Delphine Bresch-Pietri
Nicolas Petit

Optimisation

Cycle ingénieur civil
http://cas.ensmp.fr/~dbp/

MINES ParisTech
Année scolaire 2021-2022
ECUE 21.1

TABLE DES MATIÈRES

Notes On pourra se référer à [**22**] pour un exposé de certaines librairies et de certains logiciels existants pour la résolution des problèmes d'optimisation. On pourra se reporter à [**2**] pour tout détail portant sur le chapitre 1 du cours. En ce qui concerne l'analyse et l'optimisation convexe on pourra se reporter à [**15**] et [**6**] ou encore, pour plus de détails sur les notions avancées d'analyse convexe et les conditions d'optimalité, à [**17, 6, 3, 21, 28**]. On trouvera de plus amples détails sur les algorithmes de descente en dimension finie dans [**24**]. En ce qui concerne les méthodes numériques, on trouvera de plus amples détails sur les méthodes de sous-gradient dans [**4**]. Une vue générale des méthodes de faisceaux est fournie dans [**16, 19**]. Les méthodes proximales sont quant à elles exposées en détails dans [**23, 1**].

Enfin, l'ouvrage [**9**] contient une présentation très complète de l'optimisation.

CHAPITRE 1

CONDITIONS D'OPTIMALITÉ ET ANALYSE CONVEXE

Dans ce chapitre, on se place dans $\mathbb{R}^n$ ($n \in \mathbb{N}$) considéré comme un espace vectoriel normé muni de la norme euclidienne notée $\|\cdot\|$. On considérera des fonctions continues $f : \mathbb{R}^n \mapsto \mathbb{R}$. On notera Ω un ouvert de $\mathbb{R}^n$.

1.1 Notions fondamentales

***Définition* 1**. — Soit l'application $x \in \Omega \subset \mathbb{R}^n \mapsto f(x) \in \mathbb{R}$. f présente un minimum local en $x^* \in \Omega$ si et seulement s'il existe un voisinage V de x^* tel que $\forall x \in V \cap \Omega$, $f(x^*) \leq f(x)$ (on parle de minimum local strict si la précédente inégalité est stricte). On dit alors que x^* est solution locale du problème

$$\min_{x \in \Omega} f(x) \tag{1}$$

On ne considérera ici que des problèmes de minimisation (calculs de minimum), les problèmes de maximisation étant obtenus en considérant l'opposé de la fonction f.

***Définition* 2**. — La fonction $f : \Omega \subset \mathbb{R}^n \mapsto f(x) \in \mathbb{R}$ admet en $x^* \in \Omega$ un minimum global de la fonction si $\forall x \in \Omega$ on a $f(x^*) \leq f(x)$ (on parle de minimum strict global si la précédente inégalité est stricte). On dit alors que x^* est solution globale du problème (1).

***Définition* 3**. — Soit f différentiable. Pour $x = (x_1, ..., x_n) \in \mathbb{R}^n$, on note $(\nabla f(x))^T = \frac{\partial f}{\partial x}(x) = \left(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}\right)(x)$ le gradient de f en x.

***Définition* 4**. — Soit f deux fois différentiable. On appelle Hessien de f la matrice $\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{1 \leq i,j \leq n}$ qui est une matrice symétrique de $\mathcal{M}_n(\mathbb{R})$. On dit que $\nabla^2 f(x^*)$ est positive (resp. definie positive) et on note $\nabla^2 f(x^*) \geq 0$ (resp. $\nabla^2 f(x^*) > 0$) si toutes ses valeurs propres sont positives (resp. strictement positives).

***Définition* 5**. — Soit $g : \mathbb{R}^n \to \mathbb{R}^m$ différentiable. On appelle Jacobien de $g = (g_1, \dots g_m)^T$ la matrice $\frac{\partial g}{\partial x}(x) = \left(\frac{\partial g_i}{\partial x_j}(x)\right)_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathcal{M}_{(m \times n)}(\mathbb{R})$.

***Définition* 6**. — On dit que x^* est un point stationnaire de f si $\nabla f(x^*) = 0$.

1.2 Existence d'un minimum

Les théorèmes suivants garantissent l'existence d'une solution au problème de recherche de minimum.

***Théorème* 1 (Weierstrass)**. — *Si f est une fonction réelle continue sur un compact $K \subset \mathbb{R}^n$ alors le problème de recherche de minimum global*

$$\min_{x \in K} f(x)$$

possède une solution $x^* \in K$.

Démonstration. — Notons $m = \inf_{x \in K} f(x)$ (éventuellement $m = -\infty$). Par définition, $\forall x \in K, f(x) \geq m$. Soit $(x^k)_{k \in \mathbb{N}}$ suite d'éléments de K telle que $(f(x^k))_{k \in \mathbb{N}}$ converge vers m. K est compact, on peut donc extraire une sous-suite $(x^l)_{l \in \mathcal{I} \subset \mathbb{N}}$ convergeant vers $x^* \in K$. Par continuité de f, on a

$$m = \lim_{k \to +\infty} f(x^k) = \lim_{l \to +\infty, l \in \mathcal{I}} f(x^l) = f(x^*)$$

Or $f(x^*) > -\infty$ car f est continue et K compact. En conclusion, il existe $x^* \in K$ tel que $f(x^*) = \min_{x \in K} f(x)$. □

***Théorème* 2**. — *Soit $f : \mathbb{R}^n \to \mathbb{R}$ continue et telle que $\lim_{\|x\| \to \infty} f(x) = +\infty$. Alors, il existe une solution au problème $\min_{x \in \mathbb{R}^n} f(x)$.*

Démonstration. — f tendant vers $+\infty$ en l'infini, il existe $R > 0$ tel que $f(x) \geq f(0)$ pour tout x tel que $\|x\| > R$. La fonction f étant continue, elle admet et atteint un minimum sur la boule fermé $\overline{B(0,R)}$ qui est un compact. Or, $f(x) > f(0)$ sur $\mathbb{R}^n \setminus \overline{B(0,R)}$ et ce minimum est donc solution de $\min_{x \in \mathbb{R}^n} f(x)$. □

1.3 Condition nécessaire d'optimalité sur un ouvert

Théorème 3

Soit Ω un ouvert de $\mathbb{R}^n$. Une condition nécessaire pour que x^* soit un minimiseur local de $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}$ fonction deux fois différentiable est

$$\left\{\nabla f(x^*) = 0, \nabla^2 f(x^*) \geq 0\right\}$$

Démonstration. — $x^* \in \Omega$ ouvert, donc $\forall h \in \mathbb{R}^n$, $\exists \eta > 0$ tel que $(x^* + \delta h) \in \Omega$ pour $0 \leq \delta \leq \eta$. Pour η suffisamment petit on a alors $f(x^* + \delta h) \geq f(x^*)$ par hypothèse. f est différentiable, on peut utiliser un développement de Taylor

$$f(x^*) + \frac{\partial f}{\partial x}(x^*)\delta h + \circ(\delta) \geq f(x^*)$$

d'où

$$\frac{\partial f}{\partial x}(x^*)h + \circ(\delta)/\delta \geq 0$$

Par passage à la limite lorsque δ tend vers zéro on obtient

$$\frac{\partial f}{\partial x}(x^*)h \geq 0 \text{ pour tout } h \in \mathbb{R}^n$$

Nécessairement $\frac{\partial f}{\partial x}(x^*) = 0$. Utilisons maintenant un développement au deuxième ordre

$$f(x^* + \delta h) = f(x^*) + \frac{1}{2}(\delta h)^T \nabla^2 f(x^*)\delta h + \circ(\delta^2)$$

Dans le même voisinage que précédemment $f(x^* + \delta h) \geq f(x^*)$ implique après passage à la limite que pour tout $h \in \mathbb{R}^n$

$$h^T \nabla^2 f(x^*)h \geq 0$$

et donc que $\nabla^2 f(x^*) \geq 0$. □

1.4 Condition suffisante d'optimalité sur un ouvert

Théorème 4

Une condition suffisante pour que x^* soit un minimiseur local de $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}$ fonction deux fois différentiable sur Ω ouvert de $\mathbb{R}^n$ est

$$\left\{\nabla f(x^*) = 0, \nabla^2 f(x^*) > 0\right\}$$

.

Démonstration. — Le même raisonnement par un développement de Taylor du deuxième ordre s'applique. □

1.5 Importance de la convexité.

Il existe une large famille de problèmes d'optimisation pour lesquels on sait caractériser plus avant les solutions et fournir des algorithmes de résolution efficace. Cette famille est celle des problèmes convexes. Nous présentons donc ici les éléments d'analyse convexe nécessaires à leur étude.

1.5.1 Propriétés géométriques

Définition 7. — E ensemble de $\mathbb{R}^n$ est convexe si

$$\forall (x, y) \in E \times E \quad \forall \lambda \in [0, 1] \quad \lambda x + (1 - \lambda)y \in E$$

Définition 8

Soit E convexe de $\mathbb{R}^n$. On dit que l'application $f : E \to \mathbb{R}$ est convexe si

$$\forall (x,y) \in E \times E \quad \forall \lambda \in [0,1] \quad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

***Théorème* 5**. — *Soient E convexe de $\mathbb{R}^n$ et $f : E \to \mathbb{R}$ continue. Les deux propositions suivantes sont équivalentes :*

(i) *f convexe*
(ii) $\mathrm{Epi}(f) = \{(x,y) \mid x \in E\,,\ y \geq f(x)\}$ *est convexe*

Par ailleurs, dans le cas où E° (l'intérieur de E) est non vide, ceci est équivalent à

(iii) $\forall x \in E^\circ \quad \exists\, \alpha_x \in \mathbb{R}^n \quad \forall y \in E \quad f(y) \geq f(x) + \alpha_x^T (y-x)$

L'hyperplan $\big\{(y,z) \in E \times \mathbb{R} \mid z = f(x) + \alpha_x^T (y-x)\big\}$ défini en (ii) est appelé hyperplan d'appui de f au point x. L'ensemble $\mathrm{Epi}(f)$ *est appelé épigraphe de f.*

Démonstration. — On raisonne par double implication.

$\underline{(i) \Rightarrow (ii)}$: Soient (x_1, y_1) et (x_2, y_2) deux éléments de $\mathrm{Epi}(f)$ et soit $\lambda \in [0,1]$. On a alors $\lambda y_1 + (1-\lambda) y_2 \geq \lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2)$, où la dernière inégalité provient de la convexité de f. Par conséquent, on a $\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2) \in \mathrm{Epi}(f)$.

$\underline{(ii) \Rightarrow (i)}$: Direct.

$\underline{(iii) \Rightarrow (ii)}$: On considère $S = \cap_{x_0 \in E^\circ} \big\{(x,y) \in E^\circ \times \mathbb{R} \mid y \geq f(x_0) + \alpha_{x_0}^T (x - x_0)\big\}$ et $\tilde{f} : x \in E^\circ \mapsto f(x)$. S est l'intersection d'ensembles convexes et est donc convexe. Si $(x,y) \in \mathrm{Epi}(\tilde{f})$, alors $y \geq f(x)$ et donc $(x,y) \in S$ selon (ii). Respectivement, si $(x,y) \in S$, alors, pour $x_0 = x$, on a $y \geq f(x)$ et donc $(x,y) \in \mathrm{Epi}(\tilde{f})$. Ainsi, $\mathrm{Epi}(\tilde{f}) = S$ convexe. Or $\overline{\mathrm{Epi}(\tilde{f})} = \mathrm{Epi}(f)$ car f continue. Donc, $\mathrm{Epi}(f)$ convexe.

$\underline{(ii) \Rightarrow (iii) :}$ Soit $x_0 \in E$. On considère l'ensemble $S_1 = \{x_0, f(x_0)\}$ qui est un convexe non-vide. Soit S_2 l'intérieur de $\mathrm{Epi}(f)$, qui est un convexe ouvert non vide et tel que $S_1 \cap S_2 = \emptyset$. Par théorème de séparation des convexes (voir Appendice A.3), il existe $a \in \mathbb{R}^{n+1}$ $(a \neq 0)$ et b dans $\mathbb{R}$ tels que $a^T(x_0, f(x_0)) \leq b$ et $a^T(x,y) \geq b$ pour $(x,y) \in S_2$. En choisissant une suite de points (x_k, y_k) de S_2 qui converge vers $(x_0, f(x_0))$, on obtient également $a^T(x_0, f(x_0)) \geq b$, soit finalement $a^T(x_0, f(x_0)) = b$. Aussi, pour tout $(x,y) \in S_2$ et plus généralement pour tout $(x,y) \in \mathrm{Epi}(f)$ en passant à la limite, $a^T(x - x_0, f(x) - f(x_0)) \geq 0$, dont on déduit (ii) sous réserve que $a_{n+1} \neq 0$. Par l'absurde, supposons que $a_{n+1} = 0$. Alors, en notant $\tilde{a} = (a_1, \ldots, a_n)$, pour tout $x \in E$, $\tilde{a}^T(x - x_0) \geq 0$. Or, E ouvert, donc il existe $\varepsilon > 0$ telle que la boule ouverte $B(x_0, \varepsilon)$ soit incluse dans E. Ainsi, pour tout $v \in \mathbb{R}^n$, en choisissant $x = x_0 + \varepsilon v$ et $x = x_0 + \varepsilon v$, on conclut que $\tilde{a}^T \varepsilon v = 0$ soit $\tilde{a}^T v = 0$. Ceci étant vrai pour tout $v \in \mathbb{R}^n$, on conclut que $a = 0$, d'où la contradiction. □

Noter qu'une fonction continue et convexe peut ne pas être différentiable. On peut néanmoins montrer qu'elle est différentiable presque partout (cf. Appendice A.2).

1.5.2 Sous-différentiel

Définition 9

Soit $f : E \to \mathbb{R}$ convexe. Un vecteur $v \in \mathbb{R}^n$ est appelé sous-gradient de f au point $x_0 \in \mathbb{R}^n$ si

$$\forall x \in E \quad f(x) \geq f(x_0) + v^T(x - x_0) \tag{2}$$

L'ensemble de tous les sous-gradients en x_0 est appelé sous-différentiel de f en x_0 et noté $\partial f(x_0)$

$$\partial f(x_0) = \left\{ v \in \mathbb{R}^n \mid \forall x \in E \quad f(x) \geq f(x_0) + v^T(x - x_0) \right\} \tag{3}$$

Alternativement, on dispose des caractérisations équivalentes suivantes :

- $\partial f(x_0) = \text{Conv}\left\{ \lim_{i\to\infty} \nabla f(x_i) \mid x_i \to x_0 \, , x_i \notin \Omega_f \right\}$ où Ω_f est l'ensemble des points dans un voisinage de x_0 où f n'est pas différentiable et Conv l'enveloppe convexe (c.-à-d., pour un ensemble E, $\text{Conv}(E)$ est l'ensemble convexe le plus petit contenant E);
- $\partial f(x_0) = \left\{ v \in \mathbb{R}^n \mid f'(x_0, p) \geq v^T p \, , \; p \in \mathbb{R}^n \right\}$ où l'on a introduit $f'(x_0, p) = \lim_{t\to 0^+} \frac{1}{t}(f(x_0 + tp) - f(x_0))$ qui est la dérivée directionnelle de f au point x_0 dans la direction p.

***Lemme* 1**. — *Soit $f : E \to \mathbb{R}$ convexe et soit $x_0 \in E^\circ$ (l'intérieur de E). Alors $\partial f(x_0)$ est un ensemble non vide convexe fermé et borné, c.-à-d. un compact convexe non-vide de $\mathbb{R}^n$.*

Démonstration. — Par caractérisation de la convexité par hyperplan d'appui, il existe au moins un élément dans $\partial f(x_0)$, qui est un fermé par caractérisation séquentielle. Soient v_1, v_2 deux éléments de $\partial f(x_0)$ et $\lambda \in [0, 1]$. Alors, pour tout $x \in E$,

$$f(x) \geq f(x_0) + v_1^T(x - x_0) \tag{4}$$

$$f(x) \geq f(x_0) + v_2^T(x - x_0) \tag{5}$$

En multipliant la première inégalité par λ, la seconde par $1 - \lambda$ et en sommant, on conclut que $\lambda v_1 + (1 - \lambda)v_2 \in \partial f(x_0)$, c.a.d. $\partial f(x_0)$ convexe.

Enfin, pour montrer que $\partial f(x_0)$ est borné, nous utilisons le fait que f est localement lipschitzienne sur l'intérieur de son domaine de définition (cf. Appendice A.1). Soit $v \in \partial f(x_0)$. f étant localement Lipschitzienne, il existe $\varepsilon, L > 0$ tels que, pour $x \in B(x_0, \varepsilon)$, $|f(x) - f(x_0)| \leq L\|x - x_0\|$. Soit $\varepsilon_0 > 0$ tel que $x = x_0 + \varepsilon_0 v \in B(x_0, \varepsilon)$. Alors

$$\varepsilon_0\|v\|^2 = v^T(x - x_0) \leq f(x) - f(x_0) \leq L\|x - x_0\| = L\varepsilon_0\|v\| \tag{6}$$

dont on déduit $\|v\| \leq L$. □

***Lemme* 2**. — *Soit $f : E \to \mathbb{R}$ une fonction convexe et $x_0 \in E^\circ$. Si f est différentiable en x_0, alors $\partial f(x_0) = \{\nabla f(x_0)\}$. Respectivement, si $\partial f(x_0)$ est le singleton $\{y\}$, alors f est différentiable en x_0 et $\nabla f(x_0) = y$.*

Ce lemme, dont la preuve de la dernière partie repose sur la version fonctionnelle du théorème de séparation de Hahn-Banach, est admis. On pourra se reporter à [**28**] [Theorem 25.1].

***Lemme* 3**. — *Soit f convexe. Alors le sous-différentiel ∂f est monotone, i.e.*

$$\forall (x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n \quad \forall (\eta_1, \eta_2) \in \partial f(x_1) \times \partial f(x_2) \quad (\eta_1 - \eta_2)^T (x_1 - x_2) \geq 0 \tag{7}$$

Démonstration. — Soient $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$ et $(\eta_1, \eta_2) \in \partial f(x_1) \times \partial f(x_2)$. On a, par caractérisation du sous-gradient,

$$\forall x \in \mathbb{R}^n \quad f(x) \geq f(x_1) + \eta_1^T (x - x_1) \tag{8}$$

$$\forall x \in \mathbb{R}^n \quad f(x) \geq f(x_2) + \eta_2^T (x - x_2) \tag{9}$$

En appliquant la première inégalité à $x = x_2$, la seconde à $x = x_1$ et en additionnant l'une à l'autre, on obtient le résultat voulu. □

***Lemme* 4**. — *Soient $f, g : E \to \mathbb{R}$ convexes, $h : \mathbb{R} \to \mathbb{R}$ convexe croissante et $(\alpha, \beta) \in \mathbb{R}_+^2$. Alors, pour tout $x \in E^\circ$,*

- $\partial(\alpha f + \beta g)(x) = \alpha \partial f(x) + \beta \partial g(x) = \{\alpha v_1 + \beta v_2 \mid (v_1, v_2) \in \partial f(x) \times \partial g(x)\}$
- $\partial(h \circ f)(x) = \{v_1 v_2 \mid v_1 \in \partial h(f(x)),\ v_2 \in \partial f(x)\}$

1.5.3 Conditions d'optimalité d'un problème convexe

Le sous-différentiel permet notamment de caractériser les minimiseurs.

Théorème 6

Soient $f : \mathbb{R}^n \to \mathbb{R}$ convexe. Alors x^* est un minimiseur global de f si et seulement si $0 \in \partial f(x^*)$.

Démonstration. — On a $0 \in \partial f(x^*)$ si et seulement si $\forall x \in \mathbb{R}^n f(x) \geq f(x^*)$, cad $x^\star$ minimiseur global de f. □

***Théorème* 7**. — *Soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe. Tout minimum local de f est global.*

Démonstration. — Soit x^* un minimiseur local de f. Par définition, il existe $\varepsilon > 0$ tel que, pour tout x tel que $\|x - x^*\| < \varepsilon$, $f(x) \geq f(x^*)$. Par ailleurs, pour tout $x \in \mathbb{R}^n$, on peut construire, par convexité de E, $x^\alpha = (1 - \alpha)x^* + \alpha x$, avec $\alpha \in]0, 1]$ tel que $\|x^\alpha - x^*\| < \epsilon$. On alors $f(x^*) \leq f(x^\alpha)$. Or, par convexité de la fonction f on a $f(x^\alpha) \leq (1 - \alpha) f(x^*) + \alpha f(x)$. Après simplification, il vient $f(x^*) \leq f(x)$. □

1.5.4 Cas réel

Dans le cas où $n = 1$, on dispose d'une caractérisation supplémenntaire à l'aide des pentes.

***Théorème* 8.** — *Soit I intervalle de $\mathbb{R}$ et soit $f : I \to \mathbb{R}$. Alors f convexe sur I si et seulement si, pour tout $x_0 \in I$, $x \in I \setminus \{x_0\} \mapsto \frac{f(x)-f(x_0)}{x-x_0}$ est croissante.*

En particulier, si f est dérivable, f convexe sur I si et seulement si f' est croissante.

Démonstration. — Soit $x_0 \in I$. Supposons que $]x_0, +\infty[\cap I \neq \emptyset$. Il vient

$$
\begin{aligned}
& x \in I \setminus \{x_0\} \mapsto \frac{f(x) - f(x_0)}{x - x_0} \text{ croissante sur }]x_0, +\infty[\cap I \\
& \quad \Leftrightarrow \forall x_2 \in]x_0, +\infty[\cap I \quad \forall x_1 \in]x_0, x_2[\quad \frac{f(x_1) - f(x_0)}{x_1 - x_0} \leq \frac{f(x_2) - f(x_0)}{x_2 - x_0} \\
& \quad \Leftrightarrow \forall x_2 \in]x_0, +\infty[\cap I \quad \forall x_1 \in [x_0, x_2] \quad f(x_1) \leq f(x_0) + \frac{x_1 - x_0}{x_2 - x_0} \left(f(x_2) - f(x_0)\right) \\
& \quad \Leftrightarrow \forall x_2 \in]x_0, +\infty[\cap I \quad \forall x_1 \in [x_0, x_2] \quad f(x_1) \leq \frac{x_2 - x_1}{x_2 - x_0} f(x_0) + \frac{x_1 - x_0}{x_2 - x_0} f(x_2) \\
& \quad \Leftrightarrow \forall x_2 \in]x_0, +\infty[\cap I \quad \forall \lambda \in [0, 1] \quad f(\lambda x_0 + (1 - \lambda) x_2) \leq \lambda f(x_0) + (1 - \lambda) f(x_2)
\end{aligned}
$$

où la dernière équivalence est obtenue avec $\lambda = \frac{x_2 - x_1}{x_2 - x_0}$. En procédant similairement sur $]-\infty, x_0[\cap I$, on conclut que

$$
\begin{aligned}
& x \in I \setminus \{x_0\} \mapsto \frac{f(x) - f(x_0)}{x - x_0} \text{ croissante sur } I \setminus \{x_0\} \\
& \quad \Leftrightarrow \forall x_2 \in I \quad \forall \lambda \in [0, 1] \quad f(\lambda x_0 + (1 - \lambda) x_2) \leq \lambda f(x_0) + (1 - \lambda) f(x_2)
\end{aligned}
$$

et le résultat recherché s'ensuit. □

1.5.5 Cas différentiable

Avec plus de régularité, on dispose de caractéristiques supplémentaires fréquemment employées.

Théorème 9

Soit f une application différentiable de Ω dans $\mathbb{R}$. Les propositions suivantes sont équivalentes

(i) f est convexe
(ii) $\forall (x, y) \in \Omega^2$, $f(y) \geq f(x) + (\nabla f(x))^T (y - x)$
(iii) ∇f est monotone : $\forall (x, y) \in \Omega^2 \quad (\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$

Et, si f deux fois différentiable, ces propriétés sont équivalentes à

(iv) $\forall x \in \Omega$, $\nabla^2 f(x) \geq 0$

Ce théorème peut en partie se déduire des propriétés précédentes du sous-différentiel. Néanmoins, nous lui fournissons ici une preuve standard plus directe.

Démonstration. — On raisonne par implications successives.
$(i) \Rightarrow (ii)$ Soit $(x, y) \in \Omega^2$ et $\lambda \in [0, 1]$. Alors, la convexité de f implique

$$\begin{aligned} f(y) \geq \frac{f(\lambda x + (1-\lambda)y)}{1-\lambda} - \frac{\lambda}{1-\lambda} f(x) =& \frac{f(x + (1-\lambda)(y-x)) - f(x)}{1-\lambda} + f(x) \\ & \underset{\lambda \to 1-}{\to} \nabla f(x)^T (y-x) + f(x) \end{aligned}$$

$(ii) \Rightarrow (iii)$ est le Lemme 3.
$(iii) \Rightarrow (i)$ Soient $(x, y) \in \Omega^2$ et $g : \lambda \in [0, 1] \mapsto f(\lambda x + (1-\lambda)y)$. L'application f étant différentiable, g est dérivable sur $]0, 1[$ et $g'(\lambda) = \nabla f(\lambda x + (1-\lambda)y)^T (x-y)$. Pour $(\lambda_1, \lambda_2) \in [0, 1]^2$, on obtient

$$\begin{aligned} (\lambda_1 - \lambda_2)[g'(\lambda_1) - g'(\lambda_2)] =& (\lambda_1 - \lambda_2) \Big[\nabla f(\lambda_1 x + (1-\lambda_1)y)^T (x-y) \\ & - \nabla f(\lambda_2 x + (1-\lambda_2)y)^T (x-y) \Big] \\ =& (\nabla(z_1) - \nabla f(z_2))^T (z_1 - z_2) \geq 0 \end{aligned}$$

en notant $z_1 = \lambda_1 x + (1-\lambda_1)y$ et $z_2 = \lambda_2 x + (1-\lambda_2)y$. Aussi, g' croissante et, selon le Théorème 8, g convexe. Par conséquent, pour tout $\lambda \in [0, 1]$,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) = g(\lambda) = g(\lambda \times 1 + (1-\lambda) \times 0) & \\ \leq \lambda g(1) + (1-\lambda) g(0) = \lambda f(x) + (1-\lambda) f(y) & \end{aligned}$$

c.-à-d. f convexe.
$(iii) \Leftrightarrow (iv)$ découle du fait que $\nabla f(x+h) = \nabla f(x) + \nabla^2 f(x) h + o(\|h\|)$. □

Théorème 10

Soit f une fonction convexe différentiable. Une condition nécessaire et suffisante pour que $x^* \in \Omega \subset \mathbb{R}^n$ avec Ω convexe soit un minimiseur global est que

$$\nabla f(x^*) = 0$$

1.5.6 Convexité stricte et convexité forte

Enfin, des propriétés plus fortes de convexité sont souvent utiles pour l'étude de la convergence de certains algorithmes de résolution de problèmes d'optimisation.

Définition 10

Soit $E \subset \mathbb{R}^n$ convexe. On dit que $f : E \to \mathbb{R}$ est fortement convexe (ou α-convexe) s'il existe $\alpha > 0$ tel que $f - \frac{\alpha}{2} \|.\|^2$ est convexe.

***Définition* 11**. — Soit $E \subset \mathbb{R}^n$ convexe. On dit que $f : E \to \mathbb{R}$ est strictement convexe si

$$\forall (x,y) \in E \quad \forall \lambda \in]0,1[\quad f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

Un certain nombre des propriétés précédentes trouvent leur pendant naturel pour les fonctions fortement convexes (ainsi que pour celles strictement convexes). Mentionnons-en deux qui seront plus utilisées par la suite.

***Théorème* 11**. — *Soit f une application différentiable de Ω dans $\mathbb{R}$, et $\alpha > 0$. Les propositions suivantes sont équivalentes*

- *f est α-convexe sur Ω*
- $\forall (x,y) \in \Omega^2 \quad f(y) \geq f(x) + (\nabla f(x))^T (y-x) + \frac{\alpha}{2} \|x-y\|^2$
- $\forall (x,y) \in \Omega^2 \quad \left((\nabla f(x))^T - (\nabla f(y))^T\right)(x-y) \geq \alpha \|x-y\|^2$

Et si f est deux fois différentiable, la proposition suivante est aussi équivalente

- $\forall x \in \Omega \quad \nabla^2 f(x) \geq \alpha I$ *(c.-à-d. que $\nabla^2 f(x) - \alpha I$ est positive, ou encore que les valeurs propres de $\nabla^2 f(x)$ sont supérieures à α).*

***Théorème* 12**. — *Soit $f : \mathbb{R}^n \to \mathbb{R}$ fortement convexe. Alors f admet un unique minimum (global).*

Démonstration. — L'existence d'un minimum découle du point (ii) du Théorème 11 précédent et du Théorème 2. Supposons maintenant que f admette un minimum en $x_1^\star$ et $x_2^\star$. Alors, selon la propriété précédente,

$$0 = \left((\nabla f(x_1^\star))^T - (\nabla f(x_2^\star))^T\right)(x_1^\star - x_2^\star) \geq \alpha \|x_1^\star - x_2^\star\|^2$$

et, donc, $x_1^\star = x_2^\star$. □

CHAPITRE 2

MÉTHODES POUR L'OPTIMISATION DIFFÉRENTIABLE DE DIMENSION FINIE

Dans ce chapitre, on s'intéresse majoritairement à des problèmes d'optimisation convexe faisant intervenir des fonctions différentiables.

2.1 Méthodes numériques de résolution sans contraintes

On détaille dans cette première sections des algorithmes de résolution du problème de minimisation présenté précédemment, à savoir

$$\min_{x\in\Omega} f(x)$$

2.1.1 Méthodes sans dérivées

Il est possible de résoudre des problèmes (simples) d'optimisation avec peu de connaissance de la fonction. On propose ainsi une méthode n'utilisant que la possibilité d'évaluer la fonction.

***Définition* 12**. — On dit que $f : \mathbb{R} \mapsto \mathbb{R}$ est unimodale sur un intervalle $[A, B]$ si elle admet un minimiseur x^* et si $\forall x_1 < x_2$ dans $[A, B]$ on a

1. $x_2 \leq x^*$ implique $f(x_1) > f(x_2)$
2. $x_1 \geq x^*$ implique $f(x_1) < f(x_2)$

Autrement dit, f possède un minimum global unique mais n'est ni nécessairement continue ni différentiable partout. Une conséquence de la propriété d'unimodalité est que si on divise l'intervalle $[A, B]$ en 4, il existe toujours un sous-intervalle qui ne contient pas l'optimum.

2.1.1.1 Méthode de dichotomie

Parmi les 4 intervalles précédents choisis de longueur égales on peut toujours en supprimer 2. L'intervalle de recherche est alors divisé par 2 à chaque itération.

2.1.1.2 Section dorée

Au lieu de subdiviser en 4 intervalles, on utilise 3 intervalles. À chaque itération on peut exclure l'un des 3 intervalles. De manière à réutiliser les calculs entre deux itérations, il faut choisir une méthode de découpage utilisant le nombre d'or. On vérifie qu'il faut diviser l'intervalle de recherche à l'itération i $[A_i, B_i]$ suivant $[A_i, A_i + (1 - \tau)(B_i - A_i), A_i + \tau(B_i - A_i), B_i]$ où $\tau = (\sqrt{5} - 1)/2$.

2.1.1.3 Comparaison de l'effort numérique

Les deux méthodes de division en sous-intervalles convergent vers l'unique minimiseur x^* quel que soit l'intervalle de recherche de départ $[A, B]$. On peut s'intéresser à leur efficacité en terme de nombre d'itérations. Pour obtenir un effort de réduction de l'intervalle de recherche initial de 10^{-2}, 10^{-3}, 10^{-6}, on doit faire le nombre d'appel à la fonction f suivant: 17, 23, 42 pour la méthode de dichotomie et 13, 18, 31 pour la méthode de la section dorée. La méthode de la section dorée est plus efficace mais moins triviale à mettre en œuvre que la méthode de dichotomie qui lui est souvent préférée.

2.1.2 Méthode utilisant la dérivée

On suppose désormais pouvoir calculer la valeur de la fonction ainsi que son gradient. Partant d'une valeur initiale x^0, on va mettre à jour une estimation x^k de l'optimum recherché x^*.

2.1.2.1 Idée des méthodes de descente

Entre deux itérations k et $k + 1$ on fait évoluer l'estimée de l'optimum x^k suivant la formule

$$x^{k+1} = x^k + l^k p^k$$

où $l^k \in \mathbb{R}$ est appelé pas et $p^k \in \mathbb{R}^n$ est une direction. Si par exemple on choisit $p^k = -\nabla f(x^k)$ on obtient alors pour une fonction différentiable

$$f(x^{k+1}) = f(x_k) - l^k \left\|\nabla f(x^k)\right\|^2 + \circ(l^k)$$

On peut donc espérer une décroissance des valeurs de la fonction entre deux itérations, d'où le nom de méthode de descente. On constate qu'une règle simple de pas constant produit souvent une bonne décroissance au début des itérations suivi d'une relative stagnation des valeurs de la fonction coût. En cherchant à améliorer la méthode, on réalise rapidement qu'il est important de bien choisir l^k.

2.1.2.2 Gradient à pas optimal

L'algorithme suivant possède de très intéressantes propriétés de convergence.

Algorithme 1 (Gradient à pas optimal)

À partir de $x^0 \in \mathbb{R}^n$ quelconque, itérer

$$x^{k+1} = x^k - l^k \nabla f(x^k) \quad \text{avec} \quad l^k = \text{argmin}_{l \in \mathbb{R}} f(x^k - l \nabla f(x^k)).$$

Théorème 13

Si f est α-convexe, différentiable et de gradient ∇f Lipschitzien sur tout borné, alors l'algorithme 1 du gradient à pas optimal converge vers l'unique solution x^* du problème d'optimisation $\min_{x\in\mathbb{R}^n} f(x)$.

Démonstration. — La fonction $\mathbb{R} \ni l \mapsto g(l) = f(x^k - l\nabla f(x^k))$ est α-convexe et dérivable sur $\mathbb{R}$. On suppose qu'à l'itération k, $\nabla f(x^k) \neq 0$. Le problème d'optimisation $\min_{l\in\mathbb{R}} f(x^k - l\nabla f(x^k))$ a une solution unique caractérisée par $\nabla g(l) = 0$. Cette équation s'écrit

$$\nabla f(x^k)^T \nabla f(x^k - l\nabla f(x^k)) = 0$$

On s'aperçoit ainsi que deux directions de descentes successives sont orthogonales. Entre deux itérations on a donc

$$\nabla f(x^{k+1})^T (x^{k+1} - x^k) = 0$$

car $l_k \neq 0$. On déduit alors de l'α-convexité de f

$$f(x^k) \geq f(x^{k+1}) + \frac{\alpha}{2} \left\|x^{k+1} - x^k\right\|^2$$

Donc

$$\frac{\alpha}{2} \left\|x^{k+1} - x^k\right\|^2 \leq f(x^k) - f(x^{k+1}) \tag{10}$$

La suite de réels $\left(f(x^k)\right)_{k\in\mathbb{N}}$ est décroissante, minorée par $f(x^*)$, elle est donc convergente. On en déduit que, par (10),

$$\lim_{k\to+\infty} \left\|x^{k+1} - x^k\right\| = 0$$

D'autre part, $\left(f(x^k)\right)_{k\in\mathbb{N}}$ est décroissante, minorée par $f(x^*)$, elle est bornée. Il existe donc $M \in \mathbb{R}$ tel que

$$M \geq f(x^k) \geq f(x^*) + \frac{\alpha}{2} \left\|x^k - x^*\right\|^2$$

où la dernière inégalité provient de l'α-convexité de f. La suite $\left(x^k\right)_{k\in\mathbb{N}}$ est donc bornée. Utilisons maintenant que le gradient ∇f est Lipschitzien sur tout borné. $\exists C_M > 0$ tel que $\left\|\nabla f(x^k) - \nabla f(x^{k+1})\right\| \leq C_M \left\|x^{k+1} - x^k\right\|$. Par l'orthogonalité entre les directions de descente entre deux pas successifs, il vient alors

$$\left(\left\|\nabla f(x^k)\right\|^2 + \left\|\nabla f(x^{k+1})\right\|^2\right)^{1/2} \leq C_M \left\|x^{k+1} - x^k\right\|$$

Par passage à la limite on a alors

$$\lim_{k\to+\infty} \left\|\nabla f(x^k)\right\| = 0$$

Enfin, en utilisant une dernière fois l'α-convexité de f, on a

$$\left\|\nabla f(x^k)\right\| \left\|x^k - x^*\right\| \geq \left(\nabla f(x^k) - \nabla f(x^*)\right)^T (x^k - x^*) \geq \alpha \left\|x^k - x^*\right\|^2$$

D'où

$$\left\|x^k - x^*\right\| \leq \frac{1}{\alpha} \left\|\nabla f(x^k)\right\|$$

Par passage à la limite on a alors

$$\lim_{k \to +\infty} x_k = x^*$$

□

L'algorithme du gradient à pas optimal possède donc une intéressante propriété de convergence mais comporte dans chaque itération une recherche de pas optimal. C'est un problème mono-dimensionnel qui peut être traité par les méthodes de dichotomie ou de section dorée. Cette résolution itérée peut être coûteuse en calculs et on lui préférera souvent une des méthodes de recherche linéaire présentée dans la section suivante.

2.1.3 Recherche linéaire

L'idée générale consiste à proposer une méthode de sélection du pas l^k qui permette de prouver la convergence

$$\lim_{k \to +\infty} \nabla f(x_k) = 0,$$

d'avoir une bonne vitesse de convergence, et qui soit simple à implémenter.

2.1.3.1 Conditions de Wolfe

Définition 13

On appelle **condition d'Armijo** (de paramètre c_1) sur les itérations $(x^k, p^k, l^k)_{k \in \mathbb{N}}$ l'inéquation

$$f(x^k + l^k p^k) \leq f(x^k) + c_1 l^k \nabla f(x^k)^T p^k \tag{11}$$

On appelle **condition de courbure** (de paramètre c_2) sur les itérations $(x^k, p^k, l^k)_{k \in \mathbb{N}}$ l'inéquation

$$\nabla f(x^k + l^k p^k)^T p^k \geq c_2 \nabla f(x^k)^T p^k \tag{12}$$

On appelle **conditions de Wolfe** les conditions (11) et (12) avec $0 < c_1 < c_2 < 1$.

La première condition autorise des pas grands mais pas trop (en pratique ceci évite les "rebonds" au cours des itérations). La deuxième condition permet d'éviter les pas trop petits.

***Théorème* 14**. — *Soit $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}$ différentiable. Soit p^k direction de descente en x^k telle que f est bornée inférieurement sur la droite $\{x_k + lp^k, l > 0\} \subset \mathbb{R}^n$. Il existe un intervalle $[l_1, l_2]$ tel que tout $l \in [l_1, l_2]$ satisfait les conditions de Wolfe.*

Démonstration. — La fonction $\mathbb{R}_+^\star \ni l \mapsto f(x^k + lp^k)$ est bornée inférieurement donc, pour tout $0 < c_1 < 1$, l'ensemble $\{f(x^k) + lc_1\nabla f(x^k)^T p^k, l > 0\} \subset \mathbb{R}$ contient $f(x^k + lp^k)$ pour un certain $l > 0$. Notons $l' > 0$ la plus petite valeur de l telle que $f(x^k) + lc_1\nabla f(x^k)^T p^k = f(x^k + lp^k)$. La condition d'Armijo (11) est toujours satisfaite pour $l < l'$. Par le théorème des accroissements finis appliqué à la fonction différentiable $[0, l'] \ni l \mapsto f(x^k + lp^k) \in \mathbb{R}$, $\exists l^{''} \in [0, l']$ tel que $f(x^k + l'p^k) - f(x^k) = l'\nabla f(x^k + l^{''}p^k)^T p^k$. On a alors

$$\nabla f(x^k + l^{''}p^k)^T p^k = c_1\nabla f(x^k)^T p^k > c_2\nabla f(x^k)^T p^k$$

Cette inégalité étant stricte, elle reste large dans un voisinage de $l^{''}$ ce qui garantit qu'il existe un intervalle non vide autour de $l^{''}$ dans lequel l satisfait également la condition de courbure (12). □

On va maintenant chercher à prouver la convergence de l'algorithme de descente utilisant la recherche linéaire qu'on vient de présenter. A cette fin, nous allons utiliser le théorème suivant

***Théorème* 15 (Zoutendijk).** — *Soit une séquence $(x^k, p^k, l^k)_{k\in\mathbb{N}}$ vérifiant $x^{k+1} = x^k + l^k p^k$ satisfaisant les conditions de Wolfe. On définit*

$$\cos\theta^k = \frac{-\nabla f(x^k)^T p^k}{\|\nabla f(x^k)\| \, \|p^k\|} \geq 0$$

(c.-à-d. le cosinus de l'angle entre la direction de recherche et la plus grande pente). Si f (supposée différentiable) est bornée inférieurement et si ∇f est Lispchitzien alors

$$\sum_{k\in\mathbb{N}} \cos^2\theta^k \left\|\nabla f(x^k)\right\|^2 < +\infty$$

En corollaire $\lim_{k\to\infty} \cos^2\theta^k \left\|\nabla f(x^k)\right\|^2 = 0$.

Démonstration. — D'après la condition de courbure, on a

$$\nabla f(x^{k+1})^T p^k \geq c_2\nabla f(x^k)^T p^k$$

et donc

$$(\nabla f(x^{k+1}) - \nabla f(x^k))^T p^k \geq (c_2 - 1)\nabla f(x^k)^T p^k$$

Or $x \mapsto \nabla f(x)$ est Lipschitz, donc $\exists L > 0$ tel que

$$\left\|\nabla f(x^{k+1}) - \nabla f(x^k)\right\| \leq L \left\|x^{k+1} - x^k\right\|$$

Il vient donc $Ll^k \left\|p^k\right\|^2 \geq (c_2 - 1)\nabla f(x^k)^T p^k$ et donc

$$l^k \geq \frac{(c_2 - 1)\nabla f(x^k)^T p^k}{L \left\|p^k\right\|^2} \geq 0$$

D'après la condition d'Armijo, on a alors

$$f(x^k + l^k p^k) \leq f(x^k) + \frac{c_1(c_2 - 1)}{L} \frac{\left(\nabla f(x^k)^T p^k\right)^2}{\left\|p^k\right\|^2}$$

Une sommation terme à terme donne alors

$$\frac{L}{c_1(c_2-1)}\sum_{k\in\mathbb{N}}\left(f(x^{k+1})-f(x^k)\right)\geq\sum_{k\in\mathbb{N}}\cos^2\theta^k\left\|\nabla f(x^k)\right\|^2$$

La suite de réels $\left(f(x^k)\right)_{k\in\mathbb{N}}$ est décroissante, bornée inférieurement, elle converge. En conclusion

$$\sum_{k\in\mathbb{N}}\cos^2\theta^k\left\|\nabla f(x^k)\right\|^2<\infty$$

□

On utilise le théorème de Zoutendjik en choisissant $p^k=-\nabla f(x^k)$. On a alors, pour tout $k\in\mathbb{N}$, $\cos\theta^k=1$ et le résultat suivant.

Théorème 16

Soit f différentiable, bornée inférieurement et telle que ∇f Lispchitzien. Alors la série

$$\sum_{k\in\mathbb{N}}\left\|\nabla f(x^k)\right\|^2$$

converge pour tout algorithme de gradient satisfaisant les conditions de Wolfe et on a la convergence

$$\lim_{k\to\infty}\left\|\nabla f(x^k)\right\|=0$$

2.1.3.2 Résultats de convergence numérique

Si on applique l'algorithme de gradient avec règles de Wolfe à la fonction $\mathbb{R}^n\ni x\mapsto 1/2x^TQx-b^Tx$ avec Q matrice de $\mathcal{M}_n(\mathbb{R})$ symétrique définie positive et $b\in\mathbb{R}^n$ on peut montrer la convergence

$$(x^{k+1}-x^*)^TQ(x^{k+1}-x^*)\leq\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2(x^k-x^*)^TQ(x^k-x^*)$$

où λ_1 et λ_n sont respectivement la plus petite et la plus grande des valeurs propres de Q. Les performances se dégradent donc avec le conditionnement du problème.

2.1.4 Gradient stochastique.

Dans le cas où l'on ne dispose pas d'une formule analytique du gradient, l'algorithme stochastique suivant a l'avantage par rapport à un gradient déterministe de ne requérir d'approximer numériquement qu'une différence finie scalaire, ce qui est plus aisé.

Algorithme 2 (Gradient stochastique)

A partir de $x^0 \in \mathbb{R}^n$, $\theta \in]0,2[$ et une loi de probabilité discrète $(p_i)_{i=1,\dots,n}$ ($p_i > 0$ et $\sum_{i=1}^n p_i = 1$), itérer

- choisir avec une probabilité p_i l'indice i ($i \in \{1,\dots,n\}$)
- $x_j^{k+1} = \begin{cases} x_j^k - \dfrac{\theta}{L_i}\dfrac{\partial f}{\partial x_i}(x) \text{ si } j = i \\ x_j^k \text{ sinon} \end{cases}$

Théorème 17

Soit $f : \mathbb{R}^n \to \mathbb{R}$ ($n \geq 2$) convexe deux fois différentiable, telle que $\left|\frac{\partial^2 f}{\partial x_i^2}\right| \leq L_i$ (pour tout $x \in \mathbb{R}^n$ et tout $i = 1,\dots,n$) et telle que l'ensemble $\left\{x \in \mathbb{R}^n \mid f(x) \leq f(x^0)\right\}$ est borné. Alors,

$$\mathbb{E}\left(f(x^k) - f(x^*)\right) \leq \frac{2C^2}{\theta(2-\theta)}\frac{1}{k+1}$$

où $C = \sup_{f(x)\leq f(x_0)} \|x - x^*\|_{M^{-1}}$ et $\|x\|_{M^{-1}} = \sqrt{x^T M^{-1} x}$ est la norme euclidienne pondérée associée à $M^{-1} = diag\left(\frac{L_i}{p_i}\right)$.

