Programmation linéaire et Optimisation

Didier Smets

Chapitre 1

Un problème d'optimisation linéaire en dimension 2

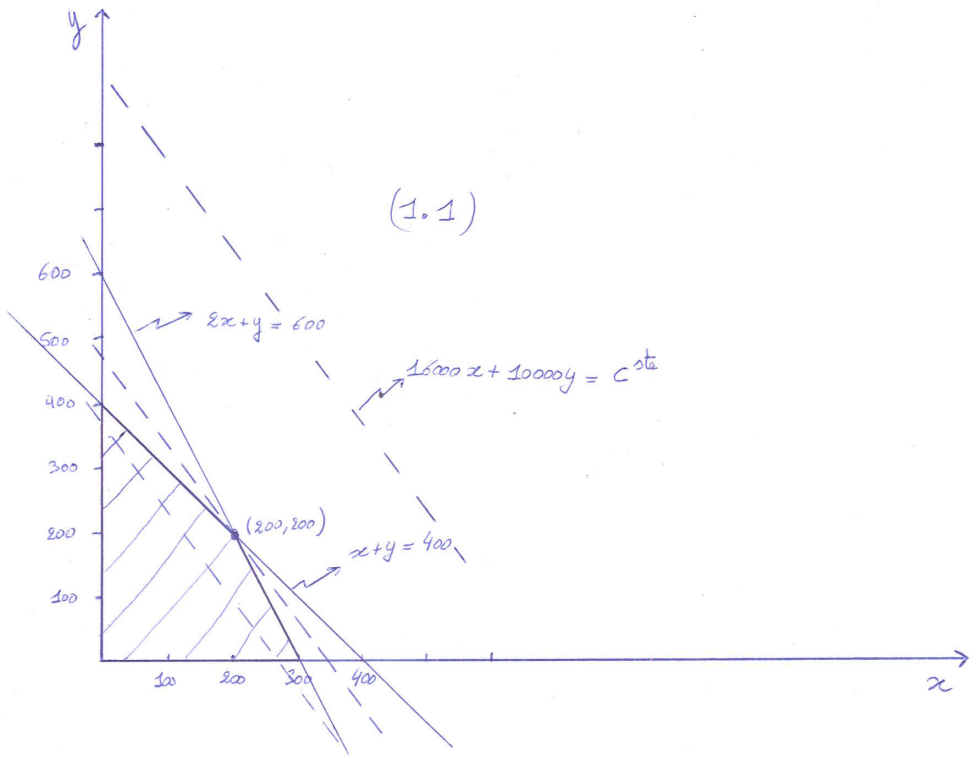
On considère le cas d'un fabricant d'automobiles qui propose deux modèles à la vente, des grosses voitures et des petites voitures. Les voitures de ce fabriquant sont tellement à la mode qu'il est certain de vendre tout ce qu'il parvient à produire, au moins au prix catalogue actuel de 16000 euros pour les grosses voitures, et 10000 euros pour les petites voitures. Son problème vient de l'approvisionnement limité en deux matières premières, le caoutchouc et l'acier. La construction d'une petite voiture nécessite l'emploi d'une unité de caoutchouc et d'une unité d'acier, tandis que celle d'une grosse voiture nécessite une unité de caoutchouc mais deux unités d'acier. Sachant que son stock de caoutchouc est de 400 unités et son stock d'acier de 600 unités, combien doit-il produire de petites et de grosses voitures au moyen de ces stocks afin de maximiser son chiffre d'affaire ?

Nous appellerons x le nombre de grosses voitures produites, y le nombre de petites voitures produites, et z le chiffre d'affaire résultant. Le problème se traduit alors sous la forme

$$\begin{array}{ll} \text{maximiser} & z = 16000x + 10000y \\ \text{sous les contraintes} & x + y \leq 400 \\ & 2x + y \leq 600 \\ & x \geq 0,\ y \geq 0. \end{array} \tag{1.1}$$

1.1 Solution graphique

Un tel système, parce qu'il ne fait intervenir que deux variables, peu se résoudre assez facilement de manière graphique, en hachurant la zone correspondant aux contraintes, et en traçant les lignes de niveaux (ici des lignes parallèles) de la fonction à maximiser (cfr. graphique ci-dessous). On obtient ainsi la solution optimale $x = 200$ et $y = 200$, qui correspond à $z = 5200000$. Elle est unique dans ce cas précis, et correspond à un "sommet" de la zone de contraintes.



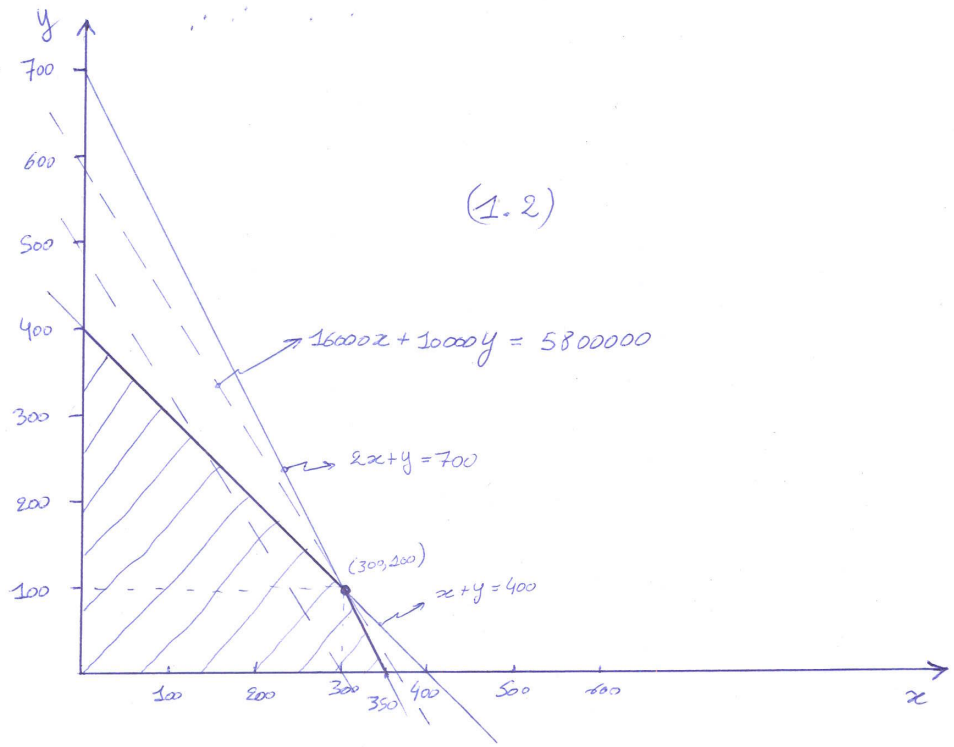
1.2 Sensibilité à la variation des stocks

Observons comment la solution du problème évolue lorsqu'on modifie certaines données de départ, par exemple une augmentation du stock de caoutchouc ou du stock d'acier.

Imaginons que le stock d'acier soit de 700 au lieu de 600, le nouveau problème s'écrit

$$\begin{array}{ll} \text{maximiser} & z = 16000x + 10000y \\ \text{sous les contraintes} & x + y \leq 400 \\ & 2x + y \leq 700 \\ & x \geq 0,\ y \geq 0. \end{array} \tag{1.2}$$

Toujours de manière graphique, on s'aperçoit que la solution optimale est maintenant donnée par $x = 300$ et $y = 100$, ce qui correspond à $z = 5800000$. Autrement dit, une augmentation de 100 unités d'acier à un impact de 600000 euros sur le chiffre d'affaire. On dira alors que le *prix marginal* de l'unité d'acier est de 6000 euros.



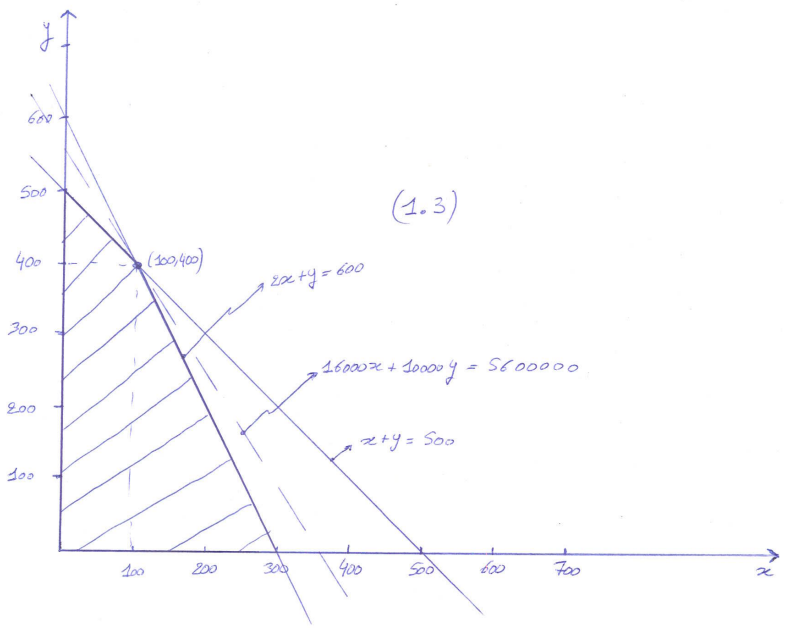
Si le stock d'acier passe à 800, la solution optimale devient $x = 400$ et $y = 0$ et le chiffre d'affaire $z = 6400000$. Augmenter le stock d'acier au-delà de 800, sans changer le stock de caoutchouc, n'a plus aucune influence sur la solution optimale, car y est contraint à rester positif.

Imaginons maintenant que le stock d'acier reste fixé à 600 mais que le stock de caoutchouc passe de 400 à 500. Le nouveau problème s'écrit

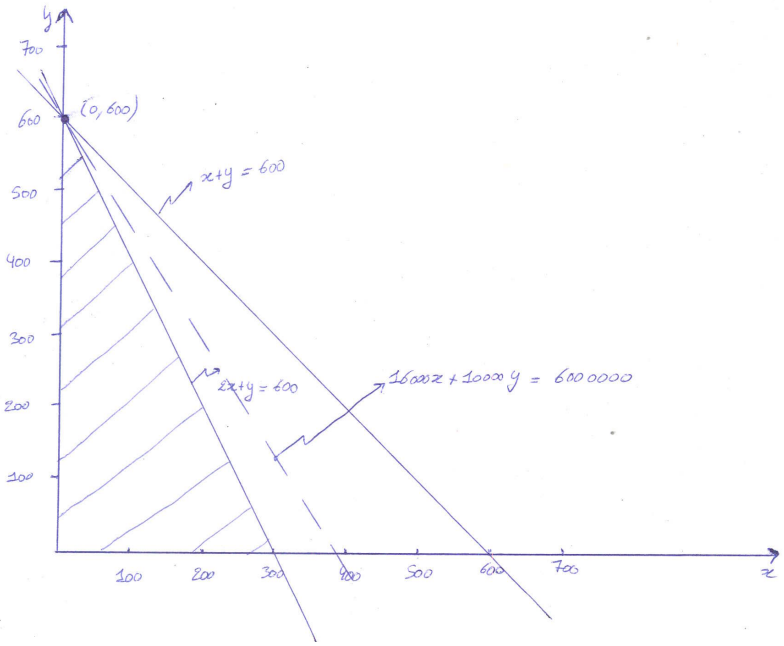
$$\begin{array}{ll} \text{maximiser} & z = 16000x + 10000y \\ \text{sous les contraintes} & x + y \leq 500 \\ & 2x + y \leq 600 \\ & x \geq 0,\ y \geq 0. \end{array} \tag{1.3}$$

Toujours de manière graphique, on s'aperçoit que la solution optimale est maintenant donnée par $x = 100$ et $y = 400$, ce qui correspond à $z = 5600000$. Autrement dit, une augmentation de 100 unités de caoutchouc à un impact de 400000 euros sur le chiffre

d'affaire. On dira alors que le *prix marginal* de l'unité de caoutchouc est de 4000 euros.



Si le stock de caoutchouc passe à 600, la solution optimale devient $x = 0$ et $y = 600$ et le chiffre d'affaire $z = 6000000$. Augmenter le stock de caoutchouc au-delà de 600, sans changer le stock d'acier, n'a plus aucune influence sur la solution optimale, car x est contraint à rester positif.

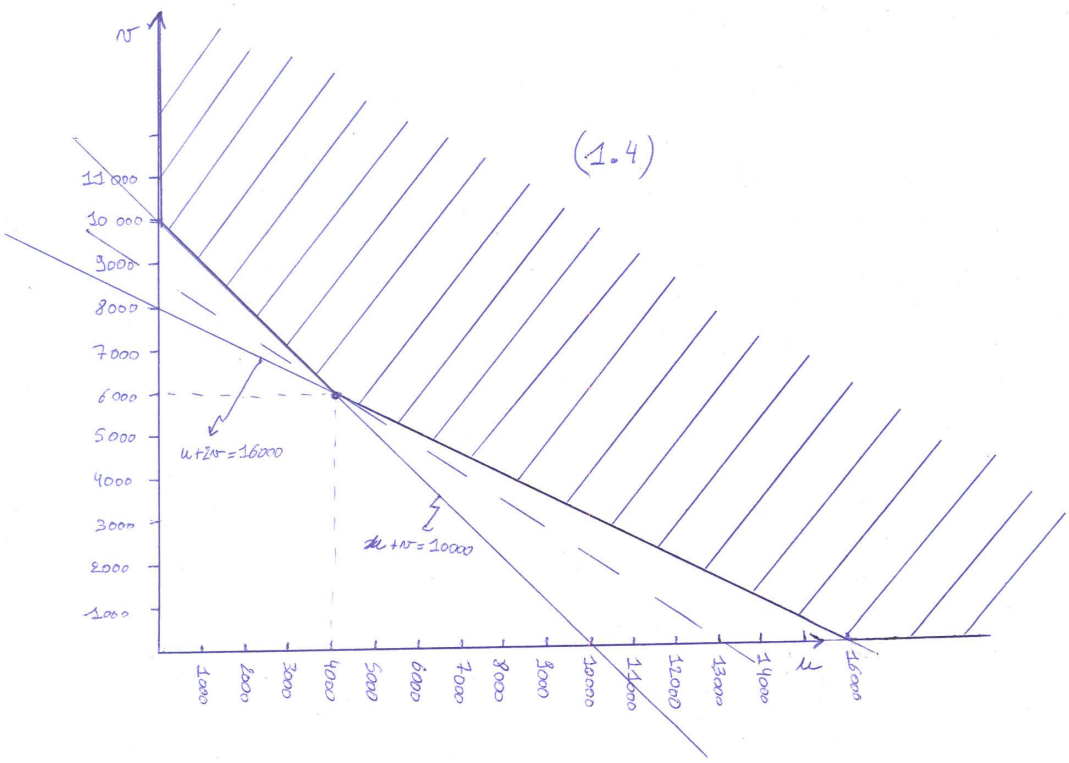


1.3 Le problème dual du concurrent

Supposons maintenant que le fabricant d'automobile possède un concurrent qui, pour honorer des commandes en trop grand nombre, se propose de lui racheter tous ses stocks. Ce dernier doit faire une offre de prix (la même, disons u) pour chaque unité de caoutchouc et une offre de prix (disons v) pour chaque unité d'acier. Pour que l'offre soit acceptée, il faut que le prix payé par le concurrent soit au moins égal à ce que le fabriquant pourrait en tirer en produisant des voitures. Le problème du concurrent s'écrit ainsi

$$\begin{array}{ll} \text{minimiser} & p = 400u + 600v \\ \text{sous les contraintes} & u + v \geq 10000 \\ & u + 2v \geq 16000 \\ & u \geq 0,\ v \geq 0. \end{array} \tag{1.4}$$

Une analyse graphique fournit la solution optimale $u = 4000$ et $v = 6000$, ce qui correspond à un prix global $p = 5200000$. On remarque (nous verrons par la suite que ce n'est pas un hasard) que la solution optimale du problème du concurrent (on parlera de *problème dual*, par opposition au *problème primal* du fabriquant) correspond aux prix marginaux du problème du fabricant, et que le prix minimal que puisse proposer le concurrent est égal au chiffre d'affaire maximal du fabricant.



Chapitre 2

Un problème d'optimisation linéaire en dimension supérieure

Dans ce chapitre, nous allons décrire un problème de transport optimal assimilable à un problème d'optimisation linéaire en dimension 6. De ce fait, il ne sera plus possible de le résoudre au moyen de la méthode graphique du chapitre précédent.

Notre fabricant d'automobiles possède trois chaînes de montage M_1, M_2 et M_3, tandis que son stock d'acier provient de deux aciéries A_1 et A_2. Les coûts de transport d'une unité d'acier d'une aciérie vers une usine de montage sont donnés par le tableau suivant :

	M_1	M_2	M_3
A_1	9	16	28
A_2	14	29	19

Les capacités de production des chaînes de montage diffèrent, ainsi que les capacités de production des aciéries, et sont données par les deux tableaux suivants :

A_1	206
A_2	394

M_1	142
M_2	266
M_3	192

Il s'agit donc pour le fabricant de déterminer le plan de transport des unités d'acier produites vers les chaînes de montage afin de minimiser le coût total de transport. Pour $i = 1, 2$ et $j = 1, 2, 3$, notons x_{ij} le nombre d'unités d'acier acheminées depuis l'aciérie A_i vers la chaîne de montage M_j. Le problème de transport optimal peut alors s'écrire :

$$
\begin{array}{lrrrrrrrrr}
\text{minimiser} & t = & 9x_{11} & +16x_{12} & +28x_{13} & +14x_{21} & +29x_{22} & +19x_{23} & & \\
\text{sous les contraintes} & & x_{11} & +x_{12} & +x_{13} & & & & \leq & 206, \\
 & & & & & x_{21} & +x_{22} & +x_{23} & \leq & 394, \\
 & & x_{11} & & & +x_{21} & & & \geq & 142, \\
 & & & x_{12} & & & +x_{22} & & \geq & 266, \\
 & & & & x_{13} & & & +x_{23} & \geq & 192, \\
 & & x_{11}, & x_{12}, & x_{13}, & x_{21}, & x_{22}, & x_{23} & \geq & 0.
\end{array}
$$

Nous verrons par la suite qu'il est possible de traiter un tel problème de manière systématique, par le biais d'une réduction à une forme standard suivie d'un algorithme

qui porte le nom de méthode du simplexe. Toutefois, dans ce cas précis, cela nous mènerait à des manipulations trop fastidieuses pour être réalisées sans l'aide d'un ordinateur. A sa place, nous allons procéder à un certain nombre de remarques ad hoc qui vont nous permettre de poursuivre les calculs à la main.

La remarque principale ici est que dans la mesure où la somme des productions des aciéries ($206 + 394 = 600$) est égale à la somme des capacités de production des trois chaînes de montage ($142 + 266 + 192 = 600$), chacune des 5 premières inégalités dans le problème d'optimisation ci-dessus doit nécessairement être une égalité. Si on omet momentanément de s'occuper des contraintes $x_{ij} \geq 0$ ($i = 1, 2$, $j = 1, 2, 3$), les contraintes restantes se réduisent à un système de 5 équations à 6 inconnues, que nous pouvons tenter de résoudre par la méthode du pivot de Gauss (cfr. Algèbre linéaire L1).

On récrit le sous-système des contraintes d'égalité sous la forme (on choisit l'ordre des équation afin de faciliter le pivot de Gauss) :

$$\begin{array}{rrrrrrrrr} x_{11} & & & +x_{21} & & & = & 142, \\ & x_{12} & & & +x_{22} & & = & 266, \\ & & x_{13} & & & +x_{23} & = & 192, \\ x_{11} & +x_{12} & +x_{13} & & & & = & 206, \\ & & & x_{21} & +x_{22} & +x_{23} & = & 394. \end{array}$$

On échelonne ensuite (méthode du tableau) :

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & 142 \\ 0 & 1 & 0 & 0 & 1 & 0 & 266 \\ 0 & 0 & 1 & 0 & 0 & 1 & 192 \\ 1 & 1 & 1 & 0 & 0 & 0 & 206 \\ 0 & 0 & 0 & 1 & 1 & 1 & 394 \end{array}\right) \rightarrow \left(\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & 142 \\ 0 & 1 & 0 & 0 & 1 & 0 & 266 \\ 0 & 0 & 1 & 0 & 0 & 1 & 192 \\ 0 & 1 & 1 & -1 & 0 & 0 & 64 \\ 0 & 0 & 0 & 1 & 1 & 1 & 394 \end{array}\right) \rightarrow$$

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & 142 \\ 0 & 1 & 0 & 0 & 1 & 0 & 266 \\ 0 & 0 & 1 & 0 & 0 & 1 & 192 \\ 0 & 0 & 1 & -1 & -1 & 0 & -202 \\ 0 & 0 & 0 & 1 & 1 & 1 & 394 \end{array}\right) \rightarrow \left(\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & 142 \\ 0 & 1 & 0 & 0 & 1 & 0 & 266 \\ 0 & 0 & 1 & 0 & 0 & 1 & 192 \\ 0 & 0 & 0 & -1 & -1 & -1 & -394 \\ 0 & 0 & 0 & 1 & 1 & 1 & 394 \end{array}\right) \rightarrow$$

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & 142 \\ 0 & 1 & 0 & 0 & 1 & 0 & 266 \\ 0 & 0 & 1 & 0 & 0 & 1 & 192 \\ 0 & 0 & 0 & 1 & 1 & 1 & 394 \end{array}\right) \rightarrow \left(\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & -1 & -1 & -252 \\ 0 & 1 & 0 & 0 & 1 & 0 & 266 \\ 0 & 0 & 1 & 0 & 0 & 1 & 192 \\ 0 & 0 & 0 & 1 & 1 & 1 & 394 \end{array}\right).$$

La forme échelonnée laisse apparaître les variables x_{22} et x_{23} comme libres, desquelles on déduit

$$\begin{array}{rcl} x_{21} & = & 394 - x_{22} - x_{23}, \\ x_{13} & = & 192 - x_{23}, \\ x_{12} & = & 266 - x_{22}, \\ x_{11} & = & -252 + x_{22} + x_{23}. \end{array} \tag{2.1}$$

On exprime ensuite le coût t uniquement en termes des variables libres x_{22} et x_{23} :

$$\begin{aligned} t &= 9(-252 + x_{22} + x_{23}) + 16(266 - x_{22}) + 28(192 - x_{23}) \\ &\quad + 14(394 - x_{22} - x_{23}) + 29x_{22} + 19x_{23} \\ &= 8x_{22} - 14x_{23} + 12880. \end{aligned} \tag{2.2}$$

Afin de minimiser t il est donc opportun de choisir x_{23} le plus grand possible, et x_{22} le plus petit possible. C'est à ce niveau qu'il nous est nécessaire de faire réapparaître les contraintes $x_{ij} \geq 0$ ($i = 1, 2$, $j = 1, 2, 3$), sans lesquelles t pourrait être rendu aussi négatif que souhaité. En examinant les équation (2.1), on se convainc assez rapidement que le meilleur choix est obtenu en prenant $x_{23} = 192$ (afin de satisfaire mais saturer la contrainte $x_{13} \geq 0$), et ensuite $x_{22} = 60$ (afin de satisfaire mais saturer la contrainte $x_{11} \geq 0$). On propose alors la solution suivante

$$\begin{array}{ll} x_{11} & = 0, \\ x_{12} & = 206, \\ x_{13} & = 0, \\ x_{21} & = 142, \\ x_{22} & = 60, \\ x_{23} & = 192, \end{array} \tag{2.3}$$

comme candidat à être le transport optimal. Pour vérifier notre intuition, on choisit d'exprimer cette fois le système (2.1) uniquement en termes des variables x_{11} et x_{13} (on comprendra ce choix dans un instant), ce qui donne (on n'a en réalité besoin que d'exprimer x_{22} et x_{33} puisqu'elles seules interviennent dans l'expression de t dans (2.2)) :

$$\begin{array}{lcl} x_{22} & = & 60 + x_{11} + x_{13}, \\ x_{23} & = & 192 - x_{13}. \end{array} \tag{2.4}$$

On obtient ainsi l'expression

$$\begin{aligned} t &= 8x_{22} - 14x_{23} + 12880 \\ &= 8(60 + x_{11} + x_{13}) - 14(192 - x_{13}) + 12880 \\ &= 8x_{11} + 22x_{13} + 10672. \end{aligned} \tag{2.5}$$

Comme $x_{11} \geq 0$ et $x_{13} \geq 0$ par contrainte, on a nécessairement $t \geq 10672$ quel que soit le choix de x_{ij} ($i = 1, 2$, $j = 1, 2, 3$) satisfaisant l'ensemble des contraintes. Par ailleurs, le choix proposé en (2.3) fournit $t = 10672$ et satisfait à l'ensemble des contraintes. Il s'agit donc effectivement de la solution optimale.

