

Projection et activation

Dans ce chapitre, nous présentons quelques méthodes d'optimisation dans lequel le critère est non linéaire et les contraintes sont suffisamment simples. Elles s'inspirent directement des techniques de descente en optimisation sans contrainte, en utilisant la notion de direction de descente et en faisant décroître le critère à chaque itération. Elles ont aussi la caractéristique de maintenir la suite des itérés $\{x_k\}$ dans l'ensemble admissible X . Si l'approche est générale, elle ne trouvera son efficacité que si les contraintes sont affines, si bien que le problème devra pouvoir être mis sous la forme suivante :

$$\begin{cases} \min f(x) \\ Ax \leq b, \end{cases} \quad (12.1)$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction régulière, A est une matrice $m \times n$ et $b \in \mathbb{R}^m$.

Une direction de descente de f en un point qui n'est pas dans l'ensemble admissible X est de peu d'utilité : si elle permet de faire décroître f , elle n'est d'aucune aide pour ramener les itérés dans l'ensemble admissible. Par conséquent, si on désire utiliser des directions de descente de f , on est forcé de générer des suites $\{x_k\}$ *admissibles*, c'est-à-dire contenues dans X . On est alors amené à introduire la notion de *direction de descente admissible* d_k . De petits pas le long de ces directions permettent de faire décroître f tout en restant dans X . Alors on peut par exemple définir la suite $\{x_k\}$ par $x_{k+1} = x_k + \alpha_k d_k$, avec α_k suffisamment petit pour que f décroisse et que x_{k+1} soit dans X .

Cette méthode est attrayante par sa simplicité, mais **elle n'est pas convergente**. On connaît en effet un contre-exemple, dû à Wolfe [625 ; 1972]. Nous n'allons pas le décrire en détail, mais il est instructif de comprendre ce qui s'y passe. Dans celui-ci (voir figure 12.1), on minimise sur l'**orthant positif** ($X = \{x \in \mathbb{R}^3 : x \geq 0\}$) une fonction convexe de 3 variables; on utilise la méthode ci-dessus, en prenant $d_k = -g_k$ comme direction de descente admissible; on prend le pas optimal si celui-ci laisse x_{k+1} dans X , sinon on prend le plus grand pas possible de manière à rester dans X . On constate que l'algorithme construit une suite $\{x_k\}$ oscillant entre deux faces de l'orthant positif et convergeant vers un point *non* stationnaire (!) situé sur l'intersection des deux faces. Que se passe-t-il ? Une façon d'interpréter ce résultat est de dire que les bornes ($x \geq 0$) de l'ensemble admissible forcent le pas à être trop petit et provoquent ainsi la « fausse » convergence de la suite. On a vu, en effet, au chapitre 6.3, qu'un des devoirs de la recherche linéaire est d'empêcher le pas d'être trop petit afin d'assurer la convergence des méthodes à directions de descente. La présence des contraintes d'inégalité empêche ici de satisfaire cette propriété.

On peut trouver plusieurs remèdes pour éviter cette fausse convergence. L'un d'eux consiste à poursuivre la recherche du pas au delà du *point d'activation* $x_k + \hat{\alpha}_k d_k$ du

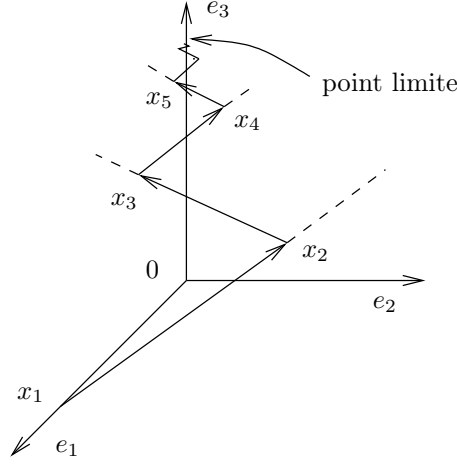


Fig. 12.1. Contre-exemple de Wolfe

chemin $\alpha \mapsto x_k + \alpha d_k$ sur le bord de X . Comme l'on veut que les itérés restent dans X , la recherche se poursuit alors, non pas le long de d_k , mais le long du chemin $\alpha \mapsto x_k + \alpha d_k$ projeté sur X . On parle de *méthodes de projection*. Cette approche sera examinée à la section 12.1. On comprend que pour que celle-ci soit efficace, il faut que la projection sur X soit peu coûteuse. C'est le cas notamment des ensembles X définis par des *contraintes de borne* :

$$X = \{x \in \mathbb{R}^n : a \leq x \leq b\}, \quad a \text{ et } b \in \mathbb{R}^n. \quad (12.2)$$

Une autre approche consiste à bloquer certaines contraintes pendant quelques itérations (cas où X est donné par des contraintes d'inégalité affines), c'est-à-dire de considérer certaines contraintes d'inégalité comme des contraintes d'égalité. On parle alors de *méthodes différentiation automatique*, aussi appelées *méthodes de pivotage* dans les problèmes de complémentarité [141] (section 12.2). Ces méthodes sont surtout utilisées en *optimisation quadratique*, c'est-à-dire pour la minimisation de fonctions quadratiques en présence de contraintes affines :

$$X = \{x \in \mathbb{R}^n : Ax \leq b\}, \quad A : m \times n, \quad b \in \mathbb{R}^m. \quad (12.3)$$

12.1 Méthode de projection

Dans cette section, nous étudions les méthodes dans lesquelles le chemin de recherche de l'itéré suivant est obtenu en projetant sur X le chemin affine $\alpha \mapsto x_k + \alpha d_k$:

$$x_k(\alpha) := P_X(x_k + \alpha d_k), \quad (12.4)$$

où $d_k \in \mathbb{R}^n$, $\alpha > 0$ et P_X est un opérateur de projection sur X (voir figure 12.2). Par conséquent, $\alpha \mapsto x_k(\alpha)$ est un chemin de recherche admissible et en prenant

$$x_{k+1} = x_k(\alpha_k), \quad \alpha_k > 0, \quad (12.5)$$

on générera une suite $\{x_k\}$ admissible, c'est-à-dire avec $x_k \in X$.

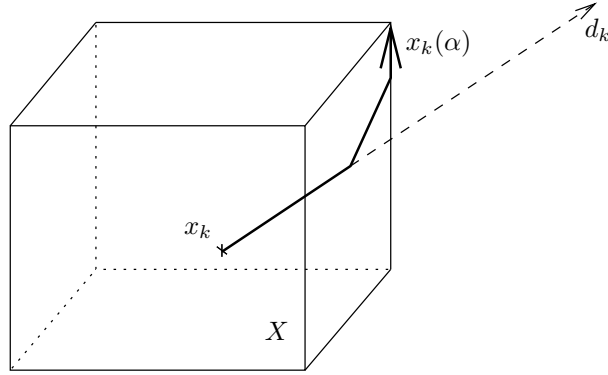


Fig. 12.2. Méthode du chemin projeté

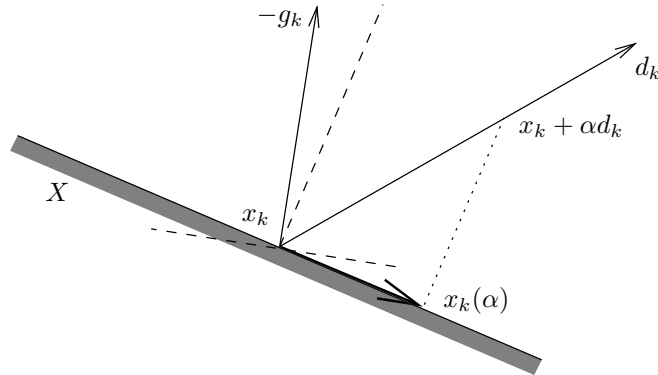
12.1.1 Méthode du gradient projeté

Venons-en maintenant à l'étude d'un algorithme de résolution du problème (12.1), basé sur (12.4)–(12.5). On suppose que X est un convexe fermé non vide de $\mathbb{R}^n$.

La première question qui se pose est de savoir si l'on fait décroître f le long du chemin $x_k(\alpha)$ défini en (12.4). En général, ce n'est pas le cas, tout au moins si d_k est une direction de descente stricte quelconque de f (c.-à-d., une direction vérifiant $g_k^T d_k < 0$, $g_k = \nabla f(x_k)$). On peut en effet construire aisément un exemple où l'on n'a pas

$$f(x_k(\alpha)) < f(x_k), \quad \text{pour } \alpha > 0 \text{ petit.} \quad (12.6)$$

La figure 12.3 illustre une telle situation. La direction d_k est de descente car elle forme avec $-g_k$ un angle plus petit que $\pi/2$ (on suppose que gradient et angle font ici référence au produit scalaire euclidien), ce qui n'est pas le cas du chemin projeté $\alpha \mapsto x_k(\alpha)$. Cependant, si l'on prend $d_k = -g_k$ et donc

Fig. 12.3. Projection d'une direction de descente d_k quelconque

$$x_k(\alpha) := P_X(x_k - \alpha g_k), \quad (12.7)$$

la condition (12.6) est vérifiée. Nous allons le voir.

La méthode utilisant le chemin (12.7) comme chemin de descente porte le nom de *méthode du gradient projecté*. Elle a été introduite par Goldstein [262 ; 1964] et Levitin et Polyak [391 ; 1966]. Cette méthode n'est pratiquement utilisable que lorsque la projection sur l'ensemble des contraintes est facile à réaliser, par exemple dans le cas de contraintes de borne (12.2). Les lemmes 12.1, 12.2 et 12.3 donnent quelques propriétés fondamentales de cette méthode.

Lemme 12.1 *Soit X un convexe fermé non vide de $\mathbb{R}^n$. Alors, le point $\bar{x} \in X$ est un point stationnaire du problème (12.1) si, et seulement si,*

$$\bar{x} = P_X(\bar{x} - \alpha \nabla f(\bar{x})), \quad (12.8)$$

pour un (ou tout) $\alpha > 0$.

DÉMONSTRATION. D'après la proposition 4.7, lorsque X est convexe, la stationnarité de f en $\bar{x}$ s'exprime par

$$\langle \nabla f(\bar{x}), y - \bar{x} \rangle \geq 0, \quad \forall y \in X.$$

Pour $\alpha > 0$, cette condition est équivalente à

$$\langle \bar{x} - (\bar{x} - \alpha \nabla f(\bar{x})), y - \bar{x} \rangle \geq 0, \quad \forall y \in X.$$

Comme $\bar{x} \in X$, la caractérisation (2.28a) implique que cette condition est équivalente à (12.8). $\square$

En conséquence de ce lemme, l'itéré courant x_k de l'algorithme (12.5) est un point stationnaire du problème (12.1) si, et seulement si, $x_{k+1} = x_k$.

Lemme 12.2 *On suppose que X est un convexe fermé non vide de $\mathbb{R}^n$ et pour $x_k \in X$, on considère le chemin $\alpha \mapsto x_k(\alpha)$ défini en (12.7). Alors, pour tout $\alpha \in \mathbb{R}$, on a*

$$\alpha \langle g_k, x_k(\alpha) - x_k \rangle \leq -\|x_k(\alpha) - x_k\|^2. \quad (12.9)$$

DÉMONSTRATION. Comme $x_k \in X$ et que $x_k(\alpha)$ est la projection de $x_k - \alpha g_k$ sur X , la caractérisation (2.28a) donne

$$\langle x_k - x_k(\alpha), x_k(\alpha) - x_k + \alpha g_k \rangle \geq 0.$$

On en déduit (12.9). $\square$

Ce résultat montre que

$$s_k(\alpha) := x_k(\alpha) - x_k$$

est une direction de descente de f en x_k , quel que soit $\alpha > 0$: l'angle entre $-g_k$ et $s_k(\alpha)$ reste toujours inférieur à $\pi/2$.

Lemme 12.3 *On suppose que X est un convexe fermé non vide de $\mathbb{R}^n$, que f est de classe C^1 et que sa dérivée est lipschitzienne, de module L . Soit $x_k \in X$. Alors, si $0 < \sigma < 1$ et $0 \leq \alpha \leq 2(1 - \sigma)/L$, on a*

$$f(x_k(\alpha)) \leq f(x_k) + \sigma \langle g_k, x_k(\alpha) - x_k \rangle. \quad (12.10)$$

DÉMONSTRATION. On a

$$\begin{aligned} f(x_k(\alpha)) - f(x_k) &= \int_0^1 f'(x_k + ts_k(\alpha)) \cdot s_k(\alpha) dt \\ &\leq \langle g_k, s_k(\alpha) \rangle + \frac{L}{2} \|s_k(\alpha)\|^2. \end{aligned}$$

En utilisant (12.9), on trouve

$$f(x_k(\alpha)) - f(x_k) \leq \left(1 - \frac{L\alpha}{2}\right) \langle g_k, s_k(\alpha) \rangle.$$

En remarquant que, d'après (12.9), $\langle g_k, s_k(\alpha) \rangle$ est négatif lorsque α est positif, on en déduit (12.10) lorsque $0 \leq \alpha \leq 2(1 - \sigma)/L$. $\square$

Dès lors, lorsque x_k n'est pas stationnaire, les inégalités (12.9) et (12.10) montrent que l'on fait décroître f en se déplaçant le long du chemin $\alpha \mapsto x_k(\alpha)$ ($\alpha > 0$ petit).

Les deux résultats suivants donnent des méthodes de détermination du pas α qui assurent la convergence de la méthode du gradient projeté : pas petit (théorème 12.4) et pas d'Armijo (théorème 12.5). Ce sont des extensions naturelles de résultats connus pour les problèmes sans contraintes.

Théorème 12.4 *Soit X un convexe fermé non vide de $\mathbb{R}^n$. Supposons que f soit bornée inférieurement sur X , que f soit de classe C^1 et que sa dérivée soit lipschitzienne de module L . Alors l'algorithme du gradient projeté, avec un pas α_k pris dans un compact de $]0, 2/L[$ génère une suite de points $\{x_k\}$ telle que*

$$\|x_{k+1} - x_k\| \rightarrow 0.$$

De plus, tous les points d'adhérence de $\{x_k\}$ sont stationnaires.

DÉMONSTRATION. Si α_k est dans un compact de $]0, 2/L[$, on peut trouver $\sigma \in]0, 1[$ tel que

$$\alpha_k \in \left[\sigma, \frac{2(1 - \sigma)}{L} \right].$$

Alors, d'après les lemmes 12.2 et 12.3, on a

$$\begin{aligned}
f(x_{k+1}) &\leq f(x_k) - \sigma \frac{\|x_{k+1} - x_k\|^2}{\alpha_k} \\
&\leq f(x_k) - \frac{\sigma L}{2(1-\sigma)} \|x_{k+1} - x_k\|^2.
\end{aligned}$$

Comme f est bornée inférieurement sur X et que $f(x_k)$ décroît, $f(x_{k+1}) - f(x_k) \rightarrow 0$. On en déduit que $\|x_{k+1} - x_k\| \rightarrow 0$.

Si $\bar{x}$ est un point d'adhérence de $\{x_k\}$, on peut trouver une sous-suite d'indices $\mathcal{K}$ telle que $x_k \rightarrow \bar{x}$, pour $k \rightarrow \infty$ dans $\mathcal{K}$. Comme $\|x_{k+1} - x_k\| \rightarrow 0$, on a aussi $x_{k+1} \rightarrow \bar{x}$, pour $k \rightarrow \infty$ dans $\mathcal{K}$. On peut également supposer que $\alpha_k \rightarrow \bar{\alpha}$, pour $k \rightarrow \infty$ dans $\mathcal{K}$ (au besoin on extrait une nouvelle sous-suite de $\mathcal{K}$). En passant à la limite dans

$$x_{k+1} = P_X(x_k - \alpha_k \nabla f(x_k))$$

(on se rappelle que P_X est continu, par (2.29)), on a

$$\bar{x} = P_X(\bar{x} - \bar{\alpha} \nabla f(\bar{x})).$$

Comme $\bar{\alpha} > 0$, le lemme 12.3 montre que $\bar{x}$ est stationnaire. $\square$

En général la constante de Lipschitz de f' n'est pas connue et il n'est pas réaliste (et il est souvent inefficace) de choisir le pas comme dans le théorème 12.4. Il est préférable de déterminer α_k selon la règle suivante qui est une extension de la règle d'Armijo utilisée en optimisation sans contrainte (voir section 6.3.3).

On se donne deux constantes $\sigma \in]0, 1[$ et $\beta \in]0, 1[$ ainsi qu'une estimation $\alpha_k^0 > 0$ du pas à l'itération k (il est important que $\{\alpha_k^0\}$ et $\{1/\alpha_k^0\}$ forment des suites bornées). On prend $\alpha_k = \beta^{j_k} \alpha_k^0$, j_k étant le plus petit entier tel qu'avec ce α_k , on ait

$$f(x_k(\alpha_k)) \leq f(x_k) + \sigma \langle g_k, x_k(\alpha_k) - x_k \rangle. \quad (12.11)$$

Lorsque f est régulière, on peut toujours trouver un tel pas α_k . En effet, d'après le lemme 12.3, cette inégalité est vérifiée lorsque α_k est plus petit que $2(1-\sigma)/L$ (où L est la constante de Lipschitz de f). Ceci montre que par la règle ci-dessus, on a

$$\alpha_k \geq \min \left(\alpha_k^0, \frac{2(1-\sigma)\beta}{L} \right). \quad (12.12)$$

On prend ensuite $x_{k+1} = x_k(\alpha_k)$.

Le théorème suivant montre que cette manière de procéder est satisfaisante.

Théorème 12.5 ([50]) *Soit X un convexe fermé non vide de $\mathbb{R}^n$. Supposons que f soit bornée inférieurement sur X , que f soit de classe C^1 , avec une dérivée lipschitzienne. Alors, l'algorithme du gradient projeté, avec un pas α_k déterminé par l'extension de la règle d'Armijo (12.11), génère une suite $\{x_k\}$ telle que*

$$\|x_{k+1} - x_k\| \rightarrow 0. \quad (12.13)$$

De plus, tous les points d'adhérence de $\{x_k\}$ sont stationnaires.

DÉMONSTRATION. D'après (12.11), le lemme 12.2 et le fait que $\alpha_k \leq \alpha_k^0 \leq C$, on a

$$f(x_{k+1}) \leq f(x_k) - \frac{\sigma}{C} \|x_{k+1} - x_k\|^2.$$

Comme $f(x_k)$ converge, on en déduit (12.13). Si $\bar{x}$ est un point d'adhérence de $\{x_k\}$, on peut trouver une sous-suite d'indices $\mathcal{K}$ telle que

$$x_k \rightarrow \bar{x}, \quad \bar{\alpha}_k \rightarrow \bar{\alpha}, \quad \text{pour } k \rightarrow \infty \text{ dans } \mathcal{K}.$$

D'après (12.12) et le fait que $\{1/\alpha_k^0\}$ est bornée, on a $\bar{\alpha} > 0$. On s'y prend ensuite comme à la fin de la preuve de le théorème 12.4. $\square$

Étant une extension de la méthode de la plus forte pente, la méthode du gradient projeté n'a pas une bonne vitesse de convergence. On peut toutefois utiliser les idées développées ci-dessus pour définir une méthode de Newton projetée : voir Bertsekas [52 ; 1982]. On peut aussi préconditionner la méthode en utilisant des produits scalaires différents pour la définition du gradient de f et pour la projection sur X : voir Gafni et Bertsekas [227 ; 1984].

12.2 Méthodes différentiation automatique ▲

La méthode que nous décrivons dans cette section est fondée sur une idée assez générale. Toutefois, d'un point de vue pratique, elle ne pourra être utilisée avec efficacité qu'en optimisation quadratique. Ces restrictions apparaîtront au cours de l'exposé.

