Mémoire de Master 2
à l'université Rennes 1

LEÇON 219 - EXTREMUMS : EXISTENCE, CARACTÉRISATION, RECHERCHE. EXEMPLES ET APPLICATIONS.

Aude Le Gluher, encadrée par Nathalie Krell

Table des matières

Introduction 3
1. Définitions et cadre de travail 4
2. Compacité et existence d'extremums 4
3. Extremums et calcul différentiel 5
3.1. Ordre un 5
3.2. Ordre deux 5
3.3. Problèmes sous contraintes 7
4. Extremums et convexité 8
4.1. Fonctions convexes 8
4.2. Projection sur un convexe fermé 12
5. Algorithmes de recherche et d'approximation 12
Annexes 15
Références 16

Introduction

La recherche d'extremums (ou extrema) occupe une place importante en mathématiques et dans les sciences en général car elle intervient dès lors qu'on cherche à optimiser une certaine quantité. Par exemple, en économie, on cherchera à maximiser les profits et minimiser les coûts ; ces deux quantités étant entre autres fonctions du temps. De même, les lois de la réflexion et de la réfraction en optique découlent d'une minimisation du temps de parcours de la lumière.

On ne considérera dans cette leçon que des fonctions définies sur un $\mathbf{R}$ espace vectoriel de dimension finie ; en particulier, on ne parlera pas de fonctions holomorphes. La première partie présente les définitions avec lesquelles nous travaillerons dans celles qui suivent. Les trois parties suivantes consistent à expliquer ce que l'on peut dire quant à l'existence et la caractérisation d'extremums en fonction des hypothèses dont on dispose sur la fonction f et sur son domaine de définition U. Enfin, la dernière aborde deux méthodes d'approximation d'extremums.

Plus précisément, dans la deuxième partie, les hypothèses sur U et f sont les suivantes : f est continue et U est compact. Dans ces conditions, la fonctions f est bornée et atteint ses bornes. On définit ensuite la notion de coercivité, qui permet en fait de se ramener au cas d'une fonction continue définie sur un compact. Dans cette partie, on retrouve aussi deux théorèmes importants de l'analyse : l'équivalence des normes en dimension finie et le théorème de d'Alembert.

Dans la troisième partie, U n'est plus compact mais ouvert. D'autre part, on rajoute de la régularité à la fonction f : différentiable ou deux fois différentiable. On remarque que dans ces conditions, la majorité des résultats abordés deviennent alors locaux (contrairement à la première partie ou les extremums étaient globaux) ce qui n'a rien de surprenant puisque la différentiabilité est une notion locale. On remarque aussi que les résultats avancés n'ont pas de réciproques ; exemples à l'appui.

La quatrième partie aborde le cas des fonctions qui en plus d'être différentiables sont convexes. On observe que dans ce cas, les théorèmes de la partie précédentes admettent des réciproques ; typiquement, être un point critique devient équivalent à être un extremum. De plus, la convexité permet aussi de récupérer de la globalité ; par exemple, un minimum local pour une fonction convexe est global. Ce minimum est d'ailleurs unique pour peu que la fonction soit strictement convexe : cette propriété est d'ailleurs utilisé dans le premier développement concernant l'ellipsoïde de John-Loewner.

La dernière partie aborde l'approximation d'un extremum par descente de gradient à pas optimal sur un exemple particulier puis la méthode de Newton, qui constitue le second développement.

1. Définitions et cadre de travail

Cadre : On se place dans un $\mathbf{R}$ espace vectoriel, noté E, de dimension finie, notée n, et muni de la norme euclidienne, notée $|| \,.\, ||$. Soit U un sous ensemble non vide de E.

Définition 1.1. *Soit a un élément de U et f une fonction de U à valeurs dans $\mathbf{R}$. On dit que f admet :*

- *un minimum global en a si pour tout x de U, $f(x) \geqslant f(a)$.*
- *un minimum local en a s'il existe un voisinage V de a tel que pour tout x dans $V \cap U$, $f(x) \geqslant f(a)$.*
- *un minimum strict global (respectivement local) en a si l'inégalité du premier point (respectivement du deuxième) est stricte dès que x est différent de a.*

Remarque 1.1.

- *Les notions de maximum (strict) et de maximum local (strict) se définissent de façon similaire en inversant le sens des inégalités dans les définitions précédentes.*
- *On nomme extremum de la fonction f tout point qui est soit un maximum soit un minimum de f.*

Exemple 1.1. *La fonction* $\sin : \left\{ \begin{array}{ccc} \mathbf{R} & \to & \mathbf{R} \\ x & \mapsto & \sin(x) \end{array} \right.$
admet un minimum global non strict en $-\frac{\pi}{2}$. Ce point est aussi un minimum local strict.

