

# Extrema

Christophe Antonini<sup>1</sup>, Olivier Teytaud<sup>2</sup>, Pierre Borgnat<sup>3</sup>, Annie Chateau<sup>4</sup>, and  
Edouard Lebeau<sup>5</sup>

<sup>1</sup>Enseignant en CPGE, Institut Stanislas, Cannes

<sup>2</sup>Chargé de recherche INRIA, Université d'Orsay, Orsay

<sup>3</sup>Chargé de recherche CNRS, ENS Lyon, Lyon

<sup>4</sup>Maitre de conférence, Université Montpellier-2, Montpellier

<sup>5</sup>Enseignant en CPGE, Lycée Henri Poincaré, Nancy

28 décembre 2021



A propos des extrema...

## 1 Extrema

Les extrema sont l'exemple naturel de mathématiques directement utiles. Si vous avez un processus industriel dont les paramètres  $\theta$  entraînent un coût  $f(\theta)$ , alors vous pouvez naturellement être intéressés par minimiser  $f(\theta)$ ; vous cherchez donc des extrema de  $f$ . Après quelques généralités (1.1), nous nous pencherons sur la compacité (1.2), utile pour garantir l'existence d'extrema, puis sur le calcul différentiel (1.3) qui nous permettra de caractériser les extrema, et enfin sur la notion de convexité (1.4), qui permet de montrer qu'un optimum local est global, sans oublier une section pour ceux qui veulent aller plus loin (1.5). Quelques éléments plus pratiques sur la recherche d'extrema sont donnés en section ??.

### 1.1 Cadre et définitions

Pour ce chapitre, on travaillera sur une application  $f$  continue d'un ouvert  $U$  d'un espace de Banach  $E$  dans  $\mathbb{R}$ .

#### DÉFINITION 0.1 **minimum relatif**

On dit que  $f$  admet un **minimum relatif** ou **minimum local** en  $x \in U$  si il existe un voisinage  $V$  de  $x$  tel que pour tout  $v$  dans  $V$   $f(x) \leq f(v)$ .

On dit que  $f$  admet un **minimum relatif strict** ou **minimum local strict** en  $x \in U$  si il existe un voisinage  $V$  de  $x$  tel que pour tout  $v \neq x$  dans  $V$   $f(x) < f(v)$ .

On dit que  $f$  admet un **minimum global** en  $x \in U$  si pour tout  $v$  dans  $U$   $f(x) \leq f(v)$ .

On dit que  $f$  admet un **minimum global strict** en  $x \in U$  si pour tout  $v \neq x$  dans  $U$   $f(x) < f(v)$ .

On définit de même les notions de **maximum relatif**, **maximum relatif strict**, **maximum global**, **maximum global strict**, en remplaçant les  $\leq$  par des  $\geq$  et les  $<$  par des  $>$ .

## 1.2 Résultats liés à la compacité

### PROPOSITION 0.1

Soit  $E$  un espace vectoriel normé de dimension finie et  $f \in C^0$  de  $E$  dans  $\mathbb{R}$  telle que

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

Alors  $f$  est minorée et atteint son minimum.

**Démonstration** On se donne  $A > 0$  tel que  $\|x\| > A$  implique  $f(x) > f(0)$ . On considère alors  $K = \overline{B}(0, A)$ .  $K$  est fermé car on est en dimension finie (les compacts d'un espace vectoriel normé de dimension finie sont les fermés bornés).  $f$  atteint donc sa borne inf (voir corollaire ??).

La condition 0.1 et d'autres conditions s'y rapportant sont parfois qualifiés de coercivité.

### COROLLAIRE 0.2

Quelques applications :

- La distance d'un point à un fermé non vide est minorée et le minimum est atteint.
- Aussi les deux applications suivantes, empruntées à [1] :

— étant donnée une application  $f \in C^0$  de  $[0, 1]$  dans  $\mathbb{R}$ , il existe un polynôme  $P$  minimisant  $\|f - P\|_\infty$  parmi les polynômes de degré  $\leq n$ .

— le **théorème de D'Alembert-Gauss**, stipulant que tout polynôme à coefficients complexes et de degré  $\geq 1$  admet une racine, en considérant  $z \mapsto |P(z)|$  (corollaire : tout polynôme à coefficients dans  $\mathbb{C}$  est scindé dans  $\mathbb{C}[X]$ ).

*Application 0.1* Voir le corollaire ?? sur la trigonalisation de matrices complexes, ou la partie ?? sur les suites récurrentes linéaires.

## 1.3 Résultats de calcul différentiel

Quand une fonction est dérivable, chercher un minimum local revient à chercher un point en lequel les dérivées (1ère et/ou seconde) vérifient certaines conditions ; d'où l'importance de caractérisations locales (par calcul différentiel) des minima ou maxima d'une fonction.

Il est à noter que l'optimisation sous contraintes (cf section ?? pour quelques mots sur ce sujet) donne lieu à des conditions bien différentes.

### 1.3.1 Résultats au premier ordre

#### THÉORÈME 0.3 Condition nécessaire du premier ordre

Si  $x$  est un minimum relatif de  $f$  et si  $f$  est différentiable en  $x$ , alors la différentielle de  $f$  en  $x$  est nulle.

#### Démonstration

$$df(x)(h) \geq 0 \text{ car } \frac{f(x+th)-f(x)}{t} \geq 0 \text{ si } t \geq 0$$

$$df(x)(h) \leq 0 \text{ car } \frac{f(x+th)-f(x)}{t} \leq 0 \text{ si } t \leq 0$$

⚠ Attention 0.2 Pas de réciproque; par exemple  $x \mapsto x^3$  de  $\mathbb{R}$  dans  $\mathbb{R}$  a une différentielle nulle en 0 et n'a ni maximum ni minimum en zéro.

