Génie Mathématique et Modélisation
Quatrième Année

OPTIMISATION NON LINÉAIRE

Philippe GUILLAUME

2002/2003

Table des matières

Chapitre 1

Introduction

Un problème d'optimisation est un problème de la forme

$$\mathcal{P}\left\{\begin{array}{l}\text{Trouver } \overline{x} \text{ tel que}\\ f(\bar{x}) \leq f(x) \quad \forall x \in S,\end{array}\right.$$

où

- $f\ : X \longmapsto \mathbb{R}$ est une fonction définie sur un espace vectoriel normé X. Elle est appelée fonction objectif, fonction coût, critère, ...
- $S \subset X$, $S \neq \emptyset$, est l'ensemble des contraintes du problème, souvent défini par des égalités ou des inégalités.

La valeur $f(\bar{x})$ (si elle existe) est appelée <u>valeur minimale</u> ou <u>minimum</u>, et par abus de langage on dit que $\bar{x}$ est un <u>minimum</u> de f sur S.

<u>Remarque</u>: si on on a un problème de maximisation, il suffit de changer f en $-f$.

<u>Exemple</u>: Soit une membrane élastique horizontale fixée sur le bord Γ d'un ouvert $\Omega \subset \mathbb{R}^2$. La membrane est repérée par sa position : $z = u(x, y)$.
Deux forces s'exercent sur cette membrane, une force $\vec{F}$ et la réaction $\vec{T}$:

$$\begin{array}{rcl}\vec{F} &=& f\,\vec{d}\\ \vec{T} &=& \Delta u\,\vec{d}\end{array}$$

où $\Delta u = \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2}$ est la courbure de la membrane et $\vec{d}$ le vecteur unitaire vertical.

A l'équilibre, on a $\vec{F} + \vec{T} = \vec{0}$, ce qui donne :

$$\begin{array}{rcl}-\Delta u &=& f \text{ dans } \Omega\\ u &=& 0 \text{ sur } \Gamma.\end{array}$$

La théorie des équations aux dérivées partielles nous dit que u minimise dans $H_0^1(\Omega)$ l'énergie

$$J(u) = \frac{1}{2}\int_\Omega |\nabla u|^2 - f\,u\,dx,$$

ainsi u est solution du problème *sans contraintes*:

$$\begin{cases} \min J(u) \\ u \in H_0^1(\Omega) \end{cases}$$

Considérons maintenant le cas où un obstacle est placé sous la membrane, repéré par $z = s(x, y)$. Le problème devient alors solution du problème *avec contraintes*:

$$\begin{cases} \min J(u) \\ u \in H_0^1(\Omega) \\ u(x, y) \geq s(x, y) \end{cases}$$

Ces problèmes peuvent être résolus par une méthode d'éléments finis. Les valeurs nodales U_i, $1 \leq i \leq N$, qui sont les valeurs de u en chaque noeud intérieur d'un maillage du domaine Ω, sont solution dans le premier cas du problème

$$\min_{U \in R^N} \frac{1}{2}(AU, U) - (B, U)$$

où (AU, U) représente une approximation de $\int_\Omega |\nabla u|^2 \, dx$ et (B, U) une approximation de $\int_\Omega f \, u \, dx$.

Avec obstacle, le problème discretisé devient :

$$\begin{cases} \min \frac{1}{2}(AU, U) - (b, U) \\ U_i \geq s_i, \quad i \leq i \leq N \end{cases}$$

où $s_i \simeq s(x_i, y_i)$.

De tels problèmes amènent les questions suivantes :

Q1) existe-t-il des solutions ? si oui, a-t-on unicité ?

Q2) peut-on déterminer des conditions nécessaires et/ou suffisantes d'optimalité ?

Q3) comment trouver ces solutions ? quels algorithmes ?

Q4) existe-t-il au moins des suites minimisantes, c.a.d. des suites (x_n), non forcément convergentes, mais qui vérifient $\lim_{n \to +\infty} f(x_n) =$ minimum de f ?

Q5) quelle est la sensibilité de la solution par rapport à de faibles variations des données f et S ?

Il y a peu de cas où l'on sait répondre à la question Q1. Voici quelques exemples.

Théorèmes d'existence :

Exemple 1 (Théorème de Weierstrass) *Soit K un compact de $\mathbb{R}^N$, et f une fonction continue de K dans $\mathbb{R}$. Alors il existe $\bar{x} \in K$ tel que*

$$f(\bar{x}) \leq f(x) \quad \forall x \in K.$$

Exemple 2 *Soit* $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ *une fonction continue, et* $S \subset \mathbb{R}^N$ *un fermé non vide.*
On suppose que f *est coercive, c.a.d.* $\lim\limits_{\substack{|x|\to+\infty \\ x\in S}} f(x) = +\infty$.
Alors il existe $\bar{x} \in S$ *tel que* $f(\bar{x}) \leq f(x) \quad \forall x \in S$.

Rappelons les propriétés suivantes sur les fonctions convexes (cf. cours d'analyse fonctionnelle):

Propriété 1.1 *Toute fonction convexe sur un ouvert convexe de* $\mathbb{R}^N$ *est continue.*

Propriété 1.2 *Toute fonction convexe sur un ouvert convexe* $C \subset \mathbb{R}^N$ *admet en tout point* $a \in C$ *un hyperplan d'appui : il existe* $h \in \mathbb{R}^N$ *tel que*

$$f(x) - f(a) \geq h.(x-a) \quad \forall x \in C.$$

Si $\nabla f(a)$ *existe, alors* h *est unique et* $h = \nabla f(a)$.

Théorème d'unicité :

Exemple 3 *Soit* f *une fonction coercive et strictement convexe sur un fermé convexe* $C \subset \mathbb{R}^N$. *Alors* f *admet un* *minimum global unique, c.a.d. il existe un unique* $\bar{x} \in C$ *tel que*

$$f(\bar{x}) \leq f(x) \quad \forall x \in C.$$

Remarque : c'est le cas particulier des fonctions du type :

$$f(x) = \frac{1}{2}Ax.x - b.x$$

où b, $x \in \mathbb{R}^N$, et $A \in \mathcal{M}_N(\mathbb{R})$ est une matrice symétrique définie positive (SDP). Par exemple, si $C = \mathbb{R}^N$, minimiser f équivaut alors à résoudre le système $Ax = b$. On rappelle qu'une matrice symétrique $A \in \mathcal{M}_N(\mathbb{R})$ est dite définie positive si $Ax.x > 0 \quad \forall x \neq 0$, et semi-définie positive si $Ax.x \geq 0 \quad \forall x$.

Dans ce cours, on traitera surtout des questions Q2 et Q3.

Chapitre 2

Optimisation sans contraintes

Ce chapitre est consacré au problème suivant :

$$\mathcal{P}\left\{\begin{array}{l}\text{Trouver } \bar{x} \in \mathbb{R}^N \text{ tel que} \\ f(\bar{x}) \leq f(x) \quad \forall x \in \mathbb{R}^N \\ f \ : \ \mathbb{R}^N \longrightarrow \mathbb{R}\end{array}\right.$$

2.1 Conditions d'optimalité

2.1.1 Notations

Le produit scalaire usuel de deux vecteurs $x = (x_1, \ldots, x_N)^T$ et $y = (y_1, \ldots, y_N)^T$ de $\mathbb{R}^N$ est noté $x.y$, et la norme associée est notée $|x|$:

$$\begin{aligned} x.y &= \sum_{i=1}^{N} x_i\, y_i, \\ |x| &= \left(\sum_{i=1}^{N} x_i^2\right)^{1/2}. \end{aligned}$$

Soit $f \ : \ \mathbb{R}^N \longrightarrow \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. Sa différentielle $Df(a) \in \mathcal{L}(\mathbb{R}^N)$ est donnée par

$$Df(a) = (\partial_1 f(a), ..., \partial_N f(a)), \text{ où } \partial_i = \frac{\partial}{\partial x_i},$$

avec pour $h \in \mathbb{R}^N$:

$$Df(a)h = \sum_{i=1}^{N} \partial_i f(a)\ h_i.$$

Soit $((\ ,\ ))$ un produit scalaire sur $\mathbb{R}^N$. Par définition, le gradient de f au point a par rapport à ce produit scalaire est l'unique vecteur $\nabla f(a) \in \mathbb{R}^N$ qui vérifie :

$$\boxed{((\nabla f(a), h)) = Df(a)h}$$

Dans le cas particulier du produit scalaire usuel de $\mathbb{R}^N$, on a

$$\nabla f(a) = \begin{pmatrix} \partial_1 f(a) \\ \vdots \\ \partial_N f(a) \end{pmatrix}$$

$$\boxed{\begin{array}{c} \text{Pour le produit scalaire usuel de } \mathbb{R}^N\text{, on a } \nabla f(a) = Df^T(a): \\ \\ Df(a)h = \nabla f(a).h. \end{array}}$$

De même, la dérivée seconde est une forme bilinéaire : si f est deux fois dérivable au point a, on a :

$$D^2 f(a) = \left(\frac{\partial^2 f(a)}{\partial x_i \partial x_j} \right)_{i,j=1,..,N}$$

et pour $h, k \in \mathbb{R}^N$,

$$\boxed{D^2 f(a)\,(h,k) = \sum_{i,j=1}^{N} \frac{\partial^2 f(a)}{\partial x_i \partial x_j}\, h_i k_j}$$

D'après le théorème de Schwartz, $D^2 f(a)$ est symétrique.

Le Hessien $\nabla^2 f(a)$ est l'unique matrice qui vérifie :

$$\boxed{((\nabla^2 f(a)\,h, k)) = D^2 f(a)\,(h,k) \quad \forall\, h, k \in \mathbb{R}^N}$$

Dans le cas du produit scalaire usuel, on a

$$\boxed{\begin{array}{c} \nabla^2 f(a) = D^2 f(a) \\ D^2 f(a)\,(h,k) = \nabla^2 f(a) h.k. \end{array}}$$

2.1.2 Conditions d'optimalité

On dit qu'un point $a \in \mathbb{R}^N$ est un minimum local de f sur $\mathbb{R}^N$ s'il existe $\varepsilon > 0$ tel que

$$f(a) \leq f(x) \quad \forall x \in B(a, \varepsilon),$$

où $B(a, \varepsilon)$ est la boule de centre a et de rayon ε. Il est dit minimum local strict si l'inégalité est stricte pour $x \neq a$, $x \in B(a, \varepsilon)$.

Proposition 2.1 (Condition nécessaire) *Soit a un minimum local de f sur $\mathbb{R}^N$. On suppose que f est dérivable au point a. Alors $\nabla f(a) = 0$.*
Si f est deux fois dérivable au point a, alors $\nabla^2 f(a)$ est une matrice semi-définie positive.

Remarque : la condition n'est pas suffisante. Contre-exemple : $f(x) = x^3$ en $a = 0$.
Démonstration
Montrons le premier point (par l'absurde). Supposons que $\nabla f(a)$ soit non nul. On a

$$f(a - t\nabla f(a)) = f(a) - t\nabla f(a).\nabla f(a) + t\varepsilon(t)$$

avec $\varepsilon(t) \underset{t\to 0}{\longrightarrow} 0$. D'où, en tenant compte de $|\nabla f(a)|^2 > 0$, on a pour t petit et $t > 0$:

$$f(a - t\nabla f(a)) - f(a) = t(\varepsilon(t) - |\nabla f(a)|^2) < 0,$$

ce qui contredit le fait que a soit un minimum local. ■

Proposition 2.2 (Condition suffisante) *Soit a un point où f admet une dérivée seconde, et tel que :*

1) $\nabla f(a) = 0$,

2) $\nabla^2 f(a)$ *est définie positive.*

Alors a est un minimum local strict.

Démonstration On a

$$f(a+h) = f(a) + \underbrace{\nabla f(a).h}_{=0} + \frac{1}{2}\nabla^2 f(a)h.h + |h|^2\varepsilon(h)$$

avec $\varepsilon(h) \underset{h\to 0}{\longrightarrow} 0$. Le Hessien $\nabla^2 f(a)$ est SDP, donc ses valeurs propres sont réelles et strictement positives. Soit $\alpha > 0$ sa plus petite valeur propre. On a $\nabla^2 f(a)h.h \geq \alpha|h|^2$, d'où

$$f(a+h) - f(a) \geq |h|^2 \quad \underbrace{\left(\frac{\alpha}{2} + \varepsilon(h)\right)}_{>0 \text{ pour } h \text{ petit}}$$

et a est bien un minimum local strict. ■

Remarque : cette condition n'est pas nécessaire. Contre-exemple : $f(x) = x^4$ en $a = 0$.

CAS DES FONCTIONS CONVEXES :

Rappels :

1. Soit f une fonction convexe sur un convexe C de $\mathbb{R}^N$. Si $a \in C$ est un minimum local de f sur C, alors a est un minimum global.

2. Si f est une fonction de classe $\mathcal{C}^2$ sur $\mathbb{R}^N$, alors :
   f est convexe $\Leftrightarrow \nabla^2 f(x)$ est semi-définie positive $\quad \forall x \in \mathbb{R}^N$.
   $\nabla^2 f(x)$ est SDP $\Rightarrow f$ est strictement convexe $\quad \forall x \in \mathbb{R}^N$.

Proposition 2.3 *Soit f un fonction convexe sur $\mathbb{R}^N$. Si $\nabla f(a)$ existe et est nul, alors a est un minimum local (et donc global) de f sur $\mathbb{R}^N$ (la condition nécessaire est devenue suffisante).*
Si de plus $\nabla^2 f(a)$existe et est définie positive, alors a est un minimum global strict.

Démonstration
- D'après la propriété 2 des fonctions convexes, aplliquée ici avec $h = \nabla f(a) = 0$, on a $f(x) - f(a) \geq \nabla f(a).(x-a) = 0$, ce qui prouve le premier point.
- Supposons de plus que $\nabla^2 f(a)$existe et soit SDP. On sait déjà que a est un minimum local strict (proposition 2.2). Montrons que c'est minimum global strict.
Soit $x \neq a \in \mathbb{R}^N$ et $d = x - a$.. Pour t petit, $t > 0$, on a $f(a) < f(a+td)$. Or $a+td = (1-t)a + t(a+d)$ et f est convexe, donc $f(a+td) \leq (1-t)f(a) + tf(a+d)$. On en déduit que $f(a) < (1-t)f(a) + tf(x)$, c.a.d. $f(a) < f(x)$. ■

Exemple :
Soit $f(x) = \frac{1}{2}Ax.x - b.x$ avec A semi-définie positive. Les minima globaux sont tous les points x qui vérifient $\nabla f(x)= 0$, c.a.d. $Ax = b$. Si A est SDP, la solution est unique, et donne un minimum global strict.

2.2 Méthodes d'optimisation dans le cas différentiable

On peut distinguer au moins trois classes d'algorithmes d'optimisation, correspondant à des approches différentes.

PREMIÈRE APPROCHE : elle consiste à diminuer la valeur du critère en cherchant le long d'une "direction de descente" un point x_{k+1} tel que $f(x_{k+1}) < f(x_k)$. On appelle direction de descente au point x un vecteur d qui vérifie :

$$\text{il existe } \varepsilon > 0 \text{ tel que } 0 < t < \varepsilon \Rightarrow f(x+td) < f(x).$$

Si f est différentiable et si $\nabla f(x)\neq 0$, alors tous les vecteurs $d \neq 0$ tels que

$$\nabla f(x).d < 0$$

sont des directions de descente. En effet $f(x+td) = f(x) + t(\nabla f(x).d + \varepsilon(t))$ où $\nabla f(x).d + \varepsilon(t)$ est du signe de $\nabla f(x).d$ pour $|t|$ petit.

Les méthodes de descente consistent alors à choisir une direction de descente d, puis $t_k > 0$ tel que

$$f(x_k + t_k d_k) < f(x_k).$$

DEUXIÈME APPROCHE : on cherche directement des points stationnaires, c.a.d. des points où le gradient s'annule. Pour résoudre $\nabla f(x) = 0$, on emploie des méthodes de type Newton.

⚠ La condition $\nabla f(x) = 0$ n'est pas suffisante, on peut tomber sur un maximum !

Ces deux approches donnent lieu à un algorithme d'optimisation de la forme suivante (avec $t_k = 1$ pour la méthode de Newton pure) :

$$\begin{cases} x_0 \in \mathbb{R}^N \text{ un point de départ} ; \\ x_k \in \mathbb{R}^N \text{ le point courant} ; \\ \text{si } \nabla f(x_k) = 0 \\ \quad \text{arrêt de l'algorithme} ; \\ \text{sinon déterminer} \\ \quad d_k \in \mathbb{R}^N \text{ une direction} ; \\ \quad t_k > 0 \text{ un déplacement dans cette direction} ; \\ \quad x_{k+1} = x_k + t_k d_k \text{ le nouveau point courant.} \end{cases}$$

Il est entendu que les tests du genre "si $\nabla f(x_k) = 0$" sont en fait programmés "si $|\nabla f(x_k)| < eps$", où eps est un paramètre du programme.

