

Optimisation linéaire algorithmes de points intérieurs

My view of interior point methods for optimization calculations with linear constraints is that it seems silly to introduce nonlinearities and iterative procedures for following central paths, because these complications are not present in the original problem. On the other hand, when the number of constraints is huge, then algorithms that treat constraints individually are also unattractive, especially if the attention to detail causes the number of iterations to be about the number of constraints. It is possible, however, to retain linear constraints explicitly, and to take advantage of the situation where the boundary of the feasible region has so many linear facets that it seems to be smooth. This is done by the TOLMIN software that I developed in 1989, for example, but the number of variables is restricted to a few hundred, because quadratic models with full second derivative matrices are employed. Therefore eventually I expect interior point methods to be best only if the number of variables is large. Another reservation about this field is that it seems to be taking far more than its share of research activity.

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

Ce chapitre se limite à l'étude de quelques aspects des algorithmes de suivi de chemin primal-dual en optimisation linéaire. Ce sont souvent les algorithmes de points intérieurs les plus efficaces en pratique. D'autres approches de points intérieurs ont été proposées et peuvent être utiles dans des contextes particuliers : algorithme de l'*ellipsoïde de Dikin*, algorithmes à réduction de potentiel, *etc.* Nous renvoyons le lecteur aux notes de fin de chapitre pour des références sur ceux-ci.

Comme leur nom l'indique les algorithmes de suivi de chemin génèrent des itérés qui suivent ce que l'on appelle aujourd'hui le *chemin central*. Celui-ci sert de guide, conduisant les itérés vers une solution particulière du problème (ou vers la solution s'il n'y en a qu'une). Nous l'étudions en détail à la section 18.1.

La partie algorithmique de ce chapitre est importante et débute par des remarques générales sur les approches primales-duales (section 18.2). Vient ensuite l'étude de trois algorithmes supposant que l'on dispose d'un premier itéré admissible (section 18.3) : les algorithmes des petits déplacements, des grands déplacements et prédicteur-correcteur. Les algorithmes sont motivés et décrits ; leur convergence et leur complexité itérative sont étudiées. Le premier algorithme a la complexité itérative théorique

la meilleure, mais ce sont les deux derniers qui sont les plus efficaces en pratique, le dernier étant le plus souvent implémenté (dans une version plus élaborée, il est vrai). Enfin, la section 18.4 décrit et analyse un algorithme n'imposant pas aux itérés d'être admissibles.

Connaissances supposées. Les bases de l'optimisation linéaire (chapitre 17); la notion de fonction asymptotique (section 3.3.4), qui est utilisée pour montrer l'existence du chemin central; l'algorithme de Newton (section 10.1.1).

Rappel et notations

On considère le problème d'optimisation linéaire dans \mathbb{R}^n suivant, dit sous forme standard

$$(P) \quad \begin{cases} \inf_x c^\top x \\ Ax = b \\ x \geq 0, \end{cases}$$

où, rappelons-le, $c \in \mathbb{R}^n$, A est une matrice $m \times n$ et $b \in \mathbb{R}^m$. Son ensemble admissible est noté

$$\mathcal{F}_P := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}.$$

L'ensemble des itérés strictement admissibles est noté

$$\mathcal{F}_P^s := \{x \in \mathbb{R}^n : Ax = b, x > 0\}.$$

Lorsqu'il est non vide, c'est l'intérieur relatif de \mathcal{F}_P (exercice 2.20). L'ensemble des solutions de (P) est noté \mathcal{S}_P .

La dualisation lagrangienne de (P) conduit au problème dual suivant (voir la section 17.3.1):

$$(D) \quad \begin{cases} \sup b^\top y \\ A^\top y + s = c \\ s \geq 0. \end{cases}$$

Son ensemble admissible est noté

$$\mathcal{F}_D := \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^\top y + s = c, s \geq 0\}.$$

On note aussi

$$\mathcal{F}_D^s := \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^\top y + s = c, s > 0\},$$

qui, lorsqu'il est non vide, est l'intérieur relatif de \mathcal{F}_D . L'ensemble des solutions de (D) est noté \mathcal{S}_D .

Les conditions d'optimalité de (P) affirment que x est solution de ce problème si, et seulement si, il existe un couple $(y, s) \in \mathbb{R}^m \times \mathbb{R}^n$ tel que l'on ait

$$\begin{cases} A^\top y + s = c, & s \geq 0 \\ Ax = b, & x \geq 0 \\ x^\top s = 0. \end{cases} \quad (18.1)$$

Ce sont aussi les conditions d'optimalité du problème dual: (y, s) est solution de (D) si, et seulement si, il existe $x \in \mathbb{R}^n$ tel que l'on ait (18.1).

Nous désignerons par $X = \text{Diag}(x_1, \dots, x_n)$ la matrice diagonale portant les composantes du vecteur x sur sa diagonale; de même $S = \text{Diag}(s_1, \dots, s_n)$, etc. On utilisera constamment le vecteur

$$e := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

dont la dimension se déduira clairement du contexte. Observons que $Xe = x$.

18.1 Le chemin central primal-dual

Le *chemin central* est une courbe située dans

$$\mathcal{F}^s := \mathcal{F}_p^s \times \mathcal{F}_D^s = \{z = (x, y, s) : Ax = b, A^T y + s = c, x > 0, s > 0\}$$

paramétrée par un scalaire $\mu > 0$. C'est donc l'image d'une application

$$\mu \in]0, +\infty[\mapsto z(\mu) := (x(\mu), y(\mu), s(\mu)) \in \mathcal{F}^s.$$

Celui-ci sert de guide aux itérés, les conduisant vers une solution primale-duale particulière du problème (P) . Pour qu'il soit défini, il faut évidemment que \mathcal{F}_p^s et \mathcal{F}_D^s soient non vides, ce qui, d'après les propositions 17.11 et 17.13, peut s'exprimer en termes primaux seulement :

$$\boxed{\mathcal{F}_p^s \neq \emptyset \text{ et } \mathcal{S}_p \text{ est non vide et borné.}} \tag{18.2}$$

On fera cette hypothèse dans toute cette section. La variable duale $y(\mu)$ ne sera définie de façon univoque que si A est surjective.

Comme son nom l'indique, les points $z(\mu)$ sont « bien centrés » dans l'ensemble admissible. On montrera que, sous certaines conditions, $z(\mu)$ émane (pour $\mu \rightarrow +\infty$) du *centre analytique* de l'ensemble admissible $\mathcal{F} := \mathcal{F}_p \times \mathcal{F}_D$ pour aboutir (lorsque $\mu \downarrow 0$) au *centre analytique de la face optimale primale-duale* $\mathcal{S} := \mathcal{S}_p \times \mathcal{S}_D$.

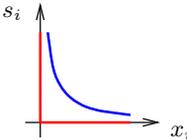
Définition et existence

Il y a plusieurs manières d'introduire le chemin central primal-dual (voir aussi l'exercice 18.5). Nous commencerons par utiliser la technique de perturbation des conditions d'optimalité, qui est celle qui se généralise le plus facilement à des problèmes ne faisant pas partie de l'optimisation (comme les problèmes de complémentarité). L'existence du chemin central est cependant plus aisément démontrable si l'on revient dans le domaine de l'optimisation en interprétant le système perturbé comme les conditions d'optimalité d'un problème pénalisé. Nous considérerons les cas de la pénalisation des problèmes dual et primal.

Les équations d'optimalité (18.1) sont formées de deux équations linéaires $A^T y + s = c$ et $Ax = b$, qui à elles seules ne présentent pas trop de difficulté, ainsi que des conditions

$$x^T s = 0, \quad x \geq 0, \quad s \geq 0.$$

Ces dernières renferment toute la *combinatoire* du problème. Il y a en effet 2^n façons de les réaliser en choisissant, pour tout i , si c'est x_i ou s_i qui est nul. On peut faire disparaître cette combinatoire en perturbant chaque condition $x_i s_i = 0$ en $x_i s_i = \mu$ ($\mu > 0$ est un paramètre) et en imposant la stricte positivité de x et s :

$$\begin{cases} A^T y + s = c, & s > 0 \\ Ax = b, & x > 0 \\ Xs = \mu e. \end{cases} \quad (18.3)$$


On a noté

$$X := \text{Diag}(x_1, \dots, x_n) \quad \text{et} \quad e := (1 \ \dots \ 1)^T.$$

Le graphique à droite dans (18.3) montre comment, dans l'espace (x_i, s_i) , la courbe *non différentiable* définie par $x_i s_i = 0$ est transformée en la courbe *différentiable* définie par $x_i s_i = \mu$. Au passage, cette observation montre le lien qu'il peut y avoir entre combinatoire et non différentiabilité. Lorsqu'elle existe, la solution de ce système perturbé (18.3) est un point central $z(\mu) = (x(\mu), y(\mu), s(\mu))$.

Nous venons d'exposer l'approche par perturbation. On peut aussi obtenir le système (18.3) par une technique de pénalisation. En effet, (18.3) sont les conditions d'optimalité du problème obtenu à partir du problème (P) par pénalisation logarithmique de sa contrainte de positivité :

$$(P_\mu) \quad \begin{cases} \inf c^T x + \mu \text{lb}(x) \\ Ax = b, \end{cases} \quad (18.4)$$

où $\text{lb} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ est la *fonction log-barrière* définie en $x \in \mathbb{R}^n$ par (c'est une fonction autoconcordante, voir l'exemple 16.2)

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

En effet, comme les contraintes de (P_μ) sont qualifiées (elles sont affines), il existe un multiplicateur de Lagrange $y \in \mathbb{R}^m$ tel que l'on ait les conditions d'optimalité suivantes :

$$\begin{cases} A^T y + \mu X^{-1} e = c \\ Ax = b. \end{cases}$$

En posant $s = \mu X^{-1} e$, on retrouve (18.3). On appelle (P_μ) le *problème barrière primal* (les logarithmes dans lb forment une barrière empêchant x de quitter l'*orthant positif*).

Une troisième manière d'introduire le chemin central est de pénaliser la contrainte de positivité $s \geq 0$ du problème dual. Lorsque $\mathcal{F}_D^s \neq \emptyset$, on peut considérer le problème :

$$(D_\mu) \quad \begin{cases} \sup b^T y - \mu \text{lb}(s) \\ A^T y + s = c. \end{cases} \quad (18.5)$$

Comme il s'agit d'un problème de maximisation, la pénalisation est obtenue en retranchant la barrière logarithmique (comparez avec le problème (P_μ)).

Intéressons-nous à présent à des conditions assurant l'existence du chemin central. Le fait que le système d'optimalité (18.1), qui est non linéaire, ait une solution n'assure

en rien qu'il en est de même pour le système perturbé (18.3). Il faut certainement que $\mathcal{F}^s \neq \emptyset$ (équivalente à (18.2)) et il se fait que cette condition est aussi suffisante. Elle joue ici le rôle de la *condition de qualification de Slater* du système (18.1).

Proposition 18.1 (existence du chemin central) *Si (18.2) a lieu et $\mu > 0$, alors le système (18.3) a une solution. Celle-ci est unique en x et s . Elle est également unique en y si A est surjective.*

DÉMONSTRATION. D'après ce qui précède (x, y, s) est solution de (18.3) si, et seulement si, x est solution de (P_μ) de multiplicateur y et si $s = \mu X^{-1}e$ (la condition est suffisante du fait de la convexité du problème (P_μ)). Il reste à montrer que (P_μ) a une solution, car du fait de la stricte convexité de $x \mapsto c^\top x - \mu \sum_{i=1}^n \log x_i$, celle-ci est nécessairement unique en x et en $s = \mu X^{-1}e$.

On considère la fonction $\psi_\mu \in \overline{\text{Conv}}(\mathbb{R}^n)$ définie par

$$\psi_\mu(x) = c^\top x + \mu \text{lb}(x) + \mathcal{I}_{\{x \in \mathbb{R}^n : Ax=b\}}(x),$$

où \mathcal{I}_P désigne la *fonction indicatrice* de l'ensemble P . Ses minimiseurs sont ceux de (P_μ) . Il suffit donc que sa fonction asymptotique ψ_μ^∞ vérifie $\psi_\mu^\infty(d) > 0$ pour tout $d \neq 0$, pour que (P_μ) ait un ensemble non vide (et borné) de solutions (proposition 3.29). On observe que

$$\text{lb}^\infty(d) = \begin{cases} 0 & \text{si } d \geq 0 \\ +\infty & \text{sinon.} \end{cases}$$

Dès lors

$$\psi_\mu^\infty(d) = \begin{cases} c^\top d & \text{si } Ad = 0 \text{ et } d \geq 0 \\ +\infty & \text{sinon.} \end{cases}$$

Si $\psi_\mu^\infty(d) \leq 0$, on a

$$d \neq 0, \quad c^\top d \leq 0, \quad Ad = 0 \quad \text{et} \quad d \geq 0.$$

La bornitude de \mathcal{S}_P entraîne alors que $d = 0$, ce qui démontre l'existence d'un minimiseur de ψ_μ , c'est-à-dire d'une solution de (P_μ) . □

On notera

$$z(\mu) := (x(\mu), y(\mu), s(\mu))$$

un triplet solution de (18.3) (unique en x et s). Le chemin central est donc l'image de l'application

$$\mu \in]0, +\infty[\mapsto z(\mu).$$

Propriétés

Nous allons montrer que, sous certaines conditions, le chemin central émane du centre analytique de l'ensemble admissible \mathcal{F}_P (lorsque $\mu \rightarrow +\infty$ et si \mathcal{F}_P^s est non vide et borné) pour aboutir au centre analytique de la face optimale (lorsque $\mu \downarrow 0$ et si \mathcal{S}_P^o est non vide et borné). Il nous faut d'abord définir ce que sont ces centres analytiques.