12.2.1 Motivation et schéma des méthodes

Pour construire une suite minimisante $\{x_k\}$ admissible, c'est-à-dire contenue dans X , on peut s'appuyer sur la notion de direction de descente admissible.

Définition 12.6 Soit $x \in X$. On dit que d est une *direction admissible* en x pour X si $x + \alpha d \in X$ pour tout $\alpha > 0$ suffisamment petit (de manière plus précise : il existe $\bar{\alpha} > 0$ tel que pour $0 \leq \alpha \leq \bar{\alpha}$ on ait $x + \alpha d \in X$). On dit que d est une *direction de descente admissible* en x pour le problème (12.1), s'il existe $\bar{\alpha} > 0$ tel que pour $0 \leq \alpha \leq \bar{\alpha}$:

$$x + \alpha d \in X \quad \text{et} \quad f(x + \alpha d) \leq f(x). \quad (12.14)$$

Il s'agit donc d'une direction de descente (non stricte) de f telle qu'avec un petit déplacement dans cette direction, on reste dans X . $\square$

On ne peut pas toujours trouver une direction vérifiant ces conditions, mais si X est convexe et x_k n'est pas stationnaire, une telle direction existe toujours. En effet, du fait de la convexité de X , les directions de la forme $d_k = y_k - x_k$, où $y_k \in X$ et $y_k \neq x_k$ sont des directions admissibles pour X . Alors, s'il n'existait pas de direction de descente admissible en x_k , cela voudrait dire que, pour $y_k \in X$ fixé, il existerait une suite de $\alpha^i > 0$, $\alpha^i \rightarrow 0$, telle que $f(x_k + \alpha^i(y_k - x_k)) > f(x_k)$. En passant à la limite, on trouverait

$$\langle g_k, y_k - x_k \rangle \geq 0, \quad \forall y_k \in X.$$

Ceci impliquerait la stationnarité de x_k (proposition 4.7). Nous supposons donc que X est **convexe** pour pouvoir trouver à chaque étape une direction de descente admissible.

Nous allons utiliser des directions admissibles dans une approche fondée sur la notion d'*activation de contraintes* (en anglais, *Active Set Methods*). L'idée est de bloquer — on dit aussi *activer* — certaines contraintes définissant X , c'est-à-dire que des contraintes d'inégalité seront considérées comme contraintes d'égalité pendant un certain nombre d'itérations (on évite ainsi l'oscillation qui se produit dans le contre-exemple de Wolfe). Ceci n'est concevable que si les contraintes sont affines, sinon la réalisation de $c_i(x_k) = 0$ est trop coûteuse. Pour simplifier l'exposé, nous supposons qu'il n'y a pas de contraintes d'égalité.

Le problème (P) à résoudre dans cette section est donc

$$(P) \quad \begin{cases} \min f(x) \\ Ax \leq b, \end{cases} \quad (12.15)$$

où A est une matrice $m \times n$ et $b \in \mathbb{R}^m$. L'ensemble admissible est convexe (clair). Les équations d'optimalité s'écrivent en $(\bar{x}, \bar{\lambda})$ (les contraintes étant affines, elles sont qualifiées; on peut donc trouver un multiplicateur $\bar{\lambda}$):

$$\begin{cases} \nabla f(\bar{x}) + A^T \bar{\lambda} = 0, \\ A\bar{x} \leq b, \\ \bar{\lambda} \geq 0, \\ \bar{\lambda}^T (A\bar{x} - b) = 0. \end{cases} \quad (12.16)$$

On note $A_{(j)}$, la j -ième ligne de A . Si $J \subseteq [1:m]$ est un ensemble d'indices, on note A_J la matrice $|J| \times n$, obtenue en extrayant de A ses lignes $j \in J$. On définit $b_{(j)}$ et b_J de manière analogue. Enfin, l'ensemble actif en x_k , $I^0(x_k)$, se notera ici simplement

$$I_k = \{j : 1 \leq j \leq m, A_{(j)}x_k = b_{(j)}\}.$$

Soit T_k une partie de I_k qu'on appelle *ensemble de travail*. On considère alors le problème

$$(P_{T_k}) \quad \begin{cases} \min f(x) \\ A_{T_k}x = b_{T_k}, \end{cases} \quad (12.17)$$

dans lequel x est contraint à appartenir à l'*espace actif*

$$E_{T_k} = \{x \in \mathbb{R}^n : A_{T_k}x = b_{T_k}\}.$$

Ce problème est en général beaucoup plus facile à résoudre que le problème (12.15). Ses équations d'optimalité du premier ordre s'écrivent : $\exists \lambda_k \in \mathbb{R}^m$ tel que

$$\begin{cases} \nabla f(x_k) + A_{T_k}^\top (\lambda_k)_{T_k} = 0 \\ A_{T_k} x_k = b_{T_k}. \end{cases} \quad (12.18)$$

On a le résultat instructif suivant.

Proposition 12.7 *Soient $\bar{x}$ une solution du problème (P) et x_k une solution du problème (P_{T_k}) , dans lequel $T_k = I^0(\bar{x})$.*

- (i) *Si f est convexe sur E_{T_k} et $x_k \in X$, alors x_k est solution de (P) .*
- (ii) *Si f est strictement convexe sur E_{T_k} , alors $x_k = \bar{x}$.*

DÉMONSTRATION. Montrons que

$$f(x_k) = f(\bar{x}).$$

D'une part, $f(x_k) \leq f(\bar{x})$, car $\bar{x}$ est solution du problème (P_{T_k}) auquel on ajoute la contrainte d'appartenance à X . D'autre part, si $d_k = x_k - \bar{x}$, la convexité de f sur E_{T_k} permet d'écrire

$$f(\bar{x}) + f'(\bar{x}) \cdot d_k \leq f(x_k).$$

Alors, si $f(x_k) < f(\bar{x})$, on aurait $f'(\bar{x}) \cdot d_k < 0$ et d_k serait une direction de descente stricte de f en $\bar{x}$:

$$f(\bar{x} + \alpha d_k) < f(\bar{x}), \quad \text{pour } \alpha > 0 \text{ petit.}$$

Mais comme $d_k \in \mathcal{N}(A_{I^0(\bar{x})})$, on a

$$\bar{x} + \alpha d_k \in X, \quad \text{pour } \alpha > 0 \text{ petit.}$$

Ces deux dernières conclusions contrediraient l'optimalité de $\bar{x}$. Donc $f(x_k) = f(\bar{x})$.

Les conditions (i) et (ii) de l'énoncé se déduisent immédiatement de ce résultat. $\square$

Ce résultat montre que pour minimiser une fonction f strictement convexe sur X , il y a deux choses à faire : identifier $I^0(\bar{x})$, c'est-à-dire trouver les indices des contraintes actives en la solution, et minimiser f sur l'espace actif correspondant.

Donc, si f est strictement convexe et le problème (P) a une solution, on pourrait s'y prendre de la manière suivante. On calcule la solution de (P_{T_k}) pour toutes les combinaisons possibles d'ensemble actif (en nombre fini) — on écarte les problèmes sans solution. Parmi ces solutions, on ne retient que celles qui sont dans X . Parmi ces dernières, celle qui donne la plus petite valeur de f est la solution de (12.15). Cependant, *cette manière de procéder est à proscrire*, du fait du nombre exponentiel de problèmes (P_{T_k}) à résoudre (il y en a 2^m si on résout aussi ceux qui ont plus de n contraintes d'égalité, ce qui n'est pas inutile car la matrice extraite A_{T_k} peut être de rang inférieur à n).

Les méthodes différentiation automatique gardent de cette démarche l'idée de résoudre un nombre fini (et restreint) de problèmes (P_{T_k}) , en les sélectionnant de manière qu'on espère appropriée. À chaque itération, on dispose d'un ensemble de travail $T_k \subseteq I_k$ servant à identifier $I^0(\bar{x})$. Celui-ci est mis à jour après une phase de minimisation (éventuellement partielle) de f sur l'espace actif E_{T_k} correspondant. Voici précisément comment ces algorithmes s'y prennent.

Schéma 12.8 (activation de contraintes)

1. Initialisation : choisir $x_1 \in X$ ($Ax_1 \leq b$) et $T_1 \subseteq I_1$;
 2. Pour $k = 1, 2, \dots$ faire :
 - 2.1. Si x_k est un point stationnaire de (P_{T_k}) et $(\lambda_k)_{T_k}$ est un multiplicateur associé, alors :
 - 2.1.1. Si $(\lambda_k)_{T_k} \geq 0$, x_k est un point stationnaire de (P) et on s'arrête;
 - 2.1.2. Sinon on désactive des contraintes correspondant à des $(\lambda_k)_{(j)} < 0$, $j \in T_k$ (mise à jour de T_k); On prend $x_{k+1} := x_k$; On passe à l'itération suivante;
 - 2.2. Trouver une direction de descente admissible d_k pour (P_{T_k}) en x_k ;
 - 2.3. Par une règle de recherche linéaire «convenable», trouver un pas $\alpha_k \geq 0$ tel que $x_{k+1} := x_k + \alpha_k d_k \in X$;
 - 2.4. Choisir T_{k+1} tel que $T_k \subseteq T_{k+1} \subseteq I_{k+1}$; Si $\alpha_k = 0$ en 2.3, alors il faut que pour un indice $l \in I_k \setminus T_k$ d'une contrainte sur laquelle on bute en x_k dans la direction d_k , on ait $l \in T_{k+1}$.
-

Cet algorithme demande quelques explications, notamment en ce qui concerne les affirmations qui y sont faites. Nous allons les donner.

Avant cela, remarquons que tout n'est pas spécifié dans cet algorithme : à l'étape 2.1.2, on ne dit pas si on désactive une ou toutes les contraintes à multiplicateur strictement négatif; à l'étape 2.2, on ne précise pas comment on doit déterminer la direction d_k ; à l'étape 2.3, la règle de recherche linéaire n'est pas très explicite; enfin, à l'étape 2.4, un certain choix est laissé à l'utilisateur quant à la détermination de T_{k+1} . Il ne s'agit donc que d'un schéma d'algorithme qu'on adaptera à chaque situation. Remarquons également qu'on ne désactive des contraintes qu'à l'étape 2.1.2, lorsque x_k est solution de (P_{T_k}) , et qu'on n'active des contraintes qu'à l'étape 2.4.

L'affirmation de l'étape 2.1.1 est justifiée parce que si x_k est point stationnaire de (P_{T_k}) , on a (12.18). On voit alors qu'en prenant $\bar{x} = x_k$ et $\bar{\lambda} \in \mathbb{R}^m$ avec

$$\bar{\lambda}_{(i)} = \begin{cases} (\lambda_k)_{(i)} & \text{si } i \in T_k \\ 0 & \text{sinon,} \end{cases}$$

le couple $(\bar{x}, \bar{\lambda})$ vérifie (12.16). Pour avoir un minimum de (P) , il suffirait que x_k soit un minimum de (P_{T_k}) avec $\nabla^2 f(x_k)$ défini positif sur E_{T_k} et que les composantes de $(\lambda_k)_{T_k}$ soient strictement positives (utiliser directement le théorème ?? et la remarque ??). Ce n'est pas ce que l'algorithme teste; il ne pourra donc trouver que des points stationnaires.

À l'étape 2.1.2, l'intérêt de désactiver des contraintes correspondant à des multiplicateurs strictement négatifs est de pouvoir, *en général*, faire décroître f à l'itération suivante. En effet, supposons que l'on désactive les contraintes d'indices $j_1 \in T_k, \dots, j_p \in T_k$, correspondant à des multiplicateurs $(\lambda_k)_{(j_1)}, \dots, (\lambda_k)_{(j_p)}$ strictement négatifs. Supposons également que A_{T_k} soit surjective; elle a donc un inverse à droite que

l'on note $A_{T_k}^-$. Alors, en prenant des coefficients positifs $\alpha_1, \dots, \alpha_p$, non tous nuls, la direction

$$d = - \sum_{i=1}^p \alpha_i A_{T_k}^- e_{T_k}^{j_i}$$

(e^j est le j -ième vecteur de base de $\mathbb{R}^m$ et $e_{T_k}^j$ est le vecteur extrait) vérifie

$$A_{T_k} d = - \sum_{i=1}^p \alpha_i e_{T_k}^{j_i}. \quad (12.19)$$

En utilisant l'équation d'optimalité de x_k , on a

$$f'(x_k) \cdot d = -(\lambda_k)_{T_k}^\top A_{T_k} d = \sum_{i=1}^p \alpha_i (\lambda_k)_{(j_i)} < 0.$$

Ceci montre que d est une direction de descente de f en $x_k = x_{k+1}$. On déduit aussi de (12.19) que $A_{(j)} d = 0$, pour $j \in T_k \setminus \{j_1, \dots, j_p\} = T_{k+1}$ et donc que d est admissible pour $(P_{T_{k+1}})$. Mais bien que l'on ait $A_{(j_i)} d = -\alpha_i \leq 0$, pour $i = 1, \dots, p$, il se peut que d ne soit pas admissible pour (P) , parce qu'il peut y avoir un indice $j \in I_k \setminus T_k$ pour lequel $A_{(j)} d > 0$. On aurait alors un pas nul à l'étape 2.3 de l'itération suivante avec un danger de *cyclage* (on retourne périodiquement au même ensemble de travail sans bouger en x). Comme notre discussion l'indique, ceci ne se produira pas si $\mathcal{R}(A_{T_k}^\top) = \mathcal{R}(A_{I_k}^\top)$ et si A_{T_k} est surjective. En particulier, si A_{I_k} est surjective, il suffira de prendre $T_k = I_k$. On est donc conduit à la définition suivante.

Définition 12.9 On dit que les contraintes $Ax \leq b$ sont *non dégénérées* si $A_{I^0(x)}$ est surjective pour tout x vérifiant $Ax \leq b$. $\square$

Comme à l'étape 2.3, x_k n'est pas stationnaire pour (P_{T_k}) , on peut trouver une direction de descente admissible pour ce problème.

Il sera utile de calculer le *pas d'activation* $\hat{\alpha}_k$, c'est-à-dire le pas α maximal laissant $x_k + \alpha d_k$ dans X . On doit avoir $\hat{\alpha}_k$ maximal tel que

$$A_{(j)}(x_k + \hat{\alpha}_k d_k) \leq b_{(j)}, \quad \forall j.$$

Comme $Ax_k \leq b$, il suffit de considérer les indices j tels que $A_{(j)} d_k > 0$. On voit que

$$\hat{\alpha}_k = \min_{j: A_{(j)} d_k > 0} \frac{b_{(j)} - A_{(j)} x_k}{A_{(j)} d_k}. \quad (12.20)$$

Par convention, $\hat{\alpha}_k = +\infty$ si $A_{(j)} d_k \leq 0$ pour tout j .

L'algorithme différentiation automatique s'utilise avec plus ou moins de succès pour la minimisation de fonctions régulières *quelconques* en présence de contraintes linéaires, avec une légère modification toutefois. Lorsque la fonction n'est pas quadratique, calculer une solution exacte de (P_{T_k}) est une tâche longue et coûteuse. On se contente alors de tester à l'étape 2.1 si x_k est « presque » stationnaire et si c'est le cas, on désactive des contraintes correspondant à des multiplicateurs négatifs.

Comme on le comprend, la nécessité de devoir désactiver des contraintes afin d'identifier l'espace actif au point optimal – ce qui ne peut se faire en toute sécurité qu'en des points stationnaires d'un problème (P_{T_k}) – fait que la méthode différentiation automatique convient préférentiellement à la minimisation de fonctions quadratiques :

$$f(x) = \frac{1}{2}x^\top Qx + q^\top x, \quad (12.21)$$

où Q est une matrice symétrique d'ordre n et $q \in \mathbb{R}^n$. Dans la suite nous allons voir comment préciser l'algorithme pour qu'il soit convergent dans les deux cas suivants :

- Q est définie positive (optimisation quadratique strictement convexe),
- Q est semi-définie positive (optimisation quadratique convexe).

Le second cas inclut le premier, mais l'introduction progressive des solutions algorithmiques aidera à la compréhension des méthodes. Lorsque Q est indéfinie le problème peut avoir de nombreuses solutions locales et la recherche d'un minimum local est plus difficile; quant à la recherche du minimum global, c'est un problème très difficile (NP-complet, dans la terminologie de la théorie de la complexité).

Notes

La méthode du gradient projeté a été à l'origine proposée par Goldstein [262 ; 1964] ainsi que Levitin et Polyak [391 ; 1966] pour minimiser une fonction différentiable sur un convexe fermé non vide de $\mathbb{R}^n$. Le théorème 12.4 est repris de ces articles. Le théorème 12.5 est repris de Bertsekas [50 ; 1976].

Pour l'algorithme du gradient projeté, voir Calamai et Moré [104 ; 1987].

Exercices

- 12.1.** *Projection sur un pavé.* Pour $l \in (\mathbb{R} \cup \{-\infty\})^n$ et $u \in (\mathbb{R} \cup \{+\infty\})^n$ tels que $l \leq u$, on définit le pavé $[l, u] := \{x \in \mathbb{R}^n : l \leq x \leq u\}$. Le projeté $\bar{x} = P_{[l, u]}(x)$ de $x \in \mathbb{R}^n$ sur $[l, u]$, pour le produit scalaire euclidien, est donnée par

$$\bar{x}_i = \begin{cases} l_i & \text{si } x_i < l_i \\ x_i & \text{si } l_i \leq x_i \leq u_i \\ u_i & \text{si } u_i < x_i. \end{cases}$$

En particulier, le projeté orthogonal de $x \in \mathbb{R}^n$ sur l'orthant positif $\mathbb{R}_+^n$ est $x^+ := \max(0, x)$ (le maximum est pris composante par composante).

- 12.2.** *Gradient projeté.* Soient X une partie convexe d'un espace euclidien $\mathbb{E}$ et $f : \mathbb{E} \rightarrow \mathbb{R}$ une fonction dérivable en $x \in X$. Le gradient projeté de f en x est la projection orthogonale du gradient $\nabla f(x)$ de f en x sur $-T_x X$. On le note

$$g^P(x) := P_{-T_x X} \nabla f(x).$$

1) On a aussi

$$g^P(x) := -P_{T_x X}(-\nabla f(x)).$$

- 2) La condition d'optimalité du premier ordre $\nabla f(x) \in (T_x X)^+$ (théorème 4.6) est équivalente à l'équation $g^P(x) = 0$.
- 3) S'il est non nul, l'opposé du gradient projeté est une direction de descente de f en x , car on a

$$\langle \nabla f(x), -g^P(x) \rangle \leq -\|g^P(x)\|^2.$$

- 4) On simplifie les notations en introduisant $g := \nabla f(x)$ et $g^P := g^P(x)$. Alors, quel que soit $\alpha \geq 0$, on a

$$P_X(x - \alpha g) = x - \alpha g^P \iff x - \alpha g^P \in X, \quad (12.22)$$

autrement dit, la projection du chemin émanant de x et porté par l'opposé du gradient se confond avec le chemin émanant de x et porté par l'opposé du gradient projeté, tant que ce dernier chemin est dans X . Par ailleurs, lorsque X est polyédrique, les propriétés dans (12.22) ont lieu pour tout $\alpha \geq 0$ petit.

Supposons maintenant que $\mathbb{E} = \mathbb{R}^n$, que l'on munit du produit scalaire euclidien. Pour $l \in (\mathbb{R} \cup \{-\infty\})^n$ et $u \in (\mathbb{R} \cup \{+\infty\})^n$ tels que $l < u$, on définit le pavé $[l, u] := \{x \in \mathbb{R}^n : l \leq x \leq u\}$ et on considère un point $x \in [l, u]$.

