Notes for the course

MATHEMATICS FOR ECONOMISTS

Guillaume CARLIER

MQEF, X-HEC, academic year 2008-2009

This set of notes gathers previous versions of courses taught at Dauphine and ENSAE together with additional material specially intended for this course. This explains why some parts are written in French and why notations may vary from one part to the other. There are certainly many typos in the current version, so feel free to make suggestions for an improved presentation.

Contents

Part I

Topology

Chapter 1

Metric spaces

1.1 Basic definitions

Définition 1.1 *Soit E un ensemble non vide. On appelle distance sur E toute application $d : E \times E \to \mathbb{R}_+$ vérifiant les propriétés:*

1. *(symétrie) $d(x, y) = d(y, x)$ pour tout $(x, y) \in E \times E$,*
2. *$d(x, y) = 0 \Leftrightarrow x = y$*
3. *(inégalité triangulaire) $d(x, z) \leq d(x, y) + d(y, z)$ pour tout $(x, y, z) \in E \times E \times E$.*

On appelle espace métrique la donnée d'un couple (E, d) où d est une distance sur E.

Bien noter que dans la définition précédente on a $d \geq 0$. Noter également que la définition précédente implique aussi $|d(x, z) - d(y, z)| \leq d(x, y)$, pour tout $(x, y, z) \in E \times E \times E$.

Exemple 1.1 *Pour $E = \mathbb{R}$, $d(x, y) := |x - y|$ est la distance usuelle. Pour $E = \mathbb{R}^n$, $x = (x_1, ..., x_n)$ et $y = (y_1, ..., y_n)$ on considère souvent les distances:*

$$d_1(x, y) := \sum_{i=1}^{n} |x_i - y_i|, \ d_\infty(x, y) := \max_{i=1,..,n} |x_i - y_i|$$

et la distance euclidienne:

$$d_2(x, y) := (\sum_{i=1}^{n} (x_i - y_i)^2)^{\frac{1}{2}}.$$

Exemple 1.2 *Soit E un ensemble non vide et définissons pour $(x, y) \in E^2$, $d(x, y) = 1$ si $x \neq y$ et $d(x, y) = 0$ si $x = y$ on vérifie aisément que d est une distance sur E (appelée distance grossière sur E).*

Nous verrons par la suite d'autres exemples dans le cadre des espaces vectoriels normés.

1.2 Topology of metric spaces

Soit (E, d) un espace métrique, $x \in E$ et $r > 0$, on notera $B_E(x, r)$ (ou simplement $B(x, r)$ s'il n'y a pas d'ambiguité) la boule ouverte de centre x et de rayon r:

$$B(x, r) := \{y \in E : d(x, y) < r\}$$

et $\overline{B}_E(x, r)$ (ou simplement $\overline{B}(x, r)$ s'il n'y a pas d'ambiguité) la boule fermée de centre x et de rayon $r \geq 0$:

$$\overline{B}(x, r) := \{y \in E : d(x, y) \leq r\}.$$

Le terme de "boule" provient du cas de la distance euclidienne (la distance d_2 définie plus haut). A titre d'exercice, dessinez dans $\mathbb{R}^2$, la boule $\overline{B}(0, 1)$ pour les trois distances d_1, d_2 et d_∞, qu'en pensez vous?

Définition 1.2 *Soit (E, d) un espace métrique et $A \subset E$, on dit que A est bornée ssi il existe $x \in E$ et $r > 0$ tels que $A \subset B(x, r)$.*

Si A est une partie de E, on définit son diamètre $\mathrm{diam}(A)$ par:

$$\mathrm{diam}(A) := \sup\{d(x, y),\ (x, y) \in A^2\}.$$

On vérifie aisément que A est bornée ssi $\mathrm{diam}(A)$ est fini.

On peut maintenant définir les ensembles ouverts de (E, d):

Définition 1.3 *Soit (E, d) un espace métrique et A une partie de E. On dit que:*

1. *A est ouvert ssi pour tout $x \in A$, $\exists r > 0$ tel que $B(x, r) \subset A$,*
2. *A est fermé ssi $E \setminus A$ est ouvert.*
3. *A est un voisinage de $x \in E$ ssi $\exists r > 0$ tel que $B(x, r) \subset A$.*

Autrement dit, un ensemble est ouvert ssi il est voisinage de chacun de ses points. L'ensemble des ouverts de (E, d) s'appelle la topologie de E induite par la distance d. On vérifie aisément qu'une boule ouverte (resp. fermée) est ouverte (resp. fermée).

Proposition 1.1 *Soit (E, d) un espace métrique, on a alors:*

1. *E et $\emptyset$ sont ouverts,*
2. *une réunion (quelconque) d'ouverts est ouverte,*
3. *une intersection FINIE d'ouverts est ouverte.*

La démonstration est élémentaire et laissée au lecteur qui s'entraienera ainsi à se familiariser avec les définitions...

Par passage au complémentaire, on obtient les énoncés correspondant aux fermés:

1. E et $\emptyset$ sont fermés,
2. une réunion FINIE de fermés est fermée,
3. une intersection (quelconque) de fermés est fermée.

Exemple 1.3 *Il est à noter l'importance du mot FINIE dans les énoncés précédents. En effet, soit pour $n \in \mathbb{N}^*$, l'intervalle ouvert $I_n :=]-1/n, 1/n[$, l'intersection de ces ouverts est $\{0\}$ qui n'est pas ouverte. La réunion des intervalles fermés $J_n := [0, 1-1/n]$ est l'intervalle $[0, 1[$ qui n'est ni ouvert ni fermé.*

Définition 1.4 *Soit (E, d) un espace métrique, A une partie de E et $x \in E$ on dit que:*

1. *x est un point intérieur à A ssi $\exists r > 0$ tel que $B(x, r) \subset A$ (autrement dit A est un voisinage de x),*
2. *x est un point adhérent à A ssi $\forall r > 0$, $B(x, r)$ rencontre A.*
3. *x est un point frontière de A ssi $\forall r > 0$, $B(x, r)$ rencontre A et $E \setminus A$.*

On appelle intérieur de A et l'on note $\mathrm{int}(A)$ l'ensemble des points intérieurs de A. On appelle adhérence de A et l'on note $\overline{A}$, l'ensemble des points adhérents à A. On appelle frontière de A et l'on note ∂A l'ensemble des points frontière de A. Enfin on dit que A est dense dans E ssi $\overline{A} = E$.

On a clairement les inclusions:

$$\text{int}(A) \subset A \subset \overline{A},$$

et il est facile de montrer (faites le en exercice...):

$$\partial A = \overline{A} \setminus \text{int}(A).$$

Let us also remark that A is dense in E iff $A \cap U \neq \emptyset$ for every open set U, or equivalently $A \cap B(x,r) \neq \emptyset$ for every $x \in E$ and $r > 0$.

Exemple 1.4 *Il convient de noter que* $\text{int}(A)$ *peut très bien être l'ensemble vide (considérer dans* $\mathbb{R}$*:* $\{0\}$*,* $\mathbb{N}$*,* $\mathbb{Q}$*, un ensemble fini...). Concernant la densité:* $\mathbb{Q}$ *et* $\mathbb{R} \setminus \mathbb{Q}$ *sont denses dans* $\mathbb{R}$*,* $]0,1[$ *est dense dans* $[0,1]$ *etc....*

On a aussi les propriétés importantes:

Proposition 1.2 *Soit* (E,d) *un espace métrique,* A *une partie de* E*, on a:*

1. $\text{int}(A)$ *est ouvert et c'est le plus grand ouvert contenu dans* A*,*
2. $\overline{A}$ *est fermé et c'est le plus petit fermé contenant* A*.*

Preuve:

Montons d'abord que $\text{int}(A)$ est ouvert: soit $x \in \text{int}(A)$ alors $\exists r > 0$ tq $B(x,r) \subset A$, donc si $y \in B(x,r/2)$ on a $B(y,r/2) \subset B(x,r) \subset A$ ce qui montre que $y \in \text{int}(A)$ et donc $B(x,r/2) \subset \text{int}(A)$. $\text{int}(A)$ est donc ouvert et évidemment $\text{int}(A) \subset A$. Montrons maintenant que $\text{int}(A)$ est le plus grand ouvert contenu dans A. Soit U ouvert avec $U \subset A$ et soit $x \in U$, comme U est ouvert $\exists r > 0$ tq $B(x,r) \subset U$ mais comme $U \subset A$ il vient $B(x,r) \subset A$ et donc $x \in \text{int}(A)$ ce qui montre $U \subset \text{int}(A)$ et achève la preuve.

La démonstration du point 2) est similaire et donc laissée au lecteur.

□

L'énoncé précédent implique en particulier les caractérisations:

$$A \text{ ouvert } \Leftrightarrow A = \text{int}(A),$$

et

$$A \text{ fermé } \Leftrightarrow A = \overline{A}.$$

Exercice 1.1 *Soit* (E,d) *un espace métrique et* A *une partie de* E*. Montrer que:*

$$\overline{A} = E \setminus \text{int}(E \setminus A), \ \text{int}(A) = E \setminus \overline{E \setminus A}.$$

Exercice 1.2 *Dans $\mathbb{R}^n$ muni de la distance d_∞ (cf Exemple 1.1), déterminer l'adhérence de $B(x,r)$ et l'intérieur de $\overline{B}(x,r)$.*

Beaucoup de propriétés topologiques dans les espaces métriques peuvent se traduire par des propriétés séquentielles (i.e. en utilisant des suites): retenez ce principe, l'utilisation de suites rend souvent les démonstrations plus simples que le maniement des définitions générales. Rappelons d'abord ce qu'est une suite convergente:

Définition 1.5 *Soit (E,d) un espace métrique et (x_n) une suite d'éléments de E, on dit que $x \in E$ est limite de la suite (x_n) (ce que l'on notera $x_n \to x$ ou $\lim_n x_n = x$) ssi : $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}^*$ t.q. $\forall n \geq N$, $d(x_n,x) \leq \varepsilon$. On dit que (x_n) est convergente si elle admet une limite.*

Quand $\lim_n x_n = x$, on dit aussi que x_n converge vers x. Remarquons que la convergence de (x_n) vers x (dans E) est équivalente à la convergence vers 0 de $d(x_n,x)$ (dans $\mathbb{R}$).

Il convient de noter que si une suite est convergente alors elle admet une UNIQUE limite (cette propriété s'exprime en disant que les espaces métriques sont séparés):

Proposition 1.3 *Soit (E,d) un espace métrique et (x_n) une suite convergente d'éléments de E, alors sa limite est unique.*

Preuve:
Supposons que (x_n) admette pour limite x et y dans E. On a $0 \leq d(x,y) \leq d(x,x_n)+d(x_n,y)$ ainsi en passant à la limite en $n \to +\infty$ on obtient $d(x,y) = 0$ i.e. $x = y$ d'où l'unicité. □

Proposition 1.4 *Soit (E,d) un espace métrique, A une partie de E, on a:*

1. *soit $x \in E$, $x \in \overline{A}$ ssi x est limite d'une suite d'éléments de A,*
2. *A est fermé ssi pour toute suite convergente (x_n) d'éléments de A, la limite de cette suite appartient à A.*

Preuve:
2) découle de 1) et du fait que A est fermé ssi $A = \overline{A}$. Supposons $x \in \overline{A}$, alors pour tout $n \in \mathbb{N}^*$, $B(x,1/n)$ rencontre A, soit donc $x_n \in A \cap B(x,1/n)$ comme $d(x,x_n) \leq 1/n$, x_n converge vers x. Réciproquement supposons que x soit la limite d'une suite (x_n) d'éléments de A montrons que $x \in \overline{A}$. Soit $r > 0$, pour n assez grand $d(x,x_n) < r$ ainsi, comme $x_n \in A$, on a $A \cap B(x,r) \neq \emptyset$. Finalement $r > 0$ étant arbitraire on a bien $x \in \overline{A}$.
□

Exercice 1.3 *En vous inspirant de la démonstration précédente montrer que $x \in \partial A$ ssi x est limite d'une suite d'éléments de A et limite d'une suite d'éléments de $E \setminus A$.*

Exercice 1.4 *Soit (E, d) un espace métrique et A une partie non vide de E. Pour tout $x \in E$ on définit la distance de x à A par:*

$$d(x, A) := \inf\{d(x, a)\ a \in A\}$$

1. *Montrer que $x \in \overline{A}$ ssi $d(x, A) = 0$.*
2. *Montrer que l'ensemble $A_n := \{x \in E : d(x, A) < 1/n\}$ est ouvert $(n \in \mathbb{N}^*)$.*
3. *Déterminer $\cap_{n\in\mathbb{N}^*} A_n$.*
4. *Déduire de ce qui précède que tout fermé peut s'écrire comme une intersection dénombrable d'ouverts.*

1.3 Cauchy sequences, complete spaces

Définition 1.6 *Soit (E, d) un espace métrique et $(x_n)_n$ une suite d'éléments de E, on dit que $(x_n)_n$ est de Cauchy ssi: $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}^*$ t.q. pour tout $(p, q) \in \mathbb{N}^2$ avec $p \geq N$ et $q \geq N$ on a: $d(x_p, x_q) \leq \varepsilon$.*

La définition précédente peut aussi s'exprimer en disant que $(x_n)_n$ est de Cauchy ssi

$$\sup_{p\geq N,\, q\geq N} d(x_p, x_q) \to 0 \text{ quand } N \to +\infty.$$

Evidemment, toute suite convergente est de Cauchy (s'en persuader!), la réciproque n'est cependant pas vraie: les espaces métriques pour lesquels cette réciproque est vraie sont dits complets:

Définition 1.7 *Soit (E, d) un espace métrique, on dit que (E, d) est complet ssi toute suite de Cauchy d'éléments de E converge dans E.*

Exemple 1.5 *Le corps des rationnels $\mathbb{Q}$ muni de la distance usuelle (induite par celle de $\mathbb{R}$) n'est pas complet (en effet, il est facile de vérifier que la suite de rationnels définie par $x_n := \sum_{k=0}^{n} 1/(k!)$ est de Cauchy, on montre par ailleurs qu'elle ne peut pas converger vers un rationnel). En revanche $\mathbb{R}$ muni de sa distance usuelle est complet. De même, $\mathbb{R}^n$ muni de n'importe laquelle des distances d_1, d_2, d_∞ est complet. Nous verrons d'autres exemples aux chapitres suivants.*

Voici une première propriété des espaces complets:

Proposition 1.5 *Soit (E, d) un espace métrique complet et (F_n) une suite décroissante de fermés non vides dont le diamètre tend vers 0, alors l'intersection des F_n est non vide.*

Preuve:
Soit $d_n := \text{diam}(F_n)$ et soit pour tout $n \in \mathbb{N}$, $x_n \in F_n$. Pour tout couple d'entiers p et q avec $p, q \geq N$ on a: $d(x_p, x_q) \leq d_N$ et comme d_N tend vers 0 quand $N \to +\infty$, ceci montre que la suite (x_n) est de Cauchy : elle converge donc, appelons x sa limite. Comme x est la limite de la suite d'éléments de F_n, $(x_p)_{p \geq n}$, et comme F_n est fermé on a $x \in F_n$ ce qui achève la preuve.
□

Notons pour clore ce paragraphe qu'une suite de Cauchy est nécessairement bornée (s'en persuader) donc en particulier les suites convergentes sont bornées.

1.4 Compactness

Rappelons d'abord quelques définitions relatives aux suites extraites et valeur d'adhérence.

Définition 1.8 *Soit E un ensemble non vide et $(x_n)_n$ une suite d'éléments de E, on appelle sous-suite (ou suite extraite) de la suite $(x_n)_n$ toute suite de la forme $(x_{\varphi(n)})_n$ avec φ une application strictement croissante de $\mathbb{N}$ dans $\mathbb{N}$.*

Définition 1.9 *Soit (E, d) un espace métrique et (x_n) une suite d'éléments de E. On dit que x est <u>valeur d'adhérence</u> de (x_n) ssi l'une des assertions équivalentes suivantes est satisfaite:*

1. *(x_n) admet une sous-suite qui converge vers x,*
2. *$\forall \varepsilon > 0$, $\forall N \in \mathbb{N}$, $\exists n \geq N$ t.q. $d(x_n, x) \leq \varepsilon$,*
3. *$\forall \varepsilon > 0$ l'ensemble $\{n \in \mathbb{N} : d(x_n, x) \leq \varepsilon\}$ est infini.*

Exercice 1.5 *Prouver l'équivalence des trois assertions précédentes.*

Exercice 1.6 *Prouver que si φ est comme dans la définition 1.8 alors $\varphi(n) \geq n$ pour tout n.*

Exemple 1.6 *La suite $(-1)^n$ admet deux valeurs d'adhérence: 1 et -1.*

Définition 1.10 *On dit que l'espace métrique (E, d) est compact ssi toute suite d'éléments de E admet une sous-suite convergente. On dit qu'une partie A de l'espace métrique (E, d) est compacte ssi toute suite d'éléments de A admet une sous-suite convergente dans A.*

Proposition 1.6 *Soit (E, d) un espace métrique. Si A est une partie compacte de E alors A est fermée et borné.*

Preuve:
Soit $(x_n)_n \in A^{\mathbb{N}}$ une suite convergente, notons $x \in E$ sa limite. Comme A est compacte, $(x_n)_n$ admet une sous suite qui converge dans A, une telle sous-suite converge nécessairement vers x (s'en persuader...) d'où $x \in A$ ce qui montre que A est fermée.

Supposons que A ne soit pas bornée on a alors $\text{diam}(A) = +\infty$ et donc il existe deux suites $(x_n) \in A^{\mathbb{N}}$, et $(y_n) \in A^{\mathbb{N}}$ telles que

$$\lim_n d(x_n, y_n) = +\infty. \tag{1.1}$$

Comme A est compacte on peut trouver des sous suites $(x_{\varphi(n)})$ et $(y_{\varphi(n)})$ convergeant respectivement vers les éléments x et y de A, on a donc

$$d(x_{\varphi(n)}, y_{\varphi(n)}) \leq d(x_{\varphi(n)}, x) + d(x, y) + d(y, y_{\varphi(n)}) \to d(x, y)$$

ce qui contredit (1.1).
□

Attention: un fermé borné n'est pas nécessairement compact, nous aurons l'occasion de revenir sur ce point.

Les parties compactes d'un métrique compact sont faciles à caractériser puisque:

Proposition 1.7 *Soit (E, d) un espace métrique compact et A une partie de E alors A est une partie compacte de E ssi A est fermé dans E.*

Preuve:
Si A est compacte alors A est fermée dans E d'après la proposition précédente. Supposons A fermée et soit $(x_n) \in A^{\mathbb{N}}$, par compacité de E, (x_n) admet une sous-suite qui converge vers une limite $x \in E$, A étant fermé $x \in A$ et donc la sous suite converge aussi vers x dans A ce qui prouve que A est compacte.
□

Notons que la notion de compacité est plus forte que celle de complétude:

Proposition 1.8 *Tout espace métrique compact est complet.*

Preuve:
Soit (E,d) un espace métrique compact et $(x_n)_n$ une suite de Cauchy dans E. Comme E est compact, (x_n) admet une valeur d'adhérence $x \in E$. Montrons que (x_n) converge vers x: soit $\varepsilon > 0$, comme la suite est de Cauchy, il existe N_1 tq pour tous entiers $p, q \geq N_1$ on a: $d(x_p, x_q) \leq \varepsilon/2$. Comme x est valeur d'adhérence, il existe $N_2 \geq N_1$ tel que $d(x_{N_2}, x) \leq \varepsilon/2$. Ainsi pour tout $p \geq N_2$ on a: $d(x_p, x) \leq d(x_p, x_{N_2}) + d(x_{N_2}, x) \leq \varepsilon$. Ce qui montre que (x_n) converge vers x et donc que (E,d) est complet.

□

Bien noter que la réciproque est fausse: $\mathbb{R}$ est complet mais pas compact (car non borné!). Remarquons aussi au passage que dans la démonstration précédente nous avons établi le résultat:

Lemme 1.1 *Soit (E,d) un espace métrique et $(x_n)_n$ une suite d' éléments de E alors $(x_n)_n$ converge ssi $(x_n)_n$ est de Cauchy et admet une valeur d'adhérence.*

Théorème 1.1 *Dans $\mathbb{R}$ muni de sa distance usuelle, tout fermé borné est compact.*

Preuve:
Soit F un fermé borné de $\mathbb{R}$, puisque F est borné, F est inclus dans un segment $[a,b]$ de $\mathbb{R}$, sans perte de généralité, nous pouvons supposer que $F \subset [0,1]$. Soit $(x_n)_n \in F^{\mathbb{N}} \subset [0,1]^{\mathbb{N}}$, on va montrer que (x_n) admet une sous-suite qui est de Cauchy en procédant comme suit. Pour tout $p \in \mathbb{N}^*$, on décompose $[0,1]$ en 2^p segments de longueur 2^{-p}:

$$[0,1] = \bigcup_{k=0}^{2^p-1} I_k^p, \;\; I_k^p := [k2^{-p}, (k+1)2^{-p}].$$

Pour $p = 1$ l'un des deux intervalles I_1^1 et I_2^1 que l'on notera J_1 est tel que l'ensemble $\{n \in \mathbb{N} : \; x_n \in J_1\}$ est infini. On écrit ensuite

$$J_1 = \bigcup_{k \in \{0,..,4\} \; : \; I_k^2 \subset J_1} I_k^2$$

et comme précédemment $\exists k \in \{0,...,4\}$ tq l'un des intervalles $I_1^2, ..., I_4^2$ que l'on notera J_2 vérifie:

$$J_2 \subset J_1, \text{ et l'ensemble } \{n \in \mathbb{N} \text{ t.q. } x_n \in J_2\} \text{ est infini.}$$

On construit ainsi par récurrence une suite décroissantes d'intervalles fermés $J_1 \supset J_2 \supset .... \supset J_p$ tel que J_p est de longueur 2^{-p} et pour tout p, l'ensemble $\{n \in \mathbb{N} : \; x_n \in J_p\}$ est infini.

Soit n_1 le premier entier k tq $x_k \in J_1$, n_2 le premier entier $k \geq n_1 + 1$ tq $x_k \in J_2$, ..., n_p le premier entier $k \geq n_{p-1} + 1$ tq $x_k \in J_p$. La suite $(x_{n_p})_p$ est une sous-suite de $(x_n)_n$. Notons maintenant que par construction, on a pour tout $r, s \geq p$, $(x_{n_s}, x_{n_r}) \in J_p$ et comme J_p est de diamètre 2^{-p}, on a $|x_{n_s} - x_{n_r}| \leq 2^{-p}$ et donc $(x_{n_p})_p$ est de Cauchy. Comme $\mathbb{R}$ est complet, $(x_{n_p})_p$ converge et comme F est fermé sa limite est dans F. Ceci montre que (x_n) admet une sous-suite convergente dans F, F est donc compact.

□

Exercice 1.7 *Soit (E, d) un espace métrique compact et $(x_n)_n \in E^{\mathbb{N}}$ montrer que la suite $(x_n)_n$ converge ssi elle admet une unique valeur d'adhérence.*

The following important Theorem (Bolzano-Weierstrass) gives a characterization of compactness in metric spaces in terms of finite open coverings:

Theorem 1.1 *Let (E, d) be a metric space. Then (E, d) is a compact metric space iff for every family of open sets $(O_i)_{i \in I}$ such that $E = \cup_{i \in I} O_i$ there is a finite set $J \subset I$ such that $E = \cup_{i \in J} O_i$ (finite covering property).*

Proof:
First let us assume that (E, d) has the finite covering property and let us remark that it implies that if F_n is a sequence of nonempty closed subsets of E such that $F_{n+1} \subset F_n$ then $\cap_n F_n$ is nonempty (otherwise $O_n = E \setminus F_n$ would be an open covering of E and there would exist a finite covering which would imply that some F_n would be empty). Now let $(x_n)_n \in E^{\mathbb{N}}$, for every n, let us set:

$$F_n := \overline{\{x_k,\ k \geq n\}}$$

we then have $F := \cap_n F_n \neq \emptyset$ and it is easy to check that F is the set of cluster points of the sequence (x_n).

Conversely let us assume that (E, d) is compact an let us prove that it has the finite covering propoerty. Let $(O_i)_{i \in I}$ be a family of open sets such that $E = \cup_{i \in I} O_i$.

Claim 1: there exists $\varepsilon > 0$ such that for every $x \in E$, there is a $i \in I$ such that $B(x, \varepsilon) \subset O_i$.

If it was not the case (taking $\varepsilon = 1/n$), for every n there would exist some x_n such that $B(x_n, 1/n)$ is not included in any O_i. Since E is compact (taking a subsequence if necessary), we may assume that x_n converges to some $\overline{x}$ that belongs to the open set O_{i_0}, but for n large enough one should have $B(x_n, 1/n) \subset O_{i_0}$ which gives the desired contradiction.

Claim 2: For every $r > 0$, E can be covered by finitely many open balls of radius r.

Otherwise, there exists $r > 0$ such that E cannot be covered by finitely many open balls of radius r. We may then choose $x_1 \in E$, $x_2 \notin B(x_1, r)$, $x_3 \notin B(x_1, r) \cup B(x_2, r)$, ..., $x_{n+1} \notin \cup_{k=1}^{n} B(x_k, r)$. By construction $d(x_p, x_q) \geq r$ for every $p \neq q$ hence (x_n) does not have any Cauchy subsequence and therefore no cluster point, which yields the contradiction.

Let $\varepsilon > 0$ be as in Claim 1 and take $r = \varepsilon$ in Claim 2, there are points $x_1, ..., x_n$ such that $E = \cup_{j=1}^{n} B(x_j, \varepsilon)$ and $B(x_j, \varepsilon) \subset O_{i_j}$ for some $i_j \in E$. This proves that the finite family $(O_{i_j})_{j=1,...,n}$ is a covering of E.

□

Compact metric spaces are *separable* i.e. admit a countable dense subset:

Proposition 1.9 *Let (E, d) be a compact metric space, then it is separable in the sense that there is a countable set $(x_n)_n \in E^{\mathbb{N}}$ that is dense in E.*

Proof:
By the Bolzano-Weierstrass theorem for every $n \in \mathbb{N}^*$ there exists finitely many points $(x_i^n)_{i \in I_n}$ such that E is covered by the union of the open balls $B(x_i^n, 1/n)$. The set $\cup_n \{x_i^n,\ i \in I_n\}$ is countable and dense in E by construction.

□

In a compact metric space, we have a convenient criterion for a sequence to converge:

Proposition 1.10 *Let (E, d) be a compact metric space and let $(x_n)_n \in E^{\mathbb{N}}$ then the sequence (x_n) converges if and only if it possesses a unique cluster point.*

Proof:
Of course if (x_n) converges it has a unique cluster point. Now assume that some subsequence $(x_{\varphi(n)})$ converges to some $\overline{x}$ but $\overline{x}$ is not the limit of (x_n). This means that : $\exists \varepsilon_0 > 0$, such that for every N, there is some $n \geq N$ such that $d(\overline{x}, x_n) \geq \varepsilon_0$. This implies that (x_n) admits a subsequence $(x_{\psi(n)})$ such that $d(x_{\psi(n)}, \overline{x}) \geq \varepsilon_0$ for all n. By compactness of (E, d), $(x_{\psi(n)})$ possesses a cluster point y and $d(y, \overline{x}) \geq \varepsilon_0 > 0$, which implies that (x_n) has two distinct cluster points.

□

Finally, we leave as an exercise, the following criterion for compactness:

Definition 1.1 *A metric space (E, d) is precompact iff for every $\varepsilon > 0$, E can be covered by finitely many open balls of radius $\varepsilon > 0$*

Theorem 1.2 *Let (E, d) be a metric space then (E, d) is compact iff it is precompact and complete*

1.5 Continuity

Définition 1.11 *Soit (E_1, d_1) et (E_2, d_2) deux espaces métriques, f une application de E_1 dans E_2 et $x \in E_1$. On dit que f est continue en x ssi $\forall \varepsilon > 0$, $\exists \delta > 0$ t.q. $d_1(x, y) \leq \delta \Rightarrow d_2(f(x), f(y)) \leq \varepsilon$. On dit que f est continue sur E_1 ssi f est continue en chacun de ses points.*

Exemple 1.7 *Soit (E, d) un espace métrique, $x_0 \in E$ et définissons $\forall x \in E$, $f(x) := d(x, x_0)$. On a alors $|f(x) - f(y)| \leq d(x, y)$ et donc (en prenant simplement "$\delta = \varepsilon$" dans la définition précédente) f est continue sur E.*

Proposition 1.11 *Soit (E_1, d_1) et (E_2, d_2) deux espaces métriques, f une application de E_1 dans E_2. Alors les assertions suivantes sont équivalentes:*

1. *f est continue sur E_1,*
2. *pour tout ouvert O de E_2, $f^{-1}(O)$ est un ouvert de E_1,*
3. *pour tout fermé F de E_2, $f^{-1}(F)$ est un fermé de E_1,*
4. *pour toute suite (x_n) d'éléments de E_1 on a:*

$$\lim_n x_n = x \text{ dans } E_1 \Rightarrow \lim_n f(x_n) = f(x) \text{ dans } E_2.$$

Preuve:
On va montrer 1) $\Rightarrow$ 2) $\Rightarrow$ 3) $\Rightarrow$ 4) $\Rightarrow$ 1).

1) $\Rightarrow$ 2): soit O un ouvert de E_2, $x \in f^{-1}(O)$ et $y := f(x) \in O$, comme O est ouvert, $\exists \varepsilon > 0$ tq $B(y, \varepsilon) \subset O$. Par continuité de f en x, $\exists \delta > 0$ tq pour tout $x' \in E_1$, $x' \in B(x, \delta) \Rightarrow f(x') \in B(f(x), \varepsilon) = B(y, \varepsilon) \subset O$, ainsi $B(x, \delta) \subset f^{-1}(O)$ donc $f^{-1}(O)$ est un voisinage de x, comme x est un point arbitraire de $f^{-1}(O)$ on en déduit que $f^{-1}(O)$ est ouvert.

2) $\Rightarrow$ 3): (par passage au complémentaire), soit F fermé de E_2, et soit O l'ouvert $O := E_2 \backslash F$, d'après 2), $f^{-1}(O)$ est ouvert mais $f^{-1}(O) = E_1 \backslash f^{-1}(F)$ donc $f^{-1}(F) = E_1 \setminus f^{-1}(O)$, ainsi $f^{-1}(F)$ est fermé.

3) $\Rightarrow$ 4): soit $(x_n) \in E^{\mathbb{N}}$ une suite convergente de limite $x \in E_1$ et supposons par l'absurde que $f(x_n)$ ne converge pas vers $f(x)$. Alors $\exists \varepsilon > 0$ tq $\forall N \in \mathbb{N}$, $\exists n_N \geq N$ tq

$$d_2(f(x_{n_N}), f(x)) \geq \varepsilon. \tag{1.2}$$

Posons $F := E_2 \setminus B(f(x), \varepsilon)$, F est fermé (complémentaire d'une boule ouverte) et donc par 3), $f^{-1}(F)$ est fermé. Notons que (1.2) signifie que $x_{n_N} \in$

$f^{-1}(F)$ pour tout N. Comme (x_{n_N}) converge vers x quand $N \to +\infty$ et comme $f^{-1}(F)$ est fermé, on en déduit que $x \in f^{-1}(F)$ i.e. $d_2(f(x), f(x)) \geq \varepsilon$ ce qui est absurde.

4) $\Rightarrow$ 1): supposons que f ne soit pas continue en un point x de E_1, alors il existe $\varepsilon > 0$ tq pour tout $\delta > 0$, il existe $x_\delta \in E_1$ tel que $d_1(x_\delta, x) \leq \delta$ et $d_2(f(x_\delta), f(x)) > \varepsilon$. En prenant $\delta_n := 1/n$ et en notant $x_n := x_{\delta_n}$ on a alors $d_1(x_n, x) \leq 1/n$ et $d_2(f(x_n), x) > \varepsilon > 0$ ce qui contredit l'assertion 4).

□

Proposition 1.12 *Soit (E_1, d_1) et (E_2, d_2) deux espaces métriques, f une application continue de E_1 dans E_2. Si E_1 est compact alors $f(E_1)$ est une partie compacte de E_2.*

Preuve:
Soit $(z_n) := f(x_n)$ (avec $x_n \in E_1$) une suite de $f(E_1)$. Comme E_1 est compact, x_n admet une sous suite $(x_{\varphi(n)})$ convergente, f étant continue la sous suite $z_{\varphi(n)} = f(x_{\varphi(n)})$ est aussi convergente. Ceci montre donc que $f(E_1)$ est compact.

□

Corollaire 1.1 *Si (E, d) est compact et f est continue de E dans $\mathbb{R}$ (muni de sa distance usuelle) alors f atteint ses bornes sur E.*

Preuve:
$f(E)$ est un compact de $\mathbb{R}$ c'est donc un fermé borné en particulier ses bornes sont finies et appartiennent à $f(E)$. □

Le corollaire précédent peut être vu comme un résultat d'existence en optimisation. Il implique en effet que lorsque E est compact, les problèmes d'optimisation:

$$\sup\{f(x),\, x \in E\} \text{ et } \inf\{f(x),\, x \in E\}$$

admettent au moins une solution, autrement dit le sup. (resp. inf.) précédent est un max. (resp. min.).

Exercice 1.8 *Soit f une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$ telle que:*

$$\lim_{|x| \to +\infty} f(x) = +\infty$$

montrer que l'infimum de f sur $\mathbb{R}$ est atteint.

Exercice 1.9 *Soit A une partie compacte d'un espace métrique (E,d), on définit:*

$$d_A(x) := \inf\{d(x,a),\ a \in A\}.$$

Montrer que l'inf précédent est atteint. Montrer que $d_A(.)$ est continue sur E.

Définition 1.12 *Soit (E_1,d_1) et (E_2,d_2) deux espaces métriques, f une application de E_1 dans E_2. On dit que f est uniformément continue sur E_1 ssi $\forall \varepsilon > 0$, $\exists \delta > 0$ t.q. pour tout $(x,y) \in E_1^2$, $d_1(x,y) \leq \delta \Rightarrow d_2(f(x),f(y)) \leq \varepsilon$.*

Attention: il convient de bien distinguer la définition précédente de celle de continuité (dans la définition de la continuité en un point, "δ" dépend de ε et du point considéré, alors que dans la définition de l'uniforme continuité δ ne dépend que de ε, c'est précisément pour cela que l'on parle d'uniformité).

Exercice 1.10 *Trouvez une fonction de $\mathbb{R}$ dans lui même qui soit uniformément continue. Trouvez une fonction de $\mathbb{R}$ dans lui même qui soit continue et non uniformément continue.*

Rappelons la définition des applications Lipschitziennes:

Définition 1.13 *Soit (E_1,d_1) et (E_2,d_2) deux espaces métriques, f une application de E_1 dans E_2 et $k \in \mathbb{R}_+$. On dit que f est k-Lipschitzienne (ou Lipschitzienne de rapport k) ssi pour tout $(x,y) \in E_1 \times E_1$ on a $d_2(f(x),f(y)) \leq kd_1(x,y)$. On dit enfin que f est Lipschitzienne ssi $\exists k \geq 0$ tel que f soit k-Lipschitzienne.*

Exercice 1.11 *Soit (E,d) un espace métrique. Montrer que:*

1. *Pour tout $x_0 \in E$, l'application $x \mapsto d(x,x_0)$ est 1-Lipschitzienne.*
2. *Pour toute partie non vide A de E l'application*

$$x \mapsto d_A(x) := \inf\{d(x,a),\ a \in A\}$$

est 1-Lipschitzienne.

Exercice 1.12 *Montrer que les applications lipschitziennes de (E_1,d_1) dans (E_2,d_2) sont uniformément continues. Trouver une application uniformément continue de $\mathbb{R}$ dans $\mathbb{R}$ qui n'est pas Lipschitzienne.*

Exercice 1.13 *Montrer qu'une fonction continue et périodique de $\mathbb{R}$ dans $\mathbb{R}$ est uniformément continue.*

Exercice 1.14 *Soit $f : \mathbb{R} \to \mathbb{R}$ uniformément continue. Montrer qu'il existe deux constantes a et b telles que $|f(x)| \leq a|x| + b$, $\forall x \in \mathbb{R}$.*

Le résultat suivant (Théorème de Heine) énonce que si l'espace de départ est compact alors les notions de continuité et de continuité uniforme coincident:

Théorème 1.2 *Soit (E_1, d_1) et (E_2, d_2) deux espaces métriques, f une application continue de E_1 dans E_2. Si (E_1, d_1) est compact alors f est uniformément continue sur E_1.*

Preuve:
Supposons, par l'absurde que f ne soit pas uniformément continue alors il existe $\varepsilon > 0$, il existe deux suites d'éléments de E_1, (x_n) et (y_n) telles que $d_1(x_n, y_n)$ tende vers 0 et $d_2(f(x_n), f(y_n)) \geq \varepsilon$ pour tout n. E_1 étant compact on peut extraire des sous-suites convergentes de (x_n) et (y_n) de limites respectives x et y. En passant à la limite on obtient $x = y$ et $d_2(f(x), f(y)) \geq \varepsilon > 0$ ce qui est absurde.
□

1.6 Banach fixed-point Theorem

Théorème 1.3 *Soit (E, d) un espace métrique complet et f une contraction de E, c'est à dire une application de E dans E telle qu'il existe $k \in]0, 1[$ tel que:*

$$d(f(x), f(y)) \leq kd(x, y),\ \forall (x, y) \in E \times E.$$

Alors f admet un unique point fixe: il existe un unique $x \in E$ tel que $f(x) = x$. De plus, pour tout $x_0 \in E$, si on définit par récurrence la suite x_n par $x_{n+1} = f(x_n)$, pour $n \geq 0$, la suite x_n converge vers x quand $n \to +\infty$.

Preuve:
On commence d'abord par montrer l'unicité, supposons que f admette deux points fixes x_1 et x_2. Comme $f(x_1) = x_1$ et $f(x_2) = x_2$, on a alors $d(x_1, x_2) = d(f(x_1), f(x_2) \leq kd(x_1, x_2)$ et comme $k < 1$, il vient $d(x_1, x_2) = 0$ donc $x_1 = x_2$ d'où l'unicité.

Montrons maintenant l'existence. Soit $x_0 \in E$, définissons la suite x_n comme dans l'énoncé et montrons que celle-ci est de Cauchy. On commence par remarquer que pour tout $n \in \mathbb{N}^*$ on a $d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq kd(x_n, x_{n-1})$ en itérant l'argument on a donc aussi:

$$d(x_{n+1}, x_n) \leq k^n d(x_1, x_0) \tag{1.3}$$

Pour $q \geq p \geq N$ on a donc:

$$d(x_p, x_q) \leq d(x_p, x_{p+1}) + ... + d(x_{q-1}, x_q) \leq d(x_1, x_0)(k^p + ... + k^{q-1})$$
$$\leq d(x_1, x_0)\frac{k^N}{1-k}$$

comme $k \in]0,1[$, k^N tend vers 0 quand $N \to +\infty$, l'inégalité précédente implique donc que (x_n) est de Cauchy et donc admet une limite x dans E puisque (E,d) est complet. On vérifie aisément que f est continue donc $x_{n+1} = f(x_n)$ converge vers $f(x)$ on a donc $x = f(x)$. □

Il faut bien retenir que le théorème précédent indique très simplement comment trouver le point fixe d'une contraction f: on part de x_0 ARBITRAIRE (c'est assez remarquable) et on calcule les itérées $x_1 = f(x_0)$, $x_2 = f(x_1)$... cette suite converge vers le point fixe de f (noter aussi que la vitesse de convergence est géométrique: $d(x, x_n) \leq k^n d(x, x_0)$).

Noter que dans le théorème précédent l'hypothèse de contraction ($k < 1$) est fondamentale. Pour s'en convaincre considérer $f(x) = x + 1$ dans $\mathbb{R}$...

1.7 Baire's Theorem

Another important property of complete metric spaces is given by the next result, due to Baire:

Theorem 1.3 *Let (E,d) be a complete metric space and O_n be a sequence of open and dense subsets of E, then $\cap_n O_n$ is a dense subset of E.*

Proof:
Let U be some open set, we have to prove that $\cap_n O_n \cap U \neq \emptyset$. First let us fix $x_0 \in E$ and $r_0 > 0$ such that $\overline{B}(x_0, r_0) \subset U$. Since $B(x_0, r_0)$ is open and O_1 is dense there exists x_1 and $0 < r_1 \leq r_0/2$ such that $\overline{B}(x_1, r_1) \subset B(x_0, r_0) \cap O_1$. Inductively, we construct a sequence x_n in E and $r_n > 0$ such that $\overline{B}(x_{n+1}, r_{n+1}) \subset B(x_n, r_n) \cap O_{n+1}$ and $r_{n+1} \leq r_n/2$. Since x_n is a Cauchy sequence, it converges to some $\overline{x}$. By construction $\overline{x} \in \overline{B}(x_0, r_0) \subset U$ and $\overline{x} \in \overline{B}(x_n, r_n)$ for all n, thus $\overline{x} \in \cap_n O_n$, which completes the proof.
□

Taking complements, we get the equivalent formulation

Corollary 1.1 *Let (E,d) be a complete metric space and F_n be a sequence of closed subsets of E, if $\text{int}(F_n) = \emptyset$ for all n then $\text{int}(\cup_n F_n) = \emptyset$.*

In particular, we have the following, which is often useful

Corollary 1.2 *Let (E, d) be a complete metric space and F_n be a sequence of closed subsets of E, such that $\cup_n F_n = E$ then there is some n_0 such that F_{n_0} has nonempty interior.*

1.8 Set-valued maps

Définition 1.14 *Soit X et Y deux espaces métriques et soit F une correspondance à valeurs* ***compactes non vides*** *de X dans Y, et soit $x \in X$ on dit que:*

1. *F est hémi-continue supérieurement (h.c.s.) en x si pour toute suite x_n convergeant vers x dans X et pour toute suite $y_n \in F(x_n)$, la suite y_n admet une valeur d'adhérence dans $F(x)$.*

2. *F est hémi-continue inférieurement (h.c.i.) en x si pour tout $y \in F(x)$ et pour toute suite x_n convergeant vers x dans X, il existe $y_n \in F(x_n)$ telle que y_n converge vers y dans Y.*

3. *F est continue si F hémi-continue supérieurement et inférieurement en chaque point de X.*

Dans le cas où X et Y sont des métriques compacts, dire que F est h.c.s. revient simplement à dire que son graphe:

$$\text{graph}(F) := \{(x, y) : x \in X, y \in F(x) \}$$

est fermé. Noter que dans ce cas F est automatiquement à valeurs compactes.

Remarquons que dans le cas *univoque* i.e. $F(x) = \{f(x)\}$ on a équivalence entre "F est h.c.s.", "F est h.c.i" et "f est continue". Si $X = Y = \mathbb{R}$ et $F(x) = [f(x), g(x)]$ avec f et g deux fonctions continues telles que $f \leq g$ alors F est une correspondance continue. Pour fixer les idées, il est bon d'avoir en mémoire les exemples suivants:

La correspondance F de $\mathbb{R}$ dans $\mathbb{R}$ définie par:

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ [0, 1] & \text{si} \quad x = 0 \\ 1 & \text{si} \quad x > 0 \end{cases}$$

est h.c.s. mais pas h.c.i. en 0.

La correspondance G de $\mathbb{R}$ dans $\mathbb{R}$ définie par:

$$G(x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ [-1, 1] & \text{si} \quad x > 0 \end{cases}$$

est quant à elle h.c.i. mais pas h.c.s. en 0.

Chapter 2

Normed spaces

2.1 Basic definitions

Définition 2.1 *Soit E un $\mathbb{R}$-espace vectoriel, on appelle norme sur E toute application : $\|.\|$: $E \to \mathbb{R}_+$ vérifiant:*

1. $\|x\| = 0 \Leftrightarrow x = 0$,
2. $\|x + y\| \leq \|x\| + \|y\|$, $\forall (x, y) \in E^2$,
3. $\|\lambda x\| = |\lambda| \|x\|$, $\forall (\lambda, x) \in \mathbb{R} \times E$.

On appelle espace vectoriel normé (evn) la donnée d'un couple $(E, \|.\|)$ avec E un espace vectoriel réel et $\|.\|$ une norme sur E.

Une norme définit une distance sur E (et donc une topologie, des fermés, des compacts...) donnée par:

$$d(x, y) := \|x - y\|, \forall (x, y) \in E^2.$$

Bien noter ici que les evn ne sont qu'un cas particulier des espaces métriques étudiés au chapitre précédent. En particulier une norme définit une distance mais une distance n'est pas nécessairement associée à une norme (prendre l'exemple de la distance grossière).

Notons aussi que $\|x\| = \| - x\|$ et $|\|x\| - \|y\|| \leq \|x - y\|$.

Exemple 2.1 *Pour $E = \mathbb{R}^n$, nous avons déja rencontré les normes:*

$$\|x\|_\infty := \max(|x_1|, .., |x_n|), \|x\|_1 := \sum_{i=1}^{n} |x_i|, \|x\|_2 := (\sum_{i=1}^{n} |x_i|^2)^{1/2}$$

Nous verrons par la suite, que pour tout $p \geq 1$:

$$\|x\|_p := (\sum_{i=1}^{n} |x_i\|^p)^{1/p} \tag{2.1}$$

définit une norme sur $\mathbb{R}^n$. On peut construire de nombreux autres exemples (par exemple en notant que la somme ou le max d'un nombre fini de normes est encore une norme).

Exemple 2.2 *$E = C^0([a,b], \mathbb{R})$ munie de la norme*

$$\|f\|_\infty := \max\{|f(t)|,\ t \in [a,b]\}.$$

Sur E on peut aussi considérer les normes:

$$\|f\|_1 := \int_a^b |f|, \|f\|_2 := (\int_a^b f^2)^{1/2}$$

ou plus généralement pour tout $p \geq 1$:

$$\|f\|_p := (\int_a^b |f|^p)^{1/p}.$$

Exemple 2.3 *$E = l^\infty := \{(x_n)_n \in \mathbb{R}^{\mathbb{N}} : (x_n)_n$ bornée$\}$ munie de la norme*

$$\|x\|_\infty := \sup |x_n|,\ n \in \mathbb{N}.$$

Définition 2.2 *Soit E un $\mathbb{R}$-ev, $\|.\|_1$ et $\|.\|_2$ deux normes sur E on dit que ces deux normes sont équivalentes ssi il existe deux constantes strictement positives a et b telles que pour tout $x \in E$ on ait:*

$$a\|x\|_1 \leq \|x\|_2 \leq b\|x\|_1.$$

La notion de normes équivalentes est importante car deux normes équivalentes ont les mêmes ouverts, les mêmes fermés, les mêmes bornés, les mêmes compacts, les mêmes suites convergentes etc... autrement dit elles définissent la même topologie (.... et le même calcul différentiel) .

