

Méthodes de points intérieurs

At the same time, we find that the nature of the [interior-point] methods, is in fact, independent of the specific properties of linear programming problems, so that these methods can be extended onto more general convex programs.

Y.E. NESTEROV, A.S. NEMIROVSKII (1994). [\[456\]](#).

16.1 Méthodes de points intérieurs en optimisation convexe

16.1.1 Vue d'ensemble

Soit \mathbb{E} un espace vectoriel de dimension finie (≥ 1), muni d'un produit scalaire $\langle \cdot, \cdot \rangle$, dit de référence, et de sa norme associée $\| \cdot \|$. Dans cette section, nous nous intéressons à la résolution numérique de problèmes d'optimisation convexes, que l'on supposera écrits sous la forme

$$(P) \quad \left\{ \begin{array}{l} \inf \langle c, x \rangle \\ x \in X. \end{array} \right. \quad (16.1)$$

Ici $c \in \mathbb{E}$ et X est un convexe de \mathbb{E} . Le fait que le critère soit linéaire n'introduit pas de perte de généralité, car on peut toujours *faire passer en contrainte* un critère, qui serait ici convexe, comme à la proposition 1.13.

La théorie que nous présentons ci-dessous, principalement développée par Nesterov et Nemirovskii [\[455 ; 1993\]](#), peut être vue comme une généralisation des méthodes de points intérieurs que nous verrons sous des formes un peu différentes dans les cas particuliers de l'optimisation linéaire (chapitre 18) et de l'optimisation semi-définie positive (chapitre 20). Chronologiquement, elle a été construite après la découverte de ces dernières, dans un but évident de généralisation et d'extension de son domaine d'application. C'est aussi pour cette raison que nous la présentons dans cette partie décrivant les méthodes générales de l'optimisation avec contraintes. Le lecteur pourra d'ailleurs étudier les chapitres 18 et 20, avant d'aborder cette section, mais rien n'impose un tel ordre de lecture.

Cette théorie est remarquable, puisqu'elle conduit au résultat impressionnant selon lequel tout problème d'optimisation convexe, écrit sous la forme ci-dessus, peut être résolu avec une précision arbitraire donnée en un nombre polynomial d'itérations ne

dépendant de la précision demandée que par son logarithme. La dépendance polynomiale ne se fait pas directement par rapport à la dimension de \mathbb{E} , car la structure de l'ensemble admissible X intervient également, mais par l'intermédiaire d'un module de complexité ϑ_f , dont la dépendance par rapport à $\dim \mathbb{E}$ pourra parfois être explicitée. Cela paraît très attrayant. Il faut toutefois que l'on dispose d'une barrière autoconcordante f dont le [domaine](#) est l'intérieur de l'ensemble admissible X pour que ce résultat ait lieu. Même si l'on peut montrer qu'une telle fonction existe toujours, on ne dispose pas nécessairement d'une forme analytique de celle-ci, si bien que le résultat peut ne pas avoir une utilité pratique immédiate. On connaît cependant de grandes classes de problèmes auxquelles sont associées des fonctions autoconcordantes : l'optimisation linéaire ([chapitre 18](#)) et l'optimisation SDP ([chapitre 20](#)). Mais on connaît aussi des problèmes convexes qui sont NP-ardus, les problèmes d'*optimisation copositive* [69; 2012] ; donc pour lesquels l'évaluation d'une barrière auto-concordante ne pourra se faire en temps polynomial que si $P = NP$.

Nous introduisons d'abord, à la section [16.1.2](#), la notion de fonction auto-concordante et en donnons les propriétés utiles à l'étude des algorithmes décrits. Nous présentons ensuite, à la section [16.1.3](#), les barrières autoconcordantes, qui sont des fonctions autoconcordantes pour lesquelles la direction de Newton est bornée (pour une métrique particulière). Ce sont ces fonctions qui permettent d'obtenir des résultats de complexité itérative polynomiale des méthodes de points intérieurs. Nous concluons la section par trois algorithmes typiques et en démontrons leur complexité polynomiale ([section 16.1.4](#)). On peut accéder à l'étude d'un algorithme polynomialement convergent avec un minimum d'effort préalable : il suffit de définir les notions de fonction autoconcordante ([définition 16.1](#) et définitions équivalentes de la [proposition 16.5](#)) et de barrière autoconcordante ([définition 16.18](#)), d'étudier les [propositions 16.8](#) et [16.11](#) et d'établir l'inégalité de droite dans ([16.15](#)) et celle de la [proposition 16.28](#); on a alors toute l'information nécessaire à la compréhension et à l'analyse du premier algorithme présenté à la [section 16.1.4](#).

La seconde partie de ce chapitre s'intéresse au cas des problèmes non linéaires et non convexes. Il n'y a alors plus de résultat de complexité polynomiale. Après avoir présenté un schéma algorithmique parfois implémenté, nous en étudions le comportement asymptotique.

16.1.2 Fonction autoconcordante

On verra aux chapitres [18](#) et [20](#), que la fonction logarithmique joue un rôle important dans la conception des algorithmes de points intérieurs et dans l'étude de leur convergence polynomiale. Les fonctions autoconcordantes sont celles qui ont les propriétés du logarithme qui y jouent un rôle-clé, avec l'intérêt supplémentaire de pouvoir être définie sur des ouverts convexes plus généraux que l'intérieur de l'[orthant positif](#) (exemple [16.2](#)) ou du cône des matrices [semi-définies positives](#) (exemple [16.3](#)).

Les fonctions autoconcordantes $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$, dont l'adhérence du [domaine](#) définit l'ensemble admissible de (P) , devront au moins vérifier les propriétés suivantes :

$$\begin{cases} \text{dom } f \text{ est un ouvert convexe non vide,} \\ f \text{ est } C^2 \text{ sur son } \text{domaine} \text{ et} \\ \nabla^2 f(x) \succ 0 \text{ pour tout } x \in \text{dom } f. \end{cases} \quad (16.2)$$

Évidemment, de telles fonctions sont strictement convexes. On note

$$g(x) := \nabla f(x) \quad \text{et} \quad H(x) := \nabla^2 f(x)$$

le gradient et hessienne de f en $x \in \text{dom } f$ pour le produit scalaire de référence.

Métrique riemannienne associée à une fonction vérifiant (16.2)

On peut introduire sur le **domaine** d'une fonction $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ vérifiant (16.2) une *métrique riemannienne* en se donnant en tout point $x \in \text{dom } f$ un produit scalaire et sa norme associée

$$\langle u, v \rangle_x := f''(x) \cdot (u, v) = \langle H(x)u, v \rangle \quad \text{et} \quad \|v\|_x := \langle v, v \rangle_x^{1/2}.$$

Cette métrique ne dépend pas du produit scalaire de référence, puisqu'elle est définie à partir de la donnée de f seule. On utilise aussi la notation $\|\cdot\|_x$ pour la norme d'un opérateur linéaire sur \mathbb{E} , **subordonnée à la norme vectorielle** définie ci-dessus. On note

$$\begin{aligned} B_x(y, 1) &:= \{y + v \in \mathbb{E} : \langle H(x)v, v \rangle < 1\} \\ \bar{B}_x(y, 1) &:= \{y + v \in \mathbb{E} : \langle H(x)v, v \rangle \leq 1\} \end{aligned}$$

les boules ouverte et fermée de centre $y \in \mathbb{E}$ et de rayon 1 pour la métrique $\langle \cdot, \cdot \rangle_x$. On peut les voir comme des *ellipsoïdes* centrés en y .

Les gradient et hessienne de f pour la métrique $\langle \cdot, \cdot \rangle_x$ sont notés g_x et H_x et valent en $y \in \text{dom } f$:

$$g_x(y) = H(x)^{-1}g(y) \quad \text{et} \quad H_x(y) = H(x)^{-1}H(y).$$

En particulier $g_x(x) = -d(x)$ est l'opposé de la direction de Newton en x pour minimiser f et $H_x(x) = I$. On notera que $H_x(y)$ est auto-adjoint pour le produit scalaire $\langle \cdot, \cdot \rangle_x$.

Il est clair que, pour v fixé, $\|v\|_x$ varie continûment avec x (car f est C^2), mais on a un peu mieux: *pour tout $\varepsilon \in]0, 1[$, il existe un voisinage V de $x \in \text{dom } f$, tel que pour tout $y \in V$ et tout $v \in \mathbb{E}$ non nul, on a*

$$1 - \varepsilon \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1 - \varepsilon}. \quad (16.3)$$

En effet, pour V suffisamment petit, on a

$$\begin{aligned} \|v\|_y^2 &= \|v\|_x^2 + \langle (H(x)^{-1/2}H(y)H(x)^{-1/2} - I)H(x)^{1/2}v, H(x)^{1/2}v \rangle \\ &\leq \|v\|_x^2 + \|H(x)^{-1/2}H(y)H(x)^{-1/2} - I\| \|v\|_x^2 \\ &\leq (1 + \varepsilon) \|v\|_x^2 \\ &\leq (1 - \varepsilon)^{-2} \|v\|_x^2, \end{aligned}$$

car $1 + \varepsilon \leq (1 + \varepsilon)^2 \leq (1 - \varepsilon)^{-2}$. On en déduit l'inégalité de droite dans (16.3). En inversant x et y , on a

$$\|v\|_x^2 \leq \|v\|_y^2 + \|H(y)^{-1/2}H(x)H(y)^{-1/2} - I\| \|v\|_y^2 \leq (1 - \varepsilon)^{-2} \|v\|_y^2.$$

On en déduit l'inégalité de gauche dans (16.3).

On aurait pu mettre $1 + \varepsilon$ à droite dans (16.3) au lieu de $(1 - \varepsilon)^{-1}$. Nous avons choisi ce dernier majorant car cela nous permet d'introduire de façon un peu plus naturelle la notion de fonction autoconcordante, laquelle vérifie cette propriété avec $V = B_x(x, 1)$ et $\varepsilon = \|y - x\|_x$.

Fonction autoconcordante

Définition 16.1 On dit qu'une fonction $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ est *autoconcordante*, si elle vérifie les conditions (16.2) et si les deux propriétés suivantes ont lieu :

- (AC1) $\forall x \in \text{dom } f, B_x(x, 1) \subseteq \text{dom } f,$
- (AC2) $\forall x \in \text{dom } f, \forall y \in B_x(x, 1), \forall v \in \mathbb{E} \setminus \{0\},$ on a

$$1 - \|y - x\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1 - \|y - x\|_x}. \quad (16.4)$$

On note $\text{AC}(\mathbb{E})$ l'ensemble des fonctions autoconcordantes sur \mathbb{E} . \square

Les inégalités (16.4) sont remarquables si on les compare à (16.3) : aucune constante indéterminée n'est utilisée et le voisinage $B_x(x, 1)$ sur lequel elles ont lieu est bien précisé. Cette absence de constante inconnue est une caractéristique que l'on s'attachera à préserver dans toutes les propriétés des fonctions autoconcordantes et qui contribuera à l'obtention des résultats de polynomialité.

La condition (AC1) donne à une fonction autoconcordante f une allure proche de la frontière de son *domaine* qui rappelle celle de l'opposé du logarithme au voisinage de 0 (proposition 16.4). En particulier, $f(x)$ tend vers $+\infty$ lorsque x tend vers un point de $\partial(\text{dom } f)$. La condition (AC2) exprime à sa manière le caractère lipschitzien de la hessienne de f (proposition 16.5). Cette propriété permet en particulier de comparer deux directions de Newton successives (proposition 16.11), ce qui sera important dans le contrôle des itérés dans les méthodes de points intérieurs.

Les fonctions autoconcordantes existent ! En voici deux exemples emblématiques, qui seront respectivement utilisées en optimisation linéaire (chapitre 18) et en optimisation semi-définie positive (chapitre 20). Le logarithme y joue un rôle important. On peut d'ailleurs voir l'autoconcordante comme un moyen d'introduire sur un ensemble convexe quelconque une fonction ayant les propriétés particulières de l'opposé du logarithme sur \mathbb{R}_{++} , qui sont décisives pour l'obtention des résultats de polynomialité.

Exemple 16.2 La fonction *log-barrière* $\text{lb} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ définie en $x \in \mathbb{R}^n$ par

$$\text{lb}(x) = \begin{cases} -\sum_{i=1}^n \log x_i & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (16.5)$$

est autoconcordante (voir l'exercice 16.1 pour une définition géométrique de cette fonction).

DÉMONSTRATION. On munit \mathbb{R}^n du produit scalaire euclidien et on note $X = \text{Diag}(x_1, \dots, x_n)$ et e le vecteur dont les composantes valent 1. Alors, pour $x \in \text{dom lb} = \mathbb{R}_{++}^n$, $g(x) = -X^{-1}e$ et $H(x) = X^{-2}$. La fonction lb vérifie donc bien les hypothèses (16.2).

Ensuite, si $x > 0$ et $y \in B_x(x, 1)$, on a $\sum_i (y_i - x_i)^2 / x_i^2 < 1$, ce qui implique que $|y_i - x_i| < x_i$ pour tout i , ou encore $y_i > 0$; donc $y \in \text{dom lb}$.

D'autre part, pour x et $y \in \text{dom lb}$,

$$\|v\|_y^2 = \sum_{i=1}^n \left(\frac{v_i}{y_i} \right)^2 = \sum_{i=1}^n \left(\frac{v_i}{x_i} \right)^2 \left(\frac{x_i}{y_i} \right)^2 \leq \|v\|_x^2 \max_{1 \leq i \leq n} \left(\frac{x_i}{y_i} \right)^2.$$

Mais

$$\frac{y_i}{x_i} \geqslant 1 - \left| \frac{y_i}{x_i} - 1 \right| \geqslant 1 - \|y - x\|_x > 0.$$

On en déduit l'inégalité à droite dans (16.4). Pour obtenir l'inégalité de gauche, on écrit

$$\|v\|_y^2 \geqslant \|v\|_x^2 \min_{1 \leqslant i \leqslant n} \left(\frac{x_i}{y_i} \right)^2$$

et on observe que

$$\frac{y_i}{x_i} \leqslant 1 + \left| \frac{y_i}{x_i} - 1 \right| \leqslant 1 + \|y - x\|_x \leqslant \frac{1}{1 - \|y - x\|_x}.$$

□

Exemple 16.3 La fonction log-déterminant $\text{ld} : \mathcal{S}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ définie en $X \in \mathcal{S}^n$ par

$$\text{ld}(X) = \begin{cases} -\log \det X & \text{si } X \in \mathcal{S}_{++}^n \\ +\infty & \text{sinon} \end{cases} \quad (16.6)$$

est autoconcordante.

DÉMONSTRATION. On munit \mathcal{S}^n du produit scalaire $\langle A, B \rangle = \text{tr } AB$. Le domaine de ld est \mathcal{S}_{++}^n . C'est bien un ouvert et sur celui-ci, ld est C^∞ . En se rappelant qu'en X inversible $\det'(X) \cdot H = (\det X)(\text{tr } X^{-1}H)$, on trouve que $\text{ld}'(X) \cdot H = -\langle X^{-1}, H \rangle$ et que $\text{ld}''(X) \cdot H^2 = \langle X^{-1}HX^{-1}, H \rangle = \|X^{-1/2}HX^{-1/2}\|^2$, qui est strictement positif pour H non nul. La fonction ld vérifie donc bien les hypothèses (16.2).