<u>Troisième approche :</u> méthodes de région de confiance. Dans ces méthodes, on choisit une approximation $\widetilde{f}(x) \simeq f(x)$. Cette approximation dépend du point courant x_k, ce peut être par exemple l'approximation par un polynôme de Taylor d'ordre 1 ou 2. On choisit ensuite $r_k > 0$ et on détermine d_k qui minimise $\widetilde{f}(x_k + d_k)$ sous la contrainte $||d_k|| \le r_k$. Cette boule $\{x,\ ||x - x_k|| \le r_k\}$ est la région de confiance, c'est à dire la région où l'on a confiance en l'approximation de $f(x)$ par $\widetilde{f}(x)$. Le nouveau point devient alors

$$x_{k+1} = x_k + d_k.$$

Pour plus de détail, se reporter à [2, 7].

<u>Remarque :</u> en général,
- La première approche fait appel à la dérivée première de f,
- La deuxième approche fait appel aux dérivées première et seconde de f,
- pour la troisième, cela dépendra de l'approximation de f choisie. A priori, la seconde approche (et éventuellement la troisième) est donc plus coûteuse en nombre d'opérations. Nous verrons cependant des méthodes efficaces qui évitent le calcul de la dérivée seconde.

2.2.1 Méthode de la plus forte pente

Soit a un point tel que $\nabla f(a) \neq 0$. On appelle <u>direction de plus forte pente</u> un vecteur $d \in \mathbb{R}^N$ non nul qui minimise $\nabla f(a).d'$ parmi tous les vecteurs d' de même norme que d. Un tel vecteur est donné par exemple par

$$d = -\nabla f(a),$$

et c'est le seul qui a pour norme $|\nabla f(a)|$.

Démonstration
Soit d' un vecteur de norme $|\nabla f(a)|$. On a $|\nabla f(a).d'| \leq |\nabla f(a)||d'| = |\nabla f(a)|^2$, avec égalité ssi $d' = \pm\nabla f(a)$. On a alors

$$\nabla f(a).d = -|\nabla f(a)|^2 \leq \nabla f(a).d',$$

avec égalité ssi $d' = d$. ■

Cette direction est celle qui diminue le plus l'approximation d'ordre 1 du critère :

$$f(a + d') \simeq f(a) + \nabla f(a).d'.$$

⚠ Ce n'est pas forcément la "meilleure direction" (considérer le cas d'une vallée allongée).

La méthode de plus forte pente s'écrit alors :

$$\left\{\begin{array}{l} x_0 \text{ donné} \\ \text{itération } k: \\ \text{si } \nabla f(x_k) = 0 \\ \quad \text{stop} \\ \text{sinon} \\ \quad d_k = -\nabla f(x_k) \\ \quad \text{choix de } t_k > 0 \\ \quad x_{k+1} = x_k + t_k d_k. \end{array}\right.$$

Théorème 2.4 *On suppose f de classe $\mathcal{C}^1$, avec $f(x) \underset{|x|\to+\infty}{\longrightarrow} +\infty$ (i.e. f est coercive). On choisit t_k tel que $f(x_k + t_k d_k) \leq f(x_k + t d_k) \quad \forall t > 0$ (recherche linéaire). Alors :*
1) la suite (x_k) est bornée (donc admet des points d'accumulation),
2) Tout point d'accumulation est un point stationnaire de f.

Remarques : 1. c'est une méthode de descente,
2. deux directions consécutives sont orthogonales.

Démonstration de la remarque 2 :
Soit $g : t \longmapsto f(x_k + t d_k)$. Le nombre t_k minimise $g(t)$, donc $g'(t_k) = 0$, c.a.d.

$$0 = g'(t_k) = \nabla f(x_k + t_k d_k).dk = \nabla f(x_{k+1}).dk = -d_{k+1}.d_k.$$

■

Démonstration du théorème
1) D'après l'hypothèse de coercivité, il existe $R > 0$ tel que $|x| > R \Rightarrow f(x) > f(x_0)$. De $f(x_k) \leq f(x_0)$ $\forall k \geq 0$ (méthode de descente), on déduit que $|x_k| \leq R$ et la suite (x_k) est bornée.

2) Soit $\bar{x}$ un point d'accumulation : il existe une sous-suite $(x_{n_k})_{k\geq 0}$ telle que $\lim_{k\to+\infty} x_{n_k} = \bar{x}$. Montrons par l'absurde que $\nabla f(\bar{x}) = 0$.
Si $\nabla f(\bar{x}) \neq 0$, on peut trouver $\bar{t} > 0$ minimisant la fonction $t \longmapsto f(\bar{x} - t\nabla f(\bar{x}))$:

$$f(\bar{x} - \bar{t}\nabla f(\bar{x})) < f(\bar{x})$$

Par continuité, on a pour k assez grand :

$$f(x_{n_k} - \bar{t}\,\nabla f(x_{n_k})) < f(\bar{x}).$$

La suite $f(x_k)$ est décroisssante, d'où, en utilisant la définition de t_{n_k},

$$\begin{aligned} f(\bar{x}) &\leq f(x_{n_k+1}) = f(x_{n_k} + t_{n_k} d_{n_k}) \\ &\leq f(x_{n_k} + \bar{t} d_{n_k}) = f(x_{n_k} - \bar{t}\,\nabla f(x_{n_k})) \\ &< f(\bar{x}), \end{aligned}$$

ce qui est absurde. ■

<u>Cas où f est strictement convexe:</u>

Corollaire 2.5 *On suppose de plus que f est strictement convexe.*
Alors la suite $(x_k)_{k\geq 0}$ converge vers l'unique minimum global de f sur $\mathbb{R}^N$.

<u>Exemple :</u>
Reprenons la fonction $f(x) = \dfrac{1}{2}Ax.x - b.x$, où A est SDP. On a $d_k = b - Ax_k$, et $t_k = |d_k|^2/Ad_k.d_k$.
L'algorithme de plus forte pente est donc le suivant :

$$\left\{ \begin{array}{l} x_0 \text{ donné} \\ \text{itération } k : \\ d_k = -Ax_k + b \\ \text{si } d_k = 0 \\ \quad \text{stop} \\ \text{sinon} \\ \quad t_k = \dfrac{|d_k|^2}{Ad_k.d_k} \\ \quad x_{k+1} = x_k + t_k d_k \end{array} \right.$$

Cette méthode peut être assez lente, surtout si la matrice A est mal conditionnée. En posant $E_k = A(x_k - \bar{x}).(x_k - \bar{x}) = ||x_k - \bar{x}||_A^2$, on montre que (cf. cours analyse numérique des grands systèmes linéaires) :

$$E_k \leq E_0 \left(\frac{\frac{\beta}{\alpha} - 1}{\frac{\beta}{\alpha} + 1} \right)^{2k}$$

où α est la plus petite valeur propre de A et β est la plus grande. On est donc conduit à abandonner les méthodes de plus forte pente.

2.2.2 Méthode du gradient conjugué

Cette méthode a été introduite par Hestenes et Stiefel en 1952. Nous commençons par le cas des fonctions quadratiques

$$f(x) = \frac{1}{2} Ax.x - b.x$$

où A est une matrice SDP. Le minimum de f est noté $\bar{x}$, c'est la solution de $Ax = b$.

L'idée générale est la suivante. Soit $x_0 \in \mathbb{R}^N$ un point initial. On suppose $Ax_0 - b \neq 0$. On pose

$$\begin{aligned} u_0 &= Ax_0 - b \\ u_1 &= Au_0 \\ &\vdots \\ u_{j+1} &= Au_j, \end{aligned}$$

et on considère la suite des espaces de Krylov définis par

$$\mathcal{U}_k = Vect\{u_0, u_1, ..., u_k\}$$
$$\mathcal{U}_0 \subset \mathcal{U}_1 \subset \ldots \subset \mathcal{U}_k \subset \ldots$$

Pour $k \geq 0$, soit x_{k+1} *le point qui minimise f sur le sous-espace affine* $x_0 + \mathcal{U}_k$. Soit $K = \max\{j; (u_0, u_1, ..., u_j) \text{ est libre}\}$. On remarque que nécessairement $K \leq N - 1$.

Lemme 2.6 *L'algorithme ci-dessus converge en exactement $K+1$ itérations :*

$$x_{K+1} = \bar{x}, \ x_k \neq \bar{x} \quad \forall\, k \leq K.$$

Démonstration

On a $x_{k+1} \in x_0 + \mathcal{U}_k$, donc x_{k+1} s'écrit $x_{k+1} = x_0 + \sum_{j=0}^{k} c_j u_j$. Le point x_{k+1} minimise f sur $x_0 + \mathcal{U}_k$, donc la dérivée de f par rapport aux c_j est nulle :

$$\nabla f(x_{k+1}).u_j = 0, \ j = 0, \ldots, k. \tag{2.1}$$

D'autre part,

$$\begin{aligned} \nabla f(x_{K+1}) &= Ax_{K+1} - b \\ &= Ax_0 - b + \sum_{j=0}^{k} c_j A u_j \\ &= u_0 + \sum_{j=0}^{k} c_j u_{j+1}. \end{aligned}$$

Compte tenu de la définition de K, on a $u_{K+1} \in \mathcal{U}_K$, donc $\nabla f(x_{K+1}) \in \mathcal{U}_K$. D'après (2.1) avec $k = K$, $\nabla f(x_{K+1})$ est orthogonal à l'espace $\mathcal{U}_K$, donc $\nabla f(x_{K+1}) = 0$, c.a.d. $Ax_{K+1} - b = 0$, et $x_{K+1} = \bar{x}$. En utilisant le fait que la famille u_k, $0 \leq k \leq K$ est libre, on montre alors que $\nabla f(x_k) \neq 0$ pour $0 \leq k \leq K$. ■

Problème : Comment calculer x_{k+1} à partir des points précédents? Pour cela, nous commençons par étudier quelques propriétés de la suite (x_k) engendrée par l'algorithme. Posons

$$\begin{aligned} r_k &= \nabla f(x_k) \\ \Delta_k &= x_{k+1} - x_k. \end{aligned}$$

On a $r_{k+1} = Ax_{k+1} - b = A\Delta_k + Ax_k - b$, i.e.

$$r_{k+1} = r_k + A\Delta_k. \tag{2.2}$$

Par construction, on a $\Delta_k \in \mathcal{U}_k$ et $A\Delta_k \in \mathcal{U}_{k+1}$. De $r_0 = u_0$ et (2.2), on déduit par récurrence que $r_k \in \mathcal{U}_k$ pour tout $k \geq 0$.

L'optimalité de x_{k+1} se traduit par (cf. (2.1)):

$$r_{k+1}.y = 0 \quad \forall\, y \in \mathcal{U}_k \tag{2.3}$$

et en particulier $r_{k+1}.r_j = 0$ pour $0 \leq j \leq k$, ce qui signifie que

$$\boxed{\text{la famille } (r_j)_{j\geq 0} \text{ est orthogonale.}}$$

On a $r_{k+1} = r_k + A\Delta_k$ et $\Delta_j \in \mathcal{U}_j \subset \mathcal{U}_k$ $(j \leq k)$, et avec (2.3), on obtient pour $j < k$

$$0 = r_{k+1}.\Delta_j = r_k.\Delta_j + A\Delta_k.\Delta_j = 0 + A\Delta_k.\Delta_j$$

d'où $A\Delta_k.\Delta_j = 0$ pour $0 \leq j < k$ ce qui signifie que

$$\boxed{\text{la famille } (\Delta_j)_{j\geq 0} \text{ est } A\text{-conjuguée.}}$$

La matrice A est symétrique donc pour $0 \leq j \leq k-1$, on a avec (2.2) :

$$0 = A\Delta_k.\Delta_j = \Delta_k.A\Delta_j = \Delta_k.(r_{j+1} - r_j)$$

donc

$$\Delta_k.r_{k-1} = \ldots = \Delta_k.r_j = \ldots = \Delta_k.r_0.$$

Ces nombres *ne dépendent pas de* j, on les note α_k:

$$\alpha_k = \Delta_k.r_j \quad \forall\, j,\ 1 \leq j \leq k-1.$$

Ces relations vont nous permettre d'obtenir des relations de récurrence permettant de déterminer x_{k+1}.

On a $r_{K+1} = 0$ (car x_{K+1} est solution), et $r_k \neq 0 \quad \forall\, k \leq K$ (car x_k n'est pas solution).

Donc $(r_0, \ldots, r_k)$ est une base orthogonale de $\mathcal{U}_k$ pour tout $k \leq K$. D'autre part, on a $\Delta_k \in \mathcal{U}_k$, donc Δ_k s'écrit sous la forme

$$\Delta_k = \sum_{j=0}^{k} \frac{\Delta_k . r_j}{||r_j||} \frac{r_j}{||r_j||} = \alpha_k \sum_{j=0}^{k} \frac{r_j}{||r_j||^2},$$

d'où la relation de récurrence :

$$\frac{\Delta_k}{\alpha_k} = \frac{\Delta_{k-1}}{\alpha_{k-1}} + \frac{r_k}{||r_k||^2}.$$

Il reste à calculer les α_k :
Posons

$$q_k = \frac{\Delta_k}{\alpha_k} \qquad (q_{-1} := 0).$$

On a donc

$$q_k = q_{k-1} + \frac{r_k}{||r_k||^2}.$$

De (2.2) et (2.3) on déduit que $(r_k + A\Delta_k).r_k = 0$ et

$$\begin{aligned} A\Delta_k . r_k &= -||r_k||^2 \\ \alpha_k A q_k . r_k &= -||r_k||^2 \\ \alpha_k &= -\frac{||r_k||^2}{A q_k . r_k}. \end{aligned}$$

La relation de récurrence cherchée est donc

$$\begin{aligned} x_{k+1} &= x_k + \Delta_k = x_k + \alpha_k q_k \\ &= x_k + \alpha_k \left(\frac{r_k}{||r_k||^2} + q_{k-1}\right). \end{aligned}$$

Pour des raisons de stabilité numérique, on effectue le changement de variable suivant :

$$\begin{aligned} d_k &= -||r_k||^2 q_k \\ t_k &= -\frac{\alpha_k}{||r_k||^2} = -\frac{||r_k||^2}{A d_k . r_k} \end{aligned}$$

On a alors avec ce changement :

$$\begin{aligned} x_{k+1} &= x_k + t_k \left(-r_k + \frac{||r_k||^2}{||r_{k-1}||^2} d_{k-1}\right) \\ &= x_k + t_k d_k. \end{aligned}$$

L'algorithme s'écrit donc :

$$\begin{cases} x_0 \in \mathbb{R}^N \text{ donné, } d_0 = -\nabla f(x_0) = b - Ax_0, \\ \text{tant que } \nabla f(x_k) \neq 0, \text{ calculer} \\ \quad t_k = -\dfrac{||\nabla f(x_k)||^2}{Ad_k.\nabla f(x_k)} = \dfrac{||\nabla f(x_k)||^2}{Ad_k.d_k} \\ \quad x_{k+1} = x_k + t_k d_k \\ \quad \nabla f(x_{k+1}) = \nabla f(x_k) + t_k A d_k \\ \quad d_{k+1} = -\nabla f(x_{k+1}) + \dfrac{||\nabla f(x_{k+1})||^2}{||\nabla f(x_k)||^2} d_k. \end{cases}$$

Théorème 2.7 *L'algorithme ci-dessus converge en au plus N itérations. La solution est le point x_{K+1} où l'indice K correspond à la première occurrence d'un gradient nul :* $\nabla f(x_{K+1}) = 0$.

Le nombre d'opérations à l'étape k est $2N^2 + O(N)$ (un produit matrice vecteur et des produits scalaires), soit en tout au plus $2N^3$ opérations. Cette méthode peut-être vue comme une méthode directe, plus coûteuse toutefois qu'une factorisation de Cholesky ($O(\frac{N^3}{3})$ opérations). Mais cette méthode est plus stable pour les grands systèmes creux, surtout si l'on utilise un préconditionnement de la matrice A (cf. cours grands systèmes).

Si A n'est pas SDP, il existe une méthode voisine : le gradient biconjugué.