Exemple 2.4 *Sur $\mathbb{R}^n$, les normes $\|.\|_1$ et $\|.\|_\infty$ sont équivalentes: en effet, on a clairement pour tout $x \in \mathbb{R}^n$, $\|x\|_1 \geq \|x\|_\infty$ et $\|x\|_1 \leq n\|x\|_\infty$. Nous verrons au paragraphe 2.2 que sur $\mathbb{R}^n$, en fait, TOUTES les normes sont équivalentes.*

Exemple 2.5 *Considérons sur $E := C^0([0,1], \mathbb{R})$ les normes $\|.\|_1$ et $\|.\|_\infty$ comme dans l'exemple 2.2. Il est facile de voir que pour tout $f \in E$ on a: $\|f\|_1 \leq \|f\|_\infty$ mais ces 2 normes ne sont pas équivalentes pour autant. En effet, considérons la suite de fonctions $f_n(t) := \max(0, n(1-nt))$, on a $\|f_n\|_\infty = n$ et $\|f_n\|_1 = 1/2$, il ne peut donc exister de constante positive a telle que $\|f_n\|_\infty \leq a\|f_n\|_1$ pour tout $n \in \mathbb{N}^*$.*

2.2 Finite dimensional spaces

En dimension finie, nous allons voir que toute les normes sont équivalentes, ce qui signifie en pratique que l'on peut utiliser sur $\mathbb{R}^n$ n'importe quelle norme sans changer de topologie, on parle alors simplement de la topologie de $\mathbb{R}^n$ sans préciser la norme.

Théorème 2.1 *Les parties compactes de* $(\mathbb{R}^k, \|.\|_\infty)$ *sont ses parties fermées bornées. En particulier toute suite bornée de* $(\mathbb{R}^k, \|.\|_\infty)$ *admet une sous-suite convergente.*

Preuve:
Soit F une partie fermée bornée de $\mathbb{R}^k$ pour la norme $\|.\|_\infty$. Il existe alors $M > 0$ telle que $F \subset \overline{B}(0, M) = [-M, M]^k$. Soit $(x_n)_n \in F^{\mathbb{N}} \subset ([-M, M]^k)^{\mathbb{N}}$, en vertu du théorème 1.1, $[-M, M]$ est un compact de $\mathbb{R}$, on peut donc extraire une sous-suite[1] $(x_{\varphi(n)})_n$ telle que pour $i = 1, ..., k$ la suite des i-èmes composantes $(x_{i,\varphi(n)})_n$ converge vers une limite $x_i \in \mathbb{R}$. Notons $x = (x_1, ..., x_k)$, pour tout i on a $|x_{i,\varphi(n)} - x_i| \to 0$ quand $n \to +\infty$ et donc

$$\lim_n \|x_{\varphi(n)} - x\|_\infty = 0.$$

Ainsi $(x_{\varphi n})$ converge vers x, enfin $x \in F$ car F est fermé ce qui achève la preuve.
□

Théorème 2.2 *Si* E *est un espace vectoriel réel* <u>*de dimension finie*</u> *alors toutes les normes sur* E *sont équivalentes.*

Preuve:
Sans perte de généralité supposons $E = \mathbb{R}^n$. Soit N une norme sur $\mathbb{R}^n$, nous allons montrer que N est équivalente à la norme $\|.\|_\infty$ de $\mathbb{R}^n$ (si toutes les normes sont équivalentes à une norme donnée alors par "transitivité" elles sont toutes équivalentes entre elles). Soit $(e_1, ..., e_n)$ une base de $\mathbb{R}^n$ et soit $x \in \mathbb{R}^n$ que l'on écrit dans cette base $x = \sum_{i=1}^n x_i e_i$, on a:

$$N(x) = N(\sum_{i=1}^n x_i e_i) \leq \sum_{i=1}^n |x_i| N(e_i) \leq (\sum_{i=1}^n N(e_i)) \|x\|_\infty$$

donc $N(x) \leq C\|x\|_\infty$ pour tout $x \in \mathbb{R}^n$ $(C = \sum N(e_i))$. On a donc pour tout x, y:

$$|N(x) - N(y)| \leq |N(x - y)| \leq C\|x - y\|_\infty \tag{2.2}$$

[1] En réalité, il faut effectuer plusieurs extractions successives, les détails sont laissés au lecteur...

ce qui montre en particulier que N est continue de $(\mathbb{R}^n, \|.\|_\infty)$ dans $\mathbb{R}$.

Soit $S := \{x \in \mathbb{R}^n : \|x\|_\infty = 1\}$, S est un fermé borné de $(\mathbb{R}^n, \|.\|_\infty)$ et donc un compact en vertu du théorème 2.1. D'après (2.2), N atteint donc son infimum sur S soit donc $x_0 \in S$ tq $N(x_0) = \min_S N$ comme $x_0 \in S$ on a $x_0 \neq 0$ et donc $N(x_0) > 0$ posons $\alpha = N(x_0)$. Pour $x \neq 0$, $x/\|x\|_\infty \in S$ et donc:

$$N\left(\frac{x}{\|x\|_\infty}\right) \geq \alpha \Rightarrow \|x\|_\infty \leq \frac{N(x)}{\alpha}.$$

La dernière inégalité étant aussi satisfaite pour $x = 0$, ceci achève de montrer que N et $\|.\|_\infty$ sont équivalentes. □

En combinant les théorèmes 2.1 et 2.2, on obtient le résultat suivant dont la preuve est laissée au lecteur:

Théorème 2.3 *Soit $(E, \|.\|)$ un evn réel de dimension finie, les parties compactes de $(E, \|.\|)$ sont ses parties fermées et bornées. En particulier, toute suite bornée de $(E, \|.\|)$ admet une sous-suite convergente.*

2.3 Banach Spaces

2.3.1 Definitions and properties

On a vu au chapitre précédent l'importance de la notion de complétude dans le cadre général des espaces métriques, ainsi les evn complets appelés espaces de Banach jouent un rôle très important en analyse:

Définition 2.3 *On appelle espace de Banach tout evn qui muni de la distance associée à sa norme est complet.*

Soit $(E, \|.\|)$ un evn et $(x_n)_n \in E^{\mathbb{N}}$, on rappelle que la série de terme général x_n (notation: $(\sum_n x_n)$, vocabulaire: série à valeurs dans E) est la suite formées par ses sommes partielles: $S_n := \sum_{k \leq n} x_k$.

Définition 2.4 *Soit $(E, \|.\|)$ un evn et $(\sum_n x_n)_n$ une série à valeurs dans E. On dit que $(\sum_n x_n)_n$ est convergente ssi la suite de ses sommes partielles converge dans $(E, \|.\|)$, on appelle somme de la série la limite des sommes partielles qu'on note simplement $\sum_{n=0}^{+\infty} x_n$. On dit que $(\sum_n x_n)_n$ est normalement convergente ssi la série $(\sum_n \|x_n\|)_n$ est convergente dans $\mathbb{R}$.*

On rappelle que la série (à termes positifs) $(\sum_n \|x_n\|)_n$ converge ssi la suite de ses sommes partielles est de Cauchy:

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ tq } \forall p \geq q \geq N, \sum_{k=q+1}^{p} \|x_k\| \leq \varepsilon \tag{2.3}$$

dans ce cas la suite des restes $\sum_{k=n}^{+\infty} \|x_k\|$ tend vers 0 quand $n \to +\infty$.

Proposition 2.1 *Soit $(E, \|.\|)$ un espace de Banach et $(\sum_n x_n)$ une série à valeurs dans E, si $(\sum_n x_n)$ est normalement convergente alors $(\sum_n x_n)$ converge dans E.*

Preuve:
Il suffit de montrer que la suite des sommes partielles $S_n := \sum_{k \leq n} x_k$ est de Cauchy, or on a pour $p \geq q$:

$$\|S_p - S_q\| \leq \sum_{k=q+1}^{p} \|x_k\| \leq \sum_{k=q+1}^{+\infty} \|x_k\| \tag{2.4}$$

comme la série est normalement convergente, le membre de droite de (2.4) tend vers 0 quand $q \to +\infty$, $(S_n)_n$ est donc de Cauchy et la série converge. □

Attention: la convergence normale est suffisante pour la convergence mais pas nécessaire (dans $\mathbb{R}$ considérer la série alternée de terme général $(-1)^n/n$ qui est convergente mais non absolument convergente).

Exercice 2.1 *Soit $(E, \|.\|)$ un evn, montrer que $(E, \|.\|)$ est un espace de Banach ssi toute série à valeurs dans E normalement convergente est absolument convergente.*

Exercice 2.2 *Soit (f_n) une suite bornée de $(C^0([0,1], \mathbb{R}), \|.\|_\infty)$ et $(\sum_n \alpha_n)$ une série à valeurs dans $\mathbb{R}$ convergente, montrer que la série $(\sum_n \alpha_n f_n)$ converge dans $(C^0([0,1], \mathbb{R}), \|.\|_\infty)$ (on pourra utiliser le théorème 2.5).*

2.3.2 Examples of Banach Spaces

Evidemment tout $\mathbb{R}$-ev de dimension finie muni d'une norme quelconque est un espace de Banach:

Proposition 2.2 *$\mathbb{R}^N$ muni de n'importe quelle norme est un espace de Banach.*

Preuve:
Soit $(x_n)_n \in (\mathbb{R}^N)^{\mathbb{N}}$ une suite de Cauchy pour la norme $\|.\|_\infty$ (le choix d'une norme n'importe pas ici puisqu'en dimension finie toutes les normes sont équivalentes). Il est clair que chaque suite formée par les composantes: $(x_n^1)_n$, ..., $(x_n^N)_n$ est de Cauchy dans $\mathbb{R}$ et comme $\mathbb{R}$ est complet, ces suites convergent respectivement vers des limites x^1, ..., x^N. Il est alors clair que $(x_n)_n$ converge dans $\mathbb{R}^N$ vers $x = (x^1, ..., x^N)$. □

Exercice 2.3 *Prouver que $\mathbb{R}^N$ est complet comme suit. Soit $(x_n)_n$ une suite de Cauchy dans $\mathbb{R}^k$, montrer que*

1. *$(x_n)_n$ est bornée et admet une sous-suite convergente,*
2. *en déduire que $(x_n)_n$ converge et conclure.*

Passons maintenant à quelques exemples d'espaces de Banach de dimension infinie. Soit X un ensemble, $(E, \|.\|)$ un espace de Banach et $B(X, E)$ l'ensemble des applications bornées de X dans E:

$$B(X, E) := \{f : X \to E \text{ tq } \sup_{x \in X} \|f(x)\| < +\infty\} \tag{2.5}$$

on vérifie trivialement que $B(X, E)$ est un ev et que sur E:

$$\|f\|_\infty := \sup_{x \in X} \|f(x)\| \tag{2.6}$$

est une norme appelée norme de la convergence uniforme (ou simplement norme uniforme).

Remarque.

Il faut bien distinguer la convergence uniforme et la convergence simple ((f_n) converge uniformément vers f lorsque $\|f_n - f\|_\infty$ tend vers 0 alors que (f_n) converge simplement vers f lorsque $(f_n(x))$ converge vers $f(x)$ dans $(E, \|.\|)$ pour tout $x \in X$). Evidemment si (f_n) converge uniformément vers f alors (f_n) converge simplement vers f mais la réciproque est fausse (trouvez des contre-exemples).

Théorème 2.4 *Soit X un ensemble, et $(E, \|.\|)$ un espace de Banach alors $(B(X, E), \|.\|_\infty)$ est un espace de Banach.*

Preuve:

Soit $(f_n)_n$ une suite de Cauchy de $(B(X, E), \|.\|_\infty)$, ce qui signifie:

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} \text{ t.q. } \forall p, q \geq N,\ \forall x \in X,\ \|f_p(x) - f_q(x)\| \leq \varepsilon. \tag{2.7}$$

Nous allons prouver que $(f_n)_n$ converge dans $(B(X, E), \|.\|_\infty)$ en passant par trois étapes.

Etape 1: identification d'une limite ponctuelle

Soit $x \in X$ (fixé), (2.7) implique en particulier que la suite $(f_n(x))_n \in E^{\mathbb{N}}$ est de Cauchy et comme $(E, \|.\|)$ est un Banach, elle converge: soit $f(x)$ sa limite.

Etape 2: $f \in B(X,E)$

D'après (2.7), il existe N tel que pour tout $p,q \geq N$, et pour tout $x \in X$ on a:

$$\|f_p(x) - f_q(x)\| \leq 1. \tag{2.8}$$

Pour $x \in X$ fixé, prenons $p = N$, faisons tendre q vers $+\infty$ dans (2.8), comme $f_q(x)$ converge vers $f(x)$ on obtient $\|f_N(x) - f(x)\| \leq 1$ mais comme $x \in X$ est arbitraire dans l'inégalité précédente, nous obtenons:

$$\sup_{x\in X} \|f(x) - f_N(x)\| \leq 1 \Rightarrow (f - f_N) \in B(X,E)$$

et comme $f_N \in B(X,E)$ on en déduit que $f \in B(X,E)$.

Etape 3: $(f_n)_n$ **converge vers** f **dans** $(B(X,E), \|.\|_\infty)$

Soit $\varepsilon > 0$, d' après (2.7), il existe N tq pour tout $p,q \geq N$ et tout $x \in X$ on a $\|f_p(x) - f_q(x)\| \leq \varepsilon$. Comme précédemment, fixons $x \in X$, prenons $p \geq N$ et faisons tendre q vers $+\infty$, on obtient alors:

$$\|f_p(x) - f(x)\| \leq \varepsilon. \tag{2.9}$$

Mais comme (2.9) a lieu pour tout $p \geq N$ et tout $x \in X$ on a:

$$\forall p \geq N,\ \|f_p - f\|_\infty \leq \varepsilon.$$

ce qui achève la preuve. □

Notons que dans ce qui précède, l'ensemble de départ X est totalement arbitraire. Un cas particulier intéressant est celui où $X = \mathbb{N}$, en effet dans ce cas $B(\mathbb{N},E) = l^\infty(E)$ est l'espace des suites bornées d'éléments de E. En munissant $l^\infty(E)$ de la norme uniforme et en appliquant le théorème précédent on a ainsi:

Corollaire 2.1 *Soit $(E, \|.\|)$ un espace de Banach, alors $l^\infty(E)$ muni de la norme uniforme est un espace de Banach.*

Un autre cas intéréssant est celui où X est muni d'une distance d, dans ce cas on peut s'intéresser à l'espace vectoriel $C_b^0(X,E)$ des applications continues et bornées[2] de X dans E. En munissant $C_b^0(X,E)$ de la norme uniforme définie par (2.6), on a:

Théorème 2.5 *Soit (X,d) un espace métrique et $(E, \|.\|)$ un espace de Banach, alors $(C_b^0(X,E), \|.\|_\infty)$ est un espace de Banach.*

[2]Notons au passage que si (X,d) est compact alors toute application f continue de (X,d) dans $(E, \|.\|)$ est bornée puisque $f(X)$ est compact donc borné.

Preuve:
Soit $(f_n)_n$ une suite de Cauchy de $(C_b^0(X,E), \|.\|_\infty)$, d'après le théorème 2.4, nous savons que $(f_n)_n$ converge vers une limite $f \in B(X,E)$ dans $(B(X,E), \|.\|_\infty)$, il nous suffit donc de montrer que f est continue pour pouvoir conclure.

Soit $\varepsilon > 0$ et soit N tq pour tout $n \geq N$ on ait:

$$\|f_n - f\|_\infty \leq \frac{\varepsilon}{3} \tag{2.10}$$

Soit $x_0 \in X$, comme f_N est continue en x_0, il existe $\delta > 0$ tel que:

$$\forall x \in B(x_0, \delta),\ \|f_N(x_0) - f_N(x)\| \leq \frac{\varepsilon}{3} \tag{2.11}$$

Pour $x \in B(x_0, \delta)$ on a:

$$\begin{aligned}\|f(x) - f(x_0)\| \leq & \|f(x) - f_N(x)\| + \|f_N(x) - f_N(x_0)\| + \|f_N(x_0) - f(x_0)\| \\ \leq & \|f - f_N\|_\infty + \varepsilon/3 + \|f - f_N\|_\infty \\ \leq & \varepsilon \text{ (avec (2.10))}\end{aligned}$$

ce qui montre que f est continue en x_0.
□

Exercice 2.4 *Soit (E_1, N_1),...(E_K, N_K) des espaces de Banach et soit $E := E_1 \times ... \times E_K$ montrer que sur E:*

$$N(x_1, .., x_K) := \sum_{i=1}^{K} N_i(x_i),\ M(x_1, .., x_k) := \max_{i=1,..,K} N_i(x_i).$$

sont des normes équivalentes et que E muni d'une de ces normes est un espace de Banach.

2.4 Hilbert Spaces

Définition 2.5 *Soit E un $\mathbb{R}$-ev, on appelle produit scalaire (ps) sur E toute application $\langle ., . \rangle : E \times E \to \mathbb{R}$ qui est:*

1. *bilinéaire: pour tout $x \in E$ (fixé) $y \mapsto \langle x, y \rangle$ est linéaire, et pour tout $y \in E$ (fixé) $x \mapsto \langle x, y \rangle$ est linéaire,*
2. *symétrique: $\langle x, y \rangle = \langle y, x \rangle$, $\forall (x, y) \in E \times E$,*

3. *définie positive:* $\langle x, x\rangle \geq 0$, $\forall x \in E$ *et* $\langle x, x\rangle = 0 \Leftrightarrow x = 0$.

Notons les identités faciles à établir en utilisant la bilinéarité et la symétrie:

$$\begin{aligned} \langle x+y, x+y\rangle &= \langle x, x\rangle + \langle y, y\rangle + 2\langle x, y\rangle, \\ \langle x-y, x-y\rangle &= \langle x, x\rangle + \langle y, y\rangle - 2\langle x, y\rangle, \ \forall (x, y) \in E \times E. \end{aligned} \tag{2.12}$$

Définition 2.6 *On appelle espace préhilbertien la donnée d'un couple* $(E, \langle ., .\rangle)$ *avec* E *un* $\mathbb{R}$*-ev et* $\langle ., .\rangle$ *un produit scalaire sur* E.

La donnée d'un produit scalaire permet de définir une norme sur E, et ce grâce à l'inégalité de Cauchy-Schwarz:

Proposition 2.3 *Soit* $(E, \langle ., .\rangle)$ *un espace préhilbertien.*

1. *pour tout* $(x, y) \in E \times E$ *on a l'inégalité de Cauchy-Schwarz:*

$$|\langle x, y\rangle| \leq (\langle x, x\rangle)^{1/2} (\langle y, y\rangle)^{1/2}. \tag{2.13}$$

De plus il y a égalité dans (2.13) ssi x *et* y *sont liés.*

l'application $x \mapsto (\langle x, x\rangle)^{1/2}$ *est une norme sur* E *appelée norme associée au ps* $\langle ., .\rangle$.

Preuve:

1): Définissons pour tout $t \in \mathbb{R}$,

$$g(t) := \langle x + ty, x + ty\rangle = t^2 \langle y, y\rangle + 2t \langle x, y\rangle + \langle x, x\rangle$$

g est un trinôme en t (non dégénéré si $y \neq 0$ mais si $y = 0$ les 2 membres de (2.13) valent 0) et $g(t) \geq 0$ $\forall t$ par positivité du ps. Le discriminant de g est donc négatif soit:

$$(\langle x, y\rangle)^2 \leq \langle x, x\rangle \langle y, y\rangle,$$

on obtient (2.13) en prenant la racine carrée de l'inégalité précédente.

Il est clair que si x et y sont liés, il y a égalité dans (2.13). Réciproquement si il y a égalité dans (2.13) alors le discriminant de g est nul et donc g admet une racine double $t_0 \in \mathbb{R}$, mais $g(t_0) = 0$ ssi $x + t_0 y = 0$ et donc x et y sont liés.

2): Notons $\|x\| := (\langle x, x\rangle)^{1/2}$, comme $\langle ., .\rangle$ est un ps, on a $\|x\| = 0$ ssi $x = 0$. La bilinéarité implique clairement $\|\lambda x\| = |\lambda|\|x\|$. Reste à montrer l'inégalité triangulaire: soit $(x, y) \in E \times E$ on a

$$\begin{aligned}\|x + y\|^2 =& \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle \\ \leq& \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \text{ (d'après Cauchy-Schwarz)} \\ =& (\|x\| + \|y\|)^2\end{aligned}$$

ce qui achève de montrer que $\|.\|$ est une norme sur E. □

Notons $\|.\|$ la norme associée au ps $\langle ., .\rangle$ remarquons que la connaissance de cette norme permet de "retrouver" le produit scalaire par l'identité suivante (identité de polarisation):

$$\langle x, y\rangle = \frac{1}{2}(\|x + y\|^2 - \|x\|^2 - \|y\|^2). \tag{2.14}$$

Mentionnons aussi l'identité du parallélogramme:

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2). \tag{2.15}$$

Remarque. Il découle de l'inégalité de Cauchy-Schwarz que pour tout $x \in H$, la forme linéaire $y \mapsto \langle x, y\rangle$ est continue sur H.

Définition 2.7 *Soit $(E, \langle ., .\rangle)$ un espace préhilbertien. On dit que $(E, \langle ., .\rangle)$ est un espace de Hilbert ssi E muni de la norme associée à $\langle ., .\rangle$ est complet.*

Exemple 2.6 *$E = \mathbb{R}^n$ muni du produit scalaire usuel: $\langle x, y\rangle := \sum_{i=1}^N x_i y_i$. Plus généralement, toute matrice carrée de taille n symétrique et définie positive A définit un ps sur $\mathbb{R}^n$ via:*

$$\langle x, y\rangle := x' A y.$$

Evidemment, le cas du ps usuel correspond à $A = I_n$.

Exemple 2.7 *l^2 l'espace des suites réelles (x_n) tq $\sum |x_n|^2 < +\infty$ muni du produit scalaire:*

$$\langle x, y\rangle := \sum_{n\geq 0} x_n y_n.$$

Exemple 2.8 *L'espace de Lebesgue $L^2([0, 1], \mathbb{R})$ muni de:*

$$(f, g) \mapsto \int_0^1 f(t)g(t)dt$$

est un Hilbert mais $C^0([0, 1], \mathbb{R})$ muni de la même structure est seulement préhilbertien.

Etant donné un espace préhilbertien $(H, \langle ., \rangle)$, on dit que deux vecteurs u et v sont orthogonaux ssi $\langle u, v\rangle = 0$. Pour $A \subset H$ on appelle orthogonal de A l'ensemble:

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

On vérifie sans peine que $A^{\perp}$ est un sev fermé de H car intersection de sev fermés.

2.5 Continuous linear and bilinear maps

Dans ce qui suit étant donnés $(E, \|.\|_E)$ et $(F, \|.\|_F)$ deux e.v.n, on notera $L(E, F)$ (resp. $L_c(E, F)$) l'espace vectoriel des applications linéaires (resp. linéaires continues) de E dans F. Pour $E = F$, on notera simplement $L(E)$ (resp. $L_c(E, F)$) l'espace vectoriel des endomorphismes (resp. endomorphismes continus) de E.

2.6 Characterization

Théorème 2.6 *Soit $(E, \|.\|_E)$ et $(F, \|.\|_F)$ deux e.v.n, et $f \in L(E, F)$, les assertions suivantes sont équivalentes:*

1. *$f \in L_c(E, F)$,*
2. *f est continue en un point,*
3. *f est bornée sur la boule unité fermée de E, $\overline{B}_E(0, 1)$,*
4. *il existe une constante $M \geq 0$ tq $\|f(x)\|_F \leq M\|x\|_E \ \forall x \in E$,*
5. *f est Lipschitzienne sur E.*

Preuve:

1) $\Rightarrow$ 2) est évident.

2) $\Rightarrow$ 3): supposons f continue en $x_0 \in E$, alors $\exists r > 0$ telle que pour tout $x \in \overline{B}_E(x_0, r)$ on ait:

$$\|f(x) - f(x_0)\|_F \leq 1. \tag{2.16}$$

Soit $u \in \overline{B}_E(0, 1)$ on a $x_0 + ru \in \overline{B}_E(x_0, r)$ et donc avec (2.16), il vient:

$$\|f(x_0 + ru) - f(x_0)\|_F = \|f(ru)\|_F = r\|f(u)\|_F \leq 1 \tag{2.17}$$

on en déduit donc que $\forall u \in \overline{B}_E(0,1)$, on a $\|f(u)\|_F \leq 1/r$.

3) $\Rightarrow$ 4): Supposons donc qu'il existe $M > 0$ telle que:

$$\|f(u)\|_F \leq M, \, \forall u \in \overline{B}_E(0,1). \tag{2.18}$$

Soit $x \in E$ avec $x \neq 0$, on a $x/\|x\|_E \in \overline{B}_E(0,1)$ ainsi avec (2.18):

$$\|f(x/\|x\|_E)\|_F = \frac{\|f(x)\|_F}{\|x\|_E} \leq M \Rightarrow \|f(x)\|_F \leq M\|x\|_E. \tag{2.19}$$

et la dernière inégalité dans (2.19) est évidente pour $x = 0$.

4) $\Rightarrow$ 5): Par linéarité, on a pour tout $(x, y) \in E \times E$:

$$\|f(x) - f(y)\|_F = \|f(x-y)\|_F \leq M\|x-y\|_E \tag{2.20}$$

ce qui montre que f est M-Lipschitzienne sur E.

5) $\Rightarrow$ 1) est évident.

□

Exemple 2.9 *Soit $E := C^0([-1,1], \mathbb{R})$. Considérons sur E, la norme 1:*

$$\|f\|_1 := \int_{-1}^{1} |f(t)|dt$$

et la norme uniforme:

$$\|f\|_\infty := \max\{|f(t)|, \, t \in [-1,1]\}.$$

Soit enfin pour tout $f \in E$, $T(f) := f(0)$. T est clairement une forme linéaire sur E (i.e. $T \in L(E, \mathbb{R})$) et pour tout $f \in E$:

$$|T(f)| \leq \|f\|_\infty$$

si bien que T est continue lorsque E est muni de la norme uniforme.

Nous allons voir que T n'est PAS continue lorsque E est munie de la norme $\|.\|_1$. Pour cela, considérons la suite de fonctions:

$$f_n(t) := \max(0, n(1 - n|t|)), \; t \in [-1,1], \; n \in \mathbb{N}^*.$$

Un calcul élémentaire montre que $\|f_n\|_1 = 1$ pour tout n et $T(f_n) = n \to +\infty$. Ainsi T n'est pas bornée sur la boule unité fermée de $(E, \|.\|_1)$ et donc T n'est pas continue.

Exercice 2.5 *Les notations étant celles de l'exercice précédent, étudier la continuité de l'application (linéaire!) S de E dans $\mathbb{R}^3$ définie pour tout $f \in E$ par*

$$S(f) := (\int_{-1}^{1} f(t)dt, \int_{0}^{1} t^2 f(t)dt, \int_{-1}^{1} e^t f(t)dt)$$

lorsque l'on munit E de la norme $\|.\|_1$ puis de la norme $\|.\|_\infty$ et $\mathbb{R}^3$ de n'importe quelle norme (cf. théorème 2.2).

Comme d'habitude, on s'attend à ce que les choses se passent bien en dimension finie, en effet:

Théorème 2.7 *Soit $(E, \|.\|_E)$ et $(F, \|.\|_F)$ deux e.v.n. Si E est de dimension finie alors $L(E,F) = L_c(E,F)$.*

Preuve:
Sans perte de généralité, on suppose que $E = \mathbb{R}^n$ et $\|.\|_E = \|.\|_\infty$. Munissons E d'une base $(e_1, .., e_n)$. Soit $f \in L(E,F)$ et $x \in E$, écrivons $x = \sum_i^n x_i e_i$, alors on a:

$$\|f(x)\|_F = \|\sum_{i=1}^{n} x_i f(e_i)\|_F \leq \sum_{i=1}^{n} |x_i| \|f(e_i)\|_F \leq \|x\|_E \sum_{i=1}^{n} \|f(e_i)\|_F$$

ce qui prouve que $f \in L_c(E,F)$. □

Remarque. Remarquons que la conclusion du théorème précédent est en général fausse si c'est F qui est de dimension finie (voir les exemples précédents).

2.6.1 Spaces of linear continuous maps

Théorème 2.8 *Soit $(E, \|.\|_E)$ et $(F, \|.\|_F)$ deux e.v.n.*

1. *Sur $L_c(E,F)$ l'application:*

$$f \mapsto \|f\|_{L_c(E,F)} := \sup\{\|f(x)\|_F \, : \, \|x\|_E \leq 1\}$$

   définit une norme.

2. *Si $(F, \|.\|_F)$ est un espace de Banach, $L_c(E,F)$ muni de la norme définie précédemment est un espace de Banach.*

Preuve:
L'assertion 1. est évidente et sa preuve laissée au lecteur.

Soit (f_n) une suite de Cauchy de $(L_c(E,F), \|.\|_{L_c(E,F)})$, soit g_n la restriction de f_n à $\overline{B}_E(0,1)$. On a $g_n \in C_b^0(\overline{B}_E(0,1), F)$ car f_n est continue et:

$$\|g_n\|_\infty = \|f_n\|_{L_c(E,F)}. \tag{2.21}$$

Par définition de g_n on a aussi pour tout $(p,q) \in \mathbb{N}^2$:

$$\|g_p - g_q\|_\infty = \|f_p - f_q\|_{L_c(E,F)}. \tag{2.22}$$

Ceci implique que (g_n) est de Cauchy dans $(C_b^0(\overline{B}_E(0,1), F), \|.\|_\infty)$, donc, grâce au théorème 2.5, g_n converge vers une limite g dans $(C_b^0(\overline{B}_E(0,1), F), \|.\|_\infty)$. Définissons alors f par $f(0) = 0$ et:

$$f(x) = \|x\| g(\frac{x}{\|x\|}). \tag{2.23}$$

Notons d'abord que $g = f$ sur $\overline{B}_E(0,1)$, en effet $g(0) = f(0) = 0$ et si $x \in \overline{B}_E(0,1) \setminus \{0\}$, pour tout n on a:

$$\|x\| g_n(\frac{x}{\|x\|}) = \|x\| f_n(\frac{x}{\|x\|}) = f_n(x) = g_n(x)$$

et donc $g = f$ sur $\overline{B}_E(0,1)$, en passant à la limite dans la relation précédente.

Montrons que f est linéaire: soit $(x_1, x_2, t) \in E \times E \times \mathbb{R}$, pour tout n, par linéarité de f_n, on a [3]:

$$\begin{aligned} 0 =& f_n(x_1 + tx_2) - f_n(x_1) - t f_n(x_2) \\ =& \|x_1 + tx_2\| g_n(\frac{x_1 + tx_2}{\|x_1 + tx_2\|}) - \|x_1\| g_n(\frac{x_1}{\|x_1\|}) - t\|x_2\| g_n(\frac{x_2}{\|x_2\|}) \end{aligned}$$

en passant à la limite on a $f(x_1 + tx_2) = f(x_1) + tf(x_2)$. On en déduit donc que f est linéaire.

Puisque $g = f$ sur $\overline{B}_E(0,1)$, f est bornée sur $\overline{B}_E(0,1)$ et donc $f \in L_c(E,F)$. Enfin, comme $g_n = f_n$ et $g = f$ sur $\overline{B}_E(0,1)$, on a:

$$\|g_n - g\|_\infty = \|f_n - f\|_{L_c(E,F)}$$

d'où l'on déduit que f_n converge vers f dans $(L_c(E,F), \|.\|_{L_c(E,F)})$.

□

[3] Dans ce qui suit, on fera un léger abus de notation, en posant $\|x\| g_n(x/\|x\|) = 0$ pour $x = 0$.

Remarque. Soit $f \in L_c(E,F)$ et $x \in E \setminus \{0\}$ puisque $x/\|x\|_E$ est de norme 1 on a:

$$\|f\left(\frac{x}{\|x\|_E}\right)\|_F \leq \|f\|_{L_c(E,F)}$$

ce qui par homogénéité donne aussi:

$$\|f(x)\|_F \leq \|f\|_{L_c(E,F)}\|x\|_E. \tag{2.24}$$

Evidemment (2.24) est aussi vérifiée par $x = 0$. Il faut retenir (2.24) qui s'avère très utile dans la pratique.

Remarque. Notons que dans le théorème précédent (comme dans le théorème 2.5) c'est l'espace d'arrivée qui doit être un Banach (l'espace de départ est un evn quelconque).

Définition 2.8 *Soit $(E, \|.\|)$ un $\mathbb{R}$-evn, on appelle dual topologique de E et l'on note E' l'espace vectoriel des formes linéaires continues sur E: $E' := L_c(E, \mathbb{R})$*

On munit E' de la norme "duale" de la norme de E:

$$\forall f \in E', \|f\|_{E'} := \sup\{|f(x)| : \|x\|_E \leq 1\} \tag{2.25}$$

Il résulte du théorème (2.8) et de la complétude de $\mathbb{R}$ que $(E', \|.\|_{E'})$ est un espace de Banach.

Attention: ne pas confondre le dual algébrique de E, $E^* := L(E, \mathbb{R})$ et son dual topologique E' (je vous renvoie aux exemples du début du chapitre).

2.6.2 Bilinear continuous maps

On va maintenant étendre les résultats précédents aux applications bilinéaires. Les preuves sont analogues à celles des paragraphes précédents et donc laissées en exercice au lecteur.

Etant donnés trois $\mathbb{R}$-ev E, F et G, on appelle application bilinéaire de $E \times F$ à valeurs dans G toute application:

$$a : \left\{ \begin{array}{ccc} E \times F & \to & G \\ (x,y) & \mapsto & a(x,y) \end{array} \right.$$

telle que:

- pour tout $y \in F$, l'application $x \mapsto a(x,y)$ est linéaire de E dans G,

- pour tout $x \in E$, l'application $y \mapsto a(x,y)$ est linéaire de F dans G.

On note $L_2(E \times F, G)$ l'ensemble des applications bilinéaires de $E \times F$ à valeurs dans G. On vérifie aisément que $L_2(E \times F, G)$ a une structure de $\mathbb{R}$-ev.

Exemple 2.10 *$E = F = \mathbb{R}^n$, $G = \mathbb{R}$ et $a(x,y) = \sum_{i=1}^n x_i y_i$.*

Exemple 2.11 *$E = M_n(\mathbb{R})$, $F = G = \mathbb{R}^n$ et l'application qui à $(A,x) \in E \times F$ associe Ax.*

Exemple 2.12 *E_0 $\mathbb{R}$-ev quelconque, $E = F = G = L(E_0)$ et l'application qui à $(u,v) \in E \times F$ associe $u \circ v$.*

Lorsque E, F et G sont munies de normes respectives $\|.\|_E$, $\|.\|_F$, et $\|.\|_G$, on peut s'intéresser à la continuité des éléments de $L_2(E \times F, G)$. On note $L_{2,c}(E \times F, G)$ l'ensemble des éléments continus de $L_2(E \times F, G)$. Notons que $L_{2,c}(E \times F, G)$ est un sev de $L_{2,c}(E \times F, G)$. On a alors la caractérisation:

Théorème 2.9 *Soit $(E, \|.\|_E)$, $(F, \|.\|_F)$ et $(G, \|.\|_G)$ trois e.v.n, et $a \in L_2(E \times F, G)$, les assertions suivantes sont équivalentes:*

1. *$a \in L_{2,c}(E \times F, G)$,*
2. *il existe une constante $M \geq 0$ tq $\|a(x,y)\|_G \leq M\|x\|_E\|y\|_F$, $\forall (x,y) \in E \times F$.*

Preuve:

Adapter la preuve du théorème 2.6.
□
Lorsque E et F sont de dimension finie, on a simplement:

Théorème 2.10 *Soit $(E, \|.\|_E)$, $(F, \|.\|_F)$ et $(G, \|.\|_G)$ trois e.v.n. Si E et F sont de dimension finie alors $L_2(E \times F, G) = L_{2,c}(E, F)$.*

Preuve:

Adapter la preuve du théorème 2.10. □

Théorème 2.11 *Soit $(E, \|.\|_E)$, $(F, \|.\|_F)$ et $(G, \|.\|_G)$ trois e.v.n*

1. *Sur $L_{2,c}(E \times F, G)$ l'application:*

$$a \mapsto \|a\|_{L_{2,c}(E\times F,G)} := \sup\{\|a(x,y)\|_G \,:\, \|x\|_E \leq 1,\ \|y\|_F \leq 1\}$$

définit une norme.

2. *Si $(G, \|.\|_G)$ est un espace de Banach, $L_{2,c}(E \times F, G)$ muni de la norme définie précédemment est un espace de Banach.*

Preuve:

Adapter la preuve du théorème 2.8. □
Noter que si $a \in L_{2,c}(E \times F, G)$, on a:

$$\|a(x,y)\|_G \leq \|a\|_{L_{2,c}(E\times F,G)}\|x\|_E\|y\|_F \ \forall (x,y) \in E \times F. \tag{2.26}$$

2.6.3 A useful isomorphism

Nous allons voir ici que l'on peut identifier $L(E, L(F, G))$ (respectivement $L_c(E, L_c(F, G))$) à $L_2(E \times F, G)$ (respectivement $L_{2,c}(E \times F, G)$). Cette identification est particulièrement utile en calcul différentiel dès lors que l'on considère des différentielles d'ordre 2 ou plus.

Plus précisément, soit $v \in L(E, L(F, G))$ et définissons pour tout $(x, y) \in E \times F$:

$$a_v(x,y) := (v(x))(y)$$

il est immédiat de vérifier que a_v est bilinéaire: $a_v \in L_2(E \times F, G)$. Soit alors Φ l'application:

$$\Phi : \left\{ \begin{array}{ccc} L(E, L(F,G)) & \rightarrow & L_2(E \times F, G) \\ v & \mapsto & a_v \end{array} \right.$$

Il est clair que Φ est linéaire (donc si on veut absolument utiliser des notations, $\Phi \in L(L(E, L(F, G)), L_2(E \times F, G))$).

Soit maintenant $a \in L_2(E \times F, G)$, alors pour tout $x \in E$, l'application $a(x, .)$ appartient à $L(F, G)$ ($a(x, .)(y) := a(x, y)$ pour tout $y \in F$). Par ailleurs par bilinéarité on a pour tout $(x_1, x_2, \lambda) \in E^2 \times F$:

$$a(\lambda x_1 + x_2, .) = \lambda a(x_1, .) + a(x_2, .).$$

Ce qui signifie que l'application:

$$A_a : \left\{ \begin{array}{ccc} E & \rightarrow & L(F, G) \\ x & \mapsto & A_a(x) := a(x, .) \end{array} \right.$$

appartient à $L(E, L(F,G))$. Soit maintenant

$$\Psi : \left\{ \begin{array}{ccc} L_2(E\times F, G) & \to & L(E, L(F,G)) \\ a & \mapsto & A_a \end{array} \right.$$

Soit $a \in L_2(E\times F, G)$ et $(x,y) \in E\times F$ on a:

$$(\Phi \circ \Psi)(a)(x,y) = (A_a(x))(y) = a(x,y)$$

a, x et y étant arbitraire on a donc

$$\Phi \circ \Psi = \text{id sur } L_2(E\times F, G)$$
$$\Psi \circ \Phi = \text{id sur } L(E, L(F,G)).$$

Autrement dit Φ est un isomorphisme et Ψ est l'inverse de Φ. L'isomorphisme Φ permet donc bien d'identifier $L(E, L(F,G))$ à $L_2(E\times F, G)$. L'identification précédente est purement algébrique. Supposons maintenant que E, F et G sont munies de normes respectives $\|.\|_E$, $\|.\|_F$, et $\|.\|_G$. On a alors

Théorème 2.12 *Soit Φ et Ψ définis comme précédemment, on a alors:*

1. *Soit $v \in L(E, L(F,G))$ alors on a:*

$$v \in L_c(E, L_c(F,G)) \Leftrightarrow \Phi(v) \in L_{2,c}(E\times F, G).$$

2. *Pour tout $v \in L_c(E, L_c(F,G))$, on a:*

$$\|v\|_{L_c(E,L_c(F,G))} = \|\Phi(v)\|_{L_{2,c}(E\times F,G)}$$

Preuve:

Par définition, $\Phi(v) \in L_{2,c}(E\times F, G)$ ssi il existe $M \geq 0$ tel que

$$\|v(x)(y)\|_G \leq M\|x\|_E\|y\|_F \ \forall (x,y) \in E\times F.$$

Ceci équivaut à: il existe $M \geq 0$ tel que

$$\forall x \in E,\ v(x) \in L_c(F,G) \text{ et } \|v(x)\|_{L_c(F,G)} \leq M\|x\|_E$$

ce qui signifie exactement que $v \in L_c(E, L_c(F,G))$.

Soit $v \in L_c(E, L_c(F,G))$, on a:

$$\begin{aligned} \|\Phi(v)\|_{L_{2,c}(E\times F,G)} &= \sup\{\|(v(x))(y)\|_G :\ \|x\|_E \leq 1,\ \|y\|_F \leq 1\} \\ &= \sup\{\|v(x)\|_{L_c(F,G)} :\ \|x\|_E \leq 1\} = \|v\|_{L_c(E,L_c(F,G))} \end{aligned}$$

□

Le théorème 2.12 exprime donc non seulement que $L_c(E, L_c(F,G)))$ et $L_{2,c}(E\times F, G)$ sont isomorphes mais en plus isométriques (Φ est une isométrie de $L_c(E, L_c(F,G))$ dans $L_{2,c}(E\times F, G)$, ces espaces étant munis de leur norme naturelle).

2.6.4 Linear maps in Banach Spaces

In Banach spaces, there are additional properties that follow from Baire's Theorem. The first one is the Banach-Steinhaus theorem, or principle of uniform boundedness:

Theorem 2.1 *Let E be a Banach Space and F be a normed vector space and let $(f_i)_{i\in I}$ be a family of $L_c(E,F)$. If,*

$$\forall x \in E, \sup_{i\in I} \|f_i(x)\|_F < +\infty$$

then

$$\sup_{i\in I} \|f_i\|_{L_c(E,F)} < +\infty.$$

Proof:
Set $E_n := \{x \in E \ : \ \|f_i(x)\|_F \leq n \ \forall i \in I\}$, then each E_n is closed and by assumption $\cup_n E_n = E$. It then follows from Baire's theorem that E_{n_0} has nonempty interior for some n_0 and then there exists $r > 0$ and $x_0 \in E$ such that

$$\|f_i(x_0 + ru)\|_F \leq n_0, \forall i \in E, \forall u \in \overline{B}_E(0,1)$$

so that for all $u \in \overline{B}_E(0,1)$ and all $j \in I$, one has

$$\|f_j(u)\|_F \leq \frac{1}{r}\left(n_0 + \sup_{i\in I} \|f_i(x_0)\|_F\right).$$

□

Another consequence of Baire's Theorem is the following open mapping principle:

Theorem 2.2 *Let E and F be two Banach Spaces and $f \in L_c(E,F)$ be continuous and surjective. Then f is an open mapping in the sense that $f(U)$ is open in F for every U open in E.*

Proof:
Due to the linearity of f it is enough to prove that there exists $r_0 > 0$ such that $B_F(0,r_0) \subset f(B_E(0,1))$. Let $F_n := \overline{nf(B_E(0,1))}$, since f is surjective, $F = \cup_n F_n$ and it follows from Baire's Theorem that there exists n_0 such that F_{n_0} has nonempty interior. There exists then $y_0 \in E$ and $\rho > 0$ such that $B_F(y_0,\rho) \subset \overline{f(B_E(0,n_0))}$, by linearity, we also have $B_F(-y_0,\rho) \subset \overline{f(B_E(0,n_0))}$ and then $B_F(0,\rho) = -y_0 + B_F(y_0,\rho) \subset \overline{f(B_E(0,2n_0))}$. By homogeneity, we then have $B_F(0,r) \subset \overline{f(B_E(0,1))}$ with $r = \rho/2n_0$.

Let us now prove that $B_F(0,r) \subset f(\overline{B}_E(0,2))$. Let $y \in B_F(0,r)$, there exists $x_1 \in B_E(0,1)$ such that $y - f(x_1) \in B_F(0,r/2)$, since $B_F(0,r/2) \subset \overline{f(B_E(0,1/2))}$, there exists $x_2 \in B_E(0,1/2)$ such that $y - f(x_1) - f(x_2) \in B_F(0,r/4)$. Iterating the argument, we find a sequence $(x_n)_n$ in E such that $\|x_n\|_E \leq 1/2^{n-1}$ and $\|y - f(x_1 + ... + x_n)\|_F \leq r/2^n$ for every n. Since $x_1 + ... + x_n$ is a Cauchy sequence, it converges to some $x \in \overline{B}_E(0,2)$ and by continuity $y = f(x)$, which proves that $B_F(0,r) \subset f(\overline{B}_E(0,2)) \subset f(B_E(0,5/2))$ and then $B_F(0,r_0) \subset f(B_E(0,1))$ with $r_0 = 2r/5$.

□

We deduce from the previous Theorem the following *automatic continuity* result due to Banach:

Theorem 2.3 *Let $(E, \|.\|_E)$ and $(F, \|.\|_F)$ be two Banach spaces and $f \in L_c(E,F)$, if f is invertible then $f^{-1} \in L_c(F,E)$.*

We end this section by the following useful result

Theorem 2.4 *Let E be a Banach space and let $f \in L_c(E)$ be such that $\|f\|_{L_c(E)} < 1$ then* $\mathrm{id} + \mathrm{f}$ *is invertible with*

$$(\mathrm{id} + f)^{-1} = \sum_{k=0}^{\infty} (-1)^k f^k$$

Proof:
Since $L_c(E)$ is a Banach Space and $\|f\|_{L_c(E)} < 1$,

$$S_n := \sum_{k=0}^{n} (-1)^k f^k$$

converges. Moreover $S_n \circ (\mathrm{id} + f) = \mathrm{id} + (-1)^n f^{n+1}$, we thus get the desired result by letting n tend to ∞.

