

Sciences du numérique  
**Optimisation numérique : aspects théoriques et algorithmes**



O. Cots & J. Gergaud & S. Gratton & D. Ruiz

5 octobre 2023

# Table des matières

<b>1</b>	<b>Outils mathématiques</b>	<b>3</b>
I	Introduction . . . . .	3
II	Algèbre linéaire . . . . .	3
II.1	rappels . . . . .	3
III	Calcul différentiel . . . . .	4
III.1	Notations . . . . .	4
III.2	Théorème des fonctions composées . . . . .	5
III.3	Formule de Taylors . . . . .	5
III.4	Courbes de niveau . . . . .	6
III.5	Surfaces et plan tangent dans $\mathbf{R}^3$ . . . . .	7
IV	Convexité des applications . . . . .	7
IV.1	Ensembles convexes - applications convexes . . . . .	7
IV.2	Convexité et dérivée première . . . . .	8
IV.3	Convexité et dérivée seconde . . . . .	9
IV.4	Illustrations . . . . .	10
V	Exercices . . . . .	10
V.1	Avec corrections . . . . .	10
<b>2</b>	<b>Existence de solution, unicité de solution</b>	<b>13</b>
I	Introduction . . . . .	13
II	Existence de solution . . . . .	13
II.1	Problèmes avec contraintes . . . . .	13
II.2	Problème sans contraintes . . . . .	14
III	Cas convexe . . . . .	14
<b>3</b>	<b>Problèmes avec contraintes</b>	<b>15</b>
I	Introduction . . . . .	15
II	Conditions du premier ordre . . . . .	16
II.1	Qualification des contraintes . . . . .	16
II.2	Théorème de Karuch, Kuhn et Tucker . . . . .	19
II.3	Cas convexe . . . . .	20
III	Conditions du second ordre . . . . .	21
III.1	Conditions Nécessaires du second ordre . . . . .	21
III.2	Conditions suffisantes . . . . .	22
IV	Exercices . . . . .	23
IV.1	Avec corrections . . . . .	23
<b>4</b>	<b>Algorithmes globalisés pour l'optimisation sans contrainte</b>	<b>25</b>
I	Algorithmes de minimisation sans contrainte . . . . .	25
I.1	La méthode de Newton . . . . .	25
I.2	Méthodes quasi-Newton . . . . .	26
I.3	Globalisation des méthodes de Newton/quasi-Newton . . . . .	28
I.4	Globalisation des moindres carrés non-linéaires . . . . .	33

<b>5</b>	<b>Quelques algorithmes pour l'optimisation avec contraintes</b>	<b>35</b>
I	Introduction . . . . .	35
II	Méthode des contraintes actives . . . . .	35
II.1	Multiplicateurs de Lagrange et sensibilité . . . . .	35
II.2	Application de la théorie des multiplicateurs de Lagrange : la méthode des contraintes actives . . . .	36
III	Pénalisation d'un problème quadratique à contraintes d'égalité . . . . .	38

# Introduction

Optimiser, c'est rechercher parmi un ensemble  $C$  de choix possibles le meilleur (s'il existe !). Si  $f$  est une application d'un ensemble  $E$  dans  $F$ . On note le problème

$$(P) \left\{ \begin{array}{l} \min f(x) \\ x \in C \subset E. \end{array} \right.$$

Il faut donc pour cela pouvoir comparer 2 choix et donc avoir une structure d'ordre sur l'ensemble  $F$ . On prendra toujours  $F = \mathbf{R}$ . Suivant les domaines d'applications :

- $E$  s'appelle l'ensemble des stratégies, des états, des paramètres, l'espace ;
- $C$  est l'ensemble des contraintes ;
- $f$  est la fonction coût, économique ou le critère, l'objectif.

Une fois le problème bien défini, il se pose deux questions. La première est de savoir si  $(P)$  admet une solution. Si la réponse est positive, il nous faut trouver la ou les solutions. Suivant la nature de l'ensemble  $E$  les réponses sont plus ou moins faciles. Si  $E$  est fini, l'existence de solution est évidente, mais le calcul est difficile si le nombre d'éléments est grand. Par contre si  $E = \mathbf{R}^n$  ou est de dimension infinie la question de l'existence de solution est moins triviale, mais si les fonctions sont dérivables il est "plus" facile de calculer la solution.



# Chapitre 1

## Outils mathématiques

### I Introduction

Les outils nécessaires pour l'optimisation continue sont :

- l'algèbre linéaire ;
- le calcul différentiel ;
- l'analyse convexe.

Nous supposons que les bases d'algèbre linéaire et de calcul différentiel sont connus. Faisons cependant quelques rappels afin de bien préciser les notations. Nous reviendrons aussi sur les formes quadratiques et leur représentation géométrique pour le cas où l'espace de départ est le plan  $\mathbf{R}^2$ . Concernant l'analyse convexe, il ne s'agira que d'une introduction extrêmement sommaire, ce sujet à lui seul pourrait faire l'objet d'un cours complet[5, 2, 3]... Enfin, dans la dernière section de ce chapitre, nous verrons une application du théorème des fonctions implicites qui nous permettra de définir proprement les courbes de niveaux ainsi que de visualiser le gradient d'une fonction  $f(x)$  définie sur  $\mathbf{R}^2$  et à valeur dans  $\mathbf{R}$ .

### II Algèbre linéaire

#### II.1 rappels

**Définition II.1.1** Soit  $A$  une matrice réelle symétrique.

- (i)  $A$  est dite semi-définie positive si et seulement si pour tout  $x$  dans  $\mathbf{R}^n$  on a  $(Ax|x) = x^T Ax \geq 0$ .
- (ii)  $A$  est dite définie positive si et seulement si pour tout  $x$  dans  $\mathbf{R}^n$  non nul on a  $(Ax|x) = x^T Ax > 0$ .

**Théorème II.1.2**

- (i)  $A$  est semi-définie positive si et seulement si toutes les valeurs propres de  $A$  sont positives ou nulles.
- (ii)  $A$  est définie positive si et seulement si toutes les valeurs propres de  $A$  sont strictement positives.

#### Exercice II.1

On considère la fonction suivante

$$\begin{aligned} q : \mathbf{R}^n &\longrightarrow \mathbf{R} \\ x &\longmapsto \frac{1}{2}x^T Ax + b^T x + c, \end{aligned}$$

où  $A \in \mathcal{M}_n(\mathbf{R})$ .

**1** Montrer que l'on peut toujours supposer que  $A$  est symétrique, c'est-à-dire que pour toute matrice  $A$  il existe une matrice  $B$  symétrique telle que  $x^T Ax = x^T Bx$  pour tout  $x$ .

On supposera donc dans toute la suite que, dans l'expression d'une forme quadratique, la matrice  $A$  est symétrique.

**2** Vérifiez que

$$q(x) = \frac{1}{2} \sum_i a_{ii} x_i^2 + \sum_{i < j} a_{ij} x_i x_j + \sum_i b_i x_i + c.$$

**3** On suppose maintenant que  $A$  est de rang  $n$ .  $x$  représente les coordonnées d'un point  $M$  dans le repère canonique  $(O, e_1, \dots, e_n)$ .

(i) Montrer qu'il existe un point  $\Omega$  tel que dans le repère  $(\Omega, e_1, \dots, e_n)$  la fonction s'écrit

$$\tilde{f}(x') = \frac{1}{2} x'^T A x' + c'$$

(ii) Montrer qu'il existe un repère orthonormé  $(\Omega, f_1, \dots, f_n)$  dans lequel  $q$  s'écrit

$$\tilde{q}(y) = \varepsilon_1 \frac{y_1^2}{a_1^2} + \dots + \varepsilon_n \frac{y_n^2}{a_n^2} + d,$$

où  $\varepsilon_i = \pm 1$ .

(iii) On considère le cas  $n = 2$  et on pose

$$Q = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad A = Q \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} Q^T, \quad b = \begin{pmatrix} 1 & 2 \end{pmatrix},$$

donner les formes des courbes de niveaux de  $q$  pour

(a)  $\lambda_1 = 1, \lambda_2 = 3/2$ ;

(b)  $\lambda_1 = 1, \lambda_2 = -3/2$ ;

(c)  $\lambda_1 = -1, \lambda_2 = -3/2$ .

4 On suppose maintenant que  $n = 2$  et que le rang de  $A$  est 1, quelles sont les formes des courbes de niveaux pour :

(i)  $\lambda_1 = 1, \lambda_2 = 0$  et  $b = \begin{pmatrix} 1 & 2 \end{pmatrix}^T$ ;

(ii)  $\lambda_1 = 1, \lambda_2 = 0$  et  $b = \begin{pmatrix} \cos(\theta) & \sin(\theta) \end{pmatrix}^T$ .

5 On prend maintenant

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad b = \begin{pmatrix} -1 & -2 \end{pmatrix} \quad \text{et} \quad c = +1/2.$$

(i) Résoudre  $\nabla q(x) = 0$ .

(ii) Calculer  $q(0, 1)$ . En déduire l'ensemble  $C = \{x \in \mathbf{R}^2, q(x) = -1/2\}$ .

(iii) Calculer  $\nabla q(0, 1)$  et vérifier que ce vecteur est orthogonal à la courbe de niveau  $C$ .

(iv) Calculer  $q(1 + \frac{\sqrt{2}}{2}, \frac{3}{2})$  et  $\nabla q(1 + \frac{\sqrt{2}}{2}, \frac{3}{2})$  et vérifier que ce vecteur est orthogonal à la courbe de niveau  $C$ .

### III Calcul différentiel

#### III.1 Notations

##### Cas général

On considère une application  $f$  d'un ouvert  $\Omega$  d'un espace vectoriel normé  $E$  à valeurs dans un espace vectoriel normé  $F$ . On note  $f'(x)$  la dérivée de l'application  $f$  en  $x \in \Omega$ . On rappelle que  $f'(x) \in \mathcal{L}(E, F)$ .

##### Matrice jacobienne

Lorsque  $E = \mathbf{R}^n$  et  $F = \mathbf{R}^m$ , on peut identifier  $f'(x)$  avec sa matrice jacobienne  $J_f(x) \in \mathcal{M}_{m,n}(\mathbf{R})$ . On a alors

$$f'(x).h = J_f(x) \times h = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.$$

##### Gradient

Lorsque  $E = H$ , espace de Hilbert et  $F = \mathbf{R}$ ,  $f'(x)$  appartient au dual topologique de  $H$  que l'on peut identifier par le théorème de Riesz à  $H$ . On peut alors écrire

$$f'(x).h = (\nabla f(x)|h).$$

où  $\nabla f(x) \in H$  est le gradient de  $f$  en  $x$ .

Lorsque  $E = \mathbf{R}^n$ , muni du produit scalaire canonique, et  $F = \mathbf{R}$  on peut donc écrire indifféremment :

$$f'(x).h = J_f(x)h = (\nabla f(x)|h).$$

Dans ce cas, en particulier on a  $\nabla f(x) = J_f(x)^T$ .

Si on considère la définition de la dérivée on peut indifféremment écrire

$$\begin{aligned} f(x+h) &= f(x) + f'(x).h + o(\|h\|) \\ &= f(x) + J_f(x)h + o(\|h\|) \\ &= f(x) + (\nabla f(x)|h) + o(\|h\|) \\ &= f(x) + \nabla f(x)^T h + o(\|h\|), \end{aligned}$$

où pour  $p \geq 1$ ,  $o(\|h\|^p) = \|h\|^p \varepsilon(h)$  avec  $\varepsilon(h) \rightarrow 0$  quand  $h \rightarrow \vec{0}$ .

### Dérivée seconde

On considère une application  $f$  d'un ouvert  $\Omega$  d'un espace vectoriel normé  $E$  à valeurs dans un espace vectoriel normé  $F$ . La dérivée seconde  $f''(x)$  est un élément de  $\mathcal{L}(E, \mathcal{L}(E, F))$  que l'on peut identifier à l'ensemble des applications bilinéaires continues  $\mathcal{L}_2(E \times E, F)$ . De plus, si  $f$  est deux fois différentiable en  $x$ , l'application bilinéaire  $f''(x)$  est toujours symétrique.

### Matrice hessienne

Lorsque  $f : \Omega \subset \mathbf{R}^n \rightarrow \mathbf{R}$  on peut alors identifier la dérivée seconde de  $f$  en  $x$  avec une matrice symétrique, appelée la matrice hessienne de  $f$  en  $x$  et notée  $\nabla^2 f(x)$  ou  $H_f(x)$ . On a alors

$$f''(x).(h, k) = (\nabla^2 f(x)h|k) = h^T \nabla^2 f(x)k = h^T H_f(x)k,$$

avec pour tout  $i, j \in \{1, \dots, n\}$

$$[\nabla^2 f(x)]_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

#### Remarque III.1.1

Lorsque  $\mathbf{R}^n$  est muni du produit scalaire canonique on a

$$\nabla^2 f(x) = J_{\nabla f}(x).$$

## III.2 Théorème des fonctions composées

### Théorème III.2.1

Soient  $E, F$  et  $G$  trois espaces vectoriels normés. Soit  $f$  une fonction de  $E$  dans  $F$  dérivable en  $x_0 \in E$  et soit  $g$  une fonction de  $F$  dans  $G$  dérivable en  $y_0 = f(x_0)$ . Alors  $g \circ f$  est dérivable en  $x_0$  et on a

$$\forall h \in E, (g \circ f)'(x_0).h = g'(f(x_0)).(f'(x_0).h). \quad (1.1)$$

### Remarque III.2.2

Si  $E, F$  et  $G$  sont respectivement  $\mathbf{R}^n, \mathbf{R}^m$  et  $\mathbf{R}^p$  alors (1.1) est équivalent à

$$J_{g \circ f}(x_0) = J_g(f(x_0)) \times J_f(x_0).$$

## III.3 Formule de Taylors

### Théorème III.3.1 (Formule de Taylor-Young)

Soit  $f$  une application d'un ouvert  $\Omega$  d'un espace vectoriel normé  $E$  à valeurs dans un espace vectoriel normé  $F$  qui admet des dérivées d'ordre  $p-1$  dans  $\Omega$  et une dérivée d'ordre  $p$  en  $x_0$  alors on a le développement de Taylor-Young

$$f(x_0 + h) = f(x_0) + f'(x_0).h + \dots + \frac{f^{(p)}(x_0)}{p!}.h^p + o(\|h\|^p), \quad (1.2)$$

où  $h^p = \underbrace{(h, \dots, h)}_{p \text{ fois}}$  et  $o(\|h\|^p) = \|h\|^p \varepsilon(h)$ , avec  $\varepsilon(h)$  qui tend vers 0 lorsque  $h$  tend vers le vecteur nul.

### Remarque III.3.2

(i) Pour  $p = 1$ , on retrouve la définition de  $f'(x_0)$ .



(ii) Pour  $p = 2$ ,  $E = \mathbf{R}^n$  et  $F = \mathbf{R}$  on obtient la meilleure approximation quadratique de la fonction ; on peut en effet écrire indifféremment

$$f(x_0 + h) = f(x_0) + f'(x_0).h + \frac{1}{2}f''(x_0).(h, h) + o(\|h\|^2) \quad (1.3)$$

$$= f(x_0) + (\nabla f(x_0)|h) + \frac{1}{2}(H_f(x_0)h|h) + o(\|h\|^2) \quad (1.4)$$

$$= f(x_0) + J_f(x_0)h + \frac{1}{2}h^T H_f(x_0)h + o(\|h\|^2). \quad (1.5)$$

$$(1.6)$$

### III.4 Courbes de niveau

#### Théorème III.4.1 (Théorème des fonctions implicites)

Soient  $E$  un espace vectoriel normé,  $F$  et  $G$  deux espaces de Banach,  $\Omega$  un ouvert de  $E \times F$ ,  $f : \Omega \rightarrow G$  une application continue et  $(x_0, y_0)$  un point de  $\Omega$ . On pose  $f(x_0, y_0) = c$  et on fait les hypothèses suivantes.

(i)  $\frac{\partial f}{\partial y}(x, y)$  existe en tout point de  $\Omega$  ;

(ii) L'application

$$\frac{\partial f}{\partial y} : \Omega \longrightarrow \mathcal{L}(F, G)$$

est continue en  $(x_0, y_0)$  ;

(iii) L'application linéaire continue

$$\frac{\partial f}{\partial y}(x_0, y_0) : F \longrightarrow G$$

est un isomorphisme de  $F$  dans  $G$

Alors il existe un voisinage ouvert  $U \times V$  de  $(x_0, y_0)$  tel que l'équation  $f(x, y) = c$  admette, pour tout  $(x, y) \in U \times V$ , une unique solution  $y = \varphi(x) \in V$ . En outre, l'application  $\varphi : U \rightarrow V$  est continue.

Si de plus  $f$  est différentiable en  $(x_0, y_0)$  alors  $\varphi$  est différentiable en  $x_0$  et

$$\varphi'(x_0) = - \left[ \frac{\partial f}{\partial y}(x_0, y_0) \right]^{-1} \frac{\partial f}{\partial x}(x_0, y_0) \quad (1.7)$$

On considère ici une application  $f : \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$  de classe  $C^1$ . On suppose que  $f'(x_0, y_0) \neq 0$  et on pose  $c = f(x_0, y_0)$ . Alors localement en  $(x_0, y_0)$  l'ensemble  $\{(x, y) \in \mathbf{R}^2, f(x, y) = c\}$  est une courbe  $C$ . En effet, supposons pour fixer les idées que  $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$ , alors le théorème des fonctions implicites s'applique et donc il existe un ouvert  $U = ]x_0 - \eta, x_0 + \eta[$  et un ouvert  $V = ]y_0 - \eta, y_0 + \eta[$  tel que

$$(\forall (x, y) \in U \times V, f(x, y) = c) \iff y = \varphi(x),$$

où  $\varphi : U \rightarrow V$  est  $C^1$  et on a

$$\varphi'(x_0) = - \left[ \frac{\partial f}{\partial y}(x_0, y_0) \right]^{-1} \frac{\partial f}{\partial x}(x_0, y_0).$$

Si  $\frac{\partial f}{\partial y}(x_0, y_0) = 0$  alors  $\frac{\partial f}{\partial x}(x_0, y_0) \neq 0$  car  $\varphi'(x_0, y_0) \neq 0$  et on a localement que  $x = \psi(y)$ .

Un vecteur directeur de la tangente en  $x_0, y_0$  à la courbe  $C$  est  $v = (1 \quad \varphi'(x_0))^T$ . Par suite on a

$$\begin{aligned} (\nabla f(x_0, y_0)|v) &= \frac{\partial f}{\partial x}(x_0, y_0) + \frac{\partial f}{\partial y}(x_0, y_0)\varphi'(x_0) \\ &= \frac{\partial f}{\partial x}(x_0, y_0) + \frac{\partial f}{\partial y}(x_0, y_0) \left( - \left[ \frac{\partial f}{\partial y}(x_0, y_0) \right]^{-1} \frac{\partial f}{\partial x}(x_0, y_0) \right) \\ &= 0. \end{aligned}$$

En d'autres termes le gradient de  $f$  en  $(x_0, y_0)$  est orthogonal à la tangente.