- 5) $T_x[l, u] = \{d \in \mathbb{R}^n : d_i \geq 0 \text{ si } x_i = l_i, \ d_i \leq 0 \text{ si } x_i = u_i\}$.
- 6) Pour tout $i \in [1 : n]$, on a

$$[g^P(x)]_i = \begin{cases} \min(0, \frac{\partial f}{\partial x_i}(x)) & \text{si } l_i = x_i \\ \frac{\partial f}{\partial x_i}(x) & \text{si } l_i < x_i < u_i \\ \max(0, \frac{\partial f}{\partial x_i}(x)) & \text{si } x_i = u_i. \end{cases}$$

- 7) Si x minimise f sur $[l, u]$, alors

$$\frac{\partial f}{\partial x_i}(x) \begin{cases} \geq 0 & \text{si } l_i = x_i \\ = 0 & \text{si } l_i < x_i < u_i \\ \leq 0 & \text{si } x_i = u_i. \end{cases}$$

12.3. Projection sur l'intersection d'un pavé et d'une contrainte affine.

12.4. Projection sur $\mathbb{R}_{\leq}^n$.

13 Pénalisation

Cette théorie et ces phénomènes nous montrent comment on peut amener autrui à modifier ses comportements sans recourir à l'autorité, ni même à quelque stratégie persuasive, mais par des moyens détournés.

J.-L. BEAUVOIS et R.-V. JOULE (1987). Petit traité de manipulation à l'usage des honnêtes gens [39].

13.1 Vue d'ensemble

La pénalisation est un concept simple qui permet de transformer un problème d'optimisation avec contraintes en *un* problème ou en *une suite* de problèmes d'optimisation sans contrainte ou avec des contraintes simples. C'est un concept qui a une utilité à la fois théorique et numérique. Ce que le qualificatif « simple » signifie dépendra du contexte, des questions que l'on se pose, de la disponibilité d'algorithmes de résolution.

En analyse, l'approche par pénalisation est parfois utilisée pour étudier un problème d'optimisation dont certaines contraintes sont difficiles à prendre en compte, alors que le problème pénalisant ces contraintes difficiles a des propriétés (l'existence de solution par exemple) mieux comprises ou plus simples à mettre en évidence. Si l'on a de la chance ou si la pénalisation est bien choisie, des *passages à la limite* parfois délicats permettent d'obtenir des propriétés du problème original. Par exemple, on peut obtenir des conditions d'optimalité d'un problème avec contraintes, à partir des conditions d'optimalité des problèmes pénalisés ([207, 99], proposition 13.11). D'autre part, comme nous allons le souligner ci-dessous, la pénalisation est un outil permettant d'étudier les problèmes d'optimisation avec et sans contrainte dans un même formalisme.

D'un point de vue numérique, cette transformation en problèmes sans contrainte (ou avec contraintes simples) permet d'utiliser des algorithmes d'optimisation sans contrainte (ou avec ces contraintes simples) pour obtenir la solution de problèmes dont l'ensemble admissible peut avoir une structure complexe. Cela semble merveilleux, inespéré, de voir que l'on puisse ainsi utiliser des algorithmes qui ne cherchent qu'à minimiser une fonction pour trouver des points qui, en plus d'être optimaux, sont admissibles. Cette approche est de ce fait très souvent utilisée. Elle permet d'obtenir une solution de qualité suffisante rapidement sans avoir à entrer dans l'algorithmique

sophistiquée de l'optimisation avec contraintes. Ce n'est cependant pas une technique universelle, car elle a ses propres inconvénients : non-différentiabilité, nécessité de minimiser une *suite* de fonctions, parfois de plus en plus mal conditionnées, paramétrage délicat. C'est cette approche qui est suivie, avec un raffinement remarquable, dans les méthodes de points intérieurs (chapitres 18, 18 et 20), conduisant ainsi à des algorithmes polynomiaux pour de grandes familles de problèmes d'optimisation.

La pénalisation peut intervenir dans un contexte très général. On désigne par $\mathbb{E}$ un ensemble arbitraire, par X une partie de $\mathbb{E}$ et par $f : \mathbb{E} \rightarrow \mathbb{R}$ une fonction. On considère le problème de minimiser f sur X . L'ensemble admissible X pourra être l'intersection de deux ensembles :

$$X := X_r \cap X_s.$$

L'ensemble X_r est celui défini par des contraintes qui seront *relaxées* ou *relâchées* par la pénalisation. Théoriquement, elles ne seront vérifiées que lorsqu'un paramètre de pénalisation atteindra sa limite ; d'un point de vue numérique, elles ne seront vérifiées qu'à la convergence des algorithmes. L'ensemble X_s est celui des contraintes *simples* ou *strictes*, qui seront maintenues sans être relaxées par la pénalisation. Formalisons cela. Le problème considéré s'écrit donc

$$(P_X) \quad \inf_{x \in X_r \cap X_s} f(x)$$

et les différentes techniques de pénalisation que nous verrons consistent souvent à transformer ce problème par un ou des problème(s) de la forme

$$(P_r) \quad \inf_{x \in X_s} \Theta_r(x),$$

où $\Theta_r(x)$ est obtenu en ajoutant à $f(x)$ le terme $rp(x)$:

$$\Theta_r := f + rp. \quad (13.1)$$

Ici, r est un scalaire strictement positif, appelé *facteur de pénalisation*, et $p : \mathbb{E} \rightarrow \mathbb{R}$ est une fonction, dénommée *fonction pénalisante*. La locution *fonction de pénalisation* sera réservée à la fonction que l'on minimise dans le problème de pénalisation, qui est Θ_r dans (P_r) . Le but de cette fonction pénalisante est de pénaliser la violation des contraintes (on parle alors de *pénalisation extérieure*, section 13.2) ou l'abord de la frontière de l'ensemble admissible (on parle dans ce cas de *pénalisation intérieure*, section 13.3). Parfois, au lieu de pénaliser le critère f , on pénalise le lagrangien du problème (c'est le cas de l'approche du lagrangien augmenté, section 13.4) ; cette technique a des avantages, mais aussi des inconvénients. Numériquement, l'intérêt de (P_r) est de pouvoir être résolu par une méthode d'optimisation sans contrainte (lorsque $X_s = \mathbb{E}$) ou avec contraintes simples ; celles définissant X_s .

La transformation du problème avec contraintes (P_X) en problème(s) (P_r) sans contrainte ou avec contraintes simples soulève deux questions : Est-ce possible ? Quel en est le prix ? La première question renvoie à celle de savoir si en résolvant (P_r) on résout (P_X) et, de manière plus précise, à celle de la détermination du lien entre les ensembles de solutions de (P_X) et (P_r) et du lien entre les valeurs optimales. La seconde question concerne l'efficacité numérique d'une telle transformation. La

réponse à ces deux questions va dépendre du choix de la fonction pénalisante p et du facteur de pénalisation r .

Par exemple, on pourrait choisir la fonction pénalisante p égale à la fonction indicatrice de X_r , $p = \mathcal{I}_{X_r}$, c'est-à-dire

$$p(x) = \begin{cases} 0 & \text{si } x \in X_r \\ +\infty & \text{si } x \notin X_r. \end{cases}$$

Il est clair que dans ce cas, les problèmes (P_X) et (P_r) sont identiques : ils ont les mêmes ensembles de solutions et la même valeur optimale. Cette fonction pénalisante est parfois utilisée dans la théorie, car elle permet de traiter en même temps les problèmes avec ou sans contrainte (voir chapitre 14). Ce choix de p n'est cependant pas très utile numériquement car les méthodes classiques d'optimisation ne peuvent pas être utilisées sur des fonctions qui prennent la valeur $+\infty$ dans des régions visitées par les itérés (dans quelle direction se déplacer pour faire décroître Θ_r si l'itéré courant n'est pas dans X_r ?). Nous allons donc, dans ce chapitre, introduire diverses fonctions pénalisantes p , autres que l'indicatrice de X_r , et en étudier les propriétés théoriques et algorithmiques.

La première question posée ci-dessus conduit à la notion de pénalisation exacte, à laquelle on fera souvent allusion dans ce chapitre.

Définition 13.1 (pénalisation exacte) On dit qu'une fonction de pénalisation Θ_r associée au problème (P_X) est *exacte* si toute « solution » de (P_X) est « solution » de (P_r) ; on dit qu'elle est *inexacte* dans le cas contraire, c'est-à-dire qu'il y a des « solutions » de (P_X) qui ne sont pas « solution » de (P_r) . $\square$

Dans cette définition, le terme « solution » est pris dans un sens ambigu et il faudra chaque fois préciser si l'on veut parler de point stationnaire, de minimum local ou de minimum global.

La structure du problème (P_r) , dont le critère est la somme pondérée de deux fonctions, permet d'énoncer d'emblée une propriété très générale sur le comportement de chaque terme *en un minimum global* $\bar{x}_r$ de (P_r) , lorsque le facteur de pénalisation r varie. Intuitivement, si r augmente, on attache moins d'importance à f et plus d'importance à p , si bien qu'il semble normal que $f(\bar{x}_r)$ croisse et que $p(\bar{x}_r)$ décroisse. La proposition suivante énonce cela de façon rigoureuse. Le résultat est très général puisqu'il ne requiert aucune hypothèse sur l'ensemble X_s , ni a fortiori d'hypothèse de convexité ou de différentiabilité; seule la structure du critère Θ_r intervient.

Proposition 13.2 (monotonie en pénalisation) Soient X_s un ensemble non vide, f et $p : X_s \rightarrow \mathbb{R}$ deux fonctions, $r \in \mathbb{R}$ et $\Theta_r := f + rp$. Si $r_1 < r_2$ sont deux réels et si $\bar{x}_{r_i} \in \arg \min\{\Theta_{r_i}(x) : x \in X_s\}$ ($i = 1, 2$), alors

- 1) $p(\bar{x}_{r_1}) \geq p(\bar{x}_{r_2})$,
- 2) $f(\bar{x}_{r_1}) \leq f(\bar{x}_{r_2})$ si $r_1 \geq 0$,
- 3) $\Theta_{r_1}(\bar{x}_{r_1}) \leq \Theta_{r_2}(\bar{x}_{r_2})$ si $p(\bar{x}_{r_2}) \geq 0$.

DÉMONSTRATION. 1) En exprimant que $\bar{x}_{r_i}$ minimise Θ_{r_i} sur X_s , on obtient :

$$\begin{aligned} f(\bar{x}_{r_1}) + r_1 p(\bar{x}_{r_1}) &\leq f(\bar{x}_{r_2}) + r_1 p(\bar{x}_{r_2}) \\ f(\bar{x}_{r_2}) + r_2 p(\bar{x}_{r_2}) &\leq f(\bar{x}_{r_1}) + r_2 p(\bar{x}_{r_1}). \end{aligned}$$

En sommant, on trouve $(r_2 - r_1)p(\bar{x}_{r_2}) \leq (r_2 - r_1)p(\bar{x}_{r_1})$. Alors $r_2 > r_1$ implique alors que $p(\bar{x}_{r_2}) \leq p(\bar{x}_{r_1})$.

2) En tenant compte du point 1 et du fait que $r_1 \geq 0$, la première inégalité exposée ci-dessus permet d'écrire

$$f(\bar{x}_{r_1}) + r_1 p(\bar{x}_{r_1}) \leq f(\bar{x}_{r_2}) + r_1 p(\bar{x}_{r_1}),$$

si bien que $f(\bar{x}_{r_1}) \leq f(\bar{x}_{r_2})$.

3) On a $\Theta_{r_1}(\bar{x}_{r_1}) \leq \Theta_{r_1}(\bar{x}_{r_2})$ par l'optimalité de $\bar{x}_{r_1}$; puis $\Theta_{r_1}(\bar{x}_{r_2}) \leq \Theta_{r_2}(\bar{x}_{r_2})$ par la positivité de $p(\bar{x}_{r_2})$. $\square$

Le même raisonnement montre que l'on a une croissance ou décroissance *stricte* des suites si $\bar{x}_r$ est l'*unique* minimum de (P_r) et si $\bar{x}_r$ change avec r .

Un second résultat général, ne dépendant principalement que de la structure de Θ_r , concerne les points d'adhérence des minimiseurs $\bar{x}_r$ des fonctions de pénalisation Θ_r lorsque $r \downarrow 0$. Dans ce cas, la fonction pénalisante p agit de moins en moins et il est naturel de se demander si les minimiseurs $\bar{x}_r$ de Θ_r ne convergeraient pas vers un minimiseur de f sur X_s , donc sans tenir compte de la contrainte d'appartenance à X_r , à savoir un élément de $S := \arg \min\{f(x) : x \in X_s\}$. La proposition suivante donne des conditions pour qu'il en soit ainsi. Puisque'il est question de convergence des $\bar{x}_r$, on a besoin cette fois d'une topologie sur $\mathbb{E}$.

Proposition 13.3 (point d'adhérence lorsque $r \downarrow 0$) *Soit $\mathbb{E}$ un espace topologique. Supposons que X_s soit fermé dans $\mathbb{E}$, que f et $p : X_s \rightarrow \mathbb{R}$ soient s.c.i., que $S := \arg \min\{f(x) : x \in X_s\}$ soit non vide et que, pour une suite de $r \downarrow 0$, (P_r) ait au moins une solution, notée $\bar{x}_r$. Alors tout point d'adhérence de $\{\bar{x}_r\}_{r \downarrow 0}$ est solution de*

$$\inf_{x \in S} p(x).$$

DÉMONSTRATION. Soient $\hat{x} \in S \neq \emptyset$ et $\bar{x}$ un point d'adhérence de $\{\bar{x}_r\}_{r \downarrow 0}$ (pour une sous-suite de $r \downarrow 0$). Il suffit de montrer que $\bar{x} \in S$ et que $p(\bar{x}) \leq p(\hat{x})$.

[Montrons que $\bar{x} \in S$] Pour cela, on observe que $\bar{x} \in X_s$, parce que X_s est fermé, que $\bar{x}_r \in X_s$ et que $\bar{x}_r \rightarrow \bar{x}$. Il reste à montrer que, pour un point $x \in X_s$ arbitraire, on a $f(\bar{x}) \leq f(x)$. L'optimalité de $\bar{x}_r$ permet d'écrire

$$\forall x \in X_s : f(\bar{x}_r) + r p(\bar{x}_r) \leq f(x) + r p(x). \quad (13.2)$$

On passe à la limite inférieure lorsque $r \downarrow 0$ dans cette inégalité :

$$\begin{aligned}
f(\bar{x}) &\leq \liminf_{r \downarrow 0} f(\bar{x}_r) \quad [f \text{ est s.c.i. et } \bar{x}_r \rightarrow \bar{x}] \\
&\leq \liminf_{r \downarrow 0} f(\bar{x}_r) + \liminf_{r \downarrow 0} rp(\bar{x}_r) \quad [(A.12), p \text{ est s.c.i. et } p(\bar{x}) \in \mathbb{R}] \\
&\leq \liminf_{r \downarrow 0} (f(\bar{x}_r) + rp(\bar{x}_r)) \quad [(A.2a)] \\
&\leq \liminf_{r \downarrow 0} (f(x) + rp(x)) \quad [(13.2)] \\
&= f(x).
\end{aligned}$$

[Montrons que $p(\bar{x}) \leq p(\hat{x})$] Observons d'abord que $f(\bar{x}_r) \in \mathbb{R}$ (car f est à valeurs réelles sur X_s et $\bar{x} \in S \subseteq X_s$). Ensuite, comme $\hat{x} \in S$, on a $f(\hat{x}) \leq f(\bar{x}_r)$, si bien que (13.2) en $x = \hat{x} \in X_s$, $f(\bar{x}_r) \in \mathbb{R}$ et $r > 0$ conduisent à

$$p(\bar{x}_r) \leq p(\hat{x}).$$

En prenant la limite inférieure lorsque $r \downarrow 0$, on trouve que $p(\bar{x}) \leq p(\hat{x})$. $\square$

Nous serons souvent amenés à considérer le problème d'optimisation sous contraintes fonctionnelles suivant :

$$(P_{EI,X}) \quad \begin{cases} \min f(x) \\ c_i(x) = 0, & i \in E \\ c_i(x) \leq 0, & i \in I \\ x \in X_s, \end{cases} \quad (13.3)$$

où f et les c_i sont des fonctions définies sur un espace euclidien $\mathbb{E}$ (produit scalaire noté $\langle \cdot, \cdot \rangle$) à valeurs dans $\mathbb{R}$, (E, I) forme une partition de $[1 : m]$ et X_s est un fermé de $\mathbb{E}$. Le nombre de contraintes d'égalité et d'inégalité se note m_E et m_I , et leur somme $m = m_E + m_I$. Le lagrangien de ce problème est l'application $\ell : \mathbb{E} \times \mathbb{R}^m \mapsto \mathbb{R}$ définie en $(x, \lambda) \in \mathbb{E} \times \mathbb{R}^m$ par

$$\ell(x, \lambda) := f(x) + \lambda^\top c(x).$$

Si $v \in \mathbb{R}^m$, on note $v^\# \in \mathbb{R}^m$ le vecteur défini par

$$(v^\#)_i = \begin{cases} v_i & \text{si } i \in E \\ v_i^+ & \text{si } i \in I. \end{cases} \quad (13.4)$$

Les contraintes de (P_{EI,X_s}) s'écrivent alors simplement $c(x)^\# = 0$ (attention : l'application $x \mapsto c(x)^\#$ n'est pas différentiable en général ; on n'a donc fait que remplacer la difficulté liée à la présence de contraintes d'inégalité par une autre !). Nous renvoyons le lecteur à la section 4.4 pour d'autres notations.

13.2 Pénalisation extérieure

13.2.1 Définition et exemples

Commençons l'exposé par un exemple, qui résume assez bien le fonctionnement de la pénalisation extérieure.

Exemple 13.4 On considère le problème à une variable et une contrainte suivant

$$\begin{cases} \inf 1 - x - \frac{1}{3}x^3 \\ x \leq 0, \end{cases} \quad (13.5)$$

auquel on associe la fonction de pénalisation suivante

$$\Theta_r(x) = 1 - x - \frac{1}{3}x^3 + r(x^+)^2.$$

où $x^+ := \max(0, x)$. Cette pénalisation est dite *quadratique*, à cause de l'exposant 2 sur x^+ (section 13.2.4). L'effet de cette pénalisation peut s'observer à la figure 13.1, dont l'abscisse est $c(x) = x$. On voit que le terme $r(x^+)^2$ ne joue un rôle qu'à l'extérieur

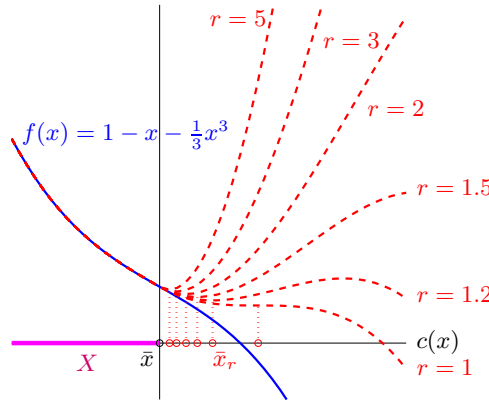


Fig. 13.1. Pénalisation quadratique du problème (13.5), avec $r = 1, 1.2, 1.5, 2, 3$ et 5 .

de l'ensemble admissible $\mathbb{R}_-$. C'est la raison pour laquelle on qualifie cette pénalisation d'*extérieure*. D'autre part, on observe que le minimiseur *local* de Θ_r (il n'existe ici que si $r > 1$) est extérieur à l'ensemble admissible. Plus r est grand, plus le minimiseur se rapproche de la solution du problème, qui est ici $\bar{x} = 0$, et la suite de ces minimiseurs converge vers cette solution lorsque r tend vers l'infini. Cependant, plus r est grand, plus le minimum local est accentué (la dérivée seconde de Θ_r si elle existait serait élevée), ce qui pourra être une source de difficultés numériques. $\square$

Dans la section 13.2.2, nous montrerons de manière rigoureuse que ce que nous venons d'observer sur cet exemple simple se produit pour une grande classe de fonctions de pénalisation. Cet exemple est en effet représentatif d'une technique de pénalisation appelée *pénalisation extérieure*, que l'on peut formaliser. Dans celle-ci, la fonction de pénalisation est de la forme $\Theta_r = f + r p$, avec une fonction pénalisante p vérifiant les propriétés suivantes :

$$p \text{ est s.c.i. sur } \mathbb{E}, \quad (13.6a)$$

$$\forall x \in \mathbb{E}: p(x) \geq 0, \quad (13.6b)$$

$$p(x) = 0 \iff x \in X. \quad (13.6c)$$

Observons que ces conditions impliquent que X est un ensemble de sous-niveau de la fonction s.c.i. p , si bien qu'il est nécessairement fermé. Le qualificatif « *extérieure* » vient de la propriété (13.6c), qui exprime que Θ_r ne modifie f qu'à l'extérieur de l'ensemble admissible.