GÉNÉRALISATION À DES FONCTIONS NON QUADRATIQUES

Localement, une fonction f suffisamment régulière s'écrit

$$f(x+h) = f(x) + \nabla f(x).h + \frac{1}{2}\nabla^2 f(x)h.h + o(h^2),$$

et est "proche" d'une fonction quadratique. En utilisant la direction de descente du gradient conjugué, on obtient les deux méthode suivantes :

MÉTHODE DE FLETCHER-REEVES (1964)

$$\begin{cases} x_0 \text{ donné} \\ d_0 = -\nabla f(x_0) \\ \text{pour } k \geq 0, \text{ tant que } \nabla f(x_k) \neq 0 \\ \text{déterminer } t_k \text{ minimisant } f(x_k + t d_k)\ ,\ t > 0 \\ x_{k+1} = x_k + t_k\, d_k \\ d_{k+1} = -\nabla f(x_{k+1}) + \frac{||\nabla f(x_{k+1})||^2}{||\nabla f(x_k)||^2} d_k \end{cases}$$

VARIANTE DE POLAK-RIBIÈRE (1969)

$$d_{k+1} = -\nabla f(x_{k+1}) + \frac{\nabla f(x_{k+1}).(\nabla f(x_{k+1}) - \nabla f(x_k))}{||\nabla f(x_k)||^2} d_k.$$

Lorsque f est quadratique, ces deux algorithmes coïncident avec le gradient conjugué.

2.2.3 Méthodes de type Newton

Pour trouver un minimum local de f, on cherche directement des points stationnaires, i.e. des points $\bar{x}$ tels que $\nabla f(\bar{x}) = 0$. En posant $F(x) = \nabla f(x)$, on est amené à chercher les zéros d'une fonction $F : \mathbb{R}^N \longrightarrow \mathbb{R}^N$.

⚠ On peut aussi tomber sur un maximum local de f, ce n'est pas forcément une méthode de descente.

Méthode de Newton pour résoudre $F(x) = 0$

Soit une fonction $F : \mathbb{R}^N \longrightarrow \mathbb{R}^N$. Etant donné un point $x_k \in \mathbb{R}^N$, on a :

$$F(x_k + d) = F(x_k) + DF(x_k)d + o(d).$$

Puisqu'on cherche d tel que $F(x_k + d) = 0$, un bon choix pour d est de prendre $d = d_k$ défini par

$$F(x_k) + DF(x_k)d_k = 0,$$

ce qui donne l'algorithme suivant.

Algorithme de Newton :

$$\begin{cases} x_0 \text{ donné} \\ \text{pour } k \geq 0, \text{ si } DF(x_k) \text{ est inversible, résoudre} \\ DF(x_k)d_k = -F(x_k), \\ x_{k+1} = x_k + d_k. \end{cases}$$

Théorème 2.8 *(Convergence locale de la méthode de Newton)*
Soit $F : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ une fonction de classe C^1 et $\bar{x} \in \mathbb{R}^N$ tel que $F(\bar{x}) = 0$. On suppose que $DF(\bar{x})$ est inversible, et que DF est Lipschitzienne, c.a.d. qu'il existe L tel que $||DF(x) - DF(x')||_{\mathcal{L}(\mathbb{R}^N)} \leq L||x - x'||_{\mathbb{R}^N}$ pour tout x et x' dans $\mathbb{R}^N$. Alors il existe $r > 0$ tel que la suite (x_k) définie par $x_0 \in B(\bar{x}, r)$ et

$$x_{k+1} = x_k - DF(x_k)^{-1}F(x_k)$$

reste dans la boule $B(\bar{x}, r)$ et converge vers $\bar{x}$. De plus, la convergence est quadratique : il existe c tel que

$$||x_{k+1} - \bar{x}|| \leq c\,||x_k - \bar{x}||^2 \quad \forall\, k \geq 0.$$

Démonstration
Par continuité de $\det(DF)$, il existe $R > 0$ tel que $DF(x)$ soit inversible sur $\bar{B}(\bar{x}, R)$. Soit

$$M = \sup_{x \in \bar{B}(\bar{x}, R)} ||DF^{-1}(x)||_{\mathcal{L}(\mathbb{R}^N)}.$$

Soit $x_0 \in B(\bar{x}, R)$. On a

$$\begin{aligned} x_1 - \bar{x} &= x_0 - \bar{x} - DF(x_0)^{-1}(F(x_0) - F(\bar{x})) \\ &= -DF(x_0)^{-1}(\underbrace{F(x_0) - F(\bar{x}) - DF(x_0)(x_0 - \bar{x})}_{A}) \end{aligned}$$

avec

$$\begin{aligned} A &= \int_0^1 DF(\bar{x} + t(x_0 - \bar{x}))(x_0 - \bar{x}) - DF(x_0)(x_0 - \bar{x})\, dt \\ |A| &\leq \int_0^1 ||DF(\bar{x} + t(x_0 - \bar{x})) - DF(x_0)||\, ||x_0 - \bar{x}||\, dt \\ &\leq \int_0^1 L||\bar{x} + t(x_0 - \bar{x}) - x_0||\, ||x_0 - \bar{x}||\, dt \\ &= L\int_0^1 (1-t)\, ||x_0 - \bar{x}||^2\, dt \\ &= \frac{L}{2}||x_0 - \bar{x}||^2, \end{aligned}$$

d'où

$$||x_1 - \bar{x}|| \leq M\frac{L}{2}||x_0 - \bar{x}||^2.$$

Prenons $\alpha < 1$ et $r = \alpha \min\{\frac{2}{ML}, R\}$. On a alors, si $||x_0 - \bar{x}|| < r$,

$$||x_1 - \bar{x}|| < \alpha\, ||x_0 - \bar{x}||.$$

Par récurrence, la suite (x_k) reste dans la boule $\bar{B}(\bar{x}, r)$ avec

$$||x_k - \bar{x}|| < \alpha^k\, ||x_0 - \bar{x}||,$$

ce qui prouve la convergence. ■

Remarque : si $F(x) \neq 0$, on peut vérifier que la direction d donnée par la méthode de Newton : $DF(x)d = -F(x)$, est une direction de descente pour la fonction "mérite" $m(x) = ||F(x)||^2$, que nous retrouverons au chapitre 4 pour rśoudre les conditions KKT.

Méthodes de Newton approchées

Appliqué à $F(x) = \nabla f(x)$, la méthode de Newton s'écrit :

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k).$$

Inconvénients : A chaque pas, on doit :
- calculer de $\nabla^2 f(x_k)$,

- résoudre un nouveau système,
- bien que l'on espère que $\nabla^2 f(\bar{x})$ soit SDP (condition suffisante d'optimalité), il n'est pas sûr que $\nabla^2 f(x_k)$ soit SDP, et l'on n'a pas forcément une méthode de descente (on peut vérifier que si H est SDP et si $\nabla f(x_k) \neq 0$, alors $-H\nabla f(x_k)$ est une direction de descente).

Solutions envisageables :
- utiliser la même matrice pour plusieurs itérations (avec factorisation LU, Cholesky, ...)
- approcher $\nabla^2 f(x_k)$ (ou même $\nabla^2 f(x_k)^{-1}$) par une matrice plus simple et SDP.

Théorème 2.9 *Soit f une fonction de classe C^2 et $\bar{x}$ un minimum local. On suppose que $\nabla^2 f(\bar{x})$ est SDP. On se donne une suite de matrices $(A_k)_{k\geq 0}$ (pas forcément distinctes) et $\lambda < \frac{1}{2}$ tels que*

$$||A_k - \nabla^2 f(\bar{x})|| \leq \frac{\lambda}{||\nabla^2 f(\bar{x})^{-1}||}.$$

Alors il existe $r > 0$ tel que la suite définie par

$$\begin{array}{rcl} x_0 & \in & B(\bar{x}, r) \\ x_{k+1} & = & x_k - A_k^{-1}\,\nabla f(x_k) \end{array}$$

soit contenue dans $B(\bar{x}, r)$ et converge vers $\bar{x}$ géométriquement, i.e il existe $\beta < 1$ tel que

$$||x_k - \bar{x})|| \leq \beta^k\,||x_0 - \bar{x})||.$$

Ce type de méthode se met en fait sous la forme suivante :

- choix de x_0 et d'une matrice H_0 qui soit SDP (par exemple $H_0 = I$),
- A l'étape k :
  - choix de la direction : $d_k = -H_k \nabla f(x_k)$,
  - recherche linéaire : trouver $t_k > 0$ tel que $f(x_k + t_k d_k) < f(x_k)$,
  - $x_{k+1} = x_k + t_k d_k$,
  - si $\nabla f(x_{k+1}) \neq 0$, déterminer H_{k+1}.

Ici, $t_k H_k$ joue le rôle de A_k^{-1}, et est donc censé approcher l'inverse du Hessien $\nabla^2 f(x_k)$.

On impose deux conditions sur la réactualisation des matrices H_k :

i) H_k doit être une matrice SDP, ce qui assure que d_k est une direction de descente si $\nabla f(x_k) \neq 0$,

ii) $H_{k+1}(\nabla f(x_{k+1}) - \nabla f(x_k)) = x_{k+1} - x_k$.

La deuxième condition est imposée du fait que l'inverse du Hessien qu'est supposé approcher H_{k+1} vérifie cette même propriété : si on effectue un développement limité de $\nabla f(x)$ au point x_{k+1}, on a

$$\nabla f(x_k) - \nabla f(x_{k+1}) = \nabla^2 f(x_{k+1})(x_k - x_{k+1}) + o(x_k - x_{k+1}),$$

d'où

$$\nabla^2 f(x_{k+1})^{-1}(\nabla f(x_{k+1} - \nabla f(x_k)) \simeq x_{k+1} - x_k.$$

Il existe différentes méthodes de mise à jour de H_k, notamment du type :

$$H_{k+1} = H_k + D_k$$

où D_k est une matrice de rang un ou deux. Les deux méthodes les plus utilisées sont les suivantes.

Méthode DFP (Davidson, Fletcher, Power 1959-63)

$$\begin{aligned} H_{k+1} &= H_k + \frac{\Delta_k\, \Delta_k^T}{\Delta_k.\sigma_k} - \frac{H_k\sigma_k\, \sigma_k^T H_k}{H_k\sigma_k.\sigma_k}, \\ \Delta_k &= x_{k+1} - x_k, \\ \sigma_k &= \nabla f(x_{k+1}) - \nabla f(x_k). \end{aligned}$$

Proposition 2.10 *On suppose $\nabla f(x_k) \neq 0$.*

i) Si H_k est SDP et si t_k minimise $f(x_k - t\, H_k\, \nabla f(x_k))$ pour $t > 0$, alors $\sigma_k.\Delta_k > 0$.

ii) Si H_k est SDP et si $\sigma_k.\Delta_k > 0$, alors H_{k+1} est SDP.

iii) $H_{k+1}\, \sigma_k = \Delta_k$.

Démonstration

i) Par définition, t_k minimise $t \mapsto f(x_k - t\, H_k \nabla f(x_k))$ donc sa dérivée est nulle en t_k, et

$$\begin{aligned} \sigma_k.\Delta_k &= \nabla f(x_{k+1}).(x_{k+1} - x_k) - \nabla f(x_k).(x_{k+1} - x_k) \\ &= -\nabla f(x_k).(x_{k+1} - x_k) \\ &= t_k \nabla f(x_k).H_k \nabla f(x_k) > 0. \end{aligned}$$

ii) Soit $x \neq 0$. On a

$$H_{k+1}x.x = H_k x.x + \frac{(\Delta_k.x)^2}{\Delta_k.\sigma_k} - \frac{(H_k\sigma_k.x)^2}{H_k\sigma_k.\sigma_k},$$

qui est bien défini car $\sigma_k.\Delta_k > 0$ et H_k est SDP. Posons $u = H_k^{1/2}x$ et $v = H_k^{1/2}\sigma_k$. On a

$$\begin{aligned} H_{k+1}x.x &= ||u||^2 + \frac{(\Delta_k.x)^2}{\Delta_k.\sigma_k} - \frac{(u.v)^2}{||v||^2} \\ &= \underbrace{\frac{||u||^2\,||v||^2 - (u.v)^2}{||v||^2}}_{\geq 0 \text{ d'après Cauchy-Schwartz}} + \underbrace{\frac{(\Delta_k.x)^2}{\Delta_k.\sigma_k}}_{\geq 0} \geq 0. \end{aligned}$$

Le premier terme du second membre est nul ssi u et v sont colinéaires, auquel cas le second terme est strictement positif. Si le premier terme est non nul, alors le tout est aussi strictement positif.
iii) Simple vérification. ■

Méthode BFGS (Broyden, Fletcher, Goldfarb, Shannon, 1969-70)

Si dans DFP, on intervertit σ_k et Δ_k, au lieu d'avoir $H_{k+1}\sigma_k = \Delta_k$, on définit une suite G_k qui vérifie $G_{k+1}\Delta_k = \sigma_k$, qui est l'inverse de la relation souhaitée. C'est donc G_k^{-1} qui nous intéresse comme approximation de l'inverse du Hessien. En inversant la relation DFP, on obtient alors la formule suivante (avec $H_k = G_k^{-1}$):

$$\boxed{H_{k+1} = H_k + \left(1 + \frac{H_k\sigma_k.\sigma_k}{\sigma_k.\Delta_k}\right)\frac{\Delta_k\ \Delta_k^T}{\Delta_k.\sigma_k} - \frac{1}{\sigma_k.\Delta_k}\left(\Delta_k\ \sigma_k^T H_k + H_k\sigma_k\ \Delta_k^T\right).}$$

Proposition 2.11 *Si H_k est SDP et si $\sigma_k.\Delta_k > 0$, alors H_{k+1} est SDP.*

Cette deuxième méthode est considérée aujourd'hui comme étant la plus performante.

Méthode de Gauss-Newton

Cette méthode s'applique aux problèmes de moindre carrés, où l'on cherche à minimiser une fonction de la forme

$$f(x) = \frac{1}{2}\sum_{j=1}^{m} r_j(x)^2 = \frac{1}{2}||r(x)||^2.$$

On suppose que r est de classe $\mathcal{C}^2$ de $\mathbb{R}^N$ dans $\mathbb{R}^m$, et l'on note $J(x) = Dr(x)$ la Jacobienne de r. Le gradient et la Hessienne de f sont donnés par

$$\begin{aligned} \nabla f(x) &= J(x)^T r(x) \\ \nabla^2 f(x) &= J(x)^T J(x) + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x). \end{aligned}$$

La méthode de Gauss-Newton consiste à appliquer la méthode de Newton en ignorant le second terme de la Hessienne de f. Ceci présente l'avantage de ne pas avoir à calculer les dérivées secondes de r, et se justifie si les termes $r_j(x)\nabla^2 r_j(x)$ sont négligeables, soit parce que $r_j(x)$ est proche de 0 (ce que l'on espère réaliser), soit parce que les variations secondes de r sont faibles. L'algorithme de Gauss-Newton est donc le suivant :

- choix de x_0,
- A l'étape k :
  - calcul de la direction par résolution du système

    $$J(x_k)^T J(x_k) d_k = -J(x_k)^T r(x_k)$$

  - recherche linéaire : trouver $t_k > 0$ tel que $f(x_k + t_k d_k) < f(x_k)$
  - $x_{k+1} = x_k + t_k d_k$
  - test d'arrêt.

Remarquons que si la matrice $J(x_k)$ est de rang N (ce qui suppose $m \geq N$), alors la matrice $J(x_k)^T J(x_k)$ est inversible et le système linéaire ci-dessus a une solution unique. Dans ce cas, d_k est une direction de descente si $\nabla f(x_k) \neq 0$: en effet, on a

$$\nabla f(x_k).d_k = J(x_k)^T r(x_k).d_k = -J(x_k)^T J(x_k) d_k.d_k = -||J(x_k)d_k||^2 \leq 0$$

avec égalité ssi $d_k = 0$, ce qui n'a pas lieu si $\nabla f(x_k) \neq 0$.

Méthode de Levenberg-Marquardt

Il s'agit là d'une variante de la méthode précédente, où la résolution de

$$J(x_k)^T J(x_k) d_k = -J(x_k)^T r(x_k)$$

est remplacée par celle d'un système de la forme

$$(J(x_k)^T J(x_k) + \lambda_k I) d_k = -J(x_k)^T r(x_k)$$

avec $\lambda_k > 0$. Cette modification est apportée lorsque l'on est pas certain que la matrice $J(x_k)^T J(x_k)$ soit inversible. Cependant, comme celle-ci est au moins semie-définie positive, la matrice $J(x_k)^T J(x_k) + \lambda_k I$ sera inversible (car symétrique définie positive) pour tout $\lambda > 0$. Dans ce cas, on peut vérifier ici encore que d_k est une direction de descente si $\nabla f(x_k) \neq 0$, et de plus tend vers la direction de plus forte pente lorsque λ_k tend vers l'infini. Tout l'art consiste alors à choisir ce λ_k de la manière la plus efficace, on pourra pour cela consulter [5, 7]. Cette méthode peut par ailleurs s'interpréter en terme de région de confiance.

2.3 Autres méthodes

On s'interesse ici aux cas où f est non différentiable, ou bien est différentiable mais $\nabla f(x)$ n'est pas disponible. Il est fréquent par exemple que l'on ait à minimiser des fonctions du type $f(x) = \max_i f_i(x)$. De telles fonctions ne sont pas partout différentiables, même si chacune des fonctions f_i est différentiable. Par contre, si chacune des fonctions f_i est convexe, alors la fonction f est aussi convexe. Cette section est essentiellement consacrée au cas des fonctions convexes.