### III.5 Surfaces et plan tangent dans $\mathbf{R}^3$

On considère une application  $\varphi : \mathbf{R}^2 \rightarrow \mathbf{R}$  de classe  $C^1$  et  $(x_0, y_0)$  un point fixé de  $\mathbf{R}^2$ . L'ensemble  $S = \{(x, y, z) \in \mathbf{R}^3, z = \varphi(x, y)\}$  est une surface de espace  $\mathbf{R}^3$ . Un vecteur de l'espace tangent à  $S$  en  $(x_0, y_0, z_0)$  est un vecteur  $v$  tel qu'il existe une fonction  $\gamma : ]-\varepsilon, \varepsilon[ \rightarrow S$  dérivable en 0 telle que  $\gamma(0) = (x_0, y_0, z_0)$  et

$$v = \frac{d\gamma}{dt}(0).$$

Si on considère par exemple

$$\begin{aligned} \gamma : ]-\varepsilon, \varepsilon[ &\longrightarrow \mathbf{R}^3 \\ t &\longmapsto \begin{pmatrix} x_0 + t \\ y_0 \\ \varphi(x_0 + t, y_0) \end{pmatrix}, \end{aligned}$$

nous obtenons le vecteur tangent  $v_1 = \begin{pmatrix} 1 & 0 & \frac{\partial \varphi}{\partial x}(x_0, y_0) \end{pmatrix}^T$ . De même  $v_2 = \begin{pmatrix} 0 & 1 & \frac{\partial \varphi}{\partial y}(x_0, y_0) \end{pmatrix}^T$  est un vecteur tangent à  $S$  en  $(x_0, y_0, z_0)$ . L'espace vectoriel tangent est donc au moins de dimension 2.

On considère maintenant le cas d'une application de classe  $C^1$

$$\begin{aligned} f : \mathbf{R}^2 \times \mathbf{R} &\longrightarrow \mathbf{R} \\ (x, y, z) &\longmapsto f(x, y, z), \end{aligned}$$

On se donne un point  $(x_0, y_0, z_0)$ , on note  $c = f(x_0, y_0, z_0)$  et on suppose que  $\frac{\partial f}{\partial z}(x_0, y_0, z_0) \neq 0$ . On peut alors, comme dans la sous-section précédente, appliquer le théorème des fonctions implicites. Par suite il existe un ouvert  $U \subset \mathbf{R}^2$  contenant  $(x_0, y_0)$ , un ouvert  $V$  de  $\mathbf{R}$  contenant  $z_0$  et une fonction  $\varphi : U \rightarrow V$  continue tels que

$$(\forall (x, y) \in U \times V, f(x, y) = c) \iff y = \varphi(x).$$

En d'autre terme localement l'ensemble des points vérifiant  $f(x, y, z) = c$  est une surface  $S$ .

Considérons maintenant un vecteur tangent  $v$  de  $S$  en  $(x_0, y_0, z_0)$ . Soit  $\gamma$  une application telle que  $v = \frac{d\gamma}{dt}(0)$ . Pour tout  $t$  suffisamment petit  $\gamma(t)$  appartient à  $S$ . On a donc  $f(\gamma(t)) = c$ . Par suite  $f'(\gamma(0)) \cdot v = f'(x_0, y_0, z_0) \cdot v = 0$  et donc  $v \in \ker f'(x_0, y_0, z_0)$ , ou encore le gradient  $\nabla f(x_0, y_0, z_0)$  est normal à l'espace tangent; cet espace est donc au plus de dimension 2.

Si nous revenons maintenant au tout premier cas de cette sous-section, on peut considérer  $f(x, y, z) = z - \varphi(x, y)$ . Les vecteurs  $v_1$  et  $v_2$  forment donc une base de l'espace tangent à  $S$  au point  $(x_0, y_0, z_0)$  (on vérifiera ici que le gradient  $\nabla f(x_0, y_0, z_0)$  est bien orthogonal à  $v_1$  et à  $v_2$ ).

## IV Convexité des applications

### IV.1 Ensembles convexes - applications convexes

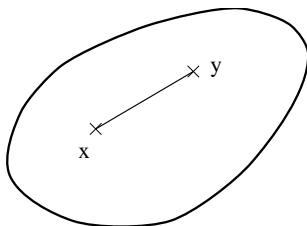
On indique dans ce paragraphe quelques propriétés de base d'une classe très importante des applications à valeurs dans  $\mathbb{R}$ .

#### Définition IV.1.1 Ensembles convexes

L'ensemble  $D_0$  est dit **convexe** si et seulement si

$$\forall \mathbf{x} \in D_0, \forall \mathbf{y} \in D_0, \forall \alpha \in [0, 1] \subset \mathbb{R} \text{ on a } \alpha \mathbf{x} + (1 - \alpha) \mathbf{y} \in D_0.$$

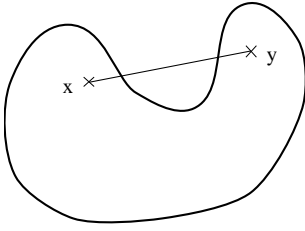
**Remarque :**



autrement dit, si  $\mathbf{x} \in D_0$  et  $\mathbf{y} \in D_0$ , alors le segment qui joint ces deux points est également contenu dans  $D_0$ , le segment  $[\mathbf{x}, \mathbf{y}]$  étant défini par

$$\mathbf{z} \in [\mathbf{x}, \mathbf{y}] \iff \exists \alpha \in [0, 1] \text{ t.q. } \mathbf{z} = \alpha \mathbf{x} + (1 - \alpha) \mathbf{y}.$$

**Exemple d'ensemble non convexe**



Remarque : la notion d'ensemble convexe correspond en fait à une propriété de régularité du domaine  $D_0$  considéré

### Définition IV.1.2 Applications convexes

Une application  $f : D_0 \subset E \rightarrow \mathbb{R}$  est **convexe** sur le domaine convexe  $D_0 \subset E$  ( $E$  espace vectoriel normé) si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \forall \alpha \in [0, 1], \quad f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}).$$

L'application  $f$  est **strictement convexe** sur le domaine convexe  $D_0$  si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \mathbf{x} \neq \mathbf{y}, \forall \alpha \in ]0, 1[, \quad f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) < \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}).$$

L'application  $f$  est **uniformément convexe** sur le domaine convexe  $D_0$  si il existe une constante  $c > 0$  telle que

$$\begin{aligned} \forall \mathbf{x}, \mathbf{y} \in D_0, \forall \alpha \in [0, 1], \\ \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) - f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \geq c \alpha (1 - \alpha) \|\mathbf{x} - \mathbf{y}\|_E^2. \end{aligned}$$

### Remarques :

- (i) Il est clair que la *convexité uniforme* entraîne la *convexité stricte* qui à son tour entraîne la *convexité*.
- (ii) La convexité indique une certaine régularité de l'application. En dimension finie, par exemple, la convexité peut induire des propriétés de continuité (c.f. proposition suivante).

### Proposition IV.1.3

Soit  $f : D_0 \subset \mathbb{R}^n \rightarrow \mathbb{R}$  une application convexe sur l'ouvert convexe  $D_0 \subset \mathbb{R}^n$ . Alors  $f$  est continue sur  $D_0$ .

Les définitions de base de la convexité (large, stricte, ou uniforme) peuvent parfois s'avérer d'un emploi peu commode. Le but des paragraphes qui suivent est de mettre en avant des propriétés qui s'y rapportent, exploitant la différentiabilité d'une application, et plus faciles à manipuler.

## IV.2 Convexité et dérivée première

### Théorème IV.2.1

#### Caractérisation de la convexité à l'aide de la dérivée première

Soit  $\Omega \subset E$  un ouvert de l'espace vectoriel normé  $E$ , et  $D_0$  un sous-ensemble convexe inclus dans  $\Omega$ . On suppose que l'application  $f : \Omega \subset E \rightarrow \mathbb{R}$  est dérivable sur le sous-ensemble convexe  $D_0 \subset \Omega$ . On a alors :

- (i)  $f$  est convexe sur  $D_0$  si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f(\mathbf{y}) - f(\mathbf{x}) \geq f'(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}).$$

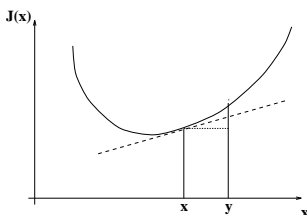
- (ii)  $f$  est strictement convexe sur  $D_0$  si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \mathbf{x} \neq \mathbf{y}, \quad f(\mathbf{y}) - f(\mathbf{x}) > f'(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}).$$

- (iii) L'application  $f$  est uniformément convexe sur  $D_0$  si et seulement si il existe une constante  $c > 0$  telle que

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f(\mathbf{y}) - f(\mathbf{x}) \geq f'(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + c \|\mathbf{y} - \mathbf{x}\|_E^2.$$

### Interprétation géométrique



L'interprétation géométrique de

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f(\mathbf{y}) - f(\mathbf{x}) \geq f'(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}),$$

est que le graphe de l'application convexe  $f$  est toujours au dessus de son plan tangent en un point quelconque du domaine  $D_0$ .

**Définition IV.2.2** Soit une application  $f : \Omega \subset E \rightarrow \mathbb{R}$  dérivable sur l'ouvert  $\Omega$ .

L'application dérivée  $f' : \Omega \subset E \rightarrow \mathcal{L}(E, \mathbb{R})$  est dite **monotone sur le sous-ensemble**  $D_0 \subset \Omega$  si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad (f'(\mathbf{y}) - f'(\mathbf{x})) \cdot (\mathbf{y} - \mathbf{x}) \geq 0.$$

L'application dérivée  $f'$  est dite **strictement monotone sur le sous-ensemble**  $D_0 \subset \Omega$  si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad \mathbf{x} \neq \mathbf{y}, \quad (f'(\mathbf{y}) - f'(\mathbf{x})) \cdot (\mathbf{y} - \mathbf{x}) > 0.$$

L'application dérivée  $f'$  est dite **fortement monotone sur le sous-ensemble**  $D_0 \subset \Omega$  si et seulement si il existe une constante  $c > 0$  telle que

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad (f'(\mathbf{y}) - f'(\mathbf{x})) \cdot (\mathbf{y} - \mathbf{x}) \geq 2c \|\mathbf{y} - \mathbf{x}\|_E^2.$$

### Proposition IV.2.3

#### Relations entre convexité et monotonie de la dérivée première

On suppose que l'application  $f : \Omega \subset E \rightarrow \mathbb{R}$  est dérivable sur l'ouvert  $\Omega$ . On a alors :

- (i) L'application  $f$  est convexe sur le sous-ensemble convexe  $D_0 \subset \Omega$  si et seulement si l'application dérivée  $f'$  est monotone sur  $D_0$ .
- (ii) L'application  $f$  est strictement convexe sur le sous-ensemble convexe  $D_0 \subset \Omega$  si et seulement si l'application dérivée  $f'$  est strictement monotone sur  $D_0$ .
- (iii) L'application  $f$  est uniformément convexe sur le sous-ensemble convexe  $D_0 \subset \Omega$  si et seulement si l'application dérivée  $f'$  est fortement monotone sur  $D_0$  (la constante  $c > 0$  intervenant dans la définition de la convexité uniforme correspondant à la constante  $c > 0$  introduite dans la définition de la forte monotonie de la dérivée).

## IV.3 Convexité et dérivée seconde

### Théorème IV.3.1

#### Relations entre convexité et positivité de la dérivée seconde

On suppose que l'application  $f : \Omega \subset E \rightarrow \mathbb{R}$  est deux fois dérivable dans un ouvert  $\Omega$  de l'espace vectoriel normé  $E$ , et soit  $D_0$  une partie convexe incluse dans  $\Omega$ .

- (i) L'application  $f$  est convexe sur le sous-ensemble convexe  $D_0 \subset \Omega$  si et seulement si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f''(\mathbf{x})(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}) \geq 0. \quad (1.8)$$

- (ii) Si

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad \mathbf{x} \neq \mathbf{y}, \quad f''(\mathbf{x})(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}) > 0, \quad (1.9)$$

alors l'application  $f$  est strictement convexe sur  $D_0$ .

- (iii) L'application  $f$  est uniformément convexe sur le sous-ensemble convexe  $D_0 \subset \Omega$  si et seulement si il existe une constante  $c > 0$  telle que

$$\forall \mathbf{x}, \mathbf{y} \in D_0, \quad f''(\mathbf{x})(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}) \geq 2c \|\mathbf{y} - \mathbf{x}\|_E^2. \quad (1.10)$$



La condition (1.9) ci-dessus n'est qu'une condition suffisante, la réciproque n'est valable qu'en dimension finie.

### Remarque IV.3.2

Si  $C = E = \mathbb{R}^n$  alors  $f''(x).(h, h) = h^T H_f(x)h$ . La relation (1.8) ( respectivement (1.9)) signifie que la matrice hessienne  $H_f(x)$  est semi-définie positive (respectivement définie positive).

### Corollaire IV.3.3

On considère un problème aux moindres carrée linéaire

$$(P) \begin{cases} \text{Min } f(\beta) = \frac{1}{2} \|y - X\beta\|^2 \\ \beta \in \mathbb{R}^p. \end{cases}$$

Alors  $f$  est convexe. Si de plus  $\text{rang}(X) = p$ , elle est alors strictement convexe.

#### Démonstration

On a  $\nabla f(\beta) = X^T X \beta - X^T y$  et  $H_f(\beta) = X^T X = H$ . Par suite la matrice hessienne est semi-définie positive car pour tout  $h \in \mathbb{R}^p$ , on a  $h^T X^T X h = (Xh | Xh) \geq 0$ . De plus si le rang de  $X$  est  $p$ , alors  $X^T X$  est aussi de rang  $p$  et donc elle est inversible. D'où sa définie positivité.  $\square$

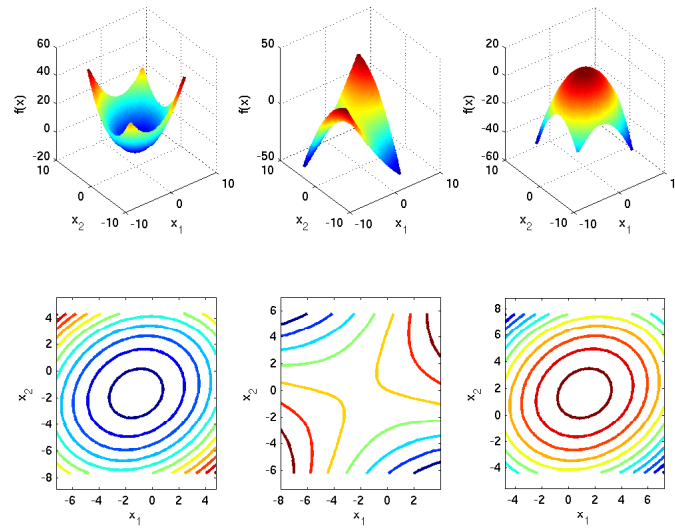
## IV.4 Illustrations

## V Exercices

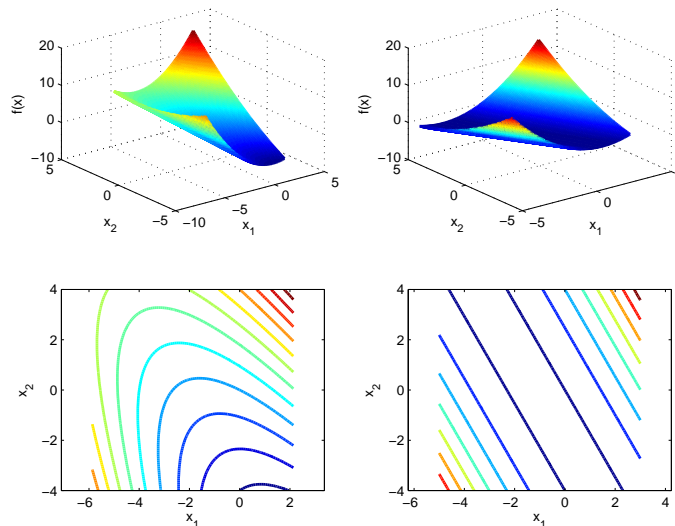
## V.1 Avec corrections

## Exercice V.1

## II.1

FIGURE 1.1 – Cas où  $\text{rang}(A) = 2$  et  $\theta = \pi/6$ .

1 (i)

FIGURE 1.2 – Cas où  $\text{rang}(A) = 1$  et  $\theta = \pi/6$ .

(ii)

**2** (i)  $\nabla q(x) = Ax + b = 0 \iff x^* = \begin{pmatrix} 1 & -1 \end{pmatrix}^T$ .

(ii)  $q(0, 1) = -1/2$ . Comme  $\lambda_1 > 0$  et  $\lambda_2 > 0$ , les courbes de niveaux sont des ellipses.  $\begin{pmatrix} 0 & -1 \end{pmatrix}^T$  appartient à cette ellipse. Par suite l'ensemble est l'unique ellipse de centre  $x^*$  d'axes  $(\Omega, e_1)$  et  $(\Omega, e_2)$  et de demi-axe  $a_1 = 1$  et  $a_2 = 1/\sqrt{2}$ .

(iii)  $\nabla q(0, 1) = \begin{pmatrix} -1 & 0 \end{pmatrix}^T$ . Ce vecteur est orthogonal à la tangente à courbe de niveau au point  $\begin{pmatrix} 0 & -1 \end{pmatrix}^T$ .



# Chapitre 2

## Existence de solution, unicité de solution

### I Introduction

Les problèmes d'optimisation où l'ensemble  $C$  est fini admettent toujours une solution, par contre, ceci n'est pas toujours le cas si  $C$  a un nombre infini d'éléments. Par exemple, le problème d'optimisation où la fonctionnelle à minimiser est  $f(x) = 1/x$  et l'ensemble des contraintes est  $C = \{x \in \mathbf{R}, x > 0\}$ , n'admet pas de solution. En effet  $f(x) > 0$  pour tout  $x$  dans  $C$  et pour tout  $\varepsilon > 0$ , il existe  $x > 1/\varepsilon$  tel que  $f(x) < \varepsilon$ . Il est donc préférable, avant de vouloir calculer la solution, de s'assurer que le problème en admet une.

### II Existence de solution

#### II.1 Problèmes avec contraintes

##### Théorème II.1.1

Soit  $(P)$  un problème d'optimisation avec contraintes  $C \subset E$ . Si  $f$  est continue et  $C$  est un compact non vide, alors le problème  $(P)$  admet une solution.

##### Démonstration

C'est une application immédiate du théorème qui dit que l'image d'un compact par une application continue dans un espace séparé est un compact.  $\square$

##### Exemple II.1.2

Considérons le problème suivant :

$$(P) \begin{cases} \text{Min } f(x) \\ x \in [0, 1] \end{cases}$$

où  $f$  est la fonction suivante :

$$\begin{aligned} f : [0, 1] &\longrightarrow \mathbf{R} \\ 0 &\longmapsto 1 \\ x &\longmapsto x \end{aligned}$$

Ce problème n'admet pas de solution. L'hypothèse du théorème (II.1.1) qui n'est pas vérifiée est la continuité de  $f$ .

##### Exemple II.1.3

Considérons le problème suivant :

$$(P) \begin{cases} \text{Min } f(x) = \frac{1}{x} \\ x \in [1, 5] \end{cases}$$

(i)  $f$  est continue ;

(ii)  $[1, 5]$  est un fermé et borné, donc un compact de  $\mathbf{R}$ .

Par suite ce problème admet une solution.

##### Exemple II.1.4

Considérons le problème suivant :

$$(P) \begin{cases} \text{Min } f(x) = \frac{1}{x} \\ x \in ]1, 5] \end{cases}$$

Ce problème a une solution, mais les hypothèses du théorème ne sont pas vérifiées.



**Exemple II.1.5**

Considérons le problème suivant :

$$(P) \begin{cases} \text{Min } f(x) = \frac{1}{x} \\ x \in [1, 5[ \end{cases}$$

Ce problème n'admet pas de solution,  $C = [1, 5[$  n'est pas fermé.

**II.2 Problème sans contraintes**

**Définition II.2.1** Une fonction  $f : E \rightarrow \mathbf{R}$ ,  $E$  espace vectoriel normé, est dite 0-coercive si et seulement si

$$f(x) \longrightarrow +\infty \text{ quand } \|x\| \longrightarrow +\infty. \quad (2.1)$$

**Théorème II.2.2**

Soit  $(P)$  un problème d'optimisation avec contraintes où  $f$  est une fonction de  $\mathbf{R}^n$  à valeurs dans  $\mathbf{R}$  et  $C$  est un fermé non vide non borné. Si  $f$  est continue et 0-coercive, alors le problème admet une solution.