Proposition 18.2 (centre analytique de l'ensemble admissible) *Si \mathcal{F}_p^s est non vide et borné, le problème*

$$\begin{cases} \inf \text{lb}(x) \\ Ax = b \end{cases} \quad (18.6)$$

a une solution unique $\tilde{x} \in \mathbb{R}^n$ caractérisée par les conditions suivantes :

$$A\tilde{x} = b, \quad \tilde{x} > 0, \quad \tilde{X}^{-1}e \in \mathcal{R}(A^\top). \quad (18.7)$$

De même, si \mathcal{F}_D^s est non vide et borné et si A est surjective, le problème

$$\begin{cases} \inf \text{lb}(s) \\ A^\top y + s = c \quad (s > 0) \end{cases} \quad (18.8)$$

a une solution unique $(\tilde{y}, \tilde{s}) \in \mathbb{R}^m \times \mathbb{R}^n$ caractérisée par les conditions suivantes :

$$A^\top \tilde{y} + \tilde{s} = c, \quad \tilde{s} > 0, \quad \tilde{S}^{-1}e \in \mathcal{N}(A). \quad (18.9)$$

DÉMONSTRATION. Le problème (18.6) est celui obtenu par pénalisation logarithmique du problème

$$\begin{cases} \inf 0 \\ Ax = b \\ x \geq 0. \end{cases}$$

Son ensemble de solution étant \mathcal{F}_p , qui par hypothèse est non vide et borné, la proposition 18.1 implique que (18.6) a une solution unique, soit $\tilde{x} \in \mathbb{R}^n$. Celle-ci est caractérisée par les conditions d'optimalité de (18.6) qui peuvent s'exprimer par (18.7).

La seconde partie se démontre de la même manière. □

Lorsqu'elles existent, les solutions uniques de (18.6) et (18.8) sont appelées les *centres analytiques* des ensembles admissibles \mathcal{F}_p et \mathcal{F}_D , respectivement. On utilise le qualificatif «analytique», car ce n'est pas une notion géométrique dans le sens où elle ne dépend pas que de la forme de \mathcal{F}_p et \mathcal{F}_D , mais également des équations que définissent ces ensembles. Ainsi, si l'on ajoute une contrainte superflue, c'est-à-dire une contrainte ne modifiant pas l'ensemble admissible, son centre analytique est déplacé (voir l'exercice 18.4). On pourrait penser que la notion géométrique de *centre de gravité* serait un concept plus attrayant, mais son calcul n'est pas aisé, même si l'ensemble est comme ici un polyèdre convexe [633 ; section 2.1.1].

Venons-en maintenant à la notion de centre analytique de la face optimale. On rappelle les notations de la section 17.2.2 :

$$\begin{aligned} \mathfrak{B} &:= \{i \in [1 : n] : \exists x \in \mathcal{S}_p \text{ vérifiant } x_i > 0\} \\ \mathfrak{N} &:= \{i \in [1 : n] : \exists (y, s) \in \mathcal{S}_D \text{ vérifiant } s_i > 0\}. \end{aligned}$$

On sait que \mathfrak{B} et \mathfrak{N} forment une partition de $[1 : n]$ (propositions 17.7 et 17.8) et que l'ensemble des solutions de (P) est la face de \mathcal{F}_p , appelée aussi *face optimale* de \mathcal{F}_p ,

définie par

$$\mathcal{S}_P = \{x \in \mathcal{F}_P : x_{\mathfrak{N}} = 0\}.$$

On note son intérieur relatif

$$\mathcal{S}_P^\circ = \{x \in \mathcal{F}_P : x_{\mathfrak{B}} > 0, x_{\mathfrak{N}} = 0\}. \quad (18.10)$$

Proposition 18.3 (centre analytique de la face optimale) 1) Si \mathcal{S}_P est non vide et borné et si $\mathfrak{B} \neq \emptyset$, le problème

$$\begin{cases} \inf \text{lb}(x_{\mathfrak{B}}) \\ Ax = b, \quad x_{\mathfrak{N}} = 0 \quad (x_{\mathfrak{B}} > 0) \end{cases} \quad (18.11)$$

a une solution unique $\hat{x} \in \mathbb{R}^n$, caractérisée par les conditions

$$A\hat{x} = b, \quad \hat{x}_{\mathfrak{B}} > 0, \quad \hat{x}_{\mathfrak{N}} = 0, \quad \hat{X}_{\mathfrak{B}}^{-1}e \in \mathcal{R}(A_{:\mathfrak{B}}^T). \quad (18.12)$$

2) De même si \mathcal{S}_D est non vide et borné (donc A est surjective) et si $\mathfrak{N} \neq \emptyset$, le problème

$$\begin{cases} \inf \text{lb}(s_{\mathfrak{N}}) \\ A^T y + s = c, \quad s_{\mathfrak{B}} = 0 \quad (s_{\mathfrak{N}} > 0) \end{cases} \quad (18.13)$$

a une solution unique $(\hat{y}, \hat{s}) \in \mathbb{R}^m \times \mathbb{R}^n$, caractérisée par les conditions

$$A^T \hat{y} + \hat{s} = c, \quad \hat{s}_{\mathfrak{B}} = 0, \quad \hat{s}_{\mathfrak{N}} > 0, \quad A_{:\mathfrak{N}} \hat{S}_{\mathfrak{N}}^{-1} e \in \mathcal{R}(A_{:\mathfrak{B}}). \quad (18.14)$$

DÉMONSTRATION. En ne considérant que les variables d'indices dans \mathfrak{B} ($x_{\mathfrak{N}}$ étant fixé à 0), le problème (18.11) est obtenu par pénalisation logarithmique du problème

$$\begin{cases} \inf 0^T x_{\mathfrak{B}} \\ A_{:\mathfrak{B}} x_{\mathfrak{B}} = b \\ x_{\mathfrak{B}} \geq 0. \end{cases}$$

Par l'hypothèse $\mathfrak{B} \neq \emptyset$, $\{x_{\mathfrak{B}} : Ax_{\mathfrak{B}} = b, x_{\mathfrak{B}} > 0\}$ est non vide et l'ensemble des solutions de ce problème (qui se confond avec son ensemble admissible, c'est-à-dire \mathcal{S}_P) est non vide et borné. Par la proposition 18.1, (18.11) a une solution unique. Celle-ci est caractérisée par les conditions d'optimalité de (18.11) qui peuvent s'exprimer par (18.12).

La seconde partie se démontre de la même manière. \square

Lorsqu'ils existent, les points \hat{x} et (\hat{y}, \hat{s}) définis par (18.12) et (18.14) sont appelés les *centres analytiques des faces optimales*.

On peut à présent décrire le chemin central. On sera amené à comparer le comportement de quantités (éventuellement vectorielles) qui dépendent du paramètre μ . Ainsi si $\mu \mapsto v(\mu)$ et $\mu \mapsto w(\mu)$ sont deux fonctions de μ , on notera :

- $v = O(w)$ s'il existe une constante positive C (indépendante de μ) telle que lorsque $\mu \downarrow 0$ on a $v(\mu) \leq Cw(\mu)$ (inégalité vectorielle),

- $v \sim w$ si $v = O(w)$ et $w = O(v)$.

Avec ces notations, $v = O(1)$ signifie que $\{v(\mu)\}_{\mu \downarrow 0}$ est bornée et $v \sim 1$ signifie que $\{v(\mu)\}_{\mu \downarrow 0}$ et $\{v(\mu)^{-1}\}_{\mu \rightarrow 0^+}$ sont bornées.

On trouvera à l'exercice 18.6 un résultat précisant le point (ii).

Proposition 18.4 (description du chemin central) *Supposons que (18.2) ait lieu. Alors*

- (i) *si A est surjective, $x > 0$ et $s > 0$, alors la matrice*

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix}$$

est inversible et l'application $\mu \mapsto z(\mu)$ définissant le chemin central est de classe C^∞ ;

- (ii) *si μ croît, $c^T x(\mu)$ et $x(\mu)^T s(\mu)$ croissent et $b^T y(\mu)$ décroît ;*
 (iii) *pour $\bar{\mu} > 0$ fixé, les ensembles $\{x(\mu) : 0 < \mu \leq \bar{\mu}\}$ et $\{s(\mu) : 0 < \mu \leq \bar{\mu}\}$ sont bornés ; si A est surjective, l'ensemble $\{y(\mu) : 0 < \mu \leq \bar{\mu}\}$ est aussi borné ;*
 (iv) *si \mathcal{F}_P est borné, $x(\mu)$ converge vers le centre analytique de \mathcal{F}_P , lorsque $\mu \rightarrow +\infty$; de même si \mathcal{F}_D est borné (donc A surjective), $(y(\mu), s(\mu))$ converge vers le centre analytique de \mathcal{F}_D , lorsque $\mu \rightarrow +\infty$;*
 (v) *si $\mu \downarrow 0$, on a $x_{\mathfrak{B}}(\mu) \sim 1$, $x_{\mathfrak{N}}(\mu) \sim \mu$, $s_{\mathfrak{B}}(\mu) \sim \mu$, $s_{\mathfrak{N}}(\mu) \sim 1$ et $(x(\mu), s(\mu)) \rightarrow (\hat{x}, \hat{s})$ le centre analytique des faces optimales.*

DÉMONSTRATION. (i) L'application $\mu \mapsto z(\mu)$ est fonction implicite de $F(z, \mu) = 0$, où

$$F(z, \mu) = \begin{pmatrix} A^T y + s - c \\ Ax - b \\ Xs - \mu e \end{pmatrix}.$$

Comme F est C^∞ , la fonction implicite sera C^∞ si la matrice $F'_z(z, \mu)$ (celle donnée au point (ii)) est inversible pour $z = z(\mu)$. C'est bien le cas, car si (dx, dy, ds) est dans son noyau, on trouve successivement

$$ds = -X^{-1}Sdx, \quad dx = S^{-1}XA^T dy \quad \text{et} \quad AS^{-1}XA^T dy = 0.$$

Comme A est surjective et $(x, s) > 0$, la matrice $AS^{-1}XA^T$ est inversible, ce qui implique que $(dx, dy, ds) = 0$.

(ii) Ce résultat se démontre comme dans la théorie de la pénalisation (proposition 13.2), en considérant les problèmes pénalisés (P_μ) et (D_μ) . Pour le saut de dualité, on utilise $x(\mu)^T s(\mu) = c^T x(\mu) - b^T y(\mu)$. (Voir aussi l'exercice 18.6 pour un résultat plus précis et une démonstration directe lorsque A est surjective.)

(iii) Comme \mathcal{S}_P est borné, l'ensemble $\{x \in \mathbb{R}^n : Ax = b, x \geq 0, c^T x \leq c^T x(\bar{\mu})\}$ est aussi borné (ils ont le même cône asymptotique). Or $x(\mu)$ est dans ce dernier ensemble lorsque $\mu \in]0, \bar{\mu}]$ (par le point (ii)), donc $\{x(\mu)\}_{0 < \mu \leq \bar{\mu}}$ est borné.

De même, comme $\mathcal{F}_p^s \neq \emptyset$, l'ensemble $\{s \in \mathbb{R}_+ : \text{il existe } y \in \mathbb{R}^m \text{ tel que } A^\top y + s = c \text{ et } b^\top y \geq b^\top y(\bar{\mu})\}$ est borné (les deux ensembles ont le même **cône asymptotique** $\{r \in \mathbb{R}_+ : \text{il existe } z \in \mathbb{R}^m \text{ tel que } A^\top z + r = 0, \text{ et } b^\top z \geq 0\}$). Comme $s(\mu)$ appartient à cet ensemble lorsque $\mu \in]0, \bar{\mu}]$ (par le point (ii)), $\{s(\mu)\}_{0 < \mu \leq \bar{\mu}}$ est bornée.

Si A est surjective, l'équation $A^\top y(\mu) + s(\mu) = c$ montre que $\{y(\mu) : 0 < \mu \leq \bar{\mu}\}$ est aussi borné.

(iv) Si \mathcal{F}_p est borné, $\{x(\mu)\} \subseteq \mathcal{F}_p$ est bornée. Il suffit donc de montrer que cette suite a un unique point d'adhérence quand $\mu \rightarrow \infty$, qui est le centre analytique \check{x} . Soit \bar{x} un point adhérent à $\{x(\mu)\}$: pour une sous-suite $x(\mu) \rightarrow \bar{x}$. Comme $x(\mu)$ est la solution de (P_μ) , on a quel que soit $x \in \mathcal{F}_p^s$:

$$c^\top x(\mu) + \mu \text{lb}(x(\mu)) \leq c^\top x + \mu \text{lb}(x).$$

En divisant par $\mu > 0$ et en passant à la limite quand $\mu \rightarrow \infty$, on trouve pour tout $x \in \mathcal{F}_p^s$:

$$\text{lb}(\bar{x}) \leq \text{lb}(x).$$

Ceci montre que $\bar{x} = \check{x}$.

On s'y prend de la même manière pour $(y(\mu), s(\mu))$. Par optimalité, pour tout couple $(y, s) \in \mathcal{F}_D^s$, on a

$$b^\top y(\mu) - \mu \text{lb}(s(\mu)) \geq b^\top y - \mu \text{lb}(s).$$

(v) Soit $(\bar{x}, \bar{y}, \bar{s})$ une solution strictement complémentaire (proposition 17.8). On réduit le saut de dualité $x^\top s$ en remplaçant x par \bar{x} :

$$\bar{x}^\top s(\mu) \leq x(\mu)^\top s(\mu) = n\mu.$$

Comme $\bar{x}_{\mathfrak{N}} = 0$, $\bar{x}_{\mathfrak{B}}^\top s_{\mathfrak{B}}(\mu) \leq n\mu$. Mais $\bar{x}_{\mathfrak{B}} > 0$, donc $s_{\mathfrak{B}}(\mu) = O(\mu)$. Alors, de $x_i(\mu)s_i(\mu) = \mu$, pour tout i , on déduit que $1 = O(x_{\mathfrak{B}}(\mu))$. Mais $\{x_{\mathfrak{B}}(\mu)\}$ est bornée, donc $x_{\mathfrak{B}}(\mu) \sim 1$. En utilisant à nouveau $x_i(\mu)s_i(\mu) = \mu$, pour tout i , on déduit que $s_{\mathfrak{B}}(\mu) \sim \mu$.

De la même manière,

$$x(\mu)^\top \bar{s} \leq x(\mu)^\top s(\mu) = n\mu.$$

Comme $\bar{s}_{\mathfrak{B}} = 0$ et $\bar{s}_{\mathfrak{N}} > 0$, on a $x_{\mathfrak{N}}(\mu) = O(\mu)$. En utilisant alors $x_i(\mu)s_i(\mu) = \mu$, on a $s_{\mathfrak{N}}(\mu) \sim 1$ ($\{s(\mu)\}$ bornée) et donc $x_{\mathfrak{N}}(\mu) \sim \mu$.