2.3.1 Méthode de relaxation

Cette méthode consiste à prendre comme directions de descente les vecteurs de la base canonique $(e_i)_{i=0}^{N-1}$ de manière cyclique (variante : on change de base à chaque cycle par rotation des axes, c'est la méthode de Rosenbrock).
L'algorithme est le suivant :

$$\begin{cases} x_0 \text{ donné} \\ x_{k+1} = x_k + t_k d_k \\ d_k = e_{[k]}, [k] = k \, mod \, N \\ t_k \text{ minimise } t \mapsto f(x_k + t d_k). \end{cases}$$

Proposition 2.12 *Si A une matrice SDP et si $f(x) = \frac{1}{2}Ax.x + bx + c$, alors l'algorithme ci-dessus converge.*

Remarque : dans ce cas, on constate qu'un cycle de cette méthode correspond à la méthode de Gauss-Seidel appliquée à la résolution de $Ax = -b$.

2.3.2 Méthode de sous-gradient pour les fonctions convexes

Définition 2.1 *Soit f une fonction convexe sur $\mathbb{R}^N$. On appelle dérivée directionnelle de f dans la direction $d \in \mathbb{R}^N$ le nombre*

$$f'(x,d) = \lim_{\substack{t \to 0 \\ t>0}} \frac{f(x+td)-f(x)}{t}.$$

Rappels :
Soit g convexe sur $\mathbb{R}$, $a < b < c$, on a :

$$\frac{g(b)-g(a)}{b-a} \leq \frac{g(c)-g(a)}{c-a} \leq \frac{g(c)-g(b)}{c-b}, \tag{2.4}$$

d'où l'on déduit que les dérivées à gauche et à droite $g'_-(a)$ et $g'_+(a)$ existent, avec $g'_-(a) \leq g'_+(a)$ et

$$g'_\pm(a)(c-a) \leq g(c)-g(a), \quad \forall c \in \mathbb{R}. \tag{2.5}$$

Appliqué à $g(t) = f(x+td)$, f convexe sur $\mathbb{R}^N$, cela donne

$$f'(x,d) \leq f(x+d)-f(x), \quad \forall d \in \mathbb{R}^N. \tag{2.6}$$

On remarque donc que si $s \in [g'_-(a), g'_+(a)]$, alors

$$s(c-a) \leq g(c)-g(a), \quad \forall c \in \mathbb{R}.$$

L'intervalle $[g'_-(a), g'_+(a)]$, est appelé sous-différentiel de g au point a et est noté $\partial g(a)$. Cette notion se généralise de la manière suivante.

Définition 2.2 *On appelle sous-différentiel au point x de f convexe sur $\mathbb{R}^N$ l'ensemble*

$$\partial f(x) = \{s \in \mathbb{R}^N \; ; \; s.d \leq f(x+d)-f(x) \quad \forall d \in \mathbb{R}^N\}$$

Un vecteur $s \in \partial f(x)$ est appelé sous-gradient de f au point x.

Proposition 2.13 *(Propriétés du sous-différentiel).*

1. *$\partial f(x) \neq \emptyset$ (existence d'un hyperplan d'appui).*
2. *Si f est différentiable au point x, alors $\partial f(x)$ est réduit à un point :*

$$\partial f(x) = \{\nabla f(x)\}.$$

3. *$\partial f(x) = \{s \in \mathbb{R}^N \; ; \; s.d \leq f'(x,d) \quad \forall d \in \mathbb{R}^N\}$.*
4. *$\partial f(x)$ est un ensemble convexe compact.*

Démonstration de 3.
Si $s \in \partial f(x)$, alors $s.td \leq f(x+td) - f(x)$ pour tout d, et pour $t > 0$, on a $s.d \leq \frac{f(x+td)-f(x)}{t}$. En passant à la limite, on obtient $s.d \leq f'(x,d)$ pour tout d. Réciproquement, si $s.d \leq f'(x,d)$ pour tout d, alors on a d'après (2.6) $s.d \leq f(x+d) - f(x)$ pour tout d, et par définition $s \in \partial f(x)$. ■

Théorème 2.14 *Soit f une fonction convexe sur $\mathbb{R}^N$. On a*

$$f'(x,d) = \max\{s.d \;;\; s \in \partial f(x)\}.$$

Proposition 2.15 *Soit f une fonction convexe sur $\mathbb{R}^N$ et $d \in \mathbb{R}^N$. Alors d est une direction de descente au point x ssi*

$$s.d < 0, \quad \forall s \in \partial f(x).$$

Démonstration
Si d est une direction de descente, il existe $t > 0$ tel que $f(x+td) < f(x)$. Soit $s \in \partial f(x)$. En utilisant la proposition 2.13 et la relation (2.6), on a $s.td \leq f'(x,td) \leq f(x+td) - f(x)$, d'où $s.d < 0$.
Réciproquement, si $s.d < 0$ pour tout $s \in \partial f(x)$, en utilisant la compacité du sous-différentiel et le théorème 2.14, on a $f'(x,d) = \max\{s.d \;;\; s \in \partial f(x)\} < 0$. Comme $f'(x,d) = \lim\limits_{t \to 0, t>0} \dfrac{f(x+td) - f(x)}{t}$ on a $f(x+td) - f(x) < 0$ pour $|t|$ petit, $t > 0$, et d est une direction de descente. ■

Algorithmes de sous-gradient
Ils sont de la forme :

$$\left\{ \begin{array}{l} \text{choix de } x_0, \\ \text{A l'étape } k \text{ déterminer un sous-gradient } s_k \in \partial f(x_k), \\ \text{si } s_k = 0, \text{ stop} \\ \text{sinon } x_{k+1} = x_k - t_k \dfrac{s_k}{||s_k||}. \end{array} \right.$$

Problème : Comment choisir un sous-gradient ?

Exemple : Soit $f(x) = \max\limits_{1 \leq i \leq p} f_i(x)$ où les f_i sont convexes différentiables.
Soit $I(x) = \{i \;;\; f_i(x) = f(x)\}$ alors

$$\nabla f_i(x) \in \partial f(x), \quad \forall i \in I(x).$$

En effet en se restreignant à la droite $\mathbb{R}\, d$, on a $f'(x,d) = \max\limits_{i \in I(x)} \nabla f_i(x).d$, donc $\nabla f_i(x).d \leq f'(x,d)$ pour tout $i \in I(x)$ et $\nabla f_i(x) \in \partial f(x)$.

Remarques :
- on peut montrer que $\partial f(x) =$ enveloppe convexe des $\nabla f_i(x)$.
- ⚠ l'opposé d'un sous gradient (donc de la forme $-\nabla f_i(x),\ i \in I(x)$) n'est pas forcément une direction de descente.

Problème : Comment choisir les t_k ? Plusieurs choix ont été proposés :

1. Méthode de la série divergente :

   on prend une suite $(t_k)_{k\geq 0}$, $t_k > 0$, telle que $\lim\limits_{k\to\infty} t_k = 0$ et $\sum\limits_{k=0}^{\infty} t_k = +\infty$.

   Proposition 2.16 *On suppose l'ensemble D des solutions non vide et borné (c'est alors un convexe compact). Soit $m = \min\limits_x f(x)$. Alors*
   - soit il existe $k \geq 0$ tel que $x_k \in D$ (donc $f(x_k) = m$),
   - soit $\lim\limits_{k\to\infty} f(x_k) = m$ et $\lim\limits_{k\to\infty} dist(x_k, D) = 0$.

   Cette méthode peut s'avérer assez lente.

2. Méthode à pas constant :

$$t_k = \text{constante}.$$

3. Méthode de la série convergente :

$$t_k = t_0\ \alpha^k,\ 0 < \alpha < 1.$$

4. Méthode de relaxation :

$$t_k = \rho \frac{f(x_k) - \bar{f}}{||s_k||}$$

   où $0 < \rho \leq 2$ et $\bar{f}$ est une estimation du minimum.

   Pour l'étude de convergence, cf. [6].

Chapitre 3

Conditions d'optimalité avec contraintes

3.1 Ensemble des contraintes

Soit $f : \mathbb{R}^N \longrightarrow \mathbb{R}$. On se propose de résoudre les problèmes de la forme :

$$\mathcal{P}\left\{\begin{array}{l}\text{Trouver } \overline{x} \in S \text{ tel que}\\ f(\bar{x}) \leq f(x) \quad \forall x \in S,\end{array}\right.$$

où $S \subset \mathbb{R}^N$ ($S \neq \emptyset$) est appelé "ensemble des contraintes" ou "ensemble des solutions admissibles", et est représenté par des contraintes qui sont du type :

1. Inégalités

   $$S = \{x \in \mathbb{R}^N \,;\ g_i(x) \leq 0,\ i = 1, 2, \ldots, p\}$$

   avec $g_i : \mathbb{R}^N \longrightarrow \mathbb{R}$.
   Avec la convention suivante :

   $$\begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \leq 0 \Longleftrightarrow y_i \leq 0 \quad \forall i = 1, 2, \ldots, p,$$

   et en posant

   $$g : \mathbb{R}^N \longrightarrow \mathbb{R}^p, \quad g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_p(x) \end{pmatrix},$$

   on a

   $$S = \{x \in \mathbb{R}^N \,;\ g(x) \leq 0\}.$$

2. <u>Egalités</u>
$$S = \{x \in \mathbb{R}^N;\ h_i(x) = 0,\ i = 1, 2, \ldots, q\}$$
avec
$$h_i \ :\ \mathbb{R}^N \ \longrightarrow\ \mathbb{R}.$$
En posant
$$h\ :\ \mathbb{R}^N \longrightarrow \mathbb{R}^q, \quad h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_q(x) \end{pmatrix},$$
on a
$$S = \{x \in \mathbb{R}^N\ ;\ h(x) = 0\}.$$

3. <u>Mixte</u>
$$S = \{x \in \mathbb{R}^N\ ;\ g(x) \leq 0\ ;\ h(x) = 0\}$$
avec
$$g\ :\ \mathbb{R}^N \ \longrightarrow\ \mathbb{R}^p$$
$$h\ :\ \mathbb{R}^N \ \longrightarrow\ \mathbb{R}^q$$

On peut d'ailleurs toujours se ramener au cas d'inégalités en considérant $h(x) \leq 0$ et $-h(x) \leq 0$ au lieu de $h(x) = 0$.

<u>Cône tangent</u>

Rappelons qu'un sous-ensemble T de $\mathbb{R}^N$ est un *cône* si pour tout $x \in T$ et tout $a > 0$, on a $ax \in T$.

Soit $S = \{x\ ;\ g(x) \leq 0\}$ et $x \in S$. On considère l'ensemble $\mathcal{A}$ des arcs paramétrés dérivables de la forme

$$\begin{array}{rcl} s\ :\ [0, a[& \longrightarrow & S \quad (a > 0), \\ t & \longmapsto & s(t), \\ s(0) & = & x. \end{array}$$

Définition 3.1 *On appelle cône tangent à S au point x, l'ensemble*
$$T(S, x) = \{s'(0);\ s \in \mathcal{A}\}.$$

$T(S, x)$ est parfois appelé *ensembles des directions admissibles.*

⚠ On n'a pas forcément $x + td \in S$ pour $t > 0$ et $d \in T(x)$.

<u>Problème :</u> Caractériser $T(S, x)$ en utilisant les $\nabla g_i(x)$.

3.2 Contraintes d'égalités

On considère ici le cas

$$S = \{x \in \mathbb{R}^N \; ; \; h(x) = 0\}, \quad h \; : \; \mathbb{R}^N \longrightarrow \mathbb{R}^q.$$

Définition 3.2 *Si h est de classe C^1 sur un voisinage de $x \in S$, et si les vecteurs $\nabla h_i(x)$, $i = 1, 2, \ldots, q$ sont linéairement indépendants, on dit que S est régulier au point x, ou que x est régulier. Si S est régulier en tout point $x \in S$, on dit que S est régulier.*

Remarques : $Dh(x)^T = (\nabla h_1(x), \ldots, \nabla h_q(x)) \in \mathcal{M}_{Nq}(\mathbb{R})$.
Si x est régulier, alors $q \leq N$.
Si $q \leq N$, alors x est régulier ssi $Dh(x)$ est de rang maximal.

$$S \text{ est régulier au point } x \quad \Leftrightarrow \left\{ \begin{array}{c} q \leq N \\ \text{rang}(Dh(x)) = q \end{array} \right.$$

Proposition 3.1 *Soit $x \in S$. Si x est régulier, alors*

$$T(S, x) = \text{Ker}\, Dh(x).$$

Démonstration

- Montrons $T(S, x) \subset \text{Ker}\, Dh(x)$.

Soit $d \in T(S, x)$. Par définition de $T(S, x)$, il existe

$$s : [0, a[\rightarrow S \text{ avec } s(0) = x, \; s'(0) = d.$$

On a $s(t) \in S$ $(0 \leq t < a)$, donc $h(s(t)) = 0$ pour tout $t \in [0, a[$. En dérivant, on a $Dh(s(t))s'(t) = 0$ et pour $t = 0$, on obtient $Dh(s(0))s'(0) = 0$, i.e. $Dh(x)d = 0$.

- Montrons $T(S, x) \supset \text{Ker}\, Dh(x)$.

Soit $d \in \text{Ker}\, Dh(x)$. Il faut construire un arc s qui a pour vitesse à l'origine d. Pour cela on utilise le théorème des fonctions implicites.
L'ensemble S est régulier en x donc $Dh(x)$ est de rang q, donc possède q colonnes indépendantes. Quitte à renuméroter les variables, on peut supposer que ce sont les q dernières. Ecrivons h de la manière suivante :

$$\begin{array}{rccc} h \; : & \mathbb{R}^{N-q} \times \mathbb{R}^q & \longrightarrow & \mathbb{R}^q \\ & y = (u, v) & \longmapsto & h(u, v) = h(y), \end{array}$$

et posons $x = (u_0, v_0)$. La fonction h est de classe $\mathcal{C}^1$ au voisinage de x, on a $h(x) = 0$, et la matrice $D_v h(u_0, v_0) = [Dh(x)]_{j=N-q+1, \ldots, N}^{i=1, \ldots, q}$ est inversible. D'après le théorème des fonctions implicites, il existe un voisinage U de u_0, un voisinage V de v_0, une application $\varphi : U \to V$ tels que

$$\left\{ \begin{array}{l} (u,v) \in U \times V \\ h(u,v) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} (u,v) \in U \times V \\ v = \varphi(u) \end{array} \right.$$

En posant $\psi(u) = (u, \varphi(u))$, on a

$$\begin{array}{rcl} \psi(u_0) & = & (u_0, \varphi(u_0)) = (u_0, v_0) = x, \\ \psi(u) & \in & S, \quad \forall u \in U, \\ h(\psi(u)) & = & 0, \quad \forall u \in U. \end{array}$$

La dérivée de $u \to h(\psi(u))$ est nulle, et en particulier

$$Dh(x)\ D\psi(u_0) = 0.$$

Donc $\operatorname{Im} D\psi(u_0) \subset \operatorname{Ker} Dh(x)$. Sachant que $D\psi(u_0) = \begin{pmatrix} I_{N-q} \\ D\varphi(u_0) \end{pmatrix}$, donc que $D\psi(u_0)$ est de rang $N-q$, et que dim(Ker $Dh(x)$)= $N-q$, on a l'égalité

$$\operatorname{Im} D\psi(u_0) = \operatorname{Ker} Dh(x).$$

Comme $d \in \operatorname{Ker} Dh(x)$, il existe $z \in \mathbb{R}^{N-q}$ tel que $D\psi(u_0)z = d$. En posant $s(t) = \psi(u_0 + tz)$ et en prenant $a > 0$ tel que $s(t) \in S$ pour $t \in]-a, a[$, on a $s(0) = x$, $s'(0) = D\psi(u_0)z = d$. ■

3.2.1 Conditions d'optimalité du premier ordre

Théorème 3.2 (Lagrange, condition nécessaire d'optimalité) *Soit $x \in S$ un minimum local de f sur S. On suppose le point x régulier et f dérivable au point x.*
Alors il existe des nombres $\lambda_1, \ldots, \lambda_q$, appelés multiplicateurs de Lagrange, tels que

$$\nabla f(x) + \sum_{i=1}^{q} \lambda_i \nabla h_i(x) = 0.$$

De plus, ces coefficients λ_i sont déterminés de manière unique.