*Démonstration*

Soit  $(x_k)_{k \in \mathbf{N}}$  une suite minimisante de points de  $C$ , c'est-à-dire une suite de point de  $C$  telle que  $\lim_{k \rightarrow +\infty} f(x_k) = \inf_{x \in \mathbf{R}^n} f(x) = \mu < +\infty$ . Montrons que cette suite est bornée. Sinon il existe une sous-suite  $(x_{n_k})_{n_k}$  telle que  $\|x_{n_k}\|$  tende vers  $+\infty$  lorsque  $n_k$  tend vers  $+\infty$  et donc, comme  $f$  est 0-coercive,  $\lim_{n_k \rightarrow +\infty} f(x_{n_k}) = +\infty$ , ce qui est impossible.

Par suite il existe un réel  $R > 0$  tel que la suite  $(x_k)_{k \in \mathbf{N}}$  soit contenue dans  $C \cap B_f(0, R)$  qui est un fermé borné de  $\mathbf{R}^n$ ; c'est donc un compact dont on peut extraire une sous-suite qui converge vers  $x^*$ . Mais  $f$  est continue, et donc  $f(x^*) = \mu$  et  $x^*$  est une solution du problème d'optimisation.  $\square$

**Remarque II.2.3**

Le théorème précédent s'applique si le problème d'optimisation est sans contraintes car dans ce cas  $C = E$ .

**III Cas convexe****Théorème III.0.1**

Si  $C$  est un convexe de  $E$  espace vectoriel normé et si  $f$  est une fonction de  $C$  à valeurs dans  $\mathbf{R}$  convexe, alors l'ensemble des solutions est vide où est un ensemble convexe de  $E$ .

*Démonstration*

Supposons que l'ensemble des solutions ne soit pas vide. Soient  $x$  et  $y$  deux solutions alors  $f(x) = f(y)$  car  $(f(x) \leq f(y)$  et  $f(y) \leq f(x))$ . Par suite, pour tout  $\alpha \in ]0, 1[$ , nous avons

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

En conséquence  $\alpha x + (1 - \alpha)y$  est aussi une solution.  $\square$

**Théorème III.0.2**

Si  $C$  est un convexe de  $E$  espace vectoriel normé et si  $f$  est une fonction de  $C$  à valeurs dans  $\mathbf{R}$  strictement convexe, alors il existe au plus un point  $x^*$  minimisant  $f$  sur  $C$ .

*Démonstration*

Supposons qu'il existe deux solutions  $x_1$  et  $x_2$ . Pour  $\alpha \in ]0, 1[$ , on pose  $x_\alpha = \alpha x_1 + (1 - \alpha)x_2$ , alors, puisque  $f$  est strictement convexe on a

$$f(x_\alpha) < \alpha f(x_1) + (1 - \alpha)f(x_2) = f(x_1) = f(x_2),$$

ce qui est impossible.  $\square$

**Théorème III.0.3**

Si  $C$  est un convexe de  $E$  espace vectoriel normé et si  $f$  est une fonction de  $C$  à valeurs dans  $\mathbf{R}$  convexe, alors tout minimum local  $x^*$  de  $f$  sur  $C$  est un minimum global de  $f$  sur  $C$ .

*Démonstration*

Soit  $x^*$  un minimum local de  $f$  sur  $C$ . Il existe donc  $\eta > 0$  tel que pour tout  $x \in C \cup B(x^*, \eta)$ ,  $f(x^*) \leq f(x)$ . Supposons maintenant qu'il existe dans  $C$  un point  $y$  tel que  $f(y) < f(x^*)$ . Alors, puisque  $f$  est convexe, on a pour tout  $\alpha \in ]0, 1[$

$$\begin{aligned} f(x^* + \alpha(y - x^*)) &= f((1 - \alpha)x^* + \alpha y) \leq (1 - \alpha)f(x^*) + \alpha f(y) \\ &< (1 - \alpha)f(x^*) + \alpha f(x^*) = f(x^*). \end{aligned}$$

Mais pour  $\alpha$  suffisamment proche de 0,  $x^* + \alpha(y - x^*) \in B(x^*, \eta)$ , d'où la contradiction.  $\square$

# Chapitre 3

## Problèmes avec contraintes

### I Introduction

On s'intéresse dans ce chapitre au problème d'optimisation en dimension finie avec des contraintes du type égalité et inégalité

$$(P) \begin{cases} \text{Min } f(x) \\ h(x) = 0 \\ g(x) \leq 0 \\ x \in \mathbf{R}^n, \end{cases}$$

où  $h$  (respectivement  $g$ ) est une application de  $\mathbf{R}^n$  à valeurs dans  $\mathbf{R}^p$  (respectivement  $\mathbf{R}^m$ ) et

$$g(x) \leq 0 \stackrel{\text{def}}{\iff} (g_j(x) \leq 0 \ \forall j = 1, \dots, m).$$

Dans le cas où il n'a pas de contraintes  $g(x) \leq 0$ , on dira que le problème est un problème du type égalité. On supposera toujours dans la suite que les fonctions  $f$ ,  $g$  et  $h$  sont dérivables autant de fois que nécessaires.

Avant d'étudier les conditions nécessaires et les conditions suffisantes de solutions faisons quelques remarques. Tout d'abord on peut toujours écrire un problème d'optimisation avec contraintes sous la forme d'un problème avec des contraintes uniquement de type inégalité. En effet une contrainte d'égalité  $h_i(x) = 0$  est équivalente aux contraintes d'inégalité  $h_i(x) \leq 0$  et  $-h_i(x) \leq 0$ . On peut aussi, en rajoutant des variables dites d'écart, toujours ramener un problème d'optimisation avec contraintes au cas où les contraintes du type inégalité s'écrivent simplement  $s \geq 0$ . En effet on a

$$g_j(x) \leq 0 \iff (g_j(x) + s_j = 0 \text{ et } s_j \geq 0).$$

Ensuite, on peut facilement voir que l'ajout de contraintes peut compliquer fortement la résolution d'un problème d'optimisation. Considérons par exemple (cf. [4] page 305) le problème d'optimisation suivant

$$(P_1) \begin{cases} \text{Min } (x_2 + 100)^2 + 0.01x_1^2 \\ \cos x_1 - x_2 \leq 0, \end{cases}$$

qui est visualisé à la FIGURE3.1.

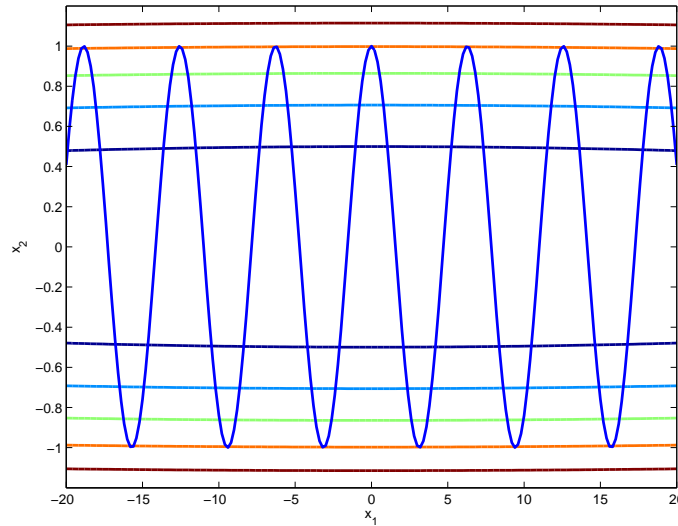
La solution de ce problème sans la contrainte est  $x^* = (0, -100)$  et tout minimum local du problème avec la contrainte satisfait celle-ci ( $f$  est convexe, donc tout minimum local de  $f$  est un minimum global, or ce dernier ne vérifie pas la contrainte). Par suite, nous avons  $x_2 = \cos x_1$  et le problème  $(P_1)$  est équivalent au problème d'optimisation dans  $\mathbf{R}$  sans contrainte

$$(P_2) \begin{cases} \text{Min } f_1(x_1) = (\cos x_1 + 100)^2 + 0.01x_1^2 \\ x_1 \in \mathbf{R} \end{cases}$$

On a alors des minima locaux proches des points  $((2k+1)\pi, -1)$ , pour  $k$  pas trop grand.

En troisième lieu, on peut remarquer que des problèmes d'optimisation non différentiables peuvent parfois s'écrire, en modifiant la modélisation, comme des problèmes d'optimisation différentiables. Une contrainte  $|x_1| + |x_2| \leq 1$  est par exemple équivalente aux quatre contraintes

$$\begin{aligned} x_1 + x_2 &\leq 1 \\ x_1 - x_2 &\leq 1 \\ -x_1 + x_2 &\leq 1 \\ -x_1 - x_2 &\leq 1. \end{aligned}$$

FIGURE 3.1 – *Problème avec et sans contraintes.*

De même le problème d'optimisation sans contrainte mais non différentiable

$$(P_3) \begin{cases} \text{Min } f(x) = \max(x^2, x) \\ x \in \mathbf{R} \end{cases}$$

s'écrit sous la forme d'un problème d'optimisation différentiable mais avec contraintes

$$(P_4) \begin{cases} \text{Min } t \\ x^2 - t \leq 0 \\ x - t \leq 0 \\ (x, t) \in \mathbf{R}^2. \end{cases}$$

Enfin, pour terminer cette introduction, on peut parfois lorsqu'il y a des contraintes simplifier le problème en supprimant des variables. Ainsi par exemple, dans le problème

$$(P_5) \begin{cases} \text{Min } f(x) \\ x_1 + x_3^2 - x_4 x_5 = 0 \\ -x_2 + x_4 + x_3^2 = 0 \\ x \in \mathbf{R}^4, \end{cases}$$

on peut éliminer les variables  $x_1$  et  $x_2$  et donc tout simplement résoudre le problème d'optimisation sans contraintes

$$(P_6) \begin{cases} \text{Min } g(x_3, x_4) = f(x_4 x_5 - x_3, x_4 + x_3^2, x_3, x_4) \\ (x_3, x_4) \in \mathbf{R}^2. \end{cases}$$

Par contre le problème suivant

$$(P_7) \begin{cases} \text{Min } f(x, y) = x^2 + y^2 \\ (x_1 - 1)^3 = y^2 \\ (x, y) \in \mathbf{R}^2, \end{cases}$$

n'est pas équivalent au problème obtenu en éliminant la variable  $y^2 = (x - 1)^3$

$$(P_8) \begin{cases} \text{Min } x^2 + (x - 1)^3 \\ x \in \mathbf{R}, \end{cases}$$

qui n'a pas de solution. En effet, on a ici "oublié" que  $y^2 = (x - 1)^3$  impliquait que  $(x - 1)^3 \geq 0 \iff x \geq 1$ .

## II Conditions du premier ordre

### II.1 Qualification des contraintes

On s'intéresse maintenant aux contraintes de type égalité et inégalité. L'idée principale est de dire que la fonction  $f$  ne peut pas croître lorsque l'on va dans une direction qui "laisse" localement dans l'ensemble des contraintes.

**Définition II.1.1 (direction tangente)** Soit  $\bar{x}$  un point d'un ensemble  $C$ ,  $d$  est une direction tangente à  $C$  en  $\bar{x}$  si et seulement si il existe une suite de points  $(x_k)_k$  de  $C$ ,  $x_k = \bar{x} + \alpha_k d_k$ ,  $\alpha_k > 0$ ,  $d_k \rightarrow d$  et  $\alpha_k \rightarrow 0$  quand  $k \rightarrow +\infty$ .

L'ensemble des directions tangentes en  $C$  en un point  $\bar{x}$  est un cône. En effet le vecteur nul est une direction tangente et si  $d$  est une direction tangente alors  $\alpha d$ ,  $\alpha \geq 0$  est une direction tangente, d'où la

**Définition II.1.2 (Cône tangent)** Soit  $\bar{x} \in C$ , on appelle cône tangent à  $C$  en  $\bar{x}$  l'ensemble noté  $T(C, \bar{x})$  des directions tangentes à  $C$  en  $\bar{x}$ .

Cependant ce cône n'est pas facile à utiliser en pratique. Une idée est de considérer le cône tangent des contraintes linéarisées en  $\bar{x}$ . Pour cela on définit tout d'abord l'ensemble des contraintes saturées en  $\bar{x}$

$$J_0(\bar{x}) = \{j \in J = \{1, \dots, m\}, g_j(\bar{x}) = 0\}. \quad (3.1)$$

Le cône tangent des contraintes linéarisées en  $\bar{x}$  s'écrit alors

$$T_L(C, \bar{x}) = \{d \in \mathbf{R}^n, (\nabla h_i(\bar{x})|d) = 0, i = 1, \dots, m, \\ (\nabla g_j(\bar{x})|d) \leq 0, j \in J_0(\bar{x})\}.$$

### Lemme II.1.3

Soit  $C = \{x \in \mathbf{R}^n, h(x) = 0, g(x) \leq 0\}$  où les fonctions  $h$  et  $g$  sont dérivables en  $\bar{x} \in C$ , on a toujours  $T(C, \bar{x}) \subset T_L(C, \bar{x})$ .

#### Démonstration

Soit  $d$  une direction tangente à  $C$  en  $\bar{x}$ , on a pour  $x_k = \bar{x} + \alpha_k d_k$  et toute contrainte  $h_i$ ,

$$h_i(x_k) = h_i(\bar{x}) + \alpha_k (\nabla h_i(\bar{x})|d_k) + \alpha_k \|\varepsilon(\alpha_k d_k)\|.$$

Or  $x_k$  et  $\bar{x}$  appartiennent à  $C$ , donc  $h_i(x_k) = h_i(\bar{x}) = 0$  et  $(\nabla h_i(\bar{x})|d_k) + \|\varepsilon(\alpha_k d_k)\| = 0$ . Par suite en faisant tendre  $k$  vers l'infini nous obtenons  $(\nabla h_i(\bar{x})|d) = 0$ .

Considérons maintenant une contrainte d'égalité saturée  $j \in J_0(\bar{x})$ . On obtient par un raisonnement identique

$$(\nabla g_j(\bar{x})|d_k) + \|\varepsilon(\alpha_k d_k)\| \leq 0,$$

pour tout  $k$ , et en passant à la limite on trouve  $(\nabla g_j(\bar{x})|d) = 0$ .  $\square$

### Exercice II.1

On considère  $C = \{x \in \mathbf{R}^2, h(x) = x_1^2 + x_2^2 = 2\}$  et on pose  $\bar{x} = (-\sqrt{2}, 0)$ .

- 1 (i) Visualiser dans le plan  $C$  et  $\nabla h(\bar{x})$ .
- (ii) Calculer  $T(C, \bar{x})$  et  $T_L(C, \bar{x})$  et visualiser ces ensembles.

- 2 On pose maintenant  $C = \{x \in \mathbf{R}^2, h(x) = (x_1^2 + x_2^2 - 2)^2 = 0\}$ . Que deviennent  $C$ ,  $T(C, \bar{x})$  et  $T_L(C, \bar{x})$  ?

**Définition II.1.4 (Hypothèse de qualification des contraintes)** On appelle hypothèse de qualification des contraintes toute condition suffisante pour avoir  $T(C, \bar{x}) = T_L(C, \bar{x})$ .

### Lemme II.1.5

- (i) Si les contraintes sont linéaires, alors la qualification des contraintes est vérifiée en tout point admissible.
- (ii) Si en  $\bar{x} \in C$ , les gradients  $(\nabla h_i(\bar{x}))_{i=1, \dots, p}$  et  $(\nabla g_j(\bar{x}))_{j \in J_0(\bar{x})}$  sont linéairement indépendants la qualification des contraintes est vérifiée en  $\bar{x}$ .
- (iii) S'il existe un vecteur  $d \in \mathbf{R}^n$  tel que
  - $(\nabla h_i(\bar{x})|d) = 0$ , pour tout  $i = 1, \dots, p$ ;
  - $(\nabla g_j(\bar{x})|d) < 0$ , pour tout  $j \in J_0(\bar{x})$ .
 et que les gradients des contraintes d'égalité sont linéairement indépendants en  $\bar{x}$ , alors la qualification des contraintes est vérifiée en  $\bar{x}$ .
- (iv) S'il n'y a pas de contrainte d'égalité, que les fonctions  $g_j$  sont convexes, et qu'il existe un vecteur  $d$  tel que  $g_j(d) < 0$  pour tout  $j \in J_0(\bar{x})$ , alors la qualification des contraintes est vérifiée en  $\bar{x}$ .

#### Démonstration

- (i) Évident.

- (ii) Considérons donc  $d \in T_L(C, \bar{x})$  différent du vecteur nul. Nous allons construire un arc de courbe  $C^1$  admissible de vecteur tangent en  $\bar{x}, d$ . On peut toujours supposer, quitte à changer l'ordre des composantes de  $g$  que  $J_0(\bar{x}) = \{1, \dots, \bar{m}\}$ . Définissons, pour des vecteurs  $b_1, \dots, b_{n-p-\bar{m}}$  fixés que nous choisirons plus tard, la fonction

$$\begin{aligned} r : \mathbf{R}^n \times \mathbf{R} &\longrightarrow \mathbf{R}^n \\ (x, \theta) &\longmapsto r(x, \theta) \end{aligned}$$

avec

$$\begin{aligned} r_i(x, \theta) &= h_i(x) - \theta(\nabla h_i(\bar{x})|d) \text{ pour } i = 1, \dots, p \\ r_{j+p}(x, \theta) &= g_j(x) - \theta(\nabla g_j(\bar{x})|d) \text{ pour } j = 1, \dots, \bar{m} \\ r_{i+p+\bar{m}}(x, \theta) &= (x - \bar{x}|b_i) - \theta(b_i|d) \text{ pour } i = 1, \dots, n - p - \bar{m} \end{aligned}$$

On a alors  $r(\bar{x}, 0) = 0$  et il existe  $\varepsilon$  strictement positifs tel que pour tout  $\theta \geq 0$ ,

$$\left. \begin{aligned} r(x, \theta) &= 0 \\ x &\in B(\bar{x}, \varepsilon) \end{aligned} \right\} \implies x \in C.$$

De plus, on a

$$\frac{\partial r}{\partial x}(\bar{x}, 0) = \begin{pmatrix} \nabla h_1(\bar{x})^T \\ \vdots \\ \nabla h_p(\bar{x})^T \\ \nabla g_1(\bar{x})^T \\ \vdots \\ \nabla g_{\bar{m}}(\bar{x})^T \\ b_1^T \\ \vdots \\ b_{n-p-\bar{m}}^T \end{pmatrix}.$$

Les  $p+\bar{m}$  premières lignes de cette matrice étant linéairement indépendantes, on peut choisir les vecteurs  $b_1, \dots, b_{n-p-\bar{m}}$  de façon à ce que cette matrice soit inversible. Dans ce cas le théorème des fonctions implicites implique qu'il existe un ouvert  $U$  contenant  $x, \theta_0 > 0$  et une fonction  $\varphi : ]-\theta_0, \theta_0[ \rightarrow U$  de classe  $C^1$  telle que  $r(x, \theta) = 0$  est équivalent dans  $U \times ]-\theta_0, \theta_0[$  à  $x = \varphi(\theta)$ . De plus on a

$$\varphi'(0) = - \left[ \frac{\partial r}{\partial x}(\bar{x}, 0) \right]^{-1} \cdot \frac{\partial r}{\partial \theta}(\bar{x}, 0) = - \begin{pmatrix} \nabla h_1(\bar{x})^T \\ \vdots \\ \nabla h_p(\bar{x})^T \\ \nabla g_1(\bar{x})^T \\ \vdots \\ \nabla g_{\bar{m}}(\bar{x})^T \\ b_1^T \\ \vdots \\ b_{n-p-\bar{m}}^T \end{pmatrix}^{-1} \begin{pmatrix} -(\nabla h_1(\bar{x})|d) \\ \vdots \\ -(\nabla h_p(\bar{x})|d) \\ -(\nabla g_1(\bar{x})|d) \\ \vdots \\ -(\nabla g_{\bar{m}}(\bar{x})|d) \\ -(b_1|d) \\ \vdots \\ -(b_{n-p-\bar{m}}|d) \end{pmatrix} = d$$

On en déduit que

$$\lim_{\theta \rightarrow 0^+} \frac{\varphi(\theta) - \bar{x}}{\theta} = d.$$

Il suffit alors de poser

$$\begin{aligned} x_k &= \varphi(\theta_k) = \varphi(0) + \theta_k(\varphi'(0) + \varepsilon(\theta_k)) \\ &= \bar{x} + \theta_k d_k, \end{aligned}$$

avec  $\theta_k > 0, \theta_k \rightarrow 0$ .