2. Compacité et existence d'extremums

Théorème 2.1. *Soit f une fonction de U dans $\mathbf{R}$. Si f est continue et U est un compact de E alors la fonction f est bornée et atteint ses bornes.*

Exemple 2.1. *Soit $N : (\mathbf{R}^n, || \,.\, ||_\infty) \to \mathbf{R}$ une norme. La fonction N est continue sur le compact $S_\infty(0,1)$ (la sphère unité pour la norme infinie) : elle y est donc bornée et elle y atteint son minimum.*

Application 2.1. *Équivalence des normes.*
Si N_1 et N_2 sont deux normes sur le même $\mathbf{R}$ espace vectoriel de dimension finie alors N_1 et N_2 sont équivalentes.

Définition 2.1. *La fonction f de U dans $\mathbf{R}$ est dite coercive si la propriété suivante est vérifiée :*

$$\forall A > 0, \exists B > 0, \exists \alpha > 0, \forall x \in U, (||x|| \geqslant B \text{ ou } d(x, U^c) \leqslant \alpha) \Rightarrow f(x) \geqslant A$$

Remarque 2.1.

- *Intuitivement, la fonction f est coercive lorsqu'elle "explose" lorsqu'on s'approche du bord de son domaine de définition. En particulier :*

- *Si $U = E$, la fonction f est coercive sur U si et seulement si*
$$\lim_{||x|| \to +\infty} f(x) = +\infty$$

Théorème 2.2. *Soit f une fonction de U dans $\mathbf{R}$. Si f est continue et coercive alors f est minorée et atteint sa borne inférieure.*

Exemple 2.2. *La fonction* $f : \left\{ \begin{array}{ccc} \mathbf{R}^2 & \to & \mathbf{R}^2 \\ (x,y) & \mapsto & x^2+y^2 \end{array} \right.$
est continue et coercive sur son domaine de définition. Elle admet effectivement un minimum en $(0,0)$.

Application 2.2. *Théorème de D'Alembert.*
Tout polynôme non constant de $\mathbf{C}[X]$ *admet une racine complexe.*

3. Extremums et calcul différentiel

Dans cette partie on suppose que U est ouvert

3.1. Ordre un.

Définition 3.1. *Soit f une fonction de U dans $\mathbf{R}$ et a un élément de U tel que la différentielle de f en a (notée $\mathrm{D}f(a)$) existe. On dit que a est un point critique de f si $\mathrm{D}f(a) = 0$.*

Théorème 3.1. *Condition nécessaire d'existence d'un extremum.*
Soit f une fonction de U dans $\mathbf{R}$ et a un élément de U. Si a est un extremum local de f et si $\mathrm{D}f(a)$ existe alors a est un point critique de f.

Contre-exemple 3.1. *Cette dernière condition est nécessaire à l'existence d'un extremum local mais pas suffisante. En effet, la fonction*
$f : \left\{ \begin{array}{ccc}]-1,1[& \to & \mathbf{R} \\ x & \mapsto & x^3 \end{array} \right.$ *vérifie* $\mathrm{D}f(0) = 0$ *mais n'a pas d'extremum en* 0.

Application 3.1. *Théorème de Rolle.*
Soit a et b deux réels tels que $a < b$. Soit f une fonction définie sur $[a,b]$ et à valeurs dans $\mathbf{R}$, continue sur $[a,b]$, dérivable sur $]a,b[$ et telle que $f(a) = f(b)$. Alors il existe $c \in]a,b[$ tel que $f'(c) = 0$.

Application 3.2. *Théorème des accroissements finis.*
Soit a et b deux réels tels que $a < b$. Soit f une fonction définie sur $[a,b]$ et à valeurs dans $\mathbf{R}$, continue sur $[a,b]$ et dérivable sur $]a,b[$. Alors il existe $c \in]a,b[$ tel que $f'(c) = \frac{f(b)-f(a)}{b-a}$.

3.2. Ordre deux.

Théorème 3.2. *Soit f une fonction de U dans $\mathbf{R}$ et a un élément de U. Si f admet un minimum (respectivement maximum) local en a et si $\mathrm{D}^2 f(a)$ existe alors $\mathrm{D}f(a) = 0$ et $\mathrm{D}^2 f(a)$ est une forme quadratique positive (respectivement négative).*

Exemple 3.1. *La fonction* $f : \left\{ \begin{array}{lll} \mathbf{R}^2 & \to & \mathbf{R} \\ (x,y) & \mapsto & x^2+y^4 \end{array} \right.$
admet un minimum en $(0,0)$ *et on a effectivement* $\mathrm{D}f((0,0)) = 0$ *et* $\mathrm{D}^2f((0,0)) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ *est une forme quadratique positive.*

Contre-exemple 3.2. *La condition 3.2 est nécessaire à l'existence d'un minimum local mais pas suffisante. En effet, la fonction*
$f : \left\{ \begin{array}{lll} \mathbf{R}^2 & \to & \mathbf{R} \\ (x,y) & \mapsto & x^2-y^3 \end{array} \right.$ *admet* $(0,0)$ *comme point critique ;*
$\mathrm{D}^2f((0,0)) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ *est bien positive, mais pourtant* $(0,0)$ *n'est pas un minimum local de* f.

