

# Analyse pour Économistes

Aymeric BARADAT



# Table des matières

<b>1</b>	<b>Rappels et compléments d'analyse réelle</b>	<b>9</b>
1	Vocabulaire : borne supérieure et inférieure, maximum, minimum . . . . .	9
2	Lorsque $E$ est l'image d'une fonction . . . . .	11
3	Extrema locaux . . . . .	12
4	Extraction de suites et théorème de Bolzano-Weierstrass . . . . .	13
5	Quand peut-on être sûrs que les extremas globaux sont atteints ? . . . . .	16
6	Condition nécessaire du premier ordre . . . . .	16
7	Classes de fonctions dérivables . . . . .	19
8	Développements limités . . . . .	19
9	Condition suffisante du second ordre . . . . .	23
<b>2</b>	<b>La structure euclidienne de <math>\mathbb{R}^d</math></b>	<b>25</b>
1	Définition . . . . .	25
2	Orthogonalité . . . . .	28
3	Projections et symétries orthogonales . . . . .	31
4	Formes linéaires et produit scalaire . . . . .	34
<b>3</b>	<b>La topologie en dimension <math>d</math></b>	<b>35</b>
1	Limite et continuité en dimension $d$ . . . . .	35
2	Les objets de base de la topologie : les ouverts de $\mathbb{R}^d$ . . . . .	37
3	Leurs complémentaires : les fermés de $\mathbb{R}^d$ . . . . .	40
4	Continuité, ouverts et fermés . . . . .	41
5	Adhérence et intérieur . . . . .	42
6	La compacité . . . . .	43
<b>4</b>	<b>Compléments d'algèbre linéaire</b>	<b>45</b>
1	Le déterminant . . . . .	45
1.1	Intuition . . . . .	45
1.2	Formalisme . . . . .	46
1.3	Règles de calcul théoriques . . . . .	48
1.4	Outils de calcul pratique . . . . .	50
1.5	Un exemple . . . . .	51
2	Diagonalisation des matrices carrées . . . . .	51
2.1	C'est quoi la diagonalisation ? . . . . .	51
2.2	Polynôme caractéristique et valeurs propres . . . . .	53
2.3	Multiplicité . . . . .	54
2.4	Comment trouve-t-on les vecteurs propres ? . . . . .	56
3	Matrices symétriques et orthogonales . . . . .	58
3.1	Définitions, liens avec le produit scalaire . . . . .	58
3.2	Théorème spectral . . . . .	59

3.3	Matrices signées . . . . .	60
<b>5</b>	<b>Différentielle des fonctions de plusieurs variables</b>	<b>63</b>
1	Dérivées partielles des fonctions de plusieurs variables . . . . .	63
1.1	Définition et interprétation . . . . .	63
1.2	Condition nécessaire d'optimalité . . . . .	64
1.3	Dérivées partielles d'ordre supérieure . . . . .	64
2	La différentielle des fonctions de plusieurs variables . . . . .	67
2.1	Définition et lien avec les dérivées partielles . . . . .	67
2.2	Formule de Taylor d'ordre 1 . . . . .	70
2.3	Gradient d'une fonction . . . . .	70
3	La différentielle des fonctions à valeurs vectorielles . . . . .	70
3.1	Définition . . . . .	70
3.2	Opérations sur les différentielles . . . . .	73
3.3	Composition de fonctions différentiables . . . . .	74
4	Caulcul différentiel à l'ordre 2 . . . . .	79
4.1	Différentielle du gradient d'une fonction . . . . .	79
4.2	Formule de Taylor du second ordre . . . . .	80
4.3	Étude des points critiques d'une fonction de plusieurs variables . . . . .	81
<b>6</b>	<b>Le théorème des fonctions implicites</b>	<b>83</b>
1	Comment définir un domaine de dimension $p$ dans un espace de dimension $n$ ? . . . . .	83
1.1	La paramétrisation . . . . .	83
1.2	Les équations . . . . .	84
1.3	Les graphes de fonctions . . . . .	85
1.4	Bilan . . . . .	86
2	Le théorème des fonctions implicites . . . . .	86
2.1	Le théorème en dimension 2 . . . . .	86
2.2	Un exemple d'application . . . . .	87
2.3	Le théorème en dimension supérieure . . . . .	88
2.4	Un résultat très proche : le théorème d'inversion locale . . . . .	89
<b>7</b>	<b>Optimisation sous contraintes</b>	<b>91</b>
1	Optimisation sous contraintes d'égalité . . . . .	91
1.1	Le théorème des extrema liés . . . . .	91
1.2	Deux exemples . . . . .	93
1.3	La formulation lagrangienne . . . . .	94
1.4	Le théorème de l'enveloppe . . . . .	95
2	Optimisation sous contraintes d'inégalité . . . . .	96
2.1	Les conditions de Kuhn et Tucker . . . . .	96
2.2	Un exemple . . . . .	98
2.3	Quand il y a des contraintes de positivité . . . . .	99
2.4	Un exemple avec ce nouveau système . . . . .	100
<b>8</b>	<b>Éléments d'analyse des fonctions convexes</b>	<b>103</b>
1	Définitions . . . . .	103
2	Propriétés en dimension 1 . . . . .	104
2.1	Croissance du taux d'accroissement . . . . .	104
2.2	Continuité . . . . .	107
3	Généralisation en dimension supérieure . . . . .	108
4	Lien avec l'optimisation . . . . .	110
4.1	Optimisation libre . . . . .	110

	4.2	Optimisation sous contraintes . . . . .	111
5		Introduction à la dualité . . . . .	112
	5.1	Définition du problème dual . . . . .	113
	5.2	Cas d'un problème convexe . . . . .	113
<b>9</b>		<b>Introduction aux équations différentielles ordinaires</b>	<b>115</b>
1		Premières définitions . . . . .	115
	1.1	C'est quoi une équation différentielle? . . . . .	115
	1.2	C'est quoi une solution? . . . . .	115
	1.3	Que veut-on faire? . . . . .	116
	1.4	Vocabulaire : équations autonomes . . . . .	117
	1.5	Équations différentielles d'ordre supérieur . . . . .	117
2		Équation linéaires en dimension 1 . . . . .	118
	2.1	Cas autonome : la fonction exponentielle . . . . .	118
	2.2	Cas général . . . . .	119
	2.3	Équation avec second membre . . . . .	120
3		Équations différentielles linéaires en dimension supérieure . . . . .	121
	3.1	Cas autonome : exponentielle de matrices . . . . .	121
	3.2	Cas non autonome . . . . .	128
4		Équations différentielles non linéaires . . . . .	130
	4.1	Le théorème de Cauchy-Lipschitz . . . . .	130
	4.2	Contre-exemples quand le champ de vecteurs n'est pas lipschitzien . . . . .	131
	4.3	Dépendance aux données : le lemme de Grönwall . . . . .	133
	4.4	Le cas autonome : notion de stabilité non-linéaire . . . . .	134
	4.5	Flots de gradient . . . . .	135



# Introduction

Ce document est le polycopié du cours d'Analyse pour Économistes que j'ai eu le plaisir de donner entre septembre 2016 et juin 2019 aux élèves de L3 du département d'économie de l'ENS Ulm. L'objectif essentiel de ce cours consistait à introduire les outils nécessaires à l'optimisation sous contraintes, puis de développer la méthode des multiplicateurs de Lagrange. Plus précisément, on discute dans ce polycopié :

- l'optimisation en dimension 1,
- la structure euclidienne de  $\mathbb{R}^d$  c'est à dire l'étude du produit scalaire canonique sur  $\mathbb{R}^d$ ,
- les bases de la topologie,
- quelques éléments d'algèbre linéaire, en particulier la notion de déterminant, quelques éléments de diagonalisation des matrices carrés et le théorème spectral,
- la notion de différentielle des fonctions de plusieurs variables,
- le théorème des fonctions implicites, qui est un outil clé pour comprendre ce qu'est un ensemble de dimension  $n \leq d$  dans  $\mathbb{R}^d$ ,
- et enfin l'optimisation sous contraintes à proprement parlé.

J'ai décidé d'ajouter à cela un chapitre sur les fonctions convexes, qui sont les fonctions que l'on sait le mieux optimiser numériquement pour plusieurs raison que je tente d'explicitier, et un chapitre sur les équations différentielles ordinaires d'usage très courant pour modéliser des systèmes économiques ou de façon plus générale lorsque l'on souhaite décrire un phénomène dépendant du temps.

Si on ajoute les TD aussi disponible sur ma page web, ce cours s'étalait sur toute l'année en 24 séances de 2h.



# Chapitre 1

## Rappels et compléments d'analyse réelle

### 1 Vocabulaire : borne supérieure et inférieure, maximum, minimum

Dans tout cette partie,  $E$  désigne un sous-ensemble de  $\mathbb{R}$ .

**Définition 1.1** (Majorant, minorant).

- On dit  $M \in \mathbb{R}$  est un **majorant** de  $E$  si :

$$\forall a \in E, \quad a \leq M.$$

- On dit  $m \in \mathbb{R}$  est un **minorant** de  $E$  si :

$$\forall a \in E, \quad a \geq m.$$

- On dit que  $E$  est **majoré** s'il admet un majorant.

- On dit que  $E$  est **minoré** s'il admet un minorant.

**Définition 1.2** (Borne supérieure, borne inférieure).

- On dit que  $S$  est la borne supérieure de  $E$  et on note :

$$S = \sup E$$

si  $S$  est le **plus petit majorant** de  $E$ , c'est à dire si pour tout majorant  $M$  de  $E$  :

$$S \leq M.$$

- On dit que  $s$  est la borne inférieure de  $E$  et on note :

$$s = \inf E$$

si  $s$  est le **plus grand minorant** de  $E$ , c'est à dire si pour tout minorant  $m$  de  $E$  :

$$s \geq m.$$

*Remarque 1.3.* Si  $E = [0, 1]$ ,  $\sup E = 1$  et donc  $\sup E \in E$ . En revanche, si  $E = [0, 1[$ , on a toujours  $\sup E = 1$  mais cette fois  $\sup E \notin E$ .

**Théorème 1.4** (Admis). *Si  $E$  est non vide et majoré, il admet une borne supérieure. S'il est non vide et minoré, il admet une borne inférieure.*

*Remarque 1.5.* Ce n'est pas un théorème évident ! Sa démonstration fait intervenir la construction de  $\mathbb{R}$ . On peut remarquer que  $\mathbb{Q}$  n'a pas cette propriété :

$$A := \{a \in \mathbb{Q} \mid x^2 \leq 2\}$$

n'a pas de borne supérieure dans  $\mathbb{Q}$ . Sa borne supérieure dans  $\mathbb{R}$  est  $\sqrt{2} \notin \mathbb{Q}$ . Donc si  $a \in \mathbb{Q}$ , ou bien  $a < \sqrt{2}$ , et alors on peut trouver  $b \in A$  tel que  $a < b$  (et  $a$  ne majore pas  $A$ ), ou bien  $a > \sqrt{2}$  et alors on peut trouver un rationnel  $b < a$  tel que  $b > \sqrt{2}$  (alors  $a$  majore  $A$ , mais ce n'est pas le plus petit majorant). Il se trouve que  $\mathbb{R}$  est exactement l'ensemble des bornes supérieures de parties de  $\mathbb{Q}$ . C'est même plus ou moins comme ça qu'il est défini.

**Convention 1.6.**

- Si  $E$  n'est pas majoré, il n'admet pas de borne supérieure dans  $\mathbb{R}$ . On pose alors :

$$\sup E = +\infty.$$

- De même, si  $E$  n'est pas minoré, on pose alors :

$$\inf E = -\infty.$$

**Proposition 1.7** (La propriété la plus importante du sup). *Si  $E$  est non-vide, il existe une suite  $(a_n)_{n \in \mathbb{N}}$  d'éléments de  $E$  (qui peut être choisie croissante) telle que :*

$$\lim_{n \rightarrow +\infty} a_n = \sup E.$$

*De même, il existe une suite  $(a_n)_{n \in \mathbb{N}}$  d'éléments de  $E$  (qui peut être choisie décroissante) telle que :*

$$\lim_{n \rightarrow +\infty} a_n = \inf E.$$

*Ces résultats sont encore vrais si  $\sup E = +\infty$  ou si  $\inf E = -\infty$ .*

*Démonstration. Dans le cas du sup lorsque  $\sup E < +\infty$ .*

Soit  $n \in \mathbb{N}^*$ . Par définition du sup,

$$\sup E - \frac{1}{n}$$

n'est pas un majorant de  $E$ . Donc il existe  $a_n \in E$  tel que :

$$a_n > \sup E - \frac{1}{n}.$$

Par ailleurs, comme  $\sup E$  majore  $E$  :

$$a_n \leq \sup E.$$

On a alors pour tout  $n$  de  $\mathbb{N}^*$  :

$$\sup E - \frac{1}{n} < a_n \leq \sup E.$$

Donc par le théorème des gendarmes :

$$\lim_{n \rightarrow +\infty} a_n = \sup E.$$

□

**Définition 1.8.**

- Si la borne supérieure  $M$  de  $E$  est un élément de  $E$ , on dit que  $M$  est le maximum de  $E$  et on note :

$$M = \max E.$$

- Si la borne inférieure  $m$  de  $E$  est un élément de  $E$ , on dit que  $m$  est le minimum de  $E$  et on note :

$$m = \min E.$$

## 2 Lorsque $E$ est l'image d'une fonction

Soit  $f$  une fonction d'un ensemble  $A$  à valeurs réelles. On rappelle que :

$$f(A) := \{f(x) \mid x \in A\} \subset \mathbb{R}.$$

**Définition 2.1.** On définit comme suit le sup de  $f$  sur  $A$ , l'inf de  $f$  sur  $A$ , et lorsqu'ils existent le maximum (global) de  $f$  sur  $A$  et le minimum (global) de  $f$  sur  $A$  (on dit alors que  $f$  **atteint son maximum (resp. minimum)** sur  $A$ ) :

$$\begin{aligned}\sup_A f &:= \sup f(A), \\ \inf_A f &:= \inf f(A), \\ \max_A f &:= \max f(A), \\ \min_A f &:= \min f(A).\end{aligned}$$

De plus, si  $f$  atteint son maximum sur  $A$  et si  $x \in A$  est tel que :

$$f(x) = \max_A f,$$

on dit que  $x$  est un point de maximum (global) de  $f$  sur  $A$ . On définit de la même façon les points de minimum (global).

- Remarque 2.2.*
1. Je mets "global" entre parenthèses parce que par défaut, un maximum est global, donc on n'est pas obligé de le dire.
  2. Il peut y avoir plusieurs points de maximum comme le montre l'exemple suivant.

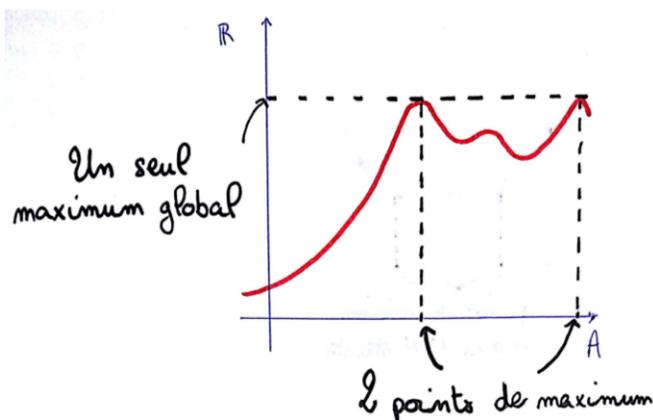


FIGURE 1.1 – Ici, il n'y a qu'un maximum, mais il est atteint deux fois.

3. En revanche, quand il n'y a qu'un point de maximum, on dit que c'est un point de maximum **strict**.

**Proposition 2.3** (La propriété la plus importante du sup d'une fonction). *Si  $A$  est non-vide, il existe une suite  $(x_n)_{n \in \mathbb{N}}$  d'éléments de  $A$  tels que :*

$$\lim_{n \rightarrow +\infty} f(x_n) = \sup_A f.$$

De même, il existe une suite  $(x_n)_{n \in \mathbb{N}}$  d'éléments de  $A$  tels que :

$$\lim_{n \rightarrow +\infty} f(x_n) = \inf_A f.$$

*Démonstration.* Dans le cas de l'inf, lorsque  $\inf_A f = -\infty$ .

Par la "propriété la plus importante du sup" et la définition de  $\inf f$ , il existe une suite  $(a_n)_{n \in \mathbb{N}}$  d'éléments de  $f(A)$  telle que :

$$\lim_{n \rightarrow +\infty} a_n = -\infty.$$

Or pour chaque  $n \in \mathbb{N}$ , je choisis  $x_n \in A$  tel que :

$$f(x_n) = a_n.$$

$(x_n)_{n \in \mathbb{N}}$  est bien une suite d'éléments de  $A$ , et :

$$\lim_{n \rightarrow +\infty} f(x_n) = \lim_{n \rightarrow +\infty} a_n = -\infty.$$

□

### 3 Extrema locaux

Maintenant, on suppose que  $A \subset \mathbb{R}$ , et même que  $A$  est un intervalle de  $\mathbb{R}$ .

**Définition 3.1.** On dit que  $f$  atteint un **maximum local** en un point  $x_0$  de  $A$  si il existe un réel strictement positif  $\varepsilon$  tel que :

$$\text{pour tout } x \in A \text{ tel que } |x - x_0| \leq \varepsilon, \text{ on a } f(x) \leq f(x_0).$$

On dit que  $f$  admet en  $x_0$  un **minimum local strict** si il existe  $\varepsilon > 0$  tel que :

$$\text{pour tout } x \in A \text{ tel que } x \neq x_0 \text{ et } |x - x_0| \leq \varepsilon, \text{ on a } f(x) < f(x_0).$$

On a une définition similaire pour les minima locaux.

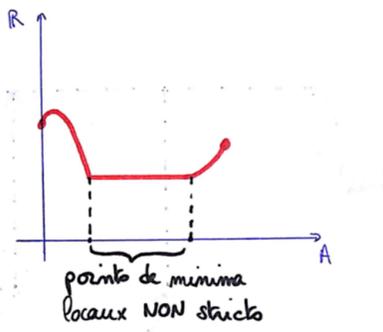


FIGURE 1.2 – Exemple de minimum local non strict.



FIGURE 1.3 – Exemples et contre-exemples d’extrema locaux d’une fonction.

## 4 Extraction de suites et théorème de Bolzano-Weierstrass

Soit  $E$  un ensemble et  $u = (u_n)_{n \in \mathbb{N}}$  une suite d’éléments de  $E$ . Une sous-suite (ou suite extraite) de la suite  $u$  est une suite  $v$  dans laquelle on ne garde que certaines valeurs de la suite  $u$ , toujours rangées dans le même ordre. On dit alors que l’on a extrait  $v$  de la suite  $u$ . Cette notion se formalise de la façon suivante.

**Définition 4.1.** La suite  $v = (v_n)_{n \in \mathbb{N}}$  est une **sous-suite** (ou suite extraite) de la suite  $u = (u_n)_{n \in \mathbb{N}}$  s’il existe une application **strictement croissante** :

$$\alpha : \mathbb{N} \longrightarrow \mathbb{N}$$

telle que pour tout  $n$  de  $\mathbb{N}$  :

$$v_n = u_{\alpha(n)}.$$

Une telle application  $\alpha$  s’appelle une extraction.

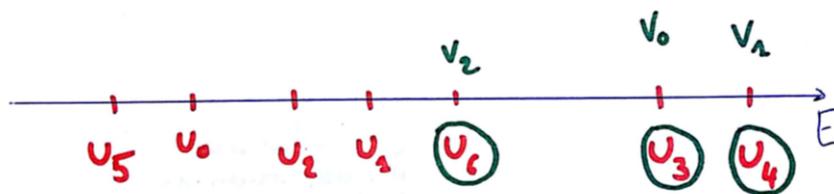


FIGURE 1.4 – Exemple de début de l’extraction d’une sous-suite : on choisit les points que l’on veut conserver et on les range dans une nouvelle suite, dans le même ordre.

*Exemple 4.2.* Soit  $u$  la suite définie pour tout  $n \in \mathbb{N}$  par :

$$u_n := (-1)^n.$$

C’est la suite qui vaut 1 quand  $n$  est pair et  $-1$  quand  $n$  est impair. Soit  $\alpha$  l’application de  $\mathbb{N}$  dans  $\mathbb{N}$  définie pour tout  $n \in \mathbb{N}$  par :

$$\alpha(n) := 2n.$$

$\alpha$  est strictement croissante donc la suite  $v$  définie pour tout  $n$  par :

$$v_n := u_{\alpha(n)} = u_{2n} = 1$$

est une sous-suite de  $u$ . De même, on aurait pu choisir  $\alpha(n) := 2n + 1$  et constater que la suite constante égale à  $-1$  est également une sous-suite de  $u$ .

**Définition 4.3.** Soit  $u = (u_n)_{n \in \mathbb{N}}$  une suite réelle. On dit que le réel  $\lambda$  est une **valeur d'adhérence** de la suite  $u$  si il existe une sous-suite  $v$  de  $u$  qui converge vers  $\lambda$ .

**Exercice 4.4** (facile). Quelles sont les valeurs d'adhérence de la suite  $u$  définie pour tout  $n \in \mathbb{N}$  par :

$$u_n = (-1)^n?$$

**Exercice 4.5** (difficile). On définit  $u = (u_n)_{n \in \mathbb{N}^*}$  de la façon suivante :

- $u_1 = 1$ .
- $u_2 = -1$  et  $u_3 = 0$ .
- $u_4 = 1$ ,  $u_5 = 1/2$ ,  $u_6 = 0$  et  $u_7 = -1/2$ .
- ...
- Pour tout  $n \in \llbracket 2^k, 2^{k+1} - 1 \rrbracket$  :

$$u_n = (-1)^k + \frac{(-1)^{k+1}}{2^{k-1}} \times (n - 2^k).$$

En particulier,  $u_{2^k} = (-1)^k$ .

- ...

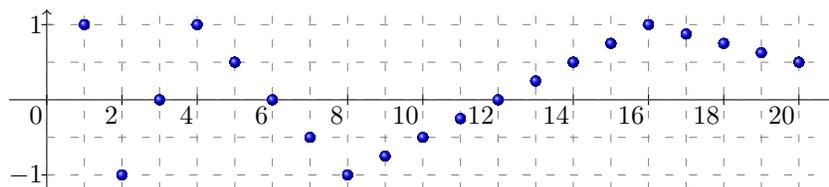


FIGURE 1.5 – Les premiers termes de la suite.

Quelles sont les valeurs d'adhérence de  $u$  ?

**Définition 4.6.** Une suite  $u = (u_n)_{n \in \mathbb{N}}$  est bornée s'il existe un réel positif  $M$  tel que pour tout  $n \in \mathbb{N}$  :

$$|u_n| \leq M.$$

**Théorème 4.7** (Bolzano-Weierstrass). *Toute suite de nombres réels bornée admet au moins une valeur d'adhérence. En d'autres termes toute suite réelle et bornée admet une sous-suite convergente.*

*Démonstration.* Soit  $u$  une suite réelle bornée. On pose  $M > 0$  telle que pour tout  $n \in \mathbb{N}$ ,

$$|u_n| \leq M.$$

Toutes les valeurs de la suite sont dans l'intervalle  $[-M, M]$ .

**Première étape.** On divise l'intervalle  $[-M, M]$ . Pour tout  $k \in \mathbb{N}$  et tout  $p \in \{-2^k, -2^k + 1, \dots, 2^k - 1\} = \llbracket -2^k, 2^k - 1 \rrbracket$ , on pose

$$I_k^p := \left[ \frac{p}{2^k} M, \frac{p+1}{2^k} M \right].$$

**Deuxième étape.** On montre par récurrence le résultat suivant : pour tout  $k \in \mathbb{N}$ , il existe  $p(k) \in \llbracket -2^k, 2^k - 1 \rrbracket$  tel que

- $I_k^{p(k)}$  contient une infinité de termes de la suite  $u$ ,
- la suite  $(p(k)M/2^k)_{k \in \mathbb{N}}$  est croissante,
- la suite  $((p(k)+1)M/2^k)_{k \in \mathbb{N}}$  est décroissante.

Initialisation. Si  $k = 0$ ,  $I_0^{-1} = [-M, 0]$  et  $I_0^0 = [0, M]$ . Donc

$$I_0^{-1} \cup I_0^0 = [-M, M].$$

Donc soit l'un soit l'autre contient une infinité de termes de la suite  $u$ . Si c'est  $I_0^{-1}$ , on pose  $p(0) = -1$ , si c'est  $I_0^0$ , on pose  $p(0) = 0$ .

Hérédité. On suppose que pour un certain  $k \in \mathbb{N}$ , on connaisse  $p(k)$  tel que  $I_k^{p(k)}$  contient une infinité de termes de la suite  $u$ . On a

$$\begin{aligned} I_k^{p(k)} &= \left[ \frac{p(k)}{2^k} M, \frac{p(k)+1}{2^k} M \right] \\ &= \left[ \frac{2p(k)}{2^{k+1}} M, \frac{2p(k)+1}{2^{k+1}} M \right] \cup \left[ \frac{2p(k)+1}{2^{k+1}} M, \frac{2(p(k)+1)}{2^{k+1}} M \right] \\ &= I_{k+1}^{2p(k)} \cup I_{k+1}^{2p(k)+1}. \end{aligned}$$

Donc parmi  $I_{k+1}^{2p(k)}$  et  $I_{k+1}^{2p(k)+1}$ , au moins l'un des deux contient une infinité de termes de la suite. Si c'est  $I_{k+1}^{2p(k)}$ , on pose  $p(k+1) := 2p(k)$ , si c'est  $I_{k+1}^{2p(k)+1}$ , on pose  $p(k+1) := 2p(k) + 1$ .

Ceci définit donc une suite  $(p(k))_{k \in \mathbb{N}}$  telle que pour tout  $k \in \mathbb{N}$ ,  $I_k^{p(k)}$  contient une infinité de termes de la suite.

Comme par construction, pour tout  $k \in \mathbb{N}$

$$I_{k+1}^{p(k+1)} = \left[ \frac{p(k+1)}{2^{k+1}} M, \frac{p(k+1)+1}{2^{k+1}} M \right] \subset \left[ \frac{p(k)}{2^k} M, \frac{p(k)+1}{2^k} M \right] = I_k^{p(k)},$$

il est clair que

$$\frac{p(k+1)}{2^{k+1}} M \geq \frac{p(k)}{2^k} M \quad \text{et} \quad \frac{p(k+1)+1}{2^{k+1}} M \leq \frac{p(k)+1}{2^k} M,$$

ce qui donne les propriétés de monotonie annoncées.

**Troisième étape.** Pour tout  $k \in \mathbb{N}$ , on définit

$$s_k := \frac{p(k)}{2^k} M \quad \text{et} \quad t_k := \frac{p(k)+1}{2^k} M.$$

En particulier, pour tout  $k \in \mathbb{N}$ ,  $I_k^{p(k)} = [s_k, t_k]$ . D'après l'étape précédente,  $s$  est croissante et  $t$  est décroissante. De plus, pour tout  $k \in \mathbb{N}$ ,

$$s_k \leq t_k \quad \text{et} \quad t_k - s_k = \frac{1}{2^k} \xrightarrow[k \rightarrow +\infty]{} 0.$$

Donc  $s$  et  $t$  sont adjacentes, et par un théorème bien connu, elles convergent vers la même limite que l'on note  $\lambda$ .

**Quatrième étape.** On construit l'extraction par récurrence.

Initialisation. On pose  $\alpha(0)$  le plus petit  $n$  tel que  $u_n \in I_0^{p(0)} = [s_0, t_0]$ .

Hérédité Si pour un certain  $k \in \mathbb{N}$ ,  $\alpha(k)$  est construit, on pose  $\alpha(k+1)$  le plus petit  $n \in \mathbb{N}$  tel que

$$\begin{aligned} n &> \alpha(k), \\ u_n &\in I_k^{p(k)} = [s_k, t_k]. \end{aligned}$$

L'application  $\alpha$  ainsi construite est strictement croissante, et pour tout  $k$ ,

$$s_k \leq u_{\alpha(k)} \leq t_k.$$

La suite  $v := (u_{\alpha(k)})_{k \in \mathbb{N}}$  est donc une suite extraite de  $u$  et par le théorème des gendarmes, elle converge vers  $\lambda$ . Le résultat est donc démontré.  $\square$

## 5 Quand peut-on être sûrs que les extremas globaux sont atteints ?

Le théorème suivant nous donne une condition nécessaire pour qu'une fonction réelle continue définie sur un intervalle atteigne ses extrema globaux.

**Théorème 5.1.** Soient deux réels  $a$  et  $b$  avec  $a < b$ , et soit  $f$  une fonction **continue** de  $[a, b]$  dans  $\mathbb{R}$ . Alors  $f$  atteint son minimum et son maximum.

*Démonstration.* On montre que  $f$  atteint son maximum, il suffira alors d'appliquer ce résultat à  $-f$  pour montrer que  $f$  atteint également son minimum.

Par la "propriété la plus importante du sup d'une fonction", il existe  $(x_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $[a, b]$  telle que :

$$\lim_{n \rightarrow +\infty} f(x_n) = \sup_{[a,b]} f.$$

Or  $(x_n)_{n \in \mathbb{N}}$  est bornée, donc par le théorème de Bolzano-Weierstrass, il existe  $\alpha$  de  $\mathbb{N}$  dans  $\mathbb{N}$  strictement croissante et un réel  $\lambda$  tels que :

$$\lim_{n \rightarrow +\infty} x_{\alpha(n)} = \lambda.$$

Le point clé, c'est que comme  $[a, b]$  est fermé en  $a$  et en  $b$ ,  $\lambda \in [a, b]$ . Je ne le démontre pas en toute rigueur mais il est **fondamental** de s'en convaincre. Et comme alors  $f$  est continue en  $\lambda$  :

$$\lim_{n \rightarrow +\infty} f(x_{\alpha(n)}) = f(\lambda).$$

Or  $(f(x_{\alpha(n)}))_{n \in \mathbb{N}}$  est une sous-suite de  $(f(x_n))_{n \in \mathbb{N}}$ , et par le un exercice du TD 2 :

$$\lim_{n \rightarrow +\infty} f(x_{\alpha(n)}) = \sup_{[a,b]} f.$$

Donc :

$$\sup_{[a,b]} f = f(\lambda),$$

et  $f$  atteint bien son maximum en  $\lambda$ .  $\square$

## 6 Condition nécessaire du premier ordre

On va voir ici le truc que tout le monde connaît : aux points d'extrema, la dérivée s'annule. Mais ce théorème a besoin d'**hypothèses**, et n'est vrai que lorsque ces hypothèses sont réalisées.

Si on demande de trouver le minimum d'une fonction, on ne peut pas se contenter de la dériver.

On choisit  $I$  un intervalle de  $\mathbb{R}$  et  $f$  une fonction de  $I$  dans  $\mathbb{R}$ .

**Définition 6.1.** Le point  $x \in I$  est un **point critique** de  $f$  si :

$$\begin{aligned} f \text{ est dérivable en } x, \\ f'(x) = 0. \end{aligned}$$

**Théorème 6.2.** Si  $x$  est un point **intérieur** de  $I$  (i.e. pas une de ses bornes), si  $f$  admet en  $x$  un **extremum local** et si  $f$  est dérivable en  $x$ , alors :

$$f'(x) = 0.$$

En particulier, si  $f$  est dérivable sur tout  $I$ , tout **extremum local** est atteint **soit** en un **point critique**, **soit** en une **borne** de  $I$ .

*Démonstration.* Soit un point  $x$  de  $I$  qui satisfait les conditions de l'énoncé, et supposons par exemple que  $x$  soit un point de minimum local de  $f$ . Soit  $\varepsilon > 0$  tel que pour tout  $y \in I$  tel que  $|x - y| \leq \varepsilon$ , on ait :

$$f(x) \leq f(y).$$

Comme  $x$  est un point intérieur de  $I$ , on peut trouver  $(u_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $I$  qui converge **en croissant** vers  $x$  (en particulier, pour tout  $n$ ,  $x \geq u_n$ ) et  $(v_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $I$  qui converge **en décroissant** vers  $x$  (en particulier, pour tout  $n$ ,  $x \leq v_n$ ). Par définition de la dérivée :

$$f'(x) = \lim_{n \rightarrow +\infty} \frac{f(x) - f(u_n)}{x - u_n} = \lim_{n \rightarrow +\infty} \frac{f(v_n) - f(x)}{v_n - x}.$$

Or si  $n$  est suffisamment grand,  $|x - u_n| \leq \varepsilon$ , donc :

$$f(x) \leq f(u_n),$$

et donc :

$$\frac{f(x) - f(u_n)}{x - u_n} \leq 0.$$

En passant à la limite, on obtient :

$$f'(x) \leq 0.$$

De la même façon, pour  $n$  suffisamment grand,  $|v_n - x| \leq \varepsilon$ , donc :

$$f(x) \leq f(v_n),$$

et donc :

$$\frac{f(v_n) - f(x)}{v_n - x} \geq 0.$$

En passant à la limite, on obtient cette fois :

$$f'(x) \leq 0.$$

Les deux conditions que l'on a obtenues ne sont compatibles que si :

$$f'(x) = 0.$$

Pour prouver la seconde partie de l'énoncé, on suppose  $f$  dérivable sur tout  $I$ , et on choisit  $x$  un point d'extremum local.

- Si  $x$  est un point intérieur de  $I$ , comme  $f$  est dérivable en  $x$ , les conditions du premier point sont remplies, donc  $f'(x) = 0$  et donc  $x$  est un point critique.
- Sinon,  $x$  est une borne de  $I$ .

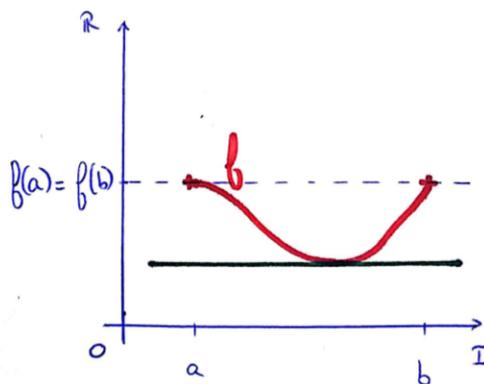


FIGURE 1.6 – Illustration du théorème de Rolle :  $f$  admet une tangente horizontale **en un point intérieur**.

Le théorème est donc démontré. □

**Exercice 6.3.** Soient  $a$  et  $b$  deux réels tels que  $a < b$ .

1. Soit  $f$  une fonction continue de l'intervalle  $[a, b]$  à valeurs dans  $\mathbb{R}$ , dérivable sur  $]a, b[$  et vérifiant :

$$f(a) = f(b).$$

- i) Montrer que ou bien  $f$  est constante, ou bien  $\max_{[a,b]} f > f(a)$ , ou bien  $\min_{[a,b]} f < f(a)$ .

- ii) En déduire le **théorème de Rolle** :

*Il existe un point  $c \in ]a, b[$  tel que  $f'(c) = 0$ .*

2. Soit  $g$  une fonction continue de l'intervalle  $[a, b]$  à valeurs dans  $\mathbb{R}$ , dérivable sur  $]a, b[$ . Montrer que la fonction  $f$  définie pour tout  $x$  de  $[a, b]$  par :

$$f(x) := g(x) - \frac{g(b) - g(a)}{b - a} \cdot (x - a),$$

satisfait les hypothèses du théorème de Rolle.

3. En déduire le **théorème des accroissements finis** :

*Il existe un point  $c \in ]a, b[$  tel que :*

$$g'(c) = \frac{g(b) - g(a)}{b - a}.$$

4. On prend maintenant  $h$  une fonction réelle sur  $[a, b]$ , continue sur  $[a, b]$  et dérivable sur  $]a, b[$ . On suppose qu'il existe un réel  $m$  tel que pour tout  $x \in ]a, b[$  :

$$h'(x) \geq m.$$

Montrer l'**inégalité des accroissements finis** :

$$h(b) - h(a) \geq m(b - a).$$

*Remarque 6.4.* On aurait bien entendu pu avoir toutes les inégalités dans l'autre sens.

5. **Application** : en déduire que si  $h$  une fonction réelle sur  $[a, b]$ , continue sur  $[a, b]$  et dérivable sur  $]a, b[$ ,  $h$  est croissante si et seulement si  $h'$  est positive.

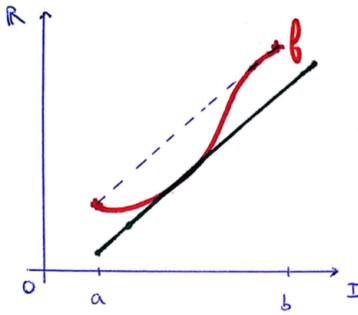


FIGURE 1.7 – Illustration du théorème des accroissements finis :  $g$  admet une tangente **en un point intérieur** dont la pente est égale à sa vitesse moyenne.

## 7 Classes de fonctions dérivables

On rappelle dans cette section les différents espaces fonctionnels usuels. On se donne  $I$  un intervalle de  $\mathbb{R}$  et  $f$  une fonction de  $I$  dans  $\mathbb{R}$ . À partir de  $k = 3$ , on note  $f^{(k)}(x)$  la dérivée  $k$ -ième de  $f$  en  $x \in I$ .

**Définition 7.1** (Fonctions dérivables). Soit  $k \in \mathbb{N}^*$ . On dit que  $f$  est de classe  $\mathcal{D}^k$  si  $f$  est dérivable sur  $I$ ,  $f'$  est dérivable sur  $I$ , ...,  $f^{(k-1)}$  est dérivable sur  $I$ .

On rappelle qu'une fonction de classe  $\mathcal{D}^1$  est nécessairement continue.

**Définition 7.2** (Fonctions continument dérivables). Soit  $k \in \mathbb{N}^*$ . On dit que  $f$  est de classe  $C^k$  si elle est de classe  $\mathcal{D}^k$  et si  $f, f', \dots, f^{(k)}$  sont des fonctions continues. On dit qu'elle est de classe  $C^\infty$  si elle est de classe  $C^k$  pour tout  $k \in \mathbb{N}^*$ .

*Remarque 7.3.* Les fonctions de classe  $\mathcal{D}^k$  sont nécessairement de classe  $C^{k-1}$ . On aurait pu remplacer la définition de  $C^k$  par " $\mathcal{D}^k$ " + " $f^{(k)}$  continue".

## 8 Développements limités

### Vocabulaire.

- Si on se donne un ensemble  $I \subset \mathbb{R}$ , et  $x$  un point de  $I$ , on appelle **voisinage** de  $x$  dans  $I$  tout sous-ensemble  $J$  de  $I$  pour lequel il existe  $\varepsilon > 0$  tel que

$$]x - \varepsilon, x + \varepsilon[ \cap I \subset J.$$

En français,  $J$  est un voisinage de  $x$  dans  $I$  si c'est un sous-ensemble de  $I$  qui contient tout les éléments de  $I$  suffisamment proches de  $x$ .

- On dit qu'une fonction  $f : I \subset \mathbb{R} \mapsto \mathbb{R}$  satisfait une propriété au voisinage de  $x \in I$  s'il existe un voisinage de  $x$  dans  $I$  sur laquelle  $f$  satisfait cette propriété. Par exemple, on dit que  $f$  est croissante au voisinage de  $x$  s'il existe un voisinage  $J$  de  $x$  dans  $I$  tel que la restriction de  $f$  à  $J$  soit croissante.
- Soit  $f$  une fonction définie sur un sous-ensemble  $I$  de  $\mathbb{R}$  et  $x$  un point de  $I$ . On dit qu'une propriété de  $f$  en  $x$  est **locale** si elle ne dépend que des valeurs de  $f$  sur un voisinage de  $x$  dans  $I$ . Par exemple, " $f$  admet un extremum local en  $x$ " est une propriété locale de  $f$  en  $x$ . Si on change les valeurs de  $f$  à l'extérieur d'un voisinage de  $x$  dans  $I$ , cette propriété sera toujours vérifiée.

**Définition 8.1** (Notation "o" et "O"). Soit  $I$  un intervalle de  $\mathbb{R}$ ,  $x$  un point de  $I$  et  $f$  et  $g$  deux fonctions de  $I$  dans  $\mathbb{R}$ . On note

$$f(y) = \underset{y \in I}{o}_{y \rightarrow x}(g(y))$$

si pour tout  $\varepsilon > 0$ , il existe  $\delta > 0$  tel que pour tout  $y \in I$ ,

$$|y - x| \leq \delta \quad \Rightarrow \quad |f(y)| \leq \varepsilon |g(y)|.$$

On note

$$f(y) = \underset{y \in I}{O}_{y \rightarrow x}(g(y))$$

si il existe  $\delta > 0$  et  $M > 0$  tels que pour tout  $y \in I$ ,

$$|y - x| \leq \delta \quad \Rightarrow \quad |f(y)| \leq M |g(y)|.$$

*Remarque 8.2.*

- Si  $x$  est un point intérieur de  $I$ , on oublie le  $y \in I$  dans la notation.
- Ce sont des propriétés locales : elles ne dépendent que des valeurs de  $f$  et  $g$  au voisinage de  $x$ .
- On peut donner des définitions équivalentes de ces notions : avec les mêmes notations que dans la définition,

$$f(y) = \underset{y \in I}{o}_{y \rightarrow x}(g(y))$$

si il existe une fonction  $\varepsilon : I \rightarrow \mathbb{R}$  telle que

$$\forall y \in I, \quad f(y) = \varepsilon(y)g(y) \quad \text{et} \quad \lim_{\substack{y \rightarrow x \\ y \in I}} \varepsilon(y) = 0,$$

et

$$f(y) = \underset{y \in I}{O}_{y \rightarrow x}(g(y))$$

si il existe une fonction  $M : I \rightarrow \mathbb{R}$  bornée au voisinage de  $x$  telle que

$$\forall y \in I, \quad f(y) = M(y)g(y).$$

- Si  $g$  ne s'annule pas (sauf éventuellement en  $x$ ) au voisinage de  $x$ , c'est encore équivalent à

$$f(y) = \underset{y \in I}{o}_{y \rightarrow x}(g(y))$$

si

$$\lim_{\substack{y \rightarrow x \\ y \in I \setminus \{x\}}} \frac{f(y)}{g(y)} = 0,$$

et

$$f(y) = \underset{y \in I}{O}_{y \rightarrow x}(g(y))$$

si

$$y \in I \setminus \{x\} \mapsto \frac{f(y)}{g(y)}$$

est bornée au voisinage de  $x$ .

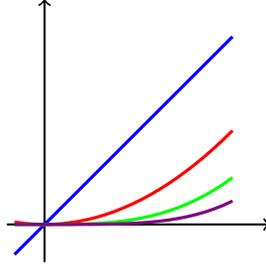


FIGURE 1.8 – Les fonction  $x$  (bleu),  $x^2$  (rouge),  $x^3$  (vert) et  $x^4$  (violet) au voisinage de 0.

Maintenant, l'idée des développements limités, c'est qu'au voisinage d'un point  $x \in \mathbb{R}$ , les fonctions

$$p_n : y \mapsto (y - x)^n, \quad n \in \mathbb{N},$$

sont de plus en plus petites quand  $n$  augmente. On s'attend donc à ce qu'on puisse écrire localement une fonction  $f$  autour de  $x$  comme une combinaison linéaire de telles fonctions quitte à faire une petite erreur.

**Définition 8.3** (Développement limité). On dit qu'une fonction  $f$  définie sur un intervalle  $I$  de  $\mathbb{R}$  admet un développement limité à l'ordre  $n \in \mathbb{N}$  en  $x \in I$  s'il existe  $a_0, \dots, a_n$  des nombres réels tels que

$$f(y) = a_0 + a_1(y - x) + \dots + a_n(y - x)^n + \underset{y \in I}{o_{y \rightarrow x}}((y - x)^n),$$

ou de façon équivalente

$$f(x + h) = a_0 + a_1h + \dots + a_nh^n + \underset{x+h \in I}{o_{h \rightarrow 0}}(h^n).$$

Une première propriété importante des développements limités est leur unicité.

**Proposition 8.4.** Soit  $I$  un intervalle de  $\mathbb{R}$ ,  $x \in I$  et  $f$  une fonction de  $I$  dans  $\mathbb{R}$ . On suppose que  $f$  admet deux développements limités d'ordre  $n \in \mathbb{N}$  en  $x$  :

$$f(x + h) = a_0 + a_1x + \dots + a_nx^n + \underset{x+h \in I}{o_{h \rightarrow 0}}(x^n) = b_0 + b_1x + \dots + b_nx^n + \underset{x+h \in I}{o_{h \rightarrow 0}}(x^n).$$

Alors  $a_0 = b_0, a_1 = b_1, \dots, a_n = b_n$ .

Le théorème fondamental que l'on généralisera un jour en plus grande dimension est la formule de Taylor-Young qui relie les coefficients du développement limité d'une fonction  $f$  en  $x$  aux dérivés de  $f$  en  $x$ .

**Théorème 8.5** (Formule de Taylor-Young). Soit  $I$  un intervalle de  $\mathbb{R}$  et  $f$  une fonction de  $I$  dans  $\mathbb{R}$  de classe  $\mathcal{D}^n$ ,  $n \in \mathbb{N}^*$ . Alors  $f$  admet le développement limité à l'ordre  $n$  suivant en tout point  $x$  de  $I$  :

$$f(y) = f(x) + f'(x)(y - x) + \dots + \frac{f^{(n)}(x)}{n!}(y - x)^n + \underset{y \in I}{o_{y \rightarrow x}}((y - x)^n),$$

ou de façon équivalente

$$f(x + h) = f(x) + f'(x)h + \dots + \frac{f^{(n)}(x)}{n!}h^n + \underset{x+h \in I}{o_{h \rightarrow 0}}(h^n),$$

*Démonstration.* On montre le résultat par récurrence. On se donne  $I$  un intervalle de  $\mathbb{R}$  et  $x$  un point de  $I$ . L'hypothèse de récurrence à l'étape  $n$  est :

Pour toute fonction  $f : I \rightarrow \mathbb{R}$  de classe  $\mathcal{D}^n$ ,

$$f(x+h) = f(x) + f'(x)h + \dots + \frac{f^{(n)}(x)}{n!}h^n + \underset{\substack{h \rightarrow 0 \\ x+h \in I}}{o}(h^n).$$

Initialisation. Si  $n = 1$ ,  $f$  est supposée de classe  $\mathcal{D}^1$ . En particulier,  $f$  est dérivable en  $x$ , donc

$$\lim_{\substack{h \rightarrow 0 \\ x+h \in I}} \frac{f(x+h) - f(x)}{h} = f'(x).$$

Ceci est équivalent à

$$\lim_{\substack{h \rightarrow 0 \\ x+h \in I}} \frac{f(x+h) - f(x) - hf'(x)}{h} = 0.$$

Comme la fonction  $h \mapsto h$  ne s'annule qu'en  $h = 0$ , par la définition du "o" par quotient, c'est la formule de Taylor-Young pour  $n = 1$ .

Hérédité. Supposons l'hypothèse de récurrence au rang  $n \in \mathbb{N}^*$ , et soit  $f$  une fonction de  $I$  dans  $\mathbb{R}$  de classe  $\mathcal{D}^{n+1}$ . La fonction  $f'$  est de classe  $\mathcal{D}^n$ . On peut donc lui appliquer l'hypothèse de récurrence :

$$f'(x+h) = f'(x) + f''(x)h + \dots + \frac{f^{(n+1)}(x)}{n!}h^n + \underset{\substack{h \rightarrow 0 \\ x+h \in I}}{o}(h^n).$$

Soit  $\varepsilon > 0$ . Par la première définition des "o" appliqué à  $(n+1)\varepsilon$ , il existe  $\delta > 0$  tel que pour tout  $h$  tel que  $x+h \in I$ ,

$$|h| \leq \delta \quad \Rightarrow \quad \left| f'(x+h) - f'(x) - f''(x)h - \dots - \frac{f^{(n+1)}(x)}{n!}h^n \right| \leq (n+1)\varepsilon|h|^n.$$

De façon équivalente, pour de tels  $h$ ,

$$\begin{aligned} f'(x+h) &\geq f'(x) + f''(x)h + \dots + \frac{f^{(n+1)}(x)}{n!}h^n - (n+1)\varepsilon|h|^n \\ \text{et } f'(x+h) &\leq f'(x) + f''(x)h + \dots + \frac{f^{(n+1)}(x)}{n!}h^n + (n+1)\varepsilon|h|^n. \end{aligned}$$

Mais  $f$  est au moins  $\mathcal{D}^2$  donc  $f'$  est au moins  $\mathcal{D}^1$  et donc continue. On peut donc l'intégrer. Soit  $l$  tel que  $|l| \leq \delta$  et  $x+l \in I$ . Par soucis de clarté, on va supposer  $l \geq 0$  mais la preuve marcherait de la même façon pour  $l \neq 0$ . Pour tout  $h$  entre 0 et  $l$ ,  $|h| \leq \delta$  et  $x+h \in I$  (car  $I$  est un intervalle). Donc en intégrant la première ligne entre 0 et  $l$ , on obtient

$$\begin{aligned} &\int_0^l f'(x+h) dh \\ &\geq f'(x) \int_0^l dh + f''(x) \int_0^l h dh + \dots + \frac{f^{(n+1)}(x)}{n!} \int_0^l h^n dh - (n+1)\varepsilon \int_0^l h^n dh. \end{aligned}$$

Cette inégalité se réécrit

$$\begin{aligned} &f(x+l) - f(x) \\ &\geq f'(x)l + f''(x)\frac{l^2}{2} + \dots + \frac{f^{(n+1)}(x)}{n!} \frac{l^{n+1}}{n+1} - (n+1)\varepsilon \frac{l^{n+1}}{n+1} \\ &= f'(x)l + \frac{f''(x)}{2}l^2 + \dots + \frac{f^{n+1}(x)}{(n+1)!}l^{n+1} - \varepsilon l^{n+1}. \end{aligned}$$

De la même façon, en intégrant la seconde ligne, on obtient

$$f(x+l) - f(x) \leq f'(x)l + \frac{f''(x)}{2}l^2 + \dots + \frac{f^{n+1}(x)}{(n+1)!}l^{n+1} + \varepsilon l^{n+1}$$

C'est exactement la formule de Taylor-Young au rang  $(n+1)$ . □

*Remarque 8.6.* En fait, on a supposé des hypothèses un peu trop fortes dans l'énoncé du théorème précédent. Pour que la formule de Taylor-Young d'ordre  $n$  de  $f$  soit valide, il suffit que  $f$  soit de classe  $\mathcal{D}^{n-1}$  et qu'elle soit  $n$  fois dérivable en  $x$ . Dans ce cadre, on ne peut plus intégrer  $f'$  dans la preuve lorsque  $n = 2$ . La preuve repose alors sur le théorème des accroissements finis.

## 9 Condition suffisante du second ordre

Il est facile de se rendre compte que l'on ne peut pas regarder la fonction au premier ordre (c'est à dire seulement sa dérivée première) et conclure qu'un point est un point de minimum local ou de maximum local. En effet, prenons l'exemple de la fonction cosinus. Quelle information nous donne le fait que sa dérivée s'annule en 0? Cela nous dit simplement que autour de 0, on a l'approximation :

$$\cos x = 1 + o(x).$$

C'est à dire ça :

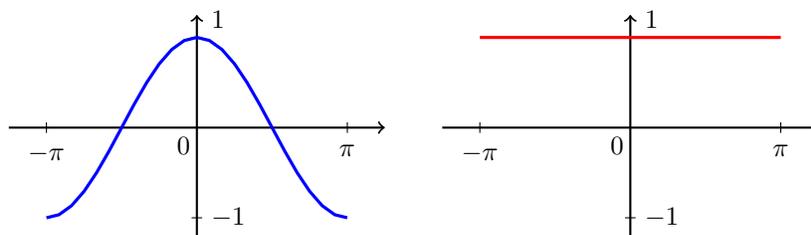


FIGURE 1.9 – La fonction cosinus (à gauche) et son approximation au premier ordre en 0 (à droite). Difficile de tirer de cette approximation le type de point critique qu'est 0...

En revanche, le fait que la dérivée seconde du cosinus en 0 soit égale à  $-1$  nous donne bien plus d'informations. On a en effet cette fois l'approximation :

$$\cos x = 1 - \frac{1}{2}x^2 + o(x^3).$$

c'est à dire ça :

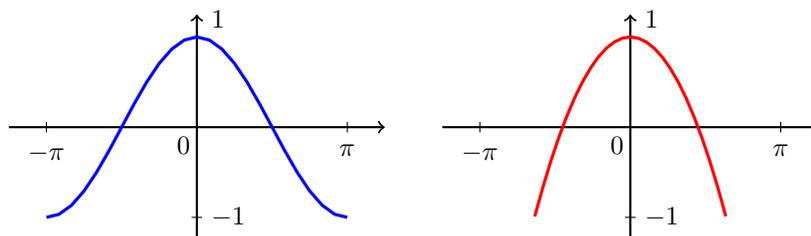


FIGURE 1.10 – La fonction cosinus (à gauche) et son approximation au second ordre en 0 (à droite). 0 semble bien être un point de maximum local.