□

2.7 One has to be cautious in infinite dimensions

Il s'agit dans ce paragraphe d'attirer votre attention sur le fait que certaines propriétés "bien commodes" des evn de dimension finie sont fausses en dimension infinie:

- dans un evn de dimension infinie, un fermé borné n'est pas automatiquement compact ou, ce qui revient au même, il peut exister des suites bornées sans sous-suite convergente,

- en dimension infinie, toutes les normes ne sont pas équivalentes (je vous renvoie à l'exemple 2.5),

- en dimension infinie, le choix d'une norme a de l'importance: muni de la norme uniforme $C^0([0,1],\mathbb{R})$ est un espace de Banach mais muni de la norme $\|.\|_1$, il n'est pas complet, comme nous allons le voir. Autrement dit, en dimension infinie il n'est pas automatique qu'un evn soit un Banach.

Bref, un certain nombre de propriétés topologiques automatiques en dimension infinie (complétude, compacité des fermés bornés, continuité des applications linéaires...) sont en défaut dès que l'on passe à la dimension infinie.

Passons maintenant en revue quelques exemples.

Une suite bornée sans valeur d'adhérence dans $(C_b^0(\mathbb{R},\mathbb{R}),\|.\|_\infty)$**:**

Posons pour $x\in\mathbb{R}$, $f_0(x):=\max(0,1-|x|)$ et pour tout $n\in\mathbb{N}$, $f_n(x):=f(x-n)$. La suite (f_n) est bornée dans $(C_b^0(\mathbb{R},\mathbb{R}),\|.\|_\infty)$, en effet: $\|f_n\|_\infty=\|f_0\|_\infty=1$. Supposons par l'absurde que la suite (f_n) admette une sous suite $(f_{\varphi(n)})_n$ qui converge vers une limite f dans $(C_b^0(\mathbb{R},\mathbb{R}),\|.\|_\infty)$. Notons d'abord que $f_n=0$ sur $]-\infty,n-1]$ on doit donc avoir $f=0$ sur $]-\infty,\varphi(n)-1]$ pour tout n et donc $f=0$ sur $\mathbb{R}$. Mais si $(f_{\varphi(n)})$ converge vers 0 (dans $(C_b^0(\mathbb{R},\mathbb{R}),\|.\|_\infty)$), alors $\|f_{\varphi(n)}\|_\infty$ tend vers 0, ce qui est absurde puisque $\|f_{\varphi(n)}\|_\infty=1$.

Une suite bornée sans valeur d'adhérence dans $(C^0([0,1],\mathbb{R}),\|.\|_\infty)$**:**

Soit pour tout $n\in\mathbb{N}^*$ et tout $t\in[0,1]$, $f_n(t):=t^{1/n}$. On a $\|f_n\|_\infty=1$ pour tout n. Supposons par l'absurde qu'une sous-suite $(f_{\varphi(n)})_n$ converge vers une limite f dans $(C^0([0,1],\mathbb{R}),\|.\|_\infty)$. En particulier pour tout $t\in[0,1]$, $f_{\varphi(n)}(t)$ converge vers $f(t)$. Comme $f_n(0)=0$ pour tout n on doit alors avoir $f(0)=0$. Pour $t\in]0,1]$, $f_n(t)$ converge vers 1 donc $f(t)=1$ pour $t\in]0,1]$ mais avec $f(0)=0$ ceci contredit la continuité supposée de f.

Une suite de Cauchy de $(C^0([-1,1],\|.\|_1)$ **qui ne converge pas:**

Pour $n\in\mathbb{N}^*$ et $t\in[-1,1]$ définissons

$$f_n(t)=\begin{cases}-1 & \text{si} \quad t\in[-1,-1/n]\\ nt & \text{si} \quad t\in[-1/n,1/n]\\ 1 & \text{si} \quad t\in[1/n,1]\end{cases}$$

Montrons d'abord que (f_n) est de Cauchy dans $(C^0([-1,1],\|.\|_1)$. Pour cela définissons la fonction (discontinue en 0):

$$f(t)=\begin{cases} -1 & \text{si} \quad t\in[-1,0[\\ 0 & \text{si} \quad t=0 \\ 1 & \text{si} \quad t\in]0,1]. \end{cases}$$

Un calcul immédiat donne:

$$\int_{-1}^{1}|f_n-f|=\frac{1}{n} \tag{2.27}$$

ainsi pour tout $p,q\geq N$ on a:

$$\|f_p-f_q\|_1\leq\int_{-1}^{1}|f_p-f|+\int_{-1}^{1}|f_q-f|\leq\frac{2}{N} \tag{2.28}$$

ce qui montre que la suite est de Cauchy.

Supposons par l'absurde que (f_n) converge dans $(C^0([-1,1],\|.\|_1)$ vers une limite g. Soit $\varepsilon>0$, pour tout $n\geq 1/\varepsilon$ on a $f_n=f$ sur $[-1,1]\setminus[-\varepsilon,\varepsilon]$ on a donc:

$$\|f_n-g\|_1\geq\int_{[-1,1]\setminus[-\varepsilon,\varepsilon]}|f-g|$$

en faisant $n\to+\infty$ on en déduit que $f=g$ sur $[-1,1]\setminus[-\varepsilon,\varepsilon]$ et comme $\varepsilon>0$ est quelconque on a $f=g$ sur $[-1,1]\setminus\{0\}$ ce qui contredit la continuité supposée de g.

Cet exemple montre que $(C^0([-1,1],\|.\|_1)$ n'est pas un espace de Banach.

Exercice 2.6 *Soit $f_n(t)=t^n$ pour $n\in\mathbb{N}^*$ et $t\in[0,1]$. Etudier la convergence simple, la convergence uniforme et la convergence en norme $\|.\|_1$ de la suite (f_n) dans $C^0([0,1],\mathbb{R})$.*

Montrer que (f_n) est de Cauchy dans $(C^0([0,1],\mathbb{R}),\|.\|_1)$. Conclure.

One has to be particularly cautious when dealing with compactness issues in infinite-dimensions. Indeed, the compactness of the closed unit ball (and then of every set with nonempty interior) is ALWAYS false in infinite dimensions as stated in the next result due to Riesz:

Theorem 2.5 *Let E be a normed space then the closed unit ball of E, $\overline{B}$ is compact if and only if E is finite dimensional.*

Proof:
Let us assume that $\overline{B}$ is compact and let us prove that E is finite dimensional. By the Bolzano-Weierstrass Theorem, since $\overline{B}$ is compact it can be covered by finitely many balls of radius $1/2$: $\overline{B} \subset \cup_{i=1}^{k} \overline{B}(x_i, 1/2)$ for some $x_1, ..., x_k$ in $\overline{B}$. We shall prove that $E = F$ with F the vector space spanned by $x_1, ..., x_k$. By homogeneity, it is enough to prove that $\overline{B} \subset F$. Let then $x \in \overline{B}$, there is some $i_0 \in \{1, ...k\}$ and $\varepsilon_0 \in \overline{B}$ such that $x = x_{i_0} + \varepsilon_0/2$. Then there exists $i_1 \in \{1, ...k\}$ and $\varepsilon_1 \in \overline{B}$ such that $\varepsilon_0 = x_{i_1} + \varepsilon_1/2$ so that $x = x_{i_1} + x_{i_2}/2 + \varepsilon_1/4$. Iterating the argument, at step n, we can write

$$x = x_{i_0} + \frac{x_{i_1}}{2} + ... + \frac{x_{i_n}}{2^n} + \frac{\varepsilon_n}{2^{n+1}}, \text{ with } \varepsilon_n \in \overline{B}$$

for some indices i_l all in $\{1, ..., k\}$. Put differently, we have

$$x = y_n + \frac{\varepsilon_n}{2^{n+1}}, \; y_n = x_{i_0} + \frac{x_{i_1}}{2} + ... + \frac{x_{i_n}}{2^n} \in F.$$

Since (y_n) is a Cauchy sequence in the finite-dimensional space F it converges to some $y \in F$ and then $x \in F$.

□

Chapter 3

Convexity

3.1 Convex sets and convex functions

In what follows, E will denote a real vector space. Let us recall the basic definitions:

Definition 3.1 *A subset C of E is convex if and only for every x and y in C and every $\lambda \in [0,1]$, $\lambda x + (1-\lambda)y \in C$.*

Basic examples of convex sets are subspaces, half spaces, balls etc... Let us also remark that intersections or convex sets are convex.

Let us remark that an intersection of convex sets is convex. Let us also remark that C is convex iff for every $p \in \mathbb{N}^*$, every $x_1, ..., x_p \in C^p$ and every $\lambda_1, ..., \lambda_p$ such that each $\lambda_i \geq 0$ and $\sum_{i=1}^p \lambda_i = 1$ one has

$$\sum_{i=1}^{p} \lambda_i x_i \in C.$$

Any vector that can be written in the form $\sum_{i=1}^p \lambda_i x_i$ with nonnegative weights λ_i that sum to 1 is called a *convex combination* of the vectors x_i.

If A is a nonempty subset of E, the intersection of all convex sets containing A is the smallest convex set containing A, it is called the *convex hull* of A and denoted $\mathrm{co}(A)$. It is easy to check that $\mathrm{co}(A)$ is the set of all convex cobinations of elements of A:

$$\mathrm{co}(A) = \left\{ \sum_{i=1}^{p} \lambda_i x_i, p \in \mathbb{N}^*, x_i \in A, \lambda_i \geq 0, \sum_{i=1}^{p} \lambda_i = 1 \right\}.$$

In the case where E has finite dimension d, Carathéodory's Theorem states that is enough to consider convex combinations of $d+1$ points:

Theorem 3.1 *If* $\dim(E) = d$ *and* A *is a nonempty subset of* E*, then*

$$\mathrm{co}(A) = \left\{ \sum_{i=1}^{d+1} \lambda_i x_i, x_i \in A, \lambda_i \geq 0, \sum_{i=1}^{d+1} \lambda_i = 1 \right\}.$$

We omit the proof of this result that we give only for the sake of completeness.

Definition 3.2 *Let* C *a convex subset of* E *and* $f : C \to \mathbb{R}$*, then* f *is said to convex iff for all* x *and* y *in* C *and* $\lambda \in [0,1]$*, one has* $f(\lambda x + (1-\lambda y)) \leq \lambda f(x) + (1-\lambda) f(y)$*. One says that* f *is concave if* $-f$ *is convex.*

Note that f is a convex function iff its epigraph

$$\mathrm{Epi}(f) := \{(x, \lambda) \in C \times \mathbb{R} : f(x) \leq \lambda\}$$

is a convex subset of $E \times \mathbb{R}$.

Note also that f is convex in C iff for every $p \in \mathbb{N}^*$, $x_1, ..., x_p$ in C and $\lambda_i \geq 0$ such that $\sum_{i=1}^{p} \lambda_i = 1$, one has:

$$f\left(\sum_{i=1}^{p} \lambda_i x_i \right) \leq \sum_{i=1}^{p} \lambda_i f(x_i).$$

Functions whose sublevels are convex are called quasiconvex:

Definition 3.3 *Let* C *be a convex subset of* E *and* $f : C \to \mathbb{R}$*, then* f *is said to be quasiconvex iff for all* $\lambda \in \mathbb{R}$*, the set* $\{x \in C : f(x) \leq \lambda\}$*, it is quasiconcave if* $-f$ *is quasiconvex.*

Quasiconcave functions are widely used in economics (where the convexity of indifference curves is a widely used assumption). Note that f is quasiconvex iff for every x and y in C and every $\lambda \in [0,1]$, one has:

$$f(\lambda x + (1-\lambda)y) \leq \max(f(x), f(y))$$

Of course convex functions are quasiconvex but the converse is not true (for instance $f(x) = x^3$ is quasiconvex on $\mathbb{R}$). On $\mathbb{R}$, a function is quasiconvex iff it is monotone or unimodal (that is, nonincreasing on some half line $(-\infty, a]$ and nondecreasing on the $[a, +\infty)$). In optimization, the interest of quasiconvex functions comes from the fact that the set where a quasiconvex function achieves its minimum is convex.

3.2 Projection on a closed convex set of a Hilbert space

Theorem 3.2 *Let $(H, \langle ., . \rangle)$ be a Hilbert space and C be a nonempty closed convex subset of H. For every $x \in H$, there exists a unique element of C called projection of x on C and denoted $p_C(x)$ s.t.:*

$$\|x - p_C(x)\| = \inf\{\|x - y\| \ , \ y \in C\}.$$

Moreover $p_C(x)$ is characterized by: $p_C(x) \in C$ and the variational inequalities:

$$\langle x - p_C(x), y - p_C(x) \rangle \leq 0, \ \forall y \in C. \tag{3.1}$$

Proof:

Let us denote $d^2(x, C)$ the squared distance of x to C:

$$d^2(x, C) := \inf\{\|x - y\|^2, \ y \in C\}.$$

Let us recall the following parallelogram identity:

$$\|\frac{u - v}{2}\|^2 + \|\frac{u + v}{2}\|^2 = \frac{1}{2}(\|u^2\| + \|v\|^2) \ , \ \forall (u, v) \in H^2. \tag{3.2}$$

Uniqueness

Suppose y_1 and y_2 belong to C and satisfy:

$$\|x - y_1\|^2 = \|x - y_2\|^2 = d^2(x, C). \tag{3.3}$$

We have $(y_1 + y_2)/2 \in C$ since C is convex, and then

$$\|x - (y_1 + y_2)/2\|^2 \geq d^2(x, C). \tag{3.4}$$

Applying (3.2) to $u = (x - y_1)$ and $v = (x - y_2)$ and using (3.3) and (3.4), we get:

$$\begin{aligned} d^2(x, C) &= \frac{1}{2}(\|x - y_1\|^2 + \|x - y_2\|^2) \\ &= \|x - (y_1 + y_2)/2\|^2 + \|(y_1 - y_2)/2\|^2 \geq d^2(x, C) + \|(y_1 - y_2)/2\|^2 \end{aligned} \tag{3.5}$$

so that $y_1 = y_2$.

Existence

For $n \in \mathbb{N}^*$, let $y_n \in C$ be such that:

$$\|x - y_n\|^2 \leq d^2(x, C) + 1/n^2 \tag{3.6}$$

Identity (3.2) applied to $u = (x - y_p)$ and $v = (x - y_q)$ gives

$$\frac{1}{2}(\|x - y_p\|^2 + \|x - y_q\|^2) = \|x - (y_p + y_q)/2\|^2 + \|(y_p - y_q)/2\|^2. \tag{3.7}$$

Since $(y_p - y_q)/2 \in C$, we have:

$$\|x - (y_p + y_q)/2\|^2 \geq d^2(x, C). \tag{3.8}$$

From (3.6), (3.7) and (3.8), we thus get

$$\|(y_p - y_q)\|^2 \leq \frac{1}{2p^2} + \frac{1}{2q^2}. \tag{3.9}$$

It follows from (3.9) that (y_n) is a Cauchy sequence thus converges to some limit denoted $p_C(x)$. Since C is closed, $p_C(x) \in C$ and passing to the limit in (3.6) yields $\|x - p_C(x)\|^2 = d^2(x, C)$.

Variational characterization

Let $y \in C$ and $t \in [0, 1]$, since $(1 - t)p_C(x) + ty \in C$, we have

$$\begin{aligned} \|x - p_C(x)\|^2 \leq \|x - ((1 - t)p_C(x) + ty)\|^2 &= \|x - p_C(x) - t(y - p_C(x))\|^2 \\ &= \|x - p_C(x)\|^2 + t^2\|y - p_C(x))\|^2 - 2t\,\langle x - p_C(x), y - p_C(x)\rangle \end{aligned} \tag{3.10}$$

Dividing by t and letting t go to 0^+ we deduce that $p_C(x)$ satisfies (3.1).

Conversely, assume that $z \in C$ satisfies:

$$\langle x - z, y - z\rangle \leq 0, \ \forall y \in C. \tag{3.11}$$

Let $y \in C$, then we have

$$\|x - y\|^2 = \|x - z\|^2 + \|z - y\|^2 + 2\,\langle x - z, z - y\rangle \geq \|x - z\|^2$$

which proves that $z = p_C(x)$.

□

Une première propriété de la projection sur un convexe fermé est donnée par:

Proposition 3.1 *Under the same assumptions and notations as in Theorem 3.2, for all $(x, y) \in H^2$, one has:*

$$\langle x - y, p_C(x) - p_C(y)\rangle \geq 0 \text{ and } \|p_C(x) - p_C(y)\| \leq \|x - y\| \tag{3.12}$$

In particular p_C is 1-Lipschitz continuous.

Proof:
Using the variational inequalities characterizing $p_C(x)$ and $p_C(y)$, we have:

$$\langle x - p_C(x), p_C(y) - p_C(x)\rangle \leq 0 \text{ et } \langle y - p_C(y), p_C(x) - p_C(y)\rangle \leq 0.$$

Summing these inequalities and using Cauchy-Schwarz inequality yields

$$\|p_C(x) - p_C(y)\|^2 \leq \langle p_C(x) - p_C(y), x - y\rangle \leq \|p_C(x) - p_C(y)\|\|x - y\| \tag{3.13}$$

which proves (3.12).
□

An important special case is when C is a closed subspace of H. In this case, p_C is linear (and continuous thanks to (3.12)): it is the orthogonal projection on C.

Proposition 3.2 *Let C be a closed subspace of the Hilbert space $(H, \langle ., .\rangle)$. Defining p_C as in theorem 3.2, for $x \in H$, $p_C(x)$ is characterized by:*

$$p_C(x) \in C \text{ and } (x - p_C(x)) \in C^{\perp}.$$

Moreover, $p_C \in L_c(H, C)$ and p_C is called the ortogonal projection on C.

Proof:
If $z \in C$ and $x - z \in C^{\perp}$ then for every $y \in C$ we have $\langle x - z, y - z\rangle = 0$ so that z satisfies (3.1). Conversely (3.1) implies that $\langle x - p_C(x), y - p_C(x)\rangle \leq 0$ for all $y \in C$, taking $y = 2p_C(x)$ and $y = p_C(x)/2$ we get $\langle x - p_C(x), p_C(x)\rangle = 0$ and then $\langle x - p_C(x), y\rangle \leq 0$ for all $y \in C$, since C is a subspace, we deduce that $(x - p_C(x)) \in C^{\perp}$. Finally, it remains to prove that p_C is linear, for x_1, x_2 in H and $t \in \mathbb{R}$, set $x = x_1 + tx_2$ and $z = p_C(x_1) + tp_C(x_2)$, we have $z \in C$ and $(x - z) \in C^{\perp}$ so that $z = p_C(x)$. □

Une conséquence importante du théorème de projection est que l'on peut identifier un Hilbert à son dual topologique. C'est l'objet du théorème de représentation de Riesz:

Théorème 3.1 *Soit $(H, \langle ., .\rangle)$ un espace de Hilbert, et $f \in H'$ alors il existe un unique $x \in H$ tel que:*

$$f(u) = \langle x, u\rangle \text{ , } \forall u \in H. \tag{3.14}$$

Preuve:
L'unicité est évidente et laissée au lecteur. Si $f = 0$, on prend $x = 0$, supposons donc $f \neq 0$, dans ce cas $F := \ker(f)$ est un hyperplan fermé de H. Soit $x_0 \in H$ tel que $f(x_0) = 1$. En appliquant la proposition 3.2, définissons y_0 la projection orthogonale de x_0 sur F, on a alors $x_0 \neq y_0$ puisque $x_0 \notin F$ et y_0 est caractérisé par:

$$y_0 \in F = \ker(f), \text{ et } (x_0 - y_0) \in F^{\perp}. \tag{3.15}$$

en particulier, comme $\langle x_0 - y_0, y_0 \rangle = 0$ on a:

$$\langle x_0 - y_0, x_0 \rangle = \|x_0 - y_0\|^2 \neq 0 \tag{3.16}$$

Définissons alors:

$$x := \frac{x_0 - y_0}{\|x_0 - y_0\|^2} = \frac{x_0 - y_0}{\langle x_0 - y_0, x_0 \rangle} \tag{3.17}$$

d'après (3.15), $x \in F^{\perp}$ et donc pour $u \in F$ on a $f(u) = \langle x, u \rangle = 0$. Par ailleurs:

$$\langle x, x_0 \rangle = \frac{\langle x_0 - y_0, x_0 \rangle}{\langle x_0 - y_0, x_0 \rangle} = 1 = f(x_0).$$

On conclut que (3.14) est vraie car $F \oplus \mathbb{R}x_0 = H$.
□

Remarque. Notons x_f la solution de:

$$\langle x, u \rangle = f(u), \forall u \in H.$$

Et considérons l'application T de H dans H' qui à f associe x_f. Il est facile de voir que $T \in L_c(H, H')$ et que T est un isomorphisme. On a même mieux (le démontrer en exercice) : T est une isométrie

$$\|T(f)\| = \|f\|_{H'}, \forall f \in H'.$$

3.3 Separation of convex sets

Théorème 3.2 *Soit $(H, \langle ., . \rangle$ un espace de Hilbert, $x_0 \in H$, C un convexe fermé tel que $x_0 \notin C$, alors il existe $p \in H$, $p \neq 0$ et $\varepsilon > 0$ tels que*

$$\langle p, x_0 \rangle \leq \langle p, y \rangle - \varepsilon, \forall y \in C. \tag{3.18}$$

Preuve:
Posons $K := C - x_0 = \{y - x_0,\ y \in C\}$, K est un convexe fermé et $0 \notin K$. Soit $p := p_K(0)$ la projection de 0 sur K, comme $O \notin K$ on a $p \neq 0$, par ailleurs p vérifie les inéquations variationnelles:

$$\left(\langle 0 - p, z - p\rangle \leq 0,\ \forall z \in K\right) \Longleftrightarrow \left(\langle p, z\rangle \geq \|p\|^2 > 0,\ \forall z \in K\right). \qquad (3.19)$$

De (3.19) et $K = C - x_0$, il vient donc:

$$\langle p, y\rangle \geq \langle p, x_0\rangle + \|p\|^2,\ \forall y \in C. \qquad (3.20)$$

□

Il faut bien comprendre géométriquement ce que signifie le théorème précédent: la séparation de x_0 et C exprime le fait que x_0 et C se situent de part et d'autre d'un hyperplan affine (parallèle à $p^\perp$) c'est à dire dans deux demi-espaces distincts. Le théorème précédent est un résultat de séparation stricte (présence du $\varepsilon > 0$), autrement dit C est inclus dans le demi-espace ouvert $\{x \in H :\ \langle p, x - x_0\rangle > \varepsilon/2\}$ tandis que x_0 est évidemment dans $\{x \in H :\ \langle p, x - x_0\rangle < \varepsilon/2\}$.

Remarque. La convexité de C est une hypothèse fondamentale : dans $\mathbb{R}^2$ soit $C = \overline{B}((0,0),1) \setminus B((0,1),1/2)$, $x_0 := (0,3/4) \notin C$ et on ne peut pas séparer x_0 de C.

Soit A et B deux parties non vides d'un ev, l'ensemble $A - B$ est défini par:

$$A - B := \{a - b,\ (a,b) \in A \times B\}.$$

Lemme 3.1 *Soit $(H, \langle .,. \rangle)$ un espace de Hilbert, A et B deux parties non vides de H. On a:*

1. *Si A et B sont convexes, alors $A - B$ est convexe,*

2. *Si A est compact et B est fermé, alors $A - B$ est fermé.*

Preuve:
La preuve de l'assertion 1. est immédiate et laissée au lecteur. Prouvons 2., supposons que la suite $x_n = a_n - b_n$ ($(a_n, b_n) \in A \times B$) converge vers une limite x. Comme A est compact, (a_n) admet une sous suite $(a_{\varphi(n)})$ qui converge vers un élément a de A. On en déduit que $b_{\varphi(n)} = a_{\varphi(n)} - x_{\varphi(n)}$ converge vers $a - x$, comme B est fermé $b := a - x$ est dans B donc $x \in A - B$.
□

Remarque. Pour A et B seulement fermés on n'a pas en général $A - B$ fermé. Dans $\mathbb{R}^2$ soit $A := \mathbb{R}_+ \times \{0\}$ et $B := \{(x, y) \in \mathbb{R}^2 : x \geq 1,\ y \geq 1/x\}$ A et B sont deux convexes fermés et $(0, 0) \notin A - B$. En considérant $a_n = (n, 0)$ et $b_n = (n, 1/n)$ il est facile de voir que $(0, 0) \in \overline{(A - B)}$.

Théorème 3.3 *Soit $(H, \langle ., . \rangle)$ un espace de Hilbert, K un convexe compact et C un convexe fermé de H tels que $K \cap C = \emptyset$, alors il existe $p \in H$, $p \neq 0$ et $\varepsilon > 0$ tels que*

$$\langle p, y \rangle \leq \langle p, x \rangle - \varepsilon,\ \forall (x, y) \in K \times C. \tag{3.21}$$

Preuve:
Posons $D := K - C$, D est un convexe fermé de H d'après le lemme 3.1 et $0 \notin D$ puisque $K \cap C = \emptyset$. En appliquant le théorème 3.2, on peut séparer (strictement) 0 de D et donc il existe $p \in H$, $p \neq 0$ et $\varepsilon > 0$ tels que

$$0 \leq \langle p, z \rangle - \varepsilon,\ \forall z \in D. \tag{3.22}$$

C'est à dire, par définition de D:

$$\langle p, y \rangle \leq \langle p, x \rangle - \varepsilon,\ \forall (x, y) \in K \times C. \tag{3.23}$$

□

Remarque. Il existe des théorèmes de séparation valables dans des cadres beaucoup plus généraux que celui des espaces de Hilbert. Les ingrédients de démonstration sont cependant plus délicats et dépassent le cadre de ce cours.

3.4 The Farkas-Minkowksi Lemma

Rappelons qu'une partie K d'un $\mathbb{R}$-ev E est un cône ssi:

$$\forall (t, x) \in \mathbb{R}_+ \times K, tx \in K.$$

Nous aurons d'abord besoin du lemme suivant:

Lemme 3.2 *Soit E un evn, $q \in \mathbb{N}^*$ et $(a_1, ..., a_q) \in E^q$ et soit :*

$$K := \left\{ \sum_{i=1}^{q} \lambda_i a_i,\ (\lambda_1, ..., \lambda_q) \in \mathbb{R}_+^q \right\}$$

Alors K est un cône convexe fermé de E.

Preuve:
Le fait que K soit un cône convexe est évident. Pour montrer qu'il est fermé, faisons une récurrence sur q. Pour $q = 1$, le résultat est évident. Faisons donc l'hypothèse au rang $q \geq 1$ que pour tout $(a_1, ..., a_q) \in E^q$, le cône:

$$\left\{ \sum_{i=1}^{q} \lambda_i a_i,\ (\lambda_1, ..., \lambda_q) \in \mathbb{R}_+^q \right\}$$

est fermé. Soit maintenant, $(a_1, ..., a_{q+1}) \in E^{q+1}$, il s'agit de montrer que le cône:

$$K := \left\{ \sum_{i=1}^{q+1} \lambda_i a_i,\ (\lambda_1, ..., \lambda_{q+1}) \in \mathbb{R}_+^{q+1} \right\}$$

est fermé. Considérons d'abord le cas, où $-a_i \in K$ pour $i = 1, ..., q+1$, dans ce cas K est le sev engendré par les vecteurs $(a_1, ..., a_{q+1})$, K est donc un sev de dimension finie de E, il est par conséquent fermé (s'en convaincre!). Supposons maintenant qu'il existe $i \in \{1, ..., q+1\}$ tel que $-a_i \notin K$, quitte à permuter les indices nous pouvons supposer

$$-a_{q+1} \notin K. \tag{3.24}$$

Montrons que K est fermé. Rappelons d'abord que par hypothèse de récurrence le cône suivant est fermé:

$$K_0 := \left\{ \sum_{i=1}^{q} \lambda_i a_i,\ (\lambda_1, ..., \lambda_{q+1}) \in \mathbb{R}_+^{q+1} \right\}.$$

Soit $y_n \in K^{\mathbb{N}}$ convergeant dans E vers y, il s'agit de montrer que $y \in K$. Pour tout $n \in \mathbb{N}$ il existe $\lambda_{i,n} \geq 0$, $i = 1, ..., q$ et $\mu_n \geq 0$ tel que:

$$y_n = \sum_{i=1}^{q} \lambda_{i,n} a_i + \mu_n a_{q+1} = z_n + \mu_n a_{q+1} \tag{3.25}$$

($z_n := \sum_{i=1}^{q} \lambda_{i,n} a_i \in K_0$). Montrons que μ_n est bornée : sinon il existerait une sous suite que nous noterons encore μ_n tendant vers $+\infty$, en divisant (3.25) par μ_n, en passant à la limite, et en utilisant le fait que K_0 est fermé, on obtiendrait:

$$-a_{q+1} = \lim_n \frac{z_n}{\mu_n} \in K_0$$

ce qui contredirait (3.24). Comme μ_n bornée on peut, à une sous-suite près, supposer que μ_n converge vers $\mu \geq 0$. Comme $y_n = z_n + \mu_n a_{q+1}$ converge vers y on en déduit que z_n converge vers $z = y - \mu a_{q+1}$, en utilisant à nouveau que K_0 est fermé on a $z \in K_0$ et donc $y = z + \mu a_{q+1}$ appartient à K.

□

Le lemme de Farkas s'énonce comme suit:

Proposition 3.3 *Soit $(H, \langle ., . \rangle)$ un espace de Hilbert, $q \in \mathbb{N}^*$ et $(a, a_1, ..., a_q) \in H^{q+1}$, alors les propriétés suivantes sont équivalentes:*

1. *pour tout $x \in H$, si $\langle a_i, x \rangle \leq 0$ pour $i = 1, ..., q$ alors $\langle a, x \rangle \leq 0$,*
2. *il existe $(\lambda_1, ..., \lambda_q) \in \mathbb{R}^q_+$ tels que $a = \sum_{i=1}^q \lambda_i a_i$.*

Preuve:
L'implication 2. $\Rightarrow$ 1. est évidente. Remarquons que 2. signifie simplement que $a \in K$ avec

$$K := \left\{ \sum_{i=1}^q \lambda_i a_i,\ (\lambda_1, ..., \lambda_q) \in \mathbb{R}^q_+ \right\}.$$

Supposons que 1. est satisfaite et $a \notin K$. En vertu du lemme 3.2 K est convexe fermé: on peut donc séparer strictement a de K. Il existe donc $x \in H$ et $\varepsilon > 0$ tel que:

$$\sup_{p \in K} \langle p, x \rangle \leq \langle a, x \rangle - \varepsilon. \tag{3.26}$$

Soit $p \in K$ comme $tp \in K$ pour tout $t > 0$, (3.26) implique en particulier:

$$\sup_{t>0} t \langle p, x \rangle < +\infty$$

ce qui implique donc $\langle p, x \rangle \leq 0$ pour tout $p \in K$. Comme $0 \in K$, il vient donc:

$$\sup_{p \in K} \langle p, x \rangle = 0$$

En reportant dans (3.26), on a donc:

$$\sup_{p \in K} \langle p, x \rangle = 0 \leq \langle a, x \rangle - \varepsilon. \tag{3.27}$$

Ceci implique enfin que $\langle a_i, x \rangle \leq 0$ pour $i = 1, ..., q$ et $\langle a, x \rangle \geq \varepsilon > 0$ ce qui contredit 1.. □

Une conséquence immédiate du lemme de Farkas est la variante suivante:

Corollaire 3.1 *Soit $(H, \langle ., . \rangle)$ un espace de Hilbert, $(p, q) \in \mathbb{N}^* \times \mathbb{N}^*$ et $(a_1, ..., a_p, a_{p+1}, ..., a_{p+q}, a) \in H^{p+q+1}$, alors les propriétés suivantes sont équivalentes:*

1. *pour tout $x \in H$, si $\langle a_i, x \rangle \leq 0$ pour $i = 1, ..., p$ et $\langle a_i, x \rangle = 0$ pour $i = p+1, ..., p+q$ alors $\langle a, x \rangle \leq 0$,*
2. *il existe $(\lambda_1, ..., \lambda_p) \in \mathbb{R}^p_+$ et $(\lambda_{p+1}, ..., \lambda_{p+q}) \in \mathbb{R}^q$ tels que $a = \sum_{i=1}^{p+q} \lambda_i a_i$.*

Chapter 4

Fixed-point theorems

4.1 Preliminaries

Let us denote by $\overline{B}^d$ the closed (euclidean) unit ball of $\mathbb{R}^d$ and S^{d-1} its boundary i.e the unit sphere of $\mathbb{R}^d$. Let us start with the following theorem which states that there is no C^1 retraction of $\overline{B}^d$. The proof uses results from differential calculus, in particular the inverse function theorem (see chapter 7), that will be proven later on.

Theorem 4.1 *There does not exist any C^1 map $f : \overline{B}^d \to S^{d-1}$ such that $f(x) = x$ for all $x \in S^{d-1}$.*

Proof:
Assume by contradiction that $f : \overline{B}^d \to S^{d-1}$ is C^1 and such that $f(x) = x$ for all $x \in S^{d-1}$. For $t \in (0, 1)$ and $x \in B$, let us set

$$f_t(x) := (1 - t)x + tf(x).$$

By convexity, we have $f_t(\overline{B}^d) \subset \overline{B}^d$. Moreover, f is M-Lipschitz with $M = \sup_{x \in \overline{B}^d} \|f'(x)\|$ and $f_t' - \mathrm{id} = t(f' - \mathrm{id})$. Thus choosing $t \in (0, t_0)$ with $t_0 = (1 + M)^{-1}$) and invoking Theorem 2.4, we have that $f_t'(x)$ is invertible for every $x \in B^d$. In particular, by the inverse function theorem, for every $x \in B^d$, f_t is a C^1-diffeomorphism from some neighbourhood of x to some neighbourhood of $f_t(x)$. By theorem 7.1, we also deduce that $f_t(B^d)$ is open. Let x and y be in $\overline{B}^d$, we have

$$\|f_t(x) - f_t(y)\| = \|(1 - t)(x - y) + t(f(x) - f(y))\| \geq ((1 - t) - tM)\|x - y\|$$

and since $1 > t(1+M)$, we deduce that f_t is injective. Therefore, by Theorem 7.1, f_t is a C^1 diffeomorphism from B^d to $f_t(B^d) \subset B^d$. Let us now prove that

$f_t(B^d) = B^d$, assume by contradiction that there exists some $y \in B^d \setminus f_t(B^d)$ and let $z \in f_t(B^d)$, since $f_t(B^d)$ is open, $y_\lambda := z + \lambda(y - z) \in f_t(B^d)$ for $\lambda > 0$ small enough. Now let us set

$$\lambda^* := \sup\{\lambda \in [0,1] \ : \ y_\lambda \in f_t(B^d)\}, y^* := y_{\lambda^*}.$$

It is clear that $y^* \in f_t(\overline{B}^d)$ let us prove that $y^* \in f_t(B^d)$. If not one would have $y^* = f_t(x)$ with $x \in S^{d-1}$ and since $f_t(x) = x$ for $x \in S^{d-1}$ we would have $y^* = x \in S^{d-1}$ contradicting the fact that $(z, y]$ is included in B^d. If $\lambda^* < 1$, since $f_t(B^d)$ is a neighbourhood of y^* in particular $y_\lambda \in f_t(B^d)$ for $\lambda > \lambda^*$ close to λ^* contradicting the maximality of λ^*. Hence $y^* = y \in f_t(B^d)$. We thus have proved that for every $t \in (0, t_0)$, f_t is a C^1 diffeomorphism of B^d into itself. Now let us define for every $t \in [0, 1]$:

$$P(t) = \int_{B^d} \det(Df_t(x))dx$$

Since f_t is linear in t, $P(t)$ is a polynomial function of t. For $t \in (0, t_0)$, by the change of variables formula, $P(t)$ is the Lebesgue measure of $f_t(B^d) = B^d$, $P(t)$ is therefore constant on $(0, t_0)$, and then

$$P(1) = \int_{B^d} \det(Df(x))dx = P(0) > 0.$$

But $\det Df(x) = 0$ everywhere, since otherwise, by the inverse function theorem, $f(B^d)$ would have nonempty interior contradicting the fact that $f(B^d) \subset S^{d-1}$. □

4.2 Brouwer, Kakutani and Schauder Theorems

In this subsection, we shall state three very important fixed-point theorems. These theorems are very useful tools in economics (and more generally in nonlinear analysis) to prove *existence* results. Let us start with Brouwer's fixed-point theorem:

Theorem 4.2 *Let C be a convex compact subset of $\mathbb{R}^d$ and $f : C \to C$ be continuous, then f possesses (at least) a fixed point: there exists $\overline{x} \in C$ such that $f(\overline{x}) = \overline{x}$.*

Proof:
We will prove the result in the case $C = \overline{B}^d$ and will deduce the result

for any C that is homeomorphic to $\overline{B}^d$ for some d (we leave as an exercise the proof of the fact that if C is a convex compact subset of $\mathbb{R}^n$ then it is homeomorphic to $\overline{B}^d$ with d the dimension of the affine space spanned by C). Indeed assume that Φ is some homeorphism from C to $\overline{B}^d$ and f is a continuous map from C into itself, then $g = \Phi \circ f \circ \Phi^{-1}$ is continuous from $\overline{B}^d$ to itself and thus possesses a fixed point : $x \in \overline{B}^d$ such that $x = \Phi(f(\Phi^{-1}(x))$ and then $y = \Phi^{-1}(x)$ is a fixed point of f.

Now let us assume that f is a continuous function from $\overline{B}^d$ into itself and let us assume by contradiction that f has no fixed point so that

$$\inf\{\|x - f(x)\|,\ x \in \overline{B}^d\} > 0. \tag{4.1}$$

Since (4.1) continues to hold for functions that are uniformly sufficiently close to f and by a suitable regularization argument (by convolution say) we may assume that in addition $f \in C^1(\overline{B}^d)$. Now for all $x \in \overline{B}^d$ let $g(x)$ be the intersection of S^{d-1} with the half-line $\{x + \lambda(f(x) - x),\ \lambda \geq 0\}$. Because of (4.1), g is well-defined and easily seen to be seen to be C^1. Moreover, by construction, g maps $\overline{B}^d$ into S^{d-1} and $g(x) = x$ for every $x \in S^{d-1}$, which, thanks to Theorem 4.1, yields the desired contradiction. □

One can deduce from Brouwer's fixed point Theorem two important generalizations: an extension to infinite dimensions (Schauder's Theorem) and an extension to set-valued maps (Kakutani's Theorem). Schauder's theorem reads as:

Theorem 4.3 *Let C be a closed and bounded convex subset of some Banach space E and $f : C \to C$ be continuous and such that $f(C)$ is relatively compact (i.e. has compact closure), then f possesses (at least) a fixed point: there exists $\overline{x} \in C$ such that $f(\overline{x}) = \overline{x}$.*

Proof:
Since $f(C)$ is relatively compact, for every $\varepsilon > 0$, it can be covered by finitely many balls of radius ε : $f(C) \subset \cup_{i=1}^{N_\varepsilon} B(f(x_i^\varepsilon), \varepsilon)$ for some x_i^ε in C. Now let E_ε be the suspace of E spanned by $\{f(x_1^\varepsilon), ..., f(x_{N_\varepsilon}^\varepsilon)\}$. Let us denote by $B^c(f(x_i^\varepsilon), \varepsilon)$ the complement of $B(f(x_i^\varepsilon), \varepsilon)$ and set for every $x \in C$ and i:

$$\alpha_i^\varepsilon(x) := \frac{d(f(x), B^c(f(x_i^\varepsilon), \varepsilon))}{\sum_{j=1}^{N_\varepsilon} d(f(x), B^c(f(x_j^\varepsilon), \varepsilon))}$$

so that $\alpha_i^\varepsilon(x) > 0$ iff $f(x) \in B(f(x_i^\varepsilon), \varepsilon)$. Now let $C_\varepsilon := C \cap E_\varepsilon$ and for every $x \in C_\varepsilon$, let us set

$$f_\varepsilon(x) := \sum_{i=1}^{N_\varepsilon} \alpha_i^\varepsilon(x) f(x_i^\varepsilon)$$

by convexity of C, we have $f_\varepsilon(C_\varepsilon) \subset C_\varepsilon$ and f_ε is obviously continuous on C_ε. Since E_ε is finite dimensional, and C_ε is convex and compact in E_ε, Brouwer's Theorem gives the existence of some $x_\varepsilon \in C_\varepsilon$ such that $x_\varepsilon = f_\varepsilon(x_\varepsilon)$. By construction for every ε, x_ε belongs to the closed convex hull of $f(C)$, $\overline{\text{co}}(f(C))$ (that is the smallest closed convex set containing $f(C)$ or, put differently, the closure of $\text{co}(f(C))$). By Lemma 4.1, $\overline{\text{co}}(f(C))$ is compact, taking $\varepsilon = 1/n$, $x_n := x_{\varepsilon_n}$ we may therefore assume that x_n converges to some $\overline{x} \in \overline{co}(f(C)) \subset C$. Now we claim that $\overline{x}$ is a fixed-point of f. Indeed for every n, we have

$$f(\overline{x}) - f_{\varepsilon_n}(x_n) = \sum_{i=1}^{N_{\varepsilon_n}} \alpha_i^{\varepsilon_n}(x_n)(f(\overline{x}) - f(x_n) + f(x_n) - f(x_i^{\varepsilon_n})) \tag{4.2}$$

In the previous sum, there are only terms such that $\|f(x_n) - f(x_i^{\varepsilon_n})\| < \varepsilon_n$, so that

$$\|f(\overline{x}) - f_{\varepsilon_n}(x_n)\| \leq \|f(\overline{x}) - f(x_n)\| + \varepsilon_n.$$

This proves that $f_{\varepsilon_n}(x_n)$ converges to $f(\overline{x})$. We thus deduce that $f(\overline{x}) = \overline{x}$ by passing to the limit in the relation $f_{\varepsilon_n}(x_n) = x_n$.

□

In the previous proof we have used the following Lemma.

Lemme 4.1 *Let E be a Banach space and let K be a relatively compact subset of E, then $\overline{\text{co}}(K)$ is compact.*

Proof:

By Theorem 1.2, it is enough to prove that $\overline{\text{co}}(K)$ is precompact (it is complete since it is closed and E is a Banach space). Let $\varepsilon > 0$, we have to prove that $\overline{\text{co}}(K)$ can be covered by finitely many balls of radius ε. Since K is relatively compact, there exist p and $x_1, ..., x_p$ in K such that $K \subset \cup_{i=1}^p B(x_i, \varepsilon/3)$. Let $C := \text{co}\{x_1, ..., x_p\}$, C is actually compact hence there is some l and some $y_1, ..., y_l$ in C such that $C \subset \cup_{j=1}^l B(y_j, \varepsilon/3)$. Now let $z \in \text{co}(K)$, i.e.

$$z = \sum_{k=1}^m \lambda_k a_k$$

for some a_k in K and nonnegative weights λ_k summing to 1. Each a_k can be written as

$$a_k = x_{i_k} + \frac{\varepsilon}{3} v_k, \text{ for some } i_k \in \{1, ..., p\}, \text{ and } v_k \in B(0, 1).$$

We then have

$$z = \sum_{k=1}^{m} \lambda_k x_{i_k} + \frac{\varepsilon}{3} v,\ v := \sum_{k=1}^{m} \lambda_k v_k \in B(0,1).$$

Now we remark that

$$x = \sum_{k=1}^{m} \lambda_k x_{i_k} \in C$$

so that there is some j such that $x \in B(y_j, \varepsilon/3)$ and then $z \in B(y_j, 2\varepsilon/3)$. This proves that $\mathrm{co}(K) \subset \cup_{j=1}^{l} B(y_j, 2\varepsilon/3)$ and then $\overline{\mathrm{co}}(K) \subset \cup_{j=1}^{l} B(y_j, \varepsilon)$.

□

Kakutani's theorem, stated below, gives sufficient conditions for a set-valued map to have a fixed-point:

Theorem 4.4 *Let C be a convex compact subset of $\mathbb{R}^d$ and $F : C \to 2^C$ be a convex-valued set-valued map with a closed graph then F possesses (at least) a fixed point: there exists $\overline{x} \in C$ such that $\overline{x} \in F(\overline{x})$.*

Proof:
Since C is compact, for every $\varepsilon > 0$, C can be covered by finitely many balls of radius ε, $C \subset \cup_{i=1}^{N_\varepsilon} B(x_i^\varepsilon, \varepsilon)$ with x_i^ε in C. Let us denote by $B^c(x_i^\varepsilon, \varepsilon)$ the complement of $B(x_i^\varepsilon, \varepsilon)$ and set for every $x \in C$ and i:

$$\alpha_i^\varepsilon(x) := \frac{d(x, B^c(x_i^\varepsilon, \varepsilon))}{\sum_{j=1}^{N_\varepsilon} d(x, B^c(x_j^\varepsilon, \varepsilon))}$$

so that $\alpha_i^\varepsilon(x) > 0$ iff $x \in B(x_i^\varepsilon, \varepsilon)$. Now let $y_i^\varepsilon \in F(x_i^\varepsilon)$ and set

$$f_\varepsilon(x) := \sum_{i=1}^{N_\varepsilon} \alpha_i^\varepsilon(x) y_i^\varepsilon, \forall x \in C.$$

Since C is convex, the continuous function f_ε maps C into C. By Brouwer's Theorem, f_ε admits a fixed-point x_ε. Since C is compact, setting $\varepsilon_n = 1/n$ and $x_n = x_{\varepsilon_n}$, we may assume that x_n converges to some $\overline{x} \in C$. Let us prove now that $\overline{x} \in F(\overline{x})$. Assume by contradiction that $\overline{x} \notin F(\overline{x})$, since $F(\overline{x})$ is convex and compact, the separation Theorem gives the existence of a $p \in \mathbb{R}^d$, $p \neq 0$ and an $\alpha \in \mathbb{R}$ such that

$$p \cdot \overline{x} < \alpha < \inf_{y \in F(\overline{x})} p \cdot y. \tag{4.3}$$

We claim, that there is some $r > 0$ such that for every $x \in B(\overline{x}, r)$, one has

$$p \cdot \overline{x} < \alpha < \inf_{y \in F(x)} p \cdot y. \tag{4.4}$$

Because otherwise, there would exist a sequence z_n converging to $\overline{x}$ and $y_n \in F(x_n)$ such that for every n

$$p \cdot y_n \leq \alpha$$

Since C is compact and F has a closed graph, there is some (not relabeled) subsequence of (z_n, y_n) converging to $\overline{x}, y$ with $y \in F(\overline{x})$, one would then have

$$p \cdot y \leq \alpha$$

a contradiction with (4.3). Now, we remark that

$$x_n = \sum_{i \,:\, d(x_n, x_i^{\varepsilon_n}) < \varepsilon_n} \alpha_i^{\varepsilon_n}(x_n) y_i^{\varepsilon_n}$$

choosing n large enough so that $d(x_n, \overline{x}) < r/2$ and $\varepsilon_n < r/2$, all the indices in the previous sum are such that $x_i^{\varepsilon_n} \in B(\overline{x}, r)$. Taking the inner product with p we have

$$p \cdot x_n = \sum_{i \,:\, d(x_n, x_i^{\varepsilon_n}) < \varepsilon_n} \alpha_i^{\varepsilon_n}(x_n) p \cdot y_i^{\varepsilon_n}.$$

For n large enough, with (4.4), the right hand side of the previous equality is larger than α which contradicts the fact that the left hand side converges to $p \cdot \overline{x} < \alpha$.