Le tableau 13.1 donne des exemples de fonctions pénalisantes satisfaisant (13.6),

Contraintes définissant X	Fonction pénalisante	Nom de la pénalisation	Références
$c(x) = 0$	$p(x) = \ c(x)\ _2^2$	quadratique	[143, 144]
$c(x) \leq 0$	$p(x) = \ c(x)^+\ _2^2$	quadratique	[2]

Tableau 13.1. Exemples de pénalisation extérieure.

si l'on suppose que c a une propriété de continuité adéquate, conduisant à ce que l'on appelle la *pénalisation quadratique* (section 13.2.4). Pour $u \in \mathbb{R}^p$, on y a noté u^+ le vecteur de $\mathbb{R}^p$ dont la i -ième composante vaut $\max(u_i, 0)$. La norme ℓ_2 y est notée $\|\cdot\|_2$. Les fonctions pénalisantes p du tableau 13.1 font de $\Theta_r = f + r p$ une fonction de pénalisation inexacte, puisque l'on trouve pour toute solution $\bar{x}$ de (P_X) :

$$\nabla \Theta_r(\bar{x}) = \nabla f(\bar{x})$$

et que rien n'impose, dans les conditions d'optimalité, que ce vecteur soit nul. Donc $\bar{x}$ n'est généralement pas solution de (P_r) .

13.2.2 Propriétés

Nous étudions dans cette section les propriétés asymptotiques des solutions (minima globaux) $\bar{x}_r$ de (P_r) , lorsque r tend vers l'infini. La proposition 13.5 énonce des conditions pour que les points d'adhérence de $\{\bar{x}_r\}$ soient solutions du problème original (P_X) . La proposition 13.6 montre que, si l'une des fonctions de pénalisation Θ_r est coercive, l'existence d'une suite bornée $\{\bar{x}_r\}$ de minimiseurs est assurée; comme cette suite a alors des points d'adhérence, le résultat de la proposition précédente 13.5 s'y applique.

Proposition 13.5 (optimalité asymptotique) *Soient X un fermé non vide, f une fonction *semi-continue inférieurement* et $p : \mathbb{E} \rightarrow \mathbb{R}$ une fonction vérifiant (13.6). Alors, tout point d'adhérence de la suite $\{\bar{x}_r\}_{r \uparrow \infty}$ est solution de (P_X) .*

DÉMONSTRATION. Soit $\bar{x}$ un point d'adhérence de $\{\bar{x}_r\}_{r \uparrow \infty}$ et $x_{r_i} \rightarrow \bar{x}$ pour une sous-suite de facteurs de pénalisation $\{r_i\}_{i \geq 0} \rightarrow \infty$. Observons que, pour i fixé,

$$\forall x \in X : \quad \Theta_{r_i}(\bar{x}_{r_i}) \leq \Theta_{r_i}(x) = f(x), \quad (13.7)$$

où la première inégalité provient de l'optimalité de $\bar{x}_{r_i}$ et la seconde de l'admissibilité de x et de (13.6c).

Montrons d'abord que $\bar{x} \in X$ ou encore, par (13.6c), que $p(\bar{x}) = 0$. Comme f est s.c.i., on peut trouver un indice i_0 tel que

$$\forall i \geq i_0 : f(\bar{x}) - 1 \leq f(\bar{x}_{r_i}).$$

En ajoutant $r_i p(\bar{x}_{r_i})$ aux deux membres de cette inégalité et en utilisant (13.7), on obtient

$$\forall x \in X \text{ et } i \geq i_0 : f(\bar{x}) + r_i p(\bar{x}_{r_i}) - 1 \leq \Theta_{r_i}(\bar{x}_{r_i}) \leq f(x).$$

Dès lors,

$$\forall x \in X \text{ et } i \geq i_0 : p(x_{r_i}) \leq \frac{f(x) - f(\bar{x}) + 1}{r_i}.$$

En fixant $x \in X$ et en prenant la limite inférieure quand $i \rightarrow \infty$, on trouve par (13.6a) et (13.6b) que $0 \leq p(\bar{x}) \leq \liminf_{i \rightarrow \infty} p(x_{r_i}) \leq 0$. Donc $p(\bar{x}) = 0$.

Montrons maintenant l'optimalité de $\bar{x}$. En minorant $\Theta_{r_i}(\bar{x}_{r_i})$ par $f(\bar{x}_{r_i})$ dans (13.7), on obtient

$$\forall x \in X : f(\bar{x}_{r_i}) \leq f(x).$$

Comme f est s.c.i., en prenant la limite inférieure, on a $f(\bar{x}) \leq \liminf f(x_{r_i}) \leq f(x)$. Comme x est arbitraire dans X et $\bar{x} \in X$, $\bar{x}$ minimise f sur X . $\square$

La proposition suivante suppose la coercivité de Θ_{r_0} pour un certain $r_0 \geq 0$, ce qui permet d'assurer l'existence de minimiseurs globaux des Θ_k lorsque $r \geq r_0$, ainsi que

Proposition 13.6 (existence d'une suite de minimiseurs) Soient X un fermé non vide, f une fonction *semi-continue inférieurement* et $p : \mathbb{E} \rightarrow \mathbb{R}$ une fonction vérifiant (13.6). Supposons aussi qu'il existe un $r_0 \geq 0$ tel que Θ_{r_0} soit *coercive*. Alors,

- 1) $\forall r \geq r_0$, (P_r) a au moins une solution, que l'on note $\bar{x}_r$,
- 2) la suite $\{\bar{x}_r\}_{r \uparrow \infty}$ est bornée,
- 3) tout point d'adhérence de la suite $\{\bar{x}_r\}_{r \uparrow \infty}$ est solution de (P_X) .

DÉMONSTRATION. 1) On voit que, pour $r \geq r_0$, Θ_r est s.c.i. sur $\mathbb{E}$ (exercice A.3) et est coercive. Par conséquent, (P_r) a au moins une solution $\bar{x}_r$ (corollaire 1.4).

2) Pour $x \in X$ et $r \geq r_0$, on a

$$\Theta_{r_0}(\bar{x}_r) \leq \Theta_r(\bar{x}_r) \leq \Theta_r(x) = f(x). \quad (13.8)$$

La première inégalité vient du fait que $(r - r_0)p(\bar{x}_r) \geq 0$ et la seconde de l'optimalité de $\bar{x}_r$. En fixant x dans X , cela montre que $\{\Theta_{r_0}(\bar{x}_r)\}_{r \uparrow \infty}$ est bornée et par la coercivité de Θ_{r_0} , on en déduit que $\{\bar{x}_r\}_{r \uparrow \infty}$ est bornée.

3) C'est une conséquence de la proposition 13.5. $\square$

Les résultats précédents ont une valeur indicative sur le comportement des minima globaux de Θ_r . Malheureusement, le même résultat ne tient plus pour les minima locaux. Par exemple, si

$$c(x) = 2x^3 - 3x^2 + 5 = (x+1)(2x^2 - 5x + 5), \quad (13.9)$$

le problème

$$\begin{cases} \inf 0 \\ c(x) = 0 \end{cases} \quad (13.10)$$

consiste à chercher l'unique racine réelle $\bar{x} = -1$ de c . Mais le problème pénalisé

$$\inf r|c(x)|^2$$

a un minimum local en $\bar{x}_r = 1$ quel que soit $r > 0$. On n'a donc pas la convergence de ces minima locaux vers $\bar{x} = -1$. La proposition 13.11 ci-dessous donne les propriétés des points d'adhérence de la suite des points stationnaires approchés de Θ_r lorsque $r \rightarrow +\infty$. Elle est donc complémentaire de la proposition précédente qui ne s'intéresse qu'aux minima globaux.

Une autre limitation du théorème est de supposer que Θ_{r_0} est bornée inférieurement sur $\mathbb{E}$. Si ce n'est pas le cas, il se peut que (P_X) ait une solution mais que (P_r) n'en ait pas. C'est le cas pour le problème suivant

$$\begin{cases} \inf x^3 \\ x \geq 0 \end{cases}$$

Alors $\Theta_r(x) = x^3 + r(x^-)^2$ n'est pas bornée inférieurement. Dans de pareils cas, on peut rajouter un terme de pénalisation plus fort à l'infini ou introduire des bornes sur les variables.

Remarque 13.7 Pour que la suite de minimiseurs $\{\bar{x}_r\}$ s'approche d'une solution de (P_X) , il faut faire croître le facteur de pénalisation r (proposition 13.5). Dans ces conditions, la proposition 13.2 nous apprend que la suite $\{f(\bar{x}_r)\}$ croît. Si la croissance de cette suite est stricte, les points $\bar{x}_r$ ne peuvent être qu'extérieurs à X , sinon les points d'adhérence de $\{\bar{x}_r\}$ ne pourraient pas être solutions de (P_X) . En effet, on aurait alors des points de X aussi proches que l'on veut d'une solution $\bar{x}$ en lesquels f prendrait une valeur strictement inférieure à $f(\bar{x})$, ce qui contredirait l'optimalité de $\bar{x}$.

Nous verrons à la remarque 13.12, une autre manière d'aboutir à la même conclusion, après avoir montré comment construire une suite convergeant vers un multiplicateur optimal. $\square$

La propriété suivante montre que la pénalisation extérieure permet d'avoir une borne inférieure du coût optimal de (P_X) . Ceci peut être utile pour certaines méthodes de résolution de problèmes d'optimisation avec variables entières, du type *branch-and-bound*. En première analyse, le résultat est un peu magique, car en minimisant la fonction Θ_r qui majore f , on obtient une valeur optimale plus petite que celle de f sur X . La magie s'estombe si l'on se rappelle que (i) la pénalisation n'opère qu'en dehors de l'ensemble admissible, là où le critère peut prendre des valeurs plus basses que $\text{val}(P_X)$ et (ii) la valeur optimale de (P_r) est obtenue par minimisation sur l'espace $\mathbb{E}$ tout entier, alors que celle de (P_X) est le résultat d'une minimisation sur X seulement. La figure 13.1 illustre ces remarques.

Proposition 13.8 (minorant de la valeur optimale)

- 1) Si $p(\cdot) \geq 0$, la suite $\{\text{val}(P_r)\}$ croît avec r .
- 2) Si $p(\cdot) \geq 0$ et $p(X) = \{0\}$, la suite $\{\text{val}(P_r)\}$ est majorée par $\text{val}(P)$.

DÉMONSTRATION. 0) Comme au point 3 de la proposition 13.2 (mais ici, on ne suppose pas l'existence de minimiseurs des Θ_r), lorsque $r_1 < r_2$, on a pour tout $x \in \mathbb{E}$:

$$\text{val}(P_{r_1}) \leq f(x) + r_1 p(x) \leq f(x) + r_2 p(x), \quad (13.11)$$

parce que $r_1 < r_2$ et $p(\cdot) \geq 0$.

1) En prenant l'infimum en $x \in \mathbb{E}$ à droite, on obtient $\text{val}(P_{r_1}) \leq \text{val}(P_{r_2})$, si bien que $\{\text{val}(P_r)\}$ croît avec r .

2) Si à droite dans (13.11), on prend l'infimum en $x \in X$, on obtient $\text{val}(P_{r_1}) \leq \text{val}(P)$, parce que $p(X) = 0$. $\square$

13.2.3 Schéma algorithmique

La proposition 13.5 conduit au schéma algorithmique suivant, qui approche des solutions de (P_X) par des solutions approchées (à ε près) des problèmes pénalisés (P_r) , avec $\varepsilon \downarrow 0$ et $r \uparrow \infty$.

Algorithme 13.9 (pénalisation extérieure) Une itération passe de l'itéré courant $(x_k, \varepsilon_k, r_k) \in \mathbb{E} \times \mathbb{R}_{++} \times \mathbb{R}$ à l'itéré suivant $(x_{k+1}, \varepsilon_{k+1}, r_{k+1})$ par les étapes suivantes.

1. *Test d'arrêt.* Arrêt si x_k est satisfaisant.
2. *Nouveaux paramètres.* Choisir $\varepsilon_{k+1} < \varepsilon_k$ (plus de précision) et $r_{k+1} > r_k$ (plus de pénalisation).
3. *Nouvel itéré.* Trouver un minimiseur *approché* (à ε_{k+1} près) x_{k+1} de $\Theta_{r_{k+1}}$ en démarrant les itérations en x_k .

Le schéma algorithmique décrit ci-dessus est simple mais peu précis. Il pose quelques problèmes de mise en œuvre qui méritent d'être discutés.

Remarques 13.10 1) Voici deux critères d'arrêt qui peuvent être utilisés au point 1.

- Pour une fonction pénalisante p vérifiant $p(X) = \{0\}$, $\bar{x}$ est solution de (P_X) , si $\bar{x}$ est admissible et minimise Θ_r . En effet, dans ce cas, en plus d'être admissible, $\bar{x}$ vérifie

$$\begin{aligned} f(\bar{x}) &= \Theta_r(\bar{x}) & [p(X) = \{0\}] \\ &\leq \Theta_r(x), \quad \forall x \in \mathbb{E} & [\text{optimalité de } \bar{x}] \\ &= f(x), \quad \forall x \in X & [p(X) = \{0\}], \end{aligned}$$

si bien que $\bar{x}$ minimise f sur X . On peut donc considérer que si x_k minimise Θ_{r_k} avec suffisamment de précision (i.e., avec un $\varepsilon_k > 0$ suffisamment petit) et si x_k est presque admissible (la vérification de cette propriété peut faire intervenir les fonctions décrivant X), alors x_k est satisfaisant.

- On peut aussi se contenter de l'optimalité au premier ordre, en ne vérifiant que la satisfaction approchée des conditions d'optimalité de KKT (4.33). Il faut alors disposer d'un multiplicateur optimal approché, ce qui est possible pour certaines fonctions pénalisantes (pour la pénalisation quadratique, voir la proposition 13.11).
- 2) Si l'on veut que les solutions approchées de (P_r) convergent vers une solution de (P_X) , il faut nécessairement que $\varepsilon_k \downarrow 0$ et $r_k \uparrow \infty$ (proposition 13.11).
 - 3) À l'étape 3, il n'est pas aisé de donner un critère d'arrêt pour la minimisation de $\Theta_{r_{k+1}}$ qui soit entièrement satisfaisant. Si l'on ne s'intéresse qu'aux points stationnaires, on pourra par exemple décider d'arrêter cette étape si

$$\|\nabla \Theta_{r_{k+1}}(x_{k+1})\| \leq \varepsilon_{k+1}.$$

On aimerait en effet ne pas passer trop de temps dans la minimisation de Θ_r si son minimiseur est éloigné de la solution du problème original, parce que r n'est pas assez grand. Nous verrons que, dans l'approche par points intérieurs en optimisation linéaire (chapitre 18), cette question est traitée avec beaucoup plus de précision qu'ici.

- 4) La manière avec laquelle l'algorithme 13.9 procède, celle de minimiser (approximativement) une suite de fonctions de pénalisation avec des facteurs de pénalisation r croissants, a un aspect étrange. Pourquoi, en effet, ne pas prendre directement r très grand et ne minimiser qu'une seule fonction Θ_r ?

Une première raison est que l'on ne sait pas ce qu'est une valeur de r très grande (cela dépend du problème traité), ce qui donne du sens à cette détermination progressive d'un bon facteur de pénalisation.

Mais il y a une autre raison, plus importante que la première, qui provient du mauvais conditionnement du problème de minimisation de Θ_r pour de grandes valeurs de r (remarque 13.12.2) et des erreurs numériques. Ainsi, si l'on commence à minimiser Θ_r avec une grande valeur de r à partir d'un itéré initial arbitraire, il n'est numériquement pas possible de trouver son minimum, parce que les itérés générés vont surtout se concentrer sur la décroissance de la fonction pénalisante p en ne voyant plus f , dont la contribution à Θ_r est alors marginale. Pour remédier à cette situation, il faut suivre le *chemin des minimiseurs* $r \mapsto \bar{x}_r$ (le tracé progressif rouge à la figure 13.2). Les chances de trouver un point minimisant Θ_r seront d'autant plus grandes que l'itéré initial n'est pas trop éloigné du chemin des minimiseurs et que le minimiseur de Θ_r n'est pas trop éloigné de cet itéré initial. Ces conditions sont remplies dans l'algorithme 13.9 si $r_{k+1} > r_k$ n'est pas pris beaucoup plus grand que r_k , car on y suppose que l'itéré initial x_{r_k} de minimisation de $\Theta_{r_{k+1}}$ est proche de celui minimisant Θ_{r_k} , donc proche du chemin des minimiseurs. $\square$

Les considérations précédentes montrent que l'algorithme 13.9 est conceptuellement simple mais coûteux (il demande beaucoup de temps de calcul), puisqu'il faut

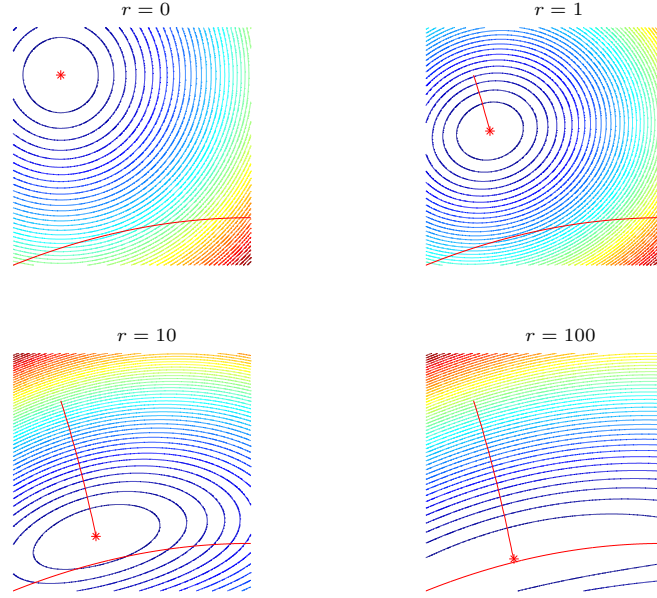


Fig. 13.2. Tracé progressif du *chemin des minimiseurs* dans la pénalisation quadratique du problème $\inf\{\|x - x_0\|_2^2 : c(x) = 0\}$, où $x_0 = (0.2, 0.8)$ et $c(x) = x_2 + 0.2(x_1 - 1)^2 - 0.2$. Celui-ci converge vers la solution du problème et l'on pourra trouver cette solution en suivant ce chemin. Les courbes de niveaux des fonctions de pénalisation, pour des facteurs de pénalisation r pris dans $\{0, 1, 10, 100\}$, montrent la détérioration du conditionnement avec l'augmentation de r (rapprochement progressif de celles-ci sur celles de $\|c(\cdot)\|_2$).

nécessairement résoudre une *suite* de problèmes d'optimisation *non linéaires*, avec l'intérêt de ne jamais devoir prendre en charge explicitement les contraintes de (P_X) .