Pour terminer cet exemple par une synthèse, observons que nous sommes parvenus à récrire le problème d'optimisation initial sous la forme d'un système linéaire augmenté de contraintes de positivité de toutes les variables. Nous avons ensuite déterminé le rang du système linéaire en question et exprimé de diverses manières possibles (deux en l'occurrence) la fonction à optimiser (ici t) en termes de variables libres pour ce système linéaire. Nous nous sommes arrêtés lorsque les coefficients des variables libres dans l'expression de la fonction à optimiser furent tous positifs ou nuls, et avons conclu que les égaler à zéro fournissait une solution assurément optimale.

Dans le chapitre qui suit, nous reprenons cette démarche de manière un peu plus systématique sur un exemple initialement en dimension 3.

Chapitre 3

Méthode du simplexe : un aperçu par l'exemple

Considérons le problème d'optimisation linéaire :

$$\begin{array}{lrrrrrrr} \text{maximiser} & z = & 5x_1 & +4x_2 & +3x_3 & & & \\ \text{sous les contraintes} & & 2x_1 & +3x_2 & +x_3 & \leq & 5, & \\ & & 4x_1 & +x_2 & +2x_3 & \leq & 11, & \\ & & 3x_1 & +4x_2 & +2x_3 & \leq & 8, & \\ & & x_1, & x_2, & x_3 & \geq & 0. & \end{array} \tag{3.1}$$

Afin de se ramener à un système d'équations plutôt que d'inéquations, on introduit les *variables d'écart* x_4, x_5, x_6 et l'on écrit le problème ci-dessus sous la forme

$$\begin{array}{rcrrrr} x_4 & = & 5 & -2x_1 & -3x_2 & -x_3, \\ x_5 & = & 11 & -4x_1 & -x_2 & -2x_3, \\ x_6 & = & 8 & -3x_1 & -4x_2 & -2x_3, \\ z & = & & 5x_1 & +4x_2 & +3x_3, \end{array} \tag{3.2}$$

avec pour but de maximiser z sous les contraintes additionnelles $x_i \geq 0$, $(i = 1, \cdots, 6)$. Il est aisé (et recommandé) de vérifier que si $(x_1, x_2, x_3, x_4, x_5, x_6)$ est une solution optimale de ce dernier problème, alors les (x_1, x_2, x_3) correspondants constituent une solution optimale du problème (3.1). Inversement, si (x_1, x_2, x_3) est une solution optimale de (3.1), alors $(x_1, x_2, x_3, 5 - 2x_1 - 3x_2 - x_3, 11 - 4x_1 - x_2 - 2x_3, 8 - 3x_1 - 4x_2 - 2x_3)$ constitue une solution optimale de (3.2).

Le système (3.2) possède la solution (non optimale) $(0, 0, 0, 5, 11, 8)$ (l'usage est d'appeler *solution réalisable* tout choix de variables satisfaisant à l'ensemble des contraintes (cfr. le chapitre suivant)).

On observe que dans l'expression $z = 5x_1 + 4x_2 + 3x_3$, une augmentation de x_1 entraîne une augmentation de z. L'idée première est alors d'augmenter x_1 autant que possible (sans modifier ni x_2 ni x_3) tant qu'aucune des variables d'écart x_4, x_5 ou x_6 ne devient négative. Le choix maximal est donc $x_1 = \min(5/2, 11/4, 8/3) = 5/2$, lorsque x_4 devient nulle, et qui fait passer à la solution réalisable $(5/2, 0, 0, 0, 3, 1/2)$.

On récrit le système (3.2) en exprimant cette fois (x_1, x_5, x_6) (ainsi que z) en termes

de (x_2, x_3, x_4), au moyen de l'équation

$$x_1 = \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4.$$

Ceci donne, après substitutions :

$$\begin{array}{rcrrrr} x_1 & = & \frac{5}{2} & -\frac{3}{2}x_2 & -\frac{1}{2}x_3 & -\frac{1}{2}x_4, \\ x_5 & = & 1 & +5x_2 & & +2x_4, \\ x_6 & = & \frac{1}{2} & +\frac{1}{2}x_2 & -\frac{1}{2}x_3 & +\frac{3}{2}x_4, \\ z & = & \frac{25}{2} & -\frac{7}{2}x_2 & +\frac{1}{2}x_3 & -\frac{5}{2}x_4. \end{array} \tag{3.3}$$

Cette fois, on observe que dans l'expression $z = 25/2 - 7/2x_2 + 1/2x_3 - 5/2x_4$, une augmentation de x_3 (c'est ici le seul choix possible) entraîne une augmentation de z. A nouveau, on augmente donc x_3 autant que possible (sans modifier ni x_2 ni x_4) tant qu'aucune des variables (dites *variables en bases* (cfr. Chapitre 5)) x_1, x_5 ou x_6 ne devient négative. Le choix maximal est donc $x_3 = \min((5/2)/(1/2), (1/2)/(1/2)) = 1$, lorsque x_6 devient nulle, et qui fait passer à la solution réalisable $(2, 0, 1, 0, 1, 0)$.

On récrit le système (3.3) en exprimant cette fois (x_1, x_3, x_5) (ainsi que z) en termes de (x_2, x_4, x_6), au moyen de l'équation

$$x_3 = 1 + x_2 + 3x_4 - 2x_6.$$

Ceci donne, après substitutions :

$$\begin{array}{rcrrrr} x_1 & = & 2 & -2x_2 & -2x_4 & +x_6, \\ x_3 & = & 1 & +x_2 & +3x_4 & -2x_6, \\ x_5 & = & 1 & +5x_2 & +2x_4 & , \\ z & = & 13 & -3x_2 & -x_4 & -x_6. \end{array} \tag{3.4}$$

Puisque les coefficients de x_2, x_4 et x_6 intervenant dans l'expression de z ci-dessus sont tous négatifs ou nuls, on déduit que la solution réalisable

$$\begin{array}{rl} x_1 & = 2, \\ x_2 & = 0, \\ x_3 & = 1, \\ x_4 & = 0, \\ x_5 & = 1, \\ x_6 & = 0, \end{array} \tag{3.5}$$

est une solution optimale, pour laquelle $z = 13$.

Avant de formaliser l'algorithme du simplexe, et d'en découvrir les bases théoriques, voyons une deuxième méthode pour l'aborder et qui consiste à placer les calculs en tableau (toutes les variables se retrouvant du même côté du signe d'égalité) plutôt que sous forme dictionnaire comme ci-dessus. L'avantage de cette deuxième façon de présenter les choses (mais qui est bien sûr équivalente à la première) et qu'elle se rapproche plus de la méthode bien connue du pivot de Gauss.

Considérons donc maintenant le problème d'optimisation linéaire

$$\begin{array}{lrrrrrrr}
\text{maximiser} & z = & x_1 & +5x_2 & +x_3 & & & \\
\text{sous les contraintes} & & x_1 & +3x_2 & +x_3 & \leq & 3, \\
& & -x_1 & & +3x_3 & \leq & 2, \\
& & 2x_1 & +4x_2 & -x_3 & \leq & 4, \\
& & x_1 & +3x_2 & -x_3 & \leq & 2, \\
& & x_1, & x_2, & x_3 & \geq & 0.
\end{array} \tag{3.6}$$

On introduit les variables d'écart x_4, x_5, x_6, x_7 (il y a une contrainte supplémentaire par rapport au problème précédent) et on récrit le problème sous la forme

$$\begin{array}{lrrrrrrrrrr}
\text{maximiser} & z = & x_1 & +5x_2 & +x_3 & & & & & & \\
\text{sous les contraintes} & & x_1 & +3x_2 & +x_3 & +x_4 & & & & = & 3, \\
& & -x_1 & & +3x_3 & & +x_5 & & & = & 2, \\
& & 2x_1 & +4x_2 & -x_3 & & & +x_6 & & = & 4, \\
& & x_1 & +3x_2 & -x_3 & & & & +x_7 & = & 2, \\
& & x_1, & x_2, & x_3 & x_4, & x_5, & x_6, & x_7 & \geq & 0.
\end{array} \tag{3.7}$$

On adopte alors la notation sous forme de tableau :

$$\begin{array}{rrrrrrrcr}
1 & 3 & 1 & 1 & 0 & 0 & 0 & | & 3 \\
-1 & 0 & 3 & 0 & 1 & 0 & 0 & | & 2 \\
2 & 4 & -1 & 0 & 0 & 1 & 0 & | & 4 \\
1 & 3 & -1 & 0 & 0 & 0 & 1 & | & 2 \\
\hline
1 & 5 & 1 & 0 & 0 & 0 & 0 & | & 0
\end{array}$$

La dernière ligne du tableau correspond à un expression possible de z comme fonction affine des variables $x_1, \cdots, x_7$, l'opposé du terme constant se trouvant en bas à droite du tableau. On part de la solution réalisable $(0, 0, 0, 3, 2, 4, 2)$ et puisque le terme en x_1 dans la dernière ligne du tableau (=1) est positif, on va augmenter x_1 (sans modifier ni x_2 ni x_3) jusqu'à ce que l'une des variables x_4, x_5, x_6 ou x_7 devienne nulle. Ceci se produit pour $x_1 = 2$ pour lequel à la fois x_6 et x_7 deviennent nuls. On choisit dont de faire rentrer x_1 en base, et de faire sortir (par exemple) x_6. Cela donne

$$\begin{array}{rrrrrrrcr}
1 & 3 & 1 & 1 & 0 & 0 & 0 & | & 3 \\
-1 & 0 & 3 & 0 & 1 & 0 & 0 & | & 2 \\
2 & 4 & -1 & 0 & 0 & 1 & 0 & | & 4 \\
1 & 3 & -1 & 0 & 0 & 0 & 1 & | & 2 \\
\hline
1 & 5 & 1 & 0 & 0 & 0 & 0 & | & 0
\end{array}
\quad \rightarrow \quad
\begin{array}{rrrrrrrcr}
1 & 3 & 1 & 1 & 0 & 0 & 0 & | & 3 \\
-1 & 0 & 3 & 0 & 1 & 0 & 0 & | & 2 \\
1 & 2 & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & | & 2 \\
1 & 3 & -1 & 0 & 0 & 0 & 1 & | & 2 \\
\hline
1 & 5 & 1 & 0 & 0 & 0 & 0 & | & 0
\end{array}
\quad \rightarrow$$

$$\begin{array}{rrrrrrrcr}
0 & 1 & \frac{3}{2} & 1 & 0 & -\frac{1}{2} & 0 & | & 1 \\
0 & 2 & \frac{5}{2} & 0 & 1 & \frac{1}{2} & 0 & | & 4 \\
1 & 2 & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & | & 2 \\
0 & 1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 & | & 0 \\
\hline
0 & 3 & \frac{3}{2} & 0 & 0 & -\frac{1}{2} & 0 & | & -2
\end{array}$$

et fournit une solution réalisable pour laquelle $x_1 = 2$, $x_4 = 1$, $x_5 = 4$, $x_7 = 0$ (et les variables hors base sont toujours nulles : $x_2 = x_3 = x_6 = 0$). Puisque le coefficient de x_2

dans la nouvelle expression de z est positive, on fait ensuite rentrer x_2 en base, et on doit faire sortir x_7 sans pouvoir en rien augmenter x_2 ! Cela donne

$$\begin{array}{ccccccc|c}
0 & 1 & \frac{3}{2} & 1 & 0 & -\frac{1}{2} & 0 & 1 \\
0 & 2 & \frac{5}{2} & 0 & 1 & \frac{1}{2} & 0 & 4 \\
1 & 2 & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 2 \\
0 & 1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 & 0 \\
\hline
0 & 3 & \frac{3}{2} & 0 & 0 & -\frac{1}{2} & 0 & -2
\end{array}
\rightarrow
\begin{array}{ccccccc|c}
0 & 0 & 2 & 1 & 0 & 0 & -1 & 1 \\
0 & 0 & \frac{7}{2} & 0 & 1 & \frac{3}{2} & -2 & 4 \\
1 & 0 & \frac{1}{2} & 0 & 0 & \frac{3}{2} & -2 & 2 \\
0 & 1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 & 0 \\
\hline
0 & 0 & 3 & 0 & 0 & 1 & -3 & -2
\end{array}$$

et fournit la solution réalisable $(2, 0, 0, 1, 4, 0, 0)$. On fait ensuite rentrer x_3 en base (son coefficient dans l'expression de z vaut maintenant 3) et on fait sortir x_4 (qui s'annule lorsque $x_3 = 1/2$). Cela donne

$$\begin{array}{ccccccc|c}
0 & 0 & 2 & 1 & 0 & 0 & -1 & 1 \\
0 & 0 & \frac{7}{2} & 0 & 1 & \frac{3}{2} & -2 & 4 \\
1 & 0 & \frac{1}{2} & 0 & 0 & \frac{3}{2} & -2 & 2 \\
0 & 1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 & 0 \\
\hline
0 & 0 & 3 & 0 & 0 & 1 & -3 & -2
\end{array}
\rightarrow
\begin{array}{ccccccc|c}
0 & 0 & 1 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & \frac{1}{2} \\
0 & 0 & \frac{7}{2} & 0 & 1 & \frac{3}{2} & -2 & 4 \\
1 & 0 & \frac{1}{2} & 0 & 0 & \frac{3}{2} & -2 & 2 \\
0 & 1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 & 0 \\
\hline
0 & 0 & 3 & 0 & 0 & 1 & -3 & -2
\end{array}
\rightarrow$$

$$\begin{array}{ccccccc|c}
0 & 0 & 1 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & -\frac{7}{4} & 1 & \frac{3}{2} & -\frac{1}{4} & \frac{9}{4} \\
1 & 0 & 0 & 0 & 0 & \frac{3}{2} & -\frac{7}{4} & \frac{7}{4} \\
0 & 1 & 0 & \frac{1}{4} & 0 & -\frac{1}{2} & \frac{3}{4} & \frac{1}{4} \\
\hline
0 & 0 & 0 & -\frac{3}{2} & 0 & 1 & -\frac{3}{2} & -\frac{7}{2}
\end{array}$$

et fournit la solution réalisable $(\frac{7}{4}, \frac{1}{4}, \frac{1}{2}, 0, 0, 0, 0)$. On fait ensuite rentrer x_6 et sortir x_1 (qui s'annule en premier lorsque lorsque $x_6 = \frac{7}{3}$). Cela donne

$$\begin{array}{ccccccc|c}
0 & 0 & 1 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & -\frac{7}{4} & 1 & \frac{3}{2} & -\frac{1}{4} & \frac{9}{4} \\
1 & 0 & 0 & 0 & 0 & \frac{3}{2} & -\frac{7}{4} & \frac{7}{4} \\
0 & 1 & 0 & \frac{1}{4} & 0 & -\frac{1}{2} & \frac{3}{4} & \frac{1}{4} \\
\hline
0 & 0 & 0 & -\frac{3}{2} & 0 & 1 & -\frac{3}{2} & -\frac{7}{2}
\end{array}
\rightarrow
\begin{array}{ccccccc|c}
0 & 0 & 1 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & -\frac{7}{4} & 1 & \frac{3}{2} & -\frac{1}{4} & \frac{9}{4} \\
\frac{2}{3} & 0 & 0 & 0 & 0 & 1 & -\frac{7}{6} & \frac{7}{6} \\
0 & 1 & 0 & \frac{1}{4} & 0 & -\frac{1}{2} & \frac{3}{4} & \frac{1}{4} \\
\hline
0 & 0 & 0 & -\frac{3}{2} & 0 & 1 & -\frac{3}{2} & -\frac{7}{2}
\end{array}
\rightarrow$$

$$\begin{array}{ccccccc|c}
0 & 0 & 1 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & \frac{1}{2} \\
-1 & 0 & 0 & -\frac{7}{4} & 1 & 0 & \frac{3}{2} & \frac{1}{2} \\
\frac{2}{3} & 0 & 0 & 0 & 0 & 1 & -\frac{7}{6} & \frac{7}{6} \\
\frac{1}{3} & 1 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{6} & \frac{5}{6} \\
\hline
-\frac{2}{3} & 0 & 0 & -\frac{3}{2} & 0 & 0 & -\frac{1}{3} & -\frac{14}{3}
\end{array}$$

et fournit finalement la solution optimale $(0, \frac{5}{6}, \frac{1}{2}, 0, \frac{1}{2}, \frac{7}{6}, 0)$ pour laquelle $z = \frac{14}{3}$.

Chapitre 4

Formes générale, canonique et standard d'un problème d'optimisation linéaire

Dans ce chapitre, nous définissons la forme générale d'un problème d'optimisation linéaire, ainsi que la forme canonique et la forme standard. Nous montrons également qu'un problème sous forme générale peut être transformé d'abord en un problème équivalent sous forme canonique, puis enfin sous un problème équivalent sous forme standard. Dans les chapitres qui suivront, nous nous restreindrons donc à fournir un algorithme de résolution pour les problèmes sous forme standard.

Définition 4.1. *On appelle problème d'optimisation linéaire sous forme générale un problème de la forme*

$$\begin{array}{ll} \textit{maximiser} & F(X) \\ \textit{sous les contraintes} & X \in \mathbb{R}^Q, G_1(X) \leq 0, \cdots, G_P(X) \leq 0, \end{array} \tag{4.1}$$

où $P, Q \in \mathbb{N}_*$, $F : \mathbb{R}^Q \to \mathbb{R}$ est une forme linéaire sur $\mathbb{R}^Q$ et $G_1, \cdots, G_P$ sont des applications affines définies sur $\mathbb{R}^Q$ et à valeurs réelles. On dit que la fonction F est la **fonction objectif** *et que les fonctions $G_1, \cdots, G_P$ sont les* **contraintes**.

Remarque 4.2. *On pourrait bien sûr traiter de manière équivalente les problèmes de minimisation. Il n'y a toutefois aucune perte de généralité à ne traiter que les problèmes de maximisation. De fait, minimiser une fonctionnelle F revient à maximiser la fonctionnelle $-F$. Dans le même esprit, on pourrait considérer des contraintes de la forme $G_i(X) \geq 0$ ou même $G_i(X) = 0$. Dans le premier cas il suffit alors de récrire la contrainte sous la forme $-G_i(X) \leq 0$, et dans le second de la dédoubler en deux contraintes : $G_j(X) \leq 0$ et $-G_i(X) \leq 0$.*

Définition 4.3. *On dit que $X \in \mathbb{R}^Q$ est une* **solution réalisable** *du problème* (4.1) *si X satisfait aux contraintes, autrement dit si $G_1(X) \leq 0, \cdots, G_P(X) \leq 0$. L'ensemble $\mathcal{P}$ de toutes les solutions réalisables d'un problème d'optimisation est appelé son* **ensemble réalisable**.

Définition 4.4. *On dit que $X \in \mathbb{R}^Q$ est une* **solution optimale** *du problème* (4.1) *si X est une solution réalisable du problème* (4.1) *et si de plus, quelle que soit la solution*

réalisable $Y \in \mathbb{R}^Q$ du problème (4.1) on a nécessairement $F(Y) \leq F(X)$. Autrement dit, une solution réalisable est optimale si elle maximise la fonction objectif sur l'ensemble réalisable.

Remarque 4.5. *Anticipant un peu sur le Chapitre 12, on affirme que l'ensemble $\mathcal{P}$ est un polyèdre dans $\mathbb{R}^n$, c'est-à-dire une intersection finie de demi-espaces fermés de $\mathbb{R}^n$. Si $\mathcal{P}$ est de plus borné (cela n'est pas nécessairement le cas), et puisque la fonction objectif est une fonction continue, l'existence d'au moins une solution optimale est alors garantie par le théorème des bornes atteintes. Dans tous les cas, nous verrons que si la fonction est majorée sur $\mathcal{P}$, alors le problème de maximisation a toujours au moins une solution.*

Un même problème d'optimisation linéaire peut se récrire de diverses manières équivalentes les unes aux autres. Parmi ces versions, nous distinguerons les formes canoniques et les formes standards.

Définition 4.6. *On appelle problème d'optimisation linéaire sous forme canonique un problème de la forme*

$$\begin{array}{lll} \textit{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \textit{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j \leq b_i & (i = 1, \cdots, p), \\ & x_j \geq 0 & (j = 1, \cdots, q), \end{array} \tag{4.2}$$

où $p, q \in \mathbb{N}_*$, et où les c_j $(1 \leq j \leq q)$, les a_{ij} $(1 \leq i \leq p, 1 \leq j \leq q)$, et les b_i $(1 \leq i \leq p)$ sont des constantes réelles.

En écriture matricielle, un problème sous forme canonique s'écrit donc

$$\begin{array}{ll} \text{maximiser} & c^T x \\ \text{sous les contraintes} & Ax \leq b, \\ & x \geq 0, \end{array}$$

où $c = (c_1, \cdots, c_q)^T$ et (la variable) $x = (x_1, \cdots, x_q)^T$ sont des vecteurs colonnes à q lignes, $A = (a_{ij})_{1 \leq i \leq p, 1 \leq j \leq q}$ est une matrice à p lignes et q colonnes, et $b = (b_1, \cdots, b_p)^T$ est un vecteur colonne à p lignes.

Il est immédiat que tout problème d'optimisation linéaire sous forme canonique est un problème d'optimisation linéaire sous forme générale. En effet, la fonction à maximiser dans (4.2) est bien une forme linéaire, les contraintes $x_j \geq 0$ s'écrivent de manière équivalente sous la forme $G_j(x) \leq 0$ avec $G_j(x) = -x_j$ qui est bien une fonction affine, et enfin les contraintes $\sum_{j=1}^{q} a_{ij} x_j \leq b_i$ se récrivent sous la forme $H_j(x) \leq 0$ où $H_j(x) = \sum_{j=1}^{q} a_{ij} x_j - b_i$ est également affine.