□

4.3 Existence of Nash equilibria

Let us consider N players indexed by $i = 1, ..., N$. For each i, K_i denotes player i's strategy set (assumed to be a convex compact subset of some finite-dimensional space) and we set $K := \Pi_{i=1}^N K_i$. Each player is characterized by a payoff function $u_i : K \to \mathbb{R}$, we typically denote by $x_i \in K_i$ player i's strategy and by $x_{-i} \in K_{-i} := \Pi_{j \neq i} K_j$ the other palyers' strategies. We further assume that each u_i is continuous on K and that for each $x_{-i} \in_{-i}$, $u_i(., x_{-i})$: $x \in K_i \to u_i(x_i, x_{-i})$ is quasi-concave.

Nash equilibria are then defined as follows

Definition 4.1 *A Nash equilibrium is an $x = (x_1, ..., x_N) \in K$ such that for every player i*

$$u_i(x_i, x_{-i}) = \max_{x \in K_i} u_i(x, x_{-i}).$$

Now, let us define for each $x \in K$, the Best-Reply set of x:

$$\mathrm{BR}(x) = (\mathrm{BR}_1(x_{-1}) \times ..... \times \mathrm{BR}_\mathrm{N}(x_{-N}))$$

where $BR_i(x)$ denotes the set of best replies to x_{-i}:

$$\mathrm{BR}_i(x_{-i}) = \{x \in K_i : u_i(x, x_{-i}) \geq u_i(y, x_{-i}),\ \forall y \in K_i\}.$$

This defines the Best-Reply set-valued map $\mathrm{BR} : K \to 2^K$. Clearly, Nash-equilibria are exactly fixed-points of the Best-Reply map. Moreover, our assumptions ensure that BR is a nonempty convex-compact valued map with a closed graph, it then admits a fixed-point as a consequence of Kakutani's Theorem. We thus have proved

Theorem 4.5 *Under the assumptions of this paragraph, there exists at least one Nash equilibrium.*

Part II

Differential calculus

Chapter 5

First-order differential calculus

5.1 Several notions of differentiability

Dans ce qui suit on se donne $(E, \|.\|_E)$ et $(F, \|.\|_F)$ deux $\mathbb{R}$-evn, Ω un ouvert de E et f une application définie sur Ω à valeurs dans F. Si $x \in \Omega$ alors il existe $r > 0$ tel que $B(x, r) \subset \Omega$ en particulier si $h \in E$ et $t \in \mathbb{R}$ est assez petit (tel que $|t|\|h\|_E < r$) alors $x + th \in \Omega$. On a alors une première notion de dérivabilité : celle de dérivabilité dans la direction h:

Définition 5.1 *Soit $x \in \Omega$ et $h \in E$, on dit que f est dérivable en x dans la direction h ssi la limite suivante existe (au sens de la topologie de $(F, \|.\|_F)$):*

$$\lim_{t \to 0,\ t \neq 0} \frac{1}{t}(f(x + th) - f(x)).$$

Si cette limite existe, on l'appelle dérivée directionnelle de f en x dans la direction h et on la note $Df(x; h)$.

Notons que f est dérivable en x dans la direction h ssi les limites (à droite et à gauche) suivantes (au sens de la topologie de $(F, \|.\|_F)$):

$$\lim_{t \to 0^+} \frac{1}{t}(f(x + th) - f(x)) \text{ et } \lim_{t \to 0^-} \frac{1}{t}(f(x + th) - f(x)).$$

existent et sont égales. Ceci conduit à la définition:

Définition 5.2 *Soit $x \in \Omega$ et $h \in E$, on dit que f est dérivable à droite en x dans la direction h ssi la limite suivante existe (au sens de la topologie de $(F, \|.\|_F)$):*

$$\lim_{t \to 0^+} \frac{1}{t}(f(x + th) - f(x)).$$

Si cette limite existe, on l'appelle dérivée à droite de f en x dans la direction h et on la note $D^+f(x; h)$.

En remarquant que si f est dérivable à droite en x dans la direction $-h$ alors:

$$D^+f(x;-h) = -\lim_{t\to 0^-}\frac{1}{t}(f(x+th)-f(x))$$

nous en déduisons que f est dérivable en x dans la direction h ssi f est dérivable à droite en x dans les directions h et $-h$ et:

$$D^+f(x;-h) = -D^+f(x;h).$$

Trois exercices (très) faciles, avant d'aller plus loin:

Exercice 5.1 *Montrer que $Df(x;0)$ existe (aucune hypothèse sur f) et vaut 0.*

Exercice 5.2 *Montrer que si f est dérivable à droite en x dans la direction h, alors pour tout $\lambda > 0$, f est dérivable à droite en x dans la direction λh et:*

$$D^+f(x;\lambda h) = \lambda D^+f(x;h).$$

Exercice 5.3 *Montrer que si f est dérivable en x dans la direction h, alors pour tout $\lambda \in \mathbb{R}$, f est dérivable en x dans la direction λh et:*

$$Df(x;\lambda h) = \lambda Df(x;h).$$

Définition 5.3 *Soit $x \in \Omega$ on dit que f est Gâteaux-dérivable en x ssi f admet une dérivée directionnelle dans la direction h pour tout $h \in E$ et l'application $h \mapsto Df(x;h)$ est linéaire et continue. On note alors $Df(x;h) := D_Gf(x)(h)$ et $D_Gf(x) \in L_c(E,F)$ s'appelle la dérivée au sens de Gâteaux de f en x. On dit que f est Gâteaux dérivable sur Ω ssi f est Gâteaux-différentiable en chaque point de $x \in \Omega$.*

Remarque. La Gâteaux différentiabilité est une notion assez faible qui n'entraine pas automatiquement la continuité. Pour s'en persuader, on étudiera avec profit le comportement au voisinage de 0 de la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x,y) = \begin{cases} 1 & \text{si} \quad y = x^2 \text{ et } x \neq 0 \\ 0 & \text{sinon} \end{cases}$$

Remarque. Le fait que les dérivées directionnelles $Df(x;h)$ existent $\forall h \in E$ n'impliquent pas que f soit Gâteaux-dérivable en x. Pour s'en persuader, étudier la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x,y) = \begin{cases} 0 & \text{si} \quad (x,y) = (0,0) \\ \frac{x^3}{x^2+|y|} & \text{sinon} \end{cases}$$

Remarque. (importante) La définition de la Gâteaux-différentiabilité dépend du choix des normes sur E et F. Il est cependant facile de voir que le choix de normes équivalentes sur E et F conduit à la même définition. En particulier, si E et F sont de dimension finie, le choix de normes particulières est sans incidence sur la définition 5.3.

Une notion plus forte est la notion de différentiabilité au sens de Fréchet:

Définition 5.4 *Soit $x \in \Omega$ on dit que f est Fréchet-dérivable (ou simplement dérivable ou différentiable) en x ssi il existe $L \in L_c(E,F)$ et une fonction ε définie sur un voisinage de 0 dans E et à valeurs dans F tels que:*

$$f(x+h) = f(x) + L(h) + \|h\|_E \cdot \varepsilon(h) \text{ avec } \lim_{h\to 0} \|\varepsilon(h)\|_F = 0. \qquad (5.1)$$

Sous forme quantifiée, (5.1) signifie exactement : $\forall \varepsilon > 0$, $\exists \delta_\varepsilon > 0$ tel que pour tout $h \in E$, on a:

$$\|h\|_E \leq \delta_\varepsilon \Rightarrow \|f(x+h) - f(x) - L(h)\|_F \leq \varepsilon \|h\|_E$$

Remarque. (importante) La définition de la (Fréchet)-différentiabilité dépend du choix des normes sur E et F. Il est cependant facile de voir que le choix de normes équivalentes sur E et F conduit à la même définition (s'en convaincre à titre d'exercice facile). En particulier, si E et F sont de dimension finie, le choix de normes particulières est sans incidence sur la définition 5.4.

Remarque. Si f est dérivable en x alors f est continue en x (noter la différence avec la Gâteaux différentiabilité).

On écrit aussi usuellement (5.1) sous la forme synthétique:

$$f(x+h) = f(x) + L(h) + o(h) \qquad (5.2)$$

la notation $o(h)$ désignant une fonction qui "tend vers 0 (dans F) plus vite que h lorsque h tend vers 0 (dans E)", c'est à dire:

$$\lim_{h\to 0,\ h\neq 0} \frac{\|o(h)\|_F}{\|h\|_E} = 0. \qquad (5.3)$$

Remarque. Lorsque E est de dimension finie, en vertu du théorème 2.10 $L_c(E,F) = L(E,F)$ et donc on peut omettre la condition "L continue" dans la définition 5.4.

Lemme 5.1 *Soit $x \in \Omega$ s'il existe $L_1 \in L_c(E,F)$ et $L_2 \in L_c(E,F)$ tels que:*

$$f(x+h) = f(x) + L_1(h) + o(h) = f(x) + L_2(h) + o(h).$$

alors $L_1 = L_2$.

Preuve:
On a $(L_1 - L_2)h = o(h)$ donc pour $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout $h \in B(0,\delta)$, on a:

$$\|(L_1 - L_2)(h)\|_F \leq \varepsilon \|h\|_E$$

et donc

$$\|L_1 - L_2\|_{L_c(E,F)} \leq \varepsilon$$

comme $\varepsilon > 0$ est arbitraire, on en déduite $L_1 = L_2$.
□

Le lemme précédent montre qu'il existe au plus un élément L de $L_c(E,F)$ vérifiant (5.2), ceci permet de définir la dérivée (au sens de Fréchet) de f en x, $f'(x)$ de manière intrinsèque:

Définition 5.5 *Soit $x \in \Omega$ et f différentiable en x, on appelle différentielle (ou dérivée) de f en x, et l'on note $f'(x)$ l'unique élément de $L_c(E,F)$ vérifiant:*

$$f(x+h) = f(x) + f'(x)(h) + o(h).$$

On dit que f est différentiable sur Ω ssi f est différentiable en chaque point de $x \in \Omega$

Si f est différentiable en x alors f est Gâteaux différentiable en x, admet des dérivées directionnelles dans toutes les directions, et:

$$f'(x) = D_G f(x),\ Df(x;h) = f'(x)(h)\ \forall h \in E.$$

Bien retenir que la dérivée de f en x (qu'elle soit au sens Gâteaux ou Fréchet) est une application linéaire continue de E vers F. Lorsque f est différentiable sur Ω (Gâteaux ou Fréchet), sa dérivée ($f' : x \mapsto f'(x)$ ou $D_G f : x \mapsto D_G f(x)$) est donc une application définie sur Ω à valeurs dans $L_c(E,F)$.

Définition 5.6 *On dit que f est de classe C^1 sur Ω (ce que l'on note $f \in C^1(\Omega,F)$) ssi f est différentiable sur Ω et $f' \in C^0(\Omega, L_c(E,F))$. On dit que f (définie sur $\overline{\Omega}$ et à valeurs dans F) est de classe C^1 sur $\overline{\Omega}$ (ce que l'on note $f \in C^1(\overline{\Omega},F)$) si et seulement s'il existe un ouvert U de E contenant $\overline{\Omega}$ et $g \in C^1(U,F)$ tel que $f = g$ sur $\overline{\Omega}$.*

On a alors le:

Théorème 5.1 *Si f est Gâteaux-différentiable sur Ω et $D_G f \in C^0(\Omega, L_c(E,F))$ alors f est de classe C^1 sur Ω .*

Nous prouverons ce résultat au chapitre suivant. Le théorème 5.1 est très utile car il permet de procéder comme suit pour montrer en pratique que f est de classe C^1:

- **Etape 1:** On calcule $Df(x;h)$ pour $(x,h) \in \Omega \times E$.
- **Etape 2:** On montre que $h \mapsto Df(x;h)$ est linéaire continue donc f est Gâteaux-différentiable en x et $Df(x;h) = D_G f(x)(h)$.
- **Etape 3:** On montre que $D_G f : x \mapsto D_G f(x)$ est continue de Ω dans $L_c(E,F)$.

Concernant les applications bijectives, on a les définitions:

Définition 5.7 *Soit Ω un ouvert de E, Ω' un ouvert de F et f une bijection de Ω sur Ω' on dit que:*

- *f est un homéomorphisme de Ω sur Ω' si $f \in C^0(\Omega,\Omega')$ et $f^{-1} \in C^0(\Omega',\Omega)$,*
- *f est un C^1-difféomorphisme (ou difféomorphisme de classe C^1) de Ω sur Ω' si $f \in C^1(\Omega,\Omega')$ et $f^{-1} \in C^1(\Omega',\Omega)$.*

Lorsque l'espace de départ E est un espace de Hilbert et que l'espace d'arrivée est $\mathbb{R}$, alors la dérivée étant une forme linéaire continue sur E, on peut en utilisant le théorème de Riesz identifier la dérivée à un élément de E, cela conduit à la notion de vecteur gradient:

Définition 5.8 *Soit $(E, \langle .,. \rangle)$ un espace de Hilbert, Ω un ouvert de E, f une application définie sur E à valeurs réelles et $x \in \Omega$. Si f est Gâteaux-différentiable en x, on appelle gradient de f en x et l'on note $\nabla f(x)$ l'unique élément de E tel que:*

$$D_G f(x)(h) = \langle \nabla f(x), h \rangle \quad \textit{pour tout } h \in E.$$

Dans le cas particulier $E = \mathbb{R}^n$ (muni de son ps usuel), nous verrons par la suite que le gradient de f en x est le vecteur formé par les dérivées partielles par rapport aux n coordonnées de f en x.

5.2 Calculus rules

Proposition 5.1 *Soit f et g deux applications définies sur Ω à valeurs dans F et $x \in \Omega$, on a:*

1. *si f est constante au voisinage de x alors f est différentiable en x et $f'(x) = 0$,*
2. *si f est différentiable en x, pour tout $\alpha \in \mathbb{R}$, αf est différentiable en x et:*
$$(\alpha f)'(x) = \alpha f'(x)$$
3. *Si f et g sont différentiables en $x \in$ alors $f + g$ aussi et:*
$$(f + g)'(x) = f'(x) + g'(x)$$
4. *Si $L \in L_c(E, F)$ alors $L \in C^1(E, F)$ et: $L'(z) = L$ pour tout $z \in E$.*

Sur la dérivabilité des applications bilinéaires continues on a le résultat dont la preuve est laissée en exercice au lecteur:

Proposition 5.2 *Soit E, F et G, trois $\mathbb{R}$-evn, $a \in L_{2,c}(E \times F, G)$ alors $a \in C^1(E \times F, G)$ et pour tout $(x, y) \in E \times F$ et $(h, k) \in E \times F$, on a:*

$$a'(x, y)(h, k) = a(x, k) + a(h, y).$$

Proposition 5.3 *Soit E, $F_1, ..., F_p$ des $\mathbb{R}$-evn, Ω un ouvert de E, et pour $i = 1, ..., p$, f_i une applications définie sur Ω à valeurs dans F_i. Pour $x \in \Omega$ on définit:*

$$f(x) = (f_1(x), ..., f_p(x)) \in \prod_{i=1}^{p} F_i,$$

alors f est différentiable en $x \in \Omega$ ssi f_i est différentiable en $x \in \Omega$ pour $i = 1, ..., p$ et l'on a dans ce cas:

$$f'(x)(h) = (f_1'(x)(h), ..., f_p'(x)(h)) \text{ pour tout } h \in E$$

On peut noter le résultat précédent sous forme synthétique:

$$f' = (f_1, ..., f_p)' = (f_1', ..., f_p')$$

qui exprime que la dérivation se fait composante par composante.

Concernant la dérivabilité d'une composeé, on a:

Théorème 5.2 *Soit E, F et G trois evn, Ω un ouvert de E, U un ouvert de F, f une application définie sur Ω à valeurs dans F, g une application définie sur U à valeurs dans G et $x \in \Omega$. Si f est différentiable en x, $f(x) \in U$ et g est différentiable en $f(x)$ alors $g \circ f$ est différentiable en x et l'on a:*

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

Corollaire 5.1 *Si, en plus des hypothèses du théorème précédent, on suppose que f est de classe C^1 sur Ω, que $f(\Omega) \subset U$ et que g est de classe C^1 sur U alors $g \circ f$ est de classe C^1 sur Ω.*

Sur la dérivabilité d'un produit (scalaire $\times$ vecteur), on a:

Proposition 5.4 *Soit E, et F et deux evn, Ω un ouvert de E, f une application définie sur Ω à valeurs dans F, u une application définie sur Ω à valeurs dans $\mathbb{R}$ et $x \in \Omega$. Si f et u sont est différentiables en x, alors $u \cdot f$ est différentiable en x et l'on a:*

$$(u \cdot f)'(x)(h) = (u'(x)(h)) \cdot f(x) + u(x) \cdot f'(x)(h) \text{ pour tout } h \in E.$$

Enfin, nous admettrons le résultat suivant sur la dérivabilité de l'inverse. Retenez que le résultat qui suit n'est valable que dans le cadre complet car sa démonstration utilise le théorème de Banach 2.3.

Théorème 5.3 *Soit E, F deux espaces de Banach, Ω un ouvert de E, U un ouvert de F, f un homéomorphisme de Ω dans U ($f^{-1} : U \to \Omega$) et $x \in \Omega$. Si f est différentiable en x et si $f'(x)$ est inversible alors f^{-1} est différentiable en $f(x)$ et:*

$$(f^{-1})'(f(x)) = [f'(x)]^{-1}.$$

5.3 Inequalities, Mean-value Theorems

Commencons par des rappels sur le cas réel. Rappelons d'abord le théorème de Rolle:

Théorème 5.4 *Soit $a < b$ deux réels et $f \in C^0([a,b], \mathbb{R})$. Si $f(a) = f(b)$ et f est dérivable sur $]a,b[$ alors il existe $c \in]a,b[$ tel que $f'(c) = 0$.*

Le théorème des accroissements finis (TAF en abrégé) s'énonce alors comme suit

Théorème 5.5 *Soit $a < b$ deux réels et $f \in C^0([a,b],\mathbb{R})$. Si f est dérivable sur $]a,b[$ alors il existe $c \in]a,b[$ tel que:*

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

Etant donnés, un evn E et $(a,b) \in E^2$ on rappelle que:

$$[a,b] = \{ta + (1-t)b,\ t \in [0,1]\} \text{ et }]a,b[= \{ta + (1-t)b,\ t \in]0,1[\}$$

On déduit alors du théorème 5.5 un TAF pour des fonctions de plusieurs variables à valeurs réelles:

Théorème 5.6 *Soit E un evn, Ω un ouvert de E, $a \neq b$ deux points de Ω tels que $[a,b] \subset \Omega$ et f une fonction définie sur Ω à valeurs réelles. Si f est continue sur $[a,b]$ et la dérivée directionnelle $Df(x;b-a)$ existe en tout $x \in]a,b[$ alors il existe $c \in]a,b[$ tel que:*

$$f(b) - f(a) = Df(c;b-a).$$

Preuve:
Définissons pour $t \in [0,1]$, $g(t) := f(a + t(b-a)$, g est continue sur $[0,1]$ et si $t \in]0,1[$ on a:

$$\lim_{h\to 0,\ h\neq 0} \frac{g(t+h)-g(t)}{h} = \lim_{h\to 0,\ h\neq 0} \frac{f(a+(t+h)(b-a)) - f(a+t(b-a))}{h}$$

et comme la dérivée directionnelle $Df(a+t(b-a);b-a)$ existe, nous en déduisons que g est différentiable sur $]a,b[$ avec:

$$g'(t) = Df(a+t(b-a);b-a)$$

on applique alors le TAF à g: il existe $t \in]0,1[$ tel que $g(1)-g(0) = g'(t)$ en posant $c = a + t(b-a)$ on a donc:

$$f(b) - f(a) = Df(c;b-a).$$

□

Pour des fonctions à valeurs dans un evn F, l'inégalité des accroissements finis (IAF) s'énonce comme suit:

Théorème 5.7 *Soit E et F deux $\mathbb{R}$-evn, Ω un ouvert de E, f une fonction définie sur Ω à valeurs dans F, $(a,b) \in \Omega^2$ tels que $a \neq b$, $[a,b] \subset \Omega$ et f est continue sur $[a,b]$. Si la dérivée directionnelle $Df(x;b-a)$ existe en tout $x \in]a,b[$ alors il existe $c \in]a,b[$ tel que:*

$$\|f(b) - f(a)\| \leq \|Df(c;b-a)\|. \tag{5.4}$$

Preuve:
Nous allons nous limiter au cas où F est un espace de Hilbert et admettrons le résultat dans le cas général. Soit $p \in F$ et pour $t \in E$, définissons:

$$g(t) := \langle p, f(a + t(b-a)\rangle$$

alors g (à valeurs réelles) vérifie les hypothèses du théorème 5.5 : il existe $t \in]0,1[$ (noter que t dépend de p) tel que:

$$\langle p, f(b) - f(a)\rangle = g(1) - g(0) = g'(t) = \langle p, Df(c; b-a)\rangle \tag{5.5}$$

(où l'on a posé $c = a + t(b-a) \in]a,b[$). Si $f(b) = f(a)$, le résultat cherché, est évident, on suppose donc $f(b) \neq f(a)$, en appliquant (5.5) à $p = (f(b) - f(a))/\|f(b) - f(a)\|$ et en utilisant l'inégalité de Cauchy-Schwarz il vient alors:

$$\begin{aligned}\left\langle f(b) - f(a), \frac{f(b)-f(a)}{\|f(b)-f(a)\|}\right\rangle &= \|f(b) - f(a)\| = \langle p, Df(c; b-a)\rangle \\ &\leq \|p\|\|Df(c; b-a)\| = \|Df(c; b-a)\|.\end{aligned}$$

□

Si f est Gâteaux-dérivable sur Ω, alors la conclusion du théorème 5.7 implique qu'il existe $c \in]a,b[$ tel que:

$$\|f(b) - f(a)\| \leq \|D_G f(c)(b-a)\| \leq \|D_G f(c)\|\|b-a\|. \tag{5.6}$$

Notons aussi que si f est Gâteaux-dérivable sur Ω et si $[a,b] \subset \Omega$ alors f est continue sur $[a,b]$ (exercice).

On en déduit plusieurs corollaires:

Corollaire 5.2 *Soit E et F deux $\mathbb{R}$-evn, Ω un ouvert de E, f une fonction définie sur Ω à valeurs dans F, $(a,b) \in \Omega^2$ tels que $a \neq b$, $[a,b] \subset \Omega$ et f est continue sur $[a,b]$. Si la dérivée directionnelle $Df(x; b-a)$ existe en tout $x \in]a,b[$ alors:*

$$\|f(b) - f(a)\| \leq \sup_{c \in]a,b[} \|Df(c; b-a)\|. \tag{5.7}$$

(le sup pouvant valoir $+\infty$.)

Si f est Gâteaux-dérivable sur Ω, alors la conclusion du corollaire 5.2 implique:

$$\|f(b) - f(a)\| \leq \sup_{c \in]a,b[} \|D_G f(c)\|\|b-a\|. \tag{5.8}$$

On en déduit donc

Corollaire 5.3 *Soit E et F deux $\mathbb{R}$-evn, Ω un ouvert convexe de E, f une fonction définie sur Ω à valeurs dans F. Si f est Gâteaux-dérivable sur Ω et si $\|D_G f(x)\| \leq k$ pour tout $x \in \Omega$ alors pour tout $(a,b) \in \Omega^2$, on a:*

$$\|f(b) - f(a)\| \leq k\|b - a\|. \tag{5.9}$$

(f est k-lipschitzienne sur Ω).

Une autre inégalité de type IAF bien utile nous est fournie par:

Corollaire 5.4 *Soit E et F deux $\mathbb{R}$-evn, Ω un ouvert de E, f une fonction définie sur Ω à valeurs dans F, $(a,b) \in \Omega^2$ tels que $a \neq b$, $[a,b] \subset \Omega$ et f est continue sur $[a,b]$. Si f est Gâteaux-dérivable sur Ω alors pour tout $z \in \Omega$ on a:*

$$\|f(b) - f(a) - D_G f(z)(b-a)\| \leq \sup_{c \in]a,b[} \|D_G f(c) - D_G f(z)\| \|b - a\|$$

en particulier:

$$\|f(b) - f(a) - D_G f(a)(b-a)\| \leq \sup_{c \in]a,b[} \|D_G f(c) - D_G f(a)\| \|b - a\|$$

Preuve:
Appliquer le corollaire 5.2 à la fonction $x \mapsto f(x) - D_G f(z)(x)$.
□

Nous sommes désormais en mesure de prouver le théorème 5.1 que nous rappelons:

Théorème 5.8 *Si f est Gâteaux-différentiable sur Ω et $D_G f \in C^0(\Omega, L_c(E,F))$ alors f est de classe C^1 sur Ω.*

Preuve:
Si nous montrons que pour tout $x \in \Omega$, et tout $\varepsilon > 0$, il existe $\delta_\varepsilon > 0$ tel que pour tout $h \in E$ tel que $\|h\| \leq \delta_\varepsilon$, on a:

$$\|f(x+h) - f(x) - D_G f(x)(h)\| \leq \delta_\varepsilon \|h\|$$

alors nous aurons montré que f est différentiable et $f'(x) = D_G f(x)$ pour tout x et donc que f est de classe C^1 sur Ω puisque $D_G f \in C^0(\Omega, L_c(E,F))$. Soit $r > 0$ tel que $B(x,r) \subset \Omega$, soit $\delta \in]0,r[$ et $h \in B(x,\delta)$, alors $[x, x+h] \subset B(x,\delta)$ et le corollaire 5.2 donne:

$$\begin{aligned}\|f(x+h) - f(x) - D_G f(x)(h)\| &\leq \sup_{c \in [x,x+h]} \|D_G f(c) - D_G f(x)\| h\| \\ &\leq \sup_{c \in B(x,\delta)} \|D_G f(c) - D_G f(x)\| \|h\|\end{aligned}$$

Comme $D_G f \in C^0(\Omega, L_c(E,F))$, il existe $\delta_\varepsilon \in]0,r[$ tel que:

$$\sup_{c\in B(x,\delta_\varepsilon)} \|D_G f(c) - D_G f(x)\| \leq \varepsilon$$

on en déduit le résultat voulu. □

Nous pouvons également prouver le théorème 5.10 que nous rappelons:

Théorème 5.9 *Soit $E_1, ..., E_p$ et F des evn, $E := E_1 \times ... \times E_p$ et Ω un ouvert de E. Si pour tout $k \in \{1, ..., p\}$ et tout $x \in \Omega$, f admet une dérivée partielle par rapport à la k-ième variable en x et si l'application $x \mapsto \partial_k f(x)$ (définie sur Ω et à valeurs dans $L_c(E_k, F)$) est continue sur Ω alors f est de classe C^1 sur Ω et la formule (5.11) est satisfaite.*

Preuve:
Nous allons démontrer le résultat pour $p = 2$, et laisser le soin au lecteur de traiter le cas général de manière analogue. Soit $x = (x_1, x_2) \in \Omega$, il nous faut montrer que pour tout $\varepsilon > 0$, il existe $\delta_\varepsilon > 0$ tel que pour tout $(h_1, h_2) \in E_1 \times E_2$ tel que $\|h_1\| + \|h_2\| \leq \delta_\varepsilon$, on a:

$$\|f(x_1 + h_1, x_2 + h_2) - f(x) - \partial_1 f(x)(h_1) - \partial_2 f(x)(h_2)\| \leq \delta_\varepsilon(\|h_1\| + \|h_2\|).$$

On commence par remarquer que:

$$f(x_1 + h_1, x_2 + h_2) - f(x) - \partial_1 f(x)(h_1) - \partial_2 f(x)(h_2) = (f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2 + h_2) - \partial_1 f(x)(h_1)) + (f(x_1, x_2 + h_2) - f(x_1, x_2) - \partial_2 f(x)(h_2))$$

Remarquons ensuite que pour tout $y \in \Omega$, f est différentiable en y dans les directions $(h_1, 0)$ et $(0, h_2)$ avec:

$$Df(y; (h_1, 0)) = \partial_1 f(y)(h_1) \text{ et } Df(y; (0, h_2)) = \partial_2 f(y)(h_2)$$

On déduit alors du théorème 5.7, qu'il existe $t_1 \in]0,1[$ tel que

$$\|f(x_1+h_1, x_2+h_2) - f(x_1, x_2+h_2) - \partial_1 f(x)(h_1)\| \leq \|(\partial_1 f(x_1+t_1h_1, x_2+h_2) - \partial_1 f(x))(h_1)\|.$$

De même il existe $t_2 \in]0,1[$ tel que

$$\|f(x_1, x_2+h_2) - f(x_1, x_2) - \partial_2 f(x)(h_2)\| \leq \|(\partial_2 f(x_1, x_2+t_2h_2) - \partial_2 f(x))(h_2))\|.$$

Comme $\partial_1 f$ et $\partial_2 f$ sont continues, il existe δ_ε tel que $B(x, \delta_\varepsilon) \subset \Omega$ et pour tout $y \in B(x, \delta_\varepsilon)$:

$$\max(\|\partial_1 f(y) - \partial_1 f(x)\|, \|\partial_2 f(y) - \partial_2 f(x)\|) \leq \varepsilon$$

Si $\|h_1\| + \|h_2\| \leq \delta_\varepsilon$, les points $(x_1 + t_1h_1, x_2 + h_2)$ et $(x_1, x_2 + t_2h_2)$ appartiennent à $B(x, \delta_\varepsilon)$ et donc:

$$\|f(x_1 + h_1, x_2 + h_2) - f(x) - \partial_1 f(x)(h_1) - \partial_2 f(x)(h_2)\| \leq \varepsilon(\|h_1\| + \|h_2\|).$$

□

5.4 Partial derivatives

Intéressons nous maintenant au cas où l'espace de départ est un produit d'evn: $E = E_1 \times ... \times E_p$ chaque E_k est muni d'une norme N_k et E est muni de la norme produit:

$$N : x = (x_1, ..., x_p) \mapsto N(x) := N_1(x_1) + ... + N_p(x_p).$$

(ou de n'importe quelle norme équivalente).

Dans ce qui suit, on considère Ω un ouvert de E de la forme $\Omega = \Pi_{k=1}^{p} \Omega_k$ avec Ω_k un ouvert de E_k, F un evn et f une application définie sur Ω à valeurs dans F.

Définition 5.9 *Soit $x = (x_1, ..., x_p) \in \Omega$ et $k \in \{1, ..., p\}$, on dit que f admet une dérivée partielle par rapport à la k-ième variable en x ssi la k-ième application partielle:*

$$y \in \Omega_k \mapsto f(x_1, ..., x_{k-1}, y, x_{k+1}, , x_p) \in F$$

est différentiable en x_k. On appelle alors dérivée partielle de f par rapport à la k-ième variable en x et l'on note $\partial_k f(x)$ (ou aussi souvent $\partial_{x_k} f(x)$, $\frac{\partial f}{\partial x_k}(x)$, $D_k f(x)$, $f_{x_k}(x)$, $f'_{x_k}(x)$) la dérivée de cette application partielle en x_k.

Remarque. La notion de dérivée partielle est reliée à celle de dérivée directionnelle. En effet, il est facile de voir que si f admet une dérivée partielle par rapport à la k-ième variable en x alors f est dérivable en x dans la direction $(0, ..., 0, h_k, 0, ..., 0)$ et l'on a:

$$Df(x; (0, ..., 0, h_k, 0, ..., 0)) = \partial_k f(x)(h_k)$$

Noter que la notion de différentiabilité de la définition précédente est celle de Fréchet et bien comprendre que $\partial_k f(x) \in L_c(E_k, F)$. Le lien entre dérivée et et dérivées partielles est donné par des formules connues et utiles:

Proposition 5.5 *Soit $x = (x_1, ..., x_p) \in \Omega$, si f est différentiable en x alors, pour tout $k \in \{1, ..., p\}$, f admet une dérivée partielle par rapport à la k-ième variable en x et on a:*

$$\partial_k f(x)(h_k) = f'(x)(0, ..., h_k, ..., 0), \forall h_k \in E_k, \tag{5.10}$$

et pour tout $h = (h_1, ..., h_p) \in E$:

$$f'(x)(h) = \sum_{k=1}^{p} \partial_k f(x)(h_k). \tag{5.11}$$

Le fait d'être différentiable implique donc d'admettre des dérivées partielles, la réciproque est fausse (cf remarque sur les dérivées directionnelles). En revanche si f admet des dérivées partielles et que celles ci dépendent continûment de x alors f est de classe C^1, c'est l'objet du:

Théorème 5.10 *Si pour tout $k \in \{1, ..., p\}$ et tout $x \in \Omega$, f admet une dérivée partielle par rapport à la k-ième variable en x et si l'application $x \mapsto \partial_k f(x)$ (définie sur Ω et à valeurs dans $L_c(E_k, F)$) est continue sur Ω alors f est de classe C^1 sur Ω et la formule (5.11) est satisfaite.*

Ce théorème très utile en pratique sera démontré au prochain chapitre.

5.5 The finite-dimensional case, the Jacobian matrix

Nous allons maintenant nous intéresser au cas où E et F sont de dimension finie. Dans ce cas, puisque la dérivée de f en x est une application linéaire de E dans F, on peut la représenter sous la forme d'une matrice. Comme d'habitude quand on fait du calcul matriciel, il est utile de représenter les vecteurs sous forme de vecteurs colonnes. On suppose donc dans ce paragraphe que $E = \mathbb{R}^n$, $F = \mathbb{R}^p$, Ω est un ouvert de $E = \mathbb{R}^n$ et f une application définie sur Ω à valeurs dans $F = \mathbb{R}^p$. On note les éléments de $\mathbb{R}^n$ sous la forme:

$$x = \begin{pmatrix} x_1 \\ . \\ . \\ . \\ . \\ x_n \end{pmatrix}$$

et f, sous la forme:

$$f(x) = \begin{pmatrix} f_1(x) \\ . \\ . \\ . \\ . \\ f_p(x) \end{pmatrix}$$

avec f_j définie sur Ω à valeurs réelles est la j-ème composante de f. Nous savons que f est différentiable en $x \in \Omega$ ssi chaque composante de f, f_1, ..., f_p est différentiable en x et dans ce cas chaque f_i admet une dérivée partielle

par rapport à chaque variable x_j. D'après la formule (5.11) on a pour tout $j \in \{1, , p\}$ et tout $h \in \mathbb{R}^n$:

$$f'_j(x)(h) = \sum_{j=1}^{n} \partial_j f_i(x) h_j$$

et donc:

$$f'(x)(h) = \begin{pmatrix} f'_1(x)(h) \\ . \\ . \\ . \\ . \\ f'_p(x)(h) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} \partial_j f_1(x) h_j \\ . \\ . \\ . \\ . \\ \sum_{j=1}^{n} \partial_j f_i(x) h_j \end{pmatrix}$$

ce qui peut se réecrire sous forme matricielle:

$$f'(x)(h) = \begin{pmatrix} \partial_1 f_1(x) & . & . & \partial_n f_1(x) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \partial_1 f_p(x) & . & . & \partial_n f_p(x) \end{pmatrix} \begin{pmatrix} h_1 \\ . \\ . \\ . \\ . \\ h_n \end{pmatrix}$$

Ainsi l'application linéaire $f'(x) \in L(\mathbb{R}^n, \mathbb{R}^p)$ est représentée dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^p$ par la matrice de format $p \times n$ de terme général $\partial_j f_i(x)$ que l'on appelle <u>matrice jacobienne</u> de f en x et que l'on note $Jf(x)$:

$$Jf(x) := \begin{pmatrix} \partial_1 f_1(x) & . & . & \partial_n f_1(x) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \partial_1 f_p(x) & . & . & \partial_n f_p(x) \end{pmatrix}$$

ainsi, sous forme matricielle l'expression de la différentielle de f en x est donnée par:

$$f'(x)(h) = Jf(x)h, \text{ pour tout } h = \begin{pmatrix} h_1 \\ . \\ . \\ . \\ . \\ h_n \end{pmatrix} \in \mathbb{R}^n$$

Notons que les règles de calcul (composition, inverse...) se traduisent matriciellement. Si f est différentiable en x et si g est une application définie sur un voisinage de $f(x)$ dans $\mathbb{R}^p$ à valeurs dans $\mathbb{R}^k$ et si g est différentiable en $f(x)$, alors $g \circ f$ est différentiable en x et l'on a:

$$J(g \circ f)(x) = Jg(f(x)) \times Jf(x).$$

Prenons par exemple $n = p = 2$ et $k = 1$, en appliquant l'expression matricielle précédente de la dérivée d'une composée, il vient:

$$\partial_1(g \circ f)(x_1, x_2) = \partial_1 g(f(x_1, x_2))\partial_1 f_1(x_1, x_2) + \partial_2 g(f(x_1, x_2))\partial_1 f_2(x_1, x_2),$$
$$\partial_2(g \circ f)(x_1, x_2) = \partial_1 g(f(x_1, x_2))\partial_2 f_1(x_1, x_2) + \partial_2 g(f(x_1, x_2))\partial_2 f_2(x_1, x_2).$$

De même, si f est un homéomorphisme d'un voisinage de x sur un voisinage de $f(x)$ et si la matrice $Jf(x)$ est inversible (ce qui implique que $n = p$) alors f^{-1} est différentiable en $f(x)$ et l'on a:

$$Jf^{-1}(f(x)) = [Jf(x)]^{-1}.$$

Dans le cas du but réel c 'est à dire $F = \mathbb{R}$, $Jf(x)$ est le vecteur ligne:

$$Jf(x) := (\partial_1 f_1(x), ..., \partial_n f_1(x))$$

ainsi:

$$f'(x)(h) = \langle \nabla f(x), h \rangle = \sum_{j=1}^{n} \partial_j f(x) h_j = \begin{pmatrix} \partial_1 f(x) \\ . \\ . \\ . \\ . \\ \partial_n f(x) \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ . \\ . \\ . \\ . \\ h_n \end{pmatrix} \text{ pour tout } h \in \mathbb{R}^n$$

en identifiant on a donc l'expression du gradient de f en x:

$$\nabla f(x) = \begin{pmatrix} \partial_1 f(x) \\ . \\ . \\ . \\ . \\ \partial_n f(x) \end{pmatrix}.$$

Dans le cas polaire où $E = \mathbb{R}$ et $F = \mathbb{R}^p$ et en notant $f = (f_1, ..., f_p)$, si f est dérivable en t, alors $f'(t) \in L(\mathbb{R}, \mathbb{R}^p)$:

$$f'(t)(h) = (f_1'(t), ..., f_p'(t))h, \text{ pour tout } h \in \mathbb{R}$$

et on identife simplement $f'(t)$ au vecteur de $\mathbb{R}^p$, $(f_1'(t), ..., f_p'(t))$.

5.6 Calculus

We end this chapter by some classical examples.Given A an $n \times n$ matrix (and denoting by A^T its transpose):

$$f_1(x) = Ax \cdot x,\ \nabla f(x) = Ax + A^T x.$$

Denoting by $\|.\|$ the usual euclidean norm:

$$f_2(x) = \|Ax - b\|^2, \nabla f_2(x) = 2A^T Ax - 2A^T b,$$

and

$$\nabla(\|.\|)(x) = \frac{x}{\|x\|}, \forall x \neq 0.$$

Given $a \in \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^n$

$$f_3(x) = a \cdot g(x),\ \nabla f_3(x) = J_g(x)^T a.$$

Given $f : \mathbb{R}^n \to \mathbb{R}$ and $u_0 \in \mathbb{R}^n$:

$$f_4(t) = f(x + tu_0),\ f_4'(t) := \nabla f(x + tu_0) \cdot u_0.$$

Given $g : \mathbb{R}^n \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}$

$$\nabla(h \circ g)(x) = J_g(x)^T \nabla h(g(x)).$$

Chapter 6

Second-order differential calculus

6.1 Definitions

Soit E et F deux evn, Ω un ouvert de E et f une application définie sur Ω à valeurs dans F, on a la:

Définition 6.1 *Soit $x \in \Omega$, on dit que f est deux fois (Fréchet) dérivable en x s'il existe un ouvert $U \subset \Omega$ tel que $x \in \Omega$ et:*

- *f est dérivable sur U,*
- *l'application $y \in U \mapsto f'(y) \in L_c(E,F)$ est dérivable en x.*

Dans ce cas, la dérivée seconde de f en x est donnée par:

$$f''(x) := (f')'(x) \in L_c(E, L_c(E,F)).$$

La définition précédente peut s'exprimer par:

$$f'(x+h) - f'(x) = f''(x)(h) + o(h) \text{ dans } L_c(E,F)$$

la notation $o(h)$ désignant une fonction telle que:

$$\frac{\|o(h)\|_{L_c(E,F)}}{\|h\|} \to 0 \text{ quand } h \to 0,\ h \neq 0$$

ce qu'on peut aussi écrire $o(h) = \|h\|\varepsilon(h)$ avec $\varepsilon(h) \in L_c(E,F)$ tel que:

$$\|\varepsilon(h)\|_{L_c(E,F)} \to 0 \text{ quand } h \to 0.$$

Grâce aux résultats du paragraphe 2.6.3, on peut identifier $L_c(E, L_c(E, F))$ à $L_{2,c}(E \times E, F)$, ceci permet d'identifier $f''(x)$ à l'application bilinéaire continue que nous notons aussi $f''(x)$:

$$f''(x)(h,k) = (f''(x)(h))(k) \text{ pour tout } (h,k) \in E^2.$$

Nous ferons systématiquement cette identification par la suite.

Si nous fixons $k \in E$ et supposons f deux fois différentiable en x, on a alors:

$$(f'(x+h) - f'(x))(k) = (f''(x)(h))(k) + o(h)(k) = f''(x)(h,k) + o(h)$$

ainsi l'application $f'(.)(k) : y \mapsto f'(y)(k)$ est différentiable en x et l'on a:

$$(f'(.)(k))'(x)(h) = f''(x)(h,k).$$

Ainsi $f''(x)(h,k)$ est la dérivée en x de $f'(.)(k)$ dans la direction h.

Remarquons que si f est deux fois différentiable en x, alors f' est continue x.

Définition 6.2 *On dit que f est de classe C^2 sur Ω (ce que l'on note $f \in C^2(\Omega, F)$) ssi $f \in C^1(\Omega, F)$, f deux fois différentiable en chaque point de Ω et $f'' \in C^0(\Omega, L_{2,c}(E \times E, F))$.*

On peut continuer par récurrence à définir la différentiabilité à des ordres plus élevés, nous en resterons cependant à la dérivée seconde dans ce cours car cela est suffisant en optimisation. Nous renvoyons le lecteur intéressé à Cartan [4] pour les dérivés d'ordre plus élevé.

6.2 Schwarz's symmetry theorem

Une propriété importante des dérivées secondes est leur symétrie, c'est l'objet du théorème de Schwarz. D'abord un résultat préliminaire:

Proposition 6.1 *Si f est deux fois différentiable en x alors la quantité:*

$$\frac{1}{(\|h\| + \|k\|)^2} (f(x+h+k) - f(x+k) - f(x+h) + f(x) - f''(x)(h,k))$$

tend vers 0 quand (h,k) tend vers $(0,0)$, $(h,k) \in E \times E \setminus \{(0,0)\}$.

