RO04/TI07 - Optimisation non-linéaire

Stéphane Mottelet

Université de Technologie de Compiègne

Printemps 2003

Sommaire

Chapitre I
Motivations et notions fondamentales

I.1 Motivations

I.1.1 Formulation générale des problèmes d'optimisation non linéaire

La forme générale d'un problème d'optimisation est la suivante :

$$(PC)\quad \begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(I.1.1)} \\ \text{sous les contraintes} & \\ g(x) \leq 0, & \text{(I.1.2)} \\ h(x) = 0, & \text{(I.1.3)} \end{cases}$$

où les fonctions f, g et h sont typiquement non-linéaires (c'est l'objet de cette deuxième partie du cours). L'équation (VI.1.2) désigne ce que nous apelleront des *contraintes d'inégalité* et l'équation (VI.1.3) des *contraintes d'égalité*.

L'objet de ce cours est la présentation de techniques permettant de résoudre le problème *(PC)*, ainsi que des problèmes où soit un seul des deux types de contraintes est présent, soit des problèmes n'y a pas de contraintes du tout. Nous noterons ces types de problèmes ainsi :

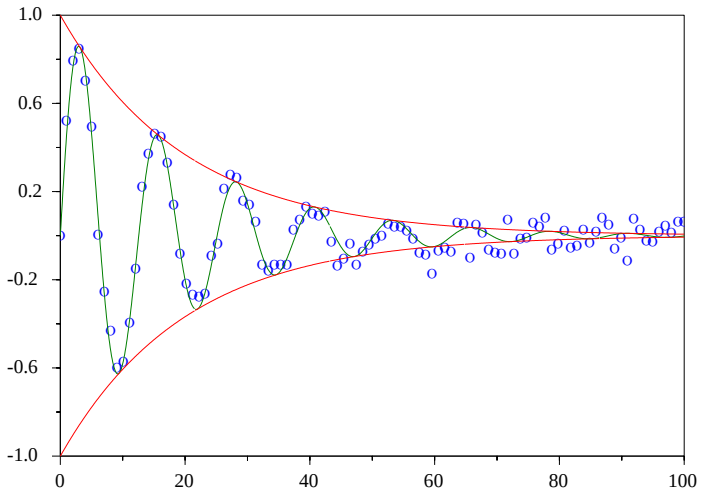
(PC)	problème général, avec contraintes d'inégalité et d'égalité,
(PCE)	problème avec contraintes d'égalité,
(PCI)	problème avec contraintes d'inégalité,
(P)	problème sans contraintes.

Il va de soi que la plupart des problèmes réels ou industriels ne sont pas initialement sous une des formes proposées. C'est pourquoi un des premiers travaux consiste en général à mettre le problème initial sous une forme standard. Par exemple, un problème donné sous la forme

$$\max_{x\in\mathbb{R}^n} g(x),$$

se mettra sous la forme standard *(P)* en posant $f(x) = -g(x)$! Cependant, la mise sous forme standard nécéssite en général un peu plus de travail, comme nous allons le voir dans les exemples qui suivent.

I.1.2 Un exemple en régression non-linéaire



On considère un problème d'identification des paramètres a, b, c et c d'un signal du type

$$f(t) = a \exp(-bt) \cos(ct + d),$$

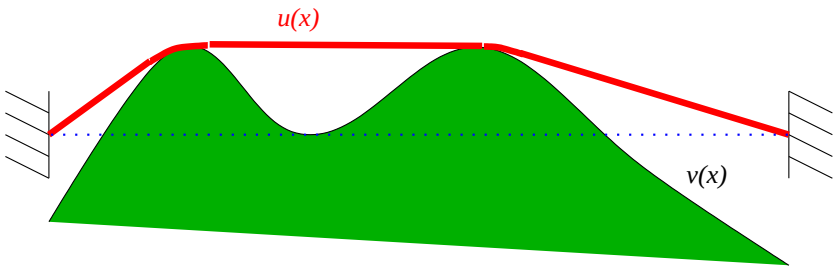
à partir d'échantillons $[t_i, y_i]_{i=1\ldots m}$ du signal $f(t)$ (ces échantillons sont représentés par les ronds sur la figure ci-dessus).

On propose de faire cette identification en minimisant la fonction

$$\begin{aligned} J(a,b,c,d) &= \frac{1}{2}\sum_{i=1}^{m}(y_i - f(t_i))^2, \\ &= \frac{1}{2}\sum_{i=1}^{m}(y_i - a\exp(-bt_i)\cos(ct_i + d))^2. \end{aligned}$$

Le choix d'élever au carré la distance entre y_i et $f(t_i)$ est bien sûr arbitraire : on aurait pu prendre la valeur absolue, mais le carré permet d'obtenir une fonction J différentiable (ceci sera bien sûr clarifié dans la suite). Si nous n'ajoutons pas de conditions sur les paramètres a, b, c, d le problème posé est donc du type (P), avec $x = [a, b, c, d]^\top \in \mathbb{R}^4$. Ce problème est communément appelé un problème de moindres carrés (non linéaire).

I.1.3 Un exemple en mécanique



On considère une corde horizontale de longueur 1 tendue à ses deux extrémités, avec une tension τ. La déviation éventuelle de la corde par rapport à sa position d'équilibre est désignée par $u(x)$, pour $x \in [0,1]$. Les extrémités étant fixées, on aura toujours $u(0) = u(1) = 0$. On négligera le poids propre de la corde par rapport à la tension τ, cela permet d'affirmer qu'en l'absence d'action extérieure, la corde est au repos et on a donc $u(x) = 0, \forall x \in [0,1]$.

Supposons maintenant que la corde est écartée de sa position d'origine. Alors on peut montrer que l'énergie potentielle associée à cette déformation (supposée petite) est

$$E(u) = \frac{1}{2}\int_0^1 \tau \left(\frac{du}{dx}\right)^2 dx. \tag{I.1.4}$$

En l'absence d'obstacle, la position de repos $u(x) = 0$ minimise cette énergie. Il peut alors être intéressant d'étudier un problème où un obstacle empêche la corde de prendre la position triviale $u(x) = 0$. Intuitivement, on voit bien que la corde va toucher l'obsctale en certains points, mais pas forcément en tous les points de l'intervalle $[0,1]$ (cela va dépendre de la forme de l'obstacle)

Supposons par exemple que cet obstacle peut être représenté par une fonction $v(x) \geq 0$. Alors la présence de l'obstacle se traduit par la condition

$$u(x) \geq v(x), x \in]0,1[. \tag{I.1.5}$$

Si on veut connaître la déformation $u(x)$ de la corde lorsque l'obstacle est présent, on peut donc penser qu'il est raisonnable de considérer le problème

$$\begin{cases} \min\limits_u \dfrac{1}{2}\displaystyle\int_0^1 \tau \left(\frac{du}{dx}\right)^2 dx, \\ u(0) = u(1) = 0, \\ u(x) \geq v(x),\ x \in]0,1[. \end{cases} \tag{I.1.6}$$

Il s'agit, techniquement parlant, d'un problème de calcul des variations, et donc l'inconnue est une fonction (la fonction $u(x)$). Il parait donc pour l'instant impossible de le mettre sous forme standard. Cependant, on peut essayer de résoudre un problème approché, en utilisant la méthode des éléments finis :

Approximation avec la méthode des éléments finis

Puisque l'on est en dimension 1 d'espace, la méthode est très simple à mettre en oeuvre. D'une part, on discrétise l'intervalle $[0,1]$: on considère les abscisses

$$x_k = \frac{k}{N},\ k = 0 \ldots N.$$

On considère le vecteur $U = [U_1, \ldots, U_{N-1}]^\top$, ainsi que la fonction $u_N(x)$ définie par :

$u_N(x_k) = U_k$, $u_N(0) = u_N(1) = 0$, de plus u_N est continue et affine par morceaux.

On peut alors montrer que

$$E(u_N) = \frac{1}{2} U^\top A U,$$

où A est la matrice (définie positive)

$$A = \tau N^2 \begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix}.$$

On peut donc proposer la version approchée du problème (I.1.6) :

$$\begin{cases} \min\limits_{U} \dfrac{1}{2} U^\top A U, \\ v(x_k) - U_k \leq 0, k = 1 \dots N-1. \end{cases} \tag{I.1.7}$$

Il s'agit donc d'un problème se mettant assurément sous la forme (PCI). De plus la fonction $f(U) = \frac{1}{2} U^\top A U$ est assez particulière : il s'agit d'une forme quadratique (nous y reviendrons plus tard). La fonction g permettant d'exprimer les contraintes d'inégalité, définie par

$$g(U) = \begin{pmatrix} v(x_1) - U_1 \\ \vdots \\ v(x_{N-1}) - U_{N-1}) \end{pmatrix},$$

est de plus *linéaire*. Nous aborderons des méthodes tenant compte de ces particularités.

I.2 Formes quadratiques

I.2.1 Définition d'une forme quadratique

Cours :
exemple en mécanique

L'exemple précédent nous donne une idée, à partir d'un problème particulier, de la forme que peut prendre la fonction f. Une telle fonction s'appelle une forme quadratique. Nous allons maintenant étudier leurs propriétés.

Définition I.2.1. *Soit A une matrice symétrique $n \times n$ et $b \in \mathbb{R}^n$. On appelle forme quadratique la fonction $f : \mathbb{R}^n \to \mathbb{R}$ définie par*

$$f(x) = \frac{1}{2} x^\top A x - b^\top x.$$

Lorsque la matrice A possède certaines propriétés, la fonction f peut prendre un nom particulier. La propriété à laquelle nous allons nous intéresser est la positivité :

Définition I.2.2. *Soit A une matrice symétrique $n \times n$ et $b \in \mathbb{R}^n$. On dit que A est semi-définie positive et on note $A \geq 0$, quand*

$$x^\top A x \geq 0, \ \forall x \in \mathbb{R}^n.$$

On dit que A est définie positive et on note $A > 0$, quand

$$x^\top A x > 0, \ \forall x \in \mathbb{R}^n, \ x \neq 0.$$

Cette définition peut être reliée aux valeurs propres de la matrice A :

Propriété I.2.3. *Soit A une matrice symétrique $n \times n$. On note $\{\lambda_i\}_{i=1\ldots n}$ ses valeurs propres (réelles). On a les équivalences suivantes :*

$$A \geq 0 \iff \lambda_i \geq 0, \ i = 1 \ldots n,$$

$$A > 0 \iff \lambda_i > 0, \ i = 1 \ldots n.$$

Lorsque la matrice A est définie positive (resp. semi-définie positive), on dira que $f(x)$ est une forme quadratique définie positive (resp. semi-définie positive). Dans le cas où A est définie positive la fonction f possède un certain nombre de propriétés. Nous nous intéressons dans un premier temps aux surfaces $f(x) = c$ où $c \in \mathbb{R}$.

I.2.2 Propriétés des formes quadratiques définies positives

Exemples :
Exemple I.1

Propriété I.2.4. *Soit A une matrice symétrique $n \times n$, définie positive et $b \in \mathbb{R}^n$. Considérons la forme quadratique*

$$f(x) = \frac{1}{2} x^\top A x - b^\top x.$$

On considère la famille de surfaces définie par

$$\gamma_c = \{x \in \mathbb{R}^n,\ f(c) = c\},$$

pour $c \in \mathbb{R}$, et on définit le vecteur $\hat{x}$ solution de

$$A\hat{x} = b.$$

Alors γ_c est définie de la façon suivante :
- *Si $c < f(\hat{x})$ alors $\gamma_c = \emptyset$.*
- *Si $c = f(\hat{x})$ alors $\gamma_c = \{\hat{x}\}$.*
- *Si $c > f(\hat{x})$ alors γ_c est un ellipsoïde centré en $\hat{x}$.*

Démonstration : La matrice A étant diagonalisable, il existe une matrice P (la matrice des vecteurs propres) orthogonale telle que

$$P^\top A P = D,$$

où $D = \text{diag}\,(\lambda_1, \ldots, \lambda_n)$ avec $\lambda_i > 0$. On fait le changement de variable $y = x - \hat{x}$: cela donne

$$f(\hat{x} + y) = f(\hat{x}) + (A\hat{x} - b)^\top y + \frac{1}{2} y^\top A y,$$

et puisque $A\hat{x} = b$, on a

$$f(x) = f(\hat{x}) + \frac{1}{2}(x - \hat{x})^\top A (x - \hat{x}).$$

On fait maintenant le changement de variable $(x - \hat{x}) = Pz$, ce qui donne

$$\begin{aligned} f(x) &= f(\hat{x}) + \frac{1}{2} z^\top P^\top A P z, \\ &= f(\hat{x}) + \frac{1}{2} z^\top D z, \\ &= f(\hat{x}) + \frac{1}{2} \sum_{i=1}^{n} \lambda_i z_i^2. \end{aligned}$$

La surface γ_c est donc définie par

$$\gamma_c = \left\{ z \in \mathbb{R}^n,\ \frac{1}{2} \sum_{i=1}^{n} \lambda_i z_i^2 = c - f(\hat{x}) \right\}.$$

Si $c - f(\hat{x}) < 0$ il est clair qu'il n'y a pas de solution à l'équation

$$\frac{1}{2} \sum_{i=1}^{n} \lambda_i z_i^2 = c - f(\hat{x}),$$

puisque le second membre est toujours positif ! Si $c = f(\hat{x})$ la seule solution est $z = 0$, c'est à dire $x = \hat{x}$. Si $c > f(\hat{x})$ l'équation définit bien un ellipsoïde, puisque les λ_i sont positifs. □

Nous avons en fait démontré un résultat très intéressant qui caractérise la valeur minimale prise par $f(x)$ quand x parcourt $\mathbb{R}^n$:

Théorème I.2.5. *Soit A une matrice symétrique $n \times n$ définie positive et $b \in \mathbb{R}^n$, et soit f la forme quadratique associée, définie par*

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

Soit $\hat{x}$ le vecteur (unique) vérifiant $A\hat{x} = b$, alors $\hat{x}$ réalise le minimum de f, c'est à dire

$$f(\hat{x}) \leq f(x),\ \forall\, x \in \mathbb{R}^n.$$

Ce résultat est une conséquence directe de la propriété I.2.4.

I.3 Rappels de calcul différentiel

I.3.1 Définition de la différentiabilité

Dans $\mathbb{R}^n$ on note x le vecteur colonne

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

et la notation $\|.\|$ désignera, sauf indication du contraire, la norme euclidienne

$$\|x\| = \left(\sum_{k=1}^{n} x_k^2 \right)^{\frac{1}{2}}.$$

Avant de donner la définition de la différentiabilité, il est important de rappeller celle de la *continuité* :

Définition I.3.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}^m$, on dit que f est continue au point $a \in \mathbb{R}^n$ si pour tout réel $\epsilon > 0$ il existe $\eta > 0$ tel que*

$$\|x - a\| < \eta \Rightarrow \|f(x) - f(a)\| < \epsilon.$$

Voici maintenant la définition de la différentiabilité :

Définition I.3.2. *Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ représentée dans la base canonique de $\mathbb{R}^m$ par le vecteur*

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}, \tag{I.3.1}$$

continue en $a \in \mathbb{R}^n$. On dit que f est différentiable en a s'il existe une application linéaire, notée $f'(a)$, telle que pour tout $h \in \mathbb{R}^n$ on ait

$$f(a + h) = f(a) + f'(a)h + \|h\| \, \epsilon(h), \tag{I.3.2}$$

où $\epsilon(.)$ est une fonction continue en 0 vérifiant $\lim_{h \to 0} \epsilon(h) = 0$. On appelle $f'(a)$ dérivée de f au point a.

La notation $f'(a)h$ doit être prise au sens "$f'(a)$ appliquée à h". Cette notation devient assez naturelle lorsque l'on représente $f'(a)$ par sa matrice dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$, comme le montre plus bas la proposition I.3.2.

I.3.2 Calcul de la dérivée première

Exemples :	Exercices :
Exemple I.3	Exercice I.2
Exemple I.2	Exercice I.1

On peut d'ores et déja donner un résultat "pratique" permettant de calculer directement la dérivée à partir du développement (I.3.2) :

Proposition I.3.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ différentiable en a, alors*

$$\lim_{t\to 0} \frac{f(a+th)-f(a)}{t} = f'(a)h.$$

Démonstration : *On a $f(a+th) = f(a) + tf'(a)h + |t|\,\|h\|\,\epsilon(th)$, d'où*

$$f'(a)h = \frac{f(a+th)-f(a)}{t} \pm \|h\|\,\epsilon(th).$$

Il suffit de noter que $\lim_{t\to 0}\epsilon(th) = 0$ pour conclure. □

La quantité $f'(a)h$ est appellée communément *dérivée directionnelle* de f au point a dans la direction h. La proposition suivante fait le lien entre la matrice de $f'(a)$ et les dérivées partielles de f au point a :

Proposition I.3.2. *Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ différentiable en a, alors on peut représenter $f'(a)$ par sa matrice dans les bases canoniques de $\mathbb{R}^n$ et de $\mathbb{R}^m$ et on a*

$$[f'(a)]_{ij} = \frac{\partial f_i}{\partial x_j}(a)$$

Démonstration : *On note $\{e^1, \ldots, e^n\}$ la base canonique de $\mathbb{R}^n$. Par définition de la matrice, la $j^{ème}$ colonne de $f'(a)$ est obtenue en appliquant $f'(a)$ au $j^{ème}$ vecteur de la base canonique de $\mathbb{R}^n$. On obtient donc le vecteur*

$$f'(a)e^j = \lim_{t\to 0} \frac{f(a+te^j)-f(a)}{t},$$

grâce à la proposition I.3.1. La définition de f donnée par (I.3.1) permet d'écrire que

$$\begin{aligned}
[f'(a)e^j]_i &= \lim_{t\to 0} \frac{f_i(a+te^j)-f_i(a)}{t}, \\
&= \lim_{t\to 0} \frac{f_i(a_1,\ldots,a_j+t,\ldots,a_n)-f_i(a_1,\ldots,a_n)}{t}, \\
&= \frac{\partial f_i}{\partial x_j}(a).
\end{aligned}$$

□

On appelle souvent $f'(a)$ la matrice *jacobienne* de f au point a. Lorsque $m = 1$ on adopte une notation et un nom particuliers : le *gradient* est le vecteur noté $\nabla f(a)$ et défini par

$$f'(a) = \nabla f(a)^\top,$$

et on a

$$f(a+h) = f(a) + \nabla f(a)^\top h + \|h\|\,\epsilon(h).$$

I.3.3 Dérivée seconde

Exemples :	Exercices :
Exemple I.4	Exercice I.4
	Exercice I.3

On se place maintenant dans le cas $m = 1$, soit $f : \mathbb{R}^n \to \mathbb{R}$.

Définition I.3.3. *L'application $f : \mathbb{R}^n \to \mathbb{R}$ est dite deux fois différentiable s'il existe une matrice symétrique $\nabla^2 f(a)$ telle que*

$$f(a+h) = f(a) + \nabla f(a)^\top h + h^\top \nabla^2 f(a) h + \|h\|^2 \epsilon(h).$$

On appelle $\nabla^2 f(a)$ matrice hessienne de f au point a. Comme l'énonce le théorème suivant (non démontré), cette matrice s'obtient à partir des dérivées secondes de f :

Théorème I.3.4. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable en un point a. Si on note $g(x) = \nabla f(x)$ alors la matrice hessienne est définie par $\nabla^2 f(a) = g'(a)$, soit*

$$[\nabla^2 f(a)]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

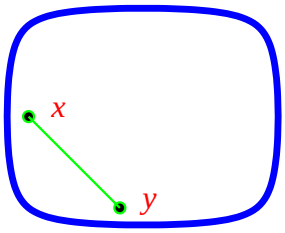
I.4 Notions sur la convexité

I.4.1 Définition de la convexité

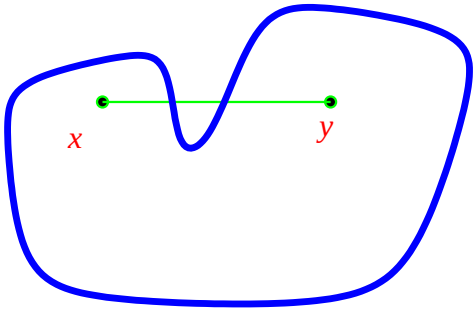
Exemples :
Exemple I.5

La convexité est à la base une propriété géométrique, assez intuitive d'ailleurs, qui permet de caractériser certains objets. On voit assez bien ce qu'est un objet convexe dans un espace à deux ou trois dimensions. Nous allons maintenant montrer comment cette propriété peut aussi s'appliquer aux fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$.

objet convexe



objet non convexe



Définition I.4.1. *Un ensemble $K \subset \mathbb{R}^n$ est dit convexe si pour tout couple $(x, y) \in K^2$ et $\forall\, \lambda \in [0, 1]$ on a*

$$\lambda x + (1 - \lambda)y \in K.$$

Cette définition peut s'interpréter en disant que le segment reliant x et y doit être dans K. Elle se généralise de la façon suivante : on dira qu'un vecteur y est une combinaison convexe des points $\{x^1, \ldots, x^p\}$ si on a

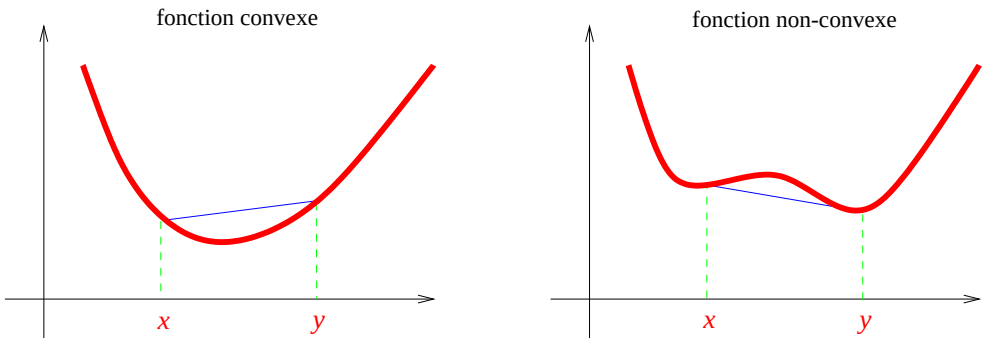
$$y = \sum_{i=1}^{p} \lambda_i x^i,$$

avec $\lambda_i \geq 0$ et $\sum_{i=1}^{p} \lambda_i = 1$.

On peut citer quelques cas particuliers : $\mathbb{R}^n$ tout entier est un ensemble convexe, de même qu'un singleton $\{a\}$.