□

## II.2 Théorème de Karuch, Kuhn et Tucker

**Définition II.2.1 (Lagrangien)** On appelle Lagrangien associé au problème d'optimisation avec des contraintes et type égalité et inégalité

$$(P) \begin{cases} \text{Min } f(x) \\ h(x) = 0 \\ g(x) \leq 0 \\ x \in \mathbf{R}^n, \end{cases}$$

la fonction

$$L : \mathbf{R}^n \times \mathbf{R}^p \times (\mathbf{R}^+)^m \longrightarrow \mathbf{R} \\ (x, \lambda, \mu) \longmapsto L(x, \lambda, \mu)$$

avec

$$L(x, \lambda, \mu) = f(x) + (\lambda|h(x)) + (\mu|g(x)) \quad (3.2)$$

### Théorème II.2.2 (Karuch-Kuhn-Tucker, 1952)

On considère le problème d'optimisation (P) avec des contraintes de type égalité et inégalité. On suppose que les fonctions  $f, g$  et  $h$  sont de classe  $C^1$  en  $x^*$ , que l'hypothèse de qualification des contraintes est vérifiée en ce point et que  $x^*$  est un minimum local de (P). Alors il existe  $(\lambda^*, \mu^*) \in \mathbf{R}^p \times (\mathbf{R}^+)^m$  vérifiant

- (i)  $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$ ;
- (ii)  $h(x^*) = 0$ ;
- (iii)  $g(x^*) \leq 0$ ;
- (iv)  $\mu^* \geq 0$ ;
- (v)  $(\mu|g(x)) = 0$ , relations de complémentarité.

### Notation II.2.3

On notera (KKT) les conditions (i),(ii),(iii),(iv) et (v) du théorème II.2.2.

Les réels  $(\lambda_1, \dots, \lambda_p)$  et  $(\mu_1, \dots, \mu_m)$  s'appellent les multiplicateurs de Lagrange (parfois appelés dans la littérature anglo-saxonne multiplicateurs de Karush-Kuhn-Tucker)

### Remarque II.2.4

Les relations (iii),(iv) et (v) sont équivalentes aux relations

- (i)  $g_j(x^*) \leq 0$  pour tout  $j$ ;
- (ii)  $\mu_j \geq 0$  pour tout  $j$ ;
- (iii)  $\mu_j^* g_j(x^*) = 0$  pour tout  $j$ .

Avant de démontrer ce théorème, démontrons trois lemmes.

### Lemme II.2.5

Soit  $S \subset \mathbf{R}^n$  un convexe fermé non vide et  $b \notin S$ . Alors, il existe  $a \in \mathbf{R}^n$  non nul et  $\alpha \in \mathbf{R}$  tel que  $(a|b) > \alpha$  et  $(a|x) \leq \alpha$  pour tout  $x \in S$ .

*Démonstration*

On considère le problème de la projection de  $b$  sur le convexe fermé qui est un problème d'optimisation convexe

$$(P_9) \begin{cases} \text{Min } \frac{1}{2} \|x - b\|^2 \\ x \in S. \end{cases}$$

Ce problème convexe admet une solution  $x^*$  qui est caractérisée par la relation  $f'(x^*) \cdot (x - x^*) \geq 0$  pour tout  $x$  dans  $S$ . Comme ici  $\nabla f(x) = x - b$ , cette relation implique

$$\begin{aligned} (x^* - b|x - x^*) &\geq 0 \text{ pour tout } x \text{ dans } S \\ (b - x^*|x - x^*) &\leq 0 \text{ pour tout } x \text{ dans } S \\ -(b - x^*|x^*) &\leq -(b - x^*|b). \end{aligned}$$

Par suite

$$\begin{aligned} \|b - x^*\|^2 &= (b|b - x^*) - (x^*|b - x^*) \\ &\leq (b - x^*|b - x^*) \end{aligned}$$

Si nous posons  $a = b - x^* \neq 0$ , alors pour tout  $x$  dans  $S$  on a  $\|b - x^*\|^2 \leq (b|a) - (x|a)$ , soit  $(b|a) \geq (x|a) + \|b - x^*\|^2$ . Il suffit maintenant de prendre  $\alpha = \sup_{x \in S} (x|a)$  pour conclure.  $\square$

**Lemme II.2.6 (Farkas et Minkowski)**

Soit  $A$  une matrice  $(n, p)$ , et  $b \in \mathbf{R}^n$  alors

$$\begin{cases} Ax = b \\ x \geq 0 \end{cases} \quad \text{admet une solution} \iff (\forall u \in \mathbf{R}^n, u^T A \geq 0 \implies (u|b) \geq 0)$$

*Démonstration*

Démontrons tout d'abord l'implication. Soit donc  $x \geq 0$  et  $u$  tels que  $Ax = b$  et  $u^T A \geq 0$ . Alors  $u^T Ax = (u|b) \geq 0$  car c'est une somme de termes positifs.

Montrons maintenant la réciproque. Supposons que  $Ax = b, x \geq 0$  n'a pas de solution et considérons le convexe fermé  $S = \{y, y = Ax, x \geq 0\}$ .  $b \notin S$ , on peut donc appliquer le lemme de séparation précédent. Il existe donc  $a \neq 0$  et  $\alpha \in \mathbf{R}$  tels que  $(a|b) > \alpha$  et  $(a|y) \leq \alpha$  pour tout  $y$  dans  $S$ . Mais  $0 \in S$ , par suite  $0 \leq \alpha$  et  $(a|b) > 0$ . D'autre part  $(a|y) = (a|Ax) \leq \alpha$  pour tout  $x \geq 0$ . Ceci implique alors que  $a^T A \leq 0$ . En effet si il existe  $(a^T A)_i > 0$  et si on prend  $x = (0 \dots x_i \dots 0)^T$  alors en faisant tendre  $x_i$  vers  $+\infty$  on obtient  $a^T Ax$  qui tend aussi vers  $+\infty$ . Il suffit maintenant de poser  $u = -a$  pour obtenir  $u^T A \geq 0$  et  $(u|b) < 0$ .  $\square$

**Lemme II.2.7**

Si  $x^*$  est un minimum local de  $(P)$  alors pour tout vecteur tangent  $d \in T(C, x^*)$ , on a  $(\nabla f(x^*)|d) \geq 0$ .

*Démonstration*

Soit  $d \in T(C, x^*)$ , alors il existe une suite de points admissibles  $x_k = x^* + \alpha_k d_k$  avec  $\alpha_k \rightarrow 0$  et  $d_k \rightarrow d$  lorsque  $k \rightarrow +\infty$ . Par suite on a

$$\begin{aligned} f(x_k) &= f(x^*) + \alpha_k (\nabla f(x^*)|d_k) + \alpha_k \|d_k\| \varepsilon(\alpha_k d_k) \geq f(x^*) \\ &\implies (\nabla f(x^*)|d_k) + \|d_k\| \varepsilon(\alpha_k d_k) \geq 0 \\ &\implies (\nabla f(x^*)|d) \geq 0. \end{aligned}$$

$\square$

*Démonstration**du théorème*

Considérons le cas où il n'y a pas de contraintes de type égalité. Si  $x^*$  est un minimum local alors le lemme précédent implique que pour tout  $d \in T(C, x^*)$ , on a  $(\nabla f(x^*)|d) \geq 0$ . L'hypothèse de qualification des contraintes en  $x^*$  implique que  $T(C, x^*) = T_L(C, x^*) = \{d \in \mathbf{R}^n, \nabla(g_j(x^*)|d) \leq 0 \text{ pour tout } j \text{ dans } J_0(x^*)\}$ . Le lemme de Farkas et Minkowski permet alors en posant  $A = (-\nabla g_{j_1}(x^*) \dots -\nabla g_{j_m}(x^*))$ ,  $b = \nabla f(x^*)$  et  $u = d$  d'écrire

$$d^T A \geq 0 \implies (d|\nabla f(x^*)) \geq 0.$$

Par suite, il existe une solution à

$$\begin{cases} A\mu = b \\ \mu \geq 0 \end{cases}$$

Soit, pour  $\mu \geq 0$ ,

$$\nabla f(x^*) = - \sum_{j \in J_0(x^*)} \mu_j \nabla g_j(x^*).$$

Il suffit alors de poser  $\mu_j = 0$  pour les  $j$  qui ne sont pas dans  $J_0(x^*)$  pour conclure.

Pour le cas général d'un problème  $(P)$  avec des contraintes de type égalité et inégalité, il suffit d'écrire les contraintes d'égalité sous la forme de deux contraintes d'inégalité :  $h_i(x) = 0 \iff h_i(x) \leq 0$  et  $-h_i(x) \leq 0$ . Nous obtenons alors pour ces deux contraintes deux multiplicateurs  $\mu_i^+ \geq 0$  et  $\mu_i^- \geq 0$ . En posant  $\lambda_i = \mu_i^- - \mu_i^+$ , on obtient alors le résultat.  $\square$

**II.3 Cas convexe****Théorème II.3.1**

On suppose que les fonctions  $f$  et  $g$  sont convexes, que  $h$  est affine et que les ces fonctions sont  $C^1$ . Alors  $(KKT)$  est une condition suffisante de solution.

*Démonstration*

Soit  $(x^*, \lambda^*, \mu^*)$  un point vérifiant  $(KKT)$  alors  $L(x^*, \lambda^*, \mu^*) = f(x^*)$  et  $L(x, \lambda, \mu)$  est convexe en  $x$ . Donc  $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$  implique que  $x^*$  est un minimum global de  $L(\cdot, \lambda^*, \mu^*)$ , c'est-à-dire que pour tout  $x \in \mathbf{R}^n$ ,  $f(x^*) \leq L(x, \lambda^*, \mu^*)$ .

Soit maintenant  $x$  un point réalisable alors  $L(x, \lambda^*, \mu^*) = f(x^*) + \sum_{i=1}^m \mu_i g_i(x) \leq f(x)$ ; on en déduit alors que

$$f(x^*) \leq L(x, \lambda^*, \mu^*) \leq f(x).$$

$\square$

**Corollaire II.3.2**

Sous les mêmes hypothèses que précédemment et si on l'hypothèse de qualification des contraintes en tout point, alors les conditions (KKT) sont des conditions nécessaires et suffisantes de solution.

*Démonstration*

C'est une conséquence immédiate des théorèmes II.2.2 et II.3.1.  $\square$

**III Conditions du second ordre****III.1 Conditions Nécessaires du second ordre****Théorème III.1.1 (Contraintes d'égalité)**

On suppose que  $f$  et  $h$  sont deux fois différentiables sur un ouvert  $\Omega$  contenant  $C$ . Si  $x^*$  est un minimum local de  $f$  sur  $C$  et si les vecteurs  $(\nabla h_1(x^*), \dots, \nabla h_p(x^*))$  sont linéairement indépendants alors il existe des multiplicateurs de Lagrange  $\lambda^* = (\lambda_1, \dots, \lambda_p) \in \mathbf{R}^p$  uniques tels que

- (i)  $\nabla_x L(x^*, \lambda^*) = 0$ ;
- (ii)  $(\nabla_{xx}^2 L(x^*, \lambda^*)d|d) \geq 0$  pour tout  $d$  dans le sous espace vectoriel tangent

$$T_L(C, x^*) = \{d \in \mathbf{R}^n, (\nabla h_i(x^*)|d) = 0 \text{ pour tout } i\}.$$

*Démonstration*

Comme les vecteurs  $\nabla h_1(x^*), \dots, \nabla h_p(x^*)$  sont linéairement indépendants on a l'hypothèse de qualification des contraintes en  $x^*$ , par suite  $T(C, x^*) = T_L(C, x^*)$ . Soit donc  $d \in T_L(C, x^*)$  et  $(x_k)_k$  une suite de point de  $C$  tels que  $x_k = x^* + \alpha_k d_k$ ,  $\alpha_k$  tend vers 0 quand  $k \rightarrow +\infty$  et  $d_k$  tend vers  $d$  quand  $k \rightarrow +\infty$ . On peut alors écrire

$$L(x^* + \alpha_k d_k, \lambda^*) = L(x^*, \lambda^*) + \alpha_k \nabla_x L(x^*, \lambda^*)^T d_k + \frac{\alpha_k^2}{2} d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k + \alpha_k^2 \|d_k\|^2 \varepsilon(\alpha_k d_k).$$

Mais lorsque  $x$  appartient à  $C$ ,  $L(x, \lambda^*) = f(x)$  et (KKT) implique que  $\nabla_x L(x^*, \lambda^*) = 0$ . On en déduit que pour tout  $k$

$$\frac{f(x^* + \alpha_k d_k) - f(x^*)}{\alpha_k^2/2} = (\nabla_{xx}^2 L(x^*, \lambda^*)d_k|d_k) + 2\|d_k\|\varepsilon(\alpha_k d_k) \geq 0.$$

Il suffit alors de faire tendre  $k$  vers  $+\infty$  pour obtenir le résultat.  $\square$

**Remarque III.1.2**

On peut exprimer la condition (ii) de la façon suivante. Dire que  $d \in T_L(C, x^*)$  est équivalent à dire que  $d \in \ker h'(x^*)$  ce qui s'écrit  $d = V\nu$  où  $V$  est une matrice dont les vecteurs colonnes forment une base de  $\ker h'(x^*)$ . Par suite la condition est équivalente à

$$\nu^T (V^T \nabla_{xx}^2 L(x^*, \lambda^*) V) \nu \geq 0.$$

Ce qui revient à dire que la matrice  $V^T \nabla_{xx}^2 L(x^*, \lambda^*) V$  est semi-définie positive.

**Théorème III.1.3 (Contraintes d'égalité et d'inégalité)**

On suppose que  $f, g$  et  $h$  sont deux fois différentiables sur un ouvert  $\Omega$  contenant  $C$ . Si  $x^*$  est un minimum local de  $f$  sur  $C$  et si les vecteurs  $(\nabla h_1(x^*), \dots, \nabla h_p(x^*))$  et  $(\nabla g_j(x^*))_{j \in J_0(x^*)}$  sont linéairement indépendants alors il existe des multiplicateurs de Lagrange  $\lambda^* = (\lambda_1, \dots, \lambda_p) \in \mathbf{R}^p$  et  $\mu^* \in (\mathbf{R}^+)^m$  uniques tels que

- (i)  $KKT(x^*, \lambda^*, \mu^*)$  soit vérifié
- (ii)  $(\nabla_{xx}^2 L(x^*, \lambda^*)d|d) \geq 0$  pour tout  $d \in \mathbf{R}^n$  tel que
  - (a)  $(\nabla h_i(x^*)|d) = 0$  pour tout  $i = 1, \dots, p$ ;
  - (b)  $(\nabla g_j(x^*)|d) = 0$  pour tout  $j \in J_0(x^*)$  et  $\mu_j^* > 0$ ;
  - (c)  $(\nabla g_j(x^*)|d) \leq 0$  pour tout  $j \in J_0(x^*)$  et  $\mu_j^* = 0$ ;

*Démonstration*

On note  $J_0^+(x^*)$  l'ensemble des indices des contraintes qui sont saturées et pour lesquelles le multiplicateur  $\mu_j^* > 0$  et on pose

$$\begin{aligned} \tilde{C} = \{x \in \mathbf{R}^n, & h_i(x) = 0, i = 1, \dots, p, \\ & g_j(x) = 0, j \in J_0^+(x), \\ & g_j(x) \leq 0, j \in J_0(x) \setminus J_0^+(x)\} \end{aligned}$$



Remarquons tout d'abord que si  $x \in \tilde{C}$  alors  $\mu_j^* g_j(x) = 0$  et que, par suite,  $L(x, \lambda^*, \mu^*) = f(x)$ .

Comme on a l'indépendance linéaire des gradients, on a, relativement à  $\tilde{C}$  l'hypothèse de qualification des contraintes et donc

$$\begin{aligned} T(\tilde{C}, x^*) &= \{d \in \mathbf{R}^n, (\nabla h_i(x^*)|d) = 0, i = 1, \dots, p, \\ &\quad (\nabla g_j(x^*)|d) = 0, j \in J_0^+(x), \\ &\quad (\nabla g_j(x^*)|d) \leq 0, j \in J_0(x) \setminus J_0^+(x)\} \end{aligned}$$

Soit maintenant  $d \in T(\tilde{C}, x^*)$ , alors il existe une suite  $(x_k)_k$  de points de  $\tilde{C}$ ,  $x_k = x^* + \alpha_k d_k$  avec  $\alpha_k$  tend vers 0 lorsque  $k$  tend vers  $+\infty$  et  $d_k$  tend vers  $d$  lorsque  $k$  tend vers  $+\infty$ . On peut maintenant écrire

$$\begin{aligned} L(x_k, \lambda^*, \mu^*) &= L(x^*, \lambda^*, \mu^*) + \alpha_k \nabla_x L(x^*, \lambda^*, \mu^*)^T d_k + \frac{\alpha_k^2}{2} d_k^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) d_k + \alpha_k^2 \|d_k\|^2 \varepsilon(\alpha_k d_k) \\ f(x_k) &= f(x^*) + \frac{\alpha_k^2}{2} [d_k^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) d_k + \alpha_k^2 \|d_k\|^2 \varepsilon(\alpha_k d_k)] \end{aligned}$$

Donc si  $d \neq 0$  et  $x^*$  est un minimum local on obtient

$$d_k^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) d_k + 2\|d_k\|^2 \varepsilon(\alpha_k d_k).$$

On en déduit alors le résultat en faisant tendre  $k$  vers  $+\infty$ .  $\square$

### III.2 Conditions suffisantes

#### Théorème III.2.1 (Contraintes d'égalité)

On suppose que  $f$  et  $h$  sont deux fois férvables sur un ouvert  $\Omega$  contenant  $C$ . Si  $(x^*, \lambda^*)$  vérifie

- (i)  $\nabla_x L(x^*, \lambda^*) = 0$ ;
- (ii)  $h(x^*) = 0$ ;
- (iii)  $(\nabla_{xx}^2 L(x^*, \lambda^*)d|d) > 0$  pour tout  $d$  dans le sous espace vectoriel tangent

$$T_L(C, x^*) = \{d \in \mathbf{R}^n, (\nabla h_i(x^*)|d) = 0 \text{ pour tout } i\}.$$

alors  $x^*$  est un minimum local de  $(P_e)$ .

#### Démonstration

Supposons que  $x^*$  ne soit pas un minimum local, nous allons alors construire un vecteur  $d \in T(C, x^*)$  qui vérifie les conditions mentionnées tel que  $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \leq 0$ , d'où la contradiction.