Soient à présent $X \succ 0$ et $Y \in B_X(X, 1)$. Alors

$$\begin{aligned} \|Y - X\|_X^2 &= \text{ld}''(X) \cdot (Y - X)^2 \\ &= \text{tr } X^{-1}(Y - X)X^{-1}(Y - X) \\ &= \text{tr}(X^{-1/2}YX^{-1/2} - I)(X^{-1/2}YX^{-1/2} - I) \\ &= \sum_{i=1}^n (1 - \lambda_i)^2, \end{aligned}$$

où les λ_i sont les valeurs propres de $X^{-1/2}YX^{-1/2}$. Comme $\|Y - X\|_X < 1$, tous les $(1 - \lambda_i)^2 < 1$ et donc toutes les valeurs propres $\lambda_i > 0$. Dès lors $X^{-1/2}YX^{-1/2} \succ 0$ ou encore $Y \succ 0$.

Il reste à montrer (16.4). Soient $X \succ 0$, $Y \in B_X(X, 1)$ et $V \in \mathcal{S}^n$. On note

$$\bar{Y} := X^{-1/2}YX^{-1/2} \quad \text{et} \quad \bar{V} := X^{-1/2}VX^{-1/2}.$$

Les valeurs propres de \bar{Y} sont notées $0 < \lambda_1 \leqslant \lambda_2 \leqslant \dots \leqslant \lambda_n$ (on a vu que $Y \succ 0$ et que $\bar{Y} \succ 0$) et sa factorisation spectrale est de la forme $\bar{Y} = Q\Lambda Q^\top$, où Q est orthogonale et $\Lambda := \text{Diag}(\lambda_1, \dots, \lambda_n)$. On a vu que

$$\|V\|_X^2 = \|\bar{V}\|^2$$

et on calcule facilement

$$\begin{aligned}
\|V\|_Y^2 &= \text{tr } Y^{-1} V Y^{-1} V \\
&= \text{tr}(X^{1/2} Y^{-1} X^{1/2})(X^{-1/2} V X^{-1/2})(X^{1/2} Y^{-1} X^{1/2})(X^{-1/2} V X^{-1/2}) \\
&= \text{tr } \bar{V}^{-1} \bar{V} \bar{V}^{-1} \bar{V} \\
&= \text{tr}(\bar{V}^{-1/2} \bar{V} \bar{V}^{-1/2})^2 \\
&= \text{tr}(\bar{\Lambda}^{-1/2} Q^\top \bar{V} Q \bar{\Lambda}^{-1/2})^2 \\
&= \sum_{i,j} \frac{1}{\lambda_i} (Q^\top \bar{V} Q)_{ij}^2 \frac{1}{\lambda_j} \\
&\leqslant \frac{1}{\lambda_1^2} \|Q^\top \bar{V} Q\|^2 \\
&= \frac{1}{\lambda_1^2} \|\bar{V}\|^2 \\
&= \frac{1}{\lambda_1^2} \|V\|_X^2
\end{aligned} \tag{16.7}$$

On obtient l'inégalité de droite dans (16.4) en observant que

$$\lambda_1 \geqslant 1 - |\lambda_1 - 1| \geqslant 1 - (\sum_i (\lambda_i - 1)^2)^{1/2} = 1 - \|Y - X\|_X.$$

L'inégalité de gauche dans (16.4) s'obtient de la même manière. On minore $\|V\|_Y^2$ ci-dessus à partir de (16.7) pour obtenir $\|Y\|_Y^2 \geqslant \|V\|_X^2 / \lambda_n^2$ et on utilise

$$\lambda_n \leqslant 1 + |\lambda_n - 1| \leqslant 1 + (\sum_i (\lambda_i - 1)^2)^{1/2} = 1 + \|Y - X\|_X \leqslant \frac{1}{1 - \|Y - X\|_X}. \quad \square$$

L'hypothèse (AC1) porte sur la courbure de f qui doit être assez importante en un point x proche de $\partial(\text{dom } f)$, la frontière du domaine de f . En exploitant cette propriété, on peut déduire le résultat de la proposition suivante selon laquelle, le long d'une droite rencontrant $\partial(\text{dom } f)$, une fonction autoconcordante croît au moins autant que l'opposé du logarithme dans le voisinage de 0. On en déduit que $f(x)$ tend vers l'infini lorsque x tend vers un point de $\partial(\text{dom } f)$.

Proposition 16.4 (croissance proche de la frontière du domaine) Soient $x \in \text{dom } f$, $\bar{x} \in \partial(\text{dom } f)$ et $x_t = (1-t)\bar{x} + tx$ avec $t \in]0, 1]$. Alors

$$f(x_t) \geqslant f(x) + f'(x) \cdot (x_t - x) - \log t + t - 1. \tag{16.8}$$

Dès lors, si une suite $\{x_k\} \subseteq \text{dom } f$ converge vers un point de $\partial(\text{dom } f)$, alors $f(x_k) \rightarrow +\infty$.

DÉMONSTRATION. Comme $\bar{x} = x_t + (\bar{x} - x_t) \notin \text{dom } f$, on a $\|\bar{x} - x_t\|_{x_t} \geqslant 1$ (par la propriété (AC1)), ce qui s'écrit $f''(x_t) \cdot (\bar{x} - x_t)^2 \geqslant 1$ ou encore

$$f''(x_t) \cdot (\bar{x} - x)^2 \geqslant \frac{1}{t^2}. \tag{16.9}$$

Pour $t \in]0, 1]$, le développement de Taylor d'ordre 2 avec reste intégral s'écrit

$$f(x_t) = f(x) + f'(x) \cdot (x_t - x) + \int_0^1 (1-\alpha) f''(x + \alpha(x_t - x)) \cdot (x_t - x)^2 d\alpha.$$

En utilisant (16.9), on a

$$f''(x + \alpha(x_t - x)) \cdot (x_t - x)^2 = (1-t)^2 f''(x_{1-\alpha(1-t)}) \cdot (\bar{x} - x)^2 \geq \frac{(1-t)^2}{(1-\alpha(1-t))^2}.$$

On a aussi en posant $\beta = 1 - \alpha(1-t)$:

$$\int_0^1 \frac{(1-t)^2(1-\alpha)}{(1-\alpha(1-t))^2} d\alpha = \int_1^t \frac{t-\beta}{\beta^2} d\beta = -1 + t - \log t.$$

En regroupant les trois dernières relations, on trouve (16.8).

Venons-en à la seconde partie de la proposition : une suite $\{x_k\} \subseteq \text{dom } f$ converge vers un point de $\partial(\text{dom } f)$. Soit $x \in \text{dom } f$, qui est un ouvert. On peut donc trouver $r > 0$ tel que $B(x, r) \subseteq \text{dom } f$. Dès que $x_k \in B(\bar{x}, r)$, $x + t(x_k - x) \notin \text{dom } f$ pour $t > 0$ assez grand. En effet, dans le cas contraire, par convexité du domaine de f , on aurait $B(x_k, r) \subseteq \text{dom } f$, ce qui n'est pas possible puisque \bar{x} appartiendrait à cette boule sans appartenir à $\text{dom } f$. Pour k assez grand, on peut donc trouver un $\bar{x}_k \in \partial(\text{dom } f)$ tel que $x_k = (1-t_k)\bar{x}_k + t_k x$, avec $0 < t_k \leq 1$. Alors, par convexité de $\text{dom } f$, $B(x_k, t_k r) \subseteq \text{dom } f$. Donc $t_k \leq \|x_k - \bar{x}\|/r$ et $t_k \downarrow 0$ lorsque $k \rightarrow \infty$. En utilisant l'inégalité (16.8), on trouve

$$f(x_k) \geq f(x) + f'(x) \cdot (x_k - x) - 1 + t_k - \log t_k.$$

Comme $x_k \rightarrow \bar{x}$ et $t_k \downarrow 0$, on trouve que $f(x_k) \rightarrow +\infty$ lorsque $k \rightarrow \infty$. \square

Ce résultat implique qu'une fonction autoconcordante est fermée :

$$\text{AC}(\mathbb{E}) \subseteq \overline{\text{Conv}}(\mathbb{E}). \quad (16.10)$$

En particulier, f a un minimum si son domaine est borné.

Si on fixe v dans les inégalités de (16.4), on voit que cette propriété essentielle des fonctions autoconcordantes exprime à sa manière que la hessienne $H(x)$ ne varie pas trop vite avec x . La proposition ci-dessous donne une autre expression de cette propriété en faisant disparaître v . Ainsi, en se rappelant que $H_x(x) = I$, on voit que la première inégalité dans (16.12) traduit une condition de Lipschitz radiale locale sur la hessienne (x est fixé et y doit être voisin de x), si bien qu'une fonction autoconcordante est un peu plus régulière que C^2 .

Proposition 16.5 (hessienne lipschitzienne) *Dans la définition d'une fonction autoconcordante, on peut remplacer (AC2) par l'une des conditions équivalentes suivantes :*

$$\forall x \in \text{dom } f, \quad \forall y \in B_x(x, 1), \quad (16.11)$$

$$\|H_x(y)\|_x, \|H_x(y)^{-1}\|_x \leq \frac{1}{(1 - \|y - x\|_x)^2}$$

ou

$$\forall x \in \text{dom } f, \quad \forall y \in B_x(x, 1),$$

$$\|H_x(y) - I\|_x, \|H_x(y)^{-1} - I\|_x \leq \frac{1}{(1 - \|y - x\|_x)^2} - 1. \quad (16.12)$$

DÉMONSTRATION. On note $0 < \lambda_1 \leq \dots \leq \lambda_n$ les valeurs propres de $H_x(y)$.

$[(\text{AC2}) \Leftrightarrow (16.11)]$ Les inégalités dans (16.11) s'obtiennent en prenant le supremum et l'infimum en $v \neq 0$ du quotient $\|v\|_y/\|v\|_x$ dans (16.4). En utilisant le fait que $H_x(y)$ est auto-adjointe et définie positive pour le produit scalaire local en x , on obtient (exercice 4.12) :

$$\begin{aligned} \max_{v \neq 0} \frac{\|v\|_y^2}{\|v\|_x^2} &= \max_{v \neq 0} \frac{\langle H(y)v, v \rangle}{\|v\|_x^2} = \max_{v \neq 0} \frac{\langle H_x(y)v, v \rangle_x}{\|v\|_x^2} = \lambda_n = \|H_x(y)\|_x, \\ \min_{v \neq 0} \frac{\|v\|_y^2}{\|v\|_x^2} &= \min_{v \neq 0} \frac{\langle H(y)v, v \rangle}{\|v\|_x^2} = \min_{v \neq 0} \frac{\langle H_x(y)v, v \rangle_x}{\|v\|_x^2} = \lambda_1 = \frac{1}{\|H_x(y)^{-1}\|_x}. \end{aligned}$$

L'équivalence entre (AC2) et (16.11) s'en déduit.

$[(16.11) \Leftrightarrow (16.12)]$ Les valeurs propres de $H_x(y) - I$ sont $\lambda_i - 1$. D'autre part, $\lambda_1 > 0$ et $(\lambda_1 - 1)^2 \geq 0$ permettent d'écrire $1 - \lambda_1 \leq \frac{1}{\lambda_1} - 1$. Dès lors (exercice 4.12)

$$\begin{aligned} \|H_x(y) - I\|_x &= \max_i |\lambda_i - 1| \\ &= \max\{\lambda_n - 1, 1 - \lambda_1\} \\ &\leq \max\{\lambda_n - 1, \frac{1}{\lambda_1} - 1\} \\ &= \max\{\|H_x(y)\|_x - 1, \|H_x(y)^{-1}\|_x - 1\}. \end{aligned}$$

On démontre de la même manière la seconde inégalité de (16.12). Réciproquement les inégalités de (16.12) impliquent directement celles de (16.11), car $\|H_x(y) - I\|_x \geq \|H_x(y)\|_x - 1$ et $\|H_x(y)^{-1} - I\|_x \geq \|H_x(y)^{-1}\|_x - 1$. \square

Règles de calcul

Nous énonçons ci-après quelques opérations qui préservent l'autoconcordance.

Proposition 16.6 (restriction à un sous-espace vectoriel) Si $f \in \text{AC}(\mathbb{E})$ et \mathbb{E}_0 est un sous-espace vectoriel de \mathbb{E} qui rencontre $\text{dom } f$, alors $f|_{\mathbb{E}_0} \in \text{AC}(\mathbb{E}_0)$.

DÉMONSTRATION. Il est clair que le **domaine** de $f|_{\mathbb{E}_0}$ ($= \mathbb{E}_0 \cap \text{dom } f$) est bien un ouvert non vide de \mathbb{E}_0 , que $f|_{\mathbb{E}_0}$ est C^2 sur \mathbb{E}_0 et que $(f|_{\mathbb{E}_0})''(x) \cdot v^2 = f''(x) \cdot v^2 > 0$,

pour tout $x \in \text{dom } f|_{\mathbb{E}_0}$ et tout v non nul dans \mathbb{E}_0 . D'autre part, les normes $\|\cdot\|_x$ associées à $f|_{\mathbb{E}_0}$ et f sont identiques, si bien que les autres propriétés requises pour que $f|_{\mathbb{E}_0}$ soit autoconcordante sont remplies. \square

Proposition 16.7 (somme) Si f_1 et $f_2 \in \text{AC}(\mathbb{E})$ et si $\text{dom } f_1 \cap \text{dom } f_2 \neq \emptyset$, alors $f_1 + f_2 \in \text{AC}(\mathbb{E})$.

DÉMONSTRATION. Les conditions (16.2) sont clairement vérifiées. D'autre part, si on note $B_{x,i}(x, 1)$ la boule unité de centre x pour la métrique $\langle u, v \rangle_{x,i} = f_i''(x) \cdot (u, v)$, la définition positivité des hessiennes $H_i(x)$ implique clairement que $B_x(x, 1) \subseteq B_{x,1}(x, 1) \cap B_{x,2}(x, 1) \subseteq \text{dom } f_1 \cap \text{dom } f_2$ [car les $f_i \in \text{AC}(\mathbb{E})$] $= \text{dom } f$. Enfin, en notant que pour des nombres réels strictement positifs a, b, c et d , on a

$$\min\left(\frac{a}{c}, \frac{b}{d}\right) \leqslant \frac{a+b}{c+d} \leqslant \max\left(\frac{a}{c}, \frac{b}{d}\right),$$

on obtient pour un vecteur $v \neq 0$:

$$\min\left(\frac{\|v\|_{y,1}^2}{\|v\|_{x,1}^2}, \frac{\|v\|_{y,2}^2}{\|v\|_{x,2}^2}\right) \leqslant \frac{\|v\|_y^2}{\|v\|_x^2} \leqslant \max\left(\frac{\|v\|_{y,1}^2}{\|v\|_{x,1}^2}, \frac{\|v\|_{y,2}^2}{\|v\|_{x,2}^2}\right).$$

En utilisant le fait que les $f_i \in \text{AC}(\mathbb{E})$, on obtient

$$\min(1 - \|y-x\|_{x,1}, 1 - \|y-x\|_{x,2}) \leqslant \frac{\|v\|_y}{\|v\|_x} \leqslant \max\left(\frac{1}{1 - \|y-x\|_{x,1}}, \frac{1}{1 - \|y-x\|_{x,2}}\right).$$

Il reste à observer que, quel que soit l'indice $i = 1, 2$, on a $1 - \|y-x\|_x \leqslant 1 - \|y-x\|_{x,i}$ et que $(1 - \|y-x\|_{x,i})^{-1} \leqslant (1 - \|y-x\|_x)^{-1}$, établissant ainsi (16.4). \square

Une fonction linéaire n'est pas une fonction autoconcordante (elle n'est pas strictement convexe), mais si on l'ajoute à une fonction autoconcordante, on ne détruit pas l'autoconcordance de cette dernière.