On voit donc bien avec ce niveau de précision que 0 est un point de maximum local de la fonction cosinus.

**Théorème 9.1.** Soit  $f$  une fonction de classe  $\mathcal{D}^2$  d'un intervalle  $I$  de  $\mathbb{R}$  à valeurs dans  $\mathbb{R}$  et  $x$  un point de  $I$  tel que :

$$\begin{aligned} f'(x) &= 0, \\ f''(x) &> 0. \end{aligned}$$

Alors  $x$  est un point de minimum local strict de  $f$ . On a évidemment une propriété analogue pour les maxima locaux.

*Remarque 9.2.* Si  $f'(x) = 0$  et  $f''(x) = 0$ , on ne peut pas conclure, en témoignent les exemples suivants.

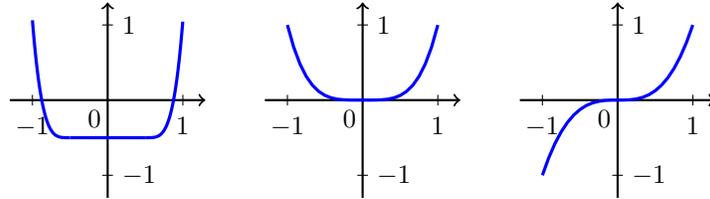


FIGURE 1.11 – Trois exemples de fonctions dont la dérivée première et la dérivée seconde s'annulent en 0. Dans le premier cas ( $f$  est constante au voisinage de 0), 0 est un point de minimum local non strict, dans le second ( $f(x) = x^4$ ), 0 est un point de minimum strict, et dans le troisième ( $f(x) = x^3$ ), 0 n'est pas un point d'extremum local.

Encore une fois, pas besoin de supposer  $\mathcal{D}^2$  : " $\mathcal{D}^1$ " + " $f'$  dérivable en  $x$ " suffit.

*Démonstration.* On écrit le développement limité d'ordre 2 de  $f$  en  $x$  :

$$f(y) = f(x) + \frac{f''(x)}{2}(y-x)^2 + o_{y \rightarrow x}((y-x)^2).$$

On choisit  $\delta > 0$  tel que pour tout  $y \in I$ , si  $|x-y| \leq \delta$ , alors

$$\left| f(y) - f(x) - \frac{f''(x)}{2}(y-x)^2 \right| \leq \frac{f''(x)}{4}(y-x)^2.$$

En particulier, si  $|y-x| \leq \delta$  et  $y \neq x$ ,

$$\begin{aligned} f(y) &\geq f(x) + \frac{f''(x)}{2}(y-x)^2 - \frac{f''(x)}{4}(y-x)^2 \\ &= f(x) + \frac{f''(x)}{4}(y-x)^2 < f(x). \end{aligned}$$

Donc  $x$  est bien un point de minimum local strict de  $f$ . □

# Chapitre 2

## La structure euclidienne de $\mathbb{R}^d$

### 1 Définition

**Définition 1.1.** Le produit scalaire usuel sur  $\mathbb{R}^d$  est l'application prenant comme argument **deux vecteurs**  $x = (x_1, \dots, x_d)$  et  $y = (y_1, \dots, y_d)$  de  $\mathbb{R}^d$  et rendant comme valeur le **réel** :

$$\langle x, y \rangle := \sum_{i=1}^d x_i y_i.$$

On dit que cette application est un produit scalaire car on peut vérifier qu'elle satisfait les propriétés suivantes.

- C'est une application de  $\mathbb{R}^d \times \mathbb{R}^d$  dans  $\mathbb{R}$  qui a pour particularité d'être linéaire en chacune de ses coordonnées : pour tous  $x, y$  et  $z$  des vecteurs de  $\mathbb{R}^d$  et tous  $\lambda$  et  $\mu$  réels :

$$\begin{aligned}\langle \lambda x + \mu y, z \rangle &= \lambda \langle x, z \rangle + \mu \langle y, z \rangle, \\ \langle x, \lambda y + \mu z \rangle &= \lambda \langle x, y \rangle + \mu \langle x, z \rangle.\end{aligned}$$

On dit que cette application est une **forme bilinéaire** sur  $\mathbb{R}^d$ .

- Elle est **symétrique**, c'est à dire que pour tous vecteurs de  $\mathbb{R}^d$   $x$  et  $y$  :

$$\langle x, y \rangle = \langle y, x \rangle.$$

- Elle est **positive**, c'est à dire que pour tout vecteur  $x$  de  $\mathbb{R}^d$  :

$$\langle x, x \rangle \geq 0.$$

- Et enfin elle est **définie**, c'est à dire que l'égalité :

$$\langle x, x \rangle = 0$$

n'est vraie que pour  $x = 0$ . En d'autres termes, pour tout  $x \in \mathbb{R}^d$  non nul :

$$\langle x, x \rangle > 0.$$

Le produit scalaire usuel défini plus haut n'est pas la seule application à satisfaire ces propriétés. Ce n'est pas le seul produit scalaire. Ce sera le seul que l'on utilisera dans ce cours, mais le lecteur motivé pourra remarquer que tous les résultats sont vrais pour tous les produits scalaires sur  $\mathbb{R}^d$ , c'est à dire que dans les preuves, on n'utilisera que les quatre propriétés listées.

Une fois qu'on a un produit scalaire, on peut définir une **norme euclidienne** de la façon suivante.

**Définition 1.2.** On définit la norme euclidienne d'un vecteur  $x$  de  $\mathbb{R}^d$  par :

$$\|x\| := \sqrt{\langle x, x \rangle}.$$

Pour le produit scalaire standard, on retrouve bien la formule usuelle :

$$\|x\| = \sqrt{\sum_{i=1}^d x_i^2}.$$

La norme euclidienne est une application de  $\mathbb{R}^d$  dans  $\mathbb{R}_+$  satisfaisant les propriétés suivantes.

- Elle est **séparante**, c'est à dire que pour tout  $x \in \mathbb{R}^d$  :

$$\|x\| = 0 \quad \Rightarrow \quad x = 0.$$

- Elle est **homogène**, c'est à dire que pour tout vecteur  $x \in \mathbb{R}^d$  et tout réel  $\lambda$  :

$$\|\lambda x\| = |\lambda| \|x\|.$$

De plus, la bilinéarité nous assure les identités remarquables suivantes :

**Proposition 1.3** (Identités remarquables). *Pour tous  $x$  et  $y$  dans  $\mathbb{R}^d$  :*

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2, \\ \|x - y\|^2 &= \|x\|^2 - 2\langle x, y \rangle + \|y\|^2, \\ \langle x + y, x - y \rangle &= \|x\|^2 - \|y\|^2, \\ \langle x, y \rangle &= \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2). \end{aligned}$$

Un théorème bien connu est l'inégalité de Cauchy-Schwarz : le produit scalaire entre deux vecteurs est inférieur au produit des normes de ces vecteurs.

**Théorème 1.4** (Inégalité de Cauchy-Schwarz). *Soient  $x$  et  $y$  deux vecteurs de  $\mathbb{R}^d$ .*

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

*Démonstration.* D'abord, le théorème est vrai quand  $x$  ou  $y$  est nul, puisque alors on a 0 de chaque côté. On peut donc supposer  $x$  et  $y$  non nuls.

Ensuite, si  $\|x\| = \|y\| = 1$ , par positivité :

$$\begin{aligned} 0 \leq \langle x - y, x - y \rangle &= \|x\|^2 - 2\langle x, y \rangle + \|y\|^2 \\ &= 2 - 2\langle x, y \rangle. \end{aligned}$$

Donc :

$$\langle x, y \rangle \leq 1.$$

Enfin, si  $x$  et  $y$  sont deux vecteurs quelconques non-nuls de  $\mathbb{R}^d$ , les deux vecteurs :

$$\frac{x}{\|x\|} \quad \text{et} \quad \frac{y}{\|y\|}$$

sont de norme euclidienne 1. Donc par ce qu'on vient de démontrer :

$$\left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \right\rangle \leq 1.$$

Mais par bilinéarité :

$$\left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \right\rangle = \frac{1}{\|x\|} \left\langle x, \frac{y}{\|y\|} \right\rangle = \frac{1}{\|x\|\|y\|} \langle x, y \rangle \leq 1.$$

Donc on a :

$$\langle x, y \rangle \leq \|x\|\|y\|.$$

Comme  $\|y\| = \|-y\|$ , la même inégalité appliquée à  $x$  et  $-y$  donne :

$$\langle x, -y \rangle = -\langle x, y \rangle \leq \|x\|\|y\|.$$

En combinant ces deux inégalités, on obtient bien :

$$|\langle x, y \rangle| \leq \|x\|\|y\|.$$

□

En conséquence, on démontre l'inégalité triangulaire.

**Corollaire 1.1** (Inégalité triangulaire). *Pour tous  $x$  et  $y$  dans  $\mathbb{R}^d$  :*

$$\|x + y\| \leq \|x\| + \|y\|.$$

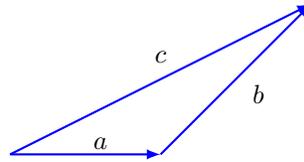


FIGURE 2.1 – Dans le dessin ci-dessus, l'inégalité triangulaire énonce que la norme euclidienne du vecteur  $c$  est plus petite que la norme euclidienne du vecteur  $a$  plus la norme euclidienne du vecteur  $b$ .

*Démonstration.* Soient  $x$  et  $y$  dans  $\mathbb{R}^d$  :

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2. \end{aligned}$$

On conclut alors grâce à la positivité de la norme euclidienne. □

**Définition 1.5.** On dit qu'une application  $N$  de  $\mathbb{R}^d$  dans  $\mathbb{R}_+$  est une **norme d'espace vectoriel** si elle est :

- homogène :  $\forall x \in \mathbb{R}^d, \forall \lambda \in \mathbb{R}, N(\lambda x) = |\lambda|N(x)$ ,
- séparante :  $\forall x \in \mathbb{R}^d, N(x) = 0 \Rightarrow x = 0$ ,
- elle satisfait l'inégalité triangulaire :

$$\forall (x, y) \in (\mathbb{R}^d)^2, N(x + y) \leq N(x) + N(y).$$

On a donc vu qu'une norme euclidienne est une norme d'espace vectoriel. Mais il y en a plein d'autres qui ne sont pas issues d'un produit scalaire...

## 2 Orthogonalité

Attention : dans cette section, on commence à considérer des familles de vecteurs. Il y aura donc une ambiguïté de notation entre le  $i$ -ième vecteur d'une famille et la  $i$ -ième coordonnée d'un vecteur. Dans la mesure du possible,  $x = (x_1, \dots, x_d)$  sera un vecteur dont les coordonnées seront les  $x_i$ , et  $u_1, u_2, \dots$  seront des vecteurs dont on n'écrira pas les coordonnées.

**Définition 2.1.** On dit que deux vecteurs  $x$  et  $y$  de  $\mathbb{R}^d$  sont orthogonaux si :

$$\langle x, y \rangle = 0.$$

On peut alors démontrer le théorème de Pythagore.

**Théorème 2.2** (Pythagore). Si  $x$  et  $y$  sont deux vecteurs orthogonaux de  $\mathbb{R}^d$  :

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Plus généralement, si  $(u_1, \dots, u_p)$  sont des vecteurs deux à deux orthogonaux,

$$\|u_1 + \dots + u_p\|^2 = \|u_1\|^2 + \dots + \|u_p\|^2.$$

*Démonstration.* On ne montre que la deuxième formule puisque c'est une généralisation de la première. Soient donc  $u_1, \dots, u_p$  des vecteurs deux à deux orthogonaux. On a

$$\begin{aligned} \|u_1 + \dots + u_p\|^2 &= \langle u_1 + \dots + u_p, u_1 + \dots + u_p \rangle \\ &= \sum_{i=1}^p \langle u_i, u_1 + \dots + u_p \rangle \\ &= \sum_{i=1}^p \sum_{j=1}^p \langle u_i, u_j \rangle. \end{aligned}$$

Mais tous les termes de cette somme correspondant à des indices  $i \neq j$  sont nuls par hypothèse. Il reste donc

$$\|u_1 + \dots + u_p\|^2 = \sum_{i=1}^p \langle u_i, u_i \rangle = \|u_1\|^2 + \dots + \|u_p\|^2.$$

□

**Définition 2.3.** Soit  $p \in \mathbb{N}$  et  $(u_1, \dots, u_p)$  une famille de vecteurs de  $\mathbb{R}^d$ . On dit que c'est une famille **orthogonale** si pour tous  $i$  et  $j$  distincts dans  $\{1, \dots, p\}$ ,  $u_i$  et  $u_j$  sont orthogonaux. On dit que c'est une famille **orthonormale** si en plus, chacun des vecteurs qui la composent est de norme euclidienne égale à 1.

En particulier, on peut parler de **bases orthogonales** et de **bases orthonormales** si la famille considérée est une base de  $\mathbb{R}^d$ .

*Exemple 2.4.* La base canonique de  $\mathbb{R}^d$ , c'est à dire celle composée des vecteurs :

$$\forall i \in \{1, \dots, d\}, \quad e_i = (0, \dots, 0, 1, 0, \dots, 0) \quad \text{où le 1, est à la } i\text{-ième position}$$

est orthonormale.

**Proposition 2.5.** Toute famille orthogonale de vecteur non-nuls est libre. En particulier, toute famille orthonormale est libre.

*Démonstration.* Soit  $p \in \mathbb{N}$  et  $(u_1, \dots, u_p)$  une famille orthogonale de vecteurs non-nuls. Soient  $(\lambda_1, \dots, \lambda_p)$  une famille de réels tels que :

$$\lambda_1 u_1 + \dots + \lambda_p u_p = 0.$$

Soit  $i \in \{1, \dots, p\}$ . Faisons le produit scalaire de cette égalité avec  $u_i$ .

$$\begin{aligned} 0 &= \langle u_i, \lambda_1 u_1 + \dots + \lambda_p u_p \rangle \\ &= \sum_{j=1}^p \lambda_j \langle u_i, u_j \rangle \quad \text{par bilinéarité,} \\ &= \lambda_i \langle u_i, u_i \rangle \quad \text{par orthogonalité,} \\ &= \lambda_i \|u_i\|^2. \end{aligned}$$

Or  $u_i$  est non nul, donc  $\lambda_i = 0$ . Comme on peut le faire pour chaque  $i$ , tous les  $\lambda$  sont nuls, et la famille est libre.  $\square$

On ne va pas le démontrer, mais on a les propriétés d'existence suivantes des bases orthogonales et orthonormales.

**Théorème 2.6.** • *Toute famille orthogonale (resp. orthonormale) de vecteurs non nuls de  $\mathbb{R}^d$  peut être complétée en une base orthogonale (resp. orthonormale). C'est à dire que si  $p \in \mathbb{N}$  et si  $(u_1, \dots, u_p)$  est une famille orthogonale (resp. orthonormale) de vecteurs non nuls, il existe  $(u_{p+1}, \dots, u_d)$  des vecteurs de  $\mathbb{R}^d$  tels que  $(u_1, \dots, u_d)$  soit une base orthogonale (resp. orthonormale).*

- *Ce résultat est encore vrai dans les sous-espaces vectoriels : toute famille orthogonale (resp. orthonormale) d'un sous-espace vectoriel  $E$  de  $\mathbb{R}^d$  peut être complétée en une base orthogonale (resp. orthonormale) de  $E$ .*
- *En particulier, tout sous-espace vectoriel de  $\mathbb{R}^d$  admet une base orthonormale.*

Dans les bases orthonormales, beaucoup de calculs sont simplifiés parce qu'on a les propositions suivantes.

**Proposition 2.7.** *Si  $(e_1, \dots, e_d)$  est une base orthonormale de  $\mathbb{R}^d$ , alors pour tout  $x \in \mathbb{R}^d$  :*

$$x = \sum_{i=1}^d \langle e_i, x \rangle e_i,$$

*c'est à dire que les coordonnées de  $x$  dans cette base sont simplement données par le produit scalaire de  $x$  avec les vecteurs de base.*

*En particulier, le vecteur colonne représentant les coordonnées de  $x$  dans cette base est :*

$$X = \begin{pmatrix} \langle e_1, x \rangle \\ \vdots \\ \langle e_d, x \rangle \end{pmatrix}$$

*Démonstration.* Soit  $x \in \mathbb{R}^d$  et  $(\lambda_1, \dots, \lambda_d)$  des réels tels que :

$$x = \sum_{i=1}^d \lambda_i e_i.$$

Soit  $j \in \{1, \dots, p\}$ . En faisant le produit scalaire de cette égalité avec le vecteur de base  $e_j$ , on obtient :

$$\langle e_j, x \rangle = \left\langle e_j, \sum_{i=1}^d \lambda_i e_i \right\rangle = \sum_{i=1}^d \lambda_i \langle e_j, e_i \rangle = \lambda_j.$$

$\square$

**Proposition 2.8.** Si  $(e_1, \dots, e_d)$  est une base orthonormale de  $\mathbb{R}^d$  et si  $x$  et  $y$  sont dans  $\mathbb{R}^d$ , alors :

$$\langle x, y \rangle = \sum_{i=1}^d \langle e_i, x \rangle \langle e_i, y \rangle.$$

En particulier, si  $X$  et  $Y$  sont les vecteurs colonnes des coordonnées de  $x$  et  $y$  dans cette base :

$$\langle x, y \rangle = X^t \times Y$$

au sens des produits de matrices, où  $X^t$  désigne la transposée de  $X$ .

*Démonstration.* Il suffit d'écrire :

$$\left\langle \sum_{i=1}^d \langle e_i, x \rangle e_i, y \right\rangle$$

et de développer en utilisant la propriété de bilinéarité. □

*Remarque 2.9.* Si on choisit comme base orthonormée la base canonique de  $\mathbb{R}^d$ , alors pour tout  $x \in \mathbb{R}^d$  et pour tout  $i \in \{1, \dots, d\}$ ,

$$\langle x, e_i \rangle = x_i.$$

On retrouve donc la formule usuel du produit scalaire canonique de  $\mathbb{R}^d$ . En fait, ceci montre que le produit scalaire canonique est le seul produit scalaire de  $\mathbb{R}^d$  pour lequel la base canonique est orthonormée.

**Définition 2.10.** Si  $E$  est un sous-ensemble de  $\mathbb{R}^d$ , on appelle l'orthogonal de  $E$  l'ensemble :

$$E^\perp := \{y \mid \forall x \in E, \langle x, y \rangle = 0\}.$$

**Proposition 2.11.** Pour tout sous-ensemble  $E$  de  $\mathbb{R}^d$ ,  $E^\perp$  est un sous-espace vectoriel de  $\mathbb{R}^d$ .

*Démonstration.* Soient  $x \in E^\perp$ ,  $y \in E^\perp$  et  $\lambda \in \mathbb{R}$ . Montrons que  $x + \lambda y \in E^\perp$ . Si  $z \in E$ , alors :

$$\langle z, x + \lambda y \rangle = \langle z, x \rangle + \lambda \langle z, y \rangle = 0.$$

Comme c'est vrai pour tout  $z \in E$ , on a le résultat. □

**Proposition 2.12.** Soient  $E$  et  $F$  deux sous-ensembles de  $\mathbb{R}^d$  avec  $F \subset E$ . Alors

$$E^\perp \subset F^\perp.$$

*Démonstration.* Soit  $x \in E^\perp$ . Montrons que  $x \in F^\perp$ . Pour ça, il faut montrer que pour tout  $y \in F$ ,  $\langle x, y \rangle = 0$ . Soit donc  $y \in F$ . Par hypothèse,  $y \in E$ . Donc comme  $x \in E^\perp$ , on a bien  $\langle x, y \rangle = 0$ . Le résultat est donc démontré. □

**Proposition 2.13.** Pour tout sous-ensemble  $E$  de  $\mathbb{R}^d$  :

$$\text{Vect}(E)^\perp = E^\perp.$$

*Démonstration.* On a  $E \subset \text{Vect}(E)$ . Donc on déduit de la proposition précédente que  $\text{Vect}(E)^\perp \subset E^\perp$ .

Montrons maintenant que  $E^\perp \subset \text{Vect}(E)^\perp$ . Soit  $u \in E^\perp$  et  $v \in \text{Vect}(E)$ . Il existe  $p \in \mathbb{N}^*$ ,  $\lambda_1, \dots, \lambda_p$  des réels et  $v_1, \dots, v_p$  dans  $E$  tels que :

$$v = \lambda_1 v_1 + \dots + \lambda_p v_p.$$

Donc :

$$\langle v, u \rangle = \lambda_1 \langle v_1, u \rangle + \dots + \lambda_p \langle v_p, u \rangle = 0.$$

On a donc démontré le résultat. □

**Proposition 2.14.** *Si  $E$  est un sous-espace vectoriel de  $\mathbb{R}^d$ , alors  $E$  et  $E^\perp$  sont des supplémentaires. En particulier,  $\dim E^\perp = d - \dim E$ .*

*Démonstration.* Soit  $p := \dim(E)$  et  $e_1, \dots, e_p$  une base orthonormée de  $E$ . On la complète en  $(e_1, \dots, e_d)$  une base de  $\mathbb{R}^d$ . Pour tout  $k = p + 1, \dots, d$ ,  $e_k \in \{e_1, \dots, e_p\}^\perp = E^\perp$ . Ceci permet de conclure que :

$$E + E^\perp = \mathbb{R}^d.$$

Montrons que la somme est directe. Soit  $x \in E \cap E^\perp$ . Alors :

$$\underbrace{\langle x, x \rangle}_{\in E} = \underbrace{\langle x, x \rangle}_{\in E^\perp} = 0.$$

Donc  $x = 0$  et on a le résultat. □

*Exemple 2.15.* Par exemple, l'orthogonal d'un vecteur  $x$  de  $\mathbb{R}^d$  est un sous-espace vectoriel de  $\mathbb{R}^d$  de dimension  $d - 1$ . On dit que c'est un hyperplan.

**Proposition 2.16.** *Soit  $E$  un sous-espace vectoriel de  $\mathbb{R}^d$ . Alors*

$$(E^\perp)^\perp = E.$$

*Démonstration.* Montrons d'abord que  $E \subset (E^\perp)^\perp$ . Soit  $x \in E$ . Pour tout  $y \in E^\perp$ , le produit scalaire  $\langle x, y \rangle = 0$ . Le vecteur  $x$  est donc orthogonal à tous les vecteurs de  $E^\perp$ , il est donc dans  $(E^\perp)^\perp$ .

Ensuite,

$$\begin{aligned} \dim((E^\perp)^\perp) &= d - \dim(E^\perp) \\ &= d - (d - \dim E) \\ &= \dim E. \end{aligned}$$

Donc on a nécessairement  $E = (E^\perp)^\perp$ . □

### 3 Projections et symétries orthogonales

**Rappels sur les projections et symétries linéaires.** Choisissons deux espaces supplémentaires  $E$  et  $F$  de  $\mathbb{R}^d$ . Pour tout  $x \in \mathbb{R}^d$ , il existe donc un unique couple  $(y, z) \in E \times F$  tel que  $x = y + z$ . C'est la décomposition de  $x$  dans  $E \oplus F$ .

**Définition 3.1** (Projections linéaires). Soit  $x$  un vecteur de  $\mathbb{R}^d$  et  $(y, z) \in E \times F$  le couple intervenant dans sa décomposition de  $x$  dans  $E \oplus F$ . La projection de  $x$  sur  $E$  parallèlement à  $F$  est le vecteur

$$p(x) := y.$$

Les projections vérifient les propriétés suivantes.

**Proposition 3.2.** *Soient  $E$  et  $F$  deux supplémentaires de  $\mathbb{R}^d$  et  $p$  la projection sur  $E$  parallèlement à  $F$ .*

- *L'application  $p$  est linéaire.*
- *Si  $x \in E$ , alors  $p(x) = x$ .*
- *Elle vérifie  $p \circ p = p$ .*
- *Son image est  $E$  et son noyau est  $F$ .*

*Démonstration.  $p$  est linéaire.* En effet, si  $u_1$  et  $u_2$  sont deux vecteurs de  $\mathbb{R}^d$ , si  $(v_1, w_1)$  et  $(v_2, w_2)$  sont les couples intervenants dans leurs décompositions dans  $E \oplus F$  et si  $\lambda$  et  $\mu$  sont des réels, alors

$$\lambda u_1 + \mu u_2 = \underbrace{\lambda v_1 + \mu v_2}_{\in E} + \underbrace{\lambda w_1 + \mu w_2}_{\in F}.$$

Donc

$$p(\lambda u_1 + \mu u_2) = \lambda v_1 + \mu v_2 = \lambda p(u_1) + \mu p(u_2).$$

$p|_E = \text{id}|_E$ . Si  $x \in E$ , sa décomposition dans  $E \oplus F$  est  $x = x + 0$ . Donc  $p(x) = x$ .

$p \circ p = p$ . En effet, si  $x \in \mathbb{R}^d$ , par définition,  $p(x) \in E$ . Donc par le point précédent appliqué à  $p(x)$ ,  $p(p(x)) = p(x)$ .

**Image.** Si  $x \in \text{im}(p)$ , alors si  $u$  est un de ses antécédants,  $x$  est l'élément de  $E$  intervenant dans la décomposition de  $u$  dans  $E \oplus F$ . Donc  $x \in E$ . Réciproquement, si  $x \in E$ , alors  $p(x) = x$ , et donc  $x \in \text{im}(p)$ .

**Noyau.** Soit  $x \in \mathbb{R}^d$ . Si  $x \in F$ , alors sa décomposition dans  $E \oplus F$  est  $x = 0 + x$ , et donc  $p(x) = 0$ . Si  $p(x) = 0$ , alors si  $(y, z) \in E \times F$  est le couple intervenant dans la décomposition de  $x$  dans  $E \oplus F$ , alors cela signifie que  $y = 0$  et donc que  $x = z \in F$ . Donc  $F = \ker(p)$ .  $\square$

En fait, on a une proposition qui donne la réciproque de cet énoncé.

**Proposition 3.3.** *Soit  $p$  une application linéaire de  $\mathbb{R}^d$  dans  $\mathbb{R}^d$  vérifiant l'égalité  $p \circ p = p$ . Alors son image et son noyau sont supplémentaires, et  $p$  est la projection sur  $\text{im}(p)$  parallèlement à  $\ker(p)$ .*

*Démonstration.* Soit  $p$  une application linéaire de  $\mathbb{R}^d$  dans  $\mathbb{R}^d$  avec  $p \circ p = p$ . On note  $E$  son image et  $F$  son noyau. Alors pour tout  $x \in E$ ,

$$x = p(x) + x - p(x).$$

Or  $p(x) \in E$ , et  $x - p(x) \in F$  car

$$p(x - p(x)) = p(x) - p(p(x)) = p(x) - p(x) = 0.$$

Donc  $E + F = \mathbb{R}^d$ . Soit  $x \in E \cap F$ . Alors il existe  $u \in \mathbb{R}^d$  tel que  $x = p(u)$ . et comme  $x \in F$ ,

$$0 = p(x) = p(p(u)) = p(u) = x.$$

Donc  $x = 0$  et  $E$  et  $F$  sont en somme directes, et la décomposition de  $x \in \mathbb{R}^d$  dans  $E \oplus F$  est donné par

$$x = p(x) + x - p(x).$$

Donc  $p$  est bien la projection sur  $E$  parallèlement à  $F$ .  $\square$

On peut faire le même travail pour les symétries. On reprend  $E$  et  $F$  deux supplémentaires de  $\mathbb{R}^d$ . On laisse les preuves en exercice.

**Définition 3.4** (Symétries linéaires). Soit  $x$  un vecteur de  $\mathbb{R}^d$  et  $(y, z) \in E \times F$  le couple intervenant dans sa décomposition de  $x$  dans  $E \oplus F$ . La symétrie de  $x$  par rapport à  $E$  parallèlement à  $F$  est le vecteur

$$s(x) := y - z.$$

Les symétries vérifient les propriétés suivantes

**Proposition 3.5.** *Soient  $E$  et  $F$  deux supplémentaires de  $\mathbb{R}^d$  et  $s$  la symétrie par rapport à  $E$  parallèlement à  $F$ .*

- *L'application  $s$  est linéaire.*
- *Elle vérifie  $s \circ s = \text{id}$ .*
- *On a  $E = \ker(s - \text{id})$  et  $F = \ker(s + \text{id})$ .*

Et réciproquement, on a la proposition suivante.

**Proposition 3.6.** *Soit  $s$  une application linéaire vérifiant  $s \circ s = \text{id}$ . Alors  $\ker(s - \text{id})$  et  $\ker(s + \text{id})$  sont supplémentaires, et  $s$  est la symétrie par rapport à  $\ker(s - \text{id})$  parallèlement à  $\ker(s + \text{id})$ .*

**Projections et symétries orthogonales.** Soit  $E$  un sous-espace vectoriel de  $\mathbb{R}^d$ . On a vu que  $\mathbb{R}^d = E \oplus E^\perp$ . On définit donc la projection orthogonale sur  $E$  comme suit.

**Définition 3.7.** La projection orthogonale sur  $E$  est la projection sur  $E$  parallèlement à  $E^\perp$ .

La projection orthogonale a les propriétés supplémentaires suivantes.

**Proposition 3.8.** • Pour tout  $x \in \mathbb{R}^d$ ,

$$\|x - p(x)\| = \min\{\|x - y\| \mid y \in E\}.$$

• Pour tous  $x$  et  $y$  dans  $\mathbb{R}^d$ ,

$$\langle x, p(y) \rangle = \langle p(x), y \rangle = \langle p(x), p(y) \rangle.$$

• Si  $p = \dim E$  et  $(e_1, \dots, e_p)$  est une base orthonormée de  $E$ , alors

$$p(x) = \sum_{i=1}^p \langle x, e_i \rangle e_i.$$

*Démonstration.* Soient  $x \in \mathbb{R}^d$  et  $y \in E$ . On rappelle que la décomposition de  $x$  dans  $E \oplus E^\perp$  est donnée par

$$x = p(x) + x - p(x),$$

et donc  $x - p(x) \in E^\perp$ . Il est donc orthogonal à  $p(x) - y$  qui est dans  $E$ . Par le théorème de Pythagore,

$$\begin{aligned} \|x - y\|^2 &= \|x - p(x) + p(x) - y\|^2 \\ &= \|x - p(x)\|^2 + \|p(x) - y\|^2 \\ &\geq \|x - p(x)\|^2. \end{aligned}$$

Donc

$$\inf\{\|x - y\| \mid y \in E\} \geq \|x - p(x)\|.$$

Comme  $p(x) \in E$  l'inf est en fait un min et on a le premier point.

Pour le second,

$$\begin{aligned} \langle x, p(y) \rangle &= \langle x - p(x) + p(x), p(y) \rangle \\ &= \underbrace{\langle x - p(x), p(y) \rangle}_{\in E^\perp} + \underbrace{\langle p(x), p(y) \rangle}_{\in E} \\ &= \langle p(x), p(y) \rangle. \end{aligned}$$

On aurait pu faire le même calcul en partant de  $\langle p(x), y \rangle$ , on obtient donc le second point.

Enfin, on prend  $(e_1, \dots, e_p)$  une base orthonormée de  $E$ . Le vecteur

$$\sum_{i=1}^p \langle x, e_i \rangle e_i$$

est dans  $E$ . Il suffit donc de montrer que

$$x - \sum_{i=1}^p \langle x, e_i \rangle e_i \in E^\perp.$$

En effet, si tel est le cas, la décomposition de  $x$  dans  $E \oplus E^\perp$  sera donné par

$$x = \sum_{i=1}^p \langle x, e_i \rangle e_i + x - \sum_{i=1}^p \langle x, e_i \rangle e_i.$$

Mais  $E = \text{Vect}(e_1, \dots, e_p)$ , donc  $E^\perp = \{e_1, \dots, e_p\}^\perp$ . Il suffit donc de montrer que pour tout  $j = 1, \dots, p$ ,

$$\sum_{i=1}^p \langle x, e_i \rangle e_i \perp e_j.$$

Mais si on calcule le produit scalaire,

$$\left\langle e_j, x - \sum_{i=1}^p \langle x, e_i \rangle e_i \right\rangle = \langle e_j, x \rangle - \sum_{i=1}^p \langle x, e_i \rangle \langle e_j, e_i \rangle = \langle e_j, x \rangle - \langle x, e_j \rangle = 0.$$

On obtient donc le résultat. □

## 4 Formes linéaires et produit scalaire

Soit  $z \in \mathbb{R}^d$ . L'application

$$x \in \mathbb{R}^d \mapsto \langle z, x \rangle$$

est une forme linéaire. On va voir dans cette section que dans  $\mathbb{R}^d$  ou plus généralement dans un espace euclidien, une forme linéaire peut toujours s'écrire de cette façon.

**Proposition 4.1.** *Soit  $\lambda : \mathbb{R}^d \rightarrow \mathbb{R}$  une forme linéaire. Il existe un unique vecteur  $z = (z_1, \dots, z_d) \in \mathbb{R}^d$  telle que pour tout  $x \in \mathbb{R}^d$ ,*

$$\lambda(x) = \langle z, x \rangle.$$

De plus, pour tout  $i = 1, \dots, d$ ,

$$z_i = \lambda(e_i).$$

Plus généralement, si  $(f_1, \dots, f_d)$  est une base orthonormée de  $\mathbb{R}^d$ , le  $i$ -ième coefficient de  $z$  dans la base  $(f_1, \dots, f_d)$  est  $\lambda(f_i)$ .

*Démonstration.* On montre directement le résultat pour une base orthonormée  $(f_1, \dots, f_d)$  générale. Montrons d'abord que si  $z$  convient, alors

$$z = \sum_{i=1}^d \lambda(f_i) f_i.$$

Soit  $z$  un vecteur qui convient. Pour  $i = 1, \dots, d$ ,

$$\langle z, f_i \rangle = \lambda(f_i)$$

par hypothèse. Donc  $z$  est bien le vecteur annoncé.

Réciproquement, on montre que ce vecteur convient. On pose donc

$$z := \sum_{i=1}^d \lambda(f_i) f_i.$$

Soit  $x \in \mathbb{R}^d$ ,

$$\lambda(x) = \lambda \left( \sum_{i=1}^d \langle f_i, x \rangle f_i \right) = \sum_{i=1}^d \langle f_i, x \rangle \lambda(f_i) = \sum_{i=1}^d \langle f_i, x \rangle \langle z, f_i \rangle = \langle z, x \rangle.$$

□

# Chapitre 3

## La topologie en dimension $d$

Dans tout ce chapitre, on se place dans  $\mathbb{R}^d$  avec  $d \in \mathbb{N}^*$ . On a introduit au chapitre précédent la norme euclidienne (que l'on notera encore dans ce chapitre  $\|\cdot\|$ ). Elle permet de mesurer la distance entre deux points de  $\mathbb{R}^d$  : si  $a$  et  $b$  sont dans  $\mathbb{R}^d$ , la **distance** entre  $a$  et  $b$  est le nombre positif  $\|a - b\|$ . Et une fois que l'on sait mesurer la distance entre des points, on peut parler de limite, de continuité etc... C'est précisément l'objectif de ce chapitre.

Je fais juste ici la remarque qu'en fait, **toutes les normes d'un espace vectoriel de dimension finie sont équivalentes**. En conséquence, toutes les notions que l'on va introduire (suites convergentes, fonctions continues, ouverts, fermés) ne dépendent pas de la norme choisie. On ne travaille avec la norme euclidienne que par commodité.

### 1 Limite et continuité en dimension $d$

Les définitions de limites en dimensions  $d$  sont globalement les mêmes qu'en dimension 1 en remplaçant les valeurs absolues par des normes. On les rappelle tout de même afin de les commenter un petit peu.

**Définition 1.1.** Soit  $(u_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $\mathbb{R}^d$ . On dit que  $(u_n)_{n \in \mathbb{N}}$  converge vers l'élément  $l \in \mathbb{R}^d$  si pour tout  $\varepsilon > 0$  il existe un entier  $N$  tel que pour tout  $n \geq N$  :

$$\|u_n - l\| \leq \varepsilon.$$

On note alors :

$$\lim_{n \rightarrow +\infty} u_n = l.$$

En fait, c'est équivalent à ce que chaque coordonnée de  $(u_n)_{n \in \mathbb{N}}$  converge vers la coordonnée de  $l$  correspondante. Ceci repose sur les inégalités suivantes que je laisse en exercice :

$$\begin{aligned} \forall a := (a_1, \dots, a_d) \in \mathbb{R}^d, \forall b := (b_1, \dots, b_d) \in \mathbb{R}^d, \forall i_0 \in \{1, \dots, d\} : \\ |b_{i_0} - a_{i_0}| \leq \|b - a\| \leq \sum_{i=1}^d |b_i - a_i|. \end{aligned} \quad (3.1)$$

**Proposition 1.2.** Soit  $(u_n := (u_n^1, \dots, u_n^d)_{n \in \mathbb{N}}$  une suite d'éléments de  $\mathbb{R}^d$ , et  $l := (l_1, \dots, l_d)$  un élément de  $\mathbb{R}^d$ . La suite  $(u_n)_{n \in \mathbb{N}}$  converge vers  $l$  si et seulement si pour tout  $i \in \{1, \dots, d\}$  :

$$\lim_{n \rightarrow +\infty} u_n^i = l_i.$$

*Démonstration.* Si  $(u_n)_{n \in \mathbb{N}}$  converge vers  $l$ .

Soit  $i \in \{1, \dots, d\}$ . Soit  $\varepsilon > 0$  et  $N \in \mathbb{N}$  tel que pour tout  $n \geq N$ ,  $\|u_n - l\| \leq \varepsilon$ . Par la première inégalité dans (3.1), pour tout  $n \geq N$  :

$$|u_n^i - l_i| \leq \|u_n - l\| \leq \varepsilon.$$

**S'il y a convergence coordonnée par coordonnée.**

Soit  $\varepsilon > 0$ . Pour chaque  $i \in \{1, \dots, d\}$ , il existe  $N_i$  tel que pour tout  $n \geq N_i$  :

$$|u_n^i - l_i| \leq \frac{\varepsilon}{d}.$$

On pose alors :

$$\max_{i \in \{1, \dots, d\}} N_i.$$

Si  $n \geq N$ , alors en particulier,  $n$  est supérieur à chaque  $N_i$ , donc en utilisant la deuxième inégalité dans (3.1) :

$$\|u_n - l\| \leq \sum_{i=1}^d |u_n^i - l_i| \leq \sum_{i=1}^d \frac{\varepsilon}{d} = \varepsilon.$$

□

La continuité également se définit facilement dans ce cadre. Si  $D$  est un sous-ensemble de  $\mathbb{R}^d$ , on définit la continuité des fonctions de  $D$  dans  $\mathbb{R}$  de la façon suivante.

**Définition 1.3.** Soit  $f$  une fonction de  $D \subset \mathbb{R}^d$  dans  $\mathbb{R}$  et  $x$  un point de  $D$ . On dit que  $f$  est continue en  $x$  si pour tout  $\varepsilon > 0$ , il existe  $\delta > 0$  ayant la propriété suivante :

$$\forall y \in D \text{ tel que } \|y - x\| < \delta, |f(y) - f(x)| < \varepsilon.$$

On dit que  $f$  est continue si pour tout  $x \in D$ ,  $f$  est continue en  $x$ .

*Exemple 1.4.* La norme euclidienne est une fonction continue de  $\mathbb{R}^d$  dans  $\mathbb{R}$ . En effet, soit  $x \in \mathbb{R}^d$  et  $\varepsilon > 0$ . En choisissant  $\delta := \varepsilon$ , on remarque que si  $y \in \mathbb{R}^d$  est tel que  $\|y - x\| < \delta$ , alors par l'inégalité triangulaire :

$$\begin{aligned} \|y\| &= \|x + y - x\| \leq \|x\| + \|y - x\| \leq \|x\| + \varepsilon, \\ \|x\| &= \|y + x - y\| \leq \|y\| + \|x - y\| \leq \|y\| + \varepsilon. \end{aligned}$$

En réunissant ces deux inégalités, on voit que :

$$|\|y\| - \|x\|| \leq \varepsilon.$$

et donc que la norme euclidienne est continue. On voit bien l'utilité de l'inégalité triangulaire !

Les applications linéaires en dimension finie sont continues. Les sommes et les produits de fonctions continues sont continus, de même on reste continu en divisant par une fonction continue ne s'annulant pas. Les compositions de fonctions continues sont continues. En particulier, les polynômes à plusieurs variables, les fractions rationnelles sur leur ensemble de définition et les compositions à gauche ou à droite de ces fonctions par les fonctions usuelles log, exp, sin, cos etc... sont continues.

Comme en dimension 1, la continuité est équivalente à la convergence des valeurs de  $f$  le long des suites convergentes.

**Proposition 1.5.** Soit  $f$  une fonction de  $D \subset \mathbb{R}^d$  dans  $\mathbb{R}$  et  $x$  un point de  $D$ . La fonction  $f$  est continue en  $x$  si et seulement si pour toute suite  $(u_n)_{n \in \mathbb{N}}$  d'éléments de  $D$  tendant vers  $x$ , la suite réelle  $(f(u_n))_{n \in \mathbb{N}}$  converge vers  $f(x)$ .

*Démonstration. Si  $f$  est continue en  $x$ .*

Soit  $(u_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $D$  tendant vers  $x$  et soit  $\varepsilon > 0$ . Soit  $\delta > 0$  tel que

$$\forall y \in D \text{ tel que } \|y - x\| < \delta, |f(y) - f(x)| < \varepsilon.$$

Soit  $N \in \mathbb{N}$  tel que pour tout  $n \geq N$ ,  $\|u_n - x\| \leq \delta$ . Alors pour tout  $n \geq N$  :

$$|f(u_n) - f(x)| < \varepsilon$$

**S'il y a convergence le long des sous-suites convergentes.** On raisonne par l'absurde et on suppose donc que  $f$  n'est pas continue en  $x$ . C'est à dire qu'il existe un  $\varepsilon > 0$  tel que pour tout  $\delta > 0$ , il existe  $y \in D$  avec les propriétés suivantes :

$$\begin{aligned} \|y - x\| &< \delta, \\ |f(y) - f(x)| &\geq \varepsilon. \end{aligned}$$

Soit  $n \in \mathbb{N}$ . On peut appliquer la proposition précédente à  $\delta := 2^{-n}$  et ainsi trouver  $u_n \in D$  tel que :

$$\begin{aligned} \|u_n - x\| &< 2^{-n} \\ |f(u_n) - f(x)| &\geq \varepsilon. \end{aligned}$$

La première inégalité montre que  $(u_n)_{n \in \mathbb{N}}$  converge vers  $x$ , la seconde montre que  $(f(u_n))_{n \in \mathbb{N}}$  ne converge pas vers  $f(x)$ . On a donc une contradiction.  $\square$

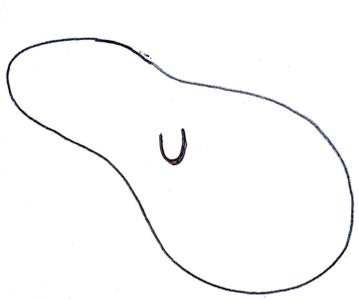
**Remarque très importante 1.6.** On peut définir de la même façon la continuité pour une fonction  $F$  d'un domaine  $D \subset \mathbb{R}^d$  à valeur dans  $\mathbb{R}^p$  avec n'importe quel  $p \geq 1$ . Il suffit pour cela de changer la valeur absolue dans la définition de la continuité en la norme euclidienne sur l'espace d'arrivée  $\mathbb{R}^p$ . Si on écrit une telle fonction coordonnée par coordonnée dans l'espace d'arrivée :

$$\forall x \in D, \quad F(x) = (f_1(x), \dots, f_p(x)),$$

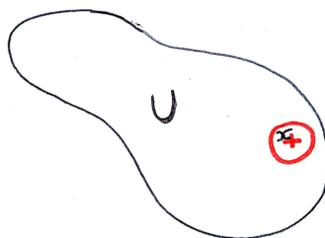
un analogue de la proposition 1.2 implique que  **$F$  est continue si et seulement si  $f_1, \dots, f_p$  sont continues.**

## 2 Les objets de base de la topologie : les ouverts de $\mathbb{R}^d$

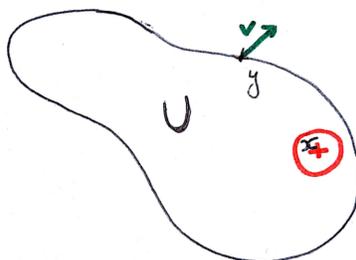
On introduit la notion d'ouvert parce que c'est le bon cadre pour faire du calcul différentiel. En effet, l'objectif du calcul différentiel est de dire : si je connais une fonction  $f$  et sa différentielle  $df$  en un point  $x$  de son domaine de définition  $U$ , alors je connais avec une bonne approximation le comportement de  $f$  dans toutes les directions à partir de  $x$ . Pour que cela ait un sens, encore faut-il qu'il soit possible d'aller dans toutes les directions à partir de  $x$  tout en restant dans  $U$  (au moins un tout petit peu). Pensez à l'ensemble suivant comme si sa frontière n'en faisait pas partie.



Alors autour de chacun de ses points, on peut tracer un petit disque restant dans  $U$ , et à l'intérieur de ce petit disque, on peut bien se promener dans toutes les directions à partir de  $x$ .



En revanche, un point du bord n'aurait pas cette propriété, dans le dessin suivant, la direction  $v$  à partir de  $y$  n'est pas accessible.



Un ouvert ne doit donc pas contenir son bord. Formalisons tout ça.

**Définition 2.1** (Les boules). Soient  $x \in \mathbb{R}^d$  et  $r \geq 0$ .

La **boule ouverte** de centre  $x$  et de rayon  $r$  est l'ensemble :

$$B(x, r) := \{y \in \mathbb{R}^d \mid \|y - x\| < r\}.$$

C'est l'ensemble des points de  $\mathbb{R}^d$  dont la distance à  $x$  est strictement inférieure à  $r$ .

La **boule fermée** de centre  $x$  et de rayon  $r$  est l'ensemble :

$$\overline{B}(x, r) := \{y \in \mathbb{R}^d \mid \|y - x\| \leq r\}.$$

C'est l'ensemble des points de  $\mathbb{R}^d$  dont la distance à  $x$  est inférieure à  $r$  (au sens large).

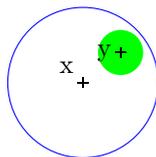
**Définition 2.2** (Les ouverts). Un sous-ensemble  $U$  de  $\mathbb{R}^d$  est dit ouvert si pour tout  $x \in U$ , il existe  $r > 0$  tel que :

$$B(x, r) \subset U.$$

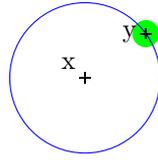
*Exemple 2.3.* • En dimension 1, on retrouve bien avec cette définition le fait que les intervalles ouverts sont de l'un des types :

$$]a, b[, \quad ]-\infty, b[, \quad ]a, +\infty[, \quad ]-\infty, +\infty[ = \mathbb{R}.$$

• Une boule ouverte est ouverte comme en témoigne le dessin suivant.



- $\mathbb{R}^d$  tout entier et  $\emptyset$  sont des ouverts.
- Une boule fermée n'est pas un ouvert comme en témoigne le dessin suivant.



- Plus généralement, les ensembles définis comme l'intérieur d'un polygone ou d'une patate sont ouverts si et seulement si on en enlève le bord. Par exemple, un terrain de basket est ouvert.

**Proposition 2.4.** • Une union quelconque d'ouverts est ouverte.

- Une intersection **finie** d'ouverts est ouverte.

*Démonstration.* **Union.**

Soient  $(U_i)_{i \in I}$  une famille d'ouverts de  $\mathbb{R}^d$  indexée par un ensemble  $I$  et soit

$$x \in U := \bigcup_{i \in I} U_i.$$

Il existe  $i_0 \in I$  tel que

$$x \in U_{i_0}.$$

Comme  $U_{i_0}$  est ouvert, il existe  $r > 0$  tel que :

$$B(x, r) \subset U_{i_0}.$$

Or :

$$U_{i_0} \subset U,$$

donc :

$$B(x, r) \subset U.$$

Comme c'est vrai pour tout  $x \in U$ ,  $U$  est ouvert.

**Intersection finie.**

Soit  $p \in \mathbb{N}^*$  et  $U_1, \dots, U_p$  des ouverts de  $\mathbb{R}^d$ . Soit

$$x \in U := \bigcap_{i=1}^p U_i.$$

pour tout  $i \in \{1, \dots, p\}$ ,  $x \in U_i$ , donc il existe  $r_i > 0$  tel que

$$B(x, r_i) \subset U_i.$$

Donc si on prend

$$r := \min_{i \in \{1, \dots, p\}} r_i,$$

alors pour tout  $i \in \{1, \dots, p\}$ ,

$$B(x, r) \subset U_i.$$

Donc :

$$B(x, r) \subset U.$$

□

**Exercice 2.5.** Trouver une famille d'ouverts dont l'intersection n'est pas ouverte.

### 3 Leurs complémentaires : les fermés de $\mathbb{R}^d$

La définition d'un fermé n'est pas très éclairante.

**Définition 3.1.** Un sous-ensemble  $F$  de  $\mathbb{R}^d$  est un fermé si son complémentaire  ${}^cF$  est un ouvert.

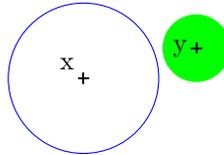
Mais l'idée qu'il faut en avoir, c'est que c'est un ensemble qui contient son bord. Donnons quelques exemples.

*Exemple 3.2.* • En dimension 1, on voit avec cette définition que les intervalles fermés sont les intervalles de l'un des types suivants :

$$[a, b], \quad [a, +\infty[, \quad ]-\infty, b], \quad ]-\infty, +\infty[ = \mathbb{R}.$$

On voit d'ores et déjà que  $\mathbb{R}$  est ouvert et fermé. Ça arrive, on n'est pas obligé de choisir !

- Les boules fermées sont fermées comme en témoigne le dessin suivant.



Mais les boules ouvertes ne sont pas fermées.

- $\mathbb{R}^d$  et  $\emptyset$  sont fermés.
- Les ensembles définis comme l'intérieur d'un polygone ou d'une patate sont fermés si et seulement si on en garde le bord. Par exemple, un terrain de tennis est fermé.

On utilise presque exclusivement la caractérisation suivante pour les fermés, qui dit essentiellement qu'une suite de  $F$  ne peut pas converger en dehors de  $F$ .

**Proposition 3.3** (Caractérisation séquentielle des fermés). *Un sous-ensemble  $F$  de  $\mathbb{R}^d$  est fermé si et seulement si pour toute suite  $(u_n)_{n \in \mathbb{N}}$  d'éléments de  $F$  convergeant dans  $\mathbb{R}^d$  vers une certaine limite  $l \in \mathbb{R}^d$ , alors  $l \in F$ .*

*Démonstration. Si  $F$  est fermé.*

Soit  $(u_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $F$  convergeant vers  $l \in \mathbb{R}^d$ . On raisonne par l'absurde. Si  $l \notin F$ , alors  $l \in {}^cF$  qui est ouvert. Il existe donc  $r > 0$  tel que  $B(l, r) \subset {}^cF$ . Soit  $N \in \mathbb{N}$  tel que pour tout  $n \geq N$ ,  $\|u_n - l\| \leq r/2$ . Alors en particulier,  $u_N \in F$ , et  $u_N \in B(l, r) \subset {}^cF$ . Il y a donc une contradiction et  $l \in F$ .

**Si les limites des suites de  $F$  sont dans  $F$ .**

Soit  $x \in {}^cF$ . Montrons par l'absurde qu'il existe  $r > 0$  tel que  $B(x, r) \subset {}^cF$ . Si ce n'est pas le cas, alors en particulier, pour tout  $n \in \mathbb{N}$  :

$$F \cap B(x, 2^{-n}) \neq \emptyset.$$

On peut donc choisir  $u_n \in F \cap B(x, 2^{-n})$ . La suite  $(u_n)_{n \in \mathbb{N}}$  d'éléments de  $F$  converge clairement vers  $x$ , ce qui contredit l'hypothèse  $x \in {}^cF$ .  $\square$

**Proposition 3.4.** • Une intersection quelconque de fermés est fermée.

- Une union **finie** de fermés est fermée.