Nous allons montrer maintenant que la résolution de n'importe quel problème d'optimisation linéaire sous forme générale peut se ramener à la résolution d'un problème d'optimisation linéaire sous forme canonique. Pour ce faire, on récrit tout d'abord (4.1) sous la forme étendue (c'est-à-dire que l'on rend explicite F et $G_1, \cdots, G_P$) :

$$\begin{array}{ll} \text{maximiser} & \sum_{l=1}^{Q} f_l X_l \\ \text{sous les contraintes} & \sum_{l=1}^{Q} G_{kl} X_l - B_k \leq 0 \ (k = 1, \cdots, P). \end{array} \tag{4.3}$$

On introduit alors les variables fictives $X_1^+, \cdots, X_Q^+$ et $X_1^-, \cdots, X_Q^-$ et on considère le problème

$$\begin{array}{lll} \text{maximiser} & \sum_{l=1}^{Q} f_l(X_l^+ - X_l^-) & \\ \text{sous les contraintes} & \sum_{l=1}^{Q} G_{kl}(X_l^+ - X_l^-) \leq B_k & (k = 1, \cdots, P), \\ & X_l^+ \geq 0, X_l^- \geq 0 & (l = 1 \cdots, Q). \end{array} \tag{4.4}$$

Le problème (4.4) est un problème d'optimisation linéaire sous forme canonique. En effet, il suffit de choisir $p = P$ et $q = 2Q$ et de poser $(x_1, \cdots, x_q) := (X_1^+, \cdots, X_Q^+, X_1^-, \cdots, X_Q^-)$. Le lecteur vérifiera que si $(X_1^+, \cdots, X_N^+, X_1^-, \cdots, X_N^-)$ est une solution réalisable (resp. optimale) de (4.4), alors $(X_1^+ - X_1^-, \cdots, X_Q^+ - X_Q^-)$ est une solution réalisable (resp. optimale) de (4.3). Inversement, si $(X_1, \cdots, X_Q)$ est une solution réalisable (resp. optimale) de (4.3), alors $(X_1^+, \cdots, X_Q^+, X_1^-, \cdots, X_Q^-)$, où pour $1 \leq l \leq Q$ on a défini $X_l^+ := \max(X_l, 0)$ et $X_l^- := -\min(X_l, 0)$, est une solution réalisable (resp. optimale) de (4.4).

Définition 4.7. *On appelle problème d'optimisation linéaire sous forme standard un problème de la forme*

$$\begin{array}{lll} \textit{maximiser} & \sum_{j=1}^{n} c_j x_j & \\ \textit{sous les contraintes} & \sum_{j=1}^{n} a_{ij} x_j = b_i & (i = 1, \cdots, m), \\ & x_j \geq 0 & (j = 1, \cdots, n), \end{array}$$

où $m, n \in \mathbb{N}_*$, et où les c_j ($1 \leq j \leq n$), les a_{ij} ($1 \leq i \leq m, 1 \leq j \leq n$), et les b_i ($1 \leq i \leq m$), sont des constantes réelles.

En écriture matricielle, un problème sous forme standard s'écrit donc

$$\begin{array}{ll} \text{maximiser} & c^T x \\ \text{sous les contraintes} & Ax = b, \\ & x \geq 0, \end{array}$$

où $c = (c_1, \cdots, c_n)^T$ et (la variable) $x = (x_1, \cdots, x_n)^T$ sont des vecteurs colonnes à n lignes, $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ est une matrice à m lignes et n colonnes, et $b = (b_1, \cdots, b_m)^T$ est un vecteur colonne à m lignes.

Remarque 4.8. *Sans perte de généralité, on peut supposer que dans un problème sous forme standard, les lignes de A sont linéairement indépendantes (si ce n'est pas le cas soit certaines contraintes sont redondantes, soit l'ensemble des contraintes est vide). Dans la suite, lorsque nous parlerons de problème sous forme standard, nous supposerons implicitement que les lignes de A sont linéairement indépendantes, autrement dit que*

$$\text{rang}(A) = m,$$

ce qui implique également que $n \geq m$.

A tout problème d'optimisation linéaire sous forme canonique, ont peut associer un problème d'optimisation linéaire sous forme standard de la manière suivante. Soit le système sous forme canonique

$$\begin{array}{lll} \text{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \text{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j \leq b_i & (i = 1, \cdots, p), \\ & x_j \geq 0 & (j = 1, \cdots, q). \end{array} \tag{4.5}$$

On pose $m = p$ et $n = p + q$, et on considère le système

$$\begin{array}{lll} \text{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \text{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j + x_{q+i} = b_i & (i = 1, \cdots, m), \\ & x_j \geq 0 & (j = 1, \cdots, n). \end{array} \tag{4.6}$$

Il est aisé (et c'est un bon exercice) de vérifier que si le vecteur $(x_1, \cdots, x_q)^T$ est une solution réalisable (resp. optimale) du problème (4.5), alors le vecteur

$$(x_1, \cdots, x_n)^T := (x_1, \cdots, x_q, b_1 - \sum_{j=1}^{q} a_{1j} x_j, \cdots, b_p - \sum_{j=1}^{q} a_{pj} x_j)^T$$

est solution réalisable (resp. optimale) du problème (4.6). Inversement, si le vecteur $(x_1, \cdots, x_q, x_{q+1}, \cdots, x_{q+p})^T$ est solution réalisable (resp. optimale) du problème (4.6), alors $(x_1, \cdots, x_q)^T$ est solution réalisable (resp. optimale) du problème (4.5). Le problème (4.6) peut se récrire sous forme matricielle comme

$$\begin{array}{ll} \text{maximiser} & \bar{c}^T \bar{x} \\ \text{sous les contraintes} & \bar{A}\bar{x} = \bar{b}, \\ & \bar{x} \geq 0, \end{array} \tag{4.7}$$

où $\bar{x} := (x_1, \cdots, x_{p+q})^T$,

$$\begin{aligned} \bar{c}^T :=& \ (c_1, \cdots, c_q, 0, \cdots, 0), \\ \bar{A} :=& \begin{pmatrix} a_{11} & \cdots & a_{1q} & 1 & 0 & 0 & \cdots & 0 \\ a_{21} & \cdots & a_{2q} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ a_{p1} & \cdots & a_{pq} & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \\ \bar{b} :=& \ b. \end{aligned}$$

De par sa structure particulière (elle contient la matrice identité de taille q dans sa partie droite), la matrice $\bar{A}$ est nécessairement de rang égal à $m = q$, quelle que soit A.

Définition 4.9. *Dans la suite, nous dirons qu'un problème d'optimisation linéaire est sous forme standard canonique s'il est de la forme* (4.7)

Remarque 4.10. *Même si les méthodes présentées ci-dessus permettent de ramener de manière systématique n'importe quel problème d'optimisation linéaire sous forme générale en des problèmes d'optimisation linéaire sous forme canonique ou standard, dans la pratique il peut arriver (c'était le cas dans le chapitre précédent) que ce ne soit pas la plus économe en nombre de variables. Dans ces cas, il est bien entendu plus avantageux d'utiliser la réduction rendant le problème standard le plus compact possible (i.e. avec le moins de contraintes ou le moins de variables possible).*

Chapitre 5

Solutions de base d'un problème sous forme standard

Dans ce chapitre, nous mettons en évidence certaines solutions réalisables (dites *de base*) pour un problème d'optimisation linéaire sous forme standard. Ces solutions se révéleront suffisantes pour la recherche d'une solution optimale.

Considérons le problème d'optimisation linéaire sous forme standard

$$\begin{array}{ll} \text{maximiser} & c^T x \\ \text{sous les contraintes} & Ax = b, \\ & x \geq 0, \end{array} \tag{5.1}$$

où $c = (c_1, \cdots, c_n)^T$ et (la variable) $x = (x_1, \cdots, x_n)^T$ sont des vecteurs colonnes à n lignes, $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ est une matrice à m lignes et n colonnes vérifiant $\text{rang}(A) = m$, et $b = (b_1, \cdots, b_m)^T$ est un vecteur colonne à m lignes. Sans perte de généralité, on peut supposer que $n > m$, car si $n = m$ l'ensemble réalisable contient au plus un point.

On note

$$A_1 := \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \ \cdots \ , \ A_k = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix}, \ \cdots \ , \ A_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix},$$

les colonnes de la matrice A. Par hypothèse sur le rang de A, on peut trouver m colonnes parmi $A_1, \cdots, A_n$ qui soient linéairement indépendantes. En général, ce choix n'est pas unique. On note

$$\Gamma := \Big\{ \gamma : \{1, \cdots, m\} \to \{1, \cdots, n\} \text{ strictement croissante } \Big\}.$$

Pour $\gamma \in \Gamma$, on note A_γ la matrice carrée de taille m

$$A_\gamma = \big(A_{\gamma(1)}, \cdots, A_{\gamma(m)}\big) = \begin{pmatrix} a_{1\gamma(1)} & \cdots & a_{1\gamma(m)} \\ \vdots & & \vdots \\ a_{m\gamma(1)} & \cdots & a_{m\gamma(m)} \end{pmatrix}.$$

Finalement, on définit

$$\mathcal{B} := \Big\{ \gamma \in \Gamma \text{ t.q. } \text{rang}(A_\gamma) = m \Big\}.$$

Remarque 5.1. *On a bien sûr*

$$\sharp\mathcal{B} \leq \sharp\Gamma = \frac{n!}{m!(n-m)!}.$$

Pour chaque $\gamma \in \Gamma$, on note $\hat{\gamma}$ l'unique application strictement croissante de $\{1, \cdots, n-m\}$ dans $\{1, \cdots, n\}$ telle que

$$\gamma\big(\{1, \cdots, m\}\big) \cup \hat{\gamma}\big(\{1, \cdots, n-m\}\big) = \{1, \cdots, n\}$$

(autrement dit, $\hat{\gamma}$ fournit en ordre croissant les indices complémentaires à ceux atteints par γ).

Définition 5.2. *Etant fixé un choix de $\gamma \in \mathcal{B}$, on dit que les variables $x_{\gamma(1)}, \cdots, x_{\gamma(m)}$ sont les* **variables en base** *(pour γ), tandis que les variables $x_{\hat{\gamma}(1)}, \cdots, x_{\hat{\gamma}(n-m)}$ sont les* **variables hors base** *(pour γ).*

Pour $x \in \mathbb{R}^n$ et $\gamma \in \mathcal{B}$, on note

$$x_B := \big(x_{\gamma(1)}, \cdots, x_{\gamma(m)}\big)^T, \quad x_N := \big(x_{\hat{\gamma}(1)}, \cdots, x_{\hat{\gamma}(n-m)}\big)^T.$$

On note aussi

$$c_B := \big(c_{\gamma(1)}, \cdots, c_{\gamma(m)}\big)^T, \quad c_N := \big(c_{\hat{\gamma}(1)}, \cdots, c_{\hat{\gamma}(n-m)}\big)^T,$$

et enfin

$$B := A_\gamma, \quad N := A_{\hat{\gamma}}.$$

On remarque alors que le système $Ax = b$ se récrit sous la forme

$$Bx_B + Nx_N = b,$$

qui est équivalent, puisque B est inversible lorsque $\gamma \in \mathcal{B}$, au système

$$x_B = B^{-1}b - B^{-1}Nx_N. \tag{5.2}$$

Définition 5.3. *On appelle solution de base du système $Ax = b$ associée au choix de base $\gamma \in \mathcal{B}$ la solution x^* définie par*

$$x^*_B = B^{-1}b, \quad x^*_N = (0, \cdots, 0).$$

Définition 5.4. *(Solution de base réalisable) On dit que la solution de base x^* du système $Ax = b$ associée au choix de base $\gamma \in \mathcal{B}$ est une* **solution de base réalisable** *si de plus elle vérifie les contraintes de (5.1), c'est-à-dire si toutes les composantes de x^* sont positives. Dans ce cas, on dit aussi que la base γ est une* **base réalisable**, *et on note $\mathcal{R}$ l'ensemble des bases réalisables. On dit que la solution de base réalisable x^* est* **non dégénérée** *si toutes les composantes de x^*_B sont strictement positives.*

Corollaire 5.5. *Etant fixée $\gamma \in \mathcal{B}$, pour $x \in \mathbb{R}^n$ solution réalisable quelconque de (5.1), on a*

$$x_B = x^*_B - B^{-1}Nx_N, \quad \text{et} \quad c^Tx = c^Tx^* + d^Tx,$$

où le vecteur d est défini par les relations

$$d^T_N = c^T_N - c^T_B B^{-1}N$$

et

$$d^T_B = (0, \cdots, 0).$$

Démonstration. La première égalité découle de manière directe de la définition de x^*. Pour la seconde, on écrit

$$c^T x = c_B^T x_B + c_N^T x_N$$

et l'on substitue x_B par $B^{-1}b - B^{-1}Nx_N$. Ceci donne

$$c^T x = c_B^T B^{-1} b + \left(c_N^T - c_B^T B^{-1} N\right) x_N = c_B^T x_B^* + \left(c_N^T - c_B^T B^{-1} N\right) x_N = c^T x^* + \left(c_N^T - c_B^T B^{-1} N\right) x_N,$$

d'où la conclusion. □

Définition 5.6. *On dit que le vecteur d est le* **vecteur des prix marginaux** *associé à la base γ.*

Remarque 5.7. *La terminologie pour vecteur des prix marginaux apparaîtra plus clairement dans le Chapitre 10 lorsque nous aborderons les problèmes duaux; il serait en fait plus juste d'attribuer ce nom à $-d$ plutôt qu'à d. Cette notion a été vaguement esquissée dans le premier chapitre, lorsque nous avons évoqué le problème du concurrent.*

Proposition 5.8. *Soit γ une base réalisable et x^* la solution de base associée à γ. Si le vecteur des prix marginaux d n'a que des composantes négatives, alors x^* est une solution optimale du problème* (5.1).

Démonstration. Si x est une solution réalisable quelconque de (5.1), on a par le Corollaire 5.5

$$c^T x = c^T x^* + d_N^T x_N \leq c^T x^*.$$

De fait, le produit matriciel $d_N^T x_N$ est négatif puisque les composantes de d (et donc d_N) sont supposées négatives et celles de x_N positives (car x est réalisable). □

La remarque qui suit est à la base de la technique du tableau (cfr. Chapitre 3) pour l'implémentation de la méthode du simplexe.

Remarque 5.9. *On peut transformer le problème 5.1 en le problème équivalent suivant*

$$\begin{array}{ll} \textit{maximiser} & -z \\ \textit{sous les contraintes} & Ax = b, \\ & c^T x + z = 0, \\ & x \geq 0, z \geq 0, \end{array} \tag{5.3}$$

où z est une variable scalaire réelle (positive). Ce dernier problème peut se récrire sous la forme matricielle

$$\begin{array}{ll} \textit{maximiser} & \bar{c}^T \bar{x} \\ \textit{sous les contraintes} & \bar{A}\bar{x} = \bar{b}, \\ & \bar{x} \geq 0, \end{array} \tag{5.4}$$

où $\bar{x}^T := (x_1, \cdots, x_n, z)$, $\bar{c}^T = (0, \cdots, 0, -1)$, et où

$$\bar{A} := \begin{pmatrix} A & 0 \\ c^T & 1 \end{pmatrix} \quad \textit{et} \quad \bar{b} := \begin{pmatrix} b \\ 0 \end{pmatrix}.$$

Si $\gamma : \{1, \cdots, m\} \to \{1, \cdots, n\}$ est un élément de $\mathcal{B}$ alors $\bar{\gamma} : \{1, \cdots, m+1\} \to \{1, \cdots, n+1\}$ définie par $\bar{\gamma}(j) = \gamma(j)$ si $j \in \{1, \cdots, m\}$ et $\bar{\gamma}(m+1) = n+1$ est

telle que les colonnes $\bar{A}_{\bar{\gamma}(1)}, \cdots, \bar{A}_{\bar{\gamma}(m+1)}$ *sont linéairement indépendantes. Par analogie avec les notations utilisées plus haut, on notera*

$$\bar{B} = \bar{A}_{\bar{\gamma}} = (\bar{A}_{\bar{\gamma}(1)}, \cdots, \bar{A}_{\bar{\gamma}(m+1)}),$$

qui est par conséquent une matrice carrée inversible de taille $m+1$*. On remarque que le système* $\bar{A}\bar{x} = \bar{b}$ *est équivalent au système*

$$\bar{B}^{-1}\bar{A}\bar{x} = \bar{B}^{-1}\bar{b},$$

et on explicite ensuite l'expression de $\bar{B}^{-1}\bar{A}$ *et* $\bar{B}^{-1}\bar{b}$*. On a*

$$\bar{B} = \begin{pmatrix} B & 0 \\ c_B^T & 1 \end{pmatrix} \quad d'où \quad \bar{B}^{-1} = \begin{pmatrix} B^{-1} & 0 \\ -c_B^T B^{-1} & 1 \end{pmatrix},$$

et donc

$$\bar{B}^{-1}\bar{A} = \begin{pmatrix} B^{-1} & 0 \\ -c_B^T B^{-1} & 1 \end{pmatrix} \begin{pmatrix} A & 0 \\ c^T & 1 \end{pmatrix} = \begin{pmatrix} B^{-1}A & 0 \\ c^T - c_B^T B^{-1} A & 1 \end{pmatrix} = \begin{pmatrix} B^{-1}A & 0 \\ d^T & 1 \end{pmatrix},$$

et

$$\bar{B}^{-1}\bar{b} = \begin{pmatrix} B^{-1}b \\ -C_B^T B^{-1} b \end{pmatrix}.$$

Finalement, pour chaque $i \in \{1, \cdots, m\}$*,*

$$(B^{-1}A)_{\gamma(i)} = (0, \cdots, 0, 1, 0, \cdots, 0)^T,$$

où l'unique 1 *est placé en ième position.*

Chapitre 6

Pivot à partir d'une solution de base réalisable : critère de Dantzig

Au chapitre précédent, nous avons associé à tout choix de base $\gamma \in \mathcal{B}$ une solution de base x^*, et nous avons fournit un critère (la Proposition 5.8) permettant de s'assurer que x^* soit une solution optimale.

Dans ce chapitre, nous présentons une méthode permettant, étant donné un choix de base $\gamma \in \mathcal{B}$ pour laquelle la solution de base x^* est réalisable mais le critère de la Proposition 5.8 n'est pas vérifié, de déterminer un autre choix de base $\delta \in \mathcal{B}$ dont la solution de base associée y^* est réalisable et vérifie de plus

$$c^T \bar{y}^* \geq c^T x^*.$$

Cette méthode opère au moyen d'un pivot, au sens où les ensembles $\gamma(\{1, \cdots, m\})$ et $\delta(\{1, \cdots, m\})$ ne diffèrent que par un élément.

Reprenons donc le problème d'optimisation linéaire sous forme standard (5.1) du chapitre précédent :

$$\begin{array}{ll} \text{maximiser} & c^T x \\ \text{sous les contraintes} & Ax = b, \\ & x \geq 0, \end{array}$$

où $n > m \in \mathbb{N}_*$, $c = (c_1, \cdots, c_n)^T$ et (la variable) $x = (x_1, \cdots, x_n)^T$ sont des vecteurs colonnes à n lignes, $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ est une matrice à m lignes et n colonnes vérifiant $\text{rang}(A) = m$, et $b = (b_1, \cdots, b_m)^T$ est un vecteur colonne à m lignes.

Soit $\gamma \in \mathcal{B}$, x^* la solution de base du système $Ax = b$ associée à γ et supposons que x^* soit réalisable mais que l'une au moins des composantes du vecteur d soit strictement positive.

On note

$$E_\gamma := \Big\{ j \in \{1, \cdots, n-m\} \text{ t.q. } d_{\hat{\gamma}(j)} > 0 \Big\}.$$

(par construction on a nécessairement $d_{\gamma(i)} = 0$ pour tout $i \in \{1, \cdots, m\}$)

Supposons $J \in E_\gamma$ fixé (nous verrons différents critères pour ce faire ci-dessous). On définit l'ensemble

$$S_{\gamma,J} := \Big\{ i \in \{1, \cdots, m\} \text{ t.q. } (B^{-1}N)_{iJ} > 0 \Big\}.$$

Lemme 6.1. *Si $S_{\gamma,J} = \emptyset$ alors le problème d'optimisation* (5.1) *n'a pas de solution optimale car la fonction objectif n'est pas bornée supérieurement sur l'ensemble réalisable.*

Démonstration. Soit $t > 0$ fixé quelconque. On définit le vecteur $x \in \mathbb{R}^n$ par les relations

$$\begin{aligned} &x_{\hat{\gamma}(j)} = 0 \text{ si } j \in \{1, \cdots, n-m\} \setminus \{J\}, \\ &x_{\hat{\gamma}(J)} = t, \\ &x_{\gamma(i)} = x^*_{\gamma(i)} - t(B^{-1}N)_{iJ} \text{ si } i \in \{1, \cdots, m\}. \end{aligned}$$

Par construction,

$$x_B = x^*_B - B^{-1}Nx_N,$$

de sorte que

$$Ax = b.$$

De plus, puisque $t \geq 0$, puisque x^* est réalisable, et puisque $(B^{-1}N)_{iJ} \leq 0$ pour tout $i \in \{1, \cdots, m\}$ par hypothèse, on a

$$x \geq 0,$$

et on déduit donc que x est une solution réalisable. Enfin, on a

$$c^T x = c^T x^* + t d_{\hat{\gamma}(J)}.$$

Comme $t \geq 0$ était quelconque et que $d_{\hat{\gamma}(J)} > 0$, on déduit que la fonction objectif n'est pas bornée supérieurement sur l'ensemble réalisable. □

Supposons maintenant que la fonction objectif soit bornée supérieurement sur l'ensemble réalisable et fixons $I \in S_{\gamma,J}$ (là aussi nous verrons différents critères de choix par la suite). Dans tous les cas, on a

Lemme 6.2. *Soit δ l'unique application strictement croissante de $\{1, \cdots, m\}$ dans $\{1, \cdots, n\}$ telle que*

$$\delta\big(\{1, \cdots, m\}\big) = \gamma\big(\{1, \cdots, m\}\big) \setminus \gamma\big(\{I\}\big) \cup \hat{\gamma}\big(\{J\}\big).$$

Alors $\delta \in \mathcal{B}$.

(Autrement dit les colonnes de A associées aux variables en base obtenues en faisant sortir de la base initiale la variable $x_{\gamma(I)}$ et en y faisant rentrer la variable $x_{\hat{\gamma}(J)}$ sont linéairement indépendantes.

Démonstration. Par définition des matrices B et N (pour la base γ), on a

$$A_{\hat{\gamma}(J)} = \sum_{i=1}^{m} (B^{-1}N)_{iJ} A_{\gamma(i)}.$$

Puisque $(B^{-1}N)_{IJ} > 0$ (et donc $\neq 0$) par hypothèse sur I, et puisque la famille $(A_{\gamma(i)})_{1 \leq i \leq m}$ est une base de $\mathbb{R}^m$, on déduit que $A_{\hat{\gamma}(J)}$ n'est pas combinaison linéaire des seuls $A_{\gamma(i)}$ avec $i \in \{1, \cdots, m\} \setminus \{I\}$. Il s'ensuit que δ définit une famille de m vecteurs libres de $\mathbb{R}^m$, et donc que $\delta \in \mathcal{B}$ par définition de $\mathcal{B}$. □

Lemme 6.3. *(Critère de Dantzig) Sous les hypothèse du lemme précédent, si de plus*

$$\frac{x^*_{\gamma(I)}}{(B^{-1}N)_{IJ}} = t_J := \min_{i \in S_{\gamma,J}} \frac{x^*_{\gamma(i)}}{(B^{-1}N)_{iJ}}$$

alors δ est une base réalisable. De plus, si y^* désigne la solution de base réalisable associée à la base δ, alors

$$c^T y^* \geq c^T x^*,$$

l'inégalité étant stricte si $t_J \neq 0$.