Démonstration. — Voir l'exercice 2.2 p.63 et son corrigé en p.78. □

2.1.5 Méthode utilisant le Hessien: méthode de Newton

Autour de x^k, une approximation de f supposée deux fois différentiable est

$$q(x) = f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k)$$

Si $\nabla^2 f(x^k) > 0$ alors $\mathbb{R}^n \ni x \mapsto q(x) \in \mathbb{R}$ est fortement convexe. Son minimum est donc atteint en un point x tel que $\nabla q(x) = 0$ c.-à-d.

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$$

qui donne $x = x^k - \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$ La méthode de Newton consiste à itérer le calcul précédent. À chaque étape on doit effectuer $0(n^3)$ calculs dont la majeure partie est due à l'inversion du Hessien de f en x^k.

Algorithme 3 (Newton)

À partir de $x^0 \in \mathbb{R}^n$ quelconque, itérer

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$$

Théorème 18

Soit $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}$ deux fois différentiable, possédant un unique minimiseur global x^* tel que $\nabla^2 f(x^*)$ est définie positive (on note $\lambda > 0$ sa plus petite valeur propre) et tel que $\mathbb{R}^n \ni x \mapsto \nabla^2 f(x) \in \mathcal{M}_n(\mathbb{R})$ est localement Lipschitz au voisinage de x^* (on note C sa constante de Lipschitz). L'algorithme 3 de Newton converge quadratiquement vers x^* si on l'initialise en un point x^0 tel que $\left\|x^0 - x^*\right\| \leq \frac{2\lambda}{3C}$.

Démonstration. — D'après la formule de Taylor avec reste intégral appliquée à la fonction $t \mapsto \nabla f(x^k + t(x^* - x^k))$, on a, en utilisant $\nabla f(x^*) = 0$,

$$-\nabla f(x^k) - \nabla^2 f(x^k)(x^* - x^k) = \int_0^1 \left(\nabla^2 f(x^k + t(x^* - x^k)) - \nabla^2 f(x^k)\right)(x^* - x^k)dt$$

D'autre part, si x^k est suffisamment proche de x^* (c'est vrai pour $k = 0$ et le restera par récurrence par l'équation (16)), la condition de Lipschitz donne qu'il existe $C > 0$ tel que

$$\left\|\nabla^2 f(x^k + t(x^* - x^k)) - \nabla^2 f(x^k)\right\| \leq C \left\|x^* - x^k\right\| t$$

On peut donc déduire

$$\left\|-\nabla f(x^k) - \nabla^2 f(x^k)(x^* - x^k)\right\| \leq \frac{C}{2} \left\|x^* - x^k\right\|^2 \tag{13}$$

L'itération de l'algorithme de Newton donne $x^k = x^{k+1} + \nabla^2 f(x^k)^{-1} \nabla f(x^k)$. Après substitution et simplification dans (13) on a

$$\left\|\nabla^2 f(x^k)(x^* - x^{k+1})\right\| \leq \frac{C}{2} \left\|x^* - x^k\right\|^2$$

Il s'agit maintenant de faire apparaître $\nabla^2 f(x^*)$ qui est la quantité sur laquelle porte l'hypothèse. Une inégalité triangulaire donne

$$\begin{aligned}\left\|\nabla^2 f(x^*)(x^* - x^{k+1})\right\| \leq &\left\|\nabla^2 f(x^k)(x^* - x^{k+1})\right\| \\ &+ \left\|(\nabla^2 f(x^*) - \nabla^2 f(x^k))(x^* - x^{k+1})\right\|\end{aligned}$$

À l'aide des majorations précédentes on a alors

$$\left\|\nabla^2 f(x^*)(x^* - x^{k+1})\right\| \leq \frac{C}{2} \left\|x^* - x^k\right\|^2 + C \left\|x^* - x^k\right\| \left\|x^{k+1} - x^*\right\| \tag{14}$$

Notons maintenant $\lambda > 0$ la plus petite valeur propre de $\nabla^2 f(x^*)$. L'inéquation (14) donne alors

$$\lambda \left\|x^* - x^{k+1}\right\| \leq \frac{C}{2} \left\|x^* - x^k\right\|^2 + C \left\|x^* - x^k\right\| \left\|x^{k+1} - x^*\right\| \tag{15}$$

On a $\left\|x^k - x^*\right\| \leq \frac{2\lambda}{3C}$. Pour tous les indices $p > k$ suivants on encore $\|x^p - x^*\| \leq \frac{2\lambda}{3C}$. En effet on peut tirer de (15) que

$$(\lambda - C \left\|x^k - x^*\right\|) \left\|x^{k+1} - x^*\right\| \leq \frac{C}{2} \left\|x^k - x^*\right\|^2$$

d'où $\left\|x^{k+1} - x^*\right\| \leq \frac{2\lambda}{3C}$. Finalement l'inéquation (15) donne donc la convergence quadratique

$$\left\|x^{k+1} - x^*\right\| \leq \frac{3C}{2\lambda} \left\|x^k - x^*\right\|^2 \tag{16}$$

□

2.1.6 Méthode de quasi-Newton

Cette méthode ne requiert que l'évaluation de la fonction et de son gradient. Elle s'avère en pratique souvent aussi performante que la méthode de Newton dans le calcul de la direction de descente. Au lieu de calculer cette direction par la formule $p^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$, on utilise une formule

$$p^k = -\left(B^k\right)^{-1} \nabla f(x^k)$$

où B^k est une matrice symétrique définie positive qu'on va mettre à jour au cours des itérations.

2.1.6.1 Approximation de la fonction coût

À l'itération k on dispose d'une estimation de l'optimum x^k, et on forme la fonction

$$\mathbb{R}^n \ni p \mapsto m^k(p) = f(x^k) + \nabla f(x^k)^T p + \frac{1}{2} p^T B^k p \in \mathbb{R}$$

Le minimum de cette fonction est atteint en $p^k = -\left(B^k\right)^{-1} \nabla f(x^k)$ et peut permettre de former une nouvelle estimée de l'optimum

$$x^{k+1} = x^k + l^k p^k$$

où l^k peut être choisi par les conditions de Wolfe (par exemple).

On va chercher à introduire les modifications du gradient entre les itérations (courbure) dans la mise à jour de B^k. Le nouveau modèle autour de l'estimation x^{k+1} est

$$\mathbb{R}^n \ni p \mapsto m^{k+1}(p) = f(x^{k+1}) + \nabla f(x^{k+1})^T p + \frac{1}{2} p^T B^{k+1} p \in \mathbb{R}$$

On va faire coïncider le gradient de m^{k+1} avec celui de f qu'on a évalué aux itérations x^k et x^{k+1}. Par construction, $\nabla m^{k+1}(0) = \nabla f(x^{k+1})$. D'autre part, l'autre gradient recherché $\nabla m^{k+1}(-l^k p^k) = \nabla f(x^{k+1}) - l^k B^{k+1} p^k$. On utilise désormais les notations

$$s^k = x^{k+1} - x^k,$$
$$y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

On a donc à résoudre

$$B^{k+1}s^k = y^k$$

Il n'est pas toujours possible de résoudre cette équation dont l'inconnue B^{k+1} est à rechercher symétrique définie positive. Une condition nécessaire et suffisante pour que cette équation possède une solution symétrique définie positive est que

$$(s^k)^T y^k > 0$$

ce qui peut se réécrire $(x^{k+1} - x^k)^T(\nabla f(x^{k+1}) - \nabla f(x^k)) > 0$. On pourra éventuellement chercher à imposer cette condition au moment de la recherche linéaire. Cette condition sera toujours remplie si la fonction f est α-convexe. Les conditions de Wolfe garantissent que cette condition est remplie. En effet on a

$$\nabla f(x^{k+1})^T s^k \geq c_2 \nabla f(x^k)^T s^k$$

d'après la condition de courbure. Ceci implique bien $(s^k)^T y^k \geq (c_2 - 1) l^k \nabla f(x^k)^T p^k > 0$ car on utilise un algorithme de descente.

2.1.6.2 Approximation optimale du Hessien

Notre but était de définir B^{k+1}, on sait qu'on peut toujours le faire par les conditions de Wolfe. On va maintenant rendre ce choix unique. En l'état actuel on a $\frac{n(n+1)}{2}$ coefficients et n équations ainsi que n inégalités (provenant de la contrainte que l'approximation doit être positive définie). On peut montrer que ces inégalités possèdent des solutions. Pour définir de manière unique notre solution, on va chercher à s'approcher le plus possible de B^k au sens d'une certaine norme. Introduisons à cet effet la définition suivante.

***Définition* 14 (Norme de Frobenius (pondérée))**. — Soit $A = (a_{ij})_{i=1\dots n, j=1\dots n}$ une matrice de $\mathcal{M}_n(\mathbb{R})$ et W une matrice définie positive de $\mathcal{M}_n(\mathbb{R})$. On définit $\|A\|_W$ la norme pondérée de Frobenius par l'égalité suivante

$$\|A\|_W = \left\|W^{1/2} A W^{1/2}\right\|_F$$

où $\|A\|_F = \sum_{i=1\dots n, j=1\dots n} a_{ij}^2$

Dans notre cas, nous allons choisir W matrice de pondérations telle que $Wy^k = s^k$, par exemple $W^{-1} = \int_0^1 \nabla^2 f(x^k + \tau l^k p^k) d\tau$ qui est la moyenne du Hessien le long du chemin de la dernière itération.

En utilisant cette norme on résout le problème d'optimisation suivant

$$\begin{aligned} \min_B \quad & \left\|B - B^k\right\|_W \\ \text{tel que } & B = B^T, \\ & Bs^k = y^k \end{aligned} \tag{17}$$

***Proposition* 1**. — *Le problème d'optimisation* (17) *possède une unique solution* B^* *qui est*

$$B^* = (I - \gamma^k y^k (s^k)^T) B^k (I - \gamma^k s^k (y^k)^T) + \gamma^k y^k (y^k)^T, \quad \text{où } \gamma^k = 1/((y^k)^T s^k)$$

La précédente formule sera utilisée comme mise à jour du Hessien dans l'algorithme DFP (Davidson, Fletcher, Powell).

2.1.6.3 Calcul itératif de l'inverse de l'approximation du Hessien

La quantité que nous utilisons dans les algorithmes de quasi-Newton est l'inverse du Hessien $(B^k)^{-1}$ et non le Hessien lui-même. Pour calculer cet inverse de manière efficace on peut utiliser la formule issue de la proposition suivante

***Proposition* 2 (Formule de Sherman-Morrison-Woodbury)**
Soient A une matrice de $\mathcal{M}_n(\mathbb{R})$ inversible, U et V deux matrices de $\mathcal{M}_{n,p}(\mathbb{R})$, $1 \leq p \leq n$. On définit $\bar{A} = A + UV^T$. Si $\bar{A}$ est inversible alors

$$\bar{A}^{(-1)} = A^{(-1)} - A^{(-1)}U\left(I_{(p)} + V^T A^{(-1)}U\right)^{(-1)} V^T A^{(-1)} \tag{18}$$

Lorsqu'on met à jour l'approximation du Hessien par la formule de la proposition 1 on a

$$B^{k+1} = (I - \gamma^k y^k (s^k)^T) B^k (I - \gamma^k s^k (y^k)^T) + \gamma^k y^k (y^k)^T \tag{19}$$

Il lui correspond la mise à jour de l'inverse

(20)
$$\left(B^{k+1}\right)^{-1} = \left(B^k\right)^{-1} - \frac{1}{(y^k)^T \left(B^k\right)^{-1} y^k}\left(\left(B^k\right)^{-1} y^k (y^k)^T \left(B^k\right)^{-1}\right) + \frac{1}{\left(y^k\right)^T s^k} s^k (s^k)^T$$

L'équation (20) redonne bien (19). Cette dernière équation fait apparaître une forme factorisée et un terme de mise à jour UV^T en notant $U = \left(-\frac{1}{a}U_1, \frac{1}{b}U_2\right)$ et $V = (U_1, U_2)$ avec

$$\begin{aligned} U_1 &= (B^k)^{-1} y^k \\ U_2 &= s^k \\ a &= (y^k)^T (B^k)^{-1} y^k \\ b &= (y^k)^T s^k \end{aligned}$$

En utilisant la formule de Sherman-Morrison-Woodbury (18), il vient alors

$$B^{k+1} = B^k - B^k U\left(I_{(2)} + V^T B^k U\right) V^T B^k$$

Après quelques lignes de développement, on obtient

$$\begin{aligned} B^{k+1} =& B^k + (\gamma^k)^2 y^k (s^k)^T B^k s^k (y^k)^T \\ & - \gamma^k y^k (s^k)^T B^k - \gamma^k B^k s^k (y^k)^T + y^k (y^k)^T \gamma^k \\ =& \left(I - \gamma^k y^k (s^k)^T\right) B^k \left(I - \gamma^k s^k (y^k)^T\right) + \gamma^k y^k (y^k)^T \end{aligned}$$

qui est bien (19).

2.1.6.4 Mise en œuvre

On retiendra deux formules possibles pour la mise à jour de l'inverse de l'approximation du Hessien: la formule DFP (20) et la formule BFGS (Broyden, Fletcher, Goldfarb, Shanno) suivante

$$(B^{k+1})^{-1} = \left(I - \gamma^k s^k (y^k)^T\right)(B^k)^{-1}\left(I - \gamma^k y^k (s^k)^T\right) + \gamma^k s^k (s^k)^T \tag{21}$$

où $\gamma^k = 1/((y^k)^T s^k)$ et qui correspond à la résolution du problème

$$\begin{aligned} \min_{B^{-1}} \quad & \left\|B^{-1} - (B^k)^{-1}\right\|_W \\ \text{tel que } & B^{-1} = B^{-T}, \\ & B^{-1} y^k = s^k \end{aligned}$$

Algorithme 4 (Algorithme de BFGS)

À partir de $x^0 \in \mathbb{R}^n$ quelconque, de $H^0 = I_n$ (d'autres choix de matrice définie positive sont possibles) et de $B^0 = (H^0)^{-1}$ itérer

$$\begin{aligned} p^k &= -H^k \nabla f(x^k) \\ x^{k+1} &= x^k + l^k p^k,\ l^k \text{ satisfaisant les conditions de Wolfe} \\ s^k &= x^{k+1} - x^k \\ y^k &= \nabla f(x^{k+1}) - \nabla f(x^k) \\ (B^{k+1})^{-1} &= \left(I - \gamma^k s^k (y^k)^T\right)(B^k)^{-1}\left(I - \gamma^k y^k (s^k)^T\right) + \gamma^k s^k (s^k)^T \\ H^{k+1} &= \left(B^{k+1}\right)^{-1} \end{aligned}$$

Théorème 19

Soit $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}$ deux fois continûment différentiable telle que la ligne de niveau $\mathcal{L} = \left\{x / f(x) \leq f(x^0)\right\}$ est convexe d'intérieur non vide et qu'il existe des constantes strictement positives m et M telles que $\forall z \in \mathbb{R}^n$ et $\forall x \in \mathcal{L}$

$$m\|z\|^2 \leq z^T \nabla^2 f(x) z \leq M\|z\|^2$$

L'algorithme 4 de BFGS converge vers l'unique minimiseur x^* de f.

***Théorème* 20**. — *Avec les hypothèses du théorème précédent, si de plus $\nabla^2 f(x^*)$ est Lipschitz au voisinage de $x^\star$, alors la convergence de l'algorithme 4 est superlinéaire.*

2.1.7 Méthode du gradient conjugué

L'algorithme que nous présentons dans cette section possède deux intérêts. Il permet de résoudre des systèmes linéaires strictement positifs de grande taille et il

sert d'algorithme de base pour la résolution itérée de problèmes d'optimisation non linéaire.

2.1.7.1 Fonctions quadratiques

Résoudre le système linéaire $Ax = b$ où $A \in \mathcal{M}_n(\mathbb{R})$ symétrique positive définie, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^n$ est équivalent à résoudre le problème de minimisation de $\phi(x) = \frac{1}{2}x^T Ax - b^T x$. Ces deux problèmes ont effectivement une même solution unique. En tout point x qui n'est pas cette solution, on note le "reste" $r(x) = Ax - b = \nabla\phi$ qui est non nul.

Considérons maintenant la définition suivante

***Définition* 15 (Directions conjuguées).** — Un ensemble $\{p^0, p^1, ..., p^l\}$ de vecteurs de $\mathbb{R}^n$ est dit A-conjugué si $\forall i \neq j$, $(p^i)^T A p^j = 0$.

Une telle famille est libre. Un exemple d'une telle famille est donnée par une famille de vecteurs propres orthogonaux de A.

Minimisation suivant les directions conjuguées Considérons une séquence

$$x^{k+1} = x^k + l^k p^k \tag{22}$$

où les suites $(x^k)_{k\in\mathbb{N}}$ et $(p^k)_{k\in\mathbb{N}}$ sont des suites de vecteurs de $\mathbb{R}^n$ et la suite $(l^k)_{k\in\mathbb{N}}$ est une suite de vecteurs de $\mathbb{R}$. On a alors

$$\phi(x^{k+1}) = \phi(x^k + l^k p^k) = \phi(x^k) + l^k (x^k)^T A p^k - b^T p^k l^k + \frac{1}{2}\left(l^k\right)^2 (p^k)^T A p^k$$

Cette expression définit une fonction convexe de la variable l^k. Son unique minimum est obtenu pour

$$l^k = \frac{b^T p^k - (x^k)^T A p^k}{(p^k)^T A p^k} = -\frac{(r^k)^T p^k}{(p^k)^T A p^k} \tag{23}$$

où $r^k = r(x^k)$.

Théorème 21

Quel que soit $x^0 \in \mathbb{R}^n$, la séquence générée par (22), où l^k est donné par (23) et où $\{p^0, p^1, ..., p^{n-1}\}$ est une famille de n directions A-conjuguées, avec $A \in \mathcal{M}_n(\mathbb{R})$ symétrique définie positive, atteint la solution x^* de $Ax = b$ en au plus n étapes.

Démonstration. — Les directions $p^0, p^1, ..., p^{n-1}$ sont linéairement indépendantes donc elles engendrent $\mathbb{R}^n$. Il existe donc n réels $\sigma^0, ..., \sigma^{n-1}$ tels que

$$x^* - x^0 = \sigma^0 p^0 + ... + \sigma^{n-1} p^{n-1}$$

En multipliant à gauche cette égalité par $(p^k)^T A$ il vient, pour tout $k = 0, ..., n-1$,

$$\sigma^k = \frac{(p^k)^T A(x^* - x^0)}{(p^k)^T A p^k}$$

Nous allons maintenant montrer que $\sigma^k = l^k$. En k étapes on calcule

$$x^k = x^0 + l^0 p^0 + ... + l^{k-1} p^{k-1}$$

tel que $(p^k)^T A(x^k - x^0) = 0$. Il vient donc

$$\begin{aligned}(p^k)^T A(x^* - x^0) &= (p^k)^T A(x^* - x^0 - x^k + x^0) \\ &= (p^k)^T A(x^* - x^k) \\ &= (p^k)^T (b - Ax^k) = -(p^k)^T r^k\end{aligned}$$

d'où $\sigma^k = -\frac{(p^k)^T r^k}{(p^k)^T A p^k} = l^k$ d'où $x^n = x^*$. Ceci prouve qu'on a trouvé la décomposition de x^* dans la base $p^0, p^1, ..., p^{n-1}$ en minimisant successivement suivant les directions conjuguées la fonction quadratique $\phi(x) = \frac{1}{2} x^T Ax - b^T x$. □

***Théorème* 22**. — *Quel que soit $x^0 \in \mathbb{R}^n$, la séquence générée par (22), où l^k est donné par (23) et où $\{p^0, p^1, ..., p^{n-1}\}$ est une famille de n directions A-conjuguées avec $A \in \mathcal{M}_n(\mathbb{R})$ symétrique définie positive, possède les propriétés suivantes*

1. *$(r^k)^T p^i = 0$ pour tout $k = 0, ..., n-1$, et tout $i = 0, ..., k-1$*
2. *x^k est le minimiseur global de $\mathbb{R}^n \ni x \mapsto \phi(x)$ sur l'espace affine $\{x = x^0 + \sum_{j=0}^{k-1} \mu^j p^j, (\mu^0, ..., \mu^{k-1}) \in \mathbb{R}^k\}$.*

Démonstration. — De manière générale, $\tilde{x}$ la combinaison linéaire

$$\tilde{x} = x^0 + \tilde{\alpha}^0 p^0 + ... + \tilde{\alpha}^{k-1} p^{k-1}$$

procurant la plus petite valeur de $\mathbb{R}^n \ni x \mapsto \phi(x)$ est telle que

$$\nabla\phi(\tilde{x})^T p^i = 0$$

pour tout $i = 0, ..., k-1$, c.-à-d. $r(\tilde{x})^T p^i = 0$.

Prouvons maintenant le premier point du théorème. Le premier point x^1 calculé par la séquence $k = 1$ est obtenu en calculant

$$l^0 = \text{argmin}_l \phi(x^0 + l p^0)$$

On a donc $\nabla\phi(x^1)^T p^0 = 0$. Raisonnons maintenant par récurrence. Supposons qu'on ait établi pour un certain $k > 1$

$$(r^{k-1})^T p^i = 0, \forall i = 0, ..., k-2$$

Calculons maintenant $(r^k)^T p^i$ pour tout $i = 0, ..., k-1$.

$$r^k = Ax^k - b = A(x^{k-1} + l^{k-1} p^{k-1}) - b = r^{k-1} + l^{k-1} Ap^{k-1}$$

d'où

$$(r^k)^T p^{k-1} = (r^{k-1})^T p^{k-1} + l^{k-1} (p^{k-1})^T Ap^{k-1}$$

or $l^{k-1} = -\frac{(r^{k-1})^T p^{k-1}}{(p^{k-1})^T Ap^{k-1}}$ d'où $(r^k)^T p^{k-1} = 0$. D'autre part, pour $i = 0, ..., k-2$,

$$(r^k)^T p^i = (r^{k-1})^T p^i + l^{k-1} (p^{k-1})^T Ap^i = 0$$

Ce qui achève de prouver le premier point du théorème.

Pour prouver le second point du théorème il suffit d'expliciter le gradient de l'application $\mathbb{R}^k \ni l \mapsto \phi(x^0 + l^0 p^0 + ... + l^{k-1} p^{k-1})$. Sa i-ème coordonnée vaut

$\nabla\phi(x^k)^T p^i = (r^k)^T p^i = 0$ quel que soit $i = 0, ...k-1$. Donc le deuxième point est prouvé. □

Calcul des directions conjuguées On peut construire une base de vecteurs propres orthogonaux car A est symétrique. Ces vecteurs sont par construction A-conjugués et constituent un choix possible de directions conjuguées. Mais le calcul de ces vecteurs est en général très coûteux en temps de calculs. En fait on peut calculer les directions conjuguées au fur et à mesure qu'on en a besoin par la récurrence suivante

$$p^k = -r^k + \beta^k p^{k-1}$$

qui s'interprète comme l'opposé du gradient au nouveau point altéré par la direction précédente. Pour que p^k et p^{k-1} soient des directions A-conjuguées on choisit

$$\beta^k = \frac{(r^k)^T A p^{k-1}}{(p^{k-1})^T A p^{k-1}} \tag{24}$$

La direction p^k est également A-conjuguée avec les directions p^i pour $i = 0, .., k-2$. Pour démontrer ce résultat on utilise la proposition suivante

***Proposition* 3**. — *Avec les notations précédentes, si $x^k \neq x^*$ alors les propriétés suivantes sont vérifiées*

1. $(r^k)^T r^i = 0$, *pour tout* $i = 0, ..., k-1$
2. $Vect(r^0, r^1, ..., r^k) = Vect(r^0, Ar^0, ..., A^k r^0)$ *(ce dernier sous espace est dénommé sous-espace de Krylov)*
3. $Vect(p^0, p^1, ..., p^k) = Vect(r^0, Ar^0, ..., A^k r^0)$
4. $(p^k)^T A p^i = 0$ *pour tout* $i = 0, ..., k-1$

On peut calculer

$$(p^i)^T A p^k = -(p^i)^T A r^k + \beta^k (p^i)^T A p^{k-1} = -(p^i)^T A r^k$$

On sait d'après le théorème 22 que $(r^k)^T p^i = 0$ pour tout $i = 0, ..., k-2$. Or d'après les points 2 et 3 de la proposition 3 on sait que

$$Ap^i \in A.\text{Vect}(r^0, Ar^0, ..., A^i r^0) = \text{Vect}(Ar^0, A^2 r^0, ..., A^{i+1} r^0) \subset \text{Vect}(p^0, p^1, ..., p^{i+1})$$

Donc il existe un vecteur $(\gamma^0, ..., \gamma^{i+1}) \in \mathbb{R}^{i+2}$ tel que

$$(p^i)^T A r^k = \sum_{j=0}^{i+1} \gamma^j (p^j)^T r^k = 0$$

quel que soit $i = 0, ..., k-2$. En conclusion, les directions $\{p^0, ..., p^k\}$ sont A-conjuguées.

On peut en outre simplifier les expressions (23) et (24) en tirant partie des propriétés de la famille $(r^i)_{i=0,...,k-1}$ avec r^k. On obtient alors

$$l^k = \frac{(r^k)^T r^k}{(p^k)^T A p^k}, \ \beta^{k+1} = \frac{(r^{k+1})^T r^{k+1}}{(r^k)^T r^k} \tag{25}$$

Une fois ces simplifications effectuées on aboutit à l'algorithme suivant.

Algorithme 5 (Algorithme du gradient conjugué)

À partir de $x^0 \in \mathbb{R}^n$ quelconque calculer $r^0 = Ax^0 - b$ et $p^0 = -r^0$. Itérer

$$\begin{aligned} l^k &= \frac{(r^k)^T r^k}{(p^k)^T A p^k} \\ x^{k+1} &= x^k + l^k p^k \\ r^{k+1} &= r^k + l^k A p^k \\ \beta^{k+1} &= \frac{\left\|r^{k+1}\right\|^2}{\left\|r^k\right\|^2} \\ p^{k+1} &= -r^{k+1} + \beta^{k+1} p^k \end{aligned}$$

On peut caractériser la vitesse de convergence de l'algorithme 5.

***Proposition* 4.** — *Soit $A \in \mathcal{M}_n(\mathbb{R})$ symétrique définie positive, $K(A) = \frac{\lambda_n}{\lambda_1}$ le rapport entre la plus grande et la plus petite des valeurs propres de A (aussi appelé nombre de conditionnement). Les itérations de l'algorithme 5 du gradient conjugué appliqué à $\mathbb{R}^n \ni x \mapsto \phi(x) = \frac{1}{2}x^T Ax - b^T x \in \mathbb{R}$ avec $b \in \mathbb{R}^n$, vérifient l'inégalité*

$$\left\|x^k - x^*\right\|_A \leq 2\left(\frac{\sqrt{K(A)} - 1}{\sqrt{K(A)} + 1}\right)^k \left\|x^0 - x^*\right\|_A$$

Il est possible d'améliorer cette vitesse de convergence par une technique de précon–ditionnement utilisant un changement de variable linéaire, on se reportera à [**22**] pour un exposé de ces techniques.

2.1.7.2 Application aux fonctions non linéaires

Lorsque la fonction à minimiser n'est pas quadratique, on ne peut pas calculer explicitement le pas optimal le long d'une direction de descente et la notion de "résidu" n'a pas le même sens. On peut néanmoins transposer les idées de l'algorithme du gradient conjugué de la manière suivante

Algorithme 6 (Algorithme de Fletcher-Reeves)

À partir de $x^0 \in \mathbb{R}^n$ quelconque calculer $f^0 = f(x^0)$, $r^0 = \nabla f(x^0)$ et $p^0 = -r^0$. Itérer

1. Choisir l^k par une recherche linéaire dans la direction p^k
2. $$\begin{aligned} x^{k+1} &= x^k + l^k p^k \\ r^{k+1} &= \nabla f(x^{k+1}) \\ \beta^{k+1} &= \frac{\left\|r^{k+1}\right\|^2}{\left\|r^k\right\|^2} \\ p^{k+1} &= -r^{k+1} + \beta^{k+1} p^k \end{aligned}$$

Tout comme l'algorithme de gradient conjugué, l'implémentation de cet algorithme ne requiert que peu de calculs et peu de mémoire quelle que soit la taille du système. Pour assurer la décroissance entre deux itérations, on utilisera dans la recherche linéaire les conditions de Wolfe (voir Définition 13) avec typiquement des paramètres $0 < c_1 < c_2 < 1/2$.

Une amélioration pratique de cet algorithme est

***Algorithme* 7 (Algorithme de Polak-Ribière).** — À partir de $x^0 \in \mathbb{R}^n$ quelconque calculer $f^0 = f(x^0)$, $r^0 = \nabla f(x^0)$ et $p^0 = -r^0$. Itérer choisir l^k par une recherche linéaire dans la direction p^k

$$\begin{aligned} x^{k+1} &= x^k + l^k p^k \\ r^{k+1} &= \nabla f(x^{k+1}) \\ \beta^{k+1} &= \frac{(r^{k+1})^T (r^{k+1} - r^k)}{\left\|r^k\right\|^2} \\ p^{k+1} &= -r^{k+1} + \beta^{k+1} p^k \end{aligned}$$

Les conditions de Wolfe peuvent également être utilisées mais n'assurent pas la décroissance.

***Proposition* 5.** — *Soit $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}$ deux fois différentiable possédant un unique minimiseur x^* tel que $\nabla^2 f(x)$ est défini positif. Les itérations des algorithmes 6 et 7 avec recherche linéaire exacte (c.-à-d. que l^k minimise $f(x^k + l^k p^k)$ à chaque itération) satisfont*

$$\lim_{k \to +\infty} \frac{\left\|x^{k+n} - x^*\right\|}{\left\|x^k - x^*\right\|} = 0$$

2.2 Principes de l'optimisation sous contraintes

2.2.1 Contraintes égalités

On s'intéresse désormais au problème de la minimisation dans $\mathbb{R}^{n+m}$, $n+m<\infty$ d'une fonction différentiable (fonction coût) $\mathbb{R}^{n+m} \ni (x,u) \mapsto f(x,u) \in \mathbb{R}$ sous la contrainte $c(x,u)=0$ définie par une fonction différentiable $c : \mathbb{R}^{n+m} \ni (x,u) \mapsto c(x,u) \in \mathbb{R}^n$ localement inversible par rapport à la variable x. En d'autres termes x est localement déterminé par u par la relation $c(x,u)=0$:

$$\begin{aligned} &\min_{x,\,u} \quad f(x,u) \\ &\text{tel que } c(x,u)=0 \end{aligned} \tag{26}$$

Une condition suffisante par le théorème des fonctions implicites est que le Jacobien partiel $\frac{\partial c}{\partial x}$ est inversible au voisinage des points tels que $c(x,u)=0$. On supposera cette condition réalisée.

Ce problème revient à chercher un minimum de f sur l'ensemble $\{(x,u)=c^{-1}(0)\}$ qui est un fermé de $\mathbb{R}^{n+m}$. Cet ensemble ne contient pas de voisinage de tous ses points. Ainsi, tout déplacement infinitésimal autour d'un des points de son bord ne fournit pas nécessairement de point admissible. Les conditions d'optimalité en seront donc modifiées.

2.2.1.1 Élimination des variables

Dans cette approche on calcule la variation de la fonction coût en préservant la contrainte. Une variation $(\delta x, \delta u) \in \mathbb{R}^{n+m}$ entraîne une variation

$$\delta f \approx \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial u}\delta u$$

Maintenir la contrainte $c(x+\delta x, u+\delta u)=0$ impose

$$0 \approx \frac{\partial c}{\partial x}\delta x + \frac{\partial c}{\partial u}\delta u$$

Par hypothèse, $\frac{\partial c}{\partial x}$ est inversible. On peut donc établir

$$\delta f \approx \left(-\frac{\partial f}{\partial x}\left(\frac{\partial c}{\partial x}\right)^{-1}\frac{\partial c}{\partial u} + \frac{\partial f}{\partial u}\right)\delta u$$

et à la limite

$$\left(\frac{\partial f}{\partial u}\right)_{c=0} = -\frac{\partial f}{\partial x}\left(\frac{\partial c}{\partial x}\right)^{-1}\frac{\partial c}{\partial u} + \frac{\partial f}{\partial u} \tag{27}$$

On remarquera que la variation de f en maintenant la contrainte est différente de la variation de f sans maintenir la contrainte (dernier terme de la précédente équation).

On peut étendre la définition 6 à notre cas.

***Définition* 16**. — On dit que (x^*,u^*) est un point stationnaire de f sous la contrainte égalité c si $c(x^*,u^*)=0$ et $\left(\frac{\partial f}{\partial u}\right)_{c=0}(x^*,u^*)=0$.

***Définition* 17**. — On dit que $(x^*, u^*) \in \mathbb{R}^n \times \mathbb{R}^m$ est un minimiseur (optimum) local de f sous la contrainte $c(x,u) = 0$ s'il existe $\varepsilon > 0$ tel que $f(x^*, u^*) \leq f(x,u)$, pour tout couple (x,u) tel que $\|(x,u) - (x^*,u^*)\| \leq \varepsilon$ et $c(x,u) = 0$ (on dit que c'est un minimiseur local strict si la précédente inégalité est stricte). On dit alors que (x^*, u^*) est solution locale du problème

$$\begin{aligned} &\min_{x,\,u} \quad f(x,u) \\ &\text{tel que } c(x,u) = 0 \end{aligned}$$

2.2.1.2 Cas des contraintes égalités linéaires, élimination des variables.

Si les contraintes égalités sont linéaires, c.-à-d. de la forme $c(z) = Az - b$ avec $A \in \mathcal{M}_{n,n+m}$ supposée de rang plein et $b \in \mathbb{R}^n$, il est alors possible de déterminer une décomposition de $z \in \mathbb{R}^{n+m}$ sous la forme $(x,u) \in \mathbb{R}^n \times \mathbb{R}^m$ précédemment évoquée, d'éliminer la contrainte et de résoudre le problème à l'aide de l'algorithme suivant.

***Algorithme* 8 (Algorithme d'élimination des variables pour contraintes linéaires)**

1. Déterminer une matrice de permutation P telle que $AP = [B \quad N]$ avec $B \in \mathcal{M}_m(\mathbb{R})$ inversible.
2. En posant $(x,u) = P^T z$, définir $x = B^{-1}(b - Nu)$.
3. Résoudre le problème $\min_{u \in \mathbb{R}^m} f\left(P \begin{bmatrix} B^{-1}(b-Nu) \\ u \end{bmatrix}\right)$

Le dernier problème n'ayant pas de contrainte, il peut être traité par les méthodes précédemments étudiées.

Cet algorithme exploite le fait que, puisque P est une matrice de permutation, $PP^T = I$ et que $b = Az = (AP)(P^T z) = Bx + Nu$. Un telle matrice P peut être calculée à l'aide de la méthode d'élimination de Gauss-Jordan. A noter que la base associée n'est pas unique et que des variantes de cet algorithme se sont donc attachées à déterminer celle qui permettrait d'obtenir la matrice B la plus judicieuse numériquement (voir [**10**]). Malgré cela, ces techniques souffrent parfois d'un mauvais conditionnement qui engendre des instabilité numériques. On leur préfère donc parfois des variantes basées sur une décomposition QR (voir Annexe B).

Ces méthodes sont également souvent utilisées dans le cas où, en plus des contraintes égalités, on impose des contraintes inégalités. Elles sont aussi très largement utilisées dans le cas où les contraintes sont non-linéaires, en résolvant successivement plusieurs problèmes où les contraintes sont approchées par leur développement de Taylor à l'ordre 1 (voir Section 2.2.2.4).

2.2.1.3 Équations adjointes

Définissons le Lagrangien du problème (26) en adjoignant les contraintes à la fonction coût:

$$\mathbb{R}^{2n+m} \ni (x,u,\lambda) \mapsto \mathcal{L}(x,u,\lambda) = f(x,u) + \lambda^T c(x,u) \in \mathbb{R} \tag{28}$$

D'après la définition 6, un point stationnaire de $\mathcal{L}$ (sans contraintes) est caractérisé par

$$\frac{\partial \mathcal{L}}{\partial x} = 0,\ \frac{\partial \mathcal{L}}{\partial u} = 0,\ \frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

Théorème 23

$(x^*, u^*, \lambda^*) \in \mathbb{R}^{2n+m}$ point stationnaire de $\mathcal{L}$ ssi (x^*, u^*) est un point stationnaire de f sous la contrainte c.

Démonstration. — Supposons $(x^*, u^*, \lambda^*) \in \mathbb{R}^{2n+m}$ point stationnaire de $\mathcal{L}$. D'une part on a $c(x^*, u^*) = 0$. On va montrer que pour toute variation $(\delta x, \delta u) \in \mathbb{R}^{n+m}$ telle que $(c(x^* + \delta x, u^* + \delta u) = 0$ au premier ordre on a $f(x^* + \delta x, u^* + \delta u) = f(x^*, u^*)$ au premier ordre. Le calcul suivant permet de conclure

$$\begin{aligned} f(x^* + \delta x, u^* + \delta u) - f(x^*, u^*) &\approx \frac{\partial f}{\partial x}(x^*, u^*)\delta x)\delta x + \frac{\partial f}{\partial u}(x^*, u^*)\delta u \\ &\approx -\lambda^* \left(\frac{\partial c}{\partial x}(x^*, u^*)\delta x + \frac{\partial c}{\partial u}(x^*, u^*)\delta u \right) \approx 0 \end{aligned}$$

Ceci prouve la première implication. Supposons maintenant que (x^*, u^*) est un point stationnaire de f sous la contrainte c. Alors,

$$\frac{\partial f}{\partial u}(x^\star, u^\star) = \frac{\partial f}{\partial x}(x^\star, u^\star) \left(\frac{\partial c}{\partial x}(x^\star, u^\star) \right)^{-1} \frac{\partial c}{\partial u}(x^\star, u^\star) = -(\lambda^\star)^T \frac{\partial c}{\partial u}(x^\star, u^\star) \tag{29}$$

en définissant $(\lambda^\star)^T = -\frac{\partial f}{\partial x}(x^\star, u^\star) \left(\frac{\partial c}{\partial x}(x^\star, u^\star) \right)^{-1}$. On a ainsi

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= \frac{\partial f}{\partial x} + (\lambda^\star)^T \frac{\partial c}{\partial x} = \frac{\partial f}{\partial x} - \frac{\partial f}{\partial x} = 0 \\ \frac{\partial \mathcal{L}}{\partial u} &= \frac{\partial f}{\partial u} + (\lambda^\star)^T \frac{\partial c}{\partial u} = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= c(x^\star, u^\star) = 0 \end{aligned}$$

par définition de $\lambda^\star$ et parce que $(x^\star, u^\star)$ vérifie la contrainte c. □

2.2.1.4 Multiplicateurs de Lagrange

Le vecteur λ dans l'équation (28) est appelé vecteur des multiplicateurs de Lagrange. Les multiplicateurs de Lagrange apportent beaucoup d'information sur le problème d'optimisation sous contraintes comme en attestent les propriétés suivantes.

Théorème 24

Si $(x^*, u^*, \lambda^*) \in \mathbb{R}^{2n+m}$ est un point stationnaire de $\mathcal{L}$, alors

$$(\nabla f)^T (x^*, u^*) = -(\lambda^*)^T (\nabla c)^T (x^*, u^*)$$

Cette proposition relie le gradient du coût et celui de la contrainte à un point stationnaire.

***Proposition* 6**. — *Si* $(x^*, u^*, \lambda^*) \in \mathbb{R}^{2n+m}$ *est un point stationnaire de* $\mathcal{L}$*, alors en ce point*

(30)
$$\left(\frac{\partial^2 f}{\partial u^2}\right)_{\mid c=0}(x^*, u^*) = \begin{pmatrix} -\frac{\partial c}{\partial u}^T \left(\frac{\partial c}{\partial x}^T\right)^{-1} & I \end{pmatrix} \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial x^2} & \frac{\partial^2 \mathcal{L}}{\partial x \partial u} \\ \frac{\partial^2 \mathcal{L}}{\partial u \partial x} & \frac{\partial^2 \mathcal{L}}{\partial u^2} \end{pmatrix} \begin{pmatrix} -\left(\frac{\partial c}{\partial x}\right)^{-1} \frac{\partial c}{\partial u} \\ I \end{pmatrix}$$

Cette dernière proposition permet souvent de lever l'ambiguïté sur la nature du point stationnaire. Les multiplicateurs de Lagrange λ^* interviennent dans la matrice centrale du calcul du Hessien (30).

Définissons maintenant un problème proche du problème (26) en utilisant $\delta c \in \mathbb{R}^m$.

(31)
$$\begin{aligned} &\min_{x,\,u} \quad f(x,u) \\ &\text{tel que } c(x,u) + \delta c = 0 \end{aligned}$$

On suppose que (26) et (31) possèdent des solutions (x^*, u^*) et $(\bar{x}, \bar{u})$ respectivement. On note $\frac{\partial f^*}{\partial c}$ le vecteur composé des limites des valeurs $(f(\bar{x}, \bar{u}) - f(x^*, u^*))/\delta c_i$ où $\delta c_i \in \mathbb{R}$ correspond à la i-ème composante de δc.

Théorème 25 (coût marginal)

Si $(x^*, u^*, \lambda^*) \in \mathbb{R}^{2n+m}$ est un point stationnaire de $\mathcal{L}$, alors
$$\frac{\partial f^*}{\partial c} = (\lambda^*)^T$$

Cette dernière propriété permet de quantifier le coût (marginal) d'une variation de la contrainte (relâchement ou durcissement) sur la valeur de l'optimum.

2.2.2 Contraintes inégalités

On s'intéresse maintenant au problème de la minimisation dans $\mathbb{R}^n$, $n < \infty$ d'une fonction différentiable (fonction coût) $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}$ sous la contrainte $c(x) \leq 0$ définie par une fonction différentiable $\mathbb{R}^n \ni x \mapsto c(x) \in \mathbb{R}^m$, $m < \infty$ (on ne spécifie plus de variable par rapport à laquelle cette fonction sera inversible, et on ne fait pas d'hypothèse sur les entiers n et m).

(32)
$$\begin{aligned} &\min_{x} \quad f(x) \\ &\text{tel que } c(x) \leq 0 \end{aligned}$$

***Définition* 18**. — On dit que $x^* \in \mathbb{R}^n$ est un minimiseur (optimum) local de f sous la contrainte $c(x) \leq 0$ si $\exists \epsilon > 0$ tel que $f(x^*) \leq f(x)$, $\forall x$ tel que $\|x - x^*\| \leq \epsilon$ et

$c(x) \leq 0$ (on dit que c'est un minimiseur local strict si $\exists \epsilon > 0$ tel que $f(x^*) < f(x)$, $\forall x$ tel que $\|x - x^*\| \leq \epsilon$ et $c(x) \leq 0$). On dit que (x^*) est solution locale du problème

$$\min_x \quad f(x)$$
$$\text{tel que } c(x) \leq 0$$

2.2.2.1 Conditions de Karush-Kuhn-Tucker

Cas des contraintes affines Considérons dans un premier temps le cas où le vecteur $c(x)$ est constitué de contraintes affines $\mathbb{R}^n \ni x \mapsto c_i(x) \in \mathbb{R}$, pour $i = 1, .., m$ telles que $c(x) \leq 0$ définit un ensemble d'intérieur non vide.

***Définition* 19**. — On appelle cône convexe engendré par $\{v_i, i = 1, ..., k\}$ un ensemble de vecteurs de $\mathbb{R}^n$, l'ensemble des combinaisons linéaires à coefficients positifs ou nuls de vecteurs de $\{v_i, i = 1, ..., k\}$.

Les cônes convexes possèdent d'intéressantes propriétés. On peut établir le résultat suivant

***Lemme* 5 (Farkas)**. — *Soient $\{a_i, i = 1, ..., k\}$ une collection de vecteurs de $\mathbb{R}^n$ et g un vecteur de $\mathbb{R}^n$, les deux propositions suivantes sont équivalentes*

1. *Il n'existe pas $p \in \mathbb{R}^n$ tel que $\left\{g^T p < 0 \text{ et } a_i^T p \geq 0 \text{ pour tout } i = 1, ..., k\right\}$*
2. *g appartient au cône convexe $\mathcal{C}$ engendré par $\{a_i, i = 1, ..., k\}$*

Démonstration. — Démontrons que la seconde proposition entraîne la première. Supposons cette seconde proposition vraie. Supposons qu'il existe $\lambda_i \geq 0$, $i = 1, ..., p$ tels que $g = \sum_{i=1}^p \lambda_i a_i$. S'il existe $h \in \mathbb{R}^n$ tel que $a_i^T h \geq 0$ pour tout $i = 1, .., p$, alors on a $g^T h \geq 0$ ce qui est une contradiction.