Propriété I.4.2. *Soit une famille $\{K_i\}_{i=1\ldots p}$ d'ensembles convexes et $S = \bigcap_{i=1}^{p} K_i$. Alors S est convexe.*

I.4.2 Fonctions convexes



Définition I.4.3. *On dit qu'une fonction $f : K \to \mathbb{R}$, définie sur un ensemble convexe K, est convexe si elle vérifie*

$$\forall (x,y) \in K^2,\ \forall \lambda \in [0,1],\ \ f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

On dira que f est strictement convexe si

$$\forall (x,y) \in K^2,\ x \neq y,\ \forall \lambda \in]0,1[,\ \ f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y).$$

Lorsque $n = 1$ cette définition s'interprète bien géométriquement : le graphe de la fonction est toujours en dessous du segment reliant les points $(x, f(x))$ et $(y, f(y))$.

Corollaire I.4.4. *On définit pour $(x,y) \in K^2$, où K est un ensemble convexe, la fonction $\varphi : [0,1] \to \mathbb{R}$ par*

$$\varphi(t) = f(tx + (1-t)y).$$

Alors on a l'équivalence

$$\varphi(t) \text{ convexe sur } [0,1],\ \forall (x,y) \in K^2 \Leftrightarrow f \text{ convexe sur } K.$$

Démonstration : *Si $\varphi(t)$ est convexe sur $[0,1]$ on a en particulier*

$$\varphi(\lambda) \leq \lambda \varphi(1) + (1-\lambda)\varphi(0),\ \forall \lambda \in [0,1],$$

ce qui donne exactement

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

La réciproque est admise. □

I.4.3 Caractérisation de la convexité en termes du hessien

Exemples :
Exemple I.6

Dans le cas où $f : K \subset \mathbb{R} \to \mathbb{R}$ on a le résultat suivant :

Propriété I.4.5. *Si $f : \mathbb{R} \to \mathbb{R}$ est 2 fois continûment dérivable sur K convexe alors f est convexe si et seulement si $f''(x) \geq 0$,$\forall x \in K$ et strictement convexe si et seulement si $f''(x) > 0$,$\forall x \in K$ (sauf éventuellement en des points isolés).*

Ce résultat se généralise pour $n > 1$: le résultat suivant fait le lien entre le hessien et la propriété de convexité :

Théorème I.4.6. *Soit $f : K \subset \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable, alors f est convexe si et seulement si $\nabla^2 f(x) \geq 0,\ \forall x \in K$, et strictement convexe si et seulement si $\nabla^2 f(x) > 0,\ \forall x \in K$.*

Démonstration : *La démonstration fait appel à un résultat obtenu dans l'exercice I.1 : si on définit $\varphi(t) = f(x + ty)$ alors on a*

$$\varphi''(t) = y^\top \nabla^2 f(x + ty) y,$$

et on sait grâce a la propriété I.4.5 que f convexe si $\varphi''(t) \geq 0$, $\forall t$. On aura donc f convexe si et seulement si

$$y^\top \nabla^2 f(x + ty) y \geq 0,\ \forall (x, y) \in K^2,$$

d'où le résultat. □

Le corrolaire suivant est immédiat :

Propriété I.4.7. *Soit f une forme quadratique définie par*

$$f(x) = \frac{1}{2} x^\top A x - b^\top x,$$

alors f est convexe si et seulement si $A \geq 0$, et strictement convexe si et seulement si $A > 0$.

Cela provient du fait que $\nabla^2 f(x) = A$ (voir l'exemple I.4).

I.4.4 Caractérisation de la convexité en termes du gradient

Dans le cas où la fonction f n'est supposée qu'une fois différentiable, on a le résultat suivant :

Théorème I.4.8. *Soit $f : K \subset \mathbb{R}^n \to \mathbb{R}$ une fonction une fois différentiable, alors f est convexe si et seulement si*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x),\ \forall (x, y) \in K^2.$$

La fonction f est strictement convexe si et seulement si

$$f(y) > f(x) + \nabla f(x)^\top (y - x),\ \forall (x, y) \in K^2,\ x \neq y.$$

On voit bien l'interprétation géométrique de ce dernier resultat quand $n = 1$: le graphe d'une fonction convexe f se trouve toujours au-dessus de la tangente en un point donné.

I.5 Résultats d'existence et d'unicité

I.5.1 Théoremes généraux d'existence

Considérons notre problème d'optimisation I.1.1 introduit au début du cours, que l'on écrira pour l'occasion un peu différemment, en mettant les contraintes sous la forme $x \in K \subset \mathbb{R}^n$:

$$\min_{x \in K} f(x). \tag{I.5.1}$$

Nous allons donner deux résultats très généraux d'existence d'une solution au problème (I.5.1). Auparavant nous avons besoin de la définition d'un ensemble compact :

Définition I.5.1. *Un ensemble $K \subset \mathbb{R}^n$ est dit compact si, de toute suite $\{x_k\}$, où $x_k \in K,\ \forall k$, on peut extraire une sous-suite convergente.*

Nous donnons le théorème suivant sans démonstration :

Théorème I.5.2. *Un ensemble $K \subset \mathbb{R}^n$ est compact si et seulement si il est fermé et borné.*

Dans $\mathbb{R}$, les intervalles fermés du type $[a, b]$ (ou des reunions de tels intervalles) sont compacts. La notion de fermeture signifie qu'une suite $\{x_k\}$, où $x_k \in K,\ \forall k$, doit converger vers une limite $x \in K$. Pour illustrer sur un exemple qu'un intervalle ouvert dans $\mathbb{R}$ ne peut pas être compact, on peut considérer l'exemple suivant.

Soit $K =]0, 1]$ et la suite $x_k = 1/k$, on a bien $x_k \in K$ mais $\lim_{k\to\infty} = 0 \notin K$.

Voici maintenant deux résultats d'existence, dont les démonstrations peuvent ètre consultées dans les documents.

Théorème I.5.3. *Si $f : K \in \mathbb{R}^n \to \mathbb{R}$ est continue et si de plus K est un ensemble compact, alors le problème (I.5.1) admet une solution optimale $\hat{x} \in K$, qui vérifie donc*

$$f(\hat{x}) \leq f(x),\ \forall x \in K.$$

Le second résultat est moins général car il considère le cas particulier $K = \mathbb{R}^n$:

Théorème I.5.4. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue sur $\mathbb{R}^n$. Si*

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

alors (I.5.1) admet une solution optimale $\hat{x}$.

Démonstration : *Soit $x_0 \in \mathbb{R}^n$. Puisque $\lim_{\|x\|\to\infty} f(x) = \infty$ il existe $M > 0$ tel que $\|x\| > M \Rightarrow f(x) > f(x_0)$, donc*

$$\exists M > 0,\ f(x) \leq f(x_0) \Rightarrow \|x\| \leq M.$$

Puisque $\hat{x}$ est caractérisé par $f(\hat{x}) \leq f(x),\ \forall x \in \mathbb{R}^n$, on a donc forcément $\|\hat{x}\| \leq M$. Donc $\hat{x}$ est solution du problème

$$\min_{\|x\| \leq M} f(x),$$

et le théorème précédent s'applique, la boule $\{x \in \mathbb{R}^n,\ \|x\| \leq M\}$ étant compacte. □

I.5.2 Unicité

L'unicité résulte en général de propriétés de convexité (de f et de K).

Théorème I.5.5. *Soit $f : K \in \mathbb{R}^n \to \mathbb{R}$ strictement convexe sur K convexe. Le minimum de f sur K, s'il existe, est unique.*

Démonstration : Soit donc $\hat{x} \in K$ tel que $f(\hat{x}) \leq f(x), \forall x \in K$. Supposons qu'il existe $\hat{y} \neq \hat{x}$ tel que $f(\hat{y}) \leq f(x), \forall x \in K$. Formons pour $\lambda \in]0,1[$ le vecteur

$$u = \lambda\hat{y} + (1-\lambda)\hat{x}.$$

D'après la stricte convexité de f et puisque nécessairement $f(\hat{y}) = f(\hat{x})$ on a

$$f(u) < \lambda f(\hat{y}) + (1-\lambda)f(\hat{x}) = f(\hat{x}),$$

ce qui contredit le fait que $\hat{x}$ soit un minimum. On a donc $\hat{x} = \hat{y}$. □

I.6 Conditions nécessaires d'optimalité en l'absence de contraintes

I.6.1 Conditions nécessaires

On va maintenant regarder de plus près le cas où $K = \mathbb{R}^n$, c'est à dire le problème sans contraintes (P). Dans le cas où f est différentiable, on a le résultat suivant :

Théorème I.6.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ différentiable et $\hat{x}$ vérifiant*

$$f(\hat{x}) \leq f(x),\ \forall x \in \mathbb{R}^n,$$

alors on a nécessairement

$$\nabla f(\hat{x}) = 0.$$

Démonstration : *Pour tout $t \in \mathbb{R}^*$ et pour tout $h \in \mathbb{R}^n$ on a*

$$f(\hat{x}) \leq f(\hat{x} + th).$$

On a donc

$$lim_{t\to 0^+} \frac{f(\hat{x}) - f(\hat{x} + th)}{t} = \nabla f(\hat{x})^\top h \leq 0,$$

et

$$lim_{t\to 0^-} \frac{f(\hat{x}) - f(\hat{x} + th)}{t} = \nabla f(\hat{x})^\top h \geq 0,$$

donc $\nabla f(\hat{x})^\top h = 0$, $\forall h \in \mathbb{R}^n$, donc $\nabla f(\hat{x}) = 0$ (prendre par exemple $h = \nabla f(\hat{x})$). □

I.6.2 Conditions nécessaires et suffisantes

La condition de gradient nul devient suffisante dans le cas où f est convexe :

Théorème I.6.2. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe et différentiable. Si $\hat{x}$ vérifie*

$$\nabla f(\hat{x}) = 0,$$

alors on a $f(\hat{x}) \leq f(x),\ \forall x \in \mathbb{R}^n$.

Démonstration : *Soient $x \in \mathbb{R}^n$ et $\lambda \in [0,1]$. Puisque f est convexe on a*

$$f(\lambda\hat{x} + (1-\lambda)x) \leq \lambda f(x) + (1-\lambda) f(\hat{x}).$$

On retranche $f(\hat{x})$ de chaque côté de l'inégalité, on note que

$$\lambda x + (1-\lambda)\hat{x} = \hat{x} + \lambda(x - \hat{x}),$$

puis in divise par λ, ce qui donne l'inégalité

$$\frac{f(\hat{x} + \lambda(x-\hat{x})) - f(\hat{x})}{\lambda} \leq f(x) - f(\hat{x}).$$

Et si on fait tendre λ vers 0 on obtient

$$\nabla f(\hat{x})^\top (x - \hat{x}) \leq f(x) - f(\hat{x}),$$

donc $0 \leq f(x) - f(\hat{x})$. □

Lorsque la fonction n'est pas convexe, on ne peut donner qu'une condition nécessaire et suffisante d'optimalité locale. On désignera par minimum local (que l'on oppose au minimum global) un vecteur vérifiant les conditions suivantes :

Définition I.6.3. *On appellera x^* minimum local de f, s'il existe $\delta > 0$ tel que*

$$f(x^*) \leq f(x),\ \forall x,\ \|x - x^*\| \leq \delta.$$

Dans le cas où f est deux fois différentiable on peut alors donner le résultat suivant :

Théorème I.6.4. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois différentiable. Si*

$$\left\{ \begin{array}{ll} \nabla f(x^*) & = 0, \\ \nabla^2 f(x^*) & > 0, \end{array} \right.$$

alors x^* est un minimum local de f.

Démonstration : *On a*

$$\begin{aligned} f(x^* + th) &= f(x^*) + t\nabla f(x^*)^\top h + \frac{t^2}{2} h^\top \nabla^2 f(x^*) h + t^2 \|h\|^2 \varepsilon(th), \\ &= f(x^*) + \frac{t^2}{2} h^\top \nabla^2 f(x^*) h + t^2 \|h\|^2 \varepsilon(h). \end{aligned}$$

On a donc pour $t > 0$

$$\frac{f(x^* + th) - f(x^*)}{t^2} = \frac{1}{2} h^\top \nabla^2 f(x^*) h + \|h\|^2 \varepsilon(th).$$

Donc si t est suffisamment petit on aura bien $f(x^* + th) - f(x^*) > 0$ puisque $\nabla^2 f(x^*) > 0$. □

Exemples du chapitre I

Exemple I.1 Courbes de niveau d'une forme quadratique dans $\mathbb{R}^2$

On considère la fonction $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ où A est une matrice symétrique 2×2 définie positive. On note P la matrice des vecteurs propres et $\lambda_1 > \lambda_2 > 0$ les deux valeurs propres. Notons $\hat{x}$ la solution du système linéaire $A\hat{x} = b$. On a montré que les courbes iso-valeurs sont définies par l'équation

$$\frac{1}{2}(\lambda_1 z_1^2 + \lambda_2 z_2^2) = c - f(\hat{x}),$$

où on a effectué le changement de variable $z = P(x - \hat{x})$. Si on a $c - f(\hat{x})$, l'équation ci-dessus définit une ellipse dans le repère (z_1, z_2), dont l'équation "canonique" est donnée par

$$\frac{{z_1}^2}{a} + \frac{{z_2}^2}{b} = 1,$$

avec

$$a = \sqrt{\frac{2(c - f(\hat{x}))}{\lambda_1}},\ b = \sqrt{\frac{2(c - f(\hat{x}))}{\lambda_2}}.$$

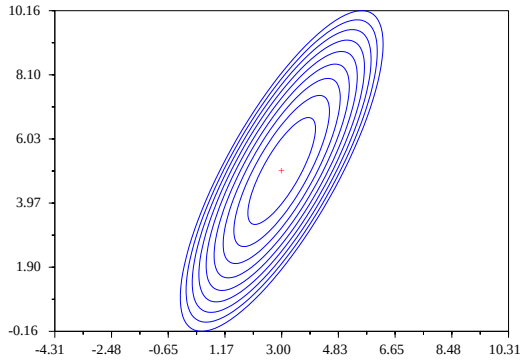
On sait que l'on peut décrire cette ellipse par la courbe paramétrique $z(t)$, $t \in [0, 2\pi]$ avec

$$z(t) = \begin{pmatrix} a\cos t \\ b\sin t \end{pmatrix},$$

donc l'équation paramétrique de la courbe $x(t)$ dans le repère original est

$$x(t) = \hat{x} + P\begin{pmatrix} a\cos t \\ b\sin t \end{pmatrix}.$$

Lancer la simulation



Exemple I.2 Gradient d'une fonction quadratique

On considère la fonction $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ où A est une matrice carrée symétrique $n \times n$. On a

$$\begin{aligned} f(x + th) &= \frac{1}{2}x^\top Ax + \frac{1}{2}t^2 h^\top Ah + tx^\top Ah + b^\top(x + th), \\ &= f(x) + t(x^\top A - b^\top)h + \frac{1}{2}t^2 h^\top Ah, \end{aligned}$$

on a donc

$$\frac{f(x + th) - f(x)}{t} = (Ax - b)^\top h + \frac{1}{2}th^\top Ah.$$

Puisque $\lim_{t\to 0} \frac{1}{2}th^\top Ah = 0$, on a donc $\nabla f(x) = Ax - b$.

Exemple I.3 Dérivée d'une fonction affine

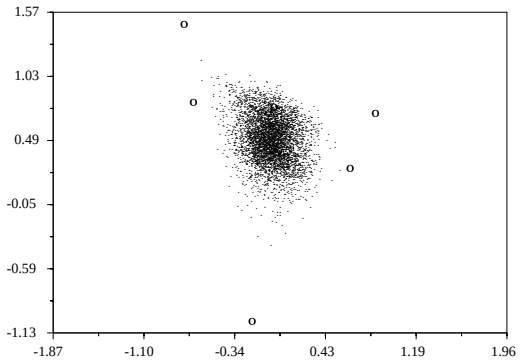
On considère la fonction $f(x) = Cx + d$ où C est une matrice $m \times n$. On a $f(x+h) = Cx + Ch + d = f(x) + Ch$. Donc $f'(x) = C$, $\forall\, x \in \mathbb{R}^n$. On notera qu'ici f est différentiable pour tout $x \in \mathbb{R}^n$, ce qui n'est pas forcément le cas quand f est quelconque.

Exemple I.4 Matrice hessienne d'une fonction quadratique

$n \times n$. L'exemple précédent nous a donné $\nabla f(x) = Ax - b$. Puisque la matrice hessienne est la dérivée du gradient on a donc $\nabla^2 f(x) = A$.

Exemple I.5 Combinaison convexe de points dans le plan

Lancer la simulation



Considérons un ensemble de points du plan $\{x^1, \ldots, x^p\}$. La simulation qui est proposée ici permet de générer aléatoirement un très grand nombre de points de la forme

$$y^k = \sum_{i=1}^{p} \lambda_i x^i,$$

en tirant aléatoirement les coefficients $\{\lambda_i\}_{i=1\ldots p}$ suivant une loi uniforme sur $[0, 1]$, renormalisés en les divisant par leur somme, de façon à ce que l'on ait toujours $\sum_{i=1}^{p} \lambda_i = 1$. Le polygone "limite" contenant tous les points générés s'appelle l'*enveloppe convexe* des points $\{x^1, \ldots, x^p\}$.

Exemple I.6 Convexité d'une fonction quadratique

On considère la fonction $f(x) = \frac{1}{2} x^\top Ax - b^\top x$ où A est une matrice carrée symétrique. Puisque $\nabla^2 f(x) = A$ (voir l'exemple précédent), f est convexe si et seulement si $A \geq 0$, strictement convexe lorsque $A > 0$

Exercices du chapitre I

Exercice I.1 Calcul d'une dérivée composée

Soit $f : \mathbb{R}^n \to \mathbb{R}$ définie par et $x : \mathbb{R} \to \mathbb{R}^n$. On définit la fonction réelle $g(t) = f(x(t))$. Calculer $g'(t)$.

Exercice I.2 Calcul du gradient d'une fonction quadratique

On considère la fonction $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ où A est une matrice $n \times n$. Montrer que l'on a

$$\nabla f(x) = \frac{1}{2}(A + A^\top)x - b.$$

Exercice I.3 Calcul d'une dérivée seconde composée

Soit $f : \mathbb{R}^n \to \mathbb{R}$ définie par et $x : \mathbb{R} \to \mathbb{R}^n$. On définit la fonction réelle $g(t) = f(x(t))$. Calculer $g''(t)$ dans le cas où $x(t) = (u + tv)$ où u et v sont deux vecteurs de $\mathbb{R}^n$, puis pour $x(t)$ quelconque.

Exercice I.4 Calcul du hessien d'une fonction quadratique

On considère la fonction $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ où A est une matrice $n \times n$. Montrer que l'on a

$$\nabla^2 f(x) = \frac{1}{2}(A + A^\top).$$

Chapitre II
Les méthodes de gradient

II.1 Les méthodes de descente

II.1.1 Principe des méthodes de descente

Définition II.1.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$. On dira qu'un vecteur d est une direction de descente en x s'il existe $\bar{t} > 0$ tel que*

$$f(x + td) < f(x),\ t \in]0, \bar{t}].$$

Le principe d'une méthode de descente consiste à faire les itérations suivantes

$$x_{k+1} = x_k + t_k d_k,\ t_k > 0, \tag{II.1.1}$$

tout en assurant la propriété

$$f(x_{k+1}) < f(x_k).$$

Le vecteur d_k est la direction de descente en x_k. Le scalaire t_k est appelé le *pas* de la méthode à l'itération k. On peut caractériser les directions de descente en x_k à l'aide du gradient :

Proposition II.1.1. *Soit $d \in \mathbb{R}^n$ vérifiant*

$$\nabla f(x)^\top d < 0,$$

alors d est une direction de descente en x.

Démonstration : *on a pour $t > 0$*

$$f(x + td) = f(x) + t\nabla f(x)^\top d + t\varepsilon(t),$$

donc si on écrit

$$\frac{f(x + td) - f(x)}{t} = \nabla f(x)^\top d + \varepsilon(t),$$

on voit bien que pour t suffisamment petit on aura $f(x + td) - f(x) < 0$. □

Dans la méthode (II.1.1) le choix de t_k est lié à la fonction

$$\varphi(t) = f(x_k + td_k),$$

en particulier, une façon de choisir t_k peut être de résoudre le problème d'optimisation (à une seule variable)

$$\min_{t>0} \varphi(t).$$

Le pas $\hat{t}_k$ obtenu ainsi s'appelle le pas optimal. La fonction $\varphi(t) = f(x_k + td_k)$ étant différentiable, on a alors nécessairement

$$\varphi'(\hat{t}_k) = \nabla f(x_k + \hat{t}_k d_k)^\top d_k = 0.$$

II.2 Les méthodes de gradient

II.2.1 Principe des méthodes de gradient

Exemples :
Exemple II.1

On cherche à déterminer la direction de descente qui fait décroitre $\varphi(t) = f(x + td)$ le plus vite possible (au moins localement). Pour cela on va essayer de minimiser la dérivée de $\varphi(t)$ en 0. On a

$$\varphi'(0) = \nabla f(x)^\top d,$$

et on cherche d solution du problème

$$\min_{d\in\mathbb{R}^n, \|d\|=1} \varphi'(0).$$

La solution est bien sûr

$$d = -\frac{\nabla f(x)}{\|\nabla f(x)\|},$$

en vertu de l'inégalité de Schwartz.

Il y a ensuite de nombreuses façon d'utiliser cette direction de descente. On peut par exemple utiliser un pas fixé a priori $t_k = \rho > 0, \forall k$.

On obtient alors la méthode du gradient simple :

$$\left\{ \begin{array}{rcl} d_k & = & -\nabla f(x_k), \\ x_{k+1} & = & x_k + \rho d_k. \end{array} \right.$$

Sous certaines hypothèses de régularité (f deux fois différentiable) cette méthode converge si ρ est choisi assez petit.

II.2.2 La méthode du gradient à pas optimal

La méthode du gradient à pas optimal consiste à faire les itérations suivantes

$$\left\{ \begin{array}{rcl} d_k & = & -\nabla f(x_k), \\ x_{k+1} & = & x_k + t_k d_k, \end{array} \right. \tag{II.2.1}$$

où t_k est choisi de manière à ce que

$$f(x_k + t_k d_k) \leq f(x_k + t d_k),\ \forall t > 0. \tag{II.2.2}$$

Cette méthode possède une propriété interessante :

Proposition II.2.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable. Les directions de descente d_k générées par la méthode (II.2.1)-(II.2.2) vérifient*

$$d_{k+1}^\top d_k = 0.$$

Démonstration : *Si on introduit la fonction $\varphi(t) = f(x_k + t d_k)$, on a*

$$\varphi'(t) = \nabla f(x_k + t d_k)^\top d_k,$$

et puisque φ est dérivable on a nécessairement $\varphi'(t_k) = 0$ donc

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

□

II.2.3 Calcul du pas optimal dans le cas quadratique

Exemples :
Exemple II.2

On a $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ avec $A > 0$ et on note $\varphi(t) = f(x_k + td_k)$. Le pas optimal t_k est caractérisé par

$$\varphi'(t_k) = 0,$$

on a donc

$$\nabla f(x_k + t_k d_k)^\top d_k = (A(x_k + t_k d_k) - b)^\top d_k = 0,$$

soit

$$(\nabla f(x_k) + t_k A d_k)^\top d_k = 0,$$

on obtient donc

$$t_k = -\frac{\nabla f(x_k)^\top d_k}{d_k^\top A d_k},$$

qui est bien positif car d_k est une direction de descente et $d_k^\top A d_k > 0$ (car $A > 0$).