Il reste à montrer que $(x(\mu), y(\mu), s(\mu))$ converge vers le centre analytique de la face optimale lorsque $\mu \downarrow 0$. Comme $\{(x(\mu), s(\mu))\}$ est bornée, il existe des sous-suites telles que

$$x(\mu) \rightarrow \tilde{x} \quad \text{et} \quad s(\mu) \rightarrow \tilde{s}.$$

Passons à la limite dans (18.3). Comme $A^\top y(\mu) = c - s(\mu)$ converge et que $\mathcal{R}(A^\top)$ est fermé, il existe $\tilde{y} \in \mathbb{R}^m$ tel que l'on ait

$$\begin{cases} A^\top \tilde{y} + \tilde{s} = c \\ A\tilde{x} = b \\ \tilde{X}\tilde{s} = 0. \end{cases}$$

Donc $\tilde{z} = (\tilde{x}, \tilde{y}, \tilde{s})$ est solution primale-duale. Dès lors $\tilde{x}_{\mathfrak{N}} = 0$ et $\tilde{s}_{\mathfrak{B}} = 0$. D'autre part, $(x_{\mathfrak{B}}(\mu), s_{\mathfrak{N}}(\mu)) \sim 1$ implique que $(\tilde{x}_{\mathfrak{B}}, \tilde{s}_{\mathfrak{N}}) > 0$. Il reste à montrer que $\tilde{X}_{\mathfrak{B}}^{-1}e \in \mathcal{R}(A_{\mathfrak{B}}^{\top})$ et $A_{\mathfrak{N}}\tilde{S}_{\mathfrak{N}}^{-1}e \in \mathcal{R}(A_{\mathfrak{B}})$ pour conclure.

Comme $(x_{\mathfrak{N}}(\mu), s_{\mathfrak{B}}(\mu)) \sim \mu$, il existe une sous-suite telle que $(x_{\mathfrak{N}}(\mu), s_{\mathfrak{B}}(\mu))/\mu \rightarrow (x'_{\mathfrak{N}}, s'_{\mathfrak{B}})$. À la limite dans $X(\mu)s(\mu)/\mu = e$, on trouve

$$\tilde{X}_{\mathfrak{B}}^{-1}e = s'_{\mathfrak{B}} \quad \text{et} \quad \tilde{S}_{\mathfrak{N}}^{-1}e = x'_{\mathfrak{N}}.$$

D'autre part $Ax(\mu) = b = A_{\mathfrak{B}}\bar{x}_{\mathfrak{B}}$ (car $\bar{x}_{\mathfrak{N}} = 0$), donc

$$-A_{\mathfrak{B}}\frac{x_{\mathfrak{B}}(\mu) - \bar{x}_{\mathfrak{B}}}{\mu} = A_{\mathfrak{N}}\frac{x_{\mathfrak{N}}(\mu)}{\mu} \rightarrow A_{\mathfrak{N}}x'_{\mathfrak{N}}.$$

Comme $\mathcal{R}(A_{\mathfrak{B}})$ est fermé, $A_{\mathfrak{N}}x'_{\mathfrak{N}} \in \mathcal{R}(A_{\mathfrak{B}})$. De même $A_{\mathfrak{B}}^{\top}y(\mu) + s_{\mathfrak{B}}(\mu) = c_{\mathfrak{B}} = A_{\mathfrak{B}}^{\top}\bar{y}$ (car $\bar{s}_{\mathfrak{B}} = 0$), donc

$$A_{\mathfrak{B}}^{\top}\frac{\bar{y} - y(\mu)}{\mu} = \frac{s_{\mathfrak{B}}(\mu)}{\mu} \rightarrow s'_{\mathfrak{B}}.$$

Comme $\mathcal{R}(A_{\mathfrak{B}}^{\top})$ est fermé, $s'_{\mathfrak{B}} \in \mathcal{R}(A_{\mathfrak{B}}^{\top})$. □

18.2 Éléments constitutifs des algorithmes

Cheminement

Résoudre le problème d'optimisation (P) revient à résoudre ses conditions d'optimalité (18.1), lesquelles sont nécessaires et suffisantes. En apparence simple, ce système d'équations et d'inéquations présente plusieurs difficultés, toutes liées aux conditions de complémentarité $0 \leq s \perp x \geq 0$. D'une part, l'équation $s^{\top}x = 0$ qui exprime la perpendicularité de s et x est non linéaire. D'autre part, elle présente une « combinatoire » importante. Elle s'écrit en effet, du fait de la positivité de s et x : $x_i s_i = 0$, pour tout $i \in [1 : n]$; il faut donc décider pour tout indice i si $x_i = 0$ ou $s_i = 0$, et il y a 2^n possibilités.

Si l'on a un premier itéré primal-dual $z := (x, y, s)$ avec $x > 0$ et $s > 0$, on pourrait songer à résoudre le système d'optimalité (18.1) directement par des itérations de Newton amorties : à chaque itération, on détermine un pas $\alpha > 0$ le long de la direction de Newton $d := (dx, dy, ds)$ de telle sorte que l'itéré suivant $z_+ := (x_+, y_+, s_+) = z + \alpha d$ vérifie encore $x_+ > 0$ et $s_+ > 0$. On sait en effet que la stricte positivité des variables x et s assure que l'équation de Newton

$$\begin{pmatrix} 0 & A^{\top} & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} dx \\ dy \\ ds \end{pmatrix} = - \begin{pmatrix} A^{\top}y + s - c \\ Ax - b \\ Xs \end{pmatrix},$$

associée au système d'optimalité est bien définie (la matrice est inversible par la proposition 18.4). L'expérience a montré que cette stratégie, qui est suivie par l'*algorithme affine* (« *affine scaling algorithm* »), ne conduit pas à des algorithmes polynomiaux. La

raison provient probablement du fait que, lorsque z est proche du bord de l'ensemble admissible primal-dual, le pas $\alpha > 0$ le long de d , assurant l'admissibilité des itérés, peut devenir très petit, empêchant tout progrès significatif vers la solution. Une des techniques mises au point pour obtenir la polynomialité consiste à forcer les itérés de rester proche du chemin central, d'une part, et à être moins gourmand, d'autre part, en ne cherchant pas à résoudre le système non linéaire (18.1) directement.

Il est difficile de renoncer à la direction de Newton, dont on connaît les qualités, si bien que le fait de faire des déplacements le long de telles directions est conservé dans les algorithmes de points intérieurs considérés dans ce chapitre. De manière à prévenir le phénomène des petits pas décrit ci-dessus, les algorithmes vont maintenir les itérés suffisamment proches du chemin central étudié à la section 18.1. Nous montrerons en effet dans chaque cas que, dans ces conditions, le pas α pris le long de la direction de Newton est borné inférieurement par une constante strictement positive en $O(n^{-\omega})$, où $\omega > 0$ dépend de l'approche algorithmique, en particulier du type de voisinage du chemin central où sont maintenus les itérés. Même si l'on peut regretter que la borne inférieure sur le pas dépende de n (il ne semble pas possible d'éviter cela), on est au moins assuré que les pas ne deviendront pas arbitrairement petits. Cette dépendance en n aura une incidence directe sur la *complexité itérative* des algorithmes, c'est-à-dire sur le nombre d'itérations qu'ils requièrent pour s'approcher d'une solution à $\varepsilon > 0$ près.

Pour des raisons évidentes, on dit que les algorithmes qui viennent d'être brièvement décrits sont des *méthodes de suivi de chemin*. La solution que l'on recherche par ces algorithmes, le centre analytique de la face optimale situé au bout de chemin central, est un point singulier parce que la jacobienne du système de Newton n'y est en général pas inversible. On ne peut donc pas appliquer directement les techniques de suivi de chemin que l'on utilise dans les méthodes d'homotopie par exemple. Celles mises en œuvre pour suivre le chemin central dans les méthodes de points intérieurs sont originales. Elles dépendent des algorithmes, mais elles relèvent presque toujours d'un principe que l'on peut voir comme la *poursuite d'un objectif fuyant*, ce que l'on a schématisé à la figure 18.1 : en l'itéré $z_k \in \mathcal{F}^s$ de l'itération k , on se fixe pour objectif

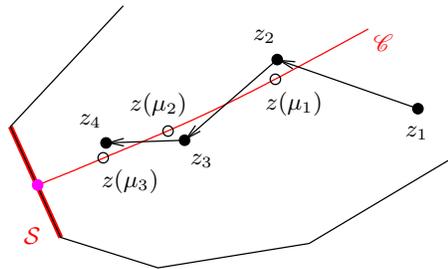


Fig. 18.1. Poursuite d'un objectif fuyant

un point $z(\mu_k)$ sur le chemin central \mathcal{C} , mais après avoir fait un pas de Newton dans sa direction (parfois plusieurs pas) conduisant à z_{k+1} , on change d'objectif en visant un autre point $z(\mu_{k+1})$ sur le chemin central, plus près de la solution ($\mu_{k+1} < \mu_k$).

On se rapproche ainsi petit à petit de la solution cherchée, laquelle est en quelque sorte inaccessible directement. Ce principe s’est avéré fécond.

Certains algorithmes imposent aux itérés d’être strictement admissibles (section 18.3). Leur complexité itérative est en $O(n^\omega \log \varepsilon^{-1})$, avec $\omega = \frac{1}{2}$ ou 1, ce qui veut dire que le nombre d’itérations pour atteindre une solution à $\varepsilon > 0$ près est majoré par une constante (indépendante de n et de ε) fois $n^\omega \log \varepsilon^{-1}$ (une définition précise de cette complexité itérative sera donnée plus loin). La complexité itérative en $O(n^{1/2} \log \varepsilon^{-1})$ est la meilleure que l’on ait obtenue ; mais les algorithmes qui la réalisent demandent que l’on dispose d’un premier itéré strictement admissible. D’autres algorithmes autorisent les itérés à ne pas satisfaire les équations linéaires de (18.1), ce qui peut être utile s’il n’y a pas de point primal-dual strictement admissible (c’est-à-dire si $\mathcal{F}^s = \emptyset$) ou si l’on ne dispose pas initialement d’un tel point. Leur complexité itérative est moins bonne ; elle est en $O(n^2 \log \varepsilon^{-1})$ pour l’algorithme étudié à la section 18.4.

Voici à présent quelques concepts qui jouent un rôle-clé dans l’étude des algorithmes de points intérieurs.

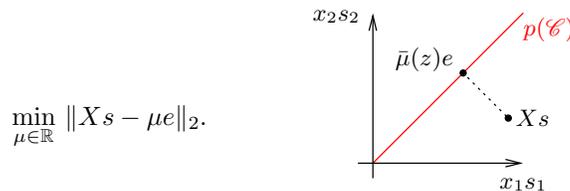
Mesures du saut de dualité et du centrage

Le contrôle des itérés dans les algorithmes de points intérieurs se fait par plusieurs « mesures » : mesure du saut de dualité, mesure du centrage et mesure de l’admissibilité.

Si l’on veut se donner une cible sur le chemin central primal-dual \mathcal{C} , il est nécessaire de savoir près de quel point central l’itéré courant z se trouve. Il n’y aurait en effet pas de sens à se donner une cible qui soit plus éloignée de la solution que ne l’est l’itéré courant. Trouver le point central le plus proche de z n’est cependant pas un problème simple, ni d’ailleurs bien posé car, le chemin central n’étant pas un convexe fermé, la projection de z sur \mathcal{C} n’est en général pas bien définie (proposition 2.25). Par contre, l’image de \mathcal{C} par l’application surjective (et bijective si A est surjective, voir l’exercice 18.2)

$$p : z = (x, y, s) \in \mathcal{F}^s \mapsto (x_1 s_1, \dots, x_n s_n) \in \mathbb{R}_{++}^n \tag{18.15}$$

est la **demi-droite** $\{\mu e : \mu > 0\}$, si bien que la projection dans l’espace d’arrivée de cette application se fait trivialement en résolvant le problème



Sa solution est la moyenne arithmétique des produits $x_i s_i$:

$$\bar{\mu} \equiv \bar{\mu}(z) := \frac{x^\top s}{n}.$$

On l’appelle la *mesure du saut de dualité* (ou plus simplement le *saut de dualité* ; il s’agit alors d’un abus de langage car, d’après le théorème 17.11, il n’y a pas de

saut de dualité en la solution d'un problème d'optimisation linéaire). Les algorithmes chercheront à faire tendre $\bar{\mu}(z)$ vers 0.

Pour $z = (x, y, s) \in \mathcal{F}^s$:

$$z \in \mathcal{C} \iff Xs = \bar{\mu}(z)e.$$

Un point $z \in \mathcal{F}^s$ est donc proche du chemin central si $\|(Xs)/\bar{\mu}(z) - e\|$ est petit devant 1 ou encore si $\|Xs - \bar{\mu}(z)e\|$ est petit devant $\bar{\mu}(z)$. On appelle *mesure du centrage* pour la **norme** ℓ_p , $p \in [1, \infty]$, la quantité

$$\gamma_p(z) := \|Xs - \bar{\mu}(z)e\|_p.$$

Divers voisinages du chemin central sont associés à ce concept de centrage.

On définit une première famille de voisinages, paramétrés par $p \in [1, \infty]$ et $\theta \in [0, 1[$, par

$$V_p(\theta) := \{z \in \mathcal{F}^s : \|Xs - \bar{\mu}(z)e\|_p \leq \theta \bar{\mu}(z)\}. \tag{18.16}$$

Dans cette famille de voisinages, on utilisera essentiellement $V_2(\theta)$, qui donne les meilleurs résultats de complexité. Cependant, il est parfois trop petit, ne laissant pas assez de liberté aux itérés. On montre en effet que pour $n \geq 2$, $\bigcup_{0 \leq \theta < 1} V_2(\theta) \neq \mathcal{F}^s$. Cette affirmation est examinée à l'exercice 18.7.

On fera aussi usage de

$$V_\infty^-(\theta) := \{z \in \mathcal{F}^s : Xs \geq (1-\theta)\bar{\mu}(z)e\}, \tag{18.17}$$

où $\theta \in [0, 1[$. Ce voisinage contient $V_\infty(\theta)$ et est assez grand puisque l'on peut montrer que $\bigcup_{0 \leq \theta < 1} V_\infty^-(\theta) = \mathcal{F}^s$. On se référera à nouveau à l'exercice 18.7 pour un examen de ces affirmations.