Démontrons maintenant que la première proposition entraîne la seconde. Supposons cette première proposition vraie. Supposons qu'il n'existe pas $p \in \mathbb{R}^n$ tel que

$$g^T p < 0 \text{ et } a_i^T p \geq 0 \ \forall i = 1, ..., p$$

Montrons que nécessairement $g \in \mathcal{C}$. Supposons que ce soit faux. Définissons h la projection de g sur $\mathcal{C}$ comme un vecteur $h \in \mathcal{C}$ qui minimise la distance au carré $f(h) = \|h - g\|^2$. Un tel h existe car il est solution d'un problème de minimisation d'une fonction continue α-convexe sur $\mathcal{C}$ convexe fermé. Il n'est pas à l'intérieur de $\mathcal{C}$ car on aurait alors $\nabla f(h) = 2(h - g) = 0$. Il est sur le bord du cône et on a nécessairement $h^T(h - g) = 0$. Posons $p = h - g \in \mathbb{R}^n$. Un calcul donne

$$g^T p = g^T(h - g) = (g - h)^T(h - g) = -\|h - g\|^2$$

Or par hypothèse $g \notin \mathcal{C}$. Donc $h \neq g$ et donc

$$g^T p < 0$$

D'autre part, $h \in \mathcal{C}$ réalise le minimum global de $f(h) = \|h - g\|^2$. Donc, quel que soit $v \in \mathcal{C}$, on a pour tout $\alpha \in [0, 1]$

$$\begin{aligned}\|h - g\|^2 &\leq \|(1 - \alpha)h + \alpha v - g\|^2 \\ &\leq \|h - g\|^2 + \alpha^2 \|v - h\|^2 + 2\alpha(h - g)^T(v - h)\end{aligned}$$

On en déduit que pour tout $\alpha \in [0,1]$

$$\alpha^2 \|v-h\|^2 + 2\alpha(h-g)^T(v-h) \geq 0$$

Nécessairement on a donc $(h-g)^T(v-h) \geq 0$. En conclusion, pour tout $v \in \mathcal{C}$, on a $p^T v \geq 0$. Cette inégalité est vraie pour le cas particulier des vecteurs a_i, $i = 1, ..., k$. On a donc trouvé $p \neq 0$ tel que $g^T p < 0$ et $p^T a_i \geq 0$, pour tout $i = 1, ..., k$. C'est en contradiction avec l'hypothèse et conclut la preuve.

□

La notion de cône convexe est utilisée pour caractériser les solutions de notre problème d'optimisation.

Proposition 7. — *Si x^* est solution du problème* (32) *alors (sous les hypothèses du paragraphe §2.2.2.1) $\nabla f(x^*)$ est dans le cône convexe engendré par $\{-\nabla c_i(x^*), i \in \mathcal{I}\}$ où $\mathcal{I} = \{i = 1, .., m \text{ tel que } c_i(x^*) = 0\}$ (famille des indices des contraintes actives).*

Démonstration. — On va montrer que si $\nabla f(x^*)$ n'appartient pas au cône convexe engendré par $\{-\nabla c_i(x^*), i \in \mathcal{I}\}$ alors x^* n'est pas minimiseur. D'après le lemme 5 de Farkas, $\exists h \in \mathbb{R}^n$ tel que $\nabla f(x^*)^T h < 0$ et $-\nabla c_i(x^*)^T h \geq 0$ pour tout $i \in \mathcal{I}$. En suivant cette direction pour $\delta \in \mathbb{R}$ il vient

$$f(x^* + h\delta) = f(x^*) + \nabla f(x^*)^T h\delta + \circ(\delta^2)$$

$$c_i(x^* + h\delta) = c_i(x^*) + \nabla c_i(x^*)^T h\delta, \forall i \in \mathcal{I}$$

Donc pour δ suffisamment petit

$$\begin{aligned} &f(x^* + h\delta) < f(x^*) \\ &c_i(x^* + h\delta) \leq c(x^*) \leq 0, \forall i = 1, ..., m \end{aligned}$$

Tout voisinage de x^* contient un meilleur point que x^* qui n'est donc pas minimiseur.

□

Cas général Dans le cas où les contraintes $\mathbb{R}^n \ni x \mapsto c_i(x) \in \mathbb{R}$ ne sont pas affines, il peut ne pas exister de direction donnée par le lemme de Farkas utilisable pour trouver un meilleur point que x^*. La courbure des contraintes peut gêner une telle construction. D'autre part, le cône convexe engendré par les contraintes actives peut dégénérer (la frontière peut ressembler à un point de rebroussement).

Néanmoins, il est possible d'étendre le résultat précédent dans le cas général sous une hypothèse supplémentaire souvent peu gênante en pratique.

Théorème 26 (Conditions KKT)

Considérons un point $x^* \in \mathbb{R}^n$. Notons la famille des indices des contraintes actives en x^* par $\mathcal{I} = \{i \in \{1, .., m \text{ tel que } c_i(x^*) = 0\}\}$. Supposons que x^* est un point régulier pour ces contraintes, c.-à-d. que la famille $(\nabla c_i(x^*))_{i\in\mathcal{I}}$ est une famille libre (on dit que les contraintes sont qualifiées). Sous ces hypothèses, si x^* est une solution du problème (32) alors $\nabla f(x^*)$ appartient au cône convexe engendré par $(-\nabla c_i(x^*))_{i\in\mathcal{I}}$. En conséquence, $\exists \lambda_i \geq 0, i = 1, ..., m$ tels que $\nabla f(x^*) = -\sum_{i=1}^m \lambda_i \nabla c_i(x^*)$ et $\lambda_i c_i(x^*) = 0$ pour $i = 1, .., m$.

Ce résultat (admis, voir [**20**]) est une condition nécessaire aux points réguliers candidats pour être minimum. On calculera par les multiplicateurs de Lagrange la combinaison linéaire qui permet d'exprimer le gradient du coût en fonction des gradients des contraintes actives et on vérifiera que les multiplicateurs sont bien tous positifs ou nuls.

En pratique on travaillera souvent avec des fonctions convexes et on exploitera le résultat suivant

***Proposition* 8.** — *Soit le problème d'optimisation* $\min_{c(x)\leq 0} f(x)$ *où les fonctions* f *et* c *sont différentiables et convexes. On suppose qu'il existe* $x \in \mathbb{R}^n$ *tel que* $c(x) < 0$. *Les conditions KKT du théorème 26 sont nécessaires et suffisantes pour que* x^* *soit un minimiseur global.*

2.2.2.2 Dualité

Considérons le problème de la minimisation dans X sous ensemble de $\mathbb{R}^n$, $n < \infty$ d'une fonction différentiable (fonction coût) $X \ni x \mapsto f(x) \in \mathbb{R}$ sous la contrainte $c(x) \leq 0$ définie par une fonction différentiable $\mathbb{R}^n \ni x \mapsto c(x) \in \mathbb{R}^m$

$$\begin{aligned} &\min_x \quad f(x) \\ &\text{tel que } x \in X, \\ &\qquad c(x) \leq 0 \end{aligned} \tag{33}$$

Considérons alors le Lagrangien associé à ce problème

$$X \times L \ni (x, \lambda) \mapsto \mathcal{L}(x, \lambda) = f(x) + \lambda^T c(x) \in \mathbb{R} \tag{34}$$

où $L \subset \mathbb{R}^m$.

Définition 20

On dit que (x^*, λ^*) est un point selle de $\mathcal{L}$ si x^* est un minimiseur pour $X \ni x \mapsto \mathcal{L}(x, \lambda^*)$ et λ^* est un maximum pour $L \ni \lambda \mapsto \mathcal{L}(x^*, \lambda)$ ou encore si

$$\sup_{\lambda\in L} \mathcal{L}(x^*, \lambda) = \mathcal{L}(x^*, \lambda^*) = \inf_{x\in X} \mathcal{L}(x, \lambda^*) \tag{35}$$

On a aussi pour tout $x \in X$ et pour tout $\lambda \in L$

$$\mathcal{L}(x^*, \lambda) \leq \mathcal{L}(x^*, \lambda^*) \leq \mathcal{L}(x, \lambda^*)$$

D'après le théorème suivant, on peut intervertir l'ordre de la maximisation et de la minimisation

Théorème 27 (Théorème du point selle)

Si (x^*, λ^*) est un point selle de $\mathcal{L}$ sur $X \times L$ alors

$$\sup_{\lambda \in L} \inf_{x \in X} \mathcal{L}(x, \lambda) = \mathcal{L}(x^*, \lambda^*) = \inf_{x \in X} \sup_{\lambda \in L} \mathcal{L}(x, \lambda)$$

Démonstration. — Par définition, on a

$$\mathcal{L}(x, \lambda) \leq \sup_{\lambda \in L} \mathcal{L}(x, \lambda)$$

et donc

$$\inf_{x \in X} \mathcal{L}(x, \lambda) \leq \inf_{x \in X} \sup_{\lambda \in L} \mathcal{L}(x, \lambda)$$

et finalement

$$\sup_{\lambda \in L} \inf_{x \in X} \mathcal{L}(x, \lambda) \leq \inf_{x \in X} \sup_{\lambda \in L} \mathcal{L}(x, \lambda)$$

Utilisons maintenant l'égalité du point selle (35) sur

$$\begin{aligned} \inf_{x \in X} \sup_{\lambda \in L} \mathcal{L}(x, \lambda) \leq \sup_{\lambda \in L} \mathcal{L}(x^*, \lambda) &= \mathcal{L}(x^*, \lambda^*) \\ &= \inf_{x \in X} \mathcal{L}(x, \lambda^*) \\ &\leq \sup_{\lambda \in L} \inf_{x \in X} \mathcal{L}(x, \lambda) \end{aligned}$$

Finalement

$$\sup_{\lambda \in L} \inf_{x \in X} \mathcal{L}(x, \lambda) = \mathcal{L}(x^*, \lambda^*) = \inf_{x \in X} \sup_{\lambda \in L} \mathcal{L}(x, \lambda)$$

□

Lorsque (x^*, λ^*) n'est pas un point selle, l'égalité précédente n'est pas vraie, on n'a que

$$\sup_{\lambda \in L} \inf_{x \in X} \mathcal{L}(x, \lambda) \leq \inf_{x \in X} \sup_{\lambda \in L} \mathcal{L}(x, \lambda)$$

on parle de "saut de dualité" lorsque les deux résultats sont différents.

On appelle problème primal le problème $\min_{x \in X} \max_{\lambda \in L} \mathcal{L}(x, \lambda)$, et problème dual $\max_{\lambda \in L} \min_{x \in X} \mathcal{L}(x, \lambda)$.

Les deux théorèmes suivants permettent de relier notre problème d'optimisation sous contraintes inégalités à l'existence d'un point selle.

***Théorème* 28 (Optimalité du point selle).** — *Si (x^*, λ^*) est un point selle de $\mathcal{L}$ sur $X \times (\mathbb{R}^+)^m$, alors x^* est solution du problème* (32).

Démonstration. — Par définition, quel que soit $\lambda \in (\mathbb{R}^+)^m$ on a

$$\mathcal{L}(x^*, \lambda) \leq \mathcal{L}(x^*, \lambda^*)$$

Ceci peut s'écrire

$$(\lambda - \lambda^*)^T c(x^*) \leq 0 \tag{36}$$

En exprimant cette inéquation pour $\lambda = 0$ et $\lambda = 2\lambda^*$ il vient

$$(\lambda^*)^T c(x^*) = 0$$

Il vient aussi en exprimant le produit scalaire (36) composante par composante

$$c(x^*) \leq 0$$

D'autre part, $\mathcal{L}(x^*, \lambda^*) \leq \mathcal{L}(x, \lambda^*)$ pour tout $x \in X$. On a donc

$$f(x^*) \leq f(x) + (\lambda^*)^T c(x)$$

Donc pour tout $x \in X$ tel que $c(x) \leq 0$ on a $f(x^*) \leq f(x)$ ce qui achève la démonstration. □

Sans hypothèse sur la régularité des f et de c, on peut énoncer un théorème analogue aux conditions KKT du théorème 26.

Théorème 29 (existence de point selle)

Supposons $X \ni x \mapsto f(x) \in \mathbb{R}$ et $\mathbb{R}^n \ni x \mapsto c(x) \in \mathbb{R}^m$ convexe. Soit x^* solution du problème (32) tel que les contraintes actives en x^* sont qualifiées, alors il existe un point selle (x^*, λ^*) du Lagrangien (34).

2.2.2.3 Algorithme de minimisation sous contraintes utilisant la dualité

Les algorithmes de cette section sont des algorithmes de recherche de point selle

Algorithme 9 (Algorithme d'Uzawa)

À partir de $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$, $\alpha \in \mathbb{R}^+$ quelconques, on note $\mathbb{R}^m \ni \lambda \mapsto P(\lambda) \in (\mathbb{R}^+)^m$ la projection sur $(\mathbb{R}^+)^m$, itérer

$$\text{résoudre } \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda^k), \text{ on note } x^{k+1} \text{ la solution}$$
$$\lambda^{k+1} = P\left(\lambda^k + \alpha c(x^{k+1})\right)$$

La seconde partie de la boucle à itérer consiste à effectuer un gradient à pas constant pour maximiser $\mathbb{R}^m \ni \lambda \mapsto \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda)$. Une variante de cet algorithme est

***Algorithme* 10 (Algorithme d'Arrow-Hurwicz).** — À partir de $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$ quelconques, $\varepsilon \in \mathbb{R}^+$ on note $\mathbb{R}^m \ni \lambda \mapsto P(\lambda) \in (\mathbb{R}^+)^m$ la projection sur $(\mathbb{R}^+)^m$,

itérer

$$x^{k+1} = x^k - \varepsilon \left(\nabla f(x^k) + \frac{\partial c}{\partial x}(x^k)\lambda^k \right)$$
$$\lambda^{k+1} = P\left(\lambda^k + \alpha c(x^{k+1})\right)$$

2.2.2.4 Méthode de contraintes actives pour les problèmes de programmation quadratique

Nous allons maintenant présenter un algorithme efficace (y compris lorsque les dimensions n et m sont grandes) de résolution de problèmes de minimisation d'une fonction quadratique sous contraintes affines.

On va traiter le cas

$$f(x) = \frac{1}{2}x^T Gx + x^T d$$

sous contraintes affines

$$c(x) = Ax - b \leq 0$$

et on cherche à résoudre le problème de minimisation (32). On suppose que G est une matrice symétrique définie positive de $\mathcal{M}_n(\mathbb{R})$ et A une matrice de $\mathcal{M}(m,n)$, avec $m > n$. On notera a_i la i-ème ligne de A et b_i la i-ème composante de $b \in \mathbb{R}^m$.

L'algorithme constitue et met à jour un ensemble de contraintes actives jusqu'à ce que cet ensemble satisfasse les conditions de Karush-Kuhn-Tucker du théorème 26. La mise à jour s'effectue en repérant des directions d'amélioration. On suit ces directions jusqu'à atteindre une nouvelle contrainte bloquante. On élimine si besoin les contraintes, pour pouvoir avancer, en vérifiant le signe de leur multiplicateur de Lagrange. À chaque étape k, on définit W^k ensemble de travail (sous-ensemble) de A_c^k ensemble des contraintes actives au point x^k. On suppose que les gradients (a_i) des contraintes de W^k sont linéairement indépendants.

***Algorithme* 11 (Algorithme de contraintes actives QP)**

À partir de W^0 ensemble de travail de départ, $x^0 \in \mathbb{R}^n$ point où les j contraintes de A constituant W^0 sont actives, itérer

1. Vérifier si x^k est un minimiseur de $f(x)$ sous les contraintes $a_i^T x - b_i \leq 0$, $i \in W^k$. Si c'est le cas l'algorithme s'achève et x^k est la solution recherchée. Sinon continuer au point 2.
2. Résoudre $\min_{a_i^T x - b_i \leq 0, i \in W^k} f(x)$:
   (a) Calculer une direction p^k solution de $\min_{a_i^T p = 0, i \in W^k} \frac{1}{2}p^T Gp + (Gx^k + d)^T p$
   (b) si $p^k \neq 0$ noter $\alpha^k = \min\left(1, \min_{i \notin W^k, a_i^T p^k > 0} \frac{b_i - a_i^T x^k}{a_i^T p^k}\right)$
   Mettre à jour $x^{k+1} = x^k + \alpha^k p^k$. Si $\alpha^k < 1$ il existe un j réalisant le précédent minimum. Mettre à jour $W^{k+1} = W^k \bigcup \{j\}$. Sinon, prendre $W^{k+1} = W^k$.
   (c) si $p^k = 0$ calculer les multiplicateurs de Lagrange $\lambda_i, i \in W^k$ du problème $\min_{a_i^T x - b_i = 0} f(x)$. Si tous ces multiplicateurs de Lagrange sont positifs

l'algorithme s'achève et x^k est la solution recherchée. Dans le cas contraire éliminer de W^k la contrainte ayant le plus petit multiplicateur de Lagrange et choisir $x^{k+1} = x^k$.

***Proposition* 9**. — *Avec les notations précédentes, l'algorithme 11 de contraintes actives possède les propriétés suivantes*

1. *On ne revisite jamais exactement l'ensemble W^k.*
2. *La suite des directions $(p^k)_k \in \mathbb{N}$ revient n-périodiquement à la valeur $p^k = 0$.*

Démonstration. — La suite des valeurs $(f(x^k))_k \in \mathbb{N}$ est décroissante. Elle est strictement décroissante aux indices tels que $\alpha^k > 0$. On ne revisite jamais exactement l'ensemble W^k car x^{k+1} est le minimiseur sous les contraintes d'indices éléments de W^k. La deuxième propriété découle du raisonnement suivant. Si $p^k \neq 0$ le point x^k est mis à jour. Soit le pas α^k utilisé dans cette mise à jour est égal à 1 et on a obtenu un minimiseur donc $p^{k+1} = 0$, soit $\alpha^k < 1$ et on doit rajouter une contrainte à l'ensemble W^k. Au bout de $l < n$ étapes, W^k contient n contraintes linéairement indépendantes. Le problème est complètement contraint et on a donc $p^{k+l} = 0$. □

***Proposition* 10**. — *Avec les notations précédentes, l'algorithme 11 de contraintes actives converge en un nombre fini d'itérations vers l'unique minimum global du problème* (32).

Démonstration. — Il existe un nombre fini de W^k possibles. L'algorithme parcourt cet ensemble jusqu'à obtenir l'optimum. □

Cet algorithme est également très utilisé pour résoudre une successions de problèmes où une fonction coût non linéaire est approchée par son développement de Taylor au deuxième ordre et où les contraintes sont approchées par leur développement de Taylor au premier ordre. De tels algorithmes SQP (successive quadratic programming) sont détaillés dans [**11, 12**].

2.2.2.5 Méthode de contraintes actives pour les problèmes de programmation linéaire : méthode du Simplexe

Enfin, nous présentons un algorithme de résolution de problèmes de minimisation d'une fonction linéaire sous contraintes affines, appelé problèmes de Programmation Linéaire (LP). Cet algorithme s'inscrit dans la famille des algorithmes de contraintes actives, présentée précédemment. Il est cependant ici décliné sous une forme spécifique aux problèmes LP, qui sont très fréquemment rencontrés en pratique.

Plus précisément, on s'intéresse au problème suivant

$$\begin{aligned} \min_z \quad & c_z^T z \\ \text{tel que } & A_{in} z \leq b_{in}, \\ & A_{eq} z = b_{eq} \end{aligned} \tag{37}$$

Introduisons $s \geq 0$ tel que $A_{in} z + s = b_{in}$. Quitte à redéfinir A_{in}, A_{eq}, b_{eq} et c avec un signe opposé sur certaines composantes, on peut, en définissant $x = (z, s)$, reformuler

le problème sous la forme générique que nous traiterons dorénavant

$$\begin{aligned} &\min_x \quad c^T x \\ &\text{tel que } Ax = b, \\ &\qquad x \geq 0 \end{aligned} \tag{38}$$

avec $x \in \mathbb{R}^n$ et $b \in \mathbb{R}^m$. On suppose $m < n$ et A de rang m (sans nuire aux généralités, puisque l'on peut toujours retirer les contraintes redondantes).

***Définition* 21**. — Soit $A_B \in \mathbb{R}^{m\times m}$ une sous-matrice de A de rang m et soit x_B l'unique solution de $A_B x_B = b$. Le vecteur dont les m coordonnées associées aux colonnes de A_B sont celles de x_B et dont les autres coordonnées sont nulles est appelé *solution de base* de $Ax = b$. Les vecteurs de la base unité de $\mathbb{R}^m$ associés aux colonnes de A_B sont appelées *variables de base*, et ceux associés aux colonnes restantes de A sont appelés *variables hors-base*.

***Définition* 22**. — Un vecteur $x \in \mathbb{R}^n$ satisfaisant les contraintes $Ax = b$ et $x \geq 0$ est appelé *solution réalisable*.

Le résultat suivant justifie que l'on s'intéresse en particulier aux solutions de base.

***Proposition* 11**. — *S'il existe une solution réalisable (resp. optimale) au Problème* (38)*, alors il en existe une solution de base réalisable (resp. optimale).*

Démonstration. — Nous prouvons ici la proposition portant sur une solution optimale (celle sur une solution réalisable se prouve de façon similaire). Notons a_i $(i = 1, \ldots, n)$ les colonnes de A et $x^\star$ une solution optimale. Supposons que p coordonnées ($p \in \{0, \ldots, n\}$) de $x^\star$ sont strictement positives et, à un réordonnancement près, que $x_i^\star > 0$ pour $i = 1, \ldots, p$ et $x_i^\star = 0$ pour $i = p+1, \ldots, n$. On distingue alors deux cas.
Cas 1: supposons $(a_1, \ldots, a_p)$ linéairement indépendantes (ce qui implique $p \leq m$). On peut alors compléter cette famille avec $m - p$ colonnes de A parmi les $n - p$ restantes pour former une base de $\mathbb{R}^m$, car A est de rang m. $x^\star$ est donc une solution optimale de base.
Cas 2: supposons $(a_1, \ldots, a_p)$ linéairement dépendantes. Alors, il existe des scalaires $\mu_1, \ldots, \mu_p$ tels que $\sum_{i=1}^p \mu_i a_i = 0$ et qu'au moins l'un des coefficients μ_i est strictement positif. Soient $\epsilon > 0$ et $\mu = (\mu_1, \ldots, \mu_p, 0, \ldots, 0) \in \mathbb{R}^n$ (quitte à réordonner les colonnes de A). Comme $x^\star$ satisfait les contraintes, il s'ensuit

$$A(x^\star \pm \epsilon\mu) = Ax^\star = b \tag{39}$$

Par ailleurs, pour ϵ suffisamment petit, $x^\star \pm \epsilon\mu \geq 0$. Ainsi, $x^\star + \epsilon\mu$ et $x^\star - \epsilon\mu$ sont deux solutions réalisables, de coûts respectifs $c^T(x^\star + \epsilon\mu)$ et $c^T(x^\star - \epsilon\mu)$. Si $c^T\mu \neq 0$, alors l'un de ces deux coûts est strictement inférieur à $c^T x^\star$, ce qui est absurde car $x^\star$ solution optimale. En conséquence, $c^T\mu = 0$ et $x^\star - \epsilon\mu$ est également une solution optimale. Enfin, pour $\epsilon = \min\left\{\frac{x_i}{\mu_i} \mid \mu_i > 0\,,\ i = 1, \ldots, p\right\}$, $x^\star - \epsilon\mu \geq 0$ et a au plus $p - 1$ coordonnées non-nulles. De ce qui précède, il s'agit également d'une solution optimale. On itère le processus jusqu'à se ramener au cas 1. □

Ces solutions réalisables de base ont par ailleurs une interprétation géométrique : il s'agit des sommets du polytope convexe des contraintes.

***Définition* 23**. — Soit C un ensemble convexe. Un point x de C est appelé point extrême de C s'il n'existe pas deux points distincts x_1 et x_2 dans C tels que $x = \alpha x_1 + (1-\alpha)x_2$ pour un certain $\alpha \in]0,1[$.

***Théorème* 30**. — *Soit $K = \{x \in \mathbb{R}^n \mid Ax = b \text{ et } x \geq 0\}$ le polytope convexe des solutions réalisables de* (38). *$x \in K$ est un point extrême si et seulement si x est une solution réalisable de base.*

Démonstration. — Supposons que x est une solution réalisable de base, que l'on écrit $x = (x_1, \ldots, x_m, 0, \ldots, 0)$ avec $(a_1, \ldots, a_m)$ colonnes linéairement indépendantes de A. Supposons par l'absurde qu'il existe $(w, z) \in K^2$ tels que $w \neq z$ et $x = \alpha w + (1-\alpha)z$ pour un certain $\alpha \in]0,1[$. Alors, comme $w \geq 0$ et $z \geq 0$, $w_i = z_i = x_i = 0$ pour $i = m+1, \ldots, n$ et

$$A_B w = A_B z = A_B x = b \tag{40}$$

avec $A_B = (a_1, \ldots, a_m)$ inversible. Ainsi, $w = z = x$, absurde.

Réciproquement, considérons x un point extrême de K, que l'on décompose sous la forme $x = (x_1, \ldots, x_k, 0, \ldots, 0)$ avec $x_i \neq 0$ pour $i = 1, \ldots, k$ et $k \geq 2$ (le cas $k \leq 1$ est trivial). Supposons par contradiction que $(a_1, \ldots, a_k)$ soient linéairement dépendants. Alors, $\sum_{i=1}^{k} \mu_i a_i = 0$ avec au moins un des $\mu_i \neq 0$. Soient $\epsilon > 0$ et $\mu = (\mu_1, \ldots, \mu_k, 0, \ldots, 0) \in \mathbb{R}^n$. Alors, pour ϵ suffisamment faible, $x \pm \epsilon\mu \in K$ et

$$x = \frac{1}{2}(x + \epsilon\mu) + \frac{1}{2}(x - \epsilon\mu) \tag{41}$$

ce qui est absurde car x est un point extrême de K. Aussi, $(a_1, \ldots, a_k)$ sont linéairement indépendants, donc $k \leq m$ car A est de rang plein. Par conséquent, x est une solution réalisable de base. □

***Corollaire* 1**. — *L'ensemble K possède un nombre fini de points extrêmes.*

Démonstration. — Il existe un nombre fini de solutions de base, puisqu'il existe un nombre fini de sous-matrices de A de rang m. Les points extrêmes étant un sous-ensemble de l'ensemble des solutions de base selon le Théorème 30 (i.e., celles réalisables), on conclut. □

Selon le Théorème 30 et la Proposition 11, il suffit donc de visiter les solutions de base réalisables pour trouver une solution optimale. Pour passer d'une solution de base à une autre, on peut utiliser la méthode du pivot de Gauss sur la matrice A. Néanmoins, rien ne garantit que la nouvelle solution de base obtenue sera une solution réalisable.

Pour obtenir cela, supposons que l'on dispose d'une solution réalisable de base non-dégénérée x, c.-à-d. avec m coordonnées non-nulles $x_1, \ldots, x_m$. On décompose $A = (A_B \ A_N)$ et $c = (c_B^T \ c_N^T)^T$ selon les variables de base et les variables hors-base. On souhaite modifier ces variables de base pour y inclure une variable x_q

($q \in \{m+1, \dots, n\}$). Soient $\mu = A_B^{-1} a_q$ et $z = \epsilon(\mu^T, e_{q-m}^T)^T$ pour un certain ϵ, avec $(e_i)_{1 \leq i \leq n-m}$ base canonique de $\mathbb{R}^{n-m}$. Alors

$$A(x - \epsilon z) = A_B x - \epsilon A_B \mu - \epsilon A_N e_q = A_B x - \epsilon a_q + \epsilon a_q = b \tag{42}$$

Supposons qu'au moins l'une des coordonnées de μ soit positive. Alors, $x - \epsilon z \geq 0$ avec $\epsilon = \min \left\{ \frac{x_i}{\mu_i} \mid \mu_i > 0\,,\ i = 1, \dots, p \right\}$, et donc $x - \epsilon z$ est alors une nouvelle variable de base réalisable.

Reste à déterminer comment choisir la coordonnée x_q à faire entrer dans la base. Un critère d'arbitrage naturel est celui du coût, que l'on cherche à faire décroître entre deux itérations. Pour déterminer une solution réalisable de base d'intérêt, on considère la variable appelée de *coût réduit*

$$\lambda = c_N - A_N^T A_B^{-T} c_B \tag{43}$$

et l'on utilise le résultat suivant.

***Théorème* 31**. — *Supposons que l'on dispose d'une solution réalisable de base non-dégénérée (c.-à-d. avec m coordonnées non-nulles) dont on note f_0 le coût.*

- *S'il existe $j \in \{1, \dots, n-m\}$ tel que $\lambda_j < 0$, alors il existe une solution réalisable de base avec un coût $f < f_0$. Soit $y_q = A_B^{-1} a_q$ pour $q \in \{m+1, \dots, n\}$, alors*
  - *(i) si $(y_q)_i > 0$ pour un certain $i \in \{1, \dots, m\}$, cette solution est obtenue en rendant x_q de base.*
  - *(ii) sinon le problème est non-borné.*
- *Si $\lambda \geq 0$, alors la solution est optimale.*

Démonstration. — Considérons $x = (x_1, \dots, x_m, 0, \dots, 0)$ une solution réalisable de base non-dégénérée de coût f_0 et supposons $\lambda_1 < 0$ (sans nuire aux généralités). On décompose $A = (A_B\ A_N)$ et $c = (c_B^T\ c_N^T)^T$ selon les variables de base et les variables hors-base. Soit $x' = (x_B'^T, x_N'^T)^T$, avec $x'_B = (x'_1, \dots, x'_m)$ et $x'_N = (x'_{m+1}, 0, \dots, 0)$, une autre variable réalisable telle que $x'_{m+1} > 0$ (obtenue par la procédure décrite précédemment par exemple). Alors, il s'ensuit

$$\begin{aligned} A_B x'_B + A_N x'_N = b = A_B x'_B + x'_{m+1} a_{m+1} \Leftrightarrow x'_B &= A_B^{-1} \left(b - x'_{m+1} a_{m+1} \right) \\ &= x_B - x'_{m+1} y_{m+1} \end{aligned} \tag{44}$$

De plus, en notant e_1 le premier vecteur de la base canonique de $\mathbb{R}^{n-m}$, le coût correspondant à x' s'écrit

$$c_B^T x'_B + c_{m+1} x'_{m+1} = c_B^T x_B + \left(-c_B^T A_B^{-1} a_{m+1} + c_{m+1} \right) x'_{m+1} \tag{45}$$

$$= c_B^T x_B + \left(-c_B^T A_B^{-1} A_N + c_N^T \right) e_1 x'_{m+1} \tag{46}$$

$$= c_B^T x_B + \lambda^T e_1 x'_{m+1} = c_B^T x_B + \lambda_1 x'_{m+1} < c_B^T x_B \tag{47}$$

puisque $\lambda_1 < 0$ et $x'_{m+1} > 0$.

Si y_{m+1} admet une composante strictement positive, alors, selon (47), une des composantes de x'_B va décroître linéairement avec x'_{m+1} et on augmente donc x'_{m+1} jusqu'à annuler cette composante et obtenir une nouvelle variable réalisable de base.

Sinon, on peut augmenter indéfiniment x'_{m+1} tout en obtenant une solution réalisable. Le problème est donc non-borné.

Maintenant, supposons que $\lambda \geq 0$. Dans la base courante, on a donc $x_B = A_B^{-1}(b - A_N x_N)$ et, en substituant, le problème se réécrit donc sous la forme

$$\min_{x_N \geq 0} c_B^T A_B^{-1}(b - A_N x_N) + c_N^T x_N = \min_{x_N \geq 0} \left(c_N - A_N^T A_B^{-T} c_B\right)^T x_N + c_B^T A_B^{-1} b \tag{48}$$

On a ainsi

$$\nabla f(x) = c_N - A_N^T A_B^{-T} c_B = \lambda \geq 0 \tag{49}$$

c.-à-d. que les conditions de KKT sont satisfaites pour la solution réalisable de base correspondante. Le problème étant convexe, on conclut qu'il s'agit d'une solution optimale. □

Ceci nous amène à considérer l'algorithme suivant.

Algorithme 12 (Algorithme du simplexe)

A partir d'une solution réalisable de base x_B^0, des décompositions $c^0 = (c_B^0\ c_N^0)$ et $A^0 = (A_B^0\ A_N^0)$ de c et A associées et de l'inverse $M^0 = (A_B^0)^{-1}$, itérer

1. calculer $\lambda = c_N^k - (A_N^k)^T (A_B^k)^{-T} c_B^k$. Si $\lambda \geq 0$, l'algorithme s'achève et renvoie x_B^k.
2. calculer $q = \text{argmin}_{i=1,\ldots,n-m} \lambda_i$ et $y_q = A_B^{-1} a_{m+q}$ où a_{m+q} q^{ieme} colonne de A_N^k.
3. si $y_q \leq 0$, l'algorithme s'achève et déclare le problème non-borné. Sinon, sélectionner $p = \text{argmin}_{i=1,\ldots,m} \frac{b_i}{(y_q)_i}$
4. Remplacer x_p par x_q dans la base, mettre à jour $M^{k+1} = (A_B^{k+1})^{-1}$, calculer $x_B^{k+1} = (A_B^{k+1})^{-1} b$, les décompositions de c et A associées et retourner à l'étape 1.

A noter que l'étape 4 peut être réalisée itérativement (en calculant M^{k+1} par la méthode du pivot appliquée sur la matrice $((A_B^k)^{-1} \mid y_q)$ avec $(y_q)_p$ comme pivot), sans nécessiter de réaliser explicitement l'opération de changement de base.

On a le résultat suivant de convergence pour cet algorithme.

Théorème 32

Si toute solution réalisable de base est non-dégénérée, l'Algorithme 12 du simplexe converge en un nombre fini d'étapes, soit en déterminant que le problème est non-borné, soit en trouvant une solution optimale.

Démonstration. — C'est une conséquence directe du Théorème 31 et du Corollaire 1. □

A noter que cet algorithme nécessite de connaître initialement une solution réalisable (que l'on peut alors rendre de base selon les discussions précédentes). Pour ce faire, il existe un certain nombre d'heuristiques ou méthodes reposant sur la résolution préalable d'un autre problème d'optimisation linéaire, facilement initialisable, comme détaillé dans [**8**].

Le cas où certaines solutions de base réalisables sont dégénérées se produit lorsque les contraintes $Ax = b$ et $x \geq 0$ sont redondantes entre elles. Dans ce cas, la pratique montre que l'algorithme du simplexe converge aussi très souvent en un nombre fini d'itérations. Néanmoins, il est également possible que l'algorithme effectue un pivot de solution de base sans améliorer le coût, et ce bien que le coût réduit associé soit négatif. Ceci peut alors amener à revisiter plusieurs fois les mêmes points extrêmes, et à entrer dans une boucle. Des méthodes d'anti-cyclage [**5, 18**] permettent d'éviter ce phénomène.

Cet algorithme très célèbre est également utilisé pour résoudre une succession de problèmes *mixtes-entiers*, c.-à-d. comprenant des variables continues, mais également des variables binaires ou entières. De tels algorithmes sont détaillés dans [**20**] ou dans [**14**] pour une fonction coût non-linéaire, qui est approchée séquentiellement par des problèmes de programmation linéaire mixtes.

CHAPITRE 3

MÉTHODES POUR L'OPTIMISATION NON-DIFFÉRENTIABLE DE DIMENSION FINIE

Dans ce chapitre, on présente les outils algorithmique permettant de traiter des problèmes d'optimisation convexe faisant intervenir des fonctions non-différentiables.

3.1 Eléments avancés d'analyse convexe

3.1.1 Tranformée de Fenchel

Définition 24

Soit $f : \mathbb{R}^n \to \mathbb{R}$. On appelle transformée de Fenchel de f la fonction f^* définie par

$$f^*(\varphi) = \sup_{x \in \mathbb{R}^n} \left(\varphi^T x - f(x)\right)$$

***Lemme* 6**. — *Pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$, $f^\star$ est convexe.*

Démonstration. — Soit $h_x : \varphi \mapsto \varphi^T x - f(x)$ fonction affine et donc convexe. Par définition, $f^\star(\varphi) = \sup_{x \in \mathbb{R}^n} h_x(\varphi)$. On a donc

$$\begin{aligned}(\varphi, y) \in \mathrm{Epi}(f^\star) \quad &\Leftrightarrow \quad y \geq \sup_{x \in \mathbb{R}^n} h_x(\varphi) \\ &\Leftrightarrow \quad \forall x \in \mathbb{R}^n \quad y \geq h_x(\varphi) \\ &\Leftrightarrow \quad (\varphi, y) \in \cap_{x \in \mathbb{R}^n} \mathrm{Epi}(h_x)\end{aligned}$$

dont on déduit $\mathrm{Epi}(f^\star) = \cap_{x \in \mathbb{R}^n} \mathrm{Epi}(h_x)$. $\mathrm{Epi}(h_x)$ étant convexe, on en déduit qu'il en est de même pour $\mathrm{Epi}(f^\star)$, c.-à-d. $f^\star$ convexe en appliquant le Théorème 5. □

***Théorème* 33**. — *Pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ continue, sa biconjuguée $f^{\star\star}$ définie par*

$$f^{**}(x) = \sup_{\varphi \in \mathbb{R}^n} \left(x^T \varphi - f^\star(\varphi)\right)$$

admet pour épigraphe l'enveloppe convexe de $\mathrm{Epi}(f)$.

Démonstration. — Soit $\Sigma \subset \mathbb{R}^n \times \mathbb{R}$ l'ensemble des paires (φ, α) telles que la fonction affine $x \mapsto \varphi^T x - \alpha$ est majorée par f. Alors,

$$\begin{aligned}(\varphi,\alpha) \in \Sigma \quad &\Leftrightarrow \quad \forall x \in \mathbb{R} \quad f(x) \geq \varphi^T x - \alpha \\ &\Leftrightarrow \quad \forall x \in \mathbb{R} \quad \alpha \geq \varphi^T x - f(x) \\ &\Leftrightarrow \quad \alpha \geq f^\star(\varphi)\end{aligned}$$

Aussi, pour $x \in \mathbb{R}^n$,

$$\sup_{(\varphi,\alpha)\in\Sigma} \left(\varphi^T x - \alpha\right) = \sup_{\varphi\in\mathbb{R}^n, \alpha \geq f^\star(\varphi)} \left(\varphi^T x - \alpha\right) = \sup_{\varphi\in\mathbb{R}^n} \left(x^T\varphi - f^\star(\varphi)\right) = f^{\star\star}(x)$$

Ainsi, la biconjuguée de f est la plus grande fonction affine inférieure à f. Son épigraphe est donc le plus petit ensemble convexe contenant $\mathrm{Epi}(f)$, soit $\mathrm{Conv}(\mathrm{Epi}(f))$. □

Une conséquence directe de ce théorème est le résultat suivant.

Théorème 34 (Théorème de Moreau-Fenchel)

Soit $f : \mathbb{R}^n \to \mathbb{R}$ continue. La biconjuguée de f satisfait $f^{\star\star} = f$ ssi f est convexe.

***Lemme* 7.** — *Soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe. Alors $\varphi \in \partial f(x)$ si et seulement si $x \in \partial f^\star(\varphi)$.*

Démonstration. — Soit $h_\varphi : s \in \mathbb{R}^n \mapsto f(s) - \varphi^T s$, fonction convexe. On a $\varphi \in \partial f(x)$ ssi $0 \in \partial h_\varphi(x)$ ssi x minimiseur global de h_φ. Par définition de la transformée de Fenchel, on obtient donc

$$x \text{ minimiseur global de } h_\varphi \quad \Leftrightarrow \quad f^\star(\varphi) = -h_\varphi(x) = \varphi^T x - f(x)$$

f étant convexe, $f = f^{\star\star}$ et cette dernière caractérisation peut se reformuler comme $f^{\star\star}(x) = \varphi^T x - f^\star(\varphi)$. Par un raisonnement en tout point similaire, en définissant $h_x(\varphi) : s \in \mathbb{R}^n \mapsto f^\star(\varphi) - x^T\varphi$, on a enfin

$$f^{\star\star}(x) = \varphi^T x - f^\star(\varphi) \;\Leftrightarrow\; \varphi \text{ min. global de } f^\star \;\Leftrightarrow\; x \in \partial f^\star(\varphi)$$

□

Théorème 35 (Régularisation des fonctions fortement convexes)

Soit $f : \mathbb{R}^n \to \mathbb{R}$ fortement convexe. Alors $f^\star$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}^n$ avec

$$\forall \varphi \in \mathbb{R}^n \quad \nabla f^\star(\varphi) = argmax_{x\in\mathbb{R}^n} \left(\varphi^T x - f(x)\right)$$

Démonstration. — Soit $\varphi \in \mathbb{R}^n$. La fonction $h_\varphi : x \in \mathbb{R}^n \mapsto \varphi^T x - f(x)$ admet un unique maximum (global) en $x^\star$ car f est fortement convexe. Ce minimiseur $x^\star$ est caractérisé par $0 \in \partial h_\varphi(x^*) = \varphi - \partial f(x^*) \Leftrightarrow \varphi \in \partial f(x^*) \Leftrightarrow x^* \in \partial f^\star(\varphi)$. Ainsi, par unicité de x^*, $\partial f^\star(\varphi)$ est un singleton, c.-à-d. $f^\star$ différentiable en φ. Par ailleurs, on en déduit $\partial f^\star(\varphi) = \{\nabla f^\star(\varphi)\} = \{\text{argmax}_{x\in\mathbb{R}^n} h_\varphi(x)\}$. □

3.1.2 Opérateur proximal

Définition 25

Soient $f : \mathbb{R}^n \to \mathbb{R}$ et $\mu > 0$. L'opérateur proximal de f (pour le paramètre μ), noté $\text{Prox}_{\mu f}$, est défini, pour tout $x \in \mathbb{R}^n$, par

$$\text{Prox}_{\mu f}(x) = \text{argmin}_{s\in\mathbb{R}^n} \left(f(s) + \frac{1}{2\mu}\|s - x\|^2 \right) \tag{50}$$

L'enveloppe de Moreau de f (pour le paramètre μ), notée $M_{\mu f}$, est définie, pour tout $x \in \mathbb{R}^n$, par

$$M_{\mu f}(x) = \min_{s\in\mathbb{R}^n} \left(f(s) + \frac{1}{2\mu}\|s - x\|^2 \right)$$

***Lemme* 8.** — *Soient f convexe et $\mu > 0$. Alors, pour tout $x \in \mathbb{R}^n$, $\text{Prox}_{\mu f}(x)$ existe, est unique et caractérisé par*

$$s = \text{Prox}_{\mu f}(x) \Leftrightarrow 0 \in \partial f(s) + \frac{1}{\mu}(s - x) \Leftrightarrow s = (\mu\partial f + Id)^{-1}(x)$$

L'enveloppe de Moreau est donnée par

$$M_{\mu f}(x) = f(\text{Prox}_{\mu f}(x)) + \frac{1}{2\mu}\|\text{Prox}_{\mu f}(x) - x\|^2$$

Démonstration. — La fonction f étant convexe, $s \mapsto f(s) + \frac{1}{2\mu}\|s - x\|^2$ est fortement convexe et admet donc un unique minimum global. $\text{Prox}_{\mu f}$ est ainsi bien défini et univoque (c.-à-d. que l'image de tout point de $\mathbb{R}^n$ est un singleton). Les caractérisations de $\text{Prox}_{\mu f}(x)$ et $M_{\mu f}$ découlent directement de leurs définitions et du fait que $\text{Prox}_{\mu f}$ est univoque. □

Noter que $\mu\partial f + Id : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ est une application dont les valeurs sont des ensembles. Ainsi, en toute rigueur, $(\mu\partial f + Id)^{-1}$ se doit d'être compris au sens de l'image réciproque entre des ensembles, c.-à-d. $(\mu\partial f + Id)^{-1}(\{x\}) = \{S \subset \mathbb{R}^n \mid S \subset \mu\partial f(x) + x\}$. Dans le cas présent, le résultat précédent implique que, pour tout $x \in \mathbb{R}^n$, $(\mu\partial f + Id)^{-1}(\{x\})$ est réduit à un singleton et nous interprétons et écrivons donc $(\mu\partial f + Id)^{-1}$ comme une fonction $\mathbb{R}^n \to \mathbb{R}^n$.

