Chapitre 6

Introduction à l'optimisation différentiable avec contraintes non linéaires

Sujets du chapitre

- Analyse des conditions d'optimalité (équations ou inéquations satisfaites pour toute solution du problème d'optimisation).
- Hypothèses sur les contraintes.
- Présentation des algorithmes de pénalité et barrière.

Introduction

Les problèmes d'optimisation différentiable avec contraintes d'égalité également différentiables constituent un sujet classique, qui a été bien étudié par Lagrange, entre autres. Le traitement de contraintes d'inégalités constitue un sujet plus moderne. Dans cette introduction, nous énonçons les conditions qui seront démontrées dans les chapitres subséquents.

Comme précédemment, nous utiliserons le concept de DED pour identifier des équations à satisfaire (conditions d'optimalité) pour les solutions locales du problème

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ &\text{sujet à } g(x) = 0, \end{aligned} \tag{6.1}$$

ou du problème

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ &\text{sujet à } g(x) \leqslant 0, \end{aligned} \tag{6.2}$$

où $g : \mathbb{R}^n \to \mathbb{R}^m$. g est un vecteur *colonne* de fonctions, et $m < n$ pour le problème 6.1. Les fonctions f g sont différentiables, et en général, supposées de classe $\mathcal{C}^3(\mathbb{R}^n)$.

6.1 Conditions d'optimalité

Avant d'introduire les conditions d'optimalité, nous devons définir quelques concepts.

Tout d'abord, l'ensemble des points x qui satisfont aux contraintes est nommé *ensemble réalisable* du problème. Nous devrons aussi définir le concept de *direction réalisable.*

6.1.1 Directions réalisables

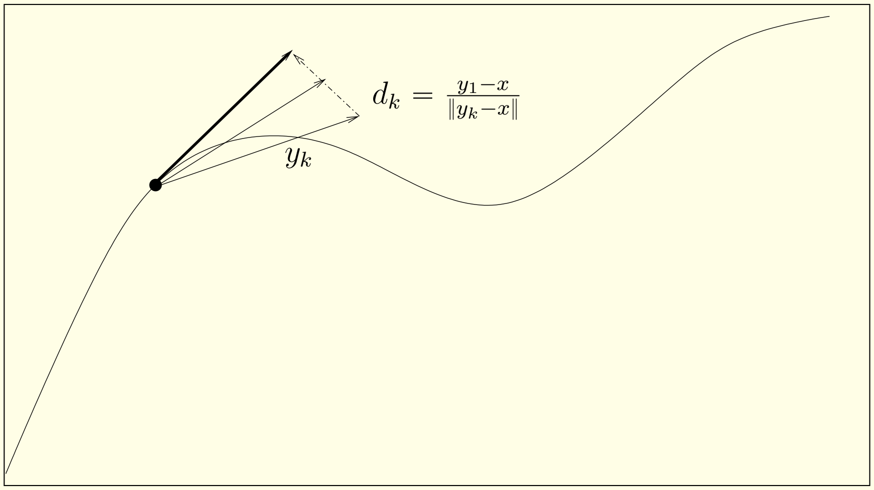
Égalités

Définition 6.1.1 *Soit x tel que $g(x) = 0$. Considérons une suite y_k, avec $g(y_k) = 0$ telle que $y_k \to x$. Considérons maintenant les vecteurs $d_k = \frac{y_k - x}{\|y_k - x\|}$. Tout point d'accumulation de la suite d_k est nommé direction réalisable en x pour les contraintes g.*

L'ensemble de ces directions, en fait, constitue le *plan tangent* aux contraintes en x.

Soit une suite $d_k = \frac{y_k - x}{\|y_k - x\|}$ convergeant vers la direction réalisable $\bar{d}$ en un point x. Alors, on a que pour chacune des composantes g_i de l'application g, le développement de Taylor limité nous donne

$$g_i(y_k) = g_i(x) + \varphi_k \nabla g_i(x) d_k + \mathcal{O}(\varphi_k^2),$$



Direction réalisable obtenue en passant à la limite de directions normalisées menant à des points voisins y_k ; ce sont des directions tangentes aux contraintes, ici une ligne courbe en deux dimensions

FIGURE 6.1 – Direction réalisable

où $\varphi_k = \|y_k - x\|$. Puisque $g_i(y_k) = g_i(x) = 0$, il reste

$$\nabla g_i(x) d_k \sim \mathcal{O}(\varphi_k),$$

et en passant à la limite,

$$\nabla g_i(x)\bar{d} = 0,$$

et ce pour toutes les composantes de l'application g, ce qui s'écrit

$$\nabla g(x)\bar{d} = 0,$$

où ∇g représente le Jacobien de l'application g, c'est-à-dire une matrice dont les lignes sont $\nabla g_i(x)$. Donc, toute direction réalisable d en x satisfait à la relation $\nabla g(x)d = 0$.

Pour aller plus loin, il faut maintenant introduire une hypothèse sur les contraintes g, à savoir une hypothèse de *qualification des contraintes* g. Cette hypothèse stipule que x^* est un point régulier du système $g(x) = 0$, c'est-à-dire, la matrice $\nabla g(x^*)$ est une matrice de plein rang-lignes, $m \leqslant n$. Cette hypothèse permet de montrer que toute direction satisfaisant $\nabla g(x)d = 0$ est une direction réalisable, complétant ainsi la caractérisation des directions réalisables. Remarquons que cette qualification dépends des fonctions g et non pas de l'ensemble des points réalisables, qui pourrait être exprimé à l'aide d'autres contraintes.

Définition 6.1.2 *La* qualification des contraintes *LICQ stipule que la matrice* $\nabla g(x^*)$ *est de plein rang ligne. Les gradients des contraintes sont donc linéairement indépendants et on dit que le point* x^* *est un point régulier des contraintes*

La courbe réalisable $x(t)$ suit la direction d, mais s'éloigne de la droite de direction d en demeurant sur la variété $g(x) = 0$. On représente cet éloignement par un multiple $u(t)$ du vecteur $\nabla g(x)$

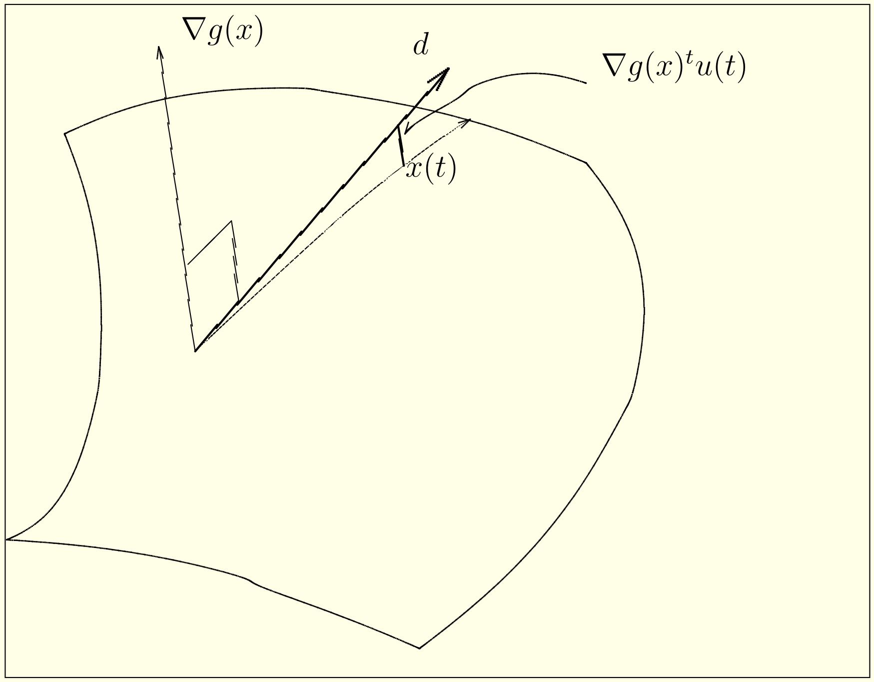


FIGURE 6.2 – Courbe réalisable

Par exemple, les contraintes $g_1(x, y) = x^2 - y$ et $g_2(x, y) = x^2 + y$ définissent un seul point réalisable, soit l'origine. Pourtant, un mauvais hasard fait que tout l'axe des x est tangent à chacune des deux contraintes alors que le plan tangent tel que défini plus haut se réduit à l'origine, un sous-espace de dimension nulle. L'origine n'est pas un point régulier de ces contraintes. Cependant, les contraintes $\bar{g}(x, y) = x + y$ et $\bar{g}_2(x, y) = x - y$ définissent le même point réalisable, qui est régulier pour les contraintes, cette fois-ci.

Théorème 6.1.1 *Soit x^* un point régulier des contraintes $g(x^*) = 0$. Alors, d est une direction réalisable en x^* si et seulement si $\nabla g(x^*)d = 0$.*

Preuve Cette preuve constitue un exercice d'application du théorème des fonctions implicites. Le même schéma de preuve sera utilisé pour analyser l'algorithme des pénalités présenté plus loin.

Il est clair que si d est une direction réalisable, alors $\nabla g(x^*)d = 0$. En fait, x^* n'a même pas besoin d'être régulier pour cette moitié du théorème.

En revanche, supposons que x^* est un point régulier qui satisfait $\nabla g(x^*)d = 0$. Il faut montrer que, quel que soit d dans le plan tangent aux contraintes en x^*, il existe une courbe réalisable $x(t)$ telle que $x(0) = x^*$ et que, $x'(0) = d$.

Pour construire une telle courbe, considérons les équations (voir la figure 6.2)

$$g(x^* + td + \nabla g(x^*)^t u(t)) = 0,$$

où u est inconnue, et t est considéré fixé. Ceci constitue un système régulier d'équations paramétrisé par t, avec $u(0) = 0$. On peut exprimer la matrice Jacobienne de ce système en u comme $\nabla g(x^*)\nabla g(x^*)^t$, qui est inversible puisque $\nabla g(x^*)$ est de plein rang, x^* étant un point régulier des contraintes. Les hypothèses du théorème des fonctions implicites sont donc satisfaites, et ce système possède une solution continûment différentiable $u(t)$; lorsque la fonction g est de classe $\mathcal{C}^p$, la courbe u l'est aussi. La courbe $x(t) = x^* + td + \nabla g(x^*)^t u(t)$ est par construction une courbe réalisable. Dérivons le système $g(x^* + td + \nabla g(x^*)^t u(t)) = 0$ par rapport à t, et évaluons le tout en $t = 0$:

$$0 = \frac{d}{dt} g(x(t))\Big|_{t=0} = \nabla g d + \nabla g(x^*)\nabla g(x^*)^t u'(0).$$

Nous avions choisi d pour que $\nabla g(x^*)d = 0$, et comme $\nabla g(x^*)\nabla g(x^*)^t$ est inversible, $u'(0) = 0$, d'où $x'(0) = d$, tel que voulu. □

On voit donc que les directions réalisables sont celles situées dans le noyau de la matrice $\nabla g(x^*)$. Considérons une base du noyau de $\nabla g(x^*)$ formée des vecteurs $b_1, b_2, \ldots, b_{m-n}$, et une matrice Z^* dont les colonnes sont les vecteurs b_i. Alors, toute direction réalisable d peut s'exprimer comme combinaison linéaire de la base du noyau de $\nabla g(x^*)$, $d = Z^* d_z$ où $d_z \in \mathbb{R}^{m-n}$ constitue l'expression de d dans la base du noyau.