Démonstration
Soit $d \in \operatorname{Ker} Dh(x)$. On a vu dans la démonstration de la proposition 3.1 qu'il existe $a > 0$ et $s\ :]-a, a[\longrightarrow S$ tel que $s(0) = x$ et $s'(0) = d$. Soit

$$g(t) = f(s(t)).$$

Le point 0 est un minimum local de g, donc $g'(0) = 0$ i.e.

$$0 = g'(0) = \nabla f(x).d.$$

Donc $\nabla f(x) \in (\mathrm{Ker}\, Dh(x))^\perp$. Comme $(\mathrm{Ker}\, Dh(x))^\perp = \mathrm{Im}\, Dh(x)^T$, il existe $\lambda \in \mathbb{R}^q$ tel que $\nabla f(x) + Dh(x)^T \lambda = 0$, ou encore

$$\nabla f(x) + \sum_{i=1}^{q} \lambda_i \nabla h_i(x) = 0.$$

De plus les λ_i sont uniques car les gradients $\nabla h_i(x)$ sont linéairement indépendants.

■

<u>Remarque</u> : un tel point x peut aussi être un <u>maximum local</u>. La condition est seulement nécessaire.

<u>Formulation Lagrangienne</u>

On pose

$$\begin{array}{rcl} L : \mathbb{R}^N \times \mathbb{R}^q & \longrightarrow & \mathbb{R} \\ (x, \lambda) & \longmapsto & L(x,\lambda) = f(x) + h(x).\lambda. \end{array}$$

Si $x \in S$, $\quad L(x,\lambda) = f(x)$ car $h(x) = 0$ sur S. On a

$$\nabla L(x,\lambda) = \begin{pmatrix} \nabla_x L(x,\lambda) \\ \nabla_\lambda L(x,\lambda) \end{pmatrix} = \begin{pmatrix} \nabla f(x) + Dh(x)^T \lambda \\ h(x) \end{pmatrix}.$$

D'où

$$\boxed{x \text{ minimum local sur } S \Rightarrow \nabla L(x,\lambda) = 0.}$$

<u>Conséquences</u> :

1. Les minima locaux sont à chercher :
   - parmi les points non réguliers,
   - parmi les points réguliers qui vérifient les conditions de Lagrange.
2. On est ramené à un problème de recherche des zéros sur $\mathbb{R}^{N+q}$ de la fonction $(x,\lambda) \mapsto \nabla L(x,\lambda)$, pour lequel on peut employer une méthode de type Newton.

3.2.2 Conditions d'optimalité du deuxième ordre

Théorème 3.3 *Soit $x \in S$ un point régulier. On suppose f et h deux fois différentiables au point x.*

a) (CN) On suppose que x est un minimum local de f sur S. Il existe donc $\lambda \in \mathbb{R}^q$ tel que $\nabla f(x) + Dh(x)^T\lambda = 0$. On a alors

$$\nabla_x^2 L(x, \lambda)d.d \geq 0 \quad \forall d \in T(S, x).$$

b) (CS) S'il existe $\lambda \in \mathbb{R}^q$ tel que $\nabla f(x) + Dh(x)^T\lambda = 0$ et si

$$\nabla_x^2 L(x, \lambda)d.d > 0 \quad \forall d \in T(S, x),\ d \neq 0,$$

alors x est un minimum local strict de f sur S.

Démonstration

On reprend les notations de la démonstration de la proposition 3.1, et on se ramène à un problème sans contraintes. Le vecteur λ étant fixé par les conditions du théorème, on pose pour $u \in U$:

$$F(u) = f(\psi(u)) = L(\psi(u), \lambda).$$

On a

$$\begin{aligned} \nabla F(u).z &= \nabla f(\psi(u)).D\psi(u)z \\ &= \nabla_x L(\psi(u), \lambda).D\psi(u)z \\ \nabla^2 F(u)z.z &= \nabla_x^2 L(\psi(u), \lambda)D\psi(u)z.D\psi(u)z + \nabla_x L(\psi(u), \lambda).D^2\psi(u)(z, z). \end{aligned}$$

Au point u_0, on a $\psi(u_0) = x$. On pose $d = D\psi(u_0)z$. On a alors

$$\begin{aligned} \nabla F(u_0).z &= \nabla_x L(x, \lambda).d \\ \nabla^2 F(u_0)z.z &= \nabla_x^2 L(x, \lambda)d.d + \nabla_x L(x, \lambda).D^2\psi(u_0)(z, z). \end{aligned}$$

De plus, on a $\nabla_x L(x, \lambda) = \nabla f(x) + Dh(x)^T\lambda = 0$, donc

$$\nabla^2 F(u_0)z.z = \nabla_x^2 L(x, \lambda)d.d. \tag{3.1}$$

a) Le point x est un minimum local de f sur S donc u_0 est un minimum local de F sur U. D'après la proposition 2.1, on a $\nabla^2 F(u_0)z.z \geq 0 \quad \forall z \in \mathbb{R}^{N-q}$. L'application linéaire $D\psi(u)$ est un isomorphisme de $\mathbb{R}^{N-q}$ sur $T(S, x) = \text{Ker}\, Dh(x)$, espace tangent à S au point x. D'où avec (3.1)

$$\nabla_x^2 L(x, \lambda)d.d \geq 0 \quad \forall d \in T(S, x).$$

b) Avec les mêmes arguments, on a $\nabla^2 F(u_0)z.z > 0$ pour tout $z \in \mathbb{R}^{N-q},\ z \neq 0$. D'après la proposition 2.2, u_0 est un minimum local strict de F sur U, donc x est minimum local strict de f sur S.

■

3.3 Contraintes d'égalités et d'inégalités

On considère ici le cas général

$$S = \{x \in \mathbb{R}^N;\ g(x) \leq 0\ ;\ h(x) = 0\}$$

$$g\ :\ \mathbb{R}^N \longmapsto \mathbb{R}^p$$
$$h\ :\ \mathbb{R}^N \longmapsto \mathbb{R}^q$$

On appelle <u>contrainte active</u> en un point $x \in S$ une contraine g_i telle que $g_i(x) = 0$. Les autres sont dites inactives. L'ensemble des indices i pour lesquels $g_i(x) = 0$ est noté $I(x)$. Par convention, les contraintes d'égalité h_i sont considérées comme étant actives.

3.3.1 Conditions d'optimalité du premier ordre

Théorème 3.4 (Karush, Kuhn, Tucker, 1951) *Soit $x \in S$ un minimum local de f sur S. On suppose que les fonctions f, g et h sont de classe C^1 sur un voisinage de x, et que les vecteurs $\nabla h_i(x)$, $i = 1, \ldots, q$ et $\nabla g_i(x)$, $i \in I(x)$ forment une famille libre.*
Alors il existe des nombres $\mu_i \geq 0$, $i \in I(x)$ et $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, q$ tels que

$$\nabla f(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) + \sum_{j=1}^{q} \lambda_j \nabla h_j(x) = 0.$$

<u>Formulation équivalente</u>
Il existe des nombres μ_i, $i = 1, \ldots, p$ et $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, q$ tels que

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} \mu_i \nabla g_i(x) + \sum_{j=1}^{q} \lambda_j \nabla h_j(x) = 0 \\ \mu_i \geq 0,\ i = 1, \ldots, p \\ \mu_i\ g_i(x) = 0 \quad i = 1, \ldots, p \end{cases}$$

ou encore tels que

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} \mu_i \nabla g_i(x) + \sum_{j=1}^{q} \lambda_j \nabla h_j(x) = 0 \\ \mu_i \geq 0,\ i = 1, \ldots, p \\ \mu . g(x) = 0. \end{cases}$$

Ces conditions sont appelées "conditions de KKT", et les nombres μ_i et λ_j sont encore appelés multiplicateurs de Lagrange.

<u>Démonstration</u>
Le point x est un minimum local de f sur S, donc aussi sur le sous-ensemble de S défini par

$$S' = \{x \in S;\ h_i(x) = 0,\ i = 1, \ldots q,\ g_i(x) = 0,\ \forall i \in I(x)\}.$$

D'après le théorème 3.2, il existe des réels μ_i et λ_i tels que

$$\nabla f(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) + \sum_{j=1}^{q} \lambda_j \nabla h_j(x) = 0.$$

Montrons par l'absurde que les μ_i sont positifs ou nuls. Supposons que l'un des μ_i soit strictement négatif, par exemple $\mu_k < 0,\ k \in I(x)$. Prenons

$$d \in \operatorname{Vect}\{\nabla h_i(x),\ i = 1, \ldots, q,\ \nabla g_i(x),\ i \in I(x)\}$$

tel que

$$\begin{array}{rcl} \nabla h_i(x).d & = & 0, i = 1, \ldots, q \\ \nabla g_i(x).d & = & 0, \quad i \neq k,\ i \in I(x) \\ \nabla g_k(x).d & < & 0. \end{array}$$

Un tel choix est possible car les vecteurs $\nabla h_i(x),\ i = 1, \ldots, q$ et $\nabla g_i(x),\ i \in I(x)$ forment une famille libre. Admettons provisoirement l'existence d'un arc

$$\begin{array}{rrcl} s\ :\ & [0, a[& \longrightarrow & S \\ & t & \longmapsto & s(t) \end{array}$$

tel que $s(0) = x$ et $s'(0) = d$. Par le choix de d, il vient que

$$\begin{array}{rcl} \dfrac{d}{dt}(f(s(t)))_{|t=0} & = & \nabla f(x).d \\ & = & -\displaystyle\sum_{i \in I(x)} \mu_i \nabla g_i(x).d - \sum_{j=1}^{q} \lambda_j \nabla h_j(x).d \\ & = & -\mu_k \nabla g_k(x).d < 0, \end{array}$$

d'où $f(s(t)) < f(s(0)) = f(x)$ pour $|t|$ petit, $t > 0$. Ceci contredit le fait que x soit un minimum local.

Il reste à montrer l'existence de l'arc s. Soit

$$S'' \ = \ \{x \in \mathbb{R}^N;\ h_i(x) = 0,\ i = 1, \ldots q,\ g_i(x) = 0,\ \forall i \in I(x) \backslash \{k\}\}.$$

Par construction, on a

$$d \in \operatorname{Ker} \begin{pmatrix} Dh(x) \\ Dg_i(x)_{i \in I(x) \backslash \{k\}} \end{pmatrix}.$$

D'après le proposition 3.1, on a $d \in T(S'', x)$, et il existe un arc

$$\begin{array}{rrcl} s\ :\ & [0, a'[& \longrightarrow & S'' \\ & t & \longmapsto & s(t), \end{array}$$

tel que $s(0) = x$ et $s'(0) = d$. Il ne reste plus qu'à montrer que $s(t) \in S$ pour $t \geq 0$ proche de 0, i.e. $g_i(x) \leq 0$ pour $i = k$ ou $i \notin I(x)$. On a

$$\frac{d}{dt}(g_k(s(t)))_{|t=0} = \nabla g_k(x).d < 0$$
$$g_k(s(0)) = g_k(x) = 0,$$

donc il existe $a'' \in]0, a']$ tel que $g_k(s(t)) \leq 0$ pour $t \in [0, a''[$. Pour les contraintes inactives $i \notin I(x)$, on a $g_i(s(0)) < 0$, et par continuité des g_i et de s, il existe $a \in]0, a'']$ tel que $g_i(s(t)) \leq 0$ pour $t \in [0, a[$. Donc $s(t) \in S$ pour $t \in [0, a[$. ■

Proposition 3.5 *L'hypothèse : les vecteurs $\nabla h_i(x)$, $i = 1, \ldots, q$ et $\nabla g_i(x)$, $i \in I(x)$ forment une famille libre, peut être remplacée par l'une des conditions suivantes :*

i) les fonctions g_i et h_i sont affines,

ii) les fonctions g_i sont convexes, les fonctions h_i sont affines, et il existe $x_0 \in S$ tel que $g_i(x_0) < 0$ pour tous les g_i non affines.

Ces hypothèses sont appelées hypothèses de qualification des contraintes, et sont notées QC(x).

Remarque : il y a une différence fondamentale entre les conditions KKT d'une part, et les conditions d'optimalité sans contraintes ou de Lagrange d'autre part : on a ici une information concernant la direction du gradient, provenant du signe des multiplicateurs. Ceci a pour conséquence que dans le cas où S est d'intérieur non vide (pas de contraintes d'égalité), le risque d'avoir un maximum local satisfaisant les conditions d'optimalité est plus "faible" que dans le cas sans contraintes ou contraintes d'égalité uniquement.

Formulation lagrangienne

$$\begin{array}{rcll} L & : & \mathbb{R}^N \times \mathbb{R}^p_+ \times \mathbb{R}^q & \longrightarrow & \mathbb{R} \\ & & (x, \mu, \lambda) & \longmapsto & L(x, \mu, \lambda) = f(x) + g(x).\mu + h(x).\lambda \end{array}$$

$$\nabla L(x, \mu, \lambda) = \begin{pmatrix} \nabla_x L(x, \mu, \lambda) & = & \nabla f(x) + Dg(x)^T \mu + Dh(x)^T \lambda \\ \nabla_\mu L(x, \mu, \lambda) & = & g(x) \\ \nabla_\lambda L(x, \mu, \lambda) & = & h(x) \end{pmatrix}$$

$$(x \in S \text{ et } KKT) \Leftrightarrow \left\{ \begin{array}{lcl} \nabla_\lambda L(x, \mu, \lambda) & = & 0 \\ \nabla_\mu L(x, \mu, \lambda) & \leq & 0 \\ \nabla_x L(x, \mu, \lambda) & = & 0 \\ \mu & \geq & 0 \\ g(x).\mu & = & 0 \end{array} \right.$$

Mis à part les contraintes de positivité sur μ, on a $N+p+q+1$ équations, avec des inégalités qui sont simples à prendre en compte. On peut donc ici aussi employer des méthodes de type Newton.

CAS DES FONCTIONS CONVEXES

On suppose ici que les fonctions h_i sont affines et que les fonctions g_i sont convexes.

Théorème 3.6 (Condition suffisante du premier ordre) *On suppose la fonction f convexe. Soit $x \in S$ tel que $\nabla f(x)$ existe. S'il existe $(\mu, \lambda) \in \mathbb{R}_+^p \times \mathbb{R}^q$ tel que*

$$\left\{ \begin{array}{lcl} \nabla_x L(x, \mu, \lambda) & = & 0 \\ g(x).\mu & = & 0 \end{array} \right.$$

alors x est un minimum global de f sur S.

Démonstration

Soit $y \in S$. En utilisant la convexité de g_i : $g_i(y) - g_i(x) \geq \nabla g_i(x).(y-x)$, l'égalité $\nabla h_j(x).(x-y) = 0$, et la convexité de f, on obtient

$$\begin{array}{rcl} f(x) & \leq & f(x) + \mu.(g(x) - g(y)) = f(x) + \sum_{i=1}^p \mu_i (g_i(x) - g_i(y)) \\ & \leq & f(x) + \sum_{i=1}^p \mu_i \nabla g_i(x).(x-y) \\ & = & f(x) + \sum_{i=1}^p \mu_i \nabla g_i(x).(x-y) + \sum_{j=1}^q \lambda_j \nabla h_j(x).(x-y) \\ & = & f(x) + \nabla f(x).(y-x) \\ & \leq & f(y). \end{array}$$

■

3.3.2 Conditions d'optimalité du deuxième ordre

Théorème 3.7 *Soit $x \in S$. On suppose que les fonctions f, g et h sont deux fois différentiables au point x.*

i) (CN) on suppose que la famille $\{\nabla h_i(x),\ i = 1, \dots, q, \nabla g_i(x),\ i \in I(x)\}$ est libre et que x est un minimum local de f sur S. Il existe donc $(\mu, \lambda) \in \mathbb{R}_+^p \times \mathbb{R}^q$ tel que :

$$\begin{cases} \nabla_x L(x, \mu, \lambda) & = & 0 \\ g(x).\mu & = & 0. \end{cases}$$

Alors

$$\nabla_x^2 L(x, \mu, \lambda) d.d \geq 0$$

pour tout d dans l'espace tangent aux contraintes actives

$$\{d \in \mathbb{R}^{\mathrm{N}};\ \nabla h_i(x).d = 0,\ i = 1, \dots, q,\ \nabla g_i(x).d = 0,\ i \in I(x)\}.$$

ii) (CS) S'il existe $(\mu, \lambda) \in \mathbb{R}_+^p \times \mathbb{R}^q$ tel que

a) $\nabla_x L(x, \mu, \lambda) = 0$

b) $g(x).\mu = 0$

c) $\nabla_x^2 L(x, \mu, \lambda) d.d > 0$ *pour tout $d \neq 0$ dans*

$$\{d \in \mathbb{R}^{\mathrm{N}};\ \nabla h_i(x).d = 0,\ i = 1 \text{ à } q,\ \nabla g_i(x).d = 0 \text{ si } i \in I(x) \text{ et } \mu_i > 0\}$$

alors x est un minimum local strict de f sur S.

