COLLECTION ENSEIGNEMENT SUP //// Mathématiques

L3M1

Optimisation et analyse convexe

EXERCICES CORRIGÉS

Jean-Baptiste Hiriart-Urruty

EDP SCIENCES

OPTIMISATION ET ANALYSE CONVEXE

Exercices et problèmes corrigés, avec rappels de cours

Jean-Baptiste Hiriart-Urruty

Collection dirigée par Daniel Guin

EDP SCIENCES

17, avenue du Hoggar
Parc d'activités de Courtabœuf, BP 112
91944 Les Ulis Cedex A, France

Illustration de couverture : un corps convexe d'épaisseur presque constante et son ombre ; reproduit avec la gracieuse permission de Christof Weber (université de Zurich).

Imprimé en France

ISBN : 978-2-7598-0373-6

TABLE DES MATIÈRES

INTRODUCTION

« *Good modern science implies good variational problems* »
M.S. Berger (1983)

Le recueil d'exercices et problèmes corrigés que nous proposons ici concerne les domaines des Mathématiques répertoriées sous les vocables d'**Optimisation** et **Analyse convexe**. L'Optimisation est traitée dans ses aspects suivants : la clé de voûte que constituent les conditions d'optimalité (chapitres II et III) ; le rôle (incontournable) de la dualisation de problèmes (chapitre IV) ; le monde particulier (et toujours en haut de l'affiche depuis ses débuts) de l'Optimisation linéaire (chapitre V). L'Analyse convexe (moderne) n'est pas traitée en tant que telle mais par l'utilisation qu'on peut en avoir en Optimisation ; il s'agit en fait d'une initiation à la manipulation de concepts et de résultats concernant essentiellement : la projection sur un convexe fermé (au chapitre VI), le calcul sous-différentiel et de transformées de Legendre-Fenchel (chapitre VII). L'Analyse linéaire et bilinéaire (ou, plutôt, l'Analyse matricielle) ainsi que le Calcul différentiel interviennent de manière harmonieuse en Optimisation et Analyse convexe : un chapitre de revision des bases leur est consacré (chapitre I). Près de 160 exercices et problèmes sont corrigés, parfois commentés et situés dans un contexte d'utilisation ou de développement historique, gradués dans leur difficulté par un, deux ou trois $*$:

$*$ Exercices plutôt faciles (applications immédiates d'un résultat du Cours, vérification d'un savoir-faire de base, etc.) ;

$**$ Exercices que le lecteur-étudiant doit pouvoir aborder après une bonne compréhension et assimilation du Cours. De difficulté moyenne, ce sont de loin les plus nombreux ;

$***$ Exercices plus difficiles, soit à cause de certains calculs à mener à bien, soit simplement en raison d'un degré de maturité plus grand que leur résolution requiert.

Comme tous les exercices de mathématiques, ceux présentés ici ne seront profitables au lecteur-étudiant que si celui-ci les travaille, un crayon à la main, sans regarder la correction dans un premier temps. Qu'il garde à l'esprit ce proverbe chinois :

« *J'entends et j'oublie,* (cours oral)
je vois et je retiens, (étude du cours)
je fais et je comprends » . (exercices)

Le *cadre de travail* choisi est volontairement simple (celui des espaces de dimension finie), et nous avons voulu insister sur les *idées* et *mécanismes de base* davantage que sur les généralisations possibles ou les techniques particulières à tel ou tel contexte. Les problèmes dits *variationnels* requièrent dans leur traitement une intervention plus grande de la Topologie et de l'Analyse fonctionnelle, à commencer par le cadre – fondamental – des espaces de Hilbert ; ils seront abordés dans un prochain recueil.

Les *connaissances mathématiques* pour tirer profit des exercices et problèmes du recueil présent sont maintenues minimales, celles normalement acquises après une formation scientifique à Bac + 2 ou Bac + 3 (suivant les cas).

Chaque chapitre débute par des rappels de résultats essentiels, ce qui ne doit pas empêcher le lecteur-étudiant d'aller consulter les références indiquées à la fin du livre. L'approche retenue est celle d'une progression en spirale plutôt que linéaire au sens strict : ainsi, par exemple, la fonction $A \in \mathcal{M}_n(\mathbb{R}) \longmapsto \ln(\text{dét} A)$ est d'abord considérée pour un calcul de différentielles, puis pour sa convexité, puis plus tard en raison de son rôle comme fonction-barrière dans des problèmes d'optimisation matricielle.

Pour ce qui est de l'*enseignement*, les aspects de l'Optimisation et Analyse convexe traités en exercices ici trouvent leur place dans les formations de niveau deuxième cycle universitaire (modules généralistes ou professionnalisés) et dans la formation mathématique des ingénieurs, sur une durée d'un semestre environ ; la connaissance de ces aspects est un préalable à des formations plus en aval, en optimisation numérique par exemple.