Proposition 16.8 (ajout d'une fonction linéaire) Si $f \in \text{AC}(\mathbb{E})$ et $c \in \mathbb{E}$, alors $f(\cdot) + \langle c, \cdot \rangle \in \text{AC}(\mathbb{E})$.

DÉMONSTRATION. Clair car $\text{dom}(f + c) = \text{dom } f$ est ouvert, $f + c$ est C^2 sur son domaine et $\nabla^2(f + c)(x) = \nabla^2 f(x)$ pour $x \in \text{dom } f$. Les autres propriétés requises s'en suivent. \square

Proposition 16.9 (précomposition par une application affine) *Si \mathbb{F} est un autre espace vectoriel de dimension finie, si $a : \mathbb{E} \rightarrow \mathbb{F}$ est affine et injective, si $f \in AC(\mathbb{F})$ et si $\text{dom } f \cap a(\mathbb{E}) \neq \emptyset$, alors $f \circ a \in AC(\mathbb{E})$.*

DÉMONSTRATION. On peut écrire $a(x) = Ax - b$, avec $A : \mathbb{E} \rightarrow \mathbb{F}$ linéaire injective et $b \in \mathbb{F}$. On note $f^a := f \circ a$ et $\|\cdot\|_x^a = \|A \cdot\|_{a(x)}$ la norme locale en x associée à f^a . Le $\text{dom } f^a = \{x \in \mathbb{E} : a(x) \in \text{dom } f\}$ est un ouvert non vide, sur lequel f^a est C^2 avec $(f^a)''(x) \cdot v^2 = f''(a(x)) \cdot (Av)^2 > 0$ pour tout v non nul, grâce à l'injectivité de A . Dès lors (16.2) a lieu.

Soit à présent $x \in \text{dom } f^a$ et $y \in \mathbb{E}$ tel que $\|y - x\|_x^a < 1$. Alors, $\|A(y - x)\|_{a(x)} < 1$ ou encore $\|a(y) - a(x)\|_{a(x)} < 1$, ce qui implique que $a(y) \in \text{dom } f$ (puisque $a(x) \in \text{dom } f$ et $f \in AC(\mathbb{F})$) ou encore $y \in \text{dom } f^a$.

Enfin, si $x \in \text{dom } f^a$ et $y \in \mathbb{E}$ est tel que $\|y - x\|_x^a < 1$, on a

$$\frac{\|v\|_y^a}{\|v\|_x^a} = \frac{\|Av\|_{a(y)}}{\|Av\|_{a(x)}} \leqslant \frac{1}{1 - \|a(y) - a(x)\|_{a(x)}} = \frac{1}{1 - \|y - x\|_x^a}.$$

L'inégalité de gauche dans (16.4) se démontre de la même manière. \square

Exemple 16.10 En composant la fonction `lb` sur \mathbb{R}^m et la fonction affine $a : \mathbb{R}^n \rightarrow \mathbb{R}^m$ définie par $a(x) = Ax - b$ (on suppose que A est injective et que $\{x : Ax > b\}$ est non vide), on obtient une fonction autoconcordante

$$\begin{cases} -\sum_{i=1}^m \log(A_i x - b_i) & \text{si } Ax > b \\ +\infty & \text{sinon,} \end{cases}$$

dont l'adhérence du [domaine](#) est le polyèdre convexe $\{x \in \mathbb{R}^n : Ax \geqslant b\}$. L'hypothèse d'injectivité de A requiert que l'on ait au moins n contraintes linéaires linéairement indépendantes. \square

Itérations de Newton sur une fonction autoconcordante

Soit

$$d(x) := -H(x)^{-1}g(x) = -g_x(x)$$

la direction de Newton en $x \in \text{dom } f$ sur une fonction $f \in AC(\mathbb{E})$. On note

$$x_+ := x + d(x) \tag{16.13}$$

l'itéré calculé par une itération de Newton en x .

La proposition ci-dessous permet de comparer, dans leurs normes locales, les directions de Newton $d(x)$ et $d(x_+)$ en deux points successifs x et $x_+ = x + d(x)$.

Proposition 16.11 (comparaison de deux directions de Newton successives) Si $f \in AC(\mathbb{E})$, $x \in \text{dom } f$ et $\|d(x)\|_x < 1$, alors $x_+ \in \text{dom } f$ et

$$\|d(x_+)\|_{x_+} \leq \left(\frac{\|d(x)\|_x}{1 - \|d(x)\|_x} \right)^2.$$

DÉMONSTRATION. On observe d'abord que si $x \in \text{dom } f$ et $\|x_+ - x\|_x = \|d(x)\|_x < 1$, alors $x_+ \in \text{dom } f$ car $B_x(x, 1) \subseteq \text{dom } f$ pour les fonction autoconcordante. Ensuite, comme la direction de Newton ne dépend pas du produit scalaire utilisé, on peut écrire $d(x_+) = -H_x(x_+)^{-1}g_x(x_+)$. La norme $\|H_x(x_+)^{-1}\|_x$ pourra se majorer grâce à l'autoconcordance de f , voir (16.11), tandis que le lien entre $g_x(x_+)$ et $g_x(x) = -d(x)$ s'obtiendra par un développement de Taylor.

On a

$$\|d(x_+)\|_{x_+}^2 = \langle g(x_+), H(x_+)^{-1}g(x_+) \rangle = \langle g_x(x_+), H(x_+)^{-1}H(x)g_x(x_+) \rangle_x.$$

Mais $H(x_+)^{-1}H(x) = H_x(x_+)^{-1}$, si bien que

$$\|d(x_+)\|_{x_+}^2 = \langle g_x(x_+), H_x(x_+)^{-1}g_x(x_+) \rangle_x \leq \|H_x(x_+)^{-1}\|_x \|g_x(x_+)\|_x^2.$$

On note $\delta := \|d(x)\|_x$. Alors (16.11) implique que $\|H_x(x_+)^{-1}\|_x \leq (1 - \delta)^{-2}$ et il reste donc à montrer que $\|g_x(x_+)\|_x \leq \delta^2(1 - \delta)^{-1}$. On utilise pour cela un développement de Taylor de $g(x_+)$ autour de x :

$$\begin{aligned} g_x(x_+) &= H(x)^{-1}g(x_+) \\ &= H(x)^{-1} \left[g(x) + \int_0^1 H(x+td(x)) d(t) dt \right] \\ &= \int_0^1 [H_x(x+td(x)) - I] d(t) dt. \end{aligned}$$

En prenant la norme locale en x et en utilisant (16.12):

$$\begin{aligned} \|g_x(x_+)\|_x &\leq \left(\int_0^1 \|H_x(x+td(x)) - I\|_x dt \right) \delta \\ &\leq \left(\int_0^1 \left[\frac{1}{(1 - \delta t)^2} - 1 \right] dt \right) \delta \\ &= \frac{\delta^2}{1 - \delta}. \end{aligned}$$

□

Soit q_x le modèle quadratique de f en $x \in \text{dom } f$, c'est-à-dire son développement de Taylor à l'ordre 2. On a pour tout $y \in \text{dom } f$:

$$q_x(y) = f(x) - \langle d(x), y - x \rangle_x + \frac{1}{2}\|y - x\|_x^2.$$

La proposition suivante donne une estimation de l'erreur entre f et son modèle quadratique q_x . On ne s'étonnera pas que l'erreur est en $O(\|y - x\|_x^3)$ puisque la dérivée seconde d'une fonction autoconcordante vérifie une condition de Lipschitz radiale locale; et on appréciera que, pour y dans le voisinage $B_x(x, 1)$ de x , aucune constante inconnue n'intervient dans l'estimation.

Proposition 16.12 (erreur du modèle quadratique) *Soient $f \in \text{AC}(\mathbb{E})$, $x \in \text{dom } f$ et $y \in B_x(x, 1)$. Alors*

$$|f(y) - q_x(y)| \leq \frac{\|y - x\|_x^3}{3(1 - \|y - x\|_x)}.$$

DÉMONSTRATION. En utilisant le développement

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(0) + \int_0^1 \int_0^t (\varphi''(s) - \varphi''(0)) \, ds \, dt$$

sur la fonction $\varphi(t) = f(x + t(y-x))$, on obtient

$$f(y) = q_x(y) + \int_0^1 \int_0^t \langle [H_x(x + s(y-x)) - I](y - x), y - x \rangle_x \, ds \, dt.$$

En prenant les normes, puis en utilisant (16.12), ce qui est licite puisque $y \in B_x(x, 1)$, on en déduit

$$\begin{aligned} |f(y) - q_x(y)| &\leq \|y - x\|_x^2 \int_0^1 \int_0^t \|H_x(x + s(y-x)) - I\|_x \, ds \, dt \\ &\leq \|y - x\|_x^2 \int_0^1 \int_0^t \left(\frac{1}{(1 - s\|y - x\|_x)^2} - 1 \right) \, ds \, dt \\ &= \|y - x\|_x^3 \int_0^1 \frac{t^2}{1 - t\|y - x\|_x} \, dt \\ &\leq \frac{\|y - x\|_x^3}{1 - \|y - x\|_x} \int_0^1 t^2 \, dt \\ &= \frac{\|y - x\|_x^3}{3(1 - \|y - x\|_x)}. \end{aligned}$$

□

La proposition suivante donne, en cas d'existence d'un minimum x_* , une estimation du progrès vers celui-ci que réalise une itération de Newton.

Proposition 16.13 (progrès vers le minimum) *Supposons que $f \in \text{AC}(\mathbb{E})$ ait un minimum en $x_* \in \text{dom } f$.*

1) Si $x \in \text{dom } f$ est tel que $x_* \in B_x(x, 1)$, alors x_+ vérifie

$$\|x_+ - x_*\|_x \leq \frac{\|x - x_*\|_x^2}{1 - \|x - x_*\|_x}.$$

2) Si $x \in \text{dom } f$ est tel que $x \in B_{x_*}(x_*, \frac{1}{4})$, alors x_+ vérifie

$$\|x_+ - x_*\|_{x_*} < 4\|x - x_*\|_{x_*}^2.$$

DÉMONSTRATION. 1) En retranchant x dans chaque membre de (16.13) et en faisant un développement de Taylor avec reste intégral de $g(x)$ autour de x_* , on obtient classiquement

$$x_+ - x_* = \int_0^1 \left[I - H_x(x_* + t(x-x_*)) \right] (x - x_*) dt.$$

En prenant les normes, puis en utilisant (16.12), ce qui est légitime puisque $x_* + t(x-x_*) \in B_x(x, 1)$, on trouve avec $\delta_+ := \|x_+ - x_*\|_x$ et $\delta := \|x - x_*\|_x$:

$$\delta_+ \leq \int_0^1 \left\| I - H_x(x_* + t(x-x_*)) \right\|_x \delta dt \leq \int_0^1 \left(\frac{\delta}{(1 - (1-t)\delta)^2} - \delta \right) dt = \frac{\delta^2}{1 - \delta}.$$

2) Afin de récrire l'inégalité obtenue au point 1 en termes de $\|x_+ - x_*\|_{x_*}$ et $\|x - x_*\|_{x_*}$, on utilise les inégalités suivantes, déduites de (16.4):

$$\begin{aligned} \|x_+ - x_*\|_{x_*} (1 - \|x - x_*\|_{x_*}) &\leq \|x_+ - x_*\|_x, \\ \|x - x_*\|_x &\leq \frac{\|x - x_*\|_{x_*}}{1 - \|x - x_*\|_{x_*}}. \end{aligned}$$

Lorsque $x \in B_{x_*}(x_*, \frac{1}{2})$, l'inégalité du point 1 devient alors

$$\|x_+ - x_*\|_{x_*} \leq \frac{\|x - x_*\|_{x_*}^2}{(1 - \|x - x_*\|_{x_*})^2 (1 - 2\|x - x_*\|_{x_*})}.$$

On obtient alors l'inégalité du point 2 lorsque, par exemple, $x \in B_{x_*}(x_*, \frac{1}{4})$. \square

Proposition 16.14 (existence d'un minimum se déduisant d'une direction de Newton petite) Si $f \in \text{AC}(\mathbb{E})$ et si $\|d(x)\|_x \leq \frac{1}{4}$ pour un $x \in \text{dom } f$, alors f a un minimum en un point $x_* \in \text{dom } f$,

$$\|x_+ - x_*\|_x \leq \frac{3\|d(x)\|_x^2}{(1 - \|d(x)\|_x)^3} \quad \text{et} \quad \|x - x_*\|_x \leq \|d(x)\|_x + \frac{3\|d(x)\|_x^2}{(1 - \|d(x)\|_x)^3}.$$

DÉMONSTRATION.

\square

Autres propriétés

Proposition 16.15 (existence d'un minimum) Si $f \in AC(\mathbb{E})$ est minorée par une constante (par exemple parce que $\text{dom } f$ est borné), alors f a un minimum.

DÉMONSTRATION. □

Proposition 16.16 (comportement frontière) Si $f \in AC(\mathbb{E})$ et si $\{x_k\} \subseteq \text{dom } f$ est une suite convergeant vers un point $\bar{x} \in \partial(\text{dom } f)$, alors $f(x_k) \rightarrow +\infty$ et $\|g(x_k)\| \rightarrow +\infty$. En particulier, $f \in \text{Conv}(\mathbb{E})$.

DÉMONSTRATION. □

Dans le voisinage d'un point $x \in \text{dom } f$, l'application $g(\cdot)$ est inversible (on peut utiliser le théorème de la fonction inverse, car $g'(x) = H(x)$ est inversible) : pour v petit, on peut écrire $g(x) + v = g(x + u)$ pour un certain $u \in \mathbb{E}$. La proposition suivante quantifie la petitesse de v pour que $g(x) + v$ soit dans l'image de g .

Proposition 16.17 (inversion du gradient) Soient $f \in AC(\mathbb{E})$, $x \in \text{dom } f$ et $r \leq \frac{1}{4}$. Si $\|v\|_x \leq r$, il existe $u \in \bar{B}_x(v, \frac{3r^2}{(1-r)^3})$ tel que $g_x(x) + v = g_x(x+u)$.

DÉMONSTRATION. On considère $\tilde{f} \in AC(\mathbb{E})$, définie en $y \in \mathbb{E}$ par

$$\tilde{f}(y) = f(y) - \langle g_x(x) + v, y \rangle_x.$$

Le minimum de \tilde{f} , s'il existe, est le point $x + u$ recherché.

On observera que $\tilde{f} \in AC(\mathbb{E})$ (proposition 16.8). Le déplacement de Newton en x sur \tilde{f} s'écrit

$$-H(x)^{-1}(g(x) - H(x)g_x(x) - H(x)v) = v.$$

Sa norme locale étant plus petite que $\frac{1}{4}$, \tilde{f} a un minimum (proposition 16.14); notons le $x + u \in \text{dom } f$. Selon la même proposition, $\|u - v\|_x \leq 3\|v\|_x^2/(1 - \|v\|_x)^3 \leq 3r^2/(1-r)^3$. □

16.1.3 Barrière autoconcordante

Pour être autoconcordante, une fonction doit vérifier les propriétés globales (16.2) et les propriétés locales (AC1) et (AC2). Pour être une *barrière* autoconcordante, il faut qu'une propriété globale supplémentaire ait lieu.

