

Frédéric Meunier

Introduction à l'optimisation

13 septembre 2023

Table des matières

1	Introduction	1
1.1	Qu'est-ce que l'optimisation ?	1
1.2	Vocabulaire et notations	2
1.3	Modélisation et algorithmes	3
1.4	Objectif du cours	5
1.5	Quelques résultats élémentaires d'optimisation sur \mathbb{R}^n	6
2	Optimisation sous contraintes	7
2.1	Quelques résultats généraux	7
2.2	Théorème de Kuhn et Tucker	8
2.3	Lagrangien et dualité	13
3	Optimisation convexe	15
3.1	Motivation	15
3.2	Convexité	15
3.3	Caractérisation de la solution optimale	19
3.4	Dualité forte	20
3.5	Algorithmes	21
4	Optimisation linéaire	27
4.1	Motivation	27
4.2	Formulations équivalentes et hypothèses pour le chapitre	27
4.3	Existence et caractérisation d'une solution optimale	28
4.4	Algorithme du simplexe	30
4.5	Interprétation géométrique	32
5	Compléments sur l'optimisation linéaire	35
5.1	Dualité en optimisation linéaire	35
5.2	Matrices totalement unimodulaires	39

Introduction

1.1 Qu'est-ce que l'optimisation ?

Optimiser consiste à chercher le “meilleur” élément d'un ensemble donné. Trouver la forme la plus ergonomique pour une centrale nucléaire, une stratégie optimale de vente ou le plus court chemin dans un réseau, sont des exemples typiques de situations où l'on veut trouver un meilleur élément. La traduction de telles problématiques en objets mathématiques, l'étude des propriétés des solutions optimales et le calcul de ces solutions optimales forment précisément les objectifs de l'*optimisation*. C'est une discipline riche, présentant des relations avec de nombreux autres domaines des mathématiques. Comme l'on peut s'en douter à la lecture des exemples mentionnés ci-dessus, les applications de l'optimisation sont nombreuses. On en trouve en physique, dans la recherche des états d'énergie minimale par exemple, en économie, où les agents cherchent souvent à maximiser leur satisfaction, en logistique, où l'on souhaite minimiser la distance parcourue, dans l'industrie, etc. Voici quelques exemples classiques, qui illustreront cette variété.

Exemple 1.1 (Le problème de Didon). Lorsque la princesse Didon arriva en Afrique du Nord, les autochtones consentirent à lui céder un territoire qu'elle pourrait délimiter sur le littoral à l'aide d'une lanière découpée dans la peau d'un bœuf. La frontière de ce territoire serait formée par la côte et par la lanière. Le problème auquel elle fut donc confrontée, en supposant qu'elle souhaitait le territoire le plus grand possible, est donc un problème de maximisation de la surface délimitée, sous une contrainte de longueur de lanière ℓ fixée.

Exemple 1.2 (Le problème de transport). Le problème suivant est un classique de l'optimisation de la logistique. On dispose d'entrepôts et de magasins. Tout entrepôt i possède a_i unités d'un certain bien et tout magasin j a une demande b_j de ce bien. En supposant $\sum_i a_i \geq \sum_j b_j$, on peut satisfaire toute la demande. Connaissant le coût c_{ij} de transport d'une unité du bien de l'entrepôt i au magasin j , il peut être souhaitable de trouver un plan de transport, précisant pour chaque magasin la quantité du bien reçu de chaque entrepôt, qui soit de coût total minimal.

Exemple 1.3 (La brachistochrone). On se donne un point a et un point b contenus dans un même plan vertical. Quelle est la courbe de ce plan telle qu'un point placé en a sans vitesse initiale et soumis à son seul poids atteigne b en un minimum de temps ? Une telle courbe est appelée *brachistochrone*.

Exemple 1.4 (Le problème du voyageur de commerce). Un camion, initialement situé dans un dépôt, doit livrer des clients situés dans des lieux différents. Toutes les distances sont supposées connues. On souhaite trouver le parcours le plus court passant par tous les clients et revenant au dépôt. Tout comme le problème de transport, c'est un classique de l'optimisation en logistique.

Exemple 1.5 (Le routage de données informatiques dans un réseau). On dispose d'un réseau informatique, formé de nœuds et de liens reliant certaines paires de nœuds. On doit router une quantité q de données d'un nœud o à un nœud d . On connaît le coût de routage de toute quantité de données, pour tout lien ℓ du réseau, coût donné sous la forme d'une fonction réelle d'une variable réelle. On veut trouver le routage qui minimise le coût total.

C'est au cours du XX^{ème} siècle que l'optimisation, appelée également parfois programmation mathématique, a été identifiée comme discipline mathématique et a connu un développement important. La plupart des résultats présentés dans ce polycopié ont été établis entre 1900 et 1950. Des méthodes et des problèmes relevant de l'optimisation ont cependant été établis bien avant. Le problème de la brachistochrone, par exemple, a été posé en 1696 par Jean Bernoulli et a été très vite résolu par Leibniz, grâce à la toute nouvelle méthode du « calcul des variations ». Le problème de transport, dû à Monge, a été formulé par ce dernier en 1791. La solution qu'il proposa était cependant erronée. Les outils mathématiques nécessaires pour sa résolution complète dans la version donnée ici datent des années 1940, et relèvent essentiellement de l'optimisation linéaire.

1.2 Vocabulaire et notations

1.2.1 Quelques notations

L'ensemble $\{1, \dots, k\}$ sera en général noté $[k]$. Le vecteur dont toutes les composantes sont des 0 (resp. des 1, resp. des ∞) sera noté $\mathbf{0}$ (resp. $\mathbf{1}$, resp. ∞). On utilisera $\widetilde{\mathbb{R}}$ pour noter l'ensemble $\mathbb{R} \cup \{+\infty\}$ et $\overline{\mathbb{R}}$ pour noter l'ensemble $\mathbb{R} \cup \{-\infty, +\infty\}$.

1.2.2 Domaine d'une fonction

Dans ce cours, on considèrera souvent des fonctions qui prennent des valeurs infinies. Etant donnée une fonction f à valeur dans $\overline{\mathbb{R}}$, l'ensemble des points où elle prend une valeur finie, noté $\text{dom } f$, est le *domaine* de f . Une fonction $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ est *continue* (resp.

différentiable) si $\text{dom } f$ est ouvert et si elle est continue (resp. différentiable) en tout point de $\text{dom } f$.

1.2.3 Minimum local, minimum global

Soit f une fonction définie sur \mathbb{R}^n et à valeur dans $\overline{\mathbb{R}}$ et soit X une partie de \mathbb{R}^n . Un élément $\mathbf{x}^* \in \mathbb{R}^n$ est un *minimum local de f sur X* s'il existe un voisinage V de \mathbf{x}^* dans \mathbb{R}^n tel que $f(\mathbf{x}^*) \leq f(\mathbf{x})$ pour tout $\mathbf{x} \in V \cap X$. Un tel élément est un *minimum global de f sur X* si $f(\mathbf{x}^*) \leq f(\mathbf{x})$ pour tout $\mathbf{x} \in X$. On définit de manière semblable *maximum local* et *maximum global*, simplement en renversant le sens des inégalités.

1.2.4 Problème d'optimisation

Un *problème d'optimisation* consiste à se donner un ensemble X et une fonction f à valeur dans $\overline{\mathbb{R}}$ et définie sur X , et à chercher un élément $x^* \in X$ minimisant ou maximisant f , i.e. à chercher $x^* \in X$ tel que $f(x^*) \leq f(x)$ pour tout $x \in X$ si l'on souhaite minimiser, ou à chercher $x^* \in X$ tel que $f(x^*) \geq f(x)$ pour tout $x \in X$ si l'on souhaite maximiser. Remarquons d'emblée que maximiser f équivaut à minimiser $-f$. Dans ce cours, les problèmes d'optimisation considérés seront en général présentés comme des problèmes de minimisation.

On écrit un tel problème sous la forme

$$\begin{array}{ll} \text{Min} & f(x) \\ \text{s.c.} & x \in X \end{array} \quad (\text{P})$$

où « s.c. » signifie « sous contraintes ». On écrit « Min » avec une majuscule pour signifier que l'on cherche l'infimum de cette fonction, sans garantie de son existence ou de sa finitude. Pour un problème de maximisation, on utilise « Max ».

La fonction f est la *fonction objectif* ou le *critère* du problème d'optimisation et « $x \in X$ » est l'ensemble des *contraintes*. Tout $x \in X$ est une *solution réalisable*. Si X est non vide, alors (P) est *réalisable*. Tout $x^* \in X$ minimum global de f sur X est une *solution optimale*. La quantité $\inf\{f(x) : x \in X\}$ est la *valeur optimale* du problème (P). Si X est vide, cette valeur est posée égale à $+\infty$ par convention. Lorsqu'on est confronté à un problème de maximisation, la valeur optimale est bien entendu définie en prenant le supremum à la place de l'infimum et en prenant $-\infty$ comme valeur optimale si X est vide.