Théorème 3.3. *Soit* f *une fonction de* U *dans* $\mathbf{R}$ *et* a *un élément de* U. *Si* $\mathrm{D}f(a) = 0$ *et* $\mathrm{D}^2f(a)$ *est définie positive (respectivement négative) alors* f *admet un minimum (respectivement maximum) local strict en* a.

Exemple 3.2. *La fonction* $f : \left\{ \begin{array}{lll} \mathbf{R}^2 & \to & \mathbf{R} \\ (x,y) & \mapsto & x^2-y^2+\frac{y^4}{4} \end{array} \right.$
admet un point critique en $(0,\sqrt{2})$ *et* $\mathrm{D}^2f((0,\sqrt{2}))$ *est définie positive donc* f *admet un minimum local strict en* $(0,\sqrt{2})$.

Contre-exemple 3.3. *La condition 3.3 est suffisante à l'existence d'un minimum local mais pas nécessaire. En effet, la fonction* $f : \left\{ \begin{array}{lll} \mathbf{R}^2 & \to & \mathbf{R} \\ (x,y) & \mapsto & x^2+y^4 \end{array} \right.$
admet un minimum local strict en $(0,0)$ *et pourtant* $\mathrm{D}^2f((0,0))$ *est positive certes, mais pas définie comme vu à l'exemple 3.1.*

Proposition 3.1. *Cas particulier de la dimension deux.*

Soit f *une fonction définie sur un ouvert* U *de* $\mathbf{R}^2$ *et à valeurs dans* $\mathbf{R}$, *de classe* C^2 *sur* U. *Soit* a *un élément de* U *tel que* $\mathrm{D}f(a) = 0$. *Alors :*

- *Si* $\det(\mathrm{D}^2f(a)) > 0$ *alors* a *est un extremum local de* f.
- *Si* $\det(\mathrm{D}^2f(a)) < 0$ *alors* f *n'a pas d'extremum en* a.
- *Si* $\det(\mathrm{D}^2f(a)) = 0$ *on ne peut rien conclure a priori.*

Application 3.3. *Principe du maximum harmonique.*

Soit f *une fonction définie sur* $\mathbf{R}^n$ *et à valeurs dans* $\mathbf{R}$ *de classe* C^2 *et telle que pour tout* x *dans* $\overset{\circ}{D}(0,1)$ *on a* $\Delta f(x) = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}(x) = 0$.

Alors pour tout $x \in \overset{\circ}{D}(0,1)$, $\min_{||y||=1} f(y) \leqslant f(x) \leqslant \max_{||y||=1} f(y)$.

3.3. Problèmes sous contraintes.

Théorème 3.4. *Théorème des extrema liés.*

Soit $f, g_1, ..., g_p$ des fonctions de U dans $\mathbf{R}$, toutes de classe C^1. On note Γ l'ensemble $\{x \in U | \forall i \in [\![1,p]\!], g_i(x) = 0\}$. Si $f_{|\Gamma}$ admet un extremum local en $a \in \Gamma$ et si les $\mathrm{D}g_i(a)$ sont linéairement indépendantes alors il existe des réels $\lambda_1, ..., \lambda_p$, nommés multiplicateurs de Lagrange, tels que

$$\mathrm{D}f(a) = \sum_{i=1}^{n} \lambda_i \mathrm{D}g_i(a)$$

Remarque 3.1. *Ce théorème s'interprète géométriquement en disant que $\nabla f(a)$ est orthogonal au plan tangent à Γ en a.*

Exemple 3.3. *(exercice posé lors de l'oral)*

Soit $P : ax + by + cz + d$ (avec a, b, c non tous nuls) un plan dans $\mathbf{R}^3$. Montrons qu'il existe un point $Q \in P$ qui minimise la distance entre 0 et Q, notée $d(0, Q)$.

Démonstration : Il s'agit ici de minimiser la fonction $N : (x, y, z) \mapsto x^2 + y^2 + z^2$ sur P (plus rigoureusement, il faut minimiser $\sqrt{N}$ sur P mais cela revient au même).

On remarque que cette fonction est continue et coercive donc par le théorème 2.2, l'existence d'un point $Q = (x_Q, y_Q, z_Q)$ minimisant $d(0, Q)$ est assurée.

De plus $DP(x_Q, y_Q, z_Q) = (a, b, c)$ est non nul ce qui assure qu'on peut utiliser le théorème des extrema liés. Ce dernier donne l'existence d'un réel λ tel que $DN(x_Q, y_Q, z_Q) = (2x_Q, 2y_Q, 2z_Q) = \lambda(a, b, c)$. On en déduit que $(x_Q, y_Q, z_Q) = \left(\frac{\lambda a}{2}, \frac{\lambda b}{2}, \frac{\lambda c}{2}\right)$.