Définition 16.18 On dit qu'une fonction $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ est une (fonction) *barrière autoconcordante*, si $f \in AC(\mathbb{E})$ et si

$$\vartheta_f := \sup_{x \in \text{dom } f} \|g_x(x)\|_x^2 \quad (16.14)$$

est fini. Le scalaire ϑ_f est appelé le *module de complexité* de f . On note $BAC(\mathbb{E})$ l'ensemble des fonctions barrières autoconcordantes sur \mathbb{E} . \square

Comme la direction de Newton s'écrit $d(x) = -g_x(x)$, une barrière auto-concordante a ses directions de Newton bornées en norme locale. On observe que ni $g_x(x)$, ni $\|\cdot\|_x$ ne dépendent du produit scalaire de référence; il en est donc de même de ϑ_f . Enfin, il est clair que $\vartheta_f > 0$ (sinon $g_x(x) = 0$ pour tout $x \in \text{dom } f$, ce qui ne serait possible que si $g(x) = 0$ pour tout $x \in \text{dom } f$ ou encore que si f était constante, en contradiction évidente avec la stricte convexité d'une fonction autoconcordante). Nous verrons que l'on a en fait $\vartheta_f \geq 1$ (proposition 16.28), ce qui peut paraître surprenant au premier abord.

Exemple 16.19 La fonction $\text{lb} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ définie en (16.5) est une barrière autoconcordante avec $\vartheta_{\text{lb}} = n$.

DÉMONSTRATION. On sait que $\text{lb} \in AC(\mathbb{R}^n)$ (exemple 16.2). D'autre part, si on note $X = \text{Diag}(x_1, \dots, x_n)$ et e le vecteur dont les composantes valent 1, le gradient de lb pour la métrique locale s'écrit $g_x(x) = H(x)^{-1}g(x) = (X^{-2})^{-1}(-X^{-1}e) = -x$. Dès lors $\vartheta_{\text{lb}} = \sup_{x > 0} \|x\|_x^2 = \sup_x x^T X^{-2} x = n$. \square

Exemple 16.20 La fonction $\text{ld} : \mathcal{S}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ définie en (16.6) est une barrière autoconcordante avec $\vartheta_{\text{ld}} = n$.

DÉMONSTRATION. On sait que $\text{ld} \in AC(\mathcal{S}^n)$ (exemple 16.3). Le gradient $g(X)$ et la hessienne $H(X)$ de ld en $X \succ 0$ vérifient $g(X) = -X^{-1}$ et $H(X)M = X^{-1}MX^{-1}$ pour tout $M \in \mathcal{S}^n$. Alors $-X^{-1} = g(X) = H(X)g_X(X) = X^{-1}g_X(X)X^{-1}$, si bien que $g_X(X) = -X$. Finalement $\|g_X(X)\|_X^2 = \langle g(X), g(X) \rangle = \text{tr } I = n$. \square

Calcul autoconcordant

Proposition 16.21 (restriction à un sous-espace vectoriel) Si $f \in BAC(\mathbb{E})$ et si \mathbb{E}_0 est un sous-espace vectoriel de \mathbb{E} qui rencontre $\text{dom } f$, alors $f|_{\mathbb{E}_0} \in BAC(\mathbb{E}_0)$ et $\vartheta_{f|_{\mathbb{E}_0}} \leq \vartheta_f$.

DÉMONSTRATION. En effet, $f|_{\mathbb{E}_0} \in AC(\mathbb{E}_0)$ (proposition 16.6). Par ailleurs, si $P_{\mathbb{E}_0, x}$ est le projecteur orthogonal sur \mathbb{E}_0 pour la métrique riemannienne en x et si $g|_{\mathbb{E}_0, x}(x)$ est le gradient de $f|_{\mathbb{E}_0}$ en x pour cette même métrique, on a $g|_{\mathbb{E}_0, x}(x) = P_{\mathbb{E}_0, x}g_x(x)$ et $\|P_{\mathbb{E}_0, x}v\|_x \leq \|v\|_x$ (quel que soit $v \in \mathbb{E}$). Dès lors

$$\vartheta_{f|_{\mathbb{E}_0}} = \sup_{x \in \mathbb{E}_0 \cap \text{dom } f} \|g|_{\mathbb{E}_0, x}(x)\|_x^2 \leqslant \sup_{x \in \mathbb{E}_0 \cap \text{dom } f} \|g_x(x)\|_x^2 \leqslant \sup_{x \in \text{dom } f} \|g_x(x)\|_x^2 = \vartheta_f.$$

□

Proposition 16.22 (somme) Si f_1 et $f_2 \in \text{BAC}(\mathbb{E})$ et si $\text{dom } f_1 \cap \text{dom } f_2 \neq \emptyset$, alors $f_1 + f_2 \in \text{BAC}(\mathbb{E})$ et $\vartheta_{f_1 + f_2} \leqslant \vartheta_{f_1} + \vartheta_{f_2}$.

DÉMONSTRATION. On sait déjà que $f := f_1 + f_2 \in \text{AC}(\mathbb{E})$ (proposition 16.7). On simplifie les écritures en notant $g_i := g_i(x)$, $H_i := H_i(x)$, $g = g_1 + g_2$ et $H = H_1 + H_2$.

Il faut estimer $\|g_x(x)\|_x$, mais comme cette quantité ne dépend pas du produit scalaire de référence, on peut supposer que celui-ci est tel que $H(x) = I$. Alors on a

$$\|g_x(x)\|_x^2 = f''(x) \cdot (H(x)^{-1}g(x))^2 = \|g\|^2 = \|g_1 + g_2\|^2.$$

On développe le carré et on introduit $v_i := H_i^{-1/2}g_i$:

$$\|g_x(x)\|_x^2 = \|g_1\|^2 + \|g_2\|^2 + 2\langle g_1, g_2 \rangle = \langle H_1 v_1, v_1 \rangle + \langle H_2 v_2, v_2 \rangle + 2\langle H_1^{1/2}v_1, H_2^{1/2}v_2 \rangle.$$

On utilise alors $H_1 + H_2 = H = I$:

$$\begin{aligned} \|g_x(x)\|_x^2 &= \langle (I - H_2)v_1, v_1 \rangle + \langle (I - H_1)v_2, v_2 \rangle + 2\langle H_1^{1/2}v_1, H_2^{1/2}v_2 \rangle \\ &= \|v_1\|^2 + \|v_2\|^2 - \|H_2^{1/2}v_1\|^2 - \|H_1^{1/2}v_2\|^2 + 2\langle H_1^{1/2}v_1, H_2^{1/2}v_2 \rangle. \end{aligned}$$

Mais H_1 et H_2 commutent (car $H_1 H_2 = H_1(I - H_1) = (I - H_1)H_1 = H_2 H_1$) et donc aussi $H_1^{1/2}$ et $H_2^{1/2}$. Alors la somme des trois derniers termes ci-dessus est la quantité négative $-\|H_2^{1/2}v_1 - H_1^{1/2}v_2\|^2$. On en déduit

$$\|g_x(x)\|_x^2 \leqslant \|v_1\|^2 + \|v_2\|^2 = \langle H_1^{-1}g_1, g_1 \rangle + \langle H_2^{-1}g_2, g_2 \rangle \leqslant \vartheta_{f_1} + \vartheta_{f_2}.$$

En prenant le supremum en $x \in \text{dom } f$ à gauche on obtient $\vartheta_f \leqslant \vartheta_{f_1} + \vartheta_{f_2}$. □

La proposition 16.8 ne s'étend pas aux barrières autoconcordantes.

Contre-exemple 16.23 La fonction $x \in \mathbb{R} \mapsto x - \log x$ n'est pas une barrière autoconcordante. En effet, $g(x) = 1 - 1/x$, $H(x) = 1/x^2$, $g_x(x) = x^2 - x$, $\|g_x(x)\|_x^2 = (x^2 - x)/x^2 = (x - 1)^2$ qui n'est pas borné sur \mathbb{R}_{++} . □

Proposition 16.24 (précomposition par une application affine) Si \mathbb{F} est un autre espace euclidien, si $a : \mathbb{E} \rightarrow \mathbb{F}$ est affine et injective, si $f \in \text{BAC}(\mathbb{F})$ et si $\text{dom } f \cap a(\mathbb{E}) \neq \emptyset$, alors $f \circ a \in \text{BAC}(\mathbb{E})$ et $\vartheta_{f \circ a} \leqslant \vartheta_f$.

DÉMONSTRATION. On sait déjà que $\tilde{f} := f \circ a \in \text{AC}(\mathbb{E})$ (proposition 16.9). Soient $x \in \mathbb{E}$ tel que $y := a(x) \in \text{dom } f$. On prend $\langle \cdot, \cdot \rangle_y$ comme produit scalaire de référence sur \mathbb{F} et on suppose que $a(x) = Ax + b$ avec A linéaire injective.

Calculons le gradient et la hessienne de \tilde{f} en x . On a $\tilde{f}'(x) \cdot u = f'(y) \cdot (Au) = \langle A^*g(y), u \rangle$ et $\tilde{f}''(x) \cdot (u, v) = f''(y) \cdot (Au, Av) = \langle A^*Au, v \rangle$ (avec le produit scalaire choisi sur \mathbb{F} , $H(y) = I$). Dès lors

$$\tilde{g}(x) = A^*g(y) \quad \text{et} \quad \tilde{H}(x) = A^*A.$$

On peut alors évaluer

$$\begin{aligned} \|\tilde{g}_x(x)\|_x^2 &= \langle \tilde{H}(x)^{-1}\tilde{g}(x), \tilde{g}(x) \rangle \\ &= \langle (A^*A)^{-1}A^*g(y), A^*g(y) \rangle \\ &= \langle A(A^*A)^{-1}A^*g(y), g(y) \rangle_y \\ &\leq \|A(A^*A)^{-1}A^*\|_y \|g_y(y)\|_y^2 \\ &\leq \vartheta_f, \end{aligned}$$

car $A(A^*A)^{-1}A^*$ étant un projecteur orthogonal pour le produit scalaire $\langle \cdot, \cdot \rangle_y$ (il est idempotent et auto-adjoint), il est de norme $\|\cdot\|_y$ inférieure à 1. \square

Autres propriétés

Beaucoup de propriétés des barrières autoconcordantes se déduisent des bornes que l'on peut obtenir sur ses dérivées directionnelles. C'est ce sur quoi se concentrent les deux propositions suivantes. On y fait usage de l'inverse de la fonction de Minkowski de $\text{dom } f$, de pôle $x \in \text{adh}(\text{dom } f)$, qui est défini en $y \in \mathbb{E}$ par

$$\alpha_x(y) = \sup\{\alpha \geq 0 : x + \alpha(y - x) \in \text{dom } f\}.$$

La *fonction de Minkowski* est l'application $y \mapsto 1/\alpha_x(y)$.

Proposition 16.25 (dérivée directionnelle première) *Soient $f \in \text{BAC}(\mathbb{E})$ et $x \in \text{dom } f$.*

1) *Si $y \in \text{dom } f$, alors*

$$\frac{-\vartheta_f}{\alpha_y(x) - 1} \leq f'(x) \cdot (y - x) < \vartheta_f. \quad (16.15)$$

2) *Si $y \in \partial(\text{dom } f)$ et $\alpha_y(x) \geq (1 + \vartheta_f^{1/2})^2$, alors*

$$1 - \frac{(1 + \vartheta_f^{1/2})^2}{\alpha_y(x)} \leq f'(x) \cdot (y - x). \quad (16.16)$$

DÉMONSTRATION. 1) Démontrons d'abord l'inégalité de droite dans (16.15). On examine f le long de la droite portée par x et $y \in \text{dom } f$, en introduisant la fonction φ définie par $\varphi(t) := f(x + t(y-x))$. Il s'agit donc de démontrer que $\varphi'(0) < \vartheta_f$ ou $1/\varphi'(0) > 1/\vartheta_f$, car on peut supposer que $\varphi'(0) > 0$ (sinon le résultat est trivial). Si

$\varphi'(0) > 0$, alors, par convexité, $\varphi'(t) > 0$ pour $t \geq 0$ dans le domaine de φ . Comme 1 est dans le dom φ , il suffit en fait de démontrer l'inégalité un peu plus forte

$$\frac{1}{\varphi'(0)} - \frac{1}{\varphi'(1)} \geq \frac{1}{\vartheta_f}. \quad (16.17)$$

Celle-ci suggère de regarder comment varie $1/\varphi'(t)$ ou encore d'intégrer la dérivée de ce quotient entre 0 et 1. On trouve

$$\left(-\frac{1}{\varphi'(t)} \right)' = \frac{\varphi''(t)}{\varphi'(t)^2} \geq \frac{1}{\vartheta_f}, \quad (16.18)$$

où l'inégalité à droite vient des estimations suivantes, dans lesquelles on a noté $x_t := x + t(y-x)$:

$$|\varphi'(t)| = |\langle g_{x_t}(x_t), y - x \rangle_{x_t}| \leq \|g_{x_t}(x_t)\|_{x_t} \|y - x\|_{x_t} \leq \vartheta_f^{1/2} \|y - x\|_{x_t},$$

$$\varphi''(t) = \|y - x\|_{x_t}^2.$$

Après intégration de (16.18) entre 0 et 1, on obtient (16.17).

Pour établir l'inégalité de gauche dans (16.15), on introduit la fonction φ définie par $\varphi(t) := f(y + t(x-y))$. Il s'agit donc de majorer $\varphi'(1) = -f'(x) \cdot (y-x)$, que l'on peut supposer strictement positif (sinon le résultat est trivial). Par convexité, on a alors $\varphi'(t) > 0$ pour $t \geq 1$ dans le domaine de φ . Comme ci-dessus, $\varphi'(t)$ vérifie (16.18). Après intégration entre 1 et t , on obtient cette fois

$$\vartheta_f - (t-1)\varphi'(1) \geq \frac{\vartheta_f \varphi'(1)}{\varphi'(t)}.$$

Le membre de droite reste strictement positif pour tout $t \geq 1$ dans dom φ ; il en est donc de même du membre de gauche. En prenant le plus grand t possible, c'est-à-dire $t = \alpha_y(x)$, on voit que $\varphi'(1) \leq \vartheta_f / (\alpha_y(x) - 1)$, qui est l'inégalité recherchée.

