

17 Optimisation linéaire : théorie et algorithme du simplexe

Connaissances requises. Une grande partie de la section 2.4 nous sera utile, en particulier pour les notions de polyèdre convexe et de sommet, ainsi que pour le résultat d'existence de solution de problème d'optimisation linéaire (théorème 2.15). Les conditions d'optimalité d'un problème d'optimalité linéaire seront directement déduites des conditions de Karush, Kuhn et Tucker (4.23). Le problème dual d'un problème d'optimisation linéaire s'obtiendra par la dualité min-max (section 13.1).

17.1 Introduction

17.1.1 Le problème à résoudre

Un *problème d'optimisation linéaire* est un problème d'optimisation dans lequel le critère et les fonctions définissant les contraintes sont linéaires (on devrait dire affines). Comme nous le verrons, la formulation suivante du problème est tout à fait générale. Il s'agit de trouver la solution $x \in \mathbb{R}^n$ du problème

$$(P_L) \quad \begin{cases} \min c^\top x \\ Ax = b \\ x \geq 0. \end{cases}$$

Un problème d'optimisation linéaire écrit de cette manière est dit sous *forme standard*. On note $f(x) = c^\top x$, que l'on appelle *critère* du problème. Les données dans (P_L) sont deux vecteurs $c \in \mathbb{R}^n$ (appelé *coût* du problème) et $b \in \mathbb{R}^m$ ($m \leq n$), et une matrice A de dimension $m \times n$. La contrainte d'inégalité $x \geq 0$ veut dire que toutes les composantes de x doivent être positives : $x_i \geq 0$ pour $i = 1, \dots, n$. On notera $x > 0$, lorsque toutes les composantes de x seront strictement positives. L'ensemble

$$\mathbb{R}_+^n = \{x \in \mathbb{R}^n : x \geq 0\}$$

est appelé l'*orthant positif* de \mathbb{R}^n .

On dit que le problème (P_L) est *réalisable* si son *ensemble admissible*¹

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

est non vide. Un point dans \mathcal{F}_P est dit *admissible*. On dit que (P_L) est *borné* si la valeur minimale du critère dans (P_L) (donc pour $x \in \mathcal{F}_P$) est finie.

¹ "feasible set" en anglais, d'où la notation \mathcal{F} .

Nous allons étudier la structure du problème (P_L) et donner un algorithme pour le résoudre : l'algorithme du simplexe. Remarquons d'abord que le problème (P_L) n'est intéressant que du fait de la présence de contraintes d'inégalité. En effet, sans ces contraintes le problème se réduit à deux cas complémentaires triviaux. Si $c \in \mathcal{N}(A)^\perp$, tout point x vérifiant $Ax = b$ est solution. Si $c \notin \mathcal{N}(A)^\perp$, il existe une direction d telle que $Ad = 0$ et $c^\top d < 0$ et donc $f(x + td) \rightarrow -\infty$ lorsque $t \rightarrow +\infty$: si le problème est réalisable (en l'occurrence, il existe un x tel que $Ax = b$), il n'est pas borné.

En ce qui concerne les algorithmes de résolution de (P_L) , on distingue deux classes de méthodes :

- Les *méthodes de points-frontière* consistent à faire des déplacements le long des arêtes du polyèdre convexe \mathcal{F}_P défini par les contraintes de (P_L) . La méthode typique de cette classe est l'*algorithme du simplexe*, introduit par Dantzig en 1947 [101; 1951]. Cet algorithme n'est pas polynomial (voir Klee et Minty [239; 1972]), c'est-à-dire que le nombre d'opérations pour résoudre le problème n'est pas borné par un polynôme dépendant des dimensions de celui-ci ou de la longueur des données le définissant. Cependant, en pratique, ces algorithmes sont souvent très efficaces. La raison tient peut-être au fait que, bien que non polynomial dans le pire des cas, il est polynomial "en moyenne" pour des données distribuées aléatoirement suivant diverses lois (voir Borgwardt [54, 55], Smale [351], Megiddo [271] et Sharmir [346]).
- Les *méthodes de points intérieurs*, développées à partir des travaux de Karmarkar (1984), génèrent des itérés dans l'intérieur relatif de l'ensemble admissible de (P_L) . Ces méthodes peuvent être polynomiales et sont particulièrement intéressantes lorsqu'il y a beaucoup de contraintes d'inégalité, parce qu'elles ne ressentent pas l'irrégularité du bord du domaine admissible.

L'algorithme du simplexe est étudié dans ce chapitre. Les méthodes de points intérieurs sont présentées au chapitre 18.

17.1.2 Formulations canoniques

Un problème d'optimisation linéaire se présente souvent sous l'une des deux formes suivantes : la forme standard (P_L) donnée ci-dessus et la *forme canonique*

$$(P'_L) \quad \begin{cases} \min (c')^\top x \\ A'x \leq b'. \end{cases}$$

Ces deux formulations sont équivalentes dans le sens où le problème (P_L) peut s'écrire sous la forme (P'_L) :

$$\begin{cases} \min c^\top x \\ \begin{pmatrix} A \\ -A \\ -I \end{pmatrix} x \leq \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix} \end{cases}$$

et inversement, le problème (P'_L) peut s'écrire sous la forme (P_L) en décomposant $x = u - v$, avec $u \geq 0$ et $v \geq 0$, et en introduisant des *variables d'écart* $z \geq 0$:

$$\begin{cases} \min (c^\top - c^\top 0) \begin{pmatrix} u \\ v \\ z \end{pmatrix} \\ A'u - A'v + z = b' \\ u \geq 0, \quad v \geq 0, \quad z \geq 0. \end{cases}$$

On a pris l'habitude de décrire et d'étudier les algorithmes de résolution des problèmes d'optimisation linéaire mis sous forme standard (P_L) et c'est ce que nous ferons dans ce chapitre et le suivant. Cependant, il peut être plus avantageux de considérer une autre formulation, en fonction du problème que l'on a à résoudre. D'autre part, les codes d'optimisation linéaire généralistes acceptent le plus souvent des contraintes écrites sous la forme :

$$Ax = b \quad \text{et} \quad l \leq Bx \leq u,$$

où les matrices A et B et les vecteurs b , l et u ont des dimensions appropriées. Il faut donc parfois adapter les algorithmes au cadre que l'on considère; ceci se fait en général sans difficulté.

17.1.3 Exemples : problèmes d'optimisation dans des réseaux

Un *réseau* n'est pas autre chose qu'un *graphe* (orienté ou non), c'est-à-dire une collection de nœuds et d'arcs qui relient ces nœuds. Disons qu'il y a m nœuds, indicés par $i = 1, \dots, m$, et n arcs, indicés par $j = 1, \dots, n$. Ces derniers peuvent aussi être spécifiés par des couples (ordonnés ou non, suivant que le graphe est ou n'est pas orienté) formés des nœuds qu'ils relient : l'arc (i_1, i_2) relie les nœuds i_1 et i_2 .