On se souvient enfin que Q est un élément de P donc que $ax_Q + by_Q + cz_Q + d = 0$ ce qui permet de calculer λ puis de trouver les coordonnées de Q uniquement en fonction de a, b, c et d.

Exemple 3.4. *(exercice posé lors de l'oral)*

On peut, en suivant la même méthode que ci-dessus minimiser la fonction
$$f : \begin{cases} \mathbf{R}^3 & \rightarrow & \mathbf{R} \\ (x, y, z) & \mapsto & 4x^2 + y^2 + z^2 \end{cases}$$
sur $C = \{(x, y, z) \in \mathbf{R}^3 | 2x + 3y + z - 12 = 0\}$.

Application 3.4. *Inégalité arithmético-géométrique.*

Pour tous réels positifs $x_1, ..., x_n$, on a $\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \leqslant \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$

Application 3.5. *Diagonalisation des endomorphismes symétriques.*

Soit u un endomorphisme symétrique de E. Alors il existe une base orthonormée de E constituée de vecteurs propres de u.

4. Extremums et convexité

On suppose ici que U est convexe.

4.1. **Fonctions convexes.**

Définition 4.1. *Soit f un fonction de U dans $\mathbf{R}$. On dit que f est convexe si la propriété suivante est vérifiée :*

$$\forall x, y \in U, \forall t \in [0,1], f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y)$$

Si de plus l'inégalité est stricte dès que $x \neq y$ et $t \in]0,1[$, on dit que f est strictement convexe.

Théorème 4.1. *Si U est ouvert en plus d'être convexe et $f : U \to \mathbf{R}$ est convexe on a :*

- *si $a \in U$ est un maximum global alors f est constante.*
- *si $a \in U$ est un minimum local de f alors ce minimum est global.*

Théorème 4.2. *Si $f : U \to \mathbf{R}$ est convexe et différentiable sur l'ouvert convexe U alors f admet un extremum local si et seulement si $\mathrm{D}f(a) = 0$.*

Exemple 4.1. *La fonction $f : \left\{ \begin{array}{ccc} \mathbf{R} & \to & \mathbf{R} \\ x & \mapsto & e^x \end{array} \right.$ est convexe et minorée mais comme sa dérivée ne s'annule en aucun point, on en déduit que f n'admet pas de minimum (ni local ni global) sur $\mathbf{R}$.*

Théorème 4.3. *Si $f : U \to \mathbf{R}$ est strictement convexe et admet un minimum alors il est unique.*

Application 4.1. *Ellipsoïde de John-Loewner* ***(DÉVELOPPEMENT 1)***

Soit K un compact de $\mathbf{R}^n$ tel que $0 \in \overset{\circ}{K}$. Il existe un unique ellipsoïde centré en 0 et de volume minimal contenant K.

Démonstration : Soit K un compact de $\mathbf{R}^n$ vérifiant les hypothèses du théorème.

On rappelle qu'un ellipsoïde est un sous ensemble de $\mathbf{R}^n$ (que l'on munit ici de la norme euclidienne) de la forme $E_S = \{X \in \mathbf{R}^n |\ {}^tXSX \leqslant 1\}$ où $S \in S_n^{++}(\mathbf{R})$ (ensemble des matrices symétriques, définies et positives).

Pour toute matrice $S \in S_n^{++}(\mathbf{R})$, notons V_S le volume de l'ellipsoïde E_S. Comme S est symétrique réelle, elle est diagonalisable en base orthonormée ; on peut donc supposer que $S = \begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & a_n \end{pmatrix}$ où les a_i sont des réels strictement positifs.

On en déduit que

$$\begin{aligned} V_S &= \int\limits_{\sum\limits_{i=1}^{n} a_i x_i^2 \leqslant 1} \mathrm{d}x_1 ... \mathrm{d}x_n \\ &= \int\limits_{\sum\limits_{i=1}^{n} t_i^2 \leqslant 1} \frac{\mathrm{d}t_1 ... \mathrm{d}t_n}{\sqrt{a_1 ... a_n}} \text{ via le changement de vraiables } x_i = \frac{t_i}{\sqrt{a_i}} \\ &= \frac{V_{I_n}}{\sqrt{\det(S)}} \end{aligned}$$

On peut donc reformuler le problème ainsi : on cherche à minimiser la fonction

$$D : \left\{ \begin{array}{lll} S_n^{++}(\mathbf{R}) & \to & \mathbf{R}^+ \\ S & \mapsto & \frac{1}{\sqrt{\det(S)}} \end{array} \right.$$

sur l'ensemble $A = \{S \in S_n^{++}(\mathbf{R}) | K \subset E_S\}$

Remarquons que $0 \in \overset{\circ}{K}$. Il existe donc $r > 0$ tel que $B(0, r) \subset K$.