Inégalités : cône tangent aux contraintes

Soit une suite $d_k = \frac{y_k - x}{\|y_k - x\|}$ convergeant vers la direction réalisable $\bar{d}$ en un point x. Alors, on a que pour chacune des composantes g_i de l'application g, le développement de Taylor limité nous donne

$$0 \geqslant g_i(y_k) = g_i(x) + \varphi_k \nabla g_i(x) d_k + \mathcal{O}(\varphi_k),$$

où $\varphi_k = \|y_k - x\|$. Puisque $g_i(y_k) \leqslant 0$ et $g_i(x) \leqslant 0$ il est commode de distinguer deux cas :

1. $g_i(x) < 0$: dans ce cas, quelle que soit la suite y_k convergeant vers x, éventuellement on aura que $g_i(y_k) < 0$, et donc aucune analyse additionnelle n'est requise.
2. $g_i(x) = 0$: dans l'expression de $g_i(y_k)$, il ne reste que

$$0 \geqslant g_i(y_k) = \nabla g_i(x) d_k \sim \mathcal{O}(\varphi_k^2),$$

et en passant à la limite,

$$\nabla g_i(x)\bar{d} \leqslant 0.$$

Donc, pour toutes les composantes i de l'application g telles que $g_i(x) = 0$, ce qui s'écrit

$$\nabla g_{I*}(x)\bar{d} \leqslant 0,$$

où ∇g_{I*} représente le Jacobien de l'application g_{I*}, c'est-à-dire une matrice dont les lignes sont $\nabla g_i(x), i \in I^*$. Donc, toute direction réalisable d en x satisfait à la relation $\nabla g_{I*}(x)d \leqslant 0$.

Comme pour les égalités, introduisons une hypothèse de qualification des contraintes g. Cette hypothèse stipule que x^* est un point régulier du système $g_{I^*}(x) = 0$, c'est-à-dire, la matrice $\nabla g_{I^*}(x^*)$ est une matrice de plein rang-lignes, $|I^*| \leqslant n$. Cette hypothèse permet de montrer que toute direction satisfaisant $\nabla g_{I^*}(x)d \leqslant 0$ est une direction réalisable, complétant ainsi la caractérisation des directions réalisables. Remarquons que cette qualification dépends des fonctions g et non pas de l'ensemble des points réalisables, qui pourrait être exprimé à l'aide d'autres contraintes. Nous approfondirons l'étude des qualifications des contraintes dans un contexte plus général dans un prochain chapitre.

Définition 6.1.3 *La* qualification des contraintes *LICQ stipule que la matrice* $\nabla g_{I^*}(x^*)$ *est de plein rang ligne. Les gradients des contraintes actives sont donc linéairement indépendants*

Une qualification moins restrictive a été proposée par Mangasarian et Fromowitz.

Définition 6.1.4 *La* qualification des contraintes *MFCQ stipule qu'il existe un vecteur* v *tel que* $\nabla g_{I^*}(x^*)v < 0$

Théorème 6.1.2 *Soit* x^* *un point satisfaisant à LICQ ou encore MFCQ. Alors,* d *est une direction réalisable en* x^* *si et seulement si* $\nabla g_{I^*}(x^*)d \leqslant 0$.

***Exercice 6.1.1** *[Preuve]* Imitez la preuve du théorème 6.1.1 pour démontrer le théorème 6.1.2.

6.1.2 Conditions de premier ordre

Pour obtenir des conditions d'optimalité de premier ordre en un point x^*, nous devons nous assurer qu'aucune direction réalisable en x^* n'appartient au $DED(x^*)$. Le $DED(x^*)$ est bien défini comme $\{d : \nabla f(x^*)d < 0\}$, et donc l'obtention de conditions d'optimalité (du premier ordre) passe par une caractérisation des directions réalisables.

Égalités

En utilisant la caractérisation des directions réalisables, on peut rephraser la condition d'optimalité comme suit : pour toute direction dans le noyau de la matrice $\nabla g(x^*)$, $\nabla f(x^*)d = 0$. Exprimons $d = Z^* d_z$ pour écrire $\nabla f(x^*)(Z^* d_z)$, quel que soit $d_z \in \mathbb{R}^{m-n}$. Ceci entraîne que $\nabla f(x^*)Z^* = 0$; cette dernière quantité se nomme *gradient réduit* de f.

Une autre manière de voir ces conditions consiste à exploiter le fait que $\nabla f(x^*)$ est un vecteur orthogonal au noyau de la matrice $\nabla g(x^*)$. Un exercice simple d'algèbre linéaire permet de conclure que ceci est équivalent à l'existence d'un vecteur $\lambda^* \in \mathbb{R}^m$ satisfaisant à l'égalité

$$\nabla f(x^*) = \lambda^* \nabla g(x^*).$$

Bon, tout ceci nous donne que $d \in DED(x^*)$ est caractérisé par $\nabla f(x^*)d < 0$ et d est une direction réalisable est caractérisé par $\nabla g(x^*)d = 0$. En observant que si d est réalisable, $-d$ l'est aussi, on retrouve que d ou $-d$ n'appartiennent pas à $DED(x^*)$ seulement si $\nabla f(x^*)d = 0$, ce qui permet d'écrire la condition d'optimalité suivante : si x^* est un minimum local de f dans l'ensemble défini par $g(x) = 0$, alors

$$\nabla g(x^*)d = 0 \Longleftrightarrow \nabla f(x^*)d = 0.$$

Résumons la précédente discussion dans le

Théorème 6.1.3 *Soit x^* un minimum local de la fonction f sous les contraintes $g(x) = 0$ tel que $\nabla g(x^*)$ est une matrice de plein rang-ligne ; alors, il existe un vecteur $\lambda^* \in \mathbb{R}^m$ tel que $\nabla f(x^*) = \lambda^* \nabla g(x^*)$; ceci est équivalent à la condition $\nabla f(x^*)Z^* = 0$, où les colonnes de Z^* forment une base du noyau de la matrice $\nabla g(x^*)$.*

Exercice 6.1.2 *[Optimalité]* Considérez le problème

$$\begin{array}{rc} \min & x_2 + x_3 \\ \text{s.à} & x_1 + x_2 + x_3 = 1 \\ & x_1^2 + x_2^2 + x_3^2 = 1 \end{array}$$

a) Par la symétrie de x_2 et x_3, reformulez le problème avec seulement deux variables, soit x_1 et $\hat{x}_2 = x_2 = x_3$. Obtenez deux points stationnaires pour ce problème en deux variables x_1 et $\hat{x}_2$. Illustrez graphiquement votre solution.

b) Vérifiez que $x' = (1, 0, 0)^t$ et $x'' = (-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})^t$ satisfont aux conditions nécessaires de Lagrange pour le problème original.

Exercice 6.1.3 *[Rectangle—bis]* Trouvez le rectangle de périmètre maximal pour une superficie donnée ($hl = 1$). Évidemment, un rectangle infiniment long et mince suggère, avec raison, que la solution n'est pas bornée. Supposons donc que la hauteur h et la longueur l sont bornées : $0.5 \leqslant h, l \leqslant 4$. Vérifiez que $l = 1$ et $h = 1$ est un point stationnaire, mais pas un maximum local ; quelle est la nature de ce point stationnaire ? Trouvez les maxima locaux, et le maximum global de ce problème.

Exercice 6.1.4 *[Projection]* Suite de l'exercice 5.3.6 page 229 Considérez le programme :

$$\begin{array}{rl} \min & gx \\ \text{s.à} & Ax = 0 \\ & \|x\|^2 = 1 \end{array}$$

Vérifiez que sa solution est un vecteur de même direction que celle trouvée à l'exercice 4.3.4.

Inégalités

On voit donc que les directions réalisables coïncident avec les directions réalisables des contraintes linéaires $0 \geqslant L(y) = g(x) + \nabla g_{I*}(x)y$. Par conséquent, nous pouvons appliquer les développements du chapitre 5 pour obtenir des conditions d'optimalité.

Théorème 6.1.4 *Soit x^* un minimum local de la fonction f sous les contraintes $g(x) \leqslant 0$ tel que LICQ ou MFCQ est satisfaite ; alors, il existe un vecteur $0 \leqslant \lambda^* \in \mathbb{R}^m$ tel que $\nabla f(x^*) + \lambda^* \nabla g(x^*) = 0$ et $\lambda_i^* g_i(x^*) = 0$; ceci est équivalent à la condition $\nabla f(x^*) Z^* = 0$, où les colonnes de Z^* forment une base du noyau de la matrice $\nabla g_{I*}(x^*)$.*

Exercice 6.1.5 *[Points stationnaires]* Considérez le problème

$$\begin{array}{rc} \min & f(x) = \frac{1}{2}\alpha(x_1 - 1)^2 - x_1 - x_2 \\ \text{sujet à} & x_1 \geqslant x_2^2 \\ & x_1^2 + x_2^2 \leqslant 1 \end{array}$$

a) Trouvez tous les points qui rendent les 2 contraintes actives ; un de ces points (à 3 décimales de précision) est $\bar{x} = (0.618, 0.786)^t$.
b) Avec 3 décimales de précision, trouvez les valeurs du paramètre α (pas forcément positif) pour lesquelles $\bar{x}$ est un point stationnaire.

6.1.3 Conditions de second ordre

Égalités

Nous venons de voir que les conditions de premier ordre expriment le fait que les directions réalisables, donc tangentes à la variété des points satisfaisant aux contraintes $g(x) = 0$, ne sont pas des directions de descente stricte.

Cette propriété n'est pas assez forte pour discriminer des minima locaux, points de selles ou maxima. Revenons à l'exemple de la fonction $f(x, y) = (x^2 - y)(x^2 - 2y)$. Nous avons déjà vérifié que l'origine est un minimum local de cette fonction restreint à toute droite passant par l'origine. En d'autres termes,

$$0 = \min_{ax+by=0} (x^2 - y)(x^2 - 2y),$$

et ce, quelles que soient les valeurs de a et b. Cependant,

$$0 = \max_{x^2-1.5y=0} (x^2 - y)(x^2 - 2y).$$

Pourtant, dans les deux cas, pour le cas particulier où $a = 0$ et $b = -1.5$, le gradient de l'unique contrainte est $\nabla g(x^*, y^*) = \begin{pmatrix} 0 & -1.5 \end{pmatrix}$, alors que le gradient de l'objectif est $\nabla f(x^*, y^*) = 0 = 0 \nabla g(x^*, y^*)$. Pour le problème sans contrainte, l'origine est un point de selle alors que pour le problème contraint, dépendamment de la contrainte, l'origine est soit un minimum local ou encore un maximum local. Cet exemple illustre de façon convainquante qu'il faut tenir compte de la courbure des contraintes (dérivées d'ordre supérieur) pour caractériser les minima et les maxima locaux contraints.

Il est remarquable que les pondérations relatives des dérivées secondes des contraintes et de la fonction objectif puissent toutes être déduites par l'analyse du *Lagrangien* associé au problème de minimisation avec contraintes. Rappelons que celui-ci est défini comme

$$L(x, \lambda) \stackrel{\text{def}}{=} f(x) + \lambda g(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x).$$

Les conditions d'optimalité du premier ordre que nous avons construite plus haut s'expriment de manière concise à l'aide de la fonction L :

$$0 = \nabla L(x^*, \lambda^*) = \begin{cases} \nabla_x L(x^*, \lambda^*) \\ \nabla_\lambda L(x^*, \lambda^*) \end{cases} .$$

En effet, $\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \lambda^* \nabla g(x^*)$ qui s'annule correspond à la condition de premier ordre alors que $\nabla_\lambda L(x^*, \lambda^*) = g(x^*)$ qui s'annule correspond simplement à la satisfaction des contraintes. Cette formulation est à la source du nom donné aux variables λ, *multiplicateurs de Lagrange* pour le problème (6.1).