Dans la discussion qui suit, le réseau est supposé être utilisé comme support à l'acheminement d'un produit (eau, gaz, électricité, voitures, trains, avions, télécommunications, messages internet, quantité abstraite, *etc*). La quantité de produit qui transite par l'arc j est notée x_j . Il est normal de supposer que c'est une grandeur positive. L'équilibre du réseau s'exprime en écrivant que la quantité de produit qui sort du nœud i est égale à la quantité qui y entre plus la quantité b_i qui y est produite ("loi de Kirchhoff"). Pour tout i , on doit avoir

$$\sum_{\substack{j : \text{arc sortant} \\ \text{du nœud } i}} x_j = \sum_{\substack{j : \text{arc entrant} \\ \text{du nœud } i}} x_j + b_i. \quad (17.1)$$

La quantité b_i produite au nœud i peut représenter ce qui y est apporté (auquel cas $b_i > 0$) ou ce qui y est consommé ou demandé (auquel cas $b_i < 0$). Les équations précédentes s'écrivent sous forme matricielle :

$$Ax = b \quad \text{et} \quad x \geq 0, \quad (17.2)$$

où la matrice A est appelée la *matrice d'incidence du réseau*. Cette matrice peut avoir des dimensions considérables en pratique, mais elle est très creuse puisqu'elle n'a que deux éléments non nuls par colonne. En effet, au vu de (17.1), la colonne j associée à l'arc j contient $A_{ij} = +1$ si i est le nœud d'origine de l'arc j , $A_{ij} = -1$ si i est le nœud de destination de l'arc j et $A_{ij} = 0$ dans les autres cas. En particulier, la somme des lignes de A est nulle, ce qui implique que A n'est pas surjective. Ceci peut poser des

difficultés algorithmiques et requière au moins une condition de compatibilité sur b , à savoir $b \in \mathcal{R}(A)$, ce qui conduit à

$$\sum_{i=1}^m b_i = 0. \quad (17.3)$$

On montre cependant que A est de rang $m - 1$ lorsque le graphe associé au réseau est *connexe*, ce qui veut dire que deux nœuds distincts i et i' peuvent être reliés par un chemin (une suite d'arcs $(i_1, i_2), (i_2, i_3), \dots, (i_{p-1}, i_p)$, avec $i_1 = i$ et $i_p = i'$); voir [40] par exemple.

Un exemple de problème d'optimisation linéaire classique dans les réseaux est le *problème du transport à coût minimal*. On cherche à acheminer des produits identiques depuis certains *nœuds d'origine* (disons pour $i \in O$) jusqu'à des *nœuds de destination* (disons pour $i \in D$, avec $O \cap D = \emptyset$). Ceci s'exprime par le positionnement des composantes de b aux valeurs désirées: $b_O > 0$, $b_D < 0$ et $b_i = 0$ si $i \notin O \cup D$, tout en respectant (17.3). D'autre part, en supposant que le coût de transport sur chaque arc est linéaire en x_j (coût c_j par unité transportée), le coût total du transport sera

$$\sum_{j=1}^n c_j x_j = c^\top x.$$

C'est la quantité que l'on cherche à minimiser. On reconnaît dans la minimisation de ce critère sous les contraintes de réseau (17.2), le problème d'optimisation linéaire sous forme standard (P_L).

Un autre exemple de problème d'optimisation linéaire classique est celui du *plus court chemin dans un graphe*. C'est un cas particulier du problème précédent: $b_{i_0} = +1$ au nœud de départ i_0 , $b_{i_1} = -1$ au nœud d'arrivée i_1 et les autres b_i sont nuls, tandis que $c_j > 0$ donne la longueur du chemin représenté par l'arc j . Si \bar{x} est la solution du problème, $\bar{x}_j = 0$ signifie que l'on ne passe pas par l'arc j et $\bar{x}_j = 1$ signifie que l'on y passe. On peut avoir des valeurs entre 0 et 1 si le problème a plusieurs solutions.

17.2 Étude du problème

17.2.1 Structure de l'ensemble admissible

L'ensemble admissible \mathcal{F}_P d'un problème d'optimisation linéaire est un polyèdre convexe (voir la section 2.4), exprimé dans une représentation duale. Les développements théoriques et algorithmiques en optimisation linéaire se font en général sur la forme suivante de l'ensemble admissible

$$P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\},$$

où A est $m \times n$ et $b \in \mathbb{R}^m$. On supposera alors toujours que

A est surjective.

Théoriquement, ce n'est pas une hypothèse restrictive (en pratique, c'est plus délicat). En effet si $b \notin \mathcal{R}(A)$, l'ensemble admissible est vide et le problème est mal posé. Sinon, on peut toujours éliminer les équations redondantes de $Ax = b$ pour se ramener à une matrice surjective (si des lignes de A sont linéairement dépendantes – $A^\top y = 0$ pour un $y \neq 0$ – les éléments de b obéissent à la même dépendance – $b^\top y = 0$).

L'optimisation linéaire a son propre jargon, que l'on doit reconnaître si l'on veut comprendre les ouvrages et articles spécialisés. Certains termes apportent pourtant d'inutiles confusions et nous les éviterons. Il en va ainsi de "solution" ou "solution admissible" pour désigner un point admissible (c'est une solution des équations qui définissent P) ou encore de "solution basique admissible" pour désigner un sommet de P .

Définition 17.1 On appelle *base d'indices* ou simplement *base* un ensemble B de m indices pris dans $\{1, \dots, n\}$ tels que la sous-matrice A_B formée des m colonnes correspondantes de A soit inversible. Si $x \in \mathbb{R}^n$, les composantes x_i avec $i \in B$ sont alors dites *basiques* de x , celles avec $i \notin B$ sont dites *non basiques*.

À une base d'indices B on associe l'ensemble d'indices complémentaire $N = \{1, \dots, n\} \setminus B$. Après permutations éventuelles des colonnes, on pourra donc écrire

$$A = (A_B \ A_N),$$

avec A_B inversible. On note (x_B, x_N) la partition correspondante d'un point $x \in \mathbb{R}^n$.

On adopte une notation déjà utilisée. Si $x \in \mathbb{R}^n$, on note

$$I^+(x) := \{i : x_i > 0\} \quad \text{et} \quad I^0(x) := \{i : x_i = 0\}.$$

On rappelle que $x \in P$ est un sommet de P si la sous matrice A_B , avec $B = I^+(x)$, est injective (proposition 2.10). On a donc nécessairement $|B| \leq m$, et on peut très bien avoir $|B| < m$.

Définition 17.2 On dit qu'un sommet $x \in P$ est *dégénéré* si $|I^+(x)| < m$ et qu'il est *non dégénéré* si $|I^+(x)| = m$.