Comme $B(0, r) = E_{\frac{I_n}{r^2}}$, si $S \in A$ on a alors $E_{\frac{I_n}{r^2}} \subset K \subset E_S$ donc $V_{\frac{I_n}{r^2}} \leqslant V_S$ et donc $D\left(\frac{I_n}{r^2}\right) \leqslant D(S)$.

On en déduit que pour minimiser D sur A, il suffit de minimiser D sur l'ensemble $C = \left\{S \in S_n^{++}(\mathbf{R}) | K \subset E_S \text{ et } D(S) \geqslant D\left(\frac{I_n}{r^2}\right)\right\}$

Montrons que D est continue sur le compact non vide C. On en déduira que D atteint son minimum sur C d'où l'existence d'un ellipsoïde de volume minimal contenant K :

- Les fonctions déterminant et racine carrée sont continues et $\sqrt{\det(S)}$ est non nulle dès que $S \in S_n^{++}(\mathbf{R})$ ce qui garantit la continuité de la fonction D sur C.
- C est fermé. En effet, soit $(S_n)_n$ une suite d'éléments de C qui converge vers S.

  Par fermeture de $S_n+(\mathbf{R})$, on sait déjà que $S \in S_n^+(\mathbf{R})$.

  D'autre part, on sait que pour tout $n \in \mathbf{N}, D(S_n) \geqslant D\left(\frac{I_n}{r^2}\right)$.

  En passant à la limite dans cette inégalité, on obtient (par continuité de D établie au point précédent) que $D(S) \geqslant D\left(\frac{I_n}{r^2}\right) > 0$ donc $S \in S_n^{++}(\mathbf{R})$.

  Enfin, on sait aussi que pour tout $X \in \mathbf{R}^n$ et pour tout $n \in \mathbf{N}$, on a ${}^tXS_nX \leqslant 1$.

  En passant à la limite on a alors que pour tout $X \in \mathbf{R}^n$, ${}^tXSX \leqslant 1$ ce qui garantit que $K \subset E_S$.

  Finalement, toutes les conditions sont remplies pour affirmer que $S \in C$.

- C est borné pour la norme $|||.||| : S \mapsto \sup_{||X|| \leqslant 1} || \, {}^tXSX||$.

  En effet, soit $S \in C$ et $X \in \mathbf{R}^n$ tel que $||X|| \leqslant 1$.
  Comme $rX \in B(0,r) \subset K \subset E_S$ on a :
  $1 \geqslant {}^t(rX)S(rX) = r^2 \, {}^tXSX$ c'est à dire $|||S||| \leqslant \frac{1}{r^2}$
- C est non vide.

  En effet, comme K est compact, il est en particulier borné.
  Il existe donc $R > 0$ tel que $K \subset B(0,R) = E_{\frac{I_n}{R^2}}$.
  Donc $\frac{I_n}{R^2} \in C$.

Montrons maintenant que D est une fonction strictement convexe sur le convexe C. Cela assurera que le minimum trouvé au point précédent de la démonstration est unique.

- C est convexe.

  En effet, soit $S, R \in C$ et $t \in [0,1]$.
  D'une part, $tS + (1-t)R \in S_n^{++}(\mathbf{R})$ par convexité de cet ensemble.
  D'autre part, pour tout $X \in \mathbf{R}^n$, on a

  $$\begin{aligned} {}^tX(tS + (1-t)R)X &\leqslant t \, {}^tXSX + (1-t) \, {}^tXRX \\ &\leqslant t + (1-t) = 1 \end{aligned}$$

  Donc $K \subset E_{tS+(1-t)R}$.
  Enfin, $B(0,r) \subset K \subset E_{tS+(1-t)R}$ donc $D(tS + (1-t)R) \geqslant D\left(\frac{I_n}{r^2}\right)$.
  On en déduit que $tS + (1-t)R \in C$.
- La fonction D est strictement convexe sur C (en fait, elle l'est sur $S_n^{++}(\mathbf{R})$).

  En effet, soit $S, R \in C$ et $t \in]0,1[$. Par le théorème de pseudo-réduction simultanée, il existe une matrice inversible P telles que $S = {}^tP \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{pmatrix} P$, où les s_i sont strictement positifs et $R = {}^tPP$.