Preuve:
Par définition pour tout $\varepsilon > 0$, il existe $\delta_\varepsilon > 0$ tel que $B(x, \delta_\varepsilon) \subset \Omega$ et pour tout $v \in B(0, \delta_\varepsilon)$:

$$\|f'(x+v) - f'(x) - f''(x)(v)\| \leq \frac{\varepsilon}{2}\|v\|. \tag{6.1}$$

Soit $r > 0$ tel que $B(x, r) \subset \Omega$ et f soit différentiable sur $B(x, r)$, définissons pour $(h, k) \in E \times E$ tels que $\|h\| + \|k\| \leq r$:

$$\Phi(h, k) := f(x+h+k) - f(x+k) - f(x+h) + f(x) - f''(x)(h, k)$$

Φ est différentiable, $\Phi(h, 0) = 0$ avec l'inégalité des accroissements finis on a alors:

$$\|\Phi(h, k)\| = \|\Phi(h, k) - \Phi(h, 0)\| \leq \sup_{u \in [0,k]} \|\partial_2 \Phi(h, u)\| \|k\|. \tag{6.2}$$

On a par ailleurs:

$$\partial_2 \Phi(h, u) = f'(x+h+u) - f'(x+u) - f''(x)(h)$$

par linéarité de $f''(x)$ on a $f''(x)(h) = f''(x)(h+u) - f''(x)(u)$ et donc:

$$\partial_2 \Phi(h, u) = (f'(x+h+u) - f'(x) - f''(x)(h+u)) - (f'(x+u) - f'(x) - f''(x)(u)).$$

Si $\|h\| + \|k\| \leq \delta_\varepsilon$, on a aussi pour tout $u \in [0, k]$, $\|u\| \leq \|h\| + \|u\| \leq \|h\| + \|k\| \leq \delta_\varepsilon$, et en utilisant (6.1) on a donc:

$$\|\partial_2 \Phi(h, u)\| \leq \frac{\varepsilon}{2} (\|h + u\| + \|u\|) \leq \varepsilon(\|h\| + \|k\|)$$

avec (6.2), il vient donc que si $\|h\| + \|k\| \leq \delta_\varepsilon$, alors:

$$\|\Phi(h, k)\| \leq \varepsilon(\|h\| + \|k\|)\|k\| \leq \varepsilon(\|h\| + \|k\|)^2$$

d'où l'on déduit le résultat voulu. □

Le théorème de Schwarz s'énonce comme suit:

Théorème 6.1 *Si f est deux fois différentiable en x alors l'application bilinéaire continue $f''(x)$ est symétrique:*

$$f''(x)(h, k) = f''(x)(k, h) \text{ pour tout } (h, k) \in E \times E.$$

Preuve:
Pour $(h, k) \in E^2$ assez petits définissons:

$$S(h, k) = (f(x+h+k) - f(x+k) - f(x+h) + f(x))$$

S est symétrique ($S(h,k) = S(k,h)$) et d'après la proposition 6.1, pour tout $\varepsilon > 0$ il existe $\delta_\varepsilon > 0$ tel que si $\|h\| + \|k\| \leq \delta_\varepsilon$ on a:

$$\|S(h,k) - f''(x)(h,k)\| \leq \frac{\varepsilon}{2}(\|h\| + \|k\|)^2.$$

Si $\|h\| + \|k\| \leq \delta_\varepsilon$, en utilisant $S(h,k) = S(k,h)$ on a donc:

$$\begin{aligned}\|f''(x)(h,k) - f''(x)(k,h)\| &\leq \|S(h,k) - f''(x)(h,k)\| + \|S(k,h) - f''(x)(k,h)\| \\ &\leq \varepsilon(\|h\| + \|k\|)^2\end{aligned}$$

Soit $(u,v) \in E \times E$ et $t > 0$ tel que $t(\|u\| + \|v\|) \leq \delta_\varepsilon$, par bilinéarité de $f''(x)$ on a $f''(x)(tu,tv) = t^2 f''(x)(u,v)$, $f''(x)(tv,tu) = t^2 f''(x)(v,u)$, et donc

$$\|f''(x)(tv,tu) - f''(x)(tu,tv)\| = t^2\|f''(x)(v,u) - f''(x)(u,v)\| \leq \varepsilon t^2(\|u\| + \|v\|)^2$$

et donc $\|f''(x)(v,u) - f''(x)(u,v)\| \leq \varepsilon(\|u\| + \|v\|)^2$, $\varepsilon > 0$ étant arbitraire on a donc $f''(x)(v,u) = f''(x)(u,v)$. □

6.3 Second-order partial derivatives

On considère maintenant le cas où E est un produit d'evn: $E = E_1 \times .... \times E_p$. Nous savons que si f est différentiable en $x = (x_1, ..., x_p) \in \Omega$, alors pour tout $k \in \{1, ..., p\}$, f admet une dérivée partielle en x, $\partial_k f(x) \in L_c(E,F)$, par rapport à sa k-ième variable. Nous savons également que pour $h = (h_1, ..., h_p) \in E$ on a:

$$f'(x)(h) = \sum_{k=1}^{p} \partial_k f(x)(h_k) \text{ et } \partial_k f(x)(h_k) = f'(x)(0, ..., 0, h_k, 0, ..., 0).$$

Pour $i \in \{1, ..., p\}$, et $j \in \{1, ..., p\}$, on peut s'intéresser à la dérivée partielle de $\partial_j f$ par rapport à sa i-ème variable, d'où la:

Définition 6.3 *Soit $x \in \Omega$, et $(i,j) \in \{1, ..., p\}^2$, on dit que f admet une dérivée partielle seconde d'indice (i,j) en x si:*

- *il existe un voisinage ouvert U de x dans Ω sur lequel f admet une dérivée partielle par rapport à sa j-ème variable,*
- *l'application $y \in U \mapsto \partial_j f(y)$ admet une dérivée partielle par rapport à sa i-ème variable en x.*

Dans ce cas, la dérivée partielle seconde d'indice (i,j) en x, notée $\partial^2_{ij}f(x)$ est donnée par:

$$\partial^2_{ij}f(x) := \partial_i(\partial_j f)(x).$$

Comme $\partial_j f(x) \in L_c(E_j, F)$, on a $\partial^2_{ij}f(x) \in L_c(E_i, L_c(E_j, F))$. En utilisant à nouveau les résultats du paragraphe 2.6.3, on peut identifier $L_c(E_i, L_c(E_j, F))$ à $L_{2,c}(E_i \times E_j, F)$, ceci permet d'identifier $\partial^2_{ij}f(x)$ à l'application bilinéaire continue que nous notons aussi $\partial^2_{ij}f(x)$:

$$\partial^2_{ij}f(x)(h_i, k_j) = (\partial^2_{ij}f(x)(h_i))(k_j) \text{ pour tout } (h_i, k_j) \in E_i \times E_j.$$

Nous ferons systématiquement cette identification par la suite.

Les relations entre dérivée seconde et dérivées secondes partielles nous sont fournies par la:

Proposition 6.2 *Si f est deux fois dérivable en x alors pour tout $(i,j) \in \{1, ..., p\}^2$, f admet une dérivée partielle seconde d'indice (i,j) en x, $\partial^2_{ij}f(x) \in L_{2,c}(E_i \times E_j, F)$ et pour tout $(h_i, k_j) \in E_i \times E_j$ on a:*

$$\partial^2_{ij}f(x)(h_i, k_j) = f''(x)((0, ...0, h_i, 0, ...0), (0, ...0, k_j, 0, ...0)) \tag{6.3}$$

et les relations de symétrie:

$$\partial^2_{ij}f(x)(h_i, k_j) = \partial^2_{ji}f(x)(k_j, h_i). \tag{6.4}$$

De plus pour tout $h = (h_1, ..., h_p)$ et $k = (k_1, ..., k_p)$ dans E on a:

$$f''(x)(h, k) = \sum_{1 \leq i,j \leq p} \partial^2_{ij}f(x)(h_i, k_j). \tag{6.5}$$

Preuve:
Si f est deux fois dérivable en x, alors f est dérivable sur un voisinage ouvert U de x et donc admet des dérivées partielles sur U, soit $y \in U$ et $k_j \in E_j$, on a:

$$\partial_j f(y)(k_j) = f'(y)(0, ..., 0, k_j, 0, ..., 0)$$

le membre de droite est dérivable en x:

$$(\partial_j f(.)(k_j))'(x)(h) = f''(x)(h)(0, ..., 0, k_j, 0, ..., 0) \text{ pour tout } h \in E$$

il admet donc en particulier une dérivée partielle par rapport à sa i-ème variable. Ainsi f admet une dérivée partielle seconde d'indice (i,j) en x et

pour $h_i \in E_i$, en prenant $h = (0, ..., 0, h_i, 0, ..., 0)$ dans l'identité précédente, il vient:

$$\partial^2_{ij} f(x)(h_i, k_j) = \partial_i(\partial_j f(.)(k_j))(x)(h_i)$$
$$= f''(x)((0, ..., 0, h_i, 0, ..., 0), (0, ..., 0, k_j, 0, ..., 0)).$$

Les relations de symétrie découlent de (6.3) et du théorème de Schwarz. Enfin, (6.5) découle de (6.3) et de la bilinéarité de $f''(x)$.

□

Dans le cas où $E = \mathbb{R}^n$, $F = \mathbb{R}$, en notant $(e_1, ..., e_n)$ la base canonique de $\mathbb{R}^n$, on a:

$\partial^2_{ij} f(x)(h_i, k_j) = f''(x)(h_i e_i, k_j e_j) = h_i \cdot k_j f''(x)(e_i, e_j)$ pour tout $(h_i, k_j) \in \mathbb{R}^2$.

On identifie alors $\partial^2_{ij} f(x)$ au réel $f''(x)(e_i, e_j)$. Les relations de symétrie prennent alors la forme $\partial^2_{ij} f(x) = \partial^2_{ji} f(x)$.
Dans ce cas, $f''(x)$ est une forme bilinéaire symétrique qui avec la formule (6.5) s'écrit:

$$f''(x)(h, k) = \sum_{1 \le i,j \le n} \partial^2_{ij} f(x) h_i \cdot k_j.$$

La matrice hessienne de f en x notée $D^2 f(x)$ (ou parfois aussi $Hf(x)$) est alors par définition la matrice de la forme bilinéaire symétrique $f''(x)$ dans la base canonique, c'est donc la matrice de terme général $f''(x)(e_i, e_j) = \partial^2_{ij} f(x)$. $D^2 f(x)$ est une matrice symétrique et son expression est

$$D^2 f(x) := \begin{pmatrix} \partial^2_{11} f(x) & . & . & \partial^2_{1n} f(x) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \partial^2_{n1} f(x) & . & . & \partial^2_{nn} f(x) \end{pmatrix}.$$

Pour tout $(h, k) \in \mathbb{R}^n \times \mathbb{R}^n$, on a alors:

$$f''(x)(h, k) = \left\langle D^2 f(x)(h), k \right\rangle == \sum_{1 \le i,j \le n} \partial^2_{ij} f(x) h_i \cdot k_j.$$

6.4 Taylor formula

Soit E et F deux evn, Ω un ouvert de E et f une application définie sur Ω à valeurs dans F, et $x \in \Omega$. La Formule de Taylor à l'ordre 2 en x est l'objet du:

Théorème 6.2 *Si f est deux fois dérivable en x, on a:*

$$f(x+h) = f(x) + f'(x)(h) + \frac{1}{2}f''(x)(h,h) + o(\|h\|^2) \tag{6.6}$$

avec:

$$\frac{o(\|h\|^2)}{\|h\|^2} \to 0 \text{ quand } h \to 0,\ h \neq 0.$$

Preuve:
Il s'agit de montrer que pour tout $\varepsilon > 0$, il existe δ_ε tel que si $\|h\| \leq \delta_\varepsilon$ alors

$$\|f(x+h) - f(x) - f'(x)(h) - \frac{1}{2}f''(x)(h,h)\| \leq \varepsilon\|h\|^2. \tag{6.7}$$

Définissons donc

$$R(h) := f(x+h) - f(x) - f'(x)(h) - \frac{1}{2}f''(x)(h,h)$$

R est bien définie et de classe C^1 sur un voisinage de 0 dans E, $R(0) = 0$ et:

$$R'(h) = f'(x+h) - f'(x) - \frac{1}{2}(f''(x)(h,.) + f''(x)(.,h))$$

et comme $f''(x)$ est symétrique, on a:

$$R'(h) = f'(x+h) - f'(x) - f''(x)(h).$$

Donc il existe δ_ε tel que pour tout u tel que $\|u\| \leq \delta_\varepsilon$, on a:

$$\|R'(u)\| \leq \varepsilon\|u\|$$

Si $\|h\| \leq \delta_\varepsilon$, l'inégalité des accroissements finis implique:

$$\begin{aligned}\|R(h)\| = \|R(h) - R(0)\| &\leq \sup_{u\in[0,h]} \|R'(u)\|\|h\| \\ &\leq \varepsilon\|h\|^2.\end{aligned}$$

on a donc établi (6.7). □

Si f est deux fois différentiable au voisinage sur Ω (ou simplement sur un voisinage de x) on a:

Théorème 6.3 *Si $h \in E$ est tel que $[x, x+h] \subset \Omega$ et f est deux fois dérivable en chaque point de Ω alors on a:*

$$\|f(x+h) - f(x) - f'(x)(h) - \frac{1}{2}f''(x)(h,h)\| \leq \sup_{z\in[x,x+h]} \|f''(z) - f''(x)\|\|h\|^2. \tag{6.8}$$

Preuve:
On définit $R(h)$ comme dans la preuve précédente, il s'agit alors de montrer que:

$$\|R(h)\| \leq \sup_{z\in[x,x+h]} \|f''(z) - f''(x)\|\|h\|^2. \tag{6.9}$$

Utilisant $R(0) = 0$ et l'inégalité des accroissements finis, on a:

$$\|R(h)\| \leq \sup_{u\in[0,h]} \|R'(u)\|\|h\|. \tag{6.10}$$

Comme, on a:

$$R'(u) = f'(x+u) - f'(x) - f''(x)(u)$$

si $u \in [0,h]$, en appliquant l'inégalité des accroissements finis du corollaire 5.4 à f' entre x et $x+u$, il vient donc:

$$\begin{aligned}\|R'(u)\| &\leq \sup_{z\in[x,x+u]} \|f''(z) - f''(x)\|\|u\| \\ &\leq \sup_{z\in[x,x+h]} \|f''(z) - f''(x)\|\|h\|\end{aligned}.$$

Reportant la majoration précédente dans (6.10), nous en déduisons exactement (6.9). □

Enfin, terminons par une formule exacte pour f à valeurs dans $\mathbb{R}^p$: la formule de Taylor à l'ordre 2 avec reste intégral:

Théorème 6.4 *Soit f définie sur Ω à valeurs dans $\mathbb{R}^p$, et $(x,h) \in \Omega\times E$ tels que $[x, x+h] \subset \Omega$. Si f est deux fois dérivable en chaque point de $[x, x+h]$ et si l'application $t \mapsto f''(x+th)$ est continue sur $[0,1]$, alors on a:*

$$f(x+h) = f(x) + f'(x)(h) + \int_0^1 (1-t)f''(x+th)(h,h)dt \tag{6.11}$$

ce qui signifie exactement en notant $f_1, ..., f_p$ les composantes de f que l'on a:

$$f_i(x+h) = f_i(x) + f_i'(x)(h) + \int_0^1 (1-t)f_i''(x+th)(h,h)dt \text{ pour } i = 1, ..., p. \tag{6.12}$$

Preuve:
Comme on veut montrer l'identité (6.12) composante par composante, on peut poser $f = f_i$ et supposer que $p = 1$. Posons pour $t \in [0,1]$,

$$g(t) := f(x+th)$$

par hypothèse g est deux fois différentiable avec:

$$g'(t) = f(x+th)(h),\ g''(t) = f''(x+th)(h,h)$$

d'où, avec une intégration par parties:

$$f(x+h) - f(x) = g(1) - g(0) = \int_0^1 g'(t)dt = \int_0^1 (1-t)g''(t)dt - [(1-t)g'(t)]_0^1$$
$$= \int_0^1 (1-t)f''(x+th)(h,h)dt + g'(0) = f'(x)(h) + \int_0^1 (1-t)f''(x+th)(h,h)dt.$$

□

Remarque. L'hypothèse $F = \mathbb{R}^p$ dans le théorème précédent nous a uniquement servi pour la définition de l'intégrale. Si l'on peut généraliser de manière satisfaisante la construction de l'intégrale à des fonctions à valeurs dans F evn de dimension infinie, alors on pourra généraliser la formule de Taylor avec reste intégral à des fonctions à valeurs dans F. Cette généralisation est possible lorsque F est un espace de Banach mais elle dépasse largement le cadre de ce cours.

6.5 Differentiable characterizations of convex functions

La convexité est une notion clé comme nous l'avons déja souligné (théorèmes de projection et de séparation dans un espace de Hilbert) qui joue un rôle fondamental en optimisation. Rappelons la définition basique:

Définition 6.4 *Soit E un $\mathbb{R}$-ev, C une partie convexe de E et f une application définie sur C à valeurs rélles, on dit que f est convexe sur C ssi $\forall (x,y) \in C^2$ et $\forall t \in [0,1]$, on a:*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

On dit que f est strictement convexe sur C ssi $\forall (x,y) \in C^2$ avec $x \neq y$ et $\forall t \in]0,1[$, on a:

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$

Notons que dans la définition précédente, puisque Ω est convexe $tx+(1-t)y \in \Omega\ \forall (x,y) \in C^2$ et $\forall t \in [0,1]$, ainsi $f(tx + (1-t)y)$ est bien défini.

Exercice 6.1 *Soit E, C et f comme précédemment, on définit, l'épigraphe de f par:*

$$\mathrm{Epi}(f) := \{(x, \lambda) \in C \times \mathbb{R} \, : \, f(x) \leq \lambda\}$$

Montrer alors que f est convexe sur C ssi $\mathrm{Epi}(f)$ *est une partie convexe de $E \times \mathbb{R}$.*

Une première application de la notion en optimisation est fournie par:

Proposition 6.3 *Soit E un $\mathbb{R}$-ev, C une partie convexe de E et f une fonction strictement convexe sur C, alors il existe au plus un point de C en lequel f atteint son minimum.*

Preuve:
Suposons au contraire qu'il existe x_1 et x_2 distincts dans C tels que:

$$f(x_1) = f(x_2) \leq f(x) \; \forall x \in C$$

par convexité de C, $\frac{1}{2}(x_1 + x_2) \in C$ et par stricte convexité de f, on aurait alors:

$$f(x_1) = \frac{f(x_1) + f(x_2)}{2} \leq f(\frac{x_1 + x_2}{2}) < \frac{f(x_1) + f(x_2)}{2}.$$

□

Dans le cadre différentiable, on a la caractérisation suivante:

Proposition 6.4 *Soit E un $\mathbb{R}$-evn, Ω un ouvert convexe de E et $f : \Omega \to \mathbb{R}$ une application différentiable sur Ω. On a les équivalences:*

1. *f est convexe sur Ω,*
2. *pour tout $(x, y) \in \Omega^2$, on a:*

$$f(x) - f(y) \geq f'(y)(x - y). \tag{6.13}$$

Preuve:
Supposons que f soit convexe, soit $(x, y) \in \Omega^2$ pour $t \in \times[0, 1]$ définissons

$$g(t) := f(tx + (1 - t)y) - tf(x) - (1 - t)f(y)$$

par convexité $g(t) \leq 0$ pour tout $t \in [0, 1]$ et $g(0) = 0$ on a donc:

$$\frac{g(t) - g(0)}{t} \leq 0 \text{ pour tout } t \in]0, 1]$$

en passant à la limite on obtient:

$$0 \geq g'(0^+) = f'(y)(x-y) - f(x) + f(y)$$

Réciproquement supposons que (6.13) soit satisfaite pour tout $(x, y) \in \Omega^2$. Soit $(z_1, z_2) \in \Omega^2$ on a:

$$f(z_1) - f(z_2) \geq f'(z_2)(z_1 - z_2) \text{ et } f(z_2) - f(z_1) \geq f'(z_1)(z_2 - z_1)$$

en sommant il vient:

$$(f'(z_1) - f'(z_2))(z_1 - z_2) \geq 0. \tag{6.14}$$

On définit, pour $t \in [0, 1]$, $g(t)$ comme précédemment. Pour établir la convexité de f il faut montrer que $g \leq 0$ sur $[0, 1]$. On a $g(0) = g(1) = 0$, g est dérivable sur $]0, 1[$:

$$g'(t) = f'(tx+(1-t)y)(x-y) - f(x) + f(y) = f'(y+t(x-y))(x-y) - f(x) + f(y)$$

Soit $1 > t > s > 0$ on a alors, en appliquant (6.14) à $z_1 = y + t(x - y)$ et $z_2 = y + s(x - y)$ $(z_1 - z_2 = (t - s)(x - y))$

$$g'(t) - g'(s) = (f'(y + t(x - y)) - f'(y + s(x - y)))\,(x - y) \geq 0$$

d'où l'on déduit que g' est croissante sur $]0, 1[$. Soit $t \in]0, 1[$, on déduit de la formule des accroissements finis qu'il existe $\theta \in]0, t[$ et $\theta' \in]t, 1[$ tels que:

$$g(t) = g(0) + g'(\theta)t = g'(\theta)t \text{ et } g(1) - g(t) = -g(t) = g'(\theta')(1 - t)$$

Comme $\theta' > t > \theta$ on a $g'(\theta') \geq g'(\theta)$ et donc:

$$-\frac{g(t)}{1-t} \geq \frac{g(t)}{t}$$

ce qui implique bien que $g(t) \leq 0$.

□

La caractérisation différentielle (6.13) de la convexité est importante et exprime géométriquement le fait que f est convexe ssi son graphe se situe au dessus de tous ses plans tangents. Lorsque E est un Hilbert, (6.13) se traduit par

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle\,, \forall (x, y) \in \Omega \times \Omega.$$

Une application de la caractérisation différentielle (6.13) de la convexité en optimisation est:

Proposition 6.5 *Soit E un $\mathbb{R}$-evn, Ω un ouvert convexe de E et $f : \Omega \to \mathbb{R}$ une application convexe différentiable sur Ω, et $x^* \in \Omega$, on a les équivalences entre*

$$f(x^*) \leq f(x),\ \forall x \in \Omega \text{ et } f'(x^*) = 0.$$

Preuve:
Supposons que f atteigne son minimum sur Ω en x^*, alors pour $h \in E$ et $t > 0$ assez petit pour que $x^* + th \in \Omega$ on a:

$$\frac{1}{t}(f(x^* + th) - f(x^*)) \geq 0$$

comme f est différentiable en x^*, en passant à la limite on obtient $f'(x^*)(h) \geq 0$, h étant arbitraire on a aussi $f'(x^*)(-h) \geq 0$ et donc $f'(x^*) = 0$.

Supposons que $f'(x^*) = 0$, alors pour tout $x \in \Omega$, en utilisant (6.13), on a:

$$f(x) \geq f(x^*) + f'(x^*)(x - x^*) = f(x^*).$$

□

Dans le cadre deux fois différentiable, on a la caractérisation:

Proposition 6.6 *Soit E un $\mathbb{R}$-evn, Ω un ouvert convexe de E et $f : \Omega \to \mathbb{R}$ une application deux fois différentiable sur Ω. On a les équivalences:*

1. *f est convexe sur Ω,*
2. *pour tout $(x, h) \in \Omega \times E$, on a:*

$$f''(x)(h, h) \geq 0. \tag{6.15}$$

($f''(x)$ est une forme quadratique semi-définie positive)

Preuve:
Supposons d'abord que f soit convexe. Soit $x \in \Omega$ et $h \in E$ tel que $x + h \in \Omega$ (ce qui implique par convexité de Ω que $x + th \in \Omega$, pour tout $t \in [0, 1]$), avec (6.13), on a pour tout $t \in [0, 1]$:

$$f(x + th) \geq f(x) + tf'(x)(h)$$

or la formule de Taylor à l'ordre 2 en x donne:

$$f(x + th) = tf'(x)(h) + t^2 f''(x)(h, h) + o(t^2)$$

et donc:

$$t^2 f''(x)(h, h) + o(t^2) \geq 0$$

divisant par t^2 et faisant tendre t vers 0 il vient bien

$$f''(x)(h,h) \geq 0.$$

Réciproquement supposons que (6.15) soit satisfaite pour tout $(x,h) \in \Omega \times E$. Soit $(x,y) \in \Omega^2$, comme dans la preuve de la proposition 6.4, pour $t \in \times[0,1]$, on définit

$$g(t) := f(tx + (1-t)y) - tf(x) - (1-t)f(y)$$

et il s'agit de montrer que $g(t) \leq 0$ pour tout $t \in [0,1]$. On remarque que g est deux fois différentiable sur $]0,1[$ avec:

$$g'(t) = f'(y+t(x-y))(x-y) - f(x) + f(y),\ g''(t) = f''(y+t(x-y))(x-y, x-y)$$

ainsi (6.15) implique $g'' \geq 0$ et donc g' est croissant sur $[0,1]$ on achève alors la preuve exactement comme pour la proposition 6.4. □

Exercice 6.2 *Soit $f \in C^2(\mathbb{R}^2, \mathbb{R})$ montrer que f est convexe sur $\mathbb{R}^2$ ssi $\forall x \in \mathbb{R}^2$, la Hessienne de f en x a une trace et un déterminant positif.*

Exercice 6.3 *Soit $f \in C^1(\mathbb{R}^d, \mathbb{R})$, montrer que f est quasiconvexe ssi pour tout x et y dans $\mathbb{R}^d$ si $f(y) \leq f(x)$ alors $f'(x)(y-x) \leq 0$.*

Exercice 6.4 *Soit $f \in C^2(\mathbb{R}^d, \mathbb{R})$*

1. *Montrer que si f est quasiconvexe alors pour tout $x \in \mathbb{R}^d$ et tout h orthogonal à $\nabla f(x)$ on a $f''(x)(h,h) \geq 0$ (i.e. $f''(x)$ est positive sur l'orthogonal de $\nabla f(x)$)*

2. *Montrer que si pour tout $x \in \mathbb{R}^d$ et tout h de norme 1 orthogonal à $\nabla f(x)$ on a $f''(x)(h,h) > 0$ alors f est quasi-convexe.*

3. *Montrer par des exemples en dimension 1 que la condition de la question 1 n'est pas suffisante et que la condition de la question 2 n'est pas nécessaire.*

Chapter 7

Local invertibility and implicit functions theorems

7.1 Local invertibility

Le théorème de l'inversion locale énoncé ci-dessous exprime que si la différentielle $f'(a)$ de l'application f au point a est inversible (en tant qu'application linéaire) alors (l'application non linéaire) f est inversible sur un voisinage de a. C'est précisément le but du calcul différentiel que de déduire d'une propriété de $f'(a)$ une information sur le comportement de f au voisinage de a.

Théorème 7.1 *Soit E et F deux espaces de Banach, Ω un ouvert de E, $f \in C^1(\Omega, F)$ et $a \in \Omega$. Si $f'(a)$ est inversible alors il existe deux voisinages ouverts U et V respectivement de a et $f(a)$ tels que la restriction $f : U \to V$ soit un difféomorphisme de classe C^1.*

Preuve:

Etape 1 : réduction

Posons pour tout $x \in \Omega - a$:

$$g(x) := [f'(a)]^{-1}(f(x+a) - f(a))$$

g est une application de classe C^1 de l'ouvert $\Omega - a \subset E$ dans E de plus $g(0) = 0$ et $g'(0) = \mathrm{id}$. Comme g est obtenue comme composée de f et d'opérations affines inversibles (et indéfiniment différentiables !) il suffit de montrer le résultat pour g.

Pour $x \in \Omega - a$, posons:

$$\Phi(x) := x - g(x).$$

Comme $\Phi'(0) = 0$, il existe $r > 0$ tel que pour tout $x \in \overline{B}(0, r)$ on a[1]:

$$\|\Phi'(x)\| \leq 1/2. \tag{7.1}$$

Etape 2 : g est une bijection de $\overline{B}(0, r)$ dans $\overline{B}(0, r/2)$

Soit $y \in \overline{B}(0, r/2)$, pour tout $x \in \overline{B}(0, r)$ définissons:

$$\Phi_y(x) := \Phi(x) + y = x - g(x) + y.$$

Remarquons alors que:

$$g(x) = y \Leftrightarrow \Phi_y(x) = x. \tag{7.2}$$

Pour x_1 et x_2 dans $\overline{B}(0, r)$ et $y \in \overline{B}(0, r/2)$, en utilisant (7.1) et le corollaire 5.2, on a:

$$\|\Phi_y(x_1) - \Phi_y(x_2)\| = \|\Phi(x_1) - \Phi(x_2)\| \leq \sup_{z \in [x_1, x_2]} \|\Phi'(z)\| \|x_1 - x_2\| \leq \frac{1}{2}\|x_1 - x_2\|$$

en particulier puisque $\Phi_y(0) = y$ on a pour tout $x \in \overline{B}(0, r)$:

$$\|\Phi_y(x)\| \leq \|y\| + \frac{1}{2}\|x\| \leq r/2 + r/2 = r.$$

Ce qui précède montre que, pour tout $y \in \overline{B}(0, r/2)$, Φ_y est une contraction de $\overline{B}(0, r)$. Il découle du théorème du point fixe pour les contractions que ϕ_y admet un unique point fixe dans $\overline{B}(0, r)$. Avec (7.2), nous en déduisons donc que pour tout $y \in \overline{B}(0, r/2)$, il existe un unique $x \in \overline{B}(0, r)$ tel que $g(x) = y$. Donc g est une bijection de $\overline{B}(0, r)$ dans $\overline{B}(0, r/2)$.

Etape 3 : g^{-1} est 2-Lipschitzienne sur $\overline{B}(0, r/2)$

Soit $(y_1, y_2) \in \overline{B}(0, r/2)^2$, $x_1 := g^{-1}(y_1)$ et $x_2 := g^{-1}(y_2)$. Avec les notations de l'étape 2, on a: $x_1 := \Phi_{y_1}(x_1)$ et $x_2 := \Phi_{y_2}(x_2)$. On a alors:

$$\begin{aligned}\|g^{-1}(y_1) - g^{-1}(y_2)\| = \|x_1 - x_2\| = \|\Phi_{y_1}(x_1) - \Phi_{y_2}(x_2)\| = \|y_1 - y_2 + \Phi(x_1) - \Phi(x_2)\| \\ \leq \|y_1 - y_2\| + \frac{1}{2}\|x_1 - x_2\| = \|y_1 - y_2\| + \frac{1}{2}\|g^{-1}(y_1) - g^{-1}(y_2)\|.\end{aligned}$$

[1]Ici la norme $\|\Phi'(x)\|$ désigne naturellement $\|\Phi'(x)\|_{L_c(E)}$.

Comme voulu, on a donc bien:

$$\|g^{-1}(y_1) - g^{-1}(y_2)\| \leq 2\|y_1 - y_2\|.$$

Etape 4 : g^{-1} est différentiable sur $B(0, r/2)$

Soit $y \in B(0, r/2)$ et $x := g^{-1}(y) \in \overline{B}(0, r)$, tout d'abord rappelons qu'avec (7.1), on a:

$$\|\Phi'(x)\| = \|\mathrm{id} - g'(x)\| \leq \frac{1}{2} < 1,$$

ceci impliquant en particulier que $g'(x)$ est inversible.[2]

Nous allons montrer que g^{-1} est différentiable en y et plus précisément que $(g^{-1})'(y) = [g'(x)]^{-1} = [g'(g^{-1}(y))]^{-1}$. Tout d'abord notons qu'il découle du théorème de Banach 2.3 que $[g'(x)]^{-1}$ est continu. Il s'agit donc de montrer que pour $\varepsilon > 0$ il existe $\delta_\varepsilon > 0$ tel que pour tout $k \in B(0, \delta_\varepsilon)$ tel que $y + k \in B(0, r/2)$ on a:

$$\|g^{-1}(y + k) - g^{-1}(y) - [g'(x)]^{-1}k\| \leq \varepsilon\|k\|. \tag{7.3}$$

Soit $k \in E$ assez petit pour que $y + k \in B(0, r/2)$ et posons $x_k := g^{-1}(y + k)$. D'après l'étape précédente, on a:

$$\|x_k - x\| = \|g^{-1}(y + k) - g^{-1}(y)\| \leq 2\|k\|. \tag{7.4}$$

Par ailleurs, puisque $k = g(x_k) - g(x) = g'(x)(x_k - x) + o(\|x_k - x\|) = g'(x)(x_k - x) + \Delta_k$, il existe $\delta > 0$ tel que:

$$\|x_k - x\| \leq \delta \Rightarrow \ \|\Delta_k\| := \|g(x_k) - g(x) - g'(x)(x_k - x)\| \leq \tfrac{\varepsilon\|x_k - x\|}{2\|[g'(x)]^{-1}\|}. \tag{7.5}$$

Si $\|k\| \leq \delta/2$ alors (7.4) entraine que $\|x_k - x\| \leq \delta$, en utilisant (7.5) et à nouveau (7.4), il vient donc:

$$\begin{aligned}\|g^{-1}(y + k) - g^{-1}(y) - [g'(x)]^{-1}k\| &= \|x_k - x - [g'(x)]^{-1}(g'(x)(x_k - x) + \Delta_k)\| \\ &= \|[g'(x)]^{-1}\Delta_k\| \leq \|[g'(x)]^{-1}\|\|\Delta_k\| \leq \|[g'(x)]^{-1}\|\frac{\varepsilon\|x_k - x\|}{2\|[g'(x)]^{-1}\|} \leq \varepsilon\|k\|.\end{aligned}$$

on a donc établi (7.3) ce qui achève la preuve. □

One can immediately deduce from the inverse function theorem the following result which is of global nature in the case where $f'(a)$ is invertible for every $a \in \Omega$:

[2]Rappelons ici que si $u \in L_c(E)$ vérifie $\|\mathrm{id} - u\| < 1$ alors u est inversible d'inverse continue (voir le poly d'exercices).

Theorem 7.1 *Let E and F be two Banach spaces, Ω be an open set of E and $f \in C^1(\Omega, F)$. If $f'(a)$ is invertible for every $a \in \Omega$ then $f(\Omega)$ is open in F. If, in addition, f is injective then f is a C^1-diffeomorphism from Ω to $f(\Omega)$.*

7.2 Implicit functions

Le théorème des fonctions implicites permet localement de passer d'une condition implicite entre les variables x et y du type $f(x, y) = c$ à une relation explicite du type $y = g(x)$.

Théorème 7.2 *Soit E, F et G trois espaces de Banach, A et B deux ouverts respectivement de E et F, $(a, b) \in A \times B$, $f \in C^1(A \times B, G)$ et $c := f(a, b)$. Si $\partial_2 f(a, b)$ est inversible (dans $L_c(F, G)$) alors il existe U un voisinage ouvert de a, V un voisinage ouvert de b et $g \in C^1(U, V)$ tel que:*

$$\{(x, y) \in U \times V \, : \, f(x, y) = c\} = \{(x, g(x)) \, : \, x \in U\}.$$

Ce qui implique en particulier : $g(a) = b$ et $f(x, g(x)) = c \; \forall x \in U$.

Preuve:
Soit $\Phi(x, y) := (x, f(x, y))$, $\forall (x, y) \in A \times B$; on a alors $\Phi \in C^1(A \times B, E \times G)$. Soit $(h, k) \in E \times F$, $\Phi'(a, b)(h, k) = (h, \partial_1 f(a, b)h + \partial_2 f(a, b)k)$. Pour $(u, v) \in E \times G$, comme $\partial_2 f(a, b)$ est inversible, on a:

$$\Phi'(a, b)(h, k) = (u, v) \Leftrightarrow (h, k) = (u, [\partial_2 f(a, b)]^{-1}(v - \partial_1 f(a, b)u)).$$

Ainsi $\Phi'(a, b)$ est inversible, avec le théorème de l'inversion locale, nous en déduisons qu'il existe M voisinage ouvert de (a, b) dans $A \times B$ et N voisinage ouvert de $\Phi(a, b) = (a, c)$ tel que Φ réalise un difféomorphisme de classe C^1 de M sur N. Sans perte de généralité, on peut supposer que $M = M_a \times M_b$, avec M_a (resp. M_b) voisinage ouvert de a dans A (resp. de b dans B). Notons alors $\Phi^{-1} : N \to M_a \times M_b$ sous la forme $\Phi^{-1}(x, z) =: (u(x, z), v(x, z)) = (x, v(x, z))$. Posons alors

$$U := \{x \in M_a \, : \, (x, c) \in N \text{ et } v(x, c) \in M_b\}, \; V := M_b.$$

Par construction U est un voisinage ouvert de a et V un voisinage ouvert de b. Pour tout $x \in U$, définissons $g(x) := v(x, c)$ on a alors $g \in C^1(U, V)$. Soit $(x, y) \in U \times V$, on alors $(x, c) \in N$, $g(x) = v(x, c) \in V$ et donc:

$$\begin{aligned} f(x, y) = c \Leftrightarrow \Phi(x, y) = (x, c) \Leftrightarrow (x, y) = \Phi^{-1}(x, c) \\ \Leftrightarrow (x, y) = (x, v(x, c)) \Leftrightarrow (x, y) = (x, g(x)). \end{aligned}$$

□

Remarque. En dérivant l'identité $f(x, g(x)) = c$ on a:

$$\partial_1 f(x, g(x)) + \partial_2 f(x, g(x)) \circ g'(x) = 0$$

et comme, au voisinage de a, $\partial_2 f(x, g(x))$ est inversible, on a:

$$g'(x) = -[\partial_2 f(x, g(x))]^{-1} \circ (\partial_1 f(x, g(x))).$$

Remarque. On a utilisé le théorème de l'inversion locale pour établir celui des fonctions implicites. On peut montrer (le faire à titre d'exercice) que ces deux énoncés sont en fait équivalents.

Remarque. L'hypothèse $\partial_2 f(a, b)$ inversible du théorème ne peut être affaiblie. Pour s'en persuader, le lecteur pourra considérer le cas du cercle trigonométrique $S^1 := \{(x, y) \in \mathbb{R}^2 : f(x, y) := x^2 + y^2 = 1\}$. On a $\partial_2 f(1, 0) = 0$ et il n'existe pas de voisinage de $(1, 0)$ sur lequel S^1 se représente localement comme un graphe $x \mapsto g(x)$.

Part III

Static Optimization

Chapter 8

Generalities and unconstrained optimization

Soit E un ensemble et f une fonction définie sur E à valeurs réelles. Résoudre le problème de minimisation:

$$\inf_{x\in E} f(x) \tag{8.1}$$

c'est trouver $x^* \in E$ tel que $f(x^*) \leq f(x)$ pour tout $x \in E$. Un tel x^* (s'il existe) est alors une solution de (8.1). La quantité $\inf_{x\in E} f(x)$ (valant éventuellement $-\infty$, voir plus bas) est appelée valeur du problème (8.1). Cette valeur est différente de $-\infty$ si et seulement si f est minorée sur E, évidemment seul ce cas présente de l'intérêt.

A ce stade, un petit rappel s'impose sur les bornes inférieures de parties de $\mathbb{R}$, en effet la valeur du problème (8.1) est par définition la borne inférieure du sous-ensemble de $\mathbb{R}$, $f(E) := \{f(x) : x \in E\}$. Si A est une partie non vide de $\mathbb{R}$, sa borne inférieure, $\inf A$ est par définition son minorant maximal dans $\mathbb{R} \cup \{-\infty\}$ ce qui signifie d'une part:

$$a \geq \inf A,\ \forall a \in A$$

et d'autre part

$$\forall b > \inf A, \text{ il existe } a \in A \text{ tel que } a \leq b.$$

Si A n'est pas minorée A alors l'ensemble de ses minorants est réduit à $\{-\infty\}$ et donc $\inf A = -\infty$. Enfin, on étend la borne inférieure à l'ensemble vide en posant $\inf \emptyset = +\infty$.

Si E est un ensemble non vide et f une fonction définie sur E à valeurs réelles et minorée, alors la valeur $\alpha := \inf_{x\in E} f(x)$ est un nombre réel, ce réel α est caractérisé par:

$$f(x) \geq \alpha,\ \forall x \in E \text{ et } \forall \varepsilon > 0,\ \exists x_\varepsilon \text{ tel que } f(x_\varepsilon) \leq \alpha + \varepsilon\ .$$

En spécifiant $\varepsilon = \varepsilon_n$ avec $(\varepsilon_n)_n$ une suite de réels strictement positifs tendant vers 0, nous en déduisons qu'il existe $x_n \in E$ tel que

$$f(x_n) \leq \inf_{x \in E} f(x) + \varepsilon_n.$$

Comme $f(x_n) \geq \inf_{x \in E} f(x)$, on a donc:

$$\lim_n f(x_n) = \inf_{x \in E} f(x). \tag{8.2}$$

Toute suite $(x_n)_n \in E^{\mathbb{N}}$ vérifiant (8.2) est appelée suite minimisante du problème (8.1). Lorsque f n'est pas minorée sur E, alors $\alpha := \inf_{x \in E} f(x) = -\infty$ et une suite minimisante est simplement une suite $(x_n)_n \in E^{\mathbb{N}}$ vérifiant:

$$\lim_n f(x_n) = \inf_{x \in E} f(x) = -\infty. \tag{8.3}$$

Notons bien que sans aucune hypothèse supplémentaire, il existe toujours des suites minimisantes du problème (8.1).

Lorsque l'ensemble E est fini, (8.1) relève des méthodes de l'optimisation combinatoire qui ne sont pas l'objet de ce cours. Nous considèrerons ici le cas de l'optimisation continue (E est un "continuum", par exemple un ouvert d'un $\mathbb{R}$-evn) ce qui nous permettra d'utiliser les résultats de topologie et de calcul différentiel vus précédemment. En particulier, E sera toujours muni d'une structure métrique. Dans le cadre métrique, on distingue naturellement les notions locales et globales de solution:

Définition 8.1 *Soit (E, d) un espace métrique et f une fonction définie sur E à valeurs réelles.*

1. *On dit que x^* est une solution globale de (8.1) ou un point de minimum global de f sur E ssi $f(x^*) \leq f(x)$ pour tout $x \in E$.*

2. *On dit que x^* est une solution locale de (8.1) ou un point de minimum local de f sur E ss'il existe $r > 0$ tel que $f(x^*) \leq f(x)$ pour tout $x \in E$ tel que $d(x, x^*) < r$.*

3. *On dit que x^* est un point de minimum global strict de f sur E ssi $f(x^*) < f(x)$ pour tout $x \in E \setminus \{x^*\}$.*

4. *On dit que x^* un point de minimum local strict de f sur E ss'il existe $r > 0$ tel que $f(x^*) < f(x)$ pour tout $x \in E$ tel que $d(x, x^*) < r$ et $x \neq x^*$.*

Enfin, on a écrit (8.1) sous la forme d'un problème de minimisation, ce qui englobe aussi les problèmes de maximisation, en effet maximiser une fonction revient à minimiser son opposé.

8.1 Existence theorems

La première question à se poser dans un problème d'optimisation est : existe-t-il (au moins) une solution? Nous allons rappeler quelques critères simples qui assurent l'existence d'une telle solution. Rappelons d'abord le théorème classique de Weierstrass (voir chapitre 1):

Théorème 8.1 *Soit (E, d) un espace métrique compact et $f \in C^0(E, \mathbb{R})$ alors il existe $x^* \in E$ tel que:*

$$f(x^*) = \inf \{f(x),\, x \in E\}\,.$$

Preuve:
Nous avons déja établi ce résultat, on se propose ici d'en donner une preuve (légèrement) différente reposant sur la notion de suite minimisante: ce type de preuve est le point de départ de ce qu'on appelle la méthode directe du calcul des variations. Soit donc $(x_n)_n \in E^{\mathbb{N}}$ une suite minimisante de f sur E:

$$\lim_n f(x_n) = \inf_{x \in E} f(x). \tag{8.4}$$

Comme E est compact, on peut, quitte à extraire une sous-suite que nous continuerons à noter $(x_n)_n$, supposer que $(x_n)_n$ converge vers un élément $x^* \in E$. Comme f est continue, $f(x_n)$ converge vers $f(x^*)$, en utilisant (8.4), il vient donc:

$$f(x^*) = \inf \{f(x),\, x \in E\}\,.$$

□

En pratique, les hypothèses de continuité et surtout celle de compacité sont assez restrictives et comme nous allons le voir, peuvent être (un peu) affaiblies.

Intuitivement, comme on s'intéresse ici à minimiser f, la situation où f n'a des sauts que "vers le bas" n'est pas gênante (considérer par exemple $f(0) = 0$, $f(x) = 1$ pour $x \in \mathbb{R} \setminus \{0\}$). Cette intuition conduit naturellement à la notion de semi-continuité inférieure:

Définition 8.2 *Soit (E, d) un espace métrique, f une application définie sur E à valeurs réelles et $x_0 \in E$, on dit que:*

1. *f est semi-continue inférieurement (s.c.i. en abrégé) en x_0 ssi:*

$$\forall \varepsilon > 0,\, \exists r > 0 \ tq\ x \in E,\, d(x, x_0) \leq r \Rightarrow f(x) \geq f(x_0) - \varepsilon. \tag{8.5}$$

2. *f est semi-continue inférieurement (s.c.i. en abrégé) sur E ssi f est s.c.i. en chaque point de E.*

Rappelons qu'étant donnée une suite de réels $(\alpha_n)_n$, on note $\liminf_n \alpha_n$ la plus petite valeur d'adhérence de $(\alpha_n)_n$ dans $\mathbb{R} \cup \{-\infty, +\infty\}$. On a alors la caractérisation suivante de la semi-continuité inférieure en un point:

Proposition 8.1 *Soit (E, d) un espace métrique, f une application définie sur E à valeurs réelles et $x_0 \in E$. Les assertions suivantes sont équivalentes:*

1. *f est semi-continue inférieurement en x_0,*
2. *pour toute suite $(x_n)_n \in E^{\mathbb{N}}$ convergeant vers x_0 on a:*

$$\liminf f(x_n) \geq f(x_0), \tag{8.6}$$

Preuve:
1. $\Rightarrow$ 2. : soit $(x_n)_n \in E^{\mathbb{N}}$ convergeant vers x_0 et $(x_{\varphi(n)})_n$ une sous-suite telle que

$$\lim_n f(x_{\varphi(n)}) = \liminf_n f(x_n)$$

Supposons par l'absurde que $\liminf_n f(x_n) < f(x_0)$ et soit $\varepsilon > 0$ tel que

$$\lim_n f(x_{\varphi(n)}) = \liminf_n f(x_n) \leq f(x_0) - \varepsilon. \tag{8.7}$$

Puisque f est s.c.i. en x_0 il existe $r > 0$ tel que pour tout $x \in B(x_0, r)$ on a : $f(x) \geq f(x_0) - \varepsilon/2$. Pour n assez grand, on a $x_{\varphi(n)} \in B(x_0, r)$ et donc : $f(x_{\varphi(n)}) \geq f(x_0) - \varepsilon/2$ en passant à la limite on a donc:

$$\lim_n f(x_{\varphi(n)}) \geq f(x_0) - \varepsilon/2$$

ce qui contredit (8.7)

2. $\Rightarrow$ 1. : Supposons que f ne soit pas s.c.i. en x_0 alors il existe ε tel que pour tout $r > 0$, il existe $x \in B(x_0, r)$ tel que $f(x) < f(x_0) - \varepsilon$. En prenant $r = 1/n$, il existe donc $x_n \in B(x_0, r)$ tel que $f(x_n) < f(x_0) - \varepsilon$, on a alors:

$$\lim_n x_n = x_0 \text{ et } \liminf f(x_n) \leq f(x_0) - \varepsilon$$

ce qui contredit 2..