La condition de second ordre s'exprime simplement comme suit :

$$\nabla^2_{xx} L(x^*, \lambda^*) = \nabla^2 f(x^*) + \lambda^* \nabla^2 g(x^*)$$

est une matrice (semi-)définie positive sur l'espace tangent $\{d : \nabla g(x^*)d = 0\}$, ce qui s'exprime à l'aide de la matrice Z^* introduite plus haut :

$$Z^{*t} \left(\nabla^2_{xx} L(x^*, \lambda^*)\right) Z^* = Z^{*t} \left(\nabla^2 f(x^*) + \lambda^* \nabla^2 g(x^*)\right) Z^*$$

est une matrice (semi-)définie positive. Cette dernière matrice porte le nom de *hessien réduit* du Lagrangien.

Pour alléger la notation, nous utiliserons fréquemment la formulation suivante dans l'avenir (tout comme ci-haut !) :

$$\lambda \nabla^2 g(x) \stackrel{\text{def}}{=} \left(\sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x) \right).$$

Grâce à cette formulation, nous pouvons énoncer les conditions d'optimalité.

Théorème 6.1.5 *1.* [Nécessaire] *Si x^*, point régulier des contraintes $g(x) = 0$, est un minimum local du programme (6.1), alors il existe un vecteur $\lambda^* \in \mathbb{R}^m$ tel que $\nabla L(x^*, \lambda^*) = 0$, et de plus $d^t \nabla^2_{xx} L(x^*, \lambda^*) d \geqslant 0$ pour tout d tel que $\nabla g(x^*)d = 0$.*

2. [Suffisante] *S'il existe un vecteur $\lambda^* \in \mathbb{R}^m$ tel que $\nabla L(x^*, \lambda^*) = 0$, et de plus $d^t \nabla^2_{xx} L(x^*, \lambda^*) d > 0$ pour tout d tel que $\nabla g(x^*)d = 0$, alors x^* est un minimum local pour le programme (6.1).*

Preuve

- Nécessité.
  Considérons une courbe différentiable $x(t)$ de classe $\mathcal{C}^2$ telle que $g(x(t)) = 0$ près de $x(0) = x^*$; pour construire une telle courbe, on procède comme au théorème 6.1.1. Alors, si x^* est un minimum local pour le programme (6.1), on a nécessairement $f(x^*) = f(x(0)) = \min_t f(x(t))$, et donc

$$\frac{d^2}{dt^2} f(x(t))\Big|_{t=0} = x'(0)^t \nabla^2 f(x^*) x'(0) + \nabla f(x^*) x''(0) \geqslant 0.$$

  En dérivant deux fois par rapport à t la relation $\lambda g(x(t))$ on obtient :

$$x'(0)^t \left(\sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x^*) \right) x'(0) + \lambda \nabla g(x^*) x''(0) = 0.$$

  La conclusion suit en additionnant les deux dernières relations, et en rappelant que $x'(0)$ est arbitraire dans le plan tangent $\nabla g(x^*) x'(0) = 0$.

- Suffisance.
  Reprenons les directions $d_k = y_k - x^*$ utilisées plus haut et procédons par contradiction en supposant que x^* n'est pas un minimum local de (6.1) bien qu'il satisfasse aux conditions suffisantes d'optimalité.
  Par la caractérisation de directions réalisables, nous savons que $\nabla g(x^*)\bar{d} = 0$. Utilisons le théorème de Taylor pour écrire

$$0 = g_i(y_k) = g_i(x^*) + \varphi_k \nabla g_i(x^*) d_k + \frac{\varphi_k^2}{2} d_k^t \nabla^2 g_i(\eta_i) d_k,$$

  et, si y_k est choisi pour que $f(y_k) < f(x^*)$,

$$0 \geqslant f(y_k) - f(x^*) = \varphi_k \nabla f(x^*) d_k + \frac{\varphi_k^2}{2} d_k^t \nabla^2 f(\eta_0) d_k,$$

  où les points η_i sont des points intermédiaires entre y_k et x^*. Maintenant, en multipliant l'avant-dernière relation par λ_i^* et en les ajoutant à la dernière relation, on obtient

$$0 \geqslant \frac{\varphi_k^2}{2} d_k^t \left(\nabla^2 f(\eta_0) + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(\eta_i) \right) d_k,$$

  ce qui fournit la contradiction désirée lorsque $k \to \infty$.

□

Remarquons que la qualification des contraintes, l'hypothèse que x^* est un point régulier des contraintes, n'est requise que pour établir les conditions nécessaires. Un exemple simple violant la qualification des contraintes est $g_1(x) = x_1^2 + (x_2 - 1)^2 = 1$ et $g_2(x) = x_1^2 + (x_2 + 1)^2 = 1$. L'origine en est le seul point réalisable, et donc solution optimale quelle que soit la fonction objectif. Les gradients des deux contraintes sont $\nabla g_1^* = (0, -2)$ et $\nabla g_2^* = (0, 2)$. Les multiplicateurs de Lagrange existe donc seulement pour des fonctions objectif dont le gradient s'aligne sur l'axe x_2.

Exercice 6.1.6 *[Optimalité]* —suite de l'exercice 6.1.2, page 281 Déterminez la nature de ces quatre points stationnaires (les deux en a), et les deux en b)).

Exercice 6.1.7 *[Conditions d'optimalité]* Considérez une hyperbole de $\mathbb{R}^2$, $y = 1/x$. Soit un point $P = \binom{P_x}{P_y}$, on veut trouver le point de l'hyperbole le plus proche de P. On ramène l'étude au programme suivant dans $\mathbb{R}^2$:

$$(A) \begin{cases} \min d(x,y) = & \frac{1}{2}((x - P_x)^2 + (y - P_y)^2) \\ \text{sujet à} & xy = 1. \end{cases}$$

a) Écrivez les conditions d'optimalité (d'ordre un et deux) de ce problème.

b) Si le point P est sur la droite $y = x$, on devine qu'il y a 2 candidats solution naturels : $\binom{1}{1}$ et $\binom{-1}{-1}$. En vous limitant aux $x, y > 0$, vérifiez cette intuition en utilisant les conditions d'optimalité ; vérifiez d'abord que $x = y = 1$ est un point stationnaire pour le programme (A). Vérifiez ensuite que l'intuition n'est pas vraie si P est assez loin de l'origine ; quantifiez ce "loin" de l'origine en utilisant les conditions d'optimalité d'ordre 2.

Inégalités

Nous venons de voir que les conditions de premier ordre expriment le fait que les directions réalisables, donc appartenant au cône tangent aux contraintes $g(x) \leqslant 0$, ne sont pas des directions de descente stricte. Nommons $I^* = \{i : g_i(x^*) = 0\}$, $\hat{I} = \{i \in I^* : \lambda_i^* > 0\}$ et aussi $I_0^* = I^* \backslash \hat{I}$.

Théorème 6.1.6 Conditions nécessaires *Si x^*, un point réalisable satisfaisant à LICQ, est un minimum local du programme (6.2), alors il existe un vecteur $0 \leqslant \lambda^* \in \mathbb{R}^m$ tel que $\nabla_x L(x^*, \lambda^*) = 0$, $\nabla_\lambda L(x^*, \lambda^*)^t \lambda^* = 0$ et de plus $d^t \nabla_{xx}^2 L(x^*, \lambda^*) d \geqslant 0$ pour tout d tel que $\nabla g_{I^*}(x^*) d = 0$.*

Théorème 6.1.7 Conditions suffisantes *Si x^* est un point réalisable pour lequel il existe un vecteur $0 \leqslant \lambda^* \in \mathbb{R}^m$ tel que $\nabla_x L(x^*, \lambda^*) = 0$, $\lambda^* g(x^*) = 0$, et de plus $d^t \nabla_{xx}^2 L(x^*, \lambda^*) d > 0$ pour tout $0 \neq d \in \ker \nabla g_{\hat{I}}$ tel que $\nabla g_{I_0^*}(x^*) d \leqslant 0$, alors x^* est un minimum local pour le programme (6.2).*

Remarquons que l'hypothèse de qualification des contraintes n'est requise que pour les conditions nécessaires, c'est-à-dire pour démontrer l'existence de multiplicateurs λ^*. Si LICQ est satisfaite, alors le vecteur de multiplicateurs λ^* est uniquement défini. La condition nécessaire de second ordre est très délicate à obtenir, nécessitant d'autres qualifications lorsqu'il existe plusieurs multiplicateurs λ satisfaisant à la condition de premier ordre.

Exercice 6.1.8 *[Conditions d'optimalité]* suite de l'exercice 5.4.8, page 245) Soient les trois fonctions suivantes :

- $f_1(x) = \frac{1}{2}((x_1 - 3)^2 + (x_2 + 1)^2)$;
- $f_2(x) = \frac{1}{2}((x_1 + 1)^2 + (x_2 - 1)^2)$;
- $f_3(x) = \frac{1}{2}((x_1 + 1)^2 + x_2^2)$.

Considérez les (trois) contraintes

$$\begin{aligned} x_1 &\geqslant 0, \\ x_2 &\geqslant 0, \\ x_1^2 + x_2^2 &\leqslant 1 \end{aligned}$$

définissant l'ensemble réalisable E, et le point candidat $x = (1, 0)^t \in E$.

a) Pour chacune des fonctions, vérifiez que le point candidat est un point stationnaire, et donnez les valeurs des multiplicateurs de Lagrange λ_1, λ_2 et λ_3.

b) Pour chacune des fonctions, déterminez si le point stationnaire est un minimum local, maximum local ou ni l'un ni l'autre.

Exercice 6.1.9 *[Optimalité]* Considérez le problème suivant ; ATTENTION, c'est un max !

$$\begin{array}{rl} \max & x_1^2 + x_2^2 \\ \text{sujet à} & (x_2 - b)^2 \geqslant 4ax_1 \quad (0 < b < 4a) \\ & x_2 \leqslant b \\ & x \geqslant 0 \end{array}$$

a) obtenez 2 maxima locaux ; lequel est global ? Une illustration graphique aide considérablement pour cet exercice.

b) Est-ce que les conditions KKT sont satisfaites en ces points ?

c) Est-ce que les qualifications des contraintes sont satisfaites en ces points ?

d) Est-ce que les conditions suffisantes d'ordre 2 sont satisfaites en ces points ?

Exercice 6.1.10 *[MinMax]* Considérez le problème de *minmax* suivant :

$$\min_{x \in \mathbb{R}^n} \max_{i \in I} f_i(x),$$

qui peut être reformulé comme :

$$\begin{array}{ll} \min\limits_{(x,u)\in\mathbb{R}^n\times\mathbb{R}} & u \\ \text{sujet à} & f_i(x) \leqslant u. \end{array}$$

a) Donnez des conditions d'optimalité (nécessaires et suffisantes) pour ce problème ; discutez également de l'utilisation des *qualifications des contraintes* dans les conditions nécessaires d'optimalité. En fait, montrez que la condition de Mangasarian-Fromovitz est toujours satisfaite.

b) Considérez les fonctions $f_1(x) = \frac{1}{2}\|x + \binom{1}{1}\|^2$ et $f_2(x) = \frac{1}{2}\|x - \binom{1}{1}\|^2$. Illustrez graphiquement les lignes de niveau constant de $\max\{f_1(x), f_2(x)\}$, obtenez la solution et vérifiez les conditions d'optimalité et de qualifications de contraintes de a).

6.2 Remarques générales sur les algorithmes

Le cas de contraintes non-linéaires ne permet pas de caractériser algébriquement les directions réalisables, comme c'était le cas pour les contraintes linéaires. La construction d'algorithmes passe donc par des perturbations ou approximations des conditions d'optimalité.