La méthode du gradient à pas optimal peut donc s'écrire (dans le cas quadratique)

$$\left\{ \begin{array}{rcl} d_k & = & b - Ax_k, \\ t_k & = & \frac{d_k^\top d_k}{d_k^\top A d_k}, \\ x_{k+1} & = & x_k + t_k d_k. \end{array} \right. \tag{II.2.3}$$

Exemples du chapitre II

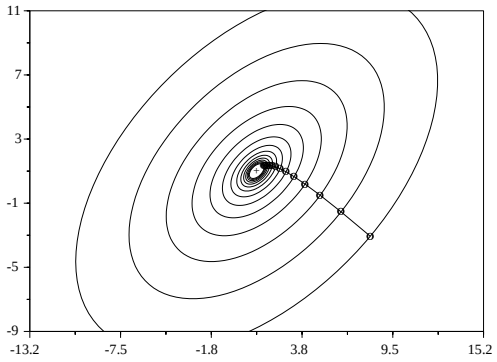
Exemple II.1 Méthode du gradient simple dans le cas quadratique

Dans le cas où $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ la méthode du gradient simple peut s'écrire

$$\left\{ \begin{array}{rcl} d_k & = & b - Ax_k, \\ x_{k+1} & = & x_k + \rho d_k, \end{array} \right. \tag{II.2.4}$$

où $\rho > 0$ est fixé a priori. Il existe bien sûr des conditions sur ρ pour que la méthode converge. Nous illustrons ici le fonctionnement de la méthode dans le cas $n = 2$ sur une petite simulation.

Lancer la simulation



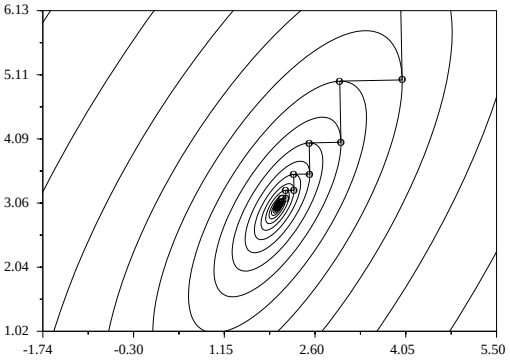
Exemple II.2 Méthode du gradient à pas optimal dans le cas quadratique

Dans le cas où $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ la méthode du gradient à pas optimal peut s'écrire

$$\left\{ \begin{array}{rcl} d_k & = & b - Ax_k, \\ t_k & = & \frac{d_k^\top d_k}{d_k^\top A d_k}, \\ x_{k+1} & = & x_k + t_k d_k, \end{array} \right. \tag{II.2.5}$$

Nous illustrons ici le fonctionnement de la méthode dans le cas $n = 2$ sur une petite simulation.

Lancer la simulation



Chapitre III
La méthode du gradient conjugué

III.1 Introduction

III.1.1 Directions conjuguées

Définition III.1.1. *Soit A une matrice symétrique $n \times n$, définie positive. On dit que deux vecteurs x et y de $\mathbb{R}^n$ sont $A-$conjugués (ou conjugués par rapport à A) s'il vérifient*

$$x^\top A y = 0. \tag{III.1.1}$$

La matrice A étant définie positive, la forme bilinéaire $a(x,y) = x^\top A y$ définit un produit scalaire et la relation (III.1.1) traduit l'orthogonalité des vecteurs x et y pour ce produit scalaire. La démonstration du théorème suivant est laissée en exercice.

Théorème III.1.2. *Si $\{d_0, d_1, \ldots, d_k\}$ sont des directions $A-$conjuguées deux à deux, soit*

$$d_i^\top A d_k = 0, \ \ \forall\, i, j,\ i < j \leq k,$$

alors elles sont linéairement indépendantes.

Considérons maintenant dans R^2 une méthode de descente appliquée à la minimisation d'une forme quadratique définie positive $f(x) = \frac{1}{2} x^\top A x - b^\top x$:

$$\begin{aligned} x_1 &= x_0 + \rho_0 d_0, \\ x_2 &= x_1 + \rho_1 d_1, \end{aligned}$$

avec d_0 et d_1 deux directions $A-$conjuguées et ρ_0 et ρ_1 déterminés de façon optimale. On a donc les relations suivantes :

$$\begin{aligned} \nabla f(x_1)^\top d_0 = (A x_1 - b)^\top d_0 &= 0, \\ \nabla f(x_2)^\top d_1 = (A x_2 - b)^\top d_1 &= 0, \end{aligned}$$

car ρ_0 et ρ_1 sont optimaux. Montrons que l'on a de plus

$$\nabla f(x_2)^\top d_0 = 0.$$

On a

$$\begin{aligned} \nabla f(x_2)^\top d_0 &= (A x_2 - b)^\top d_0 = (A(x_1 + \rho_1 d_1) - b)^\top d_0, \\ &= (A x_1 - b)^\top d_0 + \rho_1 d_1^\top A d_0, \\ &= 0. \end{aligned}$$

Puisque $\nabla f(x_2)^\top d_0 = \nabla f(x_2)^\top d_1 = 0$ et d_0, d_1 linéairement indépendants, on a $\nabla f(x_2) = 0$, x_2 réalise donc le minimum de f sur $\mathbb{R}^2$. La relation de conjugaison permet donc à la méthode de descente de converger en deux itérations (dans le cas où $n = 2$).

Définition III.1.3. *Soit $\{d_0, d_1, \ldots, d_n\}$ une famille de vecteur $A-$conjugués. On appelle alors méthode de directions conjuguées la méthode*

$$\begin{cases} x_0 & \text{donné} \\ x_{k+1} &= x_k + \rho_k d_k, \ \ \rho_k \text{ optimal} \end{cases}$$

On va maintenant montrer la propriété vérifiée pour $n = 2$, à savoir $x_n = \hat{x}$ où $\hat{x}$ réalise le minimum de $f(x) = \frac{1}{2} x^\top A x - b^\top x$, est valable pour tout n.

III.1.2 Lemme fondamental

On se donne *a priori* une famille $\{d_0, d_1, \ldots, d_n\}$ de directions conjuguées et on note

$$E_k = \text{Vect}(d_0, d_1, \ldots, d_{k-1}),$$

le sous-espace vectoriel engendré par les vecteurs $d_0, d_1, \ldots, d_{k-1}$. Par construction, l'algorithme de directions conjugué

$$\left\{ \begin{array}{rcl} x_0 & & \text{donné,} \\ x_{k+1} & = & x_k + \rho_k d_k, \ \ \rho_k \text{ optimal,} \end{array} \right. \tag{III.1.2}$$

construit itérativement un vecteur x_k vérifiant

$$x_k \in x_0 + E_k.$$

Voici l'énoncé du lemme fondamental :

Lemme III.1.4. *Le vecteur x_k défini par l'algorithme (III.1.2) réalise le minimum de $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ sur le sous espace $x_0 + E_k$, c'est à dire $x_k \in x_0 + E_k$ et*

$$f(x_k) \leq f(x), \ \ \forall\, x \in x_0 + E_k.$$

Pour la démonstration de ce lemme nous aurons besoin du théorème suivant :

Théorème III.1.5. *Une condition nécessaire et suffisante pour que $x_k \in E_k + x_0$ réalise le minimum de $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ sur le sous espace $x_0 + E_k$ est*

$$\nabla f(x_k)^\top d_i = 0, \ \ \forall\, i = 0, \ldots, k-1.$$

Démonstration : *Condition nécéssaire : supposons que $f(x_k) \leq f(x), \ \ \forall\, x \in x_0 + E_k$. On a donc pour tout $t \in \mathbb{R}$,*

$$f(x_k) \leq f(x_k + td), \ \ \forall\, d \in E_k.$$

On a donc soit

$$(f(x_k + td) - f(x_k))/t \geq 0, \ \ \text{si } t > 0,$$

soit

$$(f(x_k + td) - f(x_k))/t \leq 0, \ \ \text{si } t < 0.$$

Si l'on fait tendre t vers zéro, on en conclut que

$$\nabla f(x_k)^\top d = 0, \ \ \forall\, d \in E_k,$$

donc en particulier $\nabla f(x_k)^\top d_i = 0, \forall\, i = 0, \ldots, k-1$. On admettra que la condition est suffisante.□
□

Démonstration du lemme fondamental : Pour $k = 1$ on a

$$x_1 = x_0 + \rho_0 d_0,$$

avec ρ_0 optimal, c'est à dire $\nabla f(x_1)^\top d_0 = 0$. Puisque $d_0 \in E_1$ la propriété est donc vérifiée pour $k = 1$. Supposons maintenant que la propriété est vérifiée à l'ordre k :

$$\nabla f(x_k)^\top d_i = 0, \ \ \forall\, i = 0, \ldots, k-1.$$

D'une part ρ_k est optimal donc $\nabla f(x_{k+1})^\top d_k = 0$. D'autre part on a pour $0 \leq i < k$

$$\begin{aligned} \nabla f(x_{k+1})^\top d_i &= (A(x_k + \rho_k d_k) - b)^\top d_i, \\ &= (Ax_k - b)^\top d_i + \rho_k d_k^\top A d_i \\ &= 0, \end{aligned}$$

car ρ_k est optimal et $d_k^\top A d_i = 0$ (conjugaison). On a donc

$$\nabla f(x_{k+1})^\top d_i, \ \ \forall\, i = 0, \ldots, k,$$

ce qui démontre le lemme fondamental.□

Un corollaire direct est donc que la méthode de directions conjuguées converge en n itérations au plus, puisque $E_{n-1} = \mathbb{R}^n$.

III.2 La méthode du gradient conjugué

III.2.1 Algorithme de la méthode du gradient conjugué

L'idée de la méthode est de construire itérativement des directions $d_0, \ldots, d_k$ muutellement conjuguées. A chaque étape k la direction d_k est obtenue comme combinaison linéaire du gradient en x_k et de la direction précédente d_{k-1}, les coefficients étant choisis de telle manière que d_k soit conjuguée avec toutes les directions précédentes. Si l'on note $g_k = \nabla f(x_k)$, l'algorithme prend la forme suivante

On se donne x_0 et on pose $d_0 = -g_0$.

$$\begin{aligned} x_{k+1} &= x_k + \rho_k d_k, \text{ avec} && \text{(III.2.1)} \\ \rho_k &= -\frac{g_k^\top d_k}{d_k{}^\top A d_k}, && \text{(III.2.2)} \\ d_{k+1} &= -g_{k+1} + \beta_k d_k, \text{ avec} && \text{(III.2.3)} \\ \beta_k &= \frac{g_{k+1}^\top A d_k}{d_k{}^\top A d_k}. && \text{(III.2.4)} \end{aligned}$$

Notons d'une part que la formule (III.2.2) définit bien le pas optimal : en effet on a bien

$$\nabla f(x_{k+1})^\top d_k = g_k^\top d_k + \rho_k d_k^\top A d_k = 0.$$

On va maintenant montrer que l'algorithme ci-dessus définit bien une méthode de directions conjuguées.

Théorème III.2.1. *A une itération k quelconque de l'algorithme où l'optimum n'est pas encore atteint, c'est à dire $g_k \neq 0$, on a :*

$$\begin{aligned} \rho_k &= \frac{g_k^\top g_k}{d_k{}^\top A d_k}, && \text{(III.2.5)} \\ \beta_k &= \frac{g_{k+1}^\top (g_{k+1} - g_k)}{g_k^\top g_k} && \text{(III.2.6)} \\ , &= \frac{g_{k+1}^\top g_{k+1}}{g_k^\top g_k}, && \text{(III.2.7)} \end{aligned}$$

et les directions $d_0, \ldots, d_{k+1}$ sont mutuellement conjuguées.

Démonstration : *On raisonne par récurrence sur k en supposant que $d_0, \ldots, d_k$ sont mutuellement conjuguées.*

- Montrons d'abord l'équivalence de III.2.2 et III.2.5. Comme $d_0, \ldots, d_k$ sont mutuellement conjuguées x_k réalise le minimum de f sur $x_0 + E_k$, on a $g_k^\top d_{k-1} = 0$ d'où

$$g_k^\top d_k = g_k^\top (-g_k + \beta_k d_{k-1}) = -g_k^\top g_k.$$

- Pour montrer (III.2.6) on note que

$$g_{k+1} - g_k = A(x_{k+1} - x_k) = \rho_k A d_k, \qquad \text{(III.2.8)}$$

on a alors

$$g_{k+1}^\top A d_k = \frac{1}{\rho_k} g_{k+1}^\top (g_{k+1} - g_k),$$

et en utilisant (III.2.5) il vient bien

$$\beta_k = \frac{g_{k+1}^\top (g_{k+1} - g_k)}{g_k^\top g_k},$$

ce qui démontre (III.2.6). On a de plus $g_{k+1}^\top g_k = 0$ *car* $g_k = d_k - \beta_{k-1} d_{k-1}$ *appartient à* E_{k+1} *et que* g_{k+1} *est orthogonal à ce sous-espace (les directions* $d_0, \ldots, d_k$ *sont conjuguées, par hypothèse de récurrence), ceci démontre (III.2.7).*

- Montrons maintenant que $d_{k+1}^\top A d_i = 0$*, pour* $i = 0, \ldots, k$*. On a d'une part*

$$d_{k+1}^\top A d_k = (-g_{k+1} + \beta_k d_k)^\top A d_k = 0,$$

par définition de β_k*. D'autre part, on a pour* $i < k$

$$d_{k+1}^\top A d_i = -g_{k+1}^\top A d_i + \beta_k d_k \top A d_i,$$

avec $d_k \top A d_i = 0$ *en vertu de l'hypothèse de récurrence. On a ensuite, en utilisant la formule (III.2.8]*

$$g_{k+1}^\top A d_i = \frac{1}{\rho_i} g_{k+1}^\top (g_{i+1} - g_i),$$

et si l'on note que

$$g_{i+1} - g_i = -d_{i+1} + (\beta_i + 1) d_i - \beta_{i-1} d_{i-1},$$

on a bien

$$g_{k+1}^\top (g_{i+1} - g_i) = 0,$$

car $g_{k+1}^\top d_{i+1} = g_{k+1}^\top d_i = g_{k+1}^\top d_{i-1} = 0$*, en vertu du fait que* g_{k+1} *est orthogonal à* E_{k+1} *et que* $i < k$*. On a donc bien* $d_{k+1}^\top A d_i = 0$*, ce qui achève la démonstration.* □

III.2.2 La méthode du gradient conjugué dans le cas général

La méthode de Fletcher et Reeves est une extension directe de la méthode du Gradient conjugué pour les fonction quelconques. Appliquée à une fonction quadratique, elle se comporte comme cette dernière :

On se donne x_0 et on pose $d_0 = -\nabla f(x_0)$.

$$x_{k+1} = x_k + \rho_k d_k, \text{ avec } \rho_k \text{ optimal} \tag{III.2.9}$$

$$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k, \text{ avec} \tag{III.2.10}$$

$$\beta_k = \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2}. \tag{III.2.11}$$

Cette méthode est intéressante car elle ne nécéssite pas de stocker une matrice (contrairement aux méthodes qui seront vues dans les chapitres suivants). Sa vitesse de convergence est très supérieure à celle de la méthode du gradient (ce point sera clarifié pour le cas quadratique dans le grain suivant).

La variante dite de Polak-Ribière consiste à définir β_k par la formule (III.2.6). On peut démontrer la convergence de la méthode de Fletcher-Reeves pour une classe assez large de fonctions f, ce qu'on ne peut pas faire pour la variante de Polak-Ribière. Par contre on peut montrer que cette dernière converge plus rapidement (quand elle converge effectivement !), c'est donc la méthode qui est utilisée en général.

L'efficacité de la méthode du gradient conjugué repose essentiellement sur deux points :
- La recherche linéaire (détermination du pas optimal) doit être exacte,
- Les relations de conjugaison doivent être précises.

La recherche du pas optimal doit être réalisée à l'aide d'un algorithme spécifique (c'est l'objet du prochain chapitre) puisque f est quelconque. Par contre la notion de conjugaison n'a pas de sens dans le cas non-quadratique (sauf près de l'optimum, mais on ne le connaît pas. Il faut donc tester au cours des itérations si l'hypothèse d'approximation quadratique est vérifiée. On peut surveiller les indicateurs suivants
- $|\nabla f(x_{k+1})^\top \nabla f(x_k)|$ doit être petit
- On doit avoir
$$\frac{\nabla f(x_{k+1})^\top d_{k+1}}{\|\nabla f(x_{k+1})\| \|d_{k+1}\|} \leq -\alpha,$$
avec $0 < \alpha \leq 0$ pas trop petit, c'est à dire que d_{k+1} doit être une direction de descente «raisonnable».

Dans le cas où ces conditions ne sont pas vérifiées, on rompt la conjugaison et on redémarre l'algorithme avec $d_{k+1} = -\nabla f(x_{k+1})$. On peut aussi décider de faire ce redémarrage arbitrairement toutes les p itérations (p fixé de l'ordre de n par exemple).

III.3 Interprétation de la méthode du gradient conjugué

III.3.1 Interprétation de la méthode du gradient conjugué

Définition III.3.1. *On appelle $k$ième sous-espace de Krylov associé à la matrice A et au vecteur g_0 le sous espace*

$$\mathcal{K}_k = \text{Vect}(g_0, Ag_0, \ldots, A^{k-1}g_0).$$

Par construction, dans la méthode du gradient conjugué appliqué au cas quadratique, on a $E_k = \mathcal{K}_k$, comme le montre le résultat suivant :

Proposition III.3.1. *Dans la méthode du gradient conjugué on a*

$$E_k = \text{Vect}(d_0, d_1, \ldots, d_{k-1}) = \text{Vect}(g_0, Ag_0, \ldots, A^{k-1}g_0).$$

Démonstration : *Cette propriété est vérifiée à l'ordre $k = 1$ puisque $d_0 = -g_0$. Supposons qu'elle soit vérifiée à l'ordre k. On a alors la formule (III.2.6) qui nous permet d'écrire*

$$\begin{aligned} d_{k+1} &= A(x_k + \rho_k d_k) - b + \beta_k d_k, \\ &= g_k + \rho_k A d_k + \beta_k d_k, \\ &= d_k - \beta_{k-1} d_{k-1} + \rho_k A d_k + \beta_k d_k, \end{aligned}$$

ce qui permet de conclure que $d_{k+1} \in \mathcal{K}_{k+1}$. La propriété est donc vérifiée pour tout $k > 0$. □

Comme dans le cas de l'algorithme du gradient à pas optimal, nous choisissons maintenant de mesurer la distance séparant x_k du vecteur $\hat{x} = A^{-1}b$ à l'aide de la fonction définie par

$$E(x) = \|x - \hat{x}\|_A^2 = (x - \hat{x})^\top A(x - \hat{x}).$$

Minimiser $E(x)$ est équivalent à minimiser $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ comme le montre la proposition suivante (à démontrer en exercice)

Proposition III.3.2. *Soit $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ une forme quadratique définie positive et $\hat{x} = A^{-1}b$. On a*

$$E(x) = (x - \hat{x})^\top A(x - \hat{x}) = f(x) + c,$$

où c est une constante.

On va maintenant illustrer d'un autre point de vue la convergence particulière de l'algorithme du gradient conjugué. Tout vecteur $x_\in x_0 + E_k$ s'écrit

$$x = x_0 + \sum_{j=0}^{k-1} \gamma_j A^j g_0,$$

et comme $g_0 = Ax_0 - b = A(x_0 - \hat{x})$ on a donc

$$x - \hat{x} = x_0 - \hat{x} + \sum_{j=0}^{k-1} \gamma_j A^{j+1}(x_0 - \hat{x}) = p(A)(x_0 - \hat{x}),$$

où le polynôme

$$p(z) = 1 + \sum_{j=0}^{k-1} \gamma_j z^{j+1}$$

est de degré k et satisfait $p(0) = 1$. Puisque le vecteur x_k obtenu à l'étape k de l'algorithme du gradient conjugué vérifie

$$f(x_k) \leq f(x), \ \ \forall\, x \in E_k + x_0,$$

on a, en vertu du résultat démontré dans la proposition précédente,

$$E(x_k) = \|x_k - \hat{x}\|_A^2 \leq \|p(A)(x_0 - \hat{x})\|_A^2 \,,$$

pour tout polynome $p \in \mathcal{P}_k$ vérifiant $p(0) = 1$.

III.3.2 Convergence de la méthode du gradient conjugué

Le résultat suivant va nous permettre de retrouver d'une autre manière la propriété de convergence finie de l'algorithme du GC :

Proposition III.3.3. *Soit A une matrice définie positive et x_k le vecteur obtenu à l'étape k de l'algorithme du GC. Alors on a*

$$E(x_k) \leq E(x_0) \min_{p \in P_k, p(0)=1} \max_{z \in \sigma(A)} p(z)^2.$$

Démonstration : *Puisque la matrice A est définie positive il existe une matrice orthogonale U telle que $A = UDU^\top$ avec $D =$diag$(\lambda_1, \ldots, \lambda_n)$, où $\sigma(A) = \{\lambda_i\}_{i=1\ldots n}$ sont les valeurs propres de A. Si on définit $A^{1/2} = UD^{1/2}U^\top$ on a*

$$\|x\|_A^2 = \left\| A^{1/2} x \right\|^2,$$

donc

$$\|p(A)(x_0 - \hat{x})\|_A^2 = \left\| A^{1/2} p(A)(x_0 - \hat{x}) \right\|^2 \leq \|p(A)\|^2 \|x_0 - \hat{x}\|_A^2,$$

où on a utilisé la propriété que $p(A)$ et $A^{1/2}$ commutent (ces deux matrices ont les mêmes vecteurs propres). Puisque l'on a aussi $A^j = UD^jU^\top$ les valeurs propres de $p(A)$ sont données par les nombres $p(\lambda_i)$ pour $i = 1 \ldots n$, et donc

$$\|p(A)\|^2 = \max_{i=1\ldots n} p(\lambda_i)^2.$$

On a donc bien

$$E(x_k) \leq E(x_0) \min_{p \in P_k, p(0)=1} \max_{z \in \sigma(A)} p(z)^2.$$

□

On a le corollaire suivant, qui permet d'exhiber le polynôme optimal $p(z)$ pour $k = n$:

Théorème III.3.2. *Soit A une matrice définie positive. L'algorithme du GC converge en n itérations au plus. Plus précisément, si la matrice A possède $k \leq n$ valeurs propres distinctes, alors L'algorithme du GC converge en k itérations au plus.*

Démonstration : *Dans les deux cas possibles, notons*

$$\bar{p}(z) = \Pi_{i=1}^k \frac{\lambda_i - z}{\lambda_i}.$$

On a bien $\bar{p}(z)$ de degré k, $\bar{p}(0) = 1$ et par construction $\bar{p}(\lambda_i) = 0$ pour $i = 1 \ldots k$. En vertu du résultat montré dans la proposition III.3.3, on a donc

$$E(x_k) = 0,$$

soit $x_k = \hat{x}$. □

La méthode du gradient conjugué étant en général utilisée comme une méthode itérative, il est intéressant de la comparer à la méthode du gradient à pas optimal. Le résultat suivant sera admis (la démonstration repose sur la détermination d'un polynôme particulier $p(z)$ solution d'un problème de moindre carrés).

