Chapitre 3

Optimisation

3.1 Définitions et rappels

3.1.1 Extrema, points critiques et points selle.

L'objectif de ce chapitre est de rechercher des extrema, c'est-à-dire des minima ou des maxima d'une fonction $f \in C(\mathbb{R}^n, \mathbb{R})$ avec ou sans contrainte. Notons que la recherche d'un minimum ou d'un maximum implique que l'on ait une relation d'ordre, pour pouvoir comparer les valeurs prises par f. On insiste donc bien sur le fait que la fonction f est à valeurs dans $\mathbb{R}$ (et non pas $\mathbb{R}^n$, comme dans le chapitre précédent). Rappelons tout d'abord quelques définitions du cours de calcul différentiel.

Définition 3.1 (Extremum d'une fonction)**.** *Soit E un espace vectoriel normé et $f : E \to \mathbb{R}$.*
On dit que $\bar{x}$ est un minimum local de f s'il existe un voisinage V de $\bar{x}$ tel que

$$f(\bar{x}) \leq f(x), \forall x \in V.$$

De même, on dit que $\bar{x}$ est un maximum local de f s'il existe un voisinage V de $\bar{x}$ tel que

$$f(\bar{x}) \geq f(x), \forall x \in V.$$

On dit que $\bar{x}$ est un extremum local de f si c'est un minimum local ou un maximum local.
On dit que $\bar{x}$ est un minimum global de f si

$$f(\bar{x}) \leq f(x), \forall x \in E.$$

De même, on dit que $\bar{x}$ est un maximum global de f si

$$f(\bar{x}) \geq f(x), \forall x \in E.$$

On dit que $\bar{x}$ est un extremum global de f si c'est un minimum global ou un maximum global.

Le problème d'optimisation sans contrainte s'écrit :

$$\left\{ \begin{array}{l} \text{Trouver } \bar{x} \in \mathbb{R}^n \text{ tel que :} \\ f(\bar{x}) \leq f(y), \quad \forall y \in \mathbb{R}^n. \end{array} \right. \tag{3.1}$$

Le problème d'optimisation avec contrainte s'écrit :

$$\left\{ \begin{array}{l} \text{Trouver } \bar{x} \in K \text{tel que :} \\ f(\bar{x}) \leq f(y), \quad \forall y \in K. \end{array} \right. \tag{3.2}$$

où $K \subset \mathbb{R}^n$ et $K \neq \mathbb{R}^n$ L'ensemble K où l'on recherche la solution est donc l'ensemble qui représente les contraintes. Par exemple, si l'on cherche un miminum d'une fonction f de $\mathbb{R}$ dans $\mathbb{R}$ et que l'on demande que les points qui réalisent ce minimum soient positifs, on aura $K = \mathbb{R}_+$.
Si $\bar{x}$ est solution du problème (3.1), on dit que $\bar{x} \in \arg\min_{\mathbb{R}^n} f$, et si $\bar{x}$ est solution du problème (3.2), on dit que $\bar{x} \in \arg\min_K f$.
Vous savez déjà que si un point $\bar{x}$ réalise le minimum d'une fonction f dérivable de $\mathbb{R}$ dans $\mathbb{R}$, alors $f'(\overline{x}) = 0$. On dit que c'est un point critique (voir définition 3.2). La réciproque est évidemment fausse : la fonction $x \mapsto x^3$ est dérivable sur $\mathbb{R}$, et sa dérivée s'annule en 0 qui est donc un point critique, mais 0 n'est pas un extremum (c'est un point d'inflexion). Nous verrons plus loin que de manière générale, lorsque la fonctionnelle f est différentiable, les extrema sont des points critiques de f, au sens où ils annulent le gradient.