2) On note $T := \alpha_y(x)$ si cette quantité est finie; sinon T est un nombre arbitrairement grand. On définit à présent $\varphi \in AC(\mathbb{R})$ par $\varphi(t) := f(y + t(x-y))$. Il s'agit de majorer $\varphi'(1)$. Le domaine de φ est l'intervalle $]0, T[$. Comme φ est autoconcordante et $0 \notin \text{dom } \varphi$, on a pour tout $t \in]0, T[$:

$$\varphi''(t)t^2 \geq 1.$$

D'autre part, en appliquant à φ l'inégalité de gauche dans (16.15), qui est encore vérifiée pour $y \in \partial(\text{dom } f)$, on trouve $\varphi'(t)(-t) \geq -\vartheta_f / (\alpha_0(t) - 1)$ (proposition 16.21). Comme $\alpha_0(t) = T/t$, on obtient pour tout $t \in]0, T[$:

$$\varphi'(t) \leq \frac{\vartheta_f}{T-t}.$$

On intègre alors $\varphi''(t)$ de 1 à $t \in]t, T[$ et on tient compte des deux dernières inégalités :

$$\varphi'(1) - \frac{1}{t} + 1 \leq \varphi'(1) + \int_1^t \varphi''(s)ds = \varphi'(t) \leq \frac{\vartheta_f}{T-t}$$

ou encore

$$\varphi'(1) \leq \frac{\vartheta_f}{T-t} + \frac{1}{t} - 1. \quad (16.19)$$

Si $T := \alpha_y(x) < +\infty$, on peut prendre $t = T/(1 + \vartheta_f^{1/2})$, qui est dans $]1, T[$, par l'hypothèse sur $\alpha_y(x)$. On trouve

$$\varphi'(1) \leq \frac{\vartheta_f(1 + \vartheta_f^{1/2})}{T\vartheta_f^{1/2}} + \frac{1 + \vartheta_f^{1/2}}{T} - 1 = \frac{(1 + \vartheta_f^{1/2})^2}{T} - 1,$$

qui est bien (16.16). Si $\alpha_y(x) = +\infty$, on peut prendre T et t arbitrairement grand dans (16.19), si bien que l'on trouve $\varphi'(1) \leq -1$ et donc aussi (16.16). \square

Les dérivées directionnelles secondes peuvent s'exprimer en utilisant la norme locale. C'est ce que nous faisons dans la proposition suivante.

Proposition 16.26 (dérivée directionnelle seconde) *Soient $f \in \text{BAC}(\mathbb{E})$ et $x, y \in \text{dom } f$ vérifiant $f'(x) \cdot (y-x) \geq 0$. Alors $\|y-x\|_x \leq 4\vartheta_f + 1$.*

DÉMONSTRATION. \square

Les propositions précédentes ont de nombreuses conséquences. En voici déjà quelques-unes, toutes immédiates, apportant un éclairage sur l'allure des barrières autoconcordantes.

- Si d fait un angle plus petit que $\pi/2$ avec un gradient $g(x) \neq 0$ d'une barrière autoconcordante, alors la *demi-droite* $\{x + td : t \geq 0\}$ rencontre la frontière de son *domaine*.

Cet énoncé suppose que l'on utilise le même produit scalaire pour définir l'angle et le gradient. Il se déduit de l'inégalité de droite dans (16.15). Si on utilise aussi la proposition 16.26, la zone où doit se trouver le *domaine* de f est représenté à la figure 16.1.

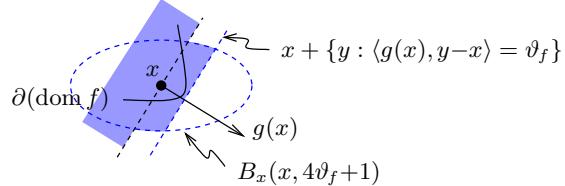


Fig. 16.1. Localisation du domaine d'une barrière autoconcordante

- Le *domaine* d'une barrière autoconcordante ne contient pas de droite. En particulier, il ne peut être \mathbb{E} tout entier.

En effet, supposons que $\{x_0 + \alpha d : \alpha \in \mathbb{R}\} \subseteq \text{dom } f$. Alors, pour tout $x = x_0 + \alpha d$, on aurait $f'(x) \cdot d = 0$ (prendre $y = x + td$ avec $t \rightarrow \pm\infty$ dans l'inégalité de droite dans (16.15)). On en déduirait que $f''(x) \cdot d^2 = 0$, ce qui ne serait pas compatible avec la stricte convexité de f .

- Si $x \in \text{dom } f$ et $d \in (\text{dom } f)^\infty(x)$, alors $f'(x) \cdot d \leq -1/\alpha_x(x-d) < 0$.

Comme $-d \notin (\text{dom } f)^\infty(x)$ (car $\text{dom } f$ ne contient pas de droite), $\alpha := \alpha_x(x-d)$ est fini et $y := x - \alpha d$ est sur la frontière de $\text{dom } f$. D'autre part $\alpha_y(x) = \infty$ et, d'après (16.16), $f'(x) \cdot (y-x) \geq 1$.

Le résultat suivant montre que, dans le cas où f a un minimum, celui-ci est bien centré dans le [domaine](#) de f . De plus, les ellipsoïdes inscrit $B_{x_*}(x_*, 1)$ et circonscrit $B_{x_*}(x_*, 4\vartheta_f+1)$ permettent d'avoir une bonne estimation de son volume.

Proposition 16.27 (existence d'un minimum) *Une fonction $f \in \text{BAC}(\mathbb{E})$ a un minimum x_* si, et seulement si, son [domaine](#) est borné. Dans ce cas*

$$B_{x_*}(x_*, 1) \subseteq \text{dom } f \subseteq B_{x_*}(x_*, 4\vartheta_f+1).$$

DÉMONSTRATION. Si $\text{dom } f$ est borné, f a un minimum (proposition 16.15). Inversement si f a un minimum $x_* \in \text{dom } f$, on a $B_{x_*}(x_*, 1) \subseteq \text{dom } f$ par la définition de l'autoconcordance et $\text{dom } f \subseteq B_{x_*}(x_*, 4\vartheta_f+1)$ par la proposition 16.26 (appliquée à $x = x_*$, avec $g(x_*) = 0$). \square

Estimation du module de complexité

Mieux vaut disposer d'une barrière autoconcordante avec un module de complexité petit car, comme nous le verrons à la section 16.1.4, la complexité itérative des algorithmes sera proportionnelle à ϑ_f^ω , avec un certain $\omega > 0$. Le module de complexité ne peut toutefois pas être arbitrairement petit. D'une part, il doit être supérieur à 1 (proposition 16.28), ce que ne suggère pas sa définition.

Proposition 16.28 (minorant de ϑ_f) *Si $f \in \text{BAC}(\mathbb{E})$, alors $\vartheta_f \geq 1$.*

DÉMONSTRATION. On peut trouver un $x \in \text{dom } f$ tel que $g_x(x) \neq 0$ car une fonction autoconcordante ne peut être constante (elle a une hessienne définie positive). Alors $y := x + tg_x(x)/\|g_x(x)\|_x$ est dans le [domaine](#) de f pour tout $t \in [0, 1[$ (on sait en effet que $B_x(x, 1) \subseteq \text{dom } f$ lorsque f est autoconcordante). En utilisant ce y dans l'inégalité de droite dans (16.15), on obtient

$$\vartheta_f > t \frac{\langle g(x), g_x(x) \rangle}{\|g_x(x)\|_x} = t \|g_x(x)\|_x.$$

Le supremum sur les $x \in \text{dom } f$ tels que $g_x(x) \neq 0$ et sur les $t \in [0, 1[$ à droite conduit à $\vartheta_f \geq \vartheta_f^{1/2}$ ou $\vartheta_f \geq 1$. \square

Existence de barrière autoconcordante

Nous admettrons le théorème fondamental suivant [455; Theorem 2.5.1]. Une *droite* fait référence ici à un *sous-espace affine* de dimension 1.

Théorème 16.29 (fondamental) *Il existe une constante $\nu > 0$ telle que pour tout entier $n \geq 1$ et pour tout ouvert convexe Ω de \mathbb{R}^n ne contenant pas de droite, il existe une fonction $f \in \text{BAC}(\mathbb{R}^n)$ telle que $\text{dom } f = \Omega$ et $\vartheta_f \leq \nu n$.*

16.1.4 Trois approches algorithmiques ▲

Problème et méthodes

On considère le problème

$$(P) \quad \left\{ \begin{array}{l} \inf_{x \in X} \langle c, x \rangle \end{array} \right.$$

dans lequel $c \in \mathbb{E}$ et $X \subseteq \mathbb{E}$ est l'adhérence du *domaine* d'une fonction $f \in \text{BAC}(\mathbb{E})$. On note

$$\text{val}(P)$$

la valeur optimale de (P) . On cherche à résoudre ce problème par des méthodes de points intérieurs, générant donc des itérés dans l'intérieur supposé non vide de X .

Comme l'ensemble admissible X n'est pas donné par des contraintes fonctionnelles, les méthodes de points intérieurs pour résoudre (P) sont nécessairement primales, dans le sens où elles ne s'intéressent qu'à la variable primaire x (à la différence des méthodes de points intérieurs primales-duales des chapitres 18 et 20, qui génèrent conjointement des itérés primaux et des multiplicateurs).

Définition 16.30 On appelle *chemin central* du problème (P) , noté \mathcal{C} , l'image de l'application $(\mu > 0) \mapsto x(\mu) \in \text{dom } f$, où $x(\mu)$ est le minimum (s'il existe) de la fonction f_μ définie par

$$f_\mu(x) = \frac{1}{\mu} \langle c, x \rangle + f(x).$$

□

S'il existe, le minimum de la fonction strictement convexe f_μ est nécessairement unique. Comme nous avons supposé que $f \in \text{BAC}(\mathbb{E})$, $f_\mu \in \text{AC}(\mathbb{E})$ (proposition 16.8), mais f_μ peut ne pas être dans $\text{BAC}(\mathbb{E})$ (contre-exemple 16.23).

Les algorithmes de points intérieurs se servent du chemin central comme d'un guide qui va conduire les itérés vers une solution du problème. Cette stratégie est motivée par le fait que la valeur du critère dans (P) décroît vers la valeur optimale le long du chemin central (lorsque $\mu \downarrow 0$). La décroissance de la valeur optimale de f_μ est assurée par la proposition 13.2, tandis que sa convergence vers la valeur optimale du problème (P) se déduit clairement de l'estimation suivante :

$$\langle c, x(\mu) \rangle \leq \text{val}(P) + \vartheta_f \mu. \tag{16.20}$$

Démontrons cette inégalité. Par optimalité, $\nabla f_\mu(x(\mu)) = 0$, ce qui s'écrit

$$-\frac{1}{\mu} c = g(x(\mu)).$$

En prenant le produit scalaire avec $y - x(\mu)$ pour $y \in \text{dom } f$, on a grâce à la proposition 16.25 :

$$-\frac{1}{\mu} \langle c, y - x(\mu) \rangle = \langle g(x(\mu)), y - x(\mu) \rangle < \vartheta_f.$$

Dès lors, $\langle c, x(\mu) \rangle < \langle c, y \rangle + \vartheta_f \mu$, pour tout $y \in \text{dom } f$. On en déduit (16.20).

C'est la direction de Newton en x sur f_μ , pour des paramètres μ décroissant vers zéro, qui joue un rôle important dans les algorithmes décrits ci-après. Elle est notée et définie par

$$d_\mu(x) := -H(x)^{-1} \left(\frac{1}{\mu} c + g(x) \right). \quad (16.21)$$

On aura aussi besoin d'imposer aux itérés x de se trouver suffisamment proche d'un point central $x(\mu)$. Le contrôle de cette proximité se fera au moyen d'un voisinage qu'il est agréable de définir à partir de la direction de Newton $d_\mu(x)$, parce que celle-ci est calculée dans les algorithmes. Pour $\mu > 0$ et $\alpha > 0$, on définit le voisinage suivant de $x(\mu)$:

$$V_\mu(\alpha) := \{x \in \text{dom } f : \|d_\mu(x)\|_x \leq \alpha\}. \quad (16.22)$$

Le voisinage $V(\alpha)$ du chemin central est alors la réunion de ces $V_\mu(\alpha)$ pour tous les $\mu > 0$.

Si f_μ n'a pas de minimum, le voisinage $V_\mu(\alpha)$ peut être vide. En fait, d'après la proposition 16.14, supposer $V_\mu(\alpha) \neq \emptyset$, avec $\alpha \in]0, \frac{1}{4}[$, assure du même coup l'existence de $x(\mu)$. Cette hypothèse sera implicite dans les algorithmes décrits ci-après.

Algorithme des petits déplacements

L'algorithme des petits déplacements relève du principe de la *poursuite d'un objectif fuyant*, ce que l'on a schématisé à la figure 16.2 : en l'itéré courant x , supposé

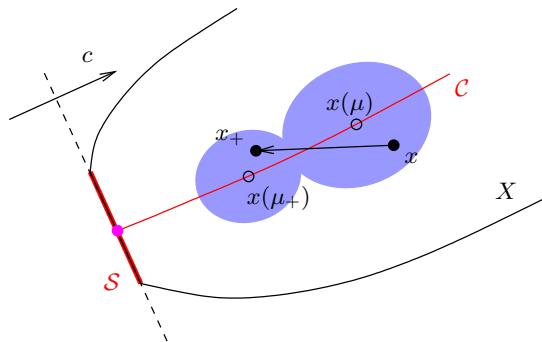


Fig. 16.2. Algorithme des petits déplacements

proche d'un point central $x(\mu)$, on se fixe pour objectif un autre point central $x(\mu_+)$ avec $\mu_+ < \mu$, et l'on fait un déplacement de Newton dans sa direction (donc sur f_{μ_+}), ce qui conduit au nouvel itéré x_+ . On n'a pas besoin de connaître $x(\mu_+)$; seuls μ_+ et f_{μ_+} sont utiles au calcul du déplacement de Newton. Le processus se répète jusqu'à ce que μ soit assez petit devant ϑ_f ou devant un majorant de ϑ_f , que l'on suppose connu. L'inégalité (16.20) permet d'avoir une estimation de l'erreur sur le coût sans avoir besoin de connaître la valeur optimale $\text{val}(P)$, pourvu que ϑ_f soit connu. La proximité du point central $x(\mu)$ se fait au moyen du voisinage $V_\mu(\alpha)$ avec $\alpha = 1/9$. Quant à la décroissance de μ , elle se fait simplement par la règle $\mu_+ = \sigma\mu$, avec la constante $\sigma \in]0, 1[$ bien choisie.

On peut à présent décrire de manière précise, l'algorithme dont on montrera la polynomialité.

Algorithme 16.31 (PIC petits déplacements — une itération)

On se donne $\mu > 0$ et $\alpha := 1/9$. On suppose que l'itéré courant $x \in V_\mu(\alpha)$. L'itéré suivant x_+ est calculé comme suit.

1. Mise à jour du paramètre $\mu_+ = \sigma\mu$, avec

$$\sigma = \frac{8\vartheta_f^{1/2}}{1 + 8\vartheta_f^{1/2}}. \quad (16.23)$$

2. Calcul de la direction de Newton $d_{\mu_+}(x)$.
3. Nouvel itéré : $x_+ = x + d_{\mu_+}(x)$.

Cet algorithme fait usage de ϑ_f , qu'il faut donc connaître. On pourrait aussi utiliser un majorant θ de ϑ_f , ce qui aurait pour effet d'augmenter σ , mais alors la complexité itérative de l'algorithme serait en $O(\theta^{1/2})$ au lieu de $O(\vartheta_f^{1/2})$.