13.2.4 Pénalisation quadratique

On considère à présent le cas plus concret de la résolution numérique par pénalisation quadratique du problème (P_{EI}) , un problème introduit à la section 4.4 et rappelé en (13.3). On parle de *pénalisation quadratique*, lorsque le problème pénalisé associée à (P_{EI}) est le suivant

$$\inf_{x \in \mathbb{E}} \left(\Theta_r(x) := f(x) + \frac{r}{2} \|c(x)^\# \|^2 \right), \quad (13.12)$$

où $v^\# \in \mathbb{R}^m$ est défini en (13.4). Lorsque c est continue, il s'agit d'une pénalisation extérieure, puisque la fonction pénalisante $x \mapsto p(x) := \frac{r}{2} \|c(x)^\# \|^2$ vérifie les conditions (13.6); dès lors les résultats obtenus à la section 13.2.2 s'appliquent. Lorsque f et c sont différentiable en $x \in \mathbb{E}$, on a

$$\nabla \Theta_r(x) = \nabla f(x) + \sum_{i \in E \cup I} r [c(x)]_i^\# \nabla c_i(x). \quad (13.13)$$

On trouvera à l'exercice 13.1 quelques informations supplémentaires sur cette fonction de pénalisation.

Bien qu'étant une méthode entièrement primale, le schéma algorithmique 13.9 utilisant la pénalisation quadratique permet d'obtenir une estimation des multiplicateurs optimaux associés aux contraintes d'égalité ou d'inégalité de (P_{EI}) . Cette estimation peut être précieuse, en particulier, pour estimer l'erreur commise sur les conditions d'optimalité du premier ordre (4.33) et ainsi mesurer la proximité d'un point stationnaire de (P_{EI}) . C'est aussi un moyen de montrer l'existence de multiplicateur optimaux.

Proposition 13.11 (approximation d'un multiplicateur optimal) *On suppose que les fonctions f et c définissant (P_{EI, X_s}) sont continûment différentiables et que pour une suite de $r \rightarrow \infty$, Θ_r a un point stationnaire approché $\bar{x}_r$, dans le sens où*

$$\|\nabla \Theta_r(\bar{x}_r)\| \leq \varepsilon_r,$$

avec $\varepsilon_r \rightarrow 0$ quand $r \rightarrow +\infty$. On suppose aussi que $\bar{x}_r \rightarrow \bar{x}$ lorsque $r \rightarrow +\infty$, que $\bar{x}$ est un point admissible de (P_{EI}) et que les conditions (QC-MF) ont lieu en $\bar{x}$. Alors

- 1) $\{r c(\bar{x}_r)^\#\}_{r \rightarrow \infty}$ est bornée,
- 2) tout point d'adhérence $\bar{\lambda}$ de $\{r c(\bar{x}_r)^\#\}$ est tel que $(\bar{x}, \bar{\lambda})$ vérifie les conditions d'optimalité (4.33) de (P_{EI, X_s}) .

DÉMONSTRATION. La différentiabilité supposée de f et c implique celle de Θ_r . Alors, en tenant compte de (13.13), on voit que l'optimalité approchée de $\bar{x}_r$ implique que

$$\nabla f(\bar{x}_r) + \sum_{i \in E \cup I} r [c(\bar{x}_r)]_i^\# \nabla c_i(\bar{x}_r) \rightarrow 0, \quad \text{lorsque } r \rightarrow \infty. \quad (13.14)$$

1) Notons $\lambda_r := r [c(\bar{x}_r)]^\#$ et montrons par l'absurde que la suite $\{\lambda_r\}_{r \rightarrow \infty}$ est bornée. S'il n'en était pas ainsi, on pourrait trouver une sous-suite de $r \rightarrow \infty$ telle que

$$\|\lambda_r\| \rightarrow \infty \quad \text{et} \quad \frac{\lambda_r}{\|\lambda_r\|} \rightarrow \mu \neq 0. \quad (13.15)$$

Après division par $\|\lambda_r\|$, la limite (13.14) donne

$$\sum_{i \in E \cup I} \mu_i \nabla c_i(\bar{x}) = 0. \quad (13.16)$$

Par la définition de λ_r , on voit que $\mu_I \geq 0$. Par ailleurs, la convergence de $\bar{x}_r \rightarrow \bar{x}$ et la définition de λ_r montrent que $\mu_{I \setminus I^0(\bar{x})} = 0$. Alors, l'identité (13.16) et (QC-MF) (définition 4.41) impliquent que $\mu = 0$, ce qui est en contradiction avec (13.15).

2) Soit $\bar{\lambda}$ un point d'adhérence de $\{\lambda_r\}$. En passant à la limite dans (13.14), on voit que le gradient du lagrangien s'annule :

$$\nabla f(\bar{x}) + \sum_{i \in E \cup I} \bar{\lambda}_i \nabla c_i(\bar{x}) = 0.$$

Par ailleurs, la définition de λ_r montre que $\bar{\lambda}_I \geq 0$ et que $\bar{\lambda}_{I \setminus I^0(\bar{x})} = 0$. Le couple $(\bar{x}, \bar{\lambda})$ vérifie donc les conditions d'optimalité du premier ordre de (P_{EI}) . $\square$

Remarques 13.12 1) Accessoirement, la convergence d'une sous-suite convergente de $\{rc(\bar{x}_r)^\#\}$ vers un multiplicateur optimal $\bar{\lambda}$ montre que, pour les contraintes d'inégalité c_i associées à des multiplicateurs $\bar{\lambda}_i > 0$, on a $rc_i(\bar{x}_r) > 0$ pour r grand. Ceci veut dire que ces contraintes d'inégalité ne sont pas vérifiées pour r grand. Autrement dit, $\bar{x}_r$ converge vers $\bar{x}$ par l'*extérieur* de l'ensemble défini par ces contraintes.

2) La pénalisation quadratique permet de mettre en évidence l'influence du facteur de pénalisation r sur le *conditionnement du problème* (P_r) . Nous entendons par là le conditionnement de la hessienne de Θ_r lorsque celle-ci existe. La formule (13.13) montre que la fonction de pénalisation quadratique Θ_r est différentiable, mais n'est pas nécessairement deux fois différentiable si $I \neq \emptyset$. Pour mettre en évidence par calcul la détérioration du conditionnement du problème (P_r) avec r , considérons alors le cas où il n'y a que des contraintes d'égalité ($I = \emptyset$ et on note $c_E = c$). Alors

$$\begin{aligned}\Theta_r(x) &= f(x) + \frac{r}{2} \|c(x)\|^2, \\ \nabla \Theta_r(x) &= \nabla f(x) + r c'(x)^\top c(x), \\ \nabla^2 \Theta_r(x) &= \nabla^2 f(x) + \sum_{i=1}^m r c_i(x) \nabla^2 c_i(x) + r c'(x)^\top c'(x).\end{aligned}$$

Intéressons-nous à la hessienne $\nabla^2 \Theta_r(\bar{x}_r)$ en un point stationnaire $\bar{x}_r$ de Θ_r . Sous les hypothèses de la proposition 13.11, les points d'adhérence des facteurs $r c_i(\bar{x}_r)$ de la hessienne $\nabla^2 c_i(\bar{x}_r)$ sont des multiplicateurs optimaux, si bien que les deux premiers termes $\nabla^2 f(\bar{x}_r) + \sum_{i=1}^m r c_i(\bar{x}_r) \nabla^2 c_i(\bar{x}_r)$ de la hessienne $\nabla^2 \Theta_r(\bar{x}_r)$ ont comme point d'adhérence des hessiennes du lagrangien en une solution, si bien que ces termes n'induisent pas un conditionnement problématique. À l'inverse le dernier terme a en général un conditionnement qui explose avec r . En effet, la forme quadratique associée à $r c'(\bar{x}_r)^\top c'(\bar{x}_r)$, à savoir $v \mapsto r \|c'(\bar{x}_r)v\|^2$, est nulle dans le noyau de $c'(\bar{x}_r)$, alors que sa courbure explose dans l'espace orthogonal (si celui-ci n'est pas de dimension nulle). $\square$

L'application de la pénalisation quadratique à l'optimisation quadratique est examinée à la section ??.

13.3 Pénalisation intérieure ▲

Dans certains problèmes, le fait que les itérés $\bar{x}_r$ générés par pénalisation extérieure ne soient pas admissibles peut être un inconvénient, par exemple, parce que f n'est pas définie à l'extérieur de X . On peut introduire des méthodes de pénalisation dans lesquelles les itérés $\bar{x}_r$ restent dans X . On parle alors de *pénalisation intérieure*. L'idée est d'utiliser un terme de pénalisation p qui tend vers l'infini lorsque x s'approche de la frontière ∂X de X .

Au problème simple (13.5), on pourra par exemple associer la fonction de pénalisation intérieure, dite logarithmique, suivante

$$\Theta_r(x) = 1 - x - \frac{1}{3}x^3 - r \log(-x).$$

L'effet de cette pénalisation peut s'observer à la figure 13.3.

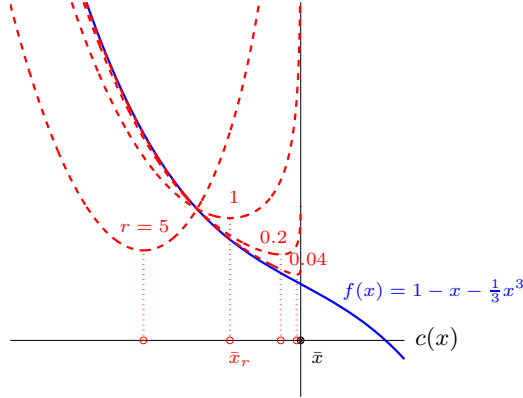


Fig. 13.3. Pénalisation logarithmique de l'exemple (13.5) pour $r = 5, 1, 0.25, 0.05$

On comprend que, dans cette section, il est nécessaire de supposer que l'intérieur X° de X est non vide :

$$X^\circ \neq \emptyset.$$

Cette hypothèse exclut d'emblée la possibilité de prendre en compte directement des contraintes d'égalité. Des artifices permettent toutefois de traiter de telles contraintes.

Les fonctions pénalisantes p considérées dans cette section satisfont les conditions suivantes :

$$p \text{ est continue sur } X^\circ \quad (13.17a)$$

$$p \geq 0, \text{ sur } X^\circ \quad (13.17b)$$

$$p(x) \rightarrow +\infty, \quad \text{quand } x \in X^\circ \text{ converge vers un point de } \partial X. \quad (13.17c)$$

On considère alors le problème de pénalisation

$$(P_r) \quad \inf_{x \in X^\circ} \Theta_r(x),$$

où $\Theta_r(x) = f(x) + rp(x)$. La condition (13.17c) crée une « barrière » au bord de l'ensemble admissible, si bien que Θ_r porte parfois le nom de *fonction barrière*.

Le tableau 13.2 donne deux exemples de fonctions p satisfaisant (13.17) lorsque l'ensemble admissible s'écrit $X = \{x \in \mathbb{E} : c(x) \leq 0\}$, avec $c : \mathbb{E} \rightarrow \mathbb{R}^m$. On suppose que $\{x \in \mathbb{E} : c(x) < 0\}$ n'est pas vide. La fonction de *pénalisation intérieure inverse* est due à Carroll [107 ; 1961]. La fonction de *pénalisation logarithmique* est due à l'économétricien norvégien R. Frisch [223 ; 1955]. Cette pénalisation a connu un renouveau avec les *algorithmes de points intérieurs*, que nous étudierons plus en détail au chapitre 18 dans le cadre de l'optimisation linéaire.

Le théorème suivant étudie la suite $\{\bar{x}_r\}$ des solutions des problèmes pénalisés. Contrairement à la pénalisation extérieure, il faut ici faire tendre r vers 0 (et non vers $+\infty$), ce qui a pour effet de diminuer l'influence de la fonction pénalisante, dont le

Contraintes définissant X	Fonction pénalisante	Nom de la pénalisation	Références
$c(x) \leq 0$	$p(x) = \sum_{i=1}^m \frac{1}{-c_i(x)}$	inverse	[107, 206]
$c(x) \leq 0$	$p(x) = -\sum_{i=1}^m \log(-c_i(x))$	logarithmique	[223]

Tableau 13.2. Exemples de pénalisation intérieure.

rôle est de repousser les points vers l'intérieur de X , et donc de permettre à $\bar{x}_r$ de se rapprocher de la frontière de l'ensemble admissible, si cela est nécessaire.

Théorème 13.13 (convergence de la pénalisation intérieure) *Supposons que f soit continue sur $\mathbb{E}$ et que l'ensemble admissible X , non vide, vérifie*

$$X = \overline{X^\circ}.$$

On suppose également que soit X est borné, soit $f(x) \rightarrow \infty$ quand $\|x\| \rightarrow \infty$. Alors, si la fonction pénalisante p vérifie (13.17), on a

- 1) $\forall r > 0$, (P_r) a au moins une solution $\bar{x}_r$,
- 2) la suite $\{\bar{x}_r\}_{r \downarrow 0}$ est bornée,
- 3) tout point d'adhérence de $\{\bar{x}_r\}_{r \downarrow 0}$ est solution de (P_X) .

Algorithme 13.14 (pénalisation intérieure) Une itération passe de l'itéré courant $(x_k, \varepsilon_k, r_k) \in \mathbb{E} \times \mathbb{R}_{++} \times \mathbb{R}$ à l'itéré suivant $(x_{k+1}, \varepsilon_{k+1}, r_{k+1})$ par les étapes suivantes.

1. *Test d'arrêt.* Arrêt si x_k est satisfaisant.
2. *Nouveaux paramètres.* Choisir $\varepsilon_{k+1} < \varepsilon_k$ (plus de précision) et $r_{k+1} \in]0, r_k[$ (moins de pénalisation).
3. *Nouvel itéré.* Trouver un minimiseur *approché* (à ε_{k+1} près) x_{k+1} de $\Theta_{r_{k+1}}$ en démarrant les itérations en x_k .

13.4 Le lagrangien augmenté

Since you ask me to mention a gratifying paper, let me pick "A method for nonlinear constraints in minimization problems", because it is regarded as one of the sources of the "augmented Lagrangian method", which is now of fundamental importance in mathematical programming. I have been very fortunate to have played a part in discoveries of this kind.

M.J.D. POWELL [497; 2003]

Le fait de devoir faire tendre r vers sa valeur limite ($+\infty$ ou 0 suivant le type de pénalisation, extérieure ou intérieure) dans les fonctions de pénalisation précédentes, pour retrouver la solution du problème original, induit, on l'a vu, un mauvais conditionnement des problèmes pénalisés. Il est raisonnable de penser que cet effet numériquement indésirable (si la pénalisation est utilisée comme outil théorique, on se fiche complètement de cette question) sera moins critique si l'on peut construire des fonctions de pénalisation qui, pour une valeur finie du paramètre de pénalisation, ont un minimum en la solution du problème original. C'est ce que l'on a appelé une pénalisation exacte.

Si une fonction de pénalisation Θ_r est exacte et différentiable en une solution $\bar{x}$, on doit avoir $\nabla \Theta_r(\bar{x}) = 0$. C'est bien ce qui manque aux fonctions de la table 13.1, puisqu'elles satisfont $\nabla \Theta_r(\bar{x}) = \nabla f(\bar{x})$, qui n'est pas nul en général. Cela se voit aussi à la figure 13.1 dans laquelle la pente $\Theta'_r(0) = f'(0)$ est non nulle et indépendante de r . Dès lors, avant d'ajouter un terme quadratique à f , il semble judicieux de lui ajouter un terme linéaire de telle sorte que la pénalisation agisse sur une fonction ayant une pente nulle en $\bar{x}$. C'est ce que suggère la figure 13.4, dans laquelle on a

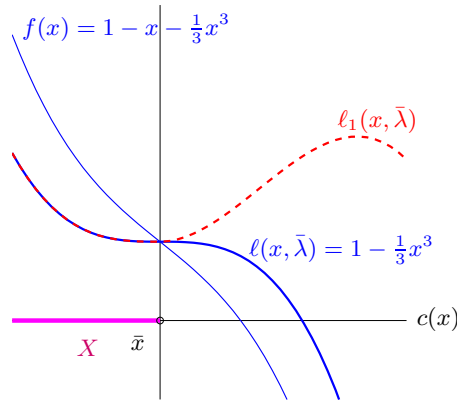


Fig. 13.4. Pénalisation quadratique du lagrangien du problème (13.5)

ajouté la fonction linéaire $x \rightarrow x$ au critère $x \mapsto 1 - x - \frac{1}{3}x^3$ de l'exemple 13.4. On voit que la pénalisation quadratique de la fonction résultante admet $\bar{x}$ comme minimiseur local, ici quel que soit le facteur de pénalisation $r > 0$; la fonction de pénalisation obtenue est donc exacte. Que prendre comme terme linéaire? On peut raisonner de deux manières différentes, qui sont d'ailleurs reliées entre elles. Le plus simple est de dire que l'on connaît une fonction dont la pente est nulle en la solution : c'est le lagrangien $\ell(\cdot, \bar{\lambda})$ avec multiplicateur optimal $\bar{\lambda}$, comme nous l'apprennent les conditions d'optimalité (4.33). C'est donc le terme $\bar{\lambda}^T c(x)$ qu'il faut ajouter à $f(x)$ avant de faire agir la pénalisation. On remarquera que le terme est linéaire en $c(x)$, pas en x (c'est la même chose pour le problème 13.5). On peut aussi raisonner en utilisant l'analyse de sensibilité de la section 4.6.1. La fonction valeur v admet comme dérivée première (sous les hypothèses fortes de la proposition 4.72): $v'(0) = \bar{\lambda}$. Si l'on

veut modifier le problème pour que cette dérivée soit nulle, il faut ajouter $\bar{\lambda}^\top c(x)$ au critère, ce qui revient à faire la transformation $f(\cdot) \curvearrowright \ell(\cdot, \bar{\lambda})$ opérée ci-dessus.

13.4.1 Conditions d'exactitude du lagrangien

Le lagrangien du problème (P_{EI, X_s}) est une fonction de pénalisation exacte si les données du problème sont convexes et si le multiplicateur utilisé est optimal. C'est en substance ce qu'affirme le résultat suivant.

Proposition 13.15 (exactitude du lagrangien d'un problème convexe)

Supposons que f et les $\{c_i\}_{i \in I}$ soient convexes et que c_E soit affine. On suppose également que f et c sont différentiables en une solution $\bar{x}$ du problème et qu'il existe un multiplicateur $\bar{\lambda}$ tel que les conditions d'optimalité (KKT) aient lieu. Alors $x \mapsto \ell(x, \bar{\lambda})$ a un minimum en $\bar{x}$.