*Démonstration.* Si  $(A_i)_{i \in I}$  est une famille de sous-ensemble de  $\mathbb{R}^d$  indexée par un ensemble  $I$ , on a les égalités suivantes :

$$\bigcup_{i \in I} {}^c A_i = {}^c \left( \bigcap_{i \in I} A_i \right),$$

$$\bigcap_{i \in I} {}^c A_i = {}^c \left( \bigcup_{i \in I} A_i \right).$$

Cette proposition est donc la même que la proposition 2.4. □

## 4 Continuité, ouverts et fermés

Il y a un théorème très satisfaisant reliant ouverts, fermés et fonctions continues. Il est tellement satisfaisant qu'on l'utilise parfois comme définition de la continuité. Cela permet de ne plus avoir besoin de la notion de norme, ni même de celle de distance, mais simplement de celle d'ouverts que l'on utilise alors comme brique fondamentale. Le théorème s'énonce de la façon suivante.

**Théorème 4.1.** *Soit  $f$  une fonction de  $\mathbb{R}^d$  dans  $\mathbb{R}^p$  où  $d$  et  $p$  sont deux entiers naturels non nuls. Les trois propositions suivantes sont équivalentes.*

(i)  $f$  est continue.

(ii) Pour tout ouvert  $U$  de  $\mathbb{R}^p$ , l'image réciproque par  $f$  de  $U$  est ouverte.

(iii) Pour tout fermé  $F$  de  $\mathbb{R}^p$ , l'image réciproque par  $f$  de  $F$  est fermée.

*Démonstration.* Le fait que (ii) et (iii) sont équivalents vient du fait que si  $A$  est un sous-ensemble de  $\mathbb{R}^p$  :

$$f^{-1}({}^c A) = {}^c f^{-1}(A).$$

Je laisse tout ça en exercice.

Montrons que (i) implique (iii). Soit  $F$  un fermé de  $\mathbb{R}^p$  et soit  $(u_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $f^{-1}(F)$  qui converge vers  $l \in \mathbb{R}^d$ . Par définition, pour chaque  $n \in \mathbb{N}$ ,  $f(u_n) \in F$  et par continuité de  $f$ ,  $(f(u_n))_{n \in \mathbb{N}}$  converge vers  $f(l) \in \mathbb{R}^p$ . Comme  $F$  est fermé,  $f(l) \in F$ , et donc  $l \in f^{-1}(F)$ .

Montrons que (ii) implique (i). Soit  $x \in \mathbb{R}^d$ . On va montrer que  $f$  est continue en  $x$ . Soit  $\varepsilon > 0$  la boule ouverte de  $\mathbb{R}^p$   $B(f(x), \varepsilon)$  (on note de la même façon les boules de  $\mathbb{R}^d$  et  $\mathbb{R}^p$  ce qui est un peu abusif mais bon) est un ouvert de  $\mathbb{R}^p$ . Donc par hypothèse,  $f^{-1}(B(f(x), \varepsilon))$  est un ouvert de  $\mathbb{R}^d$ . Or  $x \in f^{-1}(B(f(x), \varepsilon))$ . Donc il existe  $\delta > 0$  tel que :

$$B(x, \delta) \subset f^{-1}(B(f(x), \varepsilon)).$$

Soit  $y \in \mathbb{R}^d$  tel que  $\|y - x\| < \delta$ . Alors  $y \in B(x, \delta)$ , donc  $y \in f^{-1}(B(f(x), \varepsilon))$ , donc  $f(y) \in B(f(x), \varepsilon)$ , donc  $\|f(y) - f(x)\| < \varepsilon$  (et ouais, on note aussi les normes euclidiennes de  $\mathbb{R}^d$  et  $\mathbb{R}^p$  de la même façon). Pour  $\varepsilon$  fixé mais quelconque, on a trouvé  $\delta$  satisfaisant la bonne condition, donc  $f$  est continue. □

**Exemple très important 4.2.** Pour montrer qu'un ensemble est ouvert ou fermé, la façon la plus économique est très souvent d'utiliser ce théorème. Par exemple, on verra très souvent dans la suite du cours des ensembles définis par :

$$E := \{x \in \mathbb{R}^d \mid f(x) < M\}$$

où  $f$  est une certaine fonction continue de  $\mathbb{R}^d$  dans  $\mathbb{R}$ . Dans ce cas,  $E$  est ouvert, et la démonstration se fait très très rapidement :

$$E = f^{-1}(] - \infty, M[),$$

or  $] - \infty, M[$  est un ouvert de  $\mathbb{R}$  et  $f$  est continue. Donc  $E$  est ouvert de  $\mathbb{R}^d$ . De la même façon, on peut montrer très rapidement que :

$$G := \{x \in \mathbb{R}^d \mid f(x) \leq M\}$$

est un fermé de  $\mathbb{R}^d$ .

## 5 Adhérence et intérieur

Les propriétés des ouverts et des fermés nous permettent de définir ce qu'est "le plus grand ouvert inclu dans un ensemble" et "le plus petit fermé contenant un ensemble". Voyons de quoi il s'agit.

**Définition 5.1** (Intérieur). Soit  $E$  un sous-ensemble de  $\mathbb{R}^d$ . Le sous-ensemble de  $\mathcal{P}(\mathbb{R}^d)$  :

$$\mathcal{U} := \{U \subset \mathbb{R}^d, \text{ ouvert} \mid U \subset E\}$$

est non vide puisqu'il contient au moins  $\emptyset$ . On peut donc définir l'intérieur de  $E$  :

$$\overset{\circ}{E} := \bigcup_{U \in \mathcal{U}} U.$$

L'intérieur de  $E$  est une union d'ouverts, c'est donc également un ouvert, et il est facile de vérifier qu'il est inclu dans  $E$  et qu'il a la propriété que pour tout ouvert  $U$  inclu dans  $E$ ,  $U \subset \overset{\circ}{E}$ . Dans les faits, on utilise plutôt la caractérisation suivante.

**Proposition 5.2.** Soit  $E$  un sous-ensemble de  $\mathbb{R}^d$ .

$$\overset{\circ}{E} = \{x \in E \mid \exists r > 0 \text{ t.q. } B(x, r) \subset E\}.$$

*Démonstration.* On note  $V$  l'ensemble défini à droite du signe égal. Montrons que  $V \subset \overset{\circ}{E}$ . Soit  $x \in V$  et  $r > 0$  tel que  $B(x, r) \subset E$ .  $B(x, r)$  est un ouvert inclu dans  $E$ , donc il est inclu dans  $\overset{\circ}{E}$ . En particulier,  $x \in \overset{\circ}{E}$ .

Montrons que  $\overset{\circ}{E} \subset V$ . Soit  $x \in \overset{\circ}{E}$ . Il existe un ouvert  $U \in \mathcal{U}$  tel que  $x \in U \subset E$ . Il existe donc  $r > 0$  tel que  $B(x, r) \subset U \subset E$ . Donc  $x \in V$ .  $\square$

L'idée qu'il faut avoir de l'intérieur d'un ensemble, c'est qu'on prend l'ensemble et qu'on lui enlève le bord.

**Définition 5.3.** Soit  $E$  un sous-ensemble de  $\mathbb{R}^d$ . Le sous-ensemble de  $\mathcal{P}(\mathbb{R}^d)$  :

$$\mathcal{F} := \{F \subset \mathbb{R}^d, \text{ fermé} \mid E \subset F\}$$

est non vide puisqu'il contient au moins  $\mathbb{R}^d$ . On peut donc définir l'adhérence de  $E$  :

$$\overline{E} := \bigcap_{F \in \mathcal{F}} F.$$

L'adhérence de  $E$  est une intersection de fermés, c'est donc également un fermé. Il est facile de voir qu'il contient  $E$ . De plus il a la propriété que pour tout fermé  $F$  contenant  $E$ ,  $\overline{E} \subset F$ . On peut le caractériser de la façon suivante.

**Proposition 5.4.** Soit  $E$  un sous-ensemble de  $\mathbb{R}^d$ . Son adhérence  $\overline{E}$  est l'ensemble des limites de suites d'éléments de  $E$ .

*Démonstration.* On note  $F$  l'ensemble des limites de suites d'éléments de  $E$ . Montrons que  $F \subset \overline{E}$ . Soit  $x \in F$  et  $(u_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $E$  qui converge vers  $x$ . Comme  $\overline{E}$  est fermé et contient tous les éléments de la suite  $(u_n)_{n \in \mathbb{N}}$ , la limite de  $(u_n)_{n \in \mathbb{N}}$ , c'est à dire  $x$ , est dans  $\overline{E}$ .

Montrons que  $\overline{E} \subset F$ . Comme  $E \subset F$  (considérer les suites stationnaires), il suffit de montrer que  $F$  est fermé. Soit  $(x_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $F$  convergent vers  $x \in \mathbb{R}^d$ . Montrons que  $x \in F$ . Soit  $n \in \mathbb{N}$ , comme il existe une suite d'éléments de  $E$  qui converge vers  $x_n$ , il existe en particulier au moins un élément  $u_n$  de  $E$  tel que  $\|x_n - u_n\| \leq 2^{-n}$ .  $(u_n)_{n \in \mathbb{N}}$  est une suite d'éléments de  $E$ , et on observe qu'elle converge vers  $x$ . Donc  $x \in F$ , et  $F$  est fermé.  $\square$

**Exercice 5.5.** Soit  $E$  un sous-ensemble de  $\mathbb{R}^d$ . Montrer que :

$$\begin{aligned} {}^c(\overset{\circ}{E}) &= \overline{{}^c E}, \\ {}^c(\overline{E}) &= ({}^c \overset{\circ}{E}). \end{aligned}$$

## 6 La compacité

En dimension 1, on avait vu qu'une fonction continue d'un intervalle fermé et borné à valeur dans  $\mathbb{R}$  était bornée et atteignait ses bornes. Comment peut-on généraliser ça en dimension supérieure ?

**Définition 6.1.** On dit qu'un sous-ensemble  $K$  de  $\mathbb{R}^d$  est un compact si toute suite  $(u_n)_{n \in \mathbb{N}}$  d'éléments de  $K$  admet une sous-suite convergeant **dans  $K$** .

Cette définition n'est pas sans rappeler le théorème de Bolzano-Weierstrass qui était en fait l'ingrédient essentiel dans la preuve du théorème sus-mentionné. Et en effet, avec cette définition, on peut facilement montrer le théorème suivant.

**Théorème 6.2.** Soit  $f$  une fonction continue d'un compact  $K$  de  $\mathbb{R}^d$  à valeurs dans  $\mathbb{R}$ . Alors  $f$  est bornée et atteint ses bornes.

*Démonstration.* Je ne vais même pas le faire tellement c'est pareil qu'en dimension 1. □

Mais en fait, ce théorème ne sert pas à grand chose si on ne sait pas caractériser les compacts, car la définition n'est pas très manipulable. Heureusement, en dimension finie, on sait très bien le faire. Définissons d'abord ce qu'est un ensemble borné.

**Définition 6.3.** On dit qu'un sous-ensemble  $E$  de  $\mathbb{R}^d$  est borné si il existe  $R > 0$  tel que :

$$E \subset B(0, R).$$

Un sous-ensemble  $E$  de  $\mathbb{R}^d$  est donc borné si l'ensemble des normes de ses éléments est majoré. Énonçons la caractérisation des ensembles compacts en dimension finie.

**Théorème 6.4.** Un sous ensemble  $K \subset \mathbb{R}^d$  est compact si et seulement si  $K$  est fermé et borné.

*Démonstration.* **Si  $K$  est compact.**

Montrons que  $K$  est fermé. Soit  $(u_n)_{n \in \mathbb{N}}$  une suite de  $K$  qui converge dans  $\mathbb{R}^d$ . Soit  $l$  sa limite. Comme  $K$  est compact, il existe  $(v_n)_{n \in \mathbb{N}}$  une sous-suite de  $(u_n)_{n \in \mathbb{N}}$  qui converge dans  $K$ . Or la limite de  $(v_n)_{n \in \mathbb{N}}$  est  $l$  pour la même raison que dans le cas réel (si une suite converge vers une limite, chacune de ses sous-suites converge vers cette limite). Donc  $l \in K$ .

Montrons que  $K$  est borné. La norme est une fonction continue,  $K$  est compact, donc la norme est majorée sur  $K$ .

**Si  $K$  est fermé et borné.**

Soit  $R$  tel que  $K \subset B(0, R)$ . Alors on a également (voir dessin ci-dessous ou utiliser la première inégalité dans (3.1)) :

$$K \subset [-R, R]^d.$$

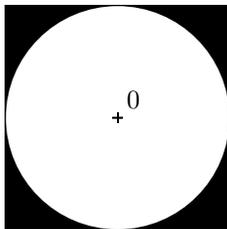


FIGURE 3.1 – Le disque blanc représente la boule de centre 0 et de rayon  $R$  de dimension 2. Le carré noir représente  $[-R, R]^2$ . Le disque est bien inclus dans le carré.

La suite  $(u_n)_{n \in \mathbb{N}}$  est donc une suite de  $[-R, R]^d$ . Montrons qu'on peut en extraire une sous-suite convergente dans  $\mathbb{R}^d$ . Pour tout  $n \in \mathbb{N}$ , on note  $(u_n^1, \dots, u_n^d)$  les coordonnées de  $u_n$ . La première coordonnée

$(u_n^1)_{n \in \mathbb{N}}$  est une suite de  $[-R, R]$ . Par le théorème de Bolzano-Weierstrass, il existe une application strictement croissante  $\alpha_1$  de  $\mathbb{N}$  dans  $\mathbb{N}$  telle que  $(u_{\alpha_1(n)}^1)_{n \in \mathbb{N}}$  converge. La première coordonnée de  $(u_{\alpha_1(n)})_{n \in \mathbb{N}}$  converge donc.

La suite  $(u_{\alpha_1(n)}^2)_{n \in \mathbb{N}}$  est dans  $[-R, R]$ . Par le théorème de Bolzano-Weierstrass, il existe une application  $\alpha_2$  de  $\mathbb{N}$  dans  $\mathbb{N}$  telle que  $(u_{\alpha_1 \circ \alpha_2(n)}^2)_{n \in \mathbb{N}}$  converge. De plus, la suite  $(u_{\alpha_1 \circ \alpha_2(n)}^1)_{n \in \mathbb{N}}$  est une sous-suite de  $(u_{\alpha_1(n)}^1)_{n \in \mathbb{N}}$  qui converge. Donc elle converge également et les deux premières coordonnées de  $(u_{\alpha_1 \circ \alpha_2(n)})_{n \in \mathbb{N}}$  convergent.

On peut itérer la construction et trouver  $\alpha_3, \dots, \alpha_d$  des applications strictement croissantes de  $\mathbb{N}$  dans  $\mathbb{N}$  ayant la propriété que pour tout  $k \in \{1, \dots, d\}$ , les  $k$  premières coordonnées de  $(u_{\alpha_1 \circ \alpha_2 \circ \dots \circ \alpha_k(n)})_{n \in \mathbb{N}}$  soient convergentes. En posant :

$$\alpha := \alpha_1 \circ \dots \circ \alpha_d,$$

$\alpha$  est une application strictement croissante de  $\mathbb{N}$  dans  $\mathbb{N}$  et toutes les coordonnées de  $(u_{\alpha(n)})_{n \in \mathbb{N}}$  convergent. Par la proposition 1.2,  $(u_{\alpha(n)})_{n \in \mathbb{N}}$  converge. C'est donc une sous-suite convergente de  $(u_n)_{n \in \mathbb{N}}$ . Soit  $l$  sa limite. Comme  $K$  est fermé,  $l \in K$ . On a donc extrait de n'importe quelle suite d'éléments de  $K$  une sous-suite qui converge dans  $K$ , donc  $K$  est compact.  $\square$

**Corollaire 6.1.**     • *En dimension finie, les boules fermées sont compactes.*

- *Si  $K$  est un compact et  $F$  est un fermé, alors  $K \cap F$  est compact.*

# Chapitre 4

## Compléments d'algèbre linéaire

On se place dans l'espace vectoriel standard de dimension  $d$ ,  $\mathbb{R}^d$  avec  $d \in \mathbb{N}^*$ .

### 1 Le déterminant

#### 1.1 Intuition

On pose le problème suivant. Y a-t-il un moyen simple (un calcul) pour déterminer si une famille de  $d$  vecteurs  $(v_1, \dots, v_d)$  de  $\mathbb{R}^d$  est libre ou liée? En réfléchissant un peu, on se rend compte que  $(v_1, \dots, v_d)$  est liée si et seulement si chacun de ses vecteurs est inclu dans le même sous espace vectoriel strict de  $\mathbb{R}^d$ , c'est à dire si le parallélogramme ( $d$ -dimensionnel) dont les arrêtes sont  $v_1, \dots, v_d$  est plat. Or ce parallélogramme est plat si et seulement si son volume ( $d$ -dimensionnel) est nul.

**Idée :** une famille de  $d$  vecteurs  $(v_1, \dots, v_d)$  est liée si et seulement si le volume du parallélogramme d'arrêtes  $v_1, \dots, v_d$  est nul.

En fait, pour des raisons qui seront claires par la suite, il vaut mieux considérer le volume *algébrique*, c'est à dire avec un signe  $+$  si la famille est orientée dans le sens direct, et avec un signe  $-$  si elle est orientée dans le sens indirect, quoi que cela veuille dire...

Plaçons-nous en dimension 2. On considère alors des parallélogrammes classiques et le volume est en fait ici l'aire. On note  $\mathcal{A}(x, y)$  l'aire algébrique du parallélogramme d'arrêtes  $x$  et  $y$ .

- Si  $x$  et  $y$  sont dans  $\mathbb{R}^2$  et si  $\lambda$  est dans  $\mathbb{R}$ , alors

$$\mathcal{A}(\lambda x, y) = \mathcal{A}(x, \lambda y) = \lambda \mathcal{A}(x, y).$$

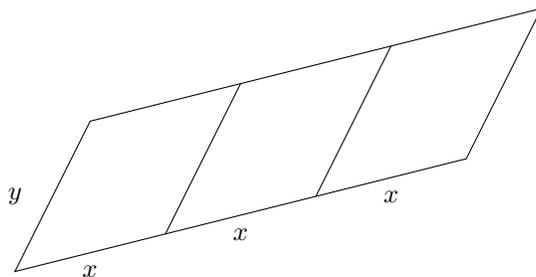


FIGURE 4.1 – Quand  $\lambda = 3$ , on voit bien que l'aire de la figure est multipliée par 3

On voit bien ici qu'il faut prendre l'aire algébrique. Pour garder l'homogénéité quand  $\lambda < 0$ , on veut multiplier l'aire par un nombre négatif. Ça tombe bien puisque justement, on change l'orientation.

- Si  $x, y$  et  $z$  sont dans  $\mathbb{R}^2$ ,

$$\mathcal{A}(x + y, z) = \mathcal{A}(x, z) + \mathcal{A}(y, z).$$

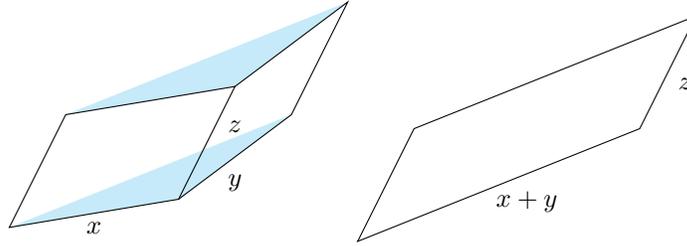


FIGURE 4.2 – Ces deux constructions suivantes ont clairement la même aire car les parties en bleu sont d'aires égales.

On aurait pu raisonner de même sur la deuxième coordonnée.

- Si  $x$  et  $y$  sont dans  $\mathbb{R}^2$ ,

$$\mathcal{A}(x, y) = -\mathcal{A}(y, x),$$

puisque échanger  $x$  et  $y$  revient simplement à changer l'orientation. En particulier,

$$\mathcal{A}(x, x) = -\mathcal{A}(x, x),$$

et donc vaut 0. On retrouve bien que l'aire d'un parallélogramme plat est nulle.

L'aire algébrique est donc **linéaire en chacune des coordonnées** et **anti-symétrique** (change de signe quand on échange les coordonnées). Ces deux propriétés suffisent à déduire une formule de l'aire algébrique en fonction des coordonnées de  $x$  et  $y$ . En effet, en posant

$$x = \begin{pmatrix} a \\ b \end{pmatrix} = ae_1 + be_2 \quad \text{et} \quad y = \begin{pmatrix} c \\ d \end{pmatrix} = ce_1 + de_2,$$

on trouve

$$\begin{aligned} \mathcal{A}(ae_1 + be_2, ce_1 + de_2) &= a\mathcal{A}(e_1, ce_1 + de_2) + b\mathcal{A}(e_2, ce_1 + de_2) \\ &= ac\mathcal{A}(e_1, e_1) + ad\mathcal{A}(e_1, e_2) + bc\mathcal{A}(e_2, e_1) + bd\mathcal{A}(e_2, e_2) \\ &= (ad - bc)\mathcal{A}(e_1, e_2). \end{aligned}$$

En convenant  $\mathcal{A}(e_1, e_2) = 1$ , ce qui revient à fixer l'orientation et l'unité de mesure de l'aire, on retrouve la formule classique du déterminant en dimension 2 :  $ad - bc$ .

## 1.2 Formalisme

On se replace en dimension  $d$ .

**Définition 1.1.** On dit qu'une application  $f$  de  $(\mathbb{R}^d)^d$  dans  $\mathbb{R}$  est une forme  $d$ -linéaire si pour tout  $i \in \llbracket 1, d \rrbracket$ , pour tous  $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_d$  dans  $\mathbb{R}^d$ , l'application qui à  $x \in \mathbb{R}^d$  associe

$$f(v_1, \dots, v_{i-1}, x, v_{i+1}, \dots, v_d) \in \mathbb{R}$$

est linéaire.

On dit qu'elle est alternée si pour tous  $i$  et  $j$  distincts dans  $\llbracket 1, d \rrbracket$ , tous  $(v_k)_{k \neq i, j}$ , et tous  $x$  et  $y$  dans  $\mathbb{R}^d$ , avec des légers abus de notations qui me semblent clairs,

$$f(v_1, \dots, x, \dots, y, \dots, v_d) = -f(v_1, \dots, y, \dots, x, \dots, v_d).$$

**Proposition 1.2.** Une forme  $d$ -linéaire  $f$  est alternée si et seulement si pour tous  $v_1, \dots, v_d$  dans  $\mathbb{R}^d$ , si il existe  $i$  et  $j$  distincts tels que  $v_i = v_j$ , alors  $f(v_1, \dots, v_d) = 0$ .

*Démonstration.* Soient  $v_1, \dots, v_d$  dans  $\mathbb{R}^d$  et supposons que  $v_i = v_j = v$  avec  $i$  et  $j$  distincts. Alors en échangeant les coordonnées  $i$  et  $j$

$$f(v_1, \dots, v \dots, v, \dots, v_d) = -f(v_1, \dots, v \dots, v, \dots, v_d),$$

donc

$$2f(v_1, \dots, v \dots, v, \dots, v_d) = 0,$$

donc  $f(v_1, \dots, v_d) = 0$ . Réciproquement, si l'affirmation est vraie, alors prenons  $v_1, \dots, v_d$  dans  $\mathbb{R}^d$ . On a alors

$$f(v_1, \dots, v_i + v_j \dots, v_i + v_j, \dots, v_d) = 0,$$

et en utilisant la  $d$ -linéarité,

$$\begin{aligned} f(v_1, \dots, v_i + v_j \dots, v_i + v_j, \dots, v_d) &= f(v_1, \dots, v_i \dots, v_i, \dots, v_d) \\ &\quad + f(v_1, \dots, v_i \dots, v_j, \dots, v_d) \\ &\quad + f(v_1, \dots, v_j \dots, v_i, \dots, v_d) \\ &\quad + f(v_1, \dots, v_j \dots, v_j, \dots, v_d), \end{aligned}$$

puis par hypothèse,

$$\begin{aligned} f(v_1, \dots, v_i + v_j \dots, v_i + v_j, \dots, v_d) &= f(v_1, \dots, v_i \dots, v_j, \dots, v_d) \\ &\quad + f(v_1, \dots, v_j \dots, v_i, \dots, v_d). \end{aligned}$$

Donc

$$f(v_1, \dots, v_i \dots, v_j, \dots, v_d) + f(v_1, \dots, v_j \dots, v_i, \dots, v_d) = 0,$$

et en passant un terme de l'autre côté du signe égal, on obtient le caractère alterné de  $f$ .  $\square$

On énonce maintenant un théorème que l'on ne démontre pas mais dont la démonstration n'est qu'une généralisation de ce qu'on a vu en dimension 2.

**Théorème 1.3.** L'ensemble des formes  $d$ -linéaires alternées sur  $\mathbb{R}^d$  est un espace vectoriel de dimension 1. En particulier, il existe une unique forme  $d$ -linéaire alternée valant 1 en  $(e_1, \dots, e_d)$ . C'est le déterminant, on le note  $\det$  et on peut le calculer (en théorie) grâce à une formule explicite qui est un polynôme en les coordonnées de  $v_1, \dots, v_d$ . Pour toute  $f$   $d$ -linéaire alternée et tous  $v_1, \dots, v_d$ , on a alors

$$f(v_1, \dots, v_d) = f(e_1, \dots, e_d) \det(v_1, \dots, v_d).$$

Je vous encourage à vous convaincre de ce théorème, par exemple en dimension 3. C'est un bon exercice même si c'est un peu pénible à rédiger. La formule du déterminant n'est pas très importante car elle est horrible à calculer à la main, et ne sert globalement que pour montrer des théorèmes, ou pour les ordinateurs. Il est par contre très important de savoir qu'elle existe! Je la donne quand même pour que vous sachiez à quoi ça ressemble. On note

$$\forall i \in \llbracket 1, d \rrbracket, \quad v_i = \begin{pmatrix} v_{1i} \\ \vdots \\ v_{di} \end{pmatrix}.$$

Ensuite, on note  $\mathcal{S}_d$  l'ensemble des permutations de  $\llbracket 1, d \rrbracket$ , c'est à dire l'ensemble des bijections de  $\llbracket 1, d \rrbracket$  dans  $\llbracket 1, d \rrbracket$ . À chaque  $\sigma \in \mathcal{S}_d$  correspond l'ordre dans lequel sont donnés les éléments de  $\llbracket 1, d \rrbracket$  par  $\sigma$  :  $(\sigma(1), \dots, \sigma(d))$ . On peut alors les remettre dans l'ordre en échangeant des termes 2 à 2. Par exemple, si  $d = 3$ ,  $\sigma(1) = 3$ ,  $\sigma(2) = 1$  et  $\sigma(3) = 2$ , on obtient  $(3, 1, 2)$ . Pour les remettre dans l'ordre, on peut par exemple

échanger le 3 et le 2, on obtient alors (2, 1, 3). On peut ensuite échanger le 1 et le 2 pour obtenir (1, 2, 3). La façon de la faire n'est pas unique, mais la parité du nombre d'étape ne dépend pas de la façon que l'on a choisie (c'est un théorème, ce n'est pas évident !). On peut alors définir la *signature* de la permutation  $\varepsilon(\sigma)$  de la façon suivante. Si on a eu besoin d'un nombre pair d'étapes, on note  $\varepsilon(\sigma) = 1$ . Si on a eu besoin d'un nombre impair d'étapes, on note  $\varepsilon(\sigma) = -1$ .

On a alors

$$\det(v_1, \dots, v_d) = \sum_{\sigma \in \mathcal{S}_d} \varepsilon(\sigma) v_{\sigma(1)1} \times \dots \times v_{\sigma(d)d}.$$

**Définition 1.4.** Si  $M$  est une matrice carrée de taille  $d$ . Son déterminant noté  $\det(M)$  est par définition le déterminant de ses vecteurs colonnes.

La matrice identité est la matrice dont les colonnes sont les vecteurs de la base canonique. Donc

$$\det(I_d) = \det(e_1, \dots, e_d) = 1.$$

Par définition de ce qu'est la matrice d'un endomorphisme, si  $F$  est un endomorphisme de  $\mathbb{R}^d$  de matrice  $M$  dans la base canonique de  $\mathbb{R}^d$ ,  $\det(M)$  est le déterminant des vecteurs  $F(e_1), \dots, F(e_d)$ . Intuitivement, on détermine le volume algébrique des images par  $F$  des vecteurs de base.

### 1.3 Règles de calcul théoriques

Voyons les propriétés essentielles du déterminant.

**Proposition 1.5.** Soit  $M$  une matrice carrée de taille  $d$ .

$$\det({}^t M) = \det(M).$$

En particulier, on aurait pu choisir comme définition du déterminant d'une matrice le déterminant de ses vecteurs lignes. Mais bon on n'aurait plus l'interprétation en termes d'endomorphismes. La preuve utilise la formule du déterminant et le fait que la signature de l'inverse d'une permutation est égale à la signature de cette permutation, mais je n'insiste pas.

La proposition suivante justifie à elle seule l'étude du déterminant.

**Proposition 1.6.** Soient  $A$  et  $B$  des matrices carrées de taille  $d$ . On a alors

$$\det(A \times B) = \det(A) \times \det(B).$$

*Démonstration.* Soit  $f$  l'application qui aux vecteurs  $v_1, \dots, v_d$  associe

$$f(v_1, \dots, v_d) := \det(Av_1, \dots, Av_d).$$

C'est une application  $d$ -linéaire et alternée qui vérifie

$$f(e_1, \dots, e_d) = \det(A)$$

car pour tout  $i$ ,  $Ae_i$  est la  $i$ -ième colonne de  $A$ , et en notant  $B_1, \dots, B_d$  les colonnes de  $B$ ,

$$f(B_1, \dots, B_d) = \det(A \times B)$$

car pour tout  $i$ ,  $AB_i$  est la  $i$ -ième colonne de  $A \times B$ . Donc par le théorème 1.3,

$$\det(A \times B) = f(B_1, \dots, B_d) = \det(A) \times \det(B).$$

□

Ce théorème a une multitude d'applications. Voyons-en quelques-unes.

**Corollaire 1.1.** • Si  $A$  et  $B$  sont deux matrices carrées de taille  $d$ , alors  $\det(A \times B) = \det(B \times A)$ .

- En particulier, si  $A$  est une matrice et  $P$  est une matrice inversible,

$$\det(PAP^{-1}) = \det(A).$$

Le déterminant ne dépend pas de la base que l'on considère. On peut donc définir le **déterminant d'un endomorphisme** comme le déterminant de sa matrice dans n'importe quelle base.

- Si  $A$  est une matrice carrée de taille  $d$ ,  $A$  est inversible si et seulement si  $\det(A) \neq 0$ . Dans ce cas,

$$\det(A^{-1}) = \det(A)^{-1}.$$

Le déterminant joue donc pleinement le rôle que l'on souhaitait. Pour déterminer si une matrice est inversible ou non, c'est à dire pour savoir si l'image des vecteurs de base forme une famille libre, il suffit de calculer le déterminant de la matrice. Si ce nombre est non-nul, la matrice est inversible.

*Démonstration.* Le premier point est simplement une conséquence du fait que le produit entre deux nombre réel est commutatif, et le deuxième point en découle directement.

Pour ce qui est du troisième point, commençons par montrer que si une matrice n'est pas inversible, alors son déterminant est nul. Soit  $A$  une matrice non inversible. On note  $A_1, \dots, A_d$  ses colonnes. Le fait que  $A$  soit non-inversible est équivalent au fait que ces colonnes soient liées, c'est à dire que l'on peut écrire une colonne comme combinaison linéaire des autres : il existe  $i$  et des réels  $(\lambda_j)_{j \neq i}$  tels que

$$A_i = \sum_{j \neq i} \lambda_j A_j.$$

On a alors

$$\begin{aligned} \det(A) &= \det(A_1, \dots, A_d) = \det\left(A_1, \dots, \sum_{j \neq i} \lambda_j A_j, \dots, A_d\right) \\ &= \sum_{j \neq i} \lambda_j \det(A_1, \dots, A_{i-1}, A_j, A_{i+1}, \dots, A_d). \end{aligned}$$

Mais chacun de ces termes est nul puisque  $A_j$  y apparait deux fois. Donc

$$\det(A) = 0.$$

Maintenant, si  $A$  est inversible, soit  $A^{-1}$  son inverse. On a

$$\det(A) \det(A^{-1}) = \det(A \times A^{-1}) = \det(I_d) = 1.$$

Donc  $\det(A)$  et  $\det(A^{-1})$  sont non-nuls, et sont inverses l'un de l'autre. □

*Remarque 1.7.* Supposons que  $D$  soit une matrice diagonale :

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_d \end{pmatrix},$$

alors les colonnes de  $D$  sont  $\lambda_1 e_1, \dots, \lambda_d e_d$ . Donc par  $d$ -linéarité, on conclut que  $\det(D) = \lambda_1 \times \dots \times \lambda_d$ . Le déterminant d'une matrice diagonale est égal au produit de ses éléments diagonaux.

Ensuite, si  $A$  est une matrice diagonalisable, soit  $P$  tel que  $D := PAP^{-1}$  soit diagonale. Alors comme  $\det(A) = \det(D)$ , qui lui-même vaut le produit des éléments diagonaux de  $D$ . Mais les éléments diagonaux de  $D$  sont les valeurs propres de  $A$  (avec multiplicité), de sorte que **si  $A$  est diagonalisable, son déterminant est le produit de ses valeurs propres comptées avec multiplicité.**

## 1.4 Outils de calcul pratique

La première chose que j'ai envie de dire dans ce paragraphe, c'est que lorsque l'on fait des calculs explicites, on note le déterminant comme ça :

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} := \det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

**Dimension 2** D'abord, en dimension 2, on connaît la formule :

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

**Ajout de lignes et colonnes** Lorsque l'on calcule un déterminant, on peut toujours ajouter à une colonne une combinaison linéaire des autres colonnes. En effet, si  $A$  est une matrice dont les colonnes sont notées  $A_1, \dots, A_d$ , et si  $(\lambda_j)_{j \neq i}$  sont des réels,

$$\begin{aligned} \det \left( A_1, \dots, A_i + \sum_{j \neq i} \lambda_j A_j, \dots, A_d \right) \\ = \det(A) + \sum_{j \neq i} \lambda_j \det(A_1, \dots, A_{i-1}, A_j, A_{i+1}, \dots, A_d) \\ = \det(A). \end{aligned}$$

On aurait pu faire pareil avec les lignes.

Par exemple, si on calcule

$$\begin{vmatrix} 4 & 1 & 5 \\ 6 & 1 & 3 \\ 4 & 1 & 2 \end{vmatrix},$$

on peut enlever à la première colonne 4 fois la deuxième, et ainsi obtenir

$$\begin{vmatrix} 4 & 1 & 5 \\ 6 & 1 & 3 \\ 4 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 0 & 1 & 5 \\ 2 & 1 & 3 \\ 0 & 1 & 2 \end{vmatrix}.$$

**Développement par rapport à une ligne ou à une colonne** La formule du déterminant permet de démontrer (exercice) la formule suivante. Soit  $A$  une matrice carrée de taille  $d > 1$  dont les colonnes sont notées  $A_1, \dots, A_d$  et les coefficients  $(a_{ij})_{i,j \in \llbracket 1, d \rrbracket}$ . Si  $i$  et  $j$  sont dans  $\llbracket 1, d \rrbracket$ , on note  $A^{ij}$  la matrice carrée de taille  $d - 1$  construite en enlevant à  $A$  sa  $i$ -ième ligne et sa  $j$ -ième colonne :

$$\text{si } A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad \text{alors } A^{12} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}.$$

Soit  $j_0 \in \llbracket 1, d \rrbracket$ . On a alors

$$\det(A) = \sum_{i=1}^d (-1)^{i+j_0} a_{ij_0} \det(A^{ij_0}).$$

On pourrait développer de la même façon par rapport à une ligne. En reprenant notre premier exemple :

$$\begin{aligned} \begin{vmatrix} 4 & 1 & 5 \\ 6 & 1 & 3 \\ 4 & 1 & 2 \end{vmatrix} &= \begin{vmatrix} 0 & 1 & 5 \\ 2 & 1 & 3 \\ 0 & 1 & 2 \end{vmatrix} \\ &= -(1)^{1+2} \times 2 \times \begin{vmatrix} 1 & 5 \\ 1 & 2 \end{vmatrix} \\ &= -2 \times (2 - 5) = 6. \end{aligned}$$

Par ailleurs si on voulait par exemple calculer

$$\begin{vmatrix} 5 & 1 & 1 \\ 4 & 2 & 3 \\ 7 & 0 & 2 \end{vmatrix},$$

ou pourrait développer par rapport à la seconde colonne et obtenir :

$$\begin{aligned} \begin{vmatrix} 5 & 1 & 1 \\ 4 & 2 & 3 \\ 7 & 0 & 2 \end{vmatrix} &= (-1)^{1+2} \begin{vmatrix} 4 & 3 \\ 7 & 2 \end{vmatrix} + (-1)^{2+2} \times 2 \begin{vmatrix} 5 & 1 \\ 7 & 2 \end{vmatrix} \\ &= -(8 - 21) + 2(10 - 7) \\ &= 13 + 6 = 19. \end{aligned}$$

## 1.5 Un exemple

On veut calculer le so-called *déterminant de Zorro* :

$$\mathcal{Z} := \begin{vmatrix} 0 & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{1} & 0 & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 \end{vmatrix}.$$

On remarque que l'on peut utiliser la première colonne pour anéantir des 1 dans la troisième, et la quatrième colonne pour enlever des 1 dans la deuxième et ainsi obtenir

$$\mathcal{Z} = \begin{vmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{vmatrix}.$$

On peut alors développer par rapport à la deuxième colonne :

$$\mathcal{Z} = (-1)^{4+2} \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

On peut alors développer par rapport à la deuxième colonne :

$$\mathcal{Z} = (-1)^{2+1} \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = -1.$$

## 2 Diagonalisation des matrices carrées

### 2.1 C'est quoi la diagonalisation ?

Les applications linéaires de  $\mathbb{R}^d$  dans  $\mathbb{R}^d$  les plus simples que l'on puisse imaginer sont celles qui sont de la forme

$$(x_1, \dots, x_d) \mapsto (\lambda_1 x_1, \dots, \lambda_d x_d)$$

où  $\lambda_1, \dots, \lambda_d$  sont des réels. Les matrices dans la base canonique de ces applications sont diagonales, c'est à dire de la forme

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_d \end{pmatrix}.$$

Cette forme est très pratique, par exemple parce qu'elle est très facile à inverser :  $D$  est inversible si et seulement si les  $\lambda_i$  sont tous non-nuls, et alors son inverse vaut

$$D^{-1} = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\lambda_d} \end{pmatrix}.$$

Il est aussi très facile de calculer n'importe quelle puissance de ces matrices, et donc n'importe quels polynômes. Si  $p$  est un polynôme,

$$p(D) = \begin{pmatrix} p(\lambda_1) & 0 & \dots & 0 \\ 0 & p(\lambda_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & p(\lambda_d) \end{pmatrix}.$$

Seules très peu de matrices sont de cette forme, mais on peut se poser la question suivante : quelles sont les matrices  $M$  pour lesquelles il existe une matrice  $P$  inversible et une matrice  $D$  diagonale telles que  $M = PDP^{-1}$ . C'est déjà très bien parce qu'une fois effectué le calcul de  $P$  et  $P^{-1}$  (qui peut être délicat), tous les calculs sur  $D$  se transposent facilement sur  $M$  :  $M$  est inversible si et seulement  $D$  est inversible et alors

$$M^{-1} = PD^{-1}P^{-1},$$

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

$$M^k = PD^kP^{-1},$$

si  $p$  est un polynôme,

$$p(M) = Pp(D)P^{-1}.$$

Une telle matrice sera dite diagonalisable.

**Interprétation en termes de changement de base.** Soit  $P$  une matrice inversible. On note  $(e_1, \dots, e_d)$  la base canonique de  $\mathbb{R}^d$ . Les colonnes de  $P$  donnent les coordonnées dans la base canonique d'une famille de vecteurs de  $\mathbb{R}^d$   $(v_1, \dots, v_d)$  qui est une base de  $\mathbb{R}^d$ . Soit  $x$  un vecteur de  $\mathbb{R}^d$  dont les coordonnées dans la base  $(v_1, \dots, v_d)$  sont  $y_1, \dots, y_d$ . On a alors

$$x = y_1v_1 + \dots + y_dv_d = P \cdot y \quad \text{où} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}.$$

Ce calcul signifie exactement que les coordonnées de  $x$  dans la base canonique sont données par le résultat de la multiplication de  $P$  par le vecteur des coordonnées de  $x$  dans la base  $(v_1, \dots, v_d)$ . De même, comme en gardant les mêmes notations,

$$y = P^{-1} \cdot x,$$

les coordonnées de  $x$  dans la base  $(v_1, \dots, v_d)$  sont données par la multiplication de  $P^{-1}$  par le vecteur des coordonnées de  $x$  dans la base canonique.

Supposons donc qu'il existe  $P$  inversible et  $D$  diagonale telle que  $M = PDP^{-1}$ . Que fait alors  $M$ ? Soit  $x \in \mathbb{R}^d$ . Dans ce cas, le vecteur  $P^{-1} \cdot x$  donne les coordonnées de  $x$  dans la base des colonnes de  $P$ . Donc  $DP^{-1} \cdot x$  est le vecteur des coordonnées de  $x$  dans la base des colonnes de  $P$  multipliées par les coefficients diagonaux de  $D$ . Enfin,  $PDP^{-1} \cdot x$  donne les coordonnées de ce dernier vecteur dans la base canonique de  $\mathbb{R}^d$ .

Donc le fait de dire qu'il existe  $P$  inversible et  $D$  diagonale telles que  $M = PDP^{-1}$  revient à dire que dans la base définie par les colonnes de  $P$ , l'endomorphisme induit par  $M$  se contente de multiplier chaque coordonnée par un scalaire.

Formalisons un peu tout ça.

**Définition 2.1.** Si  $M$  est une matrice carrée de taille  $d$  et  $\lambda$  est un réel, on dit que  $\lambda$  est une valeur propre de  $M$  s'il existe un vecteur  $x$  **non-nul** tel que

$$M \cdot x = \lambda x.$$

Dans ce cas, on dira de tout  $x$  non-nul vérifiant cette propriété que c'est un vecteur propre de  $M$  associé à la valeur propre  $\lambda$ .

On a alors une caractérisation simple du fait qu'une matrice soit diagonalisable.

**Proposition 2.2.** Une matrice  $M$  est diagonalisable si et seulement si il existe une base  $(v_1, \dots, v_d)$  de  $\mathbb{R}^d$  constituée de vecteurs propres de  $M$ .

Dans ce cas, en notant  $P$  la matrice dont les colonnes sont les coordonnées des vecteurs  $v_1, \dots, v_d$  dans la base canonique, et pour chaque  $i = 1, \dots, d$ ,  $\lambda_i$  la valeur propre associée à  $v_i$ , on a

$$M = PDP^{-1} \quad \text{avec} \quad D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_d \end{pmatrix}.$$

*Démonstration.* On garde les notations de l'énoncé. Pour montrer que  $M = PDP^{-1}$ , il suffit de vérifier que la multiplication de ces matrices par les vecteurs d'une base donnent le même résultat. Sans surprise, on va tester la base  $(v_1, \dots, v_d)$ . Soit  $i \in \{1, \dots, d\}$ . Par définition de  $\lambda_i$ ,

$$M \cdot v_i = \lambda_i v_i.$$

Par ailleurs,  $P \cdot e_i$  est la  $i$ -ième colonne de  $P$ , c'est à dire  $v_i$ . En multipliant par  $P^{-1}$ , on obtient

$$P^{-1}P \cdot e_i = e_i = P^{-1}v_i.$$

Donc

$$PDP^{-1} \cdot v_i = PD \cdot e_i = P \cdot \lambda_i e_i = \lambda_i P \cdot e_i = \lambda_i v_i.$$

On a donc le résultat. □

## 2.2 Polynôme caractéristique et valeurs propres

Soit  $M$  une matrice carrée de taille  $d$ . On note  $I_d$  la matrice identité de taille  $d$ . Supposons que  $M$  admette le réel  $\lambda$  pour valeur propre. C'est vrai si et seulement si il existe donc  $x$  un vecteur non-nul tel que

$$M \cdot x = \lambda x.$$

En reformulant,  $\lambda$  est valeur propre de  $M$  si et seulement si il existe  $x$  non nul tel que

$$(\lambda I_d - M) \cdot x = 0.$$

Donc  $\lambda$  est valeur propre de  $M$  si et seulement si la matrice  $\lambda I_d - M$  n'est pas inversible i.e. si et seulement si  $\det(\lambda I_d - M) = 0$ .

On définit alors l'application qui à  $\alpha$  réel associe

$$p_M(\alpha) := \det(\alpha I_d - M).$$

C'est une fonction qui s'annule exactement aux valeurs propres de  $M$ . De plus, la formule du déterminant nous assure que c'est un **polynôme**, de degré  $d$  de coefficient dominant 1. Il existe donc  $c_0, \dots, c_{d-1}$  des réels tels que pour tout  $\alpha$ ,

$$p_M(\alpha) = \alpha^d + c_{d-1}\alpha^{d-1} + \dots + c_1\alpha + c_0.$$

*Remarque 2.3.* Il est bon de savoir que  $c_{d-1} = -\text{tr}(M)$  et  $c_0 = (-1)^d \det(M)$ .

L'avantage de ce polynôme, c'est qu'il est invariant par changement de base : si  $P$  est une matrice inversible et  $\alpha$  est réel,

$$\det(\alpha I_d - P^{-1}MP) = \det(P^{-1}[\alpha I_d - M]P) = \det(\alpha I_d - M).$$

Donc le polynôme caractéristique d'une matrice diagonalisable est égal au polynôme caractéristique de la matrice diagonale associée.

Or on sait très facilement calculer le polynôme caractéristique d'une matrice diagonale :

$$\alpha I_d - \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_d \end{pmatrix} = \begin{pmatrix} \alpha - \lambda_1 & 0 & \dots & 0 \\ 0 & \alpha - \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \alpha - \lambda_d \end{pmatrix},$$

donc

$$\det(\alpha I_d - D) = \begin{vmatrix} \alpha - \lambda_1 & 0 & \dots & 0 \\ 0 & \alpha - \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \alpha - \lambda_d \end{vmatrix} = (\alpha - \lambda_1) \dots (\alpha - \lambda_d).$$

Le polynôme que l'on obtient est scindé dans  $\mathbb{R}$ , c'est à dire qu'il s'écrit comme un produit de polynôme de degré 1 réels. En particulier, **toute matrice diagonalisable a un polynôme caractéristique scindé dans  $\mathbb{R}$ .**

*Remarque 2.4.* Par exemple, le polynôme  $X^2 + 1$  n'est pas scindé dans  $\mathbb{R}$ . Ça ne peut pas être le polynôme caractéristique d'une matrice diagonalisable.

## 2.3 Multiplicité

Continuons nos observations. Une valeur propre  $\lambda$  d'une matrice  $M$  peut être multiple. En fait, on distingue deux types de multiplicité. D'abord,  $\lambda$  peut être multiple si  $M$  admet plusieurs vecteurs propres linéairement indépendants associés à  $\lambda$ . On parle alors de **multiplicité géométrique**, et on formalise ce concept comme suis.

**Définition 2.5.** Soit  $M$  une matrice carrée de taille  $d$  et  $\lambda$  une valeur propre de  $M$ . Le sous-espace propre de  $M$  associé à  $\lambda$  est l'ensemble des vecteurs  $x \in \mathbb{R}^d$  tels que

$$M \cdot x = \lambda x.$$

C'est un sous-espace vectoriel de  $\mathbb{R}^d$ , et sa dimension s'appelle la multiplicité géométrique de  $\lambda$ .

D'autre part, elle peut être multiple si le facteur  $(\alpha - \lambda)$  apparaît à une certaine puissance dans le polynôme caractéristique. On parle alors de **multiplicité algébrique**.

**Définition 2.6.** Soit  $M$  une matrice carrée de taille  $d$  et  $\lambda$  une valeur propre de  $M$ . La multiplicité algébrique de  $\lambda$  est l'ordre de la racine  $\lambda$  dans le polynôme caractéristique, c'est à dire la puissance du facteur  $(\alpha - \lambda)$  du polynôme caractéristique.

*Remarque 2.7.* En particulier, comme le polynôme caractéristique est de degré  $d$ , la somme des multiplicités algébriques des valeurs propres d'une matrice est toujours inférieure à  $d$ . Si le polynôme caractéristique est scindé, cette somme est égale à  $d$ .

On a alors les résultats suivants.

**Proposition 2.8.** *La multiplicité algébrique d'une valeur propre est toujours supérieure à sa multiplicité géométrique.*

*Démonstration.* Soit  $M$  une matrice carrée de taille  $d$  et  $\lambda$  une valeur propre de  $M$ . On choisit  $(v_1, \dots, v_k)$  une base du sous-espace propre associé à  $\lambda$ , que l'on complète en une base  $(v_1, \dots, v_d)$ . On note  $P$  la matrice ayant pour colonne  $(v_1, \dots, v_d)$ . On a alors

$$P^{-1}MP = \left( \begin{array}{cc|c} \lambda & 0 & N \\ & \ddots & \\ 0 & \lambda & \\ \hline & 0 & Q \end{array} \right)$$

où  $N$  est une matrice  $k \times (d - k)$  et  $Q$  est une matrice  $(d - k) \times (d - k)$ . Le calcul du déterminant de  $\alpha I_d - P^{-1}MP$  donne alors en développant par rapport aux  $k$  premières colonnes  $(\alpha - \lambda)^k \times \det(\alpha I_{d-k} - Q)$ . Donc la multiplicité algébrique de  $\lambda$  est au moins  $k$ .  $\square$

**Proposition 2.9.** *Les sous-espaces propres d'une matrice carrée  $M$  sont en somme directe.*

*Démonstration.* Soient  $\lambda_1, \dots, \lambda_p$  les valeurs propres de  $M$  (comptées sans multiplicité!) et  $E_1, \dots, E_p$  les sous-espaces propres associés. On veut montrer que si  $u_1 \in E_1, \dots, u_p \in E_p$  sont tels que

$$u_1 + \dots + u_p = 0,$$

alors  $u_1 = \dots = u_p = 0$ . Soient donc de tels  $u_1, \dots, u_p$ . En appliquant  $M$ , on observe

$$\lambda_1 u_1 + \dots + \lambda_p u_p = M \cdot 0 = 0.$$

En reformulant légèrement :

$$\begin{aligned} & \lambda_1(u_1 + \dots + u_p) + (\lambda_2 - \lambda_1)u_2 + \dots + (\lambda_p - \lambda_1)u_p \\ &= (\lambda_2 - \lambda_1)u_2 + \dots + (\lambda_p - \lambda_1)u_p \\ &= 0. \end{aligned}$$

En appliquant  $M$ , on a réussi à prendre une relation mettant en jeu  $p$  vecteurs et à fabriquer une relation mettant en jeu  $p - 1$  vecteurs avec des coefficients non nuls. Cet argument peut s'itérer pour obtenir des relations mettant en jeu  $p - 2$  puis  $p - 3$  etc... jusqu'à obtenir

$$c \times u_p = 0$$

avec  $c$  non nul. On a alors  $u_p = 0$  et en rembobinant la liste de relations, on obtient bien  $u_1 = \dots = u_p = 0$ .  $\square$

On en déduit le théorème suivant.

**Théorème 2.10.** *Soit  $M$  une matrice carrée de taille  $d$ . Les propositions suivantes sont équivalentes :*

1.  $M$  est diagonalisable,
2. la somme des multiplicités géométriques de  $M$  est égale à  $d$ ,
3. le polynôme caractéristique est scindé chaque valeur propre a une multiplicité algébrique égale à sa multiplicité géométrique.

*Démonstration.* Si  $M$  est diagonalisable, alors il existe une base de vecteurs propres. En particulier, la somme des dimensions des sous-espaces propres est égale à  $d$ .

Si la somme des multiplicité géométrique est égale à  $d$ . On note  $n_1, \dots, n_p$  les multiplicité géométriques et  $m_1, \dots, m_p$  les multiplicités algébriques. On a  $n_i \leq m_i$  pour tout  $i = 1, \dots, p$ , et

$$d = n_1 + \dots + n_p \leq m_1 + \dots + m_p \leq d.$$

Donc pour tout  $i$ ,  $n_i = m_i$  et

$$m_1 + \dots + m_p = d.$$

En particulier le polynôme caractéristique est scindé.