1.3 Modélisation et algorithmes

Le travail d'optimisation appliqué est souvent précédé d'un travail de *modélisation*, qui consiste à écrire le problème d'optimisation que l'on souhaite résoudre sous la forme (P). Effectuons cet exercice sur quatre des exemples présentés ci-dessus.

Exemple 1.1 (Le problème de Didon). Le problème peut se modéliser sous la forme suivante, en se restreignant aux fonctions \mathcal{C}^1 et en supposant dans un premier temps que les deux

points du littoral reliés par la lanrière sont des points fixés a et b . (Le littoral est supposé rectiligne).

$$\begin{aligned} \text{Max} \quad & \int_a^b f(t) dt \\ \text{s.c.} \quad & f(a) = f(b) = 0 \\ & \int_a^b \sqrt{1 + f'^2(t)} dt = \ell \\ & f \in \mathcal{C}^1(\mathbb{R}). \end{aligned}$$

En procédant ainsi, il faut justifier que, pour la solution optimale, toute perpendiculaire au littoral intersecte la lanrière en au plus un point. On peut montrer que la solution optimale est atteinte lorsque f décrit un arc de cercle. Il est alors aisé de trouver a et b maximisant la surface délimitée.

Exemple 1.2 (Le problème de transport). En posant x_{ij} la quantité de bien envoyée depuis l'entrepôt i vers le magasin j , on peut écrire le problème sous la forme

$$\begin{aligned} \text{Min} \quad & \sum_{i,j} c_{ij} x_{ij} \\ \text{s.c.} \quad & \sum_j x_{ij} \leq a_i \quad \forall i \\ & \sum_i x_{ij} \geq b_j \quad \forall j \\ & x_{ij} \in \mathbb{R}_+ \quad \forall i, j. \end{aligned}$$

Dans cette modélisation, on a supposé que le bien pouvait être subdivisé à l'infini. Si ce n'est pas le cas, i.e. que toutes les quantités transportées doivent mettre en jeu un nombre entier d'unités, il faut remplacer $\ll x_{ij} \in \mathbb{R}_+ \gg$ par $\ll x_{ij} \in \mathbb{Z}_+ \gg$.

La modélisation des deux derniers exemples s'appuie sur les *graphes*, qui sont des objets classiques de l'optimisation combinatoire, particulièrement utiles en recherche opérationnelle. Le fait que la modélisation de ces deux exemples est correcte n'est d'ailleurs pas immédiate et demande une preuve, omise dans ce cours.

Exemple 1.4 (Le problème du voyageur de commerce). Une façon de modéliser le problème est de considérer un graphe orienté dont les sommets sont formés du dépôt et des lieux où sont situés les clients et dont les arcs relient tous les couples de sommets distincts. On munit ensuite les arcs de ce graphe de poids, l'arc (u, v) ayant comme poids la distance séparant u de v . On cherche alors dans ce graphe le circuit hamiltonien de plus petit poids. (Dans un graphe orienté, un circuit hamiltonien est un circuit qui passe par chaque sommet exactement une fois).

Exemple 1.5 (Le routage de données). On modélise le réseau par un graphe orienté $D = (V, A)$ dont les sommets sont les nœuds et donc les arcs sont les liens. On note alors $c_a(x)$ le coût de routage d'une quantité de x de données sur le lien modélisé par l'arc a . Le problème s'écrit

alors

$$\begin{aligned} \text{Min} \quad & \sum_{a \in A} c_a(x_a) \\ \text{s.c.} \quad & \sum_{a \in \delta^+(o)} x_a = q \\ & \sum_{a \in \delta^+(v)} x_a = \sum_{a \in \delta^-(v)} x_a \quad \forall v \notin \{o, d\} \\ & x_a \in \mathbb{R}_+ \quad \forall a \in A. \end{aligned}$$

Seuls les exemples 1.2 et 1.5 pourront être traités par les outils du cours. Ils relèvent de l’optimisation “en dimension finie”, i.e. que l’ensemble des solutions réalisables est inclus dans un espace vectoriel de dimension finie, contrairement aux exemples 1.1 et 1.3. Ces derniers peuvent être résolus avec les outils des cours d’analyse de première année de l’ENPC. Enfin, l’exemple 1.4, si on veut le résoudre de manière exacte, nécessite des outils de la recherche opérationnelle, en particulier l’étude des polyèdres. Notons aussi que les solutions des exemples 1.1 et 1.3 peuvent être décrites à l’aide d’une formule explicite. Cette situation est plutôt rare en optimisation, et ce n’est d’ailleurs pas le cas pour les trois autres exemples. Une part importante de l’optimisation consiste par conséquent à élaborer des méthodes permettant de calculer ou d’approcher les solutions optimales. Ces méthodes se présentent en général sous la forme d’algorithmes, un *algorithme* pouvant être *grosso modo* défini comme une suite d’opérations élémentaires implémentables sur un ordinateur.

1.4 Objectif du cours

Le principal objectif du cours est de donner aux étudiants des outils pour étudier et, si possible, résoudre le cas particulier du problème (P) lorsque X est une partie de \mathbb{R}^n décrite par un nombre fini d’égalités et d’inégalités impliquant des fonctions différentiables. Plus précisément, ce sont les problèmes de la forme

$$\begin{aligned} \text{Min} \quad & f(\mathbf{x}) \\ \text{s.c.} \quad & g_i(\mathbf{x}) = 0 \quad i = 1, \dots, p \\ & h_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, q \end{aligned} \tag{Q}$$

où f , les g_i et les h_j sont des fonctions différentiables de \mathbb{R}^n dans $\tilde{\mathbb{R}}$, qui formeront l’objet principal de ce cours. On verra que dans ce cas on est capable de donner une condition nécessaire d’optimalité – le théorème de Kuhn et Tucker – qui généralise d’ailleurs le cas d’annulation de la dérivée pour une fonction à valeurs réelles et dérivable.

Une partie du cours sera consacrée au cas où la fonction f et les fonctions h_j sont convexes, et les fonctions g_i affines. On verra alors que cette condition nécessaire devient suffisante et que l’on dispose d’algorithmes efficaces pour trouver les solutions optimales. Enfin, une partie importante du cours sera consacrée au cas linéaire, i.e. au cas où ces fonctions sont des fonctions affines. C’est un cas qui possède des propriétés théoriques remarquables, des algorithmes puissants et de nombreuses applications dans l’industrie.

Les exercices et l’examen ne feront pas appel aux résultats (théorèmes, lemmes, propositions et corollaires) précédés d’une astérisque *.

1.5 Quelques résultats élémentaires d'optimisation sur \mathbb{R}^n

Dans cette section, on donne quelques résultats élémentaires d'optimisation sur \mathbb{R}^n . On rappelle qu'un *compact* de \mathbb{R}^n est une partie fermée bornée de \mathbb{R}^n .

Théorème 1.1. *Soit C un compact non vide de \mathbb{R}^n . Si $f: C \rightarrow \mathbb{R}$ est continue, alors*

$$\begin{array}{ll} \text{Min} & f(\mathbf{x}) \\ \text{s.c.} & \mathbf{x} \in C \end{array}$$

admet au moins une solution optimale.

Une fonction $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$ est *coercive* si $f(\mathbf{x}) \rightarrow +\infty$ quand $\|\mathbf{x}\| \rightarrow +\infty$.

Théorème 1.2. *Si $f: \mathbb{R}^n \rightarrow \mathbb{R}$ est continue et coercive, alors le problème d'optimisation*

$$\begin{array}{ll} \text{Min} & f(\mathbf{x}) \\ \text{s.c.} & \mathbf{x} \in \mathbb{R}^n \end{array}$$

admet au moins une solution optimale.

Démonstration. Comme f est coercive, il existe $r \in \mathbb{R}_+$ tels que si $\|\mathbf{x}\| \geq r$, alors $f(\mathbf{x}) \geq f(\mathbf{0})$. Le minimum global de f restreinte à la boule fermée centrée en $\mathbf{0}$ et de rayon r , minimum qui existe par le théorème 1.1, est donc une solution optimale du problème. \square

Si f est différentiable, on peut en dire plus sur les optima locaux (et donc globaux). Le théorème suivant est connu sous le nom de « condition d'Euler ».