  On remarque par ailleurs que l'un des s_i est différent de 1 puisque les matrices S et R sont différentes ($\star$). On a donc :

  $$\begin{aligned} D(tS + (1-t)R) &= \det\left({}^tP \begin{pmatrix} ts_1 + 1 - t & & 0 \\ & \ddots & \\ 0 & & ts_n + 1 - t \end{pmatrix} P \right)^{-\frac{1}{2}} \\ &= \det(P)^{-1} \prod_{i=1}^{n} (ts_i + 1 - t)^{-\frac{1}{2}} \\ &= \det(P)^{-1} \prod_{i=1}^{n} e^{-\frac{1}{2}\ln(ts_i+1-t)} \end{aligned}$$

  Par stricte concavité de la fonction logarithme sur $\mathbf{R}^{+\star}$, on a pour tout $i \in [\![1,n]\!]$, $\ln(ts_i + 1 - t) \geqslant t\ln(s_i) + (1-t)\ln(1)$ et l'une de ces inégalités est stricte à cause de la remarque ($\star$). Donc :

$$
\begin{aligned}
D(tS+(1-t)R) &< \det(P)^{-1}\prod_{i=1}^{n} e^{-\frac{1}{2}(t\ln(s_i)+(1-t)\ln(1)} \ (1)\\
&= \det(P)^{-1} e^{-\frac{1}{2}(t\ln(\prod_{i=1}^{n} s_i)+(1-t)\ln(1)} \ (2)\\
&\leqslant \det(P)^{-1}\left(te^{-\frac{1}{2}\ln(\prod_{i=1}^{n} s_i)}+(1-t)e^{-\frac{1}{2}\ln(1)}\right) \ (3)\\
&= \det(P)^{-1}\left(t(\prod_{i=1}^{n} s_i)^{-\frac{1}{2}}+(1-t)\right) \ (2)\\
&= t\det(S)+(1-t)\det(R)
\end{aligned}
$$

Les justifications des (in)égalités précédentes sont les suivantes :
(1) : Stricte croissance de la fonction exponentielle.
(2) : Propriétés de l'exponentielle et du logarithme.
(3) : Convexité de la fonction $x \mapsto e^{-\frac{1}{2}x}$ sur $\mathbf{R}$.

D'où le résultat annoncé.

Remarque 4.1. *La démonstration de ce théorème tient toujours si on remplace l'hypothèse "K est compact" par "K est borné". En effet, on ne se sert jamais du caractère fermé de K.*

Remarque 4.2. *(Ces remarques sont les questions ayant été posées lors de l'oral et leurs réponses)*

- *Idée de la démonstration de l'affirmation " si S est une matrice symétrique réelle alors elle est diagonalisable en base orthonormée" ?*

  On commence par montrer que les valeurs propres complexes de S sont en fait réelles. Puis, on procède par récurrence sur la dimension.

- *Pourquoi peut on faire le calcul du volume de l'ellipsoïde dans la base que l'on veut ?*

  En toute généralité, on ne peut pas ! On cache en fait ici un changement de variables.

  En effet, V_S vaut par définition $\displaystyle\int_{{}^tXSX\leqslant 1} \mathrm{d}x_1...\mathrm{d}x_n$.

  Comme $S = {}^tODO$ où D est diagonale et O est orthogonale, on a donc $V_S = \displaystyle\int_{{}^t(OX)D(OX)} \mathrm{d}x_1...\mathrm{d}x_n$.

  On effectue ensuite le changement de variable $Y = OX$; changement de variable dont le jacobien vaut 1 puisque O est orthogonale. D'où le résultat.

- *Pourquoi peut on choisir la norme que l'on veut pour montrer que C est borné ?*

  Le caractère borné d'un ensemble ne dépend pas de la norme choisie dès lors qu'elles sont toutes équivalentes. Or, c'est le cas puisqu'on est en dimension finie.

- *Quel est l'intérêt de la condition $D(S) \geqslant D\left(\frac{I_n}{r^2}\right)$ dans la définition de C ?*

  Cette condition permet de garantir le caractère fermé de C. En effet, sans elle, une suite d'éléments de C converge vers une matrice de $S_n^+(\mathbf{R})$ mais pas nécessairement une matrice de $S_n^{++}(\mathbf{R})$.

4.2. **Projection sur un convexe fermé.**

Théorème 4.4. *Projection sur un convexe fermé.*

Soit C un convexe fermé non vide de E et x un élément de E. Alors il existe un unique point de C, noté $p_C(x)$ tel que $||x - p_C(x)|| = d(x, C)$.

De plus ce point est caractérisé par $\begin{cases} p_C(x) \in C \\ \forall y \in C, \langle x - p_C(x), y - p_C(x)\rangle \leqslant 0 \end{cases}$

Application 4.2. *Soit $n \in \mathbf{N}^\star$ et, pour tout $i \in [\![1, n]\!]$ un point (x_i, y_i) de $\mathbf{R}^2$. On suppose que les x_i ne sont pas tous égaux. Alors il existe un unique couple $(\lambda, \mu) \in \mathbf{R}^2$ tel que la somme $\sum\limits_{i=1}^{n} (\lambda x_i + \mu - y_i)^2$ est minimale.*

5. Algorithmes de recherche et d'approximation

Théorème 5.1. *Descente de gradient à pas optimal*

Soit $f : \begin{cases} \mathbf{R}^n & \to & \mathbf{R} \\ x & \mapsto & \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c \end{cases}$

où A est une matrice symétrique définie positive, b est un vecteur $\mathbf{R}^n$ et c est un réel. Soit $x_0 \in \mathbf{R}^n$ et $\varepsilon > 0$.