Chapitre 4

Optimisation avec contraintes, méthodes directes

4.1 Méthode de relaxation

C'est la même méthode que celle décrite dans le cas sans contraintes, mais en se restreignant à S. On suppose ici que S est d'intérieur non vide. L'algorithme est le suivant :

$$\begin{cases} x_0 \in \text{S donné} \\ x_{k+1} = x_k + t_k e_{[k]}, \quad ([k] = k \ mod \ N) \\ t_k \text{ minimise } f \text{ sur le segment } \{x_k + \mathbb{R} d_k\} \cap S. \end{cases}$$

Cette méthode est en général déconseillée, sauf dans le cas suivant :

Proposition 4.1 *On suppose que S est de la forme*

$$S = \prod_{i=1}^{N} [a_i, \, b_i] \quad a_i, b_i \in \overline{\mathbb{R}}.$$

Si f est de classe C^2 et s'il existe $\alpha > 0$ tel que

$$\nabla^2 f(x) d.d \geq \alpha ||d||^2 \quad \forall x, \ d \in \mathbb{R}^N,$$

alors la méthode de relaxation converge.

Remarque : la fonction f est ici strictement convexe.

4.2 Méthode de projection

4.2.1 Rappels

Théorème 4.2 (Théorème de projection) *Soit C un convexe fermé non vide de $\mathbb{R}^N$. Pour tout point $x \in \mathbb{R}^N$, il existe un unique point $P_C(x) \in C$ minimisant la distance de x à C :*

$$||x - P_C(x)|| \leq ||x - y||, \quad \forall y \in C.$$

Ce point est caractérisé par

$$(y - P_C(x)).(x - P_C(x)) \leq 0, \quad \forall y \in C.$$

La projection

$$\begin{array}{lcl} \mathbb{R}^N & \longrightarrow & C \\ x & \longmapsto & P_C(x) \end{array}$$

est continue et même 1-Lipschitzienne :

$$||P_C(x) - P_C(x')|| \leq ||x - x'||, \quad \forall x,\ x' \in \mathbb{R}^N,$$

i.e. la projection réduit les distances.

4.2.2 Projection sur l'ensemble des contraintes

On suppose dans ce sous-paragraphe que S est un ensemble convexe fermé d'intérieur non vide.
Principe : soit $x_k \in S$ le point courant. On suppose que ce point ne satisfait pas les conditions KKT. Soit $x' = x_k - t_k \nabla f(x_k)$ le point obtenu par la méthode de plus forte pente. Si ce point n'est pas dans S, on prend comme nouveau point la projection de x' sur S. Ceci peut se justifier de la manière suivante : soit $s\ :\ [0, a[\rightarrow S$ un arc paramétré tel que $s(0) = x_k$ et $s'(0) = P_{T(S,x_k)}(-\nabla f(x_k))$. Alors on a généralement $(f(s(t))'(0) = \nabla f(x_k).P_{T(S,x_k)}(-\nabla f(x_k)) < 0$. Pour t_k petit, $P_S(x')$ est de la forme $s(t)$ avec t petit, et le critère décroît. L'algorithme est le suivant :

$$\begin{cases} x_0 \text{ donné} \\ x_{k+1} = P_S(x_k - t_k \nabla f(x_k)), \quad t_k > 0. \end{cases}$$

Remarque : on pourrait aussi considérer $x_{k+1} = x_k + t_k P_{T(S,x_k)}(-\nabla f(x_k))$. Ce point n'est cependant pas forcément dans S, et cette méthode peut être plus lente (exemple d'un polygone). Par contre, si on remplace $-\nabla f(x_k)$ par une direction de descente d, alors $d' = P_S(x_k + d) - x_k$ n'est pas forcément une direction de descente, et il est alors préférable de travailler directement dans l'espace tangent aux contraintes actives.

Proposition 4.3 *On suppose que f est une fonction convexe de classe C^2, et qu'il existe $\alpha > 0$ et $M > 0$ tels que*

$$\nabla^2 f(x)d.d \geq \alpha||d||^2, \quad \forall\, x,\, d \in \mathbb{R}^N$$
$$||\nabla f(x) - \nabla f(y)|| \leq M||x-y||, \quad \forall x,\, y \in \mathbb{R}^N.$$

On note $\bar{x}$ le minimum de f sur S. Soit $0 < a \leq b < 2\alpha/M^2$ et $(t_k)_{k\geq 0}$ une suite telle que $a \leq t_k \leq b$ pour tout $k \geq 0$. Alors l'algorithme ci-dessus converge géométriquement, i.e il existe $\beta < 1$ tel que

$$||x_k - \bar{x}|| \leq \beta^k ||x_0 - \bar{x}||, \quad \forall k \geq 0.$$

Démonstration
On suppose pour simplifier que $t_k = t \in [a,\, b]$ pour tout $k \geq 0$. L'ensemble S est un espace métrique complet, on utilise le théorème du point fixe. Il suffit de montrer que

$$g(x) := P_S(x - t\nabla f(x))$$

est une contraction de rapport $\beta < 1$, et que son point fixe est $\bar{x}$.

• En utilisant les hypothèses et le fait que la projection est 1-Lipschitzienne, on a

$$\begin{aligned} ||g(x) - g(y)||^2 &\leq ||x - y - t(\nabla f(x) - \nabla f(y))||^2 \\ &= ||x-y||^2 - 2t(x-y).(\nabla f(x) - \nabla f(y)) + t^2||\nabla f(y) - \nabla f(x)||^2 \\ &\leq ||x-y||^2 - 2\alpha t||x-y||^2 + M^2t^2||x-y||^2 \\ &= (1 - 2\alpha t + M^2t^2)||x-y||^2. \end{aligned}$$

Soit $P(t) = 1 - 2\alpha t + M^2t^2$. Pour $0 < a \leq t \leq b < 2\alpha/M^2$, on a $0 \leq P(t) \leq \max(P(a), P(b)) < 1$, et le résultat est obtenu en posant $\beta = \sqrt{P(t)}$.

• Soit x' le point fixe de g. Si on pose $z = x' - t\nabla f(x')$, on a $x' = g(x') = P_S(z)$. En utilisant la convexité de f et la caractérisation de la projection, on obtient

$$\begin{aligned} f(y) - f(x') &\geq \nabla f(x').(y - x') \\ &= \frac{1}{t}(x' - z).(y - x') \\ &= -\frac{1}{t}(z - P_S(z)).(y - P_S(z)) \\ &\geq 0, \quad \forall y \in S, \end{aligned}$$

ce qui prouve que x' est un minimum global. Comme f est strictement convexe, ce minimum est unique et $x' = \bar{x}$. ■

Remarque : pour $S=\mathbb{R}^N$, on obtient un résultat de convergence de la méthode du gradient à pas variable.
La projection est en général difficile à calculer, sauf dans quelques cas comme :

- ⋆ $S=\mathbb{R}^+$ (cf. les conditions de KKT) :
$$P_S(x)=\begin{pmatrix} x_1^+ \\ \vdots \\ x_N^+ \end{pmatrix} \text{ avec } x_i^+=\begin{cases} x_i \text{ si } x\geq 0 \\ 0 \text{ sinon} \end{cases}$$

- ⋆ $S=\displaystyle\prod_{i=1}^N [a_i,b_i]$:
$$P_S(x)=y \text{ avec } y_i=\begin{cases} a_i \text{ si } x_i\leq a_i \\ x_i \text{ si } a_i\leq x_i\leq b_i \\ b_i \text{ si } x_i\geq b_i \end{cases}$$

- ⋆ $S=\{x\in\mathbb{R}^N;\ Ax\leq b\}$. C'est le cas d'un polyèdre, que nous étudions dans le paragraphe suivant, où l'on combine la méthode de projection avec la méthode des surfaces actives.

4.3 Méthode des surfaces actives

4.3.1 Cas général

On considère à nouveau le problème

$$(\mathcal{P})\begin{cases} \min f(x) \\ g(x)\leq 0, \quad g\ :\ \mathbb{R}^N\longrightarrow\mathbb{R}^p. \end{cases}$$

Soit $x_0\in S$ et S_0 la *surface active* définie par

$$S_0=\{x\in\mathbb{R}^N;\ g_i(x)=0,\ \forall i\in I(x_0)\}.$$

On effectue une méthode de descente sur S_0 tout en surveillant les contraintes non actives en x_0. Deux cas peuvent se présenter :

cas 1 : on arrive à un point x_1 où une nouvelle contrainte g_j devient active (le cas de plusieurs contraintes devenant actives est similaire). On prend alors

$$S_1=\{x\in\mathbb{R}^N;\ g_i(x)=0,\ \forall i\in I(x_1)\},$$

et on continue le procédé sur S_1 qui est une surface incluse dans S_0 (une contrainte de plus).

cas 2 : les autres contraintes restent inactives, et on converge vers un point stationnaire x_1 (stationnaire par rapport à S_0), tel que $I(x_1) = I(x_0)$. D'après les conditions de Lagrange ,si x_1 est régulier, il existe des multiplicateurs λ_i réels tels que

$$\nabla f(x_1) + \sum_{i \in I(x_0)} \lambda_i \nabla g_i(x_1) = 0,$$

i.e. le gradient est orthogonal à la surface active. Il y a alors deux possibilités :

a) si $\lambda \geq 0$, les conditions KKT sont satisfaites, on a trouvé un point stationnaire (par rapport à S).

b) il existe un indice j tel que $\lambda_j < 0$. En supposant le point x_1 régulier, on peut trouver une direction de descente d, en prenant par exemple d orthogonal aux $\nabla g_i(x_1)$, $i \in I(x_0)$, $i \neq j$ et tel que $\nabla g_j(x_1).d < 0$. En prenant un arc $s\ : [0, a[\rightarrow S$ tel que $s(0) = x_1$ et $s'(0) = d$, on a

$$(f(s(t))'(0) = \nabla f(x_1).d = -\lambda_j \nabla g_j(x_1).d < 0. \tag{4.1}$$

Ceci signifie qu'en abandonnant la contrainte g_j, on peut diminuer le critère strictement. On itère alors avec

$$S_1 = \{x \in \mathbb{R}^N;\ g_i(x) = 0,\ \forall i \in I(x_0),\ i \neq j\},$$

qui est une surface contenant S_0 (une contrainte de moins).

On considère donc la stratégie suivante :

$$\left\{\begin{array}{l}\text{choix de } x_0 \in S \\ \text{à l'étape } k \geq 0 : x_{k+1} \text{ minimise } f \text{ sur } S_k \cap S \text{ (via éventuellement le cas 1)} \\ \text{cas 2a) : stop} \\ \text{cas 2b) : abandon d'une contrainte pour laquelle } \lambda_j < 0, \\ \qquad\qquad S_{k+1} = \{x \in \mathbb{R}^N;\ g_i(x) = 0,\ \forall i \in I(x_k),\ i \neq j\}.\end{array}\right.$$

Proposition 4.4 *On suppose que l'ensemble des contraintes S est régulier, et que pour tout $I \subset \{1, 2, \ldots, p\}$, le problème*

$$\inf f(x)$$
$$\left\{\begin{array}{ll} g_i(x) = 0 & \forall i \in I \\ x \in S \end{array}\right.$$

a des solutions. Alors la méthode décrite ci-dessus converge vers un point stationnaire satisfaisant les conditions de KKT.

Démonstration
Pour $x \in S$ et $I \subset \{1, 2, \ldots, p\}$, on pose $S_I(x) = \{y \in S;\ g_i(y) = 0,\ \forall i \in I(x)\}$.

Par construction, le point x_{k+1} minimise aussi f sur sur $S_I(x_{k+1})$, car $S_I(x_{k+1}) \subset S_k \cap S$. Dans le cas 2b) on a nécessairement $f(x_{k+2}) < f(x_{k+1})$ (cf. 4.1). La suite $f(x_k)$, $k \geq 1$ est strictement décroissante, et on ne peut donc pas revenir sur un S_I déjà exploré. Comme l'ensemble des $I \subset \{1, 2, \dots, p\}$ est fini, la suite des x_k est finie. Le dernier des points obtenus correspond au cas 2a), et satisfait les conditions KKT. ■

Remarque :
Cela ne définit pas vraiment un algorithme : chacun des x_i est un minimum qui demande en général un nombre infini d'itérations. En pratique, on abandonne une contrainte lorsque la distance $||x_{k+1} - x_k||$ devient petite, mais des oscillations peuvent se produire.

4.3.2 Cas des contraintes linéaires

Lorsque les contraintes d'inégalité sont affines, la méthode précédente est assez facile à mettre en oeuvre. Nous décrivons ici une méthode de plus forte pente par rapport à la surface active. Elle peut être modifiée en une méthode de type gradient conjugué. Son principe repose sur la projection du gradient sur la surface active, et met à profit les techniques qui, connaissant l'inverse d'une matrice de dimension n, permettent de calculer rapidement l'inverse de la matrice obtenue en rajoutant une ligne et une colonne (ajoût d'une contrainte) ou en supprimant une ligne et une colonne (abandon d'une contrainte).

On considère le problème

$$\min f(x)$$
$$\begin{cases} a_i.x & \leq & b_i, & \forall i \in I \\ a_j.x & = & b_j, & \forall j \in J. \end{cases}$$

Pour $x \in S$, on a $I(x) = \{i \in I \cup J;\ a_i.x = b_i\}$. La surface active peut s'écrire sous la forme

$$\begin{aligned} S(x) &= \{y \in \mathbb{R}^N;\ a_i.y = b_i,\ \forall i \in I(x)\} \\ &= \{y \in \mathbb{R}^N;\ Ay = b\} \end{aligned}$$

en notant A (respectivement b) la matrice (respectivement le vecteur) qui a pour lignes les a_i^T, $i \in I(x)$ (resp. b_i, $i \in I(x)$). Si A est de rang maximal, le projecteur sur Ker$A$ est

$$P = I - A^T(AA^T)^{-1}A.$$

La projection est ici affine ($S(x)$ est un espace affine),

$$P_{S(x)}(x - \nabla f(x)) = x + \underbrace{P_{Ker(A)}}_{P}(-\nabla f(x)),$$

on prend donc comme direction de recherche

$$d = P(-\nabla f(x)).$$

1^0 cas $d \neq 0$. On reste dans la surface active $S(x)$: on détermine $\alpha > 0$ tel que

$$x + td \in S, \quad \forall t \in [0, \alpha],$$

et on minimise f sur le segment $\{x + td\ ;\ t \in [0, \alpha]\}$, ce qui donne le nouveau point $x' = x + td$.
• Si $t < \alpha$ alors $I(x_1) = I(x)$, on itère sans avoir à recalculer le projecteur P.
• Si $t = \alpha$, une nouvelle contrainte devient active. Dans ce cas, la matrice des contraintes A' au point x' a une ligne de plus que A, et le projecteur P' sur la nouvelle surface active $S(x')$ peut se calculer rapidement à partir de P (une ligne et une colonne de plus). La situation se complique un peu si plusieurs contraintes deviennent actives, mais ce cas a peu de chance de se produire.

2^0 cas $d = 0$. On a alors

$$\nabla f(x) - A^T(AA^T)^{-1}A\nabla f(x) = 0,$$

et en posant $-\lambda = (AA^T)^{-1}A\nabla f(x)$, ceci s'écrit

$$\nabla f(x) + A^T\lambda = 0.$$

• Si tous les λ_j pour $j \in J$ sont positifs ou nuls, les conditions de KKT sont satisfaites : arrêt de l'algorithme.
• Sinon, parmi les indices $i \in J$ tels que $\lambda_i < 0$, on peut choisir un indice j pour lequel $\lambda_j||a_j||$ est le plus négatif, et on abandonne la contrainte $a_j.x \leq b_j$. Ce n'est pas forcément le choix optimal, qui serait de prendre un indice j pour lequel $-\lambda_j a_j.d_j$ est le plus négatif, les vecteurs $(d_i)_{i\in I(x)}$ étant définis de la manière suivante :

$$Vect\{d_i,\ i \in I(x)\} = Vect\{a_i,\ i \in I(x)\}, \quad ||d_i|| = 1,$$
$$d_i.a_j = 0, \ \text{ si } i \neq j, \quad d_i.a_j < 0, \ \text{ si } i = j,$$

avec

$$\nabla f(x).d_j = -A^T\lambda.d_j = -\lambda.Ad_j = -\lambda_j a_j.d_j.$$

Algorithme :

1. choix de $x \in S$,
2. déterminer $I(x)$ et A,

3. calculer $P = I - A^T(AA^T)^{-1}A$ et $d = P(-\nabla f(x))$,

4. si $||d|| > 0$, déterminer t, puis $x \leftarrow x + td$ et retour à 2.

5. sinon, calculer λ,

   - si $\lambda_i \geq 0$ pour tout $i \in I(x) \cap J$: stop,
   - sinon, choisir $j \in I(x) \cap J$ tel que $\lambda_j < 0$, $A \leftarrow A$ privé de la ligne j, $I(x) \leftarrow I(x) \backslash \{j\}$ et retour à 3.