Soit donc  $(x_k)_k$  une suite de point de  $C$  qui converge vers  $x^*$  et qui vérifie  $f(x_k) < f(x^*)$ . Posons alors  $d_k = (x_k - x^*)/\|x_k - x^*\|$ . Cette suite est une suite d'un compact ( $\|d_k\| = 1$ ), elle admet donc une sous-suite, toujours notée  $(d_k)_k$ , qui converge vers  $d \in S(0, 1)$ . Si on pose  $\alpha_k = \|x_k - x^*\|$ , on a  $x_k = x^* + \alpha_k d_k$ . Par suite  $d \in T(C, x^*) \subset T_L(C, x^*)$ . Or, on peut écrire

$$L(x^* + \alpha_k d_k, \lambda^*) = L(x^*, \lambda^*) + \alpha_k \nabla_x L(x^*, \lambda^*)^T d_k + \frac{\alpha_k^2}{2} d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k + \alpha_k^2 \|d_k\|^2 \varepsilon(\alpha_k d_k).$$

Mais  $x_k$  et  $x^*$  sont dans  $C$ , par suite  $L(x_k, \lambda^*) = f(x_k)$  et  $L(x^*, \lambda^*) = f(x^*)$ . On en déduit alors

$$\frac{1}{\alpha_k^2} (f(x_k) - f(x^*)) = d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k + \|d_k\| \varepsilon(\alpha_k d_k) < 0,$$

et donc, en passant à la limite que  $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \leq 0$ . Ce qui est contraire à la condition III.2.1.iii.  $\square$

#### Théorème III.2.2 (Contraintes d'égalité et d'inégalité)

On suppose que  $f, g$  et  $h$  sont deux fois férvables sur un ouvert  $\Omega$  contenant  $C$ . Si  $(x^*, \lambda^*, \mu^*)$  est un point de  $\mathbf{R}^n \times \mathbf{R}^p \times (\mathbf{R}^+)^r$  vérifiant

- (i)  $KKT(x^*, \lambda^*, \mu^*)$ ;
- (ii)  $(\nabla_{xx}^2 L(x^*, \lambda^*)d|d) > 0$  pour tout  $d \neq 0 \in \mathbf{R}^n$  tel que
  - (a)  $(\nabla h_i(x^*)|d) = 0$  pour tout  $i = 1, \dots, p$ ;
  - (b)  $(\nabla g_j(x^*)|d) = 0$  pour tout  $j \in J_0(x^*)$  et  $\mu_j^* > 0$ ;
  - (c)  $(\nabla g_j(x^*)|d) \leq 0$  pour tout  $j \in J_0(x^*)$  et  $\mu_j^* = 0$ ;

alors  $x^*$  est un minimum local de  $(P)$ .

#### Démonstration

On raisonne toujours par l'absurde. Supposons donc encore qu'il existe une suite  $(x_k)_k$  de point de  $C$  qui converge vers  $x^*$  et tel que  $f(x_k) < f(x^*)$  et définissons comme précédemment  $x_k = x^* + \alpha_k d_k$ . Montrons tout d'abord que  $d$  vérifie les conditions demandées. Le vecteur  $d$  appartient à  $T(C, x^*) \subset T_L(C, x^*)$ . On a donc immédiatement que  $(\nabla h_i(x^*)|d) = 0$  pour tout  $i$  et  $(\nabla g_j(x^*)|d) \leq 0$  pour tout  $j \in J_0(x^*)$ . Il reste donc à montrer que  $(\nabla g_j(x^*)|d) = 0$  pour  $j \in J_0^+(x^*)$ . Pour cela on constate tout d'abord que  $(\nabla f(x^*)|d) \leq 0$ . En effet il suffit de passer à la limite dans

$$f(x_k) - f(x^*) = L(x_k, \lambda^*) - L(x^*, \lambda^*) = \alpha_k (\nabla_x L(x^*, \lambda^*)|d_k) + \|d_k\| \varepsilon(\alpha_k d_k) \leq 0.$$

Ensuite on peut écrire

$$\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f(x^*) + \sum_{i=1}^p \lambda_i^* \nabla h_i(x^*) + \sum_{j \in J_0(x^*)} \mu_j^* \nabla g_j(x^*).$$

Et donc, pour  $d$

$$(\nabla_x L(x^*, \lambda^*, \mu^*)|d) = (\nabla f(x^*)|d) + \sum_{i=1}^p \lambda_i^* (\nabla h_i(x^*)|d) + \sum_{j \in J_0^+(x^*)} \mu_j^* (\nabla g_j(x^*)|d) = 0.$$

Ceci implique que les termes  $\mu_j^* (\nabla g_j(x^*)|d) = 0$  pour tout  $j \in J_0(x^*)$ , et donc que  $(\nabla g_j(x^*)|d) = 0$  pour  $j \in J_0(x^*)$  et  $\mu_j^* > 0$ .

Mais,

$$L(x_k, \lambda^*, \mu^*) - L(x^*, \lambda^*, \mu^*) = f(x_k) + \sum_{j \in J_0(x^*)} \mu_j g_j(x_k) - f(x^*) < 0.$$

et donc

$$\frac{1}{2} (d_k^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) d_k + 2 \|d_k\|^2 \varepsilon(\alpha_k d_k)) \leq 0.$$

On obtient alors la contradiction en passant à la limite.  $\square$

#### Remarque III.2.3

Dans les théorèmes sur les conditions suffisantes, on a pas besoin d'hypothèse de qualification des contraintes car on a l'inclusion des cônes dans le bon sens ( $T(C, x^*) \subset T_L(C, x^*)$ ). Par contre, on a besoin de cette hypothèse pour les conditions nécessaires.

## IV Exercices

### IV.1 Avec corrections

#### Exercice IV.1

On considère le cas de contraintes d'égalité affine (donc convexe) non vide :  $h(x) = Ax + b$ .

**1** Montrer que si  $x^*$  est une solution du problème alors il existe  $\lambda^* \in \mathbf{R}^p$  tel que

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i=1}^p \lambda_i^* \nabla h_i(x^*) = 0. \quad (3.3)$$

Indication : on doit avoir  $(\nabla f(x)|h) = 0$  pour tout  $h \in \ker A$ , ce qui est équivalent à  $\nabla f(x) \in (\ker A)^\perp \iff \nabla f(x) \in \text{Im} A^T$ .



## Chapitre 4

# Algorithmes globalisés pour l'optimisation sans contrainte

## I Algorithmes de minimisation sans contrainte

### I.1 La méthode de Newton

Cette méthode et ses variantes moins coûteuses, forment une des principales classes de méthode d'optimisation pour les problèmes sans contraintes. Cette méthode s'écrit :

Newton's method	
1.	Choose $x_0$
2.	For $k=0,2, \dots$ Do
3.	Compute if $\nabla^2 f(x_k)$ is nonsingular
4.	$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$
5.	EndDo

Quelques remarques sur sa mise en œuvre :

- Cette méthode nécessite d'avoir à faire à une fonction deux fois dérivable et à ses dérivées jusqu'à l'ordre 2.
- Cette méthode nécessite aussi la résolution de systèmes linéaires (on ne calcule pas l'inverse). Cette opération peut être coûteuse pour des systèmes de grande taille.
- Cette méthode jouit de propriétés de convergence locale très intéressantes comme nous allons le voir.

Soit  $x^k \in \mathbb{R}^n$ . On considère l'approximation quadratique de  $f$ , fonction deux fois dérivable, suivante :  $m(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$ . Supposons que  $\nabla^2 f(x_k)$  est symétrique et définie positive alors le minimum  $x^*$  de  $m(x)$  vérifie  $x^* - x_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$ . La méthode de Newton minimise donc à chaque pas où  $\nabla^2 f(x_k)$  est symétrique et définie positive l'approximation quadratique de  $f$ . Notez que si  $\nabla^2 f(x_k)$  a des valeurs propres négatives, l'approximation quadratique n'est pas bornée inférieurement, et le point  $x_{k+1}$  peut même dans certains cas être un maximum de  $m(x)$  (considérer  $-(x - x_k)^2$ ). Cette situation n'arrive pas si l'on est suffisamment proche de points vérifiant la condition d'optimalité suffisante du second ordre. Cela conduit aux conditions dites *standart* pour l'algorithme de Newton.

Hypothèses standart en  $\bar{x} \in \mathcal{O}$ , où  $\mathcal{O}$  est un ouvert convexe de  $\mathbb{R}^n$  :

- c1  $f$  est deux fois continûment différentiable sur  $\mathcal{O}$
- c2  $x \mapsto \nabla^2 f(x)$  est Lipschitz continue sur  $\mathcal{O}$  :  $\|\nabla^2 f(y) - \nabla^2 f(x)\| \leq \gamma \|y - x\|$
- c3  $\nabla f(\bar{x}) = 0$  et  $\nabla^2 f(\bar{x})$  est définie positive

**Exercice 4.1** Sous les hypothèses standart, il existe  $\delta > 0$  et  $K > 0$ , tels que si  $\|\bar{x} - x_0\| \leq \delta$ ,  $\|\bar{x} - x_{k+1}\| \leq K \|\bar{x} - x_k\|^2$ . Si  $K\delta < 1$ ,  $(x_k)$  converge vers  $\bar{x}$ . Une telle convergence est appelée locale quadratique.

Démonstration : 1) En utilisant le Théorème de Rouché (ou un résultat de continuité des valeurs propres), il existe un voisinage de  $\bar{x}$  inclus dans  $\mathcal{O}$  où  $\nabla^2 f(x)$  est définie positive. Les fonctions  $x \mapsto \|\nabla^2 f(x)\|$  et  $x \mapsto \|\nabla^2 f(x)^{-1}\|$  sont continues dans un voisinage de  $\bar{x}$  inclus dans  $\mathcal{O}$  car  $x \mapsto \nabla^2 f(x)$  est continue, dans  $\mathcal{O}$  et  $x \mapsto \nabla^2 f(x)^{-1}$  est continue dans un voisinage de  $\bar{x}$  car  $\nabla^2 f(\bar{x})$  est inversible. Donc il existe  $\delta$  tel que  $\|\bar{x} - x\| \leq \delta$  (noté  $x \in B(\delta)$ ) entraîne

$$\|\nabla^2 f(x)\| \leq 2\|\nabla^2 f(\bar{x})\| \text{ et } \|\nabla^2 f(x)^{-1}\| \leq 2\|\nabla^2 f(\bar{x})^{-1}\|, \text{ et } \nabla^2 f(x) \text{ est définie positive.} \quad (4.1)$$

Soit  $x_k \in B(\delta)$ . Alors on obtient par Taylor avec reste intégral,  $\nabla f(x_k) = \int_0^1 \nabla^2 f(\bar{x} + s(x_k - \bar{x}))(x_k - \bar{x}) ds$ , qui montre que

$$\|\bar{x} - x_{k+1}\| = \|\bar{x} - x_k + \nabla^2 f(x_k)^{-1} \nabla f(x_k)\| \quad (4.2)$$

$$= \|\nabla^2 f(x_k)^{-1} \left( \nabla^2 f(x_k)(\bar{x} - x_k) + \int_0^1 \nabla^2 f(\bar{x} + s(x_k - \bar{x}))(x_k - \bar{x}) ds \right)\| \quad (4.3)$$

$$= \|\nabla^2 f(x_k)^{-1} \int_0^1 (\nabla^2 f(\bar{x} + s(x_k - \bar{x})) - \nabla^2 f(x_k)) (x_k - \bar{x}) ds\| \quad (4.4)$$

$$\leq 2\gamma \|\nabla^2 f(\bar{x})^{-1}\| \int_0^1 (1-s) \|x_k - \bar{x}\|^2 ds = K \|\bar{x} - x_k\|^2. \quad (4.5)$$

Si  $K\delta < 1$ ,  $x_{k+1} \in B(\delta)$  (car  $\|\bar{x} - x_{k+1}\| \leq K \|x_k - \bar{x}\| \|x_k - \bar{x}\| \leq K\delta \|x_k - \bar{x}\|$ ) et par induction si  $x_0 \in B(\delta)$ , alors  $x_k \in B(\delta)$  pour tout  $k$ . De plus on vérifie aisément que  $\|\bar{x} - x_k\| \leq \frac{(K\delta)^{2^k}}{K}$ , ce qui montre que  $(x_k)$  converge vers  $\bar{x}$ .  $\square$

**Exercice 4.2** (Critère d'arrêt) Pour la suite  $f_n = \sum_{k=1}^n \frac{1}{k}$ , montrer que la stationnarité de  $f_n$  (i.e.  $f_{n+1} - f_n$  petit) n'indique pas la convergence. En déduire qu'arrêter une méthode d'optimisation sur  $|f(x_{k+1}) - f(x_k)| \leq \epsilon$  est dangereux. En revanche, sous les conditions standart, montrez que pour  $x_k$  suffisamment proche de  $\bar{x}$ , on a

$$\frac{\|\bar{x} - x_k\|}{4\|\bar{x} - x_0\| \text{cond}(\nabla^2 f(\bar{x}))} \leq \frac{\|\nabla f(x_k)\|}{\|\nabla f(x_0)\|} \leq \frac{4\text{cond}(\nabla^2 f(\bar{x}))\|\bar{x} - x_k\|}{\|\bar{x} - x_0\|}.$$

En déduire que la norme relative du gradient est un critère d'arrêt possible si le Hessien à l'optimum est bien conditionné.

**Démonstration :** La suite  $f_n$  diverge mais  $f_{n+1} - f_n$  tend vers 0. Par les mêmes arguments que pour la preuve de 4.1 on a pour  $x_k \in B(\delta)$ ,

$$\|\nabla f(x_k)\| = \left\| \int \nabla^2 f(\bar{x} + s(x_k - \bar{x}))(x_k - \bar{x}) ds \right\| \leq 2\|\nabla^2 f(\bar{x})\| \|\bar{x} - x_k\|.$$

On obtient alors par Taylor avec reste integral

$$\int (x_k - \bar{x})^T \nabla^2 f(\bar{x} + s(x_k - \bar{x}))(x_k - \bar{x}) ds = (x_k - \bar{x})^T \nabla f(x_k) \leq \|x_k - \bar{x}\| \|\nabla f(x_k)\|.$$

La matrice  $\nabla^2 f(\bar{x} + s(x_k - \bar{x}))$  étant définie positive dans  $B(\delta)$  l'inégalité matricielle pour une matrice  $A$  symétrique définie positive  $z^T z / \lambda_{\max}(A^{-1}) = \lambda_{\min}(A) z^T z \leq z^T A z$  montre alors que

$$\|x_k - \bar{x}\|^2 \int_0^1 \frac{1}{\|\nabla^2 f(\bar{x} + s(x_k - \bar{x}))^{-1}\|} ds \leq \|x_k - \bar{x}\| \|\nabla f(x_k)\|.$$

En utilisant  $\|\nabla^2 f(x_k)^{-1}\| \leq 2\|\nabla^2 f(\bar{x})^{-1}\|$ , on obtient  $\frac{\|x_k - \bar{x}\|^2}{2\|\nabla^2 f(\bar{x})^{-1}\|} \leq \|x_k - \bar{x}\| \|\nabla f(x_k)\|$ . Finalement, en rassemblant les majorations et minoration obtenues, on a pour  $x_k \in B(\delta)$   $\frac{\|x_k - \bar{x}\|}{2\|\nabla^2 f(\bar{x})^{-1}\|} \leq \|\nabla f(x_k)\| \leq 2\|\nabla^2 f(\bar{x})\| \|\bar{x} - x_k\|$  et la même inégalité pour  $x_0$  permet de conclure.  $\square$

Il existe des variantes inexactes de la méthode de Newton où

- le gradient  $\nabla f(x_k)$  est approximatif,
- le Hessien  $\nabla^2 f(x_k)$  est approximatif,
- la solution du système linéaire  $\nabla^2 f(x_k)s = \nabla f(x_k)$  est calculée de manière approchée,

dans le but de rendre la méthode moins coûteuse en mémoire et en temps de calcul. Pour toutes ces variantes, des théories de convergence locale existent, qui imposent un bon contrôle des approximations.

## I.2 Méthodes quasi-Newton

Une façon d'approximer la Hessienne, pour éviter de calculer et de stocker les dérivées d'ordre 2 est décrite comme suit. Pour une fonction quadratique, il est aisé de montrer que  $\nabla f(x_1) - \nabla f(x_2) = \nabla^2 f(x_1)(x_1 - x_2)$ . Cela indique que la connaissance de deux vecteurs distincts  $x_1$  et  $x_2$  et de la différence de gradient associée permet d'obtenir dans le cas quadratique -ou au voisinage de la solution sous les hypothèses standart, dans les étapes ultimes de la convergence- de l'information sur la Hessienne  $\nabla^2 f(x)$ . Plus généralement, on suppose connus,  $s = x_1 - x_2$  et  $y = \nabla f(x_1) - \nabla f(x_2)$ , ainsi qu'une approximation courante  $B$  de la Hessienne. On cherche une nouvelle approximation  $\tilde{B}$  telle que  $\tilde{B}$  soit symétrique et  $\tilde{B}s = y$ . Cela ne suffit pas pour définir de manière unique  $\tilde{B}$ , et on recherche des  $\tilde{B}$  de norme minimale (pour certaines normes) pour forcer l'unicité.