En perturbant de diverses manières les conditions d'optimalité, nous obtenons des algorithmes de pénalité, barrière ou de Lagrangien augmenté.

Une autre approche consiste à linéariser les conditions d'optimalité, à la maniere de la méthode de Newton. On se retrouve avec des conditions linéarisées qui constituent en fait les conditions d'optimalité d'un programme quadratique. Les algorithmes de programmation quadratique successive constituent une généralisation de la méthode de Newton aux problèmes contraints.

Une famille moderne est nommée "méthode primale-duale de points intérieurs" et est reliée aux méthodes de barriere. On peut aussi l'interpréter comme une extension de la méthode de Newton aux contraintes d'inégalité.

Dans ce royaume de problèmes sous contraintes non-linéaires, il n'existe actuellement aucun algorithme définitivement supérieur. Concentons-nous donc d'exposer au mieux quelques approches existantes.

6.3 Perturbations des conditions d'optimalité—pénalités et barrières

Contentons-nous pour l'instant de produire un algorithme qui calcule des points stationnaires pour le programme (6.1), c'est-à-dire des points qui satisfont aux conditions d'optimalité de premier ordre : il existe un vecteur $\lambda^* \in \mathbb{R}^m$ tel que

$$\begin{aligned} \nabla f(x^*) + \lambda^* \nabla g(x^*) &= 0 \\ g(x^*) &= 0. \end{aligned} \tag{6.3}$$

6.3.1 Solution des conditions de premier ordre

Une approche directe consiste bien sûr à tenter de résoudre le système d'équations (6.3). Par exemple, on pourrait utiliser la méthode de Newton, et espérer obtenir des résultats de convergence *locale* quadratique. Cependant, comme nous l'avons déjà discuté, le voisinage dans lequel la méthode de Newton converge est souvent assez petit. Cette constatation a motivé le développement d'une famille d'algorithmes de descente et les modifications de l'algorithme de Newton. Pour l'optimisation sans contrainte, la modification s'appuie sur le fait que la fonction objectif décroît d'une itération à l'autre (algorithme de descente). Une telle fonction, dans un contexte plus général, est nommée *fonction de mérite*.

Bien qu'il soit possible de construire des fonctions de mérite pour le problème (6.3), souvent ces fonctions de mérite ne sont pas différentiables, et nous en reportons l'étude à plus tard, lorsque nous aborderons l'optimisation de fonctions non-différentiables.

6.3.2 Égalités—pénalités

Dans l'équation (6.3), remplaçons le 0 de la seconde équation par l'approximation $\rho\lambda$, où ρ est un paramètre que nous laisserons tendre vers zéro. Le but de cette manœuvre est de calculer des solutions $x(\rho)$ dépendant du paramètre ρ en espérant pouvoir montrer que $\lim_{\rho \searrow 0} x(\rho) = x^*$. Cette manipulation permet d'écrire

$$\begin{aligned} \nabla f(x) + \lambda \nabla g(x) &= 0 \\ g(x) &= \rho \lambda^t, \end{aligned}$$

d'où l'on tire que $\lambda = \frac{g(x)^t}{\rho}$, ce qui permet de réduire ces deux équations à une seule :

$$P(x, \rho) \stackrel{\text{def}}{=} \nabla f(x) + \frac{g(x)^t}{\rho} \nabla g(x) = 0,$$

qui sont précisément les conditions d'optimalité du problème

$$\min_{x \in \mathbb{R}^n} p(x, \rho) \stackrel{\text{def}}{=} f(x) + \frac{\|g(x)\|^2}{2\rho}.$$

Ce dernier problème est nommé *pénalisation quadratique* du programme (6.1). Comme nous le verrons plus tard, d'autres perturbations des équations (6.3) conduisent à d'autres pénalisations. Dans ce qui suit, nous allons étudier le comportement de la famille de solutions (approximatives) $x(\rho)$ des problèmes pénalisés lorsque ρ s'approche de zéro, et montrer que sous des hypothèses assez faibles, les points d'accumulation de cette famille sont des points stationnaires pour le programme (6.1). Complétons la notation telle que $\nabla_x p = P$ en définissant $\nabla^2_{xx} p = \nabla_x P = H$:

$$H(x,\rho) \stackrel{\text{def}}{=} \nabla^2 f(x) + \left(\sum_{i=1}^{m} \frac{g_i(x)}{\rho} \nabla^2 g_i(x)\right) + \frac{\nabla g(x)^t \nabla g(x)}{\rho}.$$

Les *algorithmes de pénalité* consistent à définir une suite $\{\rho_k\}$ telle que $\rho_k \searrow 0$ et à calculer une suite $x(\rho_k)$ de minima approximatifs de $p(x,\rho_k)$. L'analyse de ces algorithmes consiste à montrer que $\lim_{k_i\to\infty} x(\rho_{k_i}) = x^*$, un point stationnaire de (6.1), à qualifier la vitesse de la convergence, et à quantifier la quantité de calculs requise pour identifier $x(\rho_{k+1})$ à partir de $x(\rho_k)$.

```
Pénalité
  { Donné : ρ0 > 0 et x0. }
  ρ ← ρ0
  xρ ← x0
  tantque (ρ > Precision )
    { Point de départ }
    x ← xρ
    { Corrections de Newton }
    tantque (‖P(x, ρ)‖ > tolerance )
      dN ← Direction_de_Newton(x, ρ)
      θ ← Pas_admissible(x, dN)
      x ← x + θdN
    xρ ← x
    ρ ← next(ρ)
```

Algorithme 6.1: Pénalité simple

L'approche par pénalisation est simple, on résout une suite de problèmes pénalisés selon un paramètre ρ_k sans contrainte jusqu'à ce que ρ_k soit assez petit pour que la solution du sous-problème sans contrainte soit acceptable comme solution approchée du problème original. Cette approche n'est pas nouvelle du tout, un logiciel nommé SUMT (*Sequential Unconstrained Minimization Technique*) a eu un succès important au moment de sa création vers 1967.

Bien que séduisante, cette approche nécessite une mise en œuvre soignée tel que nous le verrons dans la section dédiée aux détails d'implémentation.

Exercice 6.3.1 *[Pénalité $\|g\|^2$]* Considérez le problème

$$\begin{aligned} \min \quad & f(x) = -x_1 - x_2 + x_3 \\ \text{sujet à} \quad & x_3 = 0 \\ & x_1^2 + x_2^2 + x_3^2 = 1 \end{aligned}$$

Évidemment, x_3 est inutile dans la formulation, mais est inclus pour rendre les calculs moins triviaux. On utilise la pénalité extérieure $\|g(x)\|^2$ pour traiter chacune des égalités, et obtient les résultats suivants :

k	ρ_k	$x_1(\rho_k)$	$x_2(\rho_k)$	$x_3(\rho_k)$
1	1	0.834377	0.834377	-0.454846
2	0.1	0.728326	0.728326	-0.087925
3	0.01	0.709559	0.709559	-0.009861
4	0.001	0.707356	0.707356	-0.000999

a) En examinant la tendance des valeurs $x(\rho)$ lorsque ρ s'approche de 0, déduisez la solution optimale du problème.

b) $\lambda(\rho) = \frac{g(x)}{\rho}$; en examinant la tendance des valeurs de $\lambda(\rho)$, déduisez les multiplicateurs de Lagrange optimaux du problème.

6.3.3 Inégalités–pénalité : $g(x) \leqslant 0$

De la condition de complémentarité, nous savons que lorsque $g_i(x) < 0$, $\lambda_i = 0$. Perturbons donc l'inégalité $g(x) \leqslant 0$ par l'égalité équivalente $\max(g(x), 0) = 0$ et perturbons cette dernière en remplaçant le 0 de droite par $\rho\lambda$, où ρ est un paramètre que nous laisserons tendre vers zéro. Le but de cette manœuvre est de calculer des solutions $x(\rho)$ dépendant du paramètre ρ en espérant pouvoir montrer que $\lim_{\rho \searrow 0} x(\rho) = x^*$. Cet manipulation permet d'écrire, en notant $g^+(x) \stackrel{\text{def}}{=} \max(g(x), 0)$

$$\begin{aligned} \nabla f(x) + \lambda \nabla g(x) &= 0 \\ g^+(x) &= \rho\lambda, \end{aligned}$$

On observe la pénalisation quadratique pour des valeurs de $\rho = 1$, 0.1 et 0.01. La fonction est $f(x) = -x_1 - x_2$ et la contrainte est $1 - x_1^2 - x_2^2 = 0$. Le mauvais conditionnement s'observe par des lignes de niveau constant de plus en plus “plates”, et également de plus en plus rapprochées, ainsi que par l'échelle de l'axe $f(x)$ du graphique.
On peut également observer sur un même graphique l'effet de la pénalisation, et le mauvais conditionnement qui se traduit alors par une fonction beaucoup plus abrupte.

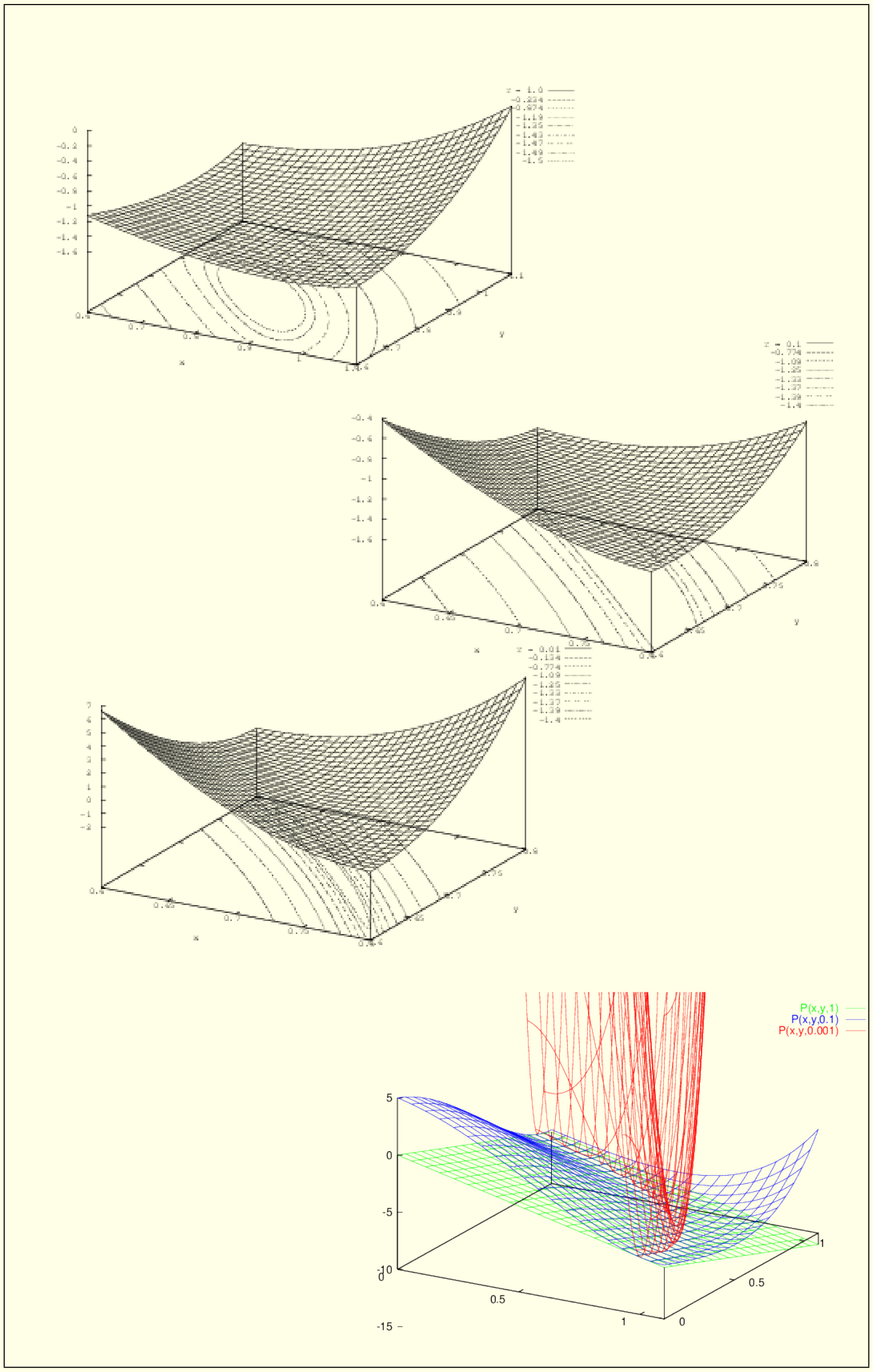


Figure 6.3 – Pénalisation

d'où l'on tire que $\lambda = \frac{g^+(x)}{\rho}$, ce qui permet de réduire ces deux équations à une seule :