Théorème III.3.3. *Soit A une matrice définie positive et x_k le vecteur obtenu à l'étape k de l'algorithme du GC. Alors on a*

$$E(x_k) \leq 4E(x_0) \left(\frac{\sqrt{\chi(A)} - 1}{\sqrt{\chi(A)} + 1} \right)^{2k},$$

où on a noté $\chi(A) = \lambda_n/\lambda_1$ le conditionnement de A pour la norme euclidienne.

Pour l'algorithme du gradient à pas optimal on avait

$$E(x_k) \leq E(x_0) \left(\frac{\chi(A) - 1}{\chi(A) + 1} \right)^{2k},$$

on voit donc que pour une même matrice A, la méthode du gradient conjugué convergera plus rapidement. Cependant cette estimation peut être très pessimiste car dans le cas où les valeurs propres sont groupées autour de valeurs distinctes, on peut être très proche du cas ou certaines valeurs propres sont multiples (et ou le nombre théorique d'itérations est inférieur à n) tout en ayant un mauvais conditionnement.

Chapitre IV
Méthodes de recherche linéaire

IV.1 introduction

IV.1.1 But de la recherche linéaire

On a vu que dans le cas non-quadratique les méthodes de descente :

$$x_{k+1} = x_k + t_k d_k,\ t_k > 0,$$

nécéssitent la recherche d'une valeur de $t_k > 0$, optimale ou non, vérfiant

$$f(x_k + t_k d_k) \leq f(x_k).$$

On définit comme précedemment la fonction $\varphi(t) = f(x_k + t d_k)$. Rappellons que si f est différentiable, le pas optimal $\hat{t}$ peut être caractérisé par

$$\begin{cases} \varphi'(\hat{t}) &= 0, \\ \varphi(\hat{t}) &\leq \varphi(t), \text{ pour } 0 \leq t \leq \hat{t}, \end{cases}$$

autrement dit, $\hat{t}$ est un minimum local de φ qui assure de plus la décroissance de f. En fait, dans la plupart des algorithmes d'optimisation modernes, on ne fait jamais de recherche linéaire exacte, car trouver $\hat{t}$ signifie qu'il va falloir calculer un grand nombre de fois la fonction φ, et cela peut être dissuasif du point de vue du temps de calcul. En pratique, on recherche plutot une valeur de t qui assure une décroissance suffisante de f. Cela conduit à la notion d'intervalle de sécurité.

IV.1.2 Intervalle de sécurité

Définition IV.1.1. *On dit que $[a, b]$ est un intervalle de sécurité s'il permet de classer les valeurs de t de la façon suivante :*

- *Si $t < a$ alors t est considéré trop petit,*
- *Si $b \geq t \geq a$ alors t est satisfaisant,*
- *Si $t > b$ alors t est considéré trop grand.*

Le problème est de traduire de façon numérique sur φ les trois conditions précédentes, ainsi que de trouver un algorithme permettant de déterminer a et b. L'idée est de partir d'un intervalle suffisament grand pour contenir $[a, b]$, et d'appliquer un bonne stratégie pour itérativement réduire cet intervalle.

Algorithme de base

Initialement, on part de $[\alpha, \beta]$ contenant $I = [a, b]$, par exemple en prenant $\alpha = 0$ et β tel que $\varphi(\beta) > \varphi(0)$ (une telle valeur de β existe avec un minimum d'hypothèses, par exemple f coercive). On fait ensuite les itérations suivantes :

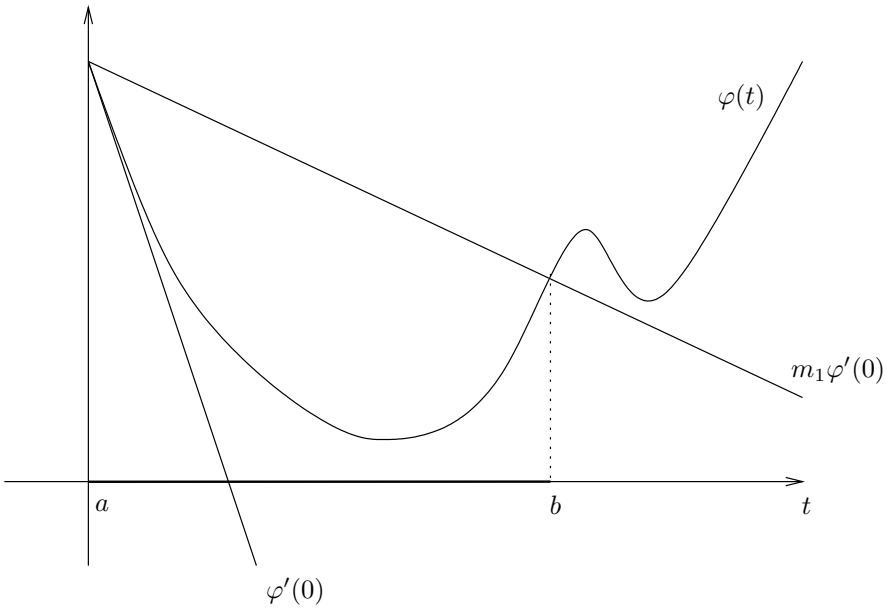
1. On choisit t dans l'intervalle $[\alpha, \beta]$.
2. Si t est trop petit on prend $\alpha = t$ et on retourne en 1.
3. Si t est trop grand on prend $\beta = t$ et on retourne en 1.
4. Si t convient on s'arrète.

Il faut maintenant préciser quelles sont les relations sur φ qui vont nous permettre de caractériser les valeurs de t convenables, ainsi que les techniques utilisées pour réduire l'intervalle (point nř1 ci-dessus).

IV.2 Caractérisation de l'intervalle de sécurité

IV.2.1 La règle d'Armijo

Dans la règle d'Armijo on prend $\alpha = 0$, un réel $0 < m < 1$. La règle est la suivante :



Règle d'Armijo

- Si $\varphi(t) \leq \varphi(0) + m\varphi'(0)t$, alors t convient.
- Si $\varphi(t) > \varphi(0) + m\varphi'(0)t$, alors t est trop grand.

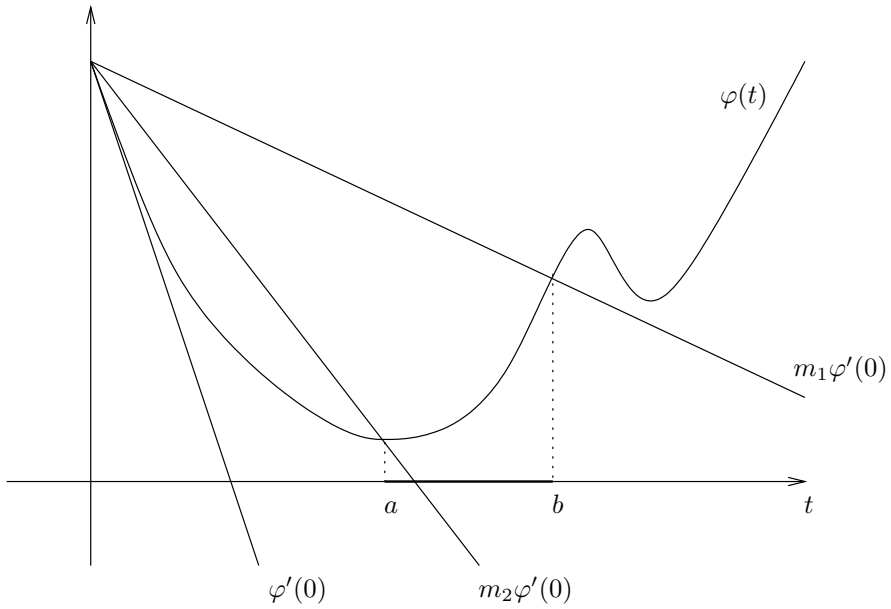
On peut noter que l'on a

$$\begin{aligned} \varphi(0) &= f(x_k), \\ \varphi'(0) &= \nabla f(x_k)^\top d_k. \end{aligned}$$

Puisque $\alpha = 0$, t n'est jamais considéré trop petit, c'est pourquoi la règle d'Armijo est peu utilisée *seule*.

IV.2.2 La règle de Goldstein

En ajoutant une deuxième inégalité à la règle d'Armijo on obtient la règle de Goldstein, où m_1 et m_2 sont deux constantes vérifiant $0 < m_1 < m_2$:



Règle de Goldstein

- Si $\varphi(t) < \varphi(0) + m_2\varphi'(0)t$, alors t est trop petit.
- Si $\varphi(t) > \varphi(0) + m_1\varphi'(0)t$, alors t est trop grand.
- si $\varphi(0) + m_1\varphi'(0)t \geq \varphi(t) \geq \varphi(0) + m_2\varphi'(0)t$, alors t convient

Le choix de m_2 doit être tel que dans le cas quadratique, le pas optimal appartienne à l'intervalle de sécurité (c'est bien la moindre des choses). Dans le cas quadratique on a

$$\varphi(t) = \frac{1}{2}at^2 + \varphi'(0)t + \varphi(0), \;\; a > 0,$$

et le pas optimal $\hat{t}$ vérifie $\varphi'(\hat{t}) = 0$, soit $\hat{t} = -\varphi'(0)/a$. On a donc (exercice)

$$\varphi(\hat{t}) \;\; = \;\; \varphi(0) + \frac{\varphi'(0)}{2}\hat{t}.$$

Donc $\hat{t}$ sera considéré comme satisfaisant si $m_2 \geq \frac{1}{2}$. Des valeurs typiques utilisées dans la pratique sont $m_1 = 0.1$ et $m_2 = 0.7$

Théorème IV.2.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ coercive, c'est à dire f continue et*

$$lim_{\|x\|\to\infty} f(x) = +\infty.$$

Soit l'algorithme de gradient

$$x_{k+1} = u_k - \rho_k g_k,$$

où $g_k = \nabla f(x_k)$ où à chaque itération le pas ρ_k satisfait à la règle de Goldstein

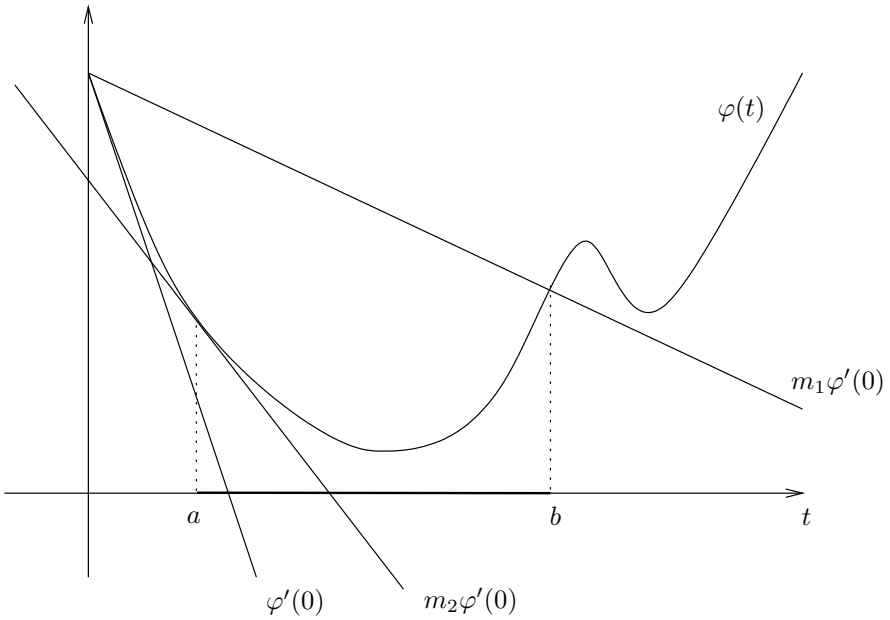
$$\varphi(0) + m_2\varphi'(0)\rho_k \leq \varphi(\rho_k) \leq \varphi(0) + m_1\varphi'(0)\rho_k,$$

où $\varphi(\rho) = f(x_k - \rho g_k)$ et $0 < m_1 < m_2 < 1$. Alors la suite x_k est bornée, la suite $f(x_k)$ est décroissante et convergente, et le vecteur g_k vérifie

$$\lim_{k\to\infty} \|g_k\| = 0.$$

IV.2.3 La règle de Wolfe

La règle de Wolfe fait appel au calcul de $\varphi'(t)$, elle est donc en théorie plus coûteuse que la règle de Goldstein. Cependant dans de nombreuses applications, le calcul du gradient $\nabla f(x)$ représente un faible coût additionnel en comparaison du coût d'évaluation de $f(x)$ (par exemple en contrôle optimal), c'est pourquoi cette règle est très utilisée. Le calcul des dérivées de φ permet de plus d'utiliser une méthode d'interpolation cubique dans la phase de réduction de l'intervalle, comme nous le verrons plus loin.



Règle de Wolfe

- Si $\varphi(t) > \varphi(0) + m_1\varphi'(0)t$, alors t est trop grand.
- Si $\varphi(t) \leq \varphi(0) + m_1\varphi'(0)t$ et $\varphi'(t) < m_2\varphi'(0)$, alors t est trop petit.
- Si $\varphi(t) \leq \varphi(0) + m_1\varphi'(0)t$ et $\varphi'(t) \geq m_2\varphi'(0)$, alors t convient.

Dans cette règle, on s'assure que t n'est pas trop petit en assurant que $\varphi'(t)$ a suffisamment augmenté.

IV.2.4 Réduction de l'intervalle

Le premier problème à résoudre est celui de la détermination d'un intervalle de départ $[\alpha, \beta]$. On peut commencer par choisir $\alpha = 0$, et utiliser une valeur initiale de t censée être une bonne valeur de départ (ce point sera clarifié plus loin).

Recherche d'un intervalle de départ

1. Si t est satisfaisant alors on s'arrête
2. Si t est trop grand, alors on prend $\beta = t$ et on s'arrête
3. Si t est trop petit, on fait $t \leftarrow ct$, $c > 1$, et on retourne en 1.

Cet algorithme donne un intervalle initial $[\alpha, \beta]$ qu'il va falloir ensuite réduire, sauf si t est admissible, auquel cas la recherche linéaire est terminée, ce peut être le cas si la valeur initiale de t est bien choisie.

Réduction de l'intervalle

On suppose maintenant que l'on dispose d'un intervalle $[\alpha, \beta]$ mais que l'on n'a pas encore de t satisfaisant. Une manière simple de faire est de procéder par exemple par dichotomie, en choisissant

$$t = \frac{\alpha + \beta}{2},$$

puis en conservant soit $[\alpha, t]$ ou $[t, \beta]$ suivant que t est trop grand ou trop petit. Le problème est que cette stratégie ne réduit pas assez rapidement l'intervalle. Cependant elle n'utilise aucune informations sur φ (dérivées ou autres). On préfère en général procéder en construisant une approximation polynomiale $p(t)$ de φ et en choisissant t réalisant le minimum (s'il existe) de $p(t)$ sur $[\alpha, \beta]$. Lorsque l'on utilise la règle de Wolfe, on peut utiliser une approximation cubique.

IV.2.5 Réduction de l'intervalle par interpolation cubique

Comme nous l'avons évoqué, un choix judicieux de t peut être fait en faisant une approximation cubique de $\varphi(t)$ sur l'intervalle $[\alpha, \beta]$ et à prendre t réalisant le minimum de cette cubique : on considère le polynôme $p(t)$ vérifiant

$$\begin{aligned} p(t_0) &= \varphi(t_0) = f_0, \\ p(t_1) &= \varphi(t_1) = f_1, \\ p'(t_0) &= \varphi'(t_0) = g_0, \\ p'(t_1) &= \varphi'(t_1) = g_1 \end{aligned}$$

où t_0 et t_1 sont quelconques (on peut bien sûr prendre $t_0 = \alpha$ et $t_1 = \beta$). On passe en variables réduites sur $[0, 1]$ ce qui conduit à définir le polynôme $q(s)$ par

$$q(s) = p(t_0 + st_1), \;\; s \in [0, 1], \;\; \tau = t_1 - t_0,$$

qui vérifie donc

$$\begin{aligned} q(0) &= f_0, \\ q(1) &= f_1, \\ q'(0) &= \tau g_0, \\ q'(1) &= \tau g_1. \end{aligned}$$

Si on cherche q de la forme

$$q(s) = as^3 + bs^2 + cs + d,$$

alors les calculs donnent

$$a = \tau(g_0 + g_1) + 2(f_0 - f_1), \; b = 3(f_1 - f_0) - \tau(2g_0 + g_1), \; c = \tau g_0, d = f_0.$$

- Si $b^2 - 3ac < 0$ alors $q(s)$ n'admet pas de minimum, et cela ne permet pas de choisir α.
- Si $b^2 - 3ac \geq 0$ il y a un minimum donné par

  $$\hat{s} = \frac{-b + \sqrt{b^2 - 3ac}}{3a},$$

  si $\hat{s} \in [0, 1]$ cela permet de donner à t la valeur

  $$t = t_0 + \hat{s}\tau,$$

  sinon, cela ne permet pas de choisir t, et on peut en dernier recours faire appel à la dichotomie.

Chapitre V
Méthodes de Quasi-Newton

V.1 Introduction

V.1.1 La méthode de Newton

La méthode de Newton permet de construire un algorithme permettant de résoudre le système d'équations non-linéaires

$$g(x) = 0,$$

où $g : \mathbb{R}^n \to \mathbb{R}^n$ est diférentiable : on se donne $x_0 \in \mathbb{R}^n$ et on fait les itérations

$$x_{k+1} = x_k - g'(x_k)^{-1} g(x_k), \tag{V.1.1}$$

où $g'(x)$ est la dérivée (ou jacobienne) de g au point x. L'application de cette méthode au problème d'optimisation

$$\min_{x \in \mathbb{R}^n} f(x), \tag{V.1.2}$$

consiste à l'utiliser pour résoudre le système d'optimalité du problème (V.1.2), c'est à dire que l'on pose $g(x) = \nabla f(x)$ dans (V.1.1) : on obtient les itérations

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k). \tag{V.1.3}$$

La méthode de Newton est intéressante car sa convergence est quadratique au voisinage de la solution, c'est à dire que l'on a

$$\|x_{k+1} - \hat{x}\| \leq \gamma \|x_k - \hat{x}\|^2, \ \ \gamma > 0,$$

mais la convergence n'est assurée que si x_0 est suffisamment proche de $\hat{x}$, ce qui en limite l'intérêt.

Pour résoudre le problème de convergence locale de la méthode de Newton, on peut penser à lui ajouter une phase de recherche linéaire, dans la direction

$$d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k).$$

Cela est possible uniquement si d_k est une direction de descente en x_k, soit

$$\nabla f(x_k)^\top d_k = -\nabla f(x_k)^\top \nabla^2 f(x_k)^{-1} \nabla f(x_k) < 0,$$

ce qui sera le cas si $\nabla^2 f(x_k)$ est une matrice définie positive, ce qui n'est pas garanti (on sait tout au plus que $\nabla^2 f(\hat{x}) > 0$).

Le principe des méthodes que nous allons voir maintenant consiste à remplacer le Hessien $\nabla^2 f(x_k)$ par une approximation H_k (si possible définie positive), construite au cours des itérations.

V.1.2 Méthodes à métrique variable

Le principe des méthodes dites «à métrique variable» consiste à faire les itérations suivantes

$$\left\{ \begin{array}{rcl} d_k & = & -B_k g_k, \\ x_{k+1} & = & x_k + \rho_k d_k, \end{array} \right. \tag{V.1.4}$$

où on a noté $g_k = \nabla f(x_k)$ et B_k est une matrice définie positive. La méthode ci-dessus coïncide avec la méthode du gradient si $B_k = I$. On peut envisager de prendre $B_k = B > 0,\ \forall k$ et cela conduit à la remarque suivante.

Remarque V.1.1. *Lorsque l'on cherche à résoudre le problème*

$$\min_{x \in \mathbb{R}^n} f(x),$$

On peut poser $x = Cy$ où C est une matrice inversible (changement de variable). Notons alors $\tilde{f}(y) = f(Cy)$. On a

$$\nabla \tilde{f}(y) = C^\top \nabla f(Cy).$$

Un pas de la méthode du gradient appliquée à la minimisation de $\tilde{f}(y)$ est donné par

$$y_{k+1} = y_k - \rho_k C^\top \nabla f(Cy_k),$$

soit en revenant à la variable originale et en posant $x_k = Cy_k$

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

On obtient bien une méthode du type (V.1.4) avec $B = CC^\top > 0$. Dans le cas où f est une forme quadratique, on voit assez facilement comment l'introduction de B permet d'accélérer la convergence de la méthode.

Théorème V.1.2. *Soit $f(x) =$ une forme quadratique définie positive et B une matrice définie positive. L'algorithme du gradient préconditionné*

$$\left\{ \begin{array}{rcl} x_0 & = & \text{donné}, \\ x_{k+1} & = & x_k - \rho_k B g_k, \ \ \rho_k \text{ optimal} \end{array} \right.$$

converge linéairement au sens où

$$\|x_{k+1} - \hat{x}\|_A \leq \gamma \|x_k - \hat{x}\|_A ,$$

avec

$$\gamma = \frac{\chi(BA) - 1}{\chi(BA) + 1}.$$

Dans cette méthode, on voit bien comment influe la matrice B sur la vitesse de convergence : plus le conditionnement de BA sera faible, plus l'accélération sera grande. On ne peut bien sûr pas poser $B = A^{-1}$, puisque cela sous-entendrait que l'on a déjà résolu le problème ! Cependant, l'idée est tout de même assez bonne, en ce sens qu'elle indique que B soit être une approximation de A^{-1} si l'on veut effectivement accélérer la méthode. Enfin, et pour terminer cette introduction avant d'étudier de plus près les méthodes de quasi-Newton pour f quelconque, on peut d'ores et déjà dire qu'un critère de bon fonctionnement de la méthode (V.1.4) serait que l'on ait au moins

$$\lim_{k \to \infty} B_k = A^{-1},$$

dans le cas quadratique.