La plupart des exercices et problèmes proposés, sinon tous, ont été posés en séances d'exercices ou examens à l'Université Paul Sabatier de Toulouse.

Je voudrais remercier les anciens étudiants ou jeunes collègues qui ont bien voulu relire une première version de ce document et y relever une multitude de petites fautes (il en reste sûrement...), parmi eux : D. Mallard, M. Torki, Y. Lucet, C. Imbert et J. Benoist. Enfin je ne voudrais pas oublier A. Andrei pour la part primordiale qui a été la sienne dans la saisie informatique de l'ouvrage.

Toulouse, 1989–1997
J.-B. Hiriart-Urruty

Depuis sa publication il y a dix ans (en mars 1998), cet ouvrage a subi les vicissitudes d'un document de formation destiné à un public (d'étudiants en sciences) en nette diminution. Il a été traduit en russe par des collègues de Kiev (Ukraine) en 2004, mais la version française originelle n'est plus disponible depuis 2006. Ainsi, pour répondre à une demande de collègues et étudiants, un nouveau tirage a été envisagé. Je remercie les éditions EDP Sciences, notamment mon collègue D. Guin (directeur de la collection Enseignement Sup – Mathématiques), d'avoir accueilli ce projet. Aude Rondepierre a donné un coup de main pour reprendre les fichiers informatiques anciens ; qu'elle soit remerciée de sa bonne volonté et efficacité.

Toulouse, printemps 2009
J.-B. Hiriart-Urruty

This page intentionally left blank

ABRÉVIATIONS ET NOTATIONS

$:=$: égal par définition.

cf. : *confer*, signifie « se reporter à ».

i.e. : *id est*, signifie « c'est-à-dire ».

ln : notation normalisée pour le logarithme népérien.

$\mathbb{R}_*^+$, $\mathbb{R}_+^*$ ou $]0, +\infty[$: ensemble des réels strictement positifs.

u^+ : partie positive du réel u.

$x = (x_1, \ldots, x_n)$ ou $x = (\xi_1, \ldots, \xi_n)$: notation générique pour un vecteur de $\mathbb{R}^n$.

u^+ signifie $(u_1^+, \ldots, u_n^+)$ lorsque $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$.

Lorque u et v sont deux vecteurs de $\mathbb{R}^n$, $u \leqslant v$ signifie « $u_i \leqslant v_i$ pour tout $i = 1, \ldots, n$ ».

$\{u_k\}$ ou (u_k) : notations utilisées pour les suites indexées par des entiers naturels.

Pour une fonction f différentiable en x (resp. deux fois différentiable en x), $Df(x)$ désigne la différentielle (première) de f en x (resp. $D^2f(x)$ désigne la différentielle seconde de f en x). Si la variable est réelle (et notée t), on utilise la notation $f'(t)$ (resp. $f''(t)$) pour la dérivée de f en t (resp. la dérivée seconde de f en t) [ce sont des éléments de l'espace d'arrivée et non des applications linéaires].

Pour une fonction numérique f définie sur un ouvert $\mathcal{O}$ de $\mathbb{R}^n$, différentiable en $x \in \mathcal{O}$ (resp. deux fois différentiable en $x \in \mathcal{O}$), $\nabla f(x)$ (resp. $\nabla^2 f(x)$) désigne le (vecteur) *gradient* de f en x (resp. la matrice *hessienne* de f en x).

Lorsqu'elle existe, la dérivée directionnelle de f en x dans la direction d est notée $f'(x, d)$.

Pour une fonction vectorielle $f : \mathcal{O} \subset \mathbb{R}^n \to \mathbb{R}^m$ différentiable en $x \in \mathcal{O}$, $Jf(x)$ désigne la matrice *jacobienne* de f en x (matrice à m lignes et n colonnes).

$\mathcal{M}_{m,n}(\mathbb{R})$: ensemble des matrices (m,n) (m lignes et n colonnes) à coefficients réels ; $\mathcal{M}_n(\mathbb{R})$ est une abréviation de $\mathcal{M}_{n,n}(\mathbb{R})$.

$[a_{ij}]$: matrice de terme général a_{ij} (à la i-ème ligne et j-ème colonne).

diag $(\lambda_1,\ldots,\lambda_n)$: matrice diagonale dont les éléments diagonaux sont $\lambda_1,\ldots,\lambda_n$.

I_n (ou I quand il n'y a pas d'ambiguïté) : matrice-unité de $\mathcal{M}_n(\mathbb{R})$, *i.e.* diag (1,...,1).