Théorème 1.3. *Soit \mathbf{x}^* un minimum local de $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$. Si f est différentiable en \mathbf{x}^* , alors $\nabla f(\mathbf{x}^*) = \mathbf{0}$.*

Démonstration. Soit $\mathbf{h} \in \mathbb{R}^n$. Puisque \mathbf{x}^* est un minimum local, on a $\frac{1}{t}(f(\mathbf{x}^* + t\mathbf{h}) - f(\mathbf{x}^*)) \geq 0$ pour tout t strictement positif suffisamment petit, ce qui implique $\nabla f(\mathbf{x}^*) \cdot \mathbf{h} \geq 0$. De même, on a $\frac{1}{t}(f(\mathbf{x}^*) - f(\mathbf{x}^* - t\mathbf{h})) \leq 0$ pour tout t strictement positif suffisamment petit, ce qui implique $\nabla f(\mathbf{x}^*) \cdot \mathbf{h} \leq 0$. On a donc $\nabla f(\mathbf{x}^*) \cdot \mathbf{h} = 0$ pour tout $\mathbf{h} \in \mathbb{R}^n$, ce qui implique le résultat. \square

On note $H(f)$ la hessienne (matrice carrée des dérivées partielles secondes) d'une fonction f .

***Théorème 1.4.** *Supposons $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$ deux fois différentiable. Si $\mathbf{x}^* \in \text{dom } f$ est tel que $\nabla f(\mathbf{x}^*) = \mathbf{0}$ et tel que $H(f)(\mathbf{x}^*)$ soit définie positive, alors \mathbf{x}^* est un minimum local.*

Démonstration. Écrivons un développement de Taylor à l'ordre 2 en \mathbf{x}^* :

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \frac{1}{2}\mathbf{h}^T H(\mathbf{x}^*)\mathbf{h} + o(\|\mathbf{h}^2\|).$$

Il existe $\alpha > 0$ tel que $\mathbf{h}^T H(\mathbf{x}^*)\mathbf{h} \geq \alpha\|\mathbf{h}\|^2$ pour tout $\mathbf{h} \in \mathbb{R}^n$. Choisissons $\eta > 0$ tel que le terme $o(\|\mathbf{h}^2\|)$ soit plus grand que $-\alpha\|\mathbf{h}^2\|$ dès que $\|\mathbf{h}\| \leq \eta$. En utilisant le développement de Taylor ci-dessus, on obtient donc $f(\mathbf{x}^* + \mathbf{h}) \geq f(\mathbf{x}^*)$ dès que $\|\mathbf{h}\| \leq \eta$. Le point \mathbf{x}^* est donc bien un minimum local. \square

Optimisation sous contraintes

2.1 Quelques résultats généraux

On s'intéresse dans cette section à la formulation très générale

$$\begin{aligned} \text{Min} \quad & f(\mathbf{x}) \\ \text{s.c.} \quad & \mathbf{x} \in X, \end{aligned} \tag{P}$$

où f est une fonction de \mathbb{R}^n à valeurs dans \mathbb{R} ou dans $\tilde{\mathbb{R}}$, et X une partie de \mathbb{R}^n .

Théorème 2.1. *Si $f: \mathbb{R}^n \rightarrow \tilde{\mathbb{R}}$ est continue et coercive et si $X \cap \text{dom } f$ est un fermé non vide de \mathbb{R}^n , alors le problème d'optimisation (P) admet au moins une solution optimale.*

La preuve est quasiment identique à celle du théorème 1.2, qui en est un cas particulier.

Démonstration du théorème 2.1. Choisissons arbitrairement \mathbf{x}_0 dans $X \cap \text{dom } f$. Il suffit alors de considérer la restriction de f à l'intersection de $X \cap \text{dom } f$ avec la boule fermée centrée en $\mathbf{0}$ et de rayon r , où r est tel que si $\|\mathbf{x}\| \geq r$ alors $f(\mathbf{x}) \geq f(\mathbf{x}_0)$, et d'appliquer le théorème 1.1. \square

Le vecteur $\mathbf{d} \in \mathbb{R}^n$ est une *direction admissible* en $\mathbf{x} \in X$ s'il existe une suite $(t_k)_k$ de réels tendant vers 0 par valeurs strictement positives et une suite $(\mathbf{d}^k)_k$ de vecteurs de \mathbb{R}^n tendant vers \mathbf{d} tels que $\mathbf{x} + t_k \mathbf{d}^k \in X$ pour tout k . On note $D(\mathbf{x})$ l'ensemble des directions admissibles en \mathbf{x} . Cet ensemble est appelé *cône tangent de X en \mathbf{x}* .

***Lemme 2.2.** *L'ensemble $D(\mathbf{x})$ des directions admissibles en \mathbf{x} est toujours fermé.*

Démonstration. Soit $(\mathbf{d}^k)_k$ une suite de directions admissibles en \mathbf{x} convergeant vers un vecteur \mathbf{d} . Par définition, pour tout k , il existe $\hat{\mathbf{d}}^k$ et t_k satisfaisant $\|\hat{\mathbf{d}}^k - \mathbf{d}^k\| < 1/k$ et $t_k < 1/k$ tels que $\mathbf{x} + t_k \hat{\mathbf{d}}^k \in X$. On a donc $\lim_{k \rightarrow +\infty} \hat{\mathbf{d}}^k = \mathbf{d}$, ce qui permet de conclure. \square

Le théorème suivant, connu parfois sous le nom de « condition d'Euler-Lagrange », généralise le théorème 1.3.

Théorème 2.3. *Soit $f: \mathbb{R}^n \rightarrow \tilde{\mathbb{R}}$. Soit \mathbf{x}^* un minimum local de f sur X et soit \mathbf{d} une direction admissible en \mathbf{x}^* . Si f est différentiable en \mathbf{x}^* , alors on a $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$.*

Démonstration. Soit donc \mathbf{d} une direction admissible en \mathbf{x}^* , minimum local de f sur X . Soit \mathbf{d}^k et t_k les suites assurées par la définition de direction admissible.

Soit $\varepsilon > 0$. En écrivant $f(\mathbf{x}^* + t_k \mathbf{d}^k) = f(\mathbf{x}^*) + t_k \nabla f(\mathbf{x}^*) \cdot \mathbf{d}^k + o(\|t_k \mathbf{d}^k\|)$, on obtient l'inégalité

$$\frac{1}{t_k} \left(f(\mathbf{x}^* + t_k \mathbf{d}^k) - f(\mathbf{x}^*) \right) \leq \nabla f(\mathbf{x}^*) \cdot \mathbf{d}^k + \varepsilon \|\mathbf{d}^k\|$$

pour k suffisamment grand. Le fait que \mathbf{x}^* est un minimum local implique que, pour k suffisamment grand, on a $0 \leq \nabla f(\mathbf{x}^*) \cdot \mathbf{d}^k + \varepsilon \|\mathbf{d}^k\|$, et donc que $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq -\varepsilon \|\mathbf{d}\|$. Comme cette inégalité est vraie pour tout $\varepsilon > 0$, on obtient l'inégalité souhaitée. \square

2.2 Théorème de Kuhn et Tucker

2.2.1 Énoncé et preuve

On s'intéresse maintenant au problème

$$\begin{aligned} \text{Min} \quad & f(\mathbf{x}) \\ \text{s.c.} \quad & g_i(\mathbf{x}) = 0 \quad i = 1, \dots, p \\ & h_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, q. \end{aligned} \tag{Q}$$

où f , les g_i et les h_j sont des fonctions de \mathbb{R}^n dans $\tilde{\mathbb{R}}$. C'est donc le problème (P) avec

$$X = \{ \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) = 0, i \in [p] \quad \text{et} \quad h_j(\mathbf{x}) \leq 0, j \in [q] \}.$$

Les contraintes de la forme $g_i(\mathbf{x}) = 0$ sont les *contraintes d'égalité* et celles de la forme $h_j(\mathbf{x}) \leq 0$ sont les *contraintes d'inégalité*. Nous allons nous restreindre au cas où f , les g_i et les h_j sont différentiables.

Soit \mathbf{x} un point de X . On définit

$$J(\mathbf{x}) = \{ j \in [q] : h_j(\mathbf{x}) = 0 \},$$

ensemble des contraintes d'inégalité *actives*, et

$$D'(\mathbf{x}) = \{ \mathbf{d} \in \mathbb{R}^n : \nabla g_i(\mathbf{x}) \cdot \mathbf{d} = 0, i \in [p] \quad \text{et} \quad \nabla h_j(\mathbf{x}) \cdot \mathbf{d} \leq 0, j \in J(\mathbf{x}) \}.$$

L'ensemble $D'(\mathbf{x})$ est appelé *cône linéarisant de X en \mathbf{x}* . Toute direction admissible en \mathbf{x} est dans $D'(\mathbf{x})$, i.e. $D(\mathbf{x}) \subseteq D'(\mathbf{x})$, comme cela peut être aisément vérifié en s'inspirant de la preuve du théorème 2.3. En revanche, l'inverse n'est pas nécessairement vrai. Les contraintes de (Q) sont *qualifiées* en \mathbf{x} si l'on a $D(\mathbf{x}) = D'(\mathbf{x})$, c'est-à-dire si l'on a égalité entre les cônes tangent et linéarisant de X en \mathbf{x} .