V.2 Les méthodes de quasi-Newton

V.2.1 Relation de quasi-Newton

Une méthode de quasi-Newton est une méthode du type :

$$\left\{ \begin{array}{rcl} d_k & = & -B_k g_k, \\ x_{k+1} & = & x_k + \rho_k d_k, \end{array} \right. \tag{V.2.1}$$

ou

$$\left\{ \begin{array}{rcl} d_k & = & -H_k^{-1} g_k, \\ x_{k+1} & = & x_k + \rho_k d_k, \end{array} \right. \tag{V.2.2}$$

où B_k (respectivement H_k) est une matrice destinée à approcher l'inverse du hessien de f (respectivement le hessien de f) en x_k. Il se pose donc un problème : quelle stratégie adopter pour faire cette approximation. On peut par exemple poser $B_0 = I$, mais comment ensuite mettre à jour l'approximation B_k au cours des itérations ? L'idée est la suivante : on sait que au point x_k, le gradient et le hessien de f vérifient la relation

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

Si on suppose que l'approximation quadratique est bonne, on peut alors négliger le reste et considérer que l'on a

$$g_{k+1} - g_k \approx \nabla^2 f(x_k)(x_{k+1} - x_k),$$

cela conduit à la notion de relation de quasi-Newton :

Définition V.2.1. *On dit que les matrice B_{k+1} et H_{k+1} vérifient une relation de quasi-Newton si on a*

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

ou

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

Il reste un problème à résoudre : comment mettre à jour B_k tout en assurant $B_k > 0$? C'est ce que nous allons voir maintenant.

V.2.2 Formules de mise à jour de l'approximation du hessien

Le principe de la mise à jour consiste, à une itération donnée de l'algorithme

$$\left\{ \begin{array}{rcl} d_k & = & -B_k g_k, \\ x_{k+1} & = & x_k + \rho_k d_k, \end{array} \right. \tag{V.2.3}$$

à appliquer une formule du type

$$B_{k+1} = B_k + \Delta_k, \tag{V.2.4}$$

avec Δ_k symétrique, assurant la relation de quasi-Newton

$$x_{k+1} - x_k = B_{k+1}(g_{k+1} - g_k),$$

ainsi que $B_{k+1} > 0$, sous l'hypothèse que $B_k > 0$.

La formule (V.2.4) permet d'utiliser les nouvelles informations obtenues lors de l'étape k de l'algorithme, c'est à dire essentiellement le gradient $g_{k+1} = \nabla f(x_{k+1})$ au point x_{k+1}, obtenu par recherche linéaire (exacte ou approchée) dans la direction d_k. Il existe différentes formules du type (V.2.4). Suivant que Δ_k est de rang 1 ou 2, on parlera de correction de rang 1 ou de rang 2.

V.2.3 Formule de Broyden

On peut chercher à déterminer une formule de correction de rang 1 de la façon suivante. On écrit B_{k+1} sous la forme

$$B_{k+1} = B_k + vv^\top,$$

et on cherche v tel que la relation de quasi-Newton

$$B_{k+1}y_k = s_k,$$

où on a posé $y_k = g_{k+1} - g_k$ et $s_k = x_{k+1} - x_k$. On a donc

$$B_k y_k + vv^\top y_k = s_k,$$

et en prenant le produit scalaire des deux membres de l'égalité précédente avec y_k on obtient

$$(y_k^\top v)^2 = (s_k - B_k y_k)^\top y_k$$

Si on utilise maintenant l'égalité

$$vv^\top = \frac{vv^\top y_k (vv^\top y_k)^\top}{(v^\top y_k)^2},$$

alors on peut écrire, en remplacant $v^\top y_k$ par $s_k - B_k y_k$ et $(v^\top y_k)^2$ par $y_k^\top (s_k - B_k y_k)$, la formule de correction

$$B_{k+1} = B_k + \frac{(s_k - B_k y_k)(s_k - B_k y_k)^\top}{(s_k - B_k y_k)^\top y_k}, \tag{V.2.5}$$

connue sous le nom de *formule de Broyden*. La validité de cette formule provient du résultat suivant :

Théorème V.2.2. *Soit f une forme quadratique définie positive. Considérons la méthode itérative qui, partant d'un point x_0 arbitraire engendre sucessivement les points*

$$x_{k+1} = x_k + s_k,$$

où les s_k sont des vecteurs linéairement indépendants. Alors la suite de matrices générée par B_0, une matrice symétrique quelconque et la formule

$$B_{k+1} = B_k + \frac{(s_k - B_k y_k)(s_k - B_k y_k)^\top}{(s_k - B_k y_k)^\top y_k},$$

où $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, converge en au plus n étapes vers A^{-1}, l'inverse du hessien de f.

Démonstration : *Puisque le hessien de f est constant et égal à A on a*

$$y_i = \nabla f(x_{i+1}) - f(x_i) = A(x_{i+1} - x_i), \forall i.$$

On a vu que B_{k+1} est construit de façon à ce que l'on ait

$$B_{k+1}y_k = s_k,$$

montrons que l'on a aussi

$$B_{k+1}y_i = s_i,\ i = 0 \dots k-1.$$

On raisonne par récurrence en supposant que cette propriété est vraie pour B_k, à savoir

$$B_k y_i = s_i,\ i = 0 \dots k-2.$$

Soit donc $i \leq k-2$ quelconque. On a

$$B_{k+1}y_i = B_k y_i + \frac{(s_k - B_k y_k)(s_k^\top y_i - B_k y_k^\top y_i)}{(s_k - B_k y_k)^\top y_k}. \tag{V.2.6}$$

Par l'hypothèse de récurrence on a $B_k y_i = s_i$ donc

$$y_k^\top B_k y_i = y_k^\top s_i,$$

mais comme $As_j = y_j, \forall j$, on obtient

$$y_k^\top s_i = s_k^\top A s_i = s_k^\top y_i,$$

donc dans (V.2.6) le numérateur est nul et on a $B_{k+1}y_i = B_k y_i = s_i$. On a donc

$$B_{k+1}y_i = s_i, \;\; i = 0 \dots k.$$

Au bout de n itérations on a donc

$$B_n y_i = s_i, \; i = 0 \dots n-1,$$

et puisque l'on a $y_i = As_i$ cette dernière formule d'écrit

$$B_n A s_i = s_i, \; i = 0 \dots n-1.$$

Comme les s_i constituent une base de $\mathbb{R}^n$ on a $B_n A = I$ ou encore

$$B_n = A^{-1},$$

ce qui montre le résultat. □

Le problème de la formule de Broyden est qu'il n'y a aucune garantie que les matrices B_k soientt définies positives même si la fonction f est quadratique et si par exemple $B_0 = I$. On peut cependant noter l'intérêt de la propriété $B_n = A^{-1}$, qui sera aussi vérifiée par les méthodes de mise à jour que nous allons voir maintenant.

V.2.4 Formule de Davidon, Fletcher et Powell

La formule de mise à jour de Davidon, Fletcher et Powell est une formule de correction de rang 2 donnée par

$$B_{k+1} = B_k + \frac{s_k s_k^\top}{s_k^\top y_k} - \frac{B_k y_k y_k^\top B_k}{y_k^\top B_k y_k} \tag{V.2.7}$$

Le résultat suivant montre que sous certaines conditions, la formule (V.2.7) conserve la définie-positivité des matrices B_k.

Théorème V.2.3. *On considère la méthode définie par*

$$\begin{aligned} d_k &= -B_k g_k, \\ x_{k+1} &= x_k + \rho_k d_k, \ \rho_k \text{ optimal} \end{aligned}$$

Où $B_0 > 0$ est donnée ainsi que x_0. Alors les matrices B_k sont définies positives, $\forall k > 0$.

Démonstration : *Soit x un vecteur de $\mathbb{R}^n$. On a*

$$\begin{aligned} x^\top B_{k+1} x &= x^\top B_k x + \frac{(s_k^\top x)^2}{s_k^\top y_k} - \frac{(y_k^\top B_k x)^2}{y_k^\top B_k y_k}, \\ &= \frac{y_k^\top B_k y_k x^\top B_k x - (y_k^\top B_k x)^2}{y_k^\top B_k y_k} + \frac{(s_k^\top x)^2}{s_k^\top y_k} \end{aligned}$$

Si on définit le produit scalaire $\langle x\,,\, y\rangle = x^\top B_k y$ alors on a

$$x^\top B_{k+1} x = \frac{\langle y_k\,,\, y_k\rangle \langle x\,,\, x\rangle - \langle y_k\,,\, x\rangle^2}{\langle y_k\,,\, y_k\rangle} + \frac{(s_k^\top x)^2}{s_k^\top y_k}. \tag{V.2.8}$$

Le premier terme du second membre est positif ou nul d'après l'inégalité de Cauchy-Schwartz. Quant au deuxième terme on peut faire l'analyse suivante : puisque le pas est optimal, on a la relation

$$g_{k+1}^\top d_k = 0,$$

et donc

$$s_k^\top y_k = +\rho_k (g_{k+1} - g_k)^\top d_k = \rho_k g_k^\top B_k g_k > 0,$$

on a donc $x^\top B_{k+1} x \geq 0$. Les deux termes dans (V.2.8) étant positifs, cette quantité ne peut s'annuler que si les deux termes sont simultanément nuls. Le premier terme ne peut s'annuler que si $x = \lambda y_k$ pour un scalaire $\lambda \neq 0$. Dans ce cas le deuxième terme est non nul car $s_k^\top x = \lambda s_k^\top y_k$. On a donc bien $B_{k+1} > 0$. □

Remarque V.2.4. *La propriété $s_k^\top y_k > 0$ est vérifiée également par des méthodes de recherche linéaire approchées comme par exemple la règle de Wolfe de Powell : en effet dans ce cas on détermine un point x_{k+1} tel que*

$$\varphi'(\rho_k) = \nabla f(x_{k+1})^\top d_k \geq m_2 \nabla f(x_k)^\top d_k, \ \ 0 < m_2 < 1,$$

d'où

$$g_{k+1}^\top \frac{x_{k+1} - x_k}{\rho_k} > g_k^\top \frac{x_{k+1} - x_k}{\rho_k},$$

et donc $(g_{k+1} - g_k)^\top (x_{k+1} - x_k) > 0$.

V.2.5 Algorithme de Davidon-Fletcher-Powel

On peut donc formuler maintenant la méthode utilisant la formule de correction (V.2.7) :

Algorithme de Davidon-Fletcher-Powel

1. Choisir x_0 et B_0 définie positive quelconque (par exemple $B_0 = I$)
2. A l'itération k, calculer la direction de déplacement

$$d_k = -B_k \nabla f(x_k),$$

déterminer le pas optimal ρ_k et poser

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

3. Poser $s_k = \rho_k d_k$ et $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ puis calculer

$$B_{k+1} = B_k + \frac{s_k s_k^\top}{s_k^\top y_k} - \frac{B_k y_k y_k^\top B_k}{y_k^\top B_k y_k}.$$

4. Faire $k \leftarrow k+1$. Retourner en 1 sauf si le critère d'arrêt est vérifié.

Comme critère d'arrêt on retiendra par exemple $\|g_{k+1}\| < \epsilon$.

Cet algorithme a un comportement remarquable dans le cas où f est une forme quadratique :

Théorème V.2.5. *Appliqué à une forme quadratique f, l'algorithme DFP engendre des directions $s_0, \ldots, s_k$ vérifiant*

$$s_i^\top A s_j = 0, \quad 0 \le i < j \le k+1, \tag{V.2.9}$$

$$B_{k+1} A s_i = s_i, \quad 0 \le i \le k. \tag{V.2.10}$$

Démonstration : *En utilisant la formule (V.2.7) on a pour tout k*

$$\begin{aligned} B_{k+1} A s_k &= B_{k+1} y_k, \\ &= s_k, \end{aligned}$$

par construction. Donc (V.2.10) est en particulier vérifiée pour $k = 0$, soit

$$B_1 A s_0 = s_0.$$

On a aussi

$$\begin{aligned} s_0^\top A s_1 &= -\rho_1 s_0^\top A B_1 g_1, \\ &= -\rho_1 s_0^\top A B_1 g_1, \\ &= -\rho_1 s_0^\top g_1, \\ &= 0, \end{aligned}$$

puisque $B_1 A s_0 = s_0$ et que x_1 est obtenu par un pas optimal dans la direction s_0. Donc (V.2.10) est vérifiée pour $k = 0$.

Supposons maintenant que (V.2.9) et (V.2.10) sont vérifiées à l'ordre $k-1$. On peut écrire d'une part pour $i = 0 \ldots k-1$.

$$\begin{aligned} g_{k+1} - g_{i+1} &= y_{i+1} + y_i + \ldots y_k, \\ &= A(s_{i+1} + s_i + \ldots s_k) \end{aligned}$$

car f est une forme quadratique de hessien A. D'autre part, puisque x_{i+1} est obtenu par un pas optimal dans la direction s_i on a $s_i^\top g_{i+1} = 0$ et donc

$$s_i^\top (g_{k+1} - g_{i+1}) = s_i^\top A(s_{i+1} + s_i + \dots s_k), \;\; i = 0 \dots k-1,$$

donc en vertu de l'hypothèse de recurrence (conjugaison des s_i) on a

$$s_i^\top g_{k+1} = 0, \;\; i = 0 \dots k-1, \tag{V.2.11}$$

Cette relation reste aussi valable pour $i = k$ puisque l'on a $s_k^\top g_{k+1} = 0$ (pas optimal). La deuxième hypothèse de récurrence permet donc d'écrire, en remplacant s_i par $B_{k+1}As_i$ dans (V.2.11)

$$s_i^\top A B_{k+1} g_{k+1} = 0, \;\; i = 0 \dots k$$

et donc, puisque $H_{k+1}g_{k+1} = -s_{k+1}/\rho_{k+1}$,

$$s_i^\top A s_{k+1} = 0, \;\; i = 0 \dots k,$$

ce qui démontre donc la propriété (V.2.9) au rang k.

Montrons maintenant que

$$B_{k+1}As_i = s_i, \;\; i = 0 \dots k-1.$$

Cette relation est vraie pour $i = k$ comme on l'a déjà montré plus haut. On a

$$B_{k+1}As_i = B_k As_i + \frac{s_k s_k^\top As_i}{s_k^\top y_k} - \frac{B_k y_k y_k^\top B_k As_i}{y_k^\top B_k y_k}.$$

Le deuxième terme du second membre est nul car $s_k^\top As_i = 0$. Si on note que par l'hypothèse de récurrence on a $B_k As_i = s_i$ pour $i = 0 \dots k-1$ et $y_k^\top = s_k^\top A$ le numérateur du troisième terme est donné par

$$B_k y_k y_k^\top B_k As_i = B_k y_k s_k^\top As_i = 0.$$

Par conséquent on a bien

$$B_{k+1}As_i = s_i, \;\; i = 0 \dots k-1,$$

ce qui démontre la propriété (V.2.10) au rang k. □

La méthode DFP se comporte donc, dans le cas quadratique, comme une méthode de directions conjuguées. Dans ce cas l'algorithme converge en au plus n itérations. On peut aussi remarquer que l'on a pour $k = n-1$ la relation

$$B_n As_i = s_i, \;\; i = 0, \dots n-1,$$

et comme les s_i sont linéairement indépendants (car mutuellement conjugués) on en déduit que

$$B_n = A^{-1}.$$

Remarque V.2.6. *On peut montrer que dans le cas général (non quadratique), sous les mêmes réserves que pour la méthode de Fletcher-Reeves (réinitialisation périodique $d_k = -g_k$), cet algorithme permet de converger vers un minimum local $\hat{x}$ de f, et que l'on a*

$$\lim_{k\to\infty} B_k = \nabla^2 f(\hat{x})^{-1},$$

ce qui montre que près de l'optimum $\hat{x}$, si la recherche linéaire est exacte, la méthode se comporte asymptotiquement comme la méthode de Newton. Cette remarque permet de justifier le choix d'une estimation du pas de déplacement donnée par

$$\rho_k = 1,$$

dans les méthodes de recherche linéaire approchée.

V.2.6 Algorithme de Broyden, Fletcher, Goldfarb et Shanno

La formule de mise à jour de Broyden, Fletcher, Goldfarb et Shanno est une formule de correction de rang 2 qui s'obtient à partir de la formule DFP en intervertissant les rôles de s_k et y_k. La formule obtenu permet de mettre à jour une approximation H_k du hessien possédant les mêmes propriétés, à savoir $H_{k+1} > 0$ si $H_k > 0$ et vérifiant la relation de quasi-Newton

$$y_k = H_k s_k.$$

La formule est donc la suivante :

$$H_{k+1} = H_k + \frac{y_k y_k^\top}{y_k^\top s_k} - \frac{H_k s_k s_k^\top H_k}{s_k^\top H_k s_k} \tag{V.2.12}$$

L'algorithme associé est le suivant :

Algorithme de Broyden, Fletcher, Goldfarb et Shanno

1. Choisir x_0 et H_0 définie positive quelconque (par exemple $H_0 = I$)
2. A l'itération k, calculer la direction de déplacement

   $$d_k = -H_k^{-1} \nabla f(x_k),$$

   déterminer le pas optimal ρ_k et poser

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

3. Poser $s_k = \rho_k d_k$ et $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ puis calculer

   $$H_{k+1} = H_k + \frac{y_k y_k^\top}{y_k^\top s_k} - \frac{H_k s_k s_k^\top H_k}{s_k^\top H_k s_k}$$

4. Faire $k \leftarrow k + 1$. Retourner en 2 sauf si le critère d'arrêt est vérifié.

Notons que la direction d_k est obtenue par résolution d'un système linéaire. En pratique la mise à jour de H_k est faite directement sur le facteur de Cholesky C_k où $H_k = C_k C_k^\top$ ce qui ramène le calcul de d_k au même coût que pour la formule de DFP. De plus, cette technique permet de contrôler précisément la définie positivité de H_k, qui peut se dégrader à cause des erreurs d'arrondi.

Remarque V.2.7. *La méthode BFGS possède les mêmes propriétés que la méthode DFP : dans le cas quadratique les directions engendrées sont conjuguées et on a $H_n = A$. Cette méthode est reconnue comme étant beaucoup moins sensible que la méthode DFP aux imprécisions dans la recherche linéaire, du point de vue de la vitesse de convergence. Elle est donc tout à fait adaptée quand la recherche linéaire est faite de façon économique, avec par exemple la règle de Goldstein ou la règle de Wolfe et Powell. Elle est par exemple utilisée dans la fonction* `fminu` *de Matlab.*

V.3 Méthodes spécifiques pour les problèmes de moindres carrés

V.3.1 La méthode de Gauss-Newton

Dans les problèmes de moindres carrés non linéaires, la fonction à minimiser prend en général la forme

$$f(x) = \frac{1}{2}\sum_{i=1}^{m} f_i(x)^2,$$

comme on peut le voir sur l'exemple vu au premier chapitre. Quand on veut appliquer la méthode de Newton à la minimisation de $f(x)$, on doit calculer le Hessien de f, qui dans ce cas précis prend une forme particulière : on a d'une part

$$\nabla f(x) = \sum_{i=1}^{m} \nabla f_i(x) f_i(x),$$

et le hessien de f est donné par

$$\nabla^2 f(x) = \sum_{i=1}^{m} \nabla f_i(x) \nabla f_i(x)^\top + \sum_{i=1}^{m} f_i(x) \nabla^2 f_i(x).$$

Si l'on se place près de l'optimum, où on supposera que les $f_i(x)$ sont petis, le deuxième terme peut alors être négligé. La matrice obtenue

$$H(x) = \sum_{i=1}^{m} \nabla f_i(x) \nabla f_i(x)^\top,$$

possède une propriété intéressante : elle est semi-définie positive. De plus dans la plupart des cas m est très supérieur à n et la matrice est la plupart du temps définie positive (nous reviendrons sur ce point). La méthode originale que l'on obtient à partir de la méthode de Newton en remplacant $\nabla^2 f(x)$ par $H(x)$ est la méthode de Gauss-Newton :

$$\left\{ \begin{array}{rcl} x_0 & & \text{donné,} \\ H_k & = & \sum_{i=1}^{m} \nabla f_i(x_k) \nabla f_i(x_k)^\top, \\ x_{k+1} & = & x_k - H_k^{-1} \nabla f(x_k). \end{array} \right.$$

V.3.2 la méthode de Levenberg-Marquardt

Pour assurer la convergence globale de la méthode de Gauss-Newton, on peut combiner l'algorithme précédent avec une recherche linéaire, et dans ce cas on peut alors faire les itérations

$$\left\{ \begin{array}{rcl} d_k & = & -H_k^{-1}\nabla f(x_k) \\ x_{k+1} & = & x_k + \rho_k d_k, \end{array} \right.$$

cependant, il n'y a aucune garantie que H_k reste défine positive, et en général on fait appel à une méthode modifiée, qui est la méthode de Levenberg-Marquardt : l'idée consiste à remplacer, dans la méthode précédente, la matrice H_k par la matrice $H_k + \lambda I$ où λ est un réel positif. Si λ est très grand, on retombe alors sur la méthode du gradient.

Méthode de Levenberg-Marquardt

$$\left\{ \begin{array}{rcl} x_0 & & \text{donné}, \\ H_k & = & \sum_{i=1}^m \nabla f_i(x_k)\nabla f_i(x_k)^\top, \\ d_k & = & -(H_k + \lambda I)^{-1}\nabla f(x_k) \\ x_{k+1} & = & x_k + \rho_k d_k, \end{array} \right.$$

Chapitre VI

Conditions d’optimalité en optimisation avec contraintes

VI.1 Les conditions de Lagrange

VI.1.1 Introduction

On s'intéresse maintenant à des problèmes d'optimisation de la forme

$$(PC)\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(VI.1.1)} \\ \text{sous les contraintes} & \\ g(x) \leq 0, & \text{(VI.1.2)} \\ h(x) = 0, & \text{(VI.1.3)} \end{cases}$$

où les fonctions f, g et h sont différentiables au moins une fois, et f est typiquement non-linéaire. Cependant nous étudierons le cas où g et h sont linéaires avec un intérêt tout particulier. Dans ce chapitre nous allons nous efforcer d'obtenir les conditions d'optimalité associées au problème (PC). Les chapitres suivants mettront ensuite l'accent sur les méthodes numériques permettant de le résoudre. Nous nous intéresserons précisément dans ce chapitre aux problèmes

(PCE) problème avec contraintes d'égalité,
(PCI) problème avec contraintes d'inégalité,

et les résultats s'étendront facilement aux problème général *(PC)*.

VI.1.2 Problème avec contraintes d'égalité

On va tout d'abord s'intéresser au problème suivant, dit problème d'optimisation avec contraintes d'égalité seulement :

$$(PCE)\ \begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(VI.1.4)} \\ \text{sous les contraintes} \\ h(x) = 0. & \text{(VI.1.5)} \end{cases}$$

La raison majeure justifiant que l'on s'intéresse en premier au problème (PCE) est que (PC) est un problème du type (PCI) dont on ne sait pas quelles sont les contraintes *actives* (nous reviendrons sur cette terminologie plus tard). Nous allons dans un premier temps nous intéresser au cas où les contraintes sont linéaires.