$A^\top$ ou ${}^tA$: *transposée* de $A \in \mathcal{M}_{m,n}(\mathbb{R})$ [les deux notations sont d'un usage très courant ; par contre A^t est à proscrire car génératrice de confusions].

Lorsque A est inversible, $A^{-\top}$ désigne l'inverse de $A^\top$ (ou, ce qui revient au même, la transposée de A^{-1}).

tr A : trace de $A \in \mathcal{M}_n(\mathbb{R})$.

dét A : déterminant de $A \in \mathcal{M}_n(\mathbb{R})$.

cof A : matrice des cofacteurs de $A \in \mathcal{M}_n(\mathbb{R})$, *i.e.* celle dont le terme (i,j) est $(-1)^{i+j}$ dét A_{ij}, où A_{ij} est obtenue à partir de A en enlevant la i-ème ligne et la j-ème colonne.

$\mathcal{S}_n(\mathbb{R})$: ensemble des matrices de $\mathcal{M}_n(\mathbb{R})$ qui sont symétriques.

$\oplus$ symbolise la somme directe de sous-espaces vectoriels.

vect$\{v_1,\ldots,v_k\}$: sous-espace vectoriel engendré par les vecteurs $v_1,\ldots,v_k$.

Sauf indication contraire, $\mathbb{R}^n$ est muni de sa base canonique ; ainsi *à* $A \in \mathcal{M}_{m,n}(\mathbb{R})$ *est canoniquement associée une application linéaire de* $\mathbb{R}^n$ *dans* $\mathbb{R}^m$, d'où les notations Ker A, Im A, etc.

L'isomorphisme canonique de $\mathbb{R}^n$ sur $\mathcal{M}_{n,1}(\mathbb{R})$ est celui qui à $x = (x_1,\ldots,x_n)$ associe la matrice unicolonne $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$; des expressions comme AX (ou Ax) ne devraient pas arrêter l'étudiant-lecteur. Si, par exemple, u et v sont deux vecteurs de $\mathbb{R}^n$, $uv^\top$ *est une matrice carrée de taille* n dont le terme général est u_iv_j, alors que $u^\top v$ est la matrice-scalaire (ou scalaire) $\sum_{i=1}^n u_iv_j$.

$\langle\cdot,\cdot\rangle$, $\ll \cdot,\cdot \gg$, $(\cdot \mid \cdot)$: notations utilisées pour les produits scalaires (dans des espaces euclidiens). Sauf indication contraire, $\langle\cdot,\cdot\rangle$ désigne dans $\mathbb{R}^n$ le produit scalaire usuel (celui qui à $x = (\xi_1,\ldots,\xi_n)$ et $y = (\eta_1,\ldots,\eta_n)$ associe $\langle x,y\rangle := \sum_{i=1}^n \xi_i\eta_i$, soit encore $x^\top y$ (*cf.* supra)). Bien des problèmes d'optimisation se posent dans des espaces de matrices : si $X := \mathcal{M}_{m,n}(\mathbb{R})$, le produit scalaire standard sur X est défini par $\ll M,N \gg :=$ tr $(M^\top N)$.

Soit $A \in \mathcal{M}_n(\mathbb{R})$, u et v des vecteurs de $\mathbb{R}^n$: $u^\top Av$ est un(e) (matrice-) scalaire égal(e) à (sa transposée) $uA^\top v$; un mécanisme plus commode d'utilisation et engendrant moins de fautes est d'écrire $\langle u, Av\rangle = \left\langle A^\top u, v\right\rangle$.

Si l est une application linéaire d'un espace euclidien $(E, \ll \cdot, \cdot \gg)$ dans un autre espace euclidien $(F, \langle\cdot,\cdot\rangle)$, l'adjointe l^* de l est l'application linéaire de F dans E définie par

$$\ll l^*(y), x \gg = \langle y, l(x)\rangle \text{ pour tout } (x, y) \in E \times F.$$

Si l'on représente l'ensemble des formes linéaires sur E par E *via* le produit scalaire $\ll \cdot, \cdot \gg$ (idem pour F), prendre l'adjointe de l ou sa transposée revient au même. Lorsque E et F sont munies de bases orthonormales (comme c'est le cas des espaces euclidiens $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ munies de leurs bases canoniques), la matrice représentant l'adjointe de l (= sa transposée) dans ces bases est la transposée de la matrice représentant l.

$\|\cdot\|$: norme dans un espace vectoriel normé X. Si X est muni d'un produit scalaire $\langle\cdot,\cdot\rangle$ (exemples typiques : $\mathbb{R}^n$, $\mathcal{M}_{m,n}(\mathbb{R})$), et en l'absence d'autres précisions, $\|\cdot\|$ désignera la norme dérivée du produit scalaire (*i.e.* $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$). Lorsqu'interviennent à la fois des normes de vecteurs et de matrices, on évite les confusions en utilisant $\|\cdot\|$ pour les vecteurs et $|||\cdot|||$ pour les matrices).

int S ou $\overset{\circ}{S}$: intérieur de S ; $\overline{S}$: adhérence de S ;
fr S : frontière de S.
$\overline{B}(x, r)$: boule fermée de centre x et de rayon r.