Démonstration. On procède de manière assez semblable à la démonstration du Lemme 6.1. Soit $y \in \mathbb{R}^n$ défini par

$$\begin{aligned} &y_{\hat{\gamma}(j)} = 0 \text{ si } j \in \{1, \cdots, n-m\} \setminus \{J\}, \\ &x_{\hat{\gamma}(J)} = t_J, \\ &y_{\gamma(i)} = x^*_{\gamma(i)} - t_J (B^{-1}N)_{iJ} \text{ si } i \in \{1, \cdots, m\}. \end{aligned}$$

Par construction,

$$y_B = x^*_B - B^{-1} N y_N,$$

de sorte que

$$Ay = b.$$

Aussi, par définition de t_J on a

$$y \geq 0$$

et on déduit donc que y est une solution réalisable. De plus

$$y_{\gamma(I)} = x^*_{\gamma(I)} - t_J (B^{-1}N)_{IJ} = 0,$$

et par construction

$$y_{\hat{\gamma}(j)} = 0 \text{ si } j \in \{1, \cdots, n-m\} \setminus \{J\}.$$

Il s'ensuit que, *relativement à la base δ*,

$$y_N = 0.$$

Par conséquent, $y = y^*$ est l'unique solution de base (on sait qu'elle est aussi réalisable) associée à la base δ, et on a

$$c^T y^* = c^T y = c^T x^* + d_{\hat{\gamma}(J)} t_J \geq c^T x^*,$$

l'inégalité étant stricte si $t_J > 0$. □

Dans la méthode présentée ci-dessus, et permettant de passer de la base réalisable γ à la base réalisable δ, il se peut que les indices J puis I des variables entrantes puis sortantes ne soient pas déterminés de manière unique (si E_γ n'est pas réduit à un élément, et si le maximum t_J est atteint pour plusieurs indices I). Dans ces cas, et avec l'optique par exemple de rendre la méthode algorithmique (et donc déterministe) en vue de la programmer sur un ordinateur, il faut y adjoindre un ou des critères additionnels permettant de déterminer I et J de manière univoque. Nous allons décrire deux tels critères (on peut imaginer beaucoup d'autres variantes) :

Définition 6.4 (Critère naturel)**.** *On appelle variable entrante selon le critère naturel la variable $x_{\hat{\gamma}(J)}$ telle que*

$$d_{\hat{\gamma}(J)} = \max_{j \in E_\gamma} d_{\hat{\gamma}(j)} \qquad et \qquad J = \min_{j \in E_\gamma, d_{\hat{\gamma}(j)} = d_{\hat{\gamma}(J)}} j.$$

(autrement dit la variable sortante est dans ce cas une de celles associées aux plus grands coefficients positifs de d, et parmi celles-ci celle de plus petit indice)

On appelle variable sortante selon le critère naturel la variable $x_{\gamma(I)}$ telle que

$$I = \min \left\{ i \in S_{\gamma,J} \ t.q. \ \frac{x^*_{\gamma(i)}}{(B^{-1}N)_{iJ}} = t_J \right\}.$$

(autrement dit la variable entrante est dans ce cas celle de plus petit indice parmi les variables satisfaisant au critère de Dantzig étant donnée la variable entrante J)

Définition 6.5 (Critère de Bland)**.** *On appelle variable entrante selon le critère de Bland la variable $x_{\hat{\gamma}(J)}$ telle que*

$$J = \min_{j \in E_\gamma} j.$$

(autrement dit la variable sortante est dans ce cas celle associée au premier coefficient strictement positif de d)

On appelle variable sortante selon le critère de Bland la variable $x_{\gamma(I)}$ telle que

$$I = \min \left\{ i \in S_{\gamma,J} \ t.q. \ \frac{x^*_{\gamma(i)}}{(B^{-1}N)_{iJ}} = t_J \right\}.$$

(autrement dit la variable entrante est dans ce cas celle de plus petit indice parmi les variables satisfaisant au critère de Dantzig étant donnée la variable entrante J)

Chapitre 7

Non cyclicité sous le critère de Bland

Le chapitre précédent nous fournit une (en réalité deux suivant que l'on applique le critère naturel ou celui de Bland) méthode pour passer d'une base réalisable γ à une base réalisable δ tout en augmentant (au sens non strict) la fonction objectif. Puisqu'il n'y a qu'un nombre fini de bases réalisables (au plus C_n^m), les itérées de cette méthode nous conduiront assurément à une solution optimale à condition que la fonction objectif soit majorée et que nous puissions démontrer que l'algorithme ne cycle pas, c'est-à-dire qu'aucune des bases réalisables obtenues par les différentes itérations ne peut être égale à la base initiale γ. Cette condition n'est pas vérifiée en général pour le critère naturel (il existe un célèbre contre-exemple du à Beale (1955)). A l'inverse, Bland (1974) à démontré que la cyclicité est proscrite suivant son critère :

Théorème 7.1. *(Bland) L'itération de la méthode du Chapitre 6 avec le critère de Bland garantit la non-cyclicité des bases réalisables produites, et par conséquent atteint une solution optimale en un nombre fini d'étapes si la fonction objectif est majorée.*

Démonstration. La démonstration se fait par l'absurde et n'est pas des plus éclairantes. On la fournit néanmoins par souci de complétude. Supposons donc que l'itération de la méthode engendre un cycle. Cela implique bien sûr que la fonction objectif soit constante tout au long du cycle, et donc qu'à chaque étape on ait $t_J = 0$, de sorte que la solution de base x^* est elle aussi constante tout au long du cycle. Appelons K le plus grand indice (parmi $\{1, \cdots, m\}$) pour lequel la variable x_K se retrouve à la fois en base et hors base dans deux itérations différentes du cycle. Appelons aussi γ la base réalisable correspondant à une étape du cycle (que nous fixons) où x_K est appelée à rentrer en base, et δ la base réalisable correspondant à une étape du cycle (que nous fixons) où x_K est appelée à sortir de la base.

Faisant référence à la remarque 5.9 du Chapitre 5, on augmente le système $Ax = b$ en le système $\bar{A}\bar{x} = \bar{b}$, que l'on récrit sous les deux formes équivalentes

$$\bar{B}_\gamma^{-1}\bar{A}\bar{x} = \bar{B}_\gamma^{-1}\bar{b} \qquad \text{et} \qquad \bar{B}_\delta^{-1}\bar{A}\bar{x} = \bar{B}_\delta^{-1}\bar{b},$$

où B_γ et B_δ se réfèrent à la matrice B du Chapitre 5 suivant que l'on choisit γ ou δ comme base réalisable. De plus,

$$\bar{B}_\gamma^{-1}\bar{A} = \begin{pmatrix} B_\gamma^{-1}A & 0 \\ d_\gamma^T & 1 \end{pmatrix} \qquad \text{et} \qquad \bar{B}_\delta^{-1}\bar{A} = \begin{pmatrix} B_\delta^{-1}A & 0 \\ d_\delta^T & 1 \end{pmatrix}.$$

On note $\mathcal{H}$ le sous-espace vectoriel de $\mathbb{R}^{n+1}$ engendré par les lignes de $\bar{A}$. Une première remarque importante est que puisque B_γ et B_δ sont inversibles, $\mathcal{H}$ est aussi le sous-espace vectoriel engendré par les lignes de $\bar{B}_\gamma^{-1}\bar{A}$ ou par celles $\bar{B}_\delta^{-1}\bar{A}$. En particulier, on a

$$h := \begin{pmatrix} d_\gamma^T & 1 \end{pmatrix} \in \mathcal{H}. \tag{7.1}$$

De plus, par définition de K et de γ, et en vertu du critère de Bland,

$$\begin{aligned} &(d_\gamma)_k \leq 0, \quad \forall k < K, \\ &(d_\gamma)_K > 0. \end{aligned} \tag{7.2}$$

Venons en maintenant à l'étape correspondant à la base δ. Puisque x_K est destinée à sortir de la base à cette étape, en particulier x_K est en base à ce moment, et on peut donc écrire $K = \delta(I)$ pour un certain $I \in \{1, \cdots, m\}$. Soit x_E la variable destinée à rentrer en base à l'étape correspondant à la base δ. Par définition de K on a donc $E \leq K$ et d'autre part

$$(d_\delta)_E > 0, \qquad \text{et} \qquad (B_\delta^{-1}A)_{IE} > 0. \tag{7.3}$$

On définit le vecteur $f \in \mathbb{R}^{n+1}$ par

$$\begin{aligned} &f_{\delta(i)} := (\bar{B}_\delta^{-1}\bar{A})_{iE} \qquad \forall i \in \{1, \cdots, m\}, \\ &f_E := -1, \\ &f_{n+1} := (d_\delta)_E, \\ &\text{les autres composantes de } f \text{ étant nulles.} \end{aligned}$$

On prétend que quel que soit $i \in \{1, \cdots, m+1\}$, on a

$$\sum_{l=1}^{n+1} (\bar{B}_\delta^{-1}\bar{A})_{il} f_l = 0,$$

autrement dit le vecteur f est orthogonal à l'espace vectoriel engendré par les lignes de $\bar{B}_\delta^{-1}\bar{A}$, c'est-à-dire $\mathcal{H}$. En effet, pour $1 \leq i \leq m$ on a, par définition de f,

$$\sum_{l=1}^{n+1} (\bar{B}_\delta^{-1}\bar{A})_{il} f_l = \sum_{j=1}^{m} (B_\delta^{-1}A)_{i\delta(j)} f_{\delta(j)} + (B_\delta^{-1}A)_{iE} f_E = (B_\delta^{-1}A)_{i\delta(i)} f_{\delta(i)} - (B_\delta^{-1}A)_{iE} = 0,$$

où on a utilisé le fait que $(B_\delta^{-1}A)_{i\delta(j)} = \delta_{ij}$ (symbole de Kronecker). D'autre par, pour $i = m+1$ on a

$$\sum_{l=1}^{n+1} (\bar{B}_\delta^{-1}\bar{A})_{il} f_l = \sum_{j=1}^{m} (d_\delta)_{\delta(j)} f_{\delta(j)} + (d_\delta)_E f_E + f_{n+1} = 0 - (d_\delta)_E + (d_\delta)_E = 0,$$

où on a ici utilisé le fait que $(d_\delta)_{\delta(j)} \equiv 0$.

Comme $h \in \mathcal{H}$ et $f \in \mathcal{H}^\perp$, on déduit que

$$\sum_{l=1}^{n+1} h_l f_l = 0.$$

Pour $l = n+1$, on a $h_{n+1}f_{n+1} = (d_\delta)_E > 0$ par (7.3), et par conséquent il existe au moins un indice $L \in \{1, \cdots, n\}$ pour lequel $h_L f_L < 0$. En particulier, $h_L \neq 0$ et donc x_L est une variable hors base γ. Aussi, $f_L \neq 0$, et par conséquent ou bien x_L est une variable en base δ ou bien $L = E$. Dans l'un ou l'autre cas, on conclut que x_L est une variable qui rentre et sort de la base à différentes étapes du cycle, et par définition de K cela implique que $L \leq K$. Enfin, $h_K = (d_\gamma)_K > 0$ par (7.2) et $f_K = (B_\delta^{-1}A)_{IE} > 0$ par (7.3). Il s'ensuit que $L < K$. Dès lors $h_L = (d_\gamma)_L \leq 0$ par (7.2), d'où on déduit que $h_L < 0$, et finalement que $f_L > 0$. Comme $f_E = -1$ on ne peut donc avoir $L = E$, et d'un argument précédent on conclut que x_L est une variable en base pour δ (rappelons que x_L est à l'inverse hors base γ, et donc que $x_L^* = 0$). Dès lors, si $L = \delta(J)$, on a alors à l'étape correspondant à la base δ,

$$\begin{aligned} &(B_\delta^{-1}A)_{JE} > 0 && \text{car } f_L > 0, \\ &x^*_{\delta(J)} = 0 && \text{car } x^*_{\delta(J)} = x_L^*, \\ &L < K. \end{aligned}$$

Le critère de Bland nous interdit alors de faire sortir la variable x_K, ce qui constitue la contradiction cherchée. □

Chapitre 8

Détermination d'une première solution de base réalisable

Le Chapitre 6 nous a fourni un moyen de passer d'une base réalisable à une autre base réalisable plus avantageuse du point de vue de la fonction objectif. Il nous reste néanmoins à déterminer, en vue d'initialiser la méthode, comment obtenir une première base réalisable. Ceci constitue l'objet du chapitre présent.

Considérons un problème d'optimisation linéaire sous forme canonique :

$$\begin{array}{ll} \text{maximiser} & c^T x \\ \text{sous les contraintes} & Ax \leq b, \\ & x \geq 0, \end{array} \tag{8.1}$$

où $c = (c_1, \cdots, c_q)^T$ et (la variable) $x = (x_1, \cdots, x_q)^T$ sont des vecteurs colonnes à q lignes, $A = (a_{ij})_{1 \leq i \leq p, 1 \leq j \leq q}$ est une matrice à p lignes et q colonnes, et $b = (b_1, \cdots, b_p)^T$ est un vecteur colonne à p lignes.

Définition 8.1. *On dit que le problème sous forme canonique* (8.1) *est un problème de première espèce si toutes les composantes du vecteur b sont positives. Dans le cas inverse, ou si le problème n'est pas sous forme canonique, on dit qu'il s'agit d'un problème de deuxième espèce.*

Comme indiqué au Chapitre 4, au problème (8.1) ont peut associer un problème équivalent sous forme standard : on explicite (8.1) comme

$$\begin{array}{lll} \text{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \text{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j \leq b_i & (i = 1, \cdots, p), \\ & x_j \geq 0 & (j = 1, \cdots, q), \end{array} \tag{8.2}$$

on pose $m = p$ et $n = p + q$, et on considère le système

$$\begin{array}{lll} \text{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \text{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j + x_{q+i} = b_i & (i = 1, \cdots, m), \\ & x_j \geq 0 & (j = 1, \cdots, n). \end{array} \tag{8.3}$$

Ce dernier problème se récrit sous forme matricielle comme

$$\begin{array}{ll} \text{maximiser} & \bar{c}^T \bar{x} \\ \text{sous les contraintes} & \bar{A}\bar{x} = \bar{b}, \\ & \bar{x} \geq 0, \end{array} \tag{8.4}$$

où $\bar{x} = (x_1, \cdots, x_{p+q})^T$,

$$\begin{aligned}
\bar{c}^T &= (c_1, \cdots, c_q, 0, \cdots, 0), \\
\bar{A} &= \begin{pmatrix} a_{11} & \cdots & a_{1q} & 1 & 0 & 0 & \cdots & 0 \\ a_{21} & \cdots & a_{2q} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ a_{p1} & \cdots & a_{pq} & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \\
\bar{b} &= b.
\end{aligned}$$

De par la forme de $\bar{A}$, on obtient directement le

Lemme 8.2. *Si toutes les composantes de b sont positives (i.e. si le problème sous forme canonique est de première espèce), alors le problème sous forme standard* (8.3) *possède comme base réalisable celle obtenue en ne retenant en base que les m variables d'écart.*

Démonstration. En effet, en posant $\gamma(\{1, \cdots, m\}) = \{n - m + 1, \cdots, n\}$, on s'aperçoit que $\bar{B} := \bar{A}_\gamma$ n'est autre que la matrice identité de taille m (en particulier $\gamma \in \mathcal{B}$), et par conséquent

$$\bar{x}^*_B = (\bar{B})^{-1}\bar{b} = \bar{b} = b$$

n'a que des composantes positives. La conclusion suit alors la définition de base réalisable, puisque d'autre part par construction $\bar{x}^*_N = (0, \cdots, 0)$. □

Venons en maintenant au cas général d'un système de deuxième espèce. Nous commençons par le cas d'un système sous forme standard (nous savons que tout problème d'optimisation peut se ramener sous cette forme). Nous verrons ensuite une deuxième manière de procéder, moins gourmande en nombre de variables additionnelles, qui s'applique lorsque l'on part d'un système sous forme canonique de deuxième espèce.

Soit donc le problème d'optimisation linéaire sous forme standard

$$\begin{array}{ll} \text{maximiser} & c^T x \\ \text{sous les contraintes} & Ax = b, \\ & x \geq 0, \end{array} \tag{8.5}$$

où $c = (c_1, \cdots, c_n)^T$ et (la variable) $x = (x_1, \cdots, x_n)^T$ sont des vecteurs colonnes à n lignes, $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ est une matrice à m lignes et n colonnes, de rang égal à m, et $b = (b_1, \cdots, b_m)^T$ est un vecteur colonne à m lignes.

Sans perte de généralité, on peut supposer que toutes les composantes de b sont positives, sinon il suffit de les multiplier, ainsi que les lignes correspondantes de A, par -1 (ce qui revient à remplacer une égalité entre deux quantités par l'égalité de leurs quantités opposées, ce qui est bien sûr équivalent).

On introduit alors la variable "fictive" $y = (y_1, \cdots, y_m)^T \in \mathbb{R}^m$ et on considère le problème d'optimisation linéaire

$$\begin{array}{ll} \text{maximiser} & -y_1 - y_2 - \cdots - y_m \\ \text{sous les contraintes} & Ax + y = b, \\ & x, y \geq 0. \end{array} \tag{8.6}$$

Notons que la fonction objectif de ce dernier problème n'a aucun lien avec la fonction objectif du problème de départ. Par contre, si le problème (8.5) possède une solution réalisable, alors le maximum du problème (8.6) est égal à zéro, et inversement. L'avantage du problème (8.6) est qu'il possède, comme dans le cas des problèmes de première espèce, une solution de base réalisable évidente donnée par $(x_1, \cdots, x_n) = (0, \cdots, 0)$ et $(y_1, \cdots, y_m) = (b_1, \cdots b_m)$ (d'où le choix d'écrire le système de départ avec b n'ayant que des composantes positives). L'itération de la méthode du chapitre 6 appliquée à (8.6) permet par conséquent de s'assurer si le maximum de (8.6) vaut 0, auquel cas le sommet optimal auquel se finit l'algorithme détermine une solution réalisable de (8.5) et l'itération de la méthode du chapitre 6 peut alors être appliquée directement à (8.5). Si par contre le maximum de (8.6) est strictement négatif, alors l'ensemble réalisable de (8.5) est vide et par conséquent il est vain de tenter de résoudre (8.5).

Revenons maintenant au cas d'un problème sous forme canonique

$$\begin{array}{ll} \text{maximiser} & c^T x \\ \text{sous les contraintes} & Ax \leq b, \\ & x \geq 0, \end{array} \tag{8.7}$$

que l'on a transformé via les variables d'écart en le problème sous forme standard canonique

$$\begin{array}{ll} \text{maximiser} & \bar{c}^T \bar{x} \\ \text{sous les contraintes} & \bar{A}\bar{x} = \bar{b}, \\ & \bar{x} \geq 0, \end{array} \tag{8.8}$$

où $\bar{x} := (x_1, \cdots, x_{p+q})^T$,

$$\begin{aligned} \bar{c}^T &:= (c_1, \cdots, c_q, 0, \cdots, 0), \\ \bar{A} &:= \begin{pmatrix} a_{11} & \cdots & a_{1q} & 1 & 0 & 0 & \cdots & 0 \\ a_{21} & \cdots & a_{2q} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ a_{p1} & \cdots & a_{pq} & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \\ \bar{b} &:= b, \end{aligned}$$

et faisons l'hypothèse que b possède au moins une composante strictement négative (de sorte que le problème soit de deuxième espèce). Plutôt que d'introduire p nouvelles variables comme dans le procédé ci-dessus, on en introduit une seule (appelons la x_{p+q+1}) et on considère le problème d'initialisation

$$\begin{array}{ll} \text{maximiser} & -x_{p+q+1} \\ \text{sous les contraintes} & \bar{\bar{A}}\bar{\bar{x}} = b, \\ & \bar{\bar{x}} \geq 0, \end{array} \tag{8.9}$$

où $\bar{\bar{x}} := (x_1, \cdots, x_{p+q+1})^T$ et

$$\bar{\bar{A}} = \begin{pmatrix} a_{11} & \cdots & a_{1q} & 1 & 0 & 0 & \cdots & 0 & -1 \\ a_{21} & \cdots & a_{2q} & 0 & 1 & 0 & \cdots & 0 & -1 \\ \vdots & & & & & & & \vdots & \vdots \\ a_{p1} & \cdots & a_{pq} & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}.$$

Comme plus haut, le problème (8.8) possède une solution de base réalisable si et seulement si le maximum du problème (8.9) est égal à zéro. Ce dernier problème possède une solution de base réalisable relativement facile à déterminer. Pour ce faire, soit $i_0 \in \{1, \cdots, p\}$ un indice tel quel

$$b_{i_0} = \min_{i \in \{1, \cdots, p\}} b_i < 0.$$

On prétend qu'en choisissant pour variables en base les variables $\{x_q + 1, \cdots, x_{q+p+1}\} \setminus \{x_{p+i_0}\}$ on obtient une base réalisable. En effet, les variables hors base étant fixées à zéro, le système (8.9) se ramène à

$$\begin{array}{ll} x_{p+q+1} = -b_{i_0} & \text{(se lit dans la ligne } i_0\text{)}, \\ x_{p+i} = b_i + x_{p+q+1} & \text{(pour tout } i \neq i_0\text{)}. \end{array}$$

Ainsi, $x_{p+q+1} > 0$ par définition de i_0 et aussi

$$x_{p+i} = b_i - b_{i_0} \geq 0$$

pour tout $i \neq i_0$. La suite de l'analyse suit alors les mêmes lignes que ci-dessus dans le cas d'un problème sous forme standard quelconque.

Chapitre 9

Description algorithmique de la méthode du simplexe

Dans ce chapitre, on traduit sous forme algorithmique la méthode développée dans les chapitres précédents afin de résoudre un problème d'optimisation linéaire que l'on supposera (le Chapitre 4 nous garantit que cela n'enlève aucune généralité) sous forme standard :

$$\begin{array}{ll} \text{maximiser} & c^T x \\ \text{sous les contraintes} & Ax = b, \\ & x \geq 0, \end{array} \tag{9.1}$$

où $c = (c_1, \cdots, c_n)^T$ et (la variable) $x = (x_1, \cdots, x_n)^T$ sont des vecteurs colonnes à n lignes, $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ est une matrice à m lignes et n colonnes, de rang égal à m, et $b = (b_1, \cdots, b_m)^T$ est un vecteur colonne à m lignes de composantes toutes positives.

Etape 1 : Soit on connaît une base réalisable pour (9.1), auquel cas on passe à l'étape 2, soit on n'en connaît pas et considère alors le problème (8.6) du Chapitre 8 pour lequel on dispose d'une base réalisable évidente. On passe alors à l'étape 2 pour ce dernier problème, et on laisse momentanément en suspend la résolution de (9.1).