Exemples 17.3 Dans \mathbb{R}^2 , $x = 0$ est un sommet dégénéré de l'ensemble défini par

$$\begin{cases} x_2 - x_1 = 0 \\ x \geq 0. \end{cases} \quad \begin{array}{c} \nearrow \\ \uparrow \\ \leftarrow x \end{array}$$

Dans \mathbb{R}^3 , le point $x = (0, 0, 1)$ est un sommet dégénéré de l'ensemble défini par

$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ -x_1 + x_2 = 0 \\ x \geq 0. \end{cases} \quad \begin{array}{c} x \\ \nearrow \\ \leftarrow e_2 \\ \nwarrow \\ e_1 \end{array}$$

17.2.2 Existence de solution et conditions d'optimalité

On considère le problème (P_L) sous la forme standard

$$(P_L) \quad \begin{cases} \min c^\top x \\ Ax = b \\ x \geq 0. \end{cases}$$

On note $f(x) := c^\top x$, $\mathcal{F}_P := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ l'ensemble admissible et

$$\inf f := \inf_{x \in \mathcal{F}_P} c^\top x.$$

Pour le problème (P_L) , l'existence de solution ne résulte pas directement du théorème sur la minimisation de fonctions continues sur un compact (théorème 1.1). En effet, l'ensemble admissible \mathcal{F}_P n'est pas nécessairement compact. De plus, $f(x)$ ne tend pas nécessairement vers $+\infty$ lorsque $\|x\| \rightarrow \infty$ dans \mathcal{F}_P (sauf si l'ensemble des solutions est borné et si \mathcal{F}_P n'est pas borné, voir l'exercice ??). Il est donc nécessaire de donner une démonstration spécifique. Une possibilité est d'utiliser la remarque 2.14 comme dans la démonstration de la proposition 2.15. Nous rappelons le résultat ci-dessous.

Théorème 17.4 (existence de solution) *Le problème (P_L) a une solution si et seulement si il est réalisable ($\mathcal{F}_P \neq \emptyset$) et borné ($\inf f > -\infty$).*

L'algorithme du simplexe repose sur le résultat suivant.

Proposition 17.5 (existence de solution-sommet) *Si le problème (P_L) a une solution, alors il a une solution en un sommet de \mathcal{F}_P .*

DÉMONSTRATION. L'ensemble des solutions de (P_L) ,

$$\text{sol}(P_L) = \{x \in \mathbb{R}^n : c^\top x = \text{val}(P_L), Ax = b, x \geq 0\},$$

est un polyèdre convexe écrit sous forme standard. Il a donc un sommet (proposition 2.10), disons \hat{x} . D'autre part, $\text{sol}(P_L)$ est une face de \mathcal{F}_P (exercice 2.12). Donc son sommet \hat{x} est aussi un sommet de \mathcal{F}_P (exercice 2.6). \square

Voici le résultat donnant les conditions d'optimalité du problème (P_L) .

Proposition 17.6 (conditions d'optimalité) *Le point x est solution de (P_L) si et seulement si il existe $y \in \mathbb{R}^m$ et $s \in \mathbb{R}^n$ tels que*

$$\begin{cases} (a) A^\top y + s = c, & s \geq 0 \\ (b) Ax = b, & x \geq 0 \\ (c) x^\top s = 0. \end{cases} \quad (17.4)$$

DÉMONSTRATION. Ce sont les conditions de KKT qui sont nécessaires (on utilise ici le fait que les contraintes de (P_L) sont qualifiées, car affines) et suffisantes (car le problème (P_L) est convexe). On les obtient en introduisant le *lagrangien*

$$\ell(x, y, s) = c^\top x - y^\top (Ax - b) - x^\top s.$$

□

La nécessité des conditions (17.4) peut aussi se voir en observant que, par optimalité de x , le critère augmente le long des directions d tangentes en x à l'ensemble admissible (voir la proposition 4.5), ce qui s'exprime par

$$c \in \{d : Ad = 0, d_I \geq 0\}^*, \quad \text{avec } I = \{i : x_i = 0\},$$

et en utilisant le lemme de Farkas. D'autre part, le fait que les conditions (17.4) soient suffisantes pour impliquer l'optimalité de x peut aussi se montrer directement : pour tout $x' \in \mathcal{F}_P$, on a

$$\begin{aligned} c^\top x &= s^\top x + y^\top Ax && \text{[par (a)]} \\ &= y^\top b && \text{[par (b) et (c)]} \\ &\leq s^\top x' + y^\top Ax' && \text{[par } Ax' = b, x' \geq 0 \text{ et } s \geq 0] \\ &= c^\top x' && \text{[par (a)].} \end{aligned}$$

Le couple (y, s) donné par la proposition précédente est appelé *solution duale* du problème linéaire (P_L) et x est alors appelé *solution primale*. La variable s est aussi appelée *variable d'écart duale*, car elle joue le rôle de variable d'écart dans la relation $A^\top y \leq c$, qui est une autre façon d'écrire l'équation (17.4)(a). On note $\mathcal{S}_P \subset \mathbb{R}^n$ l'ensemble des solutions primales, $\mathcal{S}_D \subset \mathbb{R}^m \times \mathbb{R}^n$ l'ensemble des solutions duales et \mathcal{S}_{PD} l'ensemble des triplets (x, y, s) formés de solutions primales-duales.

Proposition 17.7 (produit cartésien des solutions) *On a*

$$\mathcal{S}_{PD} = \mathcal{S}_P \times \mathcal{S}_D.$$

Autrement dit, si (x^1, y^1, s^1) et (x^2, y^2, s^2) sont des solutions primales-duales alors (x^1, y^2, s^2) et (x^2, y^1, s^1) le sont aussi.

DÉMONSTRATION. Par hypothèse, on a

$$\begin{cases} A^\top y^1 + s^1 = c, & s^1 \geq 0 \\ Ax^1 = b, & x^1 \geq 0 \\ (x^1)^\top s^1 = 0 \end{cases} \quad \text{et} \quad \begin{cases} A^\top y^2 + s^2 = c, & s^2 \geq 0 \\ Ax^2 = b, & x^2 \geq 0 \\ (x^2)^\top s^2 = 0. \end{cases}$$

On voit qu'il suffit de montrer que $(x^1)^\top s^2 = 0$ et $(x^2)^\top s^1 = 0$. Comme $c^\top x^1 = c^\top x^2$, on a

$$\begin{aligned}
(x^1)^\top s^2 &= (x^1)^\top (c - A^\top y^2) \\
&= (x^1)^\top c - b^\top y^2 \\
&= (x^2)^\top c - (Ax^2)^\top y^2 \\
&= (x^2)^\top (c - A^\top y^2) \\
&= (x^2)^\top s^2 \\
&= 0.
\end{aligned}$$

De même pour $(x^2)^\top s^1 = 0$. □

On montrera plus loin (en section 17.3) que les couples formés des solutions primales et duales sont les points-selles du lagrangien ℓ sur $\mathbb{R}^n \times (\mathbb{R}^m \times \mathbb{R}_+^n)$. Le résultat précédent peut donc aussi se déduire de cette observation puisque l'ensemble des points-selles est le produit cartésien de l'ensemble des solutions primales par l'ensemble des solutions duales (corollaire 13.5).

L'identité (17.4)(c) forme ce que l'on appelle les *conditions de complémentarité*. Comme $x \geq 0$ et $s \geq 0$, elle s'écrit de manière équivalente

$$\forall i \in \{1, \dots, n\}, \quad x_i > 0 \implies s_i = 0,$$

qui exprime que si la i -ième contrainte d'inégalité n'est pas active, le multiplicateur associé est nul. On dit qu'une solution primale-duale (x, y, s) de (P_L) est *strictement complémentaire* si

$$\forall i \in \{1, \dots, n\}, \quad x_i > 0 \iff s_i = 0.$$

Toutes les solutions ne sont pas nécessairement strictement complémentaires. Le résultat suivant montre cependant qu'en optimisation linéaire on peut toujours trouver une solution strictement complémentaire.