***Théorème* 36 (Régularisation de Moreau-Yosida).** — *Soient f convexe et $\mu > 0$. L'enveloppe de Moreau de f est de classe $\mathcal{C}^1$ sur $\mathbb{R}^n$ avec*

$$\forall x \in \mathbb{R}^n \quad \nabla M_{\mu f}(x) = \frac{1}{\mu}\left(x - \mathrm{Prox}_{\mu f}(x)\right)$$

Démonstration. — Nous commençons cette preuve par observer que l'enveloppe de Moreau peut être reformulée sous la forme suivante

$$\begin{aligned} M_{\mu f}(x) &= \min_{s\in\mathbb{R}^n}\left(f(s) + \frac{1}{2\mu}\|s-x\|^2\right) = \min_{s\in\mathbb{R}^n}\left(f(s) + \frac{1}{2\mu}(\|x\|^2 + \|s\|^2) - \frac{1}{\mu}x^T s\right) \\ &= \frac{\|x\|^2}{2\mu} - \frac{1}{\mu}\max_{s\in\mathbb{R}^n}\left(x^T s - \mu f(s) + \frac{1}{2}\|s\|^2\right) = \frac{\|x\|^2}{2\mu} - \frac{1}{\mu}\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^\star(x) \end{aligned}$$

La fonction f étant convexe, $\mu f + \frac{1}{2}\|\cdot\|^2$ est fortement convexe. Ainsi, selon le Théorème 35, $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^\star$ et donc $M_{\mu f}$ sont de classe $\mathcal{C}^1$ sur $\mathbb{R}^n$ et l'on a par ailleurs

$$\begin{aligned} \nabla M_{\mu f}(x) &= \frac{1}{\mu}x - \frac{1}{\mu}\nabla\left[\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^\star\right](x) \\ &= \frac{1}{\mu}x - \frac{1}{\mu}\mathrm{argmax}_{s\in\mathbb{R}^n}(x^T s - \mu f(s) - \frac{1}{2}\|s\|^2) \\ &= \frac{1}{\mu}x - \frac{1}{\mu}\mathrm{argmin}_{s\in\mathbb{R}^n}\left(f(s) + \frac{1}{2\mu}\|s-x\|^2\right) = \frac{1}{\mu}\left(x - \mathrm{Prox}_{\mu f}(x)\right) \end{aligned}$$

□

Théorème 37

Soit f convexe et soit $\mu > 0$. Les trois propriétés suivantes sont équivalentes:

1. $x^* \in \mathbb{R}^n$ est un minimiseur (global) de f
2. $x^* = \mathrm{Prox}_{\mu f}(x^*)$
3. x^* est un minimiseur (global) de $M_{\mu f}$

Démonstration. — Supposons que x^* est un minimiseur global de f. Alors, pour tout $x \in \mathbb{R}^n$, $f(x) \geq f(x^*)$ et donc $f(x) + \frac{1}{2\mu}\|x - x^*\|^2 \geq f(x^*) + \frac{1}{2\mu}\|x^* - x^*\|^2$. Ainsi, x^* minimise la fonction $x \mapsto f(x) + \frac{1}{2\mu}\|x - x^*\|^2$ et donc $x^* = \mathrm{Prox}_{\mu f}(x^*)$.

Supposons maintenant à l'inverse que $x^* = \mathrm{Prox}_{\mu f}(x^*)$. On a alors, par le Lemme 8, $0 \in \partial f(x^*) + \frac{1}{\mu}(x^* - x^*) = \partial f(x^*)$. En conséquence, x^* est un minimiseur de f, par le Lemme 6.

Enfin, pour $x \in \mathbb{R}^n$, on a $M_{\mu f}(x) \geq f(\mathrm{Prox}_{\mu f}(x))$ par définition de l'enveloppe de Moreau. Aussi, $\min_{x\in\mathbb{R}^n} M_{\mu f}(x) \geq \min_{x\in\mathbb{R}^n} f(x)$. Par ailleurs, pour $x \in \mathbb{R}^n$, $M_{\mu f}(x) = \min_{s\in\mathbb{R}^n}\left(f(s) + \frac{1}{2\mu}\|s-x\|^2\right) \leq f(x)$ (en prenant $s = x$). Ainsi, on conclut $\min_{x\in\mathbb{R}^n} M_{\mu f}(x) = \min_{x\in\mathbb{R}^n} f(x)$. Par ailleurs, soient $x_\mu^\star$ un minimiseur de $M_{\mu f}$ et $x^\star$ un minimiseur de f. On a prouvé que $f(x^\star) = M_{\mu f}(x_\mu^\star) = f(\mathrm{Prox}_{\mu f}(x_\mu^\star)) +$

$\frac{1}{2\mu}\left\|\mathrm{Prox}_{\mu f}(x_\mu^\star) - x_\mu^\star\right\|^2$. Or, par définition de $x^\star$, $f(\mathrm{Prox}_{\mu f}(x_\mu^\star)) \geq f(x^\star)$. Donc, $f(x^\star) \geq f(x^\star) + \frac{1}{2\mu}\left\|\mathrm{Prox}_{\mu f}(x_\mu^\star) - x_\mu^\star\right\|^2$ ce qui implique $\mathrm{Prox}_{\mu f}(x_\mu^\star) = x_\mu^\star$, soit le point 2 du théorème. On conclut que $x_\mu^\star$ minimiseur de f. □

***Lemme* 9**. — *Soient f convexe et $\mu > 0$. $\mathrm{Prox}_{\mu f}$ est 1-Lipschitz.*

Démonstration. — On a

$$u_1 = x_1 - \mathrm{Prox}_{\mu f}(x_1) \in \mu\partial f(Prox_{\mu f}(x_1))$$
$$u_2 = x_2 - \mathrm{Prox}_{\mu f}(x_2) \in \mu\partial f(Prox_{\mu f}(x_2))$$

Puisque μf est convexe, l'opérateur $\mu\partial f$ est monotone, soit

$$\forall(\eta_1,\eta_2) \in \mu\partial f(\mathrm{Prox}_{\mu f}(x_1)) \times \mu\partial f(\mathrm{Prox}_{\mu f}(x_2))$$
$$(\eta_1 - \eta_2)^T(\mathrm{Prox}_{\mu f}(x_1) - \mathrm{Prox}_{\mu f}(x_2)) \geq 0$$

Il s'ensuit

$$\begin{aligned}(51)\quad 0 \leq & (u_1 - u_2)^T(\mathrm{Prox}_{\mu f}(x_1) - \mathrm{Prox}_{\mu f}(x_2)) \\ = & (x_1 - x_2)^T(\mathrm{Prox}_{\mu f}(x_1) - \mathrm{Prox}_{\mu f}(x_2)) - \|\mathrm{Prox}_{\mu f}(x_1) - \mathrm{Prox}_{\mu f}(x_2)\|^2\end{aligned}$$

D'où, avec l'inégalité de Cauchy-Schwarz,

$$\|\mathrm{Prox}_{\mu f}(x_1) - \mathrm{Prox}_{\mu f}(x_2)\| \leq \|x_1 - x_2\|$$

□

Théorème 38 (Identité de Moreau)

Soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe. Alors,

$$\forall x \in \mathbb{R}^n \quad \forall \mu > 0 \quad x = \mathrm{Prox}_{\mu f}(x) + \mu\,\mathrm{Prox}_{\frac{1}{\mu}f^*}(x/\mu)$$

Démonstration. — On note $s = \mathrm{Prox}_{\mu f}(x)$ et $u = x - s$. Selon le Lemme 8, on a $0 \in \partial f(s) + \frac{1}{\mu}(s - x)$ soit $u \in \mu\partial f(s)$. Ainsi, $s = x - u \in \partial f^*(u/\mu)$ ou encore $0 \in \partial f^*(u/\mu) + \mu\left(\frac{u}{\mu} - \frac{x}{\mu}\right)$ soit $\frac{u}{\mu} = \mathrm{Prox}_{\frac{1}{\mu}f^*}(x/\mu)$. □

3.2 Conditions d'optimalité

Le Lemme 6 donne une caractérisation des minima d'une fonction convexe continue, dans le cas non-contraint : $0 \in \partial f(x^*)$. A l'aide de cette nouvelle caractérisation, il est possible de répliquer la démarche et les conditions obtenues dans le cas sous contraintes dans le présent contexte. Dans ce paragraphe, on se contente de les présenter brièvement.

On s'intéresse au problème de minimisation dans $\mathbb{R}^n$ d'une fonction convexe continue f (non-différentiable a priori), sous les contraintes $c(x) \leq 0$ définies par une fonction convexe continue $c : \mathbb{R}^n \to \mathbb{R}^m$ (non-différentiable a priori également)

$$\begin{aligned} &\min_x \quad f(x) \\ &\text{tel que } c(x) \leq 0 \end{aligned} \tag{52}$$

***Définition* 26**. — On appelle cône convexe engendré par un ensemble $E \subset \mathbb{R}^n$ l'ensemble des combinaisons linéaires à coefficients positifs ou nuls de vecteurs de E.

Noter que ce cône convexe n'a a priori pas de raison d'être fermé si E n'est pas de cardinal fini.

***Théorème* 39**. — *Soient $x^* \in \mathbb{R}^n$, $\mathcal{I}(x^*) = \{i = 1, \ldots, m \mid c_i(x^*) = 0\}$ la famille des indices des contraintes actives en x^* et $N_c'(x^*)$ le cône convexe engendré par $\cup_{i \in \mathcal{I}(x^*)} \partial c_i(x^*)$. Alors x^* est une solution du problème* (52) *si et seulement si $0 \in \partial f(x^*) + \overline{N_c'(x^*)}$ (l'adhérence de $N_c'(x^*)$).*

L'utilisation de cette caractérisation n'est, en tout généralité, pas aisée, notamment du fait de l'utilisation d'une adhérence. On peut néanmoins la simplifier dans le cas très fréquent ci-dessous.

***Définition* 27**. — On dit que les contraintes c du problème (52) satisfont la condition faible de Slater s'il existe un point de $\mathbb{R}^n$ où toutes les contraintes non-affines du problème sont strictement satisfaites, i.e.,

$$\exists\, x_0 \in \mathbb{R}^n \; c_i(x_0) \leq 0 \text{ si } c_i \text{ fonction affine et } c_i(x_0) < 0 \text{ sinon} \tag{53}$$

Théorème 40 (Conditions KKT non différentiable)

Supposons que les contraintes c satisfont la condition faible de Slater. Alors, x^* est une solution du problème (52) si et seulement si il existe $\lambda \in \mathbb{R}^m$ tel que

$$\begin{aligned} &0 \in \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial c_i(x^*) \\ &\lambda_i \geq 0 \text{ et } \lambda_i c_i(x^*) = 0\,,\ i = 1, \ldots, m \end{aligned}$$

3.3 Eléments numériques pour l'optimisation sans contrainte

Les problèmes d'optimisation non-lisse, même en l'absence de contraintes, sont en général très difficiles à résoudre numériquement. Nous nous focalisons donc sur le cas non-contraint dans ce paragraphe et évoquons les familles de méthodes dont l'efficacité a été le plus éprouvée à l'heure actuelle.

3.3.1 Méthodes de sous-gradients

Une idée naturelle d'adaptation des techniques de descente de gradient étudiées dans le cas différentiable est simplement de remplacer le gradient par un sous-gradient (obtenu par exemple par un oracle). Entre deux itérations k et $k+1$, on fait donc évoluer l'estimation x^k de l'optimum recherché par

$$x^{k+1} = x^k - l^k g^k\,, \quad g^k \in \partial f(x^k)$$

Si l'on ne prend pas de précaution supplémentaire dans l'application de cette technique, on va se heurter à trois différences essentielles :

- le sous-gradient ne peut être obtenu par différence finie numérique du coût, contrairement au cas différentiable. Prenons par exemple la fonction

$$f : x \mapsto \begin{cases} x^2 \text{ si } x \geq 0 \\ -x \text{ si } x < 0 \end{cases}$$

  Le taux d'accroissement en 0 vaut alors, pour $h > 0$, $\frac{f(h)-f(0)}{h} = \frac{h^2}{h} = h \notin [-1, 0] = \partial f(0)$;
- une direction oppposée à un sous-gradient ne fournit pas obligatoirement une direction de descente, comme dans le cas différentiable. On peut considérer la fonction $x \mapsto |x|$ pour s'en convaincre : cette fonction convexe admet un minimum en zéro, mais un sous-gradient en l'origine est pourtant 1. Il en découle qu'un algorithme basé sur le sous-gradient n'a a priori aucune raison de toujours décroître entre deux itérations ;
- il est possible que les sous-gradients successifs au cours des itérations ne tendent pas vers zéro puisque le minimum recherché peut admettre un sous-gradient non-nul. Ainsi, les techniques de gradient à pas fixe pourront ne pas converger a priori. De plus, des critères d'arrêt portant sur le sous-gradient, traduits directement de la condition $|\nabla f(x_k)| \leq \varepsilon$ dans le cas différentiable, pourront ne jamais être vérifiés.

Pour l'algorithme de gradient à pas fixe qui suit :

Algorithme 13

A partir de $x_0 \in \mathbb{R}^n$ quelconque, itérer

$$x^{k+1} = x^k - l g^k\,, \quad g_k \in \partial f(x^k)$$

on dispose néanmoins du résultat suivant.

Théorème 41

Si f est convexe de minimiseur $x^* \in \mathbb{R}^n$, et que son sous-gradient est borné au moins localement par $G > 0$, alors l'Algorithme 13 de sous-gradient à pas fixe garantit, pour tout $\varepsilon > 0$,

$$\exists k \in \mathbb{N} \quad |f(x^k) - f(x^*)| \leq \frac{lG^2}{2}(1+\varepsilon)$$

Démonstration. — On a

$$\|x^{k+1} - x^*\|^2 = \|x^k - lg^k - x^*\|^2 = \|x^k - x^*\|^2 + l^2\|g^k\|^2 - 2l(x^k - x^*)^T g^k$$

avec, par caractérisation du sous-gradient,

$$f(x^*) \geq f(x^k) + (g^k)^T(x^* - x^k)$$

D'où

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + l^2G^2 + 2l(f(x^*) - f(x^k))$$

Par itération, on obtient donc

$$\|x^{k+1} - x^*\|^2 \leq \|x^0 - x^*\|^2 + l^2G^2(k+1) + 2l\sum_{m=0}^{k}(f(x^*) - f(x^m))$$

dont on déduit

$$(k+1)\min_{m\in\{0,\dots,k\}}(f(x^m) - f(x^*)) \leq \sum_{m=0}^{k}(f(x^m) - f(x^*)) \leq \frac{\|x^0 - x^*\|^2 + l^2G^2(k+1)}{2l}$$

soit

$$\min_{m\in\{0,\dots,k\}}(f(x^m) - f(x^*)) \leq \frac{\|x^0 - x^*\|^2 + l^2G^2(k+1)}{2l(k+1)}$$

En notant que $f(x^m) - f(x^*) \geq 0$ par définition de x^* et en prenant la limite pour k tendant vers $+\infty$, on obtient le résultat voulu. □

Un choix adéquat des pas de descente (déterminés hors ligne et indépendamment de la fonction f à minimiser) permet néanmoins de rendre la méthode convergente. Notamment, si la suite positive (l^k) satisfait l'une des conditions suivantes :

C(i) $\lim_{k\to\infty} l^k = 0$ et $\sum_{k=0}^{\infty} l^k = +\infty$
C(ii) $\sum_{k=0}^{\infty} l^k = +\infty$ et $\sum_{k=0}^{\infty}(l^k)^2 < +\infty$

on dispose du résultat suivant associé à l'algorithme

***Algorithme* 14.** — A partir de $x_0 \in \mathbb{R}^n$ quelconque et de $f^0 = f(x_0)$, itérer

$$x^{k+1} = x^k - l^k g^k\,, \quad g_k \in \partial f(x^k)$$
$$f^{k+1} = \min\left\{f^k, f(x^{k+1})\right\}$$

Théorème 42

Si f est convexe, que son sous-gradient est borné au moins localement par $G > 0$ et que la suite (l^k) satisfait la condition C(i) ou C(ii) ci-dessus, alors l'Algorithme 14 converge vers un minimiseur de f.

Démonstration. — En reprenant la preuve précédente, on conclut dans le cas d'un pas variable l^k que

$$\min_{m\in\{0,\dots,k\}} (f(x^m) - f(x^*)) \leq \frac{\|x^0 - x^*\|^2 + G^2 \sum_{m=0}^{k} (l^m)^2}{2\sum_{m=0}^{k} l^m} \tag{54}$$

Dans le cas de la propriété (ii), le résultat de convergence voulu se déduit donc d'un simple passage à la limite.

Supposons maintenant que c'est la propriété (i) qui est satisfaite. Dans ce cas, pour $\varepsilon > 0$, il existe $N_1 \in \mathbb{N}$ et $N_2 \in \mathbb{N}$ tels que $l^k \leq \frac{\varepsilon}{G^2}$ pour $k \geq N_1$ et

$$\sum_{m=0}^{k} l^m \geq \frac{\|x^0 - x^*\|^2 + G^2 \sum_{m=0}^{N_1} (l^m)^2}{\varepsilon}, \quad k \geq N_2$$

Ainsi, pour $k \geq \max\{N_1, N_2\} + 1$, on a

$$\begin{aligned}
\frac{\|x^0 - x^*\|^2 + G^2 \sum_{m=0}^{k} (l^m)^2}{2\sum_{m=0}^{k} l^m} &= \frac{\|x^0 - x^*\|^2 + G^2 \sum_{m=0}^{N_1} (l^m)^2}{2\sum_{m=0}^{k} l^m} + \frac{G^2 \sum_{N_1+1}^{k} (l^m)^2}{2\sum_{m=0}^{k} l^m} \\
&\leq \frac{\|x^0 - x^*\|^2 + G^2 \sum_{m=0}^{N_1} (l^m)^2}{\frac{2}{\varepsilon}(\|x^0 - x^*\|^2 + G^2 \sum_{m=0}^{N_1} (l^m)^2)} + \frac{G^2 \sum_{N_1+1}^{k} \frac{\varepsilon}{G^2} l^m}{2\sum_{m=0}^{k} l^m} \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

ce qui garantit que la borne supérieure de (54) tend vers zéro et le résultat voulu. □

La condition $\sum_{k=0}^{\infty} l^k = +\infty$ indique que le pas ne doit pas converger trop rapidement vers zéro, sous peine de ne pas avoir recueilli assez d'informations sur la fonction f. Par exemple, une décroissance en $O(1/k^2)$ est trop rapide, puisque la série $\sum \frac{1}{k^2}$ converge tandis qu'une décroissance en $O(1/k)$ convient.

Bien qu'assez simples à mettre en oeuvre, ces algorithmes souffrent des inconvénients énoncés plus hauts qui engendrent souvent de mauvaises performances numériques et des phénomènes oscillatoires. Par ailleurs, leur vitesse de convergence est en général assez faible (sous-linéaire).

3.3.2 Minimisation proximale

Le théorème 36 indique que

$$\mathrm{Prox}_{\mu f}(x) = x - \mu \nabla M_{\mu f}(x) \tag{55}$$

c.-à-d. en d'autres termes que l'opérateur proximal peut être interprété comme une descente de gradient, avec un pas fixe μ, appliqué à l'enveloppe de Moreau $M_{\mu f}$, qui

n'est rien d'autre qu'une régularisation de f et qui possède les mêmes minimiseurs, selon le Théorème 37.

Ceci amène donc naturellement à considérer l'algorithme suivant.

Algorithme 15

A partir de $x^0 \in \mathbb{R}^n$ quelconque, itérer

$$x^{k+1} = \text{Prox}_{l^k f}(x^k)$$

avec (l_k) choisie hors ligne, par exemple fixe ou satisfaisant les propriétés C(i) ou C(ii).

Une autre interprétation de cet algorithme découle également du Théorème 37, le minimiseur recherché étant un point fixe de l'opérateur proximal. On peut montrer que cet algorithme d'itération du point fixe converge, bien que l'opérateur proximal ne soit pas contractant (voir Lemme 9), mais seulement fortement non-expansif c.-à-d. vérifiant l'inégalité (51).

Chaque itération de cette méthode nécessite de disposer d'une évaluation de l'opérateur proximal. En toute généralité, ceci implique donc de résoudre le problème d'optimisation (50). Ceci a une utilité quand il est difficile de minimiser la fonction f, mais plus simple de minimiser la somme de f et d'une fonction quadratique, une situation peu courante dans les faits. En pratique, on dispose néanmoins de nombreux cas où l'opérateur proximal peut être évalué analytiquement ou de méthodes numériques dédiées (notamment dans le cas scalaire, ou l'on applique une forme de dichotomie, appelée méthode de localisation).

3.3.3 Méthodes de gradient proximales

On considère dans cette section le problème suivant

$$\min_{x \in \mathbb{R}^n} [f(x) = g(x) + h(x)]$$

où $g : \mathbb{R}^n \to \mathbb{R}$ et $h : \mathbb{R}^n \to \mathbb{R}$ sont des fonctions convexes et où g est différentiable. Pour un problème et une fonction f donnés, il est toujours possible d'adopter ce formalisme, mais cette décomposition n'est bien sûr pas unique, amenant à des implémentations plus ou moins performantes.

Une caractérisation d'un minimiseur x^* est alors

$$\begin{aligned} 0 \in \partial(g+h)(x^*) = \nabla g(x^*) + \partial h(x^*) &\Leftrightarrow 0 \in l\nabla g(x^*) + l\partial h(x^*)\,, \quad l \in \mathbb{R}^* \\ &\Leftrightarrow x^* - l\nabla g(x^*) \in x^* + l\partial h(x^*)\,, \quad l \in \mathbb{R}^* \\ &\Leftrightarrow (I + l\partial h)^{-1}(x^* - l\nabla g(x^*)) = x^*\,, \quad l \in \mathbb{R}^* \\ &\Leftrightarrow \text{Prox}_{lh}(x^* - l\nabla g(x^*)) = x^*\,, \quad l \in \mathbb{R}^* \end{aligned}$$

où la dernière équivalence provient du Lemme 8. Un minimiseur x^* est ainsi le point fixe d'un opérateur proximal, ce qui conduit à l'algorithme suivant.

Algorithme 16

A partir de $x^0 \in \mathbb{R}^n$ quelconque, itérer

$$x^{k+1} = \text{Prox}_{lh}(x^k - l\nabla g(x_k))$$

Théorème 43

Soit $g : \mathbb{R}^n \to \mathbb{R}$ de gradient L-Lipschitzien. Pour $l \leq 1/L$, l'Algorithme 16 de gradient proximal à pas fixe assure que, pour x^* minimiseur de f,

$$f(x^k) - f(x^*) \leq \frac{1}{2lk}\|x_0 - x^*\|^2, \quad k \geq 1$$

Démonstration. — Nous commençons cette preuve en reformulant les itérations sous la forme

$$x^{k+1} = x^k + lp_l(x^k)$$

avec

$$p_l(x) = \frac{1}{l}\left[\text{Prox}_{lh}(x - l\nabla g(x)) - x\right]$$

Notons $s = \text{Prox}_{lh}(x - l\nabla g(x)) = x + lp_l(x)$. Par le lemme 8, on a

$$0 \in \partial h(s) + \frac{1}{l}(s - x + l\nabla g(x)) = \partial h(x + lp_l(x)) + p_l(x) + \nabla g(x)$$

$$\Leftrightarrow -(p_l(x) + \nabla g(x)) \in \partial h(x + lp_l(x)) \tag{56}$$

Par ailleurs, on note que $p_l(x) = 0$ si et seulement si $0 \in \nabla g(x) + \partial h(x) = \partial f(x)$ c.-à-d. x minimiseur de f.

On montre maintenant que, pour $l \in]0, 1/L]$ et $(x, z) \in \mathbb{R}^n \times \mathbb{R}^n$, on a

$$f(x + lp_l(x)) \leq f(z) - p_l(x)^T(x - z) - \frac{l}{2}\|p_l(x)\|^2 \tag{57}$$

En effet, le gradient de g étant L-Lipschitz,

$$\begin{aligned} g(x + lp_l(x)) &= g(x) + \int_0^1 \nabla g(x + slp_l(x))^T lp_l(x)ds \\ &= g(x) + \nabla g(x)^T lp_l(x) + \int_0^1 \left[\nabla g(x + slp_l(x)) - \nabla g(x)\right]^T lp_l(x)ds \\ &\leq g(x) + \nabla g(x)^T lp_l(x) + \frac{Ll^2}{2}\|p_l(x)\|^2 \\ &\leq g(z) + \nabla g(x)^T(x - z + lp_l(x)) + \frac{l}{2}\|p_l(x)\|^2 \qquad (58) \end{aligned}$$

où la dernière inégalité découle de la convexité de g. Par ailleurs, de (56), on a également

$$h(z) \geq h(x + lp_l(x)) - (p_l(x) + \nabla g(x))^T (z - x - lp_l(x)) \tag{59}$$

En regroupant (58) et (59), il s'ensuit

$$f(x + lp_l(x)) \leq f(z) + \frac{l}{2}\|p_l(x)\|^2 + p_l(x)^T(z - x - lp_l(x))$$

c.-à-d. (57). En appliquant (57) à $x = z = x^k$, on obtient

$$f(x^{k+1}) \leq f(x^k) - \frac{l}{2}\|p_l(x^k)\|$$

c.-à-d. que la suite $(f(x^k))$ est strictement décroissante tant que $x^k \neq x^*$ puisque $p_l(x) = 0$ ssi x minimiseur de f. Par ailleurs, en appliquant (57) à $x = x^k$ et $z = x^*$, on obtient

$$\begin{aligned} 0 \leq f(x^{k+1}) - f(x^*) &\leq -p_l(x^k)^T(x^k - x^*) - \frac{l}{2}\|p_l(x^k)\|^2 \\ &\leq \frac{1}{2l}\left(\|x^k - x^*\|^2 - \|x^k - x^* + lp_l(x^k)\|^2\right) \\ &\leq \frac{1}{2l}\left(\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2\right) \end{aligned}$$

La distance au minimiseur x^* décroît donc entre deux itérations et, puisque $(f(x^k))$ est décroissante, en sommant, on en déduit

$$\begin{aligned} f(x^{k+1}) - f(x^*) &\leq \frac{1}{k+1}\sum_{i=1}^{k+1}(f(x^i) - f(x^*)) \\ &\leq \frac{1}{2l(k+1)}\left(\|x^0 - x^*\|^2 - \|x^{k+1} - x^*\|^2\right) \end{aligned}$$

□

De la preuve précédente, on observe que l'Algorithme 16 de gradient proximal est une méthode de descente, contrairement aux méthodes de sous-gradient précédemment évoquées.

On remarque que les performances de cet algorithme sont conditionnées par la connaissance que l'on a de la constante de Lipschitz du gradient. Dans le cas où cette dernière est mauvaise, on lui préfère le choix d'un pas par la recherche linéaire suivante.

Algorithme 17

A partir de $x_0 \in \mathbb{R}^n$ et de $l^0 > 0$ quelconques, calculer $g^0 = g(x^0)$ et $r^0 = \nabla g(x^0)$. Itérer, pour $\eta \in]0,1[$,

$l^k = l^{k-1}$

repeat

$z = \mathrm{Prox}_{l^k h}(x^k - l^k r^k)$

$\hat{g}_{l^k}(z, x^k) = g^k + (r^k)^T(z - x^k) + \dfrac{1}{2l^k}\|z - x^k\|^2$

break if $g(z) \leq \hat{g}_{l^k}(z, x^k)$

$l^k = \eta l^k$

$x^{k+1} = z$

$g^{k+1} = g(z),\ r^{k+1} = \nabla g(z)$

En effet, si $l^k \in]0, 1/L]$, alors la fonction $\hat{g}_{l^k}(\cdot, x^k)$ est une borne supérieure de la fonction g qui est de gradient L-Lipschitzien. Ainsi, la recherche linéaire ci-dessus fournit toujours un pas candidat.

Une preuve similaire à celle de l'Algorithme à pas fixe garantit le résultat de convergence suivant.

Théorème 44

Soit $g : \mathbb{R}^n \to \mathbb{R}$ de gradient L-Lipschitzien. L'Algorithme 17 de gradient proximal avec recherche linéaire assure que, pour x^* minimiseur de f,

$$f(x^k) - f(x^*) \leq \frac{\eta L}{2k}\|x_0 - x^*\|^2$$

3.3.4 Méthodes de faisceaux

Les méthodes précédemment détaillées ont toutes en commun de travailler directement sur la fonction non-différentiable.

Un principe alternatif est le suivant : au lieu de travailler directement avec la fonction objectif f, on se sert des informations collectées au cours des itérations pour construire un modèle (convexe) de f, plus facile à minimiser. Etant donné un faisceau d'informations $\{(x_i, f(x_i), g_i) \mid g_i \in \partial f(x_i)\,,\ i = 1, \ldots, k\}$ obtenu après k itérations, on construit une approximation linéaire par morceaux de la fonction f

$$\forall y \in \mathbb{R}^n \quad \varphi_k(y) = \max_{i=1,\ldots,k} \left\{ f(x_i) + g_i^T(y - x_i) \right\} \tag{60}$$

Par construction, on a donc $\varphi_k(y) \leq f(y)$, $y \in \mathbb{R}^n$ et, à chaque itération, le modèle courant est enrichi d'un nouveau plan sécant. On itère la procédure jusqu'à juger que la précision fournie par le minimiseur candidat est suffisante. A noter que le problème de minimisation de la fonction (60) est toujours un problème d'optimisation non-différentiable, mais peut être réécrit comme un problème de programmation linéaire (LP) différentiable. Néanmoins, ces algorithmes ne sont pas des méthodes de descente et requièrent de stocker au cours du temps le faisceau d'informations.

CHAPITRE 4

EXERCICES TD1

Exercice 1.1 Fonction de Rosenbrock

Calculer le gradient $\nabla f(x)$ et le Hessien $\nabla^2 f(x)$ de la fonction de Rosenbrock

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Montrer que $x^* = (1, 1)^T$ est l'unique minimiseur global de cette fonction.

Exercice 1.2 Gradient et Hessien

Soit a un vecteur de $\mathbb{R}^n$ et A une matrice $n \times n$ symétrique. Calculer le gradient et le Hessien de $f_1(x) = a^T x$ et $f_2(x) = x^T Ax$.

Exercice 1.3 Moindres carrés

On admet le résultat suivant : si A matrice $p \times n$ est de rang n alors $A^T A$ est inversible. On cherche à résoudre le problème dit des "moindres carrés"

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|$$

où A est une matrice $p \times n$, $b \in \mathbb{R}^p$ et en général $p \geq n$.

1. Montrer que ce problème est convexe
2. On suppose que la norme considérée est la norme euclidienne $\|x\|^2 = x^T x$, calculer le gradient de la fonction à minimiser.
3. Quel est alors un point candidat pour être minimum ? Conclure.
4. Reprendre l'expression du minimum lorsque le problème est "pondéré" c'est à dire $\|x\| = x^T Qx$ où Q est une matrice symétrique définie positive.

Exercice 1.4 Sous-gradient et différence finie

Soit

$$f : x \in \mathbb{R} \mapsto \begin{cases} (x-1)^2 \text{ si } x \geq 1 \\ -x+1 \text{ si } x < 1 \end{cases} \tag{61}$$

Montrer que f est convexe et exprimer $\partial f(x)$. Montrer que, pour $h > 0$ petit,

$$\frac{f(1+h) - f(1)}{h} \notin \partial f(1) \tag{62}$$

CHAPITRE 5

EXERCICES TD2

Exercice 2.1 Conditions de Wolfe

Montrer en utilisant la fonction $f(x) = x^2$ et $x_0 = 1$ que si $0 < c_2 < c_1 < 1$ il peut ne pas exister de pas satisfaisant les conditions de Wolfe.

Exercice 2.2 Descente de gradient stochastique

Soit $f : \mathbb{R}^n \to \mathbb{R}$ $(n > 2)$ convexe de classe $\mathcal{C}^2$ avec $\left|\frac{\partial^2 f}{\partial x_i^2}(x)\right| \leq L_i$ pour tout $x \in \mathbb{R}^n$ et $i = 1, \ldots, n$. On introduit l'opérateur de descente

$$\left(T_l^i\right)_j(x) = \begin{cases} x_j - l\dfrac{\partial f}{\partial x_i}(x) \text{ si } j = i \\ x_j \text{ sinon} \end{cases} \tag{63}$$

1. Montrer que $f(T_l^i(x)) \leq f(x) - \frac{l}{2}(2 - lL_i)\left|\frac{\partial f}{\partial x_i}(x)\right|^2, \quad x \in \mathbb{R}^n$

On considère maintenant l'algorithme "stochastique" suivant, utilisant une loi de probabilité discrète (p_i) :

Algorithme. — A partir de $x^0 \in \mathbb{R}^n$ et $\theta \in]0, 2[$, itérer
- choisir avec une probabilité p_i l'indice i $(i \in \{1, \ldots, n\})$
- $x^{k+1} = T_{\theta/L_i}(x^k)$

2. Soient $\|x\|_M^2 = x^T M x$ $(x \in \mathbb{R}^n)$, M la matrice diagonale telle que $m_{ii} = p_i/L_i$ $(i = 1, \ldots, n)$ et $\mathbb{E}_k(f(x^{k+1}))$ est l'espérance de $f(x^{k+1})$ sachant x_k. Montrer que
$$\mathbb{E}_k(f(x^{k+1})) \leq f(x^k) - \frac{\theta(2-\theta)}{2}\|\nabla f(x^k)\|_M^2$$
3. Montrer que

$$\forall (x_1, x_2) \in \mathbb{R}^n \quad f(x_1) - f(x_2) \leq \|\nabla f(x_1)\|_M \|x_1 - x_2\|_{M^{-1}} \tag{64}$$

En déduire, pour tout $x \in \mathbb{R}^n$,

$$E_k(f(x^{k+1}) - f(x)) \leq f(x^k) - f(x) - \frac{\theta(2-\theta)}{2\|x^k - x\|^2_{M^{-1}}}(f(x^k) - f(x))^2 \tag{65}$$

4. Pour un certain $x \in \mathbb{R}^n$, suppose que l'on sait trouver une constante $C > 0$ telle que $\|x^k - x\|_{M^{-1}} \leq C$ pour tout k (et pour tous les tirages aléatoires). Montrer que

$$\Delta_{k+1}(x) \leq \Delta_k(x) - \frac{\theta(2-\theta)}{2C^2}\Delta_k(x)^2 \tag{66}$$

où $\Delta_k(x) = E(f(x^k) - f(x))$ est l'espérance de $f(x^k) - f(x)$. (On rappelle que, pour une variable aléatoire X, $E(X^2) \geq E(X)^2$.)

5. Soit $x^\star$ minimiseur de f. Justifier que $\Delta_k(x^\star) \geq 0$. En déduire

$$\Delta_k(x^\star) \leq \frac{2C^2}{\theta(2-\theta)}\frac{1}{k+1} \tag{67}$$

et conclure sur la convergence de l'algorithme.

6. Quel avantage présente cet algorithme par rapport à un algorithme déterministe de descente du gradient ? Y a-t-il un meilleur choix de la densité de probabilité $(p_i)_{i=1,\dots n}$ qu'un autre ?

Exercice 2.3 Méthode de Newton avec modification du Hessien

Loin du minimum, la matrice hessienne $\nabla^2 f(x)$ peut ne pas être définie positive, de sorte que son inverse peut ne pas être bien défini et (a fortiori) utilisé dans un algorithme de descente. Une approche pour palier ce problème consiste à ajouter un multiple de l'identité au hessien.

1. Soit A une matrice symétrique. Justifier que $A + \Delta A$ est symétrique définie positive de valeurs propres $\lambda_i \geq \delta > 0$ $(i = 1, \dots, n)$ si

$$\Delta A = \tau I \quad \text{avec} \quad \tau = \max\{0, \delta - \lambda_{\min}(A)\} \tag{68}$$

2. Commenter l'algorithme suivant (performances, avantages, inconvénients). Comment peut-il être utilisé dans le cadre d'une méthode de Newton ?

Algorithme. — Choisir $\beta > 0$.
Si $\min_i a_{ii} > 0$
$\tau_0 = 0$;
sinon
$\tau_0 = \beta - \min_i a_{ii}$
end
Itérer:
Si $A + \tau_k I$ définie positive
stop
sinon

$\tau_{k+1} = \max\{2\tau_k, \beta\}$
end

CHAPITRE 6

EXERCICES TD3

Exercice 3.1 Minima liés

Soit f la distance focale d'une lentille convergente. A quelle distance x de cette lentille faut-il placer un objet sur l'axe focal pour que son image (réelle) x' par la lentille soit la plus proche possible de lui ?

Exercice 3.2 KKT

En utilisant les conditions de Karush-Kuhn-Tucker, et en raisonnant graphiquement, trouver en fonction de la valeur de a le minimum de

$$J(x, y) = ay + \frac{1}{x}$$

sous les contraintes

$$x \geq 1/2\,, \quad 0 \leq y \leq x\,, \quad y \leq -x + 2$$

Exercice 3.3 Tente de cirque

Ce problème est issu de l'Optimization Toolbox de Matlab. On considère une tente de cirque comme une membrane élastique posée sur 5 mâts (le mât central étant plus haut que les autres). La forme prise par cette toile de tente minimise une énergie potentielle composée d'un terme de gravité auquel s'oppose un terme d'élasticité.

Après avoir utilisé un schéma aux différences pour le gradient de la surface, on obtient un problème de programmation quadratique dont l'inconnue est l'altitude des différents points de la toile. Si on représente la surface de la toile par une grille de points, on désigne par x le vecteur constitué par les lignes de cette grilles mises bout à bout. Ainsi une grille de taille 30×30 donne un vecteur de taille 900. Les contraintes expriment que la toile repose sur les mâts.

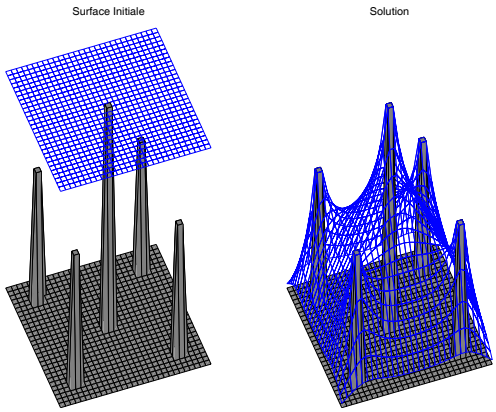
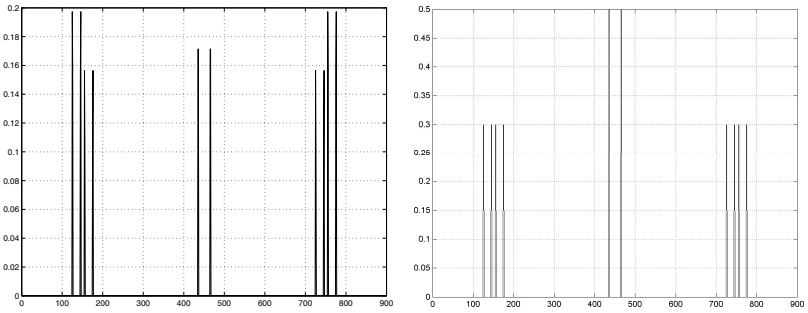


FIGURE 1. Forme prise par une tente de cirque reposant sur cinq mâts.



(a) Multiplicateurs de lagrange. (b) Contraintes.

FIGURE 2. Valeurs des multiplicateurs et contraintes à l'optimum.

Les multiplicateurs de Lagrange et les contraintes correspondantes sont représentés sur la Figure 6.

Avec les paramètres choisis, on obtient un minimum de 0.44422, de combien faut-il modifier la hauteur du mât central pour faire décroître cette valeur à 0.43767?

Exercice 3.4 Plus courte distance à un hyperplan

La plus courte distance entre un point $x_0 \in \mathbb{R}^n$ et l'intersection de m hyperplans $\{x, Ax = b\}$ où les lignes de $A \in \mathcal{M}_{m,n}$ sont linéairement indépendantes, peut s'écrire comme un problème de programmation quadratique

$$\begin{aligned} &\min_x \quad (x - x_0)^T (x - x_0) \\ &\text{tel que } Ax = b \end{aligned} \tag{69}$$

Montrer que le multiplicateur à l'optimum et la solution sont

$$\lambda^* = -(AA^T)^{-1}(b - Ax_0)$$
$$x^* = x_0 + A^T(AA^T)^{-1}(b - Ax_0)$$

Montrer que dans le cas où A est un vecteur ligne, la plus petite distance entre x_0 et l'hyperplan vaut $\frac{|b-Ax_0|}{\|A\|}$ (où $\|\cdot\|$ est la norme 2 dans $\mathbb{R}^n$).

CHAPITRE 7

EXERCICES TD4

Exercice 4.1 Pénalités intérieures

Considérer le problème pénalisé à barrière logarithmique

$$\begin{array}{ll} \min\limits_{x} & x - \mu \log x - \mu \log(1-x) \\ \text{tel que} & x \in]0,1[\end{array} \tag{70}$$

provenant du problème

$$\begin{array}{ll} \min\limits_{x} & x \\ \text{tel que} & x \geq 0, \\ & x \leq 1 \end{array} \tag{71}$$

Résoudre (70). Vérifier analytiquement que lorsque μ tend vers 0, l'optimum $x^*(\mu)$ obtenu converge vers la solution de (71).

Exercice 4.2 Dual d'une programmation quadratique

Considérer le problème de programmation quadratique

$$\begin{array}{ll} \max\limits_{x} & \frac{1}{2} x^T G x + x^T d \\ \text{tel que} & Ax \geq b \end{array}$$

où G est symétrique définie positive. Montrer que le problème dual est

$$\begin{array}{ll} \max\limits_{x,\,\lambda} & \frac{1}{2} x^T G x + x^T d - \lambda^T (Ax - b) \\ \text{tel que} & Gx + d - A^T \lambda = 0, \\ & \lambda \geq 0 \end{array}$$

et peut se simplifier en

$$\begin{array}{ll} \max\limits_{\lambda} & -\frac{1}{2}\lambda^T(AG^{-1}A^T)\lambda + \lambda^T(b + AG^{-1}d) - \frac{1}{2}d^TG^{-1} \\ \text{tel que} & \lambda \geq 0 \end{array}$$

Exercice 4.3 **Contrainte actives pour Programmation Linéaire : algorithme du simplexe**

1. On considère le problème de Programmation Linéaire

$$\begin{array}{ll} \max\limits_{x \in \mathbb{R}^2} & x_1 + 2x_2 \\ \text{tel que} & x_1 \geq 0, \\ & 0 \leq x_2 \leq 1, \\ & 2x_1 + x_2 \leq 2 \end{array}$$

   Calculer les deux premières itérations de l'algorithme du simplexe, pour la condition initiale $x_0 = (0, 0)$. Montrer que l'on aboutit à un minimum global.

2. Considérons les problèmes de Programmation Linéaire

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & c^Tx \\ \text{tel que} & Ax = b, \\ & x \geq 0 \end{array} \tag{72}$$

   avec $b \in \mathbb{R}^m_+$ (sans nuire à la généralité) et

$$\begin{array}{ll} \min\limits_{(x,y) \in \mathbb{R}^{n+m}} & \sum\limits_{i=1}^{m} y_i \\ \text{tel que} & Ax + y = b, \\ & x \geq 0, \\ & y \geq 0 \end{array} \tag{73}$$

   (a) Quelle est une solution faisable triviale de (73) ? Comment peut-on alors obtenir un minimiseur de (73) ?
   (b) A partir de ce minimiseur, proposer une procédure permettant de déterminer si le problème (72) admet des solutions et, le cas échéant, d'initialiser l'algorithme du simplexe.
   (c) Appliquer cette procédure au problème de la question précédente.

CHAPITRE 8

EXERCICES TD5

Exercice 5.1 Inégalité de Young

Soit $f : \mathbb{R}_+ \to \mathbb{R}_+$ continue strictement croissante telle que $f(0) = 0$, d'inverse g. Soient

$$F(x) = \int_0^x f(a)da\,, \quad G(y) = \int_0^y g(a)da \tag{74}$$

Montrer que G est la transformée de Fenchel de F et que

$$\forall (x, y) \in \mathbb{R}_+^2 \quad xy \leq F(x) + G(y) \tag{75}$$

En déduire que $uv \leq u \ln u + e^{v-1}\,, \quad u, v \geq 1$

Exercice 5.2 Conjuguée et opérateur proximal d'une norme

On se place dans $\mathbb{R}^n$, dont on note $\|\cdot\|$ une norme. On considère également la norme duale

$$\|y\|_D = \sup_{\|x\| \leq 1} |y^T x| \tag{76}$$

et $\mathcal{B}_D$ la boule unité fermée pour la norme duale.

1. Montrer que

$$\|y\|^* = \begin{cases} 0 & \text{si } \|y\|_D \leq 1 \\ +\infty & \text{sinon} \end{cases} \tag{77}$$

2. Soit $\Pi_{\mathcal{B}_D}$ le projecteur sur la boule unité duale. Montrer que $\text{Prox}_{\|\cdot\|^*}(x) = \Pi_{\mathcal{B}_D}(x)$.
3. En déduire une expression de $\text{Prox}_{\|\cdot\|}$.

4. Montrer que, dans le cas de la norme ℓ_1,

$$(\mathrm{Prox}_{\mu\|\cdot\|_1}(x))_i = \begin{cases} x_i - \mu & \text{si } x_i \geq \mu \\ 0 & \text{si } |x_i| \leq \mu \\ x_i + \mu & \text{si } x_i \leq -\mu \end{cases} \tag{78}$$

Exercice 5.3 Problème de Lasso

Soit le problème de minimisation

$$\min_{x \in \mathbb{R}^n} \quad \left(\|Ax - b\|_2^2 + \gamma \|x\|_1\right)$$

avec A matrice $p \times n$, $b \in \mathbb{R}^p$ $(p > n)$ et $\gamma > 0$.