Le théorème de Kuhn et Tucker, établi en 1951 [8], donne une condition nécessaire d'optimalité d'une solution réalisable du problème (Q). Cette condition est appelée *condition de Kuhn–Tucker*, ou *condition de Karush–Kuhn–Tucker*, ou encore *condition KKT*. Le nom de Karush est parfois ajouté car cette condition est déjà présente dans un mémoire de master écrit par ce dernier en 1939, ce qu'ignoraient Kuhn et Tucker.

Théorème 2.4 (Théorème de Kuhn et Tucker). *Supposons f , les g_i et les h_j différentiables. Soit $\mathbf{x}^* \in X \cap \text{dom } f$ une solution réalisable du problème (Q) en laquelle les contraintes sont qualifiées. Si \mathbf{x}^* est un minimum local de f sur X , alors il existe p réels λ_i et q réels positifs ou nuls μ_j vérifiant*

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^p \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^q \mu_j \nabla h_j(\mathbf{x}^*) = \mathbf{0} \quad \text{et} \quad \mu_j h_j(\mathbf{x}^*) = 0 \quad \forall j \in [q].$$

Les réels λ_i et μ_j sont des *multiplicateurs de Lagrange* associés à \mathbf{x}^* . Ils ne sont pas nécessairement uniques.

Démonstration du théorème 2.4. Soit $\mathbf{x}^* \in \text{dom } f$ minimum local de f sur X . Définissons la matrice A suivante :

$$A = \left((\nabla g_i(\mathbf{x}^*))_{i \in [p]}, (-\nabla g_i(\mathbf{x}^*))_{i \in [p]}, (\nabla h_j(\mathbf{x}^*))_{j \in J(\mathbf{x}^*)} \right).$$

D'après le théorème 2.3, on a $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour toute direction admissible \mathbf{d} en \mathbf{x}^* . Comme les contraintes sont qualifiées, cela signifie que $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour tout $\mathbf{d} \in \mathbb{R}^n$ tel que $\mathbf{d}^T A \leq \mathbf{0}$. En appliquant le lemme de Farkas (lemme 5.2, voir Chapitre 5), c'est équivalent à l'existence de deux vecteurs $\boldsymbol{\lambda}'$ et $\boldsymbol{\lambda}''$ dans \mathbb{R}_+^p et d'un vecteur $\boldsymbol{\mu}'$ dans $\mathbb{R}_+^{J(\mathbf{x}^*)}$ tels que

$$-A \begin{pmatrix} \boldsymbol{\lambda}' \\ \boldsymbol{\lambda}'' \\ \boldsymbol{\mu}' \end{pmatrix} = \nabla f(\mathbf{x}^*).$$

En posant $\lambda_i = \lambda'_i - \lambda''_i$ pour $i \in [p]$ et

$$\mu_j = \begin{cases} \mu'_j & \text{pour } j \in J(\mathbf{x}^*) \\ 0 & \text{sinon,} \end{cases}$$

on obtient le résultat. □

Quand il n'y a pas de contrainte d'inégalités, le théorème de Kuhn et Tucker est plus aisé à prouver et le lemme de Farkas peut être remplacé par un argument d'algèbre linéaire élémentaire. En effet, " $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour tout $\mathbf{d} \in \mathbb{R}^n$ tel que $\mathbf{d}^T A \leq \mathbf{0}$ " devient alors " $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour tout $\mathbf{d} \in \mathbb{R}^n$ tel que $\nabla g_i(\mathbf{x}^*) \cdot \mathbf{d} = 0$ pour tout i ", ce qui se réécrit " $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} = 0$ pour tout $\mathbf{d} \in \mathbb{R}^n$ tel que $\nabla g_i(\mathbf{x}^*) \cdot \mathbf{d} = 0$ pour tout i ", qui est équivalent à " $\nabla f(\mathbf{x}^*)$ appartient au sous-espace vectoriel engendré par les $\nabla g_i(\mathbf{x}^*)$ ".

Exemple 2.1. La figure 2.1 illustre le cas où $n = 2$, $p = 0$ et $q = 3$ (on est en dimension deux et il n'y a que des contraintes d'inégalité, au nombre de trois).

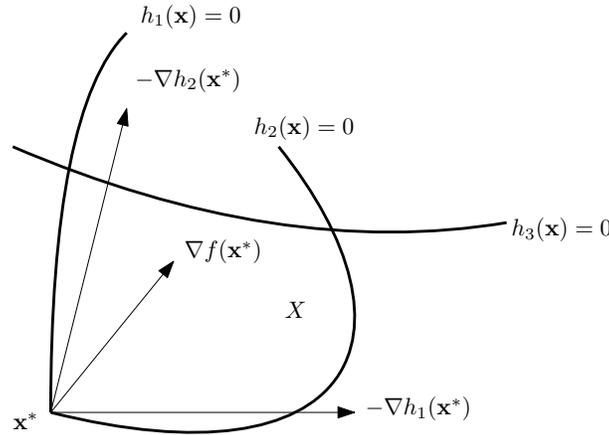


FIGURE 2.1 – Illustration de la condition KKT dans un cas particulier.

Supposons que le \mathbf{x}^* de la figure soit un minimum local de f . Le corollaire 2.8 de la section suivante assure que les contraintes sont qualifiées en \mathbf{x}^* : la troisième contrainte n'est pas active et les vecteurs $\nabla h_1(\mathbf{x}^*)$ et $\nabla h_2(\mathbf{x}^*)$ sont linéairement indépendants. Le théorème 2.4 assure alors que $\nabla f(\mathbf{x}^*)$ est dans le cône engendré par $-\nabla h_1(\mathbf{x}^*)$ et par $-\nabla h_2(\mathbf{x}^*)$.

Cela peut se comprendre sur le dessin sans invoquer le théorème 2.4 (ou le lemme de Farkas sur lequel il s'appuie). En effet, tout déplacement infinitésimal depuis \mathbf{x}^* dans une direction \mathbf{d} telle que $\nabla h_1(\mathbf{x}^*) \cdot \mathbf{d} \leq 0$ et $\nabla h_2(\mathbf{x}^*) \cdot \mathbf{d} \leq 0$ fournit une solution réalisable (avec les notions du cours, on a $D'(\mathbf{x}^*) = D(\mathbf{x}^*)$, i.e. on a qualification des contraintes en \mathbf{x}^*). Comme \mathbf{x}^* est un minimum local, on veut donc que pour une telle direction \mathbf{d} , on ait $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ (avec les notions du cours, c'est le théorème 2.3). Montrer qu'alors nécessairement la décomposition de $\nabla f(\mathbf{x}^*)$ sur la base $\{\nabla h_1(\mathbf{x}^*), \nabla h_2(\mathbf{x}^*)\}$ implique des coefficients négatifs est un exercice aisé de géométrie du plan.

Exemple 2.2. Regardons comment le théorème 2.4 permet de résoudre le problème d'optimisation

$$\begin{aligned} \text{Min} \quad & x^2 - y^2 \\ \text{s.c.} \quad & x + y \geq 4 \\ & x \leq 6 \\ & y \leq 7. \end{aligned}$$

Remarquons d'abord que l'ensemble des solutions réalisables forme un ensemble compact. En effet, cet ensemble est fermé et l'on a $x \in [-3, 6]$ et $y \in [-2, 7]$. Il existe donc une solution optimale (x^*, y^*) . D'après le théorème de Kuhn et Tucker (que l'on peut appliquer ici car les contraintes sont qualifiées partout, comme on le verra à la section suivante à l'aide de la proposition 2.5), il existe $\mu_1, \mu_2, \mu_3 \in \mathbb{R}_+$ tels que

$$2x^* - \mu_1 + \mu_2 = -2y^* - \mu_1 + \mu_3 = 0$$

et tels que

$$\mu_1(4 - x^* - y^*) = \mu_2(x^* - 6) = \mu_3(y^* - 7) = 0.$$

Supposons dans un premier temps $\mu_1 = 0$. Dans ce cas, on ne peut avoir $\mu_2 \neq 0$, car sinon, on aurait $x^* = 6$ et $\mu_2 = -2x^* < 0$. On a donc $\mu_2 = 0$, et par conséquent $x^* = 0$. Les équations $-2y^* + \mu_3 = 0$ et $\mu_3(y^* - 7) = 0$ implique que $y^* = 0$ ou $y^* = 7$.

Supposons maintenant que $\mu_1 > 0$. On a alors $x^* + y^* = 4$ et $2x^* - \mu_1 + \mu_2 = -2y^* - \mu_1 + \mu_3$ donne $\mu_3 - \mu_2 = 8$. On ne peut avoir $\mu_3 = 0$ car sinon μ_2 serait strictement négatif. On a donc $\mu_3 > 0$, qui implique $x^* = -3$, et donc $y^* = 7$.