$$P(x, \rho) \stackrel{\text{def}}{=} \nabla f(x) + \frac{g^+(x)}{\rho} \nabla g(x) = 0,$$

qui sont précisément les conditions d'optimalité du problème

$$\min_{x \in \mathbb{R}^n} p(x, \rho) \stackrel{\text{def}}{=} f(x) + \frac{\|g^+(x)\|^2}{2\rho}.$$

Ce dernier problème est nommé *pénalisation quadratique* du programme (6.2).

Un inconvénient majeur de ces fonctions de pénalité vient des termes g_i^+, qui ne sont pas différentiables aux points où la fonction $g_i = 0$. Puisque ces termes apparaîssent au carré, la fonction $p(x, \rho)$ est différentiable, mais pas de classe $\mathcal{C}^2$. Par ailleurs, au voisinage de minima satisfaisant aux conditions suffisantes d'optimalité et à la condition de stricte complémentarité, la convergence se fait de l'*extérieur* des contraintes actives, d'où le nom de *méthodes de pénalité extérieures* sous lequel ces algorithmes sont connus. En effet, si la complémentarité stricte est satisfaite au point limite, puisque $\frac{g^+(x)}{\rho}$ s'approche de λ^*, si $g_i(x)$ s'approche de 0, $\frac{g_i(x)}{\rho}$ s'approche d'une quantité positive, d'où l'interprétation que x s'approche de l'extérieur. Dans ce cas, la discontinuité des dérivées secondes n'a pas d'importance puisque les itérations demeurent dans un ensemble (l'extérieur des contraintes actives et l'intérieur des autres) où les dérivées secondes sont continues.

On observe la pénalisation quadratique pour des valeurs de $\rho = 1$, 0.1 et 0.01. La fonction est $f(x) = -x_1 - x_2$ et la contrainte est $1 - x_1^2 - x_2^2 = 0$. Un point faible de cette fomulation est apparent : à la frontière du disque, la fonction pénalisée n'est pas deux fois continûment différentiable, et la pénalisation n'a absolument aucun effet à l'intérieur du disque, créant ainsi deux zones de minimisation distinctes. En présence de nombreuses contraintes, cette faiblesse peut se traduire par une explosion combinatoire des possibilités de contraintes actives.

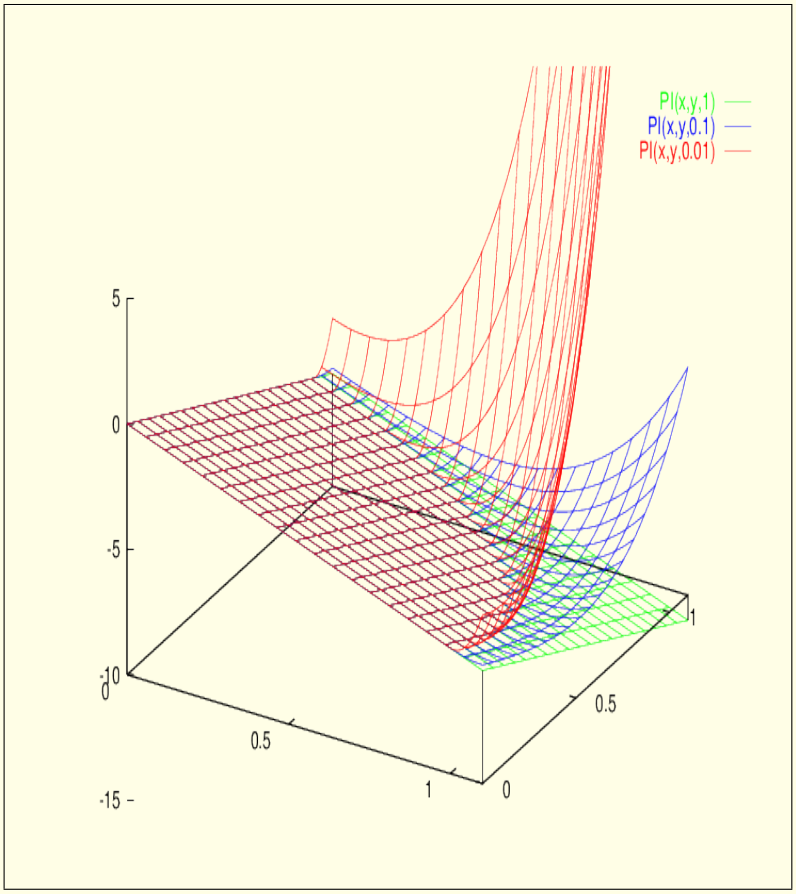


FIGURE 6.4 – Pénalisation $\|g(x)^+\|^2$

Exercice 6.3.2 *[Pénalité $\|g^+\|^2$]* Considérez le problème

$$\begin{aligned} \min \quad & f(x) = -x_1 - x_2 + x_3 \\ \text{sujet à} \quad & 0 \leqslant x_3 \leqslant 1 \\ & x_1^3 + x_3 \leqslant 1 \\ & x_1^2 + x_2^2 + x_3^2 \leqslant 1 \end{aligned}$$

On utilise la pénalité extérieure $\|g^+(x)\|^2$ pour traiter chacune des inégalités, et obtient les résultats suivants :

k	ρ_k	$x_1(\rho_k)$	$x_2(\rho_k)$	$x_3(\rho_k)$
1	1	0.834379	0.834379	-0.454846
2	0.1	0.728324	0.728324	-0.087920
3	0.01	0.709557	0.709557	-0.009864
4	0.001	0.707356	0.707356	-0.001017

a) Déduisez la solution optimale (stationnaire) ainsi que les multiplicateurs de Lagrange et indiquez quelles contraintes sont actives à la solution.

b) Est-ce que les conditions suffisantes sont satisfaites à la solution ?

6.3.4 Inégalités–barrière : $\lambda g(x) = 0$

Une autre possibilité consiste à perturber le terme de complémentarité. La perturbation la plus simple que l'on puisse imaginer est $\lambda g(x) = -\rho$, où encore une fois $\rho \searrow 0$. Supposons maintenant que les solutions des conditions d'optimalité ainsi perturbées satisfont $g(x(\rho)) < 0$ ainsi que $\lambda(\rho) > 0$, et donc l'inégalité $g(x) \leqslant 0$ est toujours satisfaite et $\lambda = \frac{-\rho}{g(x)}$ peut être substitué dans la première relation, ce qui donne

$$B(x, \rho) = \nabla f(x) - \sum \frac{\rho}{g_i(x)} \nabla g_i(x) = 0,$$

et correspond aux conditions d'optimalité de

$$\min b(x, \rho) = f(x) - \rho \sum \log(-g_i(x)).$$

Par la définition de la fonction *log*, les minima locaux de la fonction $b(x, \rho)$ seront tous des points *intérieurs* aux contraintes, et les valeur implicites $\lambda_i = -\frac{\rho}{g_i(x)}$ seront strictement positives, donc dans l'*intérieur* de l'ensemble $\lambda \geqslant 0$. Ces algorithmes sont connus sous le nom de *méthodes intérieures*.

On observe la barrière logarithmique pour des valeurs de $\rho = 1$, 0.1 et 0.01. La fonction est $f(x) = -x_1 - x_2$ et la contrainte est $1 - x_1^2 - x_2^2 = 0$. Cette formulation conduit, pour des programmes linéaires, à un algorithme polynômial en évitant les aspects combinatoires associés aux contraintes actives.

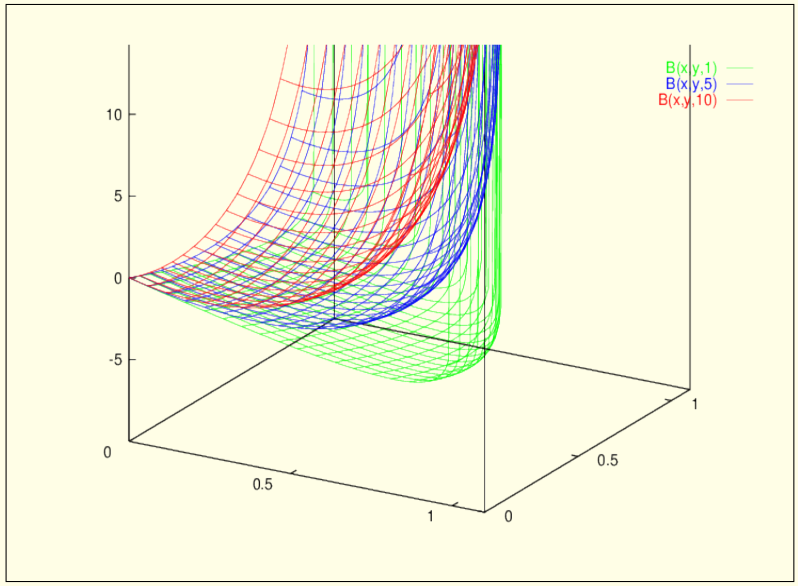


FIGURE 6.5 – Pénalisation $\sum \log(g_i(x))$

Exercice 6.3.3 *[[Algorithmes de pénalités-barrières]]* Considérez le problème

$$\begin{array}{rc} \min & f(x) = -x_1 - x_2 + x_3 \\ \text{sujet à} & 0 \leqslant x_3 \leqslant 2 \\ & x_1^3 + x_3 \leqslant 2 \\ & x_1^2 + x_2^2 + x_3^2 \leqslant 2 \end{array}$$

En supposant que toutes les contraintes sont uniformisées comme $g_i(x) \geqslant 0$, on utilise la barrière logarithmique $\rho_k \sum \log(g_i(x))$ pour traiter chacune des inégalités, et obtient les résultats suivants :

k	ρ_k	$x_1(\rho_k)$	$x_2(\rho_k)$	$x_3(\rho_k)$
1	1	0.4224168	0.6414188	0.3568029
2	0.1	0.8543391	1.0283938	0.0826567
3	0.01	0.9809094	1.0087353	0.0097632
4	0.001	0.9980099	1.0009862	0.0009975
5	0.0001	0.9998001	1.0000999	0.0001000

a) Déduisez la solution optimale (stationnaire) ainsi que les multiplicateurs de Lagrange et indiquez quelles contraintes sont actives à la solution.

b) Est-ce que la condition de stricte complémentarité est satisfaite à la solution ?

c) Est-ce que les conditions suffisantes sont satisfaites à la solution ?