1. Montrer qu'une solution x^* du problème doit vérifier, pour tout $l \geq 0$,

$$x^* - 2l(A^T A x^* - A^T b) \in x^* + l\gamma \partial \|\cdot\|_1(x^*) \tag{79}$$

2. En déduire

$$x^* = \mathrm{Prox}_{l\gamma\|\cdot\|_1}\left(x^* - 2l(A^T A x^* - A^T b)\right) \tag{80}$$

Conclure sur l'effet du terme $\gamma\|\cdot\|_1$ dans le problème de minimisation.

CHAPITRE 9

CORRIGÉS[1]

9.1 Exercices du TD1

Exercice 1.1 **Fonction de Rosenbrock**

La fonction f est de classe $\mathcal{C}^2$, son gradient s'écrit

$$\nabla f(x) := \left(\frac{\partial f}{\partial x_1}(x) \quad \frac{\partial f}{\partial x_2}(x)\right) = \left(-400x_1(x_2 - x_1^2) - 2(1 - x_1) \quad 200(x_2 - x_1^2)\right),$$

et sa hessienne

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 x_2}(x) \\ \frac{\partial^2 f}{\partial x_2 x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) \end{pmatrix} = \begin{pmatrix} -400(x_2 - x_1^2) + 800x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}.$$

Une condition nécessaire pour que x^* soit un minimum local est $\nabla f(x^*) = 0$, i.e. $x^* = (1, 1)^T$. De plus, la hessienne en x^*

$$\nabla^2 f(x^*) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}$$

vérifie $\det \nabla^2 f(x^*) = 400 > 0$ et $\operatorname{tr} \nabla^2 f(x^*) = 1002 > 0$. La hessienne est donc définie positive. Donc $(1, 1)^T$ est l'unique minimum local de f. Par ailleurs, $f(x)$ est strictement positif pour $x \neq x^*$ et $f(x^*) = 0$, donc x^* est l'unique minimum global de f.

(1) Ces corrigés sont le fruit de la collaboration de Pierre-Cyril Aubin, Delphine Bresch-Pietri, Jean-Emmanuel Deschaud, Hubert Ménou, François Pacaud et Dilshad Surroop.

Exercice 1.2 Gradient et Hessien

On calcule $f(x+h) = f(x) + \nabla(f)(x)^T h + \frac{1}{2} h^T \nabla^2 f(x) h + o(\|h\|^2)$ et on identifie.

$$
\begin{aligned}
f_1(x+h) &= a^T(x+h) = a^T x + a^T h, \\
\nabla f_1(x) &= a,\ \nabla^2 f_1(x) = 0 \\
f_2(x+h) &= (x+h)^T A(x+h) = x^T Ax + h^T Ax + x^T Ah + h^T Ah, \\
\nabla f_2(x) &= 2Ax,\ \nabla^2 f_2(x) = 2A
\end{aligned}
$$

Exercice 1.3 Moindres carrés

1. La fonction $t \mapsto \|Ax - b\|$ est convexe puisque pour $x, y \in \mathbb{R}^n$ et $t \in [0,1]$,
$$
\begin{aligned}
\|A(ty + (1-t)x) - b\| &= \|t(Ay - b) + (1-t)(Ax - b)\| \\
&\leq t\|Ay - b\| + (1-t)\|Ax - b\|,
\end{aligned}
$$
par inégalité triangulaire.
2. Minimiser $x \mapsto \|Ax - b\|$ a pour problème équivalent (au sens d'avoir les même points optimaux) la minimisation de $\varphi : x \mapsto \|Ax - b\|^2$, qui s'écrit
$$
\begin{aligned}
\|Ax - b\|^2 &= (Ax - b)^T (Ax - b) = x^T A^T Ax - x^T A^T b - b^T Ax + b^T b \\
&= x^T A^T Ax - 2b^T Ax + b^T b.
\end{aligned}
$$
D'après l'exercice *Gradient et Hessien* avec $A^T A$ symétrique, $\nabla\varphi(x) = 2A^T Ax - 2A^T b$.
3. Un point candidat x^* vérifie $\nabla\varphi(x^*) = 0$, soit $x^* = (A^T A)^{-1} A^T b$. De plus, $\nabla^2 f(x^*) = 2A^T A$, qui est définie positive, puisque $A^T A$ est inversible, et que pour $x \neq 0$, $x^T A^T Ax = \|Ax\|^2 > 0$. Le problème étant convexe, x^* est l'unique minimum global de φ.
4. Soit $\varphi_Q : x \mapsto \|Ax - b\|_Q^2$. Un calcul similaire donne
$$
\begin{aligned}
\varphi_Q(x) &= x^T A^T QAx - 2b^T QAx - b^T Qb, \\
\nabla\varphi_Q(x) &= 2A^T QAx - 2A^T Qb.
\end{aligned}
$$
La matrice $A^T QA$ étant (symétrique) définie positive, on obtient le point candidat x^* qui satisfait $x^* = (A^T QA)^{-1} A^T Qb$. En ce point la hessienne vaut $\nabla^2 f(x^*) = 2A^T QA > 0$; x^* est donc l'unique minimum global de φ_Q.

   Remarque : $(A^T A)^{-1} A^T$ est un inverse à gauche de A (au sens où appliqué à gauche de A on retrouve Id_n).

Exercice 1.4 Sous-gradient et différence finie

On remarque que la fonction f est différentiable sur $\mathbb{R} \setminus \{1\}$ et que $f^+ : x \in [1, +\infty[\mapsto (x-1)^2$ et $f^- : x \in]-\infty, 1[\mapsto -x+1$ sont convexes car dérivables et de dérivées croissantes.

On montre la convexité de f par l'épigraphe. Soient

$$\mathrm{Epi}(f^+) = \left\{(x,y) \mid x \geq 1\,,\ y \geq (x-1)^2\right\},$$
$$\mathrm{Epi}(f^-) = \left\{(x,y) \mid x < 1\,,\ y \geq -x+1\right\}.$$

On a $\mathrm{Epi}(f) = \mathrm{Epi}(f^+) \cup \mathrm{Epi}(f^-)$ et $\mathrm{Epi}(f^+), \mathrm{Epi}(f^-)$ convexes par caractérisation des fonctions convexes différentiables. Il n'y a donc qu'à considérer $A = (x^-, y^-) \in \mathrm{Epi}(f^-)$ et $B = (x^+, y^+) \in \mathrm{Epi}(f^+)$. $(x^+ - 1)(x^- - 1) \leq 0$, aussi $[A, B]$ intersecte l'axe vertical $\{1\} \times \mathbb{R}$ en un point C et

$$[A, C] \subset \mathrm{Epi}(f^-) \text{ car } \mathrm{Epi}(f^-) \text{ convexe},$$
$$[B, C[\subset \mathrm{Epi}(f^+) \text{ car } \mathrm{Epi}(f^+) \text{ convexe}.$$

Donc

$$[A, B] \in \mathrm{Epi}(f^+) \cup \mathrm{Epi}(f^-) = \mathrm{Epi}(f)$$

et f convexe.

Alternativement, on peut considérer $h^- : x \in \mathbb{R} \mapsto -x+1$ et $h^+ : x \in \mathbb{R} \mapsto (x-1)^2 1_{[1,\infty[}$ (on prolonge continûment par 0 sur $]-\infty, 1]$). Il s'agit de deux fonctions convexes dérivables et telles que $f = \max\{h^-, h^+\}$. Aussi, $\mathrm{Epi}(f) = \mathrm{Epi}(h^-) \cap \mathrm{Epi}(h^+)$ est convexe comme intersection d'ensembles convexes.

f étant dérivable sur $\mathbb{R} \setminus \{1\}$, on a

$$\partial f(x) = \begin{cases} -1\,, & x < 1 \\ 2(x-1) & x > 1 \end{cases} \tag{81}$$

Et, en 1, on a, pour $y < 1$

$$f(y) = -y + 1 \geq f(1) + v(y-1) = v(y-1) \Leftrightarrow v \in [-1, 0]$$

et pour $y \geq 1$

$$f(y) = (y-1)^2 \geq f(1) + v(y-1) \Leftrightarrow y - 1 \geq v \Rightarrow v \leq 0$$

On en déduit $\partial f(1) = [-1, 0]$. Pour $h > 0$,

$$\frac{f(1+h) - f(1)}{h} = \frac{h^2}{h} = h \notin \partial f(1) = [-1, 0]$$

9.2 Exercices du TD2

Exercice 2.1 Conditions de Wolfe

On voit sur la Figure 1 dessin ci-dessus que, en choisissant c_2 suffisamment proche de 0 et c_1 suffisamment proche de 1, les deux intervalles de pas donnés par les conditions d'Armijo et de courbure ne s'intersectent pas. Montrons le formellement.

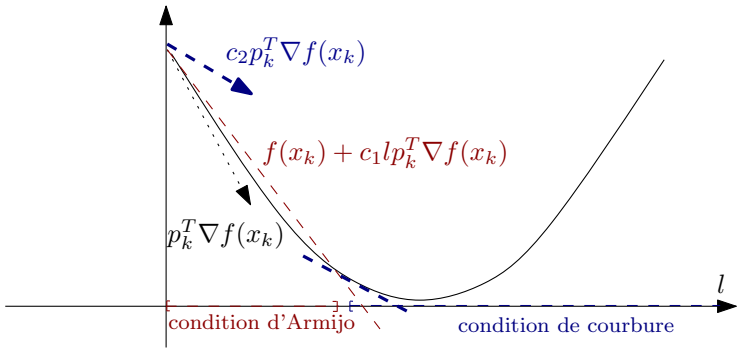


FIGURE 1. Cas où les conditions d'Armijo et de courbure ne peuvent être satisfaites simultanément.

Soient $l > 0$ et $p = \pm 1$ (les directions de descente sur $\mathbb{R}$ étant 1 ou -1). La condition d'Armijo s'écrit

$$\begin{aligned} f(x_0 + lp) \leq f(x_0) + c_1 l f'(x_0) p &\iff 1 + 2lp + (lp)^2 \leq 1 + 2c_1 lp, \\ &\iff 0 \leq (lp)^2 \leq 2lp(c_1 - 1). \end{aligned}$$

Puisque $c_1 - 1 < 0$, on en déduit $p = -1$, et donc $l \leq 2(1 - c_1)$ Quant à la condition de courbure, elle s'écrit

$$-f'(x_0 - l) \geq -c_2 f'(x_0) \iff 1 - c_2 \leq l.$$

Or par hypothèse, $0 < c_2 < c_1 < 1$, donc $1 - c_1 < 1 - c_2$. En prenant $c_1 = 3/4$ et $c_2 = 1/4$, il n'existe aucun l vérifiant les inégalités demandées.

Exercice 2.2 Descente de gradient stochastique

1. Selon le développement de Taylor avec reste intégral à l'ordre deux, on a

$$\begin{aligned} f(T_l^i(x)) = f(x) &+ \nabla f(x)^T (T_l^i(x) - x) \\ &+ \int_0^1 (T_l^i(x) - x)^T \nabla^2 (x + s(T_l^i(x) - x))(T_l^i(x) - x)(1 - s)\, ds. \end{aligned}$$

Or $T_l^i(x) - x = -l\frac{\partial f}{\partial x_i}(x)e_i$ où e_i est le i-ème vecteur de la base canonique. Ainsi,

$$\nabla f(x)^T(T_l^i(x) - x) = -l\left|\frac{\partial f}{\partial x_i}(x)\right|^2,$$

et

$$\begin{aligned}(T_l(x) - x)^T\nabla^2(x + s(T_l^i(x) - x))(T_l^i(x) - x) &= l^2\left|\frac{\partial f}{\partial x_i}(x)\right|^2\frac{\partial^2 f}{\partial x_i^2}(x + s(T_l^i(x) - x)) \\ &\leq L_i l^2\left|\frac{\partial f}{\partial x_i}(x)\right|^2.\end{aligned}$$

D'où l'inégalité

$$\begin{aligned}f(T_l^i(x)) &\leq f(x) - l\left|\frac{\partial f}{\partial x_i}(x)\right|^2 + L_i l^2\left|\frac{\partial f}{\partial x_i}(x)\right|^2\int_0^1(1-s)\,ds \\ &\leq f(x) - \frac{l}{2}(2 - lL_i)\left|\frac{\partial f}{\partial x_i}(x)\right|^2.\end{aligned}$$

2. Selon la question précédente,

$$\begin{aligned}E_k(f(x^{k+1})) &\leq \sum_{i=1}^n p_i\left(f(x^k) - \frac{\theta}{2L_i}(2-\theta)\left|\frac{\partial f}{\partial x_i}(x)\right|^2\right) \\ &\leq f(x^k) - \frac{\theta(2-\theta)}{2}\sum_{i=1}^n\frac{p_i}{L_i}\left|\frac{\partial f}{\partial x_i}(x)\right|^2 = f(x^k) - \frac{\theta(2-\theta)}{2}\|\nabla f(x^k)\|_M^2\end{aligned}$$

3. f convexe donc

$$f(x_1) - f(x_2) \leq \nabla f(x_1)^T(x_1 - x_2) = \nabla f(x_1)^T M^{1/2}M^{-1/2}(x_1 - x_2)$$

où $M^{1/2}$ racine carrée de M symétrique définie positive. En appliquant Cauchy-Schwarz, on conclut. On en déduit pour $(x_1, x_2) = (x^k, x)$

$$-\|\nabla f(x^k)\|_M \leq -\frac{f(x^k) - f(x)}{\|x^k - x\|_{M^{-1}}}$$

que l'on remplace dans l'inégalité de la question précédente pour obtenir l'inégalité voulue.

4. On a alors

$$E_k(f(x^{k+1}) - f(x)) \leq f(x^k) - f(x) - \frac{\theta(2-\theta)}{2C^2}(f(x^k) - f(x))^2$$

En utilisant le fait que

$$\begin{aligned}\mathbb{E}(f(x^{k+1}) - f(x)) &= \int_{-\infty}^{+\infty} s\underbrace{P(f(x^{k+1}) - f(x) = s)}_{=\int_{\mathbb{R}^n}P(f(x^{k+1})-f(x)=s|x^k=\tau)P(x^k=\tau)d\tau}ds \\ &= \int_{\mathbb{R}^n}E_k((f(x^{k+1}) - f(x)))P(x^k = \tau)d\tau\end{aligned}$$

on obtient

$$\begin{aligned}
&\mathbb{E}(f(x^{k+1}) - f(x)) \\
&\quad \leq \int_{\mathbb{R}^n} \left(f(x^k) - f(x) - \frac{\theta(2-\theta)}{2C^2}(f(x^k) - f(x))^2 \right) P(x^k = \tau) d\tau \\
&\quad \leq \Delta_k(x) - \frac{\theta(2-\theta)}{2C^2} E((f(x^k) - f(x))^2) \leq \Delta_k(x) - \frac{\theta(2-\theta)}{2C^2} \Delta_k(x)^2
\end{aligned}$$

5. $\Delta_k(x^\star)$ est l'espérance d'une variable aléatoire positive (car $x^\star$ minimiseur de f) et est donc positive. On en déduit $\Delta_k(x^\star) \leq 1/\alpha$ où $\alpha = \frac{\theta(2-\theta)}{2C^2}$. Montrons le résultat voulu par récurrence. Pour $k = 0$, on a bien $\Delta_k(x^\star) \leq 1/\alpha$. Supposons $\Delta_k(x^\star) \leq \frac{1}{\alpha(k+1)}$ et, par l'absurde, $\Delta_{k+1}(x^\star) > \frac{1}{\alpha(k+2)}$. Alors, de la question précédente, on conclut que

$$\frac{1}{\alpha(k+2)} < \Delta_{k+1}(x^\star) \leq \Delta_k(x^\star)(1 - \alpha\Delta_k(x^\star)) \leq \frac{1}{\alpha(k+1)} - \frac{\Delta_k(x^\star)}{k+1}$$

D'où

$$\Delta_k(x^\star) < \frac{1}{\alpha} - \frac{k+1}{\alpha(k+2)} = \frac{1}{\alpha(k+2)}$$

et ainsi $\Delta_{k+1}(x^\star) \leq \Delta_k(x^\star) < \frac{1}{\alpha(k+2)}$ ce qui contredit l'hypothèse de récurrence. On a donc une convergence sous-linéaire de l'espérance.

6. Si l'on ne dispose pas d'une formule analytique du gradient, cet algorithme a l'avantage par rapport à un gradient déterministe de ne requérir d'approximer numériquement qu'une différence finie scalaire, ce qui est plus aisé. On voit que le conditionnement du problème est lié à la constante C, qui augmente avec le conditionnement de $M^{-1} = \text{diag}(L_i/pi)$. On va donc choisir les p_i de sorte à ce que $p_i = L_i$ pour avoir la matrice la plus proche possible de l'identité (on peut toujours normaliser f pour que $\sum_i L_i = 1$). Ceci revient à privilégier les dérivées partielles pour lesquelles on suppose que les variations sont les plus importantes. Dans tous les cas, on n'a convergence qu'en espérance dans ce cas, cad en moyenne.

Exercice 2.3 Newton avec modification du hessien

1. A étant symétrique, $A + \tau I$ est également symétrique. Par ailleurs, en considérant le polynôme caractéristique, on a

$$det(A + \tau I - \lambda I) = det(A - (\lambda - \tau)I)$$

Aussi, le spectre de $A + \tau I$ est celui de A, décalé de $-\tau$. D'après la définition de τ, on déduit que toutes les valeurs propres de $A + \tau I$ sont supérieures à $\delta > 0$ et donc que $A + \tau I$ est symétrique définie positive.

2. **Avantages** : Pallie le problème de dégénérescence du hessien loin du minimum.
   Inconvénients : Nécessite de choisir β, a priori pas de bon choix et heuristique. Necessite également de disposer d'un critère de $A + \tau_k I$ définie positive (décomposition de Cholesky restant le moyen le plus direct). Vérifier la définie positivité est très coûteux... (Mais on peut réduire ce problème en multipliant d'un facteur 5 ou 10 au lieu de 2)

9.3 Exercices du TD3

Exercice 3.1 Minima liés

Dans une situation décrite par la Figure 2, la relation de Descartes exprime le lien entre l'objet, son image et la distance focale $f = \overline{OF'}$ de la lentille:

$$\frac{1}{\overline{OA'}} - \frac{1}{\overline{OA}} = \frac{1}{\overline{OF'}} \tag{82}$$

D'après les relations de Descartes, la contrainte s'écrit $c(x, y) := 1/y - 1/x - 1/f = 0$. On cherche à minimiser la distance $d(x, y) = y - x$ entre l'objet et son image par la lentille convergente. Le problème de minimisation est donc

$$\begin{aligned} \min_{x,y} \quad & d(x, y) \\ \text{tel que } & c(x, y) = 0 \end{aligned}$$

Pour $(x, y) \neq 0$, $\nabla c(x, y) \neq 0$,. La famille des contraintes actives est donc libre (la contrainte est qualifiée) et, par condition de KKT, une condition nécessaire pour que $(x^*, y^*) \neq 0$ soit solution est qu'il existe $\lambda \in \mathbb{R}^+$ tel que

$$\begin{aligned} \frac{\partial d}{\partial x}(x^*, y^*) + \lambda \frac{\partial c}{\partial x}(x^*, y^*) = 0, \\ \frac{\partial d}{\partial y}(x^*, y^*) + \lambda \frac{\partial c}{\partial y}(x^*, y^*) = 0, \\ c(x^*, y^*) = 0; \end{aligned}$$

soit donc

$$1 - \frac{\lambda}{{y^*}^2} = 0\,, \quad -1 + \frac{\lambda}{{x^*}^2} = 0, \,, \quad \frac{1}{y^*} - \frac{1}{x^*} = \frac{1}{f}.$$

Les deux premières relations donnent ${x^*}^2 = {y^*}^2$, et par la dernière on obtient $x^* = -y^*$ avec $y^* > 0$, ainsi que $y^* = 2f$, et donc $x^* = -2f$.

On ne peut en revanche pas conclure directement sur le fait que ce point est bien un minimum (il s'agit uniquement d'une condition nécessaire). Les fonction f et c

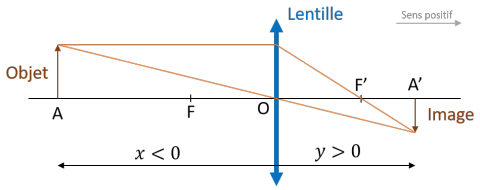


FIGURE 2. Principe de fonctionnement d'une lentille convergente.

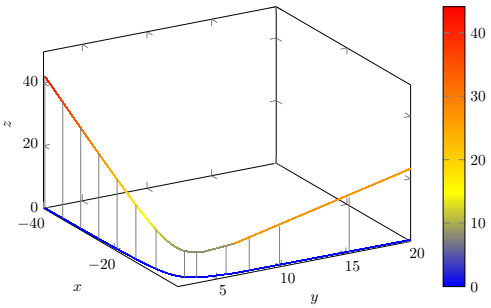


FIGURE 3. Visualisation de la structure coût sous contraintes (pour $f = 2$).

sont convexes, mais il n'existe pas $x \in \mathbb{R}$ tq $c(x) < 0$ (Proposition 8 de KKT) et l'espace de travail $c^{-1}((0,0))$ est non-convexe.

Pour montrer ceci, on peut montrer que la fonction f admet un minimum sur l'ensemble des contraintes. Alors, nécessairement, ce minimum est celui trouvé précédemment, puisqu'il est unique. Pour prouver ceci, on peut oberver que $f(x,y) \to \infty$ pour (x,y) tels que $c(x,y) = 1/y - 1/x - 1/f = 0$ et $\|(x,y)\| \to \infty$, puis invoquer le Théorème 2 du cours. Ceci est illustré sur la Figure 3.

Exercice 3.2 KKT

On note $c_1(x,y) = 1/2 - x$, $c_2(x,y) = -y$, $c_3(x,y) = y - x$ et $c_4(x,y) = y + x - 2$, de sorte à pouvoir écrire les contraintes comme $c(x,y) \leq 0$.

Si $a > 0$, J est une fonction strictement croissante de y et strictement décroissante de x. Elle est donc minimale au point $(2,0)$.

Si $a = 0$, J est une fonction strictement décroissante de x et ne dépend pas de y. Elle est donc minimale au point $(2,0)$.

Si $a < 0$, puisque $\nabla f(x,y) = (-1/x^2, a)$, graphiquement sur la Figure 4, on voit que, pour que $\nabla f(x,y)$ soit dans un cône de contraintes actives, il faut que c_4 soit active. On est donc réduit à considérer la fonction $J(x, -x+2) = -ax + 1/x + 2a$. On trouve trois cas : (i) si $a \in]-1/4, 0[$, alors $(x^*, y^*) = (2,0)$; (ii) si $a < -1$, alors $(x^*, y^*) = (1,1)$ et (iii) si $a \in [-1; -1/4]$, alors on a un minimum en $(1/\sqrt{|a|}, 2 - 1/\sqrt{|a|})$.

Exercice 3.3 Tente de cirque

Soit x^* la solution de $\min_{c(x)=0} f(x)$, et λ^* le multiplicateur de Lagrange associé. On considère $\delta c \in \mathbb{R}^{900}$ la variation de la contrainte (donc pour $1 \leq i \leq 900$, $\delta c_i = \delta h$ si i correspond à une coordonnée du mât central où δh est la variation de la hauteur du mât, et 0 sinon). Soit $\overline{x}$ la solution de $\min_{c(x)=\delta c} f(x)$. Pour des petites variations

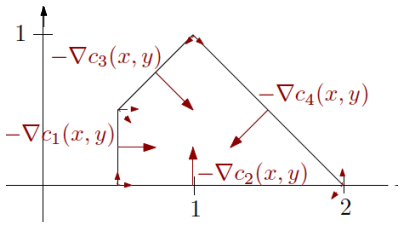


FIGURE 4. Illustration des contraintes et de l'opposé des gradients des contraintes actives.

δh, on a

$$f(\overline{x}) = f(x^*) + {\lambda^*}^T \delta c + o(\delta c).$$

Or, d'après le deuxième graphique de l'énoncé, les coordonnées correspondant au mât central sont au nombre de 4, et pour celles-ci, le multiplicateur de Lagrange vaut $\lambda_h = 0.17$. Donc ${\lambda^*}^T \delta c = 4\lambda_h \delta h$. On souhaite passer de $f(x^*) = 0.44422$ à $f(\overline{x}) = 0.43767$; pour ceci, la variation cherchée est

$$\delta h = \frac{f(\overline{x}) - f(x^*)}{4\lambda_h} = -9.63 \times 10^{-3}.$$

Remarque sur le signe de δh : il était a priori concevable d'avoir δh de signe opposé. Cependant, on cherche ici à diminuer le coût, qui correspond à une énergie potentielle. Il convient donc de diminuer la hauteur des mâts et de choisir δh négatif.

Exercice 3.4 Plus courte distance à un hyperplan

Soit $\mathcal{L}$ le lagrangien du système défini par

$$\mathcal{L} : (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \mapsto \frac{1}{2}(x - x_0)^T (x - x_0) + \lambda^T (Ax - b).$$

Un point stationnaire (x^*, λ^*) du lagrangien vérifie

$$\frac{\partial \mathcal{L}}{\partial x}^T (x^*, \lambda^*) = 0 = x^* - x_0 + A^T \lambda^*,$$
$$\frac{\partial \mathcal{L}}{\partial \lambda}^T (x^*, \lambda^*) = 0 = Ax^* - b.$$

En multipliant la première équation par A, il vient $b - Ax_0 + AA^T\lambda^* = 0$, donc $\lambda^* = -(AA^T)^{-1}(b - Ax_0)$. On obtient donc $x^* = x_0 + A^T(AA^T)^{-1}(b - Ax_0)$.

Or, on travaille avec un problème convexe f est convexe et $c^{-1}(\{0\}$ est convexe. Ainsi, le point x^* trouvé est bien un minimum global.

Si A est un vecteur ligne, alors $AA^T = \sum_{i=1}^{n} a_i^2 = \|A\|_2^2$. Puisque $b - Ax_0$ et AA^T sont des scalaires, on a

$$\begin{aligned} \|x^* - x_0\|_2 &= \|A^T\|_2 |AA^T|^{-1} |b - Ax_0|, \\ &= \frac{|b - Ax_0|}{\|A\|_2}. \end{aligned}$$

9.4 Exercices du TD4

Exercice 4.1 Pénalités intérieures

Soit $f_\mu : x \in]0,1[\mapsto x - \mu\log(x) - \mu\log(1-x)$. La fonction f_μ est de classe C^1 et convexe sur l'ouvert $]0,1[$. Ainsi, $x^*(\mu)$ est un minimum de f_μ si, et seulement si $f'_\mu(x^*(\mu)) = 0$, i.e. $1 - \mu/x + \mu/(1-x) = 0$. Cela revient donc à chercher une racine de $x^2 - x(2\mu+1) + \mu$. Ce polynôme admet deux racines

$$x_\pm = \frac{2\mu + 1 \pm \sqrt{(2\mu+1)^2 - 4\mu}}{2}$$

La plus grande racine est toujours supérieure à 1; l'autre racine est dans $]0,1[$, et est donc la solution du problème pénalisé

$$x^*(\mu) = \frac{2\mu + 1 - \sqrt{(2\mu+1)^2 - 4\mu}}{2}.$$

On a bien $\lim_{\mu\to 0} x^*(\mu) = 0$ où 0 est bien la solution du problème $\min_{0\le x\le 1} x$.

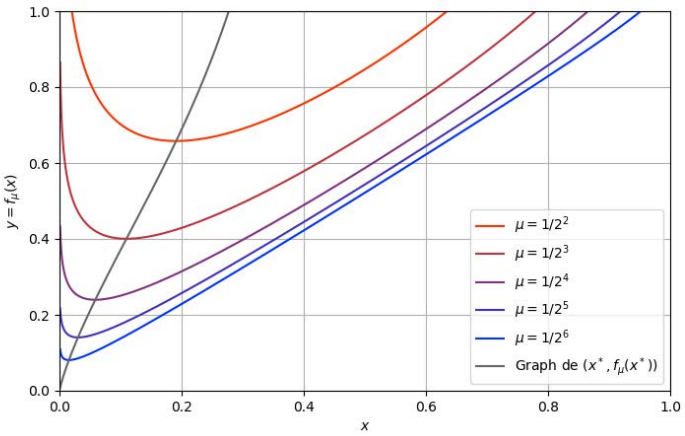


FIGURE 5. Graphique de la fonction f_μ et de son minimum pour différentes valeurs de μ.

Exercice 4.2 Dual d'une programmation quadratique

Soit $\mathcal{L}$ le lagrangien du système défini par

$$\mathcal{L} : (x,\lambda) \in \mathbb{R}^n \times (\mathbb{R}^+)^m \mapsto \frac{1}{2}x^T Gx + x^T d - \lambda^T(Ax - b).$$

Le problème dual est

$$\max_{\lambda \in \mathbb{R}^m_+} \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda).$$

Soit $\lambda \in \mathbb{R}^m_+$. La matrice G étant symétrique définie positive, $\mathcal{L}(\cdot, \lambda)$ admet un unique minimum global qui vérifie

$$\begin{aligned} x^* = \operatorname*{argmin}_{x \in \mathbb{R}^n}(\mathcal{L}(x, \lambda)) &\iff \frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda) = 0, \\ &\iff Gx^* + d - A^T\lambda = 0. \end{aligned}$$

Ainsi, le problème dual se réécrit

$$\begin{aligned} \min_{x,\,\lambda} \quad & \mathcal{L}(x, \lambda) \\ \text{tel que } & Gx + d - A^T\lambda = 0, \\ & \lambda \geq 0 \end{aligned}$$

La matrice G étant inversible car symétrique définie positive, la contrainte sur x dans le problème dual se traduit plus simplement par $x = G^{-1}(A^T\lambda - d)$. On obtient le problème simplifié en remplaçant x par cette expression dans $\mathcal{L}(x, \lambda)$ comme suit :

$$\begin{aligned} &\mathcal{L}(x^\star, \lambda) = \\ &\frac{1}{2}(A^T\lambda - d)^T G^{-1} G G^{-1}(A^T\lambda - d) + d^T G^{-1}(A^T\lambda - d) + \lambda^T(b - AG^{-1}(A^T\lambda - d)) \\ &= \frac{1}{2}(A^T\lambda - d)^T G^{-1}(A^T\lambda - d) + d^T G^{-1}(A^T\lambda - d) + \lambda^T b - \lambda^T A G^{-1}(A^T\lambda - d) \\ &= \frac{1}{2}\lambda^T A G^{-1} A^T\lambda - d^T G^{-1} A^T\lambda + \frac{1}{2} d^T G^{-1} d + d^T G^{-1} A^T\lambda - d^T G^{-1} d + \lambda^T b \\ &- \lambda^T A G^{-1} A^T\lambda + \lambda^T A G^{-1} d \\ &= -\frac{1}{2}\lambda^T A G^{-1} A^T\lambda + (b^T + d^T G^{-1} A^T)\lambda - \frac{1}{2} d^T G^{-1} d \end{aligned}$$

On peut noter que ce second problème est toujours un problème quadratique, mais que les contraintes associées ont pu être considérablement simplifiées.

Exercice 4.3 Contraintes actives pour Programmation Linéaire : algorithme du simplexe

1. On réécrit le problème avec des variables intermédiaires x_3 et x_4

$$\begin{aligned} -\min_{x \in \mathbb{R}^4} \quad & -x_1 - 2x_2 \\ \text{tel que} \quad & x_2 + x_3 = 1, \\ & 2x_1 + x_2 + x_4 = 2, \\ & x \geq 0 \end{aligned}$$

qui est bien (au signe près) de la forme LP

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^4} & c^T x \\ \text{tel que} & Ax = b, \\ & x \geq 0 \end{array}$$

avec

$$c = (-1 \ -2 \ 0 \ 0)^T \ , \ A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} \text{ et } b = (1 \ 2)^T \ . \tag{83}$$

Etape 1: On commence avec x_3 et x_4 comme variables de base, ce qui correspond à $x_B^0 = (1 \ 2)^T$, $A_B^0 = I_2$, $A_N^0 = \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}$, $c_B^0 = (0 \ 0)^T$ et $c_N^0 = (-1 \ -2)^T$. Alors:

- **Calcul du coût réduit:** $\lambda = c_N^0 - (A_N^0)^T (A_B^0)^{-T} c_B^0 = c_N^0 = (-1 \ -2)^T$
- **Indice de l'élément le plus négatif du coût réduit:** $q = \text{argmin}_{i=1,2} \lambda_i = 2$ et $y_q = (A_B^0)^{-1} a_{2+q} = (1 \ 1)^T$ (où a_{2+q} est la $2+q^{ieme}$ colonne de $(A_B^0 \ A_N^0)$).
- **Indice du ratio le plus petit:** $p = \text{argmin}_{i=1,\ldots,n-m} \frac{b_i}{(y_q)_i} = \text{argmin}_{i=1,2} \{1, 2\} = 1$. On va donc remplacer x_3 (la p^{ieme} variable de base) par x_4 (la q^{ieme} variable hors base) dans la base courante.
- **Mise à jour de la base:** on considère $((A_B^0)^{-1} \mid y_q) = \left(\begin{array}{cc|c} 1 & 0 & \boxed{1} \\ 0 & 1 & 1 \end{array} \right)$. Par un pivot réalisé sur $(y_q)_1$, l'élément encadré, on obtient la mise à jour

$$(A_B^1)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 - 1 \times 1 & 1 - 0 \times 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \tag{84}$$

qui correspond à la base (x_2, x_4) et à la solution faisable $x_B^1 = (A_B^1)^{-1} b = (1 \ 1)^T$.

Etape 2: Avec la base (x_2, x_4), on a $A_N^1 = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$, $c_B^1 = (-2 \ 0)^T$ et $c_N^1 = (-1 \ 0)^T$. Alors:

- **Calcul du coût réduit:**

$$\lambda = c_N^1 - (A_N^1)^T (A_B^1)^{-T} c_B^1 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \tag{85}$$

- **Indice de l'élément le plus négatif du coût réduit:** $q = \text{argmin}_{i=1,2} \lambda_i = 1$ et $y_q = (A_B^1)^{-1} a_{2+q} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$.
- **Indice du ratio le plus petit:** $p = \text{argmin}_{i=1,\ldots,n-m} \frac{b_i}{(y_q)_i} = \text{argmin} \{+\infty, 1\} = 2$. On va donc remplacer x_4 (la p^{ieme} variable de base) par x_1 (la q^{ieme} variable hors base) dans la base courante.

– **Mise à jour de la base:** on considère $((A_B^1)^{-1} \mid y_q) = \left(\begin{array}{cc|c} 1 & 0 & 0 \\ -1 & 1 & \boxed{2} \end{array}\right)$.

Par un pivot réalisé sur $(y_q)_2$, l'élément encadré, on obtiendrait la mise à jour

$$(A_B^2)^{-1} = \begin{pmatrix} 1 - 0 \times (-1/2) & 0 - 0 \times 1/2 \\ -1/2 & 1/2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \tag{86}$$

qui correspond à la base (x_2, x_1). En inversant l'ordre de x_1 et x_2 dans la base, on définit plutôt

$$(A_B^2)^{-1} = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 2 & 0 \end{pmatrix} \tag{87}$$

et la solution faisable $x_B^2 = (A_B^2)^{-1}b = (1/2\ 1)^T$.

Etape 3: Avec la base (x_1, x_2), on a $A_N^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $c_B^2 = (-1\ \ -2)^T$ et $c_N^2 = (0\ 0)^T$. Alors:

– **Calcul du coût réduit:**

$$\lambda = c_N^2 - (A_N^2)^T (A_B^2)^{-T} c_B^2 = -\frac{1}{2}\begin{pmatrix} -1 & 2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -1 \\ -2 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 1/2 \end{pmatrix} \geq 0 \tag{88}$$

L'optimum est donc atteint en $x_B^2 = (1/2\ 1)^T$.

En effet, le problème est convexe, avec un ensemble de contraintes non-vide. Par conditions de KKT, l'optimum trouvé est donc un minimum global du problème (qui n'est pas nécessairement unique a priori, néanmoins).

2. (a) Une solution faisable triviale de (73) est $(x, y) = (0^T, b^T)^T$. Par ailleurs, on remarque que ce problème est également de la forme LP, avec pour nouvelle variable de décision $z = (x, y)$. En effet, plus précisément, on peut le reformuler comme

$$\begin{aligned} \min_z \quad & \bar{c}^T z \\ \text{tel que } & \bar{A} z = b, \\ & z \geq 0 \end{aligned} \tag{89}$$

avec $\bar{c} = (0\ 1_m)$ et $\bar{A} = (A\ I_m)$. On peut donc le résoudre par l'algorithme du simplexe, avec la solution faisable initiale que l'on vient de déterminer, par exemple.

(b) Si le minimiseur $z^\star = (x^\star, y^\star)$ obtenu est tel que $y^\star = 0$, alors il satisfait $Ax^\star = 0$ et $x^\star \geq 0$ par définition des contraintes du problème (73). Il s'agit donc d'une solution faisable de (72), que l'on peut utiliser comme condition initiale pour sa résolution. En revanche, si $y^\star > 0$, alors il n'existe pas de solution faisable à (72).

9.5 Exercices du TD5

Exercice 5.1 Inégalité de Young

Soit $f : \mathbb{R}_+ \to \mathbb{R}_+$ continue strictement croissante, avec $f(0) = 0$ et d'inverse g. Soient:

$$F(x) := \int_0^x f(\tau)d\tau \quad \text{et} \quad G(y) := \int_0^y g(\tau)d\tau.$$

1. Montrons que $G = F^*$. Notons $J_y(x) := yx - F(x)$. Cette fonction en x est dérivable, telle que

$$J_y'(x) = y - f(x)$$

f étant croissante, J_y' est décroissante, ne s'annulant qu'en $x^* = g(y)$. Aussi, $J_y'(x) > 0$ pour $x < x^*$ et $J_y'(x) < 0$ pour $x^* > x$. Ainsi, J_y atteint un supremum (*a fortiori* un maximum). Ainsi, la transformée de Fenchel de F existe, et

$$F^*(y) = yg(y) - F(g(y))$$

Cas simple. Supposons maintenant que f soit dérivable. Comme f est strictement croissante, g est donc aussi dérivable sur $\mathbb{R}$. Ainsi, par intégration par partie

$$\begin{aligned} F^*(y) &= yg(y) - \int_0^{g(y)} f(\tau)d\tau \\ &= yg(y) - \int_0^y ug'(u)du \\ &= yg(y) - [ug(u)]_0^y + \int_0^y g(u)du \\ &= G(y) \end{aligned}$$

Cas compliqué. Supposons désormais que f n'est pas dérivable. Dans ce cas, on peut se convaincre graphiquement de la véracité du résultat avec la Figure 6.

Analytiquement, il est possible de conclure en invoquant des éléments de calcul intégral. Plus précisément, on peut montrer que l'image de la mesure de Lebesgue par une application monotone et continue $h : [a, b] \to \mathbb{R}$ est également une mesure, nommée mesure de Stieltjes et notée μ_h. On peut alors définir l'intégrale de Stieltjes de f comme étant l'intégrale de Lebesgue par rapport à la mesure μ_h et on la note

$$\int_a^b f(x)dh(x)$$

Par ailleurs, si l'on considère deux applications h_1 et h_2 monotones et continues, on dispose d'une propriété d'intégration par partie pour l'intégrale de Stieltjes,

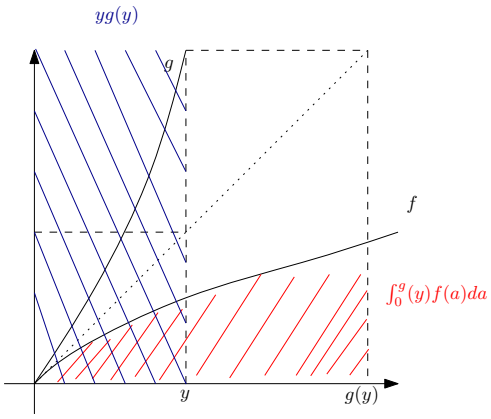


FIGURE 6. Représentation graphique des quantités manipulées dans l'exercice.

qui est

$$\int_a^b h_1(x)dh_2(x) = [h_1(x)h_2(x)]_a^b - \int_a^b h_2(x)dh_1(x)$$

En appliquant cette propriété à $h_1 = Id$ et $h_2 = g$, on peut donc suivre le même raisonnement que précédemment.

On note alors la fonction:

$$H : y \to yg(y) - F(g(y))$$

Alternativement, on peut traiter le problème plus directement (mais plus longuement...) en considérant le taux d'accroissement. On cherche à montrer que la limite du taux d'accroissement en tout point est égale à $g(y)$. On pose $y = f(x)$. En notant $(h', h) \in \mathbb{R}_+^2$, tels que:

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

on obtient:

$$\lim_{h' \to 0} \frac{H(y + h') - H(y)}{y + h' - y} = \lim_{h \to 0} \frac{H(f(x + h)) - H(f(x))}{f(x + h) - f(x)} \tag{90}$$

On cherche donc à montrer que le membre de droite de (90) tend vers x. Pour cela, on majore la quantité suivante:

$$\begin{aligned} \left| \frac{H(f(x + h)) - H(f(x))}{f(x + h) - f(x)} - x \right| &= \left| \frac{(x + h)f(x + h) - xf(x) - F(x + h) - F(x)}{f(x + h) - f(x)} - x \right| \\ &= \left| h + h \frac{f(x)}{f(x + h) - f(x)} + \frac{F(x + h) - F(x)}{f(x + h) - f(x)} \right| \end{aligned}$$

On considère ici $h > 0$. Grâce au théorème des accroissement finis:

$$|F(x + h) - F(x)| \leq \sup_{t \in [0,h]} |f(x + t)||h| = f(x + h).|h|$$

Comme les fonctions concernées sont croissantes, on obtient:

$$\left|\frac{H(f(x+h))-H(f(x))}{f(x+h)-f(x)}\right| \leq h.\left(1+\frac{f(x)}{f(x+h)-f(x)}+\frac{f(x+h)}{f(x+h)-f(x)}\right)=2h$$

La fin du raisonnement est quasi-identique dans le cas où $h<0$. Ainsi, H est dérivable en tout $y>0$ et $H'(y)=g(y)$. Donc,

$$H(y)-H(0)=\int_0^y g(\tau)d\tau=G(y)$$

D'où $G=F^*$ dans le cas où f n'est pas dérivable.

2. Montrons l'inégalité de Young $F(x)+G(y)\geq xy$. Pour un $y>0$ fixé, on note : $h_y : x \to F(x)+G(y)-xy$. Par construction, h_y est dérivable en tout $x \geq 0$ et:

$$h'_y(x)=f(x)-y$$

Comme f est croissante, pour tout $x>g(y)$, on a $h'_y(x)>0$. De même, pour tout $x<g(y)$, on a $h'_y(x)<0$. Ainsi, h_y admet un minimum en $x=g(y)$ et:

$$h_y(x) \geq h_y(g(y))=0$$

D'où: $F(x)+G(y)\geq xy$.

3. Application. On prend: $f(x)=e^x-1$. Ainsi:

$$\begin{aligned} g(y)&=\log(y+1)\\ F(x)&=e^x-1-x\\ G(x)&=(y+1)\log(y+1)-y \end{aligned}$$

Il suffit d'appliquer l'inégalité de Young et changer les variables en $u=y+1$ et $v=x+1$ pour conclure:

$$e^{v-1}+u\log u \geq uv, \quad u,v\geq 1$$

Exercice 5.2 Conjuguée et opérateur proximal d'une norme

1. Soit y tel que $\|y\|_D>1$. Alors, il existe $x_0\neq 0$ tel que $\|y\|_D=y^Tx_0>1\geq\|x_0\|$ (maximum d'une fonction continue sur un ensemble compact). Ainsi: $y^Tx_0-\|x_0\|>0$. On peut donc trouver un x tel que $y^Tx-\|x\|$ soit arbitrairement grand (les αx_0 pour $\alpha>0$ conviennent). D'où $\|y\|^*=+\infty$.

   Soit y tel que $\|y\|_D\leq 1$. Soit $x\neq 0$. Alors:

$$y^Tx-\|x\|=\|x\|\left(y^T\frac{x}{\|x\|}-1\right).$$

Comme $\frac{x}{\|x\|}$ est de norme inférieure à 1, par définition de la norme duale : $y^T\frac{x}{\|x\|}<1$. Ainsi : $y^Tx-\|x\|\leq 0$. Comme cette expression atteint 0 en $x=0$, on obtient : $\|y\|^*=0$.

2. En utilisant la question précédente, pour x dans $\mathbb{R}^n$, le minimum de:

$$y \mapsto \|y\|^* + \frac{1}{2}\|y - x\|_2^2$$

est nécessairement atteint sur $\mathcal{B}_D$, étant donné que ce terme vaut $+\infty$ ailleurs. Ainsi:

$$\text{Prox}_{\|.\|^*}(x) = \underset{s \in \mathcal{B}_D}{\text{argmin}} \left(0 + \frac{1}{2}\|s - x\|_2^2\right) = \Pi_{\mathcal{B}_D}(x)$$

On remarquera de plus que: $\text{Prox}_{\mu\|.\|^*} = \text{Prox}_{\|.\|^*}$.