□

Etant donnée f une application définie sur E à valeurs réelles, on définit son épigraphe par:

$$\mathrm{Epi}(f) := \{(x, t) \in E \times \mathbb{R} : t \geq f(x)\}$$

On a alors la caractérisation suivante de la semi-continuité inférieure:

Proposition 8.2 *Soit (E, d) un espace métrique, f une application définie sur E à valeurs réelles alors f est semi-continue inférieurement sur E ssi* $\mathrm{Epi}(f)$ *est fermé dans $E \times \mathbb{R}$.*

Preuve:
Supposons d'abord f semi-continue inférieurement sur E. Soit (x_n, t_n) une suite d'éléments de $\mathrm{Epi}(f)$ convergeant vers (x_0, t_0) dans $E \times \mathbb{R}$. Pour tout n on a $f(x_n) \leq t_n$ et comme f est s.c.i. en x_0 on a:

$$f(x_0) \leq \mathrm{liminf} f(x_n) \leq \mathrm{liminf} t_n = t_0$$

ainsi $(x_0, t_0) \in \mathrm{Epi}(f)$.

Supposons maintenant que $\mathrm{Epi}(f)$ est fermé. Soit $x_0 \in E$ et $(x_n)_n \in E^{\mathbb{N}}$ convergeant vers x_0, soit $(x_{\varphi(n)})$ une sous-suite telle que:

$$\lim_n f(x_{\varphi(n)}) = \mathrm{liminf}_n f(x_n)$$

Pour tout n, $(x_{\varphi(n)}, f(x_{\varphi(n)}) \in \mathrm{Epi}(f))$ et

$$\lim_n (x_{\varphi(n)}, f(x_{\varphi(n)}) = (x_0, \mathrm{liminf}_n f(x_n)).$$

Comme $\mathrm{Epi}(f))$ est fermé, on en déduit que $(x_0, \mathrm{liminf}_n f(x_n)) \in \mathrm{Epi}(f))$ ce qui signifie exactement:

$$\mathrm{liminf}_n f(x_n)) \geq f(x_0).$$

□

Le théorème de Weierstrass s'étend aux fonctions qui sont seulement s.c.i:

Théorème 8.2 *Soit (E, d) un espace métrique compact et f une fonction s.c.i. de E dans $\mathbb{R}$ alors il existe $x^* \in E$ tel que:*

$$f(x^*) = \inf \{f(x),\ x \in E\}.$$

Preuve:
Soit $(x_n)_n \in E^{\mathbb{N}}$ une suite minimisante de f sur E:

$$\lim_n f(x_n) = \inf_{x \in E} f(x). \tag{8.8}$$

Comme E est compact, on peut, quitte à extraire une sous-suite que nous continuerons à noter $(x_n)_n$, supposer que $(x_n)_n$ converge vers un élément $x^* \in E$. Comme f est s.c.i en x^*, on a:

$$\lim_n f(x_n) = \mathrm{liminf} f(x_n) \geq f(x^*),$$

en utilisant (8.8), il vient donc:

$$f(x^*) = \inf \{f(x),\, x \in E\}\,.$$

□

Dans un $\mathbb{R}$-ev de dimension finie, on peut remplacer l'hypothèse de compacité par l'hypothèse (8.9), appelée hypothèse de coercivité.

Théorème 8.3 *Soit E une partie non vide fermée de $\mathbb{R}^n$, f une fonction s.c.i. de E dans $\mathbb{R}$ telle que*[1]*:*

$$\lim_{x\in E,\ \|x\|\to+\infty} f(x) = +\infty \tag{8.9}$$

alors il existe $x^* \in E$ tel que:

$$f(x^*) = \inf \{f(x),\, x \in E\}\,.$$

Preuve:
Soit $(x_n)_n \in E^{\mathbb{N}}$ une suite minimisante de f sur E:

$$\lim_n f(x_n) = \inf_{x\in E} f(x) < +\infty. \tag{8.10}$$

Montrons d'abord que $(x_n)_n$ est bornée: si tel n'était pas le cas, il existerait une sous-suite $(x_{\varphi(n)})_n$ vérifiant:

$$\lim_n \|x_{\varphi(n)}\| = +\infty,$$

avec l'hypothèse de coercivité (8.9), on aurait alors:

$$\lim_n f(x_{\varphi(n)}) = +\infty,$$

ce qui contredirait (8.10). On a montré que $(x_n)_n$ est bornée dans E, fermé d'un $\mathbb{R}$-ev de dimension finie, il existe donc une sous-suite $(x_{\varphi(n)})$ convergeant vers un élément $x^* \in E$. Comme f est s.c.i en x^*, on a:

$$\text{liminf} f(x_{\varphi(n)}) \geq f(x^*),$$

en utilisant (8.10), il vient donc:

$$f(x^*) = \inf \{f(x),\, x \in E\}\,.$$

[1]Rappelons que (8.9) signifie que $\forall\ M > 0$, $\exists r > 0$ tel que pour tout $x \in E$, $\|x\| \geq r \Rightarrow f(x) \geq M$.

□

En dimension infinie, la corecivité ne suffit pas pour conclure (les suites minimisantes sont bornées mais ca ne suffit plus pour en extraire une sous suite convergente). Néanmoins, en recourant à la topologie faible on a le résultat suivant dans les Hilbert (valable plus généralement dans les Banach réflexifs):

Theorem 8.1 *Soit E un espace de Hilbert, $f : E \to \mathbb{R}$* convexe *s.c.i et coercive alors il existe $x^* \in E$*

$$f(x^*) = \inf \{f(x),\, x \in E\}\,.$$

La convexité est fondamentale dans le théorème prédédent. On renvoie par exemple à [3] pour une démonstration ainsi que la définition et les propriétés des topologies faibles.

8.2 Optimality conditions

On s'intéresse dans toute cette partie au problème:

$$\inf_{x\in\Omega} f(x) \tag{8.11}$$

Avec Ω un ouvert de $\mathbb{R}^n$ et f une fonction définie sur Ω à valeurs réelles satisfaisant certaines hypothèses de différentiabilité qui seront précisées au fur et à mesure. Le problème (8.11) avec $\Omega = \mathbb{R}^n$ par exemple est le problème-type d'optimisation sans contrainte. Les résultats de ce paragraphe sont supposés connus aussi les énoncerons-nous sans démonstration.

Rappelons d'abord la condition nécessaire du premier ordre classique (appelée aussi règle de Fermat) qui exprime que les points d'extrema locaux de f sur Ω sont des points critiques de f:

Proposition 8.3 *Si $x^* \in \Omega$ est un point de minimum local de f sur Ω et si f est différentiable en x^* alors:*

$$\nabla f(x^*) = 0.$$

Proof:
Soit $h \neq 0$ pour $t > 0$ assez petit on $f(x^* + th) - f(x^*) \geq 0$ en divisant par t et en faisant tendre t vers 0^+ on obtient $\nabla f(x^*) \cdot h \geq 0$ et comme h est arbitraire on en tire le résultat.
□

Comme nous l'avons déja vu, dans le cas convexe on a beaucoup mieux: le fait d'être point critique est une condition suffisante de minimum global:

Proposition 8.4 *Soit Ω un ouvert convexe de $\mathbb{R}^n$, f une fonction convexe sur Ω. Si f est différentiable en $x^* \in \Omega$ et $\nabla f(x^*) = 0$, alors x^* est une solution de (8.11) i.e. un point de minimum global de f sur Ω.*

La condition classique nécessaire du second-ordre nous est fournie par:

Proposition 8.5 *Si $x^* \in \Omega$ est un point de minimum local de f sur Ω et si f est deux fois différentiable en x^*, alors on a:*

$\nabla f(x^*) = 0$ *et la matrice (symétrique)* $D^2 f(x^*)$ *est semi-définie-positive.*

Proof:
On a déja vu que $\nabla f(x^*) = 0$. Soit $h \neq 0$ pour t assez petit, on a avec la formule de Taylor

$$0 \leq f(x^* + th) - f(x^*) = \frac{t^2}{2} D^2 f(x^*)h \cdot h + o(t^2)$$

en divisant par t^2 et en faisant tendre t vers 0, on obtient donc $D^2 f(x^*)h \cdot h \geq 0$.

□

Terminons par une condition suffisante de minimum local strict :

Proposition 8.6 *Si f est deux fois différentiable en $x^* \in \Omega$ et si l'on a:*

$\nabla f(x^*) = 0$ *et la matrice (symétrique)* $D^2 f(x^*)$ *est définie-positive,*

alors x^* est un point de minimum local strict de f sur Ω.

Proof:
On a

$$f(x^* + h) - f(x^*) = \frac{1}{2} D^2 f(x^*)h \cdot h + \|h\|^2 \varepsilon(h)$$

avec ε tendant vers 0 quand h tend vers 0. En utilisant le Lemme ci dessous, on déduit qu'il existe une constante $c > 0$ telle que $D^2 f(x^*)h \cdot h \geq c\|h\|^2$, et donc pour $h \neq 0$ suffisament petit pour que $c + 2\varepsilon(h) > 0$ on a

$$f(x^* + h) - f(x^*) \geq \left(\frac{c}{2} + \varepsilon(h)\right) \|h\|^2 > 0$$

ce qui montre que x^* est un point de minimum local strict de f.

□

Lemme 8.1 *Soit A une matrice définie positive alors il existe $c > 0$ telle que*

$$Ah \cdot h \geq c\|h\|^2; \forall h$$

Proof:
Soit $c := \inf\{Ah \cdot h : \|h\| = 1\}$, alors $c > 0$ et on déduit le résultat cherché par homogénéité. □

Remarque. Bien noter la différence entre la condition nécessaire de la proposition 8.5 et la condition suffisante de la proposition 8.6. Bien noter aussi que la condition suffisante de la proposition 8.6 n'est que locale mais assure que x^* est un minimum local strict.

Chapter 9

Problems with equality constraints

Dans ce chapitre nous nous intéressons à des problèmes d'optimisation sous contraintes d'égalité dans $\mathbb{R}^n$. Etant donnés Ω un ouvert de $\mathbb{R}^n$, f et $g_1, ..., g_m$ des fonctions définies sur Ω à valeurs réelles et $(c_1, ..., c_m) \in \mathbb{R}^m$, on considère donc le problème:

$$\inf_{x \in A} f(x) \tag{9.1}$$

avec:

$$A := \{x \in \Omega : \ g_j(x) = c_j, \ j = 1, ..., m\} \tag{9.2}$$

La fonction f à minimiser s'appelle fonction objectif ou coût. Les fonctions g_j et les réels c_j définissent les contraintes d'égalité de (9.1), les éléments de A s'appellent les éléments admissibles, on supposera évidemment dans ce qui suit que $A \neq \emptyset$. Comme précédemment, on distingue les solutions locales et globales, strictes et larges:

Définition 9.1 .

1. *On dit que x^* est une solution globale de (9.1) ou un point de minimum global de f sur A ssi $f(x^*) \leq f(x)$ pour tout $x \in A$.*

2. *On dit que x^* est une solution locale de (9.1) ou un point de minimum local de f sur A ss'il existe $r > 0$ tel que $f(x^*) \leq f(x)$ pour tout $x \in A$ tel que $\|x - x^*\| < r$.*

3. *On dit que x^* est un point de minimum global strict de f sur A ssi $f(x^*) < f(x)$ pour tout $x \in A \setminus \{x^*\}$.*

4. *On dit que x^* un point de minimum local strict de f sur A ss'il existe $r > 0$ tel que $f(x^*) < f(x)$ pour tout $x \in A$ tel que $\|x - x^*\| < r$ et $x \neq x^*$.*

Evidemment, la première question à se poser est celle de l'existence d'une solution de (9.1), pour cela, on utilise les résultats du paragraphe 12.1. En effet, ces derniers ont été obtenus dans des espaces métriques généraux, ils s'appliquent en particulier à la partie A de $\mathbb{R}^n$.

Il sera commode dans ce qui suit de noter sous forme plus synthétique les contraintes. Pour cela, on définit:

$$g : \left\{ \begin{array}{ccc} \Omega & \rightarrow & \mathbb{R}^m \\ x & \mapsto & g(x) := (g_1(x), ..., g_m(x)) \end{array} \right.$$

Ainsi, en posant $c = (c_1, ..., c_m)$, l'ensemble admissible s'écrit simplement $A := g^{-1}(\{c\})$.

9.1 Some linear algebra

Avant d'aller plus avant, rappelons quelques résultats d'algèbre linéaire. Tout d'abord, rappelons que si E est un $\mathbb{R}$-ev et E_1 et E_2 deux sev de E, on dit que E_1 et E_2 sont supplémentaires (ce que l'on note $E = E_1 \oplus E_2$) ssi pour tout $x \in E$ il existe un unique $(x_1, x_2) \in E_1 \times E_2$ tel que $x = x_1 + x_2$. Autrement dit $E = E_1 \oplus E_2$ ssi:

$$\Phi : \left\{ \begin{array}{ccc} E_1 \times E_2 & \rightarrow & E \\ (x_1, x_2) & \mapsto & x_1 + x_2 \end{array} \right.$$

est un isomorphisme entre $E_1 \times E_2$ et E. Cet isomorphisme permet d'identifier E au produit $E_1 \times E_2$, dans ce cas on fera l'identification $x = x_1 + x_2 = \Phi^{-1}(x) = (x_1, x_2)$. Notons enfin que si E est de dimension finie alors Φ et Φ^{-1} sont continues, dans ce cas l'identification précédente $x = \Phi^{-1}(x)$ n'altère en rien les considérations topologiques et différentielles.

Proposition 9.1 *Soit E et F deux $\mathbb{R}$-ev, $v \in L(E, F)$, $E_1 := \ker(v)$ et E_2 un supplémentaire de E_1 alors la double restriction de v à E_2 et $\mathrm{Im}(v)$ est un isomorphisme.*

Preuve:
Notons w la double restriction de v à E_2 et $\mathrm{Im}(v)$, soit $y \in \mathrm{Im}(v)$, il existe $x \in E$ tel que $y = v(x)$. Comme $E_1 \oplus E_2 = E$, il existe un unique $(x_1, x_2) \in$

$E_1 \times E_2$ tel que $x = x_1 + x_2$, par définition de E_1, on a $v(x_1) = 0$ donc $y = v(x_1+x_2) = v(x_2) = w(x_2)$ ainsi w est surjective. Supposons maintenant que $x \in E_2$ vérifie $w(x) = 0 = v(x)$ alors $x \in E_1 \cap E_2 = \{0\}$ ainsi w est injective.

□

Lemme 9.1 *Soit E un $\mathbb{R}$-ev, $u_1, ..., u_m$ m formes linéaires sur E et $u \in L(E, \mathbb{R}^m)$ défini par $u(x) := (u_1(x), ..., u_m(x)$ pour tout $x \in E$. Les assertions suivantes sont alors équivalentes:*

1. *u est surjective,*

2. *$u_1, ..., u_m$ est une famille libre.*

Preuve:
Si $u_1, ..., u_m$ est une famille liée, il existe des réels non tous nuls $\lambda_1, ..., \lambda_m$ tels que $\sum_{j=1}^m \lambda_j u_j = 0$. Ainsi pour tout $x \in E$ on a $u(x) \in H$ où H est l'hyperplan de $\mathbb{R}^m$ défini par l'équation $\sum_{j=1}^m \lambda_j y_j = 0$ ainsi $\mathrm{Im}(u) \neq \mathbb{R}^m$.

Si u n'est pas surjective $\mathrm{Im}(u) \neq \mathbb{R}^m$ et donc $\mathrm{Im}(u) \subset H$ avec H un hyperplan de $\mathbb{R}^m$. Soit $\sum_{j=1}^m \lambda_j y_j = 0$ une équation de H ($\lambda_1, ..., \lambda_m$ réels non tous nuls), comme $u(x) \in H$ pour tout $x \in E$, on a

$$\sum_{j=1}^m \lambda_j u_j(x) = 0, \, \forall x \in E.$$

Ainsi $u_1, ..., u_m$ est liée.

□

Achevons ces préliminaires avec une variante d'un corollaire du Lemme de Farkas:

Lemme 9.2 *Soit E un $\mathbb{R}$-ev, $u_1, ..., u_m$ et v $m+1$ formes linéaires sur E. Les assertions suivantes sont alors équivalentes:*

1. *$\cap_{j=1}^m \ker u_j \subset \ker(v)$,*

2. *il existe $(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ telle que:*

$$v = \sum_{j=1}^m \lambda_j u_j.$$

Preuve:
Tout d'abord, il est évident que 2. implique 1.. Supposons maintenant que 1. ait lieu, il s'agit de montrer que $v \in F := \text{vect}(u_1, ..., u_m)$. F est un sev de l'ev de dimension finie $G := \text{vect}(u_1, ..., u_m, v)$. On identifie G à $\mathbb{R}^p$ ($p = \dim(G)$) et on le munit de la structure hilbertienne usuelle de $\mathbb{R}^p$. Ainsi on identifie aussi G à son dual. Si $v \notin F$, comme F est un sev fermé de G on peut séparer v de F: il existe $x \in \mathbb{R}^p$ et $\varepsilon > 0$ tels que:

$$v(x) \leq \inf_{p \in F} p(x) - \varepsilon. \tag{9.3}$$

Comme G est un sev, nous déduisons de (9.3) que $p(x) = 0$ pour tout $p \in F$ (voir la démonstration du lemme de Farkas pour les détails), en particulier ceci implique que $x \in \cap_{j=1}^m \ker u_j$ et donc 1. implique que $v(x) = 0$. Or avec (9.3), on a $v(x) \leq -\varepsilon < 0$ ce qui constitue la contradiction recherchée. □

9.2 Lagrange first-order optimality conditions

Proposition 9.2 *Soit $x^* \in A$, une solution locale de (9.1). On suppose que:*

1. *f est différentiable en x^*,*
2. *g est de classe C^1 au voisinage de x^*,*
3. *$g'(x^*)$ est surjective*

alors pour tout $h \in \mathbb{R}^n$, on a:

$$g'(x^*)(h) = 0 \Rightarrow \nabla f(x^*) \cdot h = 0 \tag{9.4}$$

Preuve:
Il s'agit de montrer que:

$$E_1 := \ker(g'(x^*)) = \cap_{j=1}^m \ker(g_j'(x^*)) \subset \ker(f'(x^*)) = \nabla f(x^*)^\perp. \tag{9.5}$$

Soit E_2 un supplémentaire de E_1, par la suite on notera les éléments de $\mathbb{R}^n$ sous la forme $x = (x_1, x_2)$ selon le "découpage" $\mathbb{R}^n = E_1 \oplus E_2$, on notera en particulier $x^* = (x_1^*, x_2^*)$. On notera également ∂_i, $i = 1, 2$ les différentielles partielles selon le découpage $\mathbb{R}^n = E_1 \oplus E_2$. Par construction, on a: $\partial_1 g(x^*) = 0$. D'après la proposition 9.1, $\partial_2 g(x^*)$ est un isomorphisme de E_2 sur $\text{Im}(g'(x^*))$ et comme $g'(x^*)$ est surjective, $\text{Im}(g'(x^*)) = \mathbb{R}^m$ donc $\partial_2 g(x^*)$ est un isomorphisme de E_2 vers $\mathbb{R}^m$.

Puisque $g(x^*) - c = 0$ et $\partial_2 g(x^*)$ est inversible, il résulte du théorème des fonctions implicites qu'il existe un voisinage ouvert U_1 de x_1^*, un voisinage

ouvert U_2 de x_2^* et une application $\Psi \in C^1(U_1, U_2)$ tels que $U_1 \times U_2 \subset \Omega$, $\Psi(x_1^*) = x_2^*$ et:

$$A \cap (U_1 \times U_2) = \{(x_1, \Psi(x_1)) : x_1 \in U_1\}.$$

En dérivant la relation $g(x_1, \Psi(x_1)) = c$ valable sur U_1, il vient:

$$\partial_1 g(x_1, \Psi(x_1)) + \partial_2 g(x_1, \Psi(x_1)) \circ \Psi'(x_1) = 0 \ \forall x_1 \in U_1$$

en prenant $x_1 = x_1^*$, en utilisant $\partial_1 g(x^*) = 0$ et le fait que $\partial_2 g(x_1^*, \Psi(x_1^*)) = \partial_2 g(x^*)$ est inversible, on obtient donc:

$$\Psi'(x_1^*) = 0.$$

Soit $h \in E_1$, pour $t \in \mathbb{R}$ suffisamment petit, on a $x_1^* + th \in U_1$, $(x_1^* + th, \Psi(x_1^* + th)) \in A$ et:

$$f(x_1^* + th, \Psi(x_1^* + th)) \geq f(x^*) = f(x_1^*, \Psi(x_1^*)).$$

Ainsi la fonction d'une variable, $\gamma_h : t \mapsto f(x_1^* + th, \Psi(x_1^* + th))$, définie sur un voisinage ouvert de 0, présente un minimum local en $t = 0$, comme γ_h est dérivable en 0 il vient donc:

$$\dot{\gamma}_h(0) = 0 = (\partial_1 f(x^*) + \partial_2 f(x^*) \circ \Psi'(x_1^*))(h) \tag{9.6}$$

et comme $\Psi'(x_1^*) = 0$ on en déduit donc que $\partial_1 f'(x^*)(h) = 0$. Comme $h \in E_1$, on a donc:

$$\partial_1 f'(x^*)(h) = 0 = f'(x^*)(h).$$

On a donc bien établi que $E_1 := \ker(g'(x^*)) \subset \ker(f'(x^*)) = \nabla f(x^*)^\perp$. □

Retenez que l'idée essentielle de la preuve précédente est de se ramener à une minimisation sans contrainte et ce grâce au théorème des fonctions implicites.

Les conditions nécessaires du premier ordre dites de Lagrange sont alors fournies par le théorème suivant:

Théorème 9.1 *Soit $x^* \in A$, une solution locale de (9.1). On suppose que:*

1. *f est différentiable en x^*,*
2. *g est de classe C^1 au voisinage de x^*,*
3. *la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre,*

alors il existe $(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ *tels que:*

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*). \tag{9.7}$$

Preuve:
Il résulte du lemme 9.1 que $g'(x^*)$ est surjective et donc que les hypothèses de la proposition 9.2 sont satisfaites. Ainsi on a $\cap_{j=1}^m \ker(g_j'(x^*)) \subset \ker(f'(x^*))$ ce qui est équivalent à:

$$\cap_{j=1}^m \nabla g_j(x^*)^\perp \subset \nabla f(x^*)^\perp.$$

Ainsi le lemme 9.2 (ou le lemme de Farkas) permet d'en déduire qu'il existe $(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ tels que:

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*).$$

□

Les réels $\lambda_1, ..., \lambda_m$ intervenant dans (9.7) sont appelés des multiplicateurs de Lagrange associés aux contraintes de (9.1) au point de minimum local x^*.

Remarque. L'hypothèse que la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre implique que les multiplicateurs $\lambda_1, ...\lambda_m$ associés à x^* sont uniques. La conclusion du théorème 9.1 signifie simplement que le gradient de la fonction objectif en x^* appartient à l'espace vectoriel engendré par les gradients des contraintes en x^*.

Exemple 9.1 *Cherchons à minimiser* $f(x_1, ..., x_n) = x_1 + ... + x_n$ *sous la contrainte* $g(x_1, ...x_n) = \sum_{i=1}^n x_i^2 = 1$. *Tout d'abord, par compacité de la sphère il existe au moins une solution. Ensuite notons que* $\nabla g(x) = 2x$ *ainsi les conditions du thèorème de Lagrange sont remplies. Ainsi si* x^* *est un minimiseur il existe* $\lambda \in \mathbb{R}$ *tel que* $1 = \lambda x_i^*$ *pour* $i = 1, ..., n$ *ce qui implique que les composantes de* x^* *sont égales. Avec la contrainte cela laisse les deux possibilités:*

$$x^* = (n^{-1/2}, ..., n^{-1/2}) \text{ ou } x^* = -(n^{-1/2}, ..., n^{-1/2}).$$

Le premier cas correspond au point de maximum de f *sur la sphère et le second au point de minimum de* f *sur la sphère.*

Remarque. Attention à l'hypothèse "$g'(x^*)$ surjective" (équivalente, rappelons le, au fait que la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre). Cette hypothèse ne peut être affaiblie pour que la conclusion du théorème de Lagrange reste valide (voir exemple ci-dessous). Par ailleurs, cette hypothèse porte sur x^*, qui est en pratique ce que l'on cherche et donc a priori inconnu! Enfin, remarquons que si $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre alors $m \leq n$ c'est à dire qu'il y a moins de contraintes que de variables....

Exemple 9.2 *Il est facile de construire des contre exemples à (9.7) si l'hypothèse "$g'(x^*)$ surjective" n'est pas vérifiée. Cherchons à minimiser $f(x,y) = x$ sous la contrainte $x^2 + y^2 = 0$: il n'y a qu'un point admissible $(0,0)$ et $\nabla f(0,0) = (1,0) \neq \lambda \nabla g(0,0) = 0$, dans ce cas, à cause de la dégénérescence $\nabla g(0,0) = (0,0)$, il n'y a pas de multiplicateur de Lagrange.*

Remarque. Nous verrons au paragraphe suivant comment en introduisant un Lagrangien généralisé, on peut aussi traiter les cas dégénérés où $g'(x^*)$ n'est pas surjective (i.e. $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ liée)

Lorsque le problème (9.1) est convexe, la condition (9.7) est suffisante et assure que le minimum est global. Ce cas est celui où les contraintes sont affines et l'objectif convexe:

Proposition 9.3 *Supposons que Ω est un ouvert convexe de $\mathbb{R}^n$, que f est une fonction convexe sur Ω et que g_j est une fonction affine pour $j = 1, ..., m$. Si $x^* \in A$ est tel qu'il existe $(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ tels que:*

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*). \tag{9.8}$$

alors $f(x^*) \leq f(x)$ pour tout $x \in A$.

Preuve:
Soit $x \in A$, par convexité de f, on a:

$$f(x) - f(x^*) \geq \nabla f(x^*) \cdot (x - x^*)$$

avec (9.8), il vient donc:

$$f(x) - f(x^*) \geq \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*).(x - x^*). \tag{9.9}$$

Comme les g_j sont affines et $g_j(x) = g_j(x^*) = c_j$ on a aussi:

$$g_j(x) - g_j(x^*) = 0 = \nabla g_j(x^*).(x - x^*)$$

en reportant dans (9.9), il vient bien $f(x^*) \leq f(x)$. □

9.3 The Lagrangian and the generalized Lagrangian

Le lagrangien du problème (9.1) est la fonction définie sur $\Omega \times \mathbb{R}^m$ par:

$$\mathcal{L}(x, \lambda_1, ...\lambda_m) := f(x) - \sum_{j=1}^{m} \lambda_j (g_j(x) - c_j). \tag{9.10}$$

Remarquons alors que si f et les fonctions g_j sont différentiables en $x \in \Omega$ alors on a:

$$\nabla_x \mathcal{L}(x, \lambda_1, ..., \lambda_m) = \nabla f(x) - \sum_{j=1}^{m} \lambda_j \nabla g_j(x). \tag{9.11}$$

et

$$\partial_{\lambda_j} \mathcal{L}(x, \lambda_1, ..., \lambda_m) = c_j - g_j(x) \tag{9.12}$$

Ainsi le fait que $x \in A$ i.e. vérifie la contrainte $g(x) = c$ peut s'exprimer par:

$$\partial_{\lambda_j} \mathcal{L}(x, \lambda_1, ..., \lambda_m) = 0 \text{ pour } j = 1, ..., m$$

ou, sous forme plus synthétique, en posant $\lambda = (\lambda_1, ..., \lambda_m)$:

$$\nabla_\lambda \mathcal{L}(x, \lambda) = 0. \tag{9.13}$$

Avec (9.11), la condition de Lagrange se traduit par:

$$\nabla_x \mathcal{L}(x, \lambda_1, ..., \lambda_m) = 0.$$

Le théorème 9.1 peut donc se reformuler comme suit:

Théorème 9.2 *Soit $x^* \in A$, une solution locale de (9.1). On suppose que:*

1. *f est différentiable en x^*,*
2. *g est de classe C^1 au voisinage de x^*,*
3. *la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre,*

alors il existe $\lambda = (\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ tels que:

$$\mathcal{L}'(x^*, \lambda) = 0. \tag{9.14}$$

Nous avons déja discuté le caractère contraignant de la condition "$g'(x^*)$ surjective" qui porte sur le point inconnu x^*. Pour remédier à cela on peut ajouter un multiplicateur à la fonction objectif, ceci conduit à la définition du lagrangien généralisé. Le lagrangien généralisé du problème (9.1) est la fonction définie sur $\Omega \times \mathbb{R}^{m+1}$ par:

$$\mathcal{L}_0(x, \lambda_0, \lambda_1, ...\lambda_m) := \lambda_0 f(x) - \sum_{j=1}^{m} \lambda_j (g_j(x) - c_j). \tag{9.15}$$

La condition du premier ordre peut en effet se formuler par:

Théorème 9.3 *Soit $x^* \in A$, une solution locale de (9.1). On suppose que:*

1. *f est différentiable en x^*,*
2. *g est de classe C^1 au voisinage de x^*,*

alors il existe des réels $(\lambda_0, \lambda_1, ..., \lambda_m) \in \mathbb{R}^{m+1}$ non tous nuls tels que:

$$\lambda_0 \nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*). \tag{9.16}$$

Preuve:
Définissons pour tout $x \in \Omega$, $H(x) := (f(x), g(x)) \in \mathbb{R}^{m+1}$. Par hypothèse, H est différentiable en x^*: $H'(x^*) = (f'(x^*), g'(x^*))$. Distinguons alors deux cas:

Premier cas: $H'(x^*)$ est surjective. Ceci implique que $g'(x^*)$ est surjective on peut alors appliquer le théorème 9.1 et prendre $\lambda_0 = 1$ dans (9.16).

Deuxième cas: $H'(x^*)$ n'est pas surjective. Puisque

$$H'(x^*) = (f'(x^*), g_1'(x^*), ..., g_m'(x^*)),$$

il résulte alors du lemme 9.1 que la famille de formes linéaires sur $\mathbb{R}^n$, $(f'(x^*), g_1'(x^*), ..., g_m'(x^*))$ est liée ce qui revient à dire que

$$\text{la famille } (\nabla f(x^*), \nabla g_1(x^*), ..., \nabla g_m(x^*)) \text{ est liée dans } \mathbb{R}^n.$$

Il existe donc des réels $(\lambda_0, \lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ non tous nuls tels que:

$$\lambda_0 \nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*).$$

□

Remarque. Notons que la condition (9.16) est équivalente à:

$$\nabla_x \mathcal{L}_0(x^*, \lambda_0, \lambda_1, ..., \lambda_m) = 0$$

par ailleurs, $g(x^*) - c = 0$ s'exprime aussi sous la forme

$$\partial_{\lambda_j} \mathcal{L}_0(x^*, \lambda_0, \lambda_1, ..., \lambda_m) = 0 \text{ pour } j = 1, ..., m.$$

Notons enfin que

$$\partial_{\lambda_0} \mathcal{L}_0(x^*, \lambda_0, \lambda_1, ..., \lambda_m) = f(x^*)$$

donc, en général $\mathcal{L}'_0(x^*, \lambda_0, \lambda_1, ..., \lambda_m)$ est différent de 0.

9.4 Second-order optimality conditions

Les conditions nécessaires du second-ordre pour un minimum local de (9.1) sont données par:

Théorème 9.4 *Soit $x^* \in A$, une solution locale de (9.1). On suppose que:*

1. *f est deux fois différentiable en x^*,*
2. *g est de classe C^2 au voisinage de x^*,*
3. *la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre,*

alors il existe $\lambda := (\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ tel que:

$$\nabla_x \mathcal{L}(x^*, \lambda) = \nabla f(x^*) - \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*) = 0 \text{ et} \tag{9.17}$$

$$\partial^2_{xx} \mathcal{L}(x^*, \lambda)(h, h) \geq 0 \text{ pour tout } h \in \ker(g'(x^*)) \tag{9.18}$$

Preuve:
Il résulte du théorème de Lagrange 9.1 qu'il existe $\lambda \in \mathbb{R}^m$ tel que $\nabla_x \mathcal{L}(x^*, \lambda) = 0$, autrement dit:

$$f'(x^*) = \sum_{j=1}^{m} \lambda_j g'_j(x^*). \tag{9.19}$$

Posons $E_1 := \ker(g'(x^*))$ et soit E_2 un supplémentaire de E_1. Comme précédemment, on notera les éléments de $\mathbb{R}^n$ sous la forme $x = (x_1, x_2)$ selon la décomposition $\mathbb{R}^n = E_1 \oplus E_2$, on notera en particulier $x^* = (x_1^*, x_2^*)$. On notera également ∂_i, $i = 1, 2$ les différentielles partielles selon le découpage

$\mathbb{R}^n = E_1 \oplus E_2$. Par construction, on a: $\partial_1 g(x^*) = 0$. D'après la proposition 9.1, $\partial_2 g(x^*)$ est un isomorphisme de E_2 sur $\mathrm{Im}(g'(x^*))$ et comme $g'(x^*)$ est surjective, $\mathrm{Im}(g'(x^*)) = \mathbb{R}^m$ donc $\partial_2 g(x^*)$ est un isomorphisme de E_2 vers $\mathbb{R}^m$.

L'identité (9.19) implique en particulier:

$$\partial_2 f(x^*) = \sum_{j=1}^{m} \lambda_j \partial_2 g_j(x^*). \tag{9.20}$$

Puisque $g(x^*) - c = 0$ et $\partial_2 g(x^*)$ est inversible, il résulte du théorème des fonctions implicites qu'il existe un voisinage ouvert U_1 de x_1^*, un voisinage ouvert U_2 de x_2^* et une application $\Psi \in C^2(U_1, U_2)$ tels que $U_1 \times U_2 \subset \Omega$, $\Psi(x_1^*) = x_2^*$ et:

$$A \cap (U_1 \times U_2) = \{(x_1, \Psi(x_1)) : x_1 \in U_1\}.$$

En dérivant une première fois la relation $g(x_1, \Psi(x_1)) = c$ valable sur U_1, il vient:

$$\partial_1 g(x_1, \Psi(x_1)) + \partial_2 g(x_1, \Psi(x_1)) \circ \Psi'(x_1) = 0 \ \forall x_1 \in U_1 \tag{9.21}$$

en prenant $x_1 = x_1^*$, en utilisant $\partial_1 g(x^*) = 0$ et le fait que $\partial_2 g(x_1^*, \Psi(x_1^*)) = \partial_2 g(x^*)$ est inversible, on obtient donc:

$$\Psi'(x_1^*) = 0. \tag{9.22}$$

En dérivant (9.21), il vient:

$$\begin{gathered} \partial_{11}^2 g(x_1, \Psi(x_1)) + 2\partial_{12}^2 g(x_1, \Psi(x_1)) \circ \Psi'(x_1) + \\ \partial_{22}^2 g(x_1, \Psi(x_1))(\Psi'(x_1), \Psi'(x_1)) + \partial_2 g(x_1, \Psi(x_1)) \circ \Psi''(x_1) = 0. \end{gathered} \tag{9.23}$$

pour $x_1 = x_1^*$, en utilisant (9.22), il vient alors:

$$\partial_{11}^2 g(x^*) + \partial_2 g(x^*) \circ \Psi''(x_1^*) = 0. \tag{9.24}$$

Pour $x_1 \in U_1$ posons $F(x_1) := f(x_1, \Psi(x_1))$, F présente alors un minimum local en x_1^* et comme F est deux fois dérivable en x_1^* on a:

$$F'(x_1^*)(h) = 0 \text{ pour tout } h \in E_1. \tag{9.25}$$

et

$$F''(x_1^*)(h, h) \geq 0 \text{ pour tout } h \in E_1. \tag{9.26}$$

Avec des calculs semblables à (9.23) et (9.24), on a:

$$F''(x_1^*)(h,h) = \partial_{11}^2 f(x^*)(h,h) + \partial_2 f(x^*)(\Psi''(x_1^*)(h,h))$$

(9.26) devient alors:

$$\partial_{11}^2 f(x^*)(h,h) + \partial_2 f(x^*)(\Psi''(x_1^*)(h,h)) \geq 0 \text{ pour tout } h \in E_1. \qquad (9.27)$$

Avec (9.20) et (9.24), on a par ailleurs:

$$\begin{aligned}
&\partial_{11}^2 f(x^*)(h,h) + \partial_2 f(x^*)(\Psi''(x_1^*)(h,h)) \\
=&\partial_{11}^2 f(x^*)(h,h) + \sum_{j=1}^m \lambda_j \partial_2 g_j(x^*)(\Psi''(x_1^*)(h,h)) \\
=&\partial_{11}^2 f(x^*)(h,h) - \sum_{j=1}^m \lambda_j \partial_{11}^2 g_j(x^*)(h,h).
\end{aligned}$$

Si bien que (9.27) se réecrit:

$$\left(\partial_{11}^2 f(x^*) - \sum_{j=1}^m \lambda_j \partial_{11}^2 g_j(x^*)\right)(h,h) \geq 0 \text{ pour tout } h \in E_1 = \ker(g'(x^*))$$

or pour $h \in E_1$, on a:

$$\begin{aligned}
\left(\partial_{11}^2 f(x^*) - \sum_{j=1}^m \lambda_j \partial_{11}^2 g_j(x^*)\right)(h,h) =&(f''(x^*) - \sum_{j=1}^m \lambda_j g_j''(x^*))(h,h) \\
=&\partial_{xx}^2 \mathcal{L}(x^*,\lambda)(h,h)
\end{aligned}$$

ce qui achève la preuve. □

Remarque. Attention à la condition $h \in \ker(g'(x^*))$ dans (9.18). La condition du second-ordre (9.18):

$$\partial^2 \mathcal{L}(x^*,\lambda)(h,h) \geq 0 \text{ pour tout } h \in \ker(g'(x^*))$$

signifie que la forme quadratique $\partial_{xx}^2 \mathcal{L}(x^*,\lambda)$ est semi-définie positive sur $\ker(g'(x^*))$ (pas sur $\mathbb{R}^n$ en entier en général). Notons qu'on peut aussi exprimer cette forme quadratique sous la forme développée:

$$\partial_{xx}^2 \mathcal{L}(x^*,\lambda) = f''(x^*) - \sum_{j=1}^m \lambda_j g_j''(x^*)$$

Remarque. Notez bien que pour écrire la condition du second-ordre, il faut avoir déterminé d'abord les multiplicateurs de Lagrange.

Enfin, voici des conditions suffisantes pour un minimum local de (9.1):

Théorème 9.5 *Soit $x^* \in A$. On suppose que:*

1. *f est deux fois différentiable en x^*,*
2. *g est de classe C^2 au voisinage de x^*,*
3. *la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre,*

S'il existe $\lambda := (\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ tel que:

$$\nabla_x \mathcal{L}(x^*, \lambda) = \nabla f(x^*) - \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*) = 0 \text{ et} \tag{9.28}$$

$$\partial^2_{xx} \mathcal{L}(x^*, \lambda)(h, h) > 0 \text{ pour tout } h \in \ker(g'(x^*)),\ h \neq 0 \tag{9.29}$$

alors x^* est un point de minimum local strict de f sur A.

Preuve:
En reprenant les notations de la preuve du théorème 9.5, il suffit de montrer que x_1^* est un point de minimum local strict de F (rappelons que $F(x_1) := f(x_1, \Psi(x_1))$ pour $x_1 \in U_1$). On commence par remarquer:

$$F'(x_1^*) = \partial_1 f(x^*) + \partial_2 f(x^*) \circ \Psi'(x_1^*)$$

Or, nous savons que

$$\Psi'(x_1^*) = 0,\ \partial_1 g(x^*) = 0 \text{ et } \partial_1 f(x^*) = \sum_{j=1}^{m} \lambda_j \partial_1 g_j(x^*) = 0$$

et donc $F'(x_1^*) = 0$. Comme dans la preuve du théorème 9.5, on a aussi pour tout $h \in \ker(g'(x^*)) = E_1$:

$$\begin{aligned} F''(x_1^*)(h,h) =& \partial^2_{11} f(x^*)(h,h) + \partial_2 f(x^*)(\Psi''(x_1^*)(h,h)) \\ =& (f''(x^*) - \sum_{j=1}^{m} \lambda_j g_j''(x^*))(h,h) \\ =& \partial^2_{xx} \mathcal{L}(x^*, \lambda)(h,h) \end{aligned}$$

Ainsi par (9.29), $F''(x_1^*)$ est une forme quadratique définie positive sur E_1. On déduit alors de la proposition 8.6 que x_1^* est un point de minimum local strict de F et donc un point de minimum local strict de f sur A .

□

Chapter 10

Problems with equality and inequality constraints

10.1 Notations

Dans ce chapitre nous nous intéressons à des problèmes d'optimisation sous contraintes d'égalité et d'inégalité dans $\mathbb{R}^n$. Etant donnés Ω un ouvert de $\mathbb{R}^n$, f, $g_1, ..., g_m$, $k_1, ..., k_p$ des fonctions définies sur Ω à valeurs réelles et des réels c_1, ... , c_m, d_1, ..., d_p, on considère le problème:

$$\inf_{x \in A} f(x) \tag{10.1}$$

avec:

$$A := \{x \in \Omega : \ g_j(x) = c_j, \ j = 1, ..., m, , \ k_i(x) \leq d_i, \ i = 1, ..., p\} \tag{10.2}$$

La fonction f à minimiser s'appelle fonction objectif ou coût. Les fonctions g_j et les réels c_j définissent les contraintes d'égalité de (10.1), les fonctions k_i et les réels d_i définissent les contraintes d'inégalité de (10.1). Les éléments de A s'appellent les éléments admissibles, on supposera évidemment dans ce qui suit que $A \neq \emptyset$. Comme précédemment, on distingue les solutions locales et globales, strictes et larges:

Définition 10.1 .

1. *On dit que x^* est une solution globale de (9.1) ou un point de minimum global de f sur A ssi $f(x^*) \leq f(x)$ pour tout $x \in A$.*

2. *On dit que x^* est une solution locale de (9.1) ou un point de minimum local de f sur A ss'il existe $r > 0$ tel que $f(x^*) \leq f(x)$ pour tout $x \in A$ tel que $\|x - x^*\| < r$.*

3. *On dit que x^* est un point de minimum global strict de f sur A ssi $f(x^*) < f(x)$ pour tout $x \in A \setminus \{x^*\}$.*

4. *On dit que x^* un point de minimum local strict de f sur A ss'il existe $r > 0$ tel que $f(x^*) < f(x)$ pour tout $x \in A$ tel que $\|x - x^*\| < r$ et $x \neq x^*$.*

Pour étudier l'existence d'une solution de (10.1), on utilise les résultats du paragraphe 12.1.

Il sera commode dans ce qui suit de noter sous forme plus synthétique les contraintes d'égalité, on définit:

$$g : \begin{cases} \Omega & \to & \mathbb{R}^m \\ x & \mapsto & g(x) := (g_1(x), ..., g_m(x)) \end{cases}$$

Ainsi les contraintes d'égalité de (10.1) s'écrivent simplement $g(x) = c$ ($c := (c_1, ..., c_m)$).

Soit $x \in A$ si $k_i(x) = d_i$ alors on dit que la i-ème contrainte d'inégalité est saturée en x (certains disent plutôt serrée, et en anglais, on dit : binding). On note $I(x)$ l'ensemble des contraintes saturées en $x \in A$:

$$I(x) := \{i \in \{1, ..., p\} \text{ t.q. } k_i(x) = d_i\}.$$

10.2 Preliminaries

Dans tout ce paragraphe on considère $x^* \in A$ tel que les conditions suivantes sont satisfaites:

1. g est de classe C^1 au voisinage de x^*,

2. k_i est différentiable en x^* pour tout $i \in I(x^*)$,

3. la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre,

4. les contraintes d'ínégalité sont qualifiées en x^* ce qui par définition signifie:

$$\exists h_0 \in \ker(g'(x^*)) \text{ tel que } \nabla k_i(x^*) \cdot h_0 < 0 \ \forall i \in I(x^*). \tag{10.3}$$

L'hypothèse de qualification (10.3) est très importante, elle peut également s'exprimer par

$$\exists h_0 \text{ tel que } \nabla g_j(x^*) \cdot h_0 = 0 \ \forall j \in \{1, ..., m\}, \text{ et}$$
$$\nabla k_i(x^*) \cdot h_0 < 0 \ \forall i \in I(x^*).$$

Posons $E_1 := \ker(g'(x^*))$ et soit E_2 un supplémentaire de E_1. Comme dans le chapitre précédent, on notera les éléments de $\mathbb{R}^n$ sous la forme $x = (x_1, x_2)$ selon le "découpage" $\mathbb{R}^n = E_1 \oplus E_2$, on notera en particulier $x^* = (x_1^*, x_2^*)$. On notera également ∂_i, $i = 1, 2$ les différentielles partielles selon le découpage $\mathbb{R}^n = E_1 \oplus E_2$. Par construction, on a: $\partial_1 g(x^*) = 0$. Comme au chapitre précédent, $\partial_2 g(x^*)$ est un isomorphisme de E_2 vers $\mathbb{R}^m$. Puisque $g(x^*) - c = 0$ et $\partial_2 g(x^*)$ est inversible, il résulte du théorème des fonctions implicites qu'il existe un voisinage ouvert U_1 de x_1^*, un voisinage ouvert U_2 de x_2^* et une application $\Psi \in C^1(U_1, U_2)$ tels que $U_1 \times U_2 \subset \Omega$, $\Psi(x_1^*) = x_2^*$ et:

$$\{(x_1, x_2) \in U_1 \times U_2 : g(x_1, x_2) = c\} = \{(x_1, \Psi(x_1)) : x_1 \in U_1\}. \tag{10.4}$$

En dérivant la relation $g(x_1, \Psi(x_1)) = c$ valable sur U_1, il vient:

$$\partial_1 g(x_1, \Psi(x_1)) + \partial_2 g(x_1, \Psi(x_1)) \circ \Psi'(x_1) = 0 \ \forall x_1 \in U_1$$

en prenant $x_1 = x_1^*$, en utilisant $\partial_1 g(x^*) = 0$ et le fait que $\partial_2 g(x_1^*, \Psi(x_1^*)) = \partial_2 g(x^*)$ est inversible, on obtient donc:

$$\Psi'(x_1^*) = 0. \tag{10.5}$$

Fixons maintenant $\varepsilon > 0$ et $h \in \mathbb{R}^n$ vérifiant:

$$h \in E_1 = \ker(g'(x^*)) \text{ et } \nabla k_i(x^*) \cdot h \leq 0 \ \forall i \in I(x^*). \tag{10.6}$$

Définissons pour $t > 0$:

$$x_1(t) = x_1^* + t(h + \varepsilon h_0). \tag{10.7}$$

On a alors $x_1(0) = x_1^*$, $x_1(t) \in E_1$ et $x_1(t) \in U_1$ pour $t > 0$ assez petit. Définissons pour $t > 0$ assez petit pour que $x_1(t) \in U_1$:

$$x(t) := (x_1(t), \Psi(x_1(t))). \tag{10.8}$$

Par construction, notons que $x(0) = x^*$ et avec (10.4), $g(x(t)) = c$ pour $t > 0$ assez petit. On a alors:

Lemme 10.1 *Sous les hypothèses précédentes soit $x(t) \in U_1 \times U_2$ défini par (10.8) pour $t > 0$ assez petit, on a:*

$$x(t) = x^* + t(h + \varepsilon h_0) + o(t)$$

et $x(t) \in A$ pour $t > 0$ assez petit.