Δ_n : *simplexe-unité* de $\mathbb{R}^n$, c'est-à-dire l'ensemble des $(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ tels que $\alpha_1 + \ldots + \alpha_n = 1$ et $\alpha_i \geqslant 0$ pour tout i (coefficients qui servent dans la prise de combinaisons convexes).

$A \in \mathcal{S}_n(\mathbb{R})$ est dite (symétrique) *semi-définie positive* lorsque $\langle Ax, x\rangle \geqslant 0$ pour tout $x \in \mathbb{R}^n$ [cette appellation est préférable à « positive » qui peut signifier, dans certains cas, « à coefficients positifs »]. On désignera par $\mathcal{P}_n(\mathbb{R})$ l'ensemble des matrices semi-définies positives de taille n.

$A \in \mathcal{S}_n(\mathbb{R})$ est dite (symétrique) *définie positive* lorsque $\langle Ax, x\rangle > 0$ pour tout $x \neq 0$ de $\mathbb{R}^n$ (ce qui revient à : A semi-définie positive et inversible). On désignera par $\overset{\circ}{\mathcal{P}}_n(\mathbb{R})$ l'ensemble des matrices définies positives de taille n.

Le caractère de semi-définie ou définie positivité ne sera considéré dans ce recueil que pour des matrices symétriques.

This page intentionally left blank

I

RÉVISION DE BASES : CALCUL DIFFÉRENTIEL, ALGÈBRE LINÉAIRE ET BILINÉAIRE

Rappels

I.1. Algèbre linéaire et bilinéaire

Si $l : \mathbb{R}^n \to \mathbb{R}$ est linéaire, il existe un unique $v \in \mathbb{R}^n$ tel que $l(x) = \langle v, x \rangle$ pour tout $x \in \mathbb{R}^n$; les fonctions affines à valeurs réelles sont de la forme $x \longmapsto \langle v, x \rangle + c$, où $v \in \mathbb{R}^n$ et $c \in \mathbb{R}$. Même chose si l est une forme linéaire sur un espace euclidien $(E, \langle \cdot, \cdot \rangle)$: l s'écrit de manière unique sous la forme $\langle v, \cdot \rangle$, où $v \in E$.

Si $q : \mathbb{R}^n \to \mathbb{R}$ est une forme quadratique, il existe un unique $Q \in \mathcal{S}_n(\mathbb{R})$ tel que $q(x) = \dfrac{1}{2}\langle Qx, x \rangle$ pour tout $x \in \mathbb{R}^n$ (le coefficient $1/2$ est mis là pour simplifier les calculs).

Rappel du mécanisme de transposition : $\langle Ax, y \rangle = \langle x, A^\top y \rangle$ (produits scalaires dans $\mathbb{R}^m$ et $\mathbb{R}^n$ respectivement).

Si H est un sous-espace vectoriel de $\mathbb{R}^n$, $H^\perp$ désigne son sous-espace orthogonal. Rappel : si A est linéaire, $(\mathrm{Ker} A)^\perp = \mathrm{Im}(A^\top)$.

Un résultat-clé de l'Algèbre linéaire et bilinéaire : si $S \in \mathcal{S}_n(\mathbb{R})$, toutes les valeurs propres λ_i de S sont réelles et il existe une matrice orthogonale U telle que $U^\top S U = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$; ainsi il existe des vecteurs propres x_i unitaires (x_i associé à λ_i) tels que $S = \sum_{i=1}^{n} \lambda_i x_i x_i^\top$ (ceci est appelé une *décomposition spectrale* de S).

Les formulations variationnelles suivantes de la plus grande et la plus petite valeurs propres de $S \in \mathcal{S}_n(\mathbb{R})$ sont essentielles :

$$\lambda_{max}(S) = \max_{x \neq 0} \frac{\langle Sx, x\rangle}{\| x \|^2}, \quad \lambda_{min}(S) = \min_{x \neq 0} \frac{\langle Sx, x\rangle}{\| x \|^2}.$$

$\mathcal{S}_n(\mathbb{R})$ est muni de l'ordre (partiel) suivant : $A \succeq B$ dans $\mathcal{S}_n(\mathbb{R})$ lorsque $A - B$ est semi-définie positive ($A - B \in \mathcal{P}_n(\mathbb{R})$).