3. Utilisation de l'identité de Moreau:

$$\text{Prox}_{\mu\|.\|}(x) = x - \mu\,\text{Prox}_{\frac{1}{\mu}\|.\|^*}\left(\frac{x}{\mu}\right) = x - \mu\Pi_{\mathcal{B}_D}\left(\frac{x}{\mu}\right)$$

Soit pour $\mu = 1$, $\text{Prox}_{\|.\|}(x) = x - \Pi_{\mathcal{B}_D}(x)$.

4. Montrons déjà que la norme duale de la norme 1 est la norme ∞:

$$\forall x, \|x\|_D = \sup_{\|s\|_1 \leq 1} |s^T x| \leq \sup_{\|s\|_1 \leq 1} \sum |s_i x_i| \leq \left(\sup_{\|s\|_1 \leq 1} \|s\|_1\right) \|x\|_\infty = \|x\|_\infty$$

En prenant s tel que $s_i = 1$ pour un des i réalisant $|x_i| = \|x\|_\infty$ et $s_j = 0$ pour toutes les autres coordonnées de s, on a $|s^T x| = \|x\|_\infty$ et $\|s\|_1 = 1$. Donc le supremum est réalisé en ce s et $\|x\|_1^* = \|x\|_\infty$.

On peut réécrire la projection sur la boule unité duale telle que:

$$\Pi_{\mathcal{B}_D}(x) = \underset{\{s \,|\, \forall i, |s_i| \leq 1\}}{\text{argmin}} \sum_{i=1}^{n} (s_i - x_i)^2$$

Chaque composante s_i est indépendante des autres sous cette formulation. D'où:

$$\Pi_{\mathcal{B}_D}(x)|_i = \begin{cases} 1 \text{ si } x_i > 1 \\ x_i \text{ si } x_i = 1 \\ -1 \text{ si } x_i < 1 \end{cases}$$

Cette projection est représentée en Figure 7. On conclut en utilisant la question précédente:

$$\text{Prox}_{\mu\|.\|}(x) = x - \mu\Pi_{\mathcal{B}_D}\left(\frac{x}{\mu}\right) = \begin{cases} x_i - \mu \text{ si } x_i/\mu > 1 \\ x_i - \mu x_i/\mu \text{ si } |x_i/\mu| \leq 1 \\ x_i + \mu \text{ si } x_i/\mu < -1 \end{cases} = \begin{cases} x_i - \mu \text{ si } x_i \geq \mu \\ 0 \text{ si } |x_i| \leq \mu \\ x_i + \mu \text{ si } x_i < -\mu \end{cases}$$

Exercice 5.3 Problème de Lasso

1. Soit $l > 0$. Une solution x^* du problème vérifie $0 \in l\,\partial\left(\|A \cdot - b\|_2^2 + \gamma\|\cdot\|_1\right)(x^*)$. Le terme quadratique étant régulier, son sous-différentiel est simplement son

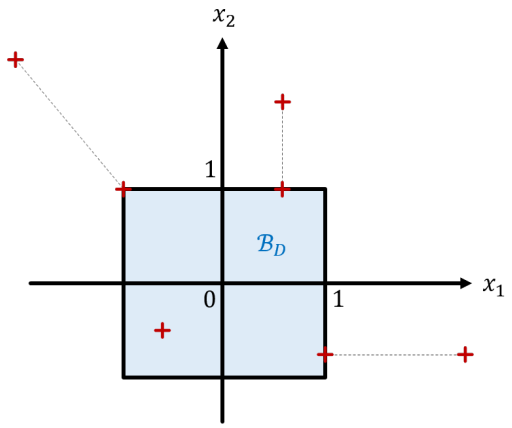


FIGURE 7. Projection de points de $\mathbb{R}^2$ sur la boule unité fermée.

gradient. En développant $\|Ax - b\|_2^2 = x^T A^T Ax - 2b^T Ax + b^T b$ et en utilisant les résultats de l'exercice 1.2 du TD1, on obtient que ce gradient en x est $2(A^T Ax - A^T b)$. Ainsi, x^* vérifie

$$0 \in 2l(A^T Ax^* - A^T b) + l\gamma \partial \|\cdot\|_1(x^*).$$

En sommant par $x^* - 2l(A^T Ax^* - A^T b)$ de chaque côté, il vient

$$x^* - 2l(A^T Ax^* - A^T b) \in x^* + l\gamma \partial \|\cdot\|_1(x^*).$$

2. La norme $\|\cdot\|_1$ étant convexe et $l\gamma > 0$, on a $s = \mathrm{Prox}_{l\gamma\|\cdot\|_1}(x^* - 2l(A^T Ax^* - A^T b))$ si, et seulement si,

$$s = (\mathrm{Id} + l\gamma\|\cdot\|_1)^{-1}(x^* - 2l(A^T Ax^* - A^T b)).$$

Et donc d'après la question précédente, $s = x^*$. Le terme $l\gamma\|\cdot\|_1$ a donc un effet régularisant.

CHAPITRE 10

EXERCICES ET PROBLÈMES COMPLÉMENTAIRES

10.1 Minimisation sous contraintes

En utilisant les conditions de Karush-Kuhn-Tucker, trouver le minimum de

$$\begin{aligned} \min_{y_1,\, y_2} \quad & J(y_1, y_2) = (\frac{y_1}{2} - a)^2 + (y_2 - b)^2 \\ \text{tel que} \quad & y_1^2 + y_2^2 \leq 1, \\ & y_1 \geq 0, \\ & y_2 \geq 0 \end{aligned}$$

Traiter tous les cas possibles pour a et b.

Résoudre le même problème en rajoutant la contrainte $y_1 + y_2 \geq \frac{1}{2}$.

10.2 Problème de Weber

Étant donné un ensemble de points dans un plan $\{y_1, ..., y_m\}$, on cherche le point dont la somme des distances pondérées est minimum. Le problème s'écrit

$$\min_{x \in \mathbb{R}^3} \sum_{i=1}^{m} m_i \parallel x - y_i \parallel$$

où les poids m_i sont des constantes positives.

1. Montrer qu'il existe un minimum global à ce problème et qu'on peut le construire par le schéma représenté Figure 1.
2. Ce minimum est-il unique?

10.3 Optimisation géométrique

Considérons le quadrilatère représenté sur la Figure 2 avec les notations telles que définies sur cette figure ($(\theta_1, \theta_2) \in]0, \pi[^2$, $a > 0$, $b > 0$, $c > 0$, $d > 0$). On cherche la configuration (θ_1, θ_2) fournissant l'aire maximale.

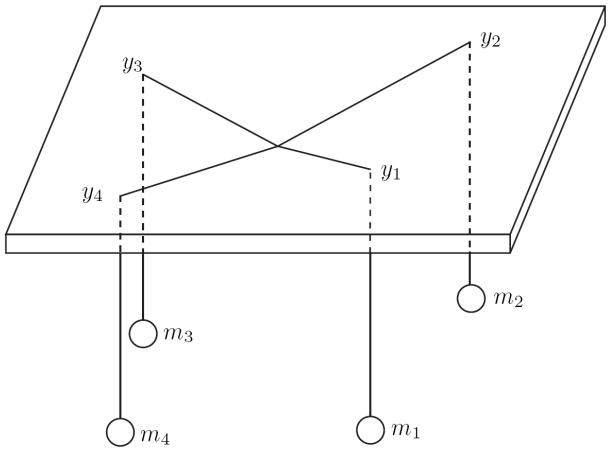


FIGURE 1. Construction de Varignon: planche percée de trous par lesquels passent des ficelles nouées en un point. On attache une masse à l'extrémité libre de chaque ficelle.

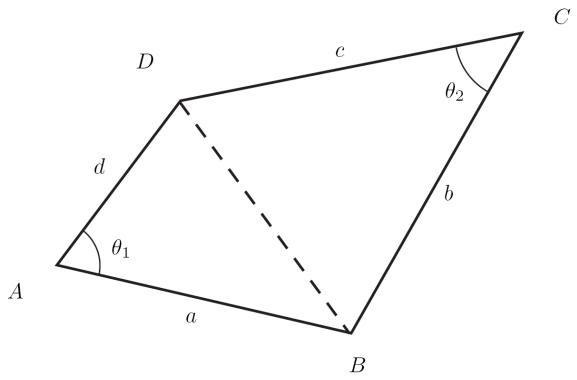


FIGURE 2. Quadrilatère.

1. Montrer que le problème revient à trouver l'extremum de

$$(\theta_1, \theta_2) \mapsto ad \sin \theta_1 + bc \sin \theta_2$$

sous la contrainte

$$a^2 + d^2 - 2ad \cos \theta_1 = b^2 + c^2 - 2bc \cos \theta_2$$

2. Écrire le Lagrangien associé à ce problème et établir des conditions d'extrémalité.
3. Calculer les valeurs de (θ_1, θ_2) extremum.

4. Quelle figure obtient-on pour $a = b = c = d = 1$? Même question pour $a = d = 1$ et $b = c = 2 + \sqrt{3}$.

10.4 Tarification de billets d'avions

Lorsque vous achetez un billet d'avion aller retour, vous pouvez bénéficier d'une importante réduction si vous acceptez de rester le samedi soir sur place. Dans cet exercice on va étudier comment, dans un cas très simple, cette ristourne est calculée.

On considère deux populations: les touristes (avec les variables z_n, α_n) et les hommes d'affaires (avec les variables z_b, α_b).

On note z_n (resp. z_b) l'agrément de passer la nuit chez soi le samedi, p_n (resp. p_b) le prix du billet, α_n (resp. α_p) la pénibilité de payer le prix du billet, et t l'agrément du séjour. On note p_n le prix d'un billet aller retour avec séjour sur place le samedi, et $p_b \geq p_n$ le prix d'un billet aller retour sans séjour sur place le samedi.

1. Expliquer le sens des hypothèses (qu'on admettra dans la suite)
$$0 = z_n < z_b, \quad \alpha_n > \alpha_b \geq 0$$
2. Expliquer pourquoi l'agrément total d'un individu s'exprime, suivant les cas, comme
$$u = z - \alpha p + t \quad \text{ou} \quad u = -\alpha p + t$$
3. On souhaite donner l'envie aux touristes et aux hommes d'affaires de voyager. On souhaite choisir une politique tarifaire telle que les touristes aient envie de voyager en passant la nuit de samedi sur place et on souhaite encourager les hommes d'affaire à prendre un billet sans séjour du samedi. Expliquer qu'on aboutit aux relations

$$-\alpha_n p_n + t \geq 0 \tag{91}$$

$$z_b - \alpha_b p_b + t \geq 0 \tag{92}$$

4. On souhaite que des deux types de voyages le plus intéressant pour les touristes soit le voyage avec séjour du samedi, et que les hommes d'affaires privilégient les séjours sans séjour du samedi. Montrer qu'on aboutit à

$$-\alpha_n p_n \geq z_n - \alpha_n p_b \tag{93}$$

$$z_b - \alpha_b p_b \geq -\alpha_b p_n \tag{94}$$

5. Étant donnés les flux supposés égaux d'hommes d'affaires et de touristes on souhaite maximiser la fonction
$$p_n/2 + p_b/2$$
Les variables de décisions étant p_n et p_b (la politique tarifaire), écrire le problème d'optimisation sous contraintes.
6. En utilisant les conditions de Karush-Kuhn-Tucker, combien y-a-t-il de cas à considérer pour résoudre ce problème?
7. Montrer que les conditions (92) et (93) sont redondantes avec (91) et (94). Réécrire le problème d'optimisation sous contraintes correspondant.

8. Résoudre ce problème d'optimisation et calculer, à l'optimum, $p_b - p_n$. Interpréter le résultat.
9. Reprendre l'étude si les flux de touristes et d'hommes d'affaires sont différents.
10. On suppose maintenant que la compagnie vend aussi des nuits d'hotel et que la nuit du samedi soir est facturée c. Comment doit-on modifier l'étude précédente pour maximiser le profit total?

10.5 Inégalité de Kantorovich

On cherche à montrer le résultat suivant. Soit Q une matrice symétrique définie positive de taille $n \times n$. Quel que soit x, on a

$$q(x) = \frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} \geq \frac{4aA}{(a+A)^2}$$

où a et A sont, respectivement, la plus petite et la plus grande des valeurs propres de Q.

1. Dans un premier temps, on considère que Q est une matrice diagonale dont les termes sont
$$0 < a = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n = A$$
Montrer que
$$q(x) = \frac{(\sum_{i=1,...,n} x_i^2)^2}{(\sum_{i=1,...,n} \lambda_i x_i^2)(\sum_{i=1,...,n} \frac{1}{\lambda_i} x_i^2)}$$
2. Montrer que
$$q(x) = \frac{1}{f(\lambda_1, ...\lambda_n, x) g(\lambda_1, ...\lambda_n, x)}$$
avec
$$f(\lambda_1, ...\lambda_n, x) = \sum_{i=1,...,n} \lambda_i \frac{x_i^2}{\sum_{i=1,...,n} x_i^2}, \quad g(\lambda_1, ...\lambda_n, x) = \sum_{i=1,...,n} \frac{1}{\lambda_i} \frac{x_i^2}{\sum_{i=1,...,n} x_i^2}$$
3. En interprétant f et g définis précédemment comme les coordonnées d'un barycentre de points de coordonnées $(\lambda_i, 1/\lambda_i)$, représenter dans un plan l'ensemble des points de coordonnées (f, g) lorsque x varie (les λ_i restant constants). Montrer, pour λ_1, ..., λ_n fixés, qu'il existe $\alpha(x) \in [0,\ 1]$ tel que
$$f(\lambda_1, ...\lambda_n, x) = (1 - \alpha(x))\lambda_1 + \alpha(x)\lambda_n$$
$$g(\lambda_1, ...\lambda_n, x) \leq (1 - \alpha(x))\frac{1}{\lambda_1} + \alpha(x)\frac{1}{\lambda_n}$$
4. Déduire de ce qui précède que
$$f(\lambda_1, ...\lambda_n, x) g(\lambda_1, ...\lambda_n, x) \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}$$
Conclure. Conclure dans le cas général où Q n'est pas diagonale.

5. On considère le problème de minimisation de la fonction quadratique
$$h(x) = \frac{1}{2}x^T Q x - b^T x$$
où b est un vecteur colonne. Montrer que l'algorithme du gradient à pas optimal engendre les itérations
$$x^{k+1} = x^k - \frac{(g^k)^T g^k}{(g^k)^T Q g^k} g^k$$
avec $g^k = Qx^k - b$.
6. On note
$$E(x) = \frac{1}{2}(x - x^*)^T Q (x - x^*)$$
où x^* est l'unique minimum de h. Calculer $E(x^{k+1})$ en fonction de $E(x^k)$ et de g^k. En utilisant l'inégalité de Kantorovich que vous venez d'établir, montrer que
$$E(x^{k+1}) \leq \left(\frac{A-a}{A+a}\right)^2 E(x^k)$$
7. Que peut-on en déduire sur la convergence de cette méthode de descente en fonction du conditionnement de la matrice Q?

10.6 Dualité

1. Considérer le problème (très simple)
$$\min_{x\in\mathbb{R},\quad x\geq 1} |x|$$
Quelle est sa solution? Former le Lagrangien du problème et considérer la fonction duale
$$f(x) = \sup_{\lambda\geq 0} \lambda(1-x)$$
Expliciter les valeurs prises par cette fonction pour tout x. En déduire, par dualité, quelle est la solution du problème.
2. Considérer le problème
$$\begin{array}{ll} \min\limits_{x\,\in\,\mathbb{R}^n} & c^T x \\ \text{tel que} & Ax = b, \\ & x \geq 0 \end{array}$$
Montrer que son dual est
$$\begin{array}{ll} \min\limits_{\lambda,\,\nu} & -b^T \nu \\ \text{tel que} & A^T\nu - \lambda + c = 0, \\ & \lambda \geq 0 \end{array}$$
Préciser les dimensions de λ et ν. Peut-on simplifier cette formulation duale?

3. Soit A une matrice de taille $m \times n$ et b un vecteur colonne. On cherche $x \in \mathbb{R}^n$ solution du problème

$$\begin{aligned} &\min_x \quad c^T x \\ &\text{tel que } Ax \leq b \end{aligned}$$

Former le Lagrangien du problème. Par dualité, montrer que ce problème revient à

$$\begin{aligned} &\min_\lambda \quad -b^T \lambda \\ &\text{tel que } A^T \lambda + c = 0, \\ &\lambda \geq 0 \end{aligned}$$

10.7 Résultats sur la minimisation d'une fonction elliptique

On considère une fonction $J : \mathbb{R}^n \to \mathbb{R}$ qu'on suppose *elliptique*, c'est à dire qu'elle est continûment différentiable, que son gradient ∇J est Lipschitzien avec comme constante M, et qu'il existe $\alpha > 0$ tel que

$$(\nabla J(v) - \nabla J(u))^T (v - u) \geq \alpha \|u - v\|^2$$

pour tout u et v. On s'intéresse à la minimisation de J sur le domaine

$$\mathcal{U} = \{u \in \mathbb{R}^n \text{ tels que } \phi_i(u) \leq 0, \quad 1 \leq i \leq m\}$$

où les fonctions $\phi_i : \mathbb{R}^n \to \mathbb{R}$ sont convexes.

1. Établir, en utilisant la formule de Taylor avec reste intégral pour évaluer $J(v) - J(u)$, et le fait que J est elliptique, la double inégalité suivante:

$$\nabla J(u)^T (v - u) + \frac{\alpha}{2} \|u - v\|^2 \leq J(v) - J(u) \leq \nabla J(u)^T (v - u) + \frac{M}{2} \|v - u\|^2$$

2. En faisant le lien avec la convexité forte, démontrer l'existence et l'unicité de la solution du problème d'optimisation

$$\min_{u \in \mathcal{U}} J(u)$$

3. On définit le Lagrangien $\mathcal{L}(u, \lambda) = J(u) + \lambda^T \phi(u)$. Montrer que si (u, λ) est point selle du Lagrangien, alors u est solution du problème.
4. On note $p : \mathbb{R} \to \mathbb{R}_+$ tel que $p(x) = \begin{cases} x \text{ si } x \geq 0 \\ 0 \text{ sinon} \end{cases}$, et $P : \mathbb{R}^m \to \mathbb{R}^m$ tel que $P(x = (x_1, ..., x_m)^T) = \begin{pmatrix} p(x_1) \\ \vdots \\ p(x_m) \end{pmatrix}$ l'opérateur de *projection* sur $(\mathbb{R}_+)^m$.

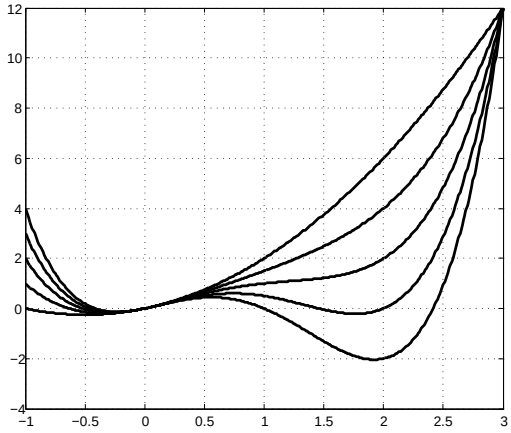


FIGURE 3. Graphe de la fonction $f(x,\theta)$ pour différentes valeurs de θ.

Établir, en utilisant les conditions Karush-Kuhn-Tucker, l'équivalence suivante

$$\{\lambda = P(\lambda + \rho\phi(u)), \rho > 0 \text{ fixé}\}$$
$$\Leftrightarrow \left\{ \begin{array}{c} \lambda \in (\mathbb{R}_+)^m \\ \sum_i \phi_i(u)(\mu_i - \lambda_i) \leq 0 \text{ pour tout } \mu = (\mu_1, ..., \mu_m)^T \in (\mathbb{R}_+)^m \end{array} \right.$$

5. En déduire (en utilisant certaines valeurs bien choisies de μ_i) l'équivalence

$$\{\lambda = P(\lambda + \rho\phi(u)), \rho > 0 \text{ fixé}\}$$
$$\Leftrightarrow \{ \ \lambda \in (\mathbb{R}_+)^m \lambda^T \phi(u) = 0$$

6. Vérifier que si (u,λ) est point selle du Lagrangien, alors $\lambda = P(\lambda + \rho\phi(u))$ pour un certain ρ, et que

$$\mathcal{L}(u,\lambda) \leq \mathcal{L}(u + \theta(v-u), \lambda)$$

pour tout v et pour tout $\theta \in [0,1]$. Faire le lien avec l'algorithme d'Uzawa.

7. En déduire, en utilisant un développement de Taylor que

$$\nabla J(u)^T (v-u) + \lambda^T \phi(v) \geq 0$$

pour tout v.

10.8 Sur un théorème de C. Berge

On s'intéresse au problème de minimisation (globale)

$$\min_{x \in [a(\theta); b(\theta)]} f(x,\theta) \tag{95}$$

où f est une fonction à valeurs dans $\mathbb{R}$, continue par rapport à chacune de ses variables et où $a, b : \mathbb{R} \to \mathbb{R}$ sont continues et vérifient en particulier la propriété suivante:

$\forall \theta \in \mathbb{R},\ a(\theta) < b(\theta)$ (pour assurer la bonne définition de l'intervalle considéré dans (95)). Le graphe d'une telle fonction f est représentée sur la figure 3.

On définit

$$x^\star(\theta) = \underset{x\in[a(\theta);b(\theta)]}{\operatorname{argmin}} f(x,\theta) \tag{96}$$

$$f^\star(\theta) = f(x,\theta)\,, \quad x \in x^\star(\theta) \tag{97}$$

Lorsque deux points donnent la même valeur minimale globale de f, alors l'ensemble $x^*(\theta)$ contient ces deux points. Le but de l'exercice est d'étudier la régularité de la fonction $f^\star$.

1. (a) Justifier que, pour tout $\theta \in \mathbb{R}$, l'ensemble $x^\star(\theta)$ est non-vide.
   (b) Illustrer par un exemple (schéma) le fait que $x^\star(\theta)$ peut varier non continûment en θ.
2. Soit $\theta \in \mathbb{R}$. On considère une suite (θ_n) convergente de limite θ et (x_n) avec $x_n \in x^\star(\theta_n)$.
   (a) Justifier que l'on puisse considérer une sous-suite convergente de (x_n), dont on notera x la limite.
   (b) Montrer par l'absurde que $x \in x^\star(\theta)$ (on pourra exhiber une suite convergente de limite $\hat{x} \in x^\star(\theta)$, dont on justifiera l'existence).
   (c) Conclure que $f^\star$ est continue en θ.

10.9 Meilleur antécédent par une matrice non inversible

On cherche le meilleur antécédent d'un vecteur $b \in \mathbb{R}^n$ par une matrice A de taille $n \times n$ *non inversible*.

1. On considère le problème d'optimisation

   $$\min_u \|Au - b\|$$

   où $\|\cdot\|$ désigne la norme Euclidienne. Montrer que ce problème est convexe. En déduire que u réalise le minimum de la fonction objectif si et seulement si

   $$A^T A u = A^T b$$

2. La matrice A étant non inversible, montrer que $A^T A$ ne l'est pas non plus.
3. Montrer que la matrice $A^T A$ est symétrique positive. On note $\lambda_1, ..., \lambda_m$ ses m valeurs propres non nulles, avec $m < n$. En déduire que

   $$A^T A = \sum_{i=1}^{m} \lambda_i v_i v_i^T$$

   où v_i sont des vecteurs propres orthonormaux de $A^T A$.
4. On considère la matrice

   $$P = A^T A \left(\sum_{i=1}^{m} \frac{1}{\lambda_i} v_i v_i^T \right)$$

Montrer que rang $P = m$, $P^2 = P$, et que Im $P =$ Im A^T. En déduire que P est la projection orthogonale sur Im A^T.

5. En déduire que le vecteur suivant

$$u^* = \left(\sum_{i=1}^{m} \frac{1}{\lambda_i} v_i v_i^T \right) A^T b$$

réalise l'optimum recherché. Montrer que si A est inversible, i.e. $m = n$, alors $Au^* = b$.

10.10 Introduction aux méthodes de points intérieurs

On s'intéresse au problème de programmation quadratique (QP) suivant

$$\begin{aligned} \min_z \quad & z^T P z + f^T z \\ \text{tel que } & Mz \leq b \end{aligned} \tag{98}$$

où P est une matrice $n \times n$ symétrique définie positive, M une matrice $m \times n$, $f \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, avec $m > n$.

L'objet de cet exercice est d'exposer la construction d'un algorithme dit de "points intérieurs".

1. Montrer que le problème (98) admet une unique solution optimale.
2. On cherche à réécrire ce problème sous une forme canonique où les inégalités sont juste des relations de positivité des inconnues. Pour cela, on introduit un vecteur s de m inconnues positives ou nulles, et un vecteur y de $2n$ inconnues positives ou nulles. Chaque composante i de z est exprimable sous la forme

$$z_i = y_i - y_{i+n}, \quad y_i \geq 0, \quad y_{i+n} \geq 0$$

Montrer qu'on peut réécrire (98) sous la forme

$$\begin{aligned} \min_z \quad & \frac{1}{2} x^T Q x + c^T x \\ \text{tel que } & Ax = b, \\ & -x \leq 0 \end{aligned} \tag{99}$$

avec

$$x = \begin{pmatrix} y \\ s \end{pmatrix}, \quad Q = \begin{pmatrix} N^T P N & 0_{2n \times m} \\ 0_{m \times 2n} & 0_{m \times m} \end{pmatrix},$$

$$c^T = \begin{pmatrix} f^T N & 0_{1 \times m} \end{pmatrix}, \quad A = \begin{pmatrix} MN & I_m \end{pmatrix}$$

où N est une matrice qu'on explicitera.

3. Former le problème dual de (99) (on l'écrira comme un problème de maximisation sous contraintes). Pour cela, on formera le Lagrangien

$$L = \frac{1}{2} x^T Q x + c^T x + \lambda^T (b - Ax) - \mu^T x$$

4. Montrer que les conditions de Karush-Kuhn-Tucker pour le problème général

$$\begin{aligned} \min_x \quad & f(x) \\ \text{tel que } & d(x) = 0, \\ & e(x) \leq 0 \end{aligned}$$

sont

$$\begin{cases} \nabla f(x^*) + \sum \lambda_i^* \nabla d_i(x^*) + \sum \mu_i^* \nabla e_i(x^*) = 0 \\ d(x^*) = 0 \\ e(x^*) \leq 0 \\ \mu^* \geq 0 \\ \mu_i^* c_i(x^*) = 0 \end{cases} \tag{100}$$

5. Former les conditions (100) pour le problème (99).
6. On va pénaliser les contraintes $-x \leq 0$ dans le côut en considérant le problème, pour $\varepsilon > 0$,

$$\begin{aligned} \min_x \quad & \frac{1}{2} x^T Q x + c^T x - \varepsilon \sum \ln(x_i) \\ \text{tel que } & Ax = b \end{aligned} \tag{101}$$

Montrer que le problème (101) possède une unique solution.

7. Former le Lagrangien associé à (101) et montrer que les conditions de Karush-Kuhn-Tucker sont

$$\begin{cases} Ax = b \\ -Qx + A^T \lambda + s = c \\ x_i s_i = \varepsilon \\ x_i \geq 0 \\ s_i \geq 0 \end{cases} \tag{102}$$

8. Montrer comment résoudre les conditions (102) par une méthode de Newton avec projection. Former pour cela le Jacobien de la fonction à annuler. Former la première itération de cet algorithme en exhibant le calcul de la direction et la procédure de projection.
9. Comment atteindre la solution de (100)?
10. Une méthode itérative consiste à considérer $\epsilon = \sigma \frac{x^T s}{2n+m}, \quad \sigma \in]0,\ 1[$.
    Comment peut-on interpréter ce choix?

10.11 Polygones inscrits du cercle de longueur maximale

Soit $f : I \to \mathbb{R}$ avec I intervalle de $\mathbb{R}$. On suppose que f est convexe sur I, et on considère $\{x_i, i = 1, ...n\}$ une famille d'éléments de I, n entier $n > 1$.

1. Montrer (par exemple par récurrence) que
$$f(x_1) + ... + f(x_n) \geq nf(\frac{x_1 + ... + x_n}{n})$$
On a égalité si et seulement si tous les x_i sont égaux.
2. En utilisant l'inégalité précédente pour $f(x) = -\sin x$ et $I = [0,\ \pi]$, montrer que parmi les polygones à n côtés inscrits dans un cercle donné, les polygones réguliers ont le plus grand périmètre.

10.12 Parallélépipède maximal

Soit un ellipsoide défini par l'équation
$$c(x, y, z) = x^2/a^2 + y^2/b^2 + z^2/c^2 - 1 = 0$$
avec a, b, c des scalaires non nuls. On cherche le parallélépipède rectangle dont les arêtes sont parallèles aux axes ayant un volume maximal en restant contenu dans l'ellipsoide. On note x, y, z les demi-longueurs de ses arêtes, si bien que son volume est
$$f(x, y, z) = 8xyz$$
1. Écrire les conditions de stationnarité pour ce problème sous contrainte. On pourra noter λ le multiplicateur associé à la contrainte c.
2. En utilisant que le volume du parallélépipède solution est de volume non nul, montrer que λ est non nul.
3. Montrer que le volume optimal s'exprime directement en fonction de λ.
4. Éliminer λ des équations pour obtenir $x = a/\sqrt{3}$. En déduire que le volume optimal est $\frac{8abc}{3\sqrt{3}}$.

10.13 Convexité

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction quadratique strictement convexe. Soit x, y_1,...,y_n des vecteurs de $\mathbb{R}^n$.

1. Montrer que la fonction $\mathbb{R}^n \ni (\lambda_1, ..., \lambda_n) \mapsto f(x + \sum_{i=1}^n \lambda_i y_i)$ est également quadratique convexe.
2. Donner l'expression de son Hessien.
3. A quelle condition est-elle strictement convexe?

10.14 Projeté sur une parabole

Soit la parabole d'équation $5y = (x - 1)^2$. On cherche le point de la parabole de distance Euclidienne minimale avec le point de coordonnées $(1, 2)$.

1. Définir un problème de minimisation sous contraintes correspondant.
2. Trouver les points satisfaisant les conditions de Lagrange.
3. Déterminer la solution.

4. Procéder par substitution directe de y en fonction de x et formuler un problème de minimisation de la seule variable x. Retrouver ainsi le résultat de la question précédente.

10.15 Existence de minimums

On considère les trois problèmes suivants. Expliquer lesquels possèdent des solutions.

$$\min x_1 + x_2 \text{ sous les contraintes } x_1^2 + x_2^2 = 1,\ 0 \leq x_1 \leq \frac{1}{2},\ 0 \leq x_2 \tag{103}$$

$$\min x_1 + 2x_2 \text{ sous les contraintes } x_1^2 + 2x_2^2 \leq 1,\ x_1 + x_2 = 10 \tag{104}$$

$$\min x_1 \text{ sous les contraintes } x_1 + x_2 = 1 \tag{105}$$

10.16 Pénalité intérieure

On s'intéresse ici à une classe de méthodes permettant de résoudre un problème de minimisation sous contraintes comme limite d'une séquence de problèmes sans contraintes. On note

$$\min_x J(x), \quad \text{sous les contraintes} \quad g_i(x) \leq 0,\ i = 1...m$$

où on suppose que les fonctions J et g_i sont régulières de $\mathbb{R}^n \to \mathbb{R}$, et que les contraintes définissent un ensemble compact non vide S, contenant une unique solution globale x^*, non isolée dans S. On veut résoudre le problème en résolvant une séquence de problèmes sans contraintes dits "pénalisés" du type

$$\min_{x \in \mathbb{R}^n} J(x) + \epsilon \sum_{i=1}^{m} \gamma(g_i(x)), \quad \epsilon > 0$$

où γ est une fonction positive, régulière sur $]-\infty, 0[$, telle que $\lim_{x \to 0-} \gamma(x) = +\infty$. Par convention, $\gamma(x) = +\infty$, pour $x \geq 0$. On considère une suite strictement décroissante (ϵ_k) de réels strictement positifs et on note, pour tout k

$$J_k(x) = J(x) + \epsilon_k p(x) \triangleq J(x) + \epsilon_k \sum_{i=1}^{m} \gamma(g_i(x))$$

On suppose que J_k admet un minimum sur $\mathbb{R}^n$ en un point unique noté x_k.

1. Montrer que pour tout k, x_k est dans l'intérieur de S.
2. Montrer que pour tout k on a

$$J_k(x_k) \leq J_k(x_{k+1}) \tag{106}$$

$$J_{k+1}(x_{k+1}) \leq J_{k+1}(x_k) \tag{107}$$

3. Déduire, par une combinaison linéaire des lignes qui précèdent, que

$$J(x_{k+1}) \leq J(x_k) \tag{108}$$

4. Montrer, en utilisant (106) et la relation précédente que
$$p(x_k) \leq p(x_{k+1})$$
5. Montrer, en utilisant (107) que
$$J_{k+1}(x_{k+1}) \leq J_k(x_k) \tag{109}$$
6. Montrer que pour tout $\delta > 0$, il existe $x^\delta \in S$ tel que $J(x^\delta) < J(x^*) + \delta/2$
7. On choisit un entier l tel que
$$\epsilon_l p(x^\delta) < \delta/2$$
Montrer, en utilisant (109), qu'on a, pour tout $k > l$
$$J_k(x_k) \leq J_l(x^\delta)$$
8. En déduire que
$$J(x^*) \leq J(x_k) < J(x^*) + \delta$$
pour tout $k > l$ et que donc
$$\lim_{k\to\infty} J(x_k) = J(x^*)$$
9. Montrer enfin que (x_k) converge vers x^*.

La séquence des problèmes non contraints mais pénalisés génère une suite qui converge donc vers la solution x^* du problème contraint.

10.17 Convexité

1. Soit f une fonction convexe définie sur un ensemble convexe E. Montrer que l'ensemble
$$L_t = \{x \in E \text{ tel que } f(x) \leq t\}$$
est un ensemble convexe.
2. Soit g_i, $i = 1, ..., k$ des fonctions convexes définies sur $\mathbb{R}^n$. Montrer que l'ensemble
$$L = \{x \text{ tel que } g_i(x) \leq 0, i = 1, ..., k\}$$
est un ensemble convexe.
3. Soit f et g deux fonctions convexes $\mathbb{R} \to \mathbb{R}$. Montrer que $h(x, y) = f(x) + g(y)$ est convexe.

10.18 Réservoir cylindrique

On souhaite concevoir un réservoir cylindrique de contenance maximale au moyen d'une quantité limitée de matériau.

1. On note d le diamètre du disque de base du cylindre et h sa hauteur. Donner l'expression du volume V contenu dans le cylindre, et l'expression de sa surface S.

2. La surface du cylindre est constituée de plaques de tôle d'une épaisseur constante (négligeable). Ainsi la grandeur S est contrainte. Former le problème d'optimisation de la contenance sous contrainte $S = S^0$ et donner son Lagrangien.
3. Former les conditions de stationnarité et montrer que la hauteur optimale et le diamètre optimal sont égaux et valent $\sqrt{\frac{2S^0}{3\pi}}$.

10.19 Convexité de l'image de petites boules par application régulière

L'objet de cet exercice est de démontrer que l'image par une fonction régulière (à Jacobien inversible) d'une boule de taille suffisamment petite est convexe. Cette propriété a été établie dans des espaces de Hilbert généraux et est utilisée notamment dans l'étude des valeurs propres de matrices et d'opérateurs perturbés (voir [**25**]).

On établit d'abord un résultat préliminaire.

Soit une fonction $P : \mathbb{R}^n \to \mathbb{R}^m$ régulière. On cherche à résoudre l'équation

$$P(x) = 0 \tag{110}$$

au moyen d'une séquence

$$x^{n+1} = x^n - Q_n(x^n) \tag{111}$$

initialisée en x^0 donné. On suppose que

(i) P' le Jacobien de P est Lipschitzien avec comme constante $L > 0$
(ii) la fonction Q_n vérifie $\|P(x) - P'(x)Q_n(x)\| \leq \gamma \|P(x)\|$, avec $0 < \gamma < 1$
(iii) $\|Q_n(x)\| \leq \lambda \|P(x)\|$, pour un certain $\lambda \geq 0$,
(iv) $\gamma + \frac{L\lambda^2}{2} \left\|P(x^0)\right\| < 1$

On va montrer que la séquence (111) converge vers x^* solution de (110). Dans la suite, on utilise la norme Euclidienne et la norme induite.

1. Dans le cas où on cherche à résoudre un système linéaire $Ax = b$, avec $P(x) = Ax - b$, et A inversible, montrer que les hypothèses ci-dessus sont vérifiées. On précisera le choix retenu pour Q_n, les valeurs de L, γ.
2. En utilisant le développement (formule de Taylor), et les propriétés (i-ii-iii),

$$P(x + y) = P(x) + \int_0^1 P'(x + ty)ydt$$

entre les points x^n et x^{n+1}, montrer que

$$\left\|P(x^{n+1})\right\| \leq \gamma \|P(x^n)\| + \frac{L\lambda^2}{2} \|P(x^n)\|^2 \tag{112}$$

3. On note (δ^n) la suite de terme général $\gamma + \frac{L\lambda^2}{2} \|P(x^n)\|$. Déduire de (112) et de $\delta^0 < 1$ que $\delta^n \leq \delta^0$ pour tout n.

4. On note $c_1 = \frac{L\lambda^2\|P(x^0)\|}{2\gamma}$. Montrer, en majorant $\|P(x^n)\|$ que
$$\delta^n \leq \gamma \exp\left(c_1(\delta^0)^n\right)$$
5. En déduire que
$$\|P(x^n)\| \leq \left\|P(x^0)\right\| \gamma^n \exp\left(\frac{c_1}{1-\delta^0}\right)$$
6. Montrer, en utilisant (iii), que la suite x^n est une suite de Cauchy. En déduire qu'elle converge vers x^* solution de (110).
7. On suppose désormais que P' est uniformément inversible si bien que il existe $m > 0$ tel que quels que soient x (localement), et y on a
$$\|P'(x)y\| \geq m\|y\| \tag{113}$$
Soit z tel que $P'(x)z = P(x)$, déduire de ce qui précède, avec $Q(x) = \alpha z$, avec $0 < \alpha < 2$, que les hypothèses (ii) et (iii) sont vérifiées.
8. En déduire que si $\left\|P(x^0)\right\|$ est suffisamment petite, alors la séquence
$$x^{n+1} = x^n - \alpha^n z^n, \quad z^n = P'(x^n)^{-1}P(x^n) \tag{114}$$
avec $0 < \alpha_n < 2$ converge.
9. En déduire, sous l'hypothèse supplémentaire précédente, par passage à la limite, l'existence d'une solution x^* de (110) et que
$$\left\|x^* - x^0\right\| \leq m\left\|P(x^0)\right\|$$

On considère maintenant une fonction régulière $f : \mathbb{R}^n \to \mathbb{R}^m$ telle que son Jacobien est Lipschitzien de constante L et inversible avec la constante m. On veut montrer que l'ensemble $F = \{f(x) \text{ pour } x \in B(a,\epsilon)\}$, avec $B(a,\epsilon) = \{x \text{ tel que } \|x-a\| \leq \epsilon\}$ est un ensemble convexe pour tout a, pour ϵ suffisamment petit.

Soient deux points x^1, x^2 de $B(a,\epsilon)$ et y^1, y^2 de F leurs images par f. Soit $y^0 = \frac{1}{2}(y^1+y^2)$. Pour montrer que F est convexe, il suffit de montrer que y^0 appartient à F.

10. On note $x^0 = \frac{1}{2}(x^1 + x^2)$. Montrer que
$$y^i = f(x^0) + f'(x^0)(x^i - x^0) + \epsilon_i, \quad i = 1,2$$
avec
$$\|\epsilon_i\| \leq \frac{L}{8}\left\|x^1 - x^2\right\|^2$$
11. Montrer que
$$y^0 = f(x^0) + \epsilon^0, \quad \left\|\epsilon^0\right\| \leq \frac{L}{8}\left\|x^1 - x^2\right\|^2$$
12. La fonction $x \mapsto f(x) - y^0$ est telle que son Jacobien est Lipschitzien de constante L et inversible avec la constante m. En déduire, d'après la première partie de l'exercice, que si ϵ est suffisamment petit, alors il existe x^* tel que $f(x^*) = y^0$ et tel que
$$\left\|x^* - x^0\right\| \leq m\left\|f(x^0) - y^0\right\|$$
13. Montrer alors que pour ϵ suffisamment petit, $x^* \in B(a,\epsilon)$.

14. Conclure que l'image par f de toute boule de taille ϵ est convexe pour ϵ suffisamment petit.

Application: On considère $f(x) = \begin{pmatrix} x_1x_2 - x_1 \\ x_1x_2 + x_2 \end{pmatrix}$.

15. Montrer que l'image par f de la boule $B(0,\epsilon)$ est convexe pour ϵ suffisamment petit (un calcul plus avancé montre que $\epsilon \leq \frac{1}{2\sqrt{2}}$ est suffisant). La situation est illustrée en figure 4.

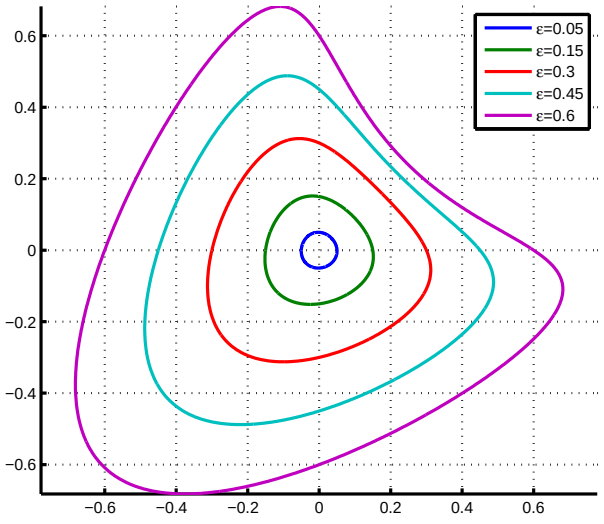


FIGURE 4. Image par f de la boule $B(0,\epsilon)$ pour différentes valeur de ϵ.

10.20 Programmation linéaire robuste

Un problème de programmation linéaire s'écrit (canoniquement) sous la forme

$$\begin{aligned} &\min_x c^T x \\ &\text{t.q.} \quad a_i^T x \leq b_i, \quad i = 1, ..., m \end{aligned} \tag{115}$$

où $c \in \mathbb{R}^n$, et pour $i = 1, ..., m > n$, $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$. On s'intéresse dans cet exercice à un problème plus général, où chaque contrainte i est définie par des ensembles E_i et F_i pour $i = 1, ..., m > n$,

$$\begin{aligned} &\min_x c^T x \\ &\text{t.q.} \quad a_i^T x \leq b_i, \quad a_i \in E_i, \quad b_i \in F_i, \quad i = 1, ..., m \end{aligned} \tag{116}$$

Les ensembles F_i et E_i sont polyhédraux (supposés non vides), c.-à-d.

$$E_i = \{a \mid D_i a \leq d_i\} \subset \mathbb{R}^n, \quad F_i = [\gamma_i, \mu_i] \subset \mathbb{R}$$

avec D_i une matrice $k_i \times n$, et $d_i \in \mathbb{R}^{k_i}$ avec $k_i > 1$ et $\gamma_i < \mu_i$. On cherche à montrer que le problème (116) est bien un problème du type (115) et à en trouver l'expression.