**Exercice 4.3** On recherche une matrice  $\tilde{B} = B + \Delta B$ , supposée mieux approcher que  $B$  la Hessienne en  $x_2$  en considérant le problème

$$\begin{aligned} \min \quad & \|\Delta B\|_F. \\ \Delta B = \Delta B^T \\ (B + \Delta B)s = y \end{aligned}$$

La solution de ce problème est donnée par la formule Powell-symmetric-Broyden :

$$\Delta B_0 = \frac{(y - Bs)s^T + s(y - Bs)^T}{s^T s} - \frac{s^T (y - Bs)s^T}{(s^T s)^2}.$$

**Démonstration** : On vérifie aisément que  $\Delta B_0 s = y - Bs$  et que  $\Delta B_0$  est symétrique. Soit  $q_1 = s / \|s\|_2$ . Pour tout  $\Delta B$  qui vérifie les contraintes (et en particulier pour  $\Delta B_0$ ), on a  $\Delta B q_1 = \Delta B_0 q_1 = \frac{y - Bs}{\|s\|_2}$ . Soient  $q_i, i = 2, \dots, n$ , qui complètent  $q_1$  en une base orthonormale de  $\mathbb{R}^n$ . Alors de  $q_i^T q_1 = 0$  pour  $i > 1$ , on tire  $\Delta B_0 q_i = \frac{s(\Delta B s)^T q_i}{s^T s} = \frac{ss^T}{s^T s} \Delta B q_i$ . D'où, en notant  $Q = [q_1, \dots, q_n]$ ,  $\|\Delta B_0 Q\|_F^2 = \sum_{i=1}^n \|\Delta B_0 q_i\|_2^2 \leq \sum_{i=1}^n \|\Delta B q_i\|_2^2 = \|\Delta B Q\|_F^2$ . En utilisant le fait que la norme de Frobenius est unitairement invariante, on obtient  $\|\Delta B_0\|_F \leq \|\Delta B\|_F$ , d'où le résultat.  $\square$

**Exercice 4.4** Soit  $f$  une fonction deux fois continûment dérivable, telle que  $\nabla^2 f(x)$  est définie positive pour tout  $x$ . Soit  $G = \int_0^1 \nabla^2 f(x_1 + s(x_2 - x_1)) ds$ . La matrice  $G$  est symétrique définie positive. Soit une matrice symétrique  $W$  telle que  $W^2 = G$ . On s'intéresse au problème

$$\begin{aligned} \min \quad & \|W^{-1} \Delta B W^{-1}\|_F. \\ \Delta B = \Delta B^T \\ (B + \Delta B)s = y \end{aligned}$$

La solution de ce problème est donnée par la formule de Davidon-Fletcher-Powell

$$\Delta B_0 = \frac{(y - Bs)y^T + y(y - Bs)^T}{s^T y} - \frac{s^T (y - Bs) \cdot yy^T}{(s^T y)^2}.$$

Noter qu'alors

$$B + \Delta B_0 = \left(I - \frac{ys^T}{s^T y}\right) B \left(I - \frac{sy^T}{s^T y}\right) + \frac{yy^T}{s^T y}.$$

**Démonstration** : Par Taylor avec reste integral, puisque  $s = x_1 - x_2$  et  $y = \nabla f(x_1) - \nabla f(x_2)$  que  $Gs = y$ . De plus  $G$  est définie positive (considérer  $\int_0^1 u^T \nabla^2 f(x_1 + s(x_2 - x_1)) u ds$  pour tout  $u$  de norme 1, et le fait que l'intégrande est continu et strictement positif). Donc  $s^T y = s^T Gs > 0$ . Soit alors  $W$  une racine carrée positive de  $G$  (en fait elle est unique). Par changement de variable  $\Delta = W^{-T} \Delta B W^{-1}$ , le problème devient

$$\begin{aligned} \min \quad & \|\Delta\|_F. \\ \Delta = \Delta^T \\ (W^{-1} B W^{-1} + \Delta) W s = W^{-1} y \end{aligned}$$

D'après l'exercice 4.3 précédent, et en notant que  $Gs = WWs = y$  et  $Ws = W^{-1}y$ , la solution s'écrit

$$\begin{aligned} \Delta_0 &= \frac{(W^{-1}y - W^{-1}BW^{-1}Ws)s^T W + s(W^{-1}y - W^{-1}BW^{-1}Ws)^T}{s^T WWs} \\ &\quad - \frac{s^T W(W^{-1}y - W^{-1}BW^{-1}Ws)W s^T W}{(s^T WWs)^2} \\ &= \frac{W^{-1}(y - Bs)y^T W^{-1} + W^{-1}y(y - Bs)^T W^{-1}}{s^T y} - \frac{s^T (y - BWs)W^{-1}yy^T W^{-1}}{(s^T y)^2}. \end{aligned}$$

En faisant le changement de variable  $\Delta B = W \Delta W$ , on obtient le résultat désiré.  $\square$

**Exercice 4.5** Nous avons vu que dans la méthode de Newton, il s'agit de résoudre des systèmes linéaires de la forme  $\nabla^2 f(x_k)s = \nabla f(x_k)$ . D'où l'idée d'approcher  $\nabla^2 f(x_k)^{-1}$  plutôt que  $\nabla^2 f(x_k)$ . Montrez que la formule BFGS (Broyden, Fletcher, Goldfarb, Shanno)

$$H + \Delta H_0 = \left(I - \frac{ys^T}{y^T s}\right) H \left(I - \frac{ys^T}{y^T s}\right) + \frac{ss^T}{y^T s},$$

est telle que  $\Delta H_0$  est solution de

$$\begin{aligned} \min_{\Delta H} \quad & \|\Delta H\|, \\ \Delta H = \Delta H^T \quad & \\ (H + \Delta H)y = s \quad & \end{aligned}$$

pour une norme  $\|\bullet\|$  que vous identifierez.

**Démonstration** : Dans la démonstration de l'exercice 4.4, on a démontré que si  $Gs = y$ , avec  $G = WW$  définie positive, alors la mise à jour DFP pour  $(B + \Delta B)s = y$  est solution du problème de mise à jour avec la norme  $\|W^{-1} \bullet W^{-1}\|_F$ . On considère maintenant l'équation  $(H + \Delta H)y = s$ . On peut appliquer DFP à ce problème en notant que  $s = G^{-1}y$  ( $G = W^{-1}W^{-1}$  est définie positive). La formule BFGS est alors la mise à jour de DFP correspondant au problème

$$\begin{aligned} \min_{\Delta H} \quad & \|W\Delta HW\|_F. \\ \Delta H = \Delta H^T \quad & \\ (H + \Delta H)y = s \quad & \end{aligned}$$

□

Deux principales difficultés sont rapportées dans la littérature sur la méthode de Newton pour la minimisation :

- (i) Son mauvais comportement lorsque le point de départ est loin de la solution sur des problèmes pour lesquels certains Hessiens  $\nabla^2 f(x_k)$  sont définis positifs.
- (ii) Son mauvais comportement lorsqu'elle rencontre des Hessiens ayant des valeurs propres négatives ou nulles.

Une amélioration possible pour le problème 1) est la mise en place de stratégies de recherches linéaires. Le point 2) est souvent appréhendé en utilisant des techniques de région de confiance.

### I.3 Globalisation des méthodes de Newton/quasi-Newton

**Exercice 4.6** Calculez quelques itérés de la méthode de Newton sur  $f(x) = -e^{-x^2}$ , pour  $x_0 = 10^{-1}$ ,  $x_0 = 1/2$  et  $x_0 = 1$ .

**Démonstration** :  $f(x) = -e^{-x^2}$ ,  $f'(x) = 2xe^{-x^2}$ ,  $f''(x) = (2 - 4x^2)e^{-x^2}$ . Alors on a  $x_{k+1} = x_k - 2x_k/(2 - 4x_k^2) = -4x_k^3/(2 - 4x_k^2)$ . Pour  $x_0 = 10^{-1}$ , on a  $x_1 \sim 2 \cdot 10^{-3}$  et  $x_2 \sim 2 \cdot 10^{-8}$ . Pour  $x_0 = 1/2$ , on a  $x_1 = -1/2$  et  $x_2 = 1/2$ . Pour  $x_0 = 1$ , on a  $x_1 \sim 2.3$  et  $x_2 \sim 2.5$ ,  $x_{23} \sim 5.4$  et  $f(x_{23}) \sim 10^{-13}$ .

□

Nous voyons dans la suite deux techniques visant à rendre la convergence moins dépendante du point de départ. Ces deux techniques sont appelées techniques de globalisation, et chercheront à approcher une convergence locale quadratique au voisinage des solutions de  $\nabla f(x) = 0$ . Ces solutions sont appelées points critiques du premier ordre.

#### Recherche linéaire

Dans cette section, on suppose que la fonction  $f$  est deux fois continûment dérivable.

**Définition I.1** Soit  $x_k \in \mathbb{R}^n$ . On dit que  $d_k$  est une direction de descente en  $x_k$  si  $\nabla f(x_k)^T d_k < 0$ .

La terminologie "direction de descente" s'explique aisément par l'exercice 4.7.

**Exercice 4.7** Si  $d_k$  est une direction de descente en  $x_k$ , alors il existe  $\eta > 0$  tel que

$$f(x_k + \alpha d_k) < f(x_k) \text{ pour tout } \alpha \in ]0, \eta].$$

**Démonstration** : Soit  $\phi(t) = f(x_k + td_k)$ . Alors  $\phi'(t) = \nabla f(x_k + td_k)^T d_k$ , donc comme  $\phi'$  est continue, et  $\phi'(0) < 0$ , il existe un intervalle  $]0, \eta]$  où  $\phi'(t) < 0$ . Alors pour  $t$  dans  $]0, \eta]$ , on a  $f(x_k + \alpha d_k) - f(x_k) = \int_{s=0}^t \phi'(s) ds < 0$ .

□

On envisage alors un premier algorithme de minimisation basé sur des directions de descente :

#### Basic linesearch (bad algorithm)

1. Choose  $x_0$
2. For  $k=0, 2, \dots$  Do
3.   Compute a descent direction such that  $\nabla f(x_k)^T d_k < 0$
4.   Compute a step such that  $f(x_k + \alpha_k d_k) < f(x_k)$ .
5.   Update  $x_{k+1} = x_k + \alpha_k d_k$ .
6. EndDo

**Exercice 4.8** L'algorithme ci-dessus ne suffit pas pour converger vers un minimum local de  $f$ . Soit  $f(x) = x^2$ ,  $x_0 = 2$ .

- (i) On choisit  $d_k = (-1)^{k+1}$  et  $\alpha_k = 2 + 3 \cdot 2^{-k-1}$ . Vérifier que  $x_k = (-1)^k(1 + 2^{-k})$  et que chaque direction  $d_k$  est de descente. Vérifier aussi que la suite ne converge pas, que  $f(x_{k+1}) < f(x_k)$  et que  $\lim_{k \rightarrow +\infty} f(x_k) = 1$ . Tracer les itérés et vérifier qu'entre deux itérés successifs, la décroissance de  $f$  est très petite par rapport au pas  $|\alpha_k d_k|$ .
- (ii) On choisit  $d_k = -1$  et  $\alpha_k = 2^{-(k+1)}$ . Vérifier que  $x_k = 1 + 2^{-k}$  et que chaque direction  $d_k$  est de descente. Vérifier aussi que la suite converge vers 1 (et pas vers 0) que  $f(x_{k+1}) < f(x_k)$  et que  $\lim_{k \rightarrow +\infty} f(x_k) = 1$ . Tracer les itérés et vérifier qu'entre deux itérés successifs, les pas  $|\alpha_k d_k|$  deviennent très petits par rapport à  $|f'(x_k) d_k|$ .

Démonstration :

- (i) Par récurrence,  $x_{k+1} = x_k + \alpha_k d_k = (-1)^k(1 + 2^{-k}) + (2 + 3 \cdot 2^{-k-1})(-1)^{k+1} = (-1)^{k+1}(1 + 2^{-(k+1)})$ . Direction de descente :  $f'(x_k) d_k = 2(-1)^k(1 + 2^{-k})(-1)^{k+1} < 0$ . La suite admet  $-1$  et  $1$  comme points d'accumulation et  $\lim_{k \rightarrow +\infty} f(x_k) = 1$ . De plus  $f(x_{k+1}) - f(x_k) = (1 + 2^{-k})^2 - (1 + 2^{-(k-1)})^2 < 0$ .
- (ii) Par récurrence,  $x_{k+1} = x_k + \alpha_k d_k = 1 + 2^{-k} - 2^{-k-1} = 1 + 2^{-(k+1)}$ . Direction de descente :  $f'(x_k) d_k = 2(1 + 2^{-k})(-1) < 0$ , et  $f(x_{k+1}) - f(x_k) < 0$ .

□

**Définition I.2** Soit  $\beta_1 \in ]0, 1[$ ,  $\beta_2 \in ]\beta_1, 1[$ , et soit  $d_k$  une direction de descente en  $x_k$ . On appelle conditions de Wolfe les deux conditions :

- (i)  $f(x_k + \alpha d_k) \leq f(x_k) + \beta_1 \alpha \nabla f(x_k)^T d_k$  (condition de diminution suffisante)
- (ii)  $\nabla f(x_k + \alpha d_k)^T d_k \geq \beta_2 \nabla f(x_k)^T d_k$  (condition de progrès suffisant)

Ces deux conditions pallient respectivement les deux types de problèmes rencontrés dans l'exercice 4.8. Si  $\alpha \rightarrow f(x_k + \alpha d_k)$  admet un minimum global, celui-ci vérifie les conditions de Wolfe (mais peut être très ou trop cher à calculer à des étapes préliminaires de convergence).

Démonstration :

- (i) Dans le cas 1.,  $f(x_k + \alpha_k d_k) - f(x_k) = (1 + 2^{-k-1})^2 - (1 + 2^{-k})^2 = -2^{-k-1}(2 + 3 \cdot 2^{-k-1})$  et  $\nabla f(x_k)^T d_k = -2(1 + 2^{-k})$ . Donc la condition de diminution suffisante n'est pas vérifiée.
- (ii) Dans le cas 2,  $\nabla f(x_k + \alpha_k d_k)^T d_k = -2x_{k+1}$  et  $\nabla f(x_k)^T d_k = -2x_k$ , et comme  $\{x_k\}$  tend vers 1, la condition de progrès suffisant n'est pas vérifiée.

□

**Exercice 4.9** Validité des conditions de Wolfe. Soient  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction différentiable, un point  $x_k \in \mathbb{R}^n$  et une direction (de descente)  $d_k \in \mathbb{R}^n$  telle que  $f$  est bornée inférieurement dans la direction  $d_k$  (c'est-à-dire il existe  $f_0$  tel que  $f(x_k + \alpha d_k) \geq f_0$  pour tout  $\alpha \geq 0$ ).

Pour  $0 < \beta_1 < 1$ , il existe  $\eta$  tel que la première condition de Wolfe soit vérifiée pour tout  $\alpha_k$ ,  $0 < \alpha_k \leq \eta$ . De plus, si  $0 < \beta_1 < \beta_2 < 1$ , il existe  $\alpha > 0$  tel que les deux conditions de Wolfe soient toutes deux vérifiées.

Démonstration : On s'intéresse aux  $\alpha > 0$  tels que  $f(x_k + \alpha d_k) = f(x_k) + \beta_1 \alpha \nabla f(x_k)^T d_k$ . Cet ensemble est non vide (car sinon  $\alpha \mapsto f(x_k + \alpha d_k)$  serait en dessous de  $\alpha \mapsto f(x_k) + \beta_1 \alpha \nabla f(x_k)^T d_k$ , ce qui est impossible car  $0 < \beta_1 < 1$  et  $f$  est bornée inférieurement), fermé (image réciproque de  $\{0\}$ ) et borné inférieurement. Donc cet ensemble admet un plus petit élément  $\alpha_1$ , qui vérifie

$$f(x_k + \alpha_1 d_k) = f(x_k) + \beta_1 \alpha_1 \nabla f(x_k)^T d_k,$$

donc qui vérifie la première condition de Wolfe.

D'après Taylor-Lagrange, appliqué à  $\alpha \mapsto f(x_k + \alpha d_k)$ , entre 0 et  $\alpha_1$ , il existe  $\alpha_2$  tel que

$$f(x_k + \alpha_1 d_k) = f(x_k) + \alpha_1 \nabla f(x_k + \alpha_2 d_k)^T d_k,$$

En rassemblant les deux résultats, on obtient

$$\nabla f(x_k + \alpha_2 d_k)^T d_k = \beta_1 \nabla f(x_k)^T d_k > \beta_2 \nabla f(x_k)^T d_k,$$

donc  $\alpha_2$  vérifie la seconde condition de Wolfe. Comme  $\alpha_2 < \alpha_1$ , on a  $f(x_k + \alpha_2 d_k) < f(x_k) + \beta_1 \alpha_2 \nabla f(x_k)^T d_k$  et donc  $\alpha_2$  vérifie la première condition de Wolfe est vérifiée.

□



## Descent algorithm with Wolfe linesearch

1. Choose  $x_0$
2. For  $k=0,2, \dots$  Do
3.   Compute a descent direction such that  $\nabla f(x_k)^T d_k < 0$
4.   Compute a step such that the Wolfe conditions hold.
5.   Update  $x_{k+1} = x_k + \alpha_k d_k$ .
6. EndDo

**Théorème I.1** *Supposons de  $f$  soit continûment différentiable, bornée inférieurement, et que son gradient vérifie  $\|\nabla f(x) - \nabla f(y)\|_2 \leq \gamma \|x - y\|_2$ . supposons qu'un algorithme de descente soit employé tel que chaque pas vérifie les conditions de Wolfe. Alors soit  $\lim_{k \rightarrow +\infty} \nabla f(x_k) = 0$ , soit  $\lim_{k \rightarrow +\infty} \frac{\nabla f(x_k)^T d_k}{\|d_k\|_2} = 0$ .*

Démonstration : Admise, voir Denis et Schnabel 1996, p.121.

□

Le théorème ci-dessus indique que si l'angle entre  $d_k$  et  $\nabla f(x_k)$  ne converge pas vers l'angle droit, la limite du gradient de l'itéré est 0 (on vérifie asymptotiquement la condition nécessaire du premier ordre) *quel que soit*  $x_0$ . C'est donc un résultat de convergence globale. Malheureusement cet algorithme peut avoir une convergence très lente si  $d_k$  n'est pas choisi avec soin. Par exemple, le choix  $d_k = -\nabla f(x_k)$  s'avère un très mauvais choix si l'algorithme converge vers un point  $x^*$  tel que  $\text{cond}(\nabla^2 f(x_k))$  est grand : la convergence est linéaire, avec une vitesse de convergence modeste.

Dans le cas d'une convergence vers un point  $x^*$  tel que  $\nabla^2 f(x^*)$  est défini positif (condition suffisante du second ordre), l'idée consiste alors à préconditionner la recherche linéaire et à la combiner avec la méthode de Newton qui est localement quadratiquement convergente, comme le fait l'algorithme ci-dessous. Il est possible de montrer que lorsque les itérés s'approchent d'une solution qui vérifie les conditions suffisantes d'optimalité au second ordre, le pas de Newton est accepté et la convergence est quadratique.

## Newton with linesearch

1. Choose  $x_0$
2. For  $k=0,2, \dots$  Do
3.   If  $\nabla^2 f(x_k)$  is SPD, compute the Newton step  $s^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$ .  
If  $s^N$  is acceptable (Wolfe) accept it. If not, perform a line search (Wolfe) in direction  $s^N$
4.   If  $\nabla^2 f(x_k)$  is not SPD, add a perturbation  $E$  so that  $\nabla^2 f(x_k) + E$  is SPD,  
and perform a line search (Wolfe) in direction  $-(\nabla^2 f(x_k)^{-1} + E) \nabla f(x_k)$
5.   Update  $x_{k+1} = x_k + \alpha_k d_k$ .
6. EndDo

## Région de confiance

**Définition I.3** *Modèle quadratique. On appelle modèle quadratique de  $f$  en  $x_k$  une fonction quadratique  $m_k(x_k + s)$  telle que  $m_k(x_k) = f(x_k)$  et  $\nabla m_k(x_k) = \nabla f_k(x_k)$ . Il existe alors une matrice  $H_k \in \mathbb{R}^{n \times n}$  telle que*

$$m_k(x_k + s) = f(x_k) + \nabla f_k(x_k)^T s + \frac{1}{2} s^T H_k s.$$

**Définition I.4** *Région de confiance. On appelle région de confiance Euclidienne centrée en  $x_k$ , de rayon  $\Delta_k > 0$  la sphère  $\mathcal{B}_k = x_k + \{s, \|s\|_2 \leq \Delta_k\}$ .*

L'idée de l'algorithme de région de confiance et de résoudre approximativement le problème

$$\min_{x_k + s \in \mathcal{B}_k} m_k(x_k + s).$$

On note  $x_{k+1} = x_k + s_k$  le point ainsi obtenu. La condition technique portant sur  $x_{k+1}$  demandée pour les résultats de convergence est la condition dite de *décroissante suffisante* :

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_{mdc} \|\nabla m_k(x_k)\|_2 \min \left( \frac{\|\nabla m_k(x_k)\|_2}{\beta_k}, \Delta_k \right), \quad (4.6)$$

où  $\kappa_{mdc} \in ]0, 1[$  et  $\beta_k = \|H_k(x)\|_2 + 1$ .

**Exercice 4.10** *Le point de Cauchy  $x_k^C$  qui est, par définition, solution de*

$$\begin{cases} t > 0 \\ x = x_k - t \nabla m(x_k) \in \mathcal{B}_k \end{cases} \min m_k(x)$$

vérifie

$$m_k(x_k) - m_k(x_k^C) \geq \frac{1}{2} \|\nabla m_k(x_k)\|_2 \min \left( \frac{\|\nabla m_k(x_k)\|_2}{\beta_k}, \Delta_k \right).$$

**Démonstration** : Posons  $g_k = \nabla_x m_k(x_k)$ . On a  $m_k(x_k - tg_k) = m_k(t_k) - t\|g_k\|^2 + \frac{1}{2}t^2 g_k^T H_k g_k$ .

(i) Supposons  $g_k^T H_k g_k > 0$ . Alors le minimum de  $m_k(x_k - tg_k)$  pour  $t \in \mathbb{R}$  est atteint en  $t^* = \frac{\|g_k\|^2}{g_k^T H_k g_k} \geq 0$ .