Théorème 16.32 (convergence et complexité de l'algorithme des petits déplacements) *L'algorithme 16.31 est bien défini, converge ($\mu_k \rightarrow 0$ *quasiment linéairement*) et pour tout $\varepsilon \in]0, 1[$, il existe un indice $K := \lceil 9\vartheta_f^{1/2} \log \varepsilon^{-1} \rceil$ tel que $\mu_k \leq \varepsilon\mu_0$, dès que $k \geq K$.*

DÉMONSTRATION. L'algorithme sera bien défini si on montre que $x_+ \in V_{\mu_+}(\alpha)$, c'est-à-dire si $\|d_{\mu_+}(x_+)\|_{x_+} \leq \alpha$. La proposition 16.11 permet d'estimer la norme précédente en fonction de $\|d_{\mu_+}(x)\|_x$. Commençons donc par estimer cette dernière en fonction de $\|d_\mu(x)\|_x$ que l'on sait être $\leq \alpha$, par récurrence. On a

$$d_{\mu_+}(x) = -H(x)^{-1} \left(\frac{1}{\mu_+} c + g(x) \right) = \frac{\mu}{\mu_+} d_\mu(x) + \left(\frac{\mu}{\mu_+} - 1 \right) g_x(x).$$

En utilisant $\mu_+ = \sigma\mu$, $\|d_\mu(x)\|_x \leq \alpha$, $\|g_x(x)\|_x \leq \vartheta_f^{1/2}$, les valeurs de $\alpha = 1/9$ et de σ utilisées par l'algorithme et $\vartheta_f \geq 1$ (proposition 16.28), on obtient

$$\|d_{\mu_+}(x)\|_x \leq \frac{\alpha}{\sigma} + \left(\frac{1}{\sigma} - 1\right) \vartheta_f^{1/2} \leq \frac{1}{9} \left(\frac{1}{8\vartheta_f^{1/2}} + 1\right) + \frac{1}{8} \leq \frac{1}{4}.$$

Comme $f_{\mu_+} \in \text{AC}(\mathbb{E})$ (proposition 16.8) et $\|d_{\mu_+}(x)\|_x < 1$, on peut utiliser la proposition 16.11, qui conduit à

$$\|d_{\mu_+}(x_+)\|_{x_+} \leq \left(\frac{\|d_{\mu_+}(x)\|_x}{1 - \|d_{\mu_+}(x)\|_x}\right)^2 \leq \frac{1}{9} = \alpha.$$

L'algorithme est donc bien défini et $\mu_{k+1} = \sigma\mu_k$, avec $\sigma < 1$, montre que $\mu_k \rightarrow 0$ linéairement.

Il reste à démontrer la polynomialité. On a

$$\begin{aligned} \log \mu_k &= \log \mu_{k-1} - \log \left(1 + \frac{1}{8\vartheta_f^{1/2}}\right) \\ &\leq \log \mu_{k-1} - \frac{1/(8\vartheta_f^{1/2})}{1 + 1/(8\vartheta_f^{1/2})} \quad [\text{car } \log(1+t) \geq \frac{t}{1+t}] \\ &\leq \log \mu_0 - \frac{k}{1 + 8\vartheta_f^{1/2}} \quad [\text{par récurrence}]. \end{aligned}$$

Dès lors, $\log(\mu_k/\mu_0) \leq \log \varepsilon$, dès que $k \geq K := \lceil 9\vartheta_f^{1/2} \log \varepsilon^{-1} \rceil \geq (1 + 8\vartheta_f^{1/2}) \log \varepsilon^{-1}$ (on se rappelle que $\vartheta_f \geq 1$). \square

Algorithme des grands déplacements

L'inconvénient de l'algorithme des petits déplacements est de réduire μ très lentement, puisqu'on le multiplie à chaque itération par le facteur σ donné par (16.23), qui est proche de 1 lorsque ϑ_f est grand. Dans l'algorithme des grands déplacements, on décide de réduire μ par un facteur σ arbitrairement petit à chaque étape de l'algorithme : $\mu_+ := \sigma\mu$. Fort bien. Mais il n'est alors plus garanti que $x + d_{\mu_+}(x)$ soit dans le voisinage $V_{\mu_+}(\alpha)$ de $x(\mu_+)$ lorsque $x \in V_\mu(\alpha)$. Or on a besoin que le nouvel itéré x_+ soit dans $V_{\mu_+}(\alpha)$ pour qu'au début de l'étape suivante on se retrouve dans les mêmes conditions qu'à l'étape courante. Dans l'algorithme des grands déplacements, on obtient $x_+ \in V_{\mu_+}(\alpha)$ en faisant plusieurs itérations de Newton amorties, avec des pas déterminés par recherche linéaire sur f_{μ_+} . L'algorithme est schématisé à la figure 16.3. Cet algorithme génère donc des itérés intermédiaires x_i . On appellera *itération* l'ensemble des opérations menant d'un itéré intermédiaire au suivant et *étape* l'ensemble des itérations conduisant de x à x_+ et correspondant à une réduction de μ en $\mu_+ = \sigma\mu$. Voici une description précise d'une étape de l'algorithme.

Algorithme 16.33 (PIC grands déplacements — une étape)

Soient $\alpha = 1/4$ et $\sigma \in]0, 1[$ deux constantes indépendantes de l'étape considérée. On suppose que l'itéré courant $x \in V_\mu(\alpha)$, pour un $\mu > 0$. L'itéré suivant $x_+ \in V_{\mu_+}(\alpha)$ est calculé comme suit.

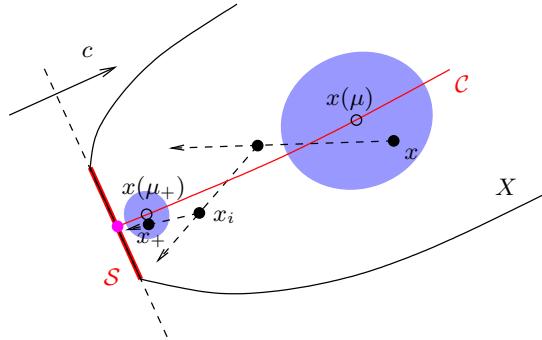


Fig. 16.3. Algorithme des grands déplacements

1. Mise à jour du paramètre: $\mu_+ = \sigma\mu$.
 2. Itérations de Newton avec recherche linéaire.
 - 2.1. $x_0 := x, i := 0$.
 - 2.2. Tant que $x_i \notin V_{\mu_+}(\alpha)$:
 - calcul du pas $t_i > 0$ qui minimise $t \mapsto f_{\mu_+}(x_i + td_{\mu_+}(x_i))$,
 - $x_{i+1} := x_i + t_id_{\mu_+}(x_i)$,
 - $i := i + 1$.
 3. Nouvel itéré: $x_+ := x_i, i(x) := i$.
-

La difficulté principale de l'analyse de l'algorithme est de montrer que le nombre $i(x)$ d'itérés intermédiaires à chaque étape est borné (par une constante ne dépendant que de ϑ_f et de σ). La complexité polynomiale de l'algorithme sera une conséquence aisée de ce fait.

Théorème 16.34 (convergence et complexité de l'algorithme des grands déplacements) *L'algorithme 16.33 est bien défini, converge ($\mu_k \rightarrow 0$ q-linéairement) et pour tout $\varepsilon \in]0, 1[$, on a $\mu_k \leq \varepsilon\mu_0$, dès que le nombre d'itérations de Newton est supérieur à $(80\vartheta_f \log_\sigma \varepsilon)/\sigma$.*

DÉMONSTRATION. 1) Première étape : la décroissance maximale que l'on peut obtenir sur f_{μ_+} à partir de $x \in V_{\mu}(\alpha)$ est bornée :

$$f_{\mu_+}(x) - f_{\mu_+}(x(\mu_+)) \leq \frac{2\vartheta_f}{\sigma}. \quad (16.24)$$

Cette inégalité s'obtient en deux temps, en estimant la décroissance de f_{μ_+} de x à $x(\mu)$ puis de $x(\mu)$ à $x(\mu_+)$:

$$f_{\mu_+}(x) - f_{\mu_+}(x(\mu)) \leq \frac{\sqrt{\vartheta_f}}{\sigma} \quad (16.25)$$

$$f_{\mu_+}(x(\mu)) - f_{\mu_+}(x(\mu_+)) \leq \left(\frac{1}{\sigma} - 1\right)\vartheta_f. \quad (16.26)$$

En sommant ces deux inégalités et en tenant compte du fait que $\vartheta_f \geq 1$ (proposition 16.28), on obtient (16.24).

Pour établir (16.25), on commence par exprimer la convexité de f_{μ_+} :

$$f_{\mu_+}(x) - f_{\mu_+}(x(\mu)) \leq f'_{\mu_+}(x) \cdot (x - x(\mu)) = \langle d_{\mu_+}(x), x(\mu) - x \rangle_x.$$

Pour estimer le membre de droite, on note que $\|x - x(\mu)\|_x \leq 3/4$ lorsque $x \in V_\mu(1/4)$ (proposition 16.14) et on exprime $d_{\mu_+}(x)$ en fonction de $d_\mu(x)$, dont on connaît la grandeur :

$$d_{\mu_+}(x) = \frac{\mu}{\mu_+} d_\mu(x) + \left(\frac{\mu}{\mu_+} - 1 \right) g_x(x) = \frac{1}{\sigma} d_\mu(x) + \left(\frac{1}{\sigma} - 1 \right) g_x(x). \quad (16.27)$$

On obtient

$$\begin{aligned} f_{\mu_+}(x) - f_{\mu_+}(x(\mu)) &\leq \|d_\mu(x)\|_x \|x - x(\mu)\|_x \\ &\leq \left(\frac{1}{\sigma} \|d_\mu(x)\|_x + \left(\frac{1}{\sigma} - 1 \right) \|g_x(x)\|_x \right) \frac{3}{4} \\ &\leq \left(\frac{1}{\sigma} \frac{1}{4} + \left(\frac{1}{\sigma} - 1 \right) \sqrt{\vartheta_f} \right) \frac{3}{4} \\ &\leq \frac{1}{\sigma} \frac{1}{4} \sqrt{\vartheta_f} + \frac{1}{\sigma} \frac{3}{4} \sqrt{\vartheta_f} \quad [\text{car } \vartheta_f \geq 1] \\ &= \frac{\sqrt{\vartheta_f}}{\sigma}. \end{aligned}$$

Pour établir (16.26), on commence également par exprimer la convexité de f_{μ_+} :

$$\begin{aligned} f_{\mu_+}(x(\mu)) - f_{\mu_+}(x(\mu_+)) &\leq f'_{\mu_+}(x(\mu)) \cdot (x(\mu) - x(\mu_+)) \\ &= \langle d_{\mu_+}(x(\mu)), x(\mu_+) - x(\mu) \rangle_{x(\mu)}. \end{aligned}$$

Comme précédemment, la direction de Newton $d_{\mu_+}(x(\mu))$ peut s'exprimer en fonction de $d_\mu(x(\mu))$, qui est nulle car $x(\mu)$ minimise f_μ , grâce à (16.27) avec x remplacé par $x(\mu)$:

$$d_{\mu_+}(x(\mu)) = \left(\frac{1}{\sigma} - 1 \right) g_{x(\mu)}(x(\mu)).$$

En utilisant l'inégalité de droite dans (16.15), on trouve bien (16.26) :

$$f_{\mu_+}(x(\mu)) - f_{\mu_+}(x(\mu_+)) \leq \left(\frac{1}{\sigma} - 1 \right) \langle g_{x(\mu)}(x(\mu)), x(\mu_+) - x(\mu) \rangle_{x(\mu)} \leq \left(\frac{1}{\sigma} - 1 \right) \vartheta_f.$$

2) Deuxième étape : tant que $x_i \notin V_{\mu_+}(\alpha)$, f_{μ_+} décroît au moins de $1/40$ lors de chaque itération de Newton :

$$x_i \notin V_{\mu_+}(1/4) \implies f_{\mu_+}(x_{i+1}) \leq f_{\mu_+}(x_i) - \frac{1}{40}.$$

En effet, $f_{\mu_+}(x_{i+1}) \leq f_{\mu_+}(x_i + \tilde{t}_i d_{\mu_+}(x_i))$ avec le pas particulier $\tilde{t}_i := (5\|d_{\mu_+}(x_i)\|_{x_i})^{-1}$ puisque le pas t_i minimise $t \mapsto f_{\mu_+}(x_i + t d_{\mu_+}(x_i))$ et que $\tilde{t}_i \|d_{\mu_+}(x_i)\|_{x_i} < 1$ implique

que $x_i + \tilde{t}_i d_{\mu_+}(x_i) \in \text{dom } f$. La proposition 16.12 permet alors de développer $f_{\mu_+}(x_i + \tilde{t}_i d_{\mu_+}(x_i))$:

$$\begin{aligned} f(x_{i+1}) &\leq f_{\mu_+}(x_i + \tilde{t}_i d_{\mu_+}(x_i)) \\ &\leq f_{\mu_+}(x_i) - \tilde{t}_i \|d_{\mu_+}(x_i)\|_{x_i}^2 + \frac{1}{2} \tilde{t}_i^2 \|d_{\mu_+}(x_i)\|_{x_i}^2 + \frac{\tilde{t}_i^3 \|d_{\mu_+}(x_i)\|_{x_i}^3}{3(1 - \tilde{t}_i \|d_{\mu_+}(x_i)\|_{x_i})} \\ &= f_{\mu_+}(x_i) - \frac{1}{5} \|d_{\mu_+}(x_i)\|_{x_i} + \frac{1}{50} + \frac{1}{300} \\ &\leq f_{\mu_+}(x_i) - \frac{1}{20} + \frac{1}{50} + \frac{1}{300} \quad [\|d_{\mu_+}(x_i)\|_{x_i} \geq \frac{1}{4} \text{ lorsque } x_i \notin V_{\mu_+}(1/4)] \\ &\leq f_{\mu_+}(x_i) - \frac{1}{40}. \end{aligned}$$

3) *Troisième étape : le nombre d’itérations de Newton par étape de l’algorithme est borné.* En effet, en utilisant les résultats obtenus dans les deux premières étapes de la démonstration, on voit que, tant que $x_{i-1} \notin V_{\mu_+}(1/4)$, on a

$$\frac{i}{40} \leq f_{\mu_+}(x) - f_{\mu_+}(x_i) \leq f_{\mu_+}(x) - f_{\mu_+}(x(\mu_+)) \leq \frac{2\vartheta_f}{\sigma}.$$

Dès lors $i(x) \leq 80\vartheta_f/\sigma$.

4) *Conclusion.* Puisque $x_+ \in V_{\mu_+}(\alpha)$ en un nombre fini d’étapes, l’algorithme est bien défini. De plus, si on note μ_k ($k \geq 0$) les paramètres utilisés à chaque étape de l’algorithme on a évidemment $\mu_k = \sigma\mu_{k-1} = \sigma^k\mu_0$, qui converge donc q-linéairement vers zéro. De plus $\mu_k/\mu_0 \leq \varepsilon$ dès que $\sigma^k \leq \varepsilon$, ce qui s’écrit

$$k \geq \log_\sigma \varepsilon \implies \mu_k \leq \varepsilon\mu_0$$

($\sigma \in]0, 1[$, donc $x \mapsto \log_\sigma x$ est décroissante). Par ailleurs, le nombre total n_k d’itérations de Newton pour terminer les k premières étapes vérifie

$$n_k := \sum_{l=0}^{k-1} i(x_l) \leq \frac{80\vartheta_f}{\sigma} k.$$

On a donc les implications

$$n_k \geq \frac{80\vartheta_f}{\sigma} \log_\sigma \varepsilon \implies k \geq \log_\sigma \varepsilon \implies \mu_k \leq \varepsilon\mu_0. \quad \square$$

La complexité itérative polynomiale de l’algorithme des grands déplacements est donc en $O(\vartheta_f \log_\sigma \varepsilon)$, ce qui est moins bien que celle de l’algorithme des petits déplacements, lequel est en $O(\sqrt{\vartheta_f} \log \varepsilon^{-1})$. Lorsqu’on le particularise à des cas concrets (voir les chapitres 18 et 20), cet algorithme est en pratique plus rapide que l’algorithme des petits déplacements. Il y a donc un écart entre la pratique et la théorie, qui n’est pas comblé à ce jour.