Sur la complexité itérative des algorithmes

Contrairement à l'algorithme du simplexe, les méthodes de points intérieurs ne sont pas des algorithmes à *terminaison finie* : ils ne trouvent pas la solution en un nombre fini d'étapes. En effet, à chaque itération, on a $x > 0$ et $s > 0$, ce qui n'est jamais le cas en une solution de (P) (voir la troisième condition dans (18.1)). On ne peut donc estimer que le nombre d'opérations pour trouver une solution à $\varepsilon > 0$ près. La proximité de la solution se mesurera ici par la petitesse du saut de dualité $\bar{\mu}(z)$. En réalité, il existe des procédures, dites de *purification* (section ??), permettant de déterminer une solution exacte strictement complémentaire par quelques opérations d'algèbre linéaire à partir d'un itéré généré par un algorithme de points intérieurs, suffisamment proche de la face optimale.

On ne s'intéressera ici qu'à la *complexité itérative* des algorithmes. On veut dire par là que l'on cherche à estimer le nombre d'*itérations* nécessaires pour obtenir une solution à $\varepsilon > 0$ près, dans le pire des cas. Les résultats que nous donnerons sur cette question feront usage du lemme suivant. On note $\{z_k\}$ la suite des itérés générés par l'algorithme considéré et $\bar{\mu}_k = \bar{\mu}(z_k)$.

Lemme 18.5 (de complexité) *Supposons qu'il existe des constantes $\delta > 0$ et $\omega > 0$, telles que pour tout $k \geq 0$, les itérés $\{z_k\}$ vérifient*

$$\bar{\mu}_{k+1} \leq \left(1 - \frac{\delta}{n^\omega}\right) \bar{\mu}_k.$$

Alors, pour tout $\varepsilon \in]0, 1]$ et tout $k \geq (n^\omega \log \varepsilon^{-1})/\delta$, on a

$$\frac{\bar{\mu}_k}{\bar{\mu}_0} \leq \varepsilon.$$

DÉMONSTRATION. En prenant le logarithme de l'inégalité donnée dans l'énoncé et en utilisant le fait que $\log(1+t) \leq t$, on a

$$\log \bar{\mu}_{k+1} \leq \log \left(1 - \frac{\delta}{n^\omega}\right) + \log \bar{\mu}_k \leq -\frac{\delta}{n^\omega} + \log \bar{\mu}_k.$$

Par récurrence

$$\log \frac{\bar{\mu}_k}{\bar{\mu}_0} \leq -\frac{k\delta}{n^\omega}.$$

Dès lors $\bar{\mu}_k/\bar{\mu}_0 \leq \varepsilon$ si $-k\delta/n^\omega \leq \log \varepsilon$, ce qui s'écrit aussi $k \geq (n^\omega \log \varepsilon^{-1})/\delta$. \square

En pratique, on cherche à avoir un saut de dualité inférieur à un seuil donné $\varepsilon > 0$:

$$\bar{\mu}_k \leq \varepsilon,$$

plutôt que l'inégalité relative $\bar{\mu}_k \leq \varepsilon \bar{\mu}_0$ fournie par le lemme. Le résultat reste le même, mais avec K qui dépend de $\bar{\mu}_0 > 0$. En effet, on peut écrire $\bar{\mu}_0 = \varepsilon^{1-\kappa}$ pour un certain $\kappa > 0$ (éventuellement grand) dépendant donc de l'itéré initial. On utilise alors le lemme avec $\varepsilon/\bar{\mu}_0 = \varepsilon^\kappa$ au lieu de ε : *pour tout $\varepsilon > 0$ et tout $k \geq K := O(n^\omega \log \varepsilon^{-1})$, on a $\bar{\mu}_k \leq \varepsilon$.*

18.3 Algorithmes avec itérés admissibles

18.3.1 Préliminaires

Les algorithmes étudiés dans cette section génèrent des itérés strictement admissibles, c'est-à-dire dans \mathcal{F}^s . Dès lors, les résidus sont toujours nuls : $r_c = 0$ et $r_b = 0$. Chaque nouvel itéré z_+ est obtenu à partir du précédent z en se déplaçant le long de la direction de Newton d , solution de (18.34), qui s'écrit ici

$$\begin{pmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} dx \\ dy \\ ds \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mu e - Xs \end{pmatrix}. \quad (18.18)$$

Du fait de la linéarité des résidus et de l'utilisation de la direction de Newton pour mettre à jour les itérés, on voit qu'il suffit que les premiers résidus r_c et r_b soient nuls pour que les suivants le soient aussi. Par exemple, si $r_b(z) = 0$, on a en $z_+ = z + \alpha d$: $r_b(z_+) = r_b(z) + \alpha Adx = 0$. De même pour r_c .

Pour que les algorithmes soient bien posés et convergent, il faudra que les itérés conservent la stricte positivité de x et s et fassent décroître le saut de dualité $\bar{\mu}(z)$ vers 0. On comprend que l'analyse des algorithmes passe par l'examen de l'évolution du saut de dualité $\bar{\mu}(z)$ et de la mesure du centrage $\gamma_2(z) := \|Xs - \bar{\mu}(z)e\|_2$ le long de la direction de Newton. C'est ce qu'étudie le lemme suivant. On suppose que le paramètre μ est de la forme $\sigma\bar{\mu}$, avec $\sigma \geq 0$.

Lemme 18.6 (évolution du saut de dualité et du centrage) Soit $d = (dx, dy, ds)$ la direction de Newton en $z = (x, y, s) \in \mathcal{F}^s$, avec $\mu := \sigma\bar{\mu}$, $\sigma \geq 0$. On note $z(\alpha) := z + \alpha d$. Alors les relations suivantes sont vérifiées :

- (i) $\bar{\mu}(z(\alpha)) = [1 - \alpha(1 - \sigma)]\bar{\mu}$,
- (ii) $\gamma_2(z(\alpha)) \leq (1 - \alpha)\gamma_2(z) + \alpha^2 \|dXds\|_2$.

DÉMONSTRATION. On a

$$x_i(\alpha)s_i(\alpha) = (x_i + \alpha dx_i)(s_i + \alpha ds_i) = x_i s_i + \alpha(s_i dx_i + x_i ds_i) + \alpha^2 dx_i ds_i. \quad (18.19)$$

En sommant et en utilisant le fait que $dx^\top ds = 0$ (car $Adx = 0$ et $A^\top dy + ds = 0$) :

$$\bar{\mu}(z(\alpha)) = \bar{\mu} + \frac{\alpha}{n}(s^\top dx + x^\top ds).$$

On utilise ensuite le fait que $s^\top dx + x^\top ds = e^\top (Sdx + Xds) = e^\top (\sigma\bar{\mu}e - Xs) = n(\sigma - 1)\bar{\mu}$ pour trouver (i).

En reprenant (18.19) et en utilisant $s_i dx_i + x_i ds_i = \sigma\bar{\mu} - x_i s_i$ ainsi que la valeur trouvée pour $\bar{\mu}(z(\alpha))$, on a

$$\begin{aligned} x_i(\alpha)s_i(\alpha) - \bar{\mu}(z(\alpha)) &= (1 - \alpha)x_i s_i + (\alpha - 1)\bar{\mu} + \alpha^2 dx_i ds_i \\ &= (1 - \alpha)(x_i s_i - \bar{\mu}) + \alpha^2 dx_i ds_i. \end{aligned}$$

On en déduit (ii). □

Le point (i) du lemme précédent montre qu'en prenant $\sigma = 0$ et en faisant un pas unité ($\alpha = 1$), on annule le saut de dualité. Il est peu probable cependant que cela conduise en une solution car il y a de fortes chances pour qu'en chemin on ait perdu la positivité des variables x et s (même si l'on a conservé la somme des $x_i s_i$ positive).

Le point (ii) montre que le contrôle du centrage $\gamma_2(z(\alpha))$ passe par celui de $\|dXds\|_2$. L'estimation de $dXds$ sera toujours un passage-clé des démonstrations de convergence des algorithmes. On utilisera à plusieurs reprises la technique suivante. On observe que la dernière équation du système de Newton (18.18) s'écrit $Sdx + Xds = \mu - Xs$. En la multipliant par $(XS)^{-\frac{1}{2}}$, elle devient

$$D^{-1}dx + Dds = (XS)^{-\frac{1}{2}}(\mu e - Xs), \quad (18.20)$$

où

$$D = X^{\frac{1}{2}} S^{-\frac{1}{2}}.$$

Le lemme de Mizuno ci-dessous donne alors un moyen de relier la quantité $dXdS = (D^{-1}dX)(Dds)$ qui nous intéresse à la somme des deux vecteurs $D^{-1}dx + Dds$, dont on a une autre expression en (18.20).

Lemme 18.7 (Mizuno [424]) Si $u, v \in \mathbb{R}^n$ vérifient $u^T v \geq 0$ et si $U := \text{Diag}(u_1, \dots, u_n)$, alors

$$\|Uv\|_2 \leq \frac{1}{\sqrt{8}} \|u + v\|_2^2.$$

DÉMONSTRATION. On a

$$\|Uv\|_2^2 = \sum_{i=1}^n (u_i v_i)^2 \leq \left(\sum_{u_i v_i < 0} u_i v_i \right)^2 + \left(\sum_{u_i v_i > 0} u_i v_i \right)^2,$$

parce que les doubles produits à droite sont tous positifs. La relation $u^T v \geq 0$ s'écrit aussi $\sum_{u_i v_i < 0} |u_i v_i| \leq \sum_{u_i v_i > 0} u_i v_i$ et comme $\alpha\beta \leq (\alpha + \beta)^2/4$, on obtient

$$\|Uv\|_2^2 \leq 2 \left(\sum_{u_i v_i > 0} u_i v_i \right)^2 \leq \frac{1}{8} \left(\sum_{u_i v_i > 0} (u_i + v_i)^2 \right)^2 = \frac{1}{8} \|u + v\|_2^4. \quad \square$$

18.3.2 Algorithme des petits déplacements

L'algorithme des petits pas est conceptuellement l'un des plus simples. Sa très faible complexité itérative est obtenue de la manière suivante (voir la figure 18.2). L'algorithme fait des déplacements de Newton complets, avec pas $\alpha = 1$, si bien

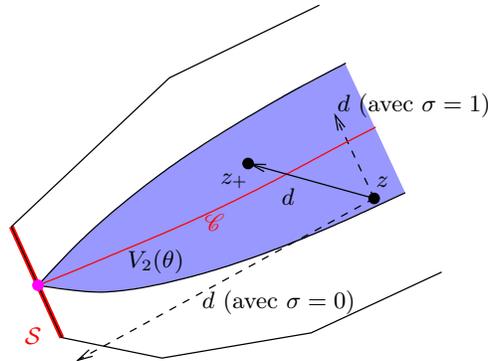
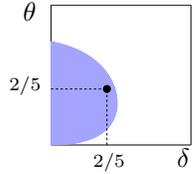


Fig. 18.2. Algorithme des petits déplacements

que $\bar{\mu}(z_+) = \sigma\bar{\mu}(z)$ (lemme 18.6). On ne peut pas alors prendre σ trop petit, sous peine de sortir de \mathcal{F}^s , comme lorsque $\sigma = 0$. En prenant $\sigma = 1 - \delta/n^{1/2}$ comme facteur de réduction de $\bar{\mu}$ ($\delta > 0$ étant une constante), on s'assure d'une part de rester dans un voisinage $V_2(\theta)$ (avec $\theta \in]0, 1[$ fixé) et d'autre part d'une complexité en $O(n^{1/2} \log \varepsilon^{-1})$ (voir le lemme de complexité 18.5). Pour que les itérés restent dans le voisinage $V_2(\theta)$, une relation doit relier θ et δ , à savoir :

$$\frac{\delta^2 + \theta^2}{(1 - \delta)(1 - \theta)\theta} \leq \sqrt{8}. \tag{18.21}$$


Elle est vérifiée pour (δ, θ) dans la zone représentée dans la figure ci-dessus, par exemple pour $\delta = \theta = 2/5$.

Algorithme 18.8 (PIL petits pas)

- On se donne θ et δ vérifiant (18.21) et :
- un voisinage du chemin central $V_2(\theta)$,
 - un facteur de réduction de $\bar{\mu}$: $\sigma = 1 - \delta/n^{1/2}$.
- L'itéré courant z est supposé dans $V_2(\theta)$.
 L'itéré suivant $z_+ \in V_2(\theta)$ s'obtient par les étapes suivantes.
1. Calcul de la direction de Newton d , solution de (18.18) avec $\mu = \sigma\bar{\mu}(z)$.
 2. Nouvel itéré : $z_+ = z + d$.

L'analyse de cet algorithme commence par l'estimation de $\|dXds\|_2$ (lemme 18.9), suivie d'une estimation de $\bar{\mu}(z_+)$ (lemme 18.10). On peut alors conclure grâce au lemme de complexité.

Lemme 18.9 Dans l'algorithme 18.8, on a

$$\|dXds\|_2 \leq \frac{\theta^2 + n(1-\sigma)^2}{\sqrt{8}(1-\theta)} \bar{\mu}.$$

DÉMONSTRATION. L'identité (18.20) et le lemme de Mizuno (qui s'applique car $(D^{-1}dx)^\top(Dds) = 0$) permettent d'écrire

$$\|dXds\|_2 = \|(D^{-1}dX)(Dds)\|_2 \leq \frac{1}{\sqrt{8}} \|(XS)^{-\frac{1}{2}}(\sigma\bar{\mu}e - Xs)\|_2^2.$$

D'une part, lorsque $z \in V_2(\theta)$, on a $x_i s_i \geq (1 - \theta)\bar{\mu}$. Dès lors

$$\|(XS)^{-\frac{1}{2}}\|_2^2 = \frac{1}{\min_i x_i s_i} \leq \frac{1}{(1-\theta)\bar{\mu}}.$$

D'autre part, on écrit $\sigma\bar{\mu}e - Xs = (\bar{\mu}e - Xs) - (1-\sigma)\bar{\mu}e$ et on observe que $e^\top(\bar{\mu}e - Xs) = 0$. Dès lors, pour des $z \in V_2(\theta)$, on a

$$\|\sigma\bar{\mu}e - Xs\|_2^2 = \|\bar{\mu}e - Xs\|_2^2 + n(1-\sigma)^2\bar{\mu}^2 \leq \theta^2\bar{\mu}^2 + n(1-\sigma)^2\bar{\mu}^2.$$

En utilisant ces deux dernières majorations dans la première, on obtient le résultat. \square

Lemme 18.10 Si $\theta \in]0, 1[$ et $\delta \in]0, 1[$ vérifient (18.21), alors $z_+ \in V_2(\theta)$ et

$$\bar{\mu}(z_+) = \left(1 - \frac{\delta}{n^{1/2}}\right) \bar{\mu}(z).$$

DÉMONSTRATION. D'après le point (i) du lemme 18.6 et $\alpha = 1$, on a

$$\bar{\mu}(z_+) = \sigma\bar{\mu}.$$

En utilisant l'expression choisie pour σ , on en déduit la formule de $\bar{\mu}(z_+)$ de l'énoncé.