Premier cas. Supposons d'abord que  $t^* \|g_k\| = \frac{\|g_k\|^3}{g_k^T H_k g_k} \leq \Delta_k$ , donc  $x_k - t^* g_k$  est dans la région de confiance et c'est  $x_k^C$ . Comme  $g_k^T H_k g_k \leq \beta_k \|g_k\|$ , on a alors

$$\begin{aligned} m_k(x_k) - m_k(x_k^C) &= t^* \|g_k\|^2 - \frac{1}{2} t^{*2} g_k^T H_k g_k \geq \frac{\|g_k\|^4}{g_k^T H_k g_k} - \frac{1}{2} \frac{\|g_k\|^4}{(g_k^T H_k g_k)^2} g_k^T H_k g_k \\ &= \frac{1}{2} \frac{\|g_k\|^4}{g_k^T H_k g_k} \geq \frac{1}{2} \frac{\|g_k\|^2}{\beta_k}. \end{aligned}$$

Deuxième cas. Supposons maintenant que  $\frac{\|g_k\|^3}{g_k^T H_k g_k} \geq \Delta_k$ . Alors  $g_k^T H_k g_k \leq \frac{\|g_k\|^3}{\Delta_k}$  et le minimum dans la région de confiance est donc atteint sur la frontière (faire un dessin). Alors  $t^* \|g_k\| = \Delta_k$  et  $x_k^C = x_k - \Delta_k g_k$  et

$$m_k(x_k) - m_k(x_k^C) = \Delta_k \|g_k\| - \frac{1}{2} \frac{\Delta_k^2}{\|g_k\|^2} g_k^T H_k g_k \geq \Delta_k \|g_k\| - \frac{1}{2} \frac{\Delta_k^2}{\|g_k\|^2} \frac{\|g_k\|^3}{\Delta_k} = \frac{1}{2} \Delta_k \|g_k\|.$$

(ii) Supposons  $g_k^T H_k g_k \leq 0$ . Le minimum est à nouveau atteint sur la frontière de la région de confiance et puisque  $-g_k^T H_k g_k \geq 0$

$$m_k(x_k) - m_k(x_k^C) = \Delta_k \|g_k\| - \frac{1}{2} \frac{\Delta_k^2}{\|g_k\|^2} g_k^T H_k g_k \geq \Delta_k \|g_k\|.$$

En regroupant les différents sous-cas, on obtient le résultat. □

Le calcul de  $x_{k+1}$  (donc de  $s_k$ ) est bien moins cher que la résolution du problème initial  $\min_x f(x)$  car

- (i)  $m_k$  est une fonction quadratique
- (ii) la décroissance suffisante est obtenue à faible coût, en calculant le point de Cauchy, et en cherchant éventuellement à diminuer encore  $m_k$  à partir de  $x_k^C$ . La méthode des régions de confiance a donc un rapport étroit avec la recherche linéaire suivant la direction  $-\nabla f_k(x_k)$ .

On introduit le ratio de la réduction observée sur  $f$  par rapport à la réduction prédite sur  $m_k$  :

$$\rho_k = \frac{f(x_k) - f(x_{k+1})}{m(x_k) - m(x_{k+1})}.$$

Si  $\rho_k$  est suffisamment proche de 1, le modèle représente la fonction de manière fiable, on accepte le pas, et on augmente éventuellement le rayon de la région de confiance. Si  $\rho_k$  est faible, voire négatif, le modèle n'est pas assez fiable, et l'on réduit la région de confiance (notez que pour  $\Delta_k$  suffisamment petit modèle et fonction sont égaux au premier ordre). Nous sommes en mesure de présenter à présent l'algorithme des régions de confiance :

Basic trust region algorithm [1]

1. Choose  $x_0$ , an initial  $\Delta_0 > 0$ , and constants  $0 < \eta_1 \leq \eta_2 < 1$  and  $0 < \gamma_1 \leq \gamma_2 < 1$
2. For  $k=0, 1, \dots$  Do
3.   Compute a step  $s_k$  that *sufficiently* reduces  $m_k$  in  $\mathcal{B}_k$  (4.6).
4.   Define  $\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m(x_k) - m(x_k + s_k)}$ .
5.   If  $\rho_k \geq \eta_1$  then define  $x_{k+1} = x_k + s_k$ ; otherwise define  $x_{k+1} = x_k$
6.   Trust region update. Set
  - $\Delta_{k+1} \in [\Delta_k, +\infty[$  if  $\rho_k \geq \eta_2$  or
  - $\Delta_{k+1} \in [\gamma_2 \Delta_k, \Delta_k]$  if  $\eta_1 \leq \rho_k < \eta_2$  or
  - $\Delta_{k+1} \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k]$  if  $\rho_k < \eta_1$
7.   If converged, exit,
8. EndDo

**Théorème I.2** On suppose que l'algorithme est appliqué à une fonction

- deux fois différentiable,
- bornée inférieurement sur  $\mathbb{R}^n$ ,
- à Hessian borné ( $\|\nabla^2 f(x)\|_2 \leq \kappa_{ufh}$  pour  $x \in \mathbb{R}^n$ ),

et que les modèles  $m_k$  sont

- quadratiques,
- ont même valeur et gradient que  $f$  en  $x_k$  (cohérence au premier ordre)
- ont des Hessian bornés ( $\|\nabla^2 f(x)\|_2 \leq \kappa_{umh}$  pour  $x \in \mathcal{B}_k$ ).

alors pour tout  $x_0$ , l'algorithme des régions de confiance produit une suite d'itérés telle que  $\lim_{k \rightarrow +\infty} \nabla f(x_k) = 0$ .

Démonstration : Admise (Conn, Gould, Toint (2000 p.136)).

□

Le théorème I.2 montre une manière aisée d'obtenir un algorithme globalement convergent : il suffit de choisir  $\nabla^2 m_k(x_k) = H_k = 0 \in \mathbb{R}^{n \times n}$  et de prendre pour itéré le point de Cauchy. Par contre on obtient alors un algorithme qui converge aussi peu rapidement que celui implantant systématiquement la recherche linéaire dans la direction  $-\nabla f(x_k)$ . Pour obtenir un algorithme plus performant et approcher la convergence locale de l'algorithme de Newton, il convient de choisir un pas  $s_k$  qui soit voisin du pas de Newton dans les étapes ultimes de la convergence.

Ceci est réalisé si l'on utilise pour algorithme de calcul de pas l'algorithme de gradient conjugué tronqué proposé par Steihaug et Toint et si le Hessian du modèle approche celui de la fonction. Cet algorithme commence par calculer le point de Cauchy puis poursuit la minimisation de la quadratique  $m(x_k + s)$  par la méthode des gradients conjugués, en s'arrêtant au premier itéré sortant de la région de confiance  $\mathcal{B}_k$ . On a ainsi minimisé davantage  $m(x_k + s)$  que  $m(x_k^C)$ , et donc on a, à la fin de cette procédure de gradient conjugué tronqué, la décroissance suffisante :

$$m(x_k) - m(x_k + s_k) \geq m(x_k) - m(x_k^C) \geq \frac{1}{2} \|\nabla_x m_k(x_k)\|_2 \min \left( \frac{\|\nabla_x m_k(x_k)\|_2}{\beta_k}, \Delta_k \right).$$

Dans le cas où la convergence a lieu vers un point  $x^*$  où le Hessian est défini positif et si  $\nabla^2 m_k(x_k) \sim \nabla^2 f_k(x_k)$ , le comportement typique de l'algorithme est alors le suivant :

- (i) les pas deviennent de plus en plus petits (on converge),
- (ii) comme le modèle et la fonction sont cohérents au premier ordre,  $\rho_k$  devient proche de 1,
- (iii) la région de confiance a un rayon qui augmente,
- (iv) l'algorithme des gradients conjugués ne rencontre plus le bord de la région de confiance,
- (v) les gradient conjugués résolvent alors le système  $\nabla^2 f(x_k)s_k + \nabla f(x_k) = 0$  ce qui correspond bien à la méthode de Newton, qui a une convergence locale quadratique.

#### Truncated Conjugate Gradient algorithm

0. Input parameters :  $x_0$ . Output :  $s$
1. Compute  $s_0 = 0$ ,  $g_0 = \nabla f(x_0)$ ,  $p_0 = -g_0$
2. For  $k=0, 1, \dots$  Do
3.  $\kappa_k = p_k^T H p_k$
4. If  $\kappa_k \leq 0$ , then
  - compute  $\sigma_k$  the root of  $\|s_k + \sigma p_k\|_2 = \Delta_k$
  - for which  $m_k(s_k + \sigma p_k)$  is the smallest.
  - $s_{k+1} = s_k + \sigma_k p_k$  and stop.
- End If
5.  $\alpha_k = g_k^T g_k / \kappa_k$
6. If  $\|s_k + \alpha_k p_k\|_2 \geq \Delta_k$ , then
  - compute  $\sigma_k$  as the positive root of  $\|s_k + \sigma p_k\|_2 = \Delta_k$
  - $s_{k+1} = s_k + \sigma_k p_k$  and stop.
- End If
4.  $s_{k+1} = s_k + \alpha_k p_k$
5.  $g_{k+1} = g_k + \alpha_k H p_k$
7.  $\beta_k = g_{k+1}^T g_{k+1} / g_k^T g_k$
8.  $p_{k+1} = -g_{k+1} + \beta_k p_k$
9. if converged then stop
10. EndDo

### I.4 Globalisation des moindres carrés non-linéaires

**Exercice 4.11** *Fonctionnelle des moindres carrés non linéaires.* Soit  $f$  définie sur un ouvert  $\mathcal{O} \subset \mathbb{R}^n$ , deux fois différentiable, à valeurs dans  $\mathbb{R}^m$ . On définit la fonction  $F(x)$  des moindres carrés non linéaires par  $F(x) = \frac{1}{2} \|f(x)\|_2^2$ . Montrez que le gradient de  $F$  en  $x$  est  $f'(x)^T f(x) = D_f(x)^T f(x)$  et que la matrice Hessienne de  $F$  en  $x$  est  $D_f(x)^T D_f(x) + \sum_{i=1}^m f_i(x) \nabla^2 f_i(x)$ .

**Démonstration** : Considérons  $\phi(x) = f_i(x)^2$ . Alors, par dérivation d'une composée,  $\frac{\partial \phi(x)}{\partial x_j} = 2f_i(x) \frac{\partial f_i(x)}{\partial x_j}$ , et donc  $\frac{\partial F(x)}{\partial x_j} = \sum_{i=1}^m \frac{\partial f_i(x)}{\partial x_j} f_i(x)$ , ce qui implique

$$\nabla F(x) = \begin{pmatrix} \frac{\partial F(x)}{\partial x_1} \\ \vdots \\ \frac{\partial F(x)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1(x)}{\partial x_n} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix} \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} = f'(x)^T f(x) = D_f(x)^T f(x).$$

Pour la dérivée seconde, si on note  $\psi(x) = 2f_i(x) \frac{\partial f_i(x)}{\partial x_j}$ , on a

$$\frac{\partial^2 \phi(x)}{\partial x_k \partial x_j} = \frac{\partial \psi(x)}{\partial x_k} = 2 \frac{\partial f_i(x)}{\partial x_k} \frac{\partial f_i(x)}{\partial x_j} + 2f_i(x) \frac{\partial^2 f_i(x)}{\partial x_k \partial x_j}.$$

On a alors  $\frac{\partial^2 F(x)}{\partial x_k \partial x_j} = \sum_{i=1}^m \frac{\partial f_i(x)}{\partial x_k} \frac{\partial f_i(x)}{\partial x_j} + f_i(x) \frac{\partial^2 f_i(x)}{\partial x_k \partial x_j}$ . Ce terme est bien le terme  $(k, l)$  de la matrice  $D_f(x)^T D_f(x) + \sum_{i=1}^m f_i(x) \nabla^2 f_i(x)$ .

□

Nous avons vu dans l'exercice 4.11 que pour la fonction des moindres carrés non linéaires,  $F(x) = \frac{1}{2} \|f(x)\|_2^2$ , le gradient de  $F$  en  $x$  est  $f'(x)^T f(x) = D_f(x)^T f(x)$  et la matrice Hessienne de  $F$  en  $x$  est  $D_f(x)^T D_f(x) + \sum_{i=1}^m f_i(x) \nabla^2 f_i(x)$ . Il est possible donc d'utiliser des variantes de la méthode de Newton pour minimiser  $F(x)$ , en utilisant une recherche linéaire ou une région de confiance.

On remarque que  $\nabla^2 f(x)$  s'écrit sous la forme d'un terme ne faisant intervenir que des dérivations ( $D_f(x)^T D_f(x)$ ) et un terme faisant intervenir des dérivations d'ordre 2 ( $\sum_{i=1}^m f_i(x) \nabla^2 f_i(x)$ ). Il est donc tentant d'approcher  $\nabla^2 f(x)$  par le terme  $D_f(x)^T D_f(x)$  pour éviter le calcul de dérivées d'ordre 2. La variante de Newton faisant cette approximation s'appelle la méthode de Gauss-Newton

$$(GN) : x_{k+1} = x_k - (D_f(x_k)^T D_f(x_k))^{-1} D_f(x_k)^T D_f(x_k) = x_k - D_f(x_k)^+ f(x_k).$$

Cette méthode n'est *même pas* toujours localement convergente (il existe des points fixes répulsifs). En la globalisant par une recherche linéaire où des régions de confiance on obtient des méthodes globalement convergentes très utilisées en pratique.



## Chapitre 5

# Quelques algorithmes pour l'optimisation avec contraintes

## I Introduction

Nous allons introduire dans ce chapitre deux idées importantes pour la résolution de problèmes contraints. La technique de pénalisation qui permet de résoudre un problème contraint en introduisant une suite de problèmes contraints sera abordée dans le cas de la programmation quadratique. Une autre idée sera la notion de détection de contraintes actives.

## II Méthode des contraintes actives

### II.1 Multiplicateurs de Lagrange et sensibilité

Nous énonçons pour commencer sans démonstration un résultat de sensibilité par rapport à une modification des contraintes. Ce résultat, important en soi, permettra de justifier la stratégie algorithmique de la méthode.

#### Proposition II.1.1

(Interprétation des multiplicateurs de Lagrange) On considère les problèmes

$$\mathcal{P}_u : \min_{\substack{h(x)=u \\ g(x) \leq v}} f(x), \text{ et } \mathcal{P} : \min_{\substack{h(x)=0 \\ g(x) \leq 0}} f(x),$$

et on pose  $\phi(u) = \inf\{f(x), h(x) = u, g(x) \leq v\}$ . On suppose que  $f$  et  $h$  sont deux fois continûment dérivables dans un voisinage de  $\bar{x}$  sachant que

- (i) le point  $\bar{x}$  est un point régulier,
- (ii) le point  $\bar{x}$ , de contraintes actives  $I$ , vérifie les conditions suivantes d'optimalité locale

$$\begin{cases} \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}) = f'(\bar{x}) + \bar{\lambda}^T h'(\bar{x}) + \bar{\mu}^T g'(\bar{x}) = 0, \\ h(\bar{x}) = 0, g(\bar{x}) \geq 0 \\ \bar{\mu}_j \geq 0, j = 1, \dots, p, \\ \bar{\mu}_j g_j(\bar{x}) = 0, j = 1, \dots, p, \bar{\mu}_j > 0, j \in I, \end{cases}$$

et pour tout  $\phi \in \text{Ker} h'(\bar{x}) \cap \text{Ker} g'_I(\bar{x})$ ,

$$\phi^T \frac{\partial^2 L(\bar{x}, \bar{\lambda})}{\partial x^2} \phi > 0.$$

Le point  $\bar{x}$  est alors une solution locale de  $\mathcal{P}$ ,

Alors il existe un voisinage de  $(u, v) = 0 \in \mathbb{R}^{m+p}$ , où  $\mathcal{P}_u$  admet une solution locale  $x(u, v)$  et des multiplicateurs de Lagrange associés  $\lambda(u, v)$  et  $\mu(u, v)$ . La fonction  $(u, v) \mapsto f(x(u, v)) = \phi(u, v)$  est alors dérivable en  $(u, v) = 0$  et on a  $\phi(u, v) = \phi(0, 0) - \lambda(0)^T u - \mu(0)^T v + o((u, v))$ .

#### Proposition II.1.2

Dans une usine, deux produits  $u_i$ ,  $i = 1, 2$  sont fabriqués, et rapportent par unité,  $e_i$  kilo euros en nécessitant  $t_i$  heures de travail machines et  $q_i$  tonnes de matières premières. On dispose de 10 heures en tout de travail machines, et de 15 tonnes de matières premières. Formaliser ce problème sous la forme d'un problème d'optimisation et le résoudre, pour  $(e_1, t_1, q_1) = (6, 2, 1)$  et  $(e_2, t_2, q_2) = (5, 1, 3)$ . Est-il intéressant, financièrement, d'augmenter la quantité de matière premières ? Jusqu'à quel point ?

Faire un dessin.

$$\begin{aligned} \min \quad & -6x_1 - 5x_2. \\ \text{s.t.} \quad & 2x_1 + x_2 \leq 10 \\ & x_1 + 3x_2 \leq 15 \\ & -x_1 \leq 0 \\ & -x_2 \leq 0 \end{aligned}$$

On voit sur un dessin que les contraintes actives à la solutions seront les deux premières contraintes. La solution du problème est donnée par les points critiques de la fonction  $L(x, \mu_1, \mu_2) = -6x_1 - 5x_2 + \mu_1(2x_1 + x_2 - 10) + \mu_2(x_1 + x_2 - 15)$ . La solution est donnée par le système linéaire

$$\begin{cases} -6 + 2\mu_1 + \mu_2 = 0 \\ -5 + \mu_1 + 3\mu_2 = 0 \\ x_1 + 3x_2 = 15 \end{cases},$$

ce qui donne  $(x_1, x_2, \mu_1, \mu_2) = (3, 4, 13/5, 4/5)$ . Si on augmente les matières premières de 15 à  $15 + M$ , le gain augmente de  $4/5M$ . Par contre pour  $M > 15$ , la seconde contrainte cesse d'être active. Dans ce cas, il ne sert à plus rien d'augmenter les matières premières, il faut augmenter aussi les 10 heures machines.

□

## II.2 Application de la théorie des multiplicateurs de Lagrange : la méthode des contraintes actives

Nous avons vu que la résolution du problème d'optimisation avec contraintes d'égalité et d'inégalités se ramène à la résolution d'un problème avec contraintes d'égalité lorsque les contraintes actives à la solution sont connues. Le principe de la méthode des contraintes actives est de créer une suite  $(x^{(k)}, I^{(k)})$  contenant un itéré et une estimation des contraintes actives. Sous des hypothèses de convexité du problème, il est possible de montrer que cette méthode est convergente.

Nous présentons ici le passage de  $(x^{(k)}, I^{(k)})$  à  $(x^{(k+1)}, I^{(k+1)})$  dans le cas d'un problème quadratique en  $x$

$$\begin{aligned} \min_{\substack{Ax=b \\ Cx-f \leq 0}} \quad & \frac{1}{2}x^T Hx + x^T g. \end{aligned}$$

(i) Résolution du problème en la variable  $d$

$$\begin{aligned} \min_{\substack{A(x^{(k)} + d) = b \\ (C(x^{(k)} + d) - f)_i = 0, \text{ pour } i \in I^{(k)}}} \quad & \frac{1}{2}(x^{(k)} + d)^T H(x^{(k)} + d) + (x^{(k)} + d)^T g, \end{aligned}$$

pour obtenir  $d^{(k)}$ ,  $\lambda^{(k)}$ , et  $\mu^{(k)}$ .

(ii) Mise à jour de  $x$  et  $I$

(a) Si  $d^{(k)} = 0$ , et  $\mu^{(k)} \geq 0$ , arrêt.