Si le polynôme caractéristique est scindé et si chaque valeur propre a une multiplicité algébrique égale à sa multiplicité géométrique, alors la somme des multiplicité géométriques vaut  $d$ , et donc  $M$  est diagonalisable.  $\square$

Ce qu'il faut retenir ce ce théorème est très simple. Par exemple, imaginons que l'on ait une matrice carrée  $M$  de taille 3 qui admette pour polynôme caractéristique  $(X - 1)(X - 3)^2$ . Dans ce cas, on est sûrs que  $M$  admet 1 et 3 pour valeurs propres. En vertu de la proposition 2.8, la dimension du sous-espace propre associé à 1 est forcément 1, et la dimension du sous-espace propre associé à 3 est 1 ou 2. Dans ce cas,  $M$  est diagonalisable si et seulement si la dimension du sous-espace propre associé à 3 est 2. Dans les faits, il suffit d'exhiber deux vecteurs propres indépendants associés à la valeur propre 3 pour montrer que  $M$  est diagonalisable.

**Multiplicité géométrique et rangs de matrices.** Grâce au théorème précédent, on a un critère pour savoir si une matrice est diagonalisable ou non, et une fois le polynôme caractéristique calculé, il ne reste qu'à calculer la multiplicité géométrique de chacune des valeurs propres. Mais on peut faire les remarques suivantes :

- si  $M$  est une matrice carré et  $\lambda$  est une valeur propre de  $M$ , alors  $x$  est dans le sous-espace propre associé à  $\lambda$  si et seulement si  $x$  est dans le noyau de  $\lambda I_d - M$  : donc le sous-espace propre associé à  $\lambda$  est exactement le noyau de  $\lambda I_d - M$ ,
- en particulier, la multiplicité géométrique de  $\lambda$  est égale à la dimension du noyau de  $\lambda I_d - M$ ,
- la dimension du noyau de  $\lambda I_d - M$  peut se calculer grâce au théorème du rang : c'est  $d - \text{rg}(\lambda I_d - M)$ .

Donc en pratique, pour déterminer la multiplicité géométrique de  $\lambda$ , on calcule la matrice  $\lambda I_d - M$ , on détermine son rang, et on applique le théorème du rang.

## 2.4 Comment trouve-t-on les vecteurs propres ?

Là, c'est le moment le moins amusant : il faut résoudre un système. Pour montrer la méthode, on traite un exemple en détail. On cherche à diagonaliser la matrice suivante :

$$M := \begin{pmatrix} 3 & 0 & -1 \\ 2 & 4 & 2 \\ -1 & 0 & 3 \end{pmatrix}.$$

On commence par calculer le polynôme caractéristique.

$$p_M(X) = \begin{vmatrix} X - 3 & 0 & -1 \\ 2 & X - 4 & 2 \\ -1 & 0 & X - 3 \end{vmatrix}.$$

En développant par rapport à la deuxième colonne,

$$\begin{aligned} p_M(X) &= (X-4) \begin{vmatrix} X-3 & -1 \\ -1 & X-3 \end{vmatrix} \\ &= (X-4)[(X-3)^2 - 1] \\ &= (X-2)(X-4)^2. \end{aligned}$$

Les valeurs propres de  $M$  sont donc 2 de multiplicité algébrique (et géométrique) 1 et 4 de multiplicité algébrique 2. Pour montrer que  $M$  est diagonalisable, il suffit de montrer que la multiplicité géométrique de 4 est 2. On calcule  $4I_d - M$  :

$$4I_d - M = \begin{pmatrix} 1 & 0 & 1 \\ -2 & 0 & -2 \\ 1 & 0 & 1 \end{pmatrix}.$$

Cette matrice est bien de rang 1, donc la multiplicité géométrique de 4 est bien  $3 - 1 = 2$ . La matrice  $M$  est donc diagonalisable, et on peut chercher ses vecteurs propres.

Cherchons un vecteur propre associé à la valeur propre 2. Soit  $x, y$  et  $z$  trois réels tels que

$$M \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2 \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

$(x, y, z)$  est alors solution du système suivant

$$\begin{aligned} 3x - z &= 2x, \\ 2x + 4y + 2z &= 2y, \\ 6x + 3z &= 2z, \end{aligned}$$

ou en manipulant un peu

$$\begin{aligned} x - z &= 0, \\ 2x + 2y + 2z &= 0, \\ 6x + z &= 0. \end{aligned}$$

Ce système se réduit clairement à

$$\begin{aligned} x &= z, \\ y &= -2x. \end{aligned}$$

On peut donc choisir  $v_1 := (1, -2, 1)$  et on vérifie  $M \cdot v_1 = 2v_1$ .

Cherchons maintenant les vecteurs propres associés à la valeur propre 4. Déjà, on observe que le vecteur de la base canonique  $v_2 := e_2 = (0, 1, 0)$  est vecteur propre de  $M$  associé à la valeur propre 4. Ensuite, on remarque que la première et la dernière colonne de  $4I_d - M$  sont égales. Cela veut exactement dire que tous les vecteurs du type  $(a, 0, -a)$  sont dans le noyau de cette matrice, et donc sont des vecteurs propres de  $M$  associés à la valeur propre 4.

En posant

$$P := \begin{pmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix},$$

on a donc

$$M = P \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix} P^{-1}$$

Ici, on n'a pas calculé  $P^{-1}$ . On aurait pu le faire en utilisant la formule de l'inverse (voir TD). Cependant, dans la plupart des problèmes, la connaissance des vecteurs propres est suffisante et le calcul de  $P^{-1}$  n'est pas utile.

### 3 Matrices symétriques et orthogonales

On confond ici  $x$  et le vecteur colonne constitué des coordonnées de  $x$ .

#### 3.1 Définitions, liens avec le produit scalaire

On définit ici deux types de matrices jouant un rôle particulier vis à vis du produit scalaire. D'abord, les matrices symétriques que l'on a déjà rencontrées.

**Définition 3.1.** Une matrice carrée  $M$  est symétrique si  ${}^tM = M$ . On note alors  $M \in \mathcal{S}_d(\mathbb{R})$ .

**Proposition 3.2.** La matrice carrée  $M$  est symétrique si et seulement si pour tous  $x$  et  $y$  dans  $\mathbb{R}^d$ ,

$$\langle M \cdot x, y \rangle = \langle x, M \cdot y \rangle.$$

*Remarque 3.3.* La deuxième affirmation fait intervenir directement l'endomorphisme engendré par  $M$ . En fait, on dit que l'endomorphisme  $f$  est symétrique si pour tous  $x$  et  $y$  dans  $\mathbb{R}^d$ ,

$$\langle f(x), y \rangle = \langle x, f(y) \rangle.$$

La proposition montre alors qu'un endomorphisme est symétrique si et seulement si sa matrice est symétrique.

*Démonstration.* Si  $M$  symétrique,

$$\begin{aligned} \langle M \cdot x, y \rangle &= {}^t(M \cdot x) \times y \\ &= {}^tx \times {}^tM \times y \\ &= {}^tx \times M \times y \\ &= {}^tx \times (M \cdot y) \\ &= \langle x, M \cdot y \rangle. \end{aligned}$$

Si pour tous  $x$  et  $y$ ,

$$\langle M \cdot x, y \rangle = \langle x, M \cdot y \rangle,$$

alors en choisissant  $i$  et  $j$  dans  $\{1, \dots, d\}$ ,  $x = e_i$  et  $y = e_j$ , la formule donne

$$\langle M \cdot e_i, e_j \rangle = M_{i,j} = M_{j,i} = \langle e_i, M \cdot e_j \rangle.$$

□

Voyons ensuite les matrices orthogonales.

**Définition 3.4.** Une matrice carrée  $P$  de taille  $d$  est dite orthogonale si

$${}^tP \times P = P \times {}^tP = I_d.$$

On note alors  $P \in \mathcal{O}_d(\mathbb{R})$ .

Une matrice est donc orthogonale si elle est inversible et si son inverse est égal à sa transposée.

**Proposition 3.5.** Soit  $P$  une matrice carrée de taille  $d$ . Les affirmations suivantes sont équivalentes :

1.  $P$  est orthogonale,
2. les colonnes de  $P$  forment une base orthonormale de  $\mathbb{R}^d$ ,
3. les lignes de  $P$  forment une base orthonormale de  $\mathbb{R}^d$ ,

4. pour tous  $x$  et  $y$  dans  $\mathbb{R}^d$ ,

$$\langle P \cdot x, P \cdot y \rangle = \langle x, y \rangle.$$

*Remarque 3.6.* La quatrième affirmation fait intervenir directement l'endomorphisme induit par  $P$ . En fait, on dit que l'endomorphisme  $f$  est orthogonal si pour tous  $x$  et  $y$  dans  $\mathbb{R}^d$ ,

$$\langle f(x), f(y) \rangle = \langle x, y \rangle.$$

La proposition montre alors que  $f$  est orthogonal si et seulement si sa matrice est orthogonale.

*Démonstration.* Montrons d'abord que 1  $\Leftrightarrow$  2. En notant  $C_1, \dots, C_d$  les colonnes de  $P$ , on observe que pour tous  $i$  et  $j$ , le coefficient  $(i, j)$  de  ${}^tP \times P$  n'est autre que  $\langle C_i, C_j \rangle$  (cette expression ne satisfierait pas trop les puristes, mais on a déjà dit que l'on confondait les vecteurs colonnes et les vecteurs de  $\mathbb{R}^d$ ). Donc on a la chaîne d'équivalences suivante :

$$\begin{aligned} P \text{ orthogonale} &\Leftrightarrow \forall (i, j) \in \{1, \dots, d\}, \quad \langle C_i, C_j \rangle = \delta_{i,j} \\ &\Leftrightarrow (C_1, \dots, C_d) \text{ forme une base orthonormale de } \mathbb{R}^d. \end{aligned}$$

On montre de la même façon que 1 est équivalent à 3 en considérant  $P \times {}^tP$  au lieu de  ${}^tP \times P$ .

Il reste donc seulement à montrer que 1 est équivalent à 4. C'est une conséquence facile du fait que pour toute matrice  $P$  et tous vecteurs  $x$  et  $y$ ,

$$\langle P \cdot x, P \cdot y \rangle = {}^t x \times ({}^tP \times P) \times y.$$

□

Les matrices orthogonales induisent donc des endomorphismes qui préservent les distances. Elles envoient des bases orthonormales sur des bases orthonormales. En dimension 2, il en existe deux types : les rotations, qui s'écrivent

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix},$$

et les symétries, qui **quitte à changer de base** s'écrivent

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Voir le TD pour plus de détails et le cas de la dimension 3.

### 3.2 Théorème spectral

On donne ici le théorème spectral qui permet de caractériser complètement les matrices symétriques.

**Théorème 3.7.** *Soit  $M$  une matrice symétrique de taille  $d$ . Il existe une matrice  $P$  orthogonale et une matrice  $D$  diagonale telles que*

$$M = PD^tP.$$

*De façon équivalente, si  $f$  est un endomorphisme symétrique de  $\mathbb{R}^d$ , il existe une base orthonormée de  $\mathbb{R}^d$  constituée de vecteurs propres de  $f$  (on dit que  $f$  est diagonalisable en base orthonormée).*

*Remarque 3.8.* La réciproque est évidemment vraie puisque si  $M = PD^tP$ , alors  ${}^tM = {}^t(PD^tP) = PD^tP = M$ .

Un endomorphisme symétrique consiste donc **dans une certaine base orthonormée** à multiplier par des constantes chacune des coordonnées.

Diagonaliser dans une base orthonormée est très facile puisqu'on calcule  $P^{-1}$  simplement en transposant.

*Preuve partielle.* Dans cette pseudo-preuve, on raisonne en termes d'endomorphisme, on laisse le lecteur se convaincre du résultat en termes de matrices.

On admet le résultat suivant : un endomorphisme d'un espace euclidien admet toujours une valeur propre réelle.

Une fois que l'on sait ça, la démarche est la suivante. On sait que  $f$  admet un vecteur propre  $u$ , que l'on peut supposer de norme 1, associé à la valeur propre  $\lambda$ . On note alors  $E$  l'orthogonal de la droite vectorielle engendrée par  $u$ . On montre alors la chose suivante :

$$\forall x \in E, \quad f(x) \in E.$$

On dit alors que  $E$  est stable par  $f$ . Soit en effet  $x \in E$ . Le but est de montrer que  $f(x)$  est orthogonal à  $u$ . On calcule donc

$$\begin{aligned} \langle f(x), u \rangle &= \langle x, f(u) \rangle \\ &= \langle x, \lambda u \rangle \\ &= \lambda \langle x, u \rangle = 0. \end{aligned}$$

On définit alors  $g$  l'application de  $E$  dans  $E$  qui à  $x \in E$  associe  $f(x)$ . Il est clair que  $g$  est encore symétrique, donc d'après ce que l'on a admis,  $g$  possède un vecteur propre, et comme il est dans  $E$ , il est orthogonal à  $u$ .

On peut alors réitérer cet argument et trouver  $d$  vecteurs propres de norme 1 deux à deux orthogonaux. Ces vecteurs forment donc une famille libre (proposition 2.5 du chapitre 2), et donc une base orthonormée. Le résultat est ainsi (partiellement) démontré.  $\square$

### 3.3 Matrices signées

Voyons maintenant comment définir un signe pour une matrice symétrique. On pourrait définir de la même façon le signe d'un endomorphisme.

**Définition 3.9.** Soit  $M$  une matrice symétrique de taille  $d$ . On dit que

- $M$  est positive ou  $M \in \mathcal{S}_d^+(\mathbb{R})$  si pour tout  $x \in \mathbb{R}^d$ ,

$$\langle M \cdot x, x \rangle \geq 0,$$

- $M$  est définie positive ou  $M \in \mathcal{S}_d^{++}(\mathbb{R})$  si pour tout  $x \in \mathbb{R}^d$  **non nul**,

$$\langle M \cdot x, x \rangle > 0,$$

- $M$  est négative ou  $M \in \mathcal{S}_d^-(\mathbb{R})$  si pour tout  $x \in \mathbb{R}^d$ ,

$$\langle M \cdot x, x \rangle \leq 0,$$

- $M$  est définie négative ou  $M \in \mathcal{S}_d^{--}(\mathbb{R})$  si pour tout  $x \in \mathbb{R}^d$  **non nul**,

$$\langle M \cdot x, x \rangle < 0.$$

*Remarque 3.10.* Il existe des matrices qui n'ont pas de signe. Il est tout à fait possible que pour certains  $x$ ,  $\langle M \cdot X, X \rangle$  soit strictement positif, et pour d'autres, il soit strictement négatif.

Une fois que l'on sait que les matrices carrés sont diagonalisables, on peut réinterpréter le signe d'une matrice en fonction de ces valeurs propres.

**Proposition 3.11.** Soit  $M$  une matrice symétrique de taille  $d$ .

- La matrice  $M$  est positive si et seulement si toutes ses valeurs propres sont positives.

- La matrice  $M$  est définie positive si et seulement si toutes ses valeurs propres sont strictement positives.
- La matrice  $M$  est négative si et seulement si toutes ses valeurs propres sont négatives.
- La matrice  $M$  est définie négative si et seulement si toutes ses valeurs propres sont strictement négatives.

*Démonstration.* Soit  $\mathcal{F} := (f_1, \dots, f_d)$  une base orthonormée de  $\mathbb{R}^d$  constituée de vecteurs propres de  $M$ . On note  $\lambda_i$  la valeur propre associée au vecteur propre  $f_i$ . On rappelle que pour tout  $x \in \mathbb{R}^d$ , la décomposition de  $x$  dans la base  $\mathcal{F}$  est donnée par

$$x = \sum_{i=1}^d \langle f_i, x \rangle f_i.$$

Donc

$$M \cdot x = \sum_{i=1}^d \lambda_i \langle f_i, x \rangle f_i.$$

En faisant le produit scalaire avec  $x$ , on obtient alors

$$\langle M \cdot x, x \rangle = \sum_{i=1}^d \lambda_i \langle f_i, x \rangle^2.$$

On en déduit aisément le résultat. □



# Chapitre 5

## Différentielle des fonctions de plusieurs variables

### 1 Dérivées partielles des fonctions de plusieurs variables

Dans tout ce paragraphe,  $U$  désigne un ouvert de  $\mathbb{R}^d$  ( $d \geq 1$ ), et  $f$  est une fonction de  $U$  dans  $\mathbb{R}$ .

#### 1.1 Définition et interprétation

Soit  $x = (x_1, \dots, x_d)$  un point de  $U$  et  $i \in \{1, \dots, d\}$ . On peut regarder les variations de  $f$  autour de  $x$  quand on fait varier la  $i$ -ième coordonnée, c'est à dire l'application

$$t \mapsto f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_d) \quad (5.1)$$

autour du point  $y = x_i$ . C'est une application d'une partie de  $\mathbb{R}$  dans  $\mathbb{R}$ , et comme  $U$  est un ouvert contenant  $x$ , il y a un intervalle autour de  $x_i$  où cette application est bien définie. On peut donc étudier sa dérivabilité en  $x_i$ .

**Définition 1.1.** On dit que  $f$  admet en  $x$  une dérivée partielle selon la coordonnée  $i$  (ou dans la direction  $i$ ) si l'application (5.1) est dérivable en  $x_i$ . On note alors

$$\frac{\partial f}{\partial x_i}(x)$$

sa dérivée en  $x_i$ .

*Exemple 1.2.* Supposons que la production d'une entreprise ne dépende que de la quantité de travail effectué et du capital introduit. C'est à dire que pour une certaine quantité de travail  $L$  et un certain capital  $K$ , on peut associer la quantité de biens produits

$$Q(L, K).$$

Supposons que l'entreprise soit actuellement dans un régime où elle utilise la quantité de travail  $L^*$  et le capital  $K^*$ , et qu'elle se demande s'il serait profitable d'augmenter la quantité de travail "toutes choses égales par ailleurs", c'est à dire en conservant le même capital. Elle veut donc regarder si pour un petit  $\Delta L$ , la quantité produite

$$Q(L^* + \Delta L, K^*)$$

ne serait pas par hasard plus grande que  $Q(L^*, K^*)$ . Mais si  $Q$  admet en  $(L^*, K^*)$  une dérivée partielle selon la coordonnée  $L$ , on peut faire l'approximation

$$Q(L^* + \Delta L, K^*) \approx Q(L^*, K^*) + \frac{\partial Q}{\partial L}(L^*, K^*)\Delta L.$$

Donc ça consiste juste à regarder le signe de la dérivée partielle

$$\frac{\partial Q}{\partial L}(L^*, K^*).$$

En fait,

$$\frac{\partial Q}{\partial L}(L^*, K^*)$$

représente **la quantité supplémentaire produite par unité de travail ajouté, à capital constant**. Et on espère pas que c'est positif, parce qu'avec El-Kohmri, on se retrouve avec un plan de compétitivité, une augmentation du temps de travail sans augmentation de salaire, tout ça tout ça...

## 1.2 Condition nécessaire d'optimalité

Avant toute chose voyons ce que signifie un extremum local en dimension  $d$ . Exceptionnellement pour cette définition, on note  $D$  le domaine de définition de  $f$ , parce que ce n'est plus nécessairement un ouvert.

**Définition 1.3.** On dit que le point  $x \in D$  est un maximum local de  $f$  s'il existe  $\varepsilon > 0$  tel que pour tout  $y \in D \cap B(x, \varepsilon)$ ,

$$f(y) \leq f(x).$$

Je vous laisse deviner ce que signifie minimum local, maximum local strict et minimum local strict.

Simplement avec la petite notion de dérivabilité que l'on a vu, on peut déduire une condition nécessaire d'optimalité. Si un extremum local est atteint, il l'est selon toutes les coordonnées, sinon, il suffirait de bouger selon cette coordonnée là!

**Proposition 1.4.** Soit  $x = (x_1, \dots, x_d) \in U$  un point d'extremum local de  $f$  et  $i \in \{1, \dots, d\}$ . Si  $f$  admet en  $x$  une dérivée partielle selon la coordonnée  $i$ , alors

$$\frac{\partial f}{\partial x_i}(x) = 0.$$

*Démonstration.* On note  $g$  l'application qui à  $t$  associe

$$f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_d).$$

la fonction  $g$  est bien définie autour de  $x_i$  et admet en  $x_i$  un extremum local. Donc

$$g'(x_i) = \frac{\partial f}{\partial x_i}(x) = 0.$$

□

## 1.3 Dérivées partielles d'ordre supérieure

Supposons que  $f$  admette des dérivées partielles dans toutes les directions en tout point de  $U$ . On dispose alors d'une famille de  $d$  applications

$$\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_d}$$

qui sont elles-mêmes des fonction de  $U$  à valeurs dans  $\mathbb{R}$ . On peut donc étudier leur régularité.

**Définition 1.5.** On dit que  $f$  est de classe  $C^1$  si elle admet en tout point de  $U$  des dérivées partielles dans toutes les directions et si ces dérivées partielles sont des fonctions **continues** de  $U$  dans  $\mathbb{R}$ .

On peut également regarder si ces applications admettent des dérivées partielles. Si pour  $i$  et  $j$  dans  $\{1, \dots, d\}$ ,

$$\frac{\partial f}{\partial x_i}$$

admet en  $x$  une dérivée partielle selon la coordonnée  $j$ , on la note

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

et on dit que  $f$  admet en  $x$  une dérivée partielle d'ordre 2 selon les coordonnées  $i, j$ .

**Définition 1.6.** On dit que  $f$  est de classe  $C^2$  si elle est de classe  $C^1$ , si elle admet en chaque point de  $U$  des dérivées partielles d'ordre 2 selon toutes les coordonnées, et si ces dérivées partielles d'ordre 2 sont continues.

C'est pareil que de dire que  $f$  est de classe  $C^1$  et que les dérivées partielles (d'ordre 1) de  $f$  sont de classe  $C^1$ .

Évidemment, on peut continuer avec les dérivées d'ordre 3, 4, etc... On a un théorème très très important mais pas évident du tous, c'est le suivant.

**Théorème 1.7.** Si  $f$  est de classe  $C^2$ , si  $x \in U$  et si  $i$  et  $j$  sont dans  $\{1, \dots, d\}$ , alors

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

*Démonstration. Réduction du problème.* D'abord, si  $i = j$ , il n'y a rien à montrer. Donc on peut supposer  $i \neq j$ . Soit

$$V := \{(s, t) \in \mathbb{R}^2 \text{ tel que } x + se_i + te_j \in U\}.$$

L'ensemble  $V$  est un ouvert de  $\mathbb{R}^2$  (image réciproque de  $U$  par l'application continue  $(s, t) \mapsto x + se_i + te_j$ ) contenant 0 (car  $x \in U$ ). Sur l'ouvert  $V$ , on définit  $g$  par

$$\forall (s, t) \in V, \quad g(s, t) := f(x + se_i + te_j).$$

La fonction  $g$  est alors  $C^2$  et on a les égalités

$$\begin{aligned} \frac{\partial^2 g}{\partial t \partial s}(0) &= \frac{\partial^2 f}{\partial x_j \partial x_i}(0), \\ \frac{\partial^2 g}{\partial s \partial t}(0) &= \frac{\partial^2 f}{\partial x_i \partial x_j}(0). \end{aligned}$$

Il s'agit donc de montrer que

$$\frac{\partial^2 g}{\partial s \partial t}(0) = \frac{\partial^2 g}{\partial t \partial s}(0).$$

On s'est donc ramené au cas où  $d = 2$  et  $x = 0$ .

**Idée.** On pose pour  $s$  et  $t$  réels non nuls tels que  $(s, t) \in V$  :

$$\begin{aligned} A(s, t) &:= \frac{g(s, t) - g(0, t) - g(s, 0) + g(0, 0)}{st} \\ &= \frac{g(s, t) - g(s, 0) - g(0, t) + g(0, 0)}{st}. \end{aligned}$$

Alors pour tout  $t$  non nul suffisamment petit,

$$\begin{aligned} \lim_{s \rightarrow 0} A(s, t) &= \frac{1}{t} \left( \lim_{s \rightarrow 0} \frac{g(s, t) - g(0, t)}{s} - \lim_{s \rightarrow 0} \frac{g(s, 0) - g(0, 0)}{s} \right) \\ &= \frac{1}{t} \left( \frac{\partial g}{\partial s}(0, t) - \frac{\partial g}{\partial s}(0, 0) \right). \end{aligned}$$

Et donc

$$\lim_{t \rightarrow 0} \lim_{s \rightarrow 0} A(s, t) = \frac{\partial^2 g}{\partial t \partial s}(0, 0).$$

Et par ailleurs de la même façon, on montre aisément que

$$\lim_{s \rightarrow 0} \lim_{t \rightarrow 0} A(s, t) = \frac{\partial^2 g}{\partial s \partial t}(0, 0).$$

Donc sous réserve que l'on puisse intervertir les limites, on a le résultat.

**Justification de l'interversion.** On utilise le lemme suivant :

*Lemme 1.8.* Soient  $a < b$  et  $\varphi$  une fonction de classe  $C^1$  de  $[a, b]$  dans  $\mathbb{R}$ . Alors

$$\frac{\varphi(b) - \varphi(a)}{b - a} = \int_0^1 \varphi'((1-t)a + tb) dt.$$

*Preuve du lemme.* On définit pour tout  $t \in [0, 1]$

$$\psi(t) := \varphi((1-t)a + tb).$$

La fonction  $\psi$  est dérivable en tout  $t$  et sa dérivée vérifie pour tout  $t$

$$\psi'(t) = (b-a)\varphi'((1-t)a + tb).$$

Ainsi,  $\psi$  est de classe  $C^1$  et

$$\psi(1) - \psi(0) = \int_0^1 \psi'(t) dt,$$

qui se réécrit

$$\varphi(b) - \varphi(a) = (b-a) \int_0^1 \varphi'((1-t)a + tb) dt,$$

et il n'y a plus qu'à diviser par  $b-a$ . □

Soit  $(s, t) \in V$  est suffisamment proche de  $(0, 0)$  pour que pour tout  $\sigma \in [0, 1]$  et  $\tau \in [0, 1]$ ,  $(\sigma s, \tau t) \in V$ . C'est possible car  $V$  est ouvert. Grâce au lemme, on peut alors réécrire (en utilisant la notation  $\partial_s g$  plutôt que  $\frac{\partial g}{\partial s}$ ) :

$$\begin{aligned} A(s, t) &= \frac{1}{t} \left( \left\{ \frac{g(s, t) - g(0, t)}{s} \right\} - \left\{ \frac{g(s, 0) - g(0, 0)}{s} \right\} \right) \\ &= \frac{1}{t} \left( \int_0^1 \partial_s g(\sigma s, t) d\sigma - \int_0^1 \partial_s g(\sigma s, 0) d\sigma \right) \\ &= \int_0^1 \frac{\partial_s g(\sigma s, t) - \partial_s g(\sigma s, 0)}{t} d\sigma \\ &= \int_0^1 \int_0^1 \partial_t \partial_s g(\sigma s, \tau t) d\tau d\sigma. \end{aligned}$$

Et puis de la même façon, en faisant la décomposition dans l'autre sens,

$$A(s, t) = \int_0^1 \int_0^1 \partial_s \partial_t g(\sigma s, \tau t) d\sigma d\tau.$$

Soit  $\varepsilon > 0$  et soit  $(s, t)$  suffisamment proche de  $(0, 0)$  pour que pour tout  $\sigma \in [0, 1]$  et  $\tau \in [0, 1]$ ,

- le point  $(\sigma s, \tau t)$  appartient à  $V$  (on utilise que  $V$  est ouvert),
- on ait l'estimation  $|\partial_t \partial_s g(\sigma s, \tau t) - \partial_t \partial_s g(0, 0)| \leq \varepsilon/2$  (on utilise que  $\partial_t \partial_s g$  est continue),

- on ait l'estimation  $|\partial_s \partial_t g(\sigma s, \tau t) - \partial_s \partial_t g(0, 0)| \leq \varepsilon/2$  (on utilise que  $\partial_s \partial_t g$  est continue).

Alors

$$\begin{aligned}
& |\partial_t \partial_s g(0, 0) - \partial_s \partial_t g(0, 0)| \\
& \leq |\partial_t \partial_s g(0, 0) - A(s, t)| + |A(s, t) - \partial_s \partial_t g(0, 0)| \\
& = \left| \partial_t \partial_s g(0, 0) - \int_0^1 \int_0^1 \partial_t \partial_s g(\sigma s, \tau t) \, d\tau \, d\sigma \right| \\
& \quad + \left| \int_0^1 \int_0^1 \partial_s \partial_t g(\sigma s, \tau t) \, d\tau \, d\sigma - \partial_s \partial_t g(0, 0) \right| \\
& = \left| \int_0^1 \int_0^1 (\partial_t \partial_s g(0, 0) - \partial_t \partial_s g(\sigma s, \tau t)) \, d\tau \, d\sigma \right| \\
& \quad + \left| \int_0^1 \int_0^1 (\partial_s \partial_t g(\sigma s, \tau t) - \partial_s \partial_t g(0, 0)) \, d\tau \, d\sigma \right| \\
& \leq \int_0^1 \int_0^1 |\partial_t \partial_s g(0, 0) - \partial_t \partial_s g(\sigma s, \tau t)| \, d\tau \, d\sigma \\
& \quad + \int_0^1 \int_0^1 |\partial_s \partial_t g(\sigma s, \tau t) - \partial_s \partial_t g(0, 0)| \, d\tau \, d\sigma \\
& \leq \varepsilon.
\end{aligned}$$

Comme on peut faire ce raisonnement pour tout  $\varepsilon > 0$ , en faisant tendre  $\varepsilon$  vers 0, on obtient bien

$$\partial_t \partial_s g(0, 0) = \partial_s \partial_t g(0, 0).$$

□

Attention, la preuve de l'interversion fait intervenir de façon cruciale le fait que les dérivées partielles d'ordre 2 sont continues. Si elles ne sont pas continues, le théorème est faux en général, comme en témoigne l'exercice très peu intéressant suivant.

**Exercice 1.9.** Soit  $f$  la fonction de  $\mathbb{R}^2$  dans  $\mathbb{R}$  définie pour tout  $(x, y) \in \mathbb{R}^2$  par

$$f(x, y) = \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{sinon.} \end{cases}$$

Montrer que  $f$  est de classe  $C^1$  et admet en 0 des dérivées partielles d'ordre 2 selon les coordonnées  $x, y$  et  $y, x$ , mais que ces dérivées ne coïncident pas.

## 2 La différentielle des fonctions de plusieurs variables

### 2.1 Définition et lien avec les dérivées partielles

La définition que les mathématiciens donnent des différentielles ne fait pas intervenir "que" les dérivées partielles, et c'est naturel puisqu'il n'y a pas de raison *a priori* que le comportement d'une fonction parallèlement aux axes de coordonnées impose son comportement partout! En fait, les dérivées partielles ne permettent d'approximer les variations de  $f$  **que dans les directions des axes**. Ce truc a des conséquences très bizarres. Par exemple, c'est possible d'avoir une fonction qui a des dérivées partielles dans toutes les directions en tout point et que ne soit même pas continue. C'est par exemple le cas de la fonction définie pour tout  $(x, y) \in \mathbb{R}^2$  par

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{sinon,} \end{cases}$$

qui n'est pas continue en 0.

Une bonne notion de différentiabilité est "l'approximabilité par une fonction linéaire", ce qui correspond à l'intuition d'avoir un "plan tangent".

On reprend une fonction  $f$  d'un ouvert  $U$  de  $\mathbb{R}^d$  à valeurs dans  $\mathbb{R}$ .

**Définition 2.1.** Soit  $x \in U$ . On dit que  $f$  est différentiable en  $x$  s'il existe **une application linéaire**  $l$  de  $\mathbb{R}^d$  dans  $\mathbb{R}$  telle que pour tout  $y \in U$ ,

$$f(y) = f(x) + l(y - x) + o_{y \rightarrow x}(\|y - x\|), \quad (5.2)$$

$$\text{i.e. } \lim_{y \rightarrow x} \frac{f(y) - f(x) - l(y - x)}{\|y - x\|} = 0. \quad (5.3)$$

Cette application linéaire est alors unique et s'appelle la différentielle de  $f$  en  $x$ , notée  $df(x)$ .

*Preuve de l'unicité.* Soient  $l$  satisfaisant (5.3) et  $v \in \mathbb{R}^d$ . Alors en particulier,  $x + tv$  tend vers  $x$  quand  $t$  tend vers 0, donc

$$\lim_{t \rightarrow 0} \frac{f(x + tv) - f(x) - tl(v)}{t\|v\|} = 0,$$

et donc

$$l(v) = \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t}$$

ce qui prescrit la valeur de  $l(v)$ . □

Là, plus question de ne pas être continue!

**Proposition 2.2.** Si  $x \in U$  et si  $f$  est différentiable en  $x$ , alors  $f$  est continue en  $x$ .

*Démonstration.* Les applications linéaires sont continues en dimension finie. En particulier,

$$\lim_{y \rightarrow x} l(y - x) = 0$$

En mettant ça dans (5.2), on obtient bien le résultat. □

Ensuite, si  $f$  est différentiable en  $x$ , alors elle admet en  $x$  des dérivées partielles dans toutes les directions.

**Proposition 2.3.** Si  $f$  est différentiable en  $x \in U$ , alors  $f$  admet en  $x$  des dérivées partielles dans toutes les directions et pour tout  $i \in \{1, \dots, d\}$ ,

$$\frac{\partial f}{\partial x_i}(x) = df(x)(e_i),$$

où  $(e_1, \dots, e_d)$  est la base canonique de  $\mathbb{R}^d$ . On a alors la formule suivante pour tout  $v = (v_1, \dots, v_d) \in \mathbb{R}^d$  :

$$df(x)(v) = v_1 \frac{\partial f}{\partial x_1}(x) + \dots + v_d \frac{\partial f}{\partial x_d}(x) \quad (5.4)$$

*Démonstration.* Soit  $i \in \{1, \dots, d\}$ . Comme  $f$  est différentiable en  $x$ , le calcul de l'unicité de la différentielle s'applique et la limite

$$\lim_{t \rightarrow 0} \frac{f(x + te_i) - f(x)}{t}$$

existe et vaut  $df(x)(e_i)$ . Mais si on y réfléchit bien, ça veut exactement dire que  $f$  admet en  $x$  une dérivée partielle dans la direction  $i$ , et que

$$\frac{\partial f}{\partial x_i}(x) = df(x)(e_i).$$

□

Ce qui est un peut plus délicat mais qui simplifie beaucoup la vie, c'est que quand  $f$  est de classe  $C^1$ , alors  $f$  est différentiable.

**Proposition 2.4.** *Soit  $f$  une fonction de classe  $C^1$ . Alors pour tout  $x \in U$ ,  $f$  est différentiable en  $x$  et sa différentielle est donnée par la formule (5.4).*

Montrons la proposition dans le cas  $d = 2$  et  $x = 0$  (la preuve est facilement adaptable dans le cas général mais un peu plus lourde à écrire). On va de nouveau utiliser le lemme 1.8.

*Preuve de la proposition.* On veut montrer que

$$\lim_{(s,t) \rightarrow 0} \left| \frac{f(s,t) - f(0,0) - s \frac{\partial f}{\partial s}(0,0) - t \frac{\partial f}{\partial t}(0,0)}{\|(s,t)\|} \right| = 0$$

On note  $A(s,t)$  cette grandeur. L'idée est de décomposer  $A$  de la façon suivante (on utilise l'inégalité triangulaire et les inégalité  $|s| \leq \|(s,t)\|$ , et  $|t| \leq \|(s,t)\|$ ) :

$$\begin{aligned} A(s,t) &= \left| \frac{f(s,t) - f(s,0) - t \frac{\partial f}{\partial t}(0,0)}{\|(s,t)\|} + \frac{f(s,0) - f(0,0) - s \frac{\partial f}{\partial s}(0,0)}{\|(s,t)\|} \right| \\ &\leq \left| \frac{f(s,t) - f(s,0) - t \frac{\partial f}{\partial t}(0,0)}{|t|} \right| + \left| \frac{f(s,0) - f(0,0) - s \frac{\partial f}{\partial s}(0,0)}{|s|} \right| \\ &= \left| \frac{f(s,t) - f(s,0)}{t} - \frac{\partial f}{\partial t}(0,0) \right| + \left| \frac{f(s,0) - f(0,0)}{s} - \frac{\partial f}{\partial s}(0,0) \right| \end{aligned}$$

Le terme de droite converge vers 0 par définition de la dérivée partielle. Quand au terme de gauche, on utilise le lemme 1.8 à l'application  $\varphi$  définie pour tout  $\tau \in [0, t]$  par  $\varphi(\tau) = f(s, \tau)$ , de dérivée définie pour tout  $\tau$  par

$$\varphi'(\tau) = \frac{\partial f}{\partial t}(s, \tau).$$

On trouve ainsi

$$\frac{f(s,t) - f(s,0)}{t} = \int_0^1 \frac{\partial f}{\partial t}(s, \tau t) d\tau.$$

Donc le membre de gauche vérifie

$$\begin{aligned} \left| \frac{f(s,t) - f(s,0)}{t} - \frac{\partial f}{\partial t}(0,0) \right| &= \left| \int_0^1 \frac{\partial f}{\partial t}(s, \tau t) d\tau - \frac{\partial f}{\partial t}(0,0) \right| \\ &= \left| \int_0^1 \left( \frac{\partial f}{\partial t}(s, \tau t) - \frac{\partial f}{\partial t}(0,0) \right) d\tau \right| \\ &\leq \int_0^1 \left| \frac{\partial f}{\partial t}(s, \tau t) - \frac{\partial f}{\partial t}(0,0) \right| d\tau, \end{aligned}$$

et on veut montrer que ça tend vers 0 quand  $(s,t)$  tend vers 0. Mais comme la dérivée partielle est continue, si on a  $\varepsilon > 0$  il existe  $\delta > 0$  tel que  $(s,t) \in B((0,0), \delta)$  implique

$$\left| \frac{\partial f}{\partial t}(s,t) - \frac{\partial f}{\partial t}(0,0) \right| \leq \varepsilon.$$

On choisit donc un tel  $\delta > 0$  et  $(s,t) \in B((0,0), \delta)$ . Alors pour tout  $\tau \in [0, 1]$ ,  $(s, \tau t) \in B((0,0), \delta)$ . Donc le terme à l'intérieur de l'intégrale est inférieur à  $\varepsilon$  pour tout  $\tau$ , et donc l'intégrale entière est inférieure à  $\varepsilon$ .  $\square$

Cette proposition donne donc un moyen standard pour montrer qu'une application  $f$  est différentiable en  $x$  :

- on vérifie que  $f$  admet des dérivées partielles dans toutes les directions dans un ouvert contenant  $x$ ,
- on vérifie que ces dérivées partielles sont continues (la continuité en  $x$  suffit mais en général, elles sont continues partout).

Par contre, ce n'est parfois (rarement) pas la méthode la plus avantageuse. Par exemple, dans l'exercice suivant, on peut aller plus vite.

**Exercice 2.5.** Soit  $l$  une application linéaire de  $\mathbb{R}^d$  dans  $\mathbb{R}$ , et  $x \in \mathbb{R}^d$ . Montrer que  $l$  est différentiable en  $x$  et calculer sa dérivée.

## 2.2 Formule de Taylor d'ordre 1

Regardons la formule (5.2) à la lumière de la formule (5.4). Si  $f$  est différentiable en un point  $x$  de  $U$ , alors pour tout  $y \in U$ ,

$$f(y) = f(x) + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(x)(y_i - x_i) + o_{y \rightarrow x}(\|y - x\|).$$

C'est la **formule de Taylor d'ordre 1** de  $f$  en  $x$ . Attention, même si la formule ne fait intervenir que les dérivées partielles de  $f$  en  $x$  il ne suffit pas que les dérivées partielles de  $f$  existent en  $x$  pour que la formule soit vérifiée, il faut bien que  $f$  soit différentiable en  $x$ .

## 2.3 Gradient d'une fonction

Revenons maintenant sur la formule (5.4). Soit  $x \in U$  un point de différentiabilité de  $f$ . L'équation (5.4) donne donc une expression de la différentielle de  $f$  au point  $x$  en fonction des dérivées partielles de  $f$  au point  $x$ . En notant

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_d}(x) \end{pmatrix},$$

on a pour tout  $v \in \mathbb{R}^d$

$$df(x)(v) = \langle \nabla f(x), v \rangle.$$

Le vecteur  $\nabla f(x)$  s'appelle le gradient de  $f$  en  $x$ . Il pointe vers la ligne de plus grande pente de  $f$  et a une norme d'autant plus grande que le graphe de  $f$  est pentu. La formule de Taylor d'ordre 1 peut se réécrire en faisant usage du gradient :

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + o_{y \rightarrow x}(\|y - x\|).$$

# 3 La différentielle des fonctions à valeurs vectorielles

## 3.1 Définition

Comme je l'ai déjà dit à de nombreuses reprises, les fonctions à valeurs vectorielles, c'est pas plus compliqué que les fonctions à valeurs réelles, parce qu'on peut regarder les coordonnées une à une. Néanmoins, ici, on peut de façon équivalente regarder les choses coordonnées par coordonnées ou de façon un peu plus algébrique. Dans ce paragraphe,  $F = (f_1, \dots, f_n)$  est une application d'un ouvert  $U$  de  $\mathbb{R}^d$  à valeurs dans  $\mathbb{R}^n$ .

**Définition 3.1.** Soit  $x \in U$ . On dit que  $F$  est différentiable en  $x$  si l'une des deux affirmations équivalentes suivantes est vérifiée :

- (i) les applications  $f_1, \dots, f_n$  sont différentiables en  $x$ ,
- (ii) il existe une application linéaire  $A$  de  $\mathbb{R}^d$  dans  $\mathbb{R}^n$  telle que pour tout  $y \in \mathbb{R}^d$ ,

$$F(y) = F(x) + A(y - x) + \underset{y \rightarrow x}{o}(\|y - x\|). \quad (5.5)$$

Dans ce cas,  $A$  est unique, on la note  $dF(x)$  et elle vérifie pour tout  $v \in \mathbb{R}^d$  :

$$dF(x)(v) = \begin{pmatrix} df_1(x)(v) \\ \vdots \\ df_n(x)(v) \end{pmatrix}. \quad (5.6)$$

*Preuve de l'équivalence.* Supposons que  $f_1, \dots, f_n$  sont différentiables. On dispose donc des applications linéaires  $df_1(x), \dots, df_n(x)$ . On choisit pour  $A$  l'application linéaire qui à  $v \in \mathbb{R}^d$  associe

$$A(v) := \begin{pmatrix} df_1(x)(v) \\ \vdots \\ df_n(x)(v) \end{pmatrix}.$$

Si  $y \in U$ , alors

$$\frac{F(y) - F(x) - A(y - x)}{\|y - x\|} = \begin{pmatrix} \frac{f_1(y) - f_1(x) - df_1(x)(y - x)}{\|y - x\|} \\ \vdots \\ \frac{f_d(y) - f_d(x) - df_d(x)(y - x)}{\|y - x\|} \end{pmatrix}.$$

Chacun des termes de ce vecteur converge vers 0 quand  $y$  tend vers  $x$ , donc on a bien (5.5).

Supposons que l'on ait une application linéaire  $A$  telle que (5.5) soit satisfaite. Alors pour tout  $v \in \mathbb{R}^d$ ,  $A(v) \in \mathbb{R}^n$ . Considérons sa  $i$ -ième coordonnée, où  $i \in \{1, \dots, n\}$ . On la note  $a_i(v)$ . L'application  $a_i$  est linéaire et à valeurs dans  $\mathbb{R}$ . Montrons que c'est la différentielle de  $f_i$  en  $x$ . On a pour tout  $y \in U$  :

$$\frac{F(y) - F(x) - A(y - x)}{\|y - x\|} = \begin{pmatrix} \frac{f_1(y) - f_1(x) - a_1(y - x)}{\|y - x\|} \\ \vdots \\ \frac{f_d(y) - f_d(x) - a_d(y - x)}{\|y - x\|} \end{pmatrix},$$

et par hypothèse ce vecteur tend vers 0 quand  $y$  tend vers  $x$ . Donc chacune de ses coordonnées tend vers 0, en particulier la  $i$ -ième, donc  $f_i$  est bien différentiable en  $x$  et  $a_i = df_i(x)$ .  $A$  est donc bien de la forme (5.6).  $\square$

*Remarque 3.2.* Par ailleurs, on dit qu'une application  $F = (f_1, \dots, f_n)$  à valeurs vectorielles est de classe  $C^k$  avec  $k \geq 1$  si chacune des applications  $f_1, \dots, f_n$  l'est. On voit que si  $F$  est  $C^1$ , alors elle est différentiable.

**Représentation par des matrices.** Soit  $x \in U$  un point de différentiabilité de  $F$ . L'application  $dF(x)$  est linéaire de  $\mathbb{R}^d$  dans  $\mathbb{R}^n$ . Elle peut donc être représentée par sa matrice (de taille  $n \times d$ ) dans les bases

canoniques de  $\mathbb{R}^d$  et  $\mathbb{R}^n$ . Si on associe (5.6) et (5.4), on obtient pour tout  $v \in \mathbb{R}^d$  :

$$\begin{aligned} dF(x)(v) &= \begin{pmatrix} v_1 \frac{\partial f_1}{\partial x_1}(x) + \dots + v_d \frac{\partial f_1}{\partial x_d}(x) \\ \vdots \\ v_1 \frac{\partial f_n}{\partial x_1}(x) + \dots + v_d \frac{\partial f_n}{\partial x_d}(x) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_d}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \dots & \frac{\partial f_n}{\partial x_d}(x) \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix}. \end{aligned}$$

Cette matrice s'appelle la matrice jacobienne de  $F$  en  $x$ , notée  $J_F(x)$ .

**Exemples.** Si  $n = 1$ .

La fonction  $F$  est alors à valeurs réelles, c'est à dire qu'on se retrouve à nouveau dans le cas déjà traité dans la section 2. Sa matrice jacobienne est simplement le vecteur ligne  ${}^t \nabla F(x)$ .

De façon générale, la  $k$ -ème ligne de la matrice jacobienne d'une fonction  $F = (f_1, \dots, f_n)$  en  $x$  est l'écriture *sous forme de ligne* du gradient de  $f_k$ .

Si  $d = 1$ .

La fonction  $F$  ne dépend que d'une variable que l'on va noter  $t \in I$ ,  $I$  étant un intervalle de  $\mathbb{R}$ . L'image qu'il faut avoir est que l'ensemble  $\{F(t), t \in I\}$  est une courbe tracée dans l'espace à  $n$  dimensions. On peut s'imaginer que  $F(t)$  est la position d'une mouche dans l'espace à l'instant  $t$ , la courbe décrite étant alors l'ensemble de sa trajectoire. On a représenté un telle courbe à la figure ci-dessous.

Chacune des  $f_1, \dots, f_d$  est une fonction réelle d'une variable réelle. Elles n'ont donc qu'une dérivée partielle : leur dérivée. En particulier, une telle fonction est différentiable si et seulement si chacune de ses coordonnées est dérivable. On a alors

$$J_F(t) = \begin{pmatrix} f'_1(t) \\ \vdots \\ f'_d(t) \end{pmatrix}.$$

Dans ce cas, on note ce vecteur  $F'(t)$  plutôt que  $J_F(t)$ . C'est le vecteur vitesse de la courbe associée à  $F$  en  $t$  (et la vitesse de la mouche à l'instant  $t$ ). Il est tangent à la courbe, pointe dans la direction du mouvement de la mouche, et est d'autant plus grand que la mouche va vite (sa longueur est égal à la vitesse instantannée de la mouche).

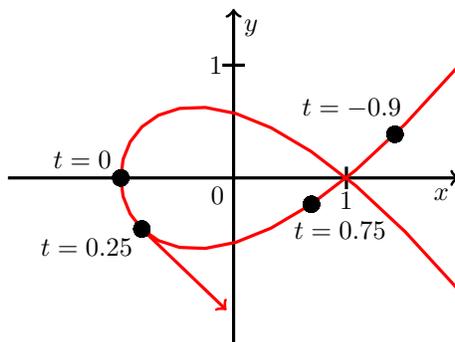


FIGURE 5.1 – On a représenté dans le plan la courbe associée à la fonction  $F$  qui à  $t \in [-1, 1]$  associe  $F(t) := (3t^2 - 1, 3t^3 - 2t)$ , c'est à dire l'ensemble  $\{(3t^2 - 1, 3t^3 - 2t), t \in [-1, 1]\}$ . On a marqué les points de cet ensemble pour différentes valeurs de  $t$ , et on a dessiné le vecteur vitesse en  $t = 0.25$ .

**Exercice 3.3.** Dans le cas où  $F$  ne dépend que d'une variable, vérifier qu'elle est différentiable en  $t$  si et seulement si  $f_1, \dots, f_d$  sont dérivables en  $t$  et que pour tout  $h \in \mathbb{R}$ ,

$$dF(t)(h) = hF'(t).$$

### 3.2 Opérations sur les différentielles

Comme dans le cas réel, il existe des formules pour la différentielle de la somme ou du produit d'applications différentiables. À chaque fois, la preuve est un peu plus pénible que ce qu'on pourrait croire car il est toujours délicat de montrer qu'une application est différentiable : il faut trouver un candidat pour la différentielle puis vérifier que le développement limité d'ordre 1 (5.5) est vérifié. En revanche, si l'on admet que les fonctions sont différentiables, les formules que l'on va donner sont juste des conséquences sur les dérivées partielles des calculs standards sur les dérivées de fonctions réelles. À chaque fois, on donnera le calcul sur les dérivées partielles donnant la formule, puis la preuve rigoureuse.

**Proposition 3.4** (linéarité). *Soient  $F$  et  $G$  deux applications d'un ouvert  $U$  de  $\mathbb{R}^d$  à valeurs dans  $\mathbb{R}^n$ , soit  $x \in U$ , et soient  $\lambda$  et  $\mu$  deux réels. Si  $F$  et  $G$  sont différentiables en  $x$ , alors  $\lambda F + \mu G$  est différentiable en  $x$  et*

$$d(\lambda F + \mu G)(x) = \lambda dF(x) + \mu dG(x).$$

On a donc également

$$J_{\lambda F + \mu G}(x) = \lambda J_F(x) + \mu J_G(x).$$

*Calcul en coordonnées.* On note  $F = (f_1, \dots, f_n)$  et  $G = (g_1, \dots, g_n)$ . Admettons que toutes les dérivées partielles que l'on va utiliser sont bien définies, et choisissons  $i \in \{1, \dots, n\}$  et  $j \in \{1, \dots, d\}$ . En dérivant en  $t = 0$  la fonction réelle d'une variable réelle

$$t \mapsto \lambda f_i(x + te_j) + \mu g_i(x + te_j),$$

et en utilisant la règle de dérivation des combinaisons linéaires de fonctions réelles, on obtient

$$\frac{\partial(\lambda f_i + \mu g_i)}{\partial x_j}(x) = \lambda \frac{\partial f_i}{\partial x_j}(x) + \mu \frac{\partial g_i}{\partial x_j}(x).$$

C'est exactement l'égalité sur les matrices de l'énoncé écrite coordonnée par coordonnée.  $\square$

*Démonstration.* Notons  $H$  l'application  $\lambda F + \mu G$ ,  $A := dF(x)$ ,  $B := dG(x)$  et  $C := \lambda A + \mu B$ . L'objectif est de montrer que  $H$  est différentiable en  $x$  et que  $dH(x) = C$ . Il s'agit donc de montrer que

$$\lim_{y \rightarrow x} \frac{H(y) - H(x) - C(y - x)}{\|y - x\|} = 0.$$

Mais on vérifie facilement que

$$\begin{aligned} \frac{H(y) - H(x) - C(y - x)}{\|y - x\|} &= \lambda \frac{F(y) - F(x) - A(y - x)}{\|y - x\|} \\ &\quad + \mu \frac{G(y) - G(x) - B(y - x)}{\|y - x\|}, \end{aligned}$$

et chacun des termes du membre de droite converge vers 0 quand  $y$  tend vers  $x$ . On a donc le résultat.

