{d^2F}{dX^2}(B)\ dB \cdot dB$$

Mais comme :

$$d\langle B \rangle = dB \cdot dB = dt$$

on a :

$$dF(B) = \frac{dF}{dX}(B)\ dB + \frac{1}{2}\frac{d^2F}{dX^2}(B)\ dt$$

et :

$$F(B(t)) - F(B(0)) = \int_0^t \frac{dF}{dX}(B(s))\ dB(s) + \frac{1}{2}\int_0^t \frac{d^2F}{dX^2}(B(s))\ ds$$

AFAIRE : PROCESSUS DE POISSON