Remarque : il n'y a pas de certitude de convergence.

4.4 Méthodes de pénalisation

L'idée consiste à remplacer un problème avec contraintes par une suite de problèmes sans contraintes, dont on espère que les solutions convergent vers celles du problème initial.

4.4.1 Pénalisation extérieure

Soit le problème

$$(\mathcal{P}) \quad \inf_{x \in S} f(x)$$

avec S fermé dans $\mathbb{R}^N$ et f continue.
On considère une fonction continue

$$\psi \ : \ \mathbb{R}^N \longrightarrow \mathbb{R}_+$$

telle que

$$\psi(x) = 0 \Leftrightarrow x \in S$$

et une suite strictement croissante de nombres positifs c_k telle que

$$\lim_{k \to \infty} c_k = +\infty.$$

Le problème pénalisé est alors :

$$(\mathcal{P}_k) \quad \inf_{x \in \mathbb{R}^N} f(x) + c_k \psi(x).$$

Algorithme :

$$\begin{cases} \text{choix de } x_0, \\ \text{pour } k \geq 1 \text{ résoudre } (\mathcal{P}_k) \text{ par une méthode} \\ \text{itérative initialisée avec } x_{k-1}, \\ x_k = \text{solution de } (\mathcal{P}_k). \end{cases}$$

⚠ Pour des raisons de stabilité numérique, il est déconseillé de commencer avec c_1 trop grand.

Proposition 4.5 *On suppose que* $(\mathcal{P})$ *admet une solution et que* $(\mathcal{P}_k)$ *admet une solution* x_k *pour tout* $k \geq 1$. *Alors toute valeur d'adhérence de la suite* (x_k) *est solution de* $(\mathcal{P})$.

Démonstration
Soit $x \in S$ solution du problème $(\mathcal{P})$. On a

$$f(x_k) \leq f(x_k) + c_k\psi(x_k) \leq f(x) + c_k\psi(x) = f(x), \tag{4.2}$$

car $c_k > 0$, $\psi(x_k) > 0$, x_k minimise $f(x) + c_k\psi(x)$ et $\psi(x) = 0$. On a également

$$f(x_{k+1}) + c_{k+1}\psi(x_{k+1}) \geq f(x_{k+1}) + c_k\psi(x_{k+1}) \geq f(x_k) + c_k\psi(x_k), \tag{4.3}$$

car $c_{k+1} > c_k$ et x_k minimise $f(x) + c_k\psi(x)$.

La suite $f(x_k) + c_k\psi(x_k)$ est croissante et majorée par $f(x)$, donc converge vers une limite $l \leq f(x)$. Soit x^* une valeur d'adhérence de la suite (x_k). Il existe une sous-suite extraite de (x_k) telle que

$$\lim_{n\to\infty} x_{k_n} = x^*.$$

On a $\lim_{n\to\infty} f(x_{k_n}) = f(x^*)$ car f est continue. D'où

$$\lim_{n\to\infty} (f(x_{k_n}) + c_{k_n}\psi(x_{k_n}) - f(x_{k_n})) = l - f(x^*),$$

c'est-à-dire la sous-suite $c_{k_n}\psi(x_{k_n})$ converge vers $l - f(x^*)$. Comme $c_{k_n} \to \infty$ et ψ est continue, on a

$$\psi(x_*) = \lim_{n\to\infty} \psi(x_{k_n}) = 0.$$

Donc $x^* \in S$. D'après (4.2), on a

$$f(x^*) = \lim_{n\to\infty} f(x_{k_n}) \leq l \leq f(x),$$

donc x^* est solution de $\mathcal{P}$. ■

Remarque : si on suppose de plus $\lim_{||x||\to\infty} f(x) = +\infty$, ou bien que S est borné et que $\lim_{||x||\to\infty} \psi(x) = +\infty$, alors d'une part, $(\mathcal{P})$ a des solutions, et d'autre part on peut montrer que la suite (x_k) a effectivement des valeurs d'adhérence.

Exemples de pénalisation :
On suppose que

$$S = \{x \in \mathbb{R}^N;\ g_i(x) \leq 0,\ i = 1, \ldots, p\},$$

où les fonctions g_i sont de classe $\mathcal{C}^1$

Pénalisation exacte

$$\boxed{\psi(x) = \sum_{i=1}^{p} g_i^+(x),}$$

avec

$$u^+(x) = \begin{cases} u(x) \text{ si } u(x) \geq 0 \\ 0 \text{ sinon.} \end{cases}$$

On a

$$\begin{aligned} \psi(x) = 0 &\Leftrightarrow \max_i g_i^+(x) = 0 \\ &\Leftrightarrow g_i(x) \leq 0 \quad \forall i \\ &\Leftrightarrow x \in S. \end{aligned}$$

La pénalisation est dite exacte car dans la plupart des cas, les solutions de $(\mathcal{P}_k)$ sont solutions de $(\mathcal{P})$ dès que k est assez grand.

⚠ Inconvénients : ψ n'est pas partout dérivable.

Pénalisation quadratique

$$\boxed{\psi(x) = \sum_{i=1}^{p} g_i^+(x)^2.}$$

Cette méthode n'est plus exacte mais par contre elle est différentiable.
En effet, soit
$u \in \mathcal{C}^1$ et $\vartheta(x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ x & \text{si} \quad x \geq 0 \end{cases}$
Notons

$$h(x) = u^+(x)^2 = \vartheta(u(x))^2.$$

Pour $u(x) \neq 0$ on a

$$\begin{aligned} h'(x) &= 2(\vartheta(u(x)))'u'(x) \\ &= 2u^+(x)u'(x), \end{aligned}$$

qui se prolonge par continuité aux points où $u(x) = 0$.

Nous allons en déduire la généralisation suivante.

Généralisation de KKT (Fritz John)
On reprend les hypothèses de la proposition 4.5. Quitte à renuméroter, on suppose que x_k est solution de $(\mathcal{P}_k)$, et converge vers x solution de $(\mathcal{P})$.

L'optimalité de x_k se traduit par

$$\nabla f(x_k) + 2c_k \sum_{i=1}^{p} g_i^+(x_k)\nabla g_i(x_k) = 0. \tag{4.4}$$

On pose

$$\Delta_k = \left(1 + 4c_k^2 \sum_{i=1}^{p} (g_i^+(x_k))^2\right)^{1/2}$$

et

$$\lambda_0^k = \frac{1}{\Delta_k}, \quad \lambda_i^k = \frac{2c_k g_i^+(x_k)}{\Delta_k}.$$

On a donc

$$\lambda_0^k \nabla(f(x_k)) + \sum_{i=1}^{p} \lambda_i^k \nabla g_i(x_k) = 0. \tag{4.5}$$

La suite $\Lambda^k = (\lambda_0^k, \ldots, \lambda_p^k)$ vérifie $||\Lambda^k|| = 1$, elle admet donc une sous-suite convergente vers un $\Lambda = \lambda_0, \ldots, \lambda_p$ avec $||\Lambda|| = 1$. En passant à la limite dans (4.5) on obtient

$$\lambda_0 \nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x) = 0,$$

où les λ_i sont non tous nuls.

Remarque : Il se peut que $\lambda_0 = 0$!

On peut en déduire le résultat suivant, où il n'y a plus de conditions de qualification de contraintes comme "les ∇g_i sont linéairement indépendants", et qui inclut le cas des contraintes d'égalité.

Proposition 4.6 *Soit x un minimum local de f sur*

$$S = \{x \in \mathbb{R}^N, \ g_i(x) \leq 0 \quad i = 1, \ldots, p\},$$

où les fonctions f et g_i sont de classe $\mathcal{C}^1$. Alors il existe des multiplicateurs $\lambda_0 \geq 0, \ \lambda_1 \geq 0, \ \ldots, \ \lambda_p \geq 0$ non tous nuls tels que

$$\lambda_0 \nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x) = 0$$
$$\lambda_i g_i(x) = 0 \quad i = 1, \ldots, p.$$

4.4.2 Pénalisation intérieure

Au lieu de pénaliser f quand x est à l'extérieur de S, on travaille à l'intérieur de S et on empêche x de s'approcher du bord de S (fonction barrière).

On considère ici le problème

$$(\mathcal{P}) \quad \min_{x \in S} f(x)$$

où l'on suppose que S est fermé, avec

$$\overset{\circ}{S} \neq \emptyset, \quad \overline{\overset{\circ}{S}} = S.$$

Soit B une fonction telle que

$$\left\{ \begin{array}{l} B(x) \geq 0 \quad \forall x \in \overset{\circ}{S} \\ \lim\limits_{x \to \partial S} B(x) = +\infty \\ B \text{ continue sur } \overset{\circ}{S}\,. \end{array} \right. \tag{4.6}$$

Par exemple si $S = \{x;\ g_i(x) \leq 0 \quad i = 1, \ldots, p\}$, les fonctions g_i étant continues, on peut prendre $B(x) = -\sum\limits_{i=1}^{p} 1/g_i(x)$. On peut également utiliser la fonction logarithme.

Le problème pénalisé est le suivant : pour un $c_k > 0$, que l'on fera tendre vers 0 :

$$(\mathcal{P}_k) \quad \inf_{x \in \overset{\circ}{S}} f(x) + c_k B(x).$$

Avantage : compte tenu de la définition de la fonction B, ce problème se traite comme un problème sans contraintes.

Si f est continue et coercive sur $\mathbb{R}^N$, le problème $(\mathcal{P}_k)$ admet au moins une solution x_k. On arrête si $c_k B(x)$ est petit, sinon on prend $0 < c_{k+1} < c_k$ et on résoud $(\mathcal{P}_{k+1})$ initialisé avec x_k solution de $(\mathcal{P}_k)$.

Proposition 4.7 *Sous les hypothèses précédentes, on suppose que f est continue, et que S est borné ou f est coercive. Soit c_k une suite positive, décroissante, telle que $\lim_{k\to\infty} c_k = 0$. Alors la suite $(x_k)_{k\geq 1}$ admet au moins une valeur d'adhérence et toute valeur d'adhérence est solution de $(\mathcal{P})$.*

Démonstration

Posons $f_k(x) = f(x) + c_k B(x)$ avec $c_k B \geq 0$. Soit $\bar{x}$ une solution de $(\mathcal{P})$. Pour $k \geq 1$, soit x_k une solution de $(\mathcal{P}_k)$. On a

$$f(\bar{x}) \leq f(x_k) \leq f_k(x_k).$$

Soit $\varepsilon > 0$. Comme f est continue et que $\overline{\overset{\circ}{S}} = S$, on peut trouver $x' \in \overset{\circ}{S}$ tel que $f(x') \leq f(\bar{x}) + \varepsilon$.

On a

$$\begin{array}{rcl} f_k(x_k) & \leq & f_k(x') = f(x') + c_k B(x') \\ & \leq & f(\bar{x}) + \varepsilon + c_k B(x') \end{array}.$$

D'où à la limite

$$f(\bar{x}) \leq \liminf_{k\to\infty} f_k(x_k) \leq \limsup_{k\to\infty} f_k(x_k) \leq f(\bar{x}) + \varepsilon,$$

ce qui prouve que $f_k(x_k)$ converge vers $f(\bar{x})$.

D'autre part la suite (x_k) est bornée, on peut donc en extraire une sous-suite $(x_{k_n})_n$ convergeant vers $y \in S$, et

$$f(y) = \lim_{n\to\infty} f_{k_n}(x_{k_n}) = f(\bar{x}).$$

■

4.5 Méthodes utilisant la résolution des conditions KKT

On se place dans le cas contraintes égalités (cf. surface active) :

$$(\mathcal{P})\left\{\begin{array}{l} \min f(x) \\ h_i(x) = 0 \quad i = 1,\ \ldots,\ q \end{array}\right.$$

Les conditions de Lagrange s'écrivent

$$(\mathcal{L})\left\{\begin{array}{r} \nabla f(x) + Dh^T(x)\lambda = 0 \\ h(x) = 0, \end{array}\right.$$

ou encore, avec $L(x,\lambda) = f(x) + h(x).\lambda$:

$$\nabla L(x,\lambda) = 0.$$

Pour résoudre ce problème, on utilise l'itération de Newton :

$$\nabla^2 L(x_k,\lambda_k)\begin{pmatrix} x_{k+1} - x_k \\ \lambda_{k+1} - \lambda_k \end{pmatrix} = -\nabla L(x_k,\lambda_k),$$

c'est à dire

$$\begin{pmatrix} \nabla_x^2 L(x_k,\lambda_k) & Dh^T(x_k) \\ Dh(x_k) & 0 \end{pmatrix}\begin{pmatrix} x_{k+1} - x_k \\ \lambda_{k+1} - \lambda_k \end{pmatrix} = -\begin{pmatrix} \nabla f(x_k) + Dh^T(x_k)\lambda_k \\ h(x_k) \end{pmatrix}$$

On pose $H_k = \nabla_x^2 L(x_k,\lambda_k)$, $D_k = Dh(x_k)$, $d_k = x_{k+1} - x_k$ et $y_k = \lambda_{k+1} - \lambda_k$. Le système précédent s'écrit alors

$$\left\{\begin{array}{lcl} H_k d_k + D_k^T y_k & = & -\nabla_x L(x_k,\lambda_k) \\ D_k d_k & = & -h(x_k) \end{array}\right. \tag{4.7}$$

Ce système est à résoudre pour chaque itération. On peut employer les méthodes de type quasi-Newton pour le calcul de H_k.

On peut vérifier que (d_k, y_k) est une direction de descente pour la *fonction mérite*

$$m(x,\lambda) = ||\nabla f(x) + Dh^T(x)\lambda||^2 + ||h(x)||^2$$

(cf. également la remarque en fin du paragraphe 2.2.3). On en déduit le résultat suivant, intéressant dans la mesure où l'on n'est pas obligé d'être proche de la solution à l'initialisation de l'algorithme :

Proposition 4.8 *On suppose que le système (4.7) est inversible pour tout k. On prend comme nouveau point*

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k d_k \\ \lambda_{k+1} &= \lambda_k + \alpha_k y_k \end{aligned}$$

où α_k minimise $m(x_k+\alpha d_k, \lambda_k+\alpha y_k)$ (Si $\alpha_k = 1$, c'est la méthode de Newton). Si la suite (x_k, λ_k) est bornée, alors toute valeur d'adhérence est solution des équations de Lagrange ($\mathcal{L}$).

Le système (4.7) s'écrit aussi

$$(2)\left\{\begin{array}{lcl} H_k d_k + D_k^T \lambda_{k+1} & = & -\nabla f(x_k) \\ D_k d_k & = & -h(x_k) \end{array}\right.$$

qui peut-être interprété comme étant les conditions de Lagrange du problème quadratique suivant :

$$(QP)\left\{\begin{array}{l} \min \frac{1}{2} H_k d.d + \nabla f(x_k).d \\ D_k d + h(x_k) = 0 \end{array}\right.$$

Méthode SQP

Cette interprétation a été généralisée au cas des contraintes d'inégalité de la manière suivante. Les conditions KKT associées au problème

$$\left\{\begin{array}{l} \min f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{array}\right.$$

sont résolues de manière itérative : la direction $d_k = x_{k+1} - x_k$ et les multiplicateurs de Lagrange μ_{k+1}, λ_{k+1} sont obtenus en résolvant les conditions KKT du sous-problème :

$$(QP)'\left\{\begin{array}{l} \min \frac{1}{2} H_k d.d + \nabla f(x_k).d \\ Dg(x_k)d + g(x_k) \leq 0 \\ Dh(x_k)d + h(x_k) = 0 \end{array}\right.$$

où $H_k = \nabla_x^2 L(x, \lambda)$, $L(x, \mu, \lambda) = f(x) + g(x).\mu + h(x).\lambda$ (la condition $h(x_k) = 0$ n'est pas forcément satisfaite). A chaque itération, on doit résoudre un sous-problème quadratique (QP)', ce qui peut être fait par une méthode de surfaces actives. Les modifications dans (QP)' par rapport au problème initial sont la linéarisation des contraintes et le remplacement de la fonction objectif par un approximation quadratique.