Définition 3.2 (Point critique)**.** *Soit E un espace vectoriel normé et $f : E \to \mathbb{R}$ différentiable. On dit que $x \in E$ est un point critique de f si $Df(x) = 0$.*

Pour illustrer un cas de point critique qui n'est pas un maximum ni un minimum, prenons un exemple en dimension 2, avec

$$f(x_1, x_2) = x_1^2 - x_2^2.$$

On a alors

$$Df(x_1, x_2)(h_1, h_2) = 2(x_1h_1 - x_2h_2) \text{ et } Df(0, 0) = 0.$$

Le point $(0, 0)$ est donc un point critique de f. Si on trace la surface $x \mapsto x_1^2 - x_2^2$, on se rend compte que le point $(0, 0)$ est minimal dans une direction et maximal dans une direction indépendante de la première. C'est ce qu'on appelle un point selle

Définition 3.3 (Point selle)**.** *Soit E un espace vectoriel normé et $f : E \to \mathbb{R}$. On dit que $\bar{x}$ est un point selle de f s'il existe F et G des sous espaces vectoriels de E tels que $E = F \oplus G$ et un voisinage V de $\bar{x}$ tel que*

$$f(\bar{x} + z) \leq f(\bar{x}), \forall z \in F\,; \bar{x} + z \in V,$$
$$f(\bar{x} + z) \geq f(\bar{x}), \forall z \in G\,; \bar{x} + z \in V.$$

3.1.2 Convexité

Définition 3.4 (Convexité)**.** *Soit E un espace vectoriel (sur $\mathbb{R}$) et $f : E \to \mathbb{R}$. On dit que f est convexe si*

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) \text{ pour tout } (x, y) \in E^2 \text{ et } t \in [0, 1].$$

On dit que f est strictement convexe si

$$f(tx + (1 - t)y) < tf(x) + (1 - t)f(y) \text{ pour tout } (x, y) \in E^2 \text{ t.q. } x \neq y \text{ et } t \in]0, 1[.$$

Proposition 3.5 (Première caractérisation de la convexité)**.** *Soit E un espace vectoriel normé (sur $\mathbb{R}$) et $f \in C^1(E, \mathbb{R})$ alors :*

1. *la fonction f est convexe si et seulement si $f(y) \geq f(x) + Df(x)(y-x)$, pour tout couple $(x,y) \in E^2$,*
2. *la fonction f est strictement convexe si et seulement si $f(y) > f(x) + Df(x)(y-x)$ pour tout couple $(x,y) \in E^2$ tel que $x \neq y$.*

DÉMONSTRATION – *Démonstration de 1.*
($\Rightarrow$) Supposons que f est convexe : soit $(x,y) \in E^2$; on veut montrer que $f(y) \geq f(x) + Df(x)(y-x)$. Soit $t \in [0,1]$, alors $f(ty + (1-t)x) \leq tf(y) + (1-t)f(x)$ grâce au fait que f est convexe. On a donc :

$$f(x + t(y-x)) - f(x) \leq t(f(y) - f(x)). \tag{3.3}$$

Comme f est différentiable, $f(x + t(y-x)) = f(x) + Df(x)(t(y-x)) + t\varepsilon(t)$ où $\varepsilon(t)$ tend vers 0 lorsque t tend vers 0. Donc en reportant dans (3.3),

$$\varepsilon(t) + Df(x)(y-x) \leq f(y) - f(x), \quad \forall t \in]0,1[.$$

En faisant tendre t vers 0, on obtient alors :

$$f(y) \geq Df(x)(y-x) + f(x).$$

($\Leftarrow$) Montrons maintenant la réciproque : Soit $(x,y) \in E^2$, et $t \in]0,1[$ (pour $t = 0$ ou $= 1$ on n'a rien à démontrer). On veut montrer que $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$. On pose $z = tx + (1-t)y$. On a alors par hypothèse :

$$\begin{aligned} f(y) &\geq f(z) + Df(z)(y-z), \\ \text{et } f(x) &\geq f(z) + Df(z)(x-z). \end{aligned}$$

En multipliant la première inégalité par $1-t$, la deuxième par t et en les additionnant, on obtient :

$$\begin{aligned} (1-t)f(y) + tf(x) &\geq f(z) + (1-t)Df(z)(y-z) + tDf(z)(x-z) \\ (1-t)f(y) + tf(x) &\geq f(z) + Df(z)((1-t)(y-z) + t(x-z)). \end{aligned}$$

Et comme $(1-t)(y-z) + t(x-z) = 0$, on a donc $(1-t)f(y) + tf(x) \geq f(z) = f(tx + (1-t)y)$.