Introduisons deux ensembles d'indices B et N , qu'il faudra se garder de confondre avec les ensembles d'indices B et N introduits au début de ce chapitre :

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

Par les conditions de complémentarité et du fait que l'ensemble des solutions primales-duales est un produit cartésien (proposition 17.7), on a $B \cap N = \emptyset$. En effet, si $i \in B$, alors $s_i = 0$ pour toute solution duale (y, s) , et donc $i \notin N$. Compte tenu de cette observation, montrer qu'il existe une solution primale-duale strictement complémentaire revient à montrer que B et N forment une partition de $\{1, \dots, n\}$.

Proposition 17.8 (existence d'une solution strictement complémentaire) *Si le problème (P_L) a une solution, alors il a une solution primale-duale strictement complémentaire, et donc*

$$B \cap N = \emptyset \quad \text{et} \quad B \cup N = \{1, \dots, n\}.$$

DÉMONSTRATION. Construisons d'abord la solution primale. Pour que celle-ci ait le plus grand nombre de composantes non nulles possible, on l'écrit comme combinaison convexe de solution ayant chacune une des composantes pouvant être strictement positive: par définition de B , on sait que pour tout $i \in B$, on peut choisir une solution \bar{x}^i de (P_L) avec $\bar{x}_i^i > 0$. On prend

$$\bar{x} = \frac{1}{|B|} \sum_{i \in B} \bar{x}^i,$$

où $|B|$ désigne le nombre d'éléments de B . C'est aussi une solution (c'est une combinaison convexe de solutions d'un problème convexe, ou on vérifie directement que $c^\top \bar{x} = \inf f$, $A\bar{x} = b$, $\bar{x} \geq 0$) et elle vérifie $\bar{x}_B > 0$ (c'est une solution avec le moins de contraintes actives possible, située dans l'intérieur relatif de la face optimale).

D'après la proposition 17.7, il reste à trouver une solution duale (\bar{y}, \bar{s}) avec $\bar{s}_i > 0$ pour $i \notin B$. Dans ce but, on considère le problème d'optimisation linéaire suivant

$$\begin{cases} \min & -\sum_{i \notin B} x_i \\ & Ax = b \\ & x \geq 0 \\ & c^\top x \leq \inf f. \end{cases}$$

Son ensemble admissible sont les solutions de (P_L) et par définition de B , sa valeur optimale est nulle. Comme \bar{x} est solution de ce problème, on peut trouver $(\hat{y}, \hat{s}, \hat{\mu})$ tel que

$$\begin{cases} A^\top \hat{y} + \hat{s} + \sum_{i \notin B} e^i = \hat{\mu}c, & \hat{s} \geq 0, \quad \hat{\mu} \geq 0 \\ \bar{x}^\top \hat{s} = 0. \end{cases}$$

On a noté e^i le i -ième vecteur de base de \mathbb{R}^n . Si $\hat{\mu}$ n'est pas nul, on peut conclure avec la solution duale $\frac{1}{\hat{\mu}}(\hat{y}, \hat{s} + \sum_{i \notin B} e^i)$, mais rien ne garantit cela. On combine alors cette solution duale avec une solution duale associée à \bar{x} : on sait en effet qu'il existe (y, s) tel que

$$\begin{cases} A^\top y + s = c, & s \geq 0 \\ \bar{x}^\top s = 0. \end{cases}$$

On prend

$$\bar{y} := \frac{1}{1 + \hat{\mu}}(y + \hat{y}) \quad \text{et} \quad \bar{s} := \frac{1}{1 + \hat{\mu}} \left(s + \hat{s} + \sum_{i \notin B} e^i \right),$$

qui est clairement une solution duale de (P_L) . De plus, comme $s \geq 0$ et $\hat{s} \geq 0$, on a $\bar{s}_i > 0$ pour $i \notin B$. \square

Les ensembles d'indices B et N permettent de donner une description simple de l'ensemble \mathcal{S}_p des solutions primales et l'ensemble \mathcal{S}_d des solutions duales de (P_L) . Pour une raison qui apparaîtra claire plus loin, on note

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

Proposition 17.9 *Si (P_L) a une solution, alors l'ensemble de ses solutions primales est la face de \mathcal{F}_P définie par*

$$\mathcal{S}_P = \{x \in \mathcal{F}_P : x_N = 0\}$$

et l'ensemble de ses solutions duales est la face de \mathcal{F}_D définie par

$$\mathcal{S}_D = \{x \in \mathcal{F}_D : s_B = 0\}.$$

DÉMONSTRATION. Si $x \in \mathcal{S}_P$, alors $x \in \mathcal{F}_P$ et $x_N = 0$ (par la proposition 17.8 et la définition de B). Inversement, si $x \in \mathcal{F}_P$ et $x_N = 0$, alors en prenant une solution duale (y, s) arbitraire, le triplet (x, y, s) vérifie les conditions d'optimalité (17.4) (car $s_B = 0$ d'après la proposition 17.8 et la complémentarité). Donc x est solution de (P_L) .

Il reste à montrer que le convexe \mathcal{S}_P est une face de \mathcal{F}_P . Pour cela, supposons que $x \in \mathcal{S}_P$ s'écrive comme suit $x = (1-t)x' + tx''$, avec $x', x'' \in \mathcal{F}_P$ et $t \in]0, 1[$. Il faut montrer que x' et $x'' \in \mathcal{S}_P$. Or on a

$$0 = x_N = (1-t)x'_N + tx''_N.$$

Comme $t > 0$, $(1-t) > 0$, $x'_N \geq 0$ et $x''_N \geq 0$, on a nécessairement $x'_N = 0$ et $x''_N = 0$, c'est-à-dire x' et $x'' \in \mathcal{S}_P$.

L'affirmation concernant \mathcal{S}_D se démontre de la même manière. □

17.3 Dualité

17.3.1 Dualité en optimisation linéaire

Nous ferons référence ici aux notions vues à la section 13.1. On part du problème d'optimisation linéaire

$$(P_L) \quad \begin{cases} \min c^\top x \\ Ax = b \\ x \geq 0 \end{cases}$$

et on introduit la fonction de couplage

$$\varphi(x, y) = c^\top x - y^\top (Ax - b),$$

par laquelle on n'a dualisé que les contraintes d'égalité. On a

$$\inf_{x \geq 0} \sup_y \varphi(x, y) = \inf_{x \geq 0} \begin{cases} c^\top x & \text{si } Ax = b \\ +\infty & \text{sinon.} \end{cases}$$

On retrouve donc le problème (P_L) , si on adopte la convention que la valeur minimale de (P_L) vaut $+\infty$ lorsque l'ensemble admissible est vide.

Alors, le problème dual de (P_L) pour la fonction φ ci-dessus s'écrit

$$\sup_y \inf_{x \geq 0} \left((c - A^\top y)^\top x + b^\top y \right) = \sup_y \begin{cases} b^\top y & \text{si } A^\top y \leq c \\ -\infty & \text{sinon.} \end{cases}$$

On trouve finalement comme problème dual

$$(D_L) \quad \begin{cases} \max b^\top y \\ A^\top y \leq c, \end{cases}$$

si on prend la convention que la valeur maximale dans (D_L) vaut $-\infty$ lorsque l'ensemble admissible de (D_L) est vide. C'est aussi un problème d'optimisation linéaire.