Etape 2 : On dispose d'une base réalisable. Si le vecteur des prix marginaux pour cette base réalisable n'a que des composantes négatives, la solution de base réalisable correspondant à la base réalisable courante est un optimum pour le problème en question, on passe alors directement à l'Etape 3. Sinon, on applique la méthode décrite dans le Chapitre 6. Celle-ci permet soit de déterminer que le problème d'optimisation en question n'est pas majoré (cfr. Lemme 6.1), auquel cas on passe directement à l'Etape 4, soit de déterminer une nouvelle base réalisable meilleure (au sens non strict) concernant la fonction objectif. Dans ce dernier cas, on applique par défaut le critère naturel pour le choix des variables entrantes et sortantes, sauf si celui-ci conduit à une augmentation non stricte de la fonction objectif, auquel cas on applique le critère de Bland. On retourne ensuite en début d'Etape 2 avec cette nouvelle base réalisable.

Etape 3 : On dispose d'une solution optimale pour le problème en question. S'il s'agit du problème initial (9.1), on passe à l'Etape 6. S'il s'agit du problème (8.6), ou bien le maximum est strictement négatif, auquel cas le problème (9.1) possède un ensemble réalisable vide et on passe à l'Etape 5, ou bien le maximum est égal à zéro, ce qui fournit une base réalisable pour (9.1), et on retourne à l'Etape 1.

Etape 4 : On sort de l'algorithme avec la conclusion que la fonction objectif du problème d'optimisation 9.1 n'est pas majorée sur l'ensemble réalisable.

Etape 5 : On sort de l'algorithme avec la conclusion que l'ensemble réalisable du problème d'optimisation 9.1 est vide.

Etape 6 : On sort de l'algorithme avec une solution optimale.

Il est vivement conseillé au lecteur de tenter d'implémenter l'algorithme ci-dessus dans son langage de programmation préféré. La procédure en question recevra en entrée la matrice A ainsi que les vecteurs b et c et fournira en sortie une solution optimale (s'il en existe, sinon elle en indiquera l'une des deux raisons possibles) ainsi que la valeur optimale de la fonction objectif.

Chapitre 10

Dualité en programmation linéaire

Considérons à nouveau un problème d'optimisation linéaire sous forme canonique

$$\begin{array}{lll} \text{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \text{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j \leq b_i & (i = 1, \cdots, p), \\ & x_j \geq 0 & (j = 1, \cdots, q), \end{array} \tag{10.1}$$

où $p, q \in \mathbb{N}_*$, et où les c_j $(1 \leq j \leq q)$, les a_{ij} $(1 \leq i \leq p, 1 \leq j \leq q)$, et les b_i $(1 \leq i \leq p)$ sont des constantes réelles.

A supposer que (10.1) possède au moins une solution optimale x^*, la méthode du simplexe permet, à chacune de ses étapes, d'obtenir (et d'améliorer) une *borne inférieure* sur la valeur optimale $c^T x^*$ de la fonction objectif.

Une question naturelle, mais qui va dans la direction opposée, est celle de savoir s'il est possible d'obtenir une *borne supérieure* sur la valeur $c^T x^*$, et cela sans être obligé de parcourir l'algorithme du simplexe jusqu'à ce qu'il aboutisse à une solution optimale. De fait, on peut imaginer que dans certains cas on puisse être intéressé à encadrer avec plus ou moins de précision la valeur $c^T x^*$, sans toutefois requérir sa valeur exacte (par exemple si les calculs sont coûteux ou si le temps est compté).

Pour ce faire, partons des p contraintes d'inégalités

$$\sum_{j=1}^{q} a_{ij} x_j \leq b_i \qquad (i = 1, \cdots, p),$$

et faisons en une somme pondérée au moyen de p coefficients positifs y_i :

$$\sum_{i=1}^{p} y_i \left(\sum_{j=1}^{q} a_{ij} x_j \right) \leq \sum_{i=1}^{p} y_i b_i.$$

En récrivant le terme de gauche de l'inégalité précédente, on obtient

$$\sum_{j=1}^{q} \left(\sum_{i=1}^{p} a_{ij} y_i \right) x_j \leq \sum_{i=1}^{p} y_i b_i,$$

de sorte que si

$$\sum_{i=1}^{p} a_{ij} y_i \geq c_j \qquad (j = 1, \cdots, q), \tag{10.2}$$

alors nécessairement pour toute solution réalisable $x \equiv (x_1, \cdots, x_q)$ de (10.1) on a

$$\sum_{j=1}^{q} c_j x_j \leq \sum_{i=1}^{p} y_i b_i.$$

En particulier,

$$c^T x^* \leq b^T y$$

quel que soit $y \equiv (y_1, \cdots, y_p)$ vérifiant (10.2) et tel que $y \geq 0$. Autrement dit, pour obtenir une borne supérieure sur la valeur optimale de la fonction objectif du problème (10.1), il suffit de connaître une solution réalisable du problème dual de (10.1), que nous définissons plus formellement maintenant :

Définition 10.1 (**Problème primal et dual).** *Le problème dual de* (10.1) *est le problème*

$$\begin{array}{lll} \textit{minimiser} & \sum_{i=1}^{p} b_i y_i & \\ \textit{sous les contraintes} & \sum_{i=1}^{p} a_{ij} y_i \geq c_j & (j = 1, \cdots, q), \\ & y_i \geq 0 & (i = 1, \cdots, p). \end{array} \tag{10.3}$$

On dit aussi que (10.1) *est le problème primal de* (10.3)*.*

Remarquons que sous forme matricielle, les problèmes primal et dual s'écrivent donc

$$\begin{array}{lll} \text{maximiser} & c^T x & \\ \text{sous les contraintes} & Ax \leq b, & \\ & x \geq 0, & \end{array} \quad \mapsto \quad \begin{array}{lll} \text{minimiser} & b^T y & \\ \text{sous les contraintes} & A^T y \geq c, & \\ & y \geq 0. & \end{array}$$

Remarquons aussi que le problème dual est équivalent au problème d'optimisation linéaire sous forme canonique

$$\begin{array}{ll} \text{maximiser} & (-b)^T y \\ \text{sous les contraintes} & (-A)^T y \leq -c, \\ & y \geq 0, \end{array}$$

l'optimum du second étant égal à l'opposé de l'optimum du premier [1]. Dès lors, on peut appliquer au problème dual tout ce que nous avons développé jusqu'ici concernant le problème primal, en particulier l'algorithme du simplexe, ce qui permet déjà de donner une réponse à la question évoquée en tête de chapitre.

Finalement, remarquons que puisque le problème dual (10.3) est lui-même (équivalent à) un problème primal, on peut considérer son problème dual à lui aussi. On s'aperçoit de suite que ce dernier n'est autre que le problème primal du départ, autrement dit **le dual du problème dual est égal au problème primal** qui est donc aussi un problème dual !

Répétant alors l'argumentaire nous ayant conduit à la définition 10.1, on obtient directement le

Théorème 10.2. *Si x est une solution réalisable du problème primal* (10.1)*, et y une solution réalisable du problème dual* (10.3)*, alors nécessairement*

$$c^T x \leq b^T y.$$

En particulier, si $c^T x = b^T y$ alors x est une solution optimale du primal et y est une solution optimale du dual.

1. En effet, cela suit la formule classique $-\max_y(-f(y)) = \min_y(f(y))$ avec $f(y) = b^T y$.

Corollaire 10.3. *Si la fonction objectif du problème primal est non majorée sur son ensemble réalisable, alors le problème dual ne possède aucune solution réalisable. Inversement, si la fonction objectif du problème dual est non minorée sur son ensemble réalisable, alors le problème primal ne possède aucune solution réalisable.*

Attention au fait que les réciproques des énoncés du Corollaire 10.3 sont fausses en général.

Un résultat plus difficile est le théorème de dualité suivant

Théorème 10.4 (**Théorème de dualité de Gale, Kuhn et Tucker).** *Le problème primal* (10.1) *possède une solution optimale x^* si et seulement si le problème dual* (10.3) *possède une solution optimale y^*. Dans ce cas, on a nécessairement*

$$\sum_{j=1}^{q} c_j x_j^* = \sum_{i=1}^{p} b_i y_i^*.$$

Démonstration. Au vu du Théorème 10.2, et de la remarque sur le dual du dual, il suffit de montrer que si $x^* \equiv (x_1^*, \cdots, x_q^*)$ est une solution optimale du problème primal, alors il existe une solution réalisable $y^* \equiv (y_1^*, \cdots, y_p^*)$ du dual pour laquelle

$$\sum_{j=1}^{q} c_j x_j^* = \sum_{i=1}^{p} b_i y_i^*.$$

Partant donc du problème primal

$$\begin{array}{lll} \text{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \text{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j \leq b_i & (i = 1, \cdots, p), \\ & x_j \geq 0 & (j = 1, \cdots, q), \end{array}$$

introduisons les p variables d'écart

$$x_{q+i} := b_i - \sum_{j=1}^{q} a_{ij} x_j, \qquad (i = 1, \cdots, p),$$

et récrivons le problème primal sous forme standard

$$\begin{array}{lll} \text{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \text{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j + x_{q+i} = b_i & (i = 1, \cdots, p), \\ & x_j \geq 0 & (j = 1, \cdots, q+p), \end{array} \tag{10.4}$$

ou sous forme matricielle

$$\begin{array}{ll} \text{maximiser} & \bar{c}^T \bar{x} \\ \text{sous les contraintes} & \bar{A}\bar{x} = \bar{b}, \\ & \bar{x} \geq 0, \end{array}$$

avec (cfr. Chapitre 4)

$$\begin{aligned}
\bar{x} &= (\bar{x}_1, \cdots, \bar{x}_q, \bar{x}_{q+1}, \cdots, \bar{x}_{q+p}),\\
\bar{c}^T &= (c_1, \cdots, c_q, 0, \cdots, 0),\\
\bar{A} &= \begin{pmatrix} a_{11} & \cdots & a_{1q} & 1 & 0 & 0 & \cdots & 0\\ a_{21} & \cdots & a_{2q} & 0 & 1 & 0 & \cdots & 0\\ \vdots & & & & & & & \vdots\\ a_{p1} & \cdots & a_{pq} & 0 & 0 & \cdots & 0 & 1 \end{pmatrix},\\
\bar{b} &= b.
\end{aligned}$$

Appelons γ la base réalisable pour (10.4) correspondant à la solution optimale $\bar{x}^* = (x_1^*, \cdots, x_q^*, b_1 - \sum_{j=1}^q a_{1j}x_j^*, \cdots, b_p - \sum_{j=1}^q a_{pj}x_j^*)$, et soit $\bar{d} \equiv (d_1, \cdots, d_q, d_{q+1}, \cdots, d_{q+p})$ le vecteur des prix marginaux correspondants. Alors pour toute solution (non nécessairement réalisable) $\bar{x}$ de (10.4) on a

$$\begin{aligned}
\bar{c}^T\bar{x} = \bar{c}^T\bar{x}^* + \bar{d}^T\bar{x} &= c^Tx^* + \sum_{j=1}^q d_j\bar{x}_j + \sum_{i=1}^p d_{q+i}\big(b_i - \sum_{j=1}^q a_{ij}\bar{x}_j\big)\\
&= \left[c^Tx^* - \sum_{i=1}^p b_i(-d_{q+i})\right] + \sum_{j=1}^q \left[d_j - \sum_{i=1}^p a_{ij}d_{q+i}\right]\bar{x}_j.
\end{aligned} \tag{10.5}$$

En choisissant pour $\bar{x}$ l'unique solution pour laquelle $\bar{x}_1 = \cdots = \bar{x}_q = 0$ on obtient de (10.5)

$$c^Tx^* = \sum_{i=1}^p b_i(-d_{q+i}). \tag{10.6}$$

En choisissant pour $\bar{x}$ l'unique solution pour laquelle $\bar{x}_1 = \cdots = \bar{x}_q = 0$ sauf pour un certain $j \in \{1, \cdots, q\}$ pour lequel $\bar{x}_j = 1$, on obtient, par (10.5) et en tenant compte de (10.6),

$$d_j - \sum_{i=1}^p a_{ij}d_{q+i} = c_j. \tag{10.7}$$

Par ailleurs, puisque $\bar{x}^*$ est une solution optimale, nécessairement le vecteur des prix marginaux $\bar{d}$ en cette solution est négatif. D'autre part $j \in \{1, \cdots, q\}$ a été choisi quelconque. Ainsi, de (10.7) on obtient

$$\sum_{i=1}^p a_{ij}(-d_{q+i}) = c_j - d_j \geq c_j, \qquad j = 1, \cdots, q. \tag{10.8}$$

Il s'ensuit que le vecteur $y^* \equiv (y_1^*, \cdots, y_p^*) := (-d_{q+1}, \cdots, -d_{q+p})$ est une solution réalisable du problème dual (10.3), et par (10.6) on a

$$c^Tx^* = b^Ty^*,$$

ce qui termine la démonstration. □

La démonstration du théorème précédent permet également d'affirmer le

Corollaire 10.5 (**Solution optimale du dual et prix marginaux**). *Si le problème primal* (10.1) *possède une solution optimale x^*, et si $\bar{d} \equiv (d_1, \cdots, d_q, d_{q+1}, \cdots, d_{q+p})$ désigne le vecteur des prix marginaux pour la base réalisable correspondant à x^*, alors une solution optimale du problème dual* (10.3) *est fournie par*

$$(y_1^*, \cdots, y_p^*) := (-d_{q+1}, \cdots, -d_{q+p}).$$

Théorème 10.6 (**Prix marginaux et problème dual**). *Supposons que le problème primal* (10.1) *possède une solution optimale x^* non dégénérée.*[2] *Alors il existe $\varepsilon > 0$ tel que si les variations $t_1, \cdots, t_p$ des contraintes vérifient la condition de petitesse*

$$|t_i| \leq \varepsilon \qquad \forall\, i \in \{1, \cdots, p\},$$

l'optimum du problème primal modifié

$$\begin{array}{lll} \textit{maximiser} & \sum_{j=1}^{q} c_j x_j & \\ \textit{sous les contraintes} & \sum_{j=1}^{q} a_{ij} x_j \leq b_i + t_i & (i = 1, \cdots, p), \\ & x_j \geq 0 & (j = 1, \cdots, q), \end{array} \tag{10.9}$$

est encore atteint et est donné par

$$c^T x^* + \sum_{i=1}^{p} t_i y_i^*,$$

où $(y_1^*, \cdots, y_p^*)$ désigne une (en fait l'unique) solution optimale du problème dual (10.3).

Démonstration. Le problème dual de (10.9) est donné par

$$\begin{array}{lll} \text{minimiser} & \sum_{i=1}^{p} (b_i + t_i) y_i & \\ \text{sous les contraintes} & \sum_{i=1}^{p} a_{ij} y_i \geq c_j & (j = 1, \cdots, q), \\ & y_i \geq 0 & (i = 1, \cdots, p). \end{array} \tag{10.10}$$

La variation des contraintes du primal se traduit donc par une variation de la fonction objectif du dual, mais les contraintes du dual ne sont pas modifiées. En particulier, les solutions de base réalisables de (10.10) et de (10.3) sont identiques. Par hypothèse, x^* est non-dégénérée, et par conséquent y^* est l'unique solution optimale de (10.3). En effet, le corollaire précédent (utilisé en renversant le rôle du primal et du dual) implique que les prix marginaux du dual correspondant aux variables d'écart du dual sont donnés par x^*, et sont donc tous strictement négatifs, il s'ensuit que l'optimum du dual est atteint et uniquement atteint en y^*. Puisque y^* est l'unique solution optimale de (10.3), il existe $r > 0$ tel que

$$\sum_{i=1}^{p} b_i y_i^* \geq \sum_{i=1}^{p} b_i y_i + r$$

quelle que soit la solution de base réalisable $(y_1, \cdots, y_p)$ de (10.3) (et donc de (10.10)) différente de y^*. Par continuité et finitude (il n'y a qu'un nombre fini de solutions de base réalisables), on déduit alors qu'il existe $\varepsilon > 0$ tel que

$$\sum_{i=1}^{p} (b_i + t_i) y_i^* > \sum_{i=1}^{p} (b_i + t_i) y_i$$

2. On rappelle qu'une solution de base est dite non dégénérée si aucune des composantes correspondant aux variables en base n'est nulle.

quelle que soit la solution de base réalisable $(y_1, \cdots, y_p)$ de (10.10) différente de y^*, et quels que soient $t_1, \cdots, t_p$ vérifiant $|t_i| \leq \varepsilon$ pour chaque $i \in \{1, \cdots, p\}$. En particulier, pour de tels t_i le problème dual (10.10) possède un optimum dont la valeur est donnée par $\sum_{i=1}^{p}(b_i + t_i)y_i^*$. Le Théorème de dualité 10.4 assure alors que (10.9) possède également un optimum, dont la valeur optimale, égale à celle du dual (10.10), est donnée par

$$\sum_{i=1}^{p}(b_i + t_i)y_i^* = \sum_{i=1}^{p} b_i y_i^* + \sum_{i=1}^{p} t_i y_i^* = \sum_{j=1}^{q} c_j x_j^* + \sum_{i=1}^{p} t_i y_i^* = c^T x^* + + \sum_{i=1}^{p} t_i y_i^*.$$

Ceci termine la démonstration. □

Il est conseillé au lecteur de relire la partie du premier chapitre consacrée au problème du concurrent à la lumière du Corollaire précédent.

Remarque 10.7 (Sur le choix du primal ou du dual pour l'application de l'algorithme du simplexe). *Dans certains cas, il est plus intéressant d'appliquer l'algorithme du simplexe à un problème dual (ou plûtot à sa forme standard associée) plutôt qu'au problème primal. Considérons en effet un primal ayant* $p = 1000$ *contraintes pour seulement* $q = 100$ *inconnues. La forme standard associée au primal (après introduction des variables d'écart) aura* $m = 1000$ *contraintes pour* $n = p + q = 1100$ *inconnues. L'algorithme du simplexe dans ce cas nécessite à chaque étape de résoudre un système de taille* 1000×1000. *A l'inverse, le problème dual possède* $p = 100$ *contraintes pour* $q = 1000$ *inconnues. Le problème standard qui lui est associé possède donc* $m = 100$ *contraintes pour* $n = p + q = 1100$ *inconnues. L'algorithme du simplexe dans ce deuxième cas ne nécessite plus donc à chaque étape "que" de résoudre un système de taille* 100×100, *ce qui constitue un gain énorme. En résumé, il est donc préférable de considérer en priorité la version (primale ou duale) qui possède le moins de contraintes.*

Théorème 10.8 (Théorème des écarts complémentaires). *Soient* $x \equiv (x_1, \cdots, x_q)$ *et* $y = (y_1, \cdots, y_p)$ *deux solutions réalisables respectivement du problème primal* (10.1) *et du problème dual* (10.3). *Une condition nécessaire et suffisante pour que* x *et* y *soient optimaux simultanément est que*

$$\begin{aligned} &\forall\ j \in \{1, \cdots, q\},\ \textit{si}\ x_j > 0\ \textit{alors}\ \sum_{i=1}^{p} a_{ij} y_i = c_j \\ &\textit{et} \\ &\forall\ i \in \{1, \cdots, p\},\ \textit{si}\ y_i > 0\ \textit{alors}\ \sum_{j=1}^{q} a_{ij} x_j = b_i. \end{aligned} \tag{10.11}$$

Démonstration. Elle reprend le schéma du début de chapitre concernant les bornes supérieures et inférieures sur la fonction objectif.

Puisque x est une solution réalisable du primal et que $y \geq 0$,

$$\sum_{i=1}^{p} \left(\sum_{j=1}^{q} a_{ij} x_j \right) y_i \leq \sum_{i=1}^{p} b_i y_i.$$

Inversement, puisque y est une solution réalisable du dual et que $x \geq 0$,

$$\sum_{j=1}^{q} \left(\sum_{i=1}^{p} a_{ij} y_i \right) x_j \geq \sum_{j=1}^{q} c_j x_j.$$

Par ailleurs, en développant on obtient bien sûr que

$$\sum_{i=1}^{p} \left(\sum_{j=1}^{q} a_{ij} x_j \right) y_i = \sum_{j=1}^{q} \left(\sum_{i=1}^{p} a_{ij} y_i \right) x_j.$$

Ainsi,

$$\sum_{j=1}^{q} c_j x_j \leq \sum_{j=1}^{q} \left(\sum_{i=1}^{p} a_{ij} y_i \right) x_j = \sum_{i=1}^{p} \left(\sum_{j=1}^{q} a_{ij} x_j \right) y_i \leq \sum_{i=1}^{p} b_i y_i. \tag{10.12}$$

Nous pouvons maintenant démontrer la condition nécessaire. Si x et y sont optimales, alors par le Théorème de dualité on a $\sum_{j=1}^{q} c_j x_j = \sum_{i=1}^{p} b_i y_i$ et par conséquent chacune des deux inégalités de (10.12) est une égalité. Les inégalités de (10.11) s'ensuivent.

Inversement pour la condition suffisante. Si les inégalités de (10.11) sont vérifiées, alors les deux inégalités de (10.12) sont des égalités et par conséquent $\sum_{j=1}^{q} c_j x_j = \sum_{i=1}^{p} b_i y_i$. Il s'ensuit du Théorème 10.2 que x et y sont optimales. □

Corollaire 10.9. *Soit x une solution réalisable du problème primal. Alors x est optimale si et seulement si il existe une solution réalisable y du problème dual pour laquelle $\sum_{j=1}^{q} a_{ij} x_j = b_i$ pour tout i vérifiant $y_i > 0$ et $\sum_{i=1}^{p} a_{ij} y_i = c_j$ pour tout j vérifiant $x_j > 0$.*

Chapitre 11

Rappels de géométrie affine

Définition 11.1 (**Sous-espace affine**). *Un sous-ensemble non vide F de $\mathbb{R}^n$ est appelé un sous-espace affine si*

$$\forall x, y \in F,\ \forall \lambda \in \mathbb{R}, \quad \lambda x + (1-\lambda) y \in F.$$

Proposition 11.2. *On a les propriétés élémentaires suivantes :*

1. *Si F_1, F_2 sont deux sous-espaces affines de $\mathbb{R}^n$ et λ_1, λ_2 deux réels, alors $\lambda_1 F_1 + \lambda_2 F_2$ est un sous-espace affine de $\mathbb{R}^n$.*
2. *Si F_1, F_2 sont deux sous-espaces affines de $\mathbb{R}^n$, alors $F_1 \cap F_2$ est un sous-espace affine de $\mathbb{R}^n$.*
3. *Si F est un sous-espace affine de $\mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}^m$ est une application affine, i.e. $f(x) = Ax + b$ pour certains $A \in \mathbb{R}^{m\times n}$ et $b \in \mathbb{R}^m$, alors $f(F)$ est un sous-espace affine de $\mathbb{R}^m$.*

Définition 11.3 (**Espace vectoriel associé et dimension**). *Soit F un sous-espace affine de $\mathbb{R}^n$. L'espace vectoriel associé à F est défini par*

$$\mathrm{lin}(F) := \{v \in \mathbb{R}^n \mid \forall x \in F,\ x + v \in F\}.$$

La dimension de F est la dimension (au sens des espaces vectoriels) de son espace vectoriel associé.