Il reste à montrer que $z_+ \in V_2(\theta)$. D'après le point (ii) du même lemme, le lemme précédent, on a pour tout $\alpha \in [0, 1]$:

$$\begin{aligned} \gamma_2(z(\alpha)) &\leq (1-\alpha)\gamma_2(z) + \alpha^2 \frac{\theta^2 + n(1-\sigma)^2}{\sqrt{8}(1-\theta)} \bar{\mu} \\ &\leq \theta \left(1 - \alpha + \alpha \frac{\theta^2 + n(1-\sigma)^2}{\sqrt{8}(1-\theta)\theta}\right) \bar{\mu}, \end{aligned} \quad (18.22)$$

où on a utilisé le fait que $z \in V_2(\theta)$ et $\alpha \leq 1$. Dès lors, compte tenu du fait que $\bar{\mu}(z(\alpha)) = (1-\alpha + \alpha\sigma)\bar{\mu}$ (voir le point (i) du lemme 18.6), on a

$$\gamma_2(z(\alpha)) \leq \theta\bar{\mu}(z(\alpha)), \quad \forall \alpha \in [0, 1], \quad (18.23)$$

si la fraction dans (18.22) est inférieure à σ . C'est ce qui impose à σ d'être une fonction de n . En prenant $\sigma = 1 - \delta/n^{1/2}$ comme dans l'algorithme, on voit qu'il faut que l'on ait

$$\frac{\theta^2 + \delta^2}{\sqrt{8}(1-\theta)\theta} \leq 1 - \frac{\delta}{n^{1/2}}.$$

Cette inégalité est la plus restrictive lorsque $n = 1$: c'est l'inégalité (18.21) qui est vérifiée. L'estimation (18.23) que nous venons de démontrer implique que $\gamma_2(z_+) \leq \theta\bar{\mu}(z_+)$ et que $x_i(\alpha)s_i(\alpha) > 0$ pour tout $\alpha \in [0, 1]$. Donc les composantes de $x(\alpha)$ et $s(\alpha)$ ne peuvent s'annuler. Comme elles sont strictement positives en $\alpha = 0$, elles le restent en $\alpha = 1$. Dès lors $z_+ \in V_2(\theta)$. \square

L'estimation obtenue dans le lemme précédent et le lemme de complexité 18.5 permettent d'obtenir facilement le résultat de convergence et de complexité itérative suivant.

Théorème 18.11 (convergence et complexité de l’algorithme des petits déplacements) *L’algorithme 18.8 converge ($\bar{\mu}_k \rightarrow 0$ q -linéairement) et pour tout $\varepsilon > 0$, il existe un indice $K = O(n^{1/2} \log \varepsilon^{-1})$ tel que $\bar{\mu}_k \leq \varepsilon \bar{\mu}_0$, dès que $k \geq K$.*

Même si la réduction du saut de dualité $\bar{\mu}_k$ par le facteur $1 - \delta/n^{1/2}$ assure à l’algorithme des petits déplacements la meilleure complexité itérative obtenue à ce jour, $\bar{\mu}_k$ tend vers zéro assez lentement lorsque n est grand. Par exemple, si on prend $\delta = 2/5$ et si $n = 10^4$, il faudra 575 itérations pour réduire le saut de dualité d’un facteur 10. C’est beaucoup. Cet algorithme n’est donc guère utilisé en pratique. Il nous a toutefois permis de présenter de manière concise un des meilleurs résultats de complexité que peuvent apporter les méthodes de points intérieurs en optimisation linéaire.

18.3.3 Algorithme des grands déplacements

On obtient de plus grands déplacements, et donc une efficacité numérique meilleure en pratique, en utilisant $V_\infty^-(\theta)$ comme voisinage du chemin central devant contrôler les itérés. La complexité itérative théorique est cependant moins bonne; elle est en $O(n \log \varepsilon^{-1})$. Dans l’algorithme proposé ci-dessous, on prend un pas α le long de la direction de Newton le plus grand possible.

Algorithme 18.12 (PII grands pas)

On se donne :

- un voisinage du chemin central $V_\infty^-(\theta)$, avec $\theta \in]0, 1[$,
- des bornes $0 < \sigma_{\min} < \sigma_{\max} < 1$ pour le facteur de réduction de $\bar{\mu}$.

L’itéré courant z est supposé dans $V_\infty^-(\theta)$.

L’itéré suivant $z_+ \in V_\infty^-(\theta)$ s’obtient par les étapes suivantes.

1. Choix de $\sigma \in [\sigma_{\min}, \sigma_{\max}]$.
 2. Calcul de la direction de Newton d , solution de (18.18) avec $\mu = \sigma \bar{\mu}(z)$.
 3. Calcul du pas α le plus grand possible dans $]0, 1[$ tel que $z + \alpha d \in V_\infty^-(\theta)$.
 4. Nouvel itéré: $z_+ = z + \alpha d$.
-

Typiquement, on prend $\theta = 0.99$, ce qui permet à $V_\infty^-(\theta)$ d’occuper une grande partie de \mathcal{F}^s . L’étape 3 demande de calculer le plus grand pas α permettant à $z(\alpha) = z + \alpha d$ de rester dans $V_\infty^-(\theta)$. Il faut donc vérifier que $x_i(\alpha)s_i(\alpha) \geq (1-\theta)\bar{\mu}(z(\alpha))$ pour tout i . Comme $x_i(\alpha)$, $s_i(\alpha)$ et $\bar{\mu}(z(\alpha))$ dépendent linéairement de α (lemme 18.6), on voit que cela revient à trouver les racines de n fonctions quadratiques.

L’analyse de l’algorithme 18.12 commence par l’estimation de $\|dXds\|_2$.

Lemme 18.13 *Dans l’algorithme 18.12, on a*

$$\|dXds\|_2 \leq \frac{n(2-\theta)}{\sqrt{8}(1-\theta)} \bar{\mu}.$$

DÉMONSTRATION. Comme $(D^{-1}dx)^\top(Dds) = 0$, on peut appliquer le lemme de Mizuno sur l'identité (18.20). Cela donne

$$\|dXds\|_2 \leq \frac{1}{\sqrt{8}} \|(XS)^{-\frac{1}{2}}(\sigma\bar{\mu}e - Xs)\|_2^2.$$

Lorsque $z \in V_\infty^-(\theta)$, on a $x_i s_i \geq (1 - \theta)\bar{\mu}$. Dès lors, en développant le carré

$$\begin{aligned} \|dXds\|_2 &\leq \frac{1}{\sqrt{8}} \|\sigma\bar{\mu}(XS)^{-\frac{1}{2}}e - (XS)^{\frac{1}{2}}e\|_2^2 \\ &\leq \frac{1}{\sqrt{8}} \left(\sigma^2 \bar{\mu}^2 \frac{n}{(1-\theta)\bar{\mu}} - 2n\sigma\bar{\mu} + n\bar{\mu} \right) \\ &\leq \frac{1}{\sqrt{8}} \left(\frac{1}{(1-\theta)} + 1 \right) n\bar{\mu}, \end{aligned}$$

où on a majoré $\sigma^2 \leq 1$ et négligé $-2n\sigma\bar{\mu} \leq 0$. \square

Lemme 18.14 *Dans l'algorithme 18.12, il existe une constante δ indépendante de n telle que*

$$\bar{\mu}(z_+) \leq \left(1 - \frac{\delta}{n}\right) \bar{\mu}(z).$$

DÉMONSTRATION. Il s'agit d'obtenir une borne inférieure sur le pas maximal α permettant de rester dans $V_\infty^-(\theta)$, qui soit en $O(n^{-1})$. Le résultat se trouve alors en appliquant le point (i) du lemme 18.6.

D'après le lemme précédent, il existe une constante $C_1 > 0$ indépendante de n telle $dx_i ds_i \geq -C_1 n \bar{\mu}$. En utilisant (18.19) comme dans le lemme 18.6, l'équation de Newton et $x_i s_i \geq (1 - \theta)\bar{\mu}$ (car $z \in V_\infty^-(\theta)$), on trouve pour $t > 0$:

$$\begin{aligned} x_i(t)s_i(t) &= (1-t)x_i s_i + t\sigma\bar{\mu} + t^2 dx_i ds_i \\ &\geq (1-t)(1-\theta)\bar{\mu} + t\sigma\bar{\mu} - C_1 t^2 n \bar{\mu}. \end{aligned}$$

D'après le point (i) du lemme 18.6, $\bar{\mu}(z(t)) = (1-t+t\sigma)\bar{\mu}$. Alors l'inégalité ci-dessus montre que l'on aura $x_i(t)s_i(t) \geq (1-\theta)\bar{\mu}(z(t)) \equiv (1-\theta)(1-t+t\sigma)\bar{\mu}$ et $x_i(t)s_i(t) > 0$ si

$$t \leq \frac{\theta\sigma}{C_1 n} \quad \text{et} \quad t < \frac{1}{1-\sigma}.$$

Comme α est le plus grand $t > 0$ tel que $x_i(t)s_i(t) \geq (1-\theta)\bar{\mu}(z(t))$, $x_i(t) > 0$ et $s_i(t) > 0$, pour tout i , on a

$$\alpha \geq \min\left(\frac{\theta\sigma}{C_1 n}, \frac{1}{2(1-\sigma)}\right).$$

On note à présent qu'avec les bornes encadrant $\sigma \in [\sigma_{\min}, \sigma_{\max}]$, on a $\sigma(1-\sigma) \geq C_2$, une constante strictement positive indépendante de n . Dès lors, la formule de $\bar{\mu}(z(\alpha))$ du lemme 18.6, conduit à l'estimation cherchée :

$$\bar{\mu}(z(\alpha)) = [1 - \alpha(1-\sigma)]\bar{\mu} \leq \left(1 - \frac{\delta}{n}\right) \bar{\mu},$$

où $\delta = \min(\theta C_2/C_1, 1/2)$. □

On peut maintenant conclure aisément en appliquant le lemme de complexité 18.5.

Théorème 18.15 (convergence et complexité de l’algorithme des grands déplacements) *L’algorithme 18.12 converge ($\bar{\mu}_k \rightarrow 0$ q-linéairement) et pour tout $\varepsilon > 0$, il existe un indice $K = O(n \log \varepsilon^{-1})$ tel que $\bar{\mu}_k \leq \varepsilon \bar{\mu}_0$, dès que $k \geq K$.*

18.3.4 Un algorithme prédicteur-correcteur

Les algorithmes précédents souffrent du fait que le facteur de réduction σ du saut de dualité est imposé (il vaut $1 - \delta/n^{1/2}$ dans l’algorithme des petits déplacements) ou minoré (par σ_{\min} dans l’algorithme des grands déplacements) a priori. Il semble en effet préférable de laisser l’algorithme choisir lui-même une réduction maximale de $\bar{\mu}(z)$. C’est ce qui motive l’algorithme prédicteur-correcteur, que nous présentons dans cette section.

Cet algorithme doit son nom au fait qu’il alterne des *phases de prédiction* et des *phases de correction* (voir la figure 18.3). La phase de prédiction en $z \in \mathcal{F}^s$ a pour

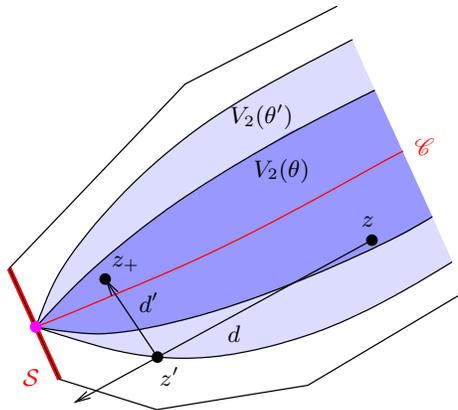


Fig. 18.3. Un algorithme prédicteur-correcteur

but de faire décroître le saut de dualité $\bar{\mu}$ le plus possible, ce qui s’obtient en suivant la direction de Newton d avec $\sigma = 0$ (voir la proposition 18.6 (i)) et en prenant le plus grand pas $\alpha > 0$ le long de cette direction, de manière toutefois à ce que $z' = z + \alpha d$ reste dans un voisinage $V_2(\theta')$ (avec un pas unitaire, z' serait vraisemblablement en dehors de l’ensemble admissible). Le point z' ne peut être le nouvel itéré car une direction de prédiction en ce point aurait de grande chance de sortir immédiatement de $V_2(\theta')$. L’étape de correction a donc pour but de revenir dans un voisinage du chemin central $V_2(\theta)$ plus petit (avec $0 < \theta < \theta' < 1$). Ce recentrage se fait par un

pas unité le long de la direction de Newton d' avec $\sigma = 1$, donc sans modification du saut de dualité. Il faut que θ et θ' vérifient les inégalités

$$0 < \theta < \theta' < 1 \quad \text{et} \quad \frac{(\theta')^2}{1 - \theta'} \leq \sqrt{8\theta}, \quad (18.24)$$

pour que le pas unité le long de d' conduise en un point z_+ dans $V_2(\theta)$. Les valeurs $\theta = \frac{1}{4}$ et $\theta' = \frac{1}{2}$ conviennent. On considère qu'une itération est formée d'une phase de prédiction suivie d'une phase de correction.

Algorithme 18.16 (PIL prédicteur-correcteur)

On considère deux voisinages $V_2(\theta)$ et $V_2(\theta')$, avec θ et θ' vérifiant (18.24).

On suppose que l'itéré courant $z := (x, y, s) \in V_2(\theta)$.

L'itéré suivant $z_+ \in V_2(\theta)$ s'obtient par les étapes suivantes.

1. *Phase de prédiction :*

- 1.1. Calcul du déplacement de Newton d , solution de (18.18) avec $\mu = 0$.
- 1.2. Calcul d'un pas α le plus grand possible dans $]0, 1]$ tel que $z + \alpha d \in V_2(\theta')$.
- 1.3. Itéré intermédiaire: $z' = z + \alpha d$.