Le résultat sur les matrices vient simplement du fait que l'application qui à une application linéaire associe sa matrice dans une base est elle-même linéaire.  $\square$

**Proposition 3.5** (produit). *Soient  $f$  et  $g$  deux applications d'un ouvert  $U$  de  $\mathbb{R}^d$  à valeurs dans  $\mathbb{R}$  et soit  $x \in U$ . Si  $f$  et  $g$  sont différentiables en  $x$ , alors  $fg$  est différentiable en  $x$ , et*

$$d(fg)(x) = f(x) dg(x) + g(x) df(x).$$

On a donc également

$$\nabla(fg)(x) = f(x)\nabla g(x) + g(x)\nabla f(x).$$

*Calcul en coordonnées.* Admettons que toutes les dérivées partielles que l'on va utiliser sont bien définies, et choisissons  $i \in \{1, \dots, d\}$ . En dérivant en  $t = 0$  la fonction réelle d'une variable réelle

$$t \mapsto f(x + te_i) \times g(x + te_i),$$

et en utilisant la règle de dérivation des produits de fonctions réelles, on obtient

$$\frac{\partial fg}{\partial x_i}(x) = f(x) \frac{\partial g}{\partial x_i}(x) + g(x) \frac{\partial f}{\partial x_i}(x).$$

C'est exactement l'égalité sur les gradients de l'énoncé écrite coordonnée par coordonnée.  $\square$

*Démonstration.* On définit  $h := fg$ , et on note  $a := df(x)$ ,  $b := dg(x)$  et  $c := f(x)a + g(x)b$ . L'objectif est de montrer que  $h$  est différentiable en  $x$  et que sa différentielle est  $c$ . On veut donc montrer que

$$\lim_{y \rightarrow x} \frac{h(y) - h(x) - c(y-x)}{\|y-x\|} = 0.$$

Mais

$$\begin{aligned} \frac{h(y) - h(x) - c(y-x)}{\|y-x\|} &= \frac{f(y)g(y) - f(x)g(x)}{\|y-x\|} - f(x) \frac{b}{\|y-x\|} - g(x) \frac{a}{\|y-x\|} \\ &= f(x) \frac{g(y) - g(x) - b(y-x)}{\|y-x\|} + g(y) \frac{f(y) - f(x) - a(y-x)}{\|y-x\|} \\ &\quad + (g(x) - g(y))a \left( \frac{y-x}{\|y-x\|} \right). \end{aligned}$$

Ça se voit facilement que les deux premiers termes tendent vers 0. En fait, le troisième se voit moins mais il tend plus vite vers 0 ! En effet,  $g(y) - g(x)$  tend vers 0 et

$$a \left( \frac{y-x}{\|y-x\|} \right)$$

reste borné donc le produit des deux tend vers 0. Pour voir ça, il y a tout un tas de façons mais celle qui a le plus de raisonance avec le cours de topologie consiste à dire que  $a$  est continue et que

$$\frac{y-x}{\|y-x\|}$$

est de norme 1. Or l'ensemble des vecteurs de  $\mathbb{R}^d$  de norme 1 (la sphère de  $\mathbb{R}^d$ , souvent notée  $S^{d-1}$ ) est un compact (car il est fermé et borné), et donc  $a$  est bornée sur  $S^{d-1}$ .  $\square$

### 3.3 Composition de fonctions différentiables

On va expliquer dans ce paragraphe comment calculer la différentielle d'une composée d'applications différentiables. C'est un point crucial pour toute la suite de l'année et je vous invite avec insistance à travailler cette partie du cours. Je vais traiter pas mal de cas différents même si tous sont impliqués par le dernier théorème, à savoir le théorème 3.9. Comme dans le paragraphe précédent, je vais insister sur les calculs qui permettent de comprendre les résultats. Par soucis d'exhaustivité, je donnerai tout de même les preuves rigoureuses des théorèmes 3.6 et 3.9, mais ces preuves peuvent être négligées en première lecture.

Commençons par rappeler la règle de dérivation des composées de fonctions réelles. On peut penser par exemple que vous vous baladiez sur l'autoroute avec votre voiture (donc globalement en ligne droite). La première fonction pourrait donner votre position en fonction du temps (par exemple le numéro de la borne kilométrique au niveau de laquelle vous êtes), et la deuxième fonction pourrait donner l'altitude en fonction de cette position. On s'interrogerait alors sur l'évolution de l'altitude de votre voiture en fonction du temps.

**Théorème 3.6.** Soit  $f : I \rightarrow J \subset \mathbb{R}$  et  $g : J \rightarrow \mathbb{R}$ , où  $I$  et  $J$  sont des intervalles de  $\mathbb{R}$ , et soit  $x \in I$ . On suppose que

- $f$  est dérivable en  $x \in I$ ,
- $g$  est dérivable en  $f(x) \in J$ .

Alors  $g \circ f$  est dérivable en  $x$  et

$$(g \circ f)'(x) = g'(f(x)) \times f'(x).$$

*Le calcul qui fait comprendre.* On rappelle que si  $h$  est une fonction de l'intervalle  $K \subset \mathbb{R}$  dans  $\mathbb{R}$ ,  $h$  est dérivable en  $t \in K$  si et seulement si elle admet un développement limité d'ordre 1 :

$$h(t + \delta) = h(t) + a\delta + \underset{\delta \rightarrow 0}{o}(\delta).$$

Dans ce cas,  $a = h'(t)$ .

On va oublier les petits  $o$  et on va essayer d'estimer  $g \circ f(x + \delta)$ . Comme  $f$  est dérivable en  $x$ , quand  $\delta$  est petit,

$$f(x + \delta) \approx f(x) + \delta f'(x).$$

De la même façon, comme  $g$  est dérivable en  $y := f(x)$ , quand  $\varepsilon$  est petit,

$$g(y + \varepsilon) \approx g(y) + \varepsilon g'(y).$$

Donc maintenant, quand  $\delta$  est petit,

$$g \circ f(x + \delta) = g(f(x + \delta)) \approx g(f(x) + \delta f'(x)).$$

Mais maintenant, avec  $y := f(x)$  et  $\varepsilon := \delta f'(x)$ ,  $\varepsilon$  est bien petit et en utilisant le développement limité d'ordre 1 de  $g$ ,

$$g \circ f(x + \delta) \approx g(f(x)) + \delta f'(x) g'(f(x)).$$

Donc la dérivée de  $g \circ f$  doit bien être

$$g'(f(x)) f'(x).$$

□

*Preuve rigoureuse.* On fait la preuve de façon volontairement un peu plus difficile que nécessaire. Ce n'est pas pour se faire du mal, c'est pour que le raisonnement s'adapte de façon presque transparente au cas multidimensionnel. Pour les plus tatillons d'entre vous, retenez que dans tous les quotients que l'on va écrire, on utilise la convention  $0/0 = 0$ .

La fonction  $g$  est dérivable en  $y := f(x)$  cela signifie que

$$\frac{g(z) - g(y) - g'(y)(z - y)}{z - y} \xrightarrow{z \rightarrow y} 0.$$

Mais quand  $h \rightarrow 0$ ,  $f(x + h) \rightarrow f(x)$  puisque comme  $f$  est dérivable en  $x$ , elle est continue en  $x$ . Donc en utilisant la formule ci-dessus avec  $z = f(x + h)$  (on utilise en fait la règle de composition des limites),

$$\frac{g(f(x + h)) - g(f(x)) - g'(f(x))(f(x + h) - f(x))}{f(x + h) - f(x)} \xrightarrow{h \rightarrow 0} 0.$$

Mais comme

$$\frac{f(x + h) - f(x)}{h} \xrightarrow{h \rightarrow 0} f'(x),$$

en multipliant ces deux limites,

$$\frac{g(f(x + h)) - g(f(x)) - g'(f(x))(f(x + h) - f(x))}{h} \xrightarrow{h \rightarrow 0} 0. \quad (5.7)$$

Par ailleurs,

$$\frac{f(x+h) - f(x) - f'(x)h}{h} \xrightarrow{h \rightarrow 0} 0. \quad (5.8)$$

Il ne reste plus qu'à calculer la limite (5.7) +  $g'(f(x)) \times$  (5.8). On obtient

$$\frac{g(f(x+h)) - g(f(x)) - g'(f(x))f'(x)h}{h} \xrightarrow{h \rightarrow 0} 0,$$

ce qui signifie bien

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

□

Je répète qu'on ne va pas démontrer directement démontrer les deux théorèmes suivant, mais ils sont des conséquences directes du théorème 3.9 qui lui sera démontré. On donne par contre les calculs heuristiques qu'il faut maîtriser !

On commence par le cas de la composition d'une fonction de  $d$  variables à valeurs réelles avec une fonction réelle. Par exemple, on peut penser que la fonction de  $d$  variables représente l'altitude d'un relief en fonction de la position horizontale et que la fonction réelle représente la température en fonction de l'altitude. On veut donc déterminer l'évolution de la température avec la position horizontale.

**Théorème 3.7.** Soit  $f : U \subset \mathbb{R}^d \rightarrow I \subset \mathbb{R}$  et  $g : I \rightarrow \mathbb{R}$  où  $U$  est un ouvert de  $\mathbb{R}^d$  et  $I$  est un intervalle de  $\mathbb{R}$ . Soit  $x \in U$ . On suppose que

- $f$  est différentiable en  $x \in U$ ,
- $g$  est dérivable en  $f(x) \in I$ .

Alors  $g \circ f : U \rightarrow \mathbb{R}$  est différentiable en  $x$  et

$$d(g \circ f)[x] = g'(f(x)) df[x].$$

En termes de gradients, cela donne

$$\nabla(g \circ f)(x) = g'(f(x)) \nabla f(x).$$

En termes de dérivées partielles, cela donne pour  $i = 1, \dots, d$ ,

$$\frac{\partial g \circ f}{\partial x_i}(x) = g'(f(x)) \frac{\partial f}{\partial x_i}(x).$$

*Le calcul qui permet de comprendre.* Si l'on admet que  $g \circ f$  est différentiable, le lien entre différentielle, gradient et dérivées partielles montre que les trois formules sont en fait les mêmes. On fait le calcul en coordonnées, c'est à dire pour les dérivées partielles. Comme dans le cas réel, si  $h : U \rightarrow \mathbb{R}$  a une approximation du genre

$$h(x + \delta e_i) \approx h(x) + a\delta$$

pour  $\delta$  petit, alors on peut conclure que

$$a = \frac{\partial h}{\partial x_i}(x).$$

Cherchons une telle approximation pour  $g \circ f$ . Si  $\delta$  est petit,

$$\begin{aligned} g \circ f(x + \delta e_i) &= g(f(x + \delta e_i)) \\ &= g(f(x_1, \dots, x_{i-1}, x_i + \delta, x_{i+1}, \dots, x_d)) \\ &\approx g\left(f(x) + \frac{\partial f}{\partial x_i}(x)\delta\right). \end{aligned}$$

Mais alors en posant  $y := f(x)$  et  $\varepsilon := \frac{\partial f}{\partial x_i}(x)\delta$ ,  $\varepsilon$  est petit, et

$$\begin{aligned} g \circ f(x + \delta e_i) &\approx g(y + \varepsilon) \approx g(y) + g'(y)\varepsilon \\ &= g(f(x)) + g'(f(x)) \frac{\partial f}{\partial x_i}(x)\delta. \end{aligned}$$

On obtient bien

$$\frac{\partial g \circ f}{\partial x_i}(x) = g'(f(x)) \frac{\partial f}{\partial x_i}(x).$$

□

On traite maintenant le cas où la première fonction ne dépend que d'une variable mais est à valeurs dans un espace de dimension  $d$ , et où la deuxième fonction dépend de  $d$  variables et est à valeurs réelles. Par exemple, la première fonction pourrait être votre position horizontale en randonnée en fonction du temps, et la seconde pourrait être l'altitude en fonction de la position horizontale. On cherche alors l'évolution de votre altitude en fonction du temps (qui est une fonction réelle!).

**Théorème 3.8.** Soit  $f : I \subset \mathbb{R} \rightarrow U \subset \mathbb{R}^d$  et  $g : U \rightarrow \mathbb{R}$  où  $I$  est un intervalle de  $\mathbb{R}$  et  $U$  est un ouvert de  $\mathbb{R}^d$ . On note pour tout  $s \in I$ ,  $f(s) = (f_1(s), \dots, f_d(s))$ . Soit  $t \in I$ . On suppose que :

- $f$  est dérivable en  $t$ ,
- $g$  est différentiable en  $f(t)$ .

Alors  $g \circ f : I \subset \mathbb{R} \rightarrow \mathbb{R}$  est dérivable en  $t$  et

$$(g \circ f)'(t) = dg[f(t)](f'(t)) = \langle \nabla g(f(t)), f'(t) \rangle = \sum_{i=1}^d \frac{\partial g}{\partial y_i}(f(t)) f'_i(t).$$

Le calcul qui fait comprendre. On pose

$$y = (y_1, \dots, y_d) := (f_1(t), \dots, f_d(t)) = f(t),$$

et on choisit  $\delta \in \mathbb{R}$  petit et  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d) \in \mathbb{R}^d$  petit. Les approximations de  $f$  et  $g$  à l'ordre 1 sont :

$$\begin{aligned} f(t + \delta) &= \begin{pmatrix} f_1(t + \delta) \\ \vdots \\ f_d(t + \delta) \end{pmatrix} \approx \begin{pmatrix} f_1(t) + f'_1(t)\delta \\ \vdots \\ f_d(t) + f'_d(t)\delta \end{pmatrix} = f(t) + f'(t)\delta, \\ g(y + \varepsilon) &= g(y_1 + \varepsilon_1, \dots, y_d + \varepsilon_d) \approx g(y) + \sum_{i=1}^d \frac{\partial g}{\partial y_i}(y)\varepsilon_i. \end{aligned}$$

Donc en combinant ces deux approximations,

$$g \circ f(t + \delta) = g\left(f(t) + \delta f'(t)\right) = g(f(t)) + \sum_{i=1}^d \frac{\partial g}{\partial y_i}(f(t)) f'_i(t)\delta.$$

On en déduit bel et bien que

$$(g \circ f)'(t) = \sum_{i=1}^d \frac{\partial g}{\partial y_i}(f(t)) f'_i(t).$$

Les autres formulations ne sont que des réécriture de ce même résultat. □

On donne (enfin) le cas général. La c'est plus compliqué de trouver une incarnation de ce résultat dans notre quotidien, mais je vous laisse le soin de chercher. Pour bien comprendre, on fera d'abord deux calculs heuristiques avant de s'attaquer à la vraie preuve.

**Théorème 3.9.** Soit  $F : U \subset \mathbb{R}^d \rightarrow V \subset \mathbb{R}^n$  et  $G : V \rightarrow \mathbb{R}^p$  où  $U$  et  $V$  sont des ouverts de  $\mathbb{R}^d$  et  $\mathbb{R}^n$ . Si  $x' \in U$ , on note  $F(x') = (f_1(x'), \dots, f_n(x'))$ . Si  $y \in V$ , on note  $G(y) = (g_1(y), \dots, g_p(y))$ . Remarquons que  $G \circ F$  est une application de  $U$  dans  $\mathbb{R}^p$  et que pour tout  $x \in U$ ,  $G \circ F(x) = (g_1 \circ F(x), \dots, g_p \circ F(x))$ .

Soit  $x \in U$ . On suppose que

- $F$  est différentiable en  $x$ ,
- $G$  est différentiable en  $F(x)$ .

Alors  $G \circ F$  est différentiable en  $x$  et on a,

$$dG \circ F[x] = dG[F(x)] \circ dF[x].$$

On a donc en termes de matrices

$$J_{G \circ F}(x) = J_G(F(x)) \cdot J_F(x).$$

Enfin, en coordonnées, cela donne pour  $i = 1, \dots, p$  et  $j = 1, \dots, d$ ,

$$\frac{g_i \circ F}{\partial x_j}(x) = \sum_{k=1}^n \frac{\partial g_i}{\partial y_k}(F(x)) \times \frac{\partial f_k}{\partial x_j}(x).$$

*Remarque 3.10.* Encore une fois, une fois que l'on sait que  $G \circ F$  est différentiable, et une fois que l'on connaît le lien entre différentielles, jacobiniennes et dérivées partielles, les trois formules de l'énoncé sont clairement équivalentes.

*Calcul pour comprendre en coordonnées.* On se propose ici de calculer la dérivée partielle

$$\frac{\partial g_i \circ F}{\partial x_j}(x)$$

où  $i \in \{1, \dots, p\}$  et  $j \in \{1, \dots, d\}$ . Si on sait le faire pour chaque  $i$  et  $j$ , on aura entièrement caractérisée  $dG \circ F[x]$  puisque l'on aura calculé toutes les coordonnées de  $J_{G \circ F}(x)$ . Pour ça, il nous faut estimer

$$g_i(F(x + \delta e_j))$$

où  $\delta$  est petit. Commençons par estimer  $F(x + \delta e_j)$  :

$$\begin{aligned} F(x + \delta e_j) &= (f_1(x + \delta e_j), \dots, f_n(x + \delta e_j)) \\ &\approx \left( f_1(x) + \frac{\partial f_1}{\partial x_j}(x)\delta, \dots, f_d(x) + \frac{\partial f_d}{\partial x_j}(x)\delta \right). \end{aligned}$$

Du coup,

$$\begin{aligned} g_i(F(x + \delta e_j)) &\approx g_i \left( f_1(x) + \frac{\partial f_1}{\partial x_j}(x)\delta, \dots, f_d(x) + \frac{\partial f_d}{\partial x_j}(x)\delta \right) \\ &\approx g_i(F(x)) + \sum_{k=1}^n \frac{\partial g_i}{\partial y_k}(F(x)) \frac{\partial f_k}{\partial x_j}(x)\delta. \end{aligned}$$

Donc on a bien

$$\frac{g_i \circ F}{\partial x_j}(x) = \sum_{k=1}^n \frac{\partial g_i}{\partial y_k}(F(x)) \times \frac{\partial f_k}{\partial x_j}(x).$$

□

*Calcul pour comprendre intrinsèque.* Une autre façon de voir ce calcul est de raisonner directement sur les différentielles. Soit  $v \in \mathbb{R}^d$  et  $\delta$  petit. Alors

$$F(x + \delta v) \approx F(x) + \delta dF[x](v).$$

Par ailleurs, si  $y := F(x)$ , si  $w \in \mathbb{R}^n$  et si  $\delta$  est petit,

$$G(y + \varepsilon w) \approx G(y) + \delta dG[y](w).$$

Donc en utilisant la formule ci-dessus avec  $w = dF[x](v)$ ,

$$\begin{aligned} G(F(x + \delta v)) &\approx G\left(F(x) + \delta dF[x](v)\right) \\ &\approx G(F(x)) + \delta dG[F(x)](dF[x](v)) \\ &= G(F(x)) + \delta dG[F(x)] \circ dF[x](v). \end{aligned}$$

Donc on a bien

$$dG \circ F[x] = dG[F(x)] \circ dF[x].$$

□

*Démonstration.* Notons  $A := dF[x]$  et  $B := dG[F(x)]$ . Il faut commencer par remarquer qu'il existe  $M > 0$  et  $\varepsilon > 0$  tels que  $y \in B(x, \varepsilon)$  implique

$$\frac{\|F(y) - F(x)\|}{\|y - x\|} \leq M.$$

En effet,

$$\frac{F(y) - F(x)}{\|y - x\|} = A \left( \frac{y - x}{\|y - x\|} \right) + \frac{F(y) - F(x) - A(y - x)}{\|y - x\|}.$$

Le premier terme du membre de droite est borné par le même argument de compacité que dans la preuve précédente, disons par un réel  $K$ , et le second tend vers 0. On obtient donc le résultat facilement en choisissant n'importe quel  $M > K$ . Ensuite, par définition,

$$\lim_{y \rightarrow x} \frac{F(y) - F(x) - A(y - x)}{\|y - x\|} = 0. \quad (5.9)$$

et par continuité de  $F$  en  $x$ ,  $F(y)$  tend vers  $F(x)$  quand  $y$  tend vers  $x$ , et donc

$$\lim_{y \rightarrow x} \frac{G(F(y)) - G(F(x)) - B(F(y) - F(x))}{\|F(y) - F(x)\|} = 0.$$

Par le petit résultat que l'on vient de montrer, on peut multiplier cette limite par  $\|F(y) - F(x)\|/\|y - x\|$  et conserver la convergence vers 0 :

$$\lim_{y \rightarrow x} \frac{G(F(y)) - G(F(x)) - B(F(y) - F(x))}{\|y - x\|} = 0. \quad (5.10)$$

En considérant (5.9) + B(5.10), on obtient le résultat par continuité de  $B$ . □

## 4 Calcul différentiel à l'ordre 2

### 4.1 Différentielle du gradient d'une fonction

Soit  $f$  une fonction de classe  $C^2$  de  $U$  dans  $R$ . Alors son gradient  $\nabla f$  est une application de  $U$  dans  $\mathbb{R}^d$ , de classe  $C^1$ , en particulier elle est différentiable pour tout  $x \in U$ . Si  $x \in U$  on peut donc considérer

sa différentielle en  $x$ . C'est une application linéaire de  $\mathbb{R}^d$  dans  $\mathbb{R}^d$ . Jetons un oeil à sa matrice dans la base canonique de  $\mathbb{R}^d$ . C'est une matrice carré dont les coefficients sont

$$J_{\nabla f}(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_1}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_d}(x) & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_d}(x) \end{pmatrix}.$$

En vertu du théorème de Schwarz, c'est une matrice **symétrique** et ceci aura un rôle fondamental dans l'étude des extrema locaux des fonctions de plusieurs variables. On l'appelle la matrice **Hessienne** de  $f$  notée  $H_f(x)$ .

## 4.2 Formule de Taylor du second ordre

Il y a un équivalent de la formule de Taylor d'ordre 1 faisant intervenir les dérivées partielles d'ordre 2 de  $f$ . Elle généralise pour les fonctions de plusieurs variables le développement limité réel valable pour les fonctions  $g$  d'une variable de classe  $C^2$

$$g(x+h) = g(x) + hg'(x) + \frac{1}{2}h^2g''(x) + o_{h \rightarrow 0}(h^2).$$

**Proposition 4.1.** Soit  $f$  une fonction de classe  $C^2$  d'un ouvert  $U$  de  $\mathbb{R}^d$  à valeurs dans  $\mathbb{R}$ . Alors pour tout  $h \in \mathbb{R}^d$  tel que  $x+h \in U$ ,

$$\begin{aligned} f(x+h) &= f(x) + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(x)h_i + \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \frac{\partial^2 f}{\partial x_j \partial x_k}(x)h_j h_k + o_{h \rightarrow 0}(\|h\|^2) \\ &= f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle H_f(x) \cdot h, h \rangle + o_{h \rightarrow 0}(\|h\|^2). \end{aligned}$$

Ici, la notation "o" est équivalente à l'existence d'une fonction  $\varepsilon_x$  sur l'ensemble des  $h$  tels que  $x+h \in U$  tendant vers 0 en 0 et telle que

$$\begin{aligned} f(x+h) &= f(x) + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(x)h_i + \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \frac{\partial^2 f}{\partial x_j \partial x_k}(x)h_j h_k + \|h\|^2 \varepsilon_x(h) \\ &= f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle H_f(x) \cdot h, h \rangle + \|h\|^2 \varepsilon_x(h). \end{aligned}$$

*Élément de preuve.* Soit  $x \in U$ ,  $r > 0$  tel que  $B(x, r) \subset U$  et  $h \in B(x, r) \setminus \{x\}$ . On définit la fonction  $\varphi$  pour tout  $t \in [-1, 1]$  par

$$\varphi(t) = f(x+th).$$

On a déjà vu à multiple reprise que pour tout  $t$ ,

$$\varphi'(t) = \langle \nabla f(x+th), h \rangle.$$

Calculons la dérivée seconde de  $\varphi$  en 0.

$$\begin{aligned}
 \varphi''(0) &= \lim_{h \rightarrow 0} \frac{\varphi'(t) - \varphi'(0)}{t} \\
 &= \lim_{h \rightarrow 0} \frac{\langle \nabla f(x + th), h \rangle - \langle \nabla f(x), h \rangle}{t} \\
 &= \lim_{h \rightarrow 0} \left\langle \frac{\nabla f(x + th) - \nabla f(x)}{h}, h \right\rangle \\
 &= \left\langle \lim_{h \rightarrow 0} \frac{\nabla f(x + th) - \nabla f(x)}{t}, h \right\rangle \\
 &= \langle H_f(x) \cdot h, h \rangle.
 \end{aligned}$$

Donc le développement limité réel donne

$$\begin{aligned}
 f(x + th) &= f(x) + t \langle \nabla f(x), h \rangle + \frac{1}{2} t^2 \langle H_f(x) \cdot h, h \rangle + o_{t \rightarrow 0}(t^2) \\
 &= f(x) + \langle \nabla f(x), th \rangle + \frac{1}{2} t^2 \langle H_f(x) \cdot th, th \rangle + o_{t \rightarrow 0}(t^2).
 \end{aligned}$$

La vraie preuve consiste alors à dire que l'on peut "uniformiser ce développement limité dans toutes les directions". On peut par exemple le faire en utilisant la formule de Taylor avec reste intégral.  $\square$

### 4.3 Étude des points critiques d'une fonction de plusieurs variables

On peut désormais voir le théorème qui clos l'étude des extrema des fonctions de plusieurs variables définies sur des ouverts. Soit donc  $U$  un ouvert de  $\mathbb{R}^d$  et  $f$  une application de classe  $C^2$  de  $U$  à valeurs réelles. Soit  $x \in U$  tel que  $\nabla f(x) = 0$ . On note  $H$  la matrice hessienne de  $f$  au point  $x$ . Par le théorème de Schwarz,  $H$  est symétrique.

**Théorème 4.2.** • Si  $H$  est définie positive, alors  $x$  est un point de minimum local strict de  $f$ .

- Si  $H$  est définie négative, alors  $x$  est un point de maximum local strict de  $f$ .
- Si  $H$  a des valeurs propres strictement positives et des valeurs propres strictement négatives, alors  $x$  n'est ni un point de minimum local ni un point de maximum local.
- Si  $H$  est seulement positives ou seulement négative, on ne peut rien dire.

*Démonstration.* On rappelle la formule de Taylor d'ordre 2 en  $x$ . Pour tout  $h$  tel que  $x + h \in U$ ,

$$f(x + h) = f(x) + \frac{1}{2} \langle H \cdot x, x \rangle + \|h\|^2 \varepsilon(h)$$

avec  $\varepsilon$  qui tend vers 0 quand  $h$  tend vers 0.

Par exemple, si  $H$  est définie positive, soit  $\lambda$  sa plus petite valeur propre. Alors pour tout  $h \in \mathbb{R}^d$ , en reprenant les notations du paragraphe précédent,

$$\langle H \cdot h, h \rangle = \sum_{i=1}^d \lambda_i \langle f_i, h \rangle^2 \geq \lambda \|h\|^2.$$

Donc si on prend  $r > 0$  tel que si  $h \in B(x, r)$ , alors  $x + h \in U$  et  $\varepsilon(h) \leq \lambda/4$ , alors si  $h \in B(x, r \setminus \{0\})$ ,

$$\begin{aligned}
 f(x + h) &\geq f(x) + \frac{\lambda}{2} \|h\|^2 - \frac{\lambda}{4} \|h\|^2 \\
 &\geq f(x) + \frac{\lambda}{4} \|h\|^2 \\
 &> f(x).
 \end{aligned}$$

La preuve est identique pour  $H$  définie négative.

Le même calcul montre que si  $u$  est un vecteur propre associé à une valeur propre strictement positive, alors pour  $\delta$  suffisamment petit,

$$f(x + \delta u) > f(x),$$

ou en d'autres termes que la restriction de  $f$  à la droite vectorielle engendrée par  $u$  admet en  $x$  un minimum local strict. Comme on peut faire la même chose avec les vecteurs propres associés à des valeurs propre strictement négatives, on obtient le troisième point.

Enfin, le dernier point est une généralisation facile du fait qu'en dimension 1, si la dérivée seconde est nulle, alors on ne peut rien dire.  $\square$

# Chapitre 6

## Le théorème des fonctions implicites

### 1 Comment définir un domaine de dimension $p$ dans un espace de dimension $n$ ?

Dans ce premier paragraphe, tout ce qui est dit n'est pas présenté de façon rigoureuse. Cependant, c'est très important d'avoir bien compris tous ces raisonnements. On raisonne sur des domaines lisses : les fonctions intervenant dans leurs définitions seront de classe au moins  $C^1$ . Elles devront aussi être définies sur des ouverts.

Il y a essentiellement trois façons pour définir un domaine de dimension  $p$  dans  $\mathbb{R}^n$ , avec  $p < n$  : donner une paramétrisation, le définir par des équations ou enfin le définir comme le graphe d'une fonction.

*Exemple 1.1.* Par exemple le cercle de centre  $O$  et de rayon 1 dans  $\mathbb{R}^2$  peut être vu comme :

- l'ensemble des solutions de l'équation  $x_1^2 + x_2^2 = 1$ ,
- l'ensemble  $\{(\cos \theta, \sin \theta), \theta \in \mathbb{R}\}$ ,
- l'union des graphes des fonctions qui à  $x_1 \in [-1, 1]$  associent respectivement  $\sqrt{1 - x_1^2}$  et  $-\sqrt{1 - x_1^2}$ .

Voyons un à un ces moyens dans le cas général.

#### 1.1 La paramétrisation

C'est sûrement la méthode la plus naturelle. La façon la plus simple de voir un domaine de dimension  $p$ , c'est de le voir comme une déformation de l'espace  $\mathbb{R}^p$  ou d'un de ses ouverts. Donc si on dispose d'un ouvert  $U \subset \mathbb{R}^p$  et d'une application  $F : U \rightarrow \mathbb{R}^n$ , on a envie de dire que l'image de  $F$ , c'est à dire l'ensemble

$$S := \{F(x), x \in U\}$$

est un sous-ensemble de  $\mathbb{R}^n$  de dimension  $p$ . Pourtant, ce n'est visiblement pas le cas. Par exemple  $F$  peut ratatiner tous les points de  $\mathbb{R}^p$  sur un point de  $\mathbb{R}^n$ . C'est pour cela qu'il faut rajouter à tout ça une condition sur  $F$ . La plus simple que l'on pourrait imposer et que  $F$  soit **injective**, c'est à dire que si  $u$  et  $v$  sont deux points distincts de  $U$ , alors  $F(u) \neq F(v)$ . Ce n'est pas facile à vérifier, et en plus, ce n'est pas exactement ce que l'on veut. Par exemple, dans l'exemple du cercle,  $(\cos \theta, \sin \theta)$  revient périodiquement au même endroit, mais définit tout de même un ensemble qui a bien l'air d'être de dimension 1. Gardons tout de même en tête que l'on ne peut pas prendre n'importe quelle fonction.

Dans ce cadre, quel est le plan tangent à  $S$  en  $z_0 = F(x_0) \in S$ ? Si  $x$  est un vecteur de  $\mathbb{R}^p$  et  $\varepsilon$  est petit,

$$F(x_0 + \varepsilon x) \approx F(x_0) + \varepsilon dF(x_0) \cdot x = x_0 + \varepsilon J_F(x_0) \cdot x.$$

Donc l'ensemble des points de  $S$  dans le voisinage de  $z_0$  est **presque** l'ensemble des points au voisinage de  $z_0$  dans

$$\{z_0 + dF(x_0) \cdot x, x \in \mathbb{R}^p\} = z_0 + \text{im}(dF(x_0)).$$

C'est donc cet ensemble que l'on appelle le plan tangent à  $S$  en  $z_0$ . Cet espace est de dimension  $p$  si le rang de  $dF(x_0)$  est égal à  $p$ . C'est le rang maximal pour une application linéaire de  $\mathbb{R}^p$  dans  $\mathbb{R}^n$ .  $dF(x_0)$  est alors injective (par le théorème du rang). Et ça c'est une bonne condition sur  $F$  pour que son image soit de dimension  $p$ . C'est facilement calculable, et cela correspond bien à ce que l'on attend d'un espace de dimension  $p$  : on peut se déplacer selon  $p$  directions dans son plan tangent (et donc sur lui même, localement).

Je fais une dernière remarque juste pour dire que différents paramétrages d'une même surface peuvent n'avoir rien à voir du tout.

## 1.2 Les équations

C'est la méthode à laquelle on a affaire le plus souvent, car cela consiste à regarder "l'ensemble des points de  $\mathbb{R}^n$  qui satisfont une certaine propriété", du moment que cette propriété est traduisible en équation. Si on demande de définir le cercle de centre  $O$  et de rayon 1 dans  $\mathbb{R}^2$ , il est bien plus naturel de dire "c'est l'ensemble des points à distance 1 de  $O$ " que de dire "c'est l'ensemble des  $(\cos \theta, \sin \theta)$ ", et il en va de même pour bien d'autres exemples.

Définir un ensemble par une équation, c'est simplement donner une fonction  $g : V \rightarrow \mathbb{R}$  où  $V$  est un ouvert de  $\mathbb{R}^n$ , et définir

$$S := \{z \in V, g(z) = 0\}.$$

Par exemple pour le cercle,  $g(x_1, x_2) = x_1^2 + x_2^2 - 1$ . On peut aussi définir un ensemble par plusieurs équations, disons  $q$ , correspondant aux fonctions  $g_1, \dots, g_q$ . Mais dans ce cas, quitte à définir

$$G : x \in V \mapsto (g_1(x), \dots, g_q(x)) \in \mathbb{R}^q,$$

on a

$$S := \{z \in V, G(z) = 0\},$$

avec  $G$  à valeurs dans un espace de grande dimension.

Ce qu'il faut retenir intuitivement, c'est que chaque nouvelle équation enlève une dimension à l'ensemble que l'on considère, donc ici,  $S$  devrait être de dimension  $n - q$ . Ce n'est évidemment pas exactement ça puisque par exemple, mettre cinq fois la même équation ne diminue pas de 5 la dimension de l'ensemble considéré, et comme précédemment, pour avoir un critère exact, il faut s'intéresser au plan tangent.

Ici,  $z$  est dans le plan tangent à  $S$  en  $z_0$  si  $z_0 + \varepsilon(z - z_0)$  est **presque** dans  $S$ , c'est à dire si  $G(z_0 + \varepsilon(z - z_0))$  vaut 0 au premier ordre. Mais au premier ordre,

$$G(z_0 + \varepsilon(z - z_0)) \approx G(z_0) + \varepsilon dG(z_0)(z - z_0) = \varepsilon dG(z_0)(z - z_0).$$

Donc  $G(z_0 + \varepsilon(z - z_0))$  vaut 0 au premier ordre signifie exactement

$$z - z_0 \in \ker dG(z_0).$$

Le plan tangent à  $S$  en  $z_0$  est donc

$$z_0 + \ker dG(z_0).$$

C'est donc bien un espace de dimension  $n - q$  si  $dG(z_0)$ , qui est une application linéaire de  $\mathbb{R}^n$  dans  $\mathbb{R}^q$ , est de rang  $q$ , qui est le rang maximal pour une application de ce type.  $dG(z_0)$  est alors surjective.

### 1.3 Les graphes de fonctions

La dernière façon de définir des ensembles de dimension  $p$  dans un espace de dimension  $n$ , c'est de le faire via des graphes de fonctions. En fait ce n'est pas très naturel, et pourtant, c'est de celle là dont j'ai le plus parlé, en disant par exemple que le graphe d'une fonction de  $\mathbb{R}^2$  dans  $\mathbb{R}$  était une surface dans  $\mathbb{R}^3$ . En fait cette méthode a un grand avantage, c'est qu'elle est en quelque sorte intermédiaire entre le paramétrage et les équations. Voyons d'abord comment on la formalise.

On commence par voir l'espace  $\mathbb{R}^n$  comme un produit :  $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^q$  dont on note les éléments  $z = (x, y)$  avec  $x \in \mathbb{R}^p$  et  $y \in \mathbb{R}^q$ . Bien évidemment,  $p + q = n$ . Soit alors  $H : U \mapsto \mathbb{R}^q$  où  $U$  est un ouvert de  $\mathbb{R}^p$ . Le graphe de  $H$  est alors l'ensemble

$$S := \{(x, y) \in \mathbb{R}^n, y = H(x)\}.$$

En définissant

$$F : x \in U \mapsto (x, H(x)) \in \mathbb{R}^n \quad \text{et} \quad G : (x, y) \in U \times \mathbb{R}^q \mapsto y - H(x) \in \mathbb{R}^q,$$

on voit que

$$S = \{F(x), x \in U\} = \{z \in U \times \mathbb{R}^q, G(z) = 0\}.$$

C'est donc à la fois l'ensemble paramétré par  $F$  (et le paramétrage par la première coordonnée a quelque chose d'intrinsèque), et l'ensemble satisfaisant l'équation associée à  $G$ .

Dans ce cadre, quel est le plan tangent. Ça on en a un peu plus l'habitude : c'est le graphe de la différentielle de  $H$  : le plan tangent à  $S$  en  $z_0 = (x_0, y_0)$  est l'ensemble

$$\mathcal{P} = \{(x, y) \in U \times \mathbb{R}^q, y = H(x_0) + dH(x_0) \cdot (x - x_0)\}.$$

Or on peut calculer la différentielle de  $F$  en  $x_0$  :

$$dF(x_0) = (\text{id}_p, dH(x_0)), \text{ de matrice } \left( \begin{array}{c} \text{id}_p \\ dH(x_0) \end{array} \right).$$

En particulier,  $dF(x_0)$  est bien de rang  $p$  et

$$\mathcal{P} = F(x_0) + \text{im}(dF(x_0)).$$

Cela correspond donc bien à la notion de plan tangent paramétrique que l'on avait énoncé, et on voit qu'il n'y a pas de condition supplémentaire pour avoir  $\dim \mathcal{P} = p$ .

On peut faire le même exercice avec la représentation de  $S$  sous forme d'équation. On calcule la différentielle de  $G$  en  $z_0$ . Sa matrice dans la base canonique est

$$\left( -dH(z_0) \mid \text{id}_q \right).$$

Elle est bien de rang  $q$ , et

$$\mathcal{P} = z_0 + \ker dG(z_0).$$

Encore une fois, cela correspond bien à la notion de plan tangent pour les ensembles définis par des équations, et on voit encore qu'il n'y a pas de condition supplémentaire sur  $H$  pour obtenir un ensemble de dimension  $p$ .

L'énorme défaut de la représentation par un graphe est que l'on ne peut définir ainsi que les surfaces dont le plan tangent n'est pas orthogonal au sous-espace  $\mathbb{R}^p$  de la première coordonnée. En fait, ce n'est pas très grave quitte à changer de coordonnée, comme on le verra dans le théorème des fonctions implicites.

## 1.4 Bilan

On récapitule toutes les informations des trois paragraphes précédents dans le tableau suivant.

$S$ est définie par	un paramétrage : $F:U \subset \mathbb{R}^p \rightarrow \mathbb{R}^n$	une équation : $G:V \subset \mathbb{R}^n \rightarrow \mathbb{R}^q$	le graphe de $H:U \subset \mathbb{R}^p \rightarrow \mathbb{R}^q$
tangent en $z_0$	$z_0 + \text{im } dF(x_0)$	$z_0 + \ker dG(z_0)$	graphe de $H(x_0)$ $+ dH(x_0)(\cdot - x_0)$
condition de non-dégénérescence	$dF(x_0)$ de rang $p$	$dG(z_0)$ de rang $q$	aucune
dimension	$p$	$n - q$	$p = n - q$

## 2 Le théorème des fonctions implicites

Le théorème des fonctions implicites affirme que si un ensemble est donné par une équation induite par une fonction  $G$  de classe  $C^1$ , alors quitte à bien choisir la première et la seconde coordonnée, elle est également le graphe d'une fonction  $H$  (et donc également une paramétrisation vu ce qui a été dit à la section précédente). Bien que ne donnant pas de formule pour  $H$  (c'est pourquoi elle est dite "implicite"), le théorème donne une formule pour sa différentielle en fonction de celle de  $G$ .

### 2.1 Le théorème en dimension 2

Ici, on a donc une équation donnée par une fonction  $G$  d'un ouvert  $V \subset \mathbb{R}^2$  dans  $\mathbb{R}$ , de classe  $C^1$  et on définit

$$S := \{z \in V, G(z) = 0\}.$$

La condition de non-dégénérescence sur  $G$  est alors que sa différentielle soit de rang 1, donc simplement qu'elle soit non-nulle. C'est équivalent à ce que son gradient soit non-nul, ou encore à ce que l'une de ses dérivées partielles soit non-nulle. On note  $z = (x, y)$  les points de  $\mathbb{R}^2$ .

**Théorème 2.1** (Théorème des fonctions implicites en dimension 2). *On se donne  $z_0 = (x_0, y_0)$  un point de  $S$  où le gradient de  $G$  est non-nul. Si c'est la dérivée partielle*

$$\frac{\partial G}{\partial y}(z_0)$$

*qui est non-nulle, alors il existe  $\delta > 0$ ,  $\varepsilon > 0$  et  $H: ]x_0 - \delta, x_0 + \delta[ \rightarrow ]y_0 - \varepsilon, y_0 + \varepsilon[$  de classe  $C^1$  telle que pour tout  $z \in ]x_0 - \delta, x_0 + \delta[ \times ]y_0 - \varepsilon, y_0 + \varepsilon[$  :*

- la dérivée partielle  $\partial_y G(z)$  est non nulle,
- $z = (x, y)$  est dans  $S$  si et seulement si  $y = H(x)$ .

On a alors pour tout  $x \in ]x_0 - \delta, x_0 + \delta[$ ,

$$H'(x) = -\frac{\partial_x G(x, H(x))}{\partial_y G(x, H(x))}.$$

Sinon, la dérivée partielle

$$\frac{\partial G}{\partial x}(z_0)$$

est non-nulle, et on peut refaire le même développement avec une fonction de  $y$ .

Enfin, si  $G$  est de classe  $C^k$ ,  $H$  l'est également.

*Quelques éléments de la preuve.* Je ne montre pas l'existence de  $\delta$ ,  $\varepsilon$  et de la fonction  $H$ , j'explique simplement la formule. Pour tout  $x \in ]x_0 - \delta, x_0 + \delta[$ ,

$$G(x, H(x)) = 0.$$

En prenant la dérivée de cette fonction par rapport à  $x$ , on obtient

$$\partial_x G(x, H(x)) + \partial_y G(x, H(x)) \times H'(x) = 0.$$

On conclut en se rappelant que  $\partial_y G(z) \neq 0$  pour tout point  $z$  dans l'ensemble  $]x_0 - \delta, x_0 + \delta[ \times ]y_0 - \varepsilon, y_0 + \varepsilon[$ .  $\square$

*Remarque 2.2.* En fait, toutes les dérivées de  $H$  (quand elles existent) se calculent en fonction des dérivées partielles de  $G$ , comme on le verra dans un exemple un peu plus loin. Cependant, les formules générales sont un peu lourdes.

## 2.2 Un exemple d'application

Le théorème des fonctions implicites est souvent utilisé pour étudier le comportement de l'ensemble des solutions d'une équation autour d'un point. Considérons l'équation sur  $\mathbb{R}^2$  :

$$2x^3y + 2x^2 + y^2 = 5.$$

On regarde une équation réelle dans  $\mathbb{R}^2$ , on s'attend donc à ce que l'ensemble de ses solutions soit de dimension 1. On note  $\mathcal{C}$  cet ensemble. On remarque que  $(1, 1) \in \mathcal{C}$ . On se demande alors à quoi ressemble  $\mathcal{C}$  autour de  $(1, 1)$  : est ce qu'il a une tangente ? Si oui, laquelle ? Est-il au-dessus ou en-dessous de cette tangente ?

Pour répondre à cette question, on commence par définir

$$f : (x, y) \in \mathbb{R}^2 \mapsto 2x^3y + 2x^2 + y^2 - 5.$$

La fonction  $f$  est de classe  $C^\infty$ . On calcule son gradient et sa hessienne en  $(x, y) \in \mathbb{R}^2$  :

$$\nabla f(x, y) = \begin{pmatrix} 6x^2y + 4x \\ 2x^3 + 2y \end{pmatrix}, \quad H_f(x, y) = \begin{pmatrix} 12xy + 4 & 6x^2 \\ 6x^2 & 2 \end{pmatrix}.$$

En particulier, en  $(1, 1)$ ,

$$\nabla f(1, 1) = \begin{pmatrix} 10 \\ 4 \end{pmatrix}, \quad H_f(1, 1) = \begin{pmatrix} 16 & 6 \\ 6 & 2 \end{pmatrix}.$$

On observe que  $\partial_y f(1, 1) \neq 0$ . On peut donc appliquer le théorème des fonctions implicites et trouver  $\delta > 0$ ,  $\varepsilon > 0$  et  $\varphi$  une fonction de classe  $C^\infty$  de  $]1 - \delta, 1 + \delta[$  dans  $]1 - \varepsilon, 1 + \varepsilon[$  telle que pour tout  $(x, y) \in ]1 - \delta, 1 + \delta[ \times ]1 - \varepsilon, 1 + \varepsilon[$ ,  $(x, y) \in \mathcal{C}$  si et seulement si  $y = \varphi(x)$ . Calculons les dérivées de  $\varphi$  en 1. Pour tout  $x \in ]1 - \delta, 1 + \delta[$ ,

$$f(x, \varphi(x)) = 0.$$

Donc en dérivant cette égalité,

$$\partial_x f(x, \varphi(x)) + \partial_y f(x, \varphi(x))\varphi'(x) = 0.$$

En  $x = 1$ , on obtient

$$10 + 4\varphi'(1) = 0,$$

c'est à dire  $\varphi'(1) = -5/2$ . Ensuite en dérivant une nouvelle fois l'égalité obtenue, pour tout  $x \in ]1 - \delta, 1 + \delta[$ ,

$$\begin{aligned} \partial_{xx} f(x, \varphi(x)) + 2\partial_{xy} f(x, \varphi(x))\varphi'(x) + \partial_{yy} f(x, \varphi(x))(\varphi'(x))^2 \\ + \partial_y f(x, \varphi(x))\varphi''(x) = 0. \end{aligned}$$

En  $x = 1$ , on obtient

$$16 + 2 \times 6 \times \left(\frac{-5}{2}\right) + 2 \times \left(\frac{-5}{2}\right)^2 + 4\varphi''(1) = 0,$$

c'est à dire  $\varphi''(1) = 3/8$ . Donc par la formule de Taylor, pour tout  $x \in ]1 - \delta, 1 + \delta[$ ,

$$\varphi(x) = 1 - \frac{5}{2}(x - 1) + \frac{3}{8}(x - 1)^2 + o((x - 1)^2).$$

Dans ce développement limité, on lit la tangente à  $\mathcal{C}$  en  $(1, 1)$  : c'est la droite d'équation

$$y = 1 - \frac{5}{2}(x - 1) = -\frac{5}{2}x + \frac{7}{2},$$

et on lit également le fait que  $\mathcal{C}$  est localement au dessus de sa tangente en 1. En effet, puisque  $3/8 > 0$ ,  $\varphi(x) \geq 1 - 5/2 \times (x - 1)$  au voisinage de 1.

### 2.3 Le théorème en dimension supérieure

En dimension supérieure, il faut faire un peu plus attention mais l'idée est exactement la même. Cette fois  $G$  est une application de classe  $C^1$  d'un ouvert  $V \subset \mathbb{R}^n$  à valeur dans  $\mathbb{R}^q$ , et on note  $p = n - q$ . On définit encore

$$S := \{z \in V, G(z) = 0\}.$$

Il faut d'abord étudier un peu la condition de non-dégénérescence. Soit  $z_0$  un point de  $V$  tel que  $dG(z_0)$  soit de rang  $q$ . Alors en particulier, il existe  $q$  indices  $k_1, \dots, k_q$  dans  $\llbracket 1, n \rrbracket$  tels que

$$(dG(z_0) \cdot e_{k_1}, \dots, dG(z_0) \cdot e_{k_q})$$

soit une famille libre (cela signifie que parmi les colonnes de la jacobienne de  $G$  en  $z_0$ , il en existe  $q$  qui soient linéairement indépendantes). Quitte à changer l'ordre des coordonnées (et donc des colonnes), on peut supposer que ces indices sont  $n - q, \dots, n$  (on met les colonnes en question dans le dernier carré de la matrice). On note alors pour tout  $z$  dans  $V$ ,  $z = (x, y)$  avec  $x \in \mathbb{R}^p$  et  $y \in \mathbb{R}^q$ . La condition de non-dégénérescence devient alors  $\partial_y G(z_0)$  est inversible.

*Remarque 2.3.* Ici, si on note  $z_0 = (x_0, y_0)$ ,  $\partial_x G(z_0)$  et  $\partial_y G(z_0)$  sont les différentielles des applications qui à  $x$  associe  $G(x, y_0)$  et qui à  $y$  associe  $G(x_0, y)$ . Ce sont des applications de  $\mathbb{R}^p$  dans  $\mathbb{R}^q$  et de  $\mathbb{R}^q$  dans  $\mathbb{R}^q$  respectivement. Leurs jacobienes, notées  $J_G^x(z_0)$  et  $J_G^y(z_0)$  sont de taille  $q \times p$  et  $q \times q$  respectivement. On peut écrire la matrice jacobienne de  $G$  en  $z_0$  en fonction de ces deux matrices :

$$J_G(z_0) = \left( J_G^x(z_0) \mid J_G^y(z_0) \right).$$

**Théorème 2.4** (Théorème des fonctions implicites). *Soit  $z_0 = (x_0, y_0)$  un point de  $S$  où  $dG(z_0)$  est de rang  $q$ . Vu ce qu'on vient de dire, on peut supposer que  $\partial_y G(z_0)$  est inversible. Il existe alors  $\delta > 0$ ,  $\varepsilon > 0$  et  $H : B(x_0, \delta) \rightarrow B(y_0, \varepsilon)$  de classe  $C^1$  telle que pour tout élément  $z = (x, y)$  de  $B(x_0, \delta) \times B(y_0, \varepsilon)$  :*

- la différentielle partielle  $\partial_y G(z)$  est inversible,
- $G(z) = 0$  si et seulement si  $y = H(x)$ .

On a alors pour tout  $x \in B(x_0, \delta)$

$$\begin{aligned} dH(x) &= -\left(\partial_y G(x, H(x))\right)^{-1} \circ \partial_x G(x, H(x)), \\ J_H(x) &= -\left(J_G^y(x, H(x))\right)^{-1} \times J_G^x(x, H(x)). \end{aligned}$$

De plus, si  $G$  est de classe  $C^k$ , alors  $H$  l'est également et ses dérivées partielles jusqu'à l'ordre  $k$  s'expriment en fonction de celles de  $G$ .

Un élément de la preuve. Encore une fois, on admet l'existence de  $H$  et on démontre la formule. Pour tout  $x \in B(x_0, \delta)$ ,

$$G(x, H(x)) = 0.$$

En différentiant par rapport à  $x$ , on obtient

$$\partial_x G(x, H(x)) + \partial_y G(x, H(x)) \cdot dH(x) = 0.$$

On conclut alors en utilisant l'inversibilité de  $\partial_y G(x, H(x))$ . □

## 2.4 Un résultat très proche : le théorème d'inversion locale

Le théorème des fonctions implicites permet de démontrer le résultat suivant qui peut également s'avérer utile. En fait, souvent, dans la littérature, c'est plutôt l'inverse qui est fait : on démontre le théorème des fonctions implicites à partir du théorème d'inversion locale. Ce qu'il faut retenir, c'est que l'on passe assez facilement de l'un à l'autre.

**Théorème 2.5** (Théorème d'inversion locale). *Soit  $H$  une application d'un ouvert  $U$  de  $\mathbb{R}^d$  dans  $\mathbb{R}^d$  et  $x_0 \in \mathbb{R}^d$  tel que  $dH(x_0)$  soit inversible. Alors il existe un ouvert  $V$  de  $\mathbb{R}^d$  contenant  $x_0$ , un ouvert  $W$  de  $\mathbb{R}^d$  contenant  $H(x_0)$  et  $K$  un "inverse local de  $H$ ", c'est à dire une application de classe  $C^1$  telle que pour tout  $x \in V$ ,  $K(H(x)) = x$  et pour tout  $y \in W$ ,  $H(K(y)) = y$ .*

*De plus, pour tout  $y \in W$ ,*

$$dK(y) = dH(K(y))^{-1}.$$

*Enfin, si  $H$  est de classe  $C^k$ , alors  $K$  l'est également.*

*Démonstration.* On définit pour tout  $(x, y) \in U \times \mathbb{R}^d$  :

$$G(x, y) = y - H(x).$$

En posant  $x = x_0$  et  $y = H(x_0)$ ,  $G(x_0, y_0) = 0$  et  $\partial_x G(x_0, y_0)$  est inversible. On peut donc appliquer le théorème des fonctions implicites et trouver  $\varepsilon > 0$ ,  $\delta > 0$  et  $K$  une application de  $B(y_0, \varepsilon)$  dans  $B(x_0, \delta)$  telle que pour tout point  $(x, y) \in B(x_0, \delta) \times B(y_0, \varepsilon)$ ,  $\partial_x G(x, y) = -dH(x)$  est inversible, et  $G(x, y) = 0$  si et seulement si  $x = K(y)$ . Donc

$$\forall (x, y) \in B(x_0, \delta) \times B(y_0, \varepsilon), \quad y = H(x) \Leftrightarrow x = K(y).$$

On conclut en choisissant  $W = B(y_0, \varepsilon)$  et  $V = K(W) = H^{-1}(W) \cap B(x_0, \delta)$ , qui est bien ouvert qui contient  $x_0$  : si  $x \in V$ ,  $(x, H(x))$  est dans  $B(x_0, \delta) \times B(y_0, \varepsilon)$ , et donc on a bien  $x = K(H(x))$ , et réciproquement, si  $y \in W$ ,  $(K(y), y)$  est dans  $B(x_0, \delta) \times B(y_0, \varepsilon)$  et donc  $y = H(K(y))$ .