L'algorithme : tant que $\nabla f(x_k) > \epsilon$, faire $x_{k+1} = x_k - t_k \nabla f(x_k)$ où t_k minimise la fonction $t \mapsto f(x_k + t\nabla f(x_k))$ permet de construire une suite $(x_k)_k$ qui converge vers le minimum $\overline{x}$ de f.

De plus, pour tout $k \in \mathbf{N}, ||x_k - \overline{x}|| \leqslant C(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n})^k$ (avec C une constante, $\lambda_1 = \max \operatorname{Sp}(A)$ et $\lambda_n = \min \operatorname{Sp}(A)$).

Exemple 5.1. *(exercice posé lors de l'oral)*

Soit f un endomorphisme symétrique et positif sur $\mathbf{R}^n$ muni du produit scalaire euclidien.

1) Montrer que pour tout $x \in \mathbf{R}^n \backslash \{0\}$ on a $\langle x, f(x)\rangle > 0$

2) On fixe $u \in \mathbf{R}^n$ et on note $g : \begin{cases} \mathbf{R}^n & \to & \mathbf{R} \\ x & \mapsto & \frac{1}{2}\langle f(x), x\rangle - \langle u, x\rangle \end{cases}$

Montrer que g a des dérivées partielles selon toutes direction et admet un unique point critique qui est un minimum global.

Idées pour la résolution :

Cet exercice reprend en fait les éléments de la preuve du théorème 5.1.

La question 1 se résout très rapidement en diagonalisant f en base orthonormée (possible puisque c'est un endomorphisme symétrique réel).

Pour la deuxième question, calculer $g(x + tv)$ pour tout vecteur v non nul de $\mathbf{R}^n$ permet de facilement montrer que g admet des dérivées dans toutes les directions. On constate en calculant la différentielle de g qu'elle admet un unique point critique. D'autre part, cette fonction est continue et coercive (le premier terme de g, quadratique, l'emporte à l'infini sur son deuxième terme, linéaire) : elle admet donc un minimum. Ce minimum se doit d'être un point critique : l'unique point critique de g est donc un minimum. Ce dernier est global par convexité stricte de g.

Théorème 5.2. *Méthode de Newton* ***(DÉVELOPPEMENT 2)***

Soit U un ouvert de $\mathbf{R}$ et f une fonction de classe C^2 sur U à valeurs dans $\mathbf{R}$. Si a est un élément de U tel que $f(a) = 0$ et $f'(a) > 0$ alors la suite $(x_n)_n$ définie par récurrence par $x_{n+1} = F(x_n)$ avec $F : x \mapsto x - \frac{f(x)}{f'(x)}$ converge vers a de façon quadratique dès que son premier terme x_0 est suffisamment proche de a.

Démonstration : Puisque f' est continue et $f'(a) > 0$ on en déduit que localement autour de a, la dérivée de f est strictement positive. La fonction f est donc strictement croissante localement autour de a ce qui assure l'existence d'un voisinage (que l'on choisit de prendre compact) V de a, inclus dans U, tel que a est le seul zéro de f dans V d'une part et tel que $f' > 0$ sur V d'autre part.

On constate que F est bien définie sur V puisque f' ne s'y annule pas. D'autre part, pour tout $x \in V$, on a

$$\begin{aligned} F(x) - a &= x - a - \frac{f(x)-f(a)}{f'(x)} \text{ puisque } f(a) = 0 \\ &= \frac{f(a)-f(x)-(a-x)f'(x)}{f'(x)} \end{aligned}$$

Or, par la formule de Taylor-Lagrange entre a et x, il existe un réel z_x entre a et x tel que $f(a) = f(x) + (a-x)f'(x) + f''(z_x)\frac{(a-x)^2}{2}$.

Donc pour tout $x \in V$, $F(x) - a = \frac{f''(z_x)}{f'x)}\frac{(x-a)^2}{2}$

Comme $(u, v) \mapsto \frac{f''(u)}{f'(v)}$ est continue sur le compact V^2, on peut introduire la constante $C = \max\limits_{u,v \in V^2} \left|\frac{f''(u)}{2f'(v)}\right|$.

Et on a donc pour tout $x \in V$, $|F(x) - a| \leqslant C|x-a|^2$ $(\star)$

Soit dès lors $\alpha > 0$ tel que $I = [a+\alpha, a-\alpha] \subset V$ et $C\alpha < 1$. Alors pour tout $x \in I$, $|F(x) - a| \leqslant C\alpha^2 < \alpha$ donc I est stable par F.