2. *Phase de correction :*

- 2.1. Calcul du déplacement de Newton d' , solution de (18.18) en $z = z'$, avec $\mu = \bar{\mu}' := \bar{\mu}(z')$.
 - 2.2. Nouvel itéré: $z_+ = z' + d'$ (pas unité).
-

L'étape 1.2 demande de calculer le plus grand pas α permettant à $z(\alpha) = z + \alpha d$ de rester dans $V_2(\theta')$. Il faut donc vérifier que $\|X(\alpha)s(\alpha) - \bar{\mu}(z(\alpha))e\|_2^2 = (\theta')^2 \bar{\mu}(z(\alpha))^2$. Comme $x(\alpha)$, $s(\alpha)$ et $\bar{\mu}(z(\alpha))$ dépendent linéairement de α (lemme 18.6), on voit que cela revient à trouver la racine maximale d'un polynôme quartique (d'ordre 4).

Nous analysons dans les deux lemmes suivants les phases de prédiction et de correction, avant de conclure par un résultat de convergence et de complexité. Dans la phase de prédiction la décroissance du saut de dualité se fait par un facteur semblable à celui de l'algorithme des petits déplacements, mais cette fois l'algorithme a la possibilité de faire décroître $\bar{\mu}$ plus rapidement.

Lemme 18.17 (phase de prédiction) *Si $0 < \theta < \theta' < 1$, il existe une constante $\delta > 0$ indépendante de n telle que, lorsque $z \in V_2(\theta)$, la phase de prédiction conduit à un point $z' \in V_2(\theta')$ avec*

$$\bar{\mu}' \leq \left(1 - \frac{\delta}{n^{1/2}}\right) \bar{\mu}.$$

DÉMONSTRATION. On cherche à estimer le pas maximal α tel que $z + \alpha d$ reste dans $V_2(\theta')$. Ceci se fait en contrôlant l'évolution du centrage le long de d .

On applique le lemme de Mizuno en considérant (18.20) avec $\sigma = 0$, ce qui est licite car $(D^{-1}dx)^\top(Dds) = 0$. On a

$$\|dX ds\|_2 = \|(D^{-1}dX)(Dds)\|_2 \leq \frac{1}{\sqrt{8}} \|(XS)^{-\frac{1}{2}} Xs\|_2^2 = \frac{n}{\sqrt{8}} \bar{\mu}.$$

Alors le point (ii) du lemme 18.6 donne pour $z \in V_2(\theta)$ et $z(t) := z + td$:

$$\gamma_2(z(t)) \leq \left[(1-t)\theta + t^2 \frac{n}{\sqrt{8}} \right] \bar{\mu}.$$

D'après le point (i) du lemme 18.6, $\bar{\mu}(z(t)) = (1-t)\bar{\mu} > 0$, si $t < 1$. Dès lors $z(t)$ sera certainement dans $V_2(\theta')$ si $t < 1$ et si le membre de droite ci-dessus est inférieur à $\theta' \bar{\mu}(z(t)) = \theta'(1-t)\bar{\mu}$, ce qui s'écrit encore

$$\frac{n}{\sqrt{8}} t^2 \leq (\theta' - \theta)(1-t).$$

Cette inégalité est vérifiée pour tout $t \in [0, \delta/\sqrt{n}]$, où $\delta > 0$ est fixé indépendamment de n par :

$$\frac{\delta^2}{\sqrt{8}} \leq (\theta' - \theta)(1 - \delta).$$

Il suffit en effet d'observer que pour $t \in [0, \delta/\sqrt{n}]$, on a $nt^2 \leq \delta^2$ et $1 - \delta \leq 1 - \delta/\sqrt{n} \leq 1 - t$. Comme la phase de prédiction détermine le plus grand pas α tel que $z(\alpha) \in V_2(\theta')$, ce pas vérifie $\alpha \geq \delta/\sqrt{n}$. Dès lors

$$\bar{\mu}(z(\alpha)) = (1 - \alpha)\bar{\mu} \leq \left(1 - \frac{\delta}{n^{1/2}}\right) \bar{\mu}. \quad \square$$

Lemme 18.18 (phase de correction) *Si θ et θ' vérifient (18.24) et si $z' \in V_2(\theta')$, alors la phase de correction donne un point $z_+ \in V_2(\theta)$ tel que $\bar{\mu}_+ = \bar{\mu}'$.*

DÉMONSTRATION. Le point (i) du lemme 18.6 et $\sigma = 1$ montrent immédiatement que $\bar{\mu}_+ = \bar{\mu}'$. Il reste à montrer que $z_+ \in V_2(\theta)$.

D'après le lemme de Mizuno et (18.20) avec $\mu = \bar{\mu}$, la direction de Newton $d' = (dx', dy', ds')$ calculée en z' vérifie

$$\|dX' ds'\|_2 \leq \frac{1}{\sqrt{8}} \|(X'S')^{-\frac{1}{2}} (\bar{\mu}' e - X's')\|_2^2 = \frac{1}{\sqrt{8}} \sum_{i=1}^n \frac{(x'_i s'_i - \bar{\mu}')^2}{x'_i s'_i}.$$

Puisque $z' \in V_2(\theta')$, on a pour tout i : $\|X's' - \bar{\mu}' e\|_2 \leq \theta' \bar{\mu}'$, donc $x'_i s'_i \geq (1 - \theta') \bar{\mu}'$, si bien que

$$\|dX' ds'\|_2 \leq \frac{1}{\sqrt{8}(1 - \theta') \bar{\mu}'} \|X's' - \bar{\mu}' e\|_2^2 \leq \frac{(\theta')^2}{\sqrt{8}(1 - \theta')} \bar{\mu}'.$$

Le point (ii) du lemme 18.6 avec $\alpha = 1$ et (18.24) donnent

$$\|X_{+s_+} - \bar{\mu}_+ e\|_2 \leq \frac{(\theta')^2}{\sqrt{8}(1 - \theta')} \bar{\mu}_+ \leq \theta \bar{\mu}_+.$$

Pour conclure que $z_+ \in V_2(\theta)$, il reste à montrer que $x_+ > 0$, $s_+ > 0$, mais ceci se déduit du fait aisément démontrable que $z' + td' \in V_2(\theta')$ pour tout $t \in [0, 1]$. \square

Théorème 18.19 (convergence et complexité de l'algorithme prédicteur-correcteur) *L'algorithme prédicteur-correcteur 18.16 converge ($\bar{\mu}_k \rightarrow 0$ q -linéairement) et pour tout $\varepsilon > 0$, il existe un indice $K = O(n^{1/2} \log \varepsilon^{-1})$ tel que $\bar{\mu}_k \leq \varepsilon \bar{\mu}_0$, dès que $k \geq K$.*

DÉMONSTRATION. D'après les lemmes 18.17 et 18.18, on a

$$\bar{\mu}_{k+1} \leq \left(1 - \frac{\delta}{n^{1/2}}\right) \bar{\mu}_k.$$

On utilise alors le lemme de complexité 18.5. □

En pratique, le pas $\alpha > 0$ le long de la direction de prédiction est souvent proche de 1, si bien que l'algorithme converge souvent très rapidement. On peut d'ailleurs montrer que la convergence de $\bar{\mu}_k \rightarrow 0$ est superlinéaire. Malgré cela, l'algorithme peut être gêné par la petitesse du voisinage $V_2(\theta)$, surtout dans les premières itérations.

18.4 Un algorithme sans admissibilité forcée

On ne dispose pas toujours d'un itéré primal-dual strictement admissible. Parfois même il n'y en a pas, alors que le problème est parfaitement bien posé. Cette situation se rencontre, par exemple, si l'on transforme un problème d'optimisation linéaire de la forme $\inf\{c^T x : Ax \leq b\}$ sous la forme standard : $\inf\{c^T u - c^T v : Au - Av + z = b, (u, v, z) \geq 0\}$, où z est un vecteur de variables d'écart et $u - v$ joue le rôle de x . Dans ce cas, les contraintes du problème dual s'écrivent : $A^T y + s^1 = c$, $-A^T y + s^2 = -c$, $y + s^3 = 0$ et $(s^1, s^2, s^3) \geq 0$. En sommant les deux premières contraintes, on voit que s^1 et s^2 sont nuls : il n'y a pas de point strictement admissible pour le dual. Il est donc intéressant de développer des algorithmes dans lesquels les itérés ne vérifient pas les contraintes linéaires de (18.1).

Dans les algorithmes n'obligeant pas les itérés à rester dans \mathcal{F}^s , quoique maintenant $x > 0$ et $s > 0$ (c'est leur aspect « points intérieurs »), il y a deux objectifs à réaliser : annuler les résidus $(r_b, r_c) = (Ax - b, A^T y + s - c)$ et le saut de dualité $\bar{\mu} = x^T s / n$. Ces algorithmes opèrent comme précédemment, en faisant des déplacements le long de la direction de Newton d , solution de (18.34).

On s'est longtemps demandé si l'on pourrait un jour trouver des algorithmes polynomiaux dans lesquels les itérés ne sont pas admissibles. Une façon d'y arriver est de ne pas faire décroître le saut de dualité plus vite que l'admissibilité. La possibilité de réaliser ce principe est fondé sur les observations de l'évolution des résidus et du saut de dualité le long de la direction de Newton, disons en $z + \alpha d$, $\alpha \geq 0$, alors que le point courant z n'annule pas les résidus. On note

$$r(z) := (r_b(z), r_c(z)) = (Ax - b, A^T y + s - c).$$

On vérifie facilement que les résidus varient de façon affine :

$$r(z + \alpha d) = (1 - \alpha)r(z).$$

De même pour les normes, pour autant que $\alpha \leq 1$:

$$\|r(z + \alpha d)\|_2 = (1 - \alpha)\|r(z)\|_2. \quad (18.25)$$

Le saut de dualité évolue quant à lui de façon non linéaire. En utilisant la troisième équation de (18.34), on a $s^\top dx + x^\top ds = n\mu - x^\top s = n(\mu - \bar{\mu})$ et donc

$$\bar{\mu}(z + \alpha d) = \frac{1}{n}(x + \alpha dx)^\top (s + \alpha ds) = \bar{\mu} + \alpha(\mu - \bar{\mu}) + \frac{\alpha^2}{n} dx^\top ds.$$

Sans admissibilité, on n'a plus $dx^\top ds = 0$, si bien que la décroissance de $\bar{\mu}(z(\alpha))$ pour de petit pas $\alpha > 0$ n'est assurée que si on prend $\mu < \bar{\mu}$ (on n'est pas maître du signe de $dx^\top ds$) et on se fera à l'idée de prendre comme précédemment

$$\mu = \sigma \bar{\mu}, \quad \text{avec un } \sigma \in]0, 1[.$$

On verra par la suite qu'il faut être plus contraignant sur la valeur de σ . On a alors

$$\bar{\mu}(z + \alpha d) = (1 - \alpha)\bar{\mu} + \alpha\sigma\bar{\mu} + \frac{\alpha^2}{n} dx^\top ds. \quad (18.26)$$

On en déduit que

$$\bar{\mu}(z + \alpha d) \geq (1 - \alpha)\bar{\mu}, \quad \text{pour } \alpha > 0 \text{ petit.} \quad (18.27)$$

En comparant (18.25) et (18.27), on voit que si $\|r\|_2 \leq \rho\bar{\mu}$ à l'itéré courant (pour une constante $\rho > 0$), on a pour $\alpha > 0$ petit :

$$\|r(z + \alpha d)\|_2 \leq (1 - \alpha)\rho\bar{\mu} \leq \rho\bar{\mu}(z + \alpha d).$$

L'inégalité $\|r\|_2 \leq \rho\bar{\mu}$ est donc conservée à l'itération suivante. Autrement dit, de petits pas le long de la direction de Newton maintiennent les itérés dans le voisinage du chemin central suivant :

$$V_\infty^- \equiv V_\infty^-(\theta, \rho) := \{z := (x, y, s) : (x, s) > 0, \|r(z)\|_2 \leq \rho\bar{\mu}(z), Xs \geq (1 - \theta)\bar{\mu}(z)e\}.$$

On prend $\theta \in]0, 1[$ et de manière à ce que le premier itéré soit dans ce voisinage, on prend

$$\rho = \rho' \frac{\|r^0\|_2}{\bar{\mu}_0}, \quad \text{avec } \rho' \geq 1. \quad (18.28)$$

Ce voisinage diffère de $V_\infty^-(\theta)$ par le fait que le résidu $r = (r_b, r_c)$ ne doit pas nécessairement être nul.

Nous donnons ci-dessous une itération d'un algorithme de suivi de chemin fondé sur les observations précédentes. Il ne demande donc pas d'avoir un premier itéré admissible et permet de faire de grands déplacements via l'utilisation du voisinage V_∞^- . L'algorithme est très simple. À chaque itération, on fait un pas $\alpha > 0$ le long de la direction de Newton d , de manière à rester dans V_∞^- , tout en faisant décroître le saut de dualité $\bar{\mu}$. Une recherche linéaire sur $\alpha \mapsto \bar{\mu}(z + \alpha d)$ assure la décroissance du saut de dualité et du même coup celle des résidus (puisque les itérés restent dans V_∞^-). Cette stratégie est possible car, comme nous l'avons montré ci-dessus, la direction de Newton rentre dans V_∞^- et est une direction de descente de $\bar{\mu}$.

Algorithme 18.20 (PIL sans admissibilité)

On se donne des constantes :

- $\omega \in]0, 1[$ est utilisé dans la recherche linéaire sur $\bar{\mu}(\cdot)$,
- $\sigma \in]0, 1-\omega[$ est le facteur de réduction de $\bar{\mu}$ dans le système de newton,
- $\theta \in]0, 1[$ et ρ vérifiant (18.28) définissent le voisinage $V_{\infty}^{-}(\theta, \rho)$.

On suppose que l'itéré courant $z \in V_{\infty}^{-}(\theta, \rho)$.

L'itéré suivant $z_+ \in V_{\infty}^{-}(\theta, \rho)$ s'obtient par les étapes suivantes.

1. *Direction de Newton.* Calcul de la direction de Newton $d = (dx, dy, ds)$, solution de (18.34), avec $\mu = \sigma\bar{\mu}(z)$.
2. *Recherche linéaire.* Prendre un pas α le plus grand possible dans $]0, 1[$ tel que $z + \alpha d \in V_{\infty}^{-}(\theta, \rho)$ et

$$\bar{\mu}(z + \alpha d) \leq (1 - \omega\alpha)\bar{\mu}(z). \quad (18.29)$$

3. *Nouvel itéré.* $z^+ := z + \alpha d$.