Algorithme prédicteur-correcteur ▲**16.2 Méthodes de points intérieurs en optimisation non linéaire ▲**

On considère, pour simplifier, le problème avec contraintes d'inégalité seulement, celles pour lesquelles les algorithmes de points intérieurs (PI) ont été conçus. On l'écrit sous la forme

$$(P_I) \quad \begin{cases} \min f(x) \\ c(x) \geq 0, \end{cases}$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ sont des fonctions régulières. Le problème peut ne pas être convexe (on ne demande pas que f et les $-c_i$ soient convexes). On note $\lambda \in \mathbb{R}^m$ le multiplicateur associé à la contrainte dans le lagrangien, lequel est la fonction $\ell : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ définie par

$$\ell(x, \lambda) = f(x) - \lambda^\top c(x),$$

où le signe « $-$ » tient compte de la positivité requise des contraintes.

Pourvu que les contraintes soient qualifiées en une solution particulière x_* , il existe un multiplicateur optimal λ_* tel que l'on ait les conditions d'optimalité (KKT) suivantes

$$\begin{cases} \nabla_x \ell(x_*, \lambda_*) = 0 \\ C(x_*) \lambda_* = 0 \\ c(x_*) \geq 0 \quad \text{et} \quad \lambda_* \geq 0. \end{cases} \quad (16.28)$$

On a noté $C(x) := \text{Diag}(c_1(x), \dots, c_m(x))$. On note $L_* := \nabla_{xx}^2 \ell(x_*, \lambda_*)$ la hessienne du lagrangien en $z_* := (x_*, \lambda_*)$ et

$$I^0 := \{i : c_i(x_*) = 0\} \quad \text{et} \quad I^- := \{i : c_i(x_*) < 0\}$$

l'ensemble des indices des contraintes actives et inactives en x_* . Évidemment ces ensembles d'indices forment une partition de $[1:m]$.

Le premier intérêt des méthodes de points intérieurs est leur propension à mieux gérer l'*aspect combinatoire* des problèmes d'optimisation avec contraintes d'inégalité, c'est-à-dire de trouver la bonne manière d'activer celles-ci parmi les 2^m possibilités. Plus exactement, elles remplacent la combinatoire par du *mauvais conditionnement*, celui des systèmes linéaires à résoudre. La gestion de ce mauvais conditionnement relève de l'algèbre linéaire et est en général plus simple à prendre en compte.

16.2.1 Un algorithme de PI primal-dual**Schéma algorithmique**

Le problème barrière associé à (P_I) s'écrit, pour un paramètre de pénalisation $\mu > 0$:

$$\min \left(f_\mu(x) := f(x) - \mu \sum_{i=1}^m \log c_i(x) \right).$$

Dans celui-ci la contrainte $c(x) > 0$ est implicite (le critère est considéré comme prenant la valeur $+\infty$ si $c(x) \not> 0$). Ce problème n'a donc de sens que s'il existe un

point strictement admissible, ce qui sera le cas sous les hypothèses 16.36 ci-dessous. Un point stationnaire x de f_μ est strictement admissible et vérifie, pour un $\lambda \in \mathbb{R}^m$, les conditions suivantes :

$$\begin{cases} \nabla_x \ell(x, \lambda) = 0 \\ C(x)\lambda = \mu e. \end{cases} \quad (16.29)$$

Ce sont les conditions d'optimalité (16.28), dans lesquelles les conditions de complémentarité ont été perturbées et les contraintes $c(x) > 0$ et $\lambda > 0$ sont implicites. L'ensemble

$$\mathcal{C} := \{z : F(z, \mu) = 0 \text{ pour un certain } \mu > 0\}$$

est appelé le *chemin central primal-dual*.

Même si les fonctions f et $-c_i$ sont convexes, la fonction f_μ ci-dessus n'est en général pas une fonction autoconcordante. Cela n'empêche pas de l'utiliser en pratique pour concevoir des algorithmes efficaces. Cependant, si l'on ne suppose que la continuité et la convexité de f et des $-c_i$, le chemin central pourra avoir une allure surprenante, parfois indésirable. Ainsi, dans \mathbb{R}^2 , avec une fonction f convexe continue et la contrainte $x_2 \geq 0$, on peut avoir les chemins centraux de la figure 16.4. Les deux

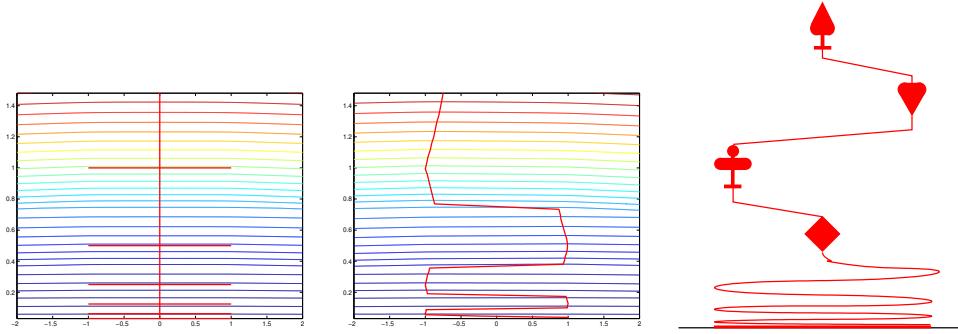


Fig. 16.4. Exemples de chemins centraux indésirables

premiers ont été dessinés avec une fonction f convexe continue, mais peuvent aussi se présenter si f est C^∞ [241]. Dans celui de gauche en forme d'antenne, il y a un nombre infini de segments horizontaux qui s'accumulent sur la droite $x_2 = 0$. Celui du milieu en zigzag a un nombre infini de tournants. De tels chemins centraux ne peuvent pas se présenter si les fonctions f et $-c_i$ sont convexes et analytiques [134, 433] ou si le lagrangien est fortement convexe pour un jeu de multiplicateurs [20]. Si la fonction f n'est que convexe continue, on peut avoir un chemin central comme dans le tracé de droite [33]. Ces exemples suggèrent que l'approche par points intérieurs n'est pas toujours appropriée.

On fera souvent référence à l'application $F : \mathbb{R}^{n+m} \times \mathbb{R} \rightarrow \mathbb{R}^{n+m}$ associée à (16.29) et définie en $z = (x, \lambda)$ et μ par

$$F(z, \mu) = \begin{pmatrix} \nabla_x \ell(x, \lambda) \\ C(x)\lambda - \mu e \end{pmatrix}.$$

Alors (16.29) se récrit $F(z, \mu) = 0$ tandis que z_* vérifie $F(z_*, 0) = 0$. Observons que

$$F'_z(z, 0) = F'_z(z, \mu) = \begin{pmatrix} \nabla_{xx}^2 \ell(x, \lambda) & -c'(x)^T \\ Ac'(x) & C(x) \end{pmatrix}.$$

L'algorithme que nous allons étudier utilise des directions de Newton en $z = (x, \lambda)$ sur l'équation $F(z, \mu) = 0$. On la note $d_\mu(z)$ et elle est définie par le système linéaire

$$F'_z(z, \mu) d_\mu(z) = -F(z, \mu).$$

D'autre part, il génère des itérés z dans le voisinage $V := \cup_{\mu>0} V_\mu$ du chemin central, où pour $\mu > 0$ fixé

$$\begin{aligned} V_\mu := \left\{ (x, \lambda) \in \mathbb{R}^{n+m} : \quad & c(x) > 0, \lambda > 0, \right. \\ & \left. \|\nabla_x \ell(x, \lambda)\|_2 \leq \varepsilon^D(\mu), \|C(x)\lambda - \mu e\|_2 \leq \varepsilon^C(\mu) \right\}. \end{aligned} \quad (16.30)$$

La dépendance en μ des seuils $\varepsilon^D(\cdot) > 0$ et $\varepsilon^C(\cdot) > 0$ sera déterminée au cours de l'analyse pour que, asymptotiquement, l'algorithme soit bien posé et génère une suite convergeant rapidement.

Précisons à présent une itération de l'algorithme.

Algorithme 16.35 (PINL asymptotique — une itération)

Asymptotiquement, une itération de l'algorithme, de $z := (x, \lambda)$ à $z_+ := (x_+, \lambda_+)$ se déroule comme suit. On suppose au départ que $z \in V_\mu$.

1. Mise à jour : $\mu_+ := \phi(\mu)$.
2. Calcul de la direction de Newton $d_{\mu_+}(z)$.
3. $z_+ := z + d_{\mu_+}(z)$.

Cet algorithme fonctionne donc de manière assez semblable à l'algorithme 16.31 des petits déplacements et la figure 16.2 peut aussi le schématiser (sauf qu'ici les hypothèses 16.36 impliquent que l'ensemble \mathcal{S} des solutions est *localement* réduit au point x_*). La différence essentielle vient de la définition de la direction de Newton qui est ici primale-duale (calculée sur $F(z, \mu_+) = 0$), alors qu'elle est primale dans l'algorithme des petits déplacements (calculée sur $\min f_{\mu_+}$). L'analyse et l'expérience ont montré qu'une direction primale-duale donnait de meilleurs résultats (théoriques et pratiques) qu'une direction primale, mais on n'avait pas d'autre choix pour résoudre le problème général (16.1), qui ne présente pas de multiplicateur de manière naturelle. Une autre différence porte sur la définition des voisinages V_μ , qui se fait ici à partir des conditions d'optimalité plutôt que de la direction de Newton. Cette différence n'est pas essentielle car, comme on le verra au lemme 16.40, ces deux quantités sont semblables (\sim) dans le voisinage d'une solution z_* vérifiant les hypothèses 16.36. Dans le cas non linéaire non convexe qui nous occupe ici, loin de z_* , on utilise souvent plusieurs itérations de Newton avec μ fixé ou les règles de mise à jour de ce paramètre sont telles que celui-ci varie peu lorsque l'itéré courant est éloigné du chemin central.

16.2.2 Analyse asymptotique

Dans cette section, le problème (P_I) et sa solution primale-duale $z_* = (x_*, \lambda_*)$ considérée sont supposés vérifier les hypothèses fortes suivantes.

Hypothèses 16.36

- (H₁) Régularité : les fonctions f et c sont C^2 dans un voisinage de la solution x_* .
- (H₂) (QC-IL) : $c'_{I^0}(x_*)$ est surjective,
- (H₃) Complémentarité stricte : pour tout i , $c_i(x_*) + (\lambda_*)_i > 0$,
- (H₄) (CS2) : $u^\top L_* u > 0$ pour toute direction non nulle u dans le noyau de $c'_{I^0}(x_*)$.

Dans l'analyse asymptotique de cette section, il sera nécessaire de comparer entre elles des grandeurs dépendant de $\mu > 0$, disons $u \equiv u(\mu)$ et $v \equiv v(\mu)$, lorsque $\mu \rightarrow 0$ (il s'agit ici d'une analyse asymptotique). On notera $u = O(v)$ s'il existe une constante C , indépendante de μ , telle que $\|u(\mu)\| \leq C\|v(\mu)\|$ pour tout $\mu > 0$ petit. On notera $u \sim v$ si $u = O(v)$ et $v = O(u)$. Enfin, on notera $u = o(v)$, s'il existe une fonction $\mu \mapsto \varepsilon(\mu)$ telle que $\varepsilon(\mu) \rightarrow 0$ et $\|u(\mu)\| \leq \varepsilon(\mu)\|v(\mu)\|$ lorsque $\mu \rightarrow 0$.

Comme le montre la proposition ci-dessous, sous les hypothèses 16.36, le chemin central se présente dans le voisinage de z_* comme une courbe régulière, que l'on note $\mu \mapsto z(\mu)$, définie pour μ voisin de zéro. En particulier, pour $\mu > 0$, la courbe est formée de points strictement admissibles. Par contre, pour de grand $\mu > 0$, rien n'assure que $F(x, \mu) = 0$ ait une unique solution (il peut ne pas y en avoir ou y en avoir plusieurs).

Proposition 16.37 (le chemin central) *Si les hypothèses 16.36 ont lieu, alors $F'_z(z_*, 0)$ est inversible et, dans un voisinage de z_* , le chemin central est une courbe de classe C^1 . De plus*

$$(z(\mu) - z_*) \sim \mu.$$

DÉMONSTRATION. On a

$$F'_z(z_*, 0) = \begin{pmatrix} L_* & -A_*^\top \\ A_* A_* & C_* \end{pmatrix},$$

où $A_* = c'(x_*)$, $A_* = \text{Diag}(\lambda_*)$ et $C_* = \text{Diag}(c(x_*))$. La première partie de la proposition sera une conséquence du théorème des fonctions implicites (théorème C.14) si l'on montre que $F'_z(z_*, 0)$ est inversible.

Soit (d, ν) dans le noyau de $F'_z(z_*, 0)$. La composante $i \in I^0$ de la seconde équation s'écrit $(\lambda_*)_i c'_i(x_*) d = 0$. Comme $(\lambda_*)_{I^0} > 0$ par complémentarité stricte, on voit que $d \in \mathcal{N}(c'_{I^0}(x_*))$. Les composantes $i \in I^-$ de la seconde équation impliquent par ailleurs que $\nu_{I^-} = 0$, si bien que la première équation devient $L_* d - c'_{I^0}(x_*)^\top \nu_{I^0} = 0$. En multipliant cette dernière par d , on trouve $d^\top L_* d = 0$ et donc $d = 0$ par la CS2. Alors, cette même équation et (QC-IL) montrent que $\nu_{I^0} = 0$, donc $\nu = 0$.

Venons-en à la seconde partie de la proposition. On note

$$\tilde{e} = (0, e) \in \mathbb{R}^n \times \mathbb{R}^m,$$

avec $e = (1, \dots, 1) \in \mathbb{R}^m$. Alors, $\mu\tilde{e} = F(z(\mu), 0) - F(z(\mu), \mu) = F(z(\mu), 0)$, car $F(z(\mu), \mu) = 0$. En faisant un développement de Taylor de $F(z(\mu), 0)$ autour de z_* , on obtient

$$\mu\tilde{e} = F'_z(z_*, 0)(z(\mu) - z_*) + o(z(\mu) - z_*).$$

Comme $F'_z(z_*, 0)$ est inversible et que $\tilde{e} \sim 1$, on en déduit que $(z(\mu) - z_*) \sim \mu$. \square

Le but de l'analyse asymptotique que nous présentons ci-après est de déterminer les fonctions $\varepsilon^D(\cdot)$ et $\varepsilon^C(\cdot)$ et de préciser la règle de décroissance de μ en $\mu_+ = \phi(\mu)$ à l'étape 1 de l'algorithme : cette dernière doit faire tendre μ vers zéro le plus rapidement possible, tout en assurant que $z_+ \in V_{\mu_+}$ lorsque $z \in V_\mu$. Le résultat principal de cette analyse est de montrer que la décroissance de μ peut être sous-quadratique dans l'algorithme 16.35, alors qu'elle était seulement linéaire dans l'algorithme 16.31 des petits déplacements. Évidemment les hypothèses ne sont pas identiques dans les deux cas. L'analyse est asymptotique dans le sens où z est supposé proche de z_* . Nous l'avons répartie sur plusieurs lemmes, de manière à bien localiser les besoins sur $\varepsilon^D(\cdot)$, $\varepsilon^C(\cdot)$ et $\phi(\cdot)$. Ces lemmes cherchent à établir la dépendance en μ de différentes grandeurs afin d'établir que $z - z_*$, tout comme $z(\mu) - z_*$ (proposition 16.37), se comporte comme μ , avec une suite de μ convergeant vers zéro sous-quadratiquement.