Définition 11.4 (**Indépendance affine**). *On dit que les points $x_1, \cdots, x_k$ de $\mathbb{R}^n$ sont affinement indépendants si quels que soient les réels $\lambda_1, \cdots, \lambda_k$ vérifiant*

$$\sum_{j=1}^{k} \lambda_j = 0,$$

si

$$\sum_{j=1}^{k} \lambda_j x_j = 0$$

alors nécessairement $\lambda_1 = \cdots = \lambda_k = 0$.

Lemme 11.5. *Les points $x_1, \cdots, x_k$ de $\mathbb{R}^n$ sont affinement indépendants si et seulement si les vecteurs $x_2 - x_1, \cdots, x_k - x_1$ sont linéairement indépendants.*

Démonstration. Pour la condition nécessaire, si $\mu_2, \cdots, \mu_k$ sont tels que

$$\sum_{j=2}^{k} \mu_j (x_j - x_1) = 0,$$

alors

$$\sum_{j=1}^{k} \lambda_j x_j = 0$$

où $\lambda_j := \mu_j$ pour $j \in \{2, \cdots, k\}$ et

$$\lambda_1 := -\sum_{j=2}^{k} \mu_j.$$

En particulier, puisque $\sum_{j=1}^{k} \lambda_j = 0$, on déduit de l'indépendance affine des x_j que $\lambda_j = 0$ pour tout $j \in \{1, \cdots, k\}$ et donc aussi que $\mu_j = 0$ pour tout $j \in \{2, \cdots, k\}$.

Pour la condition suffisante, si $\lambda_1, \cdots, \lambda_k$ sont tels que

$$\sum_{j=1}^{k} \lambda_j x_j = 0$$

alors $\lambda_1 = -\sum_{j=2}^{k} \lambda_j$ et donc

$$\sum_{j=2}^{k} \lambda_j (x_j - x_1) = 0.$$

Il suit de l'indépendance vectorielle des $(x_j - x_1)$ que $\lambda_j = 0$ pour tout $j \in \{2, \cdots, l\}$, et donc aussi que $\lambda_1 = -\sum_{j=2}^{k} \lambda_j = 0$. □

Corollaire 11.6. *Un sous-espace affine F est de dimension k si et seulement si le nombre maximal de points affinement indépendants qu'il contient est égal à $k+1$.*

Démonstration. Si F est de dimension k, et si $e_1, \cdots, e_k$ est une base de $\mathrm{lin}(F)$, alors au vu du lemme précédent quel que soit $x \in F$ les $k+1$ points $x, x-e_1, \cdots, x-e_k$ sont affinement indépendants. Inversement, si $x_1, \cdots, x_{k+1}$ sont $k+1$ points affinements indépendants dans F, alors les k vecteurs $x_2 - x_1, \cdots, x_{k+1} - x_1$ sont linéairement indépendants et appartiennent à $\mathrm{lin}(F)$. Il s'ensuit que $\mathrm{lin}(F)$ est de dimension supérieure ou égale à k. La conclusion suit ces deux implications. □

Définition 11.7 (Combinaison affine). *Soient $x_1, \cdots, x_k$ un nombre fini de points de $\mathbb{R}^n$, et $\lambda_1, \cdots, \lambda_k$ des réels tels que*

$$\sum_{j=1}^{k} \lambda_j = 1.$$

On dit que

$$x := \sum_{j=1}^{k} \lambda_j x_j$$

est une combinaison affine des points $x_1, \cdots, x_k$.

Plus généralement, si $S \subseteq \mathbb{R}^n$ est un sous-ensemble quelconque, on dit que $x \in \mathbb{R}^n$ est une combinaison affine de points de S s'il existe un nombre fini de points de S dont x soit une combinaison affine.

Proposition 11.8. *Un sous-ensemble F de $\mathbb{R}^n$ est un sous-espace affine si et seulement si il contient toutes les combinaisons affines de points de F.*

Démonstration. La condition suffisante est immédiate, puisque la condition intervenant dans la Définition 11.1 ne fait intervenir autre-chose qu'une combinaison affine à deux éléments. Pour la condition suffisante, on peut utiliser une récurrence sur le nombre d'éléments intervenant dans la combinaison affine. En effet, si l'on suppose que toute combinaison affine d'au plus k éléments de F est dans F, et si $x := \sum_{j=1}^{k+1} \lambda_j x_j$ est une combinaison affine de $k+1$ éléments de F (s'entend donc $\sum_{j=1}^{k+1} = 1$), alors on a (on peut supposer que tous les λ_j sont non nuls sinon il s'agit d'une combinaison à au plus k éléments)

$$x = \lambda_1 x_1 + (1-\lambda_1)\left(\sum_{j=2}^{k+1} \frac{\lambda_j}{1-\lambda_1} x_j\right).$$

Or,

$$\sum_{j=2}^{k+1} \frac{\lambda_j}{1-\lambda_1} = 1,$$

et dès lors par hypothèse de récurrence

$$y := \sum_{j=2}^{k+1} \frac{\lambda_j}{1-\lambda_1} x_j \in F.$$

Finalement, puisque F est affine $x = \lambda_1 x_1 + (1-\lambda_1) y \in F$ et la démonstration est complète. □

Pour un ensemble quelconque $S \subseteq \mathbb{R}^n$, on peut considérer le plus petit sous-espace affine le contenant, c'est la notion suivante d'*enveloppe affine*.

Définition 11.9 (Enveloppe affine). *L'enveloppe affine d'un sous-ensemble S de $\mathbb{R}^n$ est l'ensemble de toutes les combinaisons affines de points de S. On la note* $\mathrm{aff}(S)$.

Proposition 11.10. *L'enveloppe affine de $S \subseteq \mathbb{R}^n$ est un sous-espace affine, c'est le plus petit sous-espace affine contenant S.*

Démonstration. Si un sous-espace affine contient F, alors par la Proposition 11.8 il contient toutes les combinaisons affines de points de F, c'est-à-dire $\mathrm{aff}(F)$. Pour conclure, il suffit donc de montrer que $\mathrm{aff}(F)$ est un sous-espace affine. Soit $x := \sum_{j=1}^{k} \lambda_j x_j \in \mathrm{aff}(F)$ (s'entend que $x_j \in F$ et $\sum \lambda_j = 1$), $y := \sum_{i=1}^{\ell} \mu_i y_i \in \mathrm{aff}(F)$ (s'entend que $y_i \in F$ et $\sum \mu_i = 1$), et $\lambda \in \mathbb{R}$. Alors

$$\lambda x + (1-\lambda) y = \sum_{j=1}^{k} \lambda\lambda_j x_j + \sum_{i=1}^{\ell} (1-\lambda)\mu_i y_i \in \mathrm{aff}(F)$$

puisque $\sum_{j=1}^{k} \lambda\lambda_j x_j + \sum_{i=1}^{\ell}(1-\lambda)\mu_i = \lambda + (1-\lambda) = 1$. [Plus généralement, on montrerait de même que toute combinaison affine de combinaisons affines est une combinaison affine] □

Théorème 11.11 (**Carathéodory**). *Si S est un sous-ensemble de $\mathbb{R}^n$, tout point de* $\mathrm{aff}(S)$ *peut s'écrire comme une combinaison affine de points de S affinement indépendants.*

Démonstration. Soit $x = \sum_{j=1}^{k} \lambda_j x_j$ un point de $\mathrm{aff}(F)$, où sans perte de généralité on suppose que tous les λ_j sont non nuls. Si les x_j ne sont pas affinement indépendants, l'un d'entre eux (quitte à les renuméroter on peut supposer qu'il s'agit de x_k) peut s'écrire comme combinaison affine des $k-1$ autres :

$$x_k = \sum_{i=1}^{k-1} \mu_i x_i \qquad \text{avec } \sum_{i=1}^{k-1} \mu_i = 1.$$

Dès lors,

$$x = \sum_{j=1}^{k-1} (\lambda_j + \lambda_k \mu_j)\, x_j$$

et $\sum_{j=1}^{k-1}(\lambda_j + \lambda_k\mu_j) = 1$, de sorte que x s'écrit comme combinaison affine d'au plus $k-1$ points de F. Si les points $x_1, \cdots, x_{k-1}$ ne sont pas affinement indépendants, on recommence la procédure ci-dessus, et ainsi de suite. Au bout d'un nombre fini d'étapes (au plus $k-1$ puisqu'une famille d'un seul point est bien sûr affinement indépendante) on aboutit nécessairement à la thèse. □

Définition 11.12 (**Dimension**). *La dimension d'un sous-ensemble non vide S de $\mathbb{R}^n$, notée* $\dim(S)$*, est la dimension de son enveloppe affine (au sens de la Définition 11.3).*

Définition 11.13 (**Hyperplan affine**). *Un hyperplan affine de $\mathbb{R}^n$ est un sous-espace affine de dimension $n-1$.*

Proposition 11.14. *Tout hyperplan affine de $\mathbb{R}^n$ est de la forme*

$$H \equiv H_{a,r} := \{x \in \mathbb{R}^n \mid \langle a, x\rangle = r\}$$

où $a \in \mathbb{R}^n_*$ et $r \in \mathbb{R}$.

Démonstration. Etant donné un hyperplan affine H, il suffit de choisir pour a un vecteur de norme unité dans $H^\perp$ et de poser $r := \mathrm{dist}(H, \{0\})$, où dist se réfère à la distance signée dans la direction de a. □

Un hyperplan affine de $\mathbb{R}^n$ détermine deux demi-espaces de $\mathbb{R}^n$:

$$H^+ \equiv H^+_{a,r} := \{x \in \mathbb{R}^n \mid \langle a, x\rangle \geq r\}$$

et

$$H^- \equiv H^-_{a,r} := \{x \in \mathbb{R}^n \mid \langle a, x\rangle \leq r\}\,.$$

On a

$$H = H^+ \cap H^-.$$

Chapitre 12

Ensembles convexes, polytopes et polyèdres

12.1 Propriétés algébriques

Définition 12.1 (Ensemble convexe). *Un sous-ensemble C de $\mathbb{R}^n$ est dit convexe si*

$$\forall x, y \in C,\ \forall \lambda \in [0,1], \quad \lambda x + (1-\lambda)y \in C.$$

Dans la suite, par abus de langage et lorsque cela n'amène aucune ambiguïté, on parlera simplement d'un convexe ou d'un convexe de $\mathbb{R}^n$ pour désigner un sous-ensemble convexe de $\mathbb{R}^n$.

D'un point de vue géométrique, un convexe est donc un ensemble qui, lorsqu'il contient deux points, contient nécessairement le segment les reliant.

Proposition 12.2. *On a les propriétés élémentaires suivantes :*

1. *Si C_1, C_2 sont deux convexes de $\mathbb{R}^n$ et λ_1, λ_2 deux réels, alors $\lambda_1 C_1 + \lambda_2 C_2$ est un convexe de $\mathbb{R}^n$.*
2. *Si $(C_j)_{j\in J}$ est une famille quelconque de convexes de $\mathbb{R}^n$, alors $\cap_{j\in J} C_j$ est un convexe de $\mathbb{R}^n$.*
3. *Si C est un convexe de $\mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}^m$ est une application affine, i.e. $f(x) = Ax + b$ pour certains $A \in \mathbb{R}^{m\times n}$ et $b \in \mathbb{R}^m$, alors $f(C)$ est un convexe de $\mathbb{R}^m$.*

Définition 12.3 (Combinaison convexe). *Soient $x_1, \cdots, x_k$ un nombre fini de points de $\mathbb{R}^n$, et $\lambda_1, \cdots, \lambda_k$ des réels tels que*

$$\lambda_j \geq 0 \quad \forall j = 1, \cdots, k \qquad et \qquad \sum_{j=1}^{k} \lambda_j = 1.$$

On dit que

$$x := \sum_{j=1}^{k} \lambda_j x_j$$

est une combinaison convexe des points $x_1, \cdots, x_k$.

Plus généralement, si $S \subseteq \mathbb{R}^n$ est un sous-ensemble quelconque, on dit que $x \in \mathbb{R}^n$ est une combinaison convexe de points de S s'il existe un nombre fini de points de S dont x soit une combinaison convexe.

Dans le cas particulier de deux points x_1 et x_2, toute combinaison convexe des x_1, x_2 peut s'écrire sous la forme

$$x = \lambda x_1 + (1 - \lambda)x_2, \qquad \lambda \in [0, 1],$$

qui intervient dans la Définition 12.1 ci-dessus.

Proposition 12.4. *Un sous-ensemble C de $\mathbb{R}^n$ est convexe si et seulement si il contient toutes les combinaisons convexes de points de C.*

Démonstration. Elle est en tout point similaire à celle correspondante de la Proposition 11.8. Cfr. TD 9. □

Lorsqu'un ensemble $S \subseteq \mathbb{R}^n$ n'est pas convexe, on peut considérer le plus petit ensemble convexe le contenant, c'est la notion importante suivante d'*enveloppe convexe*.

Définition 12.5 (**Enveloppe convexe**)**.** *L'enveloppe convexe d'un sous-ensemble S de $\mathbb{R}^n$ est l'intersection de tous les sous-ensembles convexes de $\mathbb{R}^n$ contenant S. L'enveloppe convexe de S est par conséquent le plut petit sous-ensemble convexe de $\mathbb{R}^n$ qui contienne S, on le note* $\mathrm{conv}(S)$.

Proposition 12.6. *On a la définition équivalente :* $\mathrm{conv}(S)$ *est l'ensemble de toutes les combinaisons convexes de points de S.*

Démonstration. Elle est très similaire à celle correspondante de la Proposition 11.10. Cfr. TD 9. □

Théorème 12.7 (**Carathéodory** bis)**.** *Soit $S \subseteq \mathbb{R}^n$. N'importe quel point de* $\mathrm{conv}(S)$ *peut s'écrire comme combinaison convexe de points de S affinement indépendants (et par conséquent en nombre inférieur $n + 1$).*

Démonstration. Soit $x = \sum_{j=1}^{k} \lambda_j x_j$ un point de $\mathrm{conv}(S)$, où sans perte de généralité on suppose que tous les λ_j sont strictement positifs. Si les x_j ne sont pas affinement indépendants, il existe des coefficients μ_j tels que :

$$\sum_{j=1}^{k} \mu_j x_j = 0 \qquad \text{avec } \sum_{j=1}^{k} \mu_j = 1.$$

En particulier, au moins un des coefficients μ_j est strictement positif. Pour $\alpha \geq 0$ quelconque, on peut donc aussi écrire

$$x = \sum_{j=1}^{k} (\lambda_j - \alpha\mu_j)\, x_j.$$

On choisit pour α la valeur

$$\alpha := \min\{\frac{\lambda_j}{\mu_j} \mid \mu_j > 0\}.$$

Par construction, au moins un des coefficients $\lambda_j - \alpha\mu_j$ est nul, les autres restant compris entre 0 (par définition de α) et 1 (puisque tous sont positifs et leur somme totale valant 1). On obtient ainsi une expression de x comme combinaison convexe d'au plus $k-1$ points de S. Si ces $k-1$ points ne sont pas affinement indépendants, on recommence la procédure ci-dessus, et ainsi de suite. Au bout d'un nombre fini d'étapes (au plus $k-1$ puisqu'une famille d'un seul point est bien sûr affinement indépendante) on aboutit nécessairement à la thèse. □

Les cônes convexes jouent un rôle particulier dans l'étude des convexes non bornés :

Définition 12.8. *Un sous-ensemble $\hat{C}$ de $\mathbb{R}^n$ est un cône convexe si*

$$\forall x_1, x_2 \in \hat{C},\ \forall \lambda_1, \lambda_2 \in \mathbb{R}^+, \quad \lambda_1 x_1 + \lambda_2 x_2 \in \hat{C}.$$

Définition 12.9 (Combinaison positive). *Soient $x_1, \cdots, x_k$ un nombre fini de points de $\mathbb{R}^n$, et $\lambda_1, \cdots, \lambda_k$ des réels positifs. On dit alors que*

$$x := \sum_{j=1}^{k} \lambda_j x_j$$

est une combinaison positive des points $x_1, \cdots, x_k$.

Plus généralement, si $S \subseteq \mathbb{R}^n$ est un sous-ensemble quelconque, on dit que $x \in \mathbb{R}^n$ est une combinaison positive de points de S s'il existe un nombre fini de points de S dont x soit une combinaison positive.

Proposition 12.10. *Un sous-ensemble $\hat{C}$ de $\mathbb{R}^n$ est un cône convexe si et seulement si il contient toutes les combinaisons positives de points de $\hat{C}$.*

Démonstration. Elle est très similaire à celle correspondante de la Proposition 11.8. Cfr. TD 9. □

Définition 12.11 (Enveloppe conique). *L'enveloppe conique d'un sous-ensemble S de $\mathbb{R}^n$ est l'ensemble de toutes les combinaisons positives de points de S. On la note* cône(S).

Proposition 12.12. *L'enveloppe conique d'un sous-ensemble S de $\mathbb{R}^n$ est un cône convexe, c'est le plus petit cône convexe contenant S.*

Démonstration. Elle est très similaire à celle correspondante de la Proposition 11.10. Cfr. TD 9. □

Théorème 12.13 (Carathéodory ter**).** *Soit $S \subseteq \mathbb{R}^n$. N'importe quel point de* cône(S) *peut s'écrire comme combinaison positive de points de S linéairement indépendants (et par conséquent en nombre inférieur n).*

Démonstration. Elle est très similaire à celle correspondante du Théorème 12.7 ci-dessus. Cfr. TD 9. □

12.2 Propriétés topologiques

Lemme 12.14. *On a les propriétés élémentaires suivantes :*

1. *Si C est un convexe de $\mathbb{R}^n$, son adhérence $\bar{C}$ est un convexe de $\mathbb{R}^n$.*
2. *Si C est un convexe de $\mathbb{R}^n$, son intérieur $\mathring{C}$ est un convexe de $\mathbb{R}^n$.*

Pour l'étude des ensembles convexes, il est commode d'adopter la notion de topologie relative : celle-ci qui n'est autre que la topologie au sens usuel, mais considérée dans le plus petit sous-espace affine comprenant l'ensemble en question, autrement dit dans son enveloppe affine.

Définition 12.15. *Soit S un sous-ensemble quelconque de $\mathbb{R}^n$, l'intérieur relatif de S est l'ensemble*

$$\mathrm{int_r}(S) := \{x \in S \mid \exists\, \varepsilon > 0, B(x,\varepsilon) \cap \mathrm{aff}(S) \subset S\}\,.$$

Bien sûr, on a toujours $\mathring{S} \subseteq \mathrm{int}_r(S)$. Pour un ensemble convexe, on a de plus

Lemme 12.16. *Si C est un convexe de $\mathbb{R}^n$, son intérieur relatif $\mathrm{int}_r(C)$ est un convexe de $\mathbb{R}^n$. De plus, si C n'est pas vide, alors $\mathrm{int}_r(C)$ non plus.*

Démonstration. Soient x_1, x_2 deux points de $\mathrm{int}_r(C)$ et $\lambda \in (0,1)$. Par hypothèse, il existe $\varepsilon_1 > 0$ tel que $B(x_1,\varepsilon_1) \cap \mathrm{aff}(C) \subset C$ et $\varepsilon_2 > 0$ tel que $B(x_2,\varepsilon_2) \cap \mathrm{aff}(C) \subset C$. Par inégalité triangulaire, et en posant $\varepsilon := \min(\varepsilon_1, \varepsilon_2)$, on montre facilement que

$$B(\lambda x_1 + (1-\lambda)x_2, \varepsilon) = \lambda B(x_1,\varepsilon) + (1-\lambda)B(x_2,\varepsilon)$$

et par conséquent

$$B(\lambda x_1 + (1-\lambda)x_2, \varepsilon) \cap \mathrm{aff}(C) \subset C.$$

Il s'ensuit que $\mathrm{int}_r(C)$ est convexe. Montrons maintenant qu'il est non vide, si C ne l'est pas. Soit $k := \dim(C)$. Si $k = 0$ alors C est réduit à un point, de même que $\mathrm{aff}(C)$, et la conclusion suit. Si $k \geq 1$, soient $x_1, \cdots, x_{k+1}$ $k+1$ points affinement indépendants dans C et soit

$$x := \sum_{j=1}^{k+1} \frac{1}{k+1} x_k \in C.$$

On prétend que $x \in \mathrm{int}_r(C)$. En effet, si $\varepsilon > 0$ et si $y \in B(x,\varepsilon) \cap \mathrm{aff}(C)$ alors puisque les vecteurs $x_2 - x_1, \cdots, x_{k+1} - x_1$ forment une base de $\mathrm{lin}(\mathrm{aff}(C))$, on peut décomposer

$$y - x = \sum_{j=2}^{k+1} \mu_j (x_j - x_j)$$

où pour une constante C positive (toutes les normes sont équivalents en dimension finie) $|\mu_j| \leq C\varepsilon$ pour chaque j. Dès lors,

$$y = \left(\frac{1}{k+1} - \sum_{j=2}^{k+1} \mu_j\right) x_1 + \sum_{j=2}^{k+1} \left(\frac{1}{k+1} + \mu_j\right) x_j.$$

Si ε est suffisamment petit, $(\frac{1}{k+1} - \sum_{j=2}^{k+1} \mu_j) \in (0,1)$ tout comme $(\frac{1}{k+1} + \mu_j)$ et on conclut que $y \in \mathrm{int}_r(C)$. La conclusion suit. □

Définition 12.17. *Soit S un sous-ensemble quelconque de $\mathbb{R}^n$, le bord relatif de S est l'ensemble*

$$\mathrm{bd}_r(S) := \bar{S} \setminus \mathrm{int}_r(S).$$

Lemme 12.18. *Soit C un convexe de $\mathbb{R}^n$, $x_1 \in \mathrm{int}_r(C)$ et $x_2 \in \bar{C}$. Alors quel que soit $\lambda \in (0,1)$, $\lambda x_1 + (1-\lambda)x_2 \in \mathrm{int}_r(C)$.*

Démonstration. Par hypothèse, il existe $\varepsilon > 0$ tel que $B(x,\varepsilon) \cap \mathrm{aff}(C) \subset C$. Supposons dans un premier temps que $x_2 \in C$. Par inégalité triangulaire, pour tout $\lambda \in (0,1)$ on obtient

$$B(\lambda x + (1-\lambda)y, \lambda\varepsilon) \subset \lambda B(x,\varepsilon) + (1-\lambda)\{y\} \subset C.$$

Il s'ensuit que $\lambda x + (1-\lambda)y \in \mathrm{int}_r(C)$. Supposons maintenant que $y \in \bar{C}$. Soit $(y_\ell)_{\ell\in\mathbb{N}}$ une suite dans C qui converge vers y. Par la première partie, on obtient que pour chaque ℓ,$B(\lambda x + (1-\lambda)y_\ell, \lambda\varepsilon) \subset C$. Mais pour ℓ suffisamment grand,

$$B(\lambda x + (1-\lambda)y, \frac{\lambda\varepsilon}{2}) \subset B(\lambda x + (1-\lambda)y_\ell, \lambda\varepsilon)$$

et par conséquent la conclusion suit également. □

Corollaire 12.19. *Soit C un convexe de $\mathbb{R}^n$, alors $\mathrm{int}_r(C) = \mathrm{int}_r(\bar{C})$.*