Démonstration de 2
($\Rightarrow$) On suppose que f est strictement convexe, on veut montrer que $f(y) > f(x) + Df(x)(y-x)$ si $y \neq x$. Soit donc $(x,y) \in E^2$, $x \neq y$. On pose $z = \frac{1}{2}(y-x)$, et comme f est convexe, on peut appliquer la partie 1. du théorème et écrire que $f(x+z) \geq f(x) + Df(x)(z)$. On a donc $f(x) + Df(x)(\frac{y-x}{2}) \leq f(\frac{x+y}{2})$. Comme f est strictement convexe, ceci entraîne que $f(x) + Df(x)(\frac{y-x}{2}) < \frac{1}{2}(f(x) + f(y))$, d'où le résultat.
($\Leftarrow$) La méthode de démonstration est la même que pour le 1. ∎

Proposition 3.6 (Seconde caractérisation de la convexité)**.** *Soit $E = \mathbb{R}^n$ et $f \in C^2(E,\mathbb{R})$. Soit $H_f(x)$ la hessienne de f au point x, i.e. $(H_f(x))_{i,j} = \partial^2_{i,j} f(x)$. Alors*

1. *f est convexe si et seulement si $H_f(x)$ est symétrique et positive pour tout $x \in E$ (c.à.d. $H_f(x)^t = H_f(x)$ et $H_f(x)y \cdot y \geq 0$ pour tout $y \in \mathbb{R}^n$)*
2. *f est strictement convexe si $H_f(x)$ est symétrique définie positive pour tout $x \in E$. (Attention la réciproque est fausse.)*

DÉMONSTRATION – *Démonstration de 1.*
($\Rightarrow$) Soit f convexe, on veut montrer que $H_f(x)$ est symétrique positive. Il est clair que $H_f(x)$ est symétrique car $\partial^2_{i,j} f = \partial^2_{j,i} f$ car f est C^2. Par définition, $H_f(x) = D(\nabla f(x))$ et $\nabla f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Soit $(x,y) \in E^2$, comme f est convexe et de classe C^1, on a, grâce à la proposition 3.5 :

$$f(y) \geq f(x) + \nabla f(x) \cdot (y-x). \tag{3.4}$$

Soit $\varphi \in C^2(\mathbb{R}, \mathbb{R})$ définie par $\varphi(t) = f(x + t(y-x))$. Alors :

$$f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt = [\varphi'(t)(t-1)]_0^1 - \int_0^1 \varphi''(t)(t-1)dt,$$

c'est–à dire : $f(y) - f(x) = \varphi'(0) + \int_0^1 \varphi''(t)(1-t)dt$. Or $\varphi'(t) = \nabla f(x + t(y-x)) \cdot (y-x)$, et

$$\varphi''(t) = D(\nabla f(x + t(y-x))(y-x) \cdot (y-x) = H_f(x + t(y-x))(y-x) \cdot (y-x).$$

On a donc :

$$f(y) - f(x) = \nabla f(x)(y-x) + \int_0^1 H_f(x + t(y-x))(y-x) \cdot (y-x)(1-t)dt. \tag{3.5}$$

Les inégalités (3.4) et (3.5) entraînent : $\int_0^1 H_f(x+t(y-x))(y-x)\cdot(y-x)(1-t)dt \geq 0\ \forall x,y \in E$. On a donc :

$$\int_0^1 H_f(x+tz)z\cdot z(1-t)dt \geq 0 \quad \forall x, \forall z \in E. \tag{3.6}$$