VI.1.3 Contraintes d'égalité linéaires

Un problème d'optimisation avec contraintes d'égalité linéaires prend la forme

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(VI.1.6)} \\ Ax - b = 0. & \text{(VI.1.7)} \end{cases}$$

où A est une matrice $p \times n$ avec $p < n$ et $b \in R^p$. On notera

$$S = \{x \in \mathbb{R}^n,\ Ax - b = 0\}.$$

Nous allons maintenant définir le concept de direction admissible dans S.

Définition VI.1.1. *On dit que $d \in \mathbb{R}^n$ est une direction admissible en $x \in S$ s'il existe $\alpha > 0$ tel que*

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

Dans notre cas, on a $A(x + td) - b = tAd$ puisque $x \in S$, et donc les directions admissibles d sont caractérisées par

$$Ad = 0. \quad \text{(VI.1.8)}$$

Rappellons maintenant un résultat bien utile d'algèbre linéaire :

Théorème VI.1.2. *Soit A une matrice $p \times n$. On a la relation suivante*

$$(\text{Ker } A)^\perp = (\text{Im } A^\top)$$

On peut donc énoncer les conditions nécessaires d'optimalité pour le problème (VI.1.6) :

Théorème VI.1.3. *Soit $\hat{x} \in S$ solution du problème (VI.1.6), vérifiant donc*

$$f(\hat{x}) \leq f(x),\ \ \forall\, x \in S$$

Alors il existe nécessairement un vecteur $\lambda \in \mathbb{R}^p$ vérifiant

$$\nabla f(\hat{x}) + A^\top \lambda = 0.$$

Si de plus A est de rang p alors λ est unique.

Démonstration : Soit d une direction admissible, vérifiant donc $d \in \text{Ker } A$. Pour tout $t \in \mathbb{R}$ on a

$$f(\hat{x}) \leq f(\hat{x} + td),$$

soit

$$\begin{aligned} \frac{f(\hat{x} + td) - f(\hat{x})}{t} &\geq 0,\ \ t > 0, \\ \frac{f(\hat{x} + td) - f(\hat{x})}{t} &\leq 0,\ \ t < 0. \end{aligned}$$

Si on prend la limite de ces deux expressions quand t tend vers 0 en en déduit que

$$\nabla f(\hat{x})^\top d = 0,\ \ \forall d \in \text{Ker } A$$

soit $\nabla f(\hat{x}) \in (\text{Ker } A)^\perp$, donc $\nabla f(\hat{x}) \in \text{Im } A^\top$. Il existe donc un vecteur λ tel que

$$\nabla f(\hat{x}) = -A^\top \lambda,$$

ce qui démontre le résultat. Pour l'unicité, supposons qu'il existe deux vecteurs λ_1 et λ_2 vérifiant

$$\nabla f(\hat{x}) = -A^\top \lambda_1 = -A^\top \lambda_2.$$

On a donc

$$A^\top(\lambda_1 - \lambda_2) = 0,$$

et donc $\lambda_1 - \lambda_2 = 0$ si A est de rang p. □

VI.1.4 Contraintes d'égalité non-linéaires

Nous étudions maintenant le problème

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(VI.1.9)} \\ h(x) = 0. & \text{(VI.1.10)} \end{cases}$$

où $h : \mathbb{R}^n \to \mathbb{R}^p$ est différentiable. On note comme précédemment

$$S = \{x \in \mathbb{R}^n,\ h(x) = 0\}.$$

Le concept de direction admissible dans S ne peut pas se définir comme pour les contraintes linéaires, car pour $\hat{x} \in S$ il peut ne pas exister $\alpha > 0$ et $d \in \mathbb{R}^n$ tels que $\hat{x} + td \in S$. On doit donc définir le concept de courbe admissible.

Considérons une courbe $x(t)$ définie pour $t \geq 0$ vérifiant

$$\begin{cases} x(t) \in S,\ \forall t \in [-\alpha, \alpha],\ \alpha > 0 \\ x(0) = \hat{x}. \end{cases}$$

Puisque $x(t) \in S$ on a $h_i(x(t)) = 0$ pour $1 \leq i \leq p$ et on peut écrire que

$$\frac{d}{dt} h_i(x(t)) = \nabla h_i(x(t))^\top \dot{x}(t) = 0,\ \ 1 \leq i \leq p.$$

Si on note $y = \dot{x}(0)$ le vecteur tangent à la courbe $x(t)$ en $t = 0$, on a donc

$$\nabla h_i(\hat{x})^\top y = 0,\ \ 1 \leq i \leq p. \qquad \text{(VI.1.11)}$$

Cela conduit à la définition suivante :

Définition VI.1.4. *On dit que $y \in \mathbb{R}^n$ est une direction admissible en $\hat{x} \in S$ s'il existe $\alpha > 0$ et une courbe $x(t)$ vérifiant*

$$\begin{cases} x(t) \in S,\ \forall t \in [-\alpha, \alpha], \\ x(0) = \hat{x}, \\ \dot{x}(0) = y. \end{cases}$$

On notera alors $y \in T(\hat{x})$.

L'ensemble $T(\hat{x})$ définit le plan tangent à S en $\hat{x}$. L'analyse faite précédemment montre que l'on a l'implication

$$y \in T(\hat{x}) \Rightarrow \nabla h_i(\hat{x})^\top y = 0,\ \ 1 \leq i \leq p,$$

qui sera insuffisante pour montrer la condition nécéssaire d'optimalité. Nous allons donc maintenant nous attacher à montrer sous quelles conditions la relation (VI.1.11) est une condition suffisante d'appartenance à $T(\hat{x})$.

Définition VI.1.5. *On dit que $\hat{x}$ est un point régulier pour la contrainte $h(x) = 0$ si*

- $h(\hat{x}) = 0$,
- *Les vecteurs $\nabla h_i(\hat{x})$ sont linéairement indépendants.*

Si on note $\nabla h(\hat{x})$ la matrice $n \times p$

$$\nabla h(\hat{x}) = [\nabla h_1(\hat{x}) \dots \nabla h_p(\hat{x})]\,,$$

la condition d'indépendance linéaire des $\nabla h_i(\hat{x})$ peut s'écrire

$$\text{Rang}\, \nabla h(\hat{x}) = p.$$

et on a donc $\nabla h(\hat{x})^\top \dot{x}(0) = 0$ pour toute courbe admissible $x(t)$.

On a la proposition suivante :

Proposition VI.1.1. *Si $\hat{x}$ est un point régulier pour la contrainte $h(x) = 0$, alors*

$$\nabla h(\hat{x})^\top y = 0 \Rightarrow y \in T(\hat{x}).$$

Démonstration : Soit $y \in \mathbb{R}^n$ vérifiant $\nabla h(\hat{x})^\top y = 0$. On considère la courbe $x(t)$ donnée par

$$x(t) = \hat{x} + ty + \nabla h(\hat{x})u(t).$$

La fonction $u(t) \in \mathbb{R}^p$, pour l'instant inconnue, va être déterminée de telle façon que $h(x(t)) = 0$. On va pour cela poser le problème de la détermination de $u(t)$ sous la forme d'une équation implicite. On définit la fonction $F : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$ par

$$F(t, u) = h(\hat{x} + ty + \nabla h(\hat{x})u).$$

Le problème de la détermination de $u(t)$ se ramène donc à la résolution de l'équation

$$F(t, u) = 0,$$

au voisinage du point $(0, 0)$. On a d'une part $F(0, 0) = h(\hat{x}) = 0$ et

$$\frac{\partial}{\partial u}F(t, u) = \nabla h(\hat{x})^\top \nabla h(\hat{x} + ty + \nabla h(\hat{x})u),$$

soit

$$\frac{\partial}{\partial u}F(0, 0) = \nabla h(\hat{x})^\top \nabla h(\hat{x}).$$

La matrice $\frac{\partial}{\partial u}F(0, 0)$ est inversible puisque par hypothèse $\nabla h(\hat{x})$ est de rang p. On peut alors appliquer le théorème des fonctions implicites : il existe un voisinage du point $(0, 0)$ et une fonction $u(t)$ tels que

$$F(t, u) = 0 \Leftrightarrow u = u(t).$$

Notons que l'on a donc nécéssairement $u(0) = 0$ puisque $F(0, 0) = 0$.

On a donc maintenant

$$\dot{x}(t) = y + \nabla h(\hat{x})\dot{u}(t)$$

soit en $t = 0$

$$\dot{x}(0) = y + \nabla h(\hat{x})\dot{u}(0).$$

Montrons que $\dot{u}(0) = 0$. Pour cela on écrit que l'on a

$$\frac{d}{dt}h(x(t)) = \nabla h(x(t))^\top (y + \nabla h(\hat{x})\dot{u}(t)) = 0,$$

puisque $h(x(t)) = 0$, et donc en $t = 0$ la relation précédente prend la forme

$$\left.\frac{d}{dt}h(x(t))\right|_{t=0} = \nabla h(\hat{x})^\top y + \nabla h(\hat{x})^\top \nabla h(\hat{x})\dot{u}(0) = 0.$$

Le premier terme du second membre est nul par hypothèse, et donc $\dot{u}(0) = 0$ puisque $\nabla h(\hat{x})^\top \nabla h(\hat{x})$ est inversible. Donc

$$\dot{x}(0) = y,$$

soit $y \in T(\hat{x})$, ce qui démontre le résultat annoncé. □

VI.1.5 Le théorème de Lagrange

Théorème VI.1.6. *Soit $\hat{x} \in S = \{x \in \mathbb{R}^n,\ h(x) = 0\}$ un point régulier solution du problème (VI.1.9), vérifiant donc*

$$f(\hat{x}) \leq f(x),\ \ \forall\, x \in S$$

Alors il existe nécessairement un vecteur $\lambda \in \mathbb{R}^p$ unique vérifiant

$$\nabla f(\hat{x}) + \nabla h(\hat{x})\lambda = 0,$$

soit encore

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

Les composantes du vecteur λ sont appelées multiplicateurs de Lagrange.

Démonstration : Considérons une courbe $x(t)$ définie pour $t \in [-\alpha, \alpha]$ vérifiant

$$\left\{ \begin{array}{lcl} x(t) & \in & S,\ \forall t \in [-\alpha, \alpha],\ \alpha > 0 \\ x(0) & = & \hat{x}. \end{array} \right.$$

On a

$$f(x(0)) \leq f(x(t)),\ \ \forall t \in [-\alpha, \alpha],$$

donc nécessairement

$$\frac{d}{dt} f(x(t))\bigg|_{t=0} = \nabla f(\hat{x})^\top \dot{x}(0) = 0,$$

ce qui signifie que $\nabla f(\hat{x})$ se trouve dans l'orthogonal de $T(\hat{x})$ le plan tangent à S en $\hat{x}$. Si l'on utilise l'équivalence

$$T(\hat{x}) = \text{Ker}\, \nabla h(\hat{x})^\top \Leftrightarrow T(\hat{x})^\perp = \text{Im}\, \nabla h(\hat{x}),$$

il existe donc un vecteur $\lambda \in \mathbb{R}^p$ tel que

$$\nabla f(\hat{x}) = -\nabla h(\hat{x})\lambda.$$

L'unicité résulte du fait que $\nabla h(\hat{x})$ est de rang p et se montre comme dans le cas linéaire. □

VI.2 Les conditions de Kuhn et Tucker

VI.2.1 Problème avec contraintes d'inégalité

On s'intéresse maintenant au problème suivant, dit problème d'optimisation avec contraintes d'inégalité seulement :

$$(PCI)\quad \begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(VI.2.1)} \\ \text{sous les contraintes} \\ g(x) \leq 0, & \text{(VI.2.2)} \end{cases}$$

où $g : \mathbb{R}^n \to \mathbb{R}^m$, est différentiable (il n'y a ici aucune condition sur m). On notera K l'ensemble des points admissibles, c'est à dire

$$K = \{x \in \mathbb{R}^n,\ g(x) \leq 0\}.$$

Au point solution de (PCI) il va de soi que les contraintes effectivement actives vérifieront $g_i(\hat{x}) = 0$. Cependant, puisque l'on ne sait pas *a priori* quelles sont ces contraintes, le passage de (PCI) a un problème du type (PCE) n'est pas direct.

Définition VI.2.1. *On appelle contraintes saturées en $\hat{x}$ l'ensemble des indices i tel que $g_i(\hat{x}) = 0$, et on note*

$$I(\hat{x}) = \{i \mid g_i(\hat{x}) = 0\}.$$

On note alors $S(\hat{x})$ l'ensemble

$$S(\hat{x}) = \{x \in \mathbb{R}^n,\ g_i(x) = 0,\ i \in I(\hat{x})\}.$$

Le concept de direction admissible se définit comme suit :

Définition VI.2.2. *On dit que $y \in \mathbb{R}^n$ est une direction admissible en $\hat{x} \in K$ s'il existe $\alpha > 0$ et une courbe $x(t)$ vérifiant*

$$\begin{cases} x(t) & \in & K,\ \forall t \in [-\alpha, \alpha], \\ x(0) & = & \hat{x}, \\ \dot{x}(0) & = & y. \end{cases}$$

On notera alors $y \in C(\hat{x})$.

Lemme VI.2.3. *Soit $y \in \mathbb{R}^n$ une direction admissible en $\hat{x} \in K$, alors on a nécessairement*

$$\nabla g_i(\hat{x})^\top y \leq 0,\ i \in I(\hat{x}).$$

Démonstration : Considérons une courbe $x(t)$ définie pour $t \in [-\alpha, \alpha]$ vérifiant

$$\begin{cases} x(t) & \in & K,\ \forall t \in [-\alpha, \alpha],\ \alpha > 0 \\ x(0) & = & \hat{x}, \\ \dot{x}(0) & = & y. \end{cases}$$

Comme $g_i(\hat{x}) < 0$ pour $i \notin I(\hat{x})$, on aura toujours $g_i(x(t)) < 0$ pour t suffisamment petit. Par contre, pour $i \in I(\hat{x})$ on doit avoir $g_i(x(t)) \leq 0$ pour t suffisamment petit. Si on utilise le développement de Taylor de $g_i(x(t))$ en $t = 0$ on doit donc avoir

$$g_i(\hat{x}) + t\nabla g_i(\hat{x})^\top y + t\epsilon(t) \leq 0.$$

Puisque $g_i(\hat{x}) = 0$ il faut donc nécessairement que l'on ait

$$\nabla g_i(\hat{x})^\top y \leq 0.$$

□ Comme dans le cas des contraintes d'égalité, on doit définir la notion de point régulier, qui est nécessaire pour que la condition précédente soit suffisante :

Définition VI.2.4. *On dit que $\hat{x}$ est un point régulier pour la contrainte $g(x) \leq 0$ si*
- $g(\hat{x}) \leq 0$,
- *Les vecteurs $\{\nabla h_i(\hat{x})\}_{i \in I(\hat{x})}$ sont linéairement indépendants.*

Sous l'hypothèse de régularité de $\hat{x}$ on aura, comme dans le cas des contraintes d'égalité

$$\nabla g_i(\hat{x})^\top y \leq 0,\ i \in I(\hat{x}) \Rightarrow y \in C(\hat{x}).$$

La proposition suivante permet d'effectuer le premier pas vers les conditions de Kuhn et Tucker.

Proposition VI.2.1. *Soit $\hat{x}$ la solution du problème (PCI). Il existe $\eta > 0$ tel que*

$$\forall x \in B(\hat{x}, \eta),\ g_i(x) < 0,\ i \notin I(\hat{x}),$$

où on a noté $B(\hat{x}, \eta)$ la boule de centre $\hat{x}$ et de rayon η. Alors $\hat{x}$ est la solution du problème

$$\begin{cases} \min\limits_{x \in B(\hat{x},\eta)} f(x), & \text{(VI.2.3)} \\ g_i(x) = 0,\ i \in I(\hat{x}). & \text{(VI.2.4)} \end{cases}$$

Ce résultat est uniquement dû à la continuité de g, et montre que l'on est localement ramené à un problème avec contraintes d'égalité. On peut donc maintenant énoncer le résulat principal :

Théorème VI.2.5. *Soit $\hat{x} \in K$ un point régulier solution du problème (PCI). Alors il existe un unique vecteur $\lambda \in \mathbb{R}^m$ tel que*

$$\begin{aligned} \nabla f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) &= 0, && \text{(VI.2.5)} \\ \lambda_i &\geq 0,\ i = 1 \dots m, && \text{(VI.2.6)} \\ \lambda_i g_i(\hat{x}) &= 0,\ i = 1 \dots m && \text{(VI.2.7)} \end{aligned}$$

Démonstration : Les relation (VI.2.5) (VI.2.7) sont une conséquence directe du théorème de Lagrange, car il suffit de prendre $\lambda_i = 0$ pour $i \notin I(\hat{x})$. On peut ensuite montrer (VI.2.6) par l'absurde : supposons qu'il existe $k \in I(\hat{x})$ tel que $\lambda_k < 0$. On définit la surface

$$S_k = \{x \mid g_i(x) = 0,\ i \in I(\hat{x}),\ i \neq k\}.$$

On définit $y \in \mathbb{R}^n$ tel que

$$\begin{aligned} \nabla g_i(\hat{x})^\top y &= 0,\ i \in I(\hat{x}),\ i \neq k, \\ \nabla g_k(\hat{x})^\top y &= -1. \end{aligned}$$

Alors y est une direction admissible en $\hat{x}$ puisque

$$\nabla g_i(\hat{x})^\top y \leq 0,\ i \in I(\hat{x}),$$

et que $\hat{x}$ est un point régulier. Il existe donc une courbe $x(t) \in S_k$ et vérifiant de plus $x(t) \in K$, pour $t \in [\alpha, \alpha]$, telle que $\dot{x}(0) = y$. On a donc

$$\begin{aligned} \left.\frac{d}{dt} f(x(t))\right|_{t=0} &= \nabla f(\hat{x})^\top y, && \text{(VI.2.8)} \\ &= -\sum \lambda_i \nabla g_i(\hat{x})^\top y, && \text{(VI.2.9)} \\ &= -\lambda_k \nabla g_k(\hat{x})^\top y = \lambda_k < 0, && \text{(VI.2.10)} \end{aligned}$$

ce qui est impossible car f est minimum en $\hat{x}$. □

VI.2.2 Interprétation géométrique des conditions de Kuhn et Tucker

On considère un cas où $I(\hat{x}) = \{1, 2\}$. Au point $\hat{x}$, l'ensemble des directions admissibles $C(\hat{x})$ forme un cône qui est l'intersections des demi-espaces d'équation

$$\nabla g_i(\hat{x})^\top y \leq 0,\ i = 1, 2.$$

Pour que $\hat{x}$ soit un optimum local, il faut que le vecteur $-\nabla f(\hat{x})$ forme un angle obtus avec les

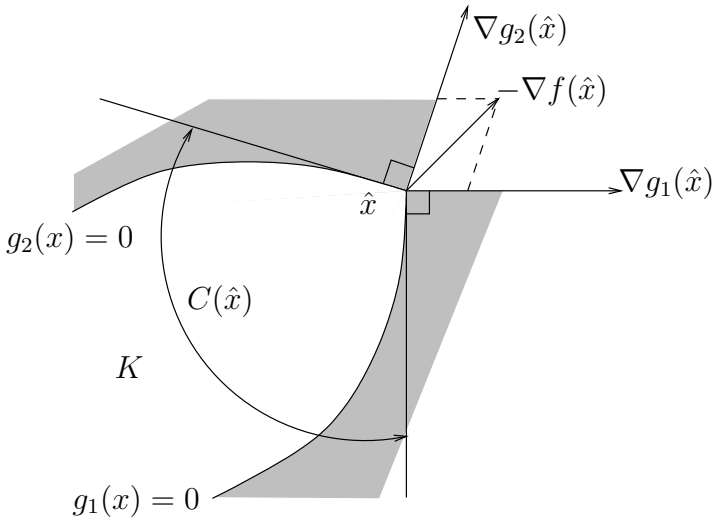


FIG. VI.2.1 – Illustration des conditions de Kuhn et Tucker sur un exemple à deux dimensions.

directions admissibles. On vérifie aussi que $-\nabla f(\hat{x})$ est combinaison linéaire (à coefficients positifs) des vecteurs $\nabla g_i(\hat{x})$, $i = 1, 2$.

VI.3 Exemples de problèmes

VI.3.1 Distance d'un point à un plan

On cherche à calculer la distance d'un point $x_0 \in \mathbb{R}^n$ au plan défini par l'équation $Ax = b$, où $A \in \mathcal{M}_{pn}$ avec $\operatorname{Rang} A = p$. Se problème se pose sous la forme

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \frac{1}{2} \|x_0 - x\|^2 \\ Ax = b. \end{aligned}$$

On pose donc $f(x) = \frac{1}{2} \|x_0 - x\|^2$. On a

$$\nabla f(x) = -(x_0 - x),$$

et donc le système d'optimalité est donné par

$$\begin{aligned} (\hat{x} - x_0) + A^\top \hat{\lambda} &= 0, && \text{(VI.3.1)} \\ A\hat{x} &= b. && \text{(VI.3.2)} \end{aligned}$$

En multipliant l'équation (VI.3.1) par A on peut exprimer $\hat{\lambda}$ par

$$\hat{\lambda} = (AA^\top)^{-1}(Ax_0 - d),$$

et on obtient en substituant $\hat{\lambda}$ dans (VI.3.2)

$$\hat{x} = (I - A^\top (AA^\top)^{-1} A)x_0 + A^\top (AA^\top)^{-1} d.$$

Un problème voisin est celui de la projection d'une direction d sur le plan $Ax = 0$. Le résultat précédent donne donc

$$\hat{d} = Pd,$$

avec $P = I - A^\top (AA^\top)^{-1}$.