La formule de la différentielle de  $K$  est la formule de la différentielle d'un inverse, et le caractère  $C^k$  de  $K$  est une conséquence du théorème des fonctions implicites. □



# Chapitre 7

## Optimisation sous contraintes

Nous voilà enfin au cœur de ce cours ! On va enfin apprendre à optimiser des fonctions sur des ensembles définis par des équations et des inéquations.

### 1 Optimisation sous contraintes d'égalité

#### 1.1 Le théorème des extrema liés

**Cadre.** Dans ce paragraphe, on se donne  $U$  un ouvert de  $\mathbb{R}^d$  ( $d \geq 1$ ),  $m < d$  et  $G$  une application de  $U$  dans  $\mathbb{R}^m$ . Pour tout  $x$  dans  $U$ , on note  $G(x) = (g_1(x), \dots, g_m(x))$  et d'après ce que l'on a vu au chapitre précédent, on s'attend à ce qu'au moins sous certaines conditions, si  $c \in \mathbb{R}^m$ , l'ensemble  $\{x \mid G(x) = c\} \subset U$  est un ensemble de dimension  $d - m$ . On se donne enfin une fonction  $F$  de  $U$  dans  $\mathbb{R}$ . L'objectif est alors étant donné  $c \in \mathbb{R}^m$  de trouver l'éventuel  $x$  dans  $U$  qui maximise  $F(x)$  tout en satisfaisant  $G(x) = c$ .

Les applications  $F$  et  $G$  seront supposées  $C^1$ . On peut considérer leurs différentielles. On rappelle que pour tout  $u \in U$ , la matrice de  $dG(u)$  dans la base canonique admet les formulations équivalentes suivantes :

$$J_G(u) = \begin{pmatrix} \partial_1 g(u) & \cdots & \cdots & \partial_d g_1(u) \\ \vdots & \ddots & \ddots & \vdots \\ \partial_1 g_m(u) & \cdots & \cdots & \partial_d g_m(u) \end{pmatrix} = \begin{pmatrix} \nabla g_1(u) \\ \vdots \\ \nabla g_m(u) \end{pmatrix},$$

où les gradients sont écrits en ligne.

Pour mettre les choses aux clairs, je vais rappeler ici les définitions de maximiseurs locaux et globaux dans ce cadre.

**Définition 1.1.** Un point de maximum local de  $F$  sous contrainte  $G = c$  est un point  $u \in U$  tel que  $G(u) = c$  et tel qu'il existe  $\varepsilon > 0$  pour lequel quelque soit  $x \in B(u, \varepsilon)$ , si  $G(x) = c$ , alors  $F(x) \leq F(u)$ .

Un point de maximum global de  $F$  sous contrainte  $G = c$  est un point  $u \in U$  tel que  $G(u) = c$  et tel que pour tout  $x \in U$ , si  $G(x) = c$ , alors  $F(x) \leq F(u)$ .

À partir de là, on définit de la même façon que d'habitude les minima locaux et minima globaux, ainsi que les extrema locaux et globaux stricts.

On énonce ici le théorème fondamental de l'optimisation sous contrainte d'égalité, à savoir le théorème des extrema liés.

**Théorème 1.2** (Théorème des extrema liés). *On suppose que  $F$  admet en  $u$  un maximum local sous contrainte  $G = c$ , et que la différentielle  $dG(u)$  est de rang  $m$  (qui est le rang maximal pour une matrice de cette taille). Alors il existe un unique vecteur ligne  $\pi = (\pi_1, \dots, \pi_m)$  tel que*

$$\nabla F(u) = {}^t(\pi \times J_G(u)) = \sum_{i=1}^m \pi_i \nabla g_i(u).$$

*Remarque 1.3.* Si  $m = 1$ , alors la conclusion du théorème est simplement que les vecteurs  $\nabla F(u)$  et  $\nabla G(u)$  sont colinéaires. En dimension supérieure, elle s'interprète de la façon suivante :  $\nabla F(u)$  est dans l'espace engendré par les gradients des contraintes. Comme les gradients des contraintes sont orthogonaux à l'espace satisfaisant les contraintes, cela signifie exactement que les directions dans lesquelles varie  $F$  sont orthogonales à l'espace satisfaisant les contraintes.

**Définition 1.4.** Les nombres  $\pi_1, \dots, \pi_m$  s'appellent les multiplicateurs de Lagrange. Il y en a autant que de contraintes.

*Remarque 1.5.* Remarquons que la condition nécessaire d'optimalité d'une fonction sur un ouvert est un jeu de  $d$  équations à  $d$  inconnues (chacunes des coordonnées du gradient s'annule). Ici, on ajoute  $m$  contraintes, ce qui rajoute  $m$  équations, mais également  $m$  inconnues : les multiplicateurs de Lagrange.

*Preuve du théorème des extrema liés.* D'abord, on se donne le paramètre donné par les hypothèses : soit  $\varepsilon > 0$  tel que pour tout  $x \in B(u, \varepsilon)$ , si  $G(x) = c$ , alors  $F(x) \leq F(u)$ .

Le principe est d'utiliser le théorème des fonctions implicites pour se ramener à un problème **sans contrainte**. Comme  $J_G(u)$  est de rang maximal, on peut supposer sans perte de généralité que ses  $m$  dernières colonnes forment une matrice inversible. On sépare les coordonnées en deux en notant  $u = (v, w)$  où  $v = (u_1, \dots, u_{d-m})$  et  $w = (u_{d-m+1}, \dots, u_d)$  et pour tout  $x \in U$ ,  $x = (y, z)$  avec  $y = (x_1, \dots, x_{d-m})$  et  $z = (x_{d-m+1}, \dots, x_d)$ . Le théorème des fonctions implicites nous garantit l'existence de  $\delta > 0$ ,  $\eta > 0$  et d'une application  $h$  de classe  $C^1$  de  $B(v, \delta)$  dans  $B(w, \eta)$  tel que pour tout  $x = (y, z) \in B(v, \delta) \times B(w, \eta)$ ,  $G(x) = c$  si et seulement si  $z = h(y)$ . De plus, on sait que

$$J_h(v) = -J_G^z(u)^{-1} \times J_G^y(u).$$

Quitte à réduire  $\delta$ , on peut supposer que pour tout  $y \in B(v, \delta)$ ,  $(y, h(y)) \in B(u, \varepsilon)$  (c'est une conséquence de la continuité de  $h$ ).

En conséquence, pour tout  $y \in B(v, \delta)$  :

- $(y, h(y)) \in B(u, \varepsilon)$ ,
- $G(y, h(y)) = c$ ,
- donc  $F(y, h(y)) \leq F(u) = F(v, h(v))$ .

Le point  $v$  est donc un point de maximum de  $y \mapsto F(y, h(y))$  sur  $B(v, \delta)$ . En calculant la différentielle de cette application en  $y = v$ , on a donc

$$J_F^y(u) + J_F^z(u) \times J_h(v) = 0.$$

En remplaçant  $J_h(v)$  par sa valeur, on obtient

$$J_F^y(u) = J_F^z(u) \times J_G^z(u)^{-1} \times J_G^y(u).$$

Par ailleurs, il est évident que

$$J_F^z(u) = J_F^z(u) \times J_G^z(u)^{-1} \times J_G^z(u).$$

On note donc

$${}^t\pi := J_F^z(u) \times J_G^z(u)^{-1}.$$

C'est bien un vecteur ligne de taille  $m$  et par les deux égalités précédentes,

$$\nabla F(u) = {}^t(\pi J_G(u)).$$

Pour ce qui est de l'unicité, c'est une conséquence directe du rang de  $J_G(u)$ . □

## 1.2 Deux exemples

*Exemple 1.6* (Existence d'une valeur propre pour une matrice symétrique). Soit  $A$  une matrice symétrique de taille  $d$ . On va montrer que la solution du problème consistant à maximiser l'application  $F$  de  $\mathbb{R}^d$  dans  $\mathbb{R}$  qui à  $x$  associe

$$\langle x, A \cdot x \rangle$$

sous la contrainte  $G(x) := \|x\|^2 = 1$  est une valeur propre de  $A$ .

D'abord, remarquons que l'ensemble  $S$  satisfaisant la contrainte est la sphère de rayon 1 de  $\mathbb{R}^d$ , qui est compacte. En conséquence,  $F$ , qui est de classe  $C^\infty$ , y atteint bien un maximum. Ensuite,  $G$  est également de classe  $C^\infty$  et pour tout  $x \in \mathbb{R}^d$ ,

$$\nabla G(x) = 2x$$

qui est bien non nul sur  $S$ . Donc la condition de non dégénérescence est bien satisfaite en tout  $x$  de  $S$ . Enfin, calculons la gradient de  $F$  en  $x \in S$ . Pour tout  $h \in \mathbb{R}^d$ , on trouve en utilisant la bilinéarité du produit scalaire et la symétrie de  $A$  :

$$\begin{aligned} F(x+h) &= \langle x+h, A \cdot (x+h) \rangle \\ &= 2\langle A \cdot x, h \rangle + \langle h, A \cdot h \rangle \\ &= 2\langle A \cdot x, h \rangle + o(\|h\|). \end{aligned}$$

Du coup,  $\nabla F(x) = 2A \cdot x$ .

On applique alors le théorème des extrema liés : au point  $u$  de maximum de  $F$  sur  $S$ , il existe un unique nombre  $\pi$  tel que

$$\nabla F(u) = 2A \cdot u = \pi u = \pi \nabla G(u)$$

Donc  $u$  est un vecteur propre de  $A$  associé à la valeur propre  $\pi/2$ . En particulier,  $A$  admet une valeur propre réelle. C'est la première étape de la démonstration du théorème spectral en dimension  $d$ .

*Exemple 1.7* (Trois variables, deux contraintes). Cherchons les extrema de la fonction  $F$  de  $\mathbb{R}^3$  dans  $\mathbb{R}$  qui à  $(x, y, z)$  associe  $xyz$  sous les deux contraintes

$$\begin{cases} x + y + z = 1, \\ x^2 + y^2 + z^2 = 1. \end{cases}$$

Notons  $\mathcal{C}$  l'ensemble des points de  $\mathbb{R}^3$  satisfaisant la contrainte. Tout d'abord, remarquons  $\mathcal{C}$  est compact. Comme  $F$  est continue, elle atteint donc son minimum et son maximum sur  $\mathcal{C}$ . Ensuite on définit  $g_1$  l'application qui à  $(x, y, z)$  associe  $x + y + z$ ,  $g_2$  celle qui à  $(x, y, z)$  associe  $x^2 + y^2 + z^2$  et  $G := (g_1, g_2)$ . Soit  $(x, y, z) \in \mathcal{C}$ . On calcule la jacobienne de  $G$  en  $(x, y, z)$  :

$$J_G(x, y, z) = \begin{pmatrix} 1 & 1 & 1 \\ 2x & 2y & 2z \end{pmatrix}.$$

Cette matrice est de rang deux **sauf** si  $x = y = z$ . Mais alors par la première contrainte,  $x = 1/3$ , et par la seconde,  $x^2 = 1/3$ , ce qui est impossible. Donc  $J_G$  est de rang 2 partout sur  $\mathcal{C}$ . On peut donc appliquer le théorème des extrema liés : soit  $(a, b, c)$  un extremum local de  $F$  sur  $\mathcal{C}$ , il existe un unique couple  $(\mu, \nu)$  tel que

$$\nabla F(a, b, c) = \mu \nabla g_1(a, b, c) + \nu \nabla g_2(a, b, c).$$

Cette égalité donne le système de 5 équations (3 pour cette égalité matricielle plus 2 pour les contraintes) à 5 inconnues (les trois coordonnées plus les deux multiplicateurs de Lagrange) suivant :

$$a + b + c = 1, \tag{7.1}$$

$$a^2 + b^2 + c^2 = 1, \tag{7.2}$$

$$bc = 2\mu a + \nu, \tag{7.3}$$

$$ac = 2\mu b + \nu, \tag{7.4}$$

$$ab = 2\mu c + \nu. \tag{7.5}$$

Si on calcule (7.4) moins (7.3), on obtient

$$(b - a)(2\mu - c) = 0.$$

On obtient également de la même façon

$$\begin{aligned}(c - b)(2\mu - a) &= 0, \\ (a - c)(2\mu - b) &= 0.\end{aligned}$$

Or si  $a = b$ , par (7.1) et (7.2), on a

$$\begin{aligned}2a + c &= 1, \\ 2a^2 + c^2 &= 1,\end{aligned}$$

et donc

$$1 - 2a^2 = (1 - 2a)^2.$$

Cette équation polynomiale d'ordre 2 a pour solution 0 et  $2/3$ .

En conséquence, si  $\mathbf{a} = \mathbf{b}$ , alors ou bien  $a = b = 0$  et alors  $c = 1$ ,  $\nu = 0$ ,  $\mu = 0$  et  $F(a, b, c) = 0$ , ou bien  $a = b = 2/3$  et alors  $c = -1/3$ ,  $\mu = 1/3$ ,  $\nu = -2/3$  et  $F(a, b, c) = -4/27$ .

On trouve les solutions au système correspondant à  $b = c$  et  $a = c$  en permutant les coordonnées.

Enfin si on n'a ni  $a = b$  ni  $b = c$  ni  $c = a$  alors les trois équations que l'on a obtenues donne  $a = b = c = 2\mu$ , et on aboutit donc à une contradiction.

Pour conclure, le maximum de  $F$  sur  $\mathcal{C}$  est 0, atteint en  $(1, 0, 0)$ ,  $(0, 1, 0)$  et  $(0, 0, 1)$ , et le minimum de  $F$  sur  $\mathcal{C}$  est  $-4/27$ , atteint en  $(-1/3, 2/3, 2/3)$ ,  $(2/3, -1/3, 2/3)$  et  $(2/3, 2/3, -1/3)$ .

### 1.3 La formulation lagrangienne

La formulation en termes de Lagrangien de ce qu'on a vu jusqu'ici ne donne aucune nouvelle information, mais peut être un bon moyen pour résoudre mécaniquement des problèmes, selon les tempéraments. Avec les notations des paragraphes précédents, supposons que  $u$  soit un point d'extremum de la fonction  $F$  sous la contrainte  $G = c$ , avec  $dG(u)$  de rang maximal, et soit  $\pi = (\pi_1, \dots, \pi_m)$  les multiplicateurs de Lagrange associés. Le théorème des extrema liés nous donne le jeu de  $d + m$  équations suivant :

$$\left\{ \begin{array}{l} \partial_1 F(u) = \sum_{i=1}^m \pi_i \partial_1 g_i(u), \\ \vdots \\ \partial_d F(u) = \sum_{i=1}^m \pi_i \partial_d g_i(u), \\ g_1(u) = c_1, \\ \vdots \\ g_m(u) = c_m. \end{array} \right.$$

On remarque alors qu'en notant pour tout  $x \in U$  et  $\pi \in \mathbb{R}^m$

$$\mathcal{L}(\pi, x) := F(x) - \sum_{i=1}^m \pi_i (g_i(x) - c_i),$$

ces équations signifient exactement

$$\nabla \mathcal{L}(\pi, x) = 0.$$

En effet, pour tout  $k = 1, \dots, d$ ,

$$\frac{\partial \mathcal{L}}{\partial x_k}(\pi, x) = \partial_k F(x) - \sum_{i=1}^d \pi_i \partial_k g_i(x),$$

et pour tout  $i = 1, \dots, m$ ,

$$\frac{\partial \mathcal{L}}{\partial \pi_i}(\pi, x) = g_i(x) - c_i.$$

Le théorème des extrema liés se reformule donc ainsi.

**Théorème 1.8** (Condition de Lagrange). *Soit  $u$  un point de  $U$  où  $F$  admet un extremum local sous contrainte  $G = c$  et tel que  $J_G(u)$  soit de rang  $m$ . Alors il existe un unique  $\pi \in \mathbb{R}^m$  tel que*

$$\nabla \mathcal{L}(\pi, u) = 0.$$

## 1.4 Le théorème de l'enveloppe

Le théorème de l'enveloppe donne une interprétation aux multiplicateurs de Lagrange : ce sont les dérivées partielles des valeurs de l'optimum par rapport à la valeur de la contrainte.

**Cadre.** On suppose toujours que  $F$  et  $G$  sont de classe  $C^1$  et que l'on cherche à trouver les extrema de la fonction  $F$  sous la contrainte  $G = (g_1, \dots, g_m) = c = (c_1, \dots, c_m)$  à l'intérieur d'un ouvert  $U$ . En revanche, cette fois,  $c$  est susceptible de bouger. Plus précisément, on suppose que l'on sait trouver  $X(c) \in U$  le minimum global de la fonction  $F$  sous contrainte  $G = c$  pour toutes les valeurs de  $c$  dans un ouvert  $V$  de  $\mathbb{R}^m$ . On note la valeur du minimum pour tout  $c \in V$

$$F^*(c) := F(X(c)).$$

On suppose également qu'en  $X(c)$ , la contrainte  $G = c$  est régulière. Pour chaque  $c \in V$ , par le théorème des extrema liés, il existe alors des multiplicateurs de Lagrange  $\pi_1(c), \dots, \pi_m(c)$  tels que

$$\nabla F(X(c)) = \sum_{i=1}^m \pi_i(c) \nabla g_i(X(c)).$$

On peut alors donner une formulation simple du théorème de l'enveloppe de la façon suivante.

**Théorème 1.9** (Théorème de l'enveloppe). *Soit  $c \in V$  tel que  $X$  soit différentiable en  $c$ . Alors  $F^*$  est différentiable en  $c$  et pour tout  $i = 1, \dots, m$ ,*

$$\frac{\partial F^*}{\partial c_i}(c) = \pi_i(c).$$

*Démonstration.* Le fait que  $F^*$  soit différentiable est une conséquence du théorème de différentiation des fonctions composées. Par la formule de la différentielle d'une composée de fonctions :

$$\nabla F^*(c) = {}^t J_X(c) \times \nabla F(X(c)).$$

Par ailleurs,  $G(X(c)) = c$ . Donc en différentiant cette égalité toujours en utilisant le théorème de différentiation des fonctions composées,

$$J_G(X(c)) \times J_X(c) = I_m.$$

Enfin, par le théorème des extrema liés, en notant  $\pi(c) := (\pi_1(c), \dots, \pi_m(c))$ ,

$$\nabla F(c) = {}^t(\pi(c) \times J_G(X(c))) = {}^t J_G(X(c)) \times {}^t \pi(c).$$

En mettant tout ça ensemble :

$$\begin{aligned} \nabla F^*(c) &= {}^t J_X(c) \times {}^t J_G(X(c)) \times {}^t \pi(c) \\ &= {}^t (J_G(X(c)) \times J_X(c)) \times {}^t \pi(c) \\ &= {}^t \pi(c). \end{aligned}$$

C'est la version matricielle du résultat. □

## 2 Optimisation sous contraintes d'inégalité

### 2.1 Les conditions de Kuhn et Tucker

**Cadre.** Comme précédemment, on a affaire à deux fonctions de classe  $C^1$  définies sur un ouvert  $U$  : une fonction valeur  $F$  de  $U$  dans  $\mathbb{R}$  que l'on veut optimiser, et une fonction contrainte  $G$  de  $U$  dans  $\mathbb{R}^m$ . On note encore pour tout  $x \in U$   $G(x) = (g_1(x), \dots, g_m(x))$ . En revanche, cette fois, on peut prendre  $m \geq d$ . Soit  $c = (c_1, \dots, c_m) \in \mathbb{R}^m$  et  $x \in U$ . On note

$$G(x) \leq c$$

si pour tout  $i = 1, \dots, m$ ,  $g_i(x) \leq c_i$ .

L'objectif est alors de trouver l'éventuel  $x \in U$  qui maximise  $F$  tout en satisfaisant  $G(x) \leq c$ . Les définitions précises de ce que sont les extrema locaux dans ce contexte sont des adaptations directes de celle de la section précédente.

Dans cette section comme dans la précédente, on présente une condition nécessaire du premier ordre pour qu'un point  $u \in U$  soit un extremum local.

Heuristiquement, le domaine où l'on veut optimiser, c'est à dire l'ensemble  $\{x \in U \mid G(x) \leq c\}$  est divisé en deux : l'**ouvert**  $\{x \in U \mid G(x) < c\}$  où l'on applique les techniques de l'optimisation **sans contrainte**, et l'ensemble (**fermé**)  $\{x \in U \mid \exists i \in \{1, \dots, m\} \text{ t.q. } g_i(x) = c_i\}$  où le problème en question est proche d'un problème d'optimisation sous contrainte d'égalité. Il y a cependant un élément supplémentaire que je vais expliquer ici.

Imaginons que l'on veuille optimiser une fonction  $F$  de  $\mathbb{R}^2$  dans  $\mathbb{R}$  sous la contrainte  $y \geq 0$  (que l'on écrit  $-y \leq 0$  pour tomber exactement dans le formalisme précédent ; on note  $G(x, y) = -y$  et  $c = 0$ ). Quelles sont les conditions nécessaires pour qu'un point de la forme  $(v, 0)$  soit un maximum local de  $F$  sous contrainte  $y \geq 0$ ? D'abord,  $(v, 0)$  doit être un maximum local de  $F$  sous contrainte  $y = 0$ , sinon, on pourra toujours trouver des points de la forme  $(x, 0)$  aussi proche que l'on veut de  $(v, 0)$  pour lesquels  $F(x, 0) > F(v, 0)$  (l'ensemble satisfaisant  $y = 0$  est inclus dans l'ensemble satisfaisant  $y \geq 0$ , donc la solution sous contrainte d'inégalité doit satisfaire plus de propriétés que la solution sous contrainte d'égalité). En particulier, il existe un multiplicateur de Lagrange  $\pi$  tel que

$$\nabla F(v, 0) = \pi \nabla G(v, 0) = (0, -\pi).$$

Le gradient de  $F$  en  $(v, 0)$  est donc vertical. Mais on peut aller plus loin : le gradient de  $F$  en  $(v, 0)$  ne peut pointer que vers le bas, c'est à dire vers l'extérieur du domaine. En d'autres termes, on ne peut avoir que  $\pi \geq 0$ . Sinon, pour  $w > 0$  petit,  $(v, w)$  satisfait encore la contrainte et

$$F(v, w) \approx F(v, 0) + w \partial_y F(v, 0) > F(v, 0).$$

Donc la condition d'inégalité aux points où il y a égalité se ramène à une condition d'égalité **plus** une condition de signe sur les multiplicateurs de Lagrange. C'est exactement ce qu'énoncent les conditions de Kuhn et Tucker que l'on donne un peu plus bas. Commençons par une définition.

**Définition 2.1** (Contraintes actives et contraintes régulières). On dit que la  $k$ -ième contrainte est active en  $x \in U$  si  $g_k(x) = c_k$ . On dit que les contraintes sont régulières en  $x \in U$  si le rang de la différentielle des contraintes actives en  $x$  est égal au nombre de ces contraintes. En d'autres termes, les contraintes sont régulières en  $x$  si en notant

$$\Lambda := \{k \mid \text{la contrainte } k \text{ est active en } x\}$$

et  $G^\Lambda$  l'application qui à  $y \in U$  associe  $(g_k(y))_{k \in \Lambda}$ , la différentielle de  $G^\Lambda$  est de rang maximal.

**Théorème 2.2** (Conditions de Kuhn & Tucker, admis). *Si  $u$  est un maximum local de  $F$  avec contrainte  $G \leq c$  et si les contraintes en  $u$  sont régulières, alors il existe un unique  $\pi = (\pi_1, \dots, \pi_m)$  tel que :*

•

$$\nabla F(u) = {}^t(\pi J_G(u)) = \sum_{k=1}^m \pi_k \nabla g_k(u),$$

- si la contrainte  $k$  n'est pas active en  $u$ , alors  $\pi_k = 0$ ,
- pour tout  $k$ ,  $\pi_k \geq 0$ .

Remarque 2.3. Il est souvent avantageux d'écrire à la place de la deuxième ligne : pour tout  $k$ ,

$$\pi_k(g_k(x) - c_k) = 0.$$

Si on cherche un minimum local, il faut changer les signes des multiplicateurs de Lagrange.

On peut à nouveau reformuler ce théorème en termes de Lagrangien. En définissant le même Lagrangien : pour tout  $x \in U$  et tout  $\pi \in \mathbb{R}^m$ ,

$$\mathcal{L}(\pi, x) := F(x) - \sum_{i=1}^m \pi_i(g_i(x) - c_i),$$

on a le théorème suivant.

**Théorème 2.4.** Soit  $u \in U$  un point de maximum local de  $F$  sous condition  $G \leq c$  et si les contraintes en  $u$  sont régulières, alors il existe un unique  $\pi \in \mathbb{R}^m$  satisfaisant

- pour tout  $k = 1, \dots, d$ ,

$$\frac{\partial \mathcal{L}}{\partial x_i}(\pi, x) = 0,$$

- pour tout  $i = 1, \dots, m$ ,

$$\pi_i \geq 0,$$

- pour tout  $i = 1, \dots, m$ ,

$$\pi_i \partial_{\pi_i} \mathcal{L}(\pi, x) = 0.$$

Remarque 2.5 (Contraintes mixtes). Une fois que l'on a fait tout ça, on peut aisément adapter nos résultats au cas où on a des contraintes mixtes, c'est à dire des contraintes d'égalité et des contraintes d'inégalité. On veut maximiser  $F$  sous l'ensemble de contraintes

$$\begin{aligned} g_1(x) = b_1, \dots, g_m(x) = b_m, \\ h_1(x) \leq c_1, \dots, h_p(x) \leq c_p. \end{aligned}$$

Soit  $x$  un maximiseur local de  $F$  sous ces contraintes.

**Condition de qualification.** Notons  $1 \leq i_1 < \dots < i_q \leq p$  les indices pour lesquels  $h_i(x) = c_i$ , c'est à dire pour lesquels la contrainte est active. La condition de qualification est alors que la famille

$$\left( \nabla g_1(x), \dots, \nabla g_m(x), \nabla h_{i_1}(x), \dots, \nabla h_{i_q}(x) \right)$$

soit de rang  $m + q$ .

**Système d'équations.** Sous la condition de qualification, la condition nécessaire du premier ordre pour que  $x$  soit un maximiseur de  $F$  sous contrainte est la suivante : il existe des multiplicateurs de Lagrange  $\pi_1, \dots, \pi_m$  et  $\lambda_1, \dots, \lambda_p$  tels que

$$\begin{cases} \nabla F(x) = \sum_{i=1}^m \pi_i \nabla g_i(x) + \sum_{i=1}^p \lambda_i \nabla h_i(x), \\ \lambda_1 \geq 0, \dots, \lambda_p \geq 0, \\ \forall i = 1, \dots, p, \quad \lambda_i (h_i(x) - c_i) = 0, \\ \forall i = 1, \dots, m, \quad g_i(x) = b_i. \end{cases}$$

## 2.2 Un exemple

On va traiter un exemple en prenant la méthode lagrangienne pour être complet.  
On souhaite minimiser la fonction

$$F : (x, y) \in \mathbb{R}^2 \mapsto x^2 - 2y$$

sous les contraintes

$$\begin{cases} x \geq 0, \\ y \geq 0, \\ x^2 + y^2 \leq 1. \end{cases}$$

**Existence d'un minimum sous contraintes.** On remarque d'abord que l'ensemble satisfaisant les contraintes est compact, donc le problème admet une solution.

On pose ensuite  $G : (x, y) \in \mathbb{R}^2 \mapsto (-x, -y, x^2 + y^2)$  et  $c = (0, 0, 1)$ . Comme les fonctions  $F$  et  $G$  sont  $C^1$ , on est exactement dans le cadre précédent. On définit le lagrangien pour tout  $(\lambda, \mu, \nu, x, y) \in \mathbb{R}^5$  :

$$\mathcal{L}(\lambda, \mu, \nu, x, y) = x^2 - 2y + \lambda x + \mu y - \nu(x^2 + y^2 - 1).$$

**Condition de qualification** Vérifions d'abord que les contraintes sont régulières en tout point  $(x, y)$  satisfaisant les contraintes. Soit donc  $(x, y)$  tel que  $x \geq 0$ ,  $y \geq 0$  et  $x^2 + y^2 \leq 1$ . On a

$$J_G(x, y) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 2x & 2y \end{pmatrix}.$$

Les trois contraintes ne peuvent clairement pas être actives en même temps. Si  $x$  est non nul, alors les deux gradients correspondant aux contraintes éventuellement actives sont  $(0, -1)$  et  $(2x, 2y)$  qui sont linéairement indépendants. On peut faire le même raisonnement si  $y$  est non nul. Enfin, si  $x$  et  $y$  sont nuls, les deux vecteurs correspondant aux contraintes actives sont  $(-1, 0)$  et  $(0, -1)$  qui sont linéairement indépendants. Donc les contraintes sont tout le temps régulières.

**Le système d'équations.** Ensuite, on cherche les points critiques de  $\mathcal{L}$  avec multiplicateurs de Lagranges positifs et  $(x, y)$  satisfaisant les contraintes (à chaque solution du problème correspond un point critique de  $\mathcal{L}$  de ce type). Cela nous mène au jeu d'équations et d'inéquations suivant :

$$\begin{cases} x \geq 0, \\ y \geq 0, \\ x^2 + y^2 \leq 1, \\ 2x + \lambda - 2x\nu = 0, \\ -2 + \mu - 2y\nu = 0, \\ \lambda x = 0, \\ \mu y = 0, \\ \nu(x^2 + y^2 - 1) = 0, \\ \lambda, \mu, \nu \leq 0. \end{cases}$$

En multipliant la quatrième ligne par  $\lambda$  et en utilisant  $\lambda x = 0$ , on obtient  $\lambda^2 = 0$ , donc  $\lambda = 0$ .

La quatrième ligne devient alors  $2x(1 - \nu) = 0$ . Comme  $\nu \leq 0$ , on a nécessairement  $x = 0$ . La première contrainte est donc active.

En multipliant la troisième ligne par  $\mu$  et en utilisant  $\mu y = 0$ , on obtient  $\mu^2 = 2\mu$ . Comme  $\mu \leq 0$ , on a nécessairement  $\mu = 0$ .

le système devient donc

$$\begin{cases} y \geq 0, \\ y^2 \leq 1, \\ y\nu = -1, \\ \nu y^2 = \nu, \\ \nu \leq 0. \end{cases}$$

Par la troisième et la quatrième ligne,  $\nu = -y$ , puis comme  $\nu y = -1$ ,  $\nu = -1$  et  $y = 1$ .

L'unique point critique de  $\mathcal{L}$  dans l'espace admissible est donc  $(0, 0, -1, 0, 1)$ . Or d'une part, chaque solution de notre problème doit nous en fournir un, et d'autre part, on sait qu'il existe une solution à notre problème. Donc cette solution est unique et correspond à  $(x, y) = (0, 1)$ . La valeur correspondante de  $F$  est  $-2$ .

### 2.3 Quand il y a des contraintes de positivité

Bien souvent en économie, les coordonnées modélisent des quantités de biens ou des prix etc... qui ne peuvent être que positives. On a donc affaire en premier lieu aux contraintes  $x_1 \geq 0, \dots, x_d \geq 0$ . Dans ce cas, on va voir que l'on peut se passer des multiplicateurs de Lagrange associés à ces contraintes. Pour voir ça, on va écrire les conditions de Kuhn et Tucker que l'on a déjà vues, puis on va modifier le système obtenu pour se débarrasser des multiplicateurs en question.

Imaginons que l'on veuille maximiser  $F$  sous contraintes

$$\begin{aligned} x_1 \geq 0, \dots, x_d \geq 0, \\ g_1(x) \leq c_1, \dots, g_m(x) \leq c_m. \end{aligned}$$

Écrivons d'abord la condition de qualification dans ce cadre.

**Condition de qualification.** Soit  $x$  satisfaisant les contraintes. Notons  $1 \leq i_1 < \dots < i_p \leq d$  les indices pour lesquels  $x_i = 0$  (les contraintes de positivités qui sont actives) et  $1 \leq j_1 < \dots < j_q \leq m$  (les autres contraintes qui sont actives). Par ailleurs si  $i = 1, \dots, d$ , on note  $e_i$  le vecteur  $(0, \dots, 0, 1, 0, \dots, 0)$  où le 1 est à la  $i$ -ième position.

Si pour  $i = 1, \dots, d$  on note  $h_i(x) = -x_i$ , la  $i$ -ième contrainte de positivité s'écrit  $h_i(x) \leq 0$ , et  $\nabla h_i(x) = -e_i$ . La condition de qualification est alors

$$\text{Rg} \left( -e_{i_1}, \dots, -e_{i_p}, \nabla g_{j_1}, \dots, \nabla g_{j_q} \right) = p + q,$$

soit de façon équivalente

$$\text{Rg} \left( e_{i_1}, \dots, e_{i_p}, \nabla g_{j_1}, \dots, \nabla g_{j_q} \right) = p + q.$$

Mais les  $p$  premiers vecteurs sont bien linéairement indépendants et engendrent tous les vecteurs ayant des coordonnées nulles pour tout  $i \notin \{i_1, \dots, i_p\}$ . En voyant cette famille de vecteurs comme une matrice, et en utilisant la méthode de Gauss pour annuler les termes des lignes  $i_1, \dots, i_p$  sur les  $q$  dernières colonnes, on se rend compte que cette condition est équivalente à ce que

$$\text{Rg} \left( \partial_i g_j(x) \right)_{i \notin \{i_1, \dots, i_p\}, j \in \{j_1, \dots, j_q\}} = q.$$

Et ça, ça se dit "le rang de la matrice composées des dérivées partielles des contraintes **actives** dans les directions où les contraintes de positivité **ne sont pas actives** est de rang maximal".

Passons maintenant au système d'équations que l'on obtient dans ce cadre.

**Système d'équations.** Si on écrit tout le système de Kuhn et Tucker, on obtient l'existence de multiplicateurs  $\lambda_1, \dots, \lambda_d$  et  $\pi_1, \dots, \pi_m$  tels que (attention au  $-$  lié au fait que les contraintes de positivité sont écrites avec un  $\geq$  et non avec un  $\leq$ )

$$\left\{ \begin{array}{l} \nabla F(x) = -\sum_{i=1}^d \lambda_i e_i + \sum_{i=1}^m \pi_i \nabla g_i(x), \\ \lambda_1 \geq 0, \dots, \lambda_d \geq 0, \\ \pi_1 \geq 0, \dots, \pi_m \geq 0, \\ \forall i = 1, \dots, d, \quad \lambda_i x_i = 0, \\ \forall i = 1, \dots, m, \quad \pi_i (g_i(x) - c_i) = 0. \end{array} \right.$$

Mais si on regarde la  $k$ -ème coordonnée dans la première équation, cela donne

$$\partial_k F(x) - \sum_{i=1}^m \pi_i \partial_i g_k(x) = -\lambda_k.$$

Donc d'une part, comme  $\lambda_k \geq 0$ ,

$$\partial_k F(x) - \sum_{i=1}^m \pi_i \partial_i g_k(x) \leq 0,$$

et d'autre part, comme  $\lambda_k x_k = 0$ ,

$$\left( \partial_k F(x) - \sum_{i=1}^m \pi_i \partial_i g_k(x) \right) x_k = 0.$$

On peut donc se débarrasser des  $\lambda$  dans le système précédent et ainsi obtenir

$$\left\{ \begin{array}{l} \forall k = 1, \dots, d, \quad \partial_k F(x) - \sum_{i=1}^m \pi_i \partial_k g_i(x) \leq 0, \\ \pi_1 \geq 0, \dots, \pi_m \geq 0, \\ \forall k = 1, \dots, d, \quad \left( \partial_k F(x) - \sum_{i=1}^m \pi_i \partial_k g_i(x) \right) x_k = 0, \\ \forall i = 1, \dots, m, \quad \pi_i (g_i(x) - c_i) = 0. \end{array} \right.$$

Réciproquement, si on a une solution de ce système, alors il suffit de définir pour tout  $k = 1, \dots, d$

$$\lambda_k := -\left( \partial_k F(x) - \sum_{i=1}^m \pi_i \partial_k g_i(x) \right)$$

pour obtenir une solution du système précédent. Ces deux systèmes sont donc équivalents et on peut travailler avec le second.

## 2.4 Un exemple avec ce nouveau système

On cherche à maximiser la fonction (de classe  $C^\infty$  sur  $\mathbb{R}^3$ )

$$F : (x, y, z) \mapsto xyz + z,$$

sous les contraintes (de classe  $C^\infty$  également)

$$\begin{array}{l} x \geq 0, y \geq 0, z \geq 0, \\ x^2 + y^2 + z \leq 6. \end{array}$$

On note  $g$  l'application qui à  $x, y$  et  $z$  associe  $x^2 + y^2 + z$  et  $\mathcal{C}$  l'ensemble des points de  $\mathbb{R}^3$  satisfaisant les quatre contraintes.

**Existence d'un maximum sous contraintes.**  $F$  est continue, il suffit donc de montrer que  $\mathcal{C}$  est compact. Il est fermé car  $g$  et les applications coordonnées sont continues. Il est borné car si  $(x, y, z) \in \mathcal{C}$ , alors comme  $z \geq 0$ ,  $x^2 + y^2 \leq 6$ , et comme  $x^2 + y^2 \geq 0$ ,  $0 \leq z \leq 6$  donc  $z^2 \leq 36$ . En conséquence, on peut majorer la norme de  $(x, y, z)$  grâce à l'estimation  $x^2 + y^2 + z^2 \leq 6 + 36 = 42$ .

**Condition de qualification.** On a pour tout  $x, y$  et  $z$

$$\nabla g(x, y, z) = \begin{pmatrix} 2x \\ 2y \\ 1 \end{pmatrix}.$$

Ce gradient a la particularité suivante : pour  $i = 1, 2$  ou  $3$ , si la  $i$ -ème coordonnée de  $(x, y, z)$  est non nulle, alors la  $i$ -ème coordonnée du gradient est non nulle. Donc la condition de qualification est satisfaite.

**Système du premier ordre.** Pour tout  $(x, y, z) \in \mathbb{R}^3$ ,

$$\nabla F(x, y, z) = \begin{pmatrix} yz \\ xz \\ xy + 1 \end{pmatrix}.$$

La condition du premier ordre pour que  $(x, y, z) \in \mathcal{C}$  soit un maximiseur local sous contrainte donne donc l'existence d'un multiplicateur  $\pi \geq 0$  tel que

$$\begin{cases} yz - 2\pi x \leq 0, \\ xz - 2\pi y \leq 0, \\ xy + 1 - \pi \leq 0, \\ \pi(x^2 + y^2 + z - 6) = 0, \\ xyz = 2\pi x^2, \\ xyz = 2\pi y^2, \\ (xy + 1 - \pi)z = 0. \end{cases}$$

D'abord, la troisième ligne donne  $\pi \geq 1 + xy \geq 1$  car  $x$  et  $y$  sont positifs. Donc  $\pi$  est non nul et la contrainte associée à  $g$  est active, c'est à dire

$$x^2 + y^2 + z = 6.$$

Dès lors que  $\pi \neq 0$ , les 5 et 6ème lignes nous montrent que  $x^2 = y^2$ , et comme ils sont positifs,

$$x = y.$$

Ensuite, par la dernière ligne,  $z = 0$  ou  $\pi = 1 + xy = 1 + x^2$ . Mais si  $z = 0$ , alors  $x = y = 0$  par les 5 et 6ème lignes, ce qui contredit la contrainte associée à  $g$ . Donc

$$\pi = 1 + x^2.$$

Il ne nous reste donc qu'à déterminer  $x$  et  $z$ , mais on a justement deux équations :

$$\begin{aligned} x^2 + y^2 + z &= 6, \\ xyz &= 2\pi y^2. \end{aligned}$$

En utilisant les équations préalablement trouvées, on en déduit

$$\begin{aligned} 2x^2 + z &= 6, \\ x^2(z - 2(1 + x^2)) &= 0. \end{aligned}$$

Par la deuxième équation, ou bien  $x = 0$ , ou bien  $z = 2(1 + x^2)$ . Si  $x = 0$ , alors  $z = 6$  et le point en question est  $(0, 0, 6)$ . Si  $z = 2(1 + x^2)$ , alors en injectant dans la première équation,

$$2x^2 + 2(1 + x^2) = 6,$$

donc

$$4x^2 = 4,$$

et donc  $x = 1$ . Le point en question est donc  $(1, 1, 4)$ . Ces deux points sont donc les seuls candidats au statut de maximum de  $F$  sous contraintes. Or on sait qu'il existe un maximiseur,

$$F(0, 0, 6) = 6 \quad \text{et} \quad F(1, 1, 4) = 8.$$

Donc l'unique point de maximum de  $F$  sous contrainte est le point  $(1, 1, 4)$ . En ce point,  $F$  vaut 8.

## Chapitre 8

# Éléments d'analyse des fonctions convexes

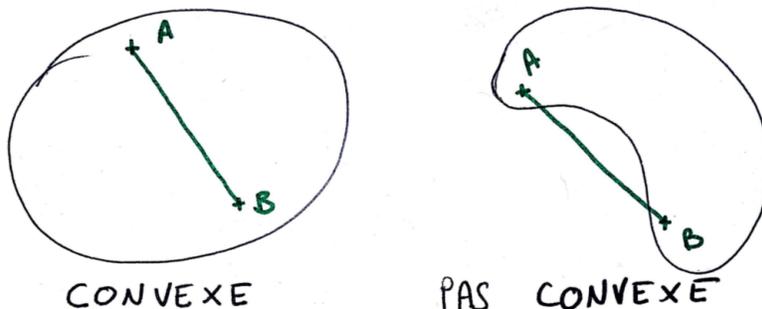
Comme on va le voir au cours de ce chapitre, le cadre convexe est le cadre idéal pour résoudre des problèmes de minimisation : beaucoup de résultats se simplifient si la fonction à minimiser est convexe. D'abord, les fonctions convexes héritent naturellement de propriétés de régularité : on verra par exemple que les fonctions convexes sont automatiquement continues à l'intérieur de leur domaine de définition, mais on pourrait dire beaucoup plus ! Par exemple des théorèmes plus subtils permettent de donner un sens au calcul différentiel d'ordre 1 dans le cas convexe sans hypothèse supplémentaire. En outre, dans le cas convexe, toutes les conditions nécessaires que l'on a développées tout au long de l'année deviennent des conditions suffisantes. Le dernier point crucial qui explique le développement de l'analyse convexe est la notion de problème dual dont nous développerons un exemple. Commençons par les définitions.

### 1 Définitions

Avant de définir la notion de fonction convexe, il faut dire ce qu'est un ensemble convexe : ce sont les ensembles sur lesquels on va pouvoir définir les fonctions convexes.

**Définition 1.1.** Soit  $C$  un sous-ensemble de  $\mathbb{R}^d$ . On dit que  $C$  est convexe si pour tout  $x$  et  $y$  de  $C$ , et tout  $t \in [0, 1]$ ,  $(1 - t)x + ty$  est dans  $C$ .

Quand  $t$  parcourt  $[0, 1]$ ,  $(1 - t)x + ty$  parcourt le segment joignant  $x$  à  $y$ . Cette définition signifie donc que pour tout  $x$  et  $y$  dans  $C$ , l'ensemble du segment joignant  $x$  à  $y$  est inclus dans  $C$ .



Donnons d'emblée un lemme qui pourra nous être utile par la suite.

*Lemme 1.2.* Soit  $(C_\alpha)_{\alpha \in I}$  une famille de sous-ensemble convexe de  $\mathbb{R}^d$ . Alors

$$C := \bigcap_{\alpha \in I} C_\alpha$$

est un sous-ensemble convexe de  $\mathbb{R}^d$ .

*Démonstration.* Soient  $x$  et  $y$  dans  $C$ ,  $t \in [0, 1]$  et  $\alpha \in I$ . Les points  $x$  et  $y$  appartiennent en particulier à  $C_\alpha$ . Par convexité de  $C_\alpha$ ,  $(1-t)x + ty \in C_\alpha$ . Comme c'est vrai pour tout  $\alpha$ , on a le résultat.  $\square$

Sur un ensemble convexe, on peut alors aisément définir ce qu'est une fonction convexe.

**Définition 1.3.** Soit  $C$  un sous-ensemble convexe de  $\mathbb{R}^d$ . On dit que la fonction  $f$  de  $C$  dans  $\mathbb{R}$  est convexe si pour tous  $x$  et  $y$  dans  $C$ , et tout  $t \in [0, 1]$  :

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

Cela a bien un sens car  $(1-t)x + ty \in C$ .

Si l'inégalité est vérifiée dans l'autre sens, on dit que la fonction est concave.

Cette définition signifie la chose suivante : si  $f$  est une fonction convexe sur  $C$  et si  $x$  et  $y$  appartiennent à  $C$ , alors le long du segment  $[x, y]$  (qui est bien dans  $C$ ),  $f$  est inférieure aux fonctions affines qui coïncident avec  $f$  en  $x$  et  $y$ .

## 2 Propriétés en dimension 1

Regardons déjà ce qu'il se passe en dimension 1.

*Remarque 2.1.* • En dimension 1, les ensembles convexes sont exactement les intervalles de  $\mathbb{R}$ .

- L'inégalité de convexité signifie dans ce cas que si  $a < b$  sont deux points du domaine de définition de  $f$ , et si on regarde la droite de  $\mathbb{R}^2$  passant par  $(a, f(a))$  et  $(b, f(b))$ , au dessus de l'intervalle  $[a, b]$ , le graphe de  $f$  est en dessous de cette droite. En d'autre terme, une fonction convexe est en dessous de sa corde à l'intérieur de l'intervalle entre les points qui définissent cette corde.

### 2.1 Croissance du taux d'accroissement

La proposition suivante est une des propriétés les plus utiles des fonctions convexes en dimension 1, et qui a même des conséquences en dimension quelconque.

**Proposition 2.2** (Monotonie du taux d'accroissement).  *$f$  est convexe si et seulement si pour tout  $a < b < c$  dans  $I$ , on a :*

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(a)}{c - a}.$$

Dans ce cas, on a aussi :

$$\frac{f(c) - f(a)}{c - a} \leq \frac{f(c) - f(b)}{c - b}$$

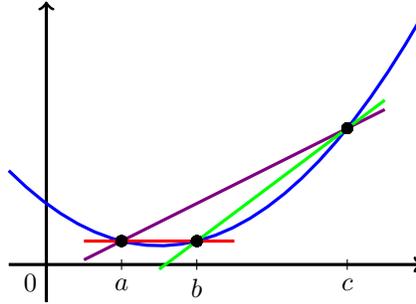


FIGURE 8.1 – Dans l'exemple ci-dessus, la pente de la droite rouge est bien inférieure à celle de la droite violette, qui est elle-même inférieure à celle de la droite verte.

*Démonstration.* **Si  $f$  est convexe.**

On a une fonction convexe et trois points, donc on n'a à peu près qu'une chose à faire : on dit que la valeur de la fonction au point du milieu est inférieure au point de la corde de même abscisse : comme  $b \in [a, c]$ , il existe  $t \in [0, 1]$  tel que :

$$b = (1 - t)a + tc.$$

On sait même que :

$$t = \frac{b - a}{c - a}$$

Comme  $f$  est convexe :

$$f(b) \leq (1 - t)f(a) + tf(c).$$

Et en trifouillant un tout petit peu :

$$f(b) - f(a) \leq t(f(c) - f(a)) = \frac{b - a}{c - a}(f(c) - f(a))$$

Donc on a bien :

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(a)}{c - a}.$$

Pour l'autre inégalité, il suffit de remarquer que :

$$(1 - t) = \frac{c - b}{c - a}$$

et qu'en trifouillant d'une façon un peu différente l'inégalité de convexité :

$$f(c) - f(b) \geq (1 - t)(f(c) - f(a)) = \frac{c - b}{c - a}(f(c) - f(a)).$$

On a donc bien :

$$\frac{f(c) - f(a)}{c - a} \leq \frac{f(c) - f(b)}{c - b}.$$

**Si l'inégalité est vérifiée.**

Pour vérifier qu'une fonction est convexe, il faut prendre  $a < c$  dans  $I$  et  $t \in [0, 1]$ , puis évaluer  $f((1-t)a+tc)$ . Donc on choisit des  $a$ ,  $c$  et  $t$  comme ça, et on note :

$$b := (1 - t)a + tc.$$

On a encore :

$$t = \frac{b - a}{c - a}$$

Par hypothèse :

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(a)}{c - a},$$

et donc :

$$\begin{aligned} f(b) &\leq f(a) + \frac{b - a}{c - a} (f(c) - f(a)) \\ &= f(a) + t(f(c) - f(a)) \\ &= (1 - t)f(a) + tf(c). \end{aligned}$$

□

*Remarque 2.3.* En particulier, une réécriture de la première inégalité est

$$f(c) \geq f(a) + \frac{f(b) - f(a)}{b - a} \cdot (c - a),$$

$f(c)$  est donc au dessus de la droite de  $\mathbb{R}^2$  passant par  $(a, f(a))$  et  $(b, f(b))$ . On aurait pu faire le même raisonnement pour montrer que  $f(a)$  est au dessus de la droite passant par  $(b, f(b))$  et  $(c, f(c))$ . En d'autres termes, une fonction convexe est au dessus de sa cordes à l'extérieur de l'intervalle entre les points qui définissent cette corde.

**Proposition 2.4** (Conséquence quand  $f$  est dérivable en un point). *On suppose  $f$  convexe et dérivable en un point  $b \in I$ . Si  $a$  est un élément de  $I$  avec  $a < b$ , alors :*

$$\frac{f(b) - f(a)}{b - a} \leq f'(b).$$

De même, si  $c$  est un élément de  $I$  avec  $b < c$ , alors :

$$f'(b) \leq \frac{f(c) - f(b)}{c - b}.$$

En particulier,  **$f$  est au dessus de sa tangente en  $b$**  : si  $x$  est n'importe quel point de  $I$ ,

$$f(x) \geq f(b) + f'(b)(x - b).$$

*Démonstration.* On montre la première inégalité, la deuxième s'obtenant de façon tout à fait similaire. On suppose donc qu'il existe  $a \in I$  tel que  $a < b$ . Soit  $(a_n)_{n \in \mathbb{N}}$  une suite croissant strictement vers  $b$ , avec  $a < a_0$ . Par la proposition précédente, pour tout  $n \in \mathbb{N}$  :

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(b) - f(a_n)}{b - a_n}.$$

En passant à la limite  $n \rightarrow +\infty$ , on obtient le résultat.

Montrons maintenant que  $f$  est au dessus de sa tangente en  $b$ . Soit par exemple  $c > b$  (le cas  $a < b$  se traite de façon similaire). L'inégalité :

$$f'(b) \leq \frac{f(c) - f(b)}{c - b}$$

se reformule :

$$f(c) \geq f(b) + f'(b)(c - b),$$

ce qui est exactement l'inégalité demandée (l'équation de la tangente à  $f$  en  $b$  est  $y = f(b) + f'(b)(x - b)$ ). □

**Théorème 2.5** (Conséquence quand  $f$  est dérivable sur tout  $I$ ). *Si  $f$  est dérivable sur tout  $I$ , alors  $f$  est convexe si et seulement si  $f'$  est croissante sur  $I$ .*

*Démonstration. Si  $f$  est convexe.*

Soient  $a$  et  $b$  deux points de  $I$  avec  $a < b$ . Par la proposition précédente :

$$f'(a) \leq \frac{f(b) - f(a)}{b - a} \leq f'(b),$$

donc  $f'$  est croissante.

**Si  $f'$  est croissante.**

D'abord, si  $f'$  est croissante, alors  $f$  est au dessus de ses tangentes. Soient par exemples  $b < c$  deux points de  $I$ . On va montrer que :

$$f(c) \geq f(b) + f'(b)(c - b),$$

le cas symétrique se traitant de façon analogue.  $f$  est continue sur  $]b, c[$ , dérivable sur  $]b, c[$ , et par croissance de  $f'$ , pour tout  $x \in ]b, c[$  :

$$f'(x) \geq f'(b).$$

L'inégalité recherchée est donc une application directe de l'inégalité des accroissements finis.