(b) Si  $d^{(k)} = 0$ , et le vecteur  $\mu^{(k)}$  a au moins une composante négative, on choisit  $j_k = \operatorname{argmin}_j \mu_j^{(k)}$ , et on pose  $I^{(k+1)} = I^{(k)} \setminus \{j_k\}$ , et  $x^{(k+1)} = x^{(k)}$ .

(c) Si  $d^{(k)} \neq 0$ .

i. Si  $x^{(k)} + d^{(k)}$  appartient à l'ensemble des contraintes,  $\mathcal{C} = \{x, h(x) = 0, \text{ et } g(x) \leq 0\}$ , on définit  $x^{(k+1)} = x^{(k)} + d^{(k)}$ .

ii. Sinon, on calcule le plus grand  $t \in [0, 1]$  tel que  $x^{(k)} + td^{(k)} \in \mathcal{C}$ . Soit  $t_{k+1}$  ce scalaire. On pose alors  $x^{(k+1)} = x^{(k)} + t_k d^{(k)}$ .

Dans ces deux cas,  $(x^{(k)} + d^{(k)})$  appartient ou non à  $\mathcal{C}$ , l'ensemble  $I^{(k+1)}$  est obtenu en rajoutant à  $I^{(k)}$  l'une des contraintes rendues nouvellement actives en  $x^{(k+1)}$ , s'il en existe une (contraintes activées par le pas). Sinon, si aucune contrainte n'est activée par le pas,  $I^{(k+1)} = I^{(k)}$ .

### Proposition II.2.1

Appliquer itérativement l'algorithme ci-dessous au problème

$$\begin{aligned} \min_{\substack{x \geq 0; y \geq 0; x \leq 2 \\ x + 2(y - 2) \leq 0}} \quad & (x - 1)^2 + (y - 2)^2, \end{aligned}$$

en partant de  $x^{(0)} = (2, 0)$  et  $I^{(0)} = \{2, 3\}$

Le problème s'écrit sous forme standard

$$\begin{aligned} \min \quad & (x-1)^2 + (y-2)^2. \\ \text{s.t.} \quad & -x \leq 0 \\ & -y \leq 0 \\ & x \leq 2 \\ & x + 2(y-2) \leq 0 \end{aligned}$$

□

1. Le lagrangien du problème s'écrit

$$L(d_1, d_2, \mu_2, \mu_3) = (1 + d_1)^2 + (d_2 - 2)^2 + \mu_2(-d_2) + \mu_3(d_1).$$

La condition d'optimalité donne

$$\begin{cases} 2d_1 + 2 + \mu_3 = 0 \\ 2d_2 - 4 - \mu_2 = 0 \\ d_2 = 0 \\ 2 + d_1 = 2 \end{cases},$$

ce qui montre que  $(d_1, d_2, \mu_2, \mu_3) = (0, 0, -4, -2)$ .

2.b  $x^{(1)} = (2, 0)^T$ ,  $I^{(1)} = \{3\}$ .

1. Le lagrangien du problème s'écrit

$$L(d_1, d_2, \mu_2, \mu_3) = (1 + d_1)^2 + (d_2 - 2)^2 + \mu_3(d_1).$$

La condition d'optimalité donne

$$\begin{cases} 2d_1 + 2 + \mu_3 = 0 \\ 2d_2 - 4 = 0 \\ 2 + d_1 = 2 \end{cases},$$

ce qui montre que  $(d_1, d_2, \mu_3) = (0, 2, -2)$ .

2.c.ii Comme  $(2, 0) + (0, 2)$  n'est pas dans le domaine, on cherche le plus grand  $0 \leq t \leq 1$  tel que

$$\begin{cases} -2 \leq 0 \\ -2t \leq 0 \\ 2 \leq 2 \\ 2 + 2(2t - 2) \leq 0 \end{cases},$$

On obtient alors  $t = 1/2$ ,  $x^{(3)} = (2, 0) + 1/2(0, 2) = (2, 1)$ , et on a activé la contrainte 4. Donc  $I^{(3)} = \{3, 4\}$ .  
item[1.] Le lagrangien du problème s'écrit

$$L(d_1, d_2, \mu_3, \mu_4) = (1 + d_1)^2 + (d_2 - 1)^2 + \mu_3(d_1) + \mu_4(d_1 + 2d_2).$$

La condition d'optimalité donne

$$\begin{cases} 2d_1 + 2 + \mu_3 + \mu_4 = 0 \\ 2d_2 - 2 + 2\mu_4 = 0 \\ d_1 = 0 \\ d_1 + 2d_2 = 0 \end{cases},$$

ce qui montre que  $(d_1, d_2, \mu_3, \mu_4) = (0, 0, -3, 1)$ .

2.b On enlève la contrainte 3.  $x^{(4)} = (2, 1)^T$ ,  $I^{(4)} = \{4\}$ . item[1.] Le lagrangien du problème s'écrit

$$L(d_1, d_2, \mu_3, \mu_4) = (1 + d_1)^2 + (d_2 - 1)^2 + \mu_4(d_1 + 2d_2).$$

La condition d'optimalité donne

$$\begin{cases} 2d_1 + 2 + \mu_4 = 0 \\ 2d_2 - 2 + 2\mu_4 = 0 \\ d_1 + 2d_2 = 0 \end{cases},$$

ce qui montre que  $(d_1, d_2, \mu_4) = 1/5(-6, 3, 2)$ .

2.c.i  $x^{(5)} = (4, 8)/5^T$ , appartient à l'ensemble des contraintes et  $I^{(5)} = \{4\}$ . Donc on ne bouge pas. item[1.] Comme on stationne, la solution est  $(d_1, d_2, \mu_4) = 1/5(0, 0, 2)$ , et on a convergé.



### III Pénalisation d'un problème quadratique à contraintes d'égalité

Nous voyons ici un algorithme servant à résoudre des problèmes quadratiques à contraintes d'égalités. Cet algorithme est loin d'être le seul possible, mais les autres techniques sortent du cadre de ce cours. On s'intéresse à

$$\mathcal{P} : \min_{Bx=0} \frac{1}{2} x^T A x - x^T b,$$

où  $A \in \mathbb{R}^{n \times n}$  est symétrique définie positive, et  $B \in \mathbb{R}^{m \times n}$  est surjective (i.e. de rang maximum  $m$ ). Ce sous-problème intervient dans les méthodes SQP où la fonction est représentée par un modèle quadratique, et les contraintes sont linéarisées.

#### Proposition III.0.1

Vérifiez l'hypothèse de qualification des contraintes et montrez que le système KKT associé à ce problème est

$$\mathcal{KKT} : \begin{cases} Ax + B^T \lambda = b \\ Bx = 0 \end{cases} \quad (5.1)$$

Montrez toute solution de ce système, est solution du problème  $\mathcal{P}$ .

Démonstration : On introduit  $L(x, \lambda) = \frac{1}{2} x^T A x - x^T b + \lambda^T Bx$ . On a alors

$$\begin{cases} \frac{\partial L}{\partial x}(x, \lambda) = x^T A - b^T + \lambda^T B = 0, \\ \frac{\partial L}{\partial \lambda}(x, \lambda) = x^T B^T = 0, \end{cases}$$

ce qui s'écrit encore en transposant

$$\begin{cases} Ax + B^T \lambda = b \\ Bx = 0 \end{cases}$$

Si une solution du système existe et est unique, elle vérifie la condition suffisante du second ordre car  $\frac{\partial^2 L}{\partial x^2}(x, \lambda) = A$  est définie positive. □

#### Proposition III.0.2

Montrez que le système KKT admet une unique solution, et donc que l'unique solution de KKT est l'unique solution de  $\mathcal{P}$ .

Démonstration : Pour cela il suffit de montrer que la matrice du système linéaire est injective et carrée, donc inversible. Supposons que

$$\begin{cases} Ax + B^T \lambda = 0 \\ Bx = 0 \end{cases}$$

En multipliant la première équation par  $x^T$ , il vient  $x^T A x + (Bx)^T \lambda = 0$ , c'est à dire  $x^T A x = 0$  puisque  $Bx = 0$ . Comme  $A$  est définie positive, il vient  $x = 0$  donc  $B^T \lambda = 0$ . Comme  $B^T$  est injective, on a  $\lambda = 0$ , dont le noyau est réduit au vecteur nul. □

Pour le reste de l'énoncé on suppose que  $\epsilon$  est un réel strictement positif.

#### Proposition III.0.3

Montrez que la solution du système

$$\begin{cases} Ax^\epsilon + B^T \lambda^\epsilon = b \\ Bx^\epsilon - \epsilon \lambda^\epsilon = 0 \end{cases} \quad (5.2)$$

existe et est unique. Par élimination de la variable  $\lambda^\epsilon$ , montrez que  $x^\epsilon$  est solution d'une équation

$$A^\epsilon x^\epsilon = b^\epsilon, \quad (5.3)$$

où  $A^\epsilon \in \mathbb{R}^{m \times n}$ , et  $b^\epsilon \in \mathbb{R}^n$ . Montrez que  $A^\epsilon$  est symétrique et définie positive. On remarquera que le système (5.3) est de dimension plus petite que le système (5.2), mais qu'il peut être plus mal conditionné pour  $\epsilon$  petit.

Démonstration : En remplaçant la seconde équation de (5.3) dans la première, on obtient  $(A + \frac{1}{\epsilon} B^T B)x^\epsilon = b^\epsilon$ . d'où  $x^{\epsilon T} A^\epsilon x^\epsilon = x^{\epsilon T} A x^\epsilon + \frac{1}{\epsilon} \|Bx^\epsilon\|_2^2 \geq 0$  car somme de termes positifs. Si  $x^{\epsilon T} A^\epsilon x^\epsilon = 0$ , alors  $x^{\epsilon T} A x^\epsilon = 0$  donc  $x^\epsilon = 0$ , ce qui prouve que  $A^\epsilon$  est définie positive. Le système est mal conditionné pour  $\epsilon$  petit, car si  $m < n$ ,  $B^T B$  est singulière, et on démontrerait que le conditionnement de  $A^\epsilon$  se comporte asymptotiquement comme celui de  $\frac{1}{\epsilon} B^T B$  (i.e. tend vers  $+\infty$ ).

□

**Proposition III.0.4**

Montrez que  $x^\epsilon$  est solution de (5.3) si et seulement si  $x^\epsilon$  est solution de

$$\min_{x^\epsilon \in \mathbb{R}^n} \frac{1}{2} x^{\epsilon T} A x^\epsilon + \frac{1}{2\epsilon} \|Bx^\epsilon\|_2^2 - x^{\epsilon T} b.$$

Interpréter ce résultat comme la résolution d'un problème d'optimisation avec contrainte par pénalisation de la contrainte.

Démonstration : Le problème d'optimisation est simplement  $\min_{x^\epsilon \in \mathbb{R}^n} \frac{1}{2} x^{\epsilon T} A x^\epsilon - x^{\epsilon T} b$ , et comme  $A^\epsilon$  est définie positive, la condition nécessaire et suffisante d'optimalité est bien  $A^\epsilon x^\epsilon = b$ . Lorsque  $\epsilon$  est petit, le minimum sera atteint vraisemblablement pour  $\|Bx^\epsilon\|_2$  petit. On dit qu'on a pénalisé la contrainte, du problème  $\mathcal{P}$ .

□

**Proposition III.0.5**

On suppose que  $(x, \lambda)$  et  $(x^\epsilon, \lambda^\epsilon)$  sont solutions respectives de (5.1) et (5.2). On s'intéresse à la limite de  $(x^\epsilon, \lambda^\epsilon)$  pour  $\epsilon \rightarrow 0$ .

(i) Montrez que  $\bar{x}^\epsilon = x^\epsilon - x$  et  $\bar{\lambda}^\epsilon = \lambda^\epsilon - \lambda$  vérifient

$$\begin{cases} A\bar{x}^\epsilon + B^T \bar{\lambda}^\epsilon &= 0 \\ B\bar{x}^\epsilon - \epsilon \bar{\lambda}^\epsilon &= \epsilon \lambda \end{cases} \quad (5.4)$$

(ii) En déduire que l'on a

$$\alpha \|\bar{x}^\epsilon\|_2^2 + \epsilon \|\bar{\lambda}^\epsilon\|_2^2 \leq \epsilon \|\lambda\|_2 \|\bar{\lambda}^\epsilon\|_2, \quad (5.5)$$

où  $\alpha > 0$  est la plus petite valeur propre de  $A$ .

(iii) Montrez que la matrice  $BA^{-1}B^T$  est définie positive. Soit  $\beta$  sa plus petite valeur propre. En repartant de (5.4), montrez que

$$(BA^{-1}B^T + \epsilon I)\bar{\lambda}^\epsilon = -\epsilon \lambda,$$

et en déduire que

$$\beta \|\bar{\lambda}^\epsilon\|_2 \leq \epsilon \|\lambda\|_2, \quad (5.6)$$

puis que

$$\sqrt{\alpha\beta} \|\bar{x}^\epsilon\|_2 \leq \epsilon \|\lambda\|_2. \quad (5.7)$$

(iv) Déduire des questions précédentes que  $\lim_{\epsilon \rightarrow 0} \lambda^\epsilon = \lambda$  et  $\lim_{\epsilon \rightarrow 0} x^\epsilon = x$ , et que l'erreur se comporte en  $O(\epsilon)$ .

Démonstration :

(i) Il suffit de faire les différences equation à equation dans les systèmes (5.1) et (5.2) pour obtenir le système (5.4).

(ii) En multipliant la première équation de (5.4) à gauche par  $\bar{x}^\epsilon$ , puis la seconde à gauche par  $-\bar{\lambda}^\epsilon$ , et en sommant, on obtient  $\bar{x}^{\epsilon T} A \bar{x}^\epsilon + \bar{x}^\epsilon B^T \bar{\lambda}^\epsilon - \bar{\lambda}^{\epsilon T} B \bar{x}^\epsilon + \epsilon \|\bar{\lambda}^\epsilon\|_2^2 = -\epsilon \bar{\lambda}^{\epsilon T} \lambda$ . On obtient que  $\alpha \|\bar{x}^\epsilon\|_2^2 \leq \bar{x}^{\epsilon T} A \bar{x}^\epsilon$  ce qui montre que

$$0 \leq \alpha \|\bar{x}^\epsilon\|_2^2 + \epsilon \|\bar{\lambda}^\epsilon\|_2^2 = -\epsilon \bar{\lambda}^{\epsilon T} \lambda = \epsilon |\bar{\lambda}^{\epsilon T} \lambda|.$$

Le résultat (5.5) est alors une conséquence de l'inégalité de Cauchy-Schwarz.

(iii) En injectant  $\bar{x}^\epsilon = -A^{-1}B^T \bar{\lambda}^\epsilon$  issu de la première équation de (5.4) dans la seconde équation de (5.4), on obtient

$$(BA^{-1}B^T + \epsilon I)\bar{\lambda}^\epsilon = -\epsilon \lambda,$$

puis en multipliant par  $\bar{\lambda}^{\epsilon T}$  à gauche, et

$$\beta \|\bar{\lambda}^\epsilon\|_2^2 \leq \bar{\lambda}^{\epsilon T} BA^{-1}B^T \bar{\lambda}^\epsilon \leq \bar{\lambda}^{\epsilon T} BA^{-1}B^T \bar{\lambda}^\epsilon + \epsilon \|\bar{\lambda}^\epsilon\|_2^2 = -\epsilon \bar{\lambda}^{\epsilon T} \lambda.$$

en utilisant à nouveau l'inégalité de Cauchy-Schwarz, on a  $\bar{\lambda}^{\epsilon T} \lambda \leq \|\bar{\lambda}^\epsilon\|_2 \|\lambda\|_2$ , puis  $\beta \|\bar{\lambda}^\epsilon\|_2 \leq \epsilon \|\lambda\|_2$ , ce qui est bien (5.6). En utilisant (5.6) dans (5.4), on obtient

$$\alpha \|\bar{x}^\epsilon\|_2^2 \leq \alpha \|\bar{x}^\epsilon\|_2^2 + \epsilon \|\bar{\lambda}^\epsilon\|_2^2 \leq \epsilon \|\lambda\|_2 \|\bar{\lambda}^\epsilon\|_2 \leq \frac{\epsilon^2}{\beta} \|\lambda\|_2^2,$$

ce qui est bien (5.8).

(iv) Le résultat est obtenu par passage à la limite dans (5.6) et (5.8).

□

**Proposition III.0.6**

Cas où  $B$  est de rang strictement inférieur à  $\min\{m, n\}$ . On suppose que  $(x, \lambda)$  et  $(x^\epsilon, \lambda^\epsilon)$  sont solutions respectives de (5.1) et (5.2). On s'intéresse à la limite de  $(x^\epsilon, \lambda^\epsilon)$  pour  $\epsilon \rightarrow 0$ .

- (i) Vérifiez que la solution du système (5.2) existe et est unique.
- (ii) Appelant  $\alpha > 0$  la plus petite valeur propre de  $A$ , montrez que  $\alpha \|x^\epsilon\|_2^2 + \epsilon \|\bar{\lambda}^\epsilon\|_2^2 \leq \frac{1}{2}\epsilon(\|\bar{\lambda}^\epsilon\|_2^2 + \|\lambda\|_2^2)$ .
- (iii) En conclure que  $\sqrt{2\alpha} \|x^\epsilon\|_2 \leq \sqrt{\epsilon} \|\lambda\|_2$  et donc que  $\lim_{\epsilon \rightarrow 0} x^\epsilon = x$ . L'erreur est donc en  $O(\sqrt{\epsilon})$ , et la convergence de  $\lambda^\epsilon$  n'est pas acquise.

Démonstration :

- (i) La démonstration est la même que pour l'exercice III.0.3.
- (ii) En reprenant (5.5) (obtenu sans supposer  $B$  de rang maximum), on obtient

$$\alpha \|x^\epsilon\|_2^2 + \epsilon \|\bar{\lambda}^\epsilon\|_2^2 \leq \epsilon \|\lambda\|_2 \|\bar{\lambda}^\epsilon\|_2 \leq \frac{1}{2}\epsilon(\|\bar{\lambda}^\epsilon\|_2^2 + \|\lambda\|_2^2), \quad (5.8)$$

ce qui implique  $\alpha \|x^\epsilon\|_2^2 + \frac{\epsilon}{2} \|\bar{\lambda}^\epsilon\|_2^2 \leq \frac{1}{2}\epsilon \|\lambda\|_2^2$ , d'où l'on tire  $\sqrt{2\alpha} \|x^\epsilon\|_2 \leq \sqrt{\epsilon} \|\lambda\|_2$ .

□

# Bibliographie

- [1] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. Number 01 in MPS-SIAM Series on Optimization. Siam, 2000.
- [2] Jean-Baptiste Hiriart-Urruty. *Analyse convexe et optimisation*. Presses Universitaires de France, 2000. ISBN :.
- [3] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of Convex Analysis*. Presses Universitaires de France, 2001. ISBN : 978-3-540-42205-1.
- [4] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer, second edition, 2006.
- [5] R. Tyrrell Rockafellar. *Convec Analysis*. Princeton University Press, 1996.

# Index

équation d'Euler, 39  
équations normales, 40

cône tangent, 59  
condition nécessaire du deuxième ordre, 40  
condition suffisante du deuxième ordre, 41

direction tangente, 59

formule de Taylor-Young, 26

hypothèse de qualification des contraintes, 59

inéquation d'Euler, 39

KKT (conditions de —), 61

Lagrangien, 61  
lagrangien, 57  
lemme de Farkas et Minkowski, 62

multiplicateur de Karush-Kuhn-Tucker, 61  
multiplicateur de Lagrange, 61

point critique, 39

théorème de Karush, Kuhn et Tucker, 61  
théorème des fonctions implicites, 26

valeur critique, 39