En fixant $x \in E$, on écrit (3.6) avec $z = \varepsilon y$, $\varepsilon > 0$, $y \in \mathbb{R}^n$. On obtient :

$$\varepsilon^2 \int_0^1 H_f(x+t\varepsilon y)y\cdot y(1-t)dt \geq 0 \quad \forall x,y \in E, \quad \forall \varepsilon > 0, \text{ et donc :}$$

$$\int_0^1 Hf(x+t\varepsilon y)y\cdot y(1-t)dt \geq 0 \quad \forall \varepsilon > 0.$$

Pour $(x,y) \in E^2$ fixé, $H_f(x+t\varepsilon y)$ tend vers $H_f(x)$ uniformément lorsque $\varepsilon \to 0$, pour $t \in [0,1]$. On a donc :

$$\int_0^1 H_f(x)y\cdot y(1-t)dt \geq 0, \text{ c.à.d. } \frac{1}{2}H_f(x)y\cdot y \geq 0.$$

Donc pour tout $(x,y) \in (\mathbb{R}^n)^2$, $H_f(x)y\cdot y \geq 0$ donc $H_f(x)$ est positive.

($\Leftarrow$) Montrons maintenant la réciproque : On suppose que $H_f(x)$ est positive pour tout $x \in E$. On veut démontrer que f est convexe ; on va pour cela utiliser la proposition 3.5 et montrer que : $f(y) \geq f(x) + \nabla f(x)\cdot(y-x)$ pour tout $(x,y) \in E^2$. Grâce à (3.5), on a :

$$f(y) - f(x) = \nabla f(x)\cdot(y-x) + \int_0^1 H_f(x+t(y-x))(y-x)\cdot(y-x)(1-t)dt.$$

Or $H_f(x+t(y-x))(y-x)\cdot(y-x) \geq 0$ pour tout couple $(x,y) \in E^2$, et $1-t \geq 0$ sur $[0,1]$. On a donc $f(y) \geq f(x) + \nabla f(x)\cdot(y-x)$ pour tout couple $(x,y) \in E^2$. La fonction f est donc bien convexe.

Démonstration de 2.

($\Leftarrow$) On suppose que $H_f(x)$ est strictement positive pour tout $x \in E$, et on veut montrer que f est strictement convexe. On va encore utiliser la caractérisation de la proposition 3.5. Soit donc $(x,y) \in E^2$ tel que $y \neq x$. Alors :

$$f(y) = f(x) + \nabla f(x)\cdot(y-x) + \int_0^1 \underbrace{H_f(x+t(y-x))(y-x)\cdot(y-x)}_{>0 \text{ si } x\neq y}\ \underbrace{(1-t)}_{\neq 0 \text{ si } t\in]0,1[}\ dt.$$

Donc $f(y) > f(x) + \nabla f(x)(y-x)$ si $x \neq y$, ce qui prouve que f est strictement convexe. ∎

Contre-exemple Pour montrer que la réciproque de 2. est fausse, on propose le contre-exemple suivant : Soit $n = 1$ et $f \in C^2(\mathbb{R},\mathbb{R})$, on a alors $H_f(x) = f''(x)$. Si f est la fonction définie par $f(x) = x^4$, alors f est strictement convexe mais $f''(0) = 0$.

3.2 Optimisation sans contrainte

3.2.1 Définition et condition d'optimalité

Soit $f \in C(E, \mathbb{R})$ et E un espace vectoriel normé. On cherche $\bar{x}$ minimum global de f, c.à.d. :

$$\bar{x} \in E \text{ tel que } f(\bar{x}) \leq f(y) \ \ \forall y \in E, \tag{3.7}$$

ou un minimum local, c.à.d. :

$$\bar{x} \text{ tel que } \exists \alpha > 0 \ \ f(\bar{x}) \leq f(y) \ \ \forall y \in B(\bar{x}, \alpha). \tag{3.8}$$

Proposition 3.7 (Condition nécessaire d'optimalité)**.**
Soit E un espace vectoriel normé, et soient $f \in C(E, \mathbb{R})$, et $\bar{x} \in E$ tel que f est différentiable en $\bar{x}$. Si $\bar{x}$ est solution de (3.8) alors $Df(\bar{x}) = 0$.