On peut alors définir correctement la suite $(x_n)_n$ de l'énoncé dès que $x_0 \in I$, et cette suite vérifie $(\star)$. On a donc pour tout $n \in \mathbf{N}$ $|x_{n+1} - a| \leqslant C|x_n - a|^2$ puis par une récurrence rapide $C|x_{n+1} - a| \leqslant (C\alpha)^{2^n} \to 0$ lorsque $n \to +\infty$ car $C\alpha < 1$. D'où la convergence quadratique de $(x_n)_n$ vers a.

Corollaire 5.1. *Si $g : [c, d] \to \mathbf{R}$ est de classe C^3 et est strictement convexe, on peut approcher son minimum par la méthode de Newton appliquée à f'.*

Remarque 5.1. *(ces questions ont été posées lors de l'oral)*

- *Comment faire si il est très difficile de calculer la dérivée d'une fonction (dans la méthode de Newton) ?*

On peut, à la place de calculer la dérivée de la fonction approcher cette dérivée par un taux d'accroissement.

- *Que se passe-t-il dans la méthode de gradient à pas optimal si on est en dimension un ?*

  On converge alors en une étape vers le minimum de f.

Annexes

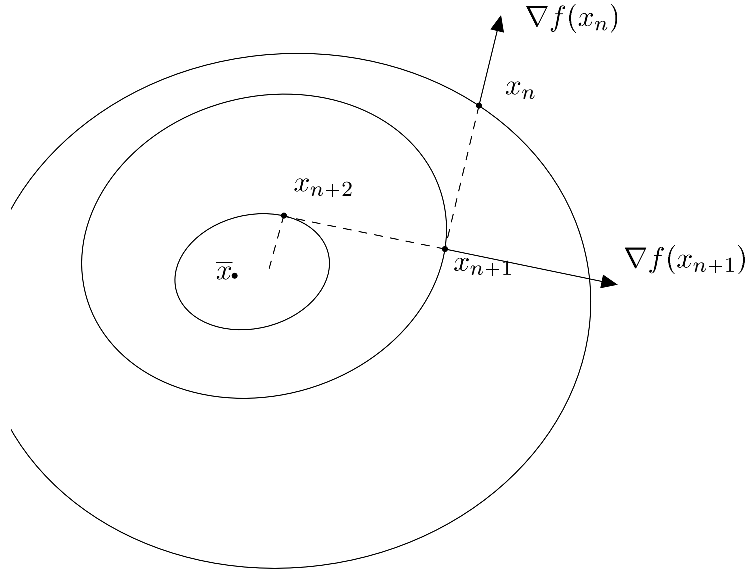
ANNEXE A :

La méthode de descente dans le cas général consiste à calculer la suite $(x_n)_n$ définie par récurrence par

$$x_{n+1} = x_n + t_n d_n$$

où t_n est appelé "pas" et où d_n est la direction de la descente.

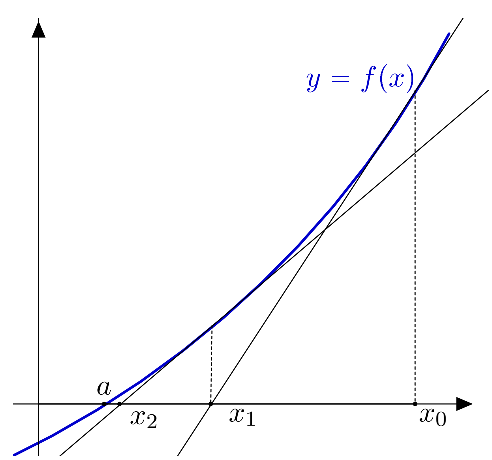
Dans le cas de la descente de gradient à pas optimal, $d_n = \nabla f(x_n)$ et le pas est optimal au sens où, à chaque étape, on minimise f sur la droite $x_n + \nabla f(x_n)$.



Remarque : les différents gradients devraient être orthogonaux à la ligne de niveau de f correspondante (ce qui n'est pas vraiment le cas sur ce dessin).

ANNEXE B

Méthode de Newton : x_{n+1} est à l'intersection de (Ox) et de la tangente à la courbe de f en x_n.



Références

[1] François Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*, Troisième édition, Cassini

[2] Xavier Gourdon, *Analyse*, Deuxième édition, Ellipses

[3] Vincent Beck, Jérôme Malick, Gabriel Peyré, *Objectif agrégation*, Deuxième édition

[4] Frédéric Testard, *Analyse mathématique : la maîtrise de l'implicite*, Calvage et Mounet

[5] Jean-Baptiste Hirriart-Urruty, *Optimisation et analyse convexe*

[6] Serge Francinou, Hervé Gianella, Serge Nicolas, *Exercices de mathématiques oraux X-ENS algèbe 3*, Deuxième édition, Cassini