On peut obtenir un problème dual de (P_L) en dualisant également la contrainte $x \geq 0$, c'est-à-dire en utilisant le lagrangien $\ell(x, y, s) = c^\top x - y^\top (Ax - b) - x^\top s$ comme fonction de couplage (il faut alors imposer $s \geq 0$ dans la dualité). On obtient alors comme problème dual

$$\begin{cases} \max b^\top y \\ A^\top y + s = c \\ s \geq 0, \end{cases}$$

qui est équivalent à (D_L) .

Remarquons enfin que les conditions d'optimalité de (D_L) sont aussi celles de (P_L) et que le dual du dual est le primal.

On dit que (y, s) est admissible pour le problème dual (D_L) , si $A^\top y + s = c$ et si $s \geq 0$. On note

$$\begin{aligned} \mathcal{F}_P &:= \{x \in \mathbb{R}^n : Ax = b, x \geq 0\} \\ \mathcal{F}_D &:= \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^\top y + s = c, s \geq 0\} \end{aligned}$$

les ensembles admissibles de (P_L) et de (D_L) respectivement.

Appliquons à présent les résultats obtenus dans le cadre général au problème d'optimisation linéaire (P_L) . De la proposition 13.2, on obtient le résultat suivant.

Proposition 17.10 (dualité faible en optimisation linéaire) *On a*

$$\sup_{A^\top y \leq c} b^\top y \leq \inf_{\substack{Ax=b \\ x \geq 0}} c^\top x. \tag{17.5}$$

Remarquons que si $x \in \mathcal{F}_P$ et si $(y, s) \in \mathcal{F}_D$, la différence entre la valeur du critère primal et celle du critère dual est donnée par la formule

$$\boxed{c^\top x - b^\top y = x^\top s.} \tag{17.6}$$

Théorème 17.11 (dualité forte en optimisation linéaire) *Les propriétés suivantes sont équivalentes :*

- (i) (P_L) et (D_L) sont réalisables,
- (ii) (P_L) a une solution,
- (iii) (D_L) a une solution.

Lorsque ces propriétés ont lieu il n'y a pas de saut de dualité, c'est-à-dire que l'on a égalité en (17.5).

DÉMONSTRATION. [(i) \Rightarrow (ii)] Si (i) a lieu, l'inégalité de dualité faible (17.5) montre que (P_L) est réalisable et borné. Ce problème a donc une solution (théorème 17.4).

[(ii) \Rightarrow (iii)] Les conditions d'optimalité (17.4), qui ont lieu si (P_L) a une solution, sont aussi celles de (D_L) , qui a donc aussi une solution.

[(iii) \Rightarrow (i)] Les conditions d'optimalité (17.4) de (D_L) montrent que (P_L) est réalisable.

Supposons que les propriétés (i)–(iii) ont lieu, que \bar{x} soit une solution de (P_L) et que \bar{y} soit une solution de (D_L) . La valeur optimale duale est donc $b^\top \bar{y}$ et la valeur optimale primale est donc $c^\top \bar{x}$. Avec $\bar{s} := c - A^\top \bar{y}$ et (17.6), on trouve que $c^\top \bar{x} - b^\top \bar{y} = \bar{x}^\top \bar{s}$, qui est nul car l'ensemble des solutions primales-duales est un produit cartésien (proposition 17.7). \square

Le point (i) est souvent un moyen commode de montrer qu'un problème d'optimisation linéaire a une solution : on montre qu'il est réalisable et que son dual l'est également. Ce théorème permet de retrouver le lemme de Farkas (voir l'exercice ??). Quelques conséquences de ce théorème sont données à l'exercice ??.

17.3.2 Une relation entre le primal et le dual

Le théorème de dualité forte a montré que les propriétés des problèmes primal et dual sont très liées. Dans cette section nous donnons d'autres exemples de relations entre les deux problèmes.

On dit que $x \in \mathcal{F}_P$ est *strictement admissible* pour le problème primal (P_L) si $x > 0$ (toutes les composantes de x sont > 0). De même, on dit que $(y, s) \in \mathcal{F}_D$ est *strictement admissible* pour le problème dual (D_L) si $s > 0$. On définit les ensembles de points strictement admissibles

$$\begin{aligned}\mathcal{F}_P^o &:= \{x \in \mathbb{R}^n : Ax = b, x > 0\} \\ \mathcal{F}_D^o &:= \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^\top y + s = c, s > 0\}.\end{aligned}$$

S'ils sont non vides, ce sont les intérieurs relatifs de \mathcal{F}_P et \mathcal{F}_D respectivement.

Pour les problèmes d'optimisation convexes, le rapprochement des propositions 4.25 [(QC-S) \iff (QC-MF)] et 4.27 [(QC-MF) \iff Λ_* borné] suggère que l'existence d'un point *strictement* admissible primal équivaut au caractère *borné* de l'ensemble des solutions duales. En optimisation linéaire, le dual du dual est le primal, si bien que l'existence d'une solution duale strictement admissible devrait aussi équivaloir au caractère borné de l'ensemble des solutions primales. C'est une manière de trouver

naturelles les équivalences énoncées dans la proposition suivante. Nous en donnons une démonstration directe ci-dessous; l'exercice ?? propose d'autres approches moins techniques et qui pourront être utiles pour étendre ce résultat à d'autres problèmes. On y a noté

$$\mathcal{S}_{D,s} = \{s : \text{il existe } y \in \mathbb{R}^m \text{ tel que } (y, s) \in \mathcal{S}_D\}$$

la projection de \mathcal{S}_D sur \mathbb{R}^n . On note aussi $e = (1 \dots 1)^\top$.

Proposition 17.12 *Supposons que les problèmes primal et dual soient réalisables ($\mathcal{F}_P \neq \emptyset$ et $\mathcal{F}_D \neq \emptyset$). Alors*

$$\mathcal{S}_P \text{ est borné} \iff \mathcal{F}_D^o \neq \emptyset, \tag{17.7}$$

$$\mathcal{S}_{D,s} \text{ est borné} \iff \mathcal{F}_P^o \neq \emptyset. \tag{17.8}$$

DÉMONSTRATION. Comme

$$\mathcal{S}_P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0, c^\top x \leq \text{val}(P_L)\}$$

est un convexe fermé non vide, \mathcal{S}_P est borné si et seulement si son cône asymptotique (exercice 2.14)

$$\mathcal{S}_P^\infty = \{d \in \mathbb{R}^n : Ad = 0, d \geq 0, c^\top d \leq 0\}$$

est réduit au singleton $\{0\}$ (proposition 2.6). Si \mathcal{F}_D^o est non vide, il existe $(y, s) : c = A^\top y + s$, avec $s > 0$. Alors pour $d \in \mathcal{S}_P^\infty$, on a $0 \geq c^\top d = y^\top Ad + s^\top d = s^\top d$, ce qui implique que $d = 0$ (car $d \geq 0$ et $s > 0$). Inversement, si $\mathcal{S}_P^\infty = \{0\}$, son cône dual est \mathbb{R}^n . Celui-ci s'écrit (corollaire 2.32)

$$(\mathcal{S}_P^\infty)^* = \{A^\top y + s - \alpha c : s \in \mathbb{R}_+^n, \alpha \in \mathbb{R}_+\}.$$

Il existe donc $y \in \mathbb{R}^m, s \in \mathbb{R}_+^n$ et $\alpha \in \mathbb{R}_+$ tels que $A^\top y + s - \alpha c = -e + c$, ce qui s'écrit aussi $A^\top y + (s + e) = (\alpha + 1)c$. Alors $(y, s+e)/(\alpha+1) \in \mathcal{F}_D^o$.