6.3.5 Existence de trajectoires différentiables

Une propriété remarquable des pénalités et barrières est que les solutions suivent des trajectoires différentiables.

Égalités

Nous sommes intéressés à exprimer, proche d'un minimum local satisfaisant aux conditions *suffisantes* de second ordre pour le problème (6.1), les solutions $x(\rho)$ et $\lambda(\rho)$ du sous-problème

$$\theta(x,\lambda) = \begin{cases} \nabla f(x) + \lambda \nabla g(x) = 0 \\ \qquad g(x) - \rho\lambda^t = 0. \end{cases}$$

Nous avons déjà vu que les solutions de ce système sont les mêmes que les solutions du système

$$\nabla f(x) + \lambda \nabla g(x) = 0$$
$$\lambda = \frac{g(x)^t}{\rho},$$

qui représentent les conditions d'optimalité de premier ordre des sous-problèmes pénalisés. Cependant, le Jacobien $\Theta(x,\lambda) = \nabla\theta(x,\lambda)$ du système équivalent se rapproche de $\nabla^2 L(x^*,\lambda^*)$ qui est une matrice inversible :

$$\lim_{\rho\to 0} \Theta(x(\rho),\lambda(\rho)) = \nabla^2 L(x^*,\lambda^*).$$

Lemme 6.3.1 *Si x^* est un point régulier des contraintes, minimum local du programme (6.1) satisfaisant aux conditions suffisantes de second ordre, alors la matrice*

$$\begin{pmatrix} \nabla^2 f(x^*) + \lambda^*\nabla^2 g(x^*) & \nabla g(x^*)^t \\ \nabla g(x^*) & 0 \end{pmatrix},$$

est inversible.

Preuve Une matrice M inversible est de plein rang et donc si $Mv = 0$, alors $v = 0$. Vérifions donc cette propriété

$$\begin{pmatrix} \nabla^2 f(x^*) + \lambda^*\nabla^2 g(x^*) & \nabla g(x^*)^t \\ \nabla g(x^*) & 0 \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix} = 0.$$

Le second bloc du système nous assure que $v \in \ker \nabla g(x^*)$. Multiplions le premier bloc pat v^t pour obtenir

$$v^t(\nabla^2 f(x^*) + \lambda^*\nabla^2 g(x^*))v + v^t\nabla g(x^*)^t u = 0.$$

Puisque $v \in \ker \nabla g(x^*)$, $v^t \nabla g(x^*)^t u = 0$ et on a que pour de tels v, $v^t(\nabla^2 f(x^*) + \lambda^* \nabla^2 g(x^*))v = 0$, ce qui entraîne que $v = 0$ selon les conditions suffisantes d'optimalité. Puisque $v = 0$, le premier bloc est réduit à $\nabla g(x^*)^t u = 0$, et donc $u = 0$ car $\nabla g(x^*)$ est de plein rang. □ Ceci nous permet d'utiliser le théorème des fonctions implicites et de déduire l'existence des fonctions $x(\rho)$ et $\lambda(\rho)$ différentiables solutions du système linéaire transformé, et donc points stationnaires du sous-problème pénalisé. La discussion précédente constitue l'essentiel de la preuve du prochain théorème.

Théorème 6.3.1 *Soit x^* un point régulier des contraintes $g(x) = 0$ satisfaisant aux conditions suffisantes de second ordre pour le programme (6.1). Si les fonctions f et g sont de classe $\mathcal{C}^p(\mathbb{R}^n)$, alors, il existe des trajectoires différentiables $x(\rho)$ et $\lambda(\rho)$ de classe $\mathcal{C}^{p-1}(\mathbb{R}^n)$ telles que*

1. *$x(0) = x^*$ et $\lambda(0) = \lambda^*$;*
2. *si ρ est assez petit, $x(\rho)$ satisfait aux conditions suffisantes de second ordre pour le problème pénalisé $\min f(x) + \sum_{i=1}^m \frac{g_i(x)^2}{2\rho}$, où $x(\rho)$ et $\lambda(\rho)$ sont solutions du système d'équations*

$$\begin{aligned} \nabla f(x) + \lambda \nabla g(x) &= 0 \\ g(x) - \rho \lambda^t &= 0. \end{aligned}$$

Preuve L'existence de trajectoires différentiables est assurée par le théorème des fonctions implicites puisque la matrice Jacobienne $\Theta(x, \lambda)$ se rapproche d'une matrice inversible.

Pour montrer que les points $x(\rho)$ sont des minima locaux des sous-problèmes pénalisés, il suffit de montrer que les matrices hessiennes des objectifs pénalisés sont définies positives proche de x^*. En effet, les points $x(\rho)$ sont des points stationnaires pour les problèmes pénalisés puisqu'ils sont solutions du système transformé équivalent à leurs conditions d'optimalité.

Pour montrer que les $x(\rho)$ sont des minima locaux de $p(x, \rho)$, considérons une suite quelconque $\rho_k \searrow 0$ et procédons par contradiction en supposant que les matrices $H(x(\rho_k), \rho_k)$ possèdent toutes une valeur propre négative. Soient v_k une suite de vecteurs tels que

$$v_k^t H(x(\rho_k), \rho_k) v_k < 0.$$

Les deux premiers termes dans l'expression de $H(x, \rho)$ s'approchent du hessien en x du Lagrangien, donc d'une matrice bornée. Le dernier terme diverge (vers $+\infty$) à moins que $\nabla g(x(\rho_k))v_k$ s'approche de zéro assez vite. Donc, le seul moyen de rendre éventuellement $v_k^t H(x(\rho_k), \rho_k) v_k$ négatif est que v_k s'approche du plan tangent, et l'hypothèse que x^* satisfait aux conditions suffisantes assure alors que

$$v_k^t \left(\nabla^2 f(x(\rho_k)) + \frac{g(x(\rho_k))^t}{\rho_k} \nabla g(x(\rho_k)) \right) v_k$$

s'approche d'une quantité positive, ce qui établit la contradiction. □

Corollaire 6.3.1 *Sous les hypothèses du théorème 6.3.1,*

1. $\|x(\rho) - x^*\| \sim \mathcal{O}(\rho)$;
2. $\|\lambda(\rho) - \lambda^*\| \sim \mathcal{O}(\rho)$;
3. $\|g(x(\rho))\| \sim \mathcal{O}(\rho)$.

Exercice 6.3.4 *[Différentiabilité des trajectoires]* Suite de l'exercice 6.3.1 page 291 Vérifiez que à chaque valeur de ρ_k, les solutions x et les $\lambda(\rho)$ du tableau de l'exercice 6.3.1 satisfont aux bornes du corollaire.

Exercice 6.3.5 *[Trajectoires]* Considérez le problème

$$\begin{array}{ll} \min & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{sujet à} & x_1 + x_2 + x_3 = 1 \end{array}$$

Obtenez les expressions explicites pour les fonctions $x(\rho)$ et $\lambda(\rho)$ et vérifiez que $x(\rho)$ satisfait aux conditions suffisantes pour le formulation pénalisée suivante ; déduisez-en aussi la solution x^* et le multiplicateur λ^*.

$$\min \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \frac{1}{2\rho}\|(1 - (x_1 + x_2 + x_3))\|^2.$$

Exercice 6.3.6 *[Trajectoires différentiables]* Considérez le programme

$$\begin{array}{ll} \min & -x_1 - x_2 \\ \text{s.à} & 1 - x_1^2 - x_2^2 = 0 \end{array}$$

a) Vérifiez que la solution $x(\rho)$ de $\min f(x) + \frac{1}{2\rho}\|g(x)\|^2$ satisfait à $x_1 = x_2$, et $2x_1^3 - x_1 - \rho/2 = 0$.

b) Vérifiez que près de $\rho = 0$, $x_1(\rho) = \frac{1}{2}\sqrt{2} + a\rho + O(\rho^2)$. Que vaut a ?

c) Fournissez la solution x^* ainsi que le multiplicateur associé λ^*.

d) Vérifiez que $\frac{1-x_1(\rho)^2-x_2(\rho)^2}{\rho}$ approche λ^* à la même vitesse que ρ s'approche de 0.

Inégalités

On a toujours la différentiabilité des trajectoires approchées en autant que l'algorithme converge vers un point régulier des contraintes actives qui satisfait aux conditions suffisantes d'optimalité et à la condition de stricte complémentarité. Ces conditions sont *suffisantes* pour démontrer la différentiabilité des trajectoires. Elles ne sont toutefois pas nécessaires, comme nous le verrons dans un chapitre ultérieur, où nous obtiendrons ce résultat sans hypothèse additionnelle pour la programmation linéaire.

Utilisons la notation $\Lambda = diag(\lambda)$ pour récrire

$$\begin{aligned} \nabla f(x) + \lambda \nabla g(x) &= 0 \\ \Lambda g(x) &= -\rho e. \end{aligned}$$

La matrice jacobienne de ce système s'écrit alors, utilisant la notation $G(x) = diag(g(x))$

$$\begin{pmatrix} \nabla^2 f(x) + \sum \lambda_i \nabla^2 g_i(x) & \nabla g^t(x) \\ \Lambda \nabla g(x) & G(x) \end{pmatrix}.$$

Or, $\Lambda_J = 0$ et $g_{I*} = 0$ ce qui fait que cette dernière matrice converge vers

$$\left(\begin{array}{cc|c} \nabla^2 f(x^*) + \sum \lambda_i \nabla^2 g_i(x^*) & \nabla g^t_{I*}(x^*) & \nabla g^t_J(x) \\ \Lambda_{I*} \nabla g_{I*}(x^*) & 0 & 0 \\ \hline 0 & 0 & g_J(x^*) \end{array}\right). \tag{6.4}$$

On peut démontrer que cette dernière matrice est inversible en utilisant la même démarche que la démonstration du lemme 6.3.1. Ceci constitue un élement important dans la preuve du prochain théorème.

Exercice 6.3.7 *[Inversibilité]* Démontrez que cette dernière matrice 6.4 est inversible en utilisant la même démarche que la démonstration du lemme 6.3.1.

Théorème 6.3.2 *Soit x^* un point régulier des contraintes $g_{I*}(x) = 0$ satisfaisant aux conditions suffisantes de second ordre pour le programme (6.2) et à la condition de stricte complémentarité $\lambda_{I*} > 0$. Si les fonctions f et g sont de classe $\mathcal{C}^p(\mathbb{R}^n)$, alors, il existe des trajectoires différentiables $x(\rho)$ et $\lambda(\rho)$ de classe $\mathcal{C}^{p-1}(\mathbb{R}^n)$ telles que*

1. *$x(0) = x^*$ et $\lambda(0) = \lambda^*$;*
2. *si ρ est assez petit, $x(\rho)$ satisfait aux conditions suffisantes de second ordre pour le problème pénalisé $\min f(x) - \rho \sum_{i=1}^m \log(-g_i(x))$, où $x(\rho)$, $\lambda(\rho)$ sont solutions du système d'équations*

$$\begin{aligned} \nabla f(x) + \lambda \nabla g(x) &= 0 \\ g_i(x)\lambda_i + \rho &= 0 \quad , i = 1, 2, \ldots, m. \end{aligned}$$

De plus, les bornes suivantes sont satisfaites :

(a) $\|x(\rho) - x^*\| \sim \mathcal{O}(\rho)$;

(b) $\|\lambda(\rho) - \lambda^*\| \sim \mathcal{O}(\rho)$;

(c) $\|g(x(\rho))\| \sim \mathcal{O}(\rho)$.