Lemme 16.38 (estimation de $c(x)$ et λ dans V_μ) Si (H_1) et (H_3) ont lieu et si z est voisin de z_* et vérifie $\|C(x)\lambda - \mu e\|_2 \leq \varepsilon^C(\mu)$, avec $\varepsilon^C(\mu) \leq \kappa\mu$ et $0 < \kappa < 1$, alors

$$\begin{aligned} i \in I^0 &\implies c_i(x) \sim \mu \quad \text{et} \quad \lambda_i \sim 1, \\ i \in I^- &\implies c_i(x) \sim 1 \quad \text{et} \quad \lambda_i \sim \mu. \end{aligned}$$

DÉMONSTRATION. Par hypothèse, on a pour tout i , $|\lambda_i c_i(x) - \mu| \leq \varepsilon^C(\mu) \leq \kappa\mu$. Dès lors

$$\forall i : \quad (1 - \kappa)\mu \leq \lambda_i c_i(x) \leq (1 + \kappa)\mu. \quad (16.31)$$

Pour $i \in I^0$, $(\lambda_*)_i > 0$ par complémentarité stricte, si bien que $\lambda_i \sim 1$ lorsque $z = (x, \lambda)$ est voisin de z_* ; de (16.31), on déduit que $c_i(x) \sim \mu$. Pour $i \in I^-$, $c_i(x_*) > 0$, si bien que $c_i(x) \sim 1$ lorsque $z = (x, \lambda)$ est voisin de z_* ; de (16.31), on déduit que $\lambda_i \sim \mu$. \square

Lemme 16.39 Si les hypothèses 16.36 ont lieu, si $\mu > 0$ est petit, si $z \in V_\mu$ est voisin de z_* , alors

- 1) $z - z(\mu) = O(\varepsilon^D(\mu)) + O(\varepsilon^C(\mu))$,
- 2) $z - z_* = O(\varepsilon^D(\mu)) + O(\varepsilon^C(\mu)) + O(\mu)$.

DÉMONSTRATION. Par définition de $z(\mu)$ on a $F(z(\mu), \mu) = 0$ et un calcul simple montre que $F'_z(z(\mu), \mu) = F'_z(z(\mu), 0)$. Alors un développement de Taylor de $F(z, \mu)$ autour de $z(\mu)$ donne

$$F(z, \mu) = F'_z(z(\mu), 0)(z - z(\mu)) + o(\|z - z(\mu)\|).$$

Comme $F'_z(z(\mu), 0)$ est d'inverse borné pour μ proche de zéro (proposition 16.37), on en déduit le point 1 :

$$(z - z(\mu)) \sim F(z, \mu) = O(\varepsilon^D(\mu)) + O(\varepsilon^C(\mu)),$$

la dernière estimation provenant du fait que $z \in V_\mu$. Le point 2 se déduit alors de

$$z - z_* = (z - z(\mu)) + (z(\mu) - z_*)$$

du point 1 et du fait que $z(\mu) - z_* = O(\mu)$ (proposition 16.37). \square

Lemme 16.40 (estimation de $d_{\mu_+}(z)$) *Si les hypothèses 16.36 ont lieu et si z est voisin de z_* , alors*

$$d_{\mu_+}(z) \sim F(z, \mu_+).$$

Si, de plus, $z \in V_\mu$ et $0 \leq \mu_+ \leq \mu$, on a

$$d_{\mu_+}(z) = O(\varepsilon^D(\mu)) + O(\varepsilon^C(\mu)) + O(\mu).$$

DÉMONSTRATION. La première estimation vient de

$$d_{\mu_+}(z) = -F'_z(z, \mu_+)^{-1}F(z, \mu_+) = -F'_z(z, 0)^{-1}F(z, \mu_+)$$

et du fait que $F'_z(z, 0)$ est borné et d'inverse borné pour z proche de z_* (proposition 16.37). Par la seconde estimation, on utilise le majorant suivant de $F(z, \mu_+)$:

$$F(z, \mu_+) = \underbrace{F(z, \mu)}_{O(\varepsilon^D(\mu)) + O(\varepsilon^C(\mu))} + \underbrace{(\mu - \mu_+)\tilde{e}}_{\|\cdot\|_2 \leq \mu m^{1/2}}.$$

\square

Le lemme ci-dessous cherche à établir des conditions pour que l'itéré suivant $z_+ := z + d_{\mu_+}(z)$ soit strictement admissible. Une de ces conditions impose que μ_+ ne soit pas trop petit devant μ , ce qui est naturel: il faut que $\mu^2 = o(\mu_+)$, ce qui interdit une décroissance quadratique du paramètre de pénalisation ($\mu_+ = O(\mu^2)$). Elle est réalisée, par exemple, en prenant $\mu_+ = \mu^{2-\tau}$, où $0 < \tau < 1$. On parle de *vitesse de convergence sous-quadratique* de la suite des μ vers zéro.

Lemme 16.41 (admissibilité stricte de z_+) *Si les hypothèses 16.36 ont lieu, si $z \in V_\mu$ est voisin de z_* , si $0 < \mu_+ \leq \mu$ avec μ assez petit et si*

$$\varepsilon^D(\mu) = O(\mu), \quad \varepsilon^C(\mu) \leq \kappa\mu \text{ avec } 0 < \kappa < 1 \quad \text{et} \quad \mu^2 = o(\mu_+),$$

alors $c(x_+) > 0$ et $\lambda_+ > 0$.

DÉMONSTRATION. Observons que

$$d_\mu(z) = -F'_z(z, 0)^{-1}F(z, \mu) = O(F(z, \mu)) = O(\varepsilon^D(\mu)) + O(\varepsilon^C(\mu)) = O(\mu). \quad (16.32)$$

Considérons d'abord les indices $i \in I^0$. On utilise systématiquement les estimations $c_i(x) \sim \mu$ et $\lambda_i \sim 1$ obtenues au lemme 16.38, ainsi que (16.32). Pour λ proche de λ_* , on a

$$(\lambda_+)_i = \underbrace{\lambda_i}_{\sim 1} + \underbrace{d\lambda_i}_{O(\mu)} > 0.$$

Par ailleurs, la i -ième composante de la seconde équation définissant $d_\mu(z) = (dx, d\lambda)$ s'écrit

$$\lambda_i c'_i(x) dx + c_i(x) d\lambda_i = \mu_+ - \lambda_i c_i(x). \quad (16.33)$$

Cette relation permet d'évaluer $c_i(x) + c'_i(x)dx$ et donc d'estimer $c_i(x_+)$ à partir de son développement de Taylor autour de x ($\gamma > 0$ est une constante indépendante de μ) :

$$\begin{aligned} c_i(x_+) &= c_i(x) + c'_i(x)dx + O(\|dx\|^2) \\ &= \underbrace{\lambda_i^{-1}}_{\sim 1} \mu_+ - \underbrace{\lambda_i^{-1}}_{\sim 1} \underbrace{c_i(x)}_{\sim \mu} \underbrace{d\lambda_i}_{O(\mu)} + \underbrace{O(\|dx\|^2)}_{O(\mu^2)} \\ &= \gamma \mu_+ + O(\mu^2) > 0, \end{aligned}$$

grâce à l'hypothèse $\mu^2 = o(\mu_+)$.

On s'y prend de manière semblable pour les indices $i \in I^-$, en utilisant systématiquement les estimations $c_i(x) \sim 1$ et $\lambda_i \sim \mu$ obtenues au lemme 16.38, ainsi que (16.32). Pour x proche de x_* , on a

$$c_i(x_+) = \underbrace{c_i(x)}_{\sim 1} + \underbrace{O(dx)}_{O(\mu)} > 0.$$

En partant de (16.33) et en utilisant l'hypothèse $\mu^2 = o(\mu_+)$, on obtient ($\gamma > 0$ est une constante indépendante de μ) :

$$(\lambda_+)_i = \lambda_i + d\lambda_i = \underbrace{c_i(x)^{-1}}_{\sim 1} \mu_+ - \underbrace{c_i(x)^{-1}}_{\sim 1} \underbrace{\lambda_i}_{\sim \mu} \underbrace{c'_i(x)dx}_{O(\mu)} \geq \gamma \mu_+ + O(\mu^2) > 0.$$

□

Lemme 16.42 (l'algorithme est bien posé) *Si les hypothèses 16.36 ont lieu, si $z \in V_\mu$ est voisin de z_* , si $0 < \mu_+ \leq \mu$ et si*

$$\varepsilon^D(\mu) \sim \mu, \quad \varepsilon^C(\mu) \sim \mu \quad \text{et} \quad \mu^2 = o(\mu_+),$$

alors $F(x_+, \mu_+) = o(\mu_+)$ et $z_+ \in V_{\mu_+}$.

DÉMONSTRATION. Un développement de Taylor, la définition de $d_{\mu_+}(z)$, le lemme 16.31, les hypothèses sur $\varepsilon^D(\mu)$ et $\varepsilon^C(\mu)$ et enfin $\mu^2 = o(\mu_+)$ donnent

$$F(z_+, \mu_+) = \underbrace{F(z, \mu_+) + F'_z(z, \mu_+) d_{\mu_+}(z)}_0 + O(\|d_{\mu_+}(z)\|^2) = O(\mu^2) = o(\mu_+).$$

ce qui démontre la première estimation annoncée. Par ailleurs, $\nabla_x \ell(x_+, \lambda_+)$ et $C(x_+) \lambda_+ - \mu_+ e$, que l'on vient d'estimer à des $o(\mu_+)$, vérifient donc certainement

$$\|\nabla_x \ell(x_+, \lambda_+)\|_2 \leq \varepsilon^D(\mu_+) \quad \text{et} \quad \|C(x_+) \lambda_+ - \mu_+ e\|_2 \leq \varepsilon^C(\mu_+),$$

ce qui veut dire que $z_+ \in V_{\mu_+}$. □

Les lemmes précédents montrent que l'algorithme est bien posé si $\varepsilon^D(\mu)$, $\varepsilon^C(\mu)$ et μ_+ vérifient les règles suivantes :

il existe des constantes $0 < \kappa_1^D \leq \kappa_2^D$ telles que $\kappa_1^D \mu \leq \varepsilon^D(\mu) \leq \kappa_2^D \mu$,

il existe des constantes $0 < \kappa_1^C \leq \kappa_2^C < 1$ telles que $\kappa_1^C \mu \leq \varepsilon^C(\mu) \leq \kappa_2^C \mu$,

$$\mu^2 = o(\mu_+).$$

Il reste à montrer que $(z_+ - z) \sim \mu_+$, ce qui permettra de conclure que la convergence de $z \rightarrow z_*$ est sous-quadratique.

Lemme 16.43 (z converge tangentielle au chemin central) *Si les hypothèses précédentes ont lieu, on a*

$$z_+ = z_* + \mu_+ z'(0) + o(\mu_+), \quad (16.34)$$

où $z'(0)$ est la dérivée de $z(\cdot)$ en zéro.

DÉMONSTRATION. Comme $F'_z(z_*, 0)$ est inversible, en dérivant $F(z(\mu), \mu)$ en $\mu = 0$, on obtient

$$z'(0) = -F'(z_*, 0)^{-1} F'_\mu(z_*, 0) = -F'(z_*, 0)^{-1} \begin{pmatrix} 0 \\ -e \end{pmatrix}. \quad (16.35)$$

Par un développement de Taylor, $F(z_+, \lambda_+) = o(\mu_+)$ (lemme 16.42), on a

$$\underbrace{F(z_+, \lambda_+)}_{o(\mu_+)} = \underbrace{F(z_*, 0)}_0 + F'_z(z_*, 0)(z_+ - z_*) + F'_\mu(z_*, 0)\mu_+ + o(\|z_+ - z_*\|) + o(\mu_+).$$

En multipliant à gauche par $F'(z_*, 0)^{-1}$, en tenant compte de (16.35) et en réarrangeant, on trouve

$$z_+ - z_* = \mu_+ z'(0) + o(\|z_+ - z_*\|) + o(\mu_+).$$

On en déduit que $z_+ - z_* = O(\mu_+)$ et donc le résultat. \square

Proposition 16.44 (convergence sous-quadratique de PINL) *Si les hypothèses 16.36 ont lieu, si $z \in V_\mu$ est voisin de z_* , si $0 < \mu_+ \leq \mu$ et si*

$$\varepsilon^D(\mu) \sim \mu, \quad \varepsilon^C(\mu) \sim \mu \quad \text{et} \quad \mu_+ = O(\mu^{2-\tau}), \quad \text{avec } 0 < \tau < 1,$$

on a $(z_+ - z_) \sim \mu_+$. En particulier, la convergence de z vers z_* est sous-quadratique.*

DÉMONSTRATION. D'après (16.34), $z_+ - z_* = O(\mu_+)$. Réciproquement, en multipliant cette équation (16.34) par $z'(0)$, on trouve

$$\|z'(0)\|_2 \|z_+ - z_*\|_2 \geq z'(0)^\top (z_+ - z_*) = \mu_+ \|z'(0)\|_2^2 + o(\mu_+).$$

D'après (16.35), $z'(0) \neq 0$, si bien que $\mu_+ = O(\|z_+ - z_*\|)$. \square

Notes

La présentation de la section 16.1 suit d'assez près l'approche de Renegar [512; 2001] et peut être vue comme une introduction à ce très bel opuscule. L'ouvrage de référence est celui de Nesterov et Nemirovskii [455; 1993] et nous en avons parfois repris certains résultats non présents chez Renegar (voir aussi Nesterov [452; 2004]). La théorie des fonctions autoconcordantes s'enrichit considérablement lorsque leur **domaine** est l'intérieur d'un cône convexe fermé K . Elle est utile lorsque le problème d'optimisation à résoudre a un ensemble admissible qui est l'intersection de K et d'un **sous-espace affine**. On parle d'*optimisation conique*. On peut alors introduire des algorithmes de points intérieurs primaux-duaux (les variables duales sont associées aux contraintes affines). Les méthodes vues aux chapitres 18 et 20 en sont des cas particuliers. Les bases de l'optimisation conique sont présentées dans les deux ouvrages précités, ainsi que dans [44].

L'analyse asymptotique de la section 16.2.2 est reprise de [271]. Elle peut se mener sous des conditions plus faibles que les hypothèses 16.36, voir [628].

Exercices

16.1. Autre expression de lb . Un *ellipsoïde non dégénéré* centré en zéro est un ensemble de la forme $\mathcal{E}_A = \{v \in \mathbb{R}^n : v^\top A v \leq 1\}$, où $A \succ 0$. Si $x \in \mathbb{R}_{++}^n$, $\text{lb}(x) = \log \sigma_n/V(x)$ où σ_n est le volume de la sphère unité dans \mathbb{R}^n et $V(x)$ est le volume du plus volumineux *ellipsoïde inscrit dans \mathbb{R}_{+}^n* et centré en x . Montrez que cet ellipsoïde est la boule unité $\bar{B}_x(0, 1)$ pour la norme locale en x associée à lb .

16.2. Exemples et contre-exemples de fonctions et de barrières autoconcordantes [452].

- 1) Une fonction quadratique strictement convexe $f : x \in \mathbb{R}^n \mapsto \frac{1}{2}x^\top Ax + b^\top x$, où $A \succ 0$ et $b \in \mathbb{R}^n$, est une fonction autoconcordante, mais n'est pas une barrière autoconcordante.