DÉMONSTRATION – Supposons qu'il existe $\alpha > 0$ tel que $f(\bar{x}) \leq f(y)$ pour tout $y \in B(\bar{x}, \alpha)$. Soit $z \in E \setminus \{0\}$, alors si $|t| < \frac{\alpha}{\|z\|}$, on a $\bar{x} + tz \in B(\bar{x}, \alpha)$ (où $B(\bar{x}, \alpha)$ désigne la boule ouverte de centre $\bar{x}$ et de rayon α) et on a donc $f(\bar{x}) \leq f(\bar{x} + tz)$. Comme f est différentiable en $\bar{x}$, on a :

$$f(\bar{x} + tz) = f(\bar{x}) + Df(\bar{x})(tz) + |t|\varepsilon_z(t),$$

où $\varepsilon_z(t) \to 0$ lorsque $t \to 0$. On a donc $f(\bar{x}) + tDf(\bar{x})(z) + |t|\varepsilon_z(t) \geq f(\bar{x})$. Et pour $\dfrac{\alpha}{\|z\|} > t > 0$, on a $Df(\bar{x})(z) + \varepsilon_z(t) \geq 0$. En faisant tendre t vers 0, on obtient que

$$Df(\bar{x})(z) \geq 0, \quad \forall z \in E.$$

On a aussi $Df(\bar{x})(-z) \geq 0 \ \ \forall z \in E$, et donc : $-Df(\bar{x})(z) \geq 0 \ \ \forall z \in E$.
On en conclut que

$$Df(\bar{x}) = 0.$$

■

Remarque 3.8. Attention, la proposition précédente donne une condition nécessaire mais non suffisante. En effet, $Df(\bar{x}) = 0$ n'entraîne pas que f atteigne un minimum (ou un maximum) même local, en $\bar{x}$. Prendre par exemple $E = \mathbb{R}$, $\overline{x} = 0$ et la fonction f définie par : $f(x) = x^3$ pour s'en convaincre.

3.2.2 Résultats d'existence et d'unicité

Théorème 3.9 (Existence)**.** *Soit $E = \mathbb{R}^n$ et $f : E \to \mathbb{R}$ une application telle que*
(i) f est continue,
(ii) $f(x) \to +\infty$ quand $\|x\| \to +\infty$.
Alors il existe $\bar{x} \in \mathbb{R}^n$ tel que $f(\bar{x}) \leq f(y)$ pour tout $y \in \mathbb{R}^n$.

DÉMONSTRATION – La condition (ii) peut encore s'écrire

$$\forall A \in \mathbb{R}, \quad \exists R \in \mathbb{R}; \|x\| \geq R \Rightarrow f(x) \geq A. \tag{3.9}$$

On écrit (3.9) avec $A = f(0)$. On obtient alors :

$$\exists R \in \mathbb{R} \text{ tel que } \|x\| \geq R \Rightarrow f(x) \geq f(0).$$

On en déduit que $\inf_{\mathbb{R}^n} f = \inf_{B_R} f$, où $B_R = \{x \in \mathbb{R}^n; |x| \leq R\}$. Or, B_R est un compact de $\mathbb{R}^n$ et f est continue donc il existe $\bar{x} \in B_R$ tel que $f(\bar{x}) = \inf_{B_R} f$ et donc $f(\bar{x}) = \inf_{\mathbb{R}^n} f$. ■

Remarque 3.10.

1. Le théorème est faux si E est un espace de Banach (c'est-à-dire un espace vectoriel normé complet) de dimension infinie car, dans ce cas, la boule fermée B_R n'est pas compacte.
2. L'hypothèse (ii) du théorème peut être remplacée par
$$(ii)' \qquad \exists b \in \mathbb{R}^n, \exists R > 0 \text{ tel que } \|x\| \geq R \Rightarrow f(x) \geq f(b).$$
3. Sous les hypothèses du théorème il n'y a pas toujours unicité de $\bar{x}$ même dans le cas $n = 1$, prendre pour s'en convaincre la fonction f définie de $\mathbb{R}$ dans $\mathbb{R}$ par $f(x) = x^2(x-1)(x+1)$.