***Exercice 6.3.8** *[Preuve du théorème 6.3.2]* Démontrez le théorème 6.3.2.

Exercice 6.3.9 *[Trajectoires]* Considérez le problème

$$\begin{array}{ll} \min & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{sujet à} & x_1 + x_2 + x_3 \geqslant 1 \end{array}$$

Obtenez les expressions explicites pour les fonctions $x(\rho)$ et $\lambda(\rho)$ et vérifiez que $x(\rho)$ satisfait aux conditions suffisantes pour les formulations pénalisées suivantes ; déduisez-en aussi la solution x^* et le multiplicateur λ^*.

a)
$$\min \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \frac{1}{2\rho}\|(1 - (x_1 + x_2 + x_3))^+\|^2.$$

b)
$$\min \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) - \rho \log(x_1 + x_2 + x_3 - 1).$$

Exercice 6.3.10 *[Pénalité $\|g^+\|^2$]* Suite de l'exercice 6.3.2 page 294 Estimez la solution optimale et les multiplicateurs de Lagrange à chaque itération k et vérifiez que la précision s'approche de 0 proportionnellement à la valeur de ρ_k.

Exercice 6.3.11 *[[Algorithmes de pénalités-barrières]]* Suite de l'exercice 6.3.3 page 296 Estimez la solution optimale et son erreur $\|x(\rho_k) - x^*\|$ et les multiplicateurs de Lagrange $\|\lambda(\rho_k) - \lambda^*\|$ à chaque itération k et vérifiez que la précision s'approche de 0 proportionnellement à la valeur de ρ_k.

Exercice 6.3.12 *[Barrière]* Considérez l'exemple (trivial) suivant ;

$$\begin{array}{rl} \min & x_1 + x_2 \\ \text{s.à} & x_1^2 + x_2^2 \leqslant 1. \end{array}$$

a) Formulez le problème de minimisation de la fonction barrière $b(x, \rho)$.

b) Obtenez la condition d'optimalité de premier ordre pour le sous-problème $\min_x b(x, \rho)$.

c) Vérifiez que la solution $x(\rho)$ de $\min \phi_\rho(x) = f(x) - \rho \ln(-g(x))$ satisfait à $x_1(\rho) = x_2(\rho)$, et fournissez-en l'expression en transformant la relation $\nabla \phi_\rho(x) = 0$ pour obtenir une équation quadratique en une seule de x_1 ou x_2.

d) Obtenez la dérivée $x'(\rho)$; que vaut $x'(0)$?

e) Fournissez la solution $x^* = x(0)$ ainsi que le multiplicateur associé λ^*.

f) Obtenez une expression pour la dérivée de $\lambda(\rho) = \frac{\rho}{1 - x_1(\rho)^2 - x_2(\rho)^2}$. Vérifiez que $\lambda(0) = \lambda^*$ et donnez la valeur de $\lambda'(0)$.

6.4 Linéarisation des conditions d'optimalité—SQP

Revenons aux conditions d'optimalité (6.3)

$$\Phi(x, \lambda) = \left\{ \begin{array}{r} \nabla f(x) + \lambda \nabla g(x) \\ g(x) \end{array} \right\} = 0. \tag{6.5}$$

Pour une valeur donnée de λ, on peut appliquer la méthode de Newton pour déterminer une solution en x de ces équations, et puisque

$$\nabla_x \Phi(x, \lambda) = \left\{ \begin{array}{r} \nabla^2 f(x) + \lambda \nabla^2 g(x) \\ \nabla g(x) \end{array} \right\}, \tag{6.6}$$

on retrouve comme équation linéarisée

$$\nabla_x \Phi(x, \lambda) d + \Phi(x, \lambda) = 0 \tag{6.7}$$

qui se ramènent à

$$\begin{array}{rrl} (\nabla^2 f(x) + \lambda \nabla^2 g(x)) d & + \nabla f(x) + \lambda \nabla g(x) & = 0 \\ \nabla g(x) d & + g(x) & = 0 \end{array}, \tag{6.8}$$

qui constituent, en utilisant $Q = (\nabla^2 f(x) + \lambda\nabla^2 g(x))$, $c = \nabla f$, les conditions d'optimalité de

$$\begin{aligned} \min_{xdin\mathbb{R}^n} q(d) &= \frac{1}{2}d^t Q d + cd \\ \text{sujet à } \nabla g(x)d + g(x) &= 0. \end{aligned} \tag{6.9}$$

Ceci inspire les méthodes SQP, consistant en la solution de sous-problèmes quadratiques (6.9). Les multiplicateurs de Lagrange du sous-problème quadratique servent de λ pour la prochaine itération.

L'approche est séduisante, mais pour établir des propriétés de convergence globale, il faut construire une *fonction de mérite*. Ceci nous ramène au contexte de solution de systèmes d'équations non linéaires.

6.4.1 Fonctions de mérite

Lorsque nous appliquons la méthode de Newton à un problème sans contrainte min $f(x)$, on peut modifier la méthode pour la rendre globalement convergence. Modification de la matrice hessienne de f et utilisation d'une recherche linéaire permettent de garantir que tout point d'accumulation sera stationnaire. La justification (démonstration) repose sur le fait que la méthode modifiée est un algorithme de descente, que la suite $f_k = f(x_k)$ est monotone décroissante.

Ici, deux objectifs sont implicites dans le problème : minimiser $f(x)$ et amener $g(x)$ à zéro. Comment faire pour témoigner du progrès de l'algorithme, comment convertir l'idée de base en un algorithme de descente ? Une solution est de construire une fonction de mérite $m(x)$ et justifier que $m_k = m(x_k)$ est monotone décroissante.

6.5 Méthodes primales-duales

En 1984, Narendra Karmarkar [24] a proposé un algorithme de complexité polynomiale pour la programmation linéaire qui a en quelque sorte révolutionné le domaine de l'optimisation mathématique. Plusieurs milliers de publications plus tard, les méthodes primales-duales irréalisables sont devenues incontournables. Les généralisations à des problèmes d'optimisation non linéaire sont délicates, cependant.

Nous présentons dans ce chapitre introductif une variante de base de telles méthodes pour la programmation linéaire. Revenons au problème 4.1

$$\begin{aligned} \min \quad & cx \\ \text{sujet à} \quad & Ax = b \\ & x \geqslant 0, \end{aligned}$$

et écrivons ses conditions d'optimalité comme si c'était un problème non linéaire, avec $x \geqslant 0$ et $s \geqslant 0$

$$\begin{array}{rcl} c + yA - s & = & 0 \\ Ax - b & = & 0 \\ sx & = & 0 \end{array}$$

où $X = diag(x)$. Les conditions d'optimalité se ramènent donc à deux systèmes linéaires et un système bi-linéaire $x_i s_i = 0$.

Par ailleurs, utilisons la barrière logarithmique pour les contraintes $x \geqslant 0$ pour obtenir le problème

$$\begin{array}{c} \min_{x>0} cx - \rho \sum \log(x_i) \\ \text{sujet à } Ax = b. \end{array} \tag{6.10}$$

et écrivons les conditions d'optimalité de ce problème (6.10) :

$$\begin{array}{rcl} c + yA - s & = & 0 \\ Ax - b & = & 0 \\ Sx & = & \rho e, \end{array}$$

où $e = (1, 1 \dots 1)^t$. Dans les méthodes de barrière, on poserait que $s_i = \frac{\rho}{x_i}$ alors que dans les méthodes primales-duales, on laisse les s_i évoluer de manière indépendante.

6.6 Tous les exercices du chapitres

***Exercice (6.1.1, page 280)** *[Preuve]* Imitez la preuve du théorème 6.1.1 pour démontrer le théorème 6.1.2.

Exercice (6.1.2, page 281) *[Optimalité]* Considérez le problème

$$\begin{array}{rc} \min & x_2 + x_3 \\ \text{s.à} & x_1 + x_2 + x_3 = 1 \\ & x_1^2 + x_2^2 + x_3^2 = 1 \end{array}$$

a) Par la symétrie de x_2 et x_3, reformulez le problème avec seulement deux variables, soit x_1 et $\hat{x}_2 = x_2 = x_3$. Obtenez deux points stationnaires pour ce problème en deux variables x_1 et $\hat{x}_2$. Illustrez graphiquement votre solution.

b) Vérifiez que $x' = (1, 0, 0)^t$ et $x'' = (-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})^t$ satisfont aux conditions nécessaires de Lagrange pour le problème original.

Exercice (6.1.3, page 282) *[Rectangle—bis]* Trouvez le rectangle de périmètre maximal pour une superficie donnée ($hl = 1$). Évidemment, un rectangle infiniment long et mince suggère, avec raison, que la solution n'est pas bornée. Supposons donc que la hauteur h et la longueur l sont bornées : $0.5 \leqslant h, l \leqslant 4$. Vérifiez que $l = 1$ et $h = 1$ est un point stationnaire, mais pas un maximum local ; quelle est la nature de ce point stationnaire ? Trouvez les maxima locaux, et le maximum global de ce problème.

Exercice (6.1.4, page 282) *[Projection]* Suite de l'exercice 5.3.6 page 229 Considérez le programme :

$$\begin{array}{rl} \min & gx \\ \text{s.à} & Ax = 0 \\ & \|x\|^2 = 1 \end{array}$$

Vérifiez que sa solution est un vecteur de même direction que celle trouvée à l'exercice 4.3.4.

Exercice (6.1.5, page 282) *[Points stationnaires]* Considérez le problème

$$\begin{array}{rc} \min & f(x) = \frac{1}{2}\alpha(x_1 - 1)^2 - x_1 - x_2 \\ \text{sujet à} & x_1 \geqslant x_2^2 \\ & x_1^2 + x_2^2 \leqslant 1 \end{array}$$

a) Trouvez tous les points qui rendent les 2 contraintes actives ; un de ces points (à 3 décimales de précision) est $\bar{x} = (0.618, 0.786)^t$.

b) Avec 3 décimales de précision, trouvez les valeurs du paramètre α (pas forcément positif) pour lesquelles $\bar{x}$ est un point stationnaire.

Exercice (6.1.6, page 285) *[Optimalité]* —suite de l'exercice 6.1.2, page 281 Déterminez la nature de ces quatre points stationnaires (les deux en a), et les deux en b)).

Exercice (6.1.7, page 285) *[Conditions d'optimalité]* Considérez une hyperbole de $\mathbb{R}^2$, $y = 1/x$. Soit un point $P = \binom{P_x}{P_y}$, on veut trouver le point de l'hyperbole le plus proche de P. On ramène l'étude au programme suivant dans $\mathbb{R}^2$:

$$(A) \left\{ \begin{array}{ll} \min d(x, y) = & \frac{1}{2}((x - P_x)^2 + (y - P_y)^2) \\ \text{sujet à} & xy = 1. \end{array} \right.$$

a) Écrivez les conditions d'optimalité (d'ordre un et deux) de ce problème.

b) Si le point P est sur la droite $y = x$, on devine qu'il y a 2 candidats solution naturels : $\binom{1}{1}$ et $\binom{-1}{-1}$. En vous limitant aux $x, y > 0$, vérifiez cette intuition en utilisant les conditions d'optimalité ; vérifiez d'abord que $x = y = 1$ est un point stationnaire pour le programme (A). Vérifiez ensuite que l'intuition n'est pas vraie si P est assez loin de l'origine ; quantifiez ce "loin" de l'origine en utilisant les conditions d'optimalité d'ordre 2.