Chapitre 5

Optimisation avec contraintes : méthodes duales

On considère toujours le problème

$$(\mathcal{P})\begin{cases}\min f(x)\\ x\in S\end{cases}$$

où $S=\{x\in\mathbb{R}^N;\ g(x)\leq 0\}$. Le Lagrangien est

$$\begin{array}{rccc} L & : & \mathbb{R}^N\times\mathbb{R}^p_+ & \longrightarrow & \mathbb{R}\\ & & (x,\lambda) & \mapsto & L((x,\lambda)=f(x)+g(x).\lambda\end{array}$$

Définition 5.1 *Soit $\bar{x}\in\mathbb{R}^N$, $\bar{\lambda}\in\mathbb{R}^p_+$. On dit que $(\bar{x},\bar{\lambda})$ est un point selle si*

$$\begin{cases} L(\bar{x},\bar{\lambda}) & \leq & L(x,\bar{\lambda}) & \forall x\in\mathbb{R}^N\\ L(\bar{x},\bar{\lambda}) & \geq & L(\bar{x},\lambda) & \forall \lambda\in\mathbb{R}^p_+\end{cases}$$

Théorème 5.1 *Si $(\bar{x},\bar{\lambda})$ est un point selle, alors $\bar{x}$ est solution de $(\mathcal{P})$. De plus, on a*

$$\bar{\lambda}.g(\bar{x})=0$$

Démonstration
• Montrons que $g(\bar{x})\leq 0$.
$(\bar{x},\bar{\lambda})$ est un point selle donc $L(\bar{x},\bar{\lambda})\geq L(\bar{x},\lambda)\quad\forall\lambda\geq 0$, d'où

$$g(\bar{x}).(\bar{\lambda}-\lambda)\geq 0\quad\forall\lambda\geq 0,$$

et

$$g_i(\bar{x})\leq 0\quad\forall i=1,\ldots,p.$$

De plus pour $\lambda = 0$, on a $g(\bar{x}).\bar{\lambda} \geq 0$ avec $g(\bar{x}) \leq 0$ et $\bar{\lambda} \geq 0$, donc

$$g(\bar{x}).\bar{\lambda} = 0$$

• $\bar{x}$ minimise f sur $\{g(x) \leq 0\}$: si $g(x) \leq 0$, alors

$$f(\bar{x}) = f(\bar{x}) + \underbrace{g(\bar{x}).\bar{\lambda}}_{=0} = L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda})$$

car c'est un point selle.
Donc

$$f(\bar{x}) \leq f(x) + \underbrace{g(x)}_{\leq 0} . \underbrace{\bar{\lambda}}_{\geq 0} \leq f(x).$$

■

Remarques :

1. on en déduit que si $(\bar{x}', \bar{\lambda}')$ est un autre point selle alors

$$L(\bar{x}, \bar{\lambda}) = L(\bar{x}', \bar{\lambda}') = f(\bar{x}) = f(\bar{x}')$$

   et même

$$L(\bar{x}, \bar{\lambda}') = L(\bar{x}', \bar{\lambda}).$$

2. il n'existe pas forcément de point-selle.

3. la réciproque n'est pas toujours vraie, sauf dans le cas particulier des fonctions convexes :

Proposition 5.2 *On suppose f et g convexes. Si $(\mathcal{P})$ a une solution $\bar{x}$ qui vérifie QC (cf. prop. 3.5), alors il existe $\bar{\lambda} \in \mathbb{R}^p_+$ tel que $(\bar{x}, \bar{\lambda})$ soit un point-selle. De plus, si f et g sont différentiables, alors*

$$\nabla f(\bar{x}) + Dg(\bar{x})^T \bar{\lambda} = 0. \tag{5.1}$$

Démonstration
Nous montrons juste l'égalité (5.1) : comme le point $\bar{x}$ minimise $L(x, \bar{\lambda})$ sur $\mathbb{R}^N$, on a $\nabla_x L(\bar{x}, \bar{\lambda}) = 0$, i.e.

$$\nabla f(\bar{x}) + Dg(\bar{x})^T \bar{\lambda} = 0.$$

■

5.1 Problème primal et problème dual

Posons

$$\bar{f}(x) = \begin{cases} f(x) \text{ si } g(x) \leq 0 \\ +\infty \text{ sinon} \end{cases}$$

On a $\sup_{\lambda \geq 0} L(x, \lambda) = \bar{f}(x)$, donc $(\mathcal{P})$ s'écrit de manière équivalente :

problème primal : $$(\mathcal{P}) \inf_{x \in \mathbb{R}^N} \sup_{\lambda \geq 0} L(x, \lambda)$$

Par définition le problème dual est le problème

problème dual : $$(\mathcal{P}') \quad \sup_{\lambda \geq 0} \inf_{x \in \mathbb{R}^N} L(x, \lambda)$$

La fonction

$$\psi(\lambda) = \inf_{x \in \mathbb{R}^N} L(x, \lambda)$$

est appelée la fonction duale. Le problème dual est donc le problème de maximisation

$$(\mathcal{P}') \quad \sup_{\lambda \geq 0} \psi(\lambda).$$

Dans le cas de contraintes d'égalité dans le problème $(\mathcal{P})$, l'approche est la même, à ceci près qu'on ne met pas de conditions de positivité sur λ.

Remarque : la fonction ψ est l'inf d'une famille de fonctions concaves (car affines) :

$$\psi(.) = \inf_{x \in \mathbb{R}^N} L(x, .)$$

C'est donc une fonction concave. Par contre ψ n'est pas forcément différentiable. On peut montrer que s'il existe x tel que $\psi(\lambda) = L(x, \lambda)$, alors ψ est dérivable au point λ si et seulement si x est un minimum unique de $L(., \lambda)$. Dans ce cas $\nabla\psi(\lambda) = g(x)$.

Question : quel est le rapport entre $(\mathcal{P})$ et $(\mathcal{P}')$? On a toujours

$$\sup_{\lambda \geq 0} \inf_{x \in \mathbb{R}^N} L(x, \lambda) \leq \inf_{x \in \mathbb{R}^N} \sup_{\lambda \geq 0} L(x, \lambda),$$

donc $(\mathcal{P}')$ minore $(\mathcal{P})$. La différence est appelée saut de dualité.

Théorème 5.3 *Théorème de la dualité.*

i) S'il existe un point-selle $(\bar{x}, \bar{\lambda})$ alors le saut de dualité est nul :

$$f(\bar{x}) = \min_{x \in S} f(x) = \max_{\lambda \geq 0} \psi(\lambda) = \psi(\bar{\lambda}).$$

ii) Réciproquement, si $(\mathcal{P})$ a une solution $\bar{x}$ et s'il existe $\bar{\lambda} \geq 0$ tel que $f(\bar{x}) = \psi(\bar{\lambda})$ alors $(\bar{x}, \bar{\lambda})$ est un point-selle.

Démonstration

i) D'après le théorème 5.1, on a $g(\bar{x}).\bar{\lambda} = 0$ et

$$f(\bar{x}) = \min_{x \in S} f(x) = L(\bar{x}, \bar{\lambda}).$$

Le point $(\bar{x}, \bar{\lambda})$ est un point-selle, on a donc

$$L(\bar{x}, \bar{\lambda}) = \inf_{x \in \mathbb{R}^N} L(x, \bar{\lambda}) = \psi(\bar{\lambda}).$$

De plus $(\mathcal{P}')$ minore $(\mathcal{P})$, donc $\psi(\lambda) \leq f(\bar{x})$ pour tout $\lambda \geq 0$. Comme $f(\bar{x}) = \psi(\bar{\lambda})$, ceci prouve que $\psi(\bar{\lambda}) = \max_{\lambda \geq 0} \psi(\lambda)$.

ii) On a pour tout x

$$f(\bar{x}) = \psi(\bar{\lambda}) \leq L(x, \bar{\lambda}) = f(x) + g(x).\bar{\lambda}.$$

En prenant $x = \bar{x}$, on obtient

$$0 \leq g(\bar{x}).\bar{\lambda}.$$

Comme $g(\bar{x}) \leq 0$ et $\bar{\lambda} \geq 0$, on a

$$g(\bar{x}).\bar{\lambda} = 0,$$

d'où

$$L(\bar{x}, \bar{\lambda}) = f(\bar{x}) \leq L(x, \bar{\lambda}) \quad \forall x \in \mathbb{R}^N. \tag{5.2}$$

D'autre part, $g(\bar{x}) \leq 0$, donc $g(\bar{x}).\lambda \leq 0$ pour $\lambda \geq 0$. D'où

$$L(\bar{x}, \lambda) = f(\bar{x}) + g(\bar{x}).\lambda \leq f(\bar{x}) = L(\bar{x}, \bar{\lambda}) \quad \forall \lambda \geq 0,$$

ce qui prouve avec (5.2) que $(\bar{x}, \bar{\lambda})$ est un point-selle.

■

5.2 Algorithme d'Uzawa

L'intérêt du problème dual est double :
- il est concave.
- les contraintes sont simples à prendre en compte : contraintes de positivité dans le cas général, pas de contraintes dans le cas de contraintes d'égalité dans $(\mathcal{P})$.
Pour résoudre le problème dual, on peut utiliser :

- une méthode de gradient projeté dans le cas des contraintes d'inégalité dans $(\mathcal{P})$,
- une méthode de gradient dans le cas des contraintes d'égalité, c'est le cas de l'algorithme d'Uzawa décrit ci-dessous,
- des méthodes de type Newton, ...

<u>Méthode d'Uzawa :</u>
Nous décrivons cette méthode dans le cas où le problème primal est :

$$(\mathcal{P}) \left\{ \begin{array}{l} \min \frac{1}{2} Ax.x - b.x = f(x) \\ Bx = c \end{array} \right.$$

avec A matrice SDP. Ce type de problème se rencontre fréquemment lors de la mise en oeuvre des méthodes d'éléments finis, les contraintes $Bx = c$ représentent les conditions aux limites du problème, par exemple des conditions de glissement (cf. également cours éléments finis mixtes de 5ème année).

On a $L(x, \lambda) = f(x) + (Bx - c).\lambda$. Pour λ fixé, le minimum de $L(x, \lambda)$ est atteint pour $\nabla_x L(x, \lambda) = 0$, i.e.

$$0 = Ax - b + B^T\lambda.$$

En posant

$$x(\lambda) = A^{-1}(b - B^T\lambda)$$

on a

$$\begin{aligned} \psi(\lambda) &= L(x(\lambda), \lambda) \\ &= (\frac{1}{2}Ax(\lambda) - b + B^T\lambda).x(\lambda) - c.\lambda \\ &= -\frac{1}{2}A^{-1}(b - B^T\lambda).(b - B^T\lambda) - c.\lambda. \end{aligned}$$

Le problème dual s'écrit alors :

$$(\mathcal{P}') \left\{ \begin{array}{l} \max \psi(\lambda) \\ \lambda \geq 0 \end{array} \right.$$

On constate ici que le problème dual est lui aussi un problème quadratique, on maximise un fonction quadratique concave.

Calculons $\nabla\psi(\lambda)$:

$$\begin{aligned} \nabla\psi(\lambda) &= -BA^{-1}B^T\lambda + BA^{-1}b - c \\ &= Bx(\lambda) - c \end{aligned}$$

On prend donc $\nabla\psi(\lambda)$ comme direction de montée, et l'algorithme s'écrit :

$$\left\{ \begin{array}{l} \text{Choix de } x_0,\ \lambda_0 \\ \text{A l'itération } k : \\ \text{- résoudre } Ax_{k+1} = b - B^T\lambda_k \\ \text{- un pas de gradient pour } (\mathcal{P}') : \lambda_{k+1} = \lambda_k + \rho(Bx_{k+1} - c). \end{array} \right.$$

Proposition 5.4 *On suppose que A est une matrice SDP et que B est de rang maximal. Soit α la plus petite valeur propre de $BA^{-1}B^T$ $(\alpha > 0)$ et β la plus grande. Alors la méthode converge si et seulement si $\rho \in]0, \frac{2}{\beta}[$. Le meilleur choix de ρ est*

$$\rho_{opt} = \frac{2}{\alpha+\beta}.$$

La convergence est géométrique, de rapport

$$\frac{\beta-\alpha}{\beta+\alpha} = \frac{cond(BA^{-1}B^T) - 1}{cond(BA^{-1}B^T) + 1}.$$

Démonstration : cf. TD.

Variante :

L'étape k nécessite la résolution d'un système, et est donc assez coûteuse. Arrow et Hurwitz ont proposé la modification suivante :

$$\left\{ \begin{array}{l} \text{Choix de } x_0,\ \lambda_0 \\ \text{A l'étape } k : \\ \text{- descente en } x \\ x_{k+1} = x_k - t\nabla_x L(x_k, \lambda_k) \quad t > 0 \\ \text{- montée en } \lambda \\ \lambda_{k+1} = \lambda_k + \rho(Bx_{k+1} - c) \quad \rho > 0. \end{array} \right.$$

Cette méthode consiste à résoudre alternativement le problème primal et le problème dual.

5.3 Lagrangien augmenté

C'est une méthode qui peut s'interpréter à la fois comme une méthode de pénalisation et comme une méthode duale. Soit le problème d'optimisation suivant :

$$(\mathcal{P}) \qquad \min_{h(x)=0} f(x)$$

où les fonctions f et h sont de classe $\mathcal{C}^2$. Comme pour la pénalisation extérieure, on choisit une suite croissante de nombres strictement positifs $(c_k)_{k\geq 0}$ telle que $\lim_{k\to+\infty} c_k = +\infty$. On définit le Lagrangien augmenté sur $\mathbb{R}^N \times \mathbb{R}^q \times \mathbb{R}_+$ par

$$L_A(x, \lambda, c) = f(x) + h(x).\lambda + \frac{c}{2}\sum_{i=1}^{q} h_i(x)^2.$$

La méthode est la suivante : on initialise x_0 et λ_0. Puis, pour $k \geq 0$, on résout approximativement le problème

$$\mathcal{Q}_k(\lambda_k) \qquad \min_{x\in\mathbb{R}^N} L_A(x, \lambda_k, c_k)$$

dont on note la solution (approchée) x_k. On passe alors au problème $\mathcal{Q}_{k+1}(\lambda_{k+1})$ en prenant

$$\lambda_{k+1} = \lambda_k + c_k h(x_k).$$

L'un des intérêts de cette méthode est le suivant, qui rappelle ce que nous avions observé avec la pénalisation exacte.

Proposition 5.5 *Soit x^* un point régulier solution du problème $(\mathcal{P})$. On suppose que ce point satisfait la condition nécessaire d'optimalité du premier ordre, le multiplicateur de Lagrange étant noté λ^*, ainsi que la condition suffisante du second ordre. Alors il existe $C > 0$ tel que pour tout $c \geq C$, x^* est également un minimum local strict de la fonction $x \mapsto L_A(x, \lambda^*, c)$.*

Démonstration : cf. TD.

A noter que si λ est proche de λ^*, on peut en déduire par continuité que $L_A(x, \lambda, c)$ aura un minimum proche de x^*. En fait, on peut montrer que la méthode du Lagrangien augmenté converge localement, c.a.d. que $x_k \to x^*$ et $\lambda_k \to \lambda^*$ si x_0 et λ_0 sont suffisamment proches de x^* et λ^* [7].

Le choix de λ_{k+1} ci-dessus peut par ailleurs être justifié de la manière suivante. Pour c_k fixé, notons $g(x)$ la fonction $f(x) + c_k/2\sum_{i=1}^{q} h_i(x)^2$. Si l'on considère la minimisation de g sous la contrainte $h(x) = 0$, problème équivalent au problème initial, la fonction duale associée à maximiser est

$$\psi(\lambda) = \inf_{x\in\mathbb{R}^N} L_A(x, \lambda, c_k).$$

Supposons alors que le problème $\mathcal{Q}_k(\lambda)$ admette un solution $x(\lambda)$ différentiable par rapport à λ. On a par définition

$$\psi(\lambda) = L_A(x(\lambda), \lambda, c_k) = g(x(\lambda)) + \lambda.h(x(\lambda)).$$

En dérivant cette expression, on obtient directement

$$D\psi(\lambda) = [Dg(x(\lambda)) + \lambda^T Dh(x(\lambda))]Dx(\lambda) + h(x(\lambda))^T.$$

Le terme entre crochets est nul car $x(\lambda)$ minimise la fonction $g(x) + \lambda.h(x)$ qui a pour dérivée $Dg(x) + \lambda^T Dh(x)$. Ainsi

$$\nabla\psi(\lambda) = h(x(\lambda))$$

est la direction de plus forte montée pour maximiser la fonction ψ, ce qui justifie une itération de la forme

$$\lambda_{k+1} = \lambda_k + ch(x_k)^T, \quad x_k = x(\lambda_k), \quad c > 0.$$

Cette méthode peut être généralisée aux contraintes d'inégalités, cf. [5, 6, 7].

Bibliographie

[1] A. Björck, *Numerical methods for least square problems*, SIAM, 1996.

[2] , A.R. Conn, N.I. Gould, Ph.L. Toint, *Trust-Region Methods*, MPS/SIAM Series on Optimization, 2000.

[3] P. G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Masson, 1994.

[4] J-B. Hiriart-Urruty, C. Lemarechal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, 1993.

[5] D.G. Luenberger, *Linear and Nolinear Programming*, Addison-Wesley, 1984.

[6] M. Minoux, *Programmation Mathématique, théorie et algorithmes*, Dunod, 1983.

[7] J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer, 1999.