La seconde partie de la proposition peut se démontrer de la même manière. \square

Évidemment, si $\mathcal{S}_{D,s}$ n'est pas borné, \mathcal{S}_D ne l'est pas non plus, ni $\mathcal{S}_{D,y} := \{y : \text{il existe } s \in \mathbb{R}^n \text{ tel que } (y, s) \in \mathcal{S}_D\}$ d'ailleurs. Par ailleurs, si A est surjective, il revient au même de dire que $\mathcal{S}_{D,s}$ est borné ou que \mathcal{S}_D est borné.

17.4 Algorithme du simplexe

On considère le problème d'optimisation linéaire sous forme standard

$$(P_L) \begin{cases} \min (f(x) = c^\top x) \\ Ax = b \\ x \geq 0, \end{cases}$$

dont les conditions d'optimalité s'écrivent (proposition 17.6) : $\exists (y, s) \in \mathbb{R}^m \times \mathbb{R}^n$, tels que

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

On sait que si ce problème a une solution il a une solution sur un sommet du polyèdre (proposition 17.5)

$$\mathcal{F}_P = \{x : Ax = b, x \geq 0\}.$$

Si m et n sont grands, il peut y avoir beaucoup de sommets, ce qui exclut de les explorer tous. L'*algorithme du simplexe* de Dantzig [101; 1951] sélectionne les sommets à explorer. L'algorithme se déroule en général en deux étapes :

- Phase I: trouver un sommet de \mathcal{F}_P ,
- Phase II: passer d'un sommet à l'autre de manière à faire décroître f , jusqu'à ce qu'on trouve une solution-sommet.

Dans la phase I, on résout un problème d'optimisation linéaire auxiliaire dont on connaît un sommet (voir la section 17.4.3). On peut donc utiliser l'algorithme du simplexe pour résoudre ce problème. Dans la phase II, on évite de parcourir tous les sommets en faisant décroître le critère à chaque itération. On ne visite donc plus les sommets dont la valeur du critère est supérieure à celle au sommet courant.

17.4.1 Description

Hypothèse. On suppose que A a plein rang $m < n$.

Nous allons décrire une itération de l'algorithme du simplexe, en partant d'un sommet \hat{x} de base B pour atteindre un nouveau sommet avec une valeur (si possible strictement) inférieure du critère. On a donc au départ de l'itération

$$\hat{x}_B \geq 0, \quad \hat{x}_N = 0 \quad \text{et} \quad A_B \hat{x}_B = b \quad (\text{donc} \quad \hat{x}_B = A_B^{-1}b).$$

1 *Reconnaître l'optimalité.* Soit x un point de \mathcal{F}_P , qui vérifie donc $Ax = b$. B étant une base d'indices, on peut exprimer les composantes basiques x_B de x en fonction de b et des composantes non basiques x_N :

$$x_B = A_B^{-1}(b - A_N x_N).$$

On peut également exprimer le coût $c^\top x$ en fonction de x_N :

$$c^\top x = c_B^\top A_B^{-1}(b - A_N x_N) + c_N^\top x_N.$$

Son gradient par rapport à x_N s'appelle le *coût réduit*. Il s'écrit

$$r = c_N - A_N^\top A_B^{-\top} c_B.$$

Proposition 17.13 *Si $r \geq 0$, alors le sommet \hat{x} est optimal. Inversement, si \hat{x} est un sommet non dégénéré optimal, alors $r \geq 0$.*

DÉMONSTRATION. Supposons que $r \geq 0$. Si l'on introduit le multiplicateur $y = A_B^{-T}c_B$, on trouve $r = c_N - A_N^T y$. En réarrangeant ces deux dernières équations et en posant $s = (0, r)$, on trouve

$$A^T y + s = c, \quad s \geq 0 \quad \text{et} \quad \hat{x}^T s = 0.$$

Inversement, si \hat{x} est un sommet non dégénéré optimal, il existe une solution duale (\hat{y}, \hat{s}) telle que $A^T \hat{y} + \hat{s} = c$, $\hat{s} \geq 0$ et $\hat{x}^T \hat{s} = 0$. Comme \hat{x} est non dégénéré, on a $\hat{x}_B > 0$ et donc $\hat{s}_B = 0$, si bien que $A_B^T \hat{y} = c_B$ ou encore $\hat{y} = A_B^{-T} c_B$. Alors le coût réduit s'écrit $r = c_N - A_N^T \hat{y} = \hat{s}_N \geq 0$. \square

Si le sommet optimal est dégénéré, il peut y avoir un coût réduit $r \not\geq 0$, comme le montre l'exemple suivant :



L'ensemble admissible est réduit au singleton $\{(0, 0)\}$. La solution du problème est donc forcément $\hat{x} = (0, 0)$, qui est un sommet dégénéré. Si on prend la base d'indices $B = \{1\}$, on a un coût réduit strictement négatif ($r = -1$). Ceci signifie que l'on peut faire décroître le critère en augmentant $x_N = x_2$ tout en satisfaisant la contrainte d'égalité (c'est le sens du coût réduit). Mais ici on ne peut pas augmenter x_2 sans sortir de l'ensemble admissible (ce ne serait pas le cas si le sommet était non dégénéré).

2 *Déplacement le long d'une arête.* Si r a une composante strictement négative, disons $r_j < 0$, cela veut dire que l'on peut faire décroître le coût en augmentant la composante j de \hat{x}_N . On est donc tenté de chercher un nouveau point admissible en faisant un déplacement suivant une direction d , c'est-à-dire

$$x(\alpha) = \hat{x} + \alpha d,$$

telle que la composante non basique de d soit

$$d_N = e_N^j.$$

On a noté e^j le j -ième vecteur de base de $\mathbb{R}^{|N|}$. Pour que ce déplacement $d = (d_B, d_N)$ soit acceptable, il faut d'abord que l'on ait $Ax(\alpha) = b$, donc $Ad = 0$, ce qui détermine sa composante basique :

$$d_B = -A_B^{-1} A_N e_N^j.$$

Sur le choix de l'indice j . Remarquons que le coût décroît bien le long de d puisque l'on a

$$c^T d = r_j < 0.$$

Si r a plusieurs composantes strictement négatives, il semble donc raisonnable de choisir l'indice j parmi ceux donnant la composante de r la plus négative. C'est ce que l'on appelle la *règle du coût réduit le plus négatif*. Cette règle, d'aspect local (on s'intéresse à une dérivée), ne garantit cependant pas l'efficacité globale de l'algorithme qui est principalement liée au nombre total d'itérations, c'est-à-dire au nombre de

sommets visités, ce qui ne peut se déduire d'un calcul de dérivée. D'autres règles existent donc et les algorithmes du simplexe diffèrent essentiellement par l'heuristique adoptée.