Théorème 3.11 (Condition suffisante d'unicité)**.** *Soit E un espace vectoriel normé et $f : E \to \mathbb{R}$ strictement convexe alors il existe au plus un $\bar{x} \in E$ tel que $f(\bar{x}) \leq f(y), \forall y \in E$.*

DÉMONSTRATION – Soit f strictement convexe, supposons qu'il existe $\bar{x}$ et $\bar{\bar{x}} \in E$ tels que $f(\bar{x}) = f(\bar{\bar{x}}) = \inf_{\mathbb{R}^n} f$. Comme f est strictement convexe, si $\bar{x} \neq \bar{\bar{x}}$ alors
$$f(\frac{1}{2}\bar{x} + \frac{1}{2}\bar{\bar{x}}) < \frac{1}{2}f(\bar{x}) + \frac{1}{2}f(\bar{\bar{x}}) = \inf_{\mathbb{R}^n} f,$$
ce qui est impossible ; donc $\bar{x} = \bar{\bar{x}}$. ∎

Ce théorème ne donne pas l'existence. Par exemple dans le cas $n = 1$ la fonction f définie par $f(x) = e^x$ n'atteint pas son minimumn ; en effet, $\inf_{\mathbb{R}^n} f = 0$ et $f(x) \neq 0$ pour tout $x \in \mathbb{R}$, et pourtant f est strictement convexe. Par contre, si on réunit les hypothèses des théorèmes 3.9 et 3.11, on obtient le résultat d'existence et unicité suivant :

Théorème 3.12 (Existence et unicité)**.** *Soit $E = \mathbb{R}^n$, et soit $f : E \to \mathbb{R}$. On suppose que :*

(i) f continue,

(ii) $f(x) \to +\infty$ quand $\|x\| \to +\infty$,

(iii) f est strictement convexe ;

alors il existe un unique $\bar{x} \in \mathbb{R}^n$ tel que $f(\bar{x}) = \inf_{\mathbb{R}^n} f$.

L'hypothèse (i) du théorème 3.12 est en fait inutile car une fonction convexe de $\mathbb{R}^n$ dans $\mathbb{R}$ est nécessairement continue.
Nous donnons maintenant des conditions suffisantes d'existence et d'unicité du minimum pour une fonction de classe C^1.

Proposition 3.13 (Conditions suffisantes d'existence et unicité)**.** *Soit $f \in C^1(\mathbb{R}^n, \mathbb{R})$. On suppose que :*

$$\exists \alpha > 0; (\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha \|x - y\|^2, \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \tag{3.10a}$$
$$\exists M > 0; \|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|, \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{3.10b}$$

Alors :

1. *f est strictement convexe,*
2. *$f(x) \to +\infty$ quand $|x| \to +\infty$,*

et en conséquence, il existe un unique $\bar{x} \in \mathbb{R}^n$ tel que $f(\bar{x}) = \inf_{\mathbb{R}^n} f$.

DÉMONSTRATION –

1. Soit φ la fonction définie de $\mathbb{R}$ dans $\mathbb{R}^n$ par : $\varphi(t) = f(x + t(y - x))$. Alors
$$f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \nabla f(x + t(y - x)) \cdot (y - x) dt,$$
On en déduit que
$$f(y) - f(x) - \nabla f(x) \cdot (y - x) = \int_0^1 (\nabla f(x + t(y - x)) \cdot (y - x) - \nabla f(x) \cdot (y - x)) dt,$$
c'est–à–dire :
$$f(y) - f(x) - \nabla f(x) \cdot (y - x) = \int_0^1 \underbrace{(\nabla f(x + t(y - x)) - \nabla f(x)) \cdot (y - x)}_{\geq \alpha t|y-x|^2} dt.$$
Grâce à l'hypothèse (3.10a) sur f, ceci entraîne :
$$f(y) - f(x) - \nabla f(x) \cdot (y - x) \geq \alpha \int_0^1 t|y - x|^2 dt = \frac{\alpha}{2}|y - x|^2 > 0 \text{ si } y \neq x. \tag{3.11}$$
On a donc, pour tout $(x, y) \in E^2$, $f(y) > f(x) + \nabla f(x) \cdot (y - x)$; d'après la première caractérisation de la convexité, voir proposition 3.5, on en déduit que f est strictement convexe.
2. Montrons maintenant que $f(y) \to +\infty$ quand $|y| \to +\infty$. On écrit (3.11) pour $x = 0$: $f(y) \geq f(0) + \nabla f(0) \cdot y + \frac{\alpha}{2}|y|^2$. Comme $\nabla f(0) \cdot y \geq -|\nabla f(0)|(y)$, on a donc
$$f(y) \geq f(0) + |y| \left(\frac{\alpha}{2}|y| - |\nabla f(0)|\right) \to +\infty \text{ quand } |y| \to +\infty.$$