Si $X \in \overset{\circ}{\mathcal{P}}_n(\mathbb{R})$, la racine carrée positive de X (notée $X^{1/2}$) est l'unique $S \in \overset{\circ}{\mathcal{P}}_n(\mathbb{R})$ telle que $S^2 = X$. L'application $X \longmapsto X^{1/2}$ a des propriétés similaires à celles de la fonction racine carrée sur $\mathbb{R}_*^+$; prudence tout de même... Mêmes remarques pour l'application $X \longmapsto X^{-1}$.

Si K est un cône convexe fermé d'un espace euclidien $(E, \langle\cdot,\cdot\rangle)$, le cône polaire K° de K est le cône convexe fermé de E défini comme suit :

$$K^\circ := \{s \in E \mid \ \langle s, x\rangle \leqslant 0 \text{ pour tout } \ x \in K\}.$$

I.2. Calcul différentiel

$f : \mathcal{O} \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ définie sur un ouvert $\mathcal{O}$ de $\mathbb{R}^n$ est dite *différentiable* en $\overline{x} \in \mathcal{O}$ s'il existe $l : \mathbb{R}^n \longrightarrow \mathbb{R}$ linéaire telle que :

$$f(\overline{x} + h) = f(\overline{x}) + l(h) + \| h \| \, \varepsilon(h), \text{ avec } \ \varepsilon(h) \to 0 \text{ quand } h \underset{\neq}{\to} 0. \tag{1.1}$$

l est représentée (dans $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ euclidien) par un unique vecteur, appelé *gradient* de f en $\overline{x}$, et noté $\nabla f(\overline{x})$ (ça se lit « nabla f de $\overline{x}$ »). Désignant par $\partial_j f(\overline{x})$ la *dérivée partielle* en $\overline{x}$ de f par rapport à la j^e variable, $\nabla f(\overline{x})$ est le vecteur de $\mathbb{R}^n$ de composantes $\partial_1 f(\overline{x}), ..., \partial_n f(\overline{x})$. Une manière équivalente d'exprimer (1.1) est :

$$\lim_{t \to 0^+} \frac{f(\overline{x} + td) - f(\overline{x})}{t} = \langle \nabla f(\overline{x}), d\rangle \tag{1.2}$$

et la convergence est uniforme par rapport à d lorsque celui-ci reste dans un ensemble borné de $\mathbb{R}^n$ (la sphère-unité par exemple).

Plus généralement, $F : \mathcal{O} \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ qui à $x \in \mathcal{O}$ associe $F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$ est dite différentiable en $\overline{x} \in \mathcal{O}$ si chacune des fonctions-composantes $f_1, ..., f_m$ l'est. On appelle alors matrice *jacobienne* de F en $\overline{x}$, et on note $JF(\overline{x})$ la matrice de $\mathcal{M}_{m,n}(\mathbb{R})$ dont les lignes sont $[\nabla f_1(\overline{x})]^\top, ..., [\nabla f_m(\overline{x})]^\top$, c'est-à-dire que le terme (i, j) de $JF(\overline{x})$ est $\partial_j f_i(\overline{x})$.

Revenant à une fonction numérique $f : \mathcal{O} \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ différentiable sur $\mathcal{O}$, on dit que f est deux fois différentiable en $\overline{x} \in \mathcal{O}$ lorsque $\nabla f : \mathcal{O} \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ est différentiable en $\overline{x}$. La matrice jacobienne de ∇f en $\overline{x}$ est appelée matrice *hessienne* de f en $\overline{x}$ et notée $\nabla^2 f(\overline{x})$; il s'agit d'une matrice symétrique de taille n dont le terme (i,j) est $\partial_i(\partial_j f)(\overline{x}) = \partial^2_{ij} f\,(\overline{x})$ (dérivée partielle d'ordre 2 en $\overline{x}$ de f par rapport à la i^{e} et j^{e} variables).

Rappel : dans le cas où f est deux fois différentiable en $\overline{x} \in \mathcal{O}$, on a le développement de Taylor-Young d'ordre deux suivant :

$$f\,(\overline{x}+h) = f\,(\overline{x}) + \langle \nabla f\,(\overline{x})\,, h\rangle + \frac{1}{2}\langle \nabla^2 f\,(\overline{x})\,h, h\rangle + \|\ h\ \|^2\,\varepsilon\,(h)\,, \tag{1.3}$$

avec $\varepsilon\,(h) \to 0$ quand $h \underset{\neq}{\longrightarrow} 0$.

Enfin deux ensembles de résultats de Calcul différentiel sont essentiels en Optimisation : le théorème de la fonction implicite et le théorème d'inversion locale ; les développements de Taylor sous leurs formes diverses. À revoir si nécessaire.