DÉMONSTRATION. Avec les hypothèses de convexité et la positivité de $\bar{\lambda}_I$, l'application $x \in \mathbb{E} \mapsto \ell(x, \bar{\lambda})$ est convexe. Selon les hypothèses, cette fonction est différentiable en $\bar{x}$ et, d'après les conditions d'optimalité (KKT), on a $\nabla_x \ell(\bar{x}, \bar{\lambda}) = 0$. On en déduit que $\ell(\cdot, \bar{\lambda})$ a un minimum (global) en $\bar{x}$. $\square$

Si l'on connaissait un multiplicateur optimal $\bar{\lambda}$, $\ell(\cdot, \bar{\lambda})$ serait une fonction de pénalisation, n'ayant peut-être pas la forme de Θ_r dans (13.1), mais qui pourrait être minimisée pour trouver une solution de (P_{EI, X_s}) . Mais $\bar{\lambda}$ est une solution duale, qui doit aussi être trouvée. On appelle *relaxation lagrangienne*, la technique qui recherche une solution de (P_{EI, X_s}) en minimisant $\ell(\cdot, \lambda)$ pour une suite de multiplicateurs λ que l'on fait converger vers un multiplicateur optimal. Il reste à préciser comment on met à jour les multiplicateurs et à étendre cette approche aux problèmes non convexes. C'est ce à quoi nous nous attacherons dans cette section 13.4.

13.4.2 Le lagrangien augmenté de (P_E)

Si le problème n'est pas convexe, le lagrangien n'est plus nécessairement une fonction de pénalisation exacte. Il peut aussi ne pas être borné inférieurement. C'est le cas dans l'exemple (13.5), pour lequel le multiplicateur optimal vaut $\bar{\lambda} = 1$ (d'après la proposition 4.72, c'est $f'(0) : x \mapsto \ell(x, \bar{\lambda}) = 1 + \frac{1}{3}x^3$ n'est pas bornée inférieurement. On comprend qu'en l'absence de convexité, il n'y a plus de sens à minimiser le lagrangien pour obtenir une solution de (P_{EI, X_s}) . Le lagrangien augmenté peut être vu comme un moyen de remédier à cet inconvénient.

Commençons par le cas où il n'y a que des contraintes d'égalité, en considérant le problème (P_E) de la page 569. Les conditions d'optimalité du premier et second ordre du théorème 4.17 et de la proposition ?? nous disent que, si l'on restreint h à être voisin de 0 et à appartenir à l'espace tangent aux contraintes en $\bar{x}$, l'application $h \mapsto \ell(\bar{x} + h, \bar{\lambda})$ est convexe et minimale en $h = 0$. Par contre, les conditions d'optimalité ne disent rien sur les valeurs prises par $\ell(\cdot, \bar{\lambda})$ dans l'espace complémentaire au plan tangent. Comme $\nabla_x \ell(\bar{x}, \bar{\lambda}) = 0$, on comprend que l'on pourra convexifier $\ell(\cdot, \bar{\lambda})$ autour de $\bar{x}$, et créer ainsi localement une cuvette minimisée en $\bar{x}$, en lui ajoutant un terme positif

qui croît transversalement à la contrainte, mais qui est sans effet longitudinalement. Un terme ayant cette propriété est $\|c(x)\|_2^2$. Il semble donc naturel de considérer la fonction

$$\ell_r(x, \mu) = f(x) + \mu^\top c(x) + \frac{r}{2} \|c(x)\|_2^2. \quad (13.18)$$

Cette fonction, définie pour $(x, \mu) \in \mathbb{E} \times \mathbb{R}^m$, est le *lagrangien augmenté* associé au problème (P_E) ¹. Le facteur de pénalisation $r > 0$ est un paramètre qu'il faudra ajuster, tout comme le multiplicateur μ qui devra tendre vers un multiplicateur optimal.

Examinons la structure du lagrangien augmenté. On vient de le construire comme un lagrangien que l'on a pénalisé par un terme semblable à ceux utilisés en pénalisation extérieure (voir le tableau 13.1). Contrairement aux fonctions de pénalisation de cette section, le lagrangien augmenté sera une fonction de pénalisation exacte lorsque $\mu = \bar{\lambda}$ et r est assez grand (mais fini!), mais cette propriété ne sera que locale (notion qui sera clarifiée ci-dessous). En ajoutant le terme $\bar{\lambda}^\top c(x)$ à la fonction de pénalisation inexacte $x \mapsto f(x) + \frac{r}{2} \|c(x)\|_2^2$, on a forcé son exactitude en corrigeant la pente de cette fonction en $\bar{x}$. En ajoutant le terme $\frac{r}{2} \|c(x)\|_2^2$ au lagrangien, on peut à présent traiter des problèmes non convexes.

L'étude de l'exactitude du lagrangien augmenté (13.18) passe par le lemme B.3 de Finsler. Celui-ci décrit en quelque sorte la version linéarisée (ou «quadratisée») du mécanisme en jeu dans le lagrangien augmenté (et c'est comme cela que nous l'utiliserons): M est définie positive dans le noyau de $A^\top A$, mais on n'a pas d'information sur M dans l'espace complémentaire $\mathcal{R}(A^\top A)$ (comparez avec ce que l'on sait sur le lagrangien dans le voisinage d'une solution). En ajoutant un multiple de $A^\top A$ à M , on peut rendre la matrice résultante définie positive. On observera que la matrice ajoutée $rA^\top A$ n'a pas d'effet dans $\mathcal{N}(A)$, puisque $u^\top A^\top A u = 0$ pour tout $u \in \mathcal{N}(A)$.

Théorème 13.16 (exactitude du lagrangien augmenté de (P_E)) *On suppose que f et c sont deux fois dérivables en un minimum local $\bar{x}$ de (P_E) . On suppose également qu'il existe un multiplicateur $\bar{\lambda}$ tel que $\nabla_x \ell(\bar{x}, \bar{\lambda}) = 0$ et tel que la condition suffisante d'optimalité du second ordre (4.28) ait lieu. Alors, il existe un réel $\bar{r}$ tel que pour tout $r \geq \bar{r}$, le lagrangien augmenté (13.18) a un minimum local strict en $\bar{x}$.*

DÉMONSTRATION. On a $\nabla_x \ell_r(x, \bar{\lambda}) = \nabla_x \ell(x, \bar{\lambda}) + rA(x)^\top c(x)$, avec $A(x) := c'(x)$. Dès lors $\nabla_x \ell_r(\bar{x}, \bar{\lambda}) = 0$. D'autre part, $\nabla_{xx}^2 \ell_r(\bar{x}, \bar{\lambda}) = \nabla_{xx}^2 \ell(\bar{x}, \bar{\lambda}) + rA(\bar{x})^\top A(\bar{x})$. D'après la condition du second ordre (4.28), $\nabla_{xx}^2 \ell(\bar{x}, \bar{\lambda})$ est définie positive dans le noyau de $A(\bar{x})$. Alors, le lemme B.3 de Finsler nous apprend que $\nabla_{xx}^2 \ell_r(\bar{x}, \bar{\lambda})$ est définie positive lorsque r est assez grand. On en déduit que, pour r assez grand, $\ell_r(\cdot, \bar{\lambda})$ a un minimum local strict en $\bar{x}$ (proposition 4.11). $\square$

¹ Ce type de pénalisation porte le nom de lagrangien *modifié* dans la littérature russe.

13.4.3 Le lagrangien augmenté de (P_{EI})

Le lagrangien augmenté associé au problème (P_{EI, X_s}) s'introduit de manière naturelle en utilisant la dualité ; nous le ferons au chapitre 14. Nous allons l'introduire ici par une approche plus intuitive, qui s'appuie sur la formule (13.18) et qui apportera des informations qualitatives intéressantes.

Une idée pourrait être de suivre la structure de (13.18) en ajoutant au lagrangien le terme pénalisant les contraintes utilisé dans (13.12), ce qui donnerait la fonction

$$x \mapsto f(x) + \mu^\top c(x) + \frac{r}{2} \|c(x)^\# \|_2^2.$$

Celle-ci a l'inconvénient de ne pas toujours être deux fois différentiable en une solution (car $t \mapsto (t^+)^2$ ne l'est pas), quel que soit la régularité de f et c . On préfère donc adopter la démarche suivante, due à Rockafellar [520 ; 1973].

Dans un premier temps, on écrit (P_{EI, X_s}) sous une forme équivalente, en introduisant des *variables d'écart* $s \in \mathbb{R}^{m_I}$:

$$\begin{cases} \inf_{(x,s)} f(x) \\ c_E(x) = 0 \\ c_I(x) + s = 0 \\ s \geq 0. \end{cases}$$

Ensuite, ce problème est *approché* en utilisant le lagrangien augmenté associé à ses contraintes d'égalité (formule (13.18)) avec un facteur de pénalisation $r > 0$:

$$\inf_x \inf_{s \geq 0} \left(f(x) + \mu_E^\top c_E(x) + \frac{r}{2} \|c_E(x)\|_2^2 + \mu_I^\top (c_I(x) + s) + \frac{r}{2} \|c_I(x) + s\|_2^2 \right).$$

Le lagrangien augmenté du problème (P_{EI, X_s}) est la fonction de (x, μ) définie comme la valeur minimale du problème d'optimisation en $s \geq 0$ ci-dessus :

$$\begin{aligned} \ell_r(x, \mu) := \inf_{s \geq 0} & \left(f(x) + \mu_E^\top c_E(x) + \frac{r}{2} \|c_E(x)\|_2^2 \right. \\ & \left. + \mu_I^\top (c_I(x) + s) + \frac{r}{2} \|c_I(x) + s\|_2^2 \right). \end{aligned} \quad (13.19)$$

La minimisation en s peut être menée explicitement puisque le problème est quadratique en s avec une hessienne diagonale et que l'on n'a que des contraintes de borne sur s . Plus précisément, comme le critère du problème ci-dessus s'écrit

$$\frac{r}{2} \left\| s + c_I(x) + \frac{\mu_I}{r} \right\|_2^2 + \text{« des termes indépendants de } s \text{ »},$$

il s'agit de projeter $-c_I(x) - \mu_I/r$ sur l'orthant positif. On trouve donc

$$s = \max \left(-c_I(x) - \frac{\mu_I}{r}, 0 \right),$$

si bien que

$$c_I(x) + s = \max \left(-\frac{\mu_I}{r}, c_I(x) \right). \quad (13.20)$$

En remplaçant $c_I(x) + s$ par cette valeur dans le critère du problème d'optimisation ci-dessus on obtient le *lagrangien augmenté* associé aux contraintes $c_E(x) = 0$ et

$c_I(x) \leq 0$ du problème (P_{EI, X_s}) . C'est la fonction $\ell_r : \mathbb{E} \times \mathbb{R}^m \rightarrow \mathbb{R}$, définie pour $(x, \mu) \in \mathbb{E} \times \mathbb{R}^m$ et $r > 0$ par

$$\ell_r(x, \mu) = f(x) + \mu^\top \tilde{c}_{\mu, r}(x) + \frac{r}{2} \|\tilde{c}_{\mu, r}(x)\|_2^2, \quad (13.21)$$

où, pour $\lambda \in \mathbb{R}^{m_I}$ et $r > 0$, $\tilde{c}_{\mu, r} : \mathbb{E} \rightarrow \mathbb{R}^m$ est définie par

$$(\tilde{c}_{\mu, r}(x))_i = \begin{cases} c_i(x) & \text{si } i \in E \\ \max\left(\frac{-\mu_i}{r}, c_i(x)\right) & \text{si } i \in I. \end{cases} \quad (13.22)$$

Ce lagrangien augmenté a donc une structure tout à fait semblable au lagrangien augmenté associé au problème avec contraintes d'égalité (P_E) , pourvu que l'on fasse intervenir la modification non différentiable de c définie par $\tilde{c}_{\mu, r}$ ci-dessus. On notera d'ailleurs que

$$\tilde{c}_{\mu, r}(x) = 0 \iff \begin{cases} c_E(x) = 0 \\ 0 \leq \mu_I \perp c_I(x) \leq 0, \end{cases}$$

où la notation $0 \leq u \perp v \leq 0$ signifie $u \geq 0$, $v \leq 0$ et $u^\top v = 0$. La nullité des composantes I de $\tilde{c}_{\mu, r}(x)$ exprime donc à la fois les conditions de signes sur la contrainte et son multiplicateur, ainsi que la complémentarité.

Les deux relations de monotonie du lemme ci-dessous nous seront utiles. La seconde provient de l'expression suivante des termes associés aux contraintes d'inégalité dans le lagrangien augmenté :

$$\mu_I^\top (\tilde{c}_{\mu, r}(x))_I + \frac{r}{2} \|(\tilde{c}_{\mu, r}(x))_I\|_2^2 = \frac{1}{2r} \sum_{i \in I} [\max(0, \mu_i + r c_i(x))^2 - \mu_i^2]. \quad (13.23)$$

Lemme 13.17 1) $\forall (x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m$, $r \in \mathbb{R}_{++} \rightarrow \ell_r(x, \mu)$ est croissante.
 2) $\forall r > 0$, $\forall \mu_I \in \mathbb{R}^{m_I}$, $v_I \in \mathbb{R}^{m_I} \rightarrow \mu_I^\top \max\left(\frac{-\mu_I}{r}, v_I\right) + \frac{r}{2} \left\| \max\left(\frac{-\mu_I}{r}, v_I\right) \right\|^2$ est croissante (pour l'ordre $v_I \leq v'_I \Leftrightarrow v_i \leq v'_i$ pour tout $i \in I$).

DÉMONSTRATION. 1) C'est une conséquence de la technique utilisée ci-dessus pour construire le lagrangien augmenté (13.21) : l'argument de l'infimum dans (13.19) croît lorsque r augmente.

2) On utilise l'expression de droite dans (13.23), dont la croissance en $c_I(x)$ est claire, quel que soit le signe de μ_I . $\square$

Malgré l'opérateur $\max$ dans (13.22), le lagrangien augmenté est différentiable en (x, λ) . Pour le voir, le plus simple est d'utiliser l'expression à droite dans (13.23) pour les termes associés aux contraintes d'inégalité. Comme la dérivée de $(t^+)^2$ est $2t^+$ et que $(a+b)^+ - a = \max(-a, b)$, on obtient

$$\nabla_x \ell_r(x, \mu) = \nabla f(x) + c'(x)^\top (\mu + r \tilde{c}_{\mu, r}(x)) \quad (13.24)$$

$$\nabla_\mu \ell_r(x, \mu) = \tilde{c}_{\mu, r}(x). \quad (13.25)$$

L'opérateur max dans (13.22) ne permet pas, en général, d'avoir la différentiabilité seconde de $\ell_r(\cdot, \mu)$: c'est le talon d'Achille de ce lagrangien augmenté. Il a été à la source de nombreux développements. En fait, pour x proche d'une solution $\bar{x}$ de (P_{EI, X_s}) et $\mu = \bar{\lambda}$ (un multiplicateur optimal), on a en utilisant la complémentarité $\bar{\lambda}_I^T c_I(\bar{x}) = 0$ et la positivité de $\bar{\lambda}_I$:

$$\ell_r(x, \bar{\lambda}) = \ell(x, \bar{\lambda}) + \frac{r}{2} \sum_{i \in E \cup I^{0+}(\bar{x})} c_i(x)^2 + \frac{r}{2} \sum_{i \in I^{00}(\bar{x})} (c_i(x)^+)^2. \quad (13.26)$$

La présence de l'opérateur $(\cdot)^+$ dans (13.26) montre que $\ell_r(\cdot, \bar{\lambda})$ peut ne pas être plus d'une fois différentiable en $\bar{x}$. En cas de complémentarité stricte, $I^{00}(\bar{x}) = \emptyset$ et la dernière somme disparaît, si bien que le lagrangien augmenté peut s'écrire (toujours pour x proche de $\bar{x}$) :

$$\ell_r(x, \bar{\lambda}) = \ell(x, \bar{\lambda}) + \frac{r}{2} \sum_{i \in E \cup I^0(\bar{x})} c_i(x)^2.$$

Localement, les contraintes d'égalité et les contraintes d'inégalité actives sont alors traitées de la même manière et $\ell_r(\cdot, \bar{\lambda})$ est régulière en $\bar{x}$ (pourvu que f et c le soient). Nous résumons ces propriétés de différentiabilité dans la proposition suivante.

Proposition 13.18 (différentiabilité du lagrangien augmenté)

- 1) Le lagrangien augmenté ℓ_r , défini en (13.21), est différentiable en μ et son gradient est donné par (13.25).
- 2) Si f et c sont différentiables en x , alors ℓ_r est différentiable en x et son gradient est donné par (13.24).
- 3) Si $(\bar{x}, \bar{\lambda})$ est un point stationnaire de (P_{EI, X_s}) vérifiant la complémentarité stricte et si $(f, c_{E \cup I^0(\bar{x})})$ est p fois dérivable ($p \in \mathbb{N}$) dans un voisinage de $\bar{x}$, alors le lagrangien augmenté est p fois dérivable dans un voisinage (éventuellement plus petit) de $\bar{x}$.

Le résultat suivant est l'analogue du théorème 13.16 pour le lagrangien augmenté (13.21). Il donne des conditions pour que ce lagrangien augmenté soit exact en $\bar{x}$. La condition suffisante d'optimalité semi-forte (4.62) qui y est utilisée est plus forte que la condition faible (4.59), mais plus faible que la condition forte (4.63). Le résultat n'a pas lieu si on ne suppose que la condition faible (4.59).

On a vu au théorème 13.16 que l'exactitude du lagrangien augmenté de (P_E) était fondée sur le lemme de Finsler. Le lecteur perspicace remarquera que la technique utilisée dans la démonstration du théorème ci-dessous est calquée sur celle mise en œuvre en annexe pour démontrer le lemme B.3 de Finsler.

Théorème 13.19 (exactitude du lagrangien augmenté de (P_{EI, X_s})) On suppose que f et $c_{E \cup I^0(\bar{x})}$ sont deux fois dérivables en un minimum local $\bar{x}$ de (P_{EI, X_s}) . On suppose également que les conditions de (KKT) ont lieu et que la

condition suffisante d'optimalité du second ordre semi-forte (4.62) a lieu pour un certain multiplicateur optimal $\bar{\lambda}$. Alors, il existe un réel $\bar{r} > 0$ et un voisinage V de $\bar{x}$ tels que pour tout $r \geq \bar{r}$, le lagrangien augmenté (13.21), avec $\mu = \bar{\lambda}$, a un minimum strict en $\bar{x}$ sur V .

DÉMONSTRATION. Il suffit de montrer qu'il existe un $\bar{r} > 0$ et un voisinage V de $\bar{x}$ dans $\mathbb{E}$ tel que

$$\ell_{\bar{r}}(\bar{x}, \bar{\lambda}) < \ell_{\bar{r}}(x, \bar{\lambda}), \quad \text{pour tout } x \in V \setminus \{\bar{x}\}.$$

En effet, si cette affirmation est vraie pour $\bar{r}$, elle le sera pour tout $r \geq \bar{r}$, avec le même voisinage V . Ceci est dû au fait que $\ell_r(\bar{x}, \bar{\lambda}) = f(\bar{x})$ est indépendant de r et que, d'après le point (i) du lemme 13.17, $r \mapsto \ell_r(x, \bar{\lambda})$ croît avec r .