On compare alors la valeur prise par la fonction objectif par les trois solutions possibles $(0, 0)$, $(0, 7)$ et $(-3, -7)$. La solution optimale du problème est donc $(0, 7)$.

Exemple 2.3. Regardons comment le théorème 2.4 permet de résoudre le problème d'optimisation

$$\begin{aligned} \text{Min} \quad & x^3 + y^3 \\ \text{s.c.} \quad & x^2 + y^2 = 1. \end{aligned}$$

L'ensemble des solutions réalisables forme un ensemble compact. Il existe donc une solution optimale (x^*, y^*) . D'après le théorème de Kuhn et Tucker (que l'on peut appliquer ici car les contraintes sont qualifiées partout, comme on le verra à la section suivante à l'aide de la proposition 2.7), il existe $\lambda \in \mathbb{R}$ tel que

$$3x^{*2} + 2\lambda x^* = 3y^{*2} + 2\lambda y^* = 0.$$

On a donc $x^*, y^* \in \{0, -\frac{2}{3}\lambda\}$. Comme (x^*, y^*) est une solution réalisable et doit donc satisfaire $x^{*2} + y^{*2} = 1$, il y a six solutions possibles : $(0, \pm 1)$, $(\pm 1, 0)$ et $\pm\sqrt{\frac{1}{2}}(1, 1)$. On compare alors la valeur prise par la fonction objectif pour ces six solutions possibles, et l'on trouve qu'il y a deux solutions optimales : $(-1, 0)$ et $(0, -1)$.

Si à la place du "Min", on avait "Max", la solution optimale serait $\sqrt{\frac{1}{2}}(1, 1)$: le raisonnement ne change pas.

Exemple 2.4. Le problème d'optimisation suivant montre que lorsque les contraintes ne sont pas qualifiées, la condition de Kuhn-Tucker n'est pas nécessairement satisfaite en la solution optimale.

$$\begin{aligned} \text{Min} \quad & x \\ \text{s.c.} \quad & x^3 \geq 0. \end{aligned}$$

La solution optimale de ce problème est clairement $x^* = 0$. Si la contrainte était qualifiée en 0, la condition de Kuhn-Tucker serait vérifiée en 0, mais elle s'écrirait alors $1 = 0$. La contrainte n'est donc pas qualifiée en 0, et on a bien pour ce problème $D(0) = \mathbb{R}_+$ et $D'(0) = \mathbb{R}$.

2.2.2 Quelques conditions suffisantes de qualification

Le théorème de Kuhn et Tucker (théorème 2.4) requiert que les contraintes soient qualifiées au point où l'on souhaite l'appliquer. Il peut être difficile de le vérifier directement. Nous donnons ci-après des conditions suffisantes, en général aisément vérifiables, qui assurent que les contraintes sont qualifiées en un point ou sur tout X .

Proposition 2.5. *Supposons les g_i et h_j affines. Alors les contraintes de (Q) sont qualifiées en tout point de X .*

Démonstration. Soit $\mathbf{x} \in X$ et soit $\mathbf{d} \in D'(\mathbf{x})$. Montrons que \mathbf{d} est admissible. Notons que l'on a clairement $g_i(\mathbf{x} + t\mathbf{d}) = 0$ pour tout $t \in \mathbb{R}$. La difficulté vient donc du côté des contraintes d'inégalités. Écrivons $h_j(\mathbf{x})$ sous la forme $\mathbf{a}_j \cdot \mathbf{x} + b_j$ et définissons K comme étant l'ensemble des indices j tels que $\mathbf{a}_j \cdot \mathbf{d} > 0$. Comme \mathbf{d} est dans $D'(\mathbf{x})$, on a $\mathbf{a}_j \cdot \mathbf{d} \leq 0$ si $\mathbf{a}_j \cdot \mathbf{x} + b_j = 0$. Par conséquent, tout j dans K vérifie $\mathbf{a}_j \cdot \mathbf{x} + b_j < 0$. Il existe donc $\varepsilon > 0$ tel que $\mathbf{a}_j \cdot (\mathbf{x} + t\mathbf{d}) + b_j \leq 0$ pour tout $t \in [0, \varepsilon]$ et tout $j \in [q]$, ce qui signifie que $\mathbf{x} + t\mathbf{d} \in X$ pour tout $t \in [0, \varepsilon]$. Le vecteur \mathbf{d} est donc admissible. \square

Dans la proposition suivante, on utilise des notions de convexité, voir Chapitre 3 et plus particulièrement la Section 3.2 pour les définitions.

***Proposition 2.6** (Condition de Slater). *Supposons les g_i affines et les h_j convexes et différentiables. S'il existe $\mathbf{y} \in X$ tel que $h_j(\mathbf{y}) < 0$ pour tout $j \in [q]$, alors les contraintes de (Q) sont qualifiées en tout point de X .*

Démonstration. Prenons donc un $\mathbf{y} \in X$ tel que $h_j(\mathbf{y}) < 0$ pour tout $j \in [q]$. Soit $\mathbf{x} \in X$ et soit $\mathbf{d} \in D'(\mathbf{x})$. Nous allons montrer que \mathbf{d} est une direction admissible, en prouvant qu'il est limite de directions admissibles. Comme $D(\mathbf{x})$ est fermé (lemme 2.2), cela permettra de conclure.

Considérons une suite $(\mathbf{x}^k)_{k \in \mathbb{Z}_+}$ de points sur le segment $[\mathbf{y}, \mathbf{z}]$, où $\mathbf{z} = \mathbf{x} + \mathbf{d}$ telle que $\lim_{k \rightarrow +\infty} \mathbf{x}^k = \mathbf{z}$ et définissons $\mathbf{d}^k = \mathbf{x}^k - \mathbf{x}$. Nous allons montrer que \mathbf{d}^k est admissible à partir d'un certain rang, ce qui prouvera que \mathbf{d} est dans l'adhérence des directions admissibles, puisque $\lim_{k \rightarrow +\infty} \mathbf{d}^k = \mathbf{d}$.

Soit $j \in J(\mathbf{x})$. Supposons que j soit tel que $\nabla h_j(\mathbf{x}) \cdot \mathbf{d} = 0$. On a alors $\nabla h_j(\mathbf{x}) \cdot \mathbf{d}^k < 0$. En effet, on a

$$\nabla h_j(\mathbf{x}) \cdot \mathbf{d}^k = \nabla h_j(\mathbf{x}) \cdot (\mathbf{x}^k - \mathbf{z} + \mathbf{z} - \mathbf{x}) = \nabla h_j(\mathbf{x}) \cdot (\mathbf{x}^k - \mathbf{z}).$$

Notons que $\nabla h_j(\mathbf{x}) \cdot (\mathbf{x}^k - \mathbf{z})$ a le même signe que $\nabla h_j(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x})$, qui est strictement négatif d'après les hypothèses et la convexité de h_j (on utilise ici la proposition 3.3). Donc $\nabla h_j(\mathbf{x}) \cdot \mathbf{d}^k < 0$. Supposons maintenant que j soit tel que $\nabla h_j(\mathbf{x}) \cdot \mathbf{d} < 0$. Il existe un rang à partir duquel on a toujours $\nabla h_j(\mathbf{x}) \cdot \mathbf{d}^k < 0$ car (\mathbf{d}^k) converge vers \mathbf{d} .

Par conséquent, pour k suffisamment grand, on a $\nabla h_j(\mathbf{x}) \cdot \mathbf{d}^k < 0$ pour tout $j \in J(\mathbf{x})$. En utilisant le caractère affine des g_i , on montre aisément que l'on a $\nabla g_i(\mathbf{x}) \cdot \mathbf{d}^k = 0$ pour tout k . Cela implique que lorsque k est suffisamment grand, on a $\mathbf{x} + t\mathbf{d}^k \in X$ pour tout t dans un voisinage de 0. Donc les \mathbf{d}^k sont dans $D(\mathbf{x})$ à partir d'un certain rang. \square

Pour les résultats suivants, on rappelle que l'ensemble vide est toujours linéairement indépendant. La condition de Mangasarian–Fromovitz s'applique donc également au cas particulier où il n'y a pas de contraintes d'égalité.

***Proposition 2.7** (Condition de Mangasarian–Fromovitz). *Supposons les g_i et les h_j continûment différentiables et soit $\mathbf{x}^* \in X$. Supposons la collection $\{\nabla g_i(\mathbf{x}^*) : i = 1, \dots, p\}$ linéairement indépendante. S'il existe $\hat{\mathbf{d}}$ tel que $\nabla g_i(\mathbf{x}^*) \cdot \hat{\mathbf{d}} = 0$ pour tout $i \in [p]$ et $\nabla h_j(\mathbf{x}^*) \cdot \hat{\mathbf{d}} < 0$ pour tout $j \in J(\mathbf{x}^*)$, alors les contraintes de (Q) sont qualifiées en \mathbf{x}^* .*

Démonstration. Prenons $\mathbf{d} \in D'(\mathbf{x}^*)$. Nous allons montrer que \mathbf{d} est une direction admissible.