Démonstration. Si C est vide la conclusion est immédiate. Sinon, il suffit bien sûr de montrer l'inclusion du deuxième ensemble dans le premier, soit donc $x \in \mathrm{int}_r(\bar{C})$. Par le Lemme 12.16 il existe $x_1 \in \mathrm{int}_r(C)$. Par définition de $\mathrm{int}_r(\bar{C})$, pour ε suffisamment petit on a $x_2 := x + \varepsilon(x - x_1) \in \bar{C}$. Comme

$$x = \frac{1}{1+\varepsilon}\left(x + \varepsilon(x - x_1)\right) + (1 - \frac{1}{1+\varepsilon})x_1 = \lambda x_1 + (1-\lambda)x_2$$

avec $\lambda = 1 - \frac{1}{1+\varepsilon} \in (0,1)$, la conclusion suit du Lemme 12.18. □

12.3 Projection orthogonale

Théorème 12.20 (Projection orthogonale). *Soit C un convexe fermé non vide de $\mathbb{R}^n$ et $x \notin C$. Il existe un unique élément $p_C(x) \in C$ qui minimise la distance de x à C :*

$$\|x - p_C(x)\| = \min_{y\in C} \|x - y\|.$$

De plus, $p_C(x)$ est caractérisé par les propriétés

$$\begin{cases} p_C(x) \in C \\ \langle x - p_C(x), y - p_C(x)\rangle \le 0 \qquad \forall y \in C. \end{cases} \tag{12.1}$$

Démonstration. Soit $(x_k)_{k\in\mathbb{N}}$ une suite minimisante pour la distance de x à C, autrement dit $x_k \in C$ et

$$\lim_{k\to+\infty} \|x_k - x\| = \inf_{y\in C} \|x - y\| := \alpha.$$

On rappelle l'inégalité du parallélogramme pour tous vecteurs a, b de $\mathbb{R}^n$,

$$\|a+b\|^2 + \|a-b\|^2 = 2\left(\|a\|^2 + \|b\|^2\right),$$

que l'on applique à $a := x - x_k$ et $b := x - x_j$ (j, k quelconques dans $\mathbb{N}$). Ceci donne

$$4\|x - \frac{x_j + x_k}{2}\|^2 + \|x_j - x_k\|^2 = 2\left(\|x - x_j\|^2 + \|x - x_k\|^2\right).$$

Puisque C est convexe, et puisque $x_j, x_k \in C$, on a $(x_j + x_k)/2 \in C$ et il suit de la définition de α que

$$\|x - \frac{x_j + x_k}{2}\|^2 \geq \alpha^2.$$

Dès lors, on déduit que

$$0 \leq \|x_j - x_k\|^2 \leq 2\left(\|x - x_j\|^2 + \|x - x_k\|^2\right) - 4\alpha^2.$$

Prenant alors la limite lorsque j, k tendent simultanément vers $+\infty$, on déduit que la suite $(x_k)_{k\in\mathbb{N}}$ est de Cauchy. Par complétude de $\mathbb{R}^n$, celle-ci possède une limite que nous notons $p_C(x)$. Par fermeture de C, $p_C(x) \in C$. Par construction et par continuité de la norme

$$\|x - p_C(x)\| = \lim_{k\to+\infty} \|x - x_k\| = \alpha,$$

ce qui prouve l'existence de $p_C(x)$. L'unicité suit de la même manière de l'identité du parallélogramme. Venons en maintenant à la condition (12.1). Pour y quelconque dans C, et $\lambda \in [0,1]$ quelconque, on a $(1-\lambda)p_C(x) + \lambda y \in C$. Par conséquent,

$$\|x - ((1-\lambda)p_C(x) + \lambda y)\|^2 \geq \|x - p_C(x)\|^2.$$

Mais

$$\|x - ((1-\lambda)p_C(x) + \lambda y)\|^2 = \|x - p_C(x) - \lambda(y - p_C(x)))\|^2$$

et

$$\|x - p_C(x) - \lambda(y - p_C(x))x)\|^2 = \|x - p_C(x)\|^2 - 2\lambda\langle x - p_C(x), y - p_C(x)\rangle + \lambda^2\|y - p_C(x)\|^2.$$

Par conséquent, pour tout $\lambda \in [0,1]$,

$$-2\lambda\langle x - p_C(x), y - p_C(x)\rangle + \lambda^2\|y - p_C(x)\|^2 \geq 0.$$

En faisant tendre λ vers 0 par les positifs (auquel cas on peut diviser par λ), on déduit (12.1). De plus, si $p_C(x)$ vérifie (12.1) alors l'argument ci-dessus appliqué à $\lambda = 1$ montre que $\|x - y\| \geq \|x - p_C(x)\|$ quel que soit $y \in C$, et par conséquent (12.1) caractérise effectivement $p_C(x)$. □

Définition 12.21. *L'application $p_C : \mathbb{R}^n \to C$, $x \mapsto p_C(x)$ décrite dans le théorème précédent est appelée la projection orthogonale sur C.*

Proposition 12.22. *La projection orthogonale p_C sur un convexe fermé non vide vérifie l'inégalité*

$$\|p_C(x) - p_C(y)\| \leq \|x - y\| \qquad \textit{pour tous } x, y \in \mathbb{R}^n.$$

En particulier, p_C est lipschitzienne de constante de Lipschitz égale à 1.

Démonstration. On écrit

$$\|p_C(x) - p_C(y)\|^2 = \langle (p_C(x) - x) + (x - y) + (y - p_C(y)), p_C(x) - p_C(y)\rangle$$

et on développe

$$\begin{aligned}
&\langle (p_C(x) - x) + (x - y) + (y - p_C(y)), p_C(x) - p_C(y)\rangle \\
=&\langle x - y, p_C(x) - p_C(y)\rangle + \langle p_C(x) - x), p_C(x) - p_C(y)\rangle + \langle y - p_C(y), p_C(x) - p_C(y)\rangle \\
\leq&\langle x - y, p_C(x) - p_C(y) \\
\leq&\|x - y\| \, \|p_C(x) - p_C(y)\|.
\end{aligned}$$

La conclusion suit. □

12.4 Séparation et Hyperplan d'appui

On rappelle (cfr. Chapitre 11) qu'un hyperplan affine $H \subset \mathbb{R}^n$ détermine deux demi-espaces fermés que nous avons notés H^+ et H^-.

Définition 12.23 (Hyperplan d'appui). *Soit S un sous-ensemble de $\mathbb{R}^n$ et $x \in S$. On dit qu'un hyperplan affine H de $\mathbb{R}^n$ est un hyperplan d'appui pour S en x si $x \in H$ et si S est entièrement contenu dans un des deux demi-espaces déterminés par H.*

Par extension, on dira qu'un hyperplan affine H de $\mathbb{R}^n$ est un hyperplan d'appui pour S si il existe $x \in S$ tel que H soit un hyperplan d'appui pour S en x. Le demi-espace déterminé par H et contenant S est appelé un demi-espace de support de S.

Proposition 12.24. *Soit C un convexe fermé de $\mathbb{R}^n$ et $x \in \mathbb{R}^n \setminus C$. Alors l'hyperplan défini par*

$$H := \{y \in \mathbb{R}^n \mid \langle x - p_C(x), y\rangle = \langle x - p_C(x), p_C(x)\rangle\},$$

où p_C désigne la projection orthogonale sur C est un hyperplan d'appui pour C en $p_C(x)$. Le demi-espace

$$H^- := \{y \in \mathbb{R}^n \mid \langle x - p_C(x), y\rangle \leq \langle x - p_C(x), p_C(x)\rangle\},$$

est un demi-espace de support pour C, de plus il ne contient pas x.

Démonstration. Il est immédiat que $p_C(x) \in H$. Pour montrer que H est un hyperplan d'appui pour C en $p_C(x)$ il suffit donc de montrer que C est entièrement contenu dans H^-. Mais pour $y \in C$, on a en vertu de Théorème 12.20

$$\langle x - p_C(x), y - p_C(x)\rangle \leq 0,$$

et par conséquent

$$\langle x - p_C(x), y\rangle \leq \langle x - p_C(x), p_C(x)\rangle,$$

ce qui montre que $y \in H^-$. □

Définition 12.25 (**Hyperplan d'appui propre**). *Soit S un sous-ensemble de $\mathbb{R}^n$ et $x \in S$. On dit qu'un hyperplan affine H de $\mathbb{R}^n$ est un hyperplan d'appui propre pour S en x si $x \in H$ et si S est entièrement contenu dans un des deux demi-espaces déterminés par H, sans être toutefois entièrement contenu dans H.*

Proposition 12.26 (**Existence d'un hyperplan d'appui propre**). *Soit C un convexe fermé de $\mathbb{R}^n$ et $x \in \mathrm{bd}_r(C)$. Alors il existe un hyperplan d'appui propre pour C en x.*

Démonstration. Puisque $x \in \mathrm{bd}_r(C)$, il existe une suite $(x_k)_{k\in\mathbb{N}}$ dans $\mathrm{aff}(C) \setminus C$ qui converge vers x lorsque $k \to +\infty$. Par la Proposition 12.24, pour chaque k l'hyperplan

$$H_k := \{y \in \mathbb{R}^n \mid \langle x_k - p_C(x_k), y\rangle = \langle x_k - p_C(x_k), p_C(x_k)\rangle\},$$

est un hyperplan d'appui pour C en $p_C(x_k)$. Remarquons que puisque $x_k \notin C$, $x_k \neq p_C(x_k)$ et par conséquent on peut récrire H_k comme

$$H_k := \left\{y \in \mathbb{R}^n \mid \langle \frac{x_k - p_C(x_k)}{\|x_k - p_C(x_k)\|}, y\rangle = \langle \frac{x_k - p_C(x_k)}{\|x_k - p_C(x_k)\|}, p_C(x_k)\rangle\right\}.$$

Quitte à passer à une sous-suite si nécessaire, on peut supposer que

$$\frac{x_k - p_C(x_k)}{\|x_k - p_C(x_k)\|} \to a \in S^{n-1} \qquad \text{lorsque } k \to +\infty.$$

Par ailleurs, par continuité de p_C (p_C est même lipschitzienne),

$$p_C(x_k) \to p_C(x) = x \qquad \text{lorsque } k \to +\infty.$$

On prétend que

$$H := \{y \in \mathbb{R}^n \mid \langle a, y\rangle = \langle a, x\rangle\}$$

est un hyperplan d'appui propre pour C en x. Premièrement, il est clair que $x \in H$. Ensuite, pour tout $y \in C$ on a

$$\langle a, y\rangle = \lim_{k\to+\infty} \langle \frac{x_k - p_C(x_k)}{\|x_k - p_C(x_k)\|}, y\rangle \leq \langle \frac{x_k - p_C(x_k)}{\|x_k - p_C(x_k)\|}, p_C(x_k)\rangle = \langle a, x\rangle,$$

où on a utilisé le fait que H_k est un hyperplan d'appui pour C en $p_C(x_k)$. Il s'ensuit que $C \subset H^-$ et donc que H est un hyperplan d'appui pour C en x. Finalement, H est un hyperplan d'appui propre. En effet, par construction $H \perp a \in \mathrm{aff}(C)$ et par conséquent si on avait $C \subset H$ alors $H \cap \mathrm{aff}(C)$ serait une sous-espace affine contenant C et strictement inclus dans $\mathrm{aff}(C)$, ce qui serait incompatible avec le fait que $\mathrm{aff}(C)$ soit le plus petit sous-espace affine contenant C. □

Définition 12.27 (**Séparation et séparation stricte**). *Soient S et T deux sous-ensembles de $\mathbb{R}^n$.*

- *On dit que l'hyperplan affine H sépare S et T au sens large si $S \subseteq H^-$ et $T \subseteq H^+$, ou inversement.*
- *On dit que l'hyperplan affine $H \equiv H_{a,r}$ sépare S et T au sens strict s'il existe $\varepsilon > 0$ tel que $S \subseteq H^-_{a,r-\varepsilon}$ et $T \subseteq H^+_{a,r+\varepsilon}$, ou inversement.*

Théorème 12.28 (**Séparation stricte d'un convexe et d'un point**). *Soit C un convexe fermé non vide de $\mathbb{R}^n$ et $x \notin C$. Il existe un hyperplan affine qui sépare C et x au sens strict, autrement dit*

$$\exists a \in \mathbb{R}^n_*, \exists r \in \mathbb{R},\ \langle a, x\rangle > r > \sup_{y\in C}\langle a, y\rangle.$$

Démonstration. Il suffit de choisir $a := x - p_C(x)$ et $r := \langle a, \frac{x+p_C(x)}{2}\rangle$, de sorte que l'hyperplan affine en question passe par le point milieu du segment joignant x et $p_C(x)$ et est orthogonal à ce segment. Les détails sont très similaires à ceux de la preuve de la Proposition 12.24. □

Proposition 12.29 (**Séparation au sens large de deux convexes**). *Soit C_1 et C_2 deux convexes disjoints de $\mathbb{R}^n$. Alors il existe un hyperplan affine H qui sépare C_1 et C_2 au sens large.*

Démonstration. Par hypothèse, $0 \notin C := C_1 - C_2$. Comme C est convexe comme différence (au sens vectoriel et non ensembliste !) de deux convexes, on déduit du Corollaire 12.19 que $0 \notin \mathrm{int}_r(\bar{C})$. Si $0 \notin \bar{C}$, alors par la Théorème 12.28 il existe un hyperplan affine qui sépare $\bar{C}$ et 0 au sens strict (en particulier au sens large). Si inversement $0 \in \bar{C}$, alors comme par ailleurs $0 \notin \mathrm{int}_r(\bar{C})$ on a nécessairement $0 \in \mathrm{bd}_r(\bar{C})$, et il s'ensuit de la Proposition 12.26 qu'il existe un hyperplan d'appui (propre) H pour $\bar{C}$ en 0. Dans un cas comme dans l'autre, il existe donc $a \in \mathbb{R}^n \setminus \{0\}$ tel que

$$\langle a, x_1\rangle \leq \langle a, x_2\rangle$$

quels que soient $x_1 \in C_1$ et $x_2 \in C_2$. Dès lors, l'hyperplan $H_{a,r}$ où $r := \sup_{x_1\in C_1}\langle a, x_1\rangle < +\infty$ sépare C_1 et C_2 au sens large. □

12.5 Représentation extérieure

Comme corollaire de la Proposition 12.24 nous obtenons :

Théorème 12.30. *Tout sous ensemble convexe fermé non vide de $\mathbb{R}^n$ est égal à l'intersection de ses demi-espaces de support.*

Démonstration. Puisque tout demi-espace de support de C contient C, il suit que C est inclus dans l'intersection de ses demi-espaces de support. D'autre part, si $x \notin C$, par la Proposition 12.24 le demi-espace défini par

$$\{y \in \mathbb{R}^n \mid \langle x - p_C(x), y\rangle \leq \langle x - p_C(x), p_C(x)\rangle,$$

où p_C désigne la projection orthogonale sur C, est un demi-espace de support de C qui ne contient pas x. D'où l'inclusion inverse. □

12.6 Faces, Points extrémaux, Cône de récession

Définition 12.31 (**Face**). *Soit C un sous-ensemble convexe fermé de $\mathbb{R}^n$ et $F \subseteq C$ convexe fermé lui aussi. On dit que F est une face de C si quels que soient $x_1, x_2 \in C$ et $\lambda \in (0,1)$, si $\lambda x_1 + (1-\lambda)x_2 \in F$ alors nécessairement $x_1 \in F$ et $x_2 \in F$.*

Définition 12.32 (**Point extrémal**). *Un point extrémal d'un convexe C de $\mathbb{R}^n$ est une face réduite à un point. Autrement dit, $x \in C$ est un point extrémal si et seulement si quels que soient $x_1, x_2 \in C$ et $\lambda \in (0,1)$, si $\lambda x_1 + (1-\lambda)x_2 = x$ alors nécessairement $x_1 = x_2 = 0$.*

Définition 12.33 (**Demi-droite et direction extrémale**). *Une demi-droite*

$$F = \{x_0 + \lambda z_0 \mid \lambda \geq 0\},$$

où $x_0, z_0 \in \mathbb{R}^n$ sont fixés est appelée une demi-droite extrémale du convexe $C \subseteq \mathbb{R}^n$ si F est une face de C. On dit alors que z_0 est une direction extrémale de C.

Définition 12.34 (**Cône de récession**). *Soit C un sous-ensemble convexe fermé de $\mathbb{R}^n$. Le cône de récession de C est l'ensemble*

$$\mathrm{rec}(C) := \{z \in \mathbb{R}^n \mid \forall x \in C,\, \forall \lambda \geq 0,\; x + \lambda z \in C\}\,.$$

Lemme 12.35. *Soit C un sous-ensemble convexe fermé de $\mathbb{R}^n$, alors* $\mathrm{rec}(C)$ *est un cône convexe fermé de $\mathbb{R}^n$ et on a l'égalité*

$$\mathrm{rec}(C) := \{z \in \mathbb{R}^n \mid \exists x \in C,\, \forall \lambda \geq 0,\; x + \lambda z \in C\}\,.$$

De plus, C est borné si et seulement si $\mathrm{rec}(C)$ *est réduit à zéro.*

Démonstration. Par définition on a

$$\mathrm{rec}(C) = \bigcap_{x \in C, \lambda > 0} \frac{C - \{x\}}{\lambda}.$$

Comme une intersection quelconque de fermés est fermée, et qu'une intersection quelconque de convexes est convexe, $\mathrm{rec}(C)$ est un convexe fermé. Supposons ensuite que $x_0 \in C$ et $z_0 \in \mathbb{R}^n$ soient tels que

$$\forall \lambda \geq 0,\; x_0 + \lambda z_0 \in C.$$

Soit $x \in C$ et $\lambda \geq 0$ quelconques. Pour tout $\varepsilon > 0$ on a

$$x + \lambda z = x + \lambda z + \varepsilon(x_0 - x) - \varepsilon(x_0 - x) = \varepsilon(x_0 + \frac{\lambda}{\varepsilon} z) + (1-\varepsilon)x - \varepsilon(x_0 - x).$$

Comme $x_0 + \frac{\lambda}{\varepsilon} z \in C$ et $x \in C$ on a

$$x + \lambda z + \varepsilon(x_0 - x) - \varepsilon(x_0 - x) = \varepsilon(x_0 + \frac{\lambda}{\varepsilon} z) + (1-\varepsilon)x \in C.$$

Comme C est fermé par hypothèse, et que $\varepsilon(x_0 - x)$ tend vers 0 lorsque ε tend vers 0, on déduit que $x + \lambda z \in C$. D'où l'égalité annoncée dans l'énoncé. Il reste à démontrer que C

est borné si son cône de récession est réduit à 0 (l'implication inverse étant immédiate). Si C n'était pas borné, il existerait dans C un suite de segments $\{\lambda y_0 + (1-\lambda)y_n \mid \lambda \in [0,1]\}$ avec $\|y_n - y_0\| \to +\infty$ quand $n \to +\infty$. Par compacité de la sphère unité de $\mathbb{R}^n$, on peut supposer, modulo un passage à une sous-suite, que

$$\frac{y_n - y_0}{\|y_n - y_0\|} \to z_1 \qquad \text{lorsque } n \to +\infty,$$

pour un certain $z_1 \in \mathbb{R}^n \setminus \{0\}$. Par fermeture de C, on déduit alors que

$$\{y_0 + \lambda z_1 \mid \lambda \geq 0\} \subseteq C$$

et par conséquent que $z_1 \in \mathrm{rec}(C) \setminus \{0\}$. □

Définition 12.36 (**Espace linéaire d'un convexe**). *Soit C un sous-ensemble convexe fermé de $\mathbb{R}^n$, l'espace linéaire de C est l'ensemble*

$$\mathrm{lin}(C) := \mathrm{rec}(C) \cap (-\mathrm{rec}(C)) = \{z \in \mathbb{R}^n \mid \forall x \in C,\, \forall \lambda \in \mathbb{R},\, x + \lambda z \in C\}\,.$$

C'est un sous-espace vectoriel de $\mathbb{R}^n$.

Définition 12.37. *On dit qu'un convexe fermé de $\mathbb{R}^n$ ne contient aucune droite si son espace linéaire est réduit à zéro.*

Proposition 12.38. *Soit C un sous-ensemble convexe fermé de $\mathbb{R}^n$, alors*

$$C = \mathrm{lin}(C) + \left(C \cap (\mathrm{lin}(C))^{\perp}\right).$$

De plus, $C \cap (\mathrm{lin}(C))^{\perp}$ est un convexe fermé ne contenant aucune droite.