Preuve:
On a $x_1(t) = x^* + t(h + \varepsilon h_0)$, posons $x_2(t) = \Psi(x_1(t))$ comme $x_1(0) = x_1^*$ et Ψ est dérivable en x_1^* avec $\Psi'(x_1^*) = 0$, x_2 est dérivable en 0 avec:

$$\dot{x}_2(t) = \Psi'(x_1^*)(h + \varepsilon h_0) = 0$$

donc $x_2(t) = o(t)$ et comme $x(t) = (x_1(t), x_2(t))$, il vient:

$$x(t) = (x_1^* + t(h + \varepsilon h_0), x_2^* + o(t)) = x^* + t(h + \varepsilon h_0) + o(t). \tag{10.9}$$

On sait déja que $g(x(t)) = c$, il s'agit de montrer que $x(t)$ satisfait aussi les contraintes d'ínégalité pour $t > 0$ assez petit, pour cela distinguons les contraintes saturées des contraintes non saturées en x^*. Si $i \notin I(x^*)$ alors $k_i(x^*) < d_i$ et puisque k_i est continue en $x^* = x(0)$ et $x(.)$ est continue en $t = 0$, on a par continuité $k_i(x(t)) < d_i$ pour $t > 0$ assez petit. Si $i \in I(x^*)$ alors on a $k_i(x^*) = k_i(x(0)) = d_i$ et avec(10.9) on a:

$$k_i(x(t)) = d_i + t\nabla k_i(x^*) \cdot (h + \varepsilon h_0) + o(t)$$

par hypothèse $\nabla k_i(x^*) \cdot h \leq 0$ et $\nabla k_i(x^*) \cdot h_0 < 0$ donc $k_i(x(t)) \leq d_i$ pour $t > 0$ assez petit.
□

10.3 Kuhn and Tucker optimality conditions

Nous sommes en mesure de prouver le théorème de Kuhn et Tucker (parfois aussi appelé théorème de Karush, Kuhn et Tucker ou KKT):

Théorème 10.1 *Soit $x^* \in A$ une solution locale de (10.1) telle que:*

1. *f est différentiable en x^*,*
2. *g est de classe C^1 au voisinage de x^*,*
3. *k_i est différentiable en x^* pour tout $i \in I(x^*)$,*
4. *la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre,*
5. *les contraintes d'ínégalité sont qualifiées en x^*:*

$$\exists h_0 \in \ker(g'(x^*)) \text{ tel que } \nabla k_i(x^*) \cdot h_0 < 0 \ \forall i \in I(x^*).$$

alors il existe $(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ *et* $\mu_i \geq 0$ *pour tout* $i \in I(x^*)$ *tel que:*

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*) - \sum_{i \in I(x^*)} \mu_i \nabla k_i(x^*). \tag{10.10}$$

Preuve:
Soit h vérifiant (10.6) et $x(t)$ défini pour $t > 0$ assez petit par (10.8), d'après le lemme 10.1, on a $x(t) \in A$, comme $x(0) = x^*$ et $x(.)$ est continu en 0 on a donc pour $t > 0$ assez petit:

$$\frac{1}{t}(f(x(t)) - f(x(0)) \geq 0. \tag{10.11}$$

Comme f est différentiable en x^* et en utilisant la première partie du lemme 10.1, on peut passer à la limite $t \to 0^+$ dans (10.11), il vient alors:

$$\nabla f(x^*) \cdot (h + \varepsilon h_0) \geq 0. \tag{10.12}$$

comme $\varepsilon > 0$ est arbitraire, on obtient aussi:

$$\nabla f(x^*) \cdot h \geq 0. \tag{10.13}$$

Comme (10.13) a lieu pour tout h vérifiant (10.6), on déduit (10.10) du lemme de Farkas. □

Les réels λ_j et μ_i sont appelés des multiplicateurs de Kuhn et Tucker associés aux contraintes de (10.1) au point de minimum local x^*.

Il faut retenir les remarques importantes et utiles en pratique:

- si toutes les contraintes sont linéaires, les hypothèses de qualification 4. et 5. sont superflues (reprendre la preuve : on n'a plus besoin du h_0!)
- de même, si les contraintes d'inégalité sont linéaires, l'hypothèse 5. est superflue,
- si les gradients des contraintes d'égalité et des contraintes d'inégalité saturées an x^* forment une famille libre, les conditions de qualification 4. et 5. sont satisfaites (utiliser le lemme 9.2).

Lorsque le problème (10.1) est convexe, la condition (10.10) est suffisante et assure que le minimum est global.

Proposition 10.1 *Supposons que Ω est un ouvert convexe de $\mathbb{R}^n$, que f et $k_1, ..., k_p$ sont des fonctions convexes sur Ω, g_j est une fonction affine pour $j = 1, ..., m$. Si $x^* \in A$ est tel qu'il existe $(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ et $\mu_i \geq 0$ pour tout $i \in I(x^*)$ tel que:*

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*) - \sum_{i \in I(x^*)} \mu_i \nabla k_i(x^*). \tag{10.14}$$

alors pour $f(x^*) \leq f(x)$ pour tout $x \in A$.

Preuve:
Soit $x \in A$, par convexité de f, on a:

$$f(x) - f(x^*) \geq \nabla f(x^*) \cdot (x - x^*)$$

avec (10.14), il vient donc:

$$f(x) - f(x^*) \geq - \sum_{i \in I(x^*)} \mu_i \nabla k_i(x^*).(x - x^*) + \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*).(x - x^*). \tag{10.15}$$

Comme les g_j sont affines et $g_j(x) = g_j(x^*) = c_j$ on a aussi:

$$g_j(x) - g_j(x^*) = 0 = \nabla g_j(x^*).(x - x^*)$$

en reportant dans (9.9), il vient donc:

$$f(x^*) - f(x) \leq \sum_{i \in I(x^*)} \mu_i \nabla k_i(x^*).(x - x^*) \tag{10.16}$$

Pour $i \in I(x^*)$, $k_i(x^*) = d_i$ et puisque $x \in A$, on a $k_i(x) \leq d_i$, par convexité de k_i il vient alors:

$$0 \geq k_i(x) - k_i(x^*) \geq \nabla k_i(x^*).(x - x^*) \tag{10.17}$$

en reportant dans (10.16), il vient bien $f(x^*) \leq f(x)$. □

10.4 Lagrangian

On peut aussi formuler les conditions d'optimalité KKT, au moyen du lagrangien associé à (10.1). L'idée est d'introduire des multiplicateurs pour toutes les contraintes d'inégalité dans (10.10), pour tout i on a:

- soit $i \notin I(x^*)$ et $\mu_i = 0$ dans (10.10),
- soit $i \in I(x^*)$ et donc $k_i(x^*) - d_i$.

ce qu'on peut résumer par la condition de complémentarité entre multiplicateurs et contraintes qui exprime que si une contrainte n'est pas saturée le multiplicateurs associé est nul:

$$\mu_i(k_i(x^*) - d_i) = 0 \text{ pour tout } i \in \{1, ..., p\}. \tag{10.18}$$

Le lagrangien du problème (9.1) est la fonction définie pour tout $(x, \lambda, \mu) \in \Omega \times \mathbb{R}^m \times (\mathbb{R}_+)^p$ par:

$$\mathcal{L}(x, \lambda, \mu) := f(x) - \sum_{j=1}^{m} \lambda_j (g_j(x) - c_j) + \sum_{i=1}^{p} \mu_i (k_i(x) - d_i). \tag{10.19}$$

Remarquons alors que si f et les fonctions g_j sont différentiables en $x \in \Omega$ alors on a:

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = \nabla f(x) - \sum_{j=1}^{m} \lambda_j \nabla g_j(x) + \sum_{i=1}^{p} \mu_i \nabla k_i(x). \tag{10.20}$$

$$\partial_{\lambda_j} \mathcal{L}(x, \lambda, \mu) = c_j - g_j(x) \tag{10.21}$$

et

$$\partial_{\mu_i} \mathcal{L}(x, \lambda, \mu) = k_i(x) - d_i \tag{10.22}$$

Ainsi le fait que $x^* \in A$ i.e. vérifie la contrainte $g(x^*) = c$ peut s'exprimer par:

$$\nabla_\lambda \mathcal{L}(x^*, \lambda, \mu) = 0. \tag{10.23}$$

Avec (9.11) et (10.18), la condition de Kuhn et Tucker se traduit par:

$$\nabla_x \mathcal{L}(x^*, \lambda, \mu) = 0.$$

Le théorème 10.10 peut donc se reformuler comme suit:

Théorème 10.2 *Soit $x^* \in A$ une solution locale de (10.1) telle que:*

1. *f est différentiable en x^*,*
2. *g est de classe C^1 au voisinage de x^*,*
3. *k_i est différentiable en x^* pour tout $i \in I(x^*)$,*
4. *la famille $\nabla g_1(x^*), ..., \nabla g_m(x^*)$ est libre,*

5. les contraintes d'ínégalité sont qualifiées en x^*:

$$\exists h_0 \in \ker(g'(x^*)) \text{ tel que } \nabla k_i(x^*) \cdot h_0 < 0 \ \forall i \in I(x^*).$$

alors il existe $\lambda = (\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$ et $(\mu_1, ..., \mu_m) \in \mathbb{R}^m_+$ tels que:

$$\mu_i(k_i(x) - d_i) = 0 \ \forall i \in \{1, , p\} \tag{10.24}$$

$$\nabla_x \mathcal{L}(x^*, \lambda, \mu) = 0. \tag{10.25}$$

$$\nabla_\lambda \mathcal{L}(x^*, \lambda, \mu) = 0. \tag{10.26}$$

Chapter 11

Problems depending on a parameter

11.1 Continuous dependence and Berge's Theorem

Nous entamons ce chapitre par le théorème de Berge. Ce résultat de dépendance continue pour les problèmes d'optimisation dépendant d'un paramètre est très utile en pratique et pas uniquement en programmation dynamique. Dans la littérature, ce théorème est souvent appelé théorème du maximum, nous éviterons soigneusement cette terminologie pour éviter toute confusion: il existe déja deux principes du maximum (le principe de Pontriaguine en contrôle que nous verons plus tard et le principe du maximum pour les équations elliptiques) qui n'ont rien à voir entre eux et encore moins avec le théorème de Berge ci-dessous!

Théorème 11.1 *Soit X et Y deux métriques, F une correspondance* ***continue, à valeurs compactes, non vides*** *de X dans Y, $f \in C^0(X \times Y, \mathbb{R})$. Pour tout $x \in X$ soit:*

$$g(x) := \max_{y \in F(x)} f(x,y) \text{ et } M(x) := \{y \in F(x) \,:\, f(x,y) = g(x)\}.$$

Alors g est continue sur X et M est une correspondance à valeurs non vides, h.c.s..

Preuve:
Le fait que M est une correspondance à valeurs compactes non vides découle immédiatement de la continuité de f et du fait que $F(x)$ est compact non vide pour tout $x \in X$.

Montrons que g est continue. Soit donc x_n une suite de X convergeant vers x. Soit $z_n \in F(x_n)$ tel que $g(x_n) = f(x_n, z_n)$. Considérons une suite extraite (x_{n_j}, z_{n_j}) vérifiant

$$\lim_j f(x_{n_j}, z_{n_j}) = \limsup_n f(x_n, z_n) = \limsup_n g(x_n).$$

Comme F est h.c.s., quitte à extraire à nouveau, on peut supposer que z_{n_j} converge vers une limite $z \in F(x)$, ainsi $g(x) \geq f(x, z)$ et par continuité de f, on a:

$$\limsup_n g(x_n) = \lim_j f(x_{n_j}, z_{n_j}) = f(x, z) \leq g(x).$$

Soit maintenant $y \in F(x)$ tel que $g(x) = f(x, y)$, comme F est h.c.i., il existe $y_n \in F(x_n)$ telle que y_n converge vers y, comme $g(x_n) \geq f(x_n, y_n)$, il vient:

$$\liminf_n g(x_n) \geq \liminf f(x_n, y_n) = f(x, y) = g(x).$$

On a donc établi la continuité de g.

Il reste à établir que M est h.c.s.. Soit $x \in X$, x_n convergeant vers x dans X et $y_n \in M(x_n)$. Comme $M(x_n) \subset F(x_n)$ et F est h.c.s., il existe une sous-suite y_{n_j} convergeant vers une limite $y \in F(x)$. Par ailleurs pour tout j, on a $f(x_{n_j}, y_{n_j}) = g(x_{n_j})$, par continuité de f et g, en passant à la limite il vient $f(x, y) = g(x)$ i.e. $y \in M(x)$; M est donc h.c.s.. □

Remarquons que si en plus $M(x)$ est réduit à un point pour tout x alors $M(x)$ dépend continûment de x.

11.2 Envelope Theorems

Now we are interested in differentiating (when it is possible), the value of some optimization problems depending on a parameter (either in the objective or in the contraints, or both). Envelope theorems basically give conditions that guarantee some differentiability of the value and explicit formulas for the derivatives. These kinds of results are particularly useful in microeconomics.

Let us start with the following, let K be some compact metric space, let g be some continuous function $\mathbb{R}^d \times K$ and set

$$v(x) = \max_{y \in K} f(x, y), \forall x \in \mathbb{R}^d$$

let us further assume that

- for every $x \in \mathbb{R}^d$ there exists a unique $y(x) \in K$ such that $v(x) = f(x, y(x))$ (so that $x \to y(x)$ is continuous as a consequence of Berge's Theorem)
- for every $y \in K$, $f(., y)$ is differentiable and $\nabla_x f$ is continuous with respect to (x, y).

Theorem 11.1 *Under the assumptions above, the value function v is of class C^1 and one has*

$$\nabla v(x) = \nabla_x f(x, y(x)), \ \forall x \in \mathbb{R}^d \tag{11.1}$$

Proof:
Let $h \in \mathbb{R}^d \setminus \{0\}$, and for $t > 0$, let us set $x_t := x + th$, $y_t := y(x_t)$, $y_0 := y(x)$, we then have

$$\frac{1}{t}(v(x_t) - v(x)) = \frac{1}{t}(f(x_t, y_t) - f(x, y_0)) \geq \frac{1}{t}(f(x + th, y_0) - f(x, y_0))$$

so that

$$\liminf \frac{1}{t}(v(x_t) - v(x)) \geq \nabla_x f(x, y_0) \cdot h. \tag{11.2}$$

Similarly

$$\frac{1}{t}(v(x_t) - v(x)) = \frac{1}{t}(f(x_t, y_t) - f(x, y_0)) \leq \frac{1}{t}(f(x + th, y_t) - f(x, y_t))$$

by the mean-value Theorem, there exists $t' \in (0, 1)$ (depending on t) such that $f(x + th, y_t) - f(x, y_t) = t\nabla_x f(x_{t'}, y_t) \cdot h$, hence

$$\limsup \frac{1}{t}(v(x_t) - v(x)) \leq \limsup \nabla_x f(x_{t'}, y_t) \cdot h.$$

By continuity of $\nabla_x f$ we thus have

$$\limsup \frac{1}{t}(v(x_t) - v(x)) \leq \nabla_x f(x, y_0) \cdot h. \tag{11.3}$$

This proves that v is Gâteaux differentiable with Gâteaux derivative $x \to \nabla_x f(x, y(x))$, since this function is continuous we deduce that v is of class C^1 and that (11.1) holds. □

Remark. If K is some open subset of $\mathbb{R}^k$, if f is C^1 and there is a C^1 function $x \to y(x)$ such that $v(x) = f(x, y(x))$ (which is not so easy to check a priori), then the same conclusion as in Theorem 11.1 holds. To prove this variant, remark that $\nabla_y f(x, y(x)) = 0$ and then use the chain rule $v'(x) = f'_x(x, y(x)) + f'_y(x, y(x))y'(x)) = f'_x(x, y(x))$.

Now let us consider the constrained case, and more precisely let us define

$$v(x) := \max_{y \in \mathbb{R}^k} \{ f(x,y) \ : g(x,y) = 0 \}$$

where f and g are of class C^1, $g = g_1, ..., g_m$ and the gradients (with respect to y) of the constraints are linearly independent so that optimal y's satisfy the Lagrange conditions:

$$\nabla_y f(x,y) = \sum_{i=1}^{m} \lambda_i \nabla_y g_i(x,y)$$

for some (unique) Lagrange multipliers $\lambda_1, ..., \lambda_m$. Let us also assume for simplicity that in a neighbourhood of some $\overline{x}$ there is an optimal $y(x)$ and that the optimal mapping $x \to y(x)$ is of class C^1. We denote by $\lambda_1, ..., \lambda_m$ the Lagrange multipliers for the value $\overline{x}$ of the parameter

Theorem 11.2 *Under the assumptions above, v is differentiable at $\overline{x}$ and one has*

$$\nabla v(\overline{x}) = \nabla_x f(\overline{x}, y(\overline{x})) - \sum_{i=1}^{m} \lambda_i \nabla_x g_i(\overline{x}, y(\overline{x})).$$

Proof:
We differentiate $v(x) = f(x, y(x))$ to get first

$$v'(x) = f'_x(x, y(x)) + f'_y(x, y(x)) \circ y'(x) \tag{11.4}$$

by the Lagrange relation $f'_y(\overline{x}, y(\overline{x})) = \sum \lambda_i g'_{iy}(\overline{x}, y(\overline{x}))$ we then get

$$v'(\overline{x}) = f'_x(\overline{x}, y(\overline{x})) + \sum_i \lambda_i g'_{iy}(\overline{x}, y(\overline{x})) \circ y'(\overline{x}). \tag{11.5}$$

Differentiating the constraint $g_i(x, y(x)) = 0$ we get

$$g'_{ix}(x, y(x)) = -g'_{iy}(x, y(x)) \circ y'(x)) \tag{11.6}$$

replacing in (11.5) we then have

$$v'(\overline{x}) = f'_x(\overline{x}, y(\overline{x})) - \sum_i \lambda_i g'_{ix}(\overline{x}, y(\overline{x})).$$

□

In the special case $f(x,y) = f(y)$ and $g(x,y) = g(y) - x$ (think of a budget constraint), the previous result gives $\partial_{x_i} v(\overline{x}) = \lambda_i$: the multiplier gives the marginal impact of x_i (an increase on the budget, say) on the value. The theory of convex duality (see [6]) gives more general results of this kind and nice interpretations of multipliers in the framework of convex programming.

Part IV

Dynamic Optimization

Chapter 12

Problems in discrete time

12.1 Examples

12.1.1 Shortest path on a graph

Il s'agit ici du problème type de programmation dynamique en horizon fini avec espace d'état fini et qui revient à un problème d'optimisation sur un graphe, sa résolution illustre de manière simple le principe de la programmation dynamique. Considérons un voyageur de commerce qui doit se rendre de la ville A à la ville E en passant par plusieurs villes intermédiaires, les chemins possibles sont donc modélisés par un graphe ayant A et E pour sommets initial et final (les autres sommets représentant les villes étapes), les arrêtes de ce graphe représentant les trajets intermédiaires. On notera $\Gamma(M)$ les successeurs de la ville M et pour $N \in \Gamma(M)$ on notera MN le temps du parcours MN. Enfin, on donne: $\Gamma(A) = \{B, B'\}$, $AB = 1 = AB'$, $\Gamma(B) = \{C, C'\}$, $(BC, BC') = (2, 1)$, $\Gamma(B') = \{C', C''\}$, $(B'C', B'C'') = (2, 4)$, $\Gamma(C'') = \{D'\}$, $C''D' = 1$, $\Gamma(C) = \{D\}$, $CD = 1$, $\Gamma(C') = \{D, D'\}$, $(C'D, C'D') = (2, 1)$, $\Gamma(D) = \Gamma(D') = \{E\}$, $(DE, D'E) = (5, 2)$. Pour déterminer le ou les chemins les plus courts on pourrait bien sûr tous les essayer mais il est bien plus judicieux d'utiliser la remarque suivante(qui est précisément le *principe de la programmation dynamique* dans sa version la plus simple):
Si un chemin optimal de A à E passe par M alors il est encore optimal entre M et E.

Introduisons la *fonction valeur* $V(M)$:= "temps de parcours minimal entre M et E". Evidemment V se calcule facilement en partant de la fin puis en procédant par rétroaction arrière ou backward induction; on a d'abord

$$V(D) = 5,\ V(D') = 2$$

on remonte ensuite aux villes précédentes, le principe de la programmation

dynamique donne en effet:

$$V(C) = 6,\ V(C') = \min(1+V(D'), 2+V(D)) = 1+V(D') = 3,\ V(C'') = 3.$$

Réitérant l'argument, il vient:

$$\begin{aligned} V(B) &= \min(2 + V(C), 1 + V(C')) = 1 + V(C') = 4, \\ V(B') &= \min(2 + V(C'), 4 + V(C'')) = 5 \end{aligned}$$

et enfin

$$V(A) = \min(1 + V(B), 1 + V(B')) = 1 + V(B) = 5.$$

Le temps de parcours minimal est donc de 5 et correspond au seul parcours $ABC'D'E$.

Cet exemple pour élémentaire qu'il soit est instructif à plusieurs égards:

1. on voit aisément comment généraliser la stratégie précédente à des problèmes plus généraux de forme: introduire les fonctions valeurs aux différentes dates, les calculer "en partant de la fin" puis par backward induction en utilisant le principe de la programmation dynamique,

2. dans l'exemple précédent, on n'a pas essayé tous les chemins possibles mais seulement les chemins optimaux à partir de M qui ont ici tous été déterminés. De fait, les raisonnements précédents montrent par exemple que si le voyageur de commerce s'égare en B' (par lequel il n'est pas optimal de passer partant de A) alors par la suite il sera optimal de passer par $C'D'E$.

3. Il peut paraitre curieux alors qu'on s'est posé un seul problème (issu du point A) de chercher à résoudre tous les problèmes issus des points intermédiaires. Donnons deux arguments pour lever cette objection: tout d'abord la stratégie de résolution précédente est robuste (si une erreur est commise à un moment donné et conduit à passer par une ville non optimale M alors on peut se rattraper par la suite en suivant le chemin optimal à partir de M), ensuite cette stratégie est naturelle (choisir la ville suivante en fonction de la ville où on se trouve maintenant plutôt que de suivre un plan établi exactement à l'avance) et permet de se ramener à une succession de problèmes statiques.

12.1.2 One sector optimal growth

On considère une économie dans laquelle à chaque période un seul bien est produit servant à la fois à la consommation et à l'investissement. On note respectivement c_t, i_t, k_t, et y_t la consommation, l'investissement, le capital et la production de période t. On suppose que $y_t = F(k_t)$, F étant la fonction de prodution, et que le capital se déprécie au taux $\delta \in [0, 1]$. On a alors;

$$c_t + i_t = y_t = F(k_t), \text{ et } k_{t+1} = (1 - \delta)k_t + i_t$$

d'où l'on tire (en posant $f(k) := F(k) + (1 - \delta)k$):

$$c_t = f(k_t) - k_{t+1}.$$

On impose évidemment à c_t et k_t d'être positifs d'où la contrainte:

$$0 \leq k_{t+1} \leq f(k_t).$$

Finalement on suppose que l'économie maximise l'utilité intertemporelle:

$$\sum_{t=0}^{\infty} \beta^t u(c_t).$$

En fonction du capital ce problème devient:

$$\sup \left\{ \sum_{t=0}^{\infty} \beta^t u(f(k_t) - k_{t+1}) \right\}$$

sous les contraintes: k_0 donnée et $0 \leq k_{t+1} \leq f(k_t)$ pour $t \geq 0$. On peut généraliser le problème précédent au cas de plusieurs secteurs, au cas d'une offre de travail inélastique, à l'introduction du capital humain etc...

12.1.3 Optimal management of a forest

On considère une forêt qui initialement est de taille x_0, x_t est sa taille à la date t (variable d'état). Un exploitant choisit à chaque période un niveau de coupe v_t (variable de contrôle), l'évolution de la forêt est supposée régie par la dynamique:

$$x_{t+1} = H(x_t) - v_t.$$

En supposant que le prix du bois est constant égal à 1 et que le coût de l'abattage est C, le profit actualisé de l'exploitant s'écrit:

$$\sum_{t=0}^{\infty} \beta^t [v_t - C(v_t)].$$

En réecrivant ce profit en fonction de la variable d'état et en imposant $v_t \geq 0$ et $x_t \geq 0$, le programme de l'exploitant se réecrit sous la forme:

$$\sup \left\{ \sum_{t=0}^{\infty} \beta^t [H(x_t) - x_{t+1} - C(H(x_t) - x_{t+1}) \right\}$$

sous les contraintes: x_0 donnée et $0 \leq x_{t+1} \leq H(x_t)$ pour $t \geq 0$.

12.2 Finite horizon

On se propose d'étudier des problèmes de programmation dynamique en temps discret et en horizon fini:

$$\sup_{(x_t)} \left\{ \sum_{t=0}^{T-1} V_t(x_t, x_{t+1}) + V_T(x_T) \right\} \tag{12.1}$$

sous les contraintes: $x_0 = x \in A$ donné (autrement dit x est la condition initiale), $x_t \in A$ pour $t = 1, .., T$ et $x_{t+1} \in \Gamma_t(x_t)$ pour tout $t = 0, ..., T-1$, T s'appelle l'horizon du problème et l'ensemble A est appelé espace d'états, Γ_t est une correspondance de A (i.e. une application de A dans l'ensemble des parties de A on dit aussi une application "multivoque") qui modélise les contraintes sur la dynamique ($\Gamma_t(x_t)$ est l'ensemble des successeurs possibles de x_t), les fonctions $V_t : A \times A \to \mathbb{R}$ sont les payoffs de chaque période et enfin V_T: $A \to \mathbb{R}$ est la fonction de payoff terminale. Sans perte de généralité nous suposerons ici que $V_T = 0$.

Nous avons résolu un problème de type (12.1) au chapitre précédent. Nous allons voir dans ce chapitre, qui se veut aussi peu technique que possible, comment généraliser la stratégie de résolution du problème de plus court chemin du paragraphe 12.1.1.

On note graph(Γ_t) le graphe de la corespondance Γ_t:

$$\text{graph}(\Gamma_t) := \{(x, y) \in A \times A : \, y \in \Gamma_t(x)\}.$$

On supposera en outre que les correspondances Γ_t sont à valeurs non vides i.e. $\Gamma_t(x) \neq \emptyset$ pour tout $x \in A$.

Concernant l'existence de solutions, remarquons que si l'on suppose que A est un espace métrique compact, que pour $t = 0, ..., T-1$, que graph(Γ_t) est fermé (donc compact dans $A \times A$) et que $V_t \in C^0(\text{graph}(\Gamma_t), \mathbb{R})$, alors il est trivial que ces conditions assurent que (12.1) admet au moins une solution; ces conditions assurent aussi que $\Gamma_t(x)$ est un compact de A pour tout $x \in A$. Notons enfin que ces conditions sont *toujours* satisfaites dans le cas où l'espace d'états A est fini. Nous n'aurons cependant pas besoin dans ce qui suit de faire ces hypothèses de compacité.

12.2.1 Dynamic programming principle

Compte tenu de la structure récursive du problème il est judicieux d'introduire les fonctions-valeur aux différentes dates. Pour $x \in A$ on définit donc:

$$\begin{aligned} v(0,x) &:= \sup\left\{\sum_{t=0}^{T-1} V_t(x_t, x_{t+1}) : x_{t+1} \in \Gamma_t(x_t),\, x_0 = x\right\} \\ v(1,x) &:= \sup\left\{\sum_{t=1}^{T-1} V_t(x_t, x_{t+1}) : x_{t+1} \in \Gamma_t(x_t),\, x_1 = x\right\} \\ &\vdots \qquad\qquad \vdots \\ v(T-1,x) &:= \sup\left\{V_{T-1}(x, x_T) : x_T \in \Gamma_{T-1}(x)\right\}. \end{aligned}$$

et enfin $v(T,x) = V_T(x) = 0$.

Dans ce qui suit nous dirons qu'une suite $(x, x_1, ..., x_T) = (x_0, x_1, ..., x_T)$ est solution du problème $v(0,x)$ si cette suite est admissible (i.e. vérifie les contraintes du problème) et

$$v(0,x) := \sum_{t=0}^{T-1} V_t(x_t, x_{t+1}).$$

On étend la définition précédente aux problèmes aux différentes dates.

Le principe de la programmation dynamique s'exprime comme suit:

Proposition 12.1 *Soit $x \in A$; si $(x_0, x_1, ..., x_T) = (x, x_1, ..., x_T)$ est une solution du problème $v(0,x)$ alors pour tout $\tau = 1, ..., T-1$, la suite $(x_\tau, ..., x_T)$ est solution du problème $v(\tau, x_\tau)$.*

Preuve:
Par définition on a:

$$v(0,x) := \sum_{t=0}^{T-1} V_t(x_t, x_{t+1}). \tag{12.2}$$

Supposons que pour une date $\tau \in \{1, ..., T-1\}$, la suite $(x_\tau, ..., x_T)$ n'est pas solution du problème $v(\tau, x_\tau)$ alors il existe $(z_\tau, z_{\tau+1}, ...z_T) = (x_\tau, z_{\tau+1}, ...z_T)$ admissible pour le problème $v(\tau, x_\tau)$ telle que:

$$\sum_{t=\tau}^{T-1} V_t(x_t, x_{t+1}) < \sum_{t=\tau}^{T-1} V_t(z_t, z_{t+1}).$$

En définissant alors la suite (admissible pour $v(0,x)$) $(y_0, ..., y_T)$ par $(y_0, ..., y_T) = (x, x_1, ..., x_\tau, z_{\tau+1}, ..., z_T)$, on obtient avec (12.2):

$$v(0,x) < \sum_{t=0}^{T-1} V_t(y_t, y_{t+1})$$

ce qui contredit la définition même de $v(0,x)$.

□

Notons bien que dans la proposition, on a supposé l'existence d'une suite optimale. Sans faire cette hypothèse (et en autorisant les fonctions-valeur à prendre éventuellement la valeur $+\infty$), on obtient des relations fonctionnelles récursives (équations de Bellman) reliant les fonctions valeurs aux dates successives.

Proposition 12.2 *Soit $x \in A$, on a:*

$$v(0,x) = \sup\left\{V_0(x,y) + v(1,y) \,:\, y \in \Gamma_0(x)\right\} \tag{12.3}$$

De même pour $t \in \{1,...,T-1\}$:

$$v(t,x) = \sup\left\{V_t(x,y) + v(t+1,y) \,:\, y \in \Gamma_t(x)\right\}. \tag{12.4}$$

Preuve:
Evidemment, il suffit d'établir (12.3). Soit $y \in \Gamma_0(x)$ et $(y_1,...,y_T) = (y,...,y_T)$ telle que $y_{t+1} \in \Gamma_t(y_t)$ pour $t \geq 1$, la suite $(x,y_1,...,y_T)$ étant admissible pour $v(0,x)$ il vient:

$$v(0,x) \geq V_0(x,y) + \sum_{t=1}^{T-1} V_t(y_t, y_{t+1})$$

passant au supremum en $(y_2,...,y_T)$, puis en $y = y_1 \in \Gamma_0(x)$ dans le membre de droite il vient:

$$v(0,x) \geq \sup\left\{V_0(x,y) + v(1,y) \,:\, y \in \Gamma_0(x)\right\}.$$

Soit $\varepsilon > 0$ et $(x_0,x_1,...,x_T) = (x,x_1,...,x_T)$ admissible pour $v(0,x)$ telle que:

$$v(0,x) - \varepsilon \leq \sum_{t=0}^{T-1} V_t(x_t, x_{t+1})$$

On a ainsi:

$$\begin{aligned}\sup\{V_0(x,y) + v(1,y) \,:\, y \in \Gamma_0(x)\} &\geq V_0(x,x_1) + v(1,x_1) \\ &\geq \sum_{t=0}^{T-1} V_t(x_t, x_{t+1}) \geq v(0,x) - \varepsilon\end{aligned}$$

Comme $\varepsilon > 0$ est arbitraire on en déduit (12.3). □

12.2.2 Backward induction

En utilisant la proposition 12.2, et la relation terminale $v(T,x) = V_T(x)$ pour tout $x \in A$, il est possible (au moins en théorie mais aussi en pratique dans certaines applications), de calculer toutes les fonctions valeurs en partant de la date finale T (backward induction). En "remontant" les équations, on calcule d'abord $v(T-1,.)$:

$$v(T-1,x) = \sup\{V_{T-1}(x,y) : y \in \Gamma_{T-1}(x)\}$$

puis $v(T-2,.)$:

$$v(T-2,x) = \sup\{V_{T-2}(x,y) + v(T-1,y) : y \in \Gamma_{T-2}(x)\}$$

et ainsi de suite jusqu'à $v(0,.)$.

Admettons maintenant que l'on connaisse $v(0,.)$, ..., $v(T-1,.)$, il est alors très facile de caractériser les suites (ou politiques) optimales:

Proposition 12.3 *La suite $(x, x_1, ..., x_T)$ est solution de $v(0,x)$ si et seulement si pour $t = 0, .., T-1$, x_{t+1} est solution de:*

$$\sup_{y \in \Gamma_t(x_t)} \{V_t(x_t, y) + v(t+1, y)\} \tag{12.5}$$

Preuve:
Application immédiate des propositions 12.1 et 12.2. □

Notons qu'en pratique pour résoudre les équations de Bellman, on a souvent déja calculé les solutions des problèmes statiques apparaissant dans (12.3).

Il convient de bien retenir la démarche en deux étapes de ce chapitre:

1. on détermine les fonctions valeur par backward induction,
2. on détermine ensuite les politiques optimales (s'il en existe) en résolvant la suite de problèmes *statiques* (12.5) qui consistent à déterminer les successeurs optimaux x_1 de x_0 puis les successeurs optimaux de x_1 etc...

Enfin notons que la méthode présentée ici (la même que celle adoptée dans le problème du plus court chemin) est robuste car elle permet aussi de résoudre tous les problèmes intermédiaires posés à n'importe quelle date intermédiaire avec n'importe quelle condition initiale à cette date.

12.3 Infinite horizon

On considère désormais des problèmes d'horizon infini avec critère escompté du type:

$$v(x) := \sup \left\{ \sum_{t=0}^{\infty} \beta^t V(x_t, x_{t+1}) : x_0 = x,\, x_t \in A,\, x_{t+1} \in \Gamma(x_t)\ \forall t \geq 0 \right\}. \tag{12.6}$$

L'interprétation de A, Γ et V est la même que précédemment et $\beta \in]0,1[$ est un facteur d'escompte. La fonction $v(.)$ est la fonction *valeur* de (12.6), son argument est la condition initiale $x \in A$. Deux différences sont à noter avec le cas de l'horizon fini du chapitre précédent. Tout d'abord ici, le critère est la somme d'une série et les politiques optimales sont des suites (infinies), des précautions sont donc à prendre d'une part pour la définition même du critère mais surtout concernant l'existence de solutions. En outre, l'approche backward induction du chapitre précédent n'a pas de sens ici; c'est la raison pour laquelle on se limite ici à un cadre "plus stationnaire" (Γ ne dépend pas de t et payoff instantané de la forme $\beta^t V(x_t, x_{t+1})$) qu'au chapitre précédent.

Un élément important dans l'étude de (12.6) est le lien étroit entre v la valeur du problème et l'équation fonctionnelle suivante appelée équation de Bellman:

$$w(x) = \sup_{y \in \Gamma(x)} \{V(x,y) + \beta w(y)\} \tag{12.7}$$

12.4 Notations and assumptions

Dans tout ce chapitre, nous supposerons que A est un espace métrique compact, nous noterons d la distance sur A. Nous ferons l'hypothèse de non vacuité: $\Gamma(x) \neq \emptyset$ pour tout $x \in A$. Nous supposerons en outre que $V(.,.)$ est continue sur $A \times A$. Pour $x \in A$, nous noterons $\mathrm{Adm}(x)$ l'ensemble des suites admissibles issues de x:

$$\mathrm{Adm}(x) := \{\widetilde{x} = (x_t)_{t\geq 0} : x_0 = x,\, x_t \in A,\, x_{t+1} \in \Gamma(x_t)\ \forall t \geq 0\} \tag{12.8}$$

Pour $\widetilde{x} = (x_t)_{t\geq 0} \in \mathrm{Adm}(x)$ ou plus généralement $\widetilde{x} = (x_t)_{t\geq 0} \in A^{\mathbb{N}}$, on pose:

$$u(\widetilde{x}) := \sum_{t=0}^{\infty} \beta^t V(x_t, x_{t+1}).$$

Ainsi le problème (12.6) consiste à maximiser u sur $\mathrm{Adm}(x)$. Notons que comme V est bornée sur $A \times A$ et $\beta \in]0,1[$, u est bien définie et bornée sur $A^{\mathbb{N}}$ donc aussi sur $\mathrm{Adm}(x)$.

Nous aurons aussi besoin d'hypothèses de continuité sur la correspondance Γ (la dynamique), ceci nécessite les définitions suivantes:

Définition 12.1 *Soit X et Y deux espaces métriques et soit F une correspondance à valeurs* ***compactes non vides*** *de X dans Y, et soit $x \in X$ on dit que:*

1. *F est hémi-continue supérieurement (h.c.s.) en x si pour toute suite x_n convergeant vers x dans X et pour toute suite $y_n \in F(x_n)$, la suite y_n admet une valeur d'adhérence dans $F(x)$.*

2. *F est hémi-continue inférieurement (h.c.i.) en x si pour tout $y \in F(x)$ et pour toute suite x_n convergeant vers x dans X, il existe $y_n \in F(x_n)$ telle que y_n converge vers y dans Y.*

3. *F est continue si F hémi-continue supérieurement et inférieurement en chaque point de X.*

Dans le cas où X et Y sont des métriques compacts, dire que F est h.c.s. revient simplement à dire que son graphe:

$$\text{graph}(F) := \{(x, y) : x \in X, y \in F(x) \}$$

est fermé. Noter que dans ce cas F est automatiquement à valeurs compactes.

Remarquons que dans le cas *univoque* i.e. $F(x) = \{f(x)\}$ on a équivalence entre "F est h.c.s.", "F est h.c.i" et "f est continue". Si $X = Y = \mathbb{R}$ et $F(x) = [f(x), g(x)]$ avec f et g deux fonctions continues telles que $f \leq g$ alors F est une correspondance continue. Pour fixer les idées, il est bon d'avoir en mémoire les exemples suivants:

La correspondance F de $\mathbb{R}$ dans $\mathbb{R}$ définie par:

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ [0, 1] & \text{si} \quad x = 0 \\ 1 & \text{si} \quad x > 0 \end{cases}$$

est h.c.s. mais pas h.c.i. en 0.

La correspondance G de $\mathbb{R}$ dans $\mathbb{R}$ définie par:

$$G(x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ [-1, 1] & \text{si} \quad x > 0 \end{cases}$$

est quant à elle h.c.i. mais pas h.c.s. en 0.

Dans toute la suite, nous suposerons que Γ est une correspondance continue de A dans A.

12.4.1 Existence

Soit $A_\infty := A^{\mathbb{N}}$ l'ensemble des suites à valeurs dans A, munissons A_∞ de:

$$d_\infty(u,v) := \sum_{t=0}^{\infty} \frac{1}{2^t} d(u_t, v_t).$$

Il est clair que d_∞ est à valeurs finies et définit une distance sur A_∞. On a alors le résultat classique de compacité (le lecteur averti reconnaitra un corollaire du théorème de Tychonov) suivant:

Proposition 12.4 *(A_∞, d_∞) est compact.*

Preuve:
La démonstration est classique (compacité de A et extraction diagonale), le détail en est donc laissé au lecteur. □

Lemme 12.1 *Pour tout $x \in A$, $\mathrm{Adm}(x)$ est un compact de (A_∞, d_∞).*

Preuve:
Avec la proposition 12.4, il suffit de vérifier que $\mathrm{Adm}(x)$ est un fermé de (A_∞, d_∞). Soit donc $\widetilde{x}^n$ une suite de $\mathrm{Adm}(x)$ convergeant vers $\widetilde{x} = (x_t)_{t\geq 0} \in A_\infty$ pour la distance d_∞. Pour tout $t \in \mathbb{N}$, $\widetilde{x}_t^n$ converge vers x_t dans A quand $n \to +\infty$, en particulier $x_0 = x$. Comme Γ est de graphe fermé et que $(\widetilde{x}_t^n, \widetilde{x}_{t+1}^n) \in \mathrm{graph}(\Gamma)$ on en déduit que $x_{t+1} \in \Gamma(x_t)$ ce qui prouve finalement que $\mathrm{Adm}(x)$ est fermé. □

Lemme 12.2 *u est continue sur (A_∞, d_∞).*

Preuve:
Soit $\widetilde{x} = (x_t)_{t\geq 0} \in A_\infty$ et $\varepsilon > 0$. Comme V est continue et $A \times A$ compact, il existe $\tau \in \mathbb{N}$ tel que

$$\max_{A\times A} |V| \sum_{t=\tau}^{\infty} \beta^t \leq \frac{\varepsilon}{4} \tag{12.9}$$

Par continuité de V, il existe δ_0 tel que pour tout $(y,z) \in A \times A$ et tout $t \leq \tau - 1$ on ait:

$$d(x_t, y) + d(x_{t+1}, z) \leq \delta_0 \Rightarrow |V(x_t, x_{t+1}) - V(y,z)| \leq \frac{\varepsilon}{2\tau} \tag{12.10}$$

Ainsi en posant $\delta := \delta_0 2^{-\tau-1}$ pour tout $\widetilde{y} \in A_\infty$ tel que $d_\infty(\widetilde{x}, \widetilde{y}) \leq \delta$ on a $|u(\widetilde{x}) - u(\widetilde{y})| \leq \varepsilon$, ce qui achève la preuve. □

Des lemmes 12.1 et 12.2, on déduit le résultat d'existence:

Théorème 12.1 *Pour tout $x \in A$, il existe $\widetilde{x} \in \mathrm{Adm}(x)$ optimale i.e. telle que: $v(x) = u(\widetilde{x})$.*

12.4.2 The value function and Bellman's equation

On rappelle que la fonction valeur de (12.6) est définie pour tout $x \in A$ par:

$$v(x) := \sup\{u(\widetilde{x}) : \widetilde{x} \in \mathrm{Adm}(x)\}. \tag{12.11}$$

Les hypothèses de ce chapitre assurent que v est bornée sur A et le théorème 12.1 assure que le sup dans (12.11) est en fait un max. Par la suite nous dirons que $\widetilde{x}$ est solution du problème $v(x)$ ssi $\widetilde{x} \in \mathrm{Adm}(x)$ et $v(x) = u(\widetilde{x})$.

On laisse comme exercice, désormais de routine, au lecteur le soin de vérifier le principe de la programmation dynamique et le fait que v est solution de l'équation de Bellman:

Proposition 12.5 *Soit $x \in A$, on a:*

1. ***Principe de la programmation dynamique:*** *si $\widetilde{x} \in \mathrm{Adm}(x)$ est solution du problème $v(x)$ alors pour tout $\tau \geq 0$ la suite $(x_t)_{t\geq\tau}$ est solution du problème $v(x_\tau)$,*

2. *$v(.)$ est solution de l'**équation de Bellman:***

$$v(x) = \sup_{y\in\Gamma(x)} \{V(x,y) + \beta v(y)\}. \tag{12.12}$$

12.4.3 Blackwell's theorem

On se propose maintenant d'examiner dans quelle mesure l'équation de Bellman (12.12) caractérise la fonction valeur. Pour cela, il est utile de définir $B(A)$ comme l'ensemble des applications bornées de A dans $\mathbb{R}$. On rappelle que muni de la norme infinie ($\|f\|_\infty := \max\{|f(x)|,\ x \in A\}$), $B(A)$ est un espace de Banach et que $C^0(A,\mathbb{R})$ est un sous-espace fermé (donc complet) de $B(A)$. Pour $f \in B(A)$ et $x \in A$ on définit:

$$Tf(x) := \sup_{y\in\Gamma(x)} \{V(x,y) + \beta f(y)\}. \tag{12.13}$$

Il est facile de voir que $Tf \in B(A)$ ainsi T définit un opérateur de $B(A)$ dans lui-même. Le fait que la fonction-valeur v soit solution de l'équation de Bellman signifie exactement que $v = Tv$ autrement dit que v est un **point fixe** de T.

Le caractère contractant de T (donc en particulier l'unicité dans $B(A)$ de la solution de l'équation de Bellman) est assuré par le théorème de Blackwell:

Théorème 12.2 *Soit H un opérateur de $B(A)$ dans lui-même vérifiant les propriétés:*

1. *H est monotone i.e. : $\forall (f,g) \in B(A) \times B(A)$, $f(x) \leq g(x)$, $\forall x \in A$ $\Rightarrow Hf(x) \leq Hg(x)$, $\forall x \in A$,*

2. *il existe $\eta \in]0,1[$ tel que, pour toute constante positive a et tout $f \in B(A)$ on ait $H(f+a) \leq Hf + \eta a$,*

alors, H est une contraction de $B(A)$ de rapport η.