Soit $\eta > 0$. Définissons $G_i(\mathbf{y}, t) = g_i(\mathbf{x}^* + t(\mathbf{d} + \eta\hat{\mathbf{d}})) + B^T \mathbf{y}$, où B est le jacobien de $\mathbf{g} : \mathbf{x} \mapsto (g_1(\mathbf{x}), \dots, g_p(\mathbf{x}))$ en \mathbf{x}^* , et considérons l'application

$$G: \begin{array}{ll} \mathbb{R}^{p+1} & \longrightarrow \mathbb{R}^p \\ (\mathbf{y}, t) & \longmapsto (G_1(\mathbf{y}, t), \dots, G_p(\mathbf{y}, t)). \end{array}$$

On a $G(\mathbf{0}, 0) = 0$ et le jacobien de $\mathbf{y} \mapsto G(\mathbf{y}, 0)$, qui n'est rien d'autre que BB^T , est inversible d'après les hypothèses. D'après le théorème des fonctions implicites, il existe donc un réel strictement positif ε_0 et une fonction continûment différentiable $\mathbf{y}:]-\varepsilon_0, \varepsilon_0[\rightarrow \mathbb{R}^p$ telle que l'on a $\mathbf{y}(0) = \mathbf{0}$ et $\mathbf{y}'(0) = \mathbf{0}$, et telle que $G(\mathbf{y}(t), t) = 0$ pour tout $t \in]-\varepsilon_0, \varepsilon_0[$. Posons maintenant

$$\bar{\mathbf{d}}(t) = \begin{cases} \mathbf{d} + \eta \hat{\mathbf{d}} & \text{si } t = 0 \\ \mathbf{d} + \eta \hat{\mathbf{d}} + \frac{1}{t} B^T \mathbf{y}(t) & \text{sinon.} \end{cases}$$

Les propriétés de \mathbf{y} assurent que $\bar{\mathbf{d}}$ est continue en 0. D'autre part, l'application $t \mapsto h_j(\mathbf{x}^* + t\bar{\mathbf{d}}(t))$ est continûment différentiable pour $t \in]-\varepsilon_0, \varepsilon_0[$. On a donc $h_j(\mathbf{x}^* + t\bar{\mathbf{d}}(t)) < 0$ pour tout $j \notin J(\mathbf{x}^*)$ et pour tout t dans un voisinage $] -\varepsilon_1, \varepsilon_1[$ de 0, avec $0 < \varepsilon_1 \leq \varepsilon_0$. Pour $j \in J(\mathbf{x}^*)$, on a par hypothèses $\frac{d}{dt} h_j(\mathbf{x}^* + t\bar{\mathbf{d}}(t))$ qui est strictement négatif quand $t = 0$ et donc on a $h_j(\mathbf{x}^* + t\bar{\mathbf{d}}(t)) < 0$ pour tout $j \in J(\mathbf{x}^*)$ et pour tout $t \in [0, \varepsilon_2[$, pour un certain $\varepsilon_2 > 0$. Enfin, on a $g_i(\mathbf{x}^* + t\bar{\mathbf{d}}(t)) = 0$ pour tout $t \in]-\varepsilon_0, \varepsilon_0[$ par définition de $\bar{\mathbf{d}}$. Par conséquent, en posant $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$, le point $\mathbf{x}^* + t\bar{\mathbf{d}}(t)$ est réalisable pour tout $t \in [0, \varepsilon[$.

Le vecteur $\mathbf{d} + \eta \hat{\mathbf{d}}$ est donc une direction admissible pour tout $\eta > 0$. Comme l'ensemble $D(\mathbf{x}^*)$ est fermé (lemme 2.2), \mathbf{d} est également une direction admissible et les contraintes sont qualifiées en \mathbf{x}^* . \square

La proposition précédente a le corollaire suivant.

Corollaire 2.8. *Supposons les g_i et les h_j continûment différentiables et soit $\mathbf{x}^* \in X$. Si la collection $\{\nabla g_i(\mathbf{x}^*) : i \in [p]\} \cup \{\nabla h_j(\mathbf{x}^*) : j \in J(\mathbf{x}^*)\}$ est linéairement indépendante, alors les contraintes de (Q) sont qualifiées en \mathbf{x}^* .*

Démonstration. Considérons la matrice $A = ((\nabla g_i(\mathbf{x}^*))_{i \in [p]}, (\nabla h_j(\mathbf{x}^*))_{j \in J(\mathbf{x}^*)})$. Les lignes de A^T sont linéairement indépendantes par hypothèses. Il existe donc $\hat{\mathbf{d}} \in \mathbb{R}^n$ tel que

$$A^T \hat{\mathbf{d}} = \begin{pmatrix} -\mathbf{1} \\ \mathbf{0} \end{pmatrix}.$$

Ce vecteur $\hat{\mathbf{d}}$ satisfait la condition de la proposition 2.7. \square

2.3 Lagrangien et dualité

Au problème d'optimisation (Q), on associe le *lagrangien*

$$\mathcal{L}: (\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}_+^q \mapsto f(\mathbf{x}) + \sum_{i=1}^p \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^q \mu_j h_j(\mathbf{x}) \in \bar{\mathbb{R}}.$$

Noter que l'on a

$$\sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} +\infty & \text{si } \mathbf{x} \notin X \\ f(\mathbf{x}) & \text{si } \mathbf{x} \in X, \end{cases} \quad (2.1)$$

où l'on travaille toujours avec

$$X = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) = 0, i \in [p] \text{ et } h_j(\mathbf{x}) \leq 0, j \in [q]\}.$$

Par conséquent la valeur optimale du problème (Q) est égale à

$$\inf_{\mathbf{x} \in \mathbb{R}^n} \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

On vérifie par ailleurs aisément que l'on a

$$\inf_{\mathbf{x} \in \mathbb{R}^n} \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \geq \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}). \quad (2.2)$$

En posant

$$d: (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q \longmapsto \inf\{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) : \mathbf{x} \in \mathbb{R}^n\} \in \overline{\mathbb{R}},$$

on peut définir un autre problème d'optimisation, le *dual* de (Q) :

$$\begin{aligned} \text{Max} \quad & d(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.c.} \quad & \boldsymbol{\lambda} \in \mathbb{R}^p, \boldsymbol{\mu} \in \mathbb{R}_+^q. \end{aligned} \quad (\text{D})$$

Le problème (Q) est alors également qualifié de *primal*. On peut montrer que la fonction d est toujours une fonction concave.

La proposition suivante, conséquence de l'inégalité (2.2), montre l'un des intérêts de ce problème dual, et donc, partant, du lagrangien : une procédure standard d'obtention de bornes inférieures à un problème de minimisation.

Proposition 2.9. *Notons v_Q la valeur optimale du problème (Q) et v_D la valeur optimale du problème (D). On a toujours*

$$v_Q \geq v_D.$$

Il est intéressant de noter que cette proposition est valable même si le primal ou le dual sont non bornés ou non réalisables.

L'inégalité donnée par la proposition 2.9 est l'inégalité de *dualité faible* et la différence $v_Q - v_D$ est le *saut de dualité*. Il arrive que le saut de dualité soit réduit à 0, auquel cas on a *dualité forte*. Dans ce cas, les solutions optimales du primal et du dual $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, si elles existent, forment un *point-selle* du lagrangien :

$$\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

En effet, on a alors la chaîne

$$\begin{aligned} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) & \geq \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) & = \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ & = \inf_{\mathbf{x} \in \mathbb{R}^n} \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) & = \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ & \geq \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*). \end{aligned}$$

La première égalité vient de la définition de $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, la seconde de la dualité forte et la troisième de la définition de \mathbf{x}^* . On a donc égalité partout.

Optimisation convexe

3.1 Motivation

Ce chapitre est consacré à certains résultats fondamentaux de l'optimisation convexe, qui s'intéresse aux problèmes d'optimisation d'une fonction convexe sur un domaine convexe. Ce sont des problèmes tout particulièrement importants car, d'une part, de nombreux problèmes naturels s'écrivent sous cette forme, et d'autre part, on dispose d'algorithmes efficaces pour les résoudre. On peut noter de plus que ce sont souvent des problèmes avec des propriétés théoriques remarquables, ce qui constitue également une motivation à leur étude. Il existe une littérature abondante sur le sujet. Boyd et Vandenberghe [1] ont par exemple proposé en 2004 un ouvrage contenant la plupart des résultats importants et les principales applications de l'optimisation convexe.