1. Montrer que pour tout i, l'ensemble E_i est convexe.
2. Montrer que (116) est équivalent au problème

$$\begin{aligned} &\min_x c^T x \\ &\text{t.q.} \quad a_i^T x \leq \gamma_i, \quad a_i \in E_i, \quad i = 1, ..., m \end{aligned} \tag{117}$$

3. Ré-écrire les contraintes de (117) en utilisant le problème auxiliaire

$$\begin{aligned} &\max_{a_i} a_i^T x \\ &\text{t.q.} \quad D_i a_i \leq d_i \end{aligned} \tag{118}$$

4. On va considérer le problème dual de (118). Former le Lagrangien $\mathcal{L}(a_i, \lambda)$ de (118). Calculer $\sup_{\lambda \geq 0} \mathcal{L}$, en distinguant le cas $D_i a_i \leq d_i$ et son complémentaire. De même, calculer $\inf_{a_i} \mathcal{L}$. Utiliser l'égalité du point selle. En supposant que inf et sup sont atteints, montrer que le problème dual de (118) s'écrit

$$\begin{aligned} &\min_{\lambda \geq 0} d_i^T \lambda \\ &\text{t.q.} \quad D_i^T \lambda = x \end{aligned} \tag{119}$$

5. En utilisant cette formulation duale, montrer que la résolution de (117) revient à la résolution de

$$\begin{aligned} &\min_x c^T x \\ &\left\{ \begin{array}{l} \min_{\lambda_i \geq 0} \lambda_i^T d_i \\ D_i^T \lambda_i = x \end{array} \right\} \leq \gamma_i \qquad i = 1, ..., m \end{aligned} \tag{120}$$

6. Conclure que le problème (116) est équivalent au problème de programmation linéaire suivant

$$\begin{aligned} &\min_{x, (\lambda_i)_{i=1,\dots,m}} c^T x \\ &\left\{ \begin{array}{ll} \lambda_i^T d_i \leq \gamma_i, & i = 1, ..., m, \\ D_i^T \lambda_i = x & i = 1, ..., m, \\ \lambda_i \geq 0 & i = 1, ..., m \end{array} \right. \end{aligned} \tag{121}$$

10.21 Dualité de Fenchel-Rockafellar

On considère deux fonctions convexes $f, g : \mathbb{R}^n \to \mathbb{R}$. Le but de cet exercice est de démontrer le principe de dualité de Fenchel-Rockafellar qui affirme que

$$\inf_{x \in \mathbb{R}^n} [f(x) + g(x)] = \sup_{\varphi \in \mathbb{R}^n} [-f^\star(\varphi) - g^\star(-\varphi)] \tag{122}$$

où $f^\star, g^\star$ sont les transformées de Fenchel de f et g. Pour ce faire, on note

$$p = \inf_{x\in\mathbb{R}^n} [f(x) + g(x)] \tag{123}$$

$$d = \sup_{\varphi\in\mathbb{R}^n} [-f^\star(\varphi) - g^\star(-\varphi)] \tag{124}$$

1. Montrer que, pour tout $(x, \varphi) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$f(x) + g(x) \geq -f^\star(\varphi) - g^\star(-\varphi) \tag{125}$$

   En déduire $p \geq d$.
2. On définit la fonction valeur

$$V(x) = \inf_{s\in\mathbb{R}^n} [f(s) + g(s + x)] \ , \quad x \in \mathbb{R}^n \tag{126}$$

   (a) Montrer que la fonction V est convexe.
   (b) Prouver que $p = V(0) = -V^\star(\varphi)$, pour tout $\varphi \in \partial V(0)$.
   (c) Montrer que $V^\star(\varphi) = f^\star(-\varphi) + g^\star(\varphi)$ pour tout $\varphi \in \mathbb{R}^n$.
   (d) En déduire $p \leq d$.
3. Conclure. Dans quel cas les inf et sup de (123)–(124) sont-ils atteints ?
4. **Application.** On considère le problème d'optimisation

$$\min_{\|x\|_2\leq 1} \frac{1}{2}\|Ax - b\|_2^2 \tag{127}$$

   où A matrice de taille $p \times n$ de rang plein ($p > n$), $b \in \mathbb{R}^p$ et $\|\cdot\|_2$ est la norme euclidienne. On définit

$$f(x) = \|Ax - b\|_2^2 \tag{128}$$

$$g(x) = \begin{cases} 0 & \text{si } \|x\|_2 \leq 1 \\ +\infty & \text{sinon} \end{cases} \tag{129}$$

   (a) Montrer que les fonctions f et g ci-dessus permettent de reformuler le problème (127) sous la forme (123).
   (b) Montrer que $g^* = \|\cdot\|_2$.
   (c) En déduire que (127) est équivalent à un problème de la forme

$$\min_{\varphi\in\mathbb{R}^n} \varphi^T Q\varphi + c^T\varphi + \|\varphi\|_2 + d \tag{130}$$

   Comment peut-on résoudre ce problème ?

10.22 Méthode de faisceaux proximale

On se propose dans cet exercice d'étudier des variantes de la méthode des faisceaux pour résoudre le problème

$$\min_{x\in\mathbb{R}^n} f(x) \tag{131}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est supposée continue et convexe.

Partie 1 : Instabilité de la méthode des faisceaux

On rappelle le principe de base des méthodes de faisceaux : on suppose que l'on dispose d'un oracle qui, pour tout $x \in \mathbb{R}$, fournit $f(x)$ et $g \in \partial f(x)$. Etant donné un faisceau d'informations $\{(x_i, f(x_i), g_i) \mid g_i \in \partial f(x_i)\,, i = 0, \ldots, k\}$, on construit une approximation affine par morceaux de f et l'on résout le problème

$$\min_{x\in\mathbb{R}^n} \varphi_k(x) \quad \text{avec} \quad \varphi_k(x) = \max_{i=0,\ldots,k} \left\{ f(x_i) + g_i^T(x - x_i) \right\} \tag{132}$$

dont on note x_{k+1} la solution. On met alors à jour le faisceau d'informations en y ajoutant la nouvelle donnée $(x_{k+1}, f(x_{k+1}), g_{k+1})$ $(g_{k+1} \in \partial f(x_{k+1}))$ et on itère.

1. **Application.** Soient $f : x \in \mathbb{R} \to \frac{1}{2}x^2$, $x_0 = 1$ et $x_1 = -\varepsilon$ avec $\varepsilon < 1/2$ un petit paramètre positif.
   (a) Construire la fonction φ_1.
   (b) Déterminer $x_2 = \text{argmin}\varphi_1$.
   (c) Comparer x_1 et x_2 au minimum de la fonction f. Quel inconvénient présente donc cette méthode ? Que se passe-t-il quand on fait tendre ε vers 0 ?

On considère maintenant une méthode alternative qui, au lieu de résoudre le problème (132), résout

$$\min_{x\in\mathbb{R}^n} \varphi_k(x) + \frac{1}{2\mu} \|x - x_k\|^2 \tag{133}$$

où $\mu > 0$ et $\|\cdot\|$ est la distance euclidienne.

2. Justifier que cette méthode porte le nom de méthode de faisceaux proximale. Quel est a priori l'effet du terme $\frac{1}{2\mu} \|x - x_k\|^2$ selon vous ?
3. **Application.** On reprend l'exemple de la question 1 et on considère maintenant le problème (133) avec $\mu = 1/3$.
   (a) Déterminer à présent la solution x_2 du nouveau problème (133).
   (b) Comparer ce nouvel x_2 à x_1 et au minimum de f. Conclure.

Partie 2 : Analyse de convergence

Afin d'assurer que l'on ne réalise pas des itérations qui ne font pas assez décroître la fonction, on modifie légèrement la méthode ci-dessus. L'algorithme de méthode de faisceaux proximale prend alors la forme finale suivante.

***Algorithme* 18**. — Soient $m \in]0, 1[$ et une suite (μ_k) de réels strictement positifs. A partir de $x_0 \in \mathbb{R}^n$ quelconque, $f(x_0)$, $s_0 \in \partial f(x_0)$, de $\varphi_0 : x \in \mathbb{R}^n \mapsto f(x_0) + s_0^T(x - x_0)$ et de $K = \emptyset$, itérer

1. Résoudre $\min_{x\in\mathbb{R}^n} \varphi_k(x) + \frac{1}{2\mu_k}\|x - x_k\|^2$, de minimiseur y_{k+1}.
2. Calculer $f(y_{k+1})$, $s_{k+1} \in \partial f(y_{k+1})$ et

$$\delta_k = f(x_k) - \varphi_k(y_{k+1}) - \frac{1}{2\mu_k}\|y_{k+1} - x_k\|^2$$

3. Si $f(y_{k+1}) \leq f(x_k) - m\delta_k$, choisir $x_{k+1} = y_{k+1}$ et $K = K \cup \{k\}$. Sinon, $x_{k+1} = x_k$.
4. Définir

$$\varphi_{k+1}(x) = \max\left\{\varphi_k(x), f(y_{k+1}) + s_{k+1}^T(x - y_{k+1})\right\}$$

On cherche maintenant à étudier les propriétés de convergence de cet algorithme. Pour ce faire, on se restreint au cas où K est de cardinal infini et on suppose que la suite (μ_k) est telle que

$$\sum_{k\in K} \mu_k = +\infty \tag{134}$$

Enfin, on suppose que la fonction convexe f est bornée inférieurement.

4. Justifier que $f(x) \geq \varphi_k(x)$ pour tout $x \in \mathbb{R}^n$ et $k \in \mathbb{N}$.
5. Montrer que

$$\sum_{k\in K} \delta_k \leq \frac{1}{m} \sum_{k=1}^{\infty} [f(x_k) - f(x_{k+1})] \tag{135}$$

Prouver que $\delta_k \geq 0$ et conclure que $\lim_{k\to\infty, k\in K} \delta_k = 0$.

6. Montrer que $\frac{1}{\mu_k}(x_k - y_{k+1}) \in \partial\varphi_k(y_{k+1})$. En déduire

$$\forall y \in \mathbb{R}^n \quad \varphi_k(y) \geq \varphi_k(y_{k+1}) + \frac{1}{\mu_k}(x_k - y_{k+1})^T(y - x_k) + \frac{1}{\mu_k}\|x_k - y_{k+1}\|^2 \tag{136}$$

7. En conclure que, si $k \in K$,

$$\forall y \in \mathbb{R}^n \quad \|y - x_{k+1}\|^2 \leq \|y - x_k\|^2 + 2\mu_k(f(y) - f(x_k) + \delta_k) \tag{137}$$

8. On suppose qu'il existe $\eta > 0$ et $z \in \mathbb{R}^n$ tels que $f(z) \leq f(x_k) - \eta$ pour tout $k \in \mathbb{N}$. Montrer qu'il existe $k_0 \in \mathbb{N}$ tel que

$$\|z - x_{k+1}\|^2 \leq \|z - x_k\|^2 - \mu_k\eta\,, \quad k_0 \leq k \in K \tag{138}$$

et obtenir une contradiction à l'aide de (134). Conclure.

10.23 Programmation sur un cône

On considère le problème de programmation sur un cône

$$\begin{aligned} &\min_x \quad \alpha^T x \\ &\text{tel que } \|A_i x + b_i\| \leq c_i^T x + d_i\,, \quad i = 1, \ldots, N \end{aligned} \tag{139}$$

où $x \in \mathbb{R}^n$ est la variable de décision et les paramètres du problème sont $\alpha \in \mathbb{R}^n$, $A_i \in \mathcal{M}_{n_i,n}(\mathbb{R})$, $b_i \in \mathbb{R}^{n_i}$, $c_i \in \mathbb{R}^n$ et $d_i \in \mathbb{R}$. La norme $\|\cdot\|$ représente la norme euclidienne dans $\mathbb{R}^{n_i}$

Partie 1 : Applications

On va dans un premier temps illustrer quelles familles de problème peuvent être traitées par un tel formalisme.

1. On s'intéresse au problème de programmation quadratique

$$\begin{aligned} &\min_{z} \quad z^T Q z + 2 r^T z \\ &\text{tel que } Mz + p \leq 0 \end{aligned} \tag{140}$$

où Q matrice symétrique définie positive.

(a) Montrer que le problème (140) est équivalent à

$$\begin{aligned} &\min_{t,\, z} \quad t + r^T Q^{-1} r \\ &\text{tel que } z^T Q z + 2 r^T z \leq t, \\ &\qquad Mz + p \leq 0 \end{aligned} \tag{141}$$

(b) On rappelle qu'une matrice symétrique définie positive Q admet une unique racine carrée R (symétrique définie positive) telle que $Q = R^2$. Reformuler le problème ci-dessus sous la forme

$$\begin{aligned} &\min_{t,\, z} \quad t + r^T Q^{-1} r \\ &\text{tel que } \|Rz + R^{-1} r\|^2 \leq t + r^T Q^{-1} r, \\ &\qquad MZ + p \leq 0 \end{aligned} \tag{142}$$

(c) En introduisant $x = (z^T, \sqrt{t + r^T Q^{-1} r})^T$, conclure que (140) est équivalent à la résolution d'un problème de la forme (139).

2. On s'intéresse maintenant au problème de programmation linéaire robuste

$$\begin{aligned} &\min_{x} \quad c^T x \\ &\text{tel que } a_i^T x + b_i \leq 0\,,\ a_i \in \mathcal{E}_i\,, \quad i = 1, \ldots, m \end{aligned} \tag{143}$$

où les paramètres a_i sont incertains et appartiennent à des ellipsoïdes donnés

$$a_i \in \mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\| \leq 1\} \tag{144}$$

avec P_i symétrique positive.

(a) Illustrer graphiquement les contraintes du problème, dans le cas $n = 2$ et $m = 1$.

(b) Montrer que le problème (143) est équivalent à

$$\begin{aligned} &\min_{x} \quad c^T x \\ &\text{tel que } \max\left\{a_i^T x \mid a_i \in \mathcal{E}_i\right\} + b_i \leq 0\,, \quad i = 1, \ldots, m \end{aligned} \tag{145}$$

(c) Exprimer $\max\left\{a_i^T x \mid a_i \in \mathcal{E}_i\right\}$ en fonction de $\bar{a}_i$ et P_i puis reformuler le problème sous la forme d'une programmation sur un cône (139).

Partie 2 : Résolution

On se propose maintenant d'étudier la résolution de ces problèmes par une méthode primale-duale de points intérieurs.

3. Montrer que le problème (139) est convexe.
4. Justifier que le gradient de $g : x \in \mathbb{R}^n \mapsto \|A_i x + b_i\|$ est $\nabla g(x) = \frac{A_i^T(A_i x + b_i)}{\|A_i x + b_i\|}$.
5. Montrer ainsi que le problème dual correspondant à (139) s'écrit

$$\begin{aligned} &\max && -\sum_{i=1}^{N}(b_i^T z_i + d_i\lambda_i) \\ &\text{sous contraintes} && \begin{cases} \sum_{i=1}^{N}(A_i^T z_i + c_i\lambda_i) = \alpha \\ \|z_i\| \le \lambda_i\,, \quad i = 1,\dots,N \end{cases} \end{aligned} \tag{146}$$

6. On définit le saut de dualité associé à la différence entre le problème primal et le problème dual comme

$$\eta(x, z, \lambda) = \alpha^T x + \sum_{i=1}^{N}(b_i^T z_i + d_i\lambda_i) \tag{147}$$

où x solution de (139) et (z, λ) solution de (146).
   (a) Montrer que $\eta(x, z, \lambda) = \sum_{i=1}^{N}\left[(A_i x + b_i)^T z_i + (c_i^T x + d_i)\lambda_i\right]$.
   (b) En déduire $\eta(x, z, \lambda) \ge 0$.
   (c) Que peut-on conclure si $\eta(x, z, \lambda) = 0$?
7. On définit la barrière logarithmique

$$\gamma(u, t) = \begin{cases} -\ln(t - \|u\|^2) & \text{si } t > \|u\|^2 \\ +\infty & \text{sinon} \end{cases} \tag{148}$$

Pour toutes solutions x et (z, λ) strictement faisables de (139) et (146), cad respectant strictement les contraintes inégalités des problèmes respectifs, on considère la fonction

$$\psi(x, z, \lambda) = (2N + 1)\ln(\eta(x, z, \lambda)) + \sum_{i=1}^{N}[\gamma(u_i, t_i) + \gamma(z_i, \lambda_i)] - 2N\ln N \tag{149}$$

avec $u_i = A_i x + b_i$ et $t_i = c_i^T x + d_i$. On admet le résultat suivant :

$$\eta(x, z, \lambda) \le \exp\left(\psi(x, z, \lambda)\right) \tag{150}$$

et on se propose de minimiser la fonction ψ à partir d'un point initial (x^0, z^0, λ^0) stritement faisable. Justifier que cette méthode permet de résoudre le problème (139) et porte le le nom de "méthode primale-duale par points intérieurs".

10.24 Optimisation robuste

On se propose dans cet exercice d'étudier la sensibilité d'une classe de problèmes d'optimisation à des incertitudes sur ses paramètres et de présenter des méthodologies de gestion dédiées.

Partie 1 : Etude de sensibilité

Soit le problème de programmation dynamique quadratique

$$\min_{Ax \leq b} \frac{1}{2} x^T Q x + r^T x \tag{151}$$

où $x \in \mathbb{R}^n$, Q matrice symétrique définie positive, $r \in \mathbb{R}^n$, A matrice $p \times n$ et $b \in \mathbb{R}^p$.

1. Justifier l'existence d'un unique minimum global pour ce problème, que l'on notera x^*.
2. Notons $\mathcal{I}(x^*)$ l'ensemble des contrainstes actives en x^* dont on suppose qu'elles sont au nombre de $m \leq p$. Justifier qu'il existe $\lambda \in \mathbb{R}_+^m$ tel que

$$Qx^* + r + A_{\mathcal{I}}^T \lambda = 0 \tag{152}$$

$$A_{\mathcal{I}} x^* = b_{\mathcal{I}} \tag{153}$$

   où $A_{\mathcal{I}} = (l_i)_{i \in \mathcal{I}(x^*)}$ matrice $m \times n$ composée des m lignes de A correspondant aux contraintes actives et $b_{\mathcal{I}} = (b_i)_{i \in \mathcal{I}(x^*)}$ vecteur de $\mathbb{R}^m$.
3. On suppose que $A_{\mathcal{I}}$ est de rang m. Montrer que

$$\lambda = -\left(A_{\mathcal{I}} Q^{-1} A_{\mathcal{I}}^T\right)^{-1} \left(A_{\mathcal{I}} Q^{-1} r + b_{\mathcal{I}}\right) \tag{154}$$

$$x^* = Q^{-1}\left(-r + A_{\mathcal{I}}^T \left(A_{\mathcal{I}} Q^{-1} A_{\mathcal{I}}^T\right)^{-1} \left(A_{\mathcal{I}} Q^{-1} r + b_{\mathcal{I}}\right)\right) \tag{155}$$

On rappelle que, si M matrice $m \times n$ est de rang m, alors $M^T M$ est inversible.

On cherche maintenant à caractériser la sensibilité de cette solution par rapport à une incertitude portant sur le vecteur b. On considère donc une variation infinitésimale de b de la forme $b + \delta$ avec $\delta \ll 1$ et le problème perturbé

$$\min_{Ax \leq b+\delta} \frac{1}{2} x^T Q x + r^T x \tag{156}$$

On admet que la solution de ce problème est continue par rapport à δ (résultat connu sous le nom de Théorème du maximum de Berge).

4. Justifier que λ est une fonction continue de x^*.
5. On suppose que la solution x^* du problème (151) est une solution dite non-dégénérée, cad telle que $\lambda_i > 0$ pour tout $i \in \mathcal{I}(x^*)$. Justifier que la solution de (156) possède le même ensemble de contraintes actives $\mathcal{I}(x^*)$.
6. Proposer un algorithme permettant de fournir la solution au problème (156) à partie de celle supposée non dégénérée du problème (151).

Partie 2 : Programmation robuste

On cherche maintenant à inclure les incertitudes dans le problème d'origine et l'on s'intéresse ainsi à

$$\begin{aligned} &\min \frac{1}{2}x^T Q x + r^T x \\ &\text{t.q.} \quad a_i^T x \leq b_i\,, \quad (a_i, b_i) \in E_i \times F_i\,, \quad i = 1, \ldots, p \end{aligned} \tag{157}$$

où $x \in \mathbb{R}^n$, Q est une matrice symétrique définie positive, $r \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ et les ensembles E_i et F_i sont des polyhèdres non-vides, cad

$$E_i = \{a \mid D_i a \leq d_i\} \subset \mathbb{R}^n\,, \quad F_i = [\gamma_i, \delta_i] \subset \mathbb{R} \tag{158}$$

avec D_i matrice de taille $k_i \times n$, $d_i \in \mathbb{R}^{k_i}$ (avec $k_i > 1$) et $\gamma_i < \delta_i$. Dans la suite du problème, on va reformuler ce problème sous une forme plus usuelle et plus simple à résoudre.

1. Montrer que (157) est équivalent au problème

$$\begin{aligned} &\min \frac{1}{2}x^T Q x + r^T x \\ &\text{t.q.} \quad a_i^T x \leq \gamma_i\,, \quad a_i \in E_i\,, \quad i = 1, \ldots, p \end{aligned} \tag{159}$$

2. Réécrire les contraintes du problème (159) à l'aide du problème auxiliaire

$$\begin{aligned} &\max_a a^T x \\ &\text{t.q.} \quad D_i a \leq d_i \end{aligned} \tag{160}$$

3. On cherche à formuler le problème dual de (160).
   (a) Former le lagrangien $\mathcal{L}(a, \lambda)$ correspondant à (160).
   (b) Calculer $\sup_{\lambda \geq 0} \mathcal{L}(a, \lambda)$ en distinguant le cas $D_i a \leq d_i$ et son complémentaire.
   (c) En utilisant de même une disjonction de cas, calculer $\inf_a \mathcal{L}(a, \lambda)$.
   (d) Utiliser l'égalité du point selle pour montrer que le problème dual de (160) s'écrit

$$\begin{aligned} &\min_{\lambda \geq 0} d_i^T \lambda \\ &\text{t.q.} \quad D_i^T \lambda = x \end{aligned} \tag{161}$$

4. En déduire que (157) est équivalent au problème de programmation quadratique

$$\min_{x, (\lambda_i)_{i=1,\ldots,p}} \frac{1}{2}x^T Q x + r^T x \tag{162}$$

$$\text{t.q.} \quad \begin{cases} d_i^T \lambda_i \leq \gamma_i \\ D_i^T \lambda_i = x \\ \lambda_i \geq 0 \end{cases}, \quad i = 1, \ldots, p \tag{163}$$

10.25 Méthode de gradient proximale accélérée

La méthode de gradient proximale présente l'inconvénient d'avoir une vitesse de convergence assez faible (convergence sous-linéaire). L'une des ses variantes, connues sous le nom de gradient proximal accéléré, permet d'améliorer ce point.

Soit le problème de minimisation

$$\min_{x\in\mathbb{R}^n} [f(x) = g(x) + h(x)] \tag{164}$$

où $g, h : \mathbb{R}^n \to \mathbb{R}$ sont des fonctions convexes avec g différentiable de gradient L-Lipschitz. L'algorithme de gradient proximal accéléré à pas constant est alors le suivant.

***Algorithme* 19**. — Choisir $l > 0$. A partir de x_0 dans $\mathbb{R}^n$, $y_1 = x_0$ et $t_1 = 1$, itérer pour $k \geq 1$

$$x_k = \mathrm{Prox}_{lh}\left(y^k - l\nabla g(y^k)\right) \tag{165}$$

$$t_{k+1} = \frac{k+2}{2} \tag{166}$$

$$\beta_{k+1} = \frac{t_k - 1}{t_{k+1}} \tag{167}$$

$$y^{k+1} = x^k + \beta_{k+1}\left(x^k - x^{k-1}\right) \tag{168}$$

1. **Application.** On considère $g : x \in \mathbb{R} \mapsto \frac{1}{2}x^2$ et $h = g$.
   (a) Calculer l'opérateur Prox_{lh} et calculer x_1, x_2 et x_3 obtenus à partir de $x_0 \in \mathbb{R}$ quelconque avec les trois premières itérations de l'algorithme de gradient proximal usuel.
   (b) Calculer x_1, x_2 et x_3 obtenus à partir de $x_0 \in \mathbb{R}$ quelconque avec les trois premières itérations de l'Algorithme 1 de gradient proximal accéléré ci-dessus.
   (c) Conclure sur l'effet du gradient proximal accéléré pour l suffisamment faible.

On étudie maintenant le taux de convergence de cet algorithme.

2. On note $r_l(x) = \mathrm{Prox}_{lh}(x - l\nabla g(x))$. Dans un premier temps, on cherche à établir la propriété suivante pour $l \leq 1/L$:

$$\forall (x,z) \in \mathbb{R}^n \times \mathbb{R}^n \quad f(z) \geq f(r_l(x)) + \frac{1}{l}(r_l(x) - x)^T(x - z) + \frac{1}{2l}\|r_l(x) - x\|^2 \tag{169}$$

   (a) En utilisant la convexité de g et le fait que ∇g est L-lipschitz, montrer que, pour $l \leq 1/L$ et $(x, z) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$g(r_l(x)) \leq g(z) + \nabla g(x)^T(r_l(x) - z) + \frac{1}{2l}\|r_l(x) - x\|^2 \tag{170}$$

   (b) Prouver que

$$\forall x \in \mathbb{R}^n \quad x - l\nabla g(x) - r_l(x) \in l\partial h(r_l(x)) \tag{171}$$

En déduire que

$$\forall (x,z) \in \mathbb{R}^n \times \mathbb{R}^n \quad h(z) \geq h(r_l(x)) + \frac{1}{l}(x - r_l(x) - l\nabla g(x))^T (z - r_l(x))$$

(c) Conclure.

3. On souhaite maintenant montrer que, pour $k \geq 1$,

$$2lt_k^2 v_k - 2lt_{k+1}^2 v_{k+1} \geq \|u_{k+1}\|^2 - \|u_k\|^2 \tag{172}$$

où $v_k = f(x_k) - f(x^*)$ et $u_k = t_k x_k - (t_k - 1)x_{k-1} - x^*$ avec x^* minimiseur de f.

(a) En appliquant (169) respectivement à $(x,z) = (y_{k+1}, x_k)$ et $(x,z) = (y_{k+1}, x^*)$ avec x^* minimum de f, montrer que

$$2l(v_k - v_{k+1}) \geq \|x_{k+1} - y_{k+1}\|^2 + 2(x_{k+1} - y_{k+1})^T (y_{k+1} - x_k) \tag{173}$$

$$-2lv_{k+1} \geq \|x_{k+1} - y_{k+1}\|^2 + 2(x_{k+1} - y_{k+1})^T (y_{k+1} - x^*) \tag{174}$$

(b) En déduire que

(175)

$$2l((t_{k+1} - 1)v_k - t_{k+1}v_{k+1}) \geq t_{k+1}\|x_{k+1} - y_{k+1}\|^2 + 2(x_{k+1} - y_{k+1})^T (t_{k+1}y_{k+1} - (t_{k+1} - 1)x_k - x^*)$$

(c) Montrer que $t_{k+1}(t_{k+1} - 1) \leq t_k^2$, $k \geq 1$. En déduire que

(176)

$$2l(t_k^2 v_k - t_{k+1}^2 v_{k+1}) \geq \|t_{k+1}(x_{k+1} - y_{k+1})\|^2 + 2t_{k+1}(x_{k+1} - y_{k+1})^T (t_{k+1}y_{k+1} - (t_{k+1} - 1)x_k - x^*)$$

(d) Conclure.

4. Conclure que

$$f(x_k) - f(x^*) \leq \frac{2}{l(k+1)^2}\|x_0 - x^*\|^2 \tag{177}$$

10.26 Autour de méthodes de pénalité

Dans ce problème, on se propose d'aborder plusieurs méthodes de résolution par pénalité pour des problème d'optimisation sous contraintes.

Partie 1: Méthode de points intérieurs

On s'intéresse ici au problème de minimisation suivant

$$\begin{aligned} &\min_x \quad f(x) \\ &\text{tel que } c(x) \leq 0 \end{aligned} \tag{178}$$

avec $f : \mathbb{R}^n \to \mathbb{R}$ et $c : \mathbb{R}^n \to \mathbb{R}^m$ des fonctions continues, $(n,m) \in \mathbb{N}^2$. On note $R = \{x \in \mathbb{R}^n \mid c(x) \leq 0\}$ dont on suppose qu'il s'agit d'un ensemble fermé borné non-vide, égal à l'adhérence de son intérieur, cad $R = \overline{R^0}$.

1. Justifier que le problème (178) admet au moins une solution, notée x^*.

On suppose que cette solution $x^\star$ est unique. On cherche à résoudre ce problème en résolvant une suite de problèmes pénalisés

$$\min_{x\in\mathbb{R}^n} f_k(x) \stackrel{\Delta}{=} f(x) + \varepsilon_k p(x) \quad \text{avec} \quad p(x) = \sum_{i=1}^{m} \gamma(c_i(x)) \tag{179}$$

où $\gamma : \mathbb{R} \to \mathbb{R}_+$ est une fonction régulière sur $]-\infty, 0[$, telle que $\lim_{s\to 0^-} \gamma(s) = +\infty$ et, par convention, $\gamma(s) = +\infty$ pour $s \geq 0$. $(\varepsilon_k)_{k\in\mathbb{N}}$ est une suite de réels strictement positifs, strictement décroissante et convergeant vers zéro . On suppose que (179) admet une solution x_k.

2. Montrer que $x_k \in R^0$ pour tout $k \in \mathbb{N}$.
3. Justifier que $f_{k+1}(x_{k+1}) \leq f_{k+1}(x_k)$. En déduire $f_{k+1}(x_{k+1}) \leq f_k(x_k)$.
4. Montrer que, pour tout $\delta > 0$, il existe $x^\delta \in R$ tel quel $f(x^\delta) < f(x^*) + \delta/2$.

On considère $K \in \mathbb{N}$ tel que $\varepsilon_k p(x^\delta) < \delta/2$ pour $k \geq K$.

5. Déduire de la question 3 que, pour tout $k \geq K$, $f_k(x_k) \leq f_K(x^\delta)$.
6. En déduire que $f(x^*) \leq f(x_k) < f(x^*) + \delta$ pour tout $k \geq K$. Conclure que $\lim_{k\to+\infty} f_k(x_k) = f(x^*)$.
7. Conclure également que la suite x_k converge vers x^*.
8. Montrer que l'étude précédente reste inchangée pour le problème

$$\begin{gathered} \min_x \quad f(x) \\ \text{tel que } c(x) \leq 0, \\ x \in C \end{gathered}$$

où C est un fermé borné de $\mathbb{R}^n$ tel que $C \cap R$ est égal à l'adhérence de son ensemble.

Partie 2: Slack variables pour contraintes égalités

On s'intéresse maintenant au problème de minimisation suivant

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{sous les contraintes} \quad c^{in}(x) \leq 0 \text{ et } c^{eq}(x) = 0 \tag{180}$$

avec $f : \mathbb{R}^n \to \mathbb{R}$, $c^{in} : \mathbb{R}^n \to \mathbb{R}^{m_1}$ et $c^{eq} : \mathbb{R}^n \to \mathbb{R}^{m_2}$ des fonctions régulières non-nulles ($(n, m_1, m_2) \in \mathbb{N}^3$ et $m_2 < n$).

9. Justifier que l'on peut réécrire ce problème comme

$$\begin{gathered} \min_x \quad f(x) \\ \text{tel que } c(x) \leq 0 \end{gathered} \tag{181}$$

avec $c : \mathbb{R}^n \to \mathbb{R}^r$ où $r = m_1 + 2m_2$ cad sous la forme du problème (178). Pourquoi ne peut-on alors pas appliquer la méthode de points intérieurs étudiée précédemment ?

10. On introduit $s \in \mathbb{R}^{m_1}$ et l'on écrit x sous la forme $x_i = y_i - y_{i+n}$ avec $y \in \mathbb{R}^{2n}_+$, cad $x = My$ avec $M = (I_n \ \ -I_n)$ matrice $n \times 2n$. Soit $z = (y, s) \in \mathbb{R}^{2n+m_1}$.

Montrer que l'on peut réécrire le problème (180) sous la forme

$$\begin{aligned} &\min_z \quad g(z) \\ &\text{tel que } d(z) = 0, \\ &\qquad -z \leq 0 \end{aligned} \tag{182}$$

où $d : z = (y, s) \mapsto (c^{in}(My) + s, c^{eq}(My)) \in \mathbb{R}^{m_1+m_2}$, et où l'on précisera la fonction g.

Pour résoudre ce problème, on pénalise les contraintes inégalités par la méthode de points intérieurs étudiée précédemment et l'on considère le problème suivant pour $\varepsilon > 0$

$$\begin{aligned} &\min_z \quad g(z) - \varepsilon \sum_{i=1}^{2n+m_1} \ln(z_i) \\ &\text{tel que } d(z) = 0 \end{aligned} \tag{183}$$

dont on suppose qu'il admet une solution.

11. Former le lagrangien associé à (183) et montrer que les conditions de Karush-Kuhn-Tucker sont

$$\begin{cases} \nabla g(z) + \displaystyle\sum_{i=1}^{m_1+m_2} \lambda_i \nabla d_i(z) - w = 0 \\ w_i z_i = \varepsilon\,, \quad i = 1, \ldots, 2n + m_1 \\ d(z) = 0 \\ z \geq 0\,,\ w \geq 0 \end{cases} \tag{184}$$

12. Formuler ces conditions sous la forme $h(z, w, \lambda) = 0$ avec $h : \mathbb{R}_+^{4n+2m_1} \times \mathbb{R}^{m_1+m_2} \mapsto \mathbb{R}^{4n+3m_1+m_2}$ que l'on précisera.
13. Proposer une méthode de résolution de cette équation et formuler l'algorithme correspondant. (On pourra employer une projection.)
14. On suppose qu'une solution de (184) fournit une solution de (183). Donner une condition suffisante pour cela.
15. On note z_k une solution associée au problème (183) pour $\epsilon_k > 0$ et $N = (I_{2n}\ 0_{m_1})$ matrice $2n \times (2n + m_1)$. On suppose que $C = \{z \mid d(z) = 0\}$ est un ensemble fermé borné et que le problème (182) admet une unique solution z^*. Montrer alors que MNz_k converge vers x^*, solution du problème (180).

10.27 Minimisation alternée

Dans cet exercice, on considère $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ ($n \in \mathbb{N}$) deux fonctions convexes continues dont on suppose que l'on connaît les opérateurs proximaux Prox_{f_1} et Prox_{f_2}. On cherche un algorithme permettant de calculer aisément

$$\mathrm{Prox}_{f_1+f_2}(x) = \arg\min_{s\in\mathbb{R}^n} \left(f_1(s) + f_2(s) + \frac{1}{2}\|s - x\|^2 \right) \tag{185}$$

Pour ce faire, on définit la convolution infimale de f_1 et f_2 comme la fonction

$$f_1 \oplus f_2 : x \in \mathbb{R}^n \mapsto \inf_{\xi \in \mathbb{R}^n} \left(f_1(\xi) + f_2(x-\xi)\right)$$

1. (a) Montrer que, pour tout $(\xi_1, \xi_2) \in \mathbb{R}^{2n}$ et tout $\lambda \in [0,1]$,

$$f_1 \oplus f_2(\lambda x + (1-\lambda)y) \leq f_1(\lambda\xi_1 + (1-\lambda)\xi_2) + f_2(\lambda(x-\xi_1) + (1-\lambda)(y-\xi_2))$$

   (b) Conclure que $f_1 \oplus f_2$ est convexe.

2. On cherche une relation entre convolution infimale et transformée de Fenchel.
   (a) Montrer que, pour tout $x \in \mathbb{R}^n$,

$$(f_1^\star \oplus f_2^\star)^\star(x) = \sup_{\varphi_2} \sup_{\varphi_1} \left(x^T(\varphi_1 + \varphi_2) - f_1^\star(\varphi_1) - f_2^\star(\varphi_2)\right)$$

   (b) En déduire que $(f_1^\star \oplus f_2^\star)^\star = f_1^{\star\star} + f_2^{\star\star}$.
   (c) Conclure que $f_1^\star \oplus f_2^\star = (f_1 + f_2)^\star$.

3. Justifier que s solution de (185) si et seulement si $s = x - \mathrm{Prox}_{f_1^\star \oplus f_2^\star}(x)$.

4. On admet que l'inf définissant $f_1^\star \oplus f_2^\star$ est atteint. Montrer que $\mathrm{Prox}_{f_1^\star \oplus f_2^\star}(x) = y + z$ où $(y,z) \in \mathbb{R}^{2n}$ est une solution de

$$\arg\min_{y,z \in \mathbb{R}^n} \left(f_1^\star(y) + f_2^\star(z) + \frac{1}{2}\|y + z - x\|^2 \right)$$

5. Montrer que l'on sait déterminer, à $z \in \mathbb{R}^n$ fixé, la solution y_z du problème

$$(P_z) \qquad \arg\min_{y \in \mathbb{R}^n} \left(f_1^\star(y) + f_2^\star(z) + \frac{1}{2}\|y + z - x\|^2 \right)$$

   et donner son expression explicite en fonction de z et x en utilisant Prox_{f_1}. De même, donner l'expression en fonction de y et x et en utilisant Prox_{f_2} de la solution z_y du problème à $y \in \mathbb{R}^n$ fixé

$$(P_y) \qquad \arg\min_{z \in \mathbb{R}^n} \left(f_1^\star(y) + f_2^\star(z) + \frac{1}{2}\|y + z - x\|^2 \right)$$

On considère l'algorithme de minimisation alternée

***Algorithme* 20**. — A partir de $y_0 \in \mathbb{R}^n$ quelconque, itérer

- z_k solution de (P_{y_k})
- y_{k+1} solution de (P_{z_k})
- définir $s_k = x - y_k - z_k$

6. Interpréter cet algorithme et son but. Justifier de son intérêt au regard de la question 5.

APPENDICE A

COMPLÉMENTS SUR L'ANALYSE CONVEXE

A.1 Une fonction convexe est localement lipschitzienne sur l'intérieur de son domaine.

***Théorème* 45**. — *Soit $f : E \to \mathbb{R}$ convexe. Alors, pour tout $x_0 \in E^\circ$ (l'intérieur de E), il existe $\varepsilon > 0$ et $L > 0$ tels que*

$$\forall (x, y) \in B(x_0, \varepsilon) \quad |f(x) - f(y)| \leq L \, \|x - y\| \tag{186}$$

Démonstration. — Soit $x_0 \in E^\circ$. Il existe un voisinage de x_0 tel que f est bornée, c.a.d. il existe $\varepsilon, M > 0$ tels que, pour $x \in B(x_0, 2\varepsilon)$, $|f(x)| \leq M$. Soient $x \in B(x_0, \varepsilon)$, $y \in B(x_0, \varepsilon)$, $\alpha = \frac{1}{\varepsilon}\|y - x\|$ et $z^+ = x + \frac{1}{\alpha}(y - x)$. Par définition, $z^+ \in B(x_0, 2\varepsilon)$ et $y = \alpha z^+ + (1 - \alpha)x$, d'où, par convexité de f, il s'ensuit

$$f(y) \leq \alpha f(z^+) + (1 - \alpha) f(x) \leq f(x) + \alpha(M - f(x)) \tag{187}$$

Par ailleurs, définissons maintenant $z^- = x - \frac{1}{\alpha}(y - x)$. De même, $z^- \in B(x_0, 2\varepsilon)$ et la convexité de f implique

$$f(x) = f\left(\frac{1}{1+\alpha} y + \frac{\alpha}{1+\alpha} z^-\right) \leq \frac{1}{1+\alpha} f(y) + \frac{\alpha}{1+\alpha} f(z^-)$$

soit

$$f(y) \geq (1 + \alpha) f(x) - \alpha f(z^-) \geq f(x) - \alpha(M - f(x)) \tag{188}$$

Ainsi, en rassemblant (187) et (188), on en déduit l'inégalité voulue, à savoir

$$|f(y) - f(x)| \leq \frac{M - f(x)}{\varepsilon} \|y - x\| \leq \frac{2M}{\varepsilon} \|y - x\| \triangleq L \|y - x\|, \quad (x, y) \in B(x_0, \varepsilon)^2$$

□

A.2 Une fonction continue convexe est différentiable presque partout.

***Proposition* 12**. — *Soient E un espace vectoriel de dimension finie et $f : E \to \mathbb{R}^n$ convexe. Si f admet des dérivées partielles en $x \in E$, alors f est différentiable en x.*

Démonstration. — Supposons que f admette des dérivées partielles en $x \in E$. Posons, pour $h \in E$ assez petit,

$$g(h) = f(x+h) + f(x) + \nabla f(x)^T h$$

qui est une fonction convexe. Soit $(e_i)_{1\leq i\leq n}$ une base de E et soit $h = \sum_{i=1}^n h_i e_i$. Alors, g étant convexe,, on a

$$g(h) = g\left(\frac{1}{N}\sum_{i=1}^n N h_i e_i\right) \leq \frac{1}{N}\sum_{i=1}^N g(N h_i e_i) = \frac{1}{N}\sum_{h_i\neq 0} h_i \frac{g(N h_i e_i)}{h_i}$$

dont on déduit

$$\frac{g(h)}{\|h\|} \leq \sum_{h_i\neq 0}\left|\frac{g(N h_i e_i)}{N h_i}\right|$$

Or, par convexité de g, on a $0 = g(0) = g\left(\frac{1}{2}h - \frac{1}{2}h\right) \leq \frac{1}{2}g(h) + \frac{1}{2}g(-h)$ et ainsi

$$-\frac{g(h)}{\|h\|} \leq \frac{g(-h)}{\|h\|} \leq \sum_{h_i\neq 0}\left|\frac{g(-N h_i e_i)}{N h_i}\right|$$

En conséquence,

$$\frac{|g(h)-g(0)|}{\|h\|} = \frac{|g(h)|}{\|h\|} \leq \sum_{h_i\neq 0} \max\left\{\left|\frac{g(N h_i e_i)}{N h_i}\right|, \left|\frac{g(-N h_i e_i)}{N h_i}\right|\right\} \underset{\|h\|\to 0}{\to} 0$$

où la limite du membre de droite existe car h admet des dérivées partielles en x. Aussi, on conclut que $g(h) \in o(\|h\|)$ et donc que $f(x+h) = f(x) + \nabla(x)^T h + o(\|h\|)$, c.-à-d. f différentiable en x. □

***Proposition* 13**. — *Soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe. Alors f est différentiable en presque tout point de $\mathbb{R}^n$.*

Démonstration. — En premier lieu, prouvons le résultat intermédiaire suivant :

Les points où $g : \mathbb{R} \to \mathbb{R}$ convexe est non-dérivable sont dénombrables.

Soit x un point de non-dérivabilité de g. Alors, par le théorème 8 d'inégalité des pentes, on a que g admet des dérivées à droite et à gauche en x et telles que $g'(x^-) < g'(x^+)$. Nécessairement, ces dérivées sont finies car g est définie sur tout $\mathbb{R}$. Définissons la famille $\{]g'(x^-), g'(x^+)[\mid x \in \Delta\}$ avec Δ l'ensemble des points de non-dérivabilité de g. Des considérations précédentes, il s'agit d'une famille d'intervalles non-vides, ouverts et disjoints. Chacun de ces intervalles contenant au moins un rationnel, on déduit que Δ est dénombrable.

Maintenant, considérons A l'ensemble des points de $\mathbb{R}^n$ où f n'est pas différentiable. Par la Proposition 12, on a

$$A \subset \cup_{i=1}^n \left\{ x \in \mathbb{R}^n \mid \frac{\partial f}{\partial e_i}(x) \text{ n'existe pas } \right\} \triangleq \cup_{i=1}^n A_i$$

Par théorème de Fubini, on a

$$\mu(A_n) = \int_{\mathbb{R}^n} \mathbb{1}_{A_n}(x)dx = \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \mathbb{1}_{A_n}(\xi, x_n)dx_n d\xi \tag{189}$$

Pour tout $\xi \in \mathbb{R}^{n-1}$, la fonction $x_n \to \mathbb{1}_{A_n}(\xi, x_n)$ est la fonction indicatrice du lieu B_ξ de non-différentiabilités de la fonction convexe $x_n \in \mathbb{R} \to f(\xi, x_n)$. Or, du résultat intermédiaire prouvé précédemment, B_ξ est dénombrable et donc négligeable. Aussi, $\int_{\mathbb{R}} \mathbb{1}_{A_n}(\xi, x_n)dx_n = 0$ et, avec (189) $\mu(A_n) = 0$. On conclut de même que $\mu(A_i) = 0$ pour tout $i = 1, \ldots, n$. □

A.3 Théorèmes de séparation des convexes

***Théorème* 46 (Théorème de Hahn-Banach, première forme géométrique)**
Soit E un espace vectoriel normé. Soient A et B deux convexes de E non vides et disjoints. On suppose A ouvert. Alors il existe un hyperplan fermé séparant A et B, i.e., il existe f une forme linéaire sur E non identiquement nulle et $\alpha \in \mathbb{R}$ tels que

$$\forall x \in A \quad f(x) \leq \alpha \quad \text{et} \quad \forall x \in B \quad f(x) \geq \alpha \tag{190}$$

***Théorème* 47 (Théorème strict de Hahn-Banach, deuxième forme géométrique)**
Soit E un espace vectoriel normé. Soient A et B deux convexes de E non vides et disjoints. On suppose A fermé et B compact. Alors il existe un hyperplan fermé séparant strictement A et B, i.e., il existe f une forme linéaire sur E non identiquement nulle, $\alpha \in \mathbb{R}$ et $\varepsilon > 0$ tels que

$$\forall x \in A \quad f(x) \leq \alpha - \varepsilon \quad \text{et} \quad \forall x \in B \quad f(x) \geq \alpha + \varepsilon \tag{191}$$

La preuve de ces théorèmes peut être trouvée dans [**7**].

APPENDICE B

DÉCOMPOSITIONS MATRICIELLES ET ALGORITHMES ASSOCIÉS

B.1 Décomposition LU

La factorisation LU d'une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est définie comme

$$PA = LU$$

où $P \in \mathcal{M}_n(\mathbb{R})$ matrice de permutation, L triangulaire inférieure unitaire (cad avec des éléments diagonaux égaux à 1) et U triangulaire supérieur.

Une telle décomposition existe pour toute matrice carrée, mais n'est pas unique. Il est par ailleurs possible de choisir $P = I$ dans le cas particulier où tous les mineurs principaux de A sont non-nuls (impliquant que A doit être inversible).

Pour obtenir cette factorisation, on recourt à l'algorithme du pivot de Gauss suivant, qui a une complexité de $2n^3/3$ opérations.