Nous verrons que l'on peut maintenir la condition $x(\alpha) \geq 0$ pour un pas $\alpha > 0$, de manière à assurer l'admissibilité de $x(\alpha)$.

Il est intéressant d'observer que le déplacement porté par la direction d se fait le long d'une arête de \mathcal{F}_P .

Proposition 17.14 *Soient d défini comme ci-dessus et $\mathcal{A} := \{\hat{x} + \alpha d \in \mathcal{F}_P : \alpha \geq 0\}$. Alors soit \mathcal{A} est réduit au sommet $\{\hat{x}\}$, soit c'est une arête de \mathcal{F}_P .*

DÉMONSTRATION. Si $\hat{x} + \alpha d \notin \mathcal{F}_P$ pour tout $\alpha > 0$, on a $\mathcal{A} = \{\hat{x}\}$. Supposons à présent que pour un certain $\alpha > 0$, $\hat{x} + \alpha d \in \mathcal{F}_P$. Alors, \mathcal{A} est de dimension 1. Pour montrer que c'est une arête de \mathcal{F}_P (c'est-à-dire une face de dimension 1), il reste à montrer que \mathcal{A} est une face de \mathcal{F}_P . Soient $y, z \in \mathcal{F}_P$ tels que

$$\hat{x} + \alpha d = \beta y + (1-\beta)z, \tag{17.9}$$

avec $\beta \in]0, 1[$. Il s'agit de montrer que $y, z \in \mathcal{A}$.

En considérant les composantes non basiques de l'identité ci-dessus et en se rappelant que $\hat{x}_N = 0$ et $(y, z) \geq 0$, on voit que $y_N = y_j e_N^j = y_j d_N$ et $z_N = z_j e_N^j = z_j d_N$. D'autre part, comme $y, z \in \mathcal{F}_P$, on a $A(z - y) = 0$. On en déduit que $z_B - y_B = (z_j - y_j)d_B$ et donc

$$z - y = (z_j - y_j)d.$$

En utilisant à nouveau (17.9), on obtient alors

$$z = [\beta y + (1-\beta)z] + \beta(z - y) = \hat{x} + [\alpha + \beta(z_j - y_j)]d = \hat{x} + z_j d,$$

par la j -ième composante de (17.9). De même, $y = \hat{x} + y_j d$. Comme $y_j \geq 0$ et $z_j \geq 0$, on a $y, z \in \mathcal{A}$. □

3 *Cas du problème non borné ($d_B \geq 0$). Si $d_B \geq 0$, alors*

$$x(\alpha) \geq 0, \quad \forall \alpha > 0$$

et comme le coût décroît strictement

$$c^\top x(\alpha) = c^\top \hat{x} + \alpha c^\top d$$

le problème n'est pas borné.

4 *Nouvelle base d'indices ($d_B \not\geq 0$). Si $d_B \not\geq 0$, on ne peut plus prendre un pas arbitrairement grand. Pour que l'on ait $x(\alpha)_B \geq 0$, il faut que $\alpha \leq \hat{\alpha}$, où*

$$\hat{\alpha} = \min \left\{ -\frac{\hat{x}_i}{d_i} : i \in B, d_i < 0 \right\}. \tag{17.10}$$

Lorsque le sommet \hat{x} est dégénéré (il y a des $\hat{x}_i = 0$ pour $i \in B$), ce pas maximal $\hat{\alpha}$ peut être nul (voir ci-après). Soit k un indice donnant le min ci-dessus. Alors, $\hat{x}_k + \hat{\alpha}d_k = 0$ et on peut faire sortir l'indice k de la base d'indices B , et y faire entrer l'indice j . La nouvelle base d'indices s'écrit

$$\bar{B} \leftarrow (B \cup \{j\}) \setminus \{k\}.$$

Proposition 17.15 \bar{B} est une base d'indices.

DÉMONSTRATION. Il s'agit de montrer que $A_{\bar{B}}$ est inversible. Si ce n'est pas le cas, l'inversibilité de A_B implique que A^j est combinaison linéaire des $\{A^i : i \in B \setminus \{k\}\}$, c'est-à-dire

$$A_N e_N^j = A_B u, \quad \text{avec } u_k = 0.$$

On en déduit que $u = -d_B$. On a alors une contradiction, car d'une part $d_k = 0$ (car $u_k = 0$) et $d_k < 0$ (car k donne le min en (17.10)). \square

5 *Progrès ou stagnation.* Deux situations peuvent se présenter.

Si $\hat{\alpha} > 0$, le coût décroît strictement et on peut passer à l'itération suivante avec $\hat{x} + \hat{\alpha}d$ comme nouveau sommet et \bar{B} comme nouvelle base d'indices.

Si $\hat{\alpha} = 0$ (ceci ne peut se produire que si le sommet \hat{x} est dégénéré), il y a un changement de base d'indices sans changer de sommet (le pas $\hat{\alpha}$ est nul). Si on ne prend pas quelques précautions, l'algorithme peut cycliser (par exemple en faisant entrer k dans la base et en faisant sortir j à l'itération suivante). Sans entrer dans des détails très techniques, retenons qu'il y a des règles simples permettant d'éviter le cyclage, par exemple, en choisissant comme indices entrant et sortant de la base les indices les plus petits parmi les indices possibles (Bland, 1977).

17.4.2 Énoncé de l'algorithme

On peut résumer l'algorithme décrit ci-dessus comme suit.

Algorithme 17.16 (du simplexe – une itération)

On suppose au départ que l'on dispose d'un sommet $x \in \mathbb{R}^n$ de base B .

1. Calculer le multiplicateur $y \in \mathbb{R}^m$, solution du système linéaire

$$A_B^\top y = c_B$$

et en déduire le coût réduit

$$r = c_N - A_N^\top y.$$

2. Si $r \geq 0$, on s'arrête: x est solution du problème.

3. Soit j un indice tel que $r_j < 0$. On définit la direction de descente d de f par

$$d_N = e_N^j \quad \text{et} \quad d_B = -A_B^{-1} A_N e_N^j,$$

où e^j est le j -ième vecteur de base de $\mathbb{R}^{|N|}$.

4. Si $d_B \geq 0$, on s'arrête: le problème n'est pas borné.
5. Sinon, on définit le pas maximal $\hat{\alpha}$ jusqu'à la frontière du domaine :

$$\hat{\alpha} := \min \left\{ -\frac{x_i}{d_i} : i \in B, d_i < 0 \right\}.$$

Soit k donnant le minimum ci-dessus.

6. Supposons que $\hat{\alpha} = 0$. De deux choses l'une :
– Soit on n'a pas encore exploité tous les indices j tels que $r_j < 0$, alors on choisit un autre indice de ce type en respectant une règle d'anti-cyclage, et on retourne au point 3.
– Soit, pour tout indice j tel que $r_j < 0$, on a eu un pas $\hat{\alpha} = 0$, alors on s'arrête: x est solution du problème.
7. Ici $\hat{\alpha} > 0$. On calcule un nouveau sommet

$$x_+ = x + \hat{\alpha}d$$

et une nouvelle base d'indices associée

$$B_+ = (B \cup \{j\}) \setminus \{k\}.$$

On a $f(x_+) < f(x)$.