Dans ce chapitre, on continuera à travailler sur les problèmes (P) et (Q), mais en supposant cette fois que f est une fonction convexe, et que X est convexe pour (P), et que les g_i sont affines et les h_j convexes pour (Q) (les définitions relatives à la convexité sont données dans la Section 3.2).

3.2 Convexité

3.2.1 Notions de base

Une partie C de \mathbb{R}^n est *convexe* si pour tout $\mathbf{x}, \mathbf{y} \in C$ et tout $t \in [0, 1]$, on a $t\mathbf{x} + (1-t)\mathbf{y} \in C$. L'espace \mathbb{R}^n tout entier, en particulier, est convexe. Il est aisé de vérifier que l'intersection d'un nombre quelconque de convexes est convexe. La figure 3.1 illustre cette notion.

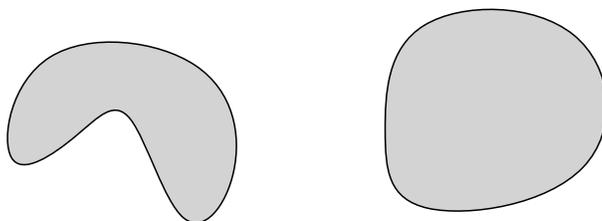


FIGURE 3.1 – Deux sous-ensembles du plan. Celui de gauche n'est pas convexe, celui de droite l'est.

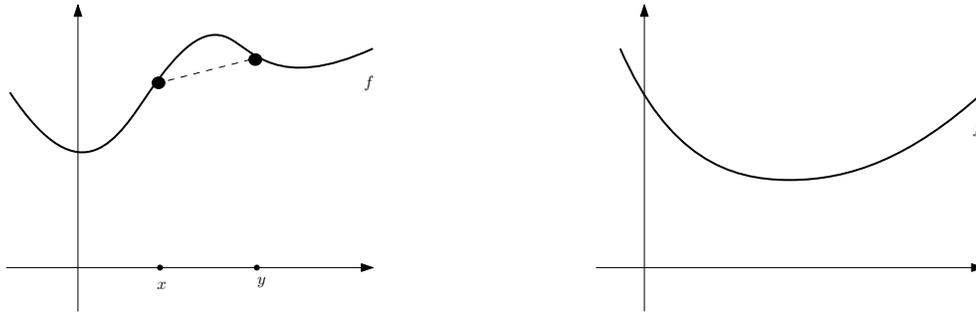


FIGURE 3.2 – Deux fonctions réelles d’une variable réelle. Celle de gauche n’est pas convexe, celle de droite l’est.

Une fonction $f: \mathbb{R}^n \rightarrow \tilde{\mathbb{R}}$ est *convexe* si

$$f(t\mathbf{x} + (1 - t)\mathbf{y}) \leq tf(\mathbf{x}) + (1 - t)f(\mathbf{y}) \tag{3.1}$$

pour tous $\mathbf{x}, \mathbf{y} \in \text{dom } f$ et tout $t \in [0, 1]$. Noter que cela implique en particulier que $\text{dom } f$ est convexe.

Exemple 3.1. La figure 3.2 donne deux exemples de fonctions réelles d’une variable réelle. Celle de gauche n’est pas convexe : on a identifié deux points $(x, f(x))$ et $(y, f(y))$ tel que le segment les reliant n’est pas tout entier au-dessus de la courbe de la fonction.

Celle de droite est convexe : le segment reliant toute paire de points de la courbe est au-dessus de cette dernière, ce qui est exactement le sens de la convexité pour les fonctions réelles d’une variable réelle.

Les fonctions convexes satisfont un grand nombre de propriétés. Une des plus classiques est la suivante, utile dans de nombreuses situations : pour tout entier $k \geq 2$, une fonction f est convexe si et seulement si

$$f\left(\sum_{i=1}^k \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^k \lambda_i f(\mathbf{x}_i) \tag{3.2}$$

pour tous $\mathbf{x}_1, \dots, \mathbf{x}_k \in \text{dom } f$ et tous $\lambda_1, \dots, \lambda_k \in \mathbb{R}_+$ sommant à 1.

Une fonction $f: \mathbb{R}^n \rightarrow \tilde{\mathbb{R}}$ est *strictement convexe* si elle est convexe et si l’inégalité dans (3.1) est stricte pour toute paire \mathbf{x}, \mathbf{y} de points distincts de $\text{dom } f$ et tout $t \in]0, 1[$. Elle est *fortement convexe* s’il existe $\alpha > 0$ tel que

$$f(t\mathbf{x} + (1 - t)\mathbf{y}) \leq tf(\mathbf{x}) + (1 - t)f(\mathbf{y}) - \frac{1}{2}\alpha t(1 - t)\|\mathbf{x} - \mathbf{y}\|_2^2 \tag{3.3}$$

pour tous $\mathbf{x}, \mathbf{y} \in \text{dom } f$ et tout $t \in [0, 1]$. Le réel α est alors le *paramètre* de la forte convexité de f . Une fonction fortement convexe est en particulier strictement convexe, et une fonction strictement convexe est en particulier convexe.

3.2.2 Convexité et continuité

Le lemme suivant permet souvent de ne pas avoir à se soucier de la continuité lorsqu'on travaille avec des fonctions convexes.

***Lemme 3.1.** *Soit $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$ une fonction convexe. Alors f est continue sur l'intérieur de $\text{dom } f$.*

Une telle fonction peut être discontinue sur la frontière de C , comme le montre l'exemple suivant.

Exemple 3.2. Considérons la fonction f définie par

$$f(x) = \begin{cases} 0 & \text{si } x \in [0, 1[\\ 1 & \text{si } x = 1 \\ +\infty & \text{sinon.} \end{cases}$$

C'est une fonction convexe, continue sur l'intérieur de $\text{dom } f = [0, 1]$, mais discontinue en $x = 1$.

Démonstration du lemme 3.1. Prenons \mathbf{y} dans l'intérieur de $\text{dom } f$ et considérons une boule ouverte B de rayon $\delta > 0$ centrée en \mathbf{y} et incluse dans $\text{dom } f$. Définissons les deux fonctions suivantes sur B privée de \mathbf{y} :

$$g: \mathbf{x} \mapsto \left(1 - \frac{\|\mathbf{x} - \mathbf{y}\|_2}{\delta}\right) f(\mathbf{y}) + \frac{\|\mathbf{x} - \mathbf{y}\|_2}{\delta} f\left(\left(1 - \frac{\delta}{\|\mathbf{x} - \mathbf{y}\|_2}\right) \mathbf{y} + \frac{\delta}{\|\mathbf{x} - \mathbf{y}\|_2} \mathbf{x}\right)$$

$$h: \mathbf{x} \mapsto \left(1 + \frac{\|\mathbf{x} - \mathbf{y}\|_2}{\delta}\right) f(\mathbf{y}) - \frac{\|\mathbf{x} - \mathbf{y}\|_2}{\delta} f\left(\left(1 + \frac{\delta}{\|\mathbf{x} - \mathbf{y}\|_2}\right) \mathbf{y} - \frac{\delta}{\|\mathbf{x} - \mathbf{y}\|_2} \mathbf{x}\right).$$

Une conséquence directe de la convexité de f est l'encadrement suivant valable en tout $\mathbf{x} \in B \setminus \{\mathbf{y}\}$:

$$h(\mathbf{x}) \leq f(\mathbf{x}) \leq g(\mathbf{x}).$$

Comme $\lim_{\mathbf{x} \rightarrow \mathbf{y}} g(\mathbf{x}) = \lim_{\mathbf{x} \rightarrow \mathbf{y}} h(\mathbf{x}) = f(\mathbf{y})$, on obtient la continuité de f en \mathbf{y} . □

Pour les fonctions fortement convexes, on a plus.

Lemme 3.2. *Soit $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$ fortement convexe. Alors f est coercive.*

Démonstration. Quitte à traduire la fonction, on peut supposer que $f(\mathbf{0}) < +\infty$. Il existe alors $\delta > 0$ telle que $f(\mathbf{x}) < +\infty$ pour tout \mathbf{x} tel que $\|\mathbf{x}\|_2 \leq \delta$. Écrivons l'équation (3.3) avec $\mathbf{y} = \mathbf{0}$ et $t = \frac{\delta}{\|\mathbf{x}\|_2}$, pour $\|\mathbf{x}\|_2 \geq 2\delta$. Un calcul direct donne

$$\left(f\left(\frac{\delta}{\|\mathbf{x}\|_2} \mathbf{x}\right) - \left(1 - \frac{\delta}{\|\mathbf{x}\|_2}\right) f(\mathbf{0})\right) \frac{\|\mathbf{x}\|_2}{\delta} + \frac{1}{2} \alpha \left(1 - \frac{\delta}{\|\mathbf{x}\|_2}\right) \|\mathbf{x}\|_2^2 \leq f(\mathbf{x}).$$

Soit r le maximum que $|f(\mathbf{z})|$ puisse prendre pour \mathbf{z} dans la boule fermée de rayon δ centrée en $\mathbf{0}$. Ce maximum existe par compacité (le lemme 3.1 assurant la continuité de f). On a alors

$$-2\frac{r}{\delta} \|\mathbf{x}\|_2 + \frac{1}{4} \alpha \|\mathbf{x}\|_2^2 \leq f(\mathbf{x})$$

et donc bien $f(\mathbf{x}) \rightarrow +\infty$ quand $\|\mathbf{x}\| \rightarrow +\infty$. □

3.2.3 Cas différentiable

Il existe une autre caractérisation de la convexité lorsque f est différentiable.