Montrons maintenant que  $f$  est convexe. Soient  $a < c$  deux éléments de  $I$  et  $t \in [0, 1]$ . On pose :

$$b := (1 - t)a + tc.$$

Comme  $f$  est au dessus de sa tangente en  $b$  :

$$f(a) \geq f(b) - f'(b)(b - a),$$

$$f(c) \geq f(b) + f'(b)(c - b).$$

En faisant  $(1 - t) \times$ (la première équation) $+t \times$ (la seconde) et en utilisant :

$$t = \frac{b - a}{c - a}, \quad (1 - t) = \frac{c - b}{c - a},$$

on obtient :

$$(1 - t)f(a) + tf(c) \geq f(b).$$

□

**Théorème 2.6** (Conséquence lorsque  $f$  est deux fois dérivable sur tout  $I$ ). *Si  $f$  est deux fois dérivable sur tout  $I$ , alors  $f$  est convexe si et seulement si :*

$$\forall x \in I, f''(x) \geq 0.$$

*Démonstration.* Si  $f$  est deux fois dérivable,  $f'$  est croissante si et seulement si  $f''$  est positive. Donc on peut appliquer le résultat précédent. □

## 2.2 Continuité

La convexité engendre nécessairement de la régularité. Dans le théorème suivant, on parle de continuité. L'ensemble  $I$  est un intervalle de  $\mathbb{R}$  et  $f : I \rightarrow \mathbb{R}$  est une fonction convexe.

**Théorème 2.7.** *Soit  $x$  un point dans l'intérieur de  $I$ . Alors  $f$  est continue en  $x$ .*

*Démonstration.* Comme  $x$  est dans l'intérieur de  $I$ , on peut trouver  $a$  et  $b$  dans  $I$  tels que  $a < x < b$ . On note

$$M := \max \left( \left| \frac{f(a) - f(x)}{a - x} \right|, \left| \frac{f(b) - f(x)}{b - x} \right| \right).$$

On va montrer que pour tout  $y \in [a, b]$ , on a

$$|f(y) - f(x)| \leq M|y - x|.$$

C'est évidemment suffisant pour que  $f$  soit continue en  $x$  (prendre  $\delta := \varepsilon/M$  dans la définition de la continuité).

Soit  $y \in [a, b]$ . Si  $y > x$ , alors par croissance du taux d'accroissement,

$$-M \leq \frac{f(x) - f(a)}{x - a} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(b) - f(x)}{b - x} \leq M.$$

En passant à la valeur absolue, on a bien

$$\left| \frac{f(y) - f(x)}{y - x} \right| \leq M,$$

et donc l'inégalité annoncée. On peut faire le même genre de raisonnement si  $y < x$  et donc on peut conclure.  $\square$

On pourrait aller beaucoup plus loin dans l'analyse de la régularité des fonctions convexes en dimension 1. Je cite ici les résultats essentiels :

- les fonctions convexes admettent des dérivées à gauche et à droite en tous leurs points de continuité (donc partout sauf au bord),
- elles sont localement lipschitziennes (leurs taux d'accroissements sont localement bornés, comme vu dans la preuve du théorème de continuité),
- elles sont dérivables partout sauf en une quantité dénombrable de points,
- elles sont deux fois dérivables presque partout au sens de la théorie de la mesure et leur dérivée seconde est positive partout où elle existe.

### 3 Généralisation en dimension supérieure

On choisit maintenant  $C$  un sous-ensemble convexe de  $\mathbb{R}^d$ . Voyons comment se généralisent les propriétés du paragraphe précédent. La première chose dont j'ai envie de parler est l'inégalité de convexité pour plus de deux points. Cette inégalité est déjà intéressante en dimension 1 mais est peut être plus visuelle quand on y pense dans des exemples multidimensionnels.

**Proposition 3.1.** *Soient  $x_1, \dots, x_n$  des points de  $C$  et  $\lambda_1, \dots, \lambda_n$  des réels de  $[0, 1]$  tels que*

$$\lambda_1 + \dots + \lambda_n = 1.$$

*Alors  $\lambda_1 x_1 + \dots + \lambda_n x_n \in C$  et*

$$f(\lambda_1 x_1 + \dots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \dots + \lambda_n f(x_n).$$

*La valeur de  $f$  en un barycentre des points  $x_1, \dots, x_n$  est inférieure ou égale à la moyenne pondérée des valeurs de  $f$  aux points  $x_1, \dots, x_n$ .*

*Exemple 3.2.* Par exemple, en dimension 2, la valeur de  $f$  au centre d'un carré est inférieure ou égale à la moyenne de ses valeurs sur les sommets du carré.

*Démonstration.* On montre ce résultat par récurrence. Il est vrai par définition si  $n = 2$ . Si il est vrai pour un certain  $n$  soient  $x_1, \dots, x_{n+1}$  dans  $C$  et  $\lambda_1, \dots, \lambda_{n+1}$  dans  $[0, 1]$  de somme 1. Il faut d'abord montrer que

$$\lambda_1 x_1 + \dots + \lambda_{n+1} x_{n+1} \in C.$$

Si  $\lambda_{n+1} = 0$ , c'est une conséquence de l'hypothèse de récurrence. Sinon, on note

$$m := \lambda_1 + \dots + \lambda_n = 1 - \lambda_{n+1} < 1$$

et pour  $i = 1, \dots, n$ ,  $\mu_i := \lambda_i/m$ . Les  $\mu_i$  sont dans  $[0, 1]$  et somment à 1, donc par l'hypothèse de récurrence,

$$X := \mu_1 x_1 + \dots + \mu_n x_n \in C.$$

Maintenant,

$$\lambda_1 x_1 + \dots + \lambda_{n+1} x_{n+1} = mX + \lambda_{n+1} x_{n+1} = (1 - \lambda_{n+1})X + \lambda_{n+1} x_{n+1} \in C$$

par définition d'un ensemble convexe. De plus,

$$\begin{aligned} f(\lambda_1 x_1 + \dots + \lambda_{n+1} x_{n+1}) &\leq (1 - \lambda_{n+1})f(X) + \lambda_{n+1}f(x_{n+1}) \\ &\leq mf(\mu_1 x_1 + \dots + \mu_n x_n) + \lambda_{n+1}f(x_{n+1}) \\ &\leq m(\mu_1 f(x_1) + \dots + \mu_n f(x_n)) + \lambda_{n+1}f(x_{n+1}) \\ &\leq \lambda_1 f(x_1) + \dots + \lambda_{n+1} f(x_{n+1}). \end{aligned}$$

□

Ensuite, on énonce la proposition suivante même si elle semble tautologique car c'est souvent de cette façon que l'on pense les fonctions convexes en dimension supérieure à 2.

**Proposition 3.3.** *Soit  $C \subset \mathbb{R}^d$  convexe et  $f : C \rightarrow \mathbb{R}$ . La fonction  $f$  est convexe si et seulement si pour tout  $a$  et  $b$  dans  $C$ , la fonction d'une variable réelle*

$$f_{a,b} : t \in [0, 1] \mapsto f((1-t)a + tb) \in \mathbb{R}$$

*est convexe.*

Je n'écris pas la preuve car elle est évidente, mais cette proposition signifie que  $f$  est convexe si et seulement si elle l'est quand on la restreint à n'importe quel segment inclus dans  $C$ . Donc les propositions du paragraphe précédent vont pouvoir s'appliquer aux différentes restrictions de  $f$ . Généralisons d'abord la proposition 2.4.

**Proposition 3.4** ( $f$  est au dessus de son plan tangent). *Soit  $C \subset \mathbb{R}^d$  un convexe et  $f : C \rightarrow \mathbb{R}$  une fonction convexe. On suppose que  $f$  est différentiable en un point  $x$  de l'intérieur de  $C$ . Alors pour tout  $y \in C$ ,*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

*Démonstration.* On applique la proposition 2.4 à  $f_{x,y}$  en 0 :

$$f_{x,y}(1) \geq f_{x,y}(0) + f'_{x,y}(0).$$

Or,  $f_{x,y}(0) = f(x)$ ,  $f_{x,y}(1) = f(y)$  et  $f'_{x,y}(0) = \langle \nabla f(x), y - x \rangle$ . □

Passons maintenant à la généralisation de la proposition de 2.6.

**Proposition 3.5** (Hessienne positive). *Soit  $C \subset \mathbb{R}^d$  un ouvert convexe et  $f : C \rightarrow \mathbb{R}$  une fonction. On suppose que  $f$  est de classe  $C^2$  sur  $C$ . Alors  $f$  est convexe si et seulement si sa hessienne (qui est symétrique) est partout positive.*

*Démonstration.* Soient  $x$  et  $y$  dans  $C$ . La fonction  $f_{x,y}$  est de classe  $C^2$ . Elle est donc convexe si et seulement si pour tout  $t \in [0, 1]$ ,  $f''_{x,y}(t) \geq 0$ . Or pour tout  $t \in [0, 1]$ ,

$$\begin{aligned} f'_{x,y}(t) &= \langle \nabla f((1-t)x + ty), y - x \rangle, \\ f''_{x,y}(t) &= \langle y - x, H_f((1-t)x + ty) \cdot (y - x) \rangle. \end{aligned}$$

Donc si  $H_f$  est positive partout, alors pour tous  $x$  et  $y$  dans  $C$ ,  $f_{x,y}$  est convexe et donc  $f$  est convexe.

Si  $f$  est convexe, il suffit de montrer que pour tout  $x \in C$  et  $u \in \mathbb{R}^d$ ,

$$\langle u, H_f(x) \cdot u \rangle \geq 0.$$

Mais dans ce cas, pour tout  $x$  et  $y$  dans  $C$ ,  $f_{x,y}$  est convexe et a donc une dérivée seconde positive. Il suffit alors d'appliquer notre égalité à  $x$ ,  $y = x + \varepsilon u$  ou  $\varepsilon$  est suffisamment petit pour que  $y \in C$  et  $t = 0$ . On obtient :

$$f''_{x,y}(0) = \varepsilon^2 \langle u, H_f(x) \cdot u \rangle \geq 0.$$

On a donc le résultat. □

Pour ce qui est de la continuité, le même résultat qu'en dimension 1 est valide. Je me contente de l'énoncer sans démonstration.

**Théorème 3.6** (Admis). *Soit  $C \subset \mathbb{R}^d$  un convexe et  $f : C \rightarrow \mathbb{R}$  une fonction convexe. Alors  $f$  est continue en tout point intérieur de  $C$ .*

## 4 Lien avec l'optimisation

Voyons maintenant ce que l'on peut dire lorsque l'on minimise une fonction convexe.

### 4.1 Optimisation libre

D'abord, dans le cas convexe, les minimiseurs locaux et globaux coïncident.

**Proposition 4.1.** *Soit  $C \subset \mathbb{R}^d$  un ensemble convexe,  $f : C \rightarrow \mathbb{R}^d$  une fonction convexe et  $x$  un point de minimum local (resp. min. local strict) de  $f$  sur  $C$ . Alors  $x$  est un point de minimum global (resp. min. global strict) de  $f$  sur  $C$ .*

*Démonstration.* C'est une conséquence de la remarque 2.3. Soit en effet  $\varepsilon > 0$  tel que pour tout  $y \in C \cap B(x, \varepsilon)$ ,  $f(y) \geq f(x)$ . Soit  $z \in C$  et

$$\gamma : t \in [0, 1] \mapsto (1 - t)x + tz \in C.$$

On a  $\gamma(0) = x$  et  $\gamma$  est continue. Donc il existe  $0 < \delta \leq 1$  tel que si  $t \leq \delta$ , alors  $\|\gamma(t) - x\| < \varepsilon$ . En particulier,  $y := \gamma(\delta) \in C \cap B(x, \varepsilon)$ . Par la remarque 2.3 appliqué à  $f_{x,z}$  avec  $a = 0$ ,  $b = \delta$  et  $c = 1$ ,

$$f(z) \geq f(x) + \frac{f(y) - f(x)}{\delta} \geq f(x)$$

car  $f(y) \geq f(x)$  (l'inégalité est stricte si le min. local est strict). □

Il suffit d'être un point critique pour être un minimum global.

**Proposition 4.2.** *Soit  $C$  un convexe et  $f : C \rightarrow \mathbb{R}$  une fonction convexe. Soit  $x$  un point intérieur de  $C$  tel que*

- $f$  est différentiable en  $x$ ,
- $\nabla f(x) = 0$ .

*Alors  $x$  est un point de minimum global.*

*Démonstration.* C'est une conséquence immédiate de la proposition 3.4. □

On connaît également la structure de l'ensemble des minimiseurs.

**Proposition 4.3.** Soit  $C \subset \mathbb{R}^d$  un convexe et  $f : C \rightarrow \mathbb{R}$  une fonction convexe. Alors l'ensemble des points de minimum de  $f$  forment un ensemble convexe. Plus généralement, si  $K \in \mathbb{R}$ ,  $f^{-1}(]-\infty, K])$  est un ensemble convexe.

*Démonstration.* Soient  $x$  et  $y$  deux points de  $f^{-1}(]-\infty, K])$  de  $f$  et soit  $t \in [0, 1]$ . On a par convexité

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) = (1-t)K + tK = K.$$

Donc  $(1-t)x + ty \in f^{-1}(]-\infty, K])$ , qui est donc convexe. Pour avoir la première affirmation, il suffit d'appliquer ce résultat à  $K = \min f$  lorsque ce minimum existe.  $\square$

Un point intéressant de la minimisation convexe est que l'on a des critères très simples pour affirmer qu'un point de minimum global est unique. On en a déjà vu un : si  $f$  admet un minimum local strict alors en fait c'est l'unique minimum global de  $f$ . En particulier, si  $x$  est un point de minimum de  $f$  et si  $f$  est deux fois différentiable en  $x$ , de hessienne définie positive, alors  $x$  est l'unique point de minimum. Un deuxième critère est souvent utilisé.

**Définition 4.4.** Soit  $C$  un sous-ensemble convexe de  $\mathbb{R}^d$  et  $f : C \rightarrow \mathbb{R}$  une fonction. On dit que  $f$  est strictement convexe si pour tous  $x$  et  $y$  distincts dans  $C$  et pour tout  $t \in ]0, 1[$ ,

$$f((1-t)x + ty) < (1-t)f(x) + tf(y).$$

**Proposition 4.5.** Soit  $C$  un sous-ensemble convexe de  $\mathbb{R}^d$  et  $f : C \rightarrow \mathbb{R}^d$  une fonction strictement convexe. Alors  $f$  admet au plus un point de minimum.

*Démonstration.* Par l'absurde, si on peut trouver  $x$  et  $y$  deux points de minimum distincts dans  $C$ , alors

$$f\left(\frac{x+y}{2}\right) < \frac{f(x) + f(y)}{2} = \min f,$$

ce qui est absurde.  $\square$

## 4.2 Optimisation sous contraintes

Dans le cas de l'optimisation sous contraintes d'inégalité, on va voir que les conditions KKT sont des conditions suffisantes d'optimalité. Le bon cadre est le cas où l'on cherche à minimiser une fonction convexe  $f$  sur un ouvert convexe  $C$  de  $\mathbb{R}^d$  sous le jeu de contraintes

$$g_1(x) \leq c_1, \dots, g_m(x) \leq c_m,$$

où  $g_1, \dots, g_m$  sont des fonctions convexes. Remarquez que l'ensemble  $K \subset C$  des points qui satisfont les contraintes. En effet,

$$K = \bigcap_{i=1}^m g_i^{-1}(]-\infty, c_i]).$$

On suppose que  $x \in C$  est un minimum local sous contraintes et que  $f, g_1, \dots, g_m$  sont différentiables en  $x$ . Dans ce cadre, le système de KKT (condition nécessaire de minimisations) est le suivant :

$$\begin{cases} \nabla f(x) = \sum_{i=1}^m \lambda_i \nabla g_i(x), \\ \lambda_i (c_i - g_i(x)) = 0, \\ \lambda_1, \dots, \lambda_m \leq 0. \end{cases} \quad (8.1)$$

**Proposition 4.6.** Si  $f, g_1, \dots, g_m$  sont des fonctions convexes et si  $x \in K$  est tel qu'il existe  $\lambda_1, \dots, \lambda_m$  tels que  $x, \lambda_1, \dots, \lambda_m$  satisfait (8.1). Alors  $x$  est un minimum global de  $f$  sous contraintes.

*Démonstration.* Soit  $y \in K$ . On a

$$\begin{aligned}
 f(y) &\geq f(x) + \langle \nabla f(x), y - x \rangle \\
 &= f(x) + \sum_{i=1}^m \lambda_i \langle \nabla g_i(x), y - x \rangle \\
 &\geq f(x) + \sum_{i=1}^m \lambda_i (g_i(y) - g_i(x)) \\
 &= f(x) + \sum_{i=1}^m \lambda_i (g_i(y) - c_i) \\
 &\geq f(x).
 \end{aligned}$$

□

## 5 Introduction à la dualité

Imaginons que l'on veuille résoudre le problème consistant à minimiser une fonction  $f : C \subset \mathbb{R}^d \rightarrow \mathbb{R}$  sous le jeu de contraintes  $g_i(x) \leq c_i$ ,  $i = 1, \dots, m$ . Pour tout  $x \in C$  et  $\lambda_1, \dots, \lambda_m$ , on peut former le lagrangien

$$L(\lambda, x) := f(x) - \sum_{i=1}^m \lambda_i (c_i - g_i(x)).$$

On suppose que le problème a une solution  $x^*$ .

On voit que quelque soient  $\lambda_1, \dots, \lambda_m$  **positifs** et quelque soit  $x$  satisfaisant les contraintes,

$$L(\lambda, x) \leq f(x).$$

En particulier

$$L(\lambda, x^*) \leq f(x^*).$$

En pratique, il peut être délicat de trouver  $x^*$ , tandis que trouver les minimiseurs de  $L(\lambda, \bullet)$  à  $\lambda$  fixé en oubliant les contraintes est souvent plus facile (par exemple si  $C$  est ouvert, dans ce cas c'est essentiellement un calcul du premier ordre "facile"). Imaginons que l'on sache trouver pour tout  $\lambda \in \mathbb{R}^m$  à coordonnées positives (on note  $\lambda \in \mathbb{R}_+^m$ ) la valeur suivante :

$$\Psi(\lambda) := \inf_{x \in C} L(\lambda, x).$$

Dans ce cas,  $\Psi(\lambda)$  nous fournit un minorant pour  $f(x^*)$  :

$$\Psi(\lambda) \leq f(x^*).$$

On peut alors se poser la question suivante : quelle est le meilleur minorant que l'on peut trouver par cette méthode ? Cela revient à calculer

$$\sup_{\lambda \in \mathbb{R}_+^m} \Psi(\lambda).$$

A priori, ce n'est pas un problème difficile car on verra que  $\Psi$  est concave. On souhaite donc maximiser une fonction concave sur un convexe relativement simple ( $\mathbb{R}_+^m$ ). Ce problème de maximisation s'appelle le problème dual (en opposition au premier, appelé problème primal).

Il se trouve que dans le cas où  $C$  est convexe (gentil pour que la méthode ait un intérêt) et  $f, g_1, \dots, g_m$  sont convexes, alors les valeurs des solutions des deux problèmes coïncident. On a donc remplacé un problème de minimisation dans lequel on aurait dû gérer des contraintes potentiellement compliqués par deux problèmes, l'un consistant à minimiser une fonction convexe sur un convexe (gentil), l'autre à maximiser une fonction concave avec de simples contraintes de positivité. Voyons cela plus en détails.

## 5.1 Définition du problème dual

Dans ce paragraphe, on se donne  $C \subset \mathbb{R}^d$  et  $f, g_1, \dots, g_m$  des fonctions réelles définies sur  $C$  **sans aucune hypothèse**. On se donne également  $c_1, \dots, c_m$  des réels. On note comme précédemment pour tout  $\lambda \in \mathbb{R}^m$

$$L(\lambda, x) := f(x) - \sum_{i=1}^m \lambda_i (c_i - g_i(x)).$$

**Définition 5.1** (Problème primal). On appelle problème primal, noté  $(P)$ , le problème consistant à minimiser  $f$  sous contraintes  $g_1 \leq c_1, \dots, g_m \leq c_m$ . On note  $p$  la valeur de la solution du problème primal.

*Remarque 5.2.* Le problème primal consiste à déterminer la valeur suivante :

$$\inf_{x \in C} \sup_{\lambda \in \mathbb{R}_+^m} L(\lambda, x).$$

en effet, pour tout  $x \in C$ , le sup vaut  $f(x)$  si les contraintes sont vérifiées, et  $+\infty$  sinon.

**Définition 5.3** (Fonction duale). Pour tout  $\lambda \in \mathbb{R}_+^m$ , on note

$$\Psi(\lambda) := \inf_{x \in C} L(\lambda, x) = \inf_{x \in C} f(x) - \sum_{i=1}^m \lambda_i (c_i - g_i(x)).$$

**Proposition 5.4.** *La fonction  $\Psi$  est concave.*

*Démonstration.* Pour tout  $x \in C$ ,

$$\lambda \in \mathbb{R}_+^m \mapsto f(x) - \sum_{i=1}^m \lambda_i (c_i - g_i(x))$$

est linéaire. La fonction  $\Psi$  est donc un inf de fonction linéaire (donc concave), elle est donc concave.  $\square$

**Définition 5.5** (Problème dual). On appelle problème dual, noté  $(D)$ , le problème consistant à maximiser  $\Psi$  sur  $\mathbb{R}_+^m$ . On note  $d$  la valeur de la solution du problème dual.

*Remarque 5.6.* Le problème dual consiste à déterminer la valeur suivante :

$$\sup_{\lambda \in \mathbb{R}_+^m} \inf_{x \in C} L(\lambda, x).$$

**Proposition 5.7.** *On a toujours*

$$d \leq p.$$

*Démonstration.* Soit  $\lambda \in \mathbb{R}_+^m$  et  $x$  satisfaisant les contraintes. On a

$$\Psi(\lambda) \leq f(x) - \sum_{i=1}^m \lambda_i (c_i - g_i(x)) \leq f(x).$$

En passant au sup à gauche et à l'inf à droite, on obtient le résultat.  $\square$

## 5.2 Cas d'un problème convexe

Maintenant, on suppose que  $C$  est convexe, et que  $f, g_1, \dots, g_m$  sont convexes.

**Définition 5.8.** On dit que les contraintes satisfont la condition de Slater s'il existe  $x \in C$  tel que pour tout  $i = 1, \dots, m$

$$g_i(x) < c_i.$$

Dans ce cas, on a le théorème suivant.

**Théorème 5.9.** *Si la condition de Slater est satisfaite, alors*

$$p = d.$$

*Remarque 5.10.* Remarquez que l'on n'a fait presque aucune hypothèse! Toute la régularité dont on a besoin est incluse dans la notion de convexité et dans la condition de Slater. Je vais faire la preuve dans un cas au contraire où on fait beaucoup d'hypothèses, à savoir dans le cas où l'on peut écrire le système KKT.

À vrai dire, on n'a même pas fait l'hypothèse que le problème primal admettait un minimiseur. En revanche, on pourrait montrer que le problème dual admet toujours un maximiseur dans le cas où la condition de Slater est vérifiée.

*Éléments de preuve.* Imaginons que le problème primal admette un minimiseur  $x^*$  et que l'on puisse écrire le système KKT. Fixons alors  $\lambda^* \in \mathbb{R}_+^m$  le multiplicateur de Lagrange donné par ce système. On a

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0,$$

ce qui signifie exactement que  $x^*$  est un point critique que

$$x \in C \mapsto f(x) - \sum_{i=1}^m \lambda_i^* (c_i - g_i(x)) = L(\lambda^*, x),$$

qui est une fonction convexe. On en déduit que  $x^*$  minimise cette fonction, c'est à dire

$$\Psi(\lambda^*) = L(\lambda^*, x^*).$$

De plus, on a également pour tout  $i = 1, \dots, m$

$$\lambda_i^* (c_i - g_i(x^*)) = 0,$$

Donc

$$L(\lambda^*, x^*) = f(x^*).$$

On en déduit

$$p = f(x^*) = \Psi(\lambda^*) \leq d \leq p,$$

donc en fait,  $p = d$ . □

# Chapitre 9

## Introduction aux équations différentielles ordinaires

### 1 Premières définitions

#### 1.1 C'est quoi une équation différentielle ?

Une équation différentielle ordinaire (d'ordre 1) est une relation formelle entre la position d'une courbe ( $y : t \mapsto y_t \in \mathbb{R}^d$ ) tracée dans l'espace euclidien de dimension  $d$ , et sa dérivée. Elle s'exprime de façon très générale sous la forme<sup>1</sup> :

$$\dot{y}_t = F(t, y_t)$$

où :

- $t$  (souvent interprété comme le temps) appartient à un intervalle  $I$  de  $\mathbb{R}$ ,
- $F$  est une application de  $I \times U$  ( $U$  étant un ouvert de  $\mathbb{R}^d$ ), à valeurs dans  $\mathbb{R}^d$  (on dit que c'est un champ de vecteurs).

Il faut bien comprendre que dans cette équation la donnée est le champs de vecteurs  $F$ . L'inconnue, quant à elle, est la courbe  $y$ . La plupart du temps, le problème de pose de la façon suivante. On veut décrire un phénomène par une courbe dans un ensemble de dimension  $d$ , obéissant à certaines lois qui prescrivent le champ de vecteurs  $F$ . Si l'on sait qu'à l'instant initial  $t_0 \in I$ , notre système se trouve dans la configuration  $y_0 \in U$ , peut-on déduire de l'équation différentielle l'évolution ultérieure de  $y$ . On ajoute donc en général à notre équation un temps initial  $t_0$  et une donnée initiale  $y_0$ , et on cherche donc à résoudre l'équation :

$$\begin{cases} \dot{y}_t = F(t, y_t), \\ y|_{t=t_0} = y_0. \end{cases} \quad (\text{EDO})$$

Le problème consistant à trouver des solutions à ce type d'équations pour différentes valeurs de  $y_0$  s'appelle un *problème de Cauchy*, ou le *problème de Cauchy associé à  $F$* .

#### 1.2 C'est quoi une solution ?

Soyons précis sur ce qu'on appelle solution de ce problème de Cauchy.

**Définition 1.1.** Étant donnés  $I, U, F, t_0$  et  $y_0$  comme précédemment, une solution du problème de Cauchy (EDO) est une application  $z : t \in J \mapsto z_t \in U$  définie sur un intervalle  $J \subset \mathbb{R}$  vérifiant  $t_0 \in J \subset I$ , et telle que :

---

1. On adopte la notation  $\dot{y}_t$  pour la dérivée de  $y$  au temps  $t$ .

- à l'instant  $t_0$ , on a  $z_{t_0} = y_0$ ,
- l'application  $z$  est dérivable sur  $J$ ,
- pour tout  $t \in J$ , on a

$$\dot{z}_t = F(t, z_t).$$

Le lecteur remarquera que dans cette définition, j'ai pris un intervalle  $J$  a priori différent de  $I$ . En effet, on verra plus tard des exemples de solutions qui cessent d'exister au bout d'un certain temps. C'est par exemple le cas lorsque cette solution atteint la frontière de  $U$  (on dit que la solution explose!). Dans ce cas, la bonne notion de solution est la notion de solution maximale, qui se définit ainsi, en deux temps.

- Définition 1.2.**
1. Soient  $z^1, z^2$  deux solutions de (EDO) définies sur  $J_1$  et  $J_2$  respectivement. On dit que  $z^2$  est un *prolongement* de  $z^1$  si  $J_1 \subset J_2$  et si pour tout  $t \in J_1$ ,  $z_t^1 = z_t^2$ .
  2. Soit  $z$  une solution de (EDO). On dit que  $z$  est une *solution maximale* si elle n'admet pas d'autres prolongements qu'elle même.

Cette définition est assez satisfaisante : on pourrait démontrer que pour toute solution  $z$  de (EDO), il existe toujours un prolongement  $\tilde{z}$  de  $z$  qui est une solution maximale. On restreint donc notre étude aux solutions qui existent pendant le plus longtemps possible.

### 1.3 Que veut-on faire ?

Les premières questions que l'on va se poser sont les suivantes :

- Le problème de Cauchy (EDO) admet-il des solutions ?
- Les solutions maximales sont elles-unicques ?

Ce sont des questions très naturelles puisque ce sont elles qui répondent à la première question que l'on s'est posé. Si la réponse à ces deux questions est affirmatives, cela signifie que l'équation différentielle caractérise bien l'évolution du modèle que l'on a choisit. Dans le cas contraire, s'il n'y a pas de solution, cela signifie que notre modèle est mauvais puisque le phénomène que l'on cherche à décrire, lui, existe bien. S'il n'y a pas unicité des solutions, cela signifie que notre modèle n'est pas suffisant pour décrire le comportement réel de notre système, qu'il n'est pas assez spécifique, puisque l'on ne sait pas quelle trajectoire le système réel choisit.

Lorsqu'il y a bien existence et unicité de solutions à notre problème de Cauchy, on peut se poser des questions cherchant à établir de façon plus précise le comportement de ces solutions. Par exemple, on peut se demander :

- Peut-on trouver des formules explicites pour ces solutions ?
- Une petite erreur sur la donnée initiale engendre-t-elle une petite erreur sur la solution aux temps ultérieurs ?
- Si les solutions maximales sont définies pour  $t \rightarrow +\infty$ , peut-on décrire l'asymptotique des solutions ?

Dans la suite de ce chapitre, on va tenter de répondre à tout ou partie de ces questions dans différentes situations (c'est à dire pour différents types de champs de vecteurs). J'ai préféré donné une grande diversité de résultats plutôt que de les démontrer en détail. Pour tout démontrer, il faudrait introduire des notions capitales en mathématiques que l'on n'aura pas le temps d'aborder, et en premier lieu la notion de complétude.

## 1.4 Vocabulaire : équations autonomes

Un cas particulier que l'on rencontrera souvent est le cas dit *autonome*, c'est à dire le cas où  $F$  ne dépend pas de la variable de temps  $t$  :

$$\begin{cases} \dot{y}_t = F(y_t), \\ y|_{t=t_0} = y_0. \end{cases} \quad (9.1)$$

Dans ce cas, on a les propriétés suivantes :

- Si  $x \in U$  est tel que  $F(x) = 0$ , alors  $y : t \mapsto x$  est une solution dite stationnaire ou constante, définie pour tout temps. Une question classique est de se poser la question de la stabilité de cette solution stationnaire : si  $y_0$  est proche de  $x$ ,  $y_t$  se rapproche-t-il de  $x$  lorsque  $t \rightarrow +\infty$  ?
- Si  $y$  est solution sur un intervalle  $J$  et si  $\bar{t} \in J$ , alors  $z : t \in \bar{t} + J \mapsto y_{t-\bar{t}}$  est également une solution. On suppose donc souvent quitte à poser  $\bar{t} = t_0$  que  $t_0 = 0$ .

## 1.5 Équations différentielles d'ordre supérieur

Parfois, on ne veut pas exprimer une relation entre la dérivée d'une courbe et sa position, mais plutôt entre sa dérivée  $n$ -ième ( $n \geq 2$ ) et ses dérivées d'ordres inférieurs. La formulation générale de ce type d'équation est alors la suivante<sup>2</sup> :

$$\begin{cases} y_t^{(n)} = F(t, y_t, \dot{y}_t, y_t^{(2)}, \dots, y_t^{(n-1)}), \\ y|_{t=t_0} = y_0, \dot{y}|_{t=t_0} = y_1, \dots, y_t^{(n-1)}|_{t=t_0} = y_{n-1}. \end{cases} \quad (9.2)$$

Cette fois,  $F$  est une application de  $I \times U \times (\mathbb{R}^d)^{n-1}$  à valeurs dans  $\mathbb{R}^d$ , et en plus de spécifier la position initiale, il faut spécifier les dérivées initiales jusqu'à l'ordre  $n-1$  pour espérer obtenir des résultats d'unicité.

Ces équations ont joué un rôle primordial dans le développement des sciences, puisque par exemple, les lois de Newton pour le mouvements des corps s'écrivent sous la forme d'équations différentielles d'ordre 2.

En fait, il est très facile d'écrire cette équation sous la forme (EDO) d'un système d'ordre 1, quitte à considérer la variable :

$$Y : t \mapsto (y_t, \dot{y}_t, \dots, y_t^{(n-1)}) \in U \times (\mathbb{R}^d)^{n-1}.$$

En effet, on vérifie très facilement que *via* ce changement de variable,  $y$  est solution de l'équation (9.2) si et seulement si  $Y$  est solution de l'équation différentielle d'ordre 1 :

$$\begin{cases} \dot{Y}_t = \tilde{F}(t, Y_t), \\ Y_0 = (y_0, \dots, y_{n-1}), \end{cases}$$

où  $\tilde{F}$  est définie par :

$$\tilde{F} : (y_0, \dots, y_{n-1}) \in U \times (\mathbb{R}^d)^{n-1} \mapsto (y_1, \dots, y_{n-1}, F(t, y_0, \dots, y_{n-1})) \in (\mathbb{R}^d)^n.$$

En conséquence, sauf dans un cas particulier, on ne considérera par la suite que des équations différentielles d'ordre 1.

À partir de maintenant, on va tenter de répondre à certaines des questions énoncées plus haut en spécifiant le type de champ de vecteurs  $F$  choisi. On commence par le cas le plus simple, le cas des équations différentielles linéaires en dimension 1.

---

2. Cette fois, la dérivée d'ordre  $k$  de  $y$  à l'instant  $t$  pour  $k \geq 2$  est notée  $y_t^{(k)}$ .

## 2 Équation linéaires en dimension 1

Dans cette section, on choisit  $d = 1$ , et on considère le cas où  $F$  est *linéaire* par rapport à la position. En d'autres termes, dans (EDO),  $U = \mathbb{R}$  et il existe  $a : I \rightarrow \mathbb{R}$  (disons continue pour pouvoir l'intégrer même si cette hypothèse n'est pas essentielle) telle que pour tout  $t \in I$  et  $x \in \mathbb{R}$  :

$$F(t, x) = a(t)x.$$

Étant donné  $y_0 \in \mathbb{R}$  et  $t_0 \in I$ , on cherche donc à résoudre le problème de Cauchy :

$$\begin{cases} \dot{y}_t = a(t)y_t, \\ y|_{t=t_0} = y_0. \end{cases} \quad (9.3)$$

### 2.1 Cas autonome : la fonction exponentielle

Le premier cas à considérer est celui où  $I = \mathbb{R}$ ,  $a \equiv 1$ ,  $t_0 = 0$  et  $y_0 = 1$  :

$$\begin{cases} \dot{y}_t = y_t, \\ y|_{t=0} = 1. \end{cases} \quad (9.4)$$

Dans ce cas, il est bien connu que l'unique solution est la fonction exponentielle (cela peut même être choisi comme définition de l'exponentielle). Ici, on va admettre les choses suivantes sur l'exponentielle :

La fonction exponentielle est une fonction de  $\mathbb{R} \rightarrow \mathbb{R}_+^*$ , valant 1 en 0, dérivable, et vérifiant pour tout  $t \in \mathbb{R}$  :

$$\frac{d}{dt} \exp(t) = \exp(t).$$

On admet donc en particulier que la fonction exponentielle est une solution du problème de Cauchy (9.4). Ce sont des propriétés que l'on peut par exemple obtenir facilement en définissant l'exponentielle comme la fonction réciproque de la fonction logarithme, elle-même définie comme la primitive de la fonction inverse s'annulant en 1. Je ne prétend pas que c'est la meilleure façon de définir la fonction exponentielle, il y a plusieurs alternatives. Mais avec ces seules propriétés, on peut démontrer le théorème suivant, dont la preuve va nous permettre de nous familiariser avec des idées clés dans l'étude des équations différentielles.

**Théorème 2.1.** *La fonction exponentielle est l'unique solution maximale du problème de Cauchy (9.4). De plus, elle définit pour tout  $s, t \in \mathbb{R}$  :*

$$\exp(s + t) = \exp(s) \exp(t), \quad (9.5)$$

et pour tout  $t \in \mathbb{R}$  :

$$\exp(t) = \sum_{k=0}^{+\infty} \frac{t^k}{k!} = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{t^k}{k!}. \quad (9.6)$$

*Démonstration.* Pour obtenir l'unicité, si  $y$  est une autre solution du problème de Cauchy, alors définissons :

$$z_t := \frac{y_t}{\exp(t)},$$

qui est bien définie pour tout  $t \in \mathbb{R}$  puisque la fonction exponentielle ne s'annule pas. En calculant la dérivée de  $z$ , on obtient :

$$\dot{z}_t = \frac{\dot{y}_t \exp(t) - y_t \exp(t)}{\exp(t)^2} = \frac{y_t \exp(t) - y_t \exp(t)}{\exp(t)^2} = 0.$$

La fonction  $z$  est donc constante, et comme elle vaut 1 en 0, on a bien  $y_t = \exp(t)$  pour tout  $t$ .

Maintenant, on peut obtenir les autres résultats en utilisant cette unicité. Pour démontrer (9.5), on fixe  $s \in \mathbb{R}$  et on définit pour tout  $t \in \mathbb{R}$  :

$$y_t := \frac{\exp(s+t)}{\exp(s)}.$$

On a  $y_0 = 1$ ,  $y$  est dérivable pour tout  $t$  et un rapide calcul montre que

$$\dot{y}_t = y_t.$$

Donc pour tout  $t$ ,  $y_t = \exp(t)$ , et on obtient le résultat.

Pour démontrer (9.6), on définit pour tout  $t$  :

$$y_t := \sum_{k=0}^{+\infty} \frac{t^k}{k!} = 1 + \sum_{k=1}^{+\infty} \frac{t^k}{k!}.$$

On a alors  $y_0 = 1$ , et on admet que l'on peut dériver cette formule terme à terme. On obtient alors :

$$\dot{y}_t = \sum_{k=1}^{+\infty} \frac{t^{k-1}}{(k-1)!} = \sum_{k=0}^{+\infty} \frac{t^k}{k!} = y_t.$$

donc pour tout  $t$ ,  $y_t = \exp(t)$  et on obtient le résultat. □

## 2.2 Cas général

On traite maintenant le cas général du problème de Cauchy (9.3) avec  $a$  continue. Dans ce cas encore, on obtient un résultat d'existence et d'unicité, ainsi qu'une formule (faisant intervenir la fonction exponentielle).

**Théorème 2.2.** *Le problème de Cauchy (9.3) admet une unique solution maximale  $y$  définie sur  $I$  tout entier. De plus, pour tout  $t \in I$ , on a :*

$$y_t = \exp\left(\int_{t_0}^t a(s) ds\right) y_0.$$

*Remarque 2.3.* • Un petit calcul permet de se souvenir facilement du résultat. On ne l'utilise pas souvent comme d'une preuve car il y aurait des détails techniques à régler alors qu'il existe une preuve très simple (que l'on va donner juste après), mais dans les faits, on utilise souvent ce type de raisonnement. Si  $y$  est une solution de (9.3) (disons strictement positive, sinon il suffirait d'ajouter des signes  $-$  aux bons endroit, je vous laisse y réfléchir), alors :

$$\frac{\dot{y}_t}{y_t} = \frac{d}{dt} \log y_t = a(t).$$

En particulier, on a :

$$\log y_t = \log y_0 + \int_{t_0}^t a(s) ds,$$

et il suffit de passer à l'exponentielle pour obtenir la formule annoncée.

- si  $a(t) \equiv a$  est constante, on observe évidemment que les solutions de  $\dot{y}_t = ay_t$  sont de la forme  $C \exp(at)$ .

*Démonstration. Existence.* Soit  $y$  comme définie dans l'énoncé. La fonction  $y$  vaut bien  $y_0$  en  $t = t_0$ , et est bien dérivable pour tout  $t \in I$ . De plus, on a pour tout  $t \in I$  :

$$\dot{y}_t = a(t) \exp\left(\int_{t_0}^t a(s) ds\right) y_0 = a(t)y_t.$$

La fonction  $y$  est donc bien une solution du problème linéaire (9.3).

**Unicité.** Soit  $\tilde{y}$  une autre solution du problème de Cauchy (9.3). Comme  $y$  ne s'annule pas, la fonction  $\tilde{y}/y$  est dérivable sur tout  $I$  et pour tout  $t \in I$  :

$$\frac{d}{dt} \left. \frac{\tilde{y}}{y} \right|_t = \frac{\dot{\tilde{y}}_t y_t - \tilde{y}_t \dot{y}_t}{y_t^2} = a(t) \frac{\tilde{y}_t y_t - \tilde{y}_t y_t}{y_t^2} = 0.$$

En  $t = 0$ ,  $\tilde{y}/y$  vaut 1, on obtient que pour tout  $t \in I$ ,  $\tilde{y}_t = y_t$ , d'où l'unicité.  $\square$

### 2.3 Équation avec second membre

On dit qu'une équation à un second membre si la fonction  $F$  du problème de Cauchy contient un terme ne dépendant que du temps. Une équation linéaire en dimension 1 avec un second membre s'écrit donc :

$$\begin{cases} \dot{y}_t = a(t)y_t + b(t), \\ y|_{t=t_0} = y_0. \end{cases} \quad (9.7)$$

Cette fois encore, on suppose que  $a$  et  $b$  sont continus, de façon à pouvoir les intégrer. On obtient également l'existence, l'unicité et une formule pour les solutions de ce problème de Cauchy.

**Théorème 2.4.** *Le problème de Cauchy (9.7) admet une unique solution maximale  $y$  définie sur  $I$  tout entier. De plus, pour tout  $t \in I$ , on a :*

$$y_t = \exp\left(\int_{t_0}^t a(\tau) d\tau\right) y_0 + \int_{t_0}^t \exp\left(\int_s^t a(\tau) d\tau\right) b(s) ds. \quad (9.8)$$

*Démonstration.* Ici encore, un calcul formel simple permet de retrouver cette formule, mais cette fois, on s'en sert comme d'une preuve.

**Unicité.** On observe que si  $y$  est une fonction dérivable, alors :

$$\frac{d}{dt} \exp\left(-\int_{t_0}^t a(\tau) d\tau\right) y_t = \exp\left(-\int_{t_0}^t a(\tau) d\tau\right) \{\dot{y}_t - a(t)y_t\}.$$

Donc si  $y$  est une solution du problème (9.7), alors :

$$\frac{d}{dt} \exp\left(-\int_{t_0}^t a(\tau) d\tau\right) y_t = \exp\left(-\int_{t_0}^t a(\tau) d\tau\right) \{\dot{y}_t - a(t)y_t\} = \exp\left(-\int_{t_0}^t a(\tau) d\tau\right) b(t).$$

Le terme de droite ne dépendant plus de  $y$ , on peut intégrer cette égalité entre  $t_0$  et  $t \in I$  et obtenir :

$$\exp\left(-\int_{t_0}^t a(\tau) d\tau\right) y_t = y_0 + \int_{t_0}^t \exp\left(-\int_{t_0}^s a(\tau) d\tau\right) b(s) ds.$$

En multipliant cette égalité par  $\exp(\int \dots)$ , et en remarquant l'identité :

$$\exp\left(\int_{t_0}^t a(\tau) d\tau\right) \exp\left(-\int_{t_0}^s a(\tau) d\tau\right) = \exp\left(\int_s^t a(\tau) d\tau\right),$$

on en déduit que si  $y$  est solution, alors elle vérifie la formule annoncée. Il y a donc bien au plus une solution. Pour conclure, il suffit donc de montrer que la formule définit bien une solution.

**Existence.** Par ailleurs, si on définit  $y$  par la formule de l'énoncé,  $y$  vaut bien  $y_0$  en  $t_0$ , elle est bien dérivable sur tout  $I$ , et pour tout  $t \in I$ , en utilisant des résultats classiques de dérivation sous le signe  $\int$  :

$$\dot{y}_t = a(t) \exp\left(\int_{t_0}^t a(\tau) d\tau\right) y_0 + a(t) \int_{t_0}^t \exp\left(\int_s^t a(\tau) d\tau\right) b(s) ds + b(t) = a(t)y_t + b(t).$$

La fonction  $y$  est donc bien une solution du problème (9.7). Le théorème est donc démontré.  $\square$

### 3 Équations différentielles linéaires en dimension supérieure

Dans cette section, on va s'intéresser au problème de Cauchy :

$$\begin{cases} \dot{y}_t = A(t)y_t, \\ y|_{t=t_0} = y_0, \end{cases} \quad (9.9)$$

où cette fois,  $y$  est une fonction à valeur dans  $\mathbb{R}^d$ , et  $A$  est une fonction continue à valeurs dans l'ensemble des matrices de tailles  $d$ .

#### 3.1 Cas autonome : exponentielle de matrices

On s'intéresse d'abord au cas où  $A(t) \equiv A$  est une matrice constante :

$$\begin{cases} \dot{y}_t = Ay_t, \\ y|_{t=t_0} = y_0, \end{cases} \quad (9.10)$$

Dans ce cas et dans ce cas seulement, on pose  $\mathbb{K} = \mathbb{R}$  ou  $\mathbb{C}$ , c'est à dire que l'on s'autorise à travailler avec des matrices et des solutions à valeurs complexes. En effet, tout ce que l'on va dire se simplifie lorsque la matrice  $A$  est diagonalisable. Or, la plupart des matrices complexes sont diagonalisables dans  $\mathbb{C}$  (l'ensemble de ces matrices contient un ouvert dense de l'ensemble des matrices, c'est à dire un ensemble très "gros"). On traitera en détails la recherche de solutions réelles dans le cas où  $A$  est diagonalisable dans  $\mathbb{C}$  mais pas dans  $\mathbb{R}$ . En suivant l'exemple de la dimension 1, on définit l'exponentielle de la matrice  $A$  par<sup>3</sup> :

$$\exp(A) := \sum_{k=0}^{+\infty} \frac{A^k}{k!} = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{A^k}{k!}. \quad (9.11)$$

On admet que cette formule a toujours un sens, à savoir que la limite dans le terme de droite existe. Commençons par donner quelques propriétés des exponentielles de matrices.

#### Propriétés de l'exponentielle

**Proposition 3.1.** *L'opérateur  $A \in \mathcal{M}_d(\mathbb{K}) \mapsto \exp(A) \in \mathcal{M}_d(\mathbb{K})$  défini par la formule (9.11) vérifie :*

1.  $\exp(A)$  commute avec  $A$ , c'est à dire  $A \times \exp(A) = \exp(A) \times A$ . Plus généralement, si  $A$  et  $B$  commutent, alors  $\exp(A)$  et  $B$  commutent.
2. On a pour tout  $t$  :

$$\frac{d}{dt} \exp(tA) = A \times \exp(tA).$$

3. Si  $A = PBP^{-1}$ , avec  $A, B, P \in \mathcal{M}_d(\mathbb{K})$ ,  $P$  inversible, alors  $\exp(A) = P \exp(B)P^{-1}$ . En particulier :

$$\text{Si } A = P \times \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} P^{-1}, \quad \text{alors } \exp(A) = P \times \begin{pmatrix} \exp(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & \exp(\lambda_d) \end{pmatrix} P^{-1}.$$

4. Si  $A$  et  $B$  commutent, alors :

$$\exp(A + B) = \exp(A) \times \exp(B) = \exp(B) \times \exp(A).$$

*Attention : ce n'est pas nécessairement vrai lorsque  $A$  et  $B$  ne commutent pas !*

---

3. Par convention,  $A^0 = \text{Id}$

*Démonstration.* Point 1. Soient  $A, B \in \mathcal{M}_d(\mathbb{K})$  deux matrices qui commutent. On a :

$$B \times \exp(A) = B \left( \sum_{k=0}^{+\infty} \frac{A^k}{k!} \right) = \sum_{k=0}^{+\infty} \frac{B \times A^k}{k!} = \sum_{k=0}^{+\infty} \frac{A^k \times B}{k!} = \left( \sum_{k=0}^{+\infty} \frac{A^k}{k!} \right) \times B = \exp(A) \times B.$$

Point 2. Soit  $A \in \mathcal{M}_d(\mathbb{K})$ . Pour tout  $t \in \mathbb{R}$ , on a :

$$\exp(tA) = \sum_{k=0}^{+\infty} t^k \times \frac{A^k}{k!} = \text{id} + \sum_{k=1}^{+\infty} t^k \times \frac{A^k}{k!}.$$

En admettant que l'on peut dériver terme à terme, on obtient pour tout  $t \in \mathbb{R}$  :

$$\frac{d}{dt} \exp(tA) = \sum_{k=1}^{+\infty} k \times t^{k-1} \times \frac{A^k}{k!} = \sum_{k=1}^{+\infty} t^{k-1} \times \frac{A^k}{(k-1)!} = A \times \left( \sum_{k=1}^{+\infty} t^{k-1} \times \frac{A^{k-1}}{(k-1)!} \right) = A \times \exp(tA).$$

Point 3. Soient  $A, B, P \in \mathcal{M}_d(\mathbb{K})$  trois matrices avec  $P$  inversible et  $A = PBP^{-1}$ . Alors :

$$\exp(A) = \sum_{k=0}^{+\infty} \frac{A^k}{k!} = \sum_{k=0}^{+\infty} \frac{(PBP^{-1})^k}{k!} = \sum_{k=0}^{+\infty} P \times \frac{B^k}{k!} \times P^{-1} = P \times \left( \sum_{k=0}^{+\infty} \frac{B^k}{k!} \right) \times P^{-1} = P \exp(B) P^{-1}.$$

Pour montrer la deuxième formule, il suffit alors de montrer que pour tout  $\lambda_1, \dots, \lambda_d \in \mathbb{K}$  :

$$\exp \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} = \begin{pmatrix} \exp(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & \exp(\lambda_d) \end{pmatrix}.$$

Mais c'est une conséquence directe de :

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix}^k = \begin{pmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_d^k \end{pmatrix}$$

et de la définition de l'exponentielle de matrice.

Point 4. Soient  $A, B \in \mathcal{M}_d(\mathbb{K})$  qui commutent. On a :

$$\exp(A + B) = \sum_{k=0}^{+\infty} \frac{(A + B)^k}{k!}.$$

Or lorsque  $A$  et  $B$  commutent (et seulement dans ce cas!), pour tout  $k \in \mathbb{N}$  :

$$(A + B)^k = \sum_{l=0}^k \binom{k}{l} A^l \times B^{k-l} = \sum_{l=0}^k \frac{k!}{l!(k-l)!} A^l \times B^{k-l}.$$

En injectant cette formule dans l'égalité précédente, on obtient :

$$\exp(A + B) = \sum_{k=0}^{+\infty} \sum_{l=0}^k \frac{1}{l!(k-l)!} A^l \times B^{k-l} = \sum_{k=0}^{+\infty} \sum_{l=0}^{+\infty} \mathbb{1}_{l \leq k} \frac{1}{l!(k-l)!} A^l \times B^{k-l}.$$

En admettant que l'on peut intervertir les deux sommes :

$$\begin{aligned}
\exp(A + B) &= \sum_{l=0}^{+\infty} \sum_{k=0}^{+\infty} \mathbb{1}_{l \leq k} \frac{1}{l!(k-l)!} A^l \times B^{k-l} \\
&= \sum_{l=0}^{+\infty} \frac{A^l}{l!} \sum_{k=l}^{+\infty} \frac{1}{(k-l)!} \times B^{k-l} \\
&= \sum_{l=0}^{+\infty} \frac{A^l}{l!} \sum_{k=0}^{+\infty} \frac{1}{k!} \times B^k = \exp(A) \times \exp(B).
\end{aligned}$$

□

### Conséquences sur l'équation différentielle

On en déduit très facilement le résultat d'existence et d'unicité suivant sur le problème (9.10).

**Théorème 3.2.** *L'unique solution au problème de Cauchy (9.10) est :*

$$y_t := \exp(tA)y_0.$$

*En particulier, l'ensemble des solutions de :*

$$\dot{y}_t = Ay_t$$

*est un sous-espace vectoriel de dimension  $d$  de l'ensemble  $C^0(\mathbb{R}; \mathbb{K}^d)$  des fonctions continues à valeurs dans  $\mathbb{K}^d$ .*

*Démonstration.* Vérifions que  $y$  tel que défini dans l'énoncé du Théorème est une solution. Pour tout  $t \in \mathbb{R}$ , par le point 2. de la Proposition 3.1, on a :

$$\dot{y}_t = \left( \frac{d}{dt} \exp(tA) \right) y_0 = A \times \exp(tA)y_0 = Ay_t.$$

Montrons que cette solution est unique. Pour ça choisissons  $z$  une solution du problème 9.10 et calculons :

$$\frac{d}{dt}(\exp(-tA) \times z_t) = -A \exp(-tA)z_t + \exp(-tA)\dot{z}_t = A \exp(-tA)(-z_t + z_t) = 0.$$

On a donc pour tout  $t$  :

$$\exp(-tA)z_t = z_t|_{t=0} = y_0.$$

On déduit le résultat du point 4. de la Proposition 3.1, en multipliant cette égalité par  $\exp(tA)$ .