I.3. Fonctions convexes

Soit C un convexe de $\mathbb{R}^n$; $f : C \to \mathbb{R}$ est dite *convexe* sur C si pour tout $(x, x') \in C \times C$ et tout $\alpha \in \,]0,1[$ on a :

$$f\left(\alpha x + (1-\alpha)\,x'\right) \leqslant \alpha f\,(x) + (1-\alpha)\,f(x'). \tag{1.4}$$

f est dite *strictement convexe* sur C quand l'inégalité (1.4) est stricte dès que $x \neq x'$. Une propriété encore plus forte est comme suit : f est dite *fortement convexe* sur C, de module de forte convexité $c > 0$, lorsque

$$f\left(\alpha x + (1-\alpha)\,x'\right) \leqslant \alpha f\,(x) + (1-\alpha)\,f(x') - \frac{1}{2}c\,\alpha\,(1-\alpha)\,\|\ x' - x\ \|^2 \tag{1.5}$$

pour tout $(x, x') \in C \times C$ et tout $\alpha \in \,]0,1[$.

Rappelons deux résultats essentiels :

Théorème. *Soit f différentiable sur un ouvert $\mathcal{O}$ de $\mathbb{R}^n$ et C un convexe de $\mathcal{O}$. Alors :*

(i) f est convexe sur C si et seulement si

$$f\,(x) \geqslant f\,(\overline{x}) + \langle \nabla f\,(\overline{x})\,, x - \overline{x}\rangle \ \text{pour tout}\ \ (x, \overline{x}) \in C \times C\ ; \tag{1.6}$$

(ii) *f est strictement convexe sur C si et seulement si l'inégalité* (1.6) *est stricte dès que $\overline{x} \neq x$;*

(iii) *f est fortement convexe sur C de module c si et seulement si*

$$f(x) \geqslant f(\overline{x}) + \langle \nabla f(\overline{x}), x - \overline{x} \rangle + \frac{1}{2} c \parallel x - \overline{x} \parallel^2, \text{ pour tout } (x, \overline{x}) \in C \times C. \quad (1.7)$$

Théorème. *Soit f deux fois différentiable sur un ouvert convexe $\mathcal{O}$. Alors :*

(i) f est convexe sur $\mathcal{O}$ si et seulement si $\nabla^2 f(x)$ est semi-définie positive pour tout $x \in \mathcal{O}$;

(ii) Si $\nabla^2 f(x)$ est définie positive pour tout $x \in \mathcal{O}$, alors f est strictement convexe sur $\mathcal{O}$;

(iii) f est fortement convexe sur $\mathcal{O}$ de module c si et seulement si la plus petite valeur propre de $\nabla^2 f(x)$ est minorée sur $\mathcal{O}$ par c , soit encore :

$$\langle \nabla^2 f(x) d, d \rangle \geqslant c \parallel d \parallel^2 \text{ pour tout } x \in \mathcal{O} \text{ et tout } d \in \mathbb{R}^n.$$

Références. Parmi les exercices de [15] figurent des applications simples à l'Analyse numérique matricielle et l'Optimisation. Outre les références suggérées en pp. 305 307 de [15] (et rééditées à présent), signalons [16], ouvrage complet et très diffusé, dans lequel l'aspect matriciel (celui qui prévaut dans les applications) est privilégié.

Pour les fonctions convexes différentiables, on pourra consulter [12], Chapitre IV, Section 4, par exemple.

* ***Exercice I.1.*** Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ continûment différentiable.

1°) Soit x_0 tel que $\nabla f(x_0) \neq 0$. Que représente $\nabla f(x_0)$ pour la surface de niveau $S := \{x \in \mathbb{R}^n \mid f(x) = f(x_0)\}$?

2°) Rappeler l'équation de l'hyperplan affine (de $\mathbb{R}^{n+1}$) tangent au graphe de f en $(x_0, f(x_0))$.

Donner à l'aide de $\nabla f(x_0)$ un vecteur normal à cet hyperplan.

3°) On suppose qu'il existe $L > 0$ tel que

$$\parallel \nabla f(x) - \nabla f(x') \parallel \leqslant L \parallel x - x' \parallel \text{ pour tout } (x, x') \in \mathbb{R}^n \times \mathbb{R}^n.$$

Montrer qu'alors

$$|f(x + d) - f(x) - \langle \nabla f(x), d \rangle| \leqslant \frac{L}{2} \parallel d \parallel^2 \text{ pour tout } (x, d) \in \mathbb{R}^n \times \mathbb{R}^n.$$

<u>*Solution :*</u> 1°) $\nabla f(x_0)$ est un vecteur normal à S en x_0 ; le sous-espace tangent à S en x_0 est orthogonal à $\nabla f(x_0)$.