Les valeurs typiques des constantes sont

$$\omega = 10^{-2}, \quad \sigma = 0.25 \quad \text{et} \quad \theta = 0.99.$$

On note z^k les itérés, α_k le pas pris à l'étape 2, $r_b^k := Ax^k - b$, $r_c^k := A^T y^k + s^k - c$, $r^k := (r_b^k, r_c^k)$, $\bar{\mu}_k := \bar{\mu}(z^k)$ et

$$D_k := X_k^{1/2} S_k^{-1/2}.$$

On simplifiera $\bar{\mu}(z^k + \alpha d^k)$ en $\bar{\mu}_k(\alpha)$ sans que cela ne porte à confusion (la dimension de l'argument change).

La preuve de convergence vise essentiellement à montrer que le pas α_k pris à l'étape 2 de l'algorithme est borné inférieurement par une constante $\bar{\alpha} > 0$. Dans ce cas, grâce à (18.29), $\bar{\mu}_{k+1} \leq (1 - \omega\bar{\alpha})\bar{\mu}_k$, ce qui montre la convergence q-linéaire de $\bar{\mu}_k$. De même, en utilisant (18.25), on a $\|r^{k+1}\|_2 \leq (1 - \bar{\alpha})\|r^k\|_2$, si bien que la convergence q-linéaire des résidus sera aussi une conséquence immédiate de l'estimation $\alpha_k \geq \bar{\alpha}$.

La convergence de cet algorithme s'obtient sans hypothèse supplémentaire, mais on n'a pu démontrer sa polynomialité qu'en supposant z^0 de la forme

$$z^0 = (\zeta_0 e, 0, \zeta_0 e), \quad \text{avec un scalaire } \zeta_0 \geq \|(x^*, s^*)\|_{\infty}, \quad (18.30)$$

où z^* est une solution primale-duale arbitraire [629]. Évidemment, en pratique, on ne peut pas réaliser (18.30) car on ne connaît pas de solution primale-duale, mais on a observé que l'algorithme est plus rapide si l'itéré initial est bien centré ($X^0 s^0 = \zeta_0^2 e$ sous l'hypothèse (18.30)) avec un rapport $\|r^0\|_2 / \bar{\mu}_0$ petit (il est de l'ordre $1/\zeta_0$ sous l'hypothèse (18.30)). En pratique, on prend donc $z^0 = (\zeta_0 e, 0, \zeta_0 e)$, avec ζ_0 « grand ».

La formule (18.26) montre que nous avons besoin d'une estimation de $|(dx^k)^T(ds^k)|$. Celle-ci découlera des estimations de $D_k^{-1} dx^k$ et $D_k ds^k$ données dans le lemme très technique suivant.

Lemme 18.21 *Il existe une constante $C_1 > 0$ telle que pour tout $k \geq 0$:*

$$\|D_k^{-1} dx^k\|_2 \leq C_1 \bar{\mu}_k^{1/2} \quad \text{et} \quad \|D_k ds^k\|_2 \leq C_1 \bar{\mu}_k^{1/2}.$$

Si z^0 vérifie (18.30), alors $C_1 = \frac{9\rho'}{(1-\theta)^{1/2}} n$ est indépendante de ζ_0 .

DÉMONSTRATION. D'après (18.25) et par récurrence, on a

$$r^k = \beta_k r^0, \quad \text{où} \quad \beta_k := \prod_{i=0}^{k-1} (1 - \alpha_i). \quad (18.31)$$

Première étape : il existe une constante $C'_1 (= 4\rho' n / \zeta_0)$ si (18.30) a lieu) telle que

$$\beta_k \|(x^k, s^k)\|_1 \leq C'_1 \bar{\mu}_k. \quad (18.32)$$

Soit z^* une solution du problème (P). On introduit

$$\bar{z} := \beta_k z^0 + (1 - \beta_k) z^* - z^k.$$

On voit facilement que $A\bar{x} = 0$ et $A^T \bar{y} + \bar{s} = 0$, si bien que $\bar{x}^T \bar{s} = 0$, ce que l'on exploite :

$$\begin{aligned} 0 &= (\beta_k x^0 + (1 - \beta_k) x^* - x^k)^T (\beta_k s^0 + (1 - \beta_k) s^* - s^k) \\ &= \beta_k^2 (x^0)^T s^0 + (1 - \beta_k)^2 (x^*)^T s^* + (x^k)^T s^k \\ &\quad + \beta_k (1 - \beta_k) [(x^0)^T s^* + (x^*)^T s^0] - \beta_k [(x^0)^T s^k + (x^k)^T s^0] \\ &\quad - (1 - \beta_k) [(x^*)^T s^k + (x^k)^T s^*]. \end{aligned}$$

La majoration que l'on cherche vient de l'avant dernier terme que l'on fait passer dans le membre de gauche. D'autre part, le dernier terme est négatif $((x^*, s^*, x^k, s^k) \geq 0$ et $\beta_k \leq 1$) et on le néglige, $(x^0)^T s^0 = n\bar{\mu}_0$, $(x^*)^T s^* = 0$ par optimalité, $(x^k)^T s^k = n\bar{\mu}_k$, $(x^0)^T s^* \leq \|x^0\|_\infty \|s^*\|_1$ et $(x^*)^T s^0 \leq \|s^0\|_\infty \|x^*\|_1$. Cela donne

$$\beta_k [(x^0)^T s^k + (x^k)^T s^0] \leq n\beta_k^2 \bar{\mu}_0 + n\bar{\mu}_k + \beta_k (1 - \beta_k) \|(x^0, s^0)\|_\infty \|(x^*, s^*)\|_1.$$

Ensuite, le membre de gauche est minoré par $\beta_k \xi_0 \|(x^k, s^k)\|_1$, où

$$\xi_0 := \min_{1 \leq i \leq n} \min\{x_i^0, s_i^0\}.$$

On obtient aussi un facteur commun $\bar{\mu}_k$ à tous les termes à droite, car, d'après (18.31) et l'appartenance de $z^k \in V_\infty^-$, on a $\beta_k = \|r_k\|_2 / \|r^0\|_2 \leq \rho \bar{\mu}_k / \|r^0\|_2$. En utilisant également $0 \leq \beta_k \leq 1$, on obtient finalement l'inégalité

$$\beta_k \xi_0 \|(x^k, s^k)\|_1 \leq n \frac{\rho \bar{\mu}_0}{\|r^0\|_2} \bar{\mu}_k + n\bar{\mu}_k + \frac{\rho}{\|r^0\|_2} \bar{\mu}_k \|(x^0, s^0)\|_\infty \|(x^*, s^*)\|_1,$$

qui est bien de la forme (18.32).

Supposons à présent que z^0 vérifie (18.30). On note

$$\rho' := \frac{\rho \bar{\mu}_0}{\|r^0\|_2},$$

qui est ≥ 1 et peut être considéré comme une constante ne dépendant pas de l'itéré initial. D'autre part, en utilisant (18.30) : $\xi_0 = \zeta_0$, $\bar{\mu}_0 = \zeta_0^2$, $\|(x^0, s^0)\|_\infty = \zeta_0$ et $\|(x^*, s^*)\|_1 \leq 2n\|(x^*, s^*)\|_\infty \leq 2n\zeta_0$. Dès lors, l'inégalité précédente devient

$$\beta_k \zeta_0 \|(x^k, s^k)\|_1 \leq n\rho' \bar{\mu}_k + n\bar{\mu}_k + 2n\rho' \bar{\mu}_k \leq 4n\rho' \bar{\mu}_k.$$

Ce qui montre que l'on peut prendre $C_1 = 4\rho'n/\zeta_0$.

Deuxième étape : démonstration du lemme. Comme dans la première partie on construit deux vecteurs orthogonaux, ici formés à partir de dx^k et ds^k . Dans ce but, on introduit $\bar{d} = (\bar{d}x, \bar{d}y, \bar{d}s)$ défini par

$$\bar{d} := d^k + \beta_k(z^0 - z^*),$$

où z^* est une solution du problème (P). On vérifie facilement que $A\bar{d}x = 0$ et $A^\top \bar{d}y + \bar{d}s = 0$, si bien que $\bar{d}x^\top \bar{d}s = 0$, ce que l'on va exploiter.

Auparavant, observons que par la troisième équation du système de (18.34)

$$\begin{aligned} S_k(dx^k + \beta_k(x^0 - x^*)) + X_k(ds^k + \beta_k(s^0 - s^*)) \\ = -(X_k S_k e - \sigma \bar{\mu}_k e) + \beta_k S_k(x^0 - x^*) + \beta_k X_k(s^0 - s^*). \end{aligned}$$

Après multiplication par $(X_k S_k)^{-1/2}$, on a

$$\begin{aligned} D_k^{-1}(dx^k + \beta_k(x^0 - x^*)) + D_k(ds^k + \beta_k(s^0 - s^*)) \\ = -(X_k S_k)^{-1/2}(X_k S_k e - \sigma \bar{\mu}_k e) + \beta_k D_k^{-1}(x^0 - x^*) + \beta_k D_k(s^0 - s^*). \end{aligned}$$

On utilise maintenant le fait démontré ci-dessus, que les deux premiers termes sont orthogonaux. En prenant la norme au carré, on obtient

$$\begin{aligned} \|D_k^{-1}(dx^k + \beta_k(x^0 - x^*))\|_2^2 + \|D_k(ds^k + \beta_k(s^0 - s^*))\|_2^2 \\ \leq \left(\|(X_k S_k)^{-1/2}(X_k S_k e - \sigma \bar{\mu}_k e)\|_2 + \beta_k \|D_k^{-1}(x^0 - x^*)\|_2 + \beta_k \|D_k(s^0 - s^*)\|_2 \right)^2. \end{aligned}$$

On traite maintenant séparément les deux termes du membre de gauche en utilisant l'inégalité triangulaire :

$$\begin{aligned} \|D_k^{-1}dx^k\|_2 \text{ et } \|D_k ds^k\|_2 \\ \leq \|(X_k S_k)^{-1/2}(X_k S_k e - \sigma \bar{\mu}_k e)\|_2 + 2\beta_k \|D_k^{-1}(x^0 - x^*)\|_2 + 2\beta_k \|D_k(s^0 - s^*)\|_2. \end{aligned}$$

Il reste à montrer que chaque terme du membre de droite de cette inégalité est en $O(\bar{\mu}_k^{-1/2})$.

Pour le premier terme, on a en utilisant l'appartenance de z^k à $V_\infty^-(\theta, \rho)$:

$$\|(X_k S_k)^{-1/2}\|_2 = \max_{1 \leq i \leq n} \frac{1}{(x_i^k s_i^k)^{1/2}} \leq \frac{1}{(1-\theta)^{1/2} \bar{\mu}_k^{1/2}}$$

D'autre part, comme $\sigma < 1$ et $(x^k)^\top s^k = n\bar{\mu}_k$:

$$\|X_k S_k e - \sigma \bar{\mu}_k e\|_2^2 = \|X_k S_k e\|_2^2 - 2\sigma \bar{\mu}_k (x^k)^\top s^k + n\sigma^2 \bar{\mu}_k^2 \leq \|X_k S_k e\|_1^2 = n^2 \bar{\mu}_k^2.$$

Le premier terme est donc majoré par $(1-\theta)^{-1/2} n \bar{\mu}_k^{-1/2}$.

Pour le second terme, on observe d'abord que

$$\|D_k^{-1}\|_2 = \max_{1 \leq i \leq n} \frac{s_i^k}{(x_i^k s_i^k)^{1/2}} \leq \frac{\|s^k\|_\infty}{(1-\theta)^{1/2} \bar{\mu}_k^{1/2}} \leq \frac{\|s^k\|_1}{(1-\theta)^{1/2} \bar{\mu}_k^{1/2}}.$$

En utilisant alors le résultat (18.32) de la première partie de la démonstration, on a

$$\beta_k \|D_k^{-1}(x^0 - x^*)\|_2 \leq \frac{\beta_k \|s^k\|_1}{(1-\theta)^{1/2} \bar{\mu}_k^{1/2}} \|x^0 - x^*\|_2 \leq \frac{C'_1}{(1-\theta)^{1/2}} \bar{\mu}_k^{1/2} \|x^0 - x^*\|_2,$$

qui est bien un majorant en $O(\bar{\mu}_k^{1/2})$. On s'y prend de la même manière pour le troisième terme, avec la majoration $\|D_k\|_2 \leq \|x^k\|_1 \bar{\mu}_k^{1/2} / (1-\theta)^{1/2}$.

Enfin, si z^0 vérifie (18.30), on s'y prend différemment pour majorer les deux derniers termes. Dans ce cas, on a

$$0 \leq x^0 - x^* \leq \zeta_0 \quad \text{et} \quad 0 \leq s^0 - s^* \leq \zeta_0.$$

On en déduit, avec (18.32), que

$$\begin{aligned} & \beta_k \|D_k^{-1}(x^0 - x^*)\|_2 + \beta_k \|D_k(s^0 - s^*)\|_2 \\ & \leq \beta_k \zeta_0 (\|D_k^{-1}e\|_2 + \|D_k e\|_2) \\ & = \beta_k \zeta_0 \left(\|(X_k S_k)^{-1/2} s^k\|_2 + \|(X_k S_k)^{-1/2} x^k\|_2 \right) \\ & \leq \zeta_0 \|(X_k S_k)^{-1/2}\|_2 (\beta_k \|(x^k, s^k)\|_1) \\ & \leq \frac{\zeta_0}{(1-\theta)^{1/2} \bar{\mu}_k^{1/2}} \frac{4\rho' n}{\zeta_0} \bar{\mu}_k \\ & = \frac{4\rho'}{(1-\theta)^{1/2}} n \bar{\mu}_k^{1/2}. \end{aligned}$$

Dès lors la constante C_1 peut être prise comme suit

$$\frac{1}{(1-\theta)^{1/2}} n + \frac{8\rho'}{(1-\theta)^{1/2}} n \leq \frac{9\rho'}{(1-\theta)^{1/2}} n =: C_1. \quad \square$$

Lemme 18.22 *Il existe une constante $\bar{\alpha} \in]0, 1]$, telle que pour tout $\alpha \in [0, \bar{\alpha}]$, on a*

$$\begin{aligned} (1-\alpha)\bar{\mu}_k & \leq \bar{\mu}(z^k + \alpha d^k) \leq (1-\omega\alpha)\bar{\mu}_k \\ (x_i^k + \alpha dx_i^k)(s_i^k + \alpha ds_i^k) & \geq (1-\theta)\bar{\mu}(z^k + \alpha d^k). \end{aligned}$$

Si z^0 vérifie (18.30), alors il existe une constante $C_2 > 0$ indépendante de n telle que $\bar{\alpha} \geq C_2 n^{-2}$.