Preuve:
Soit $(f,g) \in B(A) \times B(A)$, on a $f \leq g + \|f-g\|_\infty$, ainsi les hypothèses sur H impliquent:

$$Hf \leq H(g + \|f-g\|_\infty) \leq Hg + \eta \|f-g\|_\infty$$

inversant les rôles de f et g et en passant au sup en $x \in A$, il vient bien:

$$\|Hf - Hg\|_\infty \leq \eta \|f-g\|_\infty.$$

□

En remarquant que T vérifie les conditions du théorème de Blackwell avec $\eta = \beta$, et en utilisant le théorème du point fixe pour les contractions, on en déduit immédiatement:

Corollaire 12.1 *L'équation de Bellman 12.12 admet une unique solution qui est la fonction valeur définie par (12.11). De plus pour tout $f \in B(A)$, v est limite uniforme de la suite des itérées $T^n f$.*

L'équation de Bellman caractérise donc bien la fonction valeur: v est l'unique solution bornée de (12.12). On peut être plus précis en remarquant que T est aussi un opérateur sur les fonctions continues et par conséquent, le point fixe de T est une fonction continue.

Proposition 12.6 *Pour tout $f \in C^0(A,\mathbb{R})$, $Tf \in C^0(A,\mathbb{R})$. Ceci implique que en particulier que la fonction-valeur v est continue.*

Preuve:
La première partie du résultat précédent est une conséquence immédiate du théorème de Berge. Prouvons la seconde partie du résultat: T est une contraction de $C^0(A,\mathbb{R})$ qui est complet donc T admet un unique point fixe dans $C^0(A,\mathbb{R})$, or nous savons que l'unique point fixe de T (dans $B(A)$) est v on a donc $v \in C^0(A,\mathbb{R})$. □

Remarque: On peut établir directement (i.e. sans utiliser l'équation de Bellman ni le théorème de Berge) la continuité de v (le lecteur pourra vérifier cette affirmation sans difficulté, l'exercice étant cependant un peu fastidieux).

12.4.4 Back to optimal policies

Comme dans le cas de l'horizon fini, connaitre la fonction valeur permet de calculer les stratégies optimales. Il est en effet clair (s'en persuader) que $\widetilde{x} = (x_t)_{t\geq 0} \in \mathrm{Adm}(x)$ est solution de $v(x)$ ssi pour tout $t \geq 0$, x_{t+1} résout le problème statique:

$$\max_{y\in\Gamma(x_t)} \{V(x_t, y) + \beta v(y)\}.$$

Ainsi pour déterminer les politiques optimales on détermine d'abord v en résolvant l'équation de Bellman. On définit alors la correspondance:

$$M(x) := \{y \in \Gamma(x) : \ v(x) = V(x, y) + \beta v(y)\}.$$

$M(x)$ s'interprète naturellement comme l'ensemble des successeurs optimaux de x, et les politiques optimales issues de x sont simplement les itérées de cette correspondance.

Chapter 13

Calculus of variations

13.1 Introduction

On s'intéresse désormais à des problèmes de calcul des variations (en horizon fini pour simplifier). De tels problèmes consistent à maximiser un critère du type:

$$J(x) = \int_0^T L(t, x(t), \dot{x}(t))dt + g(x(T)) + f(x(0))$$

dans un ensemble de fonctions de $[0, T]$ dans $\mathbb{R}^n$ jouissant de certaines propriétés de différentiabilité. Le bon cadre fonctionnel est celui des espaces de Sobolev mais pour ne pas alourdir l'exposé ni décourager le lecteur qui ne serait pas familier de ces espaces, nous nous limiterons par la suite aux fonctions de classe C^1 ou "continues et C^1 par morceaux".

La fonction $(t, x, v) \mapsto L(t, x, v)$ est appelée Lagrangien, et on supposera toujours $L \in C^0([0, T] \times \mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$, g (respectivement f) est la fonction de gain terminal (respectivement initial); on supposera $g \in C^0(\mathbb{R}^n, \mathbb{R})$ (respectivement $f \in C^0(\mathbb{R}^n, \mathbb{R})$).

Une variante est le problème à conditions aux limites prescrites:

$$\sup \left\{ \int_0^T L(t, x(t), \dot{x}(t))dt : x \in C^1([0, T], \mathbb{R}^n),\ x(0) = x_0, x(T) = x_T \right\}.$$

Evidemment on peut aussi considérer le cas d'une extrémité libre et d'une extrémité prescrite. Nous n'écrirons pas les conditions d'optimalité pour tous les cas possibles, les cas "manquants" seront laissés en exercice au lecteur...

Historiquement, le calcul des variations, s'est développé depuis le 17^e siècle (problème du brachistochrone résolu par Bernoulli) conjointement au développement de la physique (la mécanique en particulier, mais aussi le

problème de la résistance minimale posé par Newton dans ses *Principia* et qui reste encore largement ouvert aujourd'hui..) et de la géométrie (problèmes de géodésiques ou d'applications harmoniques par exemple). Quelques grands noms parmi les mathématiciens des trois siècles passés ont marqué son développement: Euler, Lagrange, Hamilton, Jacobi, Legendre, Weierstrass, Noether, Carathéodory... Son usage en économie est plus récent, il devient véritablement populaire à partir des années 1960 dans les modèles de croissance, d'investissement, de gestion de stocks et, plus récemment, en théorie des incitations ou des enchères. En finance, il est aussi d'usage courant d'utiliser des modèles en temps continu, les dynamiques réalistes dans ce cadre ayant un caractère aléatoire, c'est plutôt le contrôle stochastique qui est utilisé.

13.2 Existence

Résoudre un problème de calcul des variations c'est résoudre un problème d'optimisation dans un espace fonctionnel de dimension infinie. L'existence de solutions n'a donc rien d'évident a priori et je tiens à mettre en garde le lecteur sur ce point. Il ne s'agit pas de faire ici une théorie de l'existence, pour cela on consultera par exemple le livre d' I.Ekeland et R.Temam [**?**] mais d'indiquer que la plupart des résultats d'existence demandent la concavité (si on maximise; la convexité si on minimise) du lagrangien par rapport à la variable v. Examinons maintenant un contre-exemple classique dû à Bolza:

$$\inf J(x) := \int_0^1 [(\dot{x}(t)^2 - 1)^2 + x^2(t)]dt : \ x(0) = x(1) = 0. \tag{13.1}$$

On se propose de montrer que l'infimum de ce problème est 0 et qu'il n'est pas atteint. Soit $u_0(t) := 1/2 - |t - 1/2|$ pour $t \in [0, 1]$, prolongeons u_0 à $\mathbb{R}$ par périodicité. Enfin pour $n \in \mathbb{N}^*$ soit $u_n(t) := u_0(nt)/n$, u_n vérifie les conditions aux limites du problème, $\dot{u_n}^2 = 1$ presque partout sur $[0, 1]$ et $|u_n| \leq 1/2n$ donc $J(u_n)$ tend vers 0. On en déduit donc que l'infimum de (13.1) est 0. Supposons que $J(u) = 0$ avec $u(0) = u(1) = 0$. Par définition de J on devrait avoir à la fois $u = 0$ et $\dot{u} \in \{-1, 1\}$ presque partout, ce qui est évidemment impossible.

13.3 Euler-Lagrange equations and transversality conditions

Considérons le problème:

$$\sup_{x\in C^1([0,T],\mathbb{R}^n)} J(x) = \int_0^T L(t,x(t),\dot{x}(t))dt + g(x(T)) + f(x(0)) \tag{13.2}$$

On suppose dans tout ce paragraphe que les fonctions $(t,x,v) \mapsto L(t,x,v)$, $x \mapsto g(x)$ et $x \mapsto f(x)$ sont de classe C^1. Pour $i = 1,...,n$, nous noterons L_{v_i}, L_{x_i} les dérivées partielles $\frac{\partial L}{\partial v_i}$ et $\frac{\partial L}{\partial v_i}$ et $\nabla_v L$ et $\nabla_x L$ les gradients partiels de L par rapport à x et v respectivement (i.e. $\nabla_v L = (L_{v_1},...,L_{v_n})'$, $\nabla_x L = (L_{x_1},...,L_{x_n})'$).

Proposition 13.1 *Soit $x \in C^1([0,T],\mathbb{R}^n)$, alors si x est solution de (13.2), on a*

1. *x est solution des équations d'Euler-Lagrange:*

$$\frac{d}{dt}[\nabla_v L(t,x(t),\dot{x}(t))] = \nabla_x L(t,x(t),\dot{x}(t)) \tag{13.3}$$

2. *x vérifie les conditions de transversalité:*

$$\nabla_v L(0,x(0),\dot{x}(0)) = f'(x(0)),\ \nabla_v L(T,x(T),\dot{x}(T)) = -g'(x(T)). \tag{13.4}$$

3. *Si on suppose en outre que g et f sont concaves sur $\mathbb{R}^n$ et que pour tout $t \in [0,T]$, $L(t,.,.)$ est concave sur $\mathbb{R}^n$, alors si $x \in C^1([0,T],\mathbb{R}^n)$ vérifie les équations d'Euler-Lagrange (13.3) et les conditions de transversalité (13.4) alors x est solution de (13.2).*

Preuve:

Pour $h \in C^1([0,T],\mathbb{R}^n)$ on a d'abord:

$$\lim_{t\to 0^+} \frac{1}{t}[J(x+th) - J(x)] \leq 0. \tag{13.5}$$

En utilisant la formule des accroissements finis et le théorème de convergence dominée de Lebesgue, on obtient facilement que la limite précédente vaut:

$$\begin{aligned}\int_0^T [\nabla_x L(t,x(t),\dot{x}(t)).h(t) + \nabla_v L(t,x(t),\dot{x}(t)).\dot{h}(t)]dt \\ +g'(x(T)).h(T) + f'(x(0)).h(0)\end{aligned} \tag{13.6}$$

En utilisant (13.5), (13.6) et la transformation $h \mapsto -h$, il vient que pour tout $h \in C^1([0,T],\mathbb{R}^n)$ on a:

$$\begin{aligned}\int_0^T [\nabla_x L(t,x(t),\dot{x}(t)).h(t) + \nabla_v L(t,x(t),\dot{x}(t)).\dot{h}(t)]dt \\ +g'(x(T)).h(T) + f'(x(0)).h(0) = 0\end{aligned} \tag{13.7}$$

Soit

$$E_n := \{h \in C^1([0,T],\mathbb{R}^n) : h(0) = h(T) = 0\} \tag{13.8}$$

En prenant $h \in E_n$ (13.7), et en raisonnant coordonnée par coordonnée, on obtient ainsi que pour tout $i = 1, ..., n$ et tout $h \in E_1$ on a:

$$\int_0^T [L_{x_i}(t,x(t),\dot{x}(t))h(t) + L_{v_i}(t,x(t),\dot{x}(t)).\dot{h}(t)]dt = 0 \tag{13.9}$$

Le Lemme de Dubois-Reymond rappelé plus bas implique donc que pour tout $i = 1, ..., n$, on a:

$$\frac{d}{dt}[L_{v_i}(t,x(t),\dot{x}(t))] = L_{x_i}(t,x(t),\dot{x}(t)) \tag{13.10}$$

on a donc établi (13.3). Soit maintenat $h \in C^1([0,T],\mathbb{R}^n$, en utilisant (13.3), et en intégrant par parties (13.7), on obtient ainsi:

$$(\nabla_v L(T,x(T),\dot{x}(T))+g'(x(T)))h(T)+(f'(x(0))-\nabla_v L(0,x(0),\dot{x}(0)))h(0) = 0 \tag{13.11}$$

on déduit ainsi aisément (13.4) de l'arbitrarité de h dans (13.11).

Il nous reste à vérifier que (13.3) et (13.4) sont suffisantes dans le cas concave. Soit $x \in C^1([0,T],\mathbb{R}^n)$ qui vérifie les équations d'Euler-Lagrange (13.3) et les conditions de transversalité (13.4) et $y \in C^1([0,T],\mathbb{R}^n)$. Par concavité on a:

$$\begin{aligned}J(y) - J(x) \leq \int_0^T [\nabla_x L(t,x(t),\dot{x}(t)).(y(t)-x(t))]dt \\ + \int_0^T [\nabla_v L(t,x(t),\dot{x}(t)).(\dot{y}(t)-\dot{x}(t))]dt \\ +g'(x(T)).(y(T)-x(T)) + f'(x(0))(y(0)-x(0)) \\ = 0\end{aligned}$$

la dernière égalité est obtenue en intégrant par parties et en utilisant (13.3) et (13.4).

□

Il nous reste à établir le Lemme de Dubois-Reymond:

Lemme 13.1 *Soit ϕ et ψ dans $C^0([0,T],\mathbb{R})$ et E_1 définie par (13.8), on a alors équivalence entre:*

1. *ψ et de classe C^1 et $\dot{\psi} = \phi$,*
2. *pour tout $h \in E_1$:*

$$\int_0^T (\phi h + \psi \dot{h}) = 0.$$

Preuve:
Pour démontrer 1. $\Rightarrow$ 2, il suffit d'intégrer par parties. Démontrons 2. $\Rightarrow$ 1.. Soit F une primitive de ϕ, l'hypothèse s'écrit alors:

$$\int_0^T (\psi - F)\dot{h} = 0, \, \forall h \in E_1.$$

Soit $c := T^{-1} \int_0^T (\psi - F)$ on a:

$$\int_0^T (\psi - F - c)\dot{h} = 0, \, \forall h \in E_1. \tag{13.12}$$

Il suffit de remarquer que la fonction $h(t) := \int_0^t (\psi - F - c)$ appartient à E_1, avec (13.12) il vient donc $\psi = F + c$ ce qui achève la preuve par construction de F. $\square$

13.4 An economic example

On se place en temps continu sur la période $[0,T]$ et on considère un ménage dont on note $x(t)$, $S(t)$, $c(t)$ et $e(t)$ la richesse, le salaire instantané (exogène), la consommation et l'épargne enfin on suppose que le ménage cherche à maximiser l'utilité:

$$\int_0^T e^{-\delta t} \log(c(t))dt + e^{-\delta T} V(x(T)).$$

On a les relations:

$$S(t) = c(t) + e(t), \;\; \dot{x}(t) = e(t) + rx(t)$$

avec r le taux d'intérêt exogène (et supposé constant pour simplifier). La richesse initiale du ménage x_0 étant donnée, le choix optimal consommation-épargne du ménage se ramène ainsi au problème variationnel:

$$\sup J(x) := \int_0^T e^{-\delta t} \log(S(t) + rx(t) - \dot{x}(t))dt + e^{-\delta T} V(x(T)) : \; x(0) = x_0 \tag{13.13}$$

En supposant en outre que V est concave, les conditions du premier ordre sont des conditions suffisantes d'optimalité. En posant $c(t) = S(t) + rx(t) - \dot{x}(t)$, l'équation d'Euler-Lagrange s'écrit ici:

$$-\frac{d}{dt}(\frac{e^{-\delta t}}{c(t)}) = \frac{re^{-\delta t}}{c(t)}$$

en posant $y(t) = e^{-\delta t}/c(t)$ il vient donc: $y(t) = e^{-rt}/c(0)$ et donc:

$$c(t) = e^{(r-\delta)t}c(0)$$

Il reste à déterminer la constante $c(0)$, pour cela il faut d'une part revenir à la variable (d'état) x, en intégrant $\dot{x} - rx = S - c$ d'autre part utiliser la condition de transversalité en T qui ici s'écrit:

$$V'(x(T)) = \frac{1}{c(T)}.$$

Chapter 14

Optimal control

14.1 Introduction

In this chapter, we shall study the following problem:

$$\sup_{u\in\mathcal{U}} J(u) := \int_0^T L(s, y_u(s), u(s))ds + g(y_u(T))$$

over some suitable class $\mathcal{U}$ of functions u (the control) and y_u (the state) is (indirectly) related to u via the (controlled) differential equation called the state equation

$$\dot{x}(t) = f(t, x(t), u(t)),\ x(0) = x_0.$$

Here the intial condition $x_0 \in \mathbb{R}^d$ is given, L is called the Lagrangian (or running criterion) and g the terminal criterion. In the sequel, we shall assume that K is a given compact metric space and denote by $\mathcal{U}$ the set of measurable functions from $[0, T]$ to K. The function f gives the dynamic of the system it is defined on $[0, T] \times \mathbb{R}^d \times K$ with values in the state space $\mathbb{R}^d$.

Of course, the optimal control problem above may be seen as a generalization of calculus of variations problems studied in the previous chapter: we will consider here more general state equations than the (very) special case $\dot{x} = u$ and we will be able to treat quite easily the case of pointwise constraints on the control.

14.2 Controlled differential equations

We are going to prove, that under natural assumptions on the dynamic f , for every admissible control function u (and initial condition $x_0 \in \mathbb{R}^d$), the

following state equation has a unique solution

$$\dot{x}(t) = f(t, x(t), u(t)),\ x(0) = x_0 \tag{14.1}$$

Let us assume that f is continuous $[0,T] \times \mathbb{R}^d \times K \to \mathbb{R}^d$ and satisfies the Lipschitz condition: there exists $M > 0$ such that

$$|f(t,x,u) - f(t,y,u)| \leq M|x-y|,\ \forall (t,x,y,u) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times K. \tag{14.2}$$

Under these assumptions, we have the following form of the Cauchy-Lipschitz Theorem

Theorem 14.1 *Under the assumptions above, for every control $u \in \mathcal{U}$ and initial state x_0, (14.1) admits a unique solution, simply denoted y_u.*

Proof:
Let us equip $E := C^0([0,T], \mathbb{R}^d)$ with the norm:

$$\|x\|_\lambda := \sup_{t \in [0,T]} e^{-\lambda t}|x(t)|$$

(λ is a real parameter that will be chosen later on). It is easy, to check that E equipped with this norm is a Banach Space. Now, we rewrite (14.1) in integral form as a fixed point problem $x = Tx$ where $T : E \to E$ is defined by

$$Tx(t) := x_0 + \int_0^t f(s, x(s), u(s)),\ \forall t \in [0,T],\ \forall x \in E.$$

Let x and y be in E and $t \in [0,T]$, using (14.2), we have

$$|Tx(t) - Ty(t)| \leq M \int_0^t |x(s) - y(s)| ds \leq M \int_0^t \|x-y\|_\lambda e^{\lambda s} ds \leq \frac{M}{\lambda} \|x-y\|_\lambda e^{\lambda t}.$$

so that

$$\|Tx - Ty\|_\lambda \leq \frac{M}{\lambda} \|x - y\|_\lambda$$

and then T is a strict contraction as soon as $\lambda > M$. Existence and uniqueness then follows from Banach's fixed point theorem. □

Let us also recall the following version of the classical Gronwall's Lemma

Lemme 14.1 *Let $x \in C^0([0,T], \mathbb{R}^d)$ satisfy for some constants $a \geq 0$ and $b \geq 0$*

$$|x(t)| \leq a + b \int_0^t |x(s)| ds, \forall t \in [0,T] \tag{14.3}$$

then

$$|x(t)| \leq a e^{bt}, \forall t \in [0,T]. \tag{14.4}$$

Proof:
Set $y(t) = \int_0^t |x(s)|ds$, we then have $\dot{y} - by \leq a$. Multiplying by e^{-bt} we then have

$$\frac{d}{dt}(y(t)e^{-bt} + \frac{a}{b}e^{-bt}) \leq 0$$

so that

$$y(t) \leq y(0)e^{bt} + \frac{a}{b}e^{bt} - \frac{a}{b} = \frac{a}{b}e^{bt} - \frac{a}{b}$$

with (14.3), we then get

$$|x(t)| \leq a + by(t) \leq ae^{bt}.$$

□

Let $u \in \mathcal{U}$ and denote $x = y_u$, we then have

$$\begin{aligned} |x(t)| &\leq |x_0| + \int_0^t |f(s, x(s), u(s)) - f(s, 0, u(s))|ds + \int_0^t |f(s, 0, u(s))|ds \\ &\leq |x_0| + T \max_{(s,u)\in[0,T]\times K} |f(s, 0, u)| + M \int_0^t |x(s))|ds \end{aligned}$$

we deduce from the previous inequality and Gronwall's Lemma that there is a constant C such that

$$|y_u(t)| \leq C, \ \forall u \in \mathcal{U}, \ \forall t \in [0, T]. \tag{14.5}$$

14.3 Pontryagin's principle

In the complete proof of Pontryagin's principle, we shall use Ascoli's compactness theorem that we now state and prove:

Theorem 14.2 *Let $\mathcal{F}$ be a subset of $C^0([0, T], \mathbb{R}^d)$ such that:*

$$\exists M \ : \ |f(t)| \leq M, \ \forall t \in [0, T], \ \forall f \in \mathcal{F} \tag{14.6}$$

and

$$\forall \varepsilon > 0, \ \exists \delta > 0 \ : \ |f(t) - f(s)| \leq \varepsilon, \ \forall f \in \mathcal{F}, \forall \ t, s, \text{ such that } |t - s| \leq \delta \tag{14.7}$$

then $\mathcal{F}$ is relatively compact in $C^0([0, T], \mathbb{R}^d)$ for the uniform norm.

Proof:
Let $(f_n) \in \mathcal{F}^{\mathbb{N}}$, we have to prove that it has a subsequence that converges uniformly. Let us denote by $(t^k)_k$ the dense sequence in $[0, T]$ consisting of

all points of the form mT/p with $m \leq p$ and m and p integers. For each k, the sequence $(f_n(t^k))$ is bounded hence admits a convergent subsequence. By a standard diagonal extraction argument, there is a subsequence (again denoted f_n) such that $(f_n(t^k))$ converges for every k, to some limit denoted $g(t^k)$. It follows from (14.7) that

$$\forall \varepsilon > 0,\ \exists \delta > 0\ :\ |g(t^k) - g(t^l)| \leq \varepsilon,\ \forall\ k, l, \text{ such that } |t^k - t^l| \leq \delta. \quad (14.8)$$

If $t \in [0, T]$ and $(t^{k_n})_n$ converges to t, it follows from (14.8) that $g(t^{k_n})$ is a Cauchy sequence hence has some limit, moreover, again by (14.8), this limit does not depend on the approximating sequence (t^{k_n}), we then simply denote by g this limit (which is continuous by (14.8)). Let $\varepsilon > 0$, there is a $\delta > 0$ such that for every t and s such that $|t - s| \leq \delta$, one has

$$|f_n(t) - f_n(s)| \leq \varepsilon/3,\ |g(t) - g(s)| \leq \varepsilon/3.$$

Let p be such that $T/p \leq \delta$, and N be large enough so that for all $n \geq N$ and all $m = 0, ..., p$ one has $|f_n(mT/p) - g(mT/p)| \leq \varepsilon/3$. Now let $n \geq N$ and $t \in [0, T]$, and let m be such that $|t - mT/p| \leq \delta$, we then have

$$\begin{aligned} |f_n(t) - g(t)| \leq & |f_n(t) - f_n(mT/p)| + |f_n(mT/p) - g(mT/p)| + |g(mT/p) - g(t)| \\ \leq & \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

This prove s that f_n converges uniformly to g. □

Our aim now is to give necessary optimality conditions for the optimal control problem

$$\sup_{u \in \mathcal{U}} J(u) := \int_0^T L(t, y_u(t), u(t))dt + g(y_u(T)) \quad (14.9)$$

In addition to the assumptions of the previous paragraph, we assume:

- L is continuous, differentiable with respect to x and $\nabla_x L$ is continuous with respect to (t, x, u),
- g is of class C^1
- f is differentiable with respect to x and $D_x f$ is continuous with respect to (t, x, u).

Let us assume that $\overline{u}$ is an optimal control (i.e. solves (14.9)) and $\overline{x} = y_{\overline{u}}$ is the corresponding state. For the sake of simplicity, we also assume that $\overline{u}$ is *piecewise* continuous (i.e. that there exist finitely many times $0 = t_0 < t_1 < .... < t_k = T$ such that $\overline{u}$ is continuous on every interval (t_i, t_{i+1}))

which implies that $\overline{x}$ is *piecewise* C^1 (i.e. continuous and with a piecewise continuous derivative).

Now let $t \in (0,T)$ be a continuity point of $\overline{u}$ and let $v \in K$ be an arbitrary admissible control. For $\varepsilon \in (0,t)$, let us then define

$$u_\varepsilon(s) = \begin{cases} v & \text{if } s \in (t-\varepsilon, t] \\ \overline{u}(s) & \text{otherwise.} \end{cases}$$

and denote by $x_\varepsilon := y_{u_\varepsilon}$ the associated state.

Lemma 14.1 *Let us define $z_\varepsilon := \varepsilon^{-1}(x_\varepsilon - \overline{x})$ then z_ε is bounded, z_ε converges pointwise to $z = 0$ on $[0,t)$ and z_ε converges uniformly on $[t,T]$ to the function z that solves the linearized equation:*

$$\dot{z}(s) = D_x f(s, \overline{x}(s), \overline{u}(s)) z(s) \text{ on } (t,T],\ z(t) = f(t, \overline{x}(t), v) - f(t, \overline{x}(t), \overline{u}(t)). \tag{14.10}$$

Proof:

First, by construction $z_\varepsilon = 0$ on $[0, t-\varepsilon)$ so that z_ε converges to 0 on $[0,t)$.

Step 1: z_ε is uniformly bounded.

For $s \geq t$, we have:

$$z_\varepsilon(s) = I(\varepsilon) + \frac{1}{\varepsilon}\int_t^s [f(\theta, x_\varepsilon(\theta), \overline{u}(\theta)) - f(\theta, \overline{x}(\theta), \overline{u}(\theta))] d\theta \tag{14.11}$$

where

$$I(\varepsilon) = \frac{1}{\varepsilon}\int_{t-\varepsilon}^t [f(\theta, x_\varepsilon(\theta), v) - f(\theta, \overline{x}(\theta), \overline{u}(\theta))] d\theta.$$

Thanks to (14.5), the integrand in $I(\varepsilon)$ is bounded and then $|I(\varepsilon)| \leq M_0$ for some constant M_0. With (14.2), we thus have

$$|z_\varepsilon(s)| \leq M_0 + \frac{1}{\varepsilon}\int_t^s |f(\theta, x_\varepsilon(\theta), \overline{u}(\theta)) - f(\theta, \overline{x}(\theta), \overline{u}(\theta))| d\theta \leq M_0 + M\int_t^s |z_\varepsilon|$$

with Gronwall's Lemma, we deduce that z_ε is bounded:

$$|z_\varepsilon(s)| \leq M_0 e^{M(s-t)}, \forall s \in [t,T].$$

To shorten notations, we denote by M_1 a constant such that $|x_\varepsilon - \overline{x}| \leq M_1 \varepsilon$ on $[0,T]$.

Step 2: Convergence of $z_\varepsilon(t)$.

Let us write

$$\begin{aligned} z_\varepsilon(t) = &\frac{1}{\varepsilon}\int_{t-\varepsilon}^t [f(s, \overline{x}(s), v) - f(s, \overline{x}(s), \overline{u}(s))] ds \\ &+ \frac{1}{\varepsilon}\int_{t-\varepsilon}^t [f(s, x_\varepsilon(s), v) - f(s, \overline{x}(s), v)] ds \end{aligned}$$

The second term is bounded by $MM_1\varepsilon$ hence converges to 0. Since t is a continuity point of $\overline{u}$, we deduce that

$$\lim_{\varepsilon\to 0^+} z_\varepsilon(t) = f(t,\overline{x}(t),v) - f(t,\overline{x}(t),\overline{u}(t)). \tag{14.12}$$

Step 3: z_ε is equi-Lipschitz on $[t,T]$.
Let t_1 and t_2 be such that $T \geq t_2 \geq t_1 \geq t$, we then have

$$|z_\varepsilon(t_2) - z_\varepsilon(t_1)| \leq \frac{M}{\varepsilon}\int_{t_1}^{t_2} |z_\varepsilon(s)|ds \leq MM_1(t_2-t_1). \tag{14.13}$$

The family z_ε is then equi-Lipschitz on $[t,T]$. With Ascoli's theorem (see above) we then deduce that the family $(z_\varepsilon)_\varepsilon$ is precompact in $C^0([t,T],\mathbb{R}^d)$.

Step 4: z_ε converges on $[t,T]$ to the solution of the linearized equation (14.10).
Thanks to Step 3, there is a sequence $\varepsilon_n \to 0$ such that $z_n := z_{\varepsilon_n}$ converges uniformly to some $z \in C^0([t,T])$. Let t_1 and t_2 be such that $T \geq t_2 \geq t_1 > t$, we have

$$z_n(t_2) - z_n(t_1) = \frac{1}{\varepsilon_n}\int_{t_1}^{t_2} [f(s,\overline{x}+\varepsilon_n z_n,\overline{u}(s)) - f(s,\overline{x}(s),\overline{u}(s)]ds. \tag{14.14}$$

Thanks to the differentiability of L and Lebesgue's dominated convergence theorem, passing to the limit in (14.14), yields

$$z(t_2) - z(t_1) = \int_{t_1}^{t_2} [D_x f(s,\overline{x}(s),\overline{u}(s))z(s)]ds.$$

This proves that z solves (14.10). Finally, the system (14.10) admits z as unique solution (Cauchy-Lipschitz), together with the relative compactness of the family z_ε, this implies that the whole family z_ε converges uniformly to z on $[t,T]$ as $\varepsilon \to 0^+$ and the proof is complete. □

Now we use the optimality of $\overline{u}$ to deduce:

$$0 \geq \frac{1}{\varepsilon}(J(u_\varepsilon) - J(\overline{u})) = \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} [L(s,x_\varepsilon,v) - L(s,\overline{x},\overline{u}))]ds$$
$$+\frac{1}{\varepsilon}\int_t^T [L(s,x_\varepsilon,\overline{u}) - L(s,\overline{x},\overline{u}))]ds + \frac{1}{\varepsilon}(g(x_\varepsilon(T) - g(\overline{x}(T))$$

Since $\overline{u}$ is continuous at t, it is easy to check that

$$\lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} [L(s,x_\varepsilon,v) - L(s,\overline{x},\overline{u}))]ds = L(t,\overline{x}(t),v) - L(t,\overline{x}(t),\overline{u}(t)). \tag{14.15}$$

Using lemma 14.1, and defining z as in Lemma 14.1 as the solution of the linearized equation (14.10), we also have

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \int_t^T [L(s, x_\varepsilon, \overline{u}) - L(s, \overline{x}, \overline{u}))] ds = \int_t^T \nabla_x L(s, \overline{x}(s), \overline{u}(s)) \cdot z(s) ds \tag{14.16}$$

and

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} (g(x_\varepsilon(T) - g(\overline{x}(T)) = \nabla g(\overline{x}(T)) \cdot z(T). \tag{14.17}$$

we thus have

$$\begin{aligned} 0 \geq \quad & L(t, \overline{x}(t), v) - L(t, \overline{x}(t), \overline{u}(t) + \textstyle\int_t^T \nabla_x L(s, \overline{x}(s), \overline{u}(s)) \cdot z(s) ds \\ & + \nabla g(\overline{x}(T)) \cdot z(T). \end{aligned} \tag{14.18}$$

To make the previous optimality condition (14.18) tractable, we introduce the so-called *adjoint* (or *co-state*) variable p as the solution of:

$$\dot{p}(s) = -D_x f(s, \overline{x}(s), \overline{u}(s))^T p(s) - \nabla_x L(s, \overline{x}(s), \overline{u}(s)), \ s \in [0, T], \tag{14.19}$$

(where A^T denotes the transpose of the matrix A and we recall the identity $A^T p \cdot z = p \cdot Az$) together with the transversality (terminal) condition:

$$p(T) = \nabla g(\overline{x}(T)) \tag{14.20}$$

(use the same arguments as in the proof of Theorem 14.1 for the existence and uniqueness of the solution of (14.19)-(14.20)). Using the equations defining p and z, we have:

$$\begin{aligned} & \nabla g(\overline{x}(T)) \cdot z(T) = p(T) \cdot z(T) = p(t) \cdot z(t) + \int_t^T p \cdot \dot{z} + \dot{p} \cdot z \\ & = p(t) \cdot z(t) + \int_t^T p(s) \cdot (D_x f(s, \overline{x}(s), \overline{u}(s)) z(s)) \, ds \\ & - \int_t^T (D_x f(s, \overline{x}(s), \overline{u}(s))^T p(s) - \nabla_x L(s, \overline{x}(s), \overline{u}(s))) \cdot z(s) ds \\ & = p(t) \cdot z(t) - \int_t^T \nabla_x L(s, \overline{x}(s), \overline{u}(s)) \cdot z(s) ds \\ & = p(t) \cdot (f(t, \overline{x}(t), v) - f(t, \overline{x}(t), \overline{u}(t))) - \int_t^T \nabla_x L(s, \overline{x}(s), \overline{u}(s)) \cdot z(s) ds \end{aligned}$$

inequality (14.18) then becomes

$$0 \geq L(t, \overline{x}(t), v) - L(t, \overline{x}(t), \overline{u}(t)) + p(t) \cdot (f(t, \overline{x}(t), v) - f(t, \overline{x}(t), \overline{u}(t))). \tag{14.21}$$

Since v is an arbitrary control in K, we may rewrite the previous inequality as:

$$L(t,\overline{x}(t),\overline{u}(t))+p(t)\cdot(f(t,\overline{x}(t),\overline{u}(t))) = \max_{v\in K}\{L(t,\overline{x}(t),v)+p(t)\cdot f(t,\overline{x}(t),v)\}. \tag{14.22}$$

It is therefore natural to introduce the so-called pre-Hamiltonian function $\underline{H}$: $[0,T]\times\mathbb{R}^d\times K\times\mathbb{R}^d\to\mathbb{R}$ by:

$$\underline{H}(t,x,u,p) := L(t,x,u)+p.f(t,x,u),\ \forall(t,x,u,p)\in[0,T]\times\mathbb{R}^d\times K\times\mathbb{R}^d. \tag{14.23}$$

We then define the Hamiltonian H: $[0,T]\times\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ by:

$$H(t,x,p) := \sup_{u\in K}\{L(t,x,u)+p.f(t,x,u)\} = \sup_{u\in K}\underline{H}(t,x,u,p). \tag{14.24}$$

Now we remark that $\underline{H}$ is differentiable with respect to x and p and

$$\nabla_p\underline{H}(t,x,u,p) = f(t,x,u),\ \nabla_x\underline{H}(t,x,u,p) = \nabla_x L(t,x,u)+D_x f(t,x,u)^T p.$$

Hence the state equation (14.1) can be rewritten as:

$$\dot{\overline{x}}(s) = \nabla_p\underline{H}(s,\overline{x}(s),\overline{u}(s),p(s))$$

and, more interestingly, the adjoint equation (14.19) can be rewritten as

$$\dot{p}(s) = -\nabla_x\underline{H}(s,\overline{x}(s),\overline{u}(s),p(s)).$$

Finally, the optimality condition (14.22) expresses the Hamiltonian-maximization condition:

$$\underline{H}(t,\overline{x}(t),\overline{u}(t),p(t)) = H(t,\overline{x}(t),p(t))$$

for every point of continuity t of $\overline{u}$.

We thus have proved Pontryagin's maximum principle:

Theorem 14.3 *Let $\overline{u}$ be a piecewise continuous optimal control for (14.9) and let $\overline{x}$ be the corresponding optimal state. Then for every point t of continuity of $\overline{u}$ one has*

$$\overline{u}(t)\in \mathrm{argmax}_{v\in K}\underline{H}(t,\overline{x}(t),v,p(t))$$

for the adjoint state variable p that solves

$$\dot{p}(s) = -\nabla_x\underline{H}(s,\overline{x}(s),\overline{u}(s),p(s)),\ s\in[0,T]$$

together with the transversality condition

$$p(T) = \nabla g(\overline{x}(T)).$$

14.4 Dynamic Programming and HJB equations

On définit la fonction valeur du problème de contrôle (14.9)

$$v(t,x) := \sup_u \left\{ \int_t^T L(s, y_u(s), u(s))ds + g(y_u(T)) \,:\, y_u(t) = x \right\} \qquad (14.25)$$

Clairement v vérifie la condition aux limites:

$$v(T,x) = g(x) \text{ pour tout } x \in \mathbb{R}^n \qquad (14.26)$$

Le **principe de la programmation dynamique** dit que: "si un controle u est optimal entre 0 et T pour la condition initiale x alors il est aussi optimal entre t et T avec la condition initiale $y_u(t)$ à cette date". Ce principe se traduit ici par la relation suivante:

Proposition 14.1 *La fonction valeur vérifie pour tout $x \in \mathbb{R}^n$ et tout $t \in [0,T]$:*

$$v(0,x) = \sup_u \{ \int_0^t L(s, y_u(s), u(s))ds + v(t, y_u(t)) \,:\, y(0) = x \} \qquad (14.27)$$

14.5 Hamilton-Jacobi-Bellman equations

En utilisant le principe de la programmation dynamique et en étudiant comment varie la valeur entre deux dates proches t et $t + \Delta t$ et deux états proches, nous allons voir qu'une autre propriété de v est qu'elle est solution d'une équation aux dérivées partielles du premier ordre appelée équation d'Hamilton-Jacobi-Bellman:

Proposition 14.2 *Supposons v régulière, alors v est solution de l'équation d'Hamilton-Jacobi-Bellman (H.J.B.):*

$$\partial_t v(t,x) + H(t,x,\nabla_x v(t,x)) = 0. \qquad (14.28)$$

où H est l'Hamiltonien défini par (14.24).

Preuve:
Pour simplifier, nous supposerons qu'il existe des commandes et des trajectoires optimales, i.e. que le sup dans (14.25) est atteint. Soit $[t, t+\Delta t] \subset [t,T]$, $v_0 \in V$ et soit $z(.)$ la solution de:

$$\left\{ \begin{array}{rcl} \dot{z}(s) & = & f(s, z(s), v_0) \\ z(t) & = & x \end{array} \right.$$

$u(.)$ un contrôle optimal pour le problème $v(t+\Delta t, z(t+\Delta t))$. Considérons maintenant le contrôle $w(.)$:

$$w(t) = \begin{cases} v_0 & \text{si} \quad t \in [t, t+\Delta t] \\ u(t) & \text{si} \quad t \in [t+\Delta t, T] \end{cases}$$

En notant y_w la variable d'état correspondante valant x à la date t ($y_w = z$ sur $[t, t+\Delta t]$), on a d'abord:

$$y_w(t+\Delta t) = z(t+\Delta t) = x + f(t,x,v_0)\Delta t + o(\Delta t). \tag{14.29}$$

Il vient ensuite, par définition de la valeur v:

$$\begin{aligned} v(t,x) \geq & \int_t^{t+\Delta t} L(s, y_w(s), v_0)ds + v(t+\Delta t, y_w(t+\Delta t)) \\ =& v(t,x) + \Delta t[L(t,x,v_0) + \partial_t v(t,x) + \nabla_x v(t,x).f(t,x,v_0) + o(1)] \end{aligned}$$

En divisant par Δt et en faisant $\Delta t \to 0$, il vient:

$$\partial_t v(t,x) + L(t,x,v_0) + \nabla_x v(t,x).f(t,x,v_0) \leq 0$$

comme $v_0 \in V$ est arbitraire, en passant au sup en V, on obtient que v est une sous-solution de (14.28):

$$\partial_t v(t,x) + H(t,x,\nabla_x v(t,x)) \leq 0.$$

Soit maintenant $u(.)$ un contrôle optimal pour le problème $v(t,x)$, par le principe de la programmation dynamique, notons que $u(.)$ est aussi optimal pour $v(t+\Delta t, y_u(t+\Delta t))$:

$$\begin{aligned} v(t,x) =& \int_t^T L(s, y_u(s), u(s))ds + g(y_u(T)) \\ =& \int_t^{t+\Delta t} L(s, y_u(s), u(s))ds + v(t+\Delta t, y_u(t+\Delta t) \\ =& v(t,x) + \Delta t[L(t,x,u(t)) + \partial_t v(t,x) + \nabla_x v(t,x).f(t,x,u(t)) + o(1)] \end{aligned}$$

En divisant par Δt et en faisant $\Delta t \to 0$, il vient:

$$\partial_t v(t,x) + L(t,x,u(t)) + \nabla_x v(t,x).f(t,x,u(t)) = 0$$

ainsi, par définition de H, v est aussi sur-solution de (14.28):

$$\partial_t v(t,x) + H(t,x,\nabla_x v(t,x)) \geq 0.$$

□

Notons que la démonstration précédente est heuristique (voir les remarques faites dans le cas du calcul des variations) et que pour faire une théorie satisfaisante des équations d'Hamilton-Jacobi, il faut recourir à la notion de solutions de viscosité.

14.6 Feedback control and sufficient condition

Nous allons voir pour finir ce chapitre que si l'on connait une solution (régulière) du problème aux limites pour l'équation d'H-J-B:

$$\begin{cases} \partial_t w(t,x) + H(t,x,\nabla_x w(t,x)) = 0 \text{ sur } [0,T]\times\mathbb{R}^n \\ w(T,x) = g(x)\ \forall x \in \mathbb{R}^n \end{cases} \tag{14.30}$$

alors on peut en déduire une commande optimale *en feedback*. Une commande en feedback est une fonction qui ne dépend pas seulement du temps mais aussi de l'état du sytème, c'est donc une fonction U de $[0,T]\times\mathbb{R}^n$ à valeurs dans l'espace des contrôles V. Pour un contrôle en feedback $U(.,.)$, la dynamique de la variable d'état est régie par l'équation différentielle ordinaire:

$$\dot{y}(t) = f(t,y(t),U(t,y(t))),\ y(0) = x. \tag{14.31}$$

Notons qu'il est assez naturel de s'intéresser à des contrôles en feedback i.e. dépendant de l'état instantané du système: en pratique, on conduit sa voiture en fonction de sa position et de sa vitesse plutôt qu'en fonction de l'heure qu'il est...

On dira que le contrôle en feedback $U(.,.)$ est optimal pour (14.9) si le contrôle $u(t) = U(t,y(t))$ est optimal avec $y(.)$ solution du problème de Cauchy (14.31).

Théorème 14.1 *Supposons que w est une solution de classe C^1 du problème aux limites (14.30), et que pour tout $(t,x)\in[0,T]\times\mathbb{R}^n$, il existe $U(t,x)\in V$ solution du problème:*

$$\sup_{u\in V}\{L(t,x,u) + \nabla_x w(t,x).f(t,x,u)\}$$

alors U est un contrôle optimal en feedback et donc si y est solution de

$$\dot{y}(t) = f(t,y(t),U(t,y(t))),\ y(0) = x. \tag{14.32}$$

y est une trajectoire optimale pour (14.9) et $u^*(t) = U(t,y(t))$ est un contrôle optimal. Enfin, w est la fonction valeur du problème (14.9).

Preuve:
Montrons que $u^*(t) = U(t,y(t))$ fourni par le théorème est un contrôle optimal. Pour $(t,x,u)\in[0,T]\times\mathbb{R}^n\times V$ posons:

$$F(t,x,u) := L(t,x,u) + \nabla_x w(t,x).f(t,x,u) + \partial_t w(t,x). \tag{14.33}$$

Comme w est solution de (14.30)et par définition de U, on a:

$$0 = \max_u \{F(t,x,u)\} = F(t,x,U(t,x)). \tag{14.34}$$

Définissons pour tout contrôle u la fonctionnelle:

$$K(u) := \int_0^T F(s, y_u(s), u(s))ds$$

Avec (14.34), il vient:

$$K(u^*) = 0 \geq K(v) \text{ pour tout contrôle } v(.). \tag{14.35}$$

Soit $v(.)$ un contrôle et $y_v(.)$ l'état asscoié, on a:

$$\begin{aligned} K(v) &= \int_0^T F(s, y_v(s), v(s))ds \\ &= \int_0^T L(s, y_v(s), v(s))ds + \int_0^T \partial_t w(s, y_v(s))ds + \\ &\int_0^T \nabla_x w(s, y_v(s)).f(s, y_v(s), v(s))ds \\ &= J(v) - g(y_v(T)) + \int_0^T \frac{d}{dt}[w(s, y_v(s))]ds \\ &= J(v) - w(0,x). \end{aligned}$$

Avec (14.35), il vient donc:

$$J(u^*) - J(v) = K(u^*) - K(v) \geq 0,$$

par conséquent u^* est bien un contrôle optimal et:

$$v(0,x) = J(u^*) = K(u^*) + w(0,x) = w(0,x).$$

Par le même argument que précédemment en changeant la condition de Cauchy $(0,x)$ en (t,x) on obtient de même $v(t,x) = w(t,x)$ si bien que w est la fonction valeur. □

En pratique le théorème précédent doit être vu comme une condition suffisante d'optimalité. Il permet en effet de vérifier si un candidat éventuel (fourni par le principe de Pontriaguine) est effectivement optimal.

Bibliography

[1] Aubin, *L*'Analyse non linéaire et ses motivations économiques, Masson, Paris.

[2] G. Barles, *Solutions de viscosité des équations de Hamilton-Jacobi*, Collection Mathematiques et Applications. de la SMAI.

[3] H. Brézis *Analyse fonctionnelle*, Masson, Paris.

[4] H. Cartan *Cours de calcul différentiel*, Hermann, Paris.

[5] G. Cohen *Convexité et Optimisation*, Cours de l'ENPC, disponible à `http://www-rocq.inria.fr/metalau/cohen/`

[6] Ekeland, Temam, *Analyse convexe et problmes variationnels*, Dunod-Gauthier- Villars.

[7] J.-B. Hiriart Urruty *Optimisation et analyse convexe*, PUF, Paris.

[8] J.-B. Hiriart Urruty, C. Lemaréchal *Convex Analysis and Minimization Algorithms*, tomes I et II, Springer-Verlag, Berlin.

[9] A. Mas-Colell, M. Whinston and J. Green, *M*icroeconomic Theory, Oxford, UK, Oxford Univ. Press 1995. Mathematical appendix.

[10] N. Stokey, R. E. Lucas, E. C. Prescott, *R*ecursive methods in economic dynamics, Harvard Univ. Press.