Exercice (6.1.8, page 286) *[Conditions d'optimalité]* suite de l'exercice 5.4.8, page 245)
Soient les trois fonctions suivantes :

- $f_1(x) = \frac{1}{2}((x_1 - 3)^2 + (x_2 + 1)^2)$;
- $f_2(x) = \frac{1}{2}((x_1 + 1)^2 + (x_2 - 1)^2)$;
- $f_3(x) = \frac{1}{2}((x_1 + 1)^2 + x_2^2)$.

Considérez les (trois) contraintes

$$\begin{aligned} x_1 &\geqslant 0, \\ x_2 &\geqslant 0, \\ x_1^2 + x_2^2 &\leqslant 1 \end{aligned}$$

définissant l'ensemble réalisable E, et le point candidat $x = (1, 0)^t \in E$.

a) Pour chacune des fonctions, vérifiez que le point candidat est un point stationnaire, et donnez les valeurs des multiplicateurs de Lagrange λ_1, λ_2 et λ_3.

b) Pour chacune des fonctions, déterminez si le point stationnaire est un minimum local, maximum local ou ni l'un ni l'autre.

Exercice (6.1.9, page 287) *[Optimalité]* Considérez le problème suivant ; ATTENTION, c'est un max !

$$\begin{array}{ll} \max & x_1^2 + x_2^2 \\ \text{sujet à} & (x_2 - b)^2 \geqslant 4ax_1 \quad (0 < b < 4a) \\ & x_2 \leqslant b \\ & x \geqslant 0 \end{array}$$

a) obtenez 2 maxima locaux ; lequel est global ? Une illustration graphique aide considérablement pour cet exercice.

b) Est-ce que les conditions KKT sont satisfaites en ces points ?

c) Est-ce que les qualifications des contraintes sont satisfaites en ces points ?

d) Est-ce que les conditions suffisantes d'ordre 2 sont satisfaites en ces points ?

Exercice (6.1.10, page 287) *[MinMax]* Considérez le problème de *minmax* suivant :

$$\min_{x\in\mathbb{R}^n} \max_{i\in I} f_i(x),$$

qui peut être reformulé comme :

$$\begin{array}{cc} \min\limits_{(x,u)\in\mathbb{R}^n\times\mathbb{R}} & u \\ \text{sujet à} & f_i(x) \leqslant u. \end{array}$$

a) Donnez des conditions d'optimalité (nécessaires et suffisantes) pour ce problème ; discutez également de l'utilisation des *qualifications des contraintes* dans les conditions nécessaires d'optimalité. En fait, montrez que la condition de Mangasarian-Fromovitz est toujours satisfaite.

b) Considérez les fonctions $f_1(x) = \frac{1}{2}\|x + \binom{1}{1}\|^2$ et $f_2(x) = \frac{1}{2}\|x - \binom{1}{1}\|^2$. Illustrez graphiquement les lignes de niveau constant de $\max\{f_1(x), f_2(x)\}$, obtenez la solution et vérifiez les conditions d'optimalité et de qualifications de contraintes de a).

Exercice (6.3.1, page 291) *[Pénalité $\|g\|^2$]* Considérez le problème

$$\begin{array}{rc} \min & f(x) = -x_1 - x_2 + x_3 \\ \text{sujet à} & x_3 = 0 \\ & x_1^2 + x_2^2 + x_3^2 = 1 \end{array}$$

Évidemment, x_3 est inutile dans la formulation, mais est inclus pour rendre les calculs moins triviaux. On utilise la pénalité extérieure $\|g(x)\|^2$ pour traiter chacune des égalités, et obtient les résultats suivants :

k	ρ_k	$x_1(\rho_k)$	$x_2(\rho_k)$	$x_3(\rho_k)$
1	1	0.834377	0.834377	-0.454846
2	0.1	0.728326	0.728326	-0.087925
3	0.01	0.709559	0.709559	-0.009861
4	0.001	0.707356	0.707356	-0.000999

a) En examinant la tendance des valeurs $x(\rho)$ lorsque ρ s'approche de 0, déduisez la solution optimale du problème.

b) $\lambda(\rho) = \frac{g(x)}{\rho}$; en examinant la tendance des valeurs de $\lambda(\rho)$, déduisez les multiplicateurs de Lagrange optimaux du problème.

Exercice (6.3.2, page 294) *[Pénalité $\|g^+\|^2$]* Considérez le problème

$$\begin{array}{rc} \min & f(x) = -x_1 - x_2 + x_3 \\ \text{sujet à} & 0 \leqslant x_3 \leqslant 1 \\ & x_1^3 + x_3 \leqslant 1 \\ & x_1^2 + x_2^2 + x_3^2 \leqslant 1 \end{array}$$

On utilise la pénalité extérieure $\|g^+(x)\|^2$ pour traiter chacune des inégalités, et obtient les résultats suivants :

k	ρ_k	$x_1(\rho_k)$	$x_2(\rho_k)$	$x_3(\rho_k)$
1	1	0.834379	0.834379	-0.454846
2	0.1	0.728324	0.728324	-0.087920
3	0.01	0.709557	0.709557	-0.009864
4	0.001	0.707356	0.707356	-0.001017

a) Déduisez la solution optimale (stationnaire) ainsi que les multiplicateurs de Lagrange et indiquez quelles contraintes sont actives à la solution.

b) Est-ce que les conditions suffisantes sont satisfaites à la solution ?

Exercice (6.3.3, page 296) *[[Algorithmes de pénalités-barrières]]* Considérez le problème

$$\begin{array}{rc} \min & f(x) = -x_1 - x_2 + x_3 \\ \text{sujet à} & 0 \leqslant x_3 \leqslant 2 \\ & x_1^3 + x_3 \leqslant 2 \\ & x_1^2 + x_2^2 + x_3^2 \leqslant 2 \end{array}$$

En supposant que toutes les contraintes sont uniformisées comme $g_i(x) \geqslant 0$, on utilise la barrière logarithmique $\rho_k \sum \log(g_i(x))$ pour traiter chacune des inégalités, et obtient les résultats suivants :

k	ρ_k	$x_1(\rho_k)$	$x_2(\rho_k)$	$x_3(\rho_k)$
1	1	0.4224168	0.6414188	0.3568029
2	0.1	0.8543391	1.0283938	0.0826567
3	0.01	0.9809094	1.0087353	0.0097632
4	0.001	0.9980099	1.0009862	0.0009975
5	0.0001	0.9998001	1.0000999	0.0001000

a) Déduisez la solution optimale (stationnaire) ainsi que les multiplicateurs de Lagrange et indiquez quelles contraintes sont actives à la solution.

b) Est-ce que la condition de stricte complémentarité est satisfaite à la solution ?

c) Est-ce que les conditions suffisantes sont satisfaites à la solution ?

Exercice (6.3.4, page 299) *[Différentiabilité des trajectoires]* Suite de l'exercice 6.3.1 page 291 Vérifiez que à chaque valeur de ρ_k, les solutions x et les $\lambda(\rho)$ du tableau de l'exercice 6.3.1 satisfont aux bornes du corollaire.

Exercice (6.3.5, page 299) *[Trajectoires]* Considérez le problème

$$\begin{array}{rl} \min & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{sujet à} & x_1 + x_2 + x_3 = 1 \end{array}$$

Obtenez les expressions explicites pour les fonctions $x(\rho)$ et $\lambda(\rho)$ et vérifiez que $x(\rho)$ satisfait aux conditions suffisantes pour le formulation pénalisée suivante ; déduisez-en aussi la solution x^* et le multiplicateur λ^*.

$$\min \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \frac{1}{2\rho}\|(1 - (x_1 + x_2 + x_3))\|^2.$$

Exercice (6.3.6, page 299) *[Trajectoires différentiables]* Considérez le programme

$$\begin{array}{rl} \min & -x_1 - x_2 \\ \text{s.à} & 1 - x_1^2 - x_2^2 = 0 \end{array}$$

a) Vérifiez que la solution $x(\rho)$ de $\min f(x) + \frac{1}{2\rho}\|g(x)\|^2$ satisfait à $x_1 = x_2$, et $2x_1^3 - x_1 - \rho/2 = 0$.

b) Vérifiez que près de $\rho = 0$, $x_1(\rho) = \frac{1}{2}\sqrt{2} + a\rho + O(\rho^2)$. Que vaut a ?

c) Fournissez la solution x^* ainsi que le multiplicateur associé λ^*.

d) Vérifiez que $\frac{1 - x_1(\rho)^2 - x_2(\rho)^2}{\rho}$ approche λ^* à la même vitesse que ρ s'approche de 0.

Exercice (6.3.7, page 300) *[Inversibilité]* Démontrez que cette dernière matrice 6.4 est inversible en utilisant la même démarche que la démonstration du lemme 6.3.1.

***Exercice (6.3.8, page 301)** *[Preuve du théorème 6.3.2]* Démontrez le théorème 6.3.2.

Exercice (6.3.9, page 301) *[Trajectoires]* Considérez le problème

$$\begin{array}{rl} \min & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{sujet à} & x_1 + x_2 + x_3 \geqslant 1 \end{array}$$

Obtenez les expressions explicites pour les fonctions $x(\rho)$ et $\lambda(\rho)$ et vérifiez que $x(\rho)$ satisfait aux conditions suffisantes pour les formulations pénalisées suivantes ; déduisez-en aussi la solution x^* et le multiplicateur λ^*.

a)
$$\min \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \frac{1}{2\rho}\|(1 - (x_1 + x_2 + x_3))^+\|^2.$$

b)
$$\min \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) - \rho \log(x_1 + x_2 + x_3 - 1).$$

Exercice (6.3.10, page 301) *[Pénalité $\|g^+\|^2$]* Suite de l'exercice 6.3.2 page 294 Estimez la solution optimale et les multiplicateurs de Lagrange à chaque itération k et vérifiez que la précision s'approche de 0 proportionnellement à la valeur de ρ_k.

Exercice (6.3.11, page 301) *[[Algorithmes de pénalités-barrières]]* Suite de l'exercice 6.3.3 page 296 Estimez la solution optimale et son erreur $\|x(\rho_k) - x^*\|$ et les multiplicateurs de Lagrange $\|\lambda(\rho_k) - \lambda^*\|$ à chaque itération k et vérifiez que la précision s'approche de 0 proportionnellement à la valeur de ρ_k.

Exercice (6.3.12, page 302) *[Barrière]* Considérez l'exemple (trivial) suivant ;

$$\begin{array}{rl} \min & x_1 + x_2 \\ \text{s.à} & x_1^2 + x_2^2 \leqslant 1. \end{array}$$

a) Formulez le problème de minimisation de la fonction barrière $b(x, \rho)$.

b) Obtenez la condition d'optimalité de premier ordre pour le sous-problème $\min_x b(x, \rho)$.

c) Vérifiez que la solution $x(\rho)$ de $\min \phi_\rho(x) = f(x) - \rho \ln(-g(x))$ satisfait à $x_1(\rho) = x_2(\rho)$, et fournissez-en l'expression en transformant la relation $\nabla\phi_\rho(x) = 0$ pour obtenir une équation quadratique en une seule de x_1 ou x_2.

d) Obtenez la dérivée $x'(\rho)$; que vaut $x'(0)$?

e) Fournissez la solution $x^* = x(0)$ ainsi que le multiplicateur associé λ^*.

f) Obtenez une expression pour la dérivée de $\lambda(\rho) = \frac{\rho}{1 - x_1(\rho)^2 - x_2(\rho)^2}$. Vérifiez que $\lambda(0) = \lambda^*$ et donnez la valeur de $\lambda'(0)$.

Chapitre 7

Méthodes de pénalité, barrière et Lagrangien augmenté

Sujets du chapitre

- Étude détaillée de la famille de méthodes de pénalité.
- Présentation d'un algorithme de pénalité augmentée.