On démontre cette affirmation par l'absurde en supposant qu'il existe une suite de réels $r_k \rightarrow \infty$ et une suite de points $x_k \rightarrow \bar{x}$, tels que pour $k \geq 1$: $x_k \neq \bar{x}$ et

$$\ell_{r_k}(x_k, \bar{\lambda}) \leq \ell_{r_k}(\bar{x}, \bar{\lambda}). \quad (13.27)$$

En extrayant une sous-suite au besoin, on peut supposer que pour $k \rightarrow \infty$:

$$\frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \rightarrow d, \quad \text{avec } \|d\| = 1.$$

En posant $t_k := \|x_k - \bar{x}\|$, on obtient alors

$$x_k = \bar{x} + t_k d + o(t_k).$$

Notre but à présent est de montrer que d est une direction critique. Ceci s'obtient en faisant un développement limité des fonctions intervenant dans le membre de gauche de (13.27), exprimé par la formule (13.26) : développement au deuxième ordre du lagrangien et au premier ordre des contraintes dans les deux sommes de (13.26). Pour simplifier les notations, on introduit $\bar{L} := \nabla_{xx}^2 \ell(\bar{x}, \bar{\lambda})$. De la régularité de f et c et de l'optimalité de $(\bar{x}, \bar{\lambda})$, on déduit

$$\begin{aligned} \ell(x_k, \bar{\lambda}) &= \ell(\bar{x}, \bar{\lambda}) + \frac{t_k^2}{2} d^\top \bar{L} d + o(t_k^2), \\ c_i(x_k) &= t_k c'_i(\bar{x}) \cdot d + o(t_k), \quad \text{pour } i \in E \cup I^0(\bar{x}). \end{aligned}$$

On injecte ces estimations dans (13.27), en utilisant (13.26) et $\ell_{r_k}(\bar{x}, \bar{\lambda}) = \ell(\bar{x}, \bar{\lambda})$:

$$\begin{aligned} \frac{t_k^2}{2} d^\top \bar{L} d + o(t_k^2) + \frac{r_k}{2} \sum_{i \in E \cup I^{0+}(\bar{x})} (t_k c'_i(\bar{x}) \cdot d + o(t_k))^2 \\ + \frac{r_k}{2} \sum_{i \in I^{00}(\bar{x})} \left([t_k c'_i(\bar{x}) \cdot d + o(t_k)]^+ \right)^2 \leq 0. \end{aligned} \quad (13.28)$$

La limite dans (13.28) quand $k \rightarrow \infty$, après avoir divisé par $t_k^2 r_k$, conduit à

$$\begin{aligned} c'_i(\bar{x}) \cdot d &= 0, & \text{si } i \in E \cup I^{0+}(\bar{x}) \\ c'_i(\bar{x}) \cdot d &\leq 0, & \text{si } i \in I^{00}(\bar{x}). \end{aligned}$$

Dès lors, d est une direction critique non nulle.

D'autre part, (13.28) implique également

$$\frac{t_k^2}{2} d^\top \bar{L} d + o(t_k^2) \leq 0.$$

En prenant la limite après avoir divisé par t_k^2 , on obtient $d^\top \bar{L} d \leq 0$. Cette inégalité est en contradiction avec (4.62), puisque $d \in C(\bar{x}) \setminus \{0\}$. $\square$

13.4.4 Méthode du lagrangien augmenté

Le concept de lagrangien augmenté a été introduit dans un but algorithmique. Il est toujours utilisé aujourd'hui pour résoudre de grands problèmes, surtout lorsque les techniques d'algèbre linéaire requises dans les algorithmes newtoniens comme la *programmation quadratique successive* (voir chapitre 15) ne peuvent pas être utilisées, du fait de la dimension des problèmes. On utilise aussi le lagrangien augmenté pour résoudre des problèmes structurés, comme ceux de l'optimisation quadratique convexe [161, 122; 2005-2016].

L'algorithme classique associé au lagrangien augmenté est connu sous le nom de *méthode des multiplicateurs*. Cette méthode s'apparente à la relaxation lagrangienne, dans le sens où l'on y minimise (avec plus ou moins de précision) le lagrangien augmenté pour une suite de multiplicateurs λ_k , que l'on cherche à faire converger vers un multiplicateur optimal, et de facteurs de pénalisation r_k que l'on adapte pour qu'ils soient « suffisamment grands ». L'algorithme est piloté par la recherche du multiplicateur optimal, d'où son nom.

D'un point de vue numérique, l'intérêt principal du lagrangien augmenté, par rapport aux méthodes de pénalisation extérieure et intérieure des sections 13.2 et 13.3, est de ne pas devoir faire tendre le facteur de pénalisation vers une limite qui rend la pénalisation indéfinie ($r \rightarrow +\infty$ en pénalisation extérieure) ou inopérante ($r \downarrow 0$ en pénalisation intérieure). Grâce à cette propriété, le lagrangien augmenté conserve un conditionnement raisonnable; ses courbes de niveau ne s'allongent pas trop. Son inconvénient majeur est de devoir mettre à jour un multiplicateur, en plus du facteur de pénalisation. Heureusement, on dispose d'une formule de mise à jour naturelle, que nous introduirons ici comme une heuristique, mais à laquelle nous donnerons davantage de sens après que les notions de fonction duale et de méthode proximale auront été introduites (voir chapitre 14).

Le théorème 13.19 nous apprend que si l'on connaît un multiplicateur optimal $\bar{\lambda}$ et si l'on prend r assez grand, on a quelques chances de trouver une solution de (PEI, X_s) en minimisant le lagrangien augmenté $\ell_r(\cdot, \bar{\lambda})$. On ne connaît en général aucune de ces deux informations, si bien que c'est algorithmiquement qu'elles doivent être recherchées.

Supposons que l'on dispose au début de l'itération k , d'un facteur de pénalisation $r_k > 0$ et d'un multiplicateur approché $\lambda_k \in \mathbb{R}^m$. La mise à jour de ce dernier se fera par une formule qui fait suite aux considérations suivantes (cette formule est vue ici comme une heuristique). Le théorème 13.19 suggère de minimiser $\ell_{r_k}(\cdot, \lambda_k)$. Supposons

que ce problème ait une solution, que l'on note x_k . Par optimalité, $\nabla_x \ell_{r_k}(x_k, \lambda_k) = 0$, qui par (13.24) s'écrit

$$\nabla f(x_k) + \sum_{i \in E \cup I} \left(\lambda_k + r_k \tilde{c}_{\lambda_k, r_k}(x_k) \right)_i \nabla c_i(x_k) = 0,$$

où $\tilde{c}_{\lambda_k, r_k}$ est défini en (13.22). Comme, pour résoudre (P_{EI, X_s}) , on cherche à annuler le gradient du lagrangien de ce problème, les facteurs de $\nabla c_i(x_k)$ ci-dessus semblent être de bons candidats pour être la nouvelle approximation du multiplicateur optimal. On prend donc

$$\lambda_{k+1} = \lambda_k + r_k \tilde{c}_{\lambda_k, r_k}(x_k).$$

Dans certains cas, comme dans le schéma algorithmique 13.20 ci-dessous, on remplace r_k par un « pas » $\alpha_k > 0$. Donnons une autre expression de $\lambda_k + \alpha_k \tilde{c}_{\lambda_k, \alpha_k}(x_k)$, qui nous informera sur le signe de ses composantes correspondant aux inégalités. Pour $i \in I$ et en laissant tomber l'indice d'itération k , on a $\lambda_i + \alpha \tilde{c}_{\lambda, \alpha}(x)_i = \lambda_i + \alpha \max(-\lambda_i/\alpha, c_i(x)) = \max(0, \lambda_i + \alpha c_i(x))$. On en déduit que

$$\lambda_k + \alpha_k \tilde{c}_{\lambda_k, \alpha_k}(x_k) = (\lambda_k + \alpha_k c(x_k))^\#.$$

Les composantes d'indice $i \in I$ de λ_{k+1} sont donc positives.

On manque d'arguments solides pour trouver une règle de mise à jour du facteur de pénalisation r_k qui soit entièrement satisfaisante (le cas des fonctions quadratiques convexes est une exception [161, 122; 2005-2016]). On se contente en général d'augmenter r_k si $\tilde{c}_{\lambda_k, r_k}$ ne décroît pas suffisamment vite vers zéro. Cela paraît raisonnable, étant donné que r_k apparaît en facteur de la norme des contraintes dans le lagrangien augmenté (sous la forme (13.19)) et que c'est effectivement son rôle lorsque λ_k reste constant (on est alors proche de la pénalisation extérieure). Mais il ne faut pas oublier que λ_k joue un rôle tout aussi important que r_k pour forcer l'admissibilité des itérés. D'ailleurs dès que r_k est supérieur à un certain seuil, le bon réglage de λ_k devrait suffire pour obtenir l'admissibilité.

Nous donnons ci-dessous un schéma algorithmique inspiré de celui utilisé dans LANCELOT, un code d'optimisation généraliste qui est fondé sur le lagrangien augmenté [137; 1992]. Fletcher [212; 1987, page 292] propose un algorithme semblable.

Algorithme 13.20 (méthode des multiplicateurs) Une itération passe de l'itéré courant $(\lambda_k, r_k) \in \mathbb{R}^m \times \mathbb{R}_{++}$, vérifiant $(\lambda_k)_I \geq 0$, à l'itéré suivant $(\lambda_{k+1}, r_{k+1}) \in \mathbb{R}^m \times \mathbb{R}_{++}$, vérifiant $(\lambda_{k+1})_I \geq 0$, par les étapes suivantes.

1. *Nouvel itéré primal.* Avec ℓ_{r_k} défini en (13.21), calculer

$$x_k \in \arg \min \ell_{r_k}(\cdot, \lambda_k).$$

2. *Test de convergence.* Arrêt si $\tilde{c}_{\lambda_k, r_k}(x_k) \simeq 0$, où $\tilde{c}_{\mu, r}$ est défini en (13.22).
3. *Nouveau multiplicateur.* Choisir un pas $\alpha_k > 0$ et prendre

$$\lambda_{k+1} = \lambda_k + \alpha_k \tilde{c}_{\lambda_k, \alpha_k}(x_k) = (\lambda_k + \alpha_k c(x_k))^\#.$$

3. Mise à jour du paramètre de pénalisation r_k si nécessaire.

Ce schéma algorithmique mérite quelques éclaircissements.

- À l'étape 1, il est rare que l'on minimise complètement le lagrangien augmenté ; de toute façon, un test d'arrêt pour la minimisation de ce problème non linéaire doit être introduit, si le problème original est lui-même non linéaire. Une possibilité est d'utiliser le principe général selon lequel il ne faut pas être plus exigeant dans la minimisation *interne* (l'étape 1), que ce que l'algorithme a obtenu par la boucle *externe* (ensemble des étapes 1 à 4). Autrement dit, on peut utiliser un test qui compare les normes des quantités qui doivent tendre vers zéro dans les problèmes interne ($\nabla_x \ell_{r_k}(x_k, \lambda_k)$, voir (13.24)) et externe ($\tilde{c}_{\lambda_k, r_k}(x_k)$, voir (13.25)) ; l'algorithme se contente d'un x_k vérifiant

$$\|\nabla_x \ell_{r_k}(x_k, \lambda_k)\| \leq \sigma \|\tilde{c}_{\lambda_k, r_k}(x_k)\|, \quad (13.29)$$

où $\sigma > 0$ est une constante « bien choisie ». À notre connaissance, on n'a pas réussi jusqu'à présent (2021) à démontrer la convergence de l'algorithme avec un critère d'arrêt des itérations internes aussi simple que celui-là, même pour les problèmes convexes, sans pour autant avoir d'argument écartant l'opportunité d'un tel critère. Pour diverses contributions sur ce sujet, on pourra consulter [53, 136, 299, 135, 55, 176, 557, 558, 178, 179, 559, 177, 323, 175, 13, 187, 205, 188 ; 1982-2013].

- Le test d'arrêt de l'étape 2 est fondé sur le fait que l'on a bien optimalité au premier ordre si $\tilde{c}_{\lambda_k, r_k}(x_k) = 0$. En effet, l'optimalité à l'étape 1 et (13.24) impliquent que $\nabla_x \ell(x_k, \lambda_k) = 0$. Par ailleurs, $\tilde{c}_{\lambda_k, r_k}(x_k) = 0$ implique que

$$c_E(x_k) = 0 \quad \text{et} \quad \min \left(\frac{(\lambda_k)_I}{r_k}, -c_I(x_k) \right) = 0.$$

Cette dernière relation est équivalente aux relations de *complémentarité* $0 \leq (\lambda_k)_I \perp -c_I(x_k) \geq 0$, qui avec $c_E(x_k) = 0$, montrent que les autres conditions du système d'optimalité (4.33) de KKT sont vérifiées.

- À l'étape 3, on prend souvent le pas $\alpha_k = r_k$ comme le suggère la discussion qui précède.
- La seconde partie de la formule de mise à jour de λ_k à l'étape 3 montre que $(\lambda_{k+1})_I \geq 0$, si bien que l'algorithme se retrouve au début de l'itération suivante dans les mêmes conditions qu'à l'itération courante.
- La mise-à-jour du facteur de pénalisation r_k à l'étape 4 est une opération délicate, car on manque souvent de connaissance pour en déterminer une valeur correcte, si bien que les concepteurs de solveur utilisent des heuristiques variées. Par exemple, dans [13 ; 2007], r_k n'est mis à jour que lorsqu'au cours de l'itération précédente, une amélioration suffisante de l'admissibilité ou de la complémentarité n'est pas observée. Mentionnons le cas particulier de l'optimisation quadratique convexe, qui est bien mieux compris, pour lequel l'algorithme peut régler la valeur de r_k en fonction de la vitesse de convergence prescrite [161, 122].

Algorithme du lagrangien augmenté pour résoudre un problème convexe non réalisable. Lorsque le problème est convexe, l'algorithme du lagrangien augmenté est identique à l'algorithme proximal sur la fonction duale (proposition 14.39), si bien que son comportement sur des problèmes non réalisables peut se déduire de celui de l'algorithme proximal ; il est décrit dans [97, 510, 567 ; 1977-1987]. On montre que l'algorithme a un comportement maîtrisé :

- il trouve la plus petite translation $\bar{s}$, au sens de la norme ℓ_2 , rendant le problème réalisable,
- il minimise l'objectif sur les contraintes translatées par ce $\bar{s}$.

On peut être plus précis lorsque le problème est quadratique convexe [218 ; 1982, remarque 5.6] et [159 ; 2006] (pour les contraintes d'égalité seulement) et [122 ; 2016] (pour les contraintes d'inégalité) et montrer un résultat de convergence linéaire globale.

13.4.5 Méthode du lagrangien augmenté à directions alternées ▲

L'algorithme du *lagrangien augmenté à directions alternées*, souvent référencé par son sigle anglo-saxon ADMM abrégéant *Alternating Direction Method of Multipliers*, s'est montré sur le devant de la scène ces dernières années, du fait de son intérêt dans la résolution des *problèmes à données massives* (apprentissage approfondi, acquisition comprimée², traitement d'images, etc) [300, 85]. Dans ces problèmes aux milliards de variables, on ne cherche pas une solution précise, hors d'atteinte, mais l'accent est mis sur l'obtention *rapide* d'une solution *approchée*. L'algorithme ADMM convient à cette situation. Cet algorithme a été introduit dans les années 1970 pour résoudre certains types d'équations aux dérivées partielles [254, 226]. Il s'agit d'une méthode de lagrangien augmenté, dans laquelle la minimisation de ce dernier, à multiplicateur fixé, se fait de manière très approchée par un unique cycle de Gauss-Seidel (section 7.3.3). L'évocation de ce couple de mathématiciens suggère que le problème puisse être décrit par (au moins) deux groupes de variables, notées $x \in \mathbb{E}$ et $y \in \mathbb{F}$, des groupes parfois introduits artificiellement (par duplication de variables par exemple) de manière à pouvoir appliquer la méthode et bénéficier de ses caractéristiques.

L'algorithme ADMM est en réalité adapté à la résolution de problèmes qui peuvent s'écrire de la manière suivante :

$$\begin{cases} \inf_{(x,y)} f(x) + g(y) \\ Ax + By = c, \end{cases} \quad (13.30)$$

où $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ et $g : \mathbb{F} \rightarrow \mathbb{R} \cup \{+\infty\}$ sont des fonctions définies sur des espaces vectoriels $\mathbb{E}$ et $\mathbb{F}$ pouvant prendre la valeur $+\infty$ (c'est-à-dire pouvant prendre en compte implicitement des contraintes supplémentaires), $A \in \mathcal{L}(\mathbb{E}, \mathbb{G})$ et $B \in \mathcal{L}(\mathbb{F}, \mathbb{G})$ sont des applications linéaires à valeurs dans un espace euclidien $\mathbb{G}$, et $c \in \mathbb{G}$. Le lagrangien augmenté associé est la fonction $\ell_r : \mathbb{E} \times \mathbb{F} \times \mathbb{G} \rightarrow \mathbb{R}$ définie en $(x, y, \lambda) \in \mathbb{E} \times \mathbb{F} \times \mathbb{G}$ par

² Traduction française de *compressed sensing* [35]. Il s'agit de trouver des solutions (de systèmes linéaires sous-déterminés par exemple) avec le plus de zéros. On exprime ce problème en utilisant la norme $\ell_1 : \inf\{\|x\|_1 : x \in X\}$.

$$\ell_r(x, y, \lambda) = f(x) + g(y) + \langle \lambda, Ax + By - c \rangle + \frac{r}{2} \|Ax + By - c\|^2,$$

où $r > 0$ est le facteur d'augmentation, $\langle \cdot, \cdot \rangle$ est le produit scalaire de $\mathbb{G}$ et $\|\cdot\|$ est la norme associée à ce dernier.

Algorithme 13.21 (ADMM) Une itération passe de l'itéré courant $(y_k, \lambda_k) \in \mathbb{F} \times \mathbb{G}$ à l'itéré suivant (y_{k+1}, λ_{k+1}) par les étapes ci-dessous :

$$x_{k+1} \in \arg \min_{x \in \mathbb{E}} \ell_r(x, y_k, \lambda_k) \quad (13.31)$$

$$y_{k+1} \in \arg \min_{y \in \mathbb{F}} \ell_r(x_{k+1}, y, \lambda_k) \quad (13.32)$$

$$\lambda_{k+1} := \lambda_k + r(Ax_{k+1} + By_{k+1} - c). \quad (13.33)$$

Comme annoncé, l'algorithme ressemble très fort à l'algorithme du lagrangien augmenté 13.20, si ce n'est que la minimisation complète de $\ell_r(\cdot, \cdot, \lambda_k)$ est remplacée par un cycle gauss-seidien (section 7.3.3) : une première minimisation en x dans (13.31), avec y fixé à y_k , suivie d'une minimisation en y dans (13.32), avec x fixé à x_{k+1} ; c'est tout, alors que l'algorithme de Gauss-Seidel itérerait ce cycle jusqu'à minimisation complète du lagrangien augmenté. C'est à cette double minimisation que l'on doit le vocable *directions alternées* utilisé dans le nom de l'algorithme. On notera que x_k est une variable auxiliaire qui ne doit pas exister au début de l'itération. Par ailleurs, cet algorithme n'est pas une méthode de dualité (chapitre 14), comme l'est l'algorithme du lagrangien augmenté, car l'itéré contient la variable primale y_k .

On peut montrer que la convergence du coût est en $O(1/k)$ [307, 430]. Des efforts importants ont été faits pour améliorer cette vitesse de convergence du coût, pour qu'elle soit en $O(1/k^2)$, notamment par le schéma de Nesterov [453, 264].

13.5 Pénalisation exacte non différentiable ▲

Le lagrangien augmenté étudié à la section 13.4 est une première manière d'obtenir une pénalisation exacte (voir la définition 13.1), pourvu que l'on connaisse un multiplicateur optimal (comme ce n'est généralement pas le cas, la méthode des multiplicateurs – l'algorithme 13.20 – approche celui-ci de manière itérative). L'idée sous-jacente est de pénaliser quadratiquement une fonction dont la dérivée est nulle en la solution, le lagrangien $\ell(\cdot, \bar{\lambda})$.