***Algorithme* 21 (Elimination de Gauss-Jordan avec pivot partiel des colonnes)**

A partir de $A \in \mathcal{M}_n(\mathbb{R})$, de $P = I$ et $L = 0$, itérer pour j entre 1 et n:

1. trouver l'indice $i \in \{j, \ldots, n\}$ tel que $|A_{ij}| = \max_{k=j,\ldots,n} |A_{kj}|$.
2. si $A_{ij} = 0$, arrêt (matrice singulière)
3. sinon, mettre à jour P pour inverser les colonnes i et j de la matrice A
4. $L_{jj} = 1$
5. for $k = j + 1, \ldots, n$
   $L_{kj} = A_{kj}/A_{jj}$
   for $l = j + 1, \ldots, n$
   $A_{kl} = A_{kl} - L_{kj} A_{jl}$
6. Définir U comme la partie triangulaire supérieur de A.

Cet algorithme est utilisé de façon standard pour inverser une matrice. En effet, une fois cette décomposition effectuée, il est alors aisé de résoudre le système $AX = I$, cad d'inverser la matrice A, en résolvant tour à tour $LY = I$ et $UX = Y$. La

complexité finale de l'algorithme est plus faible que celle d'une méthode de pivot de Gauss directement appliquée à la matrice A.

B.2 Décomposition de Cholesky

Quand la matrice A est symétrique positive définie, il est alors possible d'écrire la décomposition ci-dessus sous une forme plus forte, à savoir

$$A = LL^T$$

où L triangulaire supérieure. Par ailleurs, si l'on impose à L d'avoir des coefficients diagonaux positifs, cette décomposition est unique.

Une variante de l'algorithme précédent, exploitant la symétrie de A et n'impliquant pas d'inversion de colonnes, permet de réaliser une telle décomposition en $n^3/3$ opérations. Elle est également utilisée pour inverser A.

B.3 Décomposition QR

Une autre décomposition fréquemment utilisée pour une matrice $A \in \mathcal{M}_{m\times n}\mathbb{R}$ est

$$AP = QR$$

où $P \in \mathcal{M}_n(\mathbb{R})$ matrice de permutation, $Q \in \mathcal{M}_m(\mathbb{R})$ matrice orthogonale et $R \in \mathcal{M}_{m\times n}(\mathbb{R})$ matrice triangulaire supérieure.

Cette factorisation peut être réalisée en appliquant une suite de matrices orthogonales à A (méthode de Householder, voir [**13**] pour une description complète). Le coût de cet algorithme est $4m^2n/3$. Dans le cas d'une matrice carrée, on voit donc que cet algorithme est plus complexe que les précédents, mais il présente l'avantage d'être plus stable numériquement. On l'utilise notamment pour déterminer le noyau de l'application linéaire associée à A car les $n-m$ dernières colonnes de Q en forment une base.

APPENDICE C

OPTIMISATION DE TRAJECTOIRES

C.1 Calcul des variations

C.1.1 Historique

Au lieu d'un nombre fini de paramètres, on peut considérer qu'il s'agit de l'optimisation d'un nombre infini de valeurs successives de la fonction recherchée.

Ces problèmes sont très anciens. On pourra se référer à [**29**] pour un exposé complet de leur historique. Le problème de Dido est le calcul du contour de périmètre donné enfermant l'aire maximale. Il date du 9ème siècle avant JC. Galilée (1588) considéra des problèmes célèbres, mais la réelle percée vint de Johann Bernoulli qui mit au défi les plus illustres mathématiciens de son époque (son propre frère Jakob, Leibnitz, Newton). Le problème en question était le Brachistochrone: le calcul de la courbe permettant à un point matériel soumis à la gravité de rallier son point initial à un point d'arrivée donné.

D'autres problèmes importants sont le calcul des géodésiques, les problèmes iso–périmétriques, et de manière générale les problèmes avec contraintes.

Les "sciences naturelles" (mécanique, optique, ...) ont été profondément marquées par le calcul des variations. De nombreuses lois ne la nature se déduisent de principes variationnels, énoncant que parmi tous les mouvements possibles, celui qui se réalise est un extremum. Cette vision de la nature, indiqua que les réalisations humaines devaient elles aussi être des extrema: Brachistochrone ou problème de Newton.

C.1.2 Notions fondamentales

***Définition* 28 (Fonctionnelle)**. — Soit X un espace vectoriel normé. On appelle fonctionnelle une application de X dans $\mathbb{R}$.

Dans ce qui suit nous nous intéressons à l'espace vectoriel normé $X = D$ des fonctions continues à dérivées continues $\mathbb{R} \supset [t_1, t_2] \longrightarrow \mathbb{R}^n$, $n < \infty$ muni de la norme $X \ni u \mapsto \|u\|_D = \max_{t\in[t_1,t_2]} \|u(t)\| + \max_{t\in[t_1,t_2]} \|\dot{u}(t)\|$. On considère les

fonctionnelles du type

$$X \ni u \mapsto J(u) = \int_{t_1}^{t_2} L(u(t), \dot{u}(t), t)dt \in \mathbb{R}$$

où L est une fonction $\mathbb{R}^{2n+1} \ni (u, v, t) \mapsto L(u, v, t) \in \mathbb{R}$ continue possédant des dérivées partielles continues.

On cherchera à minimiser J sous certaines contraintes, en restreignant l'ensemble des fonctions considérées à $X \supset U = \{u \in X, u(t_1) = a, u(t_2) = b\}$ où a et b sont des réels donnés. C'est un ensemble convexe. On cherchera à résoudre le problème suivant

$$\min_{u \in U} J(u) \tag{192}$$

***Définition* 29**. — On dit que u^* est un minimum (local) de J (c.-à-d. une solution du problème (192) s'il existe un voisinage $\mathcal{V}(u^*)$ tel que $J(u^*) \leq J(u)$ pour tout $u \in \mathcal{V}(u^*)$.

***Définition* 30 (Variation admissible)**. — On dit que $h \in X$ est une variation admissible au point $u \in U$ si $(u + h) \in U$.

***Définition* 31 (Différentielle de Gâteaux)**. — Soient J une fonctionnelle sur X espace vectoriel normé, $u^* \in U$ et $h \in X$ variation admissible. On appelle différentielle de Gâteaux de J au point u^* dans la direction h la limite

$$\delta J(u^*, h) = \lim_{\delta \longrightarrow 0} \frac{1}{\delta} \left(J(u^* + \delta h) - J(u^*)\right)$$

C.1.3 Conditions nécessaires d'extrémalité

La notion de différentielle de Gâteaux remplace pour les fonctionnelles la notion de différentielle pour les fonctions. Une condition nécessaire pour que $u^* \in U$ soit un minimum est que la différentielle de Gâteaux $\delta J(u^*, h)$ doit être nulle pour toute variation admissible h. On va réécrire cette proposition en une équation exploitable à l'aide du résultat suivant.

***Lemme* 10 (Lemme de duBois-Reymond)**. — *Soit $\mathbb{R} \supset [t_1, t_2] \ni t \mapsto \phi(t) \in \mathbb{R}$ une fonction continue. Si on a, pour tout $\mathbb{R} \supset [t_1, t_2] \ni t \mapsto h(t) \in \mathbb{R}$ continue à dérivée continue telle que $h(t_1) = h(t_2) = 0$,*

$$\int_{t_1}^{t_2} \phi(t)h(t)dt = 0$$

alors $\phi = 0$.

Démonstration. — Supposons que les hypothèses du lemme sont valides et que $\phi \neq 0$. Sans perte de généralité on peut supposer qu'elle est strictement positive en un point $t \in [t_1, t_2]$. Par continuité, elle est donc strictement positive sur $[t'_1, t'_2] \subset [t_1, t_2]$. Définissons h comme

$$h(t) = (t - t'_1)^2(t - t'_2)^2 \mathbb{I}_{[t'_1, t'_2]}(t)$$

Avec cette variation, on obtient $\int_{t_1}^{t_2} \phi(t)h(t)dt > 0$. D'où une contradiction et donc $\phi = 0$. □

L'équation $\delta J(u^*, h) = 0$ pour toute variation admissible se réécrit

$$\int_{t_1}^{t_2} \left(\frac{\partial L}{\partial u}(u^*, \dot{u}^*, t)h(t) + \frac{\partial L}{\partial \dot{u}}(u^*, \dot{u}^*, t)\dot{h}(t) \right) dt = 0$$

Par intégration par parties il vient

$$\int_{t_1}^{t_2} \left(\frac{\partial L}{\partial u}(u^*, \dot{u}^*, t) - \frac{d}{dt}\frac{\partial L}{\partial \dot{u}}(u^*, \dot{u}^*, t) \right) h(t)dt + \left[\frac{\partial L}{\partial \dot{u}}(u^*, \dot{u}^*, t)h(t) \right]_{t_1}^{t_2} = 0$$

h est une variation admissible donc $h(t_1) = h(t_2) = 0$. La précédente équation se simplifie

$$\int_{t_1}^{t_2} \left(\frac{\partial L}{\partial u}(u^*, \dot{u}^*, t) - \frac{d}{dt}\frac{\partial L}{\partial \dot{u}}(u^*, \dot{u}^*, t) \right) h(t)dt = 0$$

Cette équation doit être vraie pour toute variation admissible h, et le lemme 10 de duBois-Reymond permet de conclure

$$\frac{\partial L}{\partial u}(u^*, \dot{u}^*, t) - \frac{d}{dt}\frac{\partial L}{\partial \dot{u}}(u^*, \dot{u}^*, t) = 0 \tag{193}$$

L'équation (193) est appelée équation d'Euler-Lagrange et est une condition nécessaire qui doit être satisfaite par toute solution du problème (192).

Dans de nombreux cas pratique, le Lagrangien ne dépend pas explicitement de t. On pourra alors utiliser la proposition suivante.

***Proposition* 14**. — *Lorsque L ne dépend pas explicitement de t, l'équation d'Euler-Lagrange* (193) *implique*

$$\frac{d}{dt}\left(L - \dot{u}\frac{\partial L}{\partial \dot{u}} \right) = 0$$

Démonstration. — On calcule

$$\frac{d}{dt}\left(L - \dot{u}\frac{\partial L}{\partial \dot{u}} \right) = \dot{u}\left(\frac{\partial L}{\partial u} - \frac{d}{dt}\frac{\partial L}{\partial \dot{u}} \right) = 0$$

en utilisant (193). □

On peut étendre le calcul des variations aux fonctionnelles du type

$$X \ni u \mapsto J(u) = \int_{t_1}^{t_2} L(u(t), \dot{u}(t), ..., u^{(n)}, t)dt \in \mathbb{R}$$

(en utilisant un espace vectoriel normé X adapté) et on obtient l'équation d'Euler-Poisson

$$\frac{\partial L}{\partial u} - \frac{d}{dt}\frac{\partial L}{\partial \dot{u}} + \frac{d^2}{dt^2}\frac{\partial L}{\partial \ddot{u}} + ... + (-1)^n \frac{d^n}{dt^n}\frac{\partial L}{\partial u^{(n)}} = 0$$

On peut également appliquer les mêmes règles de calcul ainsi que la formule de Green pour les fonctionnelles

$$u \mapsto J(u) = \int_{t_1}^{t_2} L\left(u(x,y), \frac{\partial}{\partial x}u(x,y), \frac{\partial}{\partial y}u(x,y), x, y\right) dx\ dy \in \mathbb{R}$$

pour obtenir l'équation d'Ostrogradski

$$\frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial L}{\partial \frac{\partial}{\partial x}u}u(x,y)\right) - \frac{\partial}{\partial y}\left(\frac{\partial L}{\partial \frac{\partial}{\partial y}u}u(x,y)\right) = 0$$

C.2 Optimisation de systèmes dynamiques

On s'intéresse désormais au problème de la minimisation d'une fonctionnelle $X \times U \ni (x,u) \mapsto J(x,u) \in \mathbb{R}$ où $X \times U$ est un espace vectoriel normé. Ici, X est l'espace des fonctions continues à dérivées continues $\mathbb{R} \supset [t_1,t_2] \longrightarrow \mathbb{R}^n$, $n < \infty$ muni de la norme $X \ni x \mapsto \|x\|_D = \max_{t\in[t_1,t_2]} \|x(t)\| + \max_{t\in[t_1,t_2]} \|\dot{x}(t)\|$. U est l'espace des fonctions définies sur $[t_1,t_2] \longrightarrow \mathbb{R}^m$. Les fonctions x et u sont contraintes par une équation différentielle

$$\dot{x} = f(x,u)$$

où $\mathbb{R}^{n+m} \ni (x,u) \mapsto f(x,u) \in \mathbb{R}^n$ est une fonction de classe $\mathcal{C}^1$. En outre on impose la contrainte

$$x(t_1) = x^0 \in \mathbb{R}^n \tag{194}$$

On va chercher une caractérisation des solutions du problème

$$\begin{aligned} &\min_{x,\,u} \quad J(x,u) \\ &\text{tel que } \dot{x} = f(x,u), \\ &\qquad x(t_1) = x^0 \end{aligned} \tag{195}$$

On considérera des fonctionnelles du type

$$X \times U \ni (x,u) \mapsto J(x,u) = \varphi(x(t_2),t_2) + \int_{t_1}^{t_2} L(x(t),u(t),t)dt$$

où $\mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto \varphi(x,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$et $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \ni (x,t) \mapsto L(x,u,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$. Pour tenir compte de la condition initiale (194), on considère qu'une variation admissible $(\delta x, \delta u) \in X \times U$ doit vérifier $\delta x(t_1) = 0$. Comme dans le cas de l'optimisation continue de dimension finie, on va chercher à résoudre le problème d'optimisation sous contrainte en adjoignant les contraintes à la fonction coût.

On forme alors $X \times U \times M \ni (x,u,\lambda) \mapsto \bar{J}(x,u,\lambda)$ où M est l'espace des fonctions $\mathbb{R} \longrightarrow \mathbb{R}^n$ différentiable par

$$\bar{J}(x,u,\lambda) = J(x,u) + \int_{t_1}^{t_2} \lambda(t)^T (f(x,u,t) - \dot{x})(t)dt$$

En notant

$$H(x(t),u(t),\lambda(t),t) = L(x(t),u(t),t) + \lambda(t)^T f(x,u,t) \tag{196}$$

qu'on appellera Hamiltonien du problème (195), il vient

$$\begin{aligned}\bar{J}(x,u,\lambda) =&\varphi(x(t_2),t_2)+\lambda(t_1)^T x(t_1)-\lambda(t_2)^T x(t_2)\\ &+\int_{t_1}^{t_2}\Big(H(x(t),u(t),\lambda(t),t)+\dot{\lambda}(t)^T x(t)\Big)\,dt\end{aligned}\tag{197}$$

***Proposition* 15**. — *Si* $(x^*,u^*,\lambda^*)\in X\times U\times M$ *est un point stationnaire de* $\bar{J}$ *défini par* (197) *alors* (x^*,u^*) *est un point stationnaire de* J *sous les contraintes* $\dot{x}=f(x,u,t)$, $x(t_1)=x^0$.

Démonstration. — Nous allons montrer que, autour de (x^*,u^*), pour toute variation admissible du premier ordre $[t_1,t_2]\ni(\delta x,\delta u)(t)\in\mathbb{R}^n\times\mathbb{R}^m$ telle que $\|(\delta x,\delta u)\|_{X\times U}=\circ(\delta)$ satisfaisant la contrainte $\dot{x}=f(x,u,t)$, la variation δJ de la fonctionnelle J est du deuxième ordre, c.-à-d. $\delta J=\circ(\delta)$.

La stationnarité de (x^*,u^*,λ^*) pour $\bar{J}$ est caractérisée par 3 relations. La première est

$$\begin{aligned}&\frac{\partial}{\partial x(t_2)}\varphi(x(t_2),t_2)\delta x(t_2)-\lambda(t_2)^T\delta x(t_2)\\ &+\int_{t_1}^{t_2}(\frac{\partial}{\partial x}L(x,u,t)\delta x+\lambda(t)^T\frac{\partial}{\partial x}f(x,u,t)\delta x+\dot{\lambda}(t)^T\delta x)dt=\circ(\delta)\end{aligned}$$

pour toute variation admissible $\mathbb{R}\ni t\mapsto\delta x(t)\in\mathbb{R}^n$. Ce calcul des variations donne

$$\frac{\partial}{\partial x}L(x,u,t)+\lambda(t)^T\frac{\partial}{\partial x}f(x,u,t)+\dot{\lambda}(t)^T=0\tag{198}$$

$$\lambda^T(t_2)=\frac{\partial}{\partial x(t_2)}\varphi(x(t_2),t_2)\tag{199}$$

On réécrira la condition (198) sous la forme $\dot{\lambda}^T=-\frac{\partial H}{\partial x}$. La seconde condition de stationnarité est que pour toute variation $\mathbb{R}\ni t\mapsto\delta u(t)\in\mathbb{R}^m$ on doit avoir

$$\int_{t_1}^{t_2}\left(\frac{\partial}{\partial u}L(x,u,t)+\lambda(t)^T\frac{\partial}{\partial u}f(x,u,t)\right)\delta u(t)\;dt=\circ(\delta)$$

Ce calcul des variations donne

$$\frac{\partial}{\partial u}L(x,u,t)+\lambda(t)^T\frac{\partial}{\partial u}f(x,u,t)=\circ(\delta)\tag{200}$$

On réécrira cette condition comme $\frac{\partial}{\partial u}H=0$.

La dernière condition de stationnarité est

$$\int_{t_1}^{t_2}\delta\lambda(t)^T(f(x,u,t)-\dot{x})dt=0$$

Elle redonne la contrainte

$$\dot{x}=f(x,u,t)\tag{201}$$

□

Supposons que les équations (198), (199), (200) et (201) soient vérifiées. Calculons la variation de la fonctionnelle J.

$$\begin{aligned}
\delta J =& \frac{\partial}{\partial x(t_2)}\varphi(x(t_2),t_2)\delta x(t_2) + \int_{t_1}^{t_2}\left(\frac{\partial}{\partial x}L(x,u,t)\delta x + \frac{\partial}{\partial u}L(x,u,t)\delta u\right)dt \\
=& \frac{\partial}{\partial x(t_2)}\varphi(x(t_2),t_2)\delta x(t_2) \\
&+ \int_{t_1}^{t_2}\left(-\lambda(t)^T\frac{\partial}{\partial x}f(x,u,t)\delta x(t) - \dot\lambda(t)^T\delta x(t) - \lambda(t)^T\frac{\partial}{\partial u}f(x,u,t)\delta u(t)\right)dt \\
=& \frac{\partial}{\partial x(t_2)}\varphi(x(t_2),t_2)\delta x(t_2) - \lambda(t_2)^T\delta x(t_2) + \lambda(t_1)^T\delta x(t_1) \\
&+ \int_{t_1}^{t_2}\lambda(t)^T\left(\dot{\delta x}(t) - \frac{\partial}{\partial x}f(x,u,t)\delta x(t) - \frac{\partial}{\partial u}f(x,u,t)\delta u(t)\right)dt \\
&= \circ(\delta)
\end{aligned}$$

C.2.1 Problème aux deux bouts

Les conditions de stationnarité (198) (199) de la proposition 15 forment, avec les contraintes (201) et (194), le problème "aux deux bouts" suivant

$$\left\{\begin{array}{r}
\dot x = f(x,u) \\
x(t_1) = x^0 \\
\dot\lambda^T = -\dfrac{\partial H}{\partial x}(x,u,\lambda) \\
\lambda^T(t_2) = \dfrac{\partial}{\partial x(t_2)}\varphi(x(t_2),t_2) \\
\text{où } u \text{ est solution de } \dfrac{\partial H}{\partial u} = 0
\end{array}\right. \tag{202}$$

Le long de l'extrémale on élimine u des équations en résolvant les équations $\frac{\partial}{\partial u}H = 0$. Le problème aux deux bouts a pour seules inconnues les fonctions $\mathbb{R} \supset [t_1,t_2] \ni t \mapsto x(t) \in \mathbb{R}^n$ et $\mathbb{R} \supset [t_1,t_2] \ni t \mapsto \lambda(t) \in \mathbb{R}^n$. Le système d'équations différentielles et de conditions limites (202) ne définit pas un problème de Cauchy à cause de la séparation des conditions de bords aux deux extrémités du domaine. En outre, la condition portant sur λ dépend de l'inconnue x. C'est un problème difficile à résoudre en général.

***Proposition* 16 (Conservation de l'Hamiltonien).** — *Lorsque L ne dépend pas explicitement de t, les conditions de stationnarité constituant le système d'équations différentielles aux deux bouts* (202) *impliquent*

$$\frac{d}{dt}H = 0$$

Démonstration. — Un calcul direct donne

$$\frac{d}{dt}H = \frac{\partial H}{\partial x}\dot x + 0 + \frac{\partial H}{\partial \lambda}\dot\lambda = \frac{\partial H}{\partial x}f(x,u) - \frac{\partial H}{\partial x}f(x,u) = 0$$

□

En pratique on se servira souvent de la conservation de l'Hamiltonien pour éliminer une variable et essayer de résoudre le problème aux deux bouts.

C.2.2 Contraintes finales

On cherche ici une caractérisation des solutions du problème

$$\begin{aligned} &\min_{x,\,u} \quad J(x,u) \\ &\text{tel que } \dot{x} = f(x,u), \\ &\qquad\quad x(t_1) = x^0, \\ &\qquad\quad \psi(x(t_2),t_2) = 0 \end{aligned} \tag{203}$$

où $\mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto \psi(x,t) \in \mathbb{R}^q$, $1 < q \leq n$ est une fonction de classe $\mathcal{C}^1$. Ces contraintes portent sur l'état final et le temps final du système qui est ici fixe, on les nomme contraintes de "rendez-vous".

On considérera des fonctionnelles du type

$$X \times U \ni (x,u) \mapsto J(x,u) = \varphi(x(t_2),t_2) + \int_{t_1}^{t_2} L(x(t),u(t),t)dt$$

où $\mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto \varphi(x,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$et $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \ni (x,u,t) \mapsto L(x,u,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$.

Le problème aux deux bouts correspondant à ce problème d'optimisation est

$$\begin{cases} \dot{x} = f(x,u) \\ x(t_1) = x^0 \\ \dot{\lambda}^T = -\dfrac{\partial H}{\partial x}(x,u,\lambda) \\ \lambda^T(t_2) = \dfrac{\partial}{\partial x(t_2)}\varphi(x(t_2),t_2) + \nu^T \dfrac{\partial \psi}{\partial x(t_2)} \\ \psi(x(t_2),t_2) = 0 \\ \qquad \text{où } u \text{ est solution de } \dfrac{\partial H}{\partial u} = 0 \end{cases} \tag{204}$$

où $H(x,u,\lambda,t) = L(x,u,t) + \lambda^T f(x,u,t)$ et $\nu \in \mathbb{R}^q$. On obtient ce résultat en adjoignant les contraintes $\dot{x} = f(x,u,t)$ et $\psi(x(t_2),t_2) = 0$ à la fonctionnelle et en explicitant le calcul des variations. On a ainsi

$$\bar{J}(x,u,\lambda,\nu) = J(x,u) + \int_{t_1}^{t_2} \lambda(t)^T (f(x,u,t) - \dot{x})(t)dt + \nu^T \psi(x(t_2),t_2) \tag{205}$$

On peut alors établir la proposition suivante

***Proposition* 17**. — *Si* $(x^*, u^*, \lambda^*, \nu^*) \in X \times U \times M \times \mathbb{R}^q$ *est un point stationnaire de* $\bar{J}$ *défini par* (205) *alors* (x^*, u^*) *est un point stationnaire de* J *sous les contraintes* $\dot{x} = f(x, u, t)$, $x(t_1) = x^0$, $\psi(x(t_2), t_2) = 0$.

C.2.3 Résolution numérique du problème aux deux bouts

C.2.3.1 Calcul du gradient par l'adjoint

On va chercher à étendre les calculs menés en dimension finie établissant la variation de la fonction coût par rapport aux inconnues en satisfaisant les contraintes (27). Dans le cas présent, on revient au problème

$$\begin{aligned} \min_{x,\,u} \quad & J(x,u) \\ \text{tel que } & \dot{x} = f(x,u), \\ & x(t_1) = x^0 \end{aligned}$$

On considérera des fonctionnelles du type

$$X \times U \ni (x,u) \mapsto J(x,u) = \varphi(x(t_2),t_2) + \int_{t_1}^{t_2} L(x(t),u(t),t)dt$$

où $\mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto \varphi(x,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$et $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \ni (x,t) \mapsto L(x,u,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$.

On va chercher à calculer la variation de J lorsqu'on fait varier l'inconnue u tout en maintenant la contrainte $\dot{x} = f(x,u)$, $x(t_1) = x^0$. On va noter δJ la variation de la valeur de J engendrée par une telle variation δu. Il vient

$$\delta J = \frac{\partial \varphi}{\partial x}(x(t_2),t_2)\delta x(t_2) + \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x}(x(t),u(t),t)\delta x(t) + \frac{\partial L}{\partial u}\delta u(t) \right) dt + \circ(\delta) \tag{206}$$

Dans cette équation δx est liée à δu par la satisfaction de la contrainte $\dot{x} = f(x,u)$ qui donne

$$\dot{\delta} x(t) = \frac{\partial f}{\partial x}(x,u,t)\delta x(t) + \frac{\partial f}{\partial u}(x,u,t)\delta u(t) + \circ(\delta)$$

Il est possible d'intégrer cette équation linéaire en utilisant la matrice de passage M définie par $\frac{d}{dt}M(t,t_1) = \frac{\partial f}{\partial x}(x,u,t)M(t,t_1)$, $M(t_1,t_1) = I$ ce qui donne alors

$$\delta x(t) = M(t,t_1)\delta x(t_1) + \int_{t_1}^{t} M(t,\tau)\frac{\partial f}{\partial u}(x,u,\tau)\delta u(\tau)d\tau + \circ(\delta)$$

On peut donc éliminer δx dans l'équation (206) et obtenir explicitement δJ en fonction de la variation $\mathbb{R} \supset [t_1,t_2] \in t \mapsto u(t) \in \mathbb{R}^m$. Cette expression est très compliquée à calculer car elle nécessite la résolution complète d'un système d'équations différentielles. En fait, le lemme suivant va nous permettre de simplifier énormément les calculs

***Lemme* 11 (Lemme de l'adjoint).** — *Les solutions du système d'équations différentielles*

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ \dot{\lambda}(t) = -A^T(t)\lambda(t) + \Gamma(t) \end{cases}$$

où pour tout $t \in [t_1,t_2] \subset \mathbb{R}$, on a $A(t) \in \mathcal{M}_n(\mathbb{R})$, $B(t) \in \mathcal{M}_{(n\times m)}(\mathbb{R})$, $\Gamma(t) \in \mathbb{R}^n$, satisfont l'égalité

$$\lambda^T(t_2)x(t_2) - \lambda^T(t_1)x(t_1) = \int_{t_1}^{t_2} \left(\lambda^T(t)B(t)u(t) + \Gamma^T(t)x(t) \right) dt$$

Le système d'équations que nous devons résoudre dans le problème aux deux bouts implique

$$\dot{\delta}x(t) = \frac{\partial f}{\partial x}(x,u,t)\delta x(t) + \frac{\partial f}{\partial u}(x,u,t)\delta u(t) + \circ(\delta)$$
$$\dot{\lambda}(t) = -\left(\frac{\partial f}{\partial x}(x,u,t)\right)^T \lambda(t) - \left(\frac{\partial L}{\partial x}(x,u,t)\right)^T$$

avec comme conditions limites $\delta x(t_1) = 0$, $\lambda(t_2) = \left(\frac{\partial}{\partial x(t_2)}\varphi(x(t_2),t_2)\right)^T$. Le lemme 11 de l'adjoint donne

$$\delta J = \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial u}(x,u,t) + \lambda^T \frac{\partial f}{\partial u}(x,u,t)\right)\delta u(t)dt + \circ(\delta) = \int_{t_1}^{t_2} \frac{\partial H}{\partial u}(x,u,\lambda,t)\delta u(t)dt + \circ(\delta)$$

Par abus de notation on retiendra la formule analogue à (27)

$$\left(\frac{\partial J}{\partial u}\right)_{\substack{\dot{x} = f(x,u,t) \\ x(t_1) = x^0}} = \frac{\partial H}{\partial u} \tag{207}$$

Autrement dit, la variation de la valeur de J en satisfaisant les contraintes est calculée en formant $\frac{\partial H}{\partial u} = \frac{\partial L}{\partial u}(x,u,t) + \lambda^T \frac{\partial f}{\partial u}(x,u,t)$, évaluée à partir des valeurs de x, u, et de l'adjoint λ. La formule (207) est appelée formule du calcul du gradient par l'adjoint.

***Algorithme* 22 (Résolution du problème aux deux bouts par le calcul du gradient par l'adjoint)**

À partir de u^0 fonction $\mathbb{R} \supset [t_1,t_2] \mapsto u^0(t) \in \mathbb{R}^m$ quelconque, itérer

- Résoudre $\dot{x}^k = f(x^k,u^k,t)$, $x^k(t_1) = x^0$ pour $t \in [t_1,t_2]$
- Résoudre $\dot{\lambda}^k(t) = -\left(\frac{\partial f}{\partial x}(x^k,u^k,t)\right)^T \lambda(t) - \left(\frac{\partial L}{\partial x}(x,u,t)\right)^T$ avec comme condition limite $\lambda^k(t_2) = \left(\frac{\partial}{\partial x(t_2)}\varphi(x^k(t_2),t_2)\right)^T$
- Calculer $\frac{\partial H^k}{\partial u} = \frac{\partial L}{\partial u}(x^k,u^k,t) + \left(\lambda^k\right)^T \frac{\partial f}{\partial u}(x^k,u^k,t)$
- Mettre à jour $u^{k+1} = u^k - l^k \left(\frac{\partial H^k}{\partial u}\right)$ avec l^k satisfaisant les règles de Wolfe (11) et (12).

En pratique la fonction u^k sera représentée par un vecteur (de coefficients représentant une décomposition dans une base de fonctions) de dimension finie et le problème sera ainsi résolu de manière approchée. Les étapes de résolution des équations différentielles satisfaites par x et λ (équation en temps rétrograde) seront effectuées de manière numérique (par exemple par des schémas de Runge-Kutta) avec une précision finie. On peut adapter l'algorithme 22 pour tenir compte de contraintes finales. Il suffit d'introduire les équations du problème aux deux bouts correspondant et d'ajouter une étape de mise à jour des inconnues ν définies précédemment à la section C.2.2.

Afin d'obtenir une vitesse de convergence supérieure à celle de l'algorithme 22, on peut utiliser l'algorithme suivant

***Algorithme* 23 (Algorithme de tir).** — À partir de $\lambda^0(t_1) \in \mathbb{R}$ quelconque, itérer

– Calculer formellement u^{k+1} solution de $\frac{\partial H}{\partial u}(x^{k+1}, u^{k+1}, \lambda^{k+1}, t) = 0$ en fonction de x^{k+1}, λ^{k+1}, t.
– Résoudre le système

$$\begin{cases} \dot{x}^{k+1} = f(x^{k+1}, u^{k+1}, t) \\ \dot{\lambda}^{k+1}(t) = -\dfrac{\partial f}{\partial x}(x^{k+1}, u^{k+1}, t)\lambda^{k+1}(t) - \left(\dfrac{\partial L}{\partial x}(x^{k+1}, u^{k+1}, t)\right)^T \end{cases} \tag{208}$$

pour $t \in [t_1, t_2]$ avec comme condition initiale $x^k(t_1) = x^0$, $\lambda^k(t_1) = \lambda^k(t_1)$.
– Calculer la fonction de sensibilité $F \in \mathcal{M}_n(\mathbb{R})$ définie comme $F = \frac{\partial \lambda(t_2)}{\partial \lambda(t_1)}$
– Mettre à jour $\lambda^{k+1}(t_1) = \lambda^k(t_1) - l^k F^{-1}\left(\lambda(t_2) - \left(\frac{\partial}{\partial x(t_2)}\varphi(x(t_2), t_2)\right)^T\right)$ avec l^k satisfaisant les règles de Wolfe (11) et (12).

Cet algorithme a pour inconnue la seule valeur de l'état adjoint λ à l'instant initial. La commande u est calculée à chaque itération. Une valeur $\lambda(t_1)$ satisfait les conditions de stationnarité si elle fournit par intégration une valeur $\lambda(t_2) = \left(\frac{\partial}{\partial x(t_2)}\varphi(x(t_2), t_2)\right)^T$. Si ce n'est pas le cas, l'algorithme modifie la valeur candidate de l'inconnue pour tenter de satisfaire cette contrainte. Par rapport à l'algorithme 22, on constate en général une convergence plus rapide en pratique lorsque la valeur $\lambda^0(t_1) \in \mathbb{R}$ est proche de la valeur optimale. Si ce n'est pas le cas, l'algorithme 23 souffre souvent de l'instabilité numérique de la résolution simultanée des équations différentielles (208) satisfaites par x et λ. Cette instabilité est liée à l'instabilité du système linéarisé tangent. L'algorithme 23 peut être amélioré de nombreuses façons, on pourra se reporter à [**27**] pour un exposé sur les méthodes de tirs multiples.

C.2.4 Principe du minimum

Les calculs des variations menés jusqu'à présent ont consisté à calculer la variation de la valeur d'une fonctionnelle de deux variables x, u lorsque u était libre et x était liée à u par une équation différentielle $\dot{x} = f(x, u, t)$. Maintenant nous allons considérer que la variable u n'est pas libre mais contrainte. Nous regardons maintenant les problèmes de la forme

$$\begin{aligned} \min_{x,\,u} \quad & J(x, u) \\ \text{tel que } & \dot{x} = f(x, u), \\ & x(t_1) = x^0, \\ & C(u, t) \leq 0 \end{aligned} \tag{209}$$

où $\mathbb{R}^m \times \mathbb{R} \ni (u, t) \mapsto C(u, t) \in \mathbb{R}^l$, $l \in \mathbb{N}$ est une fonction de classe $\mathcal{C}^1$. On considérera des fonctionnelles du type

$$X \times U \ni (x, u) \mapsto J(x, u) = \varphi(x(t_2), t_2) + \int_{t_1}^{t_2} L(x(t), u(t), t)dt$$

où $\mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto \varphi(x,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$et $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \ni (x,t) \mapsto L(x,u,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$.

La preuve du résultat que nous allons énoncer est difficile. Nous nous contentons d'en suggérer quelques grandes lignes. Comme nous l'avons vu, la variation de la valeur de la fonctionnelle s'écrit en fonction d'une variation δu,

$$\delta J = \int_{t_1}^{t_2} \frac{\partial H}{\partial u} \delta u(t) dt \tag{210}$$

Autour d'un optimum (x^*, u^*), il ne doit pas exister de variation admissible, compatible avec les contraintes $C(u,t) \leq 0$ fournissant une variation négative δJ. En reprenant le point de vue utilisé dans la présentation des conditions de Karush, Kuhn et Tucker du théorème 26, on doit avoir que le gradient de chaque contribution ponctuelle dans l'intégrale (210) doit être dans le cône convexe des contraintes actives. Autrement dit, pour tout $t \in [t_1, t_2]$, il existe un vecteur $\mathbb{R} \supset [t_1, t_2] \ni t \mapsto \mu(t) \in \mathbb{R}^l$ dont les composantes vérifient $\mu_i(t) = 0$ si $C_i(x(t), u(t) < 0$, $\mu_i(t) \geq 0$ si $C_i(x(t), u(t) = 0$ tel que

$$\frac{\partial H}{\partial u} = -\mu^T \frac{\partial C}{\partial u}$$

Plus formellement le principe du minimum s'énonce

***Théorème* 48 (Principe du minimum de Pontryagin [26])**
Soient $\mathbb{R} \supset [t_1, t_2] \ni t \mapsto (x, u, \lambda)(t) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ optimum du problème (209)*. Pour tout $t \in [t_1, t_2]$, $u(t)$ réalise le minimum de $\mathbb{R}^m \ni u \mapsto H(x(t), u, \lambda(t), t) \in \mathbb{R}$ sous la contrainte $C(u,t) \leq 0$.*

C.3 Champs d'extrémales

Il est possible d'exhiber une structure reliant les trajectoires optimales de problèmes d'optimisation voisins. Le but de cette section est de présenter cette structure au travers de l'équation aux dérivées partielles de Hamilton-Jacobi-Bellman.

On considère une fois de plus le problème suivant (d'autres généralisations sont possibles)

$$\begin{aligned} \min_{x,u} \quad & J(x,u) \\ \text{tel que } & \dot{x} = f(x,u), \\ & x(t_1) = x^0, \\ & \psi(x(t_2), t_2) = 0 \end{aligned} \tag{211}$$

où $\mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto \psi(x,t) \in \mathbb{R}^q$, $1 < q \leq n$ est une fonction de classe $\mathcal{C}^1$. On considérera des fonctionnelles du type

$$X \times U \ni (x,u) \mapsto J(x,u) = \varphi(x(t_2), t_2) + \int_{t_1}^{t_2} L(x(t), u(t), t) dt$$

où $\mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto \varphi(x,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$et $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \ni (x,u,t) \mapsto L(x,u,t) \in \mathbb{R}$ est de classe $\mathcal{C}^1$.

***Définition* 32 (Extrémale).** — Une extrémale est l'ensemble de $\mathbb{R}^n$ des points $(x(t),t)$, $t \in [t_1,t_2]$ où $\mathbb{R} \supset [t_1,t_2] \ni t \mapsto x(t) \in \mathbb{R}^n$ est solution du problème d'optimisation (211).

***Définition* 33.** — Soit $\mathcal{E}$ une famille d'extrémales obtenue en faisant varier x^0, t_1. Soit U un ensemble compact de $\mathbb{R}^n \times \mathbb{R}$. On dit que $\mathcal{E}$ est un champs d'extrémales pour U si $\mathcal{E} \supset U$.

En conséquence, il existe une extrémale passant par tous les points de U.

***Définition* 34 (Fonction de retour optimal).** — On appelle fonction de retour optimal $\mathbb{R}^n \times \mathbb{R} \ni (x^0,t_1) \mapsto \mathcal{J}(x^0,t_1)$ (à l'ensemble $\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \text{ t.q. } \psi(x,t)=0\}$), la fonction ayant pour valeur la valeur optimale pour le problème d'optimisation (211) avec pour condition initiale (x^0,t_1).

Les extrémales vérifient la propriété suivante.

***Proposition* 18 (Principe d'optimalité de Bellman).** — *Une suite de décisions est optimale si, quel que soit l'état et l'instant considérés sur la trajectoire qui lui est associée, les décisions ultérieures constituent une suite optimale de décisions pour le sous problème dynamique ayant cet état et cet instant comme conditions initiales.*

Grâce à cette propriété, il est possible d'établir une équation aux dérivées partielles satisfaites par $\mathcal{J}$. On suppose disposer d'un champs d'extrémales $\mathcal{E}$ pour le problème d'optimisation (211). Considérons $(x,t) \in \mathcal{E}$. Il correspond à la fonction de retour optimal évaluée en ce point $J^0(x,t)$ une commande optimale $\mathbb{R} \supset [t,t_2] \ni l \mapsto u^0(l) \in \mathbb{R}^m$ que nous supposerons (pour simplifier) unique. Pour toute commande u proche de u^0 on peut calculer une fonction coût à partir de la trajectoire $\mathbb{R} \supset [t_1,t_2] \ni t \mapsto x(t) \in \mathbb{R}^n$ (proche de la trajectoire optimale) issue de $x(t_1)=x^0$. Il vient

$$J(x,u) \geq \mathcal{J}(x^0,t_1)$$

avec

$$\min_u J(x,u) = \mathcal{J}(x^0,t_1) \tag{212}$$

Choisissons comme candidat d'appliquer u une commande différente de la commande optimale u^0 pendant δt puis optimale sur $[t_1+\delta t, t_2]$. On peut évaluer

$$J(x,u) = \int_{t_1}^{t_1+\delta t} L(x(t),u(t),t)dt + \mathcal{J}(x(t_1+\delta t), t_1+\delta t)$$

Un développement au premier ordre donne

$$\begin{aligned} J(x,u) =& L(x^0,u(t_1),t_1)\delta t + \mathcal{J}(x^0,t_1) + \frac{\partial \mathcal{J}}{\partial x}(x^0,t_1) f(x^0,u(t_1),t_1)\delta t \\ &+ \frac{\partial \mathcal{J}}{\partial t}(x^0,t_1)\delta t + \circ(\delta t) \end{aligned}$$

L'équation (212) donne

$$\mathcal{J}(x^0,t_1) = \mathcal{J}(x^0,t_1) + \circ(\delta t) + \left(\min_u \left(L(x^0,u(t_1),t_1) + \frac{\partial \mathcal{J}}{\partial x}(x^0,t_1) f(x^0,u(t_1),t_1) \right) + \frac{\partial \mathcal{J}}{\partial t}(x^0,t_1) \right) \delta t$$

Un passage à la limite lorsque $\delta t \to 0$ donne

$$-\frac{\partial \mathcal{J}}{\partial t} = \min_u \left(L(x,u,t) + \frac{\partial \mathcal{J}}{\partial x} f(x,u,t) \right)$$

On reconnaît ici l'Hamiltonien (196) et on écrit finalement l'équation de Hamilton-Jacobi-Bellman

$$-\frac{\partial \mathcal{J}}{\partial t} = \min_u \left(H(x,u,\frac{\partial \mathcal{J}}{\partial x},t) \right) \tag{213}$$

Il s'agit d'une équation aux dérivées partielles hyperbolique dont la condition limite est donnée sur l'ensemble $\mathcal{V} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \text{ t.q. } \psi(x,t) = 0\}$ par $\mathbb{R}^n \times \mathbb{R} \ni (x,t) \mapsto \mathcal{J}(x,t) = \varphi(x,t)$. On la résoudra en général par une propagation (rétrograde) depuis l'ensemble $\mathcal{V}$. Cette résolution sera par exemple menée le long d'un réseau quadrillant le plan (lorsque x est à valeur dans $\mathbb{R}^2$) par la méthode de la programmation dynamique.

BIBLIOGRAPHIE

[1] A. Beck and M. Teboulle. Gradient-based algorithms with applications to signal recovery problems. *Convex Optimization in Signal Processing and Communications*, pages 42–88, 2010.

[2] Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.

[3] Dimitri P Bertsekas. *Convex optimization theory*. Athena Scientific Belmont, 2009.

[4] Dimitri P Bertsekas. *Convex optimization algorithms*. Athena Scientific Belmont, 2015.

[5] Robert G Bland. New finite pivoting rules for the simplex method. *Mathematics of operations Research*, 2(2):103–107, 1977.

[6] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[7] Haïm Brezis. *Analyse fonctionnelle: théorie et applications*, volume 91. Éditions Masson, 1983.

[8] John W Chinneck. *Feasibility and Infeasibility in Optimization:: Algorithms and Computational Methods*, volume 118. Springer Science & Business Media, 2007.

[9] Jean-Christophe Culioli. *Introduction à l'optimisation*. Ellipses, 1994.

[10] Iain S Duff and John K Reid. The design of ma48: a code for the direct solution of sparse unsymmetric linear systems of equations. *ACM Transactions on Mathematical Software (TOMS)*, 22(2):187–226, 1996.

[11] Philip E Gill, Walter Murray, and Michael A Saunders. Large-scale sqp methods and their application in trajectory optimization. In *Computational optimal control*, pages 29–42. Springer, 1994.

[12] Philip E Gill, Walter Murray, Michael A Saunders, and Margaret H Wright. User's guide for npsol 5.0: A fortran package for nonlinear programming. Technical report, Technical Report NA 98-2, Department of Mathematics, 1998.

[13] G. H. Golub and C. F. Van Loan. *Matrix computations*. The John Hopkins University Press, Baltimore, Third Ed.

[14] Ignacio E Grossmann. Review of non-linear mixed integer and disjunctive programming techiques for process systems engineering. *Optim. Eng*, 3, 2001.

[15] Jean-Baptiste Hiriart-Urruty. *Optimisation et analyse convexe*. Presses Universitaires de France, 1998.

[16] James E Kelley, Jr. The cutting-plane method for solving convex programs. *Journal of the society for Industrial and Applied Mathematics*, 8(4):703–712, 1960.

[17] Claude Lemaréchal. Chapter vii nondifferentiable optimization. *Handbooks in operations research and management science*, 1:529–572, 1989.

[18] David G Luenberger, Yinyu Ye, et al. *Linear and nonlinear programming*, volume 2. Springer, 1984.

[19] Marko Mäkelä. Survey of bundle methods for nonsmooth optimization. *Optimization methods and software*, 17(1):1–29, 2002.

[20] Michel Minoux. *Programmation mathématique. Théorie et algorithmes*. Dunod, 1983.

[21] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[22] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[23] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and trends in Optimization*, 1(3):123–231, 2013.

[24] Elijah Polak. *Optimization: algorithms and consistent approximations*, volume 124. Springer Verlag, 1998.

[25] BT Polyak. Convexity of nonlinear image of a small ball with applications to optimization. *Set-Valued Analysis*, 9(1-2):159–168, 2001.

[26] L. S. Pontryagin. The mathematical theory of optimal processes. 1962.

[27] S. M. Roberts and J. S. Shipman. Two-point boundary value problems: shooting methods. 1972.

[28] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 1970.

[29] V.M. Tokhomirov. Mathematical world series: Stories about maxima and minima. vol. 1. *The Mathematics Teacher*, 91(3):269, 1998.