La fonction f vérifie donc bien les hypothèses du théorème 3.30, et on en déduit qu'il existe un unique $\bar{x}$ qui minimise f.

■

Remarque 3.14 (Généralisation à un espace de Hilbert)**.** Le théorème 3.12 reste vrai si E est un espace de Hilbert ; on a besoin dans ce cas pour la partie existence des hypothèses (i), (ii) et de la convexité de f.

Proposition 3.15 (Caractérisation des points tels que $f(\bar{x}) = \inf_E f$)**.** *Soit E espace vectoriel normé et f une fonction de E dans $\mathbb{R}$. On suppose que $f \in C^1(E, \mathbb{R})$ et que f est convexe. Soit $\bar{x} \in E$. Alors :*
$$f(\bar{x}) = \inf_E f \Leftrightarrow Df(\bar{x}) = 0.$$
En particulier si $E = \mathbb{R}^n$ alors $f(\bar{x}) = \inf_{x \in \mathbb{R}^n} f(x) \Leftrightarrow \nabla f(\bar{x}) = 0$.

Démonstration

($\Rightarrow$) Supposons que $f(\bar{x}) = \inf_E f$ alors on sait (voir Proposition 3.7) que $Df(\bar{x}) = 0$ (la convexité est inutile).

($\Leftarrow$) Si f est convexe et différentiable, d'après la proposition 3.5, on a : $f(y) \geq f(\bar{x}) + Df(\bar{x})(y - x)$ pour tout $y \in E$ et comme par hypothèse $Df(\bar{x}) = 0$, on en déduit que $f(y) \geq f(\bar{x})$ pour tout $y \in E$. Donc $f(\bar{x}) = \inf_E f$.

Cas d'une fonction quadratique On appelle fonction quadratique une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par
$$\boldsymbol{x} \mapsto f(\boldsymbol{x}) = \frac{1}{2} A\boldsymbol{x} \cdot \boldsymbol{x} - \boldsymbol{b} \cdot \boldsymbol{x} + c, \tag{3.12}$$
où $A \in \mathcal{M}_n(\mathbb{R})$, $\boldsymbol{b} \in \mathbb{R}^n$ et $c \in \mathbb{R}$. On peut vérifier facilement que $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$. Calculons le gradient de f et sa hessienne : on a
$$\begin{aligned}
f(\boldsymbol{x} + \boldsymbol{h}) &= \frac{1}{2} A(\boldsymbol{x} + \boldsymbol{h}) \cdot (\boldsymbol{x} + \boldsymbol{h}) - \boldsymbol{b} \cdot (\boldsymbol{x} + \boldsymbol{h}) + c \\
&= \frac{1}{2} A\boldsymbol{x} \cdot \boldsymbol{x} + \frac{1}{2} A\boldsymbol{x} \cdot \boldsymbol{h} + \frac{1}{2} A\boldsymbol{h} \cdot \boldsymbol{x} + \frac{1}{2} A\boldsymbol{h} \cdot \boldsymbol{h} - \boldsymbol{b} \cdot \boldsymbol{x} - \boldsymbol{b} \cdot \boldsymbol{h} + c \\
&= f(\boldsymbol{x}) + \frac{1}{2}(A\boldsymbol{x} \cdot \boldsymbol{h} + A\boldsymbol{h} \cdot \boldsymbol{x}) - \boldsymbol{b} \cdot \boldsymbol{h} + \frac{1}{2} A\boldsymbol{h} \cdot \boldsymbol{h} \\
&= f(\boldsymbol{x}) + \frac{1}{2}(A\boldsymbol{x} + A^t\boldsymbol{x}) \cdot \boldsymbol{h} - \boldsymbol{b} \cdot \boldsymbol{h} + \frac{1}{2} A\boldsymbol{h} \cdot \boldsymbol{h}.
\end{aligned}$$