VI.3.2 Pseudo-inverse de Moore et Penrose

On cherche à résoudre le système

$$Ax = b,$$

avec $A \in \mathcal{M}_{pn}$, $p < n$ et A de rang p. Il s'agit donc d'un système sous-déterminé. La pseudo-inverse de Moore-Penrose est par définition la matrice $A^\dagger$ telle que le vecteur

$$\hat{x} = A^\dagger b,$$

est la solution de norme minimale du système

$$Ax = b.$$

Le problème d'optimisation à résoudre est donc :

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \frac{1}{2} \|x\|^2 \\ Ax = b, \end{aligned}$$

et le système d'optimalité est donné par

$$\hat{x} + A^\top \hat{\lambda} = 0, \tag{VI.3.3}$$

$$A\hat{x} = b. \tag{VI.3.4}$$

Il suffit de substituer $\hat{x}$ dans la deuxième équation et puisque AA^top est de rang p on obtient

$$\hat{x} = A^\top (AA^\top)^{-1} b,$$

et donc la pseudo-inverse est donnée par

$$A^\dagger = A^\top (AA^\top)^{-1}.$$

VI.3.3 Exemple de programme quadratique

On cherche à résoudre le problème

$$\min_{x\in\mathbb{R}^2} \frac{1}{2}\|x-x_0\|^2$$

$$\begin{aligned} x_1 &\geq 0,\\ x_2 &\geq 0,\\ x_1+x_2 &\leq 1, \end{aligned}$$

où $x_0 = (1, \frac{1}{2})$. Il s'agit d'un problème avec contraintes d'inégalité se mettant sous la forme $g(x) \leq 0$ avec

$$g(x) = \begin{pmatrix} -x_1 \\ -x_2 \\ x_1+x_2-1 \end{pmatrix}.$$

Sur le dessin, on peut s'assurer que très probablement seule la contrainte numéro 3 est active. On peut

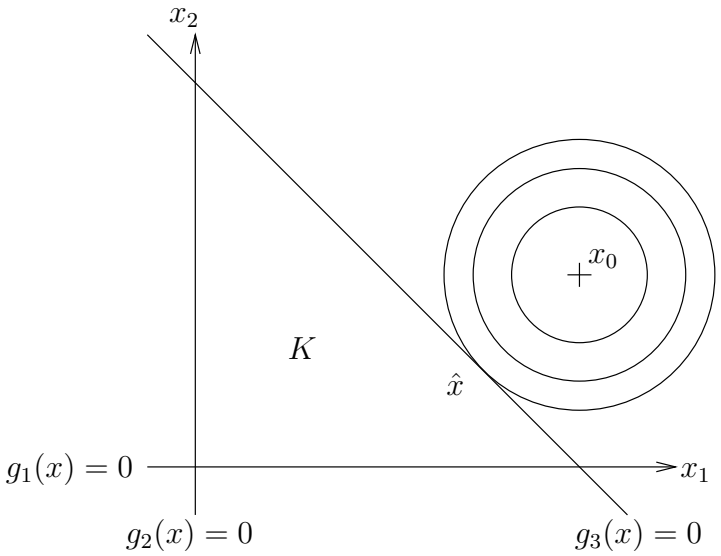


FIG. VI.3.2 – Exemple de programme quadratique

s'en persuader par le calcul de la façon suivante : on peut tenter de résoudre le système

$$\begin{aligned} \nabla f(x) + \lambda_3 \nabla g_3(x) &= 0,\\ g_3(x) &= 0, \end{aligned}$$

soit ici

$$\begin{aligned} x - x_0 + \lambda_3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= 0,\\ x_1 + x_2 &= 1, \end{aligned}$$

ou bien encore

$$\begin{aligned} x_1 + \lambda_3 &= 1,\\ x_2 + \lambda_3 &= \frac{1}{2},\\ x_1 + x_2 &= 1, \end{aligned}$$

dont la solution est donnée par

$$x_1 = \frac{3}{4},\ x_2 = \frac{1}{4}, \lambda_3 = \frac{1}{4}.$$

On a bien $\lambda_3 \geq 0$ ce qui justifie *a posteriori* le choix de saturer la contrainte numéro 3.

VI.4 Conditions suffisantes d'optimalité

VI.4.1 Définition du lagrangien

Considérons le problème (PCE) avec contraintes d'égalité

$$\begin{cases} \min_{x\in\mathbb{R}^n} f(x), \\ h(x) = 0, \end{cases}$$

où $h : \mathbb{R}^n \to \mathbb{R}^p$.

Définition VI.4.1. *On appelle* lagrangien *associé au problème* (PCE) *la fonction* $L : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ *définie par*

$$L(x, \lambda) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x).$$

Les conditions de Lagrange peuvent se reformuler à l'aide du lagrangien : soit $\hat{x}$ solution de (PCE). Alors il existe $\hat{\lambda}$ tel que

$$\nabla_x L(\hat{x}, \hat{\lambda}) = 0,$$

où on a noté ∇_x le gradient partiel par rapport à la variable x. Dans la suite nous ferons l'hypothèse que h et f sont deux fois continûment différentiables.

VI.4.2 Condition nécéssaire du second ordre

Théorème VI.4.2. *Soit $\hat{x}$ un point régulier solution de (PCE). Alors il existe $\hat{\lambda}$ tel que*

$$\nabla_x L(\hat{x}, \hat{\lambda}) = 0,$$

et de plus pour tout $y \in T(\hat{x})$, $y \neq 0$, on a

$$y^\top \nabla^2_{xx} L(\hat{x}, \hat{\lambda}) y \geq 0.$$

Démonstration : Soit $y \in T(\hat{x})$. On sait qu'il existe une courbe $x(t)$ définie pour $t \in [-\alpha, \alpha]$ vérifiant

$$\left\{ \begin{array}{rcl} x(t) & \in & S,\ \forall t \in [-\alpha, \alpha],\ \alpha > 0 \\ x(0) & = & \hat{x}, \\ \dot{x}(0) & = & y. \end{array} \right.$$

Puisque $\hat{x}$ est optimal on a

$$f(x(0)) \leq f(x(t)),\ \forall t,$$

et puisque la fonction f est deux fois différentiable, on a nécessairement

$$\frac{d^2}{dt^2} f(x(t))\bigg|_{t=0} \geq 0.$$

On a ici d'une part

$$\frac{d}{dt} f(x(t)) = \nabla f(x(t))^\top \dot{x}(t),$$

et donc

$$\frac{d^2}{dt^2} f(x(t)) = \dot{x}(t)^\top \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t))^\top \ddot{x}(t), \tag{VI.4.1}$$

$$\frac{d^2}{dt^2} f(x(t))\bigg|_{t=0} = y^\top \nabla^2 f(\hat{x}) y + \nabla f(\hat{x})^\top \ddot{x}(0) \geq 0 \tag{VI.4.2}$$

D'autre part on a $h_i(x(t)) = 0$ donc

$$\frac{d^2}{dt^2} h(x(t))\bigg|_{t=0} = y^\top \nabla^2 h_i(\hat{x}) y + \nabla h_i(\hat{x})^\top \ddot{x}(0) = 0,\ i = 1, \ldots, p.$$

On peut multiplier chacune de ces égalités par $\hat{\lambda}_i$ et en faire la somme, ce qui donne

$$y^\top \left(\sum_{i=1}^p \hat{\lambda}_i \nabla^2 h_i(\hat{x}) \right) y + \left(\sum_{i=1}^p \hat{\lambda}_i \nabla h_i(\hat{x})^\top \right) \ddot{x}(0) = 0.$$

En additionnant cette dernière égalité à (VI.4.2) on obtient

$$y^\top \left(\nabla^2 f(\hat{x}) + \sum_{i=1}^p \hat{\lambda}_i \nabla^2 h_i(\hat{x}) \right) y + \left(\nabla f(\hat{x}) + \sum_{i=1}^p \hat{\lambda}_i \nabla h_i(\hat{x}) \right)^\top \ddot{x}(0) \geq 0,$$

et puisque le deuxième terme est nul (condition de Lagrange) on obtient bien l'inégalité annonçée. □

Le résultat suivant est une généralisation du thèorème précédent dont la démonstration sera admise.

Théorème VI.4.3. *Soit $\hat{x} \in \mathbb{R}^n$ et $\hat{\lambda} \in \mathbb{R}^p$ vérifiant les conditions*

$$\begin{array}{rcl} h(\hat{x}) & = & 0, \\ \nabla f(\hat{x}) + \sum_{i=1}^p \hat{\lambda}_i \nabla h_i(\hat{x}) & = & 0, \\ y^\top \nabla^2_{xx} L(\hat{x}, \hat{\lambda}) y & \geq 0 & ,\ \forall y \in T(\hat{x}),\ y \neq 0, \end{array}$$

alors $\hat{x}$ est un minimum local du problème (PCE).

VI.4.3 Condition nécéssaire du second ordre

Théorème VI.4.4. *Soit $\hat{x} \in \mathbb{R}^n$ et $\hat{\lambda} \in \mathbb{R}^p$ vérifiant les conditions*

$$\begin{aligned} g(\hat{x}) &\leq 0, \\ \nabla f(\hat{x}) + \sum_{i=1}^{p} \hat{\lambda}_i \nabla g_i(\hat{x}) &= 0, \\ \hat{\lambda}_i &\geq 0,\ i = 1 \dots m, \\ \hat{\lambda}_i g_i(\hat{x}) &= 0,\ i = 1 \dots m, \\ y^\top \nabla^2_{xx} L(\hat{x}, \hat{\lambda}) y &\geq 0 \quad ,\ \forall y \in T^+(\hat{x}),\ y \neq 0, \end{aligned}$$

où on a noté $T^+(\hat{x})$ le plan tangent en $\hat{x}$ à la surface

$$S^+ = \{x \in \mathbb{R}^n,\ g_i(\hat{x}) = 0,\ i \in I(\hat{x}) \text{ et } \lambda_i > 0\}.$$

Alors $\hat{x}$ est un minimum local du problème (PCE).

Chapitre VII
Méthodes primales

VII.1 Contraintes d'égalité linéaires

VII.1.1 La méthode du gradient projeté

On s'intéresse à un problème avec contraintes d'égalité lineaires

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(VII.1.1)} \\ Ax - b = 0, & \text{(VII.1.2)} \end{cases}$$

et nous ferons l'hypothèse que $A \in \mathcal{M}_{pn}$ est de rang maximal. Une idée assez naturelle consiste à appliquer une méthode de descente qui prenne en compte la contrainte $Ax - b = 0$. Supposons que nous disposons d'un point $x_0 \in K = \{x \in \mathbb{R}^n,\ Ax - b = 0\}$. On sait qu'une direction admissible doit vérifier

$$Ad = 0. \tag{VII.1.3}$$

On peut chercher la meilleure direction de descente respectant (VII.2.3) en résolvant le problème

$$\begin{cases} \min \nabla f(x)^\top d, & \text{(VII.1.4)} \\ Ad = 0, & \text{(VII.1.5)} \\ \|d\| = 1. & \text{(VII.1.6)} \end{cases}$$

Proposition VII.1.1. *Le vecteur d solution du problème (VII.1.4),(VII.1.5),(VII.1.6) est donné par $d = y/\|y\|$ où y est la projection orthogonale de $-\nabla f(x)$ sur $\operatorname{Ker} A$.*

Démonstration : On peut écrire que

$$-\nabla f(x) = y + z,$$

où $y \in \operatorname{Ker} A$ et $z \in (\operatorname{Ker} A)^\perp$, ces deux sous-espaces étant complémentaires dans $\mathbb{R}^n$. On a donc

$$-\nabla f(x)^\top d = -y^\top d.$$

Comme d est un vecteur unitaire quelconque $y^\top d$ sera maximal pour

$$d = \frac{y}{\|y\|},$$

d'où le résultat. On remarquera que si $y \neq 0$, le vecteur d est bien une direction de descente car on a

$$\nabla f(x)^\top = -y^\top(y + z) = -y^\top y < 0.$$

□

Pour former la matrice de projection sur $\operatorname{Ker} A$ on utilise en général la factorisation QR de la matrice $A^\top$, qui s'exprime sous la forme

$$A^\top = Q \begin{pmatrix} R \\ 0 \end{pmatrix},$$

où $R \in \mathcal{M}_{pp}$ est triangulaire supérieure et $Q \in \mathcal{M}_{nn}$ est orthogonale, et se décompose en $Q = [U\ \ V]$ où les colonnes de $U \in \mathcal{M}_{n,p}$ forment une base orthogonale de $\operatorname{Im} A^\top$ et les colonnes de $V \in \mathcal{M}_{n,n-p}$ une base orthogonale de $(\operatorname{Im} A^\top)^\perp = \operatorname{Ker} A$. Dans ce cas la matrice de la projection orthogonale sur $\operatorname{Ker} A$ s'écrit

$$P = I - UU^\top = VV^\top.$$

Remarque VII.1.1. *Dans l'algorithme que nous allons étudier, la matrice de projection peut être calculée une fois pour toutes puisque A est donnée. Cependant, pour les problèmes avec contraintes d'inégalité lineéaires, on sera amené à considérer une succession de problèmes avec contraintes d'égalité, et la matrice A pourra évoluer à chaque itération, par ajout ou supression d'une ligne. Le choix de la factorisation QR est tout indiqué car il existe des techniques de mise à jour particulièrement économiques, ce qui n'est pas le cas quand on exprime la matrice P sous la forme classique*

$$P = I - A^\top [AA^\top]^{-1} A.$$

La méthode du gradient projeté consiste tout simplement à mettre en oeuvre une méthode de descente utilisant à chaque pas la direction $d_k = -VV^\top \nabla f(x_k)$. Les itérations sont poursuivies jusqu'à ce que $d_k = 0$. Cela signifie alors que $\nabla f(x) \in \operatorname{Im} A^\top$ et donc qu'il existe λ tel que

$$\nabla f(x_k) = -A\top\lambda.$$

On peut utiliser la factorisation de $A^\top$ pour obtenir λ par résolution du système linéaire

$$R\lambda = -U^\top \nabla f(x).$$

Algorithme du gradient projeté

1. Poser $k = 0$ et choisir x_0 admissible.
2. Calculer la projection $d_k = -VV^\top \nabla f(x_k)$,
3. Si $d_k = 0$
   - Calculer $\lambda = -R^{-1}U^\top \nabla f(x_k)$
   - Arrêter les itérations.
4. Déterminer $\rho_k > 0$ réalisant le minimum de $f(x_k + \rho d_k)$.
5. Poser $x_{k+1} = x_k + \rho_k d_k$, faire $k \leftarrow k + 1$ et retourner en 2.

VII.1.2 La méthode de Newton projetée

La méthode du gradient projeté souffrant des mêmes problèmes que la méthode du gradient (vitesse de convergence très sensible au conditionnement), on lui préfère souvent les méthodes de quasi-Newton adaptées au cas des contraintes linéaires. Il est plus facile de comprendre comment fonctionnent ces méthodes en faisant l'analyse suivante

Supposons que l'on dispose d'un point x_0 admissible. L'idée est de poser $x = x_0 + Vz$ et de considérer une nouvelle fonction $\tilde{f}$ définie par

$$\tilde{f}(z) = f(x_0 + Vz),$$

où les colonnes de V forment une base orthogonale de $\mathrm{Ker}\, A$ (on a vu comment obtenir une telle matrice). Alors par construction le problème (**??**) est équivalent au problème sans contraintes

$$\min_{z \in \mathbb{R}^p} \tilde{f}(z), \tag{VII.1.7}$$

puisque

$$A(x_0 + Vz) - b = Ax_0 - b + AVz = 0.$$

On peut donc appliquer n'importe quelle méthode de descente à la résolution de (VII.1.7). Notons que l'on a

$$\nabla \tilde{f}(z) = V^\top \nabla f(x_0 + Vz),$$

donc la méthode du gradient appliquée à la minimisation de $\tilde{f}(z)$ s'écrit

$$z_{k+1} = z_k - \rho_k V^\top \nabla f(x_0 + Vz_k),$$

et si on pose $x_k = x_0 + Vz_k$, les itérations précédentes s'écrivent

$$x_{k+1} = x_k - \rho_k VV^\top \nabla f(x_k),$$

ce qui redonne exactement la méthode du gradient projeté. On peut de la même manière écrire la méthode de Newton appliquée à la résolution de (VII.1.7) : le hessien de $\tilde{f}$ s'écrit

$$\nabla^2 \tilde{f}(z) = V^\top \nabla^2 f(x_0 + Vz) V,$$

si si on note $G_k = \nabla^2 \tilde{f}(z_k)$ la direction de Newton en z_k s'écrit

$$p_k = -G_k^{-1} \nabla \tilde{f}(z_k).$$

Si la matrice G_k est définie positive alors p_k sera une direction de descente pour $\tilde{f}$ et le vecteur Vp_k sera une direction de descente pour f puisque

$$\nabla f(x_k)^\top V p_k = \nabla \tilde{f}(z_k)^\top p_k < 0.$$

Remarque VII.1.2. *On sait que dans le cas général un optimum local du problème (PCE) est caractérisé par*

$$y^\top \nabla^2_{xx} L(\hat{x}, \hat{\lambda}) y \geq 0, \ \forall y \in T(\hat{x}), \ y \neq 0.$$

Or dans le cas des contraintes linéaires on a

$$\nabla^2_{xx} L(x, \lambda) = \nabla^2 f(x), \tag{VII.1.8}$$

et le sous espace $T(\hat{x})$ n'est autre que $\mathrm{Ker}\, A$. Et donc si l'on dispose d'une matrice V dont les colonnes forment une base orthogonale de $\mathrm{Ker}\, A$, tout vecteur $y \in T(\hat{x})$ s'exprime sous la forme $y = Vz$ et la condition (VII.1.8) s'écrit

$$zV^\top \nabla^2 f(\hat{x}) Vz > 0, \ \ \forall\, z.$$

On est donc assuré que le hessien projeté est défini positif à l'optimum, ce qui justifie l'utilisation des méthodes de quasi-Newton.

On peut donc envisager une méthode de quasi-Newton ou la mise à jour opère non pas sur le hessien de f mais sur le hessien projeté. Voici l'algorithme correspondant pour la méthode BFGS :

Algorithme de la méthode BFGS projetée

1. Poser $k = 0$, choisir x_0 admissible et poser $H_0 = I$.
2. Poser $g_k = V^\top \nabla f(x_k)$.
3. Si $g_k = 0$
   - Calculer $\lambda = -R^{-1}U^\top \nabla f(x_k)$
   - Arrêter les itérations.
4. Calculer la direction $p_k = -H_k^{-1} g_k$.
5. Déterminer $\rho_k > 0$ réalisant le minimum de $f(x_k + \rho V p_k)$.
6. Poser $x_{k+1} = x_k + \rho_k V p_k$.
7. Calculer $g_{k+1} = V^\top \nabla f(x_{k+1})$ et $y_k = g_{k+1} - g_k$.
8. Mise à jour du hessien projeté
   $$H_{k+1} = H_k + \frac{y_k y_k^\top}{\rho_k y_k^\top p_k} + \frac{g_k g_k^\top}{p_k^\top g_k}$$
9. faire $k \leftarrow k + 1$ et retourner en 2.

VII.2 Contraintes d'inégalité linéaires

VII.2.1 Méthode de directions réalisables

On s'intéresse maintenant à un problème avec contraintes d'inégalités lineaires

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(VII.2.1)} \\ Ax - b \leq 0. & \text{(VII.2.2)} \end{cases}$$

On peut essayer de voir comment adapter la stratégie de l'algorithme du gradient projeté. Supposons que nous disposons d'un point initial admissible $x_0 \in K = \{x \in \mathbb{R}^n,\ Ax - b \leq 0\}$. Notons I_0 l'ensemble des indices des contraintes saturées, soit

$$I_0 = \{i \mid A_i x_0 - b_i = 0\}.$$

On peut chercher une direction de descente d qui permette, au moins pour un petit déplacement, de rester dans K. Si on note $A_0 \in \mathcal{M}_{pn}$ la matrice composée des lignes $i \in I_0$ on doit donc avoir

$$A_{I_0} d = 0. \qquad \text{(VII.2.3)}$$

Après calcul de la factorisation $(U \ \ V)\begin{pmatrix} R \\ 0 \end{pmatrix}$ de $A_{I_0}^\top$, une direction admissible d peut être obtenue par $d = -VV^\top \nabla f(x_0)$.

Il y a ensuite deux cas à envisager :

1. Si $d \neq 0$, il faut déterminer le déplacement maximal autorisé par les contraintes non saturées, c'est à dire ρ_{max} tel que

   $$\rho_{max} = \{\rho \,|\rho \geq 0,\ A_i(x_0 + \rho d) - b_i \leq 0,\ i \notin I_0\}.$$

   Ensuite, on cherche le pas optimal ρ_{opt} dans direction d. Ce pas pouvant faire sortir du domaine admissible, on prendra donc toujours

   $$\rho = \min(\rho_{opt}, \rho_{max}),$$

   en notant bien que lorsque $\rho = \rho_{max}$, cela signifie qu'une nouvelle contrainte sera saturée.
2. Si $d = 0$ cela signifie que $\nabla f(x) \in \operatorname{Im} A_{I_0}^\top$ et donc qu'il existe λ tel que

   $$\nabla f(x) = -A_{I_0}^\top \lambda,$$

   et qui s'obtient par résolution du système linéaire

   $$R\lambda = -U^\top \nabla f(x),$$

   et il faut ensuite considérer deux cas

   (a) Si $\lambda \geq 0$, alors x satisfait les condition de Kuhn et Tucker. Le point x est donc un optimum local du problème.

   (b) Sinon, on supprime dans I_0 une des contraintes pour lesquelles $\lambda_i < 0$ (par exemple la plus négative). On obtient alors une nouvelle matrice A_1 qui permet de déterminer une nouvelle direction de descente en x_0. On peut ensuite poursuivre les itérations.

On peut donc résumer l'algorithme de la façon suivante :

Algorithme du gradient projeté (contraintes d'inégalité)

1. Poser $k = 0$ et choisir x_0.
2. Déterminer $I_k = \{i \mid A_i x_k - b_i = 0\}$.
3. Former la matrice $A_{I_k} = \{A_i\}_{i \in I_k}$.
4. Calculer ou mettre à jour la factorisation $A_{I_k}^\top = [U_k \;\; V_k] \begin{pmatrix} R_k \\ 0 \end{pmatrix}$
5. Calculer la projection $d_k = -V_k V_k^\top \nabla f(x_k)$
6. Si $d_k = 0$
   - Calculer $\lambda = -(R_k)^{-1} U_k^\top \nabla f(x_k)$
   - Si $\lambda \geq 0$ alors on s'arrête
   - Sinon, choisir j tel que $\lambda_j \leq \lambda_i, \forall i$, faire $I_k = I_k - \{j\}$ et retourner en 3.
7. Calculer $\rho_{max} = \{\rho \,|\rho \geq 0, \; A_i(x_k + \rho d_k)a - b_i \leq 0, \; i \notin I_k\}$.
8. Déterminer ρ_k réalisant le minimum de $f(x_k + \rho d_k)$ sur $[0, \rho_{max}]$.
9. Poser $x_{k+1} = x_k + \rho_k d_k$, faire $k \leftarrow k + 1$ et retourner en 2.