Démonstration. Soit p la projection orthogonale de $\mathbb{R}^n$ sur $\mathrm{lin}(C)$. Alors tout $x \in C$ se décompose comme

$$x = p(x) + (x - p(x))$$

et il suit de la caractérisation (12.1) de $p(x)$, et du caractère linéaire (et pas seulement convexe) de $\mathrm{lin}(C)$, que $x - p(x) \in C \cap (\mathrm{lin}(C))^{\perp}$. □

12.7 Représentation intérieure

Théorème 12.39 (**Minkowski**). *Tout convexe fermé et borné de $\mathbb{R}^n$ est égal à l'enveloppe convexe de ses points extrémaux.*

Nous déduisons le Théorème de Minkowski du théorème plus général suivant :

Théorème 12.40. *Tout convexe fermé de $\mathbb{R}^n$ ne contenant aucune droite est égal à l'enveloppe convexe de ses points extrémaux et de ses demi-droites extrémales.*

Démonstration. La démonstration se fait par récurrence sur la dimension du convexe en question. Remarquons d'abord qu'un convexe de dimension zéro est réduit à un point, ce point est nécessairement extrémal et le convexe est donc égal à l'enveloppe convexe de lui-même ! Ensuite, un convexe fermé de dimension un est soit *a)* un segment, auquel cas il est égal à l'enveloppe convexe de ses deux extrémités (qui sont des points extrémaux) ; *b)* une demi-droite fermée, auquel cas il est égal à l'enveloppe convexe de ses demi-droites extrémales (il n'en a qu'une, égale à lui-même) ; *c)* une droite, mais ce dernier cas est exclu par hypothèse. Supposons maintenant avoir montré la thèse pour tout sous-ensemble convexe fermé de $\mathbb{R}^n$ de dimension au plus k, $k \geq 1$, et considérons un sous-ensemble convexe fermé C de $\mathbb{R}^n$ de dimension $k+1$. Soit $x \in C$. On distingue deux cas.

i) Cas où $x \in \mathrm{bd}_r(C)$. Dans ce cas, par la Proposition 12.26, il existe un hyperplan d'appui propre H pour C en x. L'intersection $C \cap H$ est un convexe fermé de dimension au plus k de $\mathbb{R}^n$. Par notre hypothèse de récurrence, x est donc contenu dans l'enveloppe convexe des points extrémaux et des demi-droites extrémales de $C \cap H$. Mais comme H est un hyperplan d'appui, les points extrémaux et les demi-droites extrémales de $C \cap H$ sont aussi des points extrémaux et des demi-droites extrémales de C (vérifier !). La conclusion suit dans ce cas.

ii) Cas où $x \in \mathrm{int}_r(C)$. Par hypothèse sur k, $E := \mathrm{lin}(\mathrm{aff}(C))$ est un espace vectoriel de dimension supérieure ou égale à deux. Par hypothèse sur C, $\mathrm{lin}(C) = \{0\}$, et donc $\mathrm{rec}(C) \cap (-\mathrm{rec}(C)) = \{0\}$. Par conséquent,

$$(\mathrm{rec}(C) \cap \{e \in E \mid \|e\| = 1\}) \cap (-\mathrm{rec}(C) \cap \{e \in E \mid \|e\| = 1\}) = \emptyset.$$

Comme d'autre part les deux ensembles ci-dessus sont fermés, et que la sphère unité de E est connexe (E est de dimension supérieure à deux), on déduit que

$$\mathrm{rec}(C) \cup (-\mathrm{rec}(C)) \neq E.$$

Soit $z \in E \setminus (\mathrm{rec}(C) \cup (-\mathrm{rec}(C)))$. Par construction, l'ensemble

$$\tilde{C} := C \cap \{x + \lambda z \mid \lambda \in \mathbb{R}\}$$

est un convexe borné de dimension 1 ($x \in \mathrm{int}_r(C)$). Dès lors c'est un segment. On peut ainsi écrire x comme combinaison convexe des extrémités (appelons les x_1 et x_2) de ce segment. Puisque $x_1, x_2 \in \mathrm{bd}_r(C)$, par le Cas i) x_1 et x_2 sont contenues dans l'enveloppe convexe des points extrémaux et des demi-droites extrémales de C. Comme ce dernier ensemble est un convexe, il contient donc x lui aussi. □

12.8 Polytopes et Cônes finiment engendrés

Définition 12.41 (Polytope). *Un polytope dans $\mathbb{R}^n$ est l'enveloppe convexe d'un nombre fini de points de $\mathbb{R}^n$.*

Définition 12.42 (Simplexe). *Un simplexe dans $\mathbb{R}^n$ est l'enveloppe convexe d'un nombre fini de points de $\mathbb{R}^n$ affinement indépendants.*

Proposition 12.43. *Tout polytope de $\mathbb{R}^n$ peut s'écrire comme une union finie de simplexes de $\mathbb{R}^n$.*

Démonstration. C'est une conséquence directe du Théorème de Carathéodory version bis (Théorème 12.7). □

Proposition 12.44. *Tout polytope de $\mathbb{R}^n$ est fermé et borné.*

Démonstration. Si $P = \text{conv}(\{x_1, \cdots, x_k\})$, alors P est l'image par l'application linéaire (de $\mathbb{R}^k$ dans $\mathbb{R}^n$)

$$(\lambda_1, \cdots, \lambda_k) \longmapsto \sum_{j=1}^{j} \lambda_j x_j$$

du simplexe standard de $\mathbb{R}^k$

$$S_k := \{(\lambda_1, \cdots, \lambda_k) \in \mathbb{R}^k \mid \sum_{j=1}^{k} \lambda_j = 1,\ \lambda_j \geq 0\, \forall j \in \{1, \cdots, k\}\}.$$

Comme S_k est un fermé borné de $\mathbb{R}^k$ et que l'image d'un fermé borné par une application continue est un fermé borné, la conclusion suit. □

Définition 12.45 (Cône finiment engendré). *Un cône finiment engendré dans $\mathbb{R}^n$ est l'enveloppe conique d'un nombre fini de points de $\mathbb{R}^n$.*

Définition 12.46 (Cône simplicial). *Un cône simplicial dans $\mathbb{R}^n$ est l'enveloppe conique d'un nombre fini de points de $\mathbb{R}^n$ linéairement indépendants.*

Proposition 12.47. *Tout cône finiment engendré de $\mathbb{R}^n$ peut s'écrire comme une union finie de cônes simpliciaux de $\mathbb{R}^n$*

Démonstration. Il s'agit cette fois d'une conséquence directe du Théorème de Carathéodory version ter (Théorème 12.13). □

Corollaire 12.48. *Tout cône finiment engendré de $\mathbb{R}^n$ est fermé.*

Démonstration. Contrairement à la démonstration de la Proposition 12.44, on ne peut se baser ici sur la compacité. Toutefois, en vertu de la Proposition 12.47, il suffit de montrer que tout cône simplicial est fermé (car une union finie de fermés est fermée). Soit donc cône$(\{x_1, \cdots, x_k\})$ un cône simplicial dans $\mathbb{R}^n$, avec $x_1, \cdots, x_k$ linéairement indépendants. L'application linéaire (de $\mathbb{R}^k$ dans $\mathbb{R}^n$)

$$(\lambda_1, \cdots, \lambda_k) \longmapsto \sum_{j=1}^{j} \lambda_j x_j$$

est par conséquent injective. Son inverse (bien définie sur son image) est linéaire et donc continue ; il suffit dès lors de remarquer que cône$(\{x_1, \cdots, x_k\})$ est l'image réciproque, par cette application inverse, du fermé $\{(\lambda_1, \cdots, \lambda_k) \in \mathbb{R}^k \mid \lambda_j \geq 0\, \forall j \in \{1, \cdots, k\}\}$ de $\mathbb{R}^k$. Comme l'image réciproque d'un fermé par une application continue est un fermé, la conclusion suit. □

12.9 Lemme de Farkas

Théorème 12.49 (Lemme de Farkas). *Soient $a_1, \cdots, a_k$ des points de $\mathbb{R}^n$ et $b \in \mathbb{R}^n$. Alors*

$$\cap_{j=1}^k \{x \in \mathbb{R}^n \mid \langle a_j, x\rangle \leq 0\} \subseteq \{x \in \mathbb{R}^n \mid \langle b, x\rangle \leq 0\}$$

si et seulement si

$$b \in \text{cône}(\{a_1, \cdots, a_k\}).$$

Démonstration. Commençons par la condition suffisante, la plus aisée. Si

$$b = \sum_{j=1}^k \beta_j a_j \in \text{cône}(\{a_1, \cdots, a_k\}),$$

avec les β_j tous positifs, et si

$$x_0 \in \cap_{j=1}^k \{x \in \mathbb{R}^n \mid \langle a_j, x\rangle \leq 0\},$$

alors

$$\langle b, x_0\rangle = \sum_{j=1}^k \beta_j \langle a_j, x_0\rangle \leq 0,$$

de sorte que $x_0 \in \{x \in \mathbb{R}^n \mid \langle b, x\rangle \leq 0\}$.

Passons à la condition nécessaire, où plutôt à sa contraposée.
Supposons que $b \notin \hat{C} := \text{cône}(\{a_1, \cdots, a_k\})$. Comme $\hat{C}$ est un convexe fermé (Corollaire12.48), par le théorème de séparation stricte (Théorème 12.28) il existe $a \in \mathbb{R}^n$ et $r \in R$ tels que

$$\sup_{y \in \hat{C}} \langle a, y\rangle < r < \langle a, b\rangle.$$

Comme $\hat{C}$ est un cône, pour tout $\lambda > 0$

$$r > \sup_{y \in \hat{C}} \langle a, y\rangle = \sup_{\lambda y \in \hat{C}} \langle a, \lambda y\rangle = \lambda \sup_{y \in \hat{C}} \langle a, y\rangle.$$

Il s'ensuit que $\sup_{y \in \hat{C}} \langle a, y\rangle = 0$. Par conséquent, comme chaque $a_j \in \hat{C}$,

$$a \in \cap_{j=1}^k \{x \in \mathbb{R}^n \mid \langle a_j, x\rangle \leq 0\}.$$

Mais $\langle b, a\rangle > r \geq 0$, de sorte que

$$a \notin \{x \in \mathbb{R}^n \mid \langle b, x\rangle \leq 0\}.$$

Ceci termine la démonstration. □

12.10 Polyèdres

Définition 12.50 (Polyèdre). *Un polyèdre dans $\mathbb{R}^n$ est un ensemble de la forme*

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

où $A \in \mathbb{R}^{m\times n}(\mathbb{R})$ et $b \in \mathbb{R}^m$ sont fixés.

Soit $P := \{x \in \mathbb{R}^n \mid Ax \leq b\}$ un polyèdre de $\mathbb{R}^n$. Pour $j \in \{1, \cdots, m\}$, notons A^j la j-ième ligne de A (considérée comme un élément de $\mathbb{R}^n$) et b^j la j-ième composante de b. On peut dès lors décrire de manière équivalente

$$P := \left\{x \in \mathbb{R}^n \mid \langle A^j, x\rangle \leq b^j, \quad \forall j \in \{1, \cdots, m\}\right\}.$$

Autrement dit, un sous-ensemble de $\mathbb{R}^n$ est un polyèdre si et seulement si c'est une intersection *finie* de demi-espaces fermés de $\mathbb{R}^n$. En particulier, tout polyèdre dans $\mathbb{R}^n$ est un convexe fermé.

Proposition 12.51. *Soit P un polyèdre dans $\mathbb{R}^n$ (décrit par A et b) et $F \subseteq P$ un sous-ensemble de dimension k. Les deux affirmations suivantes sont équivalentes :*

1. *F est une face de P.*
2. *Il existe $n-k$ lignes linéairement indépendantes de A, notées $\{A^j\}_{j\in J}$ avec $\sharp J = n-k$, telles que*
$$F = \left\{x \in P \mid \langle A^j, x\rangle = b^j \quad \forall\, j \in J\right\}.$$

Démonstration. Commençons par la condition suffisante, la plus simple. Montrons à cet effet que quel que soit le sous-ensemble J de $\{1, \cdots, m\}$, l'ensemble $G := \{x \in P \mid \langle A^j, x\rangle = b^j \quad \forall\, j \in J\}$ est une face de P. De fait, si x_1 et x_2 sont des éléments de P et $\lambda \in (0,1)$ sont tels que $\lambda x_1 + (1-\lambda)x_2 \in G$, alors pour $i \in G$ on a à la fois $\langle A^i, x_1\rangle \leq b^i$ (car $x_1 \in P$), $\langle A^i, x_2\rangle \leq b^i$ (car $x_2 \in P$) et $\lambda\langle A^i, x_1\rangle + (1-\lambda)\langle A^i, x_2\rangle = b^i$ (car $x \in G$). Ceci implique que $\langle A^i, x_1\rangle = \langle A^i, x_2\rangle = b^i$, et par conséquent que $x_1 \in G$ et $x_2 \in G$, d'où la conclusion.

Venons en à la condition nécessaire. Soit F une face de P de dimension k. Soit $x_0 \in \mathrm{int}_r(F)$ fixé (nous avons que celui-ci est non-vide puisque F l'est). On désigne par I l'ensemble des indices i de lignes de A pour lesquelles $\langle A^i, x_0\rangle = b^i$. Soit $V := \mathrm{Vect}(\{A^i\}_{i\in I})$. Pour $v \in V^\perp$ quelconque et pour tout $i \in I$, on a

$$\langle A^i, x_0 + v\rangle = \langle A^i, x\rangle = b^i.$$

D'autre part, pour tout $i \notin I$ on a $\langle A^i, x_0\rangle < b^i$ et par conséquent, par continuité,

$$\exists \delta > 0 \mid \langle A^i, x_0 + v\rangle < b^i, \quad \forall i \notin I,\ \forall v \in V^\perp \text{ t.q. } \|v\| \leq \delta.$$

Il s'ensuit que

$$x_0 + v \in P \quad \forall v \in V^\perp \text{ t.q. } \|v\| \leq \delta. \tag{12.2}$$

D'autre part, puisque $x_0 = (x_0+v)/2 + (x_0-v)/2$ et que F est une face, il s'ensuit aussi que

$$x_0 + v \in F \quad \forall v \in V^\perp \text{ t.q. } \|v\| \leq \delta. \tag{12.3}$$

Dès lors, comme F est de dimension k, $\dim(V^\perp) \leq k$, et aussi $\dim(V) \geq n-k$. Supposons par l'absurde que $\dim(V) > n-k$. Si $E := \mathrm{lin}(\mathrm{aff}(F))$ désigne l'espace vectoriel associé à l'enveloppe affine de F, alors $\dim(E) = k$ et par conséquent $V \cap E \neq \{0\}$. Soit $0 \neq w = \sum_{i\in I} \alpha_i A^i \in V \cap E$. Comme $x_0 \in \mathrm{int}_r(F)$ par hypothèse, on a $x_0 + \varepsilon w \in F \subset P$ pour tout $\varepsilon \in \mathbb{R}$ suffisamment petit en valeur absolue, et en particulier, pour de tels ε

$$b^j \geq \langle A^j, x_0 + \varepsilon \sum_{i\in I} \alpha_i a_i \rangle = b^j + \langle A^j, \varepsilon \sum_{i\in I} \alpha_i a_i \rangle$$

pour tout $j \in I$, et donc, en variant le signe de ε,

$$0 = \langle A^j, \varepsilon \sum_{i\in I} \alpha_i A^i \rangle$$

pour tout $j \in I$. Par sommation, on obtient alors

$$0 = \sum_{j\in I} \alpha_j \langle A^j, \sum_{i\in I} \alpha_i A^i \rangle = \|w\|^2,$$

ce qui est absurde. D'où $\dim(V) = n-k$ et $E = V^\perp$. Montrons maintenant que

$$F = G := \left\{ x \in P \mid \langle A^i, x \rangle = b^i \quad \forall\, i \in I \right\}.$$

Soit $x \in F \cup G$ avec $x \neq x_0$. On peut écrire

$$x_0 = \lambda \left(x_0 + \frac{\delta}{\|x_0 - x\|}(x_0 - x) \right) + (1-\lambda)x$$

où

$$\lambda := \left(1 + \frac{\delta}{\|x - x_0\|} \right)^{-1} \in (0,1).$$

Comme $x - x_0 \in E = V^\perp$, on a $x_0 + \frac{\delta}{\|x_0 - x\|}(x_0 - x) \in P$ par (12.2). Mais comme F (par hypothèse) et G (par la première partie) sont des faces de P, on déduit que $x \in F \cap G$, et donc que $F = G$.

Pour terminer, on remarque que puisque $\dim(V) = n-k$ (et $\emptyset \neq F \ni x_0$),

$$\left\{ x \in P \mid \langle A^i, x \rangle = b^i \quad \forall\, i \in I \right\} = \left\{ x \in P \mid \langle A^j, x \rangle = b^j \quad \forall\, j \in J \right\},$$

où J désigne n'importe quel sous-ensemble de I correspondant à $n-k$ lignes linéairement indépendantes de A. □

Corollaire 12.52. *Un polyèdre de $\mathbb{R}^n$ possède au plus un nombre fini de faces. En particulier, il contient un nombre fini de points extrémaux et de demi-droites extrémales.*

Démonstration. Il suit de la proposition 12.51 que pour $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$, où $A \in \mathbb{R}^{m\times n}(\mathbb{R})$ et $b \in \mathbb{R}^m$, le nombre de faces de dimension k de P est majoré par le nombre de combinaisons de $n-k$ lignes parmi m, soit $m!/((n-k)!(m-n+k)!)$. □

Définition 12.53 (Cône polyédrique). *Un cône polyédrique dans $\mathbb{R}^n$ est un ensemble de la forme*

$$\hat{C} := \{x \in \mathbb{R}^n \mid Ax \leq 0\},$$

où $A \in \mathbb{R}^{m\times n}(\mathbb{R})$.

La démonstration de la proposition suivante est immédiate.

Proposition 12.54. *Un cône polyédrique est un cône convexe fermé.*

12.11 Correspondances polyèdres/polytopes

Les polytopes et les cônes finiment engendrés correspondent à une description intérieure de certains convexes (en tant que combinaisons convexes ou de combinaisons positives d'un nombre fini de points). A l'inverse, les polyèdres et les cônes polyédriques correspondent à une description extérieure (en tant qu'intersection d'un nombre fini de demi-espaces). Dans cette section, on montre que ces différentes notions coïncident cependant.

Théorème 12.55. *Tout polyèdre borné est un polytope.*

Démonstration. C'est une conséquence directe du Théorème de Minkowski (Théorème 12.39) et du Corollaire 12.52. □

Plus généralement, on a

Théorème 12.56. *Soit $P \subseteq \mathbb{R}^n$ un polyèdre. Il existe des ensembles de cardinal fini E et D dans $\mathbb{R}^n$ tels que*

$$P = \mathrm{conv}(E) + \mathrm{cône}(D).$$

Si de plus P ne contient aucune droite, alors on peut choisir pour E l'ensemble des points extrémaux de P et pour D l'ensemble des directions extrémales de P.

Démonstration. On comment par écrire, au moyen de la Proposition 12.38,

$$P = \mathrm{lin}(P) + \left(P \cap (\mathrm{lin}(P))^\perp\right)$$

où $P \cap (\mathrm{lin}(P))^\perp$ ne contient aucune droite. Soit $(e_1, \cdots, e_\ell)$ une base de $\mathrm{lin}(P)$. Alors si

$$P = \left\{x \in \mathbb{R}^n \mid \langle A^j, x\rangle \leq b^j, \quad \forall j \in \{1, \cdots, m\}\right\}.$$

on a

$$P \cap (\mathrm{lin}(P))^\perp = \left\{x \in \mathbb{R}^n \mid \left\{ \begin{array}{ll} \langle A^j, x\rangle \leq b^j, & \forall j \in \{1, \cdots, m\} \\ \langle e_j, x\rangle \leq 0, & \forall j \in \{1, \cdots, \ell\} \\ \langle -e_j, x\rangle \leq 0, & \forall j \in \{1, \cdots, \ell\} \end{array} \right. \right\}$$

et ce dernier est donc un polyèdre dans $\mathbb{R}^n$ ne contenant aucune droite. Soient $\{x_1, \cdots, x_k\}$ les points extrémaux de $P \cap (\mathrm{lin}(P))^\perp$ et $\{z_1, \cdots, z_p\}$ ses directions extrémales. Alors par le Théorème 12.40 on a

$$P \cap (\mathrm{lin}(P))^\perp = \mathrm{conv}(\{x_1, \cdots, x_k\}) + \mathrm{cône}(\{z_1, \cdots, z_p\}),$$

et finalement, puisque

$$\mathrm{lin}(P) = \mathrm{Vect}(\{e_1, \cdots, e_\ell)\} = \mathrm{cône}(\{e_1, \cdots, e_\ell, -e_1, \cdots, -e_\ell\}),$$

on obtient

$$P = \mathrm{conv}(\{x_1, \cdots, x_k\}) + \mathrm{cône}(\{z_1, \cdots, z_p, e_1, \cdots, e_\ell, -e_1, \cdots, -e_\ell\}).$$

Ceci termine la démonstration. □

Enfin, dans le cas particulier où P est un cône polyédrique, on déduit le

Corollaire 12.57. *Tout cône polyédrique est un cône finiment engendré.*

Démonstration. Il suffit de reprendre la démonstration du théorème précédent et de remarquer que $P \cap (\mathrm{lin}(P))^{\perp}$ est un cône polyédrique ne contenant aucune droite. Par conséquent, $\{0\}$ est son unique point extrémal, et dès lors

$$P = \mathrm{conv}(\{0\}) + \mathrm{cône}(D) = \mathrm{cône}(D),$$

ce qui termine la démonstration. □

Nous allons maintenant montrer la réciproque du Corollaire 12.57, ce qui permettra d'affirmer :

Théorème 12.58. *Un cône est finiment engendré si et seulement si il est polyédrique.*

Démonstration. Il ne nous reste bien sûr qu'à démontrer la condition nécessaire. Soit $\hat{C} = \mathrm{cône}(\{x_1, \cdots, x_\ell\})$ un cône finiment engendré. On définit son cône polaire $\hat{C}^o$ par

$$\hat{C}^o := \{y \in \mathbb{R}^n \mid \langle x_i, y\rangle \leq 0 \ \forall i \in \{1, \cdots, \ell\}\}.$$

Par construction, $\hat{C}^o$ est un cône polyédrique, et il suit dès lors du Corollaire 12.57 que $\hat{C}^o = \mathrm{cône}(\{y_1, \cdots, y_k\})$ pour certains $y_1, \cdots, y_k$ dans $\mathbb{R}^n$. On prétend que

$$\hat{C} = (\hat{C}^o)^o := \{x \in \mathbb{R}^n \mid \langle y_j, x\rangle \leq 0 \ \forall j \in \{1, \cdots, k\}\},$$

ce qui terminera la démonstration. Afin de montrer une première inclusion, prenons $x \in C$ quelconque. Alors $x = \sum_{i=1}^{\ell} \alpha_i x_i$ pour certains $\alpha_i \geq 0$. Pour chaque $j \in \{1, \cdots, k\})$, on a alors

$$\langle y_j, x\rangle = \sum_{i=1}^{\ell} \alpha_i \langle y_j, x_i\rangle = \sum_{i=1}^{\ell} \alpha_i \langle x_i, y_j\rangle \leq 0$$

car les α_i sont positifs et $y_j \in \hat{C}^o$. Donc $x \in (\hat{C}^o)^o$. Inversément, soit $x \in (\hat{C}^o)^o$. Si $z = \sum_{j=1}^{k} \beta_j y_j \in \hat{C}^o$ avec tous les β_j positifs, alors

$$\langle x, z\rangle = \sum_{j=1}^{k} \beta_j \langle x, y_j\rangle \leq 0.$$

Il suit du Lemme de Farkas (Théorème 12.49), appliqué pour $a_j := x_j$ et $b := x$, que $x \in \mathrm{cône}(\{x_1, \cdots, x_k\})$, c'est-à-dire $x \in \hat{C}$, d'où l'inclusion opposée. □

De même, le théorème qui suit est une réciproque du Théorème 12.56.

Théorème 12.59. *Soient E et D des ensembles finis de $\mathbb{R}^n$. Alors le convexe*

$$P := \mathrm{conv}(E) + \mathrm{cône}(D)$$

est un polyèdre.

Démonstration. On écrit $E = \{e_1, \cdots, e_k\}$, $D = \{z_1, \cdots, z_\ell\}$, et on pose

$$\hat{C} := \text{cône}(\{(e_1, 1), \cdots, (e_k, 1), (z_1, 0), \cdots, (z_\ell, 0)\}) \subseteq \mathbb{R}^{n+1}.$$

Comme $\hat{C}$ est un cône finiment engendré dans $\mathbb{R}^{n+1}$, par le théorème précédent, il existe $\{c_1, \cdots, c_p\} \subset \mathbb{R}^{n+1}$ tels que

$$\hat{C} = \{z \in \mathbb{R}^{n+1} \mid \langle a_i, z\rangle \leq 0 \ \forall i \in \{1, \cdots, p\}\}.$$

On observe enfin que $x \in P$ si et seulement si $z := (x, 1) \in \hat{C}$. Dès lors,

$$P = \{x \in \mathbb{R}^n \mid \langle a_i, x\rangle \leq b_i \ \forall i \in \{1, \cdots, p\}\},$$

où, pour chaque $i \in \{1, \cdots, p\}$, on a écrit $c_i \equiv (a_i, -b_i)$ avec $a_i \in \mathbb{R}^n$ et $b_i \in \mathbb{R}$. Par conséquent P est polyédrique. □

Corollaire 12.60. *Un sous-ensemble de $\mathbb{R}^n$ est un polytope si et seulement si c'est un polyèdre borné.*

Démonstration. C'est une conséquence du Théorème 12.55, du théorème précédent appliqué avec $D = \emptyset$, et de la Proposition 12.44. □