17.4.3 Démarrage de l'algorithme du simplexe

Utilisation d'une phase préliminaire (phase I)

Pour utiliser l'algorithme du simplexe, il faut disposer d'un itéré initial qui est un sommet de $X := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$. La stratégie suivante permet de trouver un tel sommet si $X \neq \emptyset$. Il s'agit de résoudre le problème d'optimisation linéaire suivant :

$$\begin{cases} \min \sum_{i=1}^m z_i \\ Ax + Dz = b \\ x \geq 0, \quad z \geq 0, \end{cases} \quad (17.11)$$

où D est la matrice diagonale définie par

$$D_{ii} = \begin{cases} 1 & \text{si } b_i \geq 0 \\ -1 & \text{si } b_i < 0. \end{cases}$$

Proposition 17.17 *Le point $(0, Db)$ est un sommet de l'ensemble admissible du problème (17.11), lequel a toujours une solution. Si ce problème est résolu par l'algorithme du simplexe en partant de ce point, il obtient pour solution un point (\hat{x}, \hat{z}) . Si $\hat{z} \neq 0$, le problème (P_L) n'est pas réalisable. Si $\hat{z} = 0$, \hat{x} est un sommet de l'ensemble admissible de (P_L) .*

DÉMONSTRATION. Il est clair que $(0, Db)$ est un sommet de l'ensemble admissible de (17.11), car les colonnes de D sont linéairement indépendantes (proposition 2.10). Comme (17.11) est borné ($\sum_{i=1}^m z_i \geq 0$), il a une solution (proposition 17.4). Si l'algorithme du simplexe est démarré en $(0, Db)$, il trouve une solution (\hat{x}, \hat{z}) .

Si $\hat{z} \neq 0$, la valeur optimale de (17.11) est non nulle et le problème (P_L) n'est pas réalisable (car si $x \in X$, $(x, 0)$ est admissible pour (17.11) et donne une valeur nulle au critère).

Si $\hat{z} = 0$, on a $A\hat{x} = b$ et $\hat{x} \geq 0$, donc $\hat{x} \in X$. D'autre part, $(\hat{x}, 0)$ étant un sommet de l'ensemble admissible de (17.11), les colonnes $\{A^j : \hat{x}_j > 0\}$ sont linéairement indépendantes; donc \hat{x} est un sommet de X . \square

En pratique, l'utilisation de l'algorithme du simplexe se fait le plus souvent en deux phases. Dans la première phase (qui n'est pas nécessaire si on connaît un sommet de X), on résout (17.11) par l'algorithme du simplexe à partir de $(0, Db)$ pour trouver un sommet \hat{x} de X (si $X \neq \emptyset$). Dans la seconde phase, on utilise à nouveau l'algorithme du simplexe pour résoudre (P_L) à partir du sommet \hat{x} ainsi trouvé.

Technique "big M"

17.4.4 Algorithme du simplexe dual**

17.4.5 Non polynomialité

Bien que l'algorithme du simplexe soit souvent efficace en pratique, ce n'est pas un algorithme polynomial. Klee et Minty [239; 1972] ont en effet construit un problème, dans lequel l'ensemble admissible est un cube de \mathbb{R}^n légèrement déformé, pour lequel l'algorithme du simplexe avec la règle du coût réduit le plus négatif visite les 2^n sommets de l'ensemble admissible. Des variantes de l'exemple de Klee et Minty existent pour la plupart des règles de pivotage autres que celle du coût réduit le plus négatif, voir [366] et ses références. Ce contre-exemple a stimulé la recherche d'algorithmes pouvant être polynomiaux en optimisation linéaire, un problème jugé suffisamment "simple" pour admettre un tel algorithme. Ceci a conduit aux algorithmes de points intérieurs, que nous verrons au chapitre 18.

17.5 Résolution de grands problèmes d'optimisation linéaire**

17.5.1 Décomposition de Dantzig-Wolfe

La situation est la suivante. On cherche à résoudre un problème de la forme

$$\begin{cases} \min c^\top x \\ Ax = a \\ Bx = b \\ x \geq 0. \end{cases}$$

Il s'agit d'un problème avec la même structure que le problème original (P_L), mais dans lequel la partition des contraintes en deux parties, définies par les matrices A et B , correspond à une sélection de contraintes "faciles" $Ax = a$ et "difficiles" $Bx = b$.

La référence originale est [102; 1961]. Voir [51; section 10.3.2] pour une présentation fondée sur la relaxation lagrangienne (ou décomposition par les prix) avec une résolution des problèmes de Lagrange par l'algorithme des plans-coupants [80, 237] dont l'instabilité bien connue rend compte du ralentissement de la convergence au cours des itérations que l'on observe dans le cas non linéaire (effet "tailing-off").

17.5.2 Décomposition de Benders

La référence originale est [30; 1962]. On cherche à résoudre un problème de la forme

$$\begin{cases} \min c^\top x + f^\top y \\ Ax + By \geq b \\ x \geq 0 \\ y \in Y, \end{cases}$$

dans lequel la matrice A a une structure simple, si bien que à y fixé, le problème est facile à résoudre.

L'approche peut s'étendre à des problèmes non linéaires; voir [51; section 10.5].

Lasdon [248] a montré que, dans le cas linéaire (comme ici), la décomposition de Benders est duale de la décomposition de Dantzig-Wolfe appliquée au dual.

17.5.3 Problèmes en nombres entiers

Branch-and-bound, branch-and-cut.

Notes

L'énoncé du problème dual d'un problème d'optimisation linéaire est attribué à J. Von Neumann. L'existence de solutions strictement complémentaires (proposition 17.8) a été montré par Goldman et Tucker (1956).

On pourra trouver d'autres résultats sur l'optimisation linéaire dans les références suivantes : Minoux [274; 1983]; Chvátal [82; 1983] a une approche très opérationnelle de l'optimisation linéaire et de l'algorithme du simplexe et s'intéresse à la modélisation de problèmes concrets sous forme de problèmes d'optimisation linéaire, notamment des problèmes de transport dans les réseaux; Schrijver [342; 1986]; Fletcher [139; 1987]; Ciarlet [84; 1988]; Goldfarb et Todd [173; 1989]; Nering et Tucker [289; 1993]; Saigal [340; 1995], que nous apprécions pour la précision de ses énoncés, toujours démontrés; Helgason et Kennington [209; 1995] discutent de l'utilisation de l'algorithme du simplexe dans divers problèmes d'optimisation linéaire rencontrés

dans les réseaux; le même commentaire s'applique à l'ouvrage de Bertsimas et Tsitsiklis [40; 1997], qui contient des chapitres sur les problèmes de grande taille avec leurs méthodes de décomposition, sur les problèmes d'OL dans les réseaux et sur les problèmes en nombres entiers; Vanderbei [379; 1997]; Martin [263; 1999] aborde, dans la partie IV de son ouvrage, les méthodes de décomposition pour résoudre les grands problèmes linéaires (décomposition de Benders, de Dantzig-Wolfe et lagrangienne) et traite, dans le chapitre 14, de l'optimisation dans les réseaux; Padberg [298; 1999].

Extrait de “*Éléments d’Optimisation Différentiable : Théorie et Algorithmes*”,
J. Ch. Gilbert (Jean-Charles.Gilbert@inria.fr), à paraître. Version du 23
janvier 2007.