DÉMONSTRATION. On laisse tomber l'indice k . Avec la constante C_1 donnée par le lemme 18.21 et l'inégalité de Cauchy-Schwarz, on a $dx^\top ds \geq -C_1^2 \bar{\mu}$. Alors (18.26) conduit à

$$\bar{\mu}(z + \alpha d) \geq (1-\alpha)\bar{\mu} + \alpha \bar{\mu} \left(\sigma - \frac{\alpha C_1^2}{n} \right) \geq (1-\alpha)\bar{\mu},$$

dès que

$$\alpha \leq \bar{\alpha}_a := \frac{n\sigma}{C_1^2}.$$

En majorant cette fois $dx^\top dx \leq C_1^2 \bar{\mu}$, on obtient de (18.26)

$$\begin{aligned} \bar{\mu}(z(\alpha)) &\leq \bar{\mu} - \alpha \bar{\mu} + \alpha \sigma \bar{\mu} + \frac{\alpha^2 C_1^2}{n} \bar{\mu} \\ &\leq \bar{\mu} - \omega \alpha \bar{\mu}, \end{aligned} \quad (18.33)$$

dès que

$$\alpha \leq \bar{\alpha}_b := \frac{n(1-\omega-\sigma)}{C_1^2}.$$

En utilisant le fait que $x_i s_i \geq (1-\theta)\bar{\mu}$ lorsque $z \in V_\infty^-(\theta, \rho)$, on trouve

$$\begin{aligned} (x_i + \alpha dx_i)(s_i + \alpha ds_i) &= x_i s_i + \alpha \underbrace{(x_i ds_i + s_i dx_i)}_{\sigma \bar{\mu} - x_i s_i} + \alpha^2 dx_i ds_i \\ &\geq (1-\alpha)(1-\theta)\bar{\mu} + \alpha \sigma \bar{\mu} + \alpha^2 dx_i ds_i, \quad [\text{si } \alpha \leq 1] \\ &\geq ((1-\alpha)(1-\theta) + \alpha \sigma - \alpha^2 C_1^2) \bar{\mu} \\ &\geq (1-\theta) \left(1 - \alpha + \alpha \sigma + \frac{1}{n} \alpha^2 C_1^2 \right) \bar{\mu} \\ &\geq (1-\theta) \bar{\mu}(z(\alpha)), \quad [\text{par (18.33)}], \end{aligned}$$

où l'avant-dernière inégalité est vraie dès que

$$\alpha \leq \bar{\alpha}_c := \min \left(1, \frac{\theta \sigma}{C_1^2 (1 + \frac{1-\theta}{n})} \right).$$

Il suffit alors de prendre $\bar{\alpha} := \min(\bar{\alpha}_a, \bar{\alpha}_b, \bar{\alpha}_c) \leq 1$. On observe aussi que si z^0 vérifie (18.30), $C_1 = O(n)$ par le lemme 18.21. On en déduit alors que $\bar{\alpha} \geq C_2/n^2$. \square

Théorème 18.23 (convergence de l'algorithme 18.20) *L'algorithme 18.20 converge : $\bar{\mu} \rightarrow 0$ et $(r_b^k, r_c^k) \rightarrow 0$ q-linéairement. De plus, si z^0 vérifie (18.30), alors, pour tout $\varepsilon > 0$, il existe un indice $K = O(n^2 \log \varepsilon^{-1})$ tel que $\bar{\mu}_k \leq \varepsilon \bar{\mu}_0$, dès que $k \geq K$.*

DÉMONSTRATION. Comme l'algorithme choisit le plus grand pas tel que l'on ait (18.29), α_k est certainement supérieur à $\bar{\alpha} > 0$ donné par le lemme 18.22. Alors l'inégalité (18.29) montre que l'on a

$$\bar{\mu}_{k+1} \leq (1 - \omega \bar{\alpha}) \bar{\mu}_k.$$

Ce qui implique la convergence q-linéaire de $\bar{\mu}_k$ vers zéro. Le même raisonnement et

$$\|(r_b^{k+1}, r_c^{k+1})\|_2 = (1 - \alpha_k) \|(r_b^k, r_c^k)\|_2 \leq (1 - \bar{\alpha}) \|(r_b^k, r_c^k)\|_2$$

montre que (r_b^k, r_c^k) converge vers zéro q-linéairement.

Enfin, si z^0 vérifie (18.30), $\bar{\alpha} \geq C_2/n^2$ pour une constante $C_2 > 0$ indépendante de n , si bien que

$$\bar{\mu}_{k+1} \leq \left(1 - \frac{\omega C_2}{n^2}\right) \bar{\mu}_k.$$

On conclut en utilisant le lemme de complexité 18.5. \square

18.5 Mise en œuvre

18.5.1 Calcul du déplacement de Newton

Les algorithmes primaux-duaux calculent l'itéré suivant en se déplaçant le long de la direction de Newton obtenue par linéarisation du système d'optimalité perturbé (18.3). Cette direction est solution de l'équation de Newton

$$\begin{pmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} dx \\ dy \\ ds \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ \mu e - Xs \end{pmatrix}. \quad (18.34)$$

On a noté les *résidus*

$$r_c \equiv r_c(z) := A^\top y + s - c \quad \text{et} \quad r_b \equiv r_b(z) := Ax - b.$$

Intéressons-nous aux méthodes de calcul d'une solution du système de Newton (18.34). On peut éliminer ds grâce à la dernière équation, qui donne

$$ds = -X^{-1}Sdx + \mu X^{-1}e - s.$$

On obtient alors ce que l'on appelle le *système linéaire augmenté*

$$\begin{pmatrix} -X^{-1}S & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} -r_c - \mu X^{-1}e + s \\ -r_b \end{pmatrix}.$$

La matrice de ce système est symétrique, mais indéfinie, et garde le caractère creux éventuel de A . On peut le résoudre par la *factorisation de Bunch et Parlett* ou de *Bunch et Kaufman*.

On peut poursuivre la réduction de la dimension du système linéaire à résoudre en éliminant dx grâce à la première équation

$$dx = XS^{-1}A^\top dy + \mu S^{-1}e - x + XS^{-1}r_c.$$

On obtient ce que l'on appelle le *système normal* ou l'*équation normale*

$$(AXS^{-1}A^\top)dy = -r_b - A(\mu S^{-1}e - x + XS^{-1}r_c).$$

Si A est surjective, la matrice de ce système est symétrique définie positive. En pratique, c'est souvent ce système qui est résolu, par *factorisation de Cholesky creuse*.

Le mauvais conditionnement dû au facteur XS^{-1} requiert toutefois une factorisation adaptée [630]. Par ailleurs, si A a une colonne dense, la matrice $AXS^{-1}A^T$ perd son caractère creux. Si celles-ci ne sont pas trop nombreuses, on peut faire face à la difficulté de la manière suivante. Si C (resp. D) désigne l'ensemble des indices des colonnes creuses (resp. denses) de A , on écrit

$$AXS^{-1}A^T = A_C X_C S_C^{-1} A_C^T + A_D X_D S_D^{-1} A_D^T.$$

On calcule les facteurs de Cholesky creux de $A_C X_C S_C^{-1} A_C^T$ et on prend en compte le second terme $A_D X_D S_D^{-1} A_D^T$ par la technique de Sherman-Morrison-Woodbury.

18.5.2 Logiciels

La difficulté principale des méthodes de points intérieurs est due à la résolution des systèmes linéaires mal conditionnés que l'on y rencontre (section 18.5.1). Si l'on dispose de solveurs linéaires adéquats, les algorithmes peuvent s'écrire facilement et être ainsi adaptés à des situations particulières.

Logiciels généralistes : CPLEX [536], HOPDM [268], LIPSOL (en Matlab avec solveurs linéaires en Fortran) [642], LOQO, MOSEK [12, 537], OSX, PCx (en C) [150] et SEDUMI [575, 538].

Notes

L'algorithme des petits déplacements de la section 18.3.2 a été mis au point par Kojima, Mizuno et Yoshise [359 ; 1989] et par Monteiro et Adler [427 ; 1989] ; celui des grands déplacements de la section 18.3.3 est dû à Kojima, Mizuno et Yoshise [358 ; 1989]. L'algorithme prédicteur-correcteur de la section 18.3.4 fut énoncé et analysé par Mizuno, Todd et Ye [425 ; 1993], mais son principe est fondé sur les travaux antérieurs de Monteiro et Adler [427 ; 1989] et de Sonnevend, Stoer et Zhao [560, 561 ; 1989-1991]. L'algorithme prédicteur-correcteur de Mehrotra [419 ; 1992], que nous n'avons pas présenté, en est une version plus élaborée. Il suit le chemin central avec plus de précision en construisant une approximation d'ordre plus élevé ; c'est important pour ne pas perdre du temps dans le suivi des coudes que peut former le chemin central. Il permet aussi d'adapter à l'itération courante le facteur σ de réduction de $\bar{\mu}$, alors qu'il est donné *a priori* dans l'algorithme 18.16. C'est cette méthode qui est la plus souvent implémentée dans les codes de points intérieurs pour l'optimisation linéaire.

Le lecteur trouvera des compléments à cette brève introduction dans la monographie de S. Wright [629], qui est une très bonne référence sur les algorithmes de points intérieurs primaux-duaux en optimisation linéaire ; nous nous en sommes souvent inspiré dans la partie algorithmique de ce chapitre. D'autres aspects de l'approche par points intérieurs en optimisation linéaire sont exposés par den Hertog [165], Saigal [534] (algorithmes affines), Terlaky [580], Jansen [334], Roos, Terlaky et Vial [529], Vanderbei [604] et Ye [633].

Les méthodes de points intérieurs sont utilisées pour résoudre d'autres classes de problèmes convexes, ayant un côté combinatoire provenant de relations de complémentarité. On pourra consulter [73] pour les problèmes de complémentarité linéaire

monotone (incluant l'optimisation quadratique convexe), [626] pour l'optimisation semi-définie positive (voir aussi le chapitre 20), [512] pour l'optimisation conique (qui permet une extension à la dimension infinie), [456] pour l'optimisation convexe (voir aussi la section 16.1). Ces sujets sont aussi abordés par Ben-Tal et Nemirovski [44], qui présentent de nombreux exemples et applications.

L'utilisation des points intérieurs pour résoudre les problèmes d'optimisation non linéaire remonte à l'ouvrage pionnier de Fiacco et McCormick [207 ; 1968]. Il faut également citer l'article parfois oublié de McLinden [416 ; 1980], qui contient des résultats qualitatifs sur les problèmes convexes. En optimisation non linéaire, les techniques sont moins bien maîtrisées, les algorithmes ne sont pas encore stabilisés et l'intérêt de l'approche est débattu [241]. On trouvera dans [101] une étude de convergence d'un algorithme se ramenant à une suite de problèmes-barrières sous contraintes d'égalité non linéaires globalisés par régions de confiance (voir aussi la section 16.2) ; les codes KNITRO [459] et OPINL [225] s'en inspirent chacun à leur manière. Un état de l'art est présenté dans [217].

Exercices

18.1. On suppose que c , A et b sont donnés comme en optimisation linéaire (avec A surjective) et que (18.2) a lieu. Soit $\mu > 0$. Démontrez les affirmations suivantes.

- (i) Le problème $(D_\mu) \equiv (18.5)$ a une solution et une seule.
- (ii) Si l'on note (x_μ, y_μ, s_μ) le point central correspondant à μ , c'est-à-dire la solution unique de (18.3), la solution de (D_μ) n'est autre que (y_μ, s_μ) et x_μ est le multiplicateur optimal associé à la contrainte de (D_μ) .

18.2. On suppose que c , A et b sont donnés comme en optimisation linéaire et que (18.2) a lieu. Soit $w \in \mathbb{R}_{++}^n$. Montrez qu'il existe un unique triplet (x, y, s) , unique si A est surjective, vérifiant

$$\begin{cases} A^\top y + s = c, & s > 0 \\ Ax = b, & x > 0 \\ Xs = w. \end{cases}$$

En conséquence, l'application p définie en (18.15) est surjective (et bijective si A est surjective).

18.3. *Problèmes singuliers.* Supposons que (P) ait une solution \bar{x} telle que $\bar{x} > 0$. Alors

- (i) $c \perp \mathcal{N}(A)$,
- (ii) tout point admissible de (P) est solution primale,
- (iii) si (18.2) a lieu et $\mu \mapsto (x_\mu, s_\mu, s_\mu)$ est le chemin central, alors $x_\mu = \bar{x}$ pour tout $\mu > 0$, où \bar{x} est le centre analytique de \mathcal{F}_P .

Supposons maintenant que (D) ait une solution (\bar{y}, \bar{s}) telle que $\bar{s} > 0$. Alors

- (iv) $b = 0$,
- (v) tout point admissible de (D) est solution duale,
- (vi) si (18.2) a lieu, si A est surjective et si $\mu \mapsto (x_\mu, y_\mu, s_\mu)$ est le chemin central, alors $(y_\mu, s_\mu) = (\bar{y}, \bar{s})$ pour tout $\mu > 0$, où (\bar{y}, \bar{s}) est le centre analytique de \mathcal{F}_D .

18.4. *Le centre analytique n'est pas géométrique.* Montrez que le centre analytique de l'ensemble $\{x \in \mathbb{R}^2 : x \geq 0, x_1 + x_2 \leq 1\}$ est $(\frac{1}{3}, \frac{1}{3})$, alors que si le même ensemble est décrit par $\{x \in \mathbb{R}^2 : x \geq 0, x_1 + x_2 \leq 1, x_1 \leq 1\}$, on trouve $(\frac{1}{4}, \frac{3}{8})$.

18.5. *Autre définition du chemin central.* On considère le problème d'optimisation linéaire (P) et on suppose que (18.2) a lieu. Montrez que pour $\alpha > \text{val}(P)$, $X_\alpha := \{x \in \mathcal{F}_P :$