Et comme $|A\boldsymbol{h}\cdot\boldsymbol{h}| \leq \|A\|_2\,|\boldsymbol{h}|^2$, on en déduit que :

$$\nabla f(\boldsymbol{x}) = \frac{1}{2}(A\boldsymbol{x} + A^t\boldsymbol{x}) - \boldsymbol{b}. \tag{3.13}$$

Si A est symétrique, on a donc $\nabla f(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}$. Calculons maintenant la hessienne de f. D'après (3.13), on a :

$$\nabla f(\boldsymbol{x}+\boldsymbol{h}) = \frac{1}{2}(A(\boldsymbol{x}+\boldsymbol{h}) + A^t(\boldsymbol{x}+\boldsymbol{h})) - \boldsymbol{b} = \nabla f(\boldsymbol{x}) + \frac{1}{2}(A\boldsymbol{h} + A^t\boldsymbol{h})$$

et donc $H_f(\boldsymbol{x}) = D(\nabla f(\boldsymbol{x})) = \frac{1}{2}(A + A^t)$. On en déduit que si A est symétrique, $H_f(\boldsymbol{x}) = A$. Dans le cas où A est symétrique définie positive, f est donc strictement convexe.
De plus on a $f(\boldsymbol{x}) \to +\infty$ quand $|\boldsymbol{x}| \to +\infty$. (On note comme d'habitude $|\cdot|$ la norme euclidienne de $\boldsymbol{x}$.) En effet,

$$A\boldsymbol{x}\cdot\boldsymbol{x} \geq \alpha|\boldsymbol{x}|^2 \text{ où } \alpha \text{ est la plus petite valeur propre de } A, \text{ et } \alpha > 0.$$

Donc

$$f(\boldsymbol{x}) \geq \frac{\alpha}{2}|\boldsymbol{x}|^2 - |\boldsymbol{b}\cdot\boldsymbol{x}| - |c|;$$

Mais comme $|\boldsymbol{b}\cdot\boldsymbol{x}| \leq |\boldsymbol{b}||\boldsymbol{x}|$, on a

$$f(\boldsymbol{x}) \geq |\boldsymbol{x}|\left(\frac{\alpha|\boldsymbol{x}|}{2} - |\boldsymbol{b}|\right) - |c| \longrightarrow +\infty \text{ quand } |\boldsymbol{x}| \to +\infty.$$

On en déduit l'existence et l'unicité de $\bar{\boldsymbol{x}}$ qui minimise f. On a aussi :

$$\nabla f(\bar{\boldsymbol{x}}) = 0 \Leftrightarrow f(\bar{\boldsymbol{x}}) = \inf_{\mathbb{R}^n} f$$

et donc $\bar{\boldsymbol{x}}$ est l'unique solution du système $A\boldsymbol{x} = \boldsymbol{b}$.
On en déduit le théorème suivant, très important, puisqu'il va nous permettre en particulier le lien entre certains algorithmes d'optimisation et les méthodes de résolution de systèmes linéaires vues au chapitre 1.

Théorème 3.16 (Minimisation d'une fonction quadratique)**.** *Soit f une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par* (3.12) *où $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice symétrique définie positive et $\boldsymbol{b} \in \mathbb{R}^n$. Alors il existe un unique $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ qui minimise f, et $\bar{\boldsymbol{x}}$ est l'unique solution du système linéaire $A\boldsymbol{x} = \boldsymbol{b}$.*