VII.3 Méthodes de pénalisation

VII.3.1 Méthode de pénalisation externe

Exemples :
Exemple VII.1

On considère un problème avec contraintes d'inégalité non-linéaires :

$$(PCI)\ \begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x), & \text{(VII.3.1)} \\ \text{sous les contraintes} \\ g(x) \leq 0, & \text{(VII.3.2)} \end{cases}$$

Le but des méthodes de pénalisation est de résoudre (PCI) de façon approchée de la façon suivante : on définit la fonction $\varphi(x)$ par

$$\varphi(x) = \sum_{i=1}^{m} (g_i^+(x))^2,$$

où $[.]^+$ est la fonction *partie positive* définie par

$$y^+ = \max(0, y).$$

Si on note $K = \{x \in \mathbb{R}^n,\ g(x) \leq 0\}$, la fonction φ vérifie par construction

$$\begin{cases} \varphi(x) = 0, \text{ pour } x \in K, \\ \varphi(x) > 0, \text{ pour } x \notin K. \end{cases}$$

On introduit alors le problème P_ϵ

$$(P_\epsilon)\ \begin{cases} \min\limits_{x\in\mathbb{R}^n} f_\epsilon(x), & \text{(VII.3.3)} \\ f_\epsilon(x) = f(x) + \frac{1}{\epsilon}\varphi(x), & \text{(VII.3.4)} \end{cases}$$

dont on notera x_ϵ la solution, vérifiant

$$f_\epsilon(x_\epsilon) \leq f_\epsilon(x)\ \forall x \in \mathbb{R}^N.$$

Le nom de pénalité *extérieure* provient du fait que x_ϵ est toujours à l'extérieur (au sens large) de K comme le montre le résultat suivant :

Proposition VII.3.1. *S'il existe au moins une contrainte saturée à l'optimum $\hat{x}$ du problème (PCI) alors le vecteur solution du problème pénalisé (P_ϵ) verifie nécessairement*

$$\exists\, i_0,\ g_{i_0}(x_\epsilon) \geq 0.$$

Démonstration : Montrons la contraposée : si $g_i(x_\epsilon) < 0,\ \forall i$ on a par définition $x_\epsilon \in K$. Puisque

$$f_\epsilon(x_\epsilon) \leq f_\epsilon(x),\ \forall\, x \in \mathbb{R}^n,$$

donc en particulier pour $x = \hat{x}$, on a

$$f_\epsilon(x_\epsilon) \leq f_\epsilon(\hat{x}),$$

mais commme $x_\epsilon \in K$ et $\hat{x} \in K$ on a

$$\varphi(x_\epsilon) = \varphi(\hat{x}) = 0,$$

et donc

$$f(x_\epsilon) \leq f(\hat{x}).$$

D'où $x_\epsilon = \hat{x}$. On a donc $g_i(\hat{x}) < 0$, $\forall i$ et aucune contrainte n'est saturée en $\hat{x}$. □ En général, on a toujours $x_\epsilon \notin K$ comme le montre l'**??** mais sous des hypothèses assez peu restrictives, x_ϵ tend vers une solution du problème (PCI) quand ϵ tend vers 0.

Théorème VII.3.1. *Soit $\varphi : \mathbb{R}^n \to \mathbb{R}$ une fonction de pénalisation extérieure vérifiant :*

- $\varphi(x) \geq 0$,
- $\varphi(x) = 0 \Leftrightarrow x \in K$,
- φ *continue.*

On suppose d'autre part que f est continue, que K est fermé et que l'une des deux conditions suivantes est vérifieé :

- $f(x) \to +\infty$ *quand* $\|x\| \to \infty$,
- K *est borné et* $\varphi(x) \to +\infty$ *quand* $\|x\| \to \infty$.
- φ *continue.*

Alors, quand ϵ_k tend vers 0, la suite x_{ϵ_k} admet au moins un point d'accumulation qui est alors une solution optimale du problème (PCI).

Lorsqu'on met en oeuvre cette méthode de façon pratique, on ne peut pas prendre tout de suite ϵ_k très petit, à cause des problèmes de conditionnement que cela peut causer. On commence donc avec une valeur du type $\epsilon_0 = 1$, et chaque solution x_{ϵ_k} est prise comme vecteur initial pour résoudre le problème avec $\epsilon_{k+1} = \epsilon_k/100$ (par exemple). On peut bien sûr utiliser n'importe quelle méthode pour résoudre le problème $\min_x f_{\epsilon_k}(x)$ (BFGS, gradient conjugué, ...).

Algorithme de la méthode de pénalisation

1. Choisir x_0, $\epsilon_1 = 1$ et poser $k = 1$
2. Trouver x_k solution du problème $\min\limits_{x \in \mathbb{R}^n} f_{\epsilon_k}(x)$ en partant de x_{k-1}.
3. Poser $\epsilon_{k+1} = \epsilon_k/100$
4. faire $k \leftarrow k + 1$ et retourner en 2

VII.3.2 Méthode de pénalisation interne

Dans le cas des méthodes internes, en général, x_ϵ n'est jamais dans K (sauf cas particulier) : cela peut poser des problèmes si par exemple la fonction f n'est pas définie hors de K. Les méthodes internes permettent d'éviter cet inconvénient. Leur principe est le même que pour les méthodes externes : on considère une fonction

$$f_\epsilon(x) = f(x) + \epsilon\psi(x),$$

mais ici la fonction $\psi(x)$ est défine pour $x \in K$ et est du type

$$\psi(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)^2}.$$

Puisque l'on a $\psi(x) \to \infty$ quand on s'approche de la frontière de K, on qualifie souvent ψ de fonction *barrière*. Les propriété de convergence sont les même que pour les méthodes externes, mais il faut ici disposer d'un $x_0 \in K$, ce qui peut être difficile dans certains cas.

VII.3.3 Estimation des multiplicateurs

Les méthodes de pénalisation ne sont en général jamais utilisées pour obtenir la solution du problème avec contraintes, car cela nécessitérait d'utiliser des paramètres de pénalisation beaucoup trop petits. En revanche, elles permettent de calculer des estimations correctes des multiplicateurs.

Pour les méthodes externes, le point x_k est solution du problème $\min f_{\epsilon_k}(x)$ où

$$f_\epsilon(x) = f(x) + \frac{1}{\epsilon}\sum_{i=1}^{m}[g_i^+(x)]^2,$$

et vérifie donc les conditions d'optimalité

$$\nabla f(x_k) + \frac{2}{\epsilon}\sum_{i=1}^{m} g_i^+(x_k)\nabla g_i(x_k) = 0.$$

Sous les hypothèses du théorème VII.3.1 $x_k \to \hat{x}$ et donc pour les contraintes non saturées, puisque $g_i(\hat{x}) < 0$, il existe k_0 tel que

$$k > k_0 \Rightarrow g_i(x_k) < 0,\ i \notin I(\hat{x}).$$

Si on suppose que $\hat{x}$ est régulier, les conditions de Kuhn et Tucker sont vérifiées et on a

$$\nabla f(\hat{x}) + \sum_{i\in I}\lambda_i \nabla g_i(\hat{x}) = 0.$$

Si on note maintenant que pour $k > k_0$,

$$\nabla f(x_k) + \frac{2}{\epsilon}\sum_{i\in I} g_i^+(x_k)\nabla g_i(x_k) = 0,$$

alors par continuité de ∇f et ∇g on en déduit que pour $i \in I$

$$\lim_{k\to\infty}\frac{2}{\epsilon}g_i^+(x_k) = \lambda_i.$$

On peut bien sûr faire le même type de raisonnement pour la méthode de pénalité interne.

VII.4 Méthodes par résolution des équations de Kuhn et Tucker

VII.4.1 Cas des contraintes d'égalité

On cherche à résoudre le problème :

$$\begin{array}{l} \min\limits_{x\in\mathbb{R}^n} f(x), \\ h_i(x) = 0,\ i = 1 \dots p \end{array} \tag{VII.4.1}$$

On sait que la recherche d'un point de Kuhn et Tucker revient à résoudre le système à $n+p$ inconnues et $n+p$ inconnues

$$\left\{ \begin{array}{rcl} \nabla_x L(x,\lambda) & = & 0, \\ h(x) & = & 0, \end{array} \right. \tag{VII.4.2}$$

où on a noté $L(x,\lambda) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x)$ le lagrangien associé à (VII.4.1). La méthode de Newton consiste, à partir d'un point (x_k, λ_k), à linéariser (VII.4.2) au voisinage de ce point, et à définir (x_{k+1}, λ_{k+1}) comme la solution du système obtenu. On peut écrire les équations suivantes :

$$\begin{array}{rcl} \nabla_x L(x_k,\lambda_k) + \nabla_x^2 L(x_k,\lambda_k)(x_{k+1}-x_k) + \nabla h(x_k)(\lambda_{k+1}-\lambda_k) & = & 0, \\ h(x_k) + \nabla h(x_k)^\top (x_{k+1}-x_k) & = & 0, \end{array}$$

où $\nabla_x L(x_k,\lambda_k) = \nabla f(x_k) + \nabla h(x_k)\lambda_k$. Si on pose

$$J_k = \nabla h(x_k)^\top = \frac{\partial h}{\partial x}(x_k),$$

et $H_k = \nabla_x^2 L(x_k,\lambda_k)$, on obtient le système

$$\left(\begin{array}{cc} H_k & J_k^\top \\ J_k & 0 \end{array} \right) \left(\begin{array}{c} x_{k+1}-x_k \\ \lambda_{k+1} \end{array} \right) = \left(\begin{array}{c} -\nabla f(x_k) \\ -h(x_k) \end{array} \right). \tag{VII.4.3}$$

Une méthode basée sur la résolution itérative de (VII.4.3) présentera les inconvénients habituels de la méthode de Newton : la convergence est locale. De plus, les équations de Kuhn et Tucker sont aussi vérifiées pour les maximums. Si on veut remédier à ces inconvénients il faut diposer d'une bonne estimation initiale de $(\hat{x}, \hat{\lambda})$, qui peut par exemple être fournie par une méthode de pénalisation.

VII.4.2 Méthode de Wilson

Dans la méthode précédente, pour éviter les points stationnaires qui ne sont pas des minimum, on peut faire l'analyse suivante : si on note $s_k = x_{k+1} - x_k$ on observe que le système (VII.4.3) s'écrit

$$H_k y_k + J_k^\top \lambda_{k+1} = -\nabla f(x_k).$$

Le vecteur y_k est la solution du problème d'optimisation quadratique suivant :

$$\begin{cases} \min_y \frac{1}{2} y^\top H_k y + \nabla f(x_k)^\top y, \\ J_k y + h(x_k) = 0, \end{cases} \tag{VII.4.4}$$

et λ_{k+1} est le multiplicateur associé. Au lieu de résoudre le système (VII.4.3) on peut donc résoudre le problème (VII.4.4), ce qui permet d'éviter les points stationnaires qui ne sont pas des minima. La résolution de ce problème peut se faire avec toute méthode adaptée aux problèmes quadratiques. Cette extension de la méthode de Newton est due à Wilson.

VII.4.3 Cas des contraintes d'inégalité

La méthode de Wilson vue au grain précédent se généralise très facilement au cas des contraintes d'inégalité. Si le problème original est de la forme :

$$\begin{array}{l} \min\limits_{x\in\mathbb{R}^n} f(x), \\ g_i(x) \leq 0,\ i = 1 \dots m, \end{array} \tag{VII.4.5}$$

les contraintes linéarisées prennent la forme

$$\nabla g(x_k)^\top y + g(x_k) \leq 0.$$

On peut alors utiliser une méthode consistant à résoudre itérativement le problème quadratique

$$\left\{ \begin{array}{r} \min_y \frac{1}{2} y^\top H_k y + \nabla f(x_k)^\top y, \\ J_k y + g(x_k) \leq 0, \end{array} \right. \tag{VII.4.6}$$

Remarque VII.4.1. *Comme on l'a déjà dit la méthode de Wilson (pour les contraintes d'égalité et d'inégalité) ne converge que localement. La globalisation de cette méthode peut se faire en utilisant une approximation de quasi-Newton pour la matrice $H_k = \nabla^2_x L(x_k, \lambda_k)$ et en faisant une recherche linéaire dans la direction s_k pour définir $x_{k+1} = x_k + \rho_k s_k$. Lors de la recherche linéaire, on cherche alors à minimiser une* fonction de mérite *du type*

$$\theta(x) = f(x) + c \sum_{k=1}^{p} |h_i(x)|,$$

dans le cas des contraintes d'égalité, ou

$$\sigma(x) = f(x) + c \sum_{k=1}^{m} g_i^+(x),$$

dans le cas des contraintes d'inégalité (dans ce dernier cas c doit être un majorant des multiplicateurs optimaux). Les fonctions $\sigma(x)$ et $\theta(x)$ sont des fonctions de pénalisation exacte : cette terminologie traduit le fait que contrairement aux fonctions de pénalisation différentiables *que l'on a vu précédemment, le minimum de θ ou σ peut coïncider avec $\hat{x}$ pour des valeurs* finies *de c.*

Exemples du chapitre VII

Exemple VII.1 Un problème pénalisé

On considère le problème

$$\begin{cases} \min \frac{1}{2}x^2, \\ x \geq 1. \end{cases}$$

La fonction pénalisée s'écrit

$$f_\epsilon(x) = \frac{1}{2}x^2 + \frac{1}{\epsilon}([1-x]^+)^2.$$

Pour $x \notin K$ on a

$$\nabla f_\epsilon(x) = x - \frac{2}{\epsilon}(1-x).$$

Si on fait l'hypothèse a priori que $x_\epsilon \notin K$ alors on a

$$x_\epsilon - \frac{2}{\epsilon}(1-x_\epsilon) = 0,$$

et donc $x_\epsilon = (1+\epsilon/2)^{-1}$. On a bien $x_\epsilon \notin K$ et

$$\lim_{\epsilon \to 0} x_\epsilon = 1.$$

Chapitre VIII
Méthodes utilisant la notion de dualité

VIII.1 Elements sur la dualité

VIII.1.1 Le problème dual

On s'intéresse ici aux problèmes avec contrainte d'inégalité du type

$$\begin{array}{l} \min\limits_{x\in\mathbb{R}^n} f(x), \\ g(x) \leq 0, \end{array} \tag{VIII.1.1}$$

et on note comme d'habitude $K = \{x \in \mathbb{R}^n,\ g(x) \leq 0\}$. Le problème (VIII.1.1) est appellé *problème primal* par opposition au *problème dual* que l'on va maintenant définir.

Soit $\varphi(x)$ une fonction indicatrice de K :

$$\varphi(x) = 0, \text{ si } x \in K, \tag{VIII.1.2}$$

$$\varphi(x) = +\infty, \text{ sinon.} \tag{VIII.1.3}$$

Alors le problème primal est équivalent à

$$\min_{x\in\mathbb{R}^n} f(x) + \varphi(x).$$

On peut construire la fonction φ de la façon suivante :

$$\varphi(x) = \max_{\lambda\geq 0} \lambda^\top g(x) = \max_{\lambda\geq 0} \sum_{i=1}^{m} \lambda_i g_i(x).$$

On peut vérifier que la fonction ainsi définie a bien les caractéristiques données par (VIII.1.2)-(VIII.1.3) : si $x \in K$ on a $g_i(x) \leq 0$ et donc $\lambda^\top g(x) \leq 0$, le max est donc atteint pour $\lambda = 0$. Si $x \notin K$ il existe j tel que $g_j(x) > 0$, et donc $\lambda^\top g(x)$ peut être rendu arbitrairement grand en faisant tendre λ_j vers $+\infty$.

Le problème primal est donc équivalent au problème

$$\min_{x\in\mathbb{R}^n} \left(f(x) + \max_{\lambda\geq 0} \lambda^\top g(x) \right),$$

et si on utilise le lagrangien $L(x, \lambda) = f(x) + \lambda^\top g(x)$, on peut alors noter que le problème primal s'écrit

$$\min_{x\in\mathbb{R}^n} \max_{\lambda\geq 0} L(x, \lambda). \tag{VIII.1.4}$$

Définition VIII.1.1. *On appelle problème dual du problème (VIII.1.1) le problème*

$$\max_{\lambda\geq 0} \min_{x\in\mathbb{R}^n} L(x, \lambda), \tag{VIII.1.5}$$

et appelle $w(\lambda) = \min_{x\in\mathbb{R}^n} L(x, \lambda)$ *la fonction duale.*

Proposition VIII.1.1. *La fonction duale* $w(\lambda)$ *est concave.*

Démonstration : Soient $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\theta \in [0, 1]$ et $\lambda = \theta\lambda_1 + (1 - \theta)\lambda_2$. Il existe x_1 ,x_2 et x tels que

$$\begin{aligned} w(\lambda_1) &= L(x_1, \lambda_1), \\ w(\lambda_2) &= L(x_2, \lambda_2), \\ w(\lambda) &= L(x, \lambda). \end{aligned}$$

On a donc par définition de la fonction duale :

$$\begin{aligned} w(\lambda_1) &\leq L(x, \lambda_1), \\ w(\lambda_2) &\leq L(x, \lambda_2). \end{aligned}$$

Si on multiplie la première inéquation par θ et la deuxième par $(1-\theta)$ il vient

$$\theta w(\lambda_1) + (1-\theta) w(\lambda_2) \leq f(x) + [\theta\lambda_1 + (1-\theta)\lambda_2]^\top g(x) = w(\lambda).$$

□ Ce qui est remarquable dans cette propriété est que le résultat ne suppose absolument rien sur la convexité des fonctions f et g_i.

VIII.1.2 Point-col du lagrangien

On montre facilement la proposition suivante :

Proposition VIII.1.2. *On a*

$$\max_{\lambda\geq 0}\left\{\min_{x\in\mathbb{R}^n} L(x,\lambda)\right\} \leq \min_{x\in\mathbb{R}^n}\left\{\max_{\lambda\geq 0} L(x,\lambda)\right\}.$$

Démonstration : On a $L(x,\lambda)\leq \max_{\lambda\geq 0} L(x,\lambda)$ et donc par définition de $w(\lambda)$

$$w(\lambda)\leq \min_{x\in\mathbb{R}^n}\max_{\lambda\geq 0} L(x,\lambda).$$

On a donc

$$\max_{\lambda\geq 0} w(\lambda)\leq \min_{x\in\mathbb{R}^n}\max_{\lambda\geq 0} L(x,\lambda),$$

ce qui montre le résultat. □ Si l'on note que par construction

$$\min_{x\in\mathbb{R}^n}\max_{\lambda\geq 0} L(x,\lambda) = f(\hat{x}),$$

où $\hat{x}$ est la solution du problème primal, on a donc

$$\max_{\lambda\geq 0} w(\lambda)\leq f(\hat{x}).$$

Alors s'il existe bien un maximum de la fonction duale atteint pour $\lambda=\bar{\lambda}$, la valeur $w(\bar{\lambda})$ est un minorant de $f(\hat{x})$ et il existe un point $x(\bar{\lambda})$ tel que

$$w(\bar{\lambda}) = L(x(\bar{\lambda}),\bar{\lambda})\leq f(\hat{x}).$$

Le théorème suivant précise dans quelles conditions on a $x(\bar{\lambda})=\hat{x}$:

Théorème VIII.1.2. *S'il existe un couple $(\hat{x},\hat{\lambda})$ tel que*

$$L(\hat{x},\lambda)\leq L(\hat{x},\hat{\lambda})\leq L(x,\hat{\lambda}),\ \forall x\in\mathbb{R}^n,\ \forall\lambda\in\mathbb{R}^m,$$

alors $\hat{x}$ est une solution du problème primal et $\hat{\lambda}$ est le multiplicateur de Kuhn et Tucker associé.

Un point vérifiant cette propriété est appelé un *point-col* du lagrangien. On a dans ce cas

$$L(\hat{x},\hat{\lambda}) = \max_{\lambda\geq 0} w(\lambda) = \min_{x\in K} f(x).$$

Lorsque ce point existe, on peut donc résoudre le problème dual à la place du problème primal : l'intérêt principal est la concavité de la fonction duale ainsi que la simplicité des contraintes. On voit aussi que même lorsqu'il n'existe pas de point col, le maximum de la fonction duale fournit un minorant de $f(\hat{x})$, ce qui peut être utile dans certaines circonstances. On appelle alors la différence $f(\hat{x})-w(\hat{\lambda})$ le *saut de dualité*.

Théorème VIII.1.3. *Si f est strictement convexe, si les g_i sont convexes et si K est d'intérieur non-vide, l'existence de $\hat{x}$ est équivalente à l'existence de $\hat{\lambda}$ et on a*

$$w(\hat{\lambda}) = L(\hat{x},\hat{\lambda}) = f(\hat{x}).$$

Il existe cependant des cas où il existe un point-col et les conditions précédentes ne sont pas vérifiées. Quand il n'y a pas de point-col, on peut faire alors appel à des techniques où on utilise un lagrangien *augmenté* du type

$$L(x,\lambda,r) = f(x)+\lambda^\top g(x)+r\sum_{i=1}^{m}(g_i^+(x))^2,$$

pour définit la fonction duale. Ce type d'approche permet de généraliser les méthodes duales pour les cas typiquement non-convexes.

VIII.2 Methodes duales

VIII.2.1 Méthode d'Uzawa

Le principe de la méthode d'Uzawa est d'utiliser la méthode du gradient pour maximiser la fonction duale, tout en tenant compte de la contrainte $\lambda \geq 0$: cela donne la méthode

$$\lambda_{k+1} = [\lambda_k + \rho_k \nabla w(\lambda_k)]^+ .$$

L'utilisation de cette méthode suppose que la fonction duale est différentiable (au moins a l'optimum). Ce sera le cas si le minimum en x de $L(x, \hat{\lambda})$ est unique. Dans ce cas si on note $x(\lambda)$ le vecteur tel que

$$w(\lambda) = L(x(\lambda), \lambda),$$

on peut écrire que

$$\begin{aligned} \nabla w(\lambda) &= \nabla_x L(x(\lambda), \lambda) \frac{dx(\lambda)}{d\lambda} + \nabla_\lambda L(x(\lambda), \lambda), \\ &= g(x(\lambda)), \end{aligned}$$

puisque $x(\lambda)$ est par définition le minimum en x de $L(x, \lambda)$. L'algorithme de la méthode est donc le suivant :

Algorithme d'Uzawa

1. Poser $k = 0$ et $\lambda_0 = 0$.
2. Déterminer x_k solution du problème $\min_{x \in \mathbb{R}^n} f(x) + \lambda_k^\top g(x)$
3. Si $\max_i g_i(x_k) < \epsilon$ alors on s'arrête.
4. Sinon, calculer $\lambda_{k+1} = [\lambda_k + \rho_k g(x_k)]^+$
5. Faire $k \leftarrow k + 1$ et retourner en 2.

Au point 4 on peut choisir ρ_k fixe ou bien faire une recherche linéaire. Lorsque la fonction duale est mal conditionnée, on peut aussi utiliser une méthode de quasi-Newton. Dans le test d'arrêt choisi la valeur de $\epsilon > 0$ devra être choisie prudemment : en effet, s'il n'existe pas de point-col on ne peut avoir $x_k \in K$ et donc si ϵ est trop petit l'algorithme ne s'arrêtera pas.

VIII.2.2 Méthode d'Arrow et Hurwicz

Cette méthode est très voisine de la méthode d'Uzawa. Au lieu de déterminer x_k comme le minimum de $L(x, \lambda_k)$ on se contente d'un pas dans la direction $-\nabla_x L(x, \lambda_k)$: on définit x_{k+1} par

$$x_{k+1} = x_k - \alpha_k \nabla_x L(x_k, \lambda_k),$$

et λ_{k+1} par

$$\lambda_{k+1} = [\lambda_k + \rho_k g(x_k)]^+ .$$

Index des concepts

Le gras indique un grain où le concept est défini ; l'italique indique un renvoi à un exercice ou un exemple, le gras italique à un document, et le romain à un grain où le concept est mentionné.

M

P

R

T

U

Index des notions