#### DÉFINITION 0.2

Si  $df(x) = 0$ , on dit que  $x$  est un point critique.

Pour aller plus loin on peut s'intéresser aux extréma liés; voir pour cela [2].

### 1.3.2 Résultats du second ordre

#### THÉORÈME 0.4 Condition nécessaire du second ordre

Soit  $f$  deux fois différentiable en  $x$ .

Alors si  $f$  admet un minimum local en  $x$ ,  $f''(x)$  est positive.

#### Démonstration

Supposons  $f''(x)$  non positive; alors il existe  $v$  tel que  $f''(x)(v, v)$  soit  $< 0$ . On se donne un voisinage  $V$  de  $x$  sur lequel  $f \leq f(x)$  et sur lequel la formule de Taylor-Young ?? donne

$$f(x + t.v) \leq f(x) + \frac{1}{4}f''(x)(tv, tv)$$

(rappelons que par le résultat précédent  $f'(x)$  est nul)

on peut toujours choisir un tel voisinage  $V$  car

$$f(x + t.v) = f(x) + \frac{1}{2}f''(x)(tv, tv) + o(t^2)$$

pour  $t$  assez petit

et donc  $f(x + tv) - f(x)$  est alors négatif pour ces valeurs de  $t$  (à part pour  $t = 0$ ).

#### THÉORÈME 0.5 Condition suffisante du second ordre

Supposons  $E$  de dimension finie. Soit  $f$  deux fois différentiables en  $x$ ,  $f'(x) = 0$ , et  $f''(x)$  définie positive. Alors  $f$  admet un minimum relatif strict en  $x$ .

**Démonstration** On considère simplement un minimum de  $f''(x)(u, u)$  pour  $\|u\| = 1$  et la conclusion vient rapidement.

**Intuition** On peut se passer d'hypothèse de dimension finie à condition d'imposer que  $f''(x)$  vérifie  $\exists \alpha > 0; f''(x)(u, u) > \alpha \|u\|^2$ .

## 1.4 La convexité

Pour une introduction à la convexité, voir la partie ??.

Les résultats liés à la convexité sont très intuitifs, et se justifient rigoureusement sans trop de difficulté : on pourra consulter [2] pour moult développements. Le théorème suivant est très naturel :

### THÉORÈME 0.6

Un minimum local d'une fonction convexe définie sur une partie convexe est en fait un minimum global.

Une fonction strictement convexe définie sur une partie convexe admet au plus un minimum et si un tel minimum existe il est strict.

Bien sûr, cela ne nous dit pas comment traiter de problèmes non-convexes. Divers algorithmes permettent de trouver des minima de fonctions en très grande dimension et en temps raisonnable, lorsque le problème est très convexe; le cas général, non convexe, est bien sûr beaucoup plus délicat. Il est facile de montrer que l'algorithme consistant à simplement tirer au sort  $n$  points dans  $\mathbb{R}^d$  suivant une distribution gaussienne et à sélectionner le meilleur de ces points trouve l'essentiel minimum de toute fonction mesurable à mesure que  $n \rightarrow \infty$  (au sens où  $\inf_{i \in \{1, \dots, n\}} f(X_i)$  où les  $X_i$  sont aléatoires simples tend presque sûrement vers le minimum essentiel de  $f$ , quelle que soit la fonction  $f$ ). L'inconvénient de cet algorithme est bien sûr sa lenteur. Une façon de concilier cet algorithme et des méthodes plus usuelles de descente de gradient ou de Newton est d'effectuer plusieurs algorithmes de Newton (ou quasi-Newton) à partir de points initiaux différents. On peut aussi se référer à l'important champ des algorithmes évolutionnaires qui essaient de faire mieux (plus rapide) que la recherche aléatoire ci-dessus, tout en en préservant la robustesse.

## 1.5 Pour aller plus loin

La partie ?? contient quelques développements, dont l'indispensable méthode de Newton.

La notion de quasi-convexité, plus générale que la convexité, est parfois utile en recherche d'extrema; une fonction  $f$  est quasi-convexe si  $\{x; f(x) < t\}$  est convexe quel que soit  $t$ . En particulier, certaines méthodes d'optimisation, dites méthodes de recherche directe (on trouvera plus facilement "direct search methods" sur internet), parviennent à trouver les minima de fonctions quasi-convexes sous des hypothèses très générales.

On peut aussi s'intéresser aux extrêmes d'une fonction, sous contraintes. On pourra pour cela consulter [2]. On n'a pas ici formulé les caractérisations par calcul différentiel de l'optimalité locale d'un point en termes de conditions sur le jacobien ou la hessienne. Le lecteur est encouragé à se référer à la section ?? pour y découvrir la hessienne, à la section ?? pour y découvrir jacobienne et jacobien, et à [2] pour en apprendre plus sur le sujet de l'énorme lien entre algèbre linéaire (et en particulier matricielle) et optimisation. En résumé :

- le gradient est nul si la jacobienne est nulle.
- la dérivée seconde  $f''(x)$  est définie positive si la matrice hessienne de  $f$  en  $x$  est définie positive (ce qui, rappelons-le, est beaucoup plus fort que le fait que le hessien soit strictement positif).

## Références

- [1] A. Pommellet, *Cours d'analyse*, Ellipses 1994.
- [2] P.G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Dunod, 1998.