Pour le dernier point, il suffit de remarquer que l'application :

$$y_0 \in \mathbb{K}^d \mapsto \left( t \mapsto \exp(tA)y_0 \right) \in C^0(\mathbb{R}; \mathbb{K}^d)$$

est linéaire, injective (car si  $y_0$  est tel que pour tout  $t$ ,  $\exp(tA)y_0 = 0$ , alors en particulier à  $t = 0$ ,  $y_0 = 0$ ) et surjective sur l'ensemble des solutions de  $\dot{y}_t = Ay_t$  par la première partie du théorème. Son image est donc bien un sous-espace vectoriel de dimension  $d$  de  $C^0(\mathbb{R}; \mathbb{K}^d)$ . □

### Lorsque la matrice est diagonalisable

Lorsque  $A$  est diagonalisable, on peut résoudre (9.10) de façon très simple. Pour cela, deux moyens (évidemment équivalents).

En utilisant la réduction : On trouve  $P \in \mathcal{M}_d(\mathbb{K})$  inversible et  $\lambda_1, \dots, \lambda_d \in \mathbb{K}$  tels que :

$$A = P \times \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} P^{-1}.$$

Dans ce cas, l'unique solution du problème de Cauchy (9.10) est :

$$P \times \begin{pmatrix} \exp(\lambda_1 t) & & 0 \\ & \ddots & \\ 0 & & \exp(\lambda_d t) \end{pmatrix} P^{-1} y_0.$$

Comme combinaison linéaire : Soient  $\lambda_1, \dots, \lambda_d \in \mathbb{K}$  les valeurs propres de  $A$ , et  $u_1, \dots, u_d$  les vecteurs propres correspondants. Remarquons que pour  $i = 1, \dots, d$ , la fonction :

$$y^i : t \in \mathbb{R} \mapsto \exp(\lambda_i t) u_i$$

est une solution de (9.10) car pour tout  $t$  :

$$\dot{y}_t^i = \exp(\lambda_i t) \lambda_i u_i = \exp(\lambda_i t) A u_i = A \exp(\lambda_i t) u_i = A y_t^i.$$

Or on a vu que l'ensemble des solutions de  $\dot{y}_t = A y_t$  était un espace vectoriel de dimension  $d$ , et  $y^1, \dots, y^d$  sont clairement linéairement indépendants. Ils forment donc une base de l'ensemble des solutions. Il existe donc  $\alpha_1, \dots, \alpha_d \in \mathbb{K}$  tels que l'unique solution de (9.10) s'écrivent :

$$\forall t \in \mathbb{R}, \quad y_t = \sum_{i=1}^d \alpha_i y_t^i = \sum_{i=1}^d \alpha_i \exp(\lambda_i t) u_i.$$

Pour trouver  $\alpha_1, \dots, \alpha_d$  il suffit de résoudre à l'instant  $t = 0$  le système de  $d$  équations à  $d$  inconnues :

$$\sum_{i=1}^d \alpha_i u_i = y_0.$$

**Exercice 3.3.** Expliciter complètement le lien entre ces deux méthodes.

### Un petit lemme bien pratique

Dans les deux paragraphes suivants, on va voir besoin du :

*Lemme 3.4.* Soient  $\mu_1, \dots, \mu_k \in \mathbb{C}$  distincts. Alors les fonctions  $t \mapsto \exp(\mu_i t)$ ,  $i = 1, \dots, k$  forment une famille linéairement indépendante de  $C^0(\mathbb{R}; \mathbb{C})$ . En particulier, le résultat est encore vrai quand on remplace  $\mathbb{C}$  par  $\mathbb{R}$ .

*Démonstration.* On suppose  $\Re(\mu_1) \geq \dots \geq \Re(\mu_k)$ . Soient  $\alpha_1, \dots, \alpha_k$  tels que pour tout  $t \in \mathbb{R}$  :

$$\sum_{i=1}^k \alpha_i \exp(\mu_i t) = 0.$$

Montrons que  $\alpha_1 = 0$  (évidemment les autres viendront de la même façon). Multiplions l'égalité précédente par  $\exp(-\mu_1 t)$ . On obtient :

$$\alpha_1 + \sum_{i=2}^k \alpha_i \exp(-(\mu_1 - \mu_i)t) = 0$$

Or par l'hypothèse  $\Re(\mu_1) \geq \dots \geq \Re(\mu_k)$ , les fonctions  $\exp(-(\mu_1 - \mu_i)t)$  restent bornées quand  $t \rightarrow +\infty$  (soit elles tendent vers 0 si  $\Re(\mu_1) > \Re(\mu_i)$ , soit elles sont de norme 1 si  $\Re(\mu_1) = \Re(\mu_i)$ ). Soit  $T > 0$ . Appliquons l'opération  $\frac{1}{T} \int_0^T$  à notre égalité. On obtient :

$$\alpha_1 + \frac{1}{T} \sum_{i=2}^k \frac{\alpha_i}{\mu_1 - \mu_i} \left(1 - \exp(-(\mu_1 - \mu_i)T)\right) = 0.$$

En passant à la limite  $T \rightarrow +\infty$ , on trouve bien  $\alpha_1 = 0$ , et donc le résultat.  $\square$

### Un exemple en dimension 2 : recherche de solutions réelles

On va chercher l'ensemble des solutions réelles du système d'équations :

$$\begin{cases} \dot{x}_t = x_t + 5y_t \\ \dot{y}_t = -5x_t + y_t. \end{cases}$$

qui se reformule en définissant  $X_t = (x_t, y_t)$  :

$$\dot{X}_t = \underbrace{\begin{pmatrix} 1 & 5 \\ -5 & 1 \end{pmatrix}}_{=:A} X_t. \quad (9.12)$$

En vertu du théorème 3.2, ces solutions doivent former une sous-espace vectoriel de dimension 2 de  $C^0(\mathbb{R}; \mathbb{R}^2)$ . Un rapide examen de la matrice  $A$  montre que  $A$  est diagonalisable dans  $\mathbb{C}$  mais pas dans  $\mathbb{R}$ , et que les vecteurs propres (complexes) de  $A$  sont  $(1, i)$  et  $(1, -i)$  associés aux valeurs propres  $1 + 5i$  et  $1 - 5i$  respectivement.

*Remarque 3.5.* On remarque que les deux vecteurs propres et les deux valeurs propres sont conjuguées. C'est normal : comme  $A$  est réel, si  $u \in \mathbb{C}^2$  est un vecteur propre de  $A$  associé à la valeur propre  $\lambda \in \mathbb{C}$  i.e. :

$$Au = \lambda u, \quad \text{alors} \quad A\bar{u} = \overline{Au} = \overline{\lambda u} = \bar{\lambda}\bar{u},$$

donc  $\bar{u}$  est également un vecteur propre de  $A$ , associé à la valeur propre  $\bar{\lambda}$ . Dans le cas des matrices réelles, les valeurs propres qui ne sont pas réelles vont nécessairement par deux et sont conjuguées l'une de l'autre.

Selon le paragraphe précédent, l'ensemble des solutions complexes de (9.12) est le sous-espace vectoriel de dimension 2 du  $\mathbb{C}$ -espace vectoriel  $C^0(\mathbb{R}; \mathbb{C}^d)$  formé des fonctions de la forme :

$$X_{\alpha, \beta} : t \in \mathbb{R} \mapsto \alpha \exp((1 + 5i)t) \begin{pmatrix} 1 \\ i \end{pmatrix} + \beta \exp((1 - 5i)t) \begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad \alpha, \beta \in \mathbb{C}.$$

Mais les solutions réelles sont en particulier des solutions complexes. Il doit donc exister des  $\alpha, \beta \in \mathbb{C}$  tels que  $X_{\alpha, \beta}$  défini par la formule précédente soit en fait une fonction réelle. Soient  $\alpha, \beta \in \mathbb{C}$ . La fonction  $X_{\alpha, \beta}$  est réelle si pour tout  $t \in \mathbb{R}$ ,  $\overline{X_{\alpha, \beta}(t)} = X_{\alpha, \beta}(t)$ . Or pour tout  $t \in \mathbb{R}$ ,

$$X_{\alpha, \beta}(t) = \begin{pmatrix} \alpha \exp((1 + 5i)t) + \beta \exp((1 - 5i)t) \\ i\alpha \exp((1 + 5i)t) - i\beta \exp((1 - 5i)t) \end{pmatrix} = \exp(t) \begin{pmatrix} \alpha \exp(5it) + \beta \exp(-5it) \\ i\alpha \exp(5it) - i\beta \exp(-5it) \end{pmatrix},$$

et donc :

$$\overline{X_{\alpha, \beta}(t)} = \exp(t) \begin{pmatrix} \bar{\alpha} \exp(-5it) + \bar{\beta} \exp(5it) \\ i\bar{\alpha} \exp(-5it) + i\bar{\beta} \exp(5it) \end{pmatrix}.$$

Par le Lemme 3.4,  $X_{\alpha, \beta}$  est donc réel si et seulement si  $\beta = \bar{\alpha}$ . On note alors  $\alpha = (a + ib)/2$ ,  $\beta = (a - ib)/2$  et  $Y_{a, b} = X_{(a+ib)/2, (a-ib)/2}$ ,  $a, b \in \mathbb{R}$ . On a alors pour tout  $t$  :

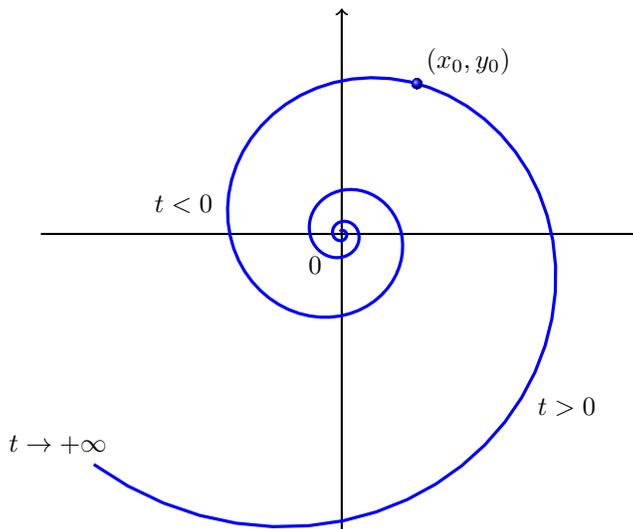
$$Y_{a, b} = \exp(t) \begin{pmatrix} a \cos(5t) + b \sin(5t) \\ -a \sin(5t) + b \cos(5t) \end{pmatrix} = a \exp(t) \begin{pmatrix} \cos(5t) \\ -\sin(5t) \end{pmatrix} + b \exp(t) \begin{pmatrix} \sin(5t) \\ \cos(5t) \end{pmatrix}.$$

On a donc bien trouvé un ensemble de solutions réelles, de dimension 2, engendré par :

$$Y_{1,0} = \exp(t) \begin{pmatrix} \cos(5t) \\ -\sin(5t) \end{pmatrix} \quad \text{et} \quad Y_{0,1} = \exp(t) \begin{pmatrix} \sin(5t) \\ \cos(5t) \end{pmatrix}.$$

## Notion de stabilité linéaire

Traçons la solution partant de  $(x_0, y_0) \in \mathbb{R}^2$ .



La solution tourne autour de l'origine à la vitesse angulaire  $\omega$ , en s'en éloignant comme  $\exp(t)$ . A contrario, quand  $t \rightarrow -\infty$ , la solution tend vers 0. Plus généralement, si les valeurs propres de  $A$  étaient  $\lambda \pm i\omega$ , les solutions tourneraient à la vitesse angulaire  $\omega$ , et auraient une distance à l'origine proportionnelle à  $\exp(\lambda t)$ . En particulier, si  $\lambda > 0$ , quand  $t \rightarrow +\infty$ , les solutions partent à l'infini, si  $\lambda < 0$ , elle tendent vers 0, et si  $\lambda = 0$ , les solutions sont périodiques. En fait, on pourrait démontrer le théorème suivant (dans lequel on n'a même pas besoin de supposer que  $A$  est diagonalisable).

**Théorème 3.6.** • Si toutes les valeurs propres de  $A$  ont une partie réelle strictement négative, alors toutes les solutions de  $\dot{y}_t = Ay_t$  tendent vers 0 exponentiellement vite quand  $t \rightarrow +\infty$ . On dit alors que le problème de Cauchy est **stable**.

- Si une valeur propre de  $A$  a une partie réelle strictement positive, il existe au moins une solution de  $\dot{y}_t = Ay_t$  qui part à l'infini quand  $t \rightarrow +\infty$ . On dit que le problème de Cauchy est **instable**.

## Cas des équations différentielles linéaires autonomes d'ordre supérieur

On s'intéresse à l'équation :

$$\begin{cases} y_t^{(n)} = a_{n-1}y_t^{(n-1)} + \dots + a_0y_t, \\ y_t|_{t=0} = y_0, \dots, y_t^{(n-1)}|_{t=0} = y_{n-1}. \end{cases} \quad (9.13)$$

Ici,  $a_0, \dots, a_{n-1}, y_0, \dots, y_{n-1} \in \mathbb{K}$  et l'inconnue  $t \mapsto y_t$  est une fonction à valeur dans  $\mathbb{K}$ , l'espace de dimension 1. Grâce au raisonnement de la Sous-section 1.5, quitte à prendre pour inconnue  $t \mapsto (y_t, \dots, y_t^{(n-1)})$ , ce problème est un problème de Cauchy linéaire et autonome en dimension  $n$ . Donc grâce aux paragraphes précédent, l'ensemble des solutions de  $y_t^{(n)} = a_{n-1}y_t^{(n-1)} + \dots + a_0y_t$  est une sous-ensemble de dimension  $n$  de  $C^0(\mathbb{R}; \mathbb{K})$ . Pour résoudre complètement cette équation, il suffit de trouver  $n$  solutions linéairement indépendantes. Pour cela, on pose :

$$P(X) := a_n X^n - \sum_{k=0}^{n-1} a_k X^k.$$

Montrons le résultat suivant :

**Proposition 3.7.** Si  $P(\lambda) = 0$ , alors  $y^\lambda : t \mapsto \exp(\lambda t)$  vérifie  $y_t^{(n)} = a_{n-1}y_t^{(n-1)} + \dots + a_0y_t$ .

*Démonstration.* Il s'agit d'un calcul immédiat. □

On en déduit le Corollaire suivant.

**Corollaire 3.1.** Supposons que  $P(X) = (X - \lambda_1) \dots (X - \lambda_n)$  a  $n$  racines distinctes. Dans ce cas, il existe  $\alpha_1, \dots, \alpha_n \in \mathbb{K}$  tels que l'unique solution de (9.13) s'écrive pour tout  $t$  :

$$y_t = \sum_{k=1}^n \alpha_k \exp(\lambda_k t).$$

Pour trouver  $\alpha_1, \dots, \alpha_n$ , il suffit alors de résoudre le système de  $n$  équations à  $n$  inconnues :

$$\forall i \in \{1, \dots, n\}, \quad \sum_{k=1}^n \alpha_k (\lambda_k)^i = y_i.$$

*Démonstration.* Grâce au Lemme 3.4, les solutions  $y^{\lambda_1}, \dots, y^{\lambda_n}$  forment bien une famille indépendante de  $n$  éléments, et donc une base de l'ensemble des solutions. □

Voici quelques exercices pour aller plus loin : Que se passe-t-il quand le polynôme a des racines multiples ? Quelles sont les solutions réelles ? Quelle est le lien avec les exponentielles de matrices ?

**Exercice 3.8** (Cas des polynômes à racines multiples). 1. Montrer que  $\lambda$  est une racine de  $P$  de multiplicité  $k$  si et seulement si  $P(\lambda) = P'(\lambda) = \dots = P^{(k)}(\lambda) = 0$ .

2. En déduire que  $\lambda$  est une racine de multiplicité  $k$  si et seulement si les  $k$  fonctions  $t \mapsto \exp(\lambda t)$ ,  $t \mapsto t \exp(\lambda t)$ ,  $\dots$ ,  $t \mapsto t^k \exp(\lambda t)$  sont solutions de  $y_t^{(n)} = a_{n-1}y_t^{(n-1)} + \dots + a_0y_t$ .

3. On écrit  $P$  sous la forme  $P(X) = (X - \lambda_1)^{k_1} \dots (X - \lambda_p)^{k_p}$ . Donner une base de l'ensemble des solutions de  $y_t^{(n)} = a_{n-1}y_t^{(n-1)} + \dots + a_0y_t$ .

**Exercice 3.9** (Solutions réelles). On suppose que  $a_0, \dots, a_{n-1}$  sont des réels. Le polynôme  $P$  est donc lui aussi réel. Soit  $\lambda$  une racine de  $P$  qui n'est pas réelle.

1. Montrer que  $\bar{\lambda}$  est également une racine de  $P$ , de même multiplicité que  $\lambda$ .

On note  $k$  cette multiplicité, et  $\lambda = a + ib$ ,  $a, b \in \mathbb{R}$ .

2. Montrer que pour  $l = 1, \dots, k$ ,  $t \mapsto t^l \exp(at) \cos(bt)$  et  $t \mapsto t^l \exp(at) \sin(bt)$  sont solutions.

3. En déduire une base de solutions réelles de  $y_t^{(n)} = a_{n-1}y_t^{(n-1)} + \dots + a_0y_t$ .

**Exercice 3.10** (En passant par les exponentielles de matrices). Dans cet exercice, on suppose à nouveau que les racines (complexes) de  $P$  sont toutes distinctes. On les note  $\lambda_1, \dots, \lambda_n$ .

1. Montrer que pour tout  $k = 1, \dots, n$  :

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ a_0 & \dots & \dots & \dots & a_{n-1} \end{pmatrix} \begin{pmatrix} 1 \\ \lambda_k \\ \lambda_k^2 \\ \vdots \\ \lambda_k^{n-1} \end{pmatrix} = \lambda_k \begin{pmatrix} 1 \\ \lambda_k \\ \lambda_k^2 \\ \vdots \\ \lambda_k^{n-1} \end{pmatrix}.$$

2. Retrouver grâce au Théorème 3.2 les résultats du Corollaire 3.1. On pourra admettre que la matrice :

$$\begin{pmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_{n-1} \\ \vdots & & \vdots \\ \lambda_1^{n-1} & \dots & \lambda_{n-1}^{n-1} \end{pmatrix}$$

est inversible.

### 3.2 Cas non autonome

On revient à l'équation (9.9), et cette fois, on autorise  $A = A(t)$  à dépendre du temps. On suppose (même si ce n'est pas essentiel) que  $t \mapsto A(t)$  est continue (c'est à dire que tous les coefficients de la matrice sont continus). On repasse dans le cas réel pour se fixer les idées, même si ça ne changerait rien d'être dans l'ensemble des complexes.

#### Semi-groupe associé

Avant toute chose, on cherche à résoudre les problèmes de Cauchy suivants, indexés par  $s \in \mathbb{R}$ , à valeur dans l'ensemble des matrices :

$$\begin{cases} \frac{d}{dt}S(t, s) = A(t)S(t, s), \\ S(s, s) = \text{id}. \end{cases} \quad (9.14)$$

C'est également un problème de Cauchy linéaire, mais à valeurs dans l'ensemble des matrices carrées, de taille  $d$ . Le théorème de Cauchy-Lipschitz (Théorème 4.1) que l'on verra plus tard garantit un résultat d'existence et d'unicité pour ce problème de Cauchy.

**Théorème 3.11.** *Soit  $s \in \mathbb{R}$ , le problème de Cauchy (9.14) admet une unique solution maximale, et elle est définie pour tout  $t \in \mathbb{R}$ .*

*Démonstration.* C'est une conséquence directe du Théorème 4.1. □

*Remarque 3.12.* Lorsque  $A(t) \equiv A$  est constante, on vérifie facilement que pour tout  $s, t \in \mathbb{R}$  :

$$S(t, s) = \exp((t - s)A).$$

Sauf cas très particuliers, on ne peut pas espérer obtenir des formules explicites pour la matrice  $S(t, s)$ ,  $s, t \in \mathbb{R}$ . En revanche, la famille  $(S(t, s))_{s, t \in \mathbb{R}}$  satisfait la proposition suivante (on dit que c'est un *semi-groupe*) :

**Proposition 3.13.** *Soient  $t_0, t_1, t_2 \in \mathbb{R}$ . On a :*

$$S(t_2, t_1) \times S(t_1, t_0) = S(t_2, t_0).$$

*Démonstration.* Considérons les deux courbes de matrices :

$$T_1 : t \mapsto S(t, t_1) \times S(t_1, t_0) \quad \text{et} \quad T_2 : t \mapsto S(t, t_0).$$

On a pour  $i = 1, 2$  par des calculs directs :

$$T_i(t_1) = S(t_1, t_0) \quad \text{et} \quad \dot{T}_i(t) = A(t)T_i(t).$$

$T_1$  et  $T_2$  sont donc deux solutions du même problème de Cauchy, et par la partie unicité du Théorème 4.1, on a pour tout  $t \in \mathbb{R}$ ,  $T_1(t) = T_2(t)$ . On obtient donc le résultat en choisissant  $t = t_2$ . □

*Remarque 3.14.* Dans le cas où  $A(t) \equiv A$  est constante, on retrouve simplement :

$$\exp((t_2 - t_1)A) \times \exp((t_1 - t_0)A) = \exp((t_2 - t_0)A).$$

*Remarque 3.15.* En particulier, pour tout  $s, t \in \mathbb{R}$  :

$$S(s, t) \times S(t, s) = S(s, s) = \text{id}.$$

Donc pour tout  $s, t \in \mathbb{R}$ ,  $S(s, t)$  est inversible et  $S(s, t)^{-1} = S(t, s)$ .

Dérivons l'identité précédente par rapport à  $t$ . On obtient :

$$\left(\frac{d}{dt}S(s, t)\right) \times S(t, s) + S(s, t) \times \left(\frac{d}{dt}S(t, s)\right) = 0.$$

En utilisant la définition de  $S$  dans le deuxième terme, on obtient :

$$\left(\frac{d}{dt}S(s, t)\right) \times S(t, s) + S(s, t) \times A(t) \times S(t, s) = 0.$$

Enfin, en multipliant par  $S(s, t)$  à droite, on déduit :

$$\frac{d}{dt}S(s, t) = -S(s, t) \times A(t). \quad (9.15)$$

### Conséquence sur le problème de Cauchy

De ces considérations sur le problème de Cauchy (9.14), on déduit<sup>4</sup> le théorème suivant d'existence et d'unicité pour le problème de Cauchy (9.9).

**Théorème 3.16.** *L'unique solution maximale du problème de Cauchy (9.9) est la courbe :*

$$y : t \in \mathbb{R} \mapsto S(t, 0)y_0,$$

qui est bien définie pour tout  $t \in \mathbb{R}$ .

En particulier, l'ensemble des solutions de  $\dot{y}_t = A(t)y_t$  est un sous-espace vectoriel de dimension  $d$  de  $C^0(\mathbb{R}; \mathbb{R}^d)$ .

*Démonstration.* On commence à avoir l'habitude. La formule donnée dans l'énoncé du théorème est clairement solution. Concernant l'unicité, si  $y = y_t$  est une solution, posons  $z_t := S(0, t)y_t$ . En utilisant (9.15), on trouve :

$$\dot{z}_t = \left(\frac{d}{dt}S(0, t)\right) y_t + S(0, t)\dot{y}_t = -S(0, t)A(t)y_t + S(0, t)A(t)y_t = 0.$$

Donc pour tout  $t$ ,  $z_t = z_0 = y_0$ , et on obtient  $y_t = S(t, 0)y_0$  en multipliant par  $S(t, 0)$ .

La propriété de sous-espace vectoriel se fait comme dans le Théorème 3.2. □

### Équation avec second membre

Dans ce paragraphe, on va chercher à généraliser la formule (9.8). Il se trouve que l'on s'est placé dans le bon cadre pour pouvoir le faire, et on va donner une formule pour les solutions de l'équation avec second membre faisant intervenir le semi-groupe. On montre le Théorème suivant :

**Théorème 3.17.** *L'unique solution du problème de Cauchy :*

$$\begin{cases} \dot{y}_t = A(t)y_t + B(t), \\ y|_{t=t_0} = y_0, \end{cases}$$

où  $t \in \mathbb{R} \mapsto B(t) \in \mathbb{R}^d$  est continue est donnée par la formule dite formule de Duhamel :

$$y_t = S(t, 0)y_0 + \int_0^t S(t, s)B(s) ds.$$

---

4. En fait, on pourrait aussi le faire directement avec Cauchy-Lipschitz, mais il me semble qu'il est instructif de voir ce qu'est le semi-groupe.

*Remarque 3.18.* Évidemment, quand  $A(t) \equiv A$  est constante, on trouve la formule de Duhamel :

$$y_t = \exp(tA)y_0 + \int_0^t \exp((t-s)A)B(s) ds.$$

*Démonstration.* On vérifie que la formule fonctionne. Si  $y_t$  est donnée par la formule et  $t \in \mathbb{R}$  :

$$\begin{aligned} \dot{y}_t &= A(t)S(t,0)y_0 + \int_0^t A(t)S(t,s)B(s) ds + B(t) \\ &= A(t) \left( S(t,0)y_0 + \int_0^t S(t,s)B(s) ds \right) + B(t) \\ &= A(t)y_t + B(t), \end{aligned}$$

d'où le résultat.

Si  $y$  est une solution, posons  $z : t \mapsto S(0,t)y_t$ . On a alors par les mêmes calculs qu'au paragraphe précédent :

$$\begin{aligned} \dot{z}_t &= -S(0,t)A(t)y_t + S(0,t) \{ A(t)y_t + B(t) \} \\ &= S(0,t)B(t). \end{aligned}$$

Donc pour tout  $t \geq 0$  :

$$z_t = z_0 + \int_0^t S(0,s)B(s) ds = y_0 + \int_0^t S(0,s)B(s) ds.$$

Puis en multipliant par  $S(t,0)$  et en utilisant la propriété de semi-groupe :

$$\begin{aligned} y_t &= S(t,0)y_0 + S(t,0) \int_0^t S(0,s)B(s) ds \\ &= S(t,0)y_0 + \int_0^t S(t,0) \times S(0,s)B(s) ds \\ &= S(t,0)y_0 + \int_0^t S(t,s)B(s) ds, \end{aligned}$$

ce qui conclut la preuve. □

## 4 Équations différentielles non linéaires

Dans cette section, on reprend en toute généralité le problème de Cauchy (EDO). On commence par montrer que lorsque le champ de vecteur  $F$  est lipschitzien, alors il y a toujours existence et unicité des solutions.

### 4.1 Le théorème de Cauchy-Lipschitz

Dans cette sous-section, on suppose donc que  $F : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  est une fonction continue en ses deux variables, et qu'il existe  $K > 0$  tels que pour tout  $t \in \mathbb{R}$  et pour tous  $x, y \in \mathbb{R}^d$ , on a :

$$\|F(t, y) - F(t, x)\| \leq K\|y - x\| \tag{9.16}$$

Dans ce cas, il y a toujours existence et unicité de solutions globales (c'est à dire définies sur tout  $\mathbb{R}$ ). C'est l'objet du théorème de Cauchy-Lipschitz, probablement un des théorèmes les plus importants de l'analyse.

**Théorème 4.1.** *Si  $F$  est comme ci-dessus, il existe une unique solution globale au problème de Cauchy (EDO).*

*Si  $F$  n'est défini que sur  $I \times U$ ,  $I \subset \mathbb{R}$  intervalle contenant  $t_0$ ,  $U$  ouvert, et où si (9.16) n'est vérifiée que localement au voisinage de  $(t_0, y_0) \in I \times U$  alors le résultat reste vrai en remplaçant solution globale par solution maximale.*

*Idée de la preuve.* On ne fait qu'esquisser la preuve car elle demande des notions que l'on n'a pas abordé. On ne s'embêtera pas non plus avec les détails concernant la régularité des solutions. Cependant, il me semble que ça vaut le coup de retenir l'idée, car elle peut s'adapter à une quantité incroyable de situations différentes. On se place dans le premier cas, où  $F$  est défini et lipschitzienne partout sur  $\mathbb{R} \times \mathbb{R}^d$ .

Le premier point consiste à remarquer que  $y : \mathbb{R} \rightarrow \mathbb{R}^d$  est solution si et seulement si pour tout  $t \in \mathbb{R}$  :

$$y_t = y_0 + \int_{t_0}^t F(s, y_s) ds.$$

On introduit donc un schéma itératif :

- On note  $y^0$  la fonction constante, égale à  $y_0$ ,
- Si  $y^k : \mathbb{R} \rightarrow \mathbb{R}^d$  est définie et continue pour  $k \in \mathbb{N}$ , on définit pour  $t \in \mathbb{R}$  :

$$y_t^{k+1} = y_0 + \int_{t_0}^t F(s, y_s^k) ds. \quad (9.17)$$

Il suffit alors de montrer que sur tout intervalle  $J$  borné, la suite  $(y^k)$  converge uniformément vers une fonction  $y$ . En effet, dans ce cas, on peut passer à la limite dans (9.17) et en déduire le résultat.

Pour le faire, il suffit de remarquer que si l'intervalle  $J$  est borné, de longueur  $L$ , et si  $k \geq 1$  :

$$\begin{aligned} \sup_{t \in J} \|y_t^{k+1} - y_t^k\| &= \sup_{t \in J} \left\| \int_{t_0}^t (F(s, y_s^k) - F(s, y_s^{k-1})) ds \right\| \\ &\leq \sup_{t \in J} \int_{t_0}^t \|F(s, y_s^k) - F(s, y_s^{k-1})\| ds \\ &\leq \sup_{t \in J} \int_{t_0}^t K \|y_s^k - y_s^{k-1}\| ds \\ &\leq KL \sup_{t \in J} \|y_t^k - y_t^{k-1}\|. \end{aligned}$$

En conséquence, si  $KL \leq 1/2$ , les courbes  $(y^k)$  vont se rapprocher de plus en plus, et finiront par converger.

Si on n'a pas  $KL \leq 1/2$ , il suffit de découper l'intervalle en sous-intervalles de taille suffisamment petite, puis de raisonner de la même façon sur chaque sous intervalle.  $\square$

## 4.2 Contre-exemples quand le champ de vecteurs n'est pas lipschitzien

On va montrer deux types de phénomènes quand  $F$  ne satisfait pas (9.16). D'abord, il se peut que l'unicité ne perdure pas. Ensuite, il se peut que la solution explose en temps fini. Notez que je ne donne pas de contre-exemple à l'existence locale (c'est à dire sur un petit intervalle de temps). C'est parce qu'en fait, on peut obtenir des résultats d'existence sous des hypothèses beaucoup plus faibles (typiquement, avec la même preuve mais en remplaçant  $F$  par des approximations lipschitziennes). Par exemple, il y a toujours existence de solutions sous la seule hypothèse  $F$  continue.

*Exemple 4.2* (Contre-exemple à l'unicité). On se place en dimension 1, dans le cas autonome, avec pour tout  $x \in \mathbb{R}$  :

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ \frac{\sqrt{x}}{2} & \text{si } x \geq 0. \end{cases}$$

On cherche alors à résoudre le problème de Cauchy :

$$\begin{cases} \dot{y}_t = F(y_t), \\ y_0 = 0. \end{cases}$$

On remarque alors que pour tout  $t_c \geq 0$ , la fonction :

$$y : t \in \mathbb{R} \mapsto \begin{cases} 0 & \text{si } t \leq t_c, \\ (t - t_c)^2 & \text{si } t \geq t_c, \end{cases}$$

est solution, violant ainsi l'unicité.

*Exemple 4.3* (Explosion en temps fini). Ici, on se place également en dimension 1 dans le cas autonome, mais cette fois avec  $F$  définie pour  $x \in \mathbb{R}$  par :

$$F(x) = x^2.$$

On cherche à résoudre le problème de Cauchy :

$$\begin{cases} \dot{y}_t = y_t^2, \\ y_0 = 1. \end{cases}$$

Notons que  $F$  n'est pas lipschitzienne, mais elle est bien localement lipschitzienne (sa restriction à tout intervalle borné est lipschitzienne). En vertu du théorème de Cauchy-Lipschitz, ce problème admet une unique solution maximale.

Soit donc  $y : J \rightarrow \mathbb{R}$  l'unique solution maximale de notre problème de Cauchy. D'abord, on est sûrs que  $y$  ne s'annule pas. En effet, si  $y$  s'annulait au temps  $t_0$ , alors  $y$  serait aussi solution du problème de Cauchy :

$$\begin{cases} \dot{y}_t = y_t^2, \\ y_{t_0} = 0, \end{cases}$$

dont l'unique solution maximale est la fonction nulle. On aboutit donc à une absurdité.

Comme  $y$  ne s'annule pas, pour tout  $t \in J$ , on a :

$$\frac{d}{dt} \frac{1}{y_t} = -\frac{\dot{y}_t}{y_t^2} = -1.$$

En intégrant cette égalité entre 0 et  $t$ , on obtient donc pour tout  $t \in J$  :

$$1 - \frac{1}{y_t} = t,$$

qui se réécrit :

$$y_t = \frac{1}{1 - t}.$$

On a donc  $J = ] - \infty, 1[$ , et pour tout  $t \in ] - \infty, 1[$  :

$$y_t = \frac{1}{1 - t},$$

qui diverge quand  $t \rightarrow 1^-$ .

### 4.3 Dépendance aux données : le lemme de Grönwall

Dans ce paragraphe, on se demande comment on peut quantifier l'écart entre deux solutions de :

$$\dot{y}_t = F(t, y_t),$$

dans le cas où  $F$  est  $K$ -lipschitzienne en espace, uniformément en temps. On se donne  $y = y_t$  et  $z = z_t$  deux solutions de cette équation différentielle. On a donc pour tout  $t \geq 0$  :

$$y_t = y_0 + \int_0^t F(s, y_s) ds \quad \text{et} \quad z_t = z_0 + \int_0^t F(s, z_s) ds.$$

Si on fait la différence et que l'on passe à la norme, on obtient :

$$\begin{aligned} |z_t - y_t| &= \left| z_0 - y_0 + \int_0^t \{F(s, z_s) - F(s, y_s)\} ds \right| \\ &\leq |z_0 - y_0| + \int_0^t |F(s, z_s) - F(s, y_s)| ds \\ &\leq |z_0 - y_0| + K \int_0^t |z_s - y_s| ds. \end{aligned}$$

Il se trouve qu'en vertu du lemme suivant, on peut alors conclure que :

$$|z_t - y_t| \leq \exp(Kt) |z_0 - y_0|. \quad (9.18)$$

*Lemme 4.4* (Lemme de Grönwall). Soit  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  une fonction continue vérifiant pour tout  $t \geq 0$  :

$$f(t) \leq f(0) + K \int_0^t f(s) ds \quad (9.19)$$

Alors

$$f(t) \leq f(0) \exp(Kt).$$

*Démonstration.* On pose pour  $t \geq 0$  :

$$h(t) := \left( \int_0^t f(s) ds \right) \exp(-Kt).$$

La fonction  $h$  est dérivable et pour tout  $t \geq 0$  :

$$\begin{aligned} h'(t) &= \left\{ f(t) - K \left( \int_0^t f(s) ds \right) \right\} \exp(-Kt) \\ &\leq f(0) \exp(-Kt). \end{aligned}$$

Comme  $h(0) = 0$ , on en déduit que :

$$h(t) \leq f(0) \frac{1 - \exp(-Kt)}{K}.$$

En multipliant par  $\exp(Kt)$ , on en déduit :

$$\int_0^t f(s) ds \leq f(0) \frac{\exp(Kt) - 1}{K}.$$

Enfin, on utilise (9.19) pour conclure. □

Revenons sur l'inégalité (9.18). D'une part, cette estimation est bonne : elle signifie que la position à l'instant  $t$  de la solution du problème de Cauchy (EDO), que l'on note  $\Phi(t, y_0) := y_t$  est elle même lipschitzienne par rapport à  $y_0$ . En particulier, si  $y^n \rightarrow y$ , on a pour tout  $t$ ,  $\Phi(t, y^n) \rightarrow \Phi(t, y)$ . En revanche, la constante de Lipschitz est ici  $\exp(Kt)$ , et devient donc très vite catastrophique ! En pratique, deux solutions qui partent de points très proches peuvent engendrer des solutions qui s'éloignent très vite !

#### 4.4 Le cas autonome : notion de stabilité non-linéaire

Dans ce paragraphe, on se place dans le cas autonome, c'est à dire que  $F : x \in \mathbb{R}^d \mapsto F(x)$  ne dépend pas du temps. On suppose que  $F$  est régulière (disons de classe  $C^2$ , mais on ne verra pas vraiment pourquoi), et lipschitzienne pour être sûrs que les solutions sont définies globalement, même si on pourrait tout localiser en faisant attention. D'après le théorème de Cauchy-Lipschitz, pour tout  $x \in U$ , le problème de Cauchy :

$$\begin{cases} \dot{y}_t = F(y_t), \\ y_0 = x. \end{cases}$$

admet une unique solution globale, que l'on note  $t \in \mathbb{R} \mapsto \Phi(t, x)$ .

On se donne  $x_0 \in U$  tel que  $F(x_0) = 0$ . Dans ce cas, la fonction  $t \in \mathbb{R} \mapsto x_0$  est solution, et donc pour tout  $t \in \mathbb{R}$ ,  $\Phi(t, x_0) = x_0$ .

Pour des temps  $t$  petit et des  $x$  proches de  $x_0$ , on s'attend à ce que :

$$\|\Phi(t, x) - x_0\| \ll 1.$$

On note donc  $r(t, x) := \Phi(t, x) - x_0$ , et on admet que l'on peut justifier le calcul au premier ordre suivant :

$$\frac{d}{dt}r(t, x) = F(\Phi(t, x)) = F(x_0 + r(t, x)) \approx DF(x_0) \cdot r(t, x).$$

Au premier ordre,  $r$  est donc solution du problème de Cauchy linéaire et autonome<sup>5</sup> :

$$\begin{cases} \frac{d}{dt}r(t, x) = DF(x_0) \cdot r(t, x), \\ r(0, x) = x - x_0. \end{cases} \quad (9.20)$$

En conséquence, à l'instar de ce que l'on a vu au Théorème 3.6, trois cas peuvent survenir :

- Ou bien toutes les valeurs propres (complexe) de  $DF(x_0)$  ont une partie réelle strictement négative. Dans ce cas, si  $x$  est suffisamment proche de  $x_0$  pour que l'approximation soit justifiée,  $r(t, x) \rightarrow 0$  quand  $t \rightarrow +\infty$ .
- Ou bien au moins une valeur propre (complexe) de  $DF(x_0)$  est strictement positive, et alors pour des  $x$  bien choisis aussi proches que l'on veut de  $x_0$ ,  $\Phi(t, x)$  commence par s'éloigner de  $x_0$ , jusqu'à ce que l'approximation ne soit plus valable.
- Ou bien  $DF(x_0)$  n'a que des valeurs propres (complexes) de partie réelle strictement négative et au moins une valeur propre de partie réelle nulle, et dans ce cas, au niveau non linéaire, tout peut arriver (existence de solutions périodiques, qui se rapprochent, qui s'éloignent).

On pourrait alors justifier rigoureusement le théorème suivant.

**Théorème 4.5.** Soit  $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$  une fonction lisse et  $x_0 \in \mathbb{R}^d$  avec  $F(x_0) = 0$ .

- Si toutes les valeurs propres de  $DF(x_0)$  ont une partie réelle strictement négative, alors il existe un voisinage  $V$  de  $x_0$  tel que :

$$\forall x \in V, \quad \lim_{t \rightarrow +\infty} \Phi(t, x) = x_0.$$

On dit que le point d'équilibre  $x_0$  est **asymptotiquement stable**.

- S'il existe une valeur propre de  $DF(x_0)$  qui a une partie réelle strictement positive, alors il existe un voisinage  $V$  de  $x_0$ , une suite  $(x_n)$  convergeant vers  $x_0$  et une suite de temps  $(t_n > 0)$  tels que :

$$\Phi(t_n, x_n) \notin V.$$

On dit que le point d'équilibre  $x_0$  est **instable au sens de Lyapounov**.

5. La signification de cette phrase n'est pas claire, je l'accorde. En fait,  $r$  est solution d'une équation différentielle qui est très proche, et ce que l'on admet, c'est que deux équations proches admettent des solutions proches.

*Remarque 4.6* (En dimension 1). Si  $d = 1$ , la condition de stabilité est  $F'(x_0) < 0$  (en particulier  $F$  est positive puis négative), et la condition d'instabilité est  $F'(x_0) > 0$  (et donc  $F$  est négative puis positive). Cela se comprend très bien : dans le premier cas, si  $\Phi(t, x)$  est inférieur et proche de  $x_0$ ,  $F(\Phi(t, x)) > 0$  et donc  $\Phi(t, x)$  croît avec  $t$  de sorte qu'il se rapproche de  $x_0$ . De même, si  $\Phi(t, x)$  est à droite et proche de  $x_0$ ,  $F(\Phi(t, x)) < 0$  et donc  $\Phi(t, x)$  décroît avec  $t$  de sorte qu'une fois encore, il se rapproche de  $x_0$ . Dans le second cas, on peut faire le même type de raisonnement, mais dans les deux cas, si  $\Phi(t, x)$  est proche de  $x_0$ , alors  $\Phi(t, x)$  s'éloigne de  $x_0$ .

*Remarque 4.7* (Dépendance aux données dans le cas stable). Dans le cas où  $DF(x_0)$  n'a que des valeurs propres de partie réelle strictement négative, comme dans le cas linéaire, les solutions convergent exponentiellement vite vers l'état stationnaire. De plus, comme au voisinage de  $x_0$ , les solutions sont de la forme  $x_0 + r$  où  $r$  est une solution approchée de l'équation linéaire limite (9.20), la différence entre deux solutions est également solution approchée de (9.20), et converge donc vers 0 exponentiellement vite. En particulier, au voisinage d'un tel point stationnaire les solutions se rapprochent exponentiellement vite, ce qui tranche avec la Sous-section 4.3.

## 4.5 Flots de gradient

Soit  $G : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction régulière (au moins de classe  $C^1$ , et de gradient lipschitzien, disons globalement pour ne pas se poser de questions quant au temps d'existence des solutions maximales). Le flot de gradient associé à  $G$  est l'équation différentielle :

$$\dot{y}_t = -\nabla G(y_t). \quad (9.21)$$

Cela s'interprète de la façon suivante : à chaque instant, la vitesse de la courbe  $y$  est orientée sur la ligne de plus grande pente, dans le sens de la descente, et est d'autant plus grande que la pente est grande. Il semble légitime lorsque l'on cherche les minima de  $G$  de résoudre une telle équation : si vous êtes sur une montagne, il semble légitime pour atteindre le fond de la vallée de toujours descendre selon la ligne de plus grande pente. Si vous ne restez pas piégés dans un minimum local (au fond d'un lac d'altitude par exemple) cela devrait marcher... C'est précisément dans cet objectif que l'on utilise les flots de gradients.

En vertu du Théorème de Cauchy-Lipschitz 4.1, pour tout  $x \in \mathbb{R}^d$ , l'équation différentielle (9.21) admet une unique solution partant de  $x$ , et on note comme précédemment  $\Phi(t, x)$  sa position à l'instant  $t \in \mathbb{R}$ . La première chose à vérifier, c'est que pour tout  $x$ , la fonction  $t \mapsto G(\Phi(t, x))$  est décroissante.

**Proposition 4.8.** *Soit  $x \in \mathbb{R}^d$ . La fonction  $t \mapsto G(\Phi(t, x))$  est de classe  $C^1$ , et pour tout  $t \in \mathbb{R}$  :*

$$\frac{d}{dt}G(\Phi(t, x)) = -\|\nabla G(\Phi(t, x))\|^2 = -\left\|\frac{d}{dt}\Phi(t, x)\right\|^2.$$

*En particulier, elle est décroissante, et admet donc une limite dans  $[-\infty, G(x)]$  quand  $t \rightarrow +\infty$ .*

*Démonstration.* Il s'agit d'un calcul direct en utilisant la formule de dérivation d'une composée :

$$\frac{d}{dt}G(\Phi(t, x)) = \left\langle \nabla G(\Phi(t, x)), \frac{d}{dt}\Phi(t, x) \right\rangle,$$

et en substituant le vecteur de gauche ou le vecteur de droite grâce à l'équation différentielle (9.21). □

**Les points d'accumulation sont des points critiques** Rie ne garantit *a priori* que  $\Phi(t, x)$  converge (d'ailleurs, si  $d = 1$  et  $G(x) = -x$ , toute solution tend vers  $+\infty$  quand  $t \rightarrow +\infty$ ). Dans ce paragraphe, on va montrer que si la solution admet un point d'accumulation, alors ce point d'accumulation est un point critique de  $G$ , qui ne peut être un maximum local. Je ne vais pas le montrer, mais bien que les points selles soient des limites possibles de flots de gradient, il faut très mal choisir  $x$  pour que  $\Phi(t, x)$  converge vers un point selle (typiquement, il faut le choisir sur un ensemble de codimension 1). Donc par exemple si on lance un calcul de flot de gradient sur l'ordinateur et que l'on observe que le programme converge, on peut légitimement considérer le point limite comme un minimum local de la fonction  $G$ .

**Proposition 4.9.** Soient  $G$  et  $\Phi$  comme précédemment. Si  $x_\infty \in \mathbb{R}^d$  est un point d'accumulation de  $t \mapsto \Phi(t, x)$  quand  $t \rightarrow +\infty$ , alors  $\nabla G(x_\infty) = 0$ . De plus  $x_\infty$  n'est pas un point de maximum local de  $G$ .

*Démonstration.* D'abord, on a par la Proposition 4.8 :

$$\int_0^t \left\| \frac{d}{dt} \Phi(s, x) \right\|^2 ds = G(x) - G(\Phi(t, x)),$$

qui converge quand  $t \rightarrow +\infty$ . Soit  $(t_n) \in \mathbb{R}_+^{\mathbb{N}}$  telle que  $\Phi(t_n, x) \rightarrow x_\infty$ . On voudrait dire qu'à cause de la convergence de l'intégrale, on devrait avoir :

$$\left\| \frac{d}{dt} \Phi(t_n, x) \right\|^2 = \|\nabla G(\Phi(t_n, x))\|^2 \xrightarrow{n \rightarrow +\infty} 0,$$

ce qui permettrait de conclure  $\nabla G(x_\infty) = \lim_{n \rightarrow +\infty} \nabla G(\Phi(t_n, x)) = 0$ . Mais on ne peut pas déduire de la convergence de l'intégrale la convergence vers 0 des vitesses : une fonction même positive dont l'intégrale converge ne converge pas nécessairement vers 0. En revanche, on va construire pour  $n \in \mathbb{N}$   $\tilde{t}_n$  proche de  $t_n$  tels que  $\Phi(\tilde{t}_n, x) \rightarrow x_\infty$  et  $\nabla G(\Phi(\tilde{t}_n, x)) \rightarrow 0$ , ce qui permettra de conclure.

Sans perte de généralité, on peut supposer  $(t_n)$  croissante et pour tout  $n$ ,  $t_{n+1} - t_n \geq 1$ . Dans ce cas, la convergence de l'intégrale implique :

$$\int_{t_n}^{t_{n+1}} \left\| \frac{d}{dt} \Phi(s, x) \right\|^2 ds \xrightarrow{n \rightarrow +\infty} 0.$$

Il existe donc  $\tilde{t}_n \in [t_n, t_n + 1]$  tel que :

$$\left\| \frac{d}{dt} \Phi(\tilde{t}_n, x) \right\|^2 \leq \int_{t_n}^{t_{n+1}} \left\| \frac{d}{dt} \Phi(s, x) \right\|^2 ds \xrightarrow{n \rightarrow +\infty} 0.$$

Il suffit donc de vérifier que  $\Phi(\tilde{t}_n, x) \rightarrow x_\infty$ . Mais on a :

$$\begin{aligned} \|\Phi(\tilde{t}_n, x) - x_\infty\| &\leq \|\Phi(\tilde{t}_n, x) - \Phi(t_n, x)\| + \|\Phi(t_n, x) - x_\infty\| \\ &\leq \int_{t_n}^{\tilde{t}_n} \left\| \frac{d}{dt} \Phi(s, x) \right\| ds + \|\Phi(t_n, x) - x_\infty\| \\ &\leq \sqrt{\int_{t_n}^{t_{n+1}} \left\| \frac{d}{dt} \Phi(s, x) \right\|^2 ds} + \|\Phi(t_n, x) - x_\infty\| \\ &\xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

Enfin,  $x_\infty$  ne peut pas être un point de maximum local de  $G$ , car d'une part  $\Phi(t_n, x) \rightarrow x_\infty$ , et d'autre part, par la Proposition 4.8,  $G(x_\infty) = \lim_{t \rightarrow +\infty} G(\Phi(t, x)) \leq G(\Phi(t_n, x))$ , pour tout  $n$ .  $\square$

**Propriété de stabilité des solutions dans le cas convexe** Dans le cas où la fonction  $G$  est convexe, on obtient gratuitement le résultat de stabilité suivant :

**Proposition 4.10.** En plus des hypothèses et définitions des paragraphes précédents, on suppose que  $G$  est convexe. Alors pour tout  $x, y \in \mathbb{R}^d$  :

$$t \mapsto \|\Phi(t, y) - \Phi(t, x)\|$$

est une fonction décroissante.

*Démonstration.* Il suffit de calculer pour  $x, y \in \mathbb{R}^d$  et  $t \in \mathbb{R}$  :

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \|\Phi(t, y) - \Phi(t, x)\|^2 &= \left\langle \frac{d}{dt} \Phi(t, y) - \Phi(t, x), \Phi(t, y) - \Phi(t, x) \right\rangle \\ &= - \left\langle \nabla G(\Phi(t, y)) - \nabla G(\Phi(t, x)), \Phi(t, y) - \Phi(t, x) \right\rangle \\ &\leq 0, \end{aligned}$$

par la propriété de monotonie bien connue du gradient des fonctions convexes.  $\square$

Les trajectoire du flot de gradient d'une fonction convexe se rapprochent donc à mesure que le temps avance, ce qui une fois encore tranche radicalement avec le cas général des équations différentielles où des trajectoires peuvent s'éloigner exponentiellement vite (Sous-section 4.3).

**Conséquence : convergence du flot vers les min globaux** La propriété de stabilité permet de démontrer le résultat suivant :

**Théorème 4.11.** *Soit  $G$  une fonction de classe  $C^1$ , convexe, de gradient lipschitzien<sup>6</sup>, et  $\Phi = \Phi(t, x)$  le flot associé. On suppose que minimum de  $G$  est atteint. Alors pour tout  $x \in \mathbb{R}^d$ ,  $t \mapsto \Phi(t, x)$  converge, et sa limite est un point de minimum global de  $G$ .*

*Démonstration.* Soit  $x \in \mathbb{R}^d$ . D'abord, on montre que  $t \in \mathbb{R}_+ \mapsto \Phi(t, x)$  a ses valeurs dans un ensemble borné. Soit  $\bar{x}$  un point de minimum de  $G$ . On a  $\nabla G(\bar{x}) = 0$ , donc  $t \mapsto \bar{x}$  est une solution du flot de gradient, ou en d'autres termes, pour tout  $t \in \mathbb{R}$ ,  $\Phi(t, \bar{x}) = \bar{x}$ . Donc par la Proposition 4.10, on a pour tout  $t \geq 0$  :

$$\|\Phi(t, x) - \bar{x}\| \leq \|x - \bar{x}\|,$$

d'où le fait que  $t \mapsto \Phi(t, x)$  est bornée.

Par le théorème de Bolzano-Weierstrass, il existe donc  $(t_n) \in \mathbb{R}_+^{\mathbb{N}}$  et  $x_\infty$  tels que :

$$\Phi(t_n, x) \xrightarrow{n \rightarrow +\infty} x_\infty.$$

Par conséquent, d'après la Proposition 4.9,  $\nabla G(x_\infty) = 0$ . Par un résultat que l'on a vu dans le Chapitre sur les fonctions convexes,  $x_\infty$  est donc un point de minimum global de  $G$ .

La dernière chose à faire est de montrer que  $\Phi(t, x) \rightarrow x_\infty$ . On va utiliser le fait que  $\Phi(t, x_\infty) = x_\infty$ , pour tout  $t \in \mathbb{R}$  et la Proposition 4.10. Soit  $\varepsilon > 0$  et  $n \in \mathbb{N}$  tel que  $\|\Phi(t_n, x) - x_\infty\| \leq \varepsilon$ . Pour tout  $t \geq t_n$  :

$$\|\Phi(t, x) - x_\infty\| = \|\Phi(t, x) - \Phi(t, x_\infty)\| \leq \|\Phi(t_n, x) - \Phi(t_n, x_\infty)\| = \|\Phi(t_n, x) - x_\infty\| \leq \varepsilon,$$

d'où le résultat.  $\square$

Il y aurait encore bien des choses à dire, on pourrait être beaucoup plus précis par biens des aspects, mais j'espère avec ce chapitre avoir pu donner un premier aperçu des questions que l'on se pose lorsque l'on étudie des équations différentielles, et le type de réponses générales que l'on peut apporter.

---

6. Localement lipschitzien suffit.