2°) Soit $\mathrm{gr}f := \{(x, f(x)) \mid x \in \mathbb{R}^n\}$ le graphe de f (partie de $\mathbb{R}^{n+1}$ donc). Le sous-espace affine tangent à $\mathrm{gr}f$ en $(x_0, f(x_0))$ a pour équation

$$y = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle.$$

Un vecteur normal à cet hyperplan est fourni par $(\nabla f(x_0), -1) \in \mathbb{R}^n \times \mathbb{R}$.

On peut voir aussi $\mathrm{gr}f$ comme la surface de niveau 0 de la fonction

$$g : (x, r) \in \mathbb{R}^n \times \mathbb{R} \longmapsto g(x, r) := f(x) - r.$$

Comme $\nabla g\,(x_0, f(x_0)) = (\nabla f(x_0), -1)$ n'est jamais nul, le vecteur $(\nabla f(x_0), -1)$ est normal à $\mathrm{gr}f$ en $(x_0, f(x_0))$.

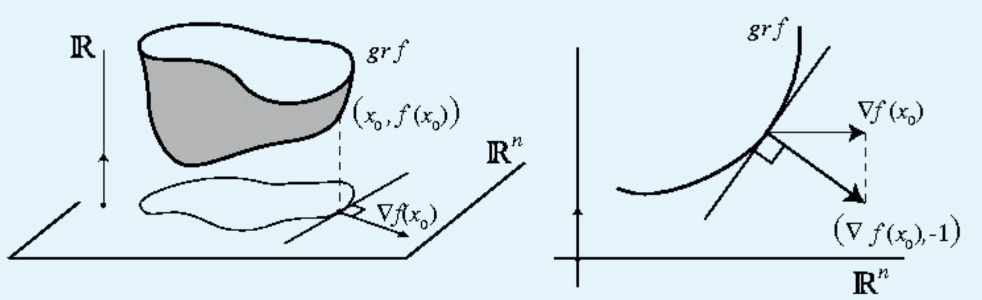

FIGURE 1.

3°) f étant continûment différentiable, on a grâce à la formule de Taylor à l'ordre zéro avec reste sous forme d'intégrale

$$f(x + d) = f(x) + \int_0^1 \langle \nabla f(x + td), d \rangle dt.$$

Sachant que $\langle \nabla f(x), d \rangle = \displaystyle\int_0^1 \langle \nabla f(x), d \rangle dt$, on transforme l'expression précédente en

$$f(x + d) - f(x) - \langle \nabla f(x), d \rangle = \int_0^1 \langle \nabla f(x + td) - \nabla f(x), d \rangle dt.$$

Il s'ensuit grâce à l'hypothèse faite sur ∇f :

$$\begin{aligned}|\langle \nabla f\,(x+td) - \nabla f(x), d\rangle| &\leqslant \|\, \nabla f\,(x+td) - \nabla f(x)\,\| \cdot \|\, d\,\| \\ &\leqslant tL\,\|\, d\,\|^2,\end{aligned}$$

d'où :

$$|f\,(x+d) - f(x) - \langle \nabla f(x), d\rangle| \leqslant \int_0^1 (tL\,\|\, d\,\|^2)dt = \frac{L}{2}\,\|\, d\,\|^2 .$$

Commentaire :

– Prolongement de l'exercice. On suppose à présent que f est deux fois différentiable sur $\mathbb{R}^n$. Comment traduire alors en termes de $\nabla^2 f$ l'hypothèse dans la 3e question ? Proposer ensuite une démonstration de l'inégalité de la 3e question qui s'appuie sur une autre formule de Taylor.

– Il importe de bien assimiler ce que représentent géométriquement les objets $\nabla f(x)$ et $\nabla^2 f(x)$ vis-à-vis des surfaces de niveau et du graphe de f.

**** Exercice I.2.***

Situations :	À déterminer :
1°) $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ $x \longmapsto f(x) := \langle c, x\rangle + \gamma.$	$\nabla f(x), \nabla^2 f(x)$ en tout point x.
2°) $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ $x \longmapsto F(x) := Lx + b$, où $L \in \mathcal{M}_{m,n}(\mathbb{R})$.	$JF(x)$.
3°) $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ $x \longmapsto f(x) := \frac{1}{2}\langle Ax, x\rangle + \langle d, x\rangle + \delta$, où $A \in \mathcal{M}_n(\mathbb{R})$.	$\nabla f(x), \nabla^2 f(x)$.
4°) $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ $x \longmapsto g(x) := \sum_{i=1}^{m} [r_i(x)]^2$, où les $r_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ sont deux fois différentiables.	$\nabla g(x), \nabla^2 g(x)$.

Commentaire : Ne pas se lancer tête baissée dans le calcul de dérivées partielles !