Une autre manière d'obtenir une pénalisation exacte est de le faire au moyen d'une fonction pénalisante non différentiable. Illustrons cela sur l'exemple 13.4. La figure 13.5 montre que si l'on ajoute le terme rx^+ à la fonction $1 - x - \frac{1}{3}x^3$, on obtient une fonction de pénalisation exacte dès que $r > 1$. Ce seuil $r = 1$ vient bien sûr ici de la pente de f en zéro et en toute généralité de la pente de la fonction valeur en zéro, si bien que c'est à nouveau le multiplicateur optimal $\bar{\lambda}$ qui jouera un rôle-clé dans la détermination de ce seuil.

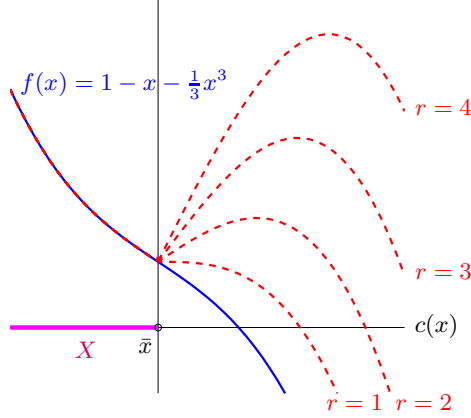


Fig. 13.5. Pénalisation non différentiable du problème (13.5) avec $r = 1, 2, 3$ et 4 .

De manière plus générale, si l'on enlève les carrés aux fonctions pénalisantes de la table 13.1, on obtient des fonctions de pénalisation exactes : si r est pris assez grand, les solutions *locales* de (P_X) sont solutions *locales* du problème pénalisé. Cela semble très intéressant, puisque l'on remplace un problème d'optimisation avec contraintes par un *unique* problème d'optimisation sans contrainte. On ne peut cependant pas gagner sur tous les plans : si la fonction de pénalisation est exacte, elle est aussi non différentiable et donc plus délicate à minimiser. La minimisation de ces fonctions se fait en général de manière détournée : on obtient une direction de descente en résolvant un problème auxiliaire et non pas en calculant quelque chose ressemblant à un gradient. Nous verrons au chapitre 15 une utilisation de cette idée.

Dans cette section, on considère le problème d'optimisation avec contraintes explicites

$$(P_{EI}) \quad \begin{cases} \inf f(x) \\ c_i(x) = 0, & i \in E \\ c_i(x) \leq 0, & i \in I, \end{cases}$$

décrit dans l'introduction de ce chapitre, auquel on associe la fonction de pénalisation suivante :

$$\Theta_r(x) = f(x) + r\|c(x)^\# \|_p, \quad (13.34)$$

où $r > 0$ et $\|\cdot\|_p$ est une norme quelconque. La première pénalisation de ce type a été introduite par Ablow et Brigham [2 ; 1955], avec la norme ℓ_1 . Le résultat fondamental est donné à la proposition 13.24, qui énonce des conditions pour que Θ_r soit une fonction de pénalisation exacte.

Sa démonstration utilise les deux lemmes suivants. Le premier étudie la différentiabilité de Θ_r et utilise l'opérateur $P_v : \mathbb{R}^m \rightarrow \mathbb{R}^m$, défini pour u et $v \leq 0$ dans $\mathbb{R}^m$ par

$$(P_v u)_i = \begin{cases} u_i & \text{si } i \in E \\ u_i^+ & \text{si } i \in I \text{ et } v_i = 0 \\ 0 & \text{si } i \in I \text{ et } v_i < 0. \end{cases}$$

Lemme 13.22 *Si f et c admettent des dérivées directionnelles en un point $x \in \mathbb{E}$, alors Θ_r admet des dérivées directionnelles en x et si $c(x)^\# = 0$, on a*

$$\Theta'_r(x; d) = f'(x; d) + r \|P_{c(x)} c'(x; d)\|_{\mathbb{P}}.$$

DÉMONSTRATION. Le fait que Θ_r admette des dérivées directionnelles en x vient de ce que $(\cdot)^\#$ et $\|\cdot\|_{\mathbb{P}}$ sont lipschitziennes et admettent, ainsi que f et c , des dérivées directionnelles en x (proposition C.3).

Si $c(x)^\# = 0$, on a

$$\Theta'_r(x; d) = f'(x; d) + r(\|\cdot\|_{\mathbb{P}})'(0; (c^\#)'(x; d)).$$

Mais

$$(\|\cdot\|_{\mathbb{P}})'(0; v) = \lim_{t \downarrow 0} \frac{1}{t} \|tv\|_{\mathbb{P}} = \|v\|_{\mathbb{P}}$$

et

$$(c^\#)'(x; d) = (\cdot^\#)'(c(x); c'(x; d)) = P_{c(x)} c'(x; d).$$

On en déduit le résultat. $\square$

Le second lemme montre qu'avec un facteur de pénalisation suffisamment grand, Θ_r domine le *lagrangien ordinaire* $\ell(\cdot, \bar{\lambda})$ sur $\mathbb{E}$. On rappelle de (A.8) que la *norme duale* de $\|\cdot\|_{\mathbb{P}}$ pour le produit scalaire euclidien est la norme $\|\cdot\|_{\mathbb{D}}$ définie par

$$\|y\|_{\mathbb{D}} = \sup_{\|x\|_{\mathbb{P}} \leq 1} y^{\mathsf{T}} x.$$

Lemme 13.23 *Si $\lambda \in \mathbb{R}^m$ vérifie $r \geq \|\lambda\|_{\mathbb{D}}$ et $\lambda_I \geq 0$, alors pour tout $x \in \mathbb{E}$ on a $\ell(x, \lambda) \leq \Theta_r(x)$.*

DÉMONSTRATION. Comme $\lambda_I \geq 0$, on a

$$\ell(x, \lambda) \leq f(x) + \lambda^{\mathsf{T}} c(x)^\#.$$

On en déduit que

$$\ell(x, \lambda) \leq f(x) + \|\lambda\|_{\mathbb{D}} \|c(x)^\#\|_{\mathbb{P}} \leq f(x) + r \|c(x)^\#\|_{\mathbb{P}} = \Theta_r(x). \quad \square$$

Voici le résultat annoncé donnant des conditions suffisantes pour que Θ_r soit une fonction de pénalisation exacte en $\bar{x}$. L'hypothèse (13.35) ne peut avoir lieu que si l'ensemble des multiplicateurs optimaux est borné, ce qui revient à dire que la condition suffisante de qualification des contraintes de Mangasarian-Fromovitz (QC-MF) (définition 4.41) a lieu.

Proposition 13.24 (exactitude de Θ_r) Soit $\bar{x}$ un minimum local du problème (P_{EI}) . On suppose que f et c sont deux fois dérivables en $\bar{x}$ et lipschitziennes dans un voisinage de $\bar{x}$. On suppose également que $\bar{x}$ vérifie les conditions suffisantes d'optimalité du second ordre faible (4.59). On suppose enfin que

$$r \geq \sup_{\bar{\lambda} \in \Lambda(\bar{x})} \|\bar{\lambda}\|_{\mathcal{D}} \quad \text{et} \quad r > \|\hat{\lambda}\|_{\mathcal{D}}, \text{ pour un } \hat{\lambda} \in \Lambda(\bar{x}), \quad (13.35)$$

où $\Lambda(\bar{x})$ est l'ensemble non vide des multiplicateurs optimaux associés à $\bar{x}$. Alors, $\bar{x}$ est un minimum local strict de la fonction de pénalisation Θ_r donnée en (13.34).

DÉMONSTRATION. On raisonne par l'absurde en supposant que $\bar{x}$ n'est pas un minimum local strict de Θ_r . Alors, il existe une suite $\{x_k\}$ telle que $x_k \neq \bar{x}$, $x_k \rightarrow \bar{x}$ et

$$\forall k \geq 1 : \quad \Theta_r(x_k) \leq \Theta_r(\bar{x}). \quad (13.36)$$

1) *Construction d'une direction critique non nulle.* Comme la suite $\{(x_k - \bar{x})/\|x_k - \bar{x}\|\}$ est bornée ($\|\cdot\|$ est une norme arbitraire), on peut en extraire une sous-suite convergente : $(x_k - \bar{x})/\|x_k - \bar{x}\| \rightarrow d$, où $\|d\| = 1$. En notant $\alpha_k = \|x_k - \bar{x}\|$, on a

$$x_k = \bar{x} + \alpha_k d + o(\alpha_k).$$

Montrons que d est la direction critique recherchée.

- Comme Θ_r est lipschitzienne dans un voisinage de $\bar{x}$, on a

$$\Theta_r(x_k) = \Theta_r(\bar{x} + \alpha_k d) + o(\alpha_k).$$

Alors (13.36) montre que $\Theta'_r(\bar{x}; d) \leq 0$. Alors, grâce au lemme 13.22, on peut écrire

$$f'(\bar{x}) \cdot d + r \|P_{c(\bar{x})}(c'(\bar{x}) \cdot d)\|_{\mathcal{P}} \leq 0. \quad (13.37)$$

On a donc certainement

$$f'(\bar{x}) \cdot d \leq 0. \quad (13.38)$$

- D'autre part, d'après les hypothèses, on peut trouver un multiplicateur optimal $\hat{\lambda}$ tel que $r > \|\hat{\lambda}\|_{\mathcal{D}}$. En utilisant les conditions d'optimalité du premier ordre, notamment la positivité de $\hat{\lambda}_I$ et la complémentarité $\hat{\lambda}_I^T c_I(x_*) = 0$, on a

$$\begin{aligned} f'(\bar{x}) \cdot d &= - \sum_{i \in E \cup I} \hat{\lambda}_i c'_i(\bar{x}) \cdot d \\ &\geq - \hat{\lambda}^T P_{c(\bar{x})}(c'(\bar{x}) \cdot d) \\ &\geq - \|\hat{\lambda}\|_{\mathcal{D}} \|P_{c(\bar{x})}(c'(\bar{x}) \cdot d)\|_{\mathcal{P}}. \end{aligned}$$

Alors (13.37) implique que $P_{c(\bar{x})}(c'(\bar{x}) \cdot d) = 0$, c'est-à-dire

$$\begin{cases} c'_i(\bar{x}) \cdot d = 0 & \text{pour } i \in E \\ c'_i(\bar{x}) \cdot d \leq 0 & \text{pour } i \in I^0(\bar{x}). \end{cases}$$

Ces relations et (13.38) montrent que d est une direction critique non nulle.

Soit à présent $\bar{\lambda}$ le multiplicateur dépendant de d , déterminé par (4.59). D'après la proposition 4.49, on a

$$d^T \nabla_{xx}^2 \ell(\bar{x}, \bar{\lambda}) d > 0.$$

Alors le développement suivant (on utilise le fait que $\nabla_x \ell(\bar{x}, \bar{\lambda}) = 0$)

$$\ell(x_k, \bar{\lambda}) = \ell(\bar{x}, \bar{\lambda}) + \frac{\alpha_k^2}{2} d^T \nabla_{xx}^2 \ell(\bar{x}, \bar{\lambda}) d + o(\alpha_k^2)$$

permet de voir que, pour k suffisamment grand, on a

$$\ell(x_k, \bar{\lambda}) > \ell(\bar{x}, \bar{\lambda}). \quad (13.39)$$

On conclut en obtenant une contradiction. Pour k grand, on a

$$\begin{aligned} \Theta_r(x_k) &\leq \Theta_r(\bar{x}) && [\text{par (13.36)}] \\ &= f(\bar{x}) \\ &= \ell(\bar{x}, \bar{\lambda}) \\ &< \ell(x_k, \bar{\lambda}) && [\text{par (13.39)}] \\ &\leq \Theta_r(x_k) && [\text{par le lemme 13.23}], \end{aligned}$$

ce qui est absurde. □

Notes

Origine du lagrangien augmenté. Le lagrangien augmenté (13.18) associé aux problèmes avec contraintes d'égalité remonte au moins à Arrow et Solow [24; 1958] qui l'utilisent pour localiser des points-selles au moyen d'équations différentielles modifiant les variables primales et duales simultanément. Il a été redécouvert et utilisé indépendamment par Hestenes [311; 1969] et Powell [486; 1969] dans le même contexte algorithmique qu'ici, celui de la méthode des multiplicateurs du schéma 13.20, qui modifie les variables primales et duales séquentiellement. Le lagrangien augmenté (13.21) permettant de prendre en compte des contraintes d'inégalité est attribué à Rockafellar [518, 520, 521, 524; 1971-76]; voir aussi Buys [100; 1972] et Arrow, Gould et Howe [23; 1973]. On pourra consulter la revue de Bertsekas [51; 1976] pour plus de références bibliographiques avant 1976.

Lagrangiens augmentés plus réguliers. Même en l'absence de complémentarité stricte, on peut obtenir des lagrangiens augmentés plus réguliers que celui (13.19) de la section 13.4.3 (voir ce qu'il en est dit à la proposition 13.18), si le terme de pénalisation quadratique est remplacé par d'autres fonctions (*divergence de Bregman*, *φ -divergence*, etc); sur ces questions, on pourra consulter [111, 576, 186, 30, 31, 551; 1992-2006]. Il n'est pas clair, cependant, qu'une telle fonction plus régulière soit intéressante numériquement [59, 187].

Algorithme du lagrangien augmenté pour résoudre un problème quadratique convexe. La proposition ?? étend un résultat de [252, 253, 218; 1976-1982] au cas où H n'est définie positive que dans le noyau de A ; la démonstration est identique. La

convergence linéaire globale de l'algorithme du lagrangien augmenté (ou méthode des multiplicateurs) sur les problèmes quadratiques convexes (éventuellement non réalisables) est étudiée dans [161, 122].

Minimisation inexacte du lagrangien augmenté. On gagne en efficacité en minimisant de manière approchée le lagrangien augmenté à l'étape 1 de la méthode des multiplicateurs (schéma 13.20), avec une précision qui dépend de la proximité de la solution. Pour une étude locale dans le cas de l'optimisation non linéaire, on pourra lire [205].

Pénalisation exacte. Nous avons montré que l'on peut obtenir des fonctions de pénalisation qui sont exactes en des points vérifiant les conditions d'optimalité du second ordre (proposition 13.24). On peut suivre la démarche inverse et retrouver les conditions d'optimalité à partir du concept de pénalisation exacte. Voir [99, 640].

Exercices

13.1. *Pénalisation quadratique.* Soit $\Theta_r : \mathbb{E} \rightarrow \mathbb{R}$ la fonction de pénalisation quadratique définie en (13.12).

1) Montrez qu'en $x \in \mathbb{E}$:

$$\Theta_r(x) = \inf_{s \in \mathbb{R}_+^m} f(x) + \frac{r}{2} \|c_E(x)\|_2^2 + \frac{r}{2} \|c_I(x) + s\|_2^2.$$

Remarque. La fonction Θ_r s'obtient donc en remplaçant les contraintes $c(x)^\# = 0$ de (P_{EI}) par les contraintes équivalentes $(c_E(x) = 0, c_I(x) + s = 0 \text{ et } s \geq 0)$ et en prenant une pénalisation quadratique des contraintes d'égalité tout en gardant explicite les contraintes de positivité $s \geq 0$.

2) Montrez que, si f et c définissant (P_{EI}) sont différentiables en x , alors

$$\nabla \Theta_r(x) = \nabla f(x) + \sum_{i \in E \cup I} r[c(x)]_i^\# \nabla c_i(x).$$

13.2. *Conditionnement limite de la hessienne du lagrangien augmenté.* Soient H une matrice d'ordre n symétrique et A une matrice de type $m \times n$ telles que $\sigma := \inf\{v^\top H v : v \in \mathcal{N}(A), \|v\|_2 = 1\} > 0$. On note $\kappa_2(H_r)$ le conditionnement ℓ_2 de la matrice $H_r := H + rA^\top A$. Montrez que

$$\lim_{r \rightarrow +\infty} \frac{\kappa_2(H_r)}{r} = \frac{\|A\|_2^2}{\sigma}.$$

13.3. *Pénalisation exacte ℓ_1 .* On suppose que f et c sont régulières dans un voisinage d'une solution $\bar{x}$ de (P_{EI}) . On se donne des scalaires positifs $r_i, i \in E \cup I$, et on considère la fonction de pénalisation suivante

$$\Theta_r^1(x) = f(x) + \sum_{i \in E} r_i |c_i(x)| + \sum_{i \in I} r_i c_i(x)^+.$$

On suppose également qu'il existe un multiplicateur $\bar{\lambda} = \{\bar{\lambda}_i\}_{i \in E \cup I}$ tel que $(\bar{x}, \bar{\lambda})$ vérifie les conditions suffisantes d'optimalité du deuxième ordre. Montrez que si

$$r_i > |\bar{\lambda}_i|, \quad \text{pour } i \in E \cup I,$$

alors $\bar{x}$ est un minimum local strict de Θ_r^1 .

- 13.4.** *Lagrangien augmenté non différentiable* [70; 1989]. On suppose que f et c sont régulières dans un voisinage d'une solution $\bar{x}$ du problème $\inf\{f(x) : c(x) = 0\}$. On suppose également qu'il existe un multiplicateur $\bar{\lambda}$ tel que $(\bar{x}, \bar{\lambda})$ vérifie les conditions suffisantes d'optimalité du deuxième ordre de ce problème. Montrez que

$$\Theta_{\mu,r}(x) = f(x) + \mu^\top c(x) + r\|c(x)\|_p$$

est une fonction de pénalisation exacte du problème d'optimisation considéré si $r > \|\bar{\lambda} - \mu\|_p$.

- 13.5.** *Courbes de niveaux de fonctions de pénalisation.* On considère un problème de minimisation sur $\mathbb{R}^2$

$$\begin{cases} \inf f(x) \\ x \in \mathbb{R}^2 \\ c(x) = 0, \end{cases}$$

en présence d'une unique contrainte d'égalité. Les fonctions f et $c : \mathbb{R}^2 \rightarrow \mathbb{R}$ sont supposées régulières. La figure 13.6 donne les tracés des courbes de niveau de trois

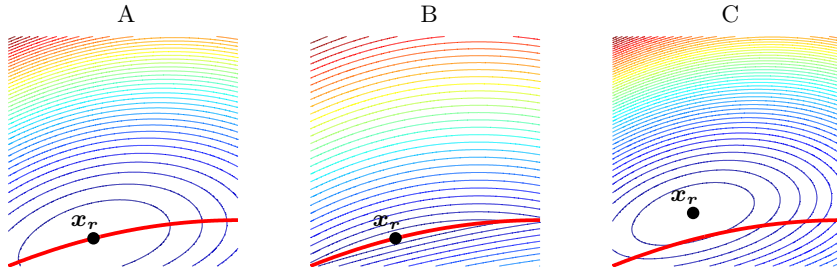


Fig. 13.6. Courbes de niveaux de 3 fonctions de pénalisation

fonctions de pénalisation associées au problème ci-dessus. L'ensemble admissible du problème est représenté dans chaque tracé par la courbe en trait large du bas. Déterminez pour chaque tracé A, B et C, laquelle des fonctions de pénalisation Θ_r^1 , Θ_r^2 ou Θ_r^3 données ci-après qui a été utilisée pour dessiner ces courbes de niveau, sachant que toutes les trois ont été utilisées. Dans celles-ci r est un scalaire strictement positif et λ_* est le multiplicateur optimal du problème. Le minimum de ces fonctions est repéré par le point x_r dans chaque dessin.

$$\begin{aligned} \Theta_r^1(x) &= f(x) + \frac{r}{2}\|c(x)\|_2^2 \\ \Theta_r^2(x) &= f(x) + \lambda_*^\top c(x) + \frac{r}{2}\|c(x)\|_2^2 \\ \Theta_r^3(x) &= f(x) + r\|c(x)\|_2 \end{aligned}$$