Pour les fonctions de base que sont les fonctions affines, quadratiques, sommes de carrés, etc., il faut savoir déterminer $\nabla f(x), \nabla^2 f(x)$ rapidement et sans erreur, « mécaniquement » presque.

<u>*Solution :*</u>

1°) $\nabla f(x) = c$, $\nabla^2 f(x) = 0$ pour tout $x \in \mathbb{R}^n$.

2°) $JF(x) = L$ (L est la partie linéaire de la fonction affine F).

3°) $\nabla f(x) = \frac{1}{2}\left(A + A^\top\right)x + d$, $\nabla f(x) = Ax + d$ si A est symétrique.

$\nabla^2 f(x) = \frac{1}{2}\left(A + A^\top\right)$, $\nabla^2 f(x) = A$ si A est symétrique.

4°) $\nabla g(x) = 2\sum_{i=1}^{m} r_i(x)\nabla r_i(x)$,

$$\nabla^2 g(x) = 2\sum_{i=1}^{m}\left\{\nabla r_i(x)\left[\nabla r_i(x)\right]^\top + r_i(x)\nabla^2 r_i(x)\right\}.$$

Ainsi, si $(r_1(x), \ldots, r_m(x))$ est « proche de 0 dans $\mathbb{R}^m$ », on peut considérer que $2\sum_{i=1}^{m}\nabla r_i(x)\left[\nabla r_i(x)\right]^\top$ est une « bonne » approximation de $\nabla^2 g(x)$; approximation qui, de plus, ne fait appel qu'aux gradients des fonctions r_i.

*** *Exercice* I.3.**

Situations :

1°) $g : \mathbb{R}^n \longrightarrow \mathbb{R}$

$$x \longmapsto g(x) := \sum_{i=1}^{m}\left[f_i^+(x)\right]^2,$$

où les $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ sont deux fois différentiables et a^+ désigne max $(0, a)$.

Questions :

g est-elle différentiable deux fois différentiable ?

2°)

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "A"] \arrow[dr, "g := f \circ A"'] & \mathbb{R}^m \arrow[d, "f"] \\ & \mathbb{R} \end{tikzcd}$$

où $A : u \longmapsto Au = A_0 u + b$, avec $A_0 \in \mathcal{M}_{m,n}(\mathbb{R})$, et f deux fois différentiable.

Déterminer $\nabla g, \nabla^2 g$ en fonction des éléments correspondants ∇f et $\nabla^2 f$.

<u>*Application*</u> : Soit : $\mathbb{R}^n \longrightarrow \mathbb{R}$ deux fois différentiable, $x_0 \in \mathbb{R}^n$, $d \in \mathbb{R}^n$. On pose $\varphi(t) := f(x_0 + td)$.

Calculer $\varphi'(t)$ (resp. $\varphi''(t)$) en fonction de $\nabla f(x_0 + td)$ (resp. $\nabla^2 f(x_0 + td)$).

J. JENSEN (1859–1925). Mathématicien danois, autodidacte. Il n'occupa aucune position universitaire mais fut longtemps expert dans une compagnie téléphonique de Copenhague. L'inégalité de convexité qui porte son nom (*cf.* Chapitre VI) date de 1906. La tradition de convexité chez les mathématiciens scandinaves est commémorée par la flamme de la poste danoise ci-dessous.

F. JOHN (1910–1994) (on trouve souvent écrit Fritz John) est aussi un mathématicien de ce siècle qui a établi les conditions qui portent son nom (*cf.* Chapitre III) en 1948. Le travail de F. John était motivé par des questions de nature géométrique ; les ellipsoïdes pleins de volume extrémal mis en évidence dans les Exercices III.21 et III.22 sont d'ailleurs appelés ellipsoïdes de John (ou de Loewner-John). On doit aussi au mathématicien d'origine tchèque Ch. Loewner (ou K. Löwner) l'ordre partiel $\succeq$ dans $\mathcal{S}_n(\mathbb{R})$.

H. MINKOWSKI (1864–1909). Mathématicien allemand considéré comme le fondateur de la géométrie des nombres entiers ; il s'est intéressé aussi à la Physique mathématique (les fameux espaces de Minkowski de dimension 4) et a, le premier, procédé à une étude systématique de la convexité dans les espaces de dimension finie.

J. (VON) NEUMANN (1903–1957). Il y a plusieurs mathématiciens du nom de Neumann. Celui-ci, Von Neumann (John) est un mathématicien américain d'origine hongroise, mort relativement jeune, et qui fut pionnier dans ses idées et recherches en Mathématiques appliquées. Et quand il attaquait un problème, il n'enfonçait pas des portes ouvertes ! On lui doit notamment un théorème de mini-maximisation (*cf.* Chapitre IV).

INDEX