Proposition 3.3. *Soit $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$ différentiable. Alors f est convexe si et seulement si l'on a $\text{dom } f$ convexe et*

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})$$

pour tout $\mathbf{x}, \mathbf{y} \in \text{dom } f$.

Démonstration. Si f est convexe, on a par définition $f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y})$ pour tout $\mathbf{x}, \mathbf{y} \in \text{dom } f$ et tout $t \in [0, 1]$. Cette inégalité peut se réécrire $\frac{1}{t}(f(\mathbf{y} + t(\mathbf{x} - \mathbf{y})) - f(\mathbf{y})) \leq f(\mathbf{x}) - f(\mathbf{y})$. En faisant tendre t vers 0, on obtient $\nabla f(\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) \leq f(\mathbf{x}) - f(\mathbf{y})$ qui est ce qu'il fallait montrer.

Réciproquement, supposons que $\text{dom } f$ est convexe et que f satisfait l'inégalité de l'énoncé. Prenons \mathbf{x}, \mathbf{y} dans $\text{dom } f$ et posons $\mathbf{z} = t\mathbf{x} + (1-t)\mathbf{y}$. Comme $\text{dom } f$ est convexe, on a $\mathbf{z} \in \text{dom } f$. L'inégalité de l'énoncé donne alors les deux inégalités suivantes :

$$f(\mathbf{x}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z}) \cdot (\mathbf{x} - \mathbf{z}) \quad \text{et} \quad f(\mathbf{y}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z}) \cdot (\mathbf{y} - \mathbf{z}).$$

En prenant une combinaison convexe de ces deux inégalités, on obtient $tf(\mathbf{x}) + (1-t)f(\mathbf{y}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z}) \cdot \mathbf{0}$, ce qui est exactement ce qu'il fallait prouver. \square

Les propositions suivantes, dont les preuves sont omises, peuvent en particulier être utiles pour établir la convexité d'une fonction. On rappelle qu'on note $H(f)$ la hessienne de f .

Proposition 3.4. *Soit f deux fois différentiable. Alors f est convexe si et seulement si l'on a $\text{dom } f$ convexe et sa hessienne est semi-définie positive en tout point de $\text{dom } f$.*

Noter que la condition suivante est simplement une condition suffisante. La fonction $x \in \mathbb{R} \mapsto x^4 \in \mathbb{R}$ a sa hessienne nulle en $x = 0$ et est pourtant bien strictement convexe.

Proposition 3.5. *Soit f deux fois différentiable. Alors si l'on a $\text{dom } f$ convexe et la hessienne de f est définie positive en tout point de $\text{dom } f$, alors f est strictement convexe.*

Proposition 3.6. *Soit f deux fois différentiable. Alors f est fortement convexe de paramètre $\alpha > 0$ si et seulement si l'on a $\text{dom } f$ convexe et $H(f) - \alpha I_n$ semi-définie positive en tout point de $\text{dom } f$.*

Proposition 3.7. *Soit C un convexe de \mathbb{R}^n et f différentiable partout sur C . Alors f est fortement convexe de paramètre $\alpha > 0$ si et seulement si*

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})) \cdot (\mathbf{x} - \mathbf{y}) \geq \alpha \|\mathbf{x} - \mathbf{y}\|_2^2$$

pour tout $\mathbf{x}, \mathbf{y} \in C$.

3.3 Caractérisation de la solution optimale

Dans le cas où la fonction à minimiser est convexe et que le domaine est convexe, le théorème 2.3 devient une condition suffisante d'optimalité. Rappelons le problème auquel il s'applique.

$$\begin{aligned} \text{Min} \quad & f(\mathbf{x}) \\ \text{s.c.} \quad & \mathbf{x} \in X. \end{aligned} \tag{P}$$

Théorème 3.8. *Soit $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$. Supposons f convexe et différentiable et X convexe. Alors $\mathbf{x}^* \in X \cap \text{dom } f$ est une solution optimale du problème (P) si et seulement si on a $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour toute direction admissible \mathbf{d} en \mathbf{x}^* .*

Démonstration. D'après le théorème 2.3, si \mathbf{x}^* est une solution optimale du problème (P), alors on a $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour toute direction admissible \mathbf{d} en \mathbf{x}^* . Il reste donc à prouver l'implication réciproque. Supposons que $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour toute direction admissible \mathbf{d} en \mathbf{x}^* et montrons qu'alors \mathbf{x}^* est une solution optimale. D'après la proposition 3.3, on a

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*)$$

pour tout $\mathbf{x} \in X$. Comme X est convexe, $\mathbf{x} - \mathbf{x}^*$ est une direction admissible en \mathbf{x}^* si l'on a $\mathbf{x} \neq \mathbf{x}^*$. Cette inégalité implique que $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ pour tout $\mathbf{x} \in X$. \square

Dans le cas où il n'y a pas de contrainte, i.e. que l'on cherche simplement le minimum global d'une fonction $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$ convexe différentiable, le théorème 3.8 se spécialise en la condition bien connue d'« annulation de la dérivée », qui s'exprime ainsi

$$\mathbf{x}^* \in \text{dom } f \text{ est le minimum global de } f \text{ si et seulement si } \nabla f(\mathbf{x}^*) = \mathbf{0}.$$

Cet énoncé se retrouve bien à partir de ce théorème. Si $\nabla f(\mathbf{x}^*) = \mathbf{0}$, alors on a bien $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour toute direction admissible. Pour la réciproque, notons que sur \mathbb{R}^n tout \mathbf{d} est direction admissible. Par conséquent, si $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ pour toute direction admissible \mathbf{d} , on a $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$ et $\nabla f(\mathbf{x}^*) \cdot (-\mathbf{d}) \geq 0$ pour tout $\mathbf{d} \in \mathbb{R}^n$, et donc $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

En ce qui concerne l'unicité de la solution optimale, on a le résultat suivant.

Proposition 3.9. *Soit $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$. Supposons f strictement convexe et X convexe. Si le problème (P) admet une solution optimale de valeur finie, alors cette solution est unique.*

Démonstration. Supposons que l'on ait deux solutions optimales distinctes \mathbf{x}^* et $\mathbf{x}^\#$. On a alors

$$f\left(\frac{1}{2}(\mathbf{x}^* + \mathbf{x}^\#)\right) < \frac{1}{2}(f(\mathbf{x}^*) + f(\mathbf{x}^\#)),$$

ce qui contredit l'optimalité de \mathbf{x}^* et $\mathbf{x}^\#$. \square

Si f est fortement convexe, on est de plus assuré de l'existence d'une solution optimale.

Proposition 3.10. *Soit $f: \mathbb{R}^n \rightarrow \widetilde{\mathbb{R}}$. Supposons f fortement convexe et continue, et $X \cap \text{dom } f$ convexe, fermé et non vide. Alors le problème (P) admet une et une seule solution optimale.*

Démonstration. Toute fonction fortement convexe est strictement convexe. D'après la proposition 3.9, il suffit donc de prouver que si f est fortement convexe, alors le problème (P) admet une solution optimale. Toute fonction fortement convexe est coercive (lemme 3.2. Le théorème 2.1 permet alors de conclure. \square

Considérons maintenant le problème (Q) dans le cas où l'ensemble des solutions réalisables forme un ensemble convexe, dont on rappelle la forme :

$$\begin{aligned} \text{Min} \quad & f(\mathbf{x}) \\ \text{s.c.} \quad & g_i(\mathbf{x}) = 0 \quad i = 1, \dots, p \\ & h_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, q, \end{aligned} \quad (\text{Q})$$

où f , les g_i et les h_j sont des fonctions de \mathbb{R}^n dans $\widetilde{\mathbb{R}}$. Il est aisé de vérifier que l'ensemble des solutions réalisables de ce problème forme un ensemble convexe lorsque les fonctions g_i sont affines et les fonctions h_j convexes. Un résultat fondamental de l'optimisation convexe est que pour un tel problème, la condition nécessaire du théorème de Kuhn et Tucker (théorème 2.4 de la Section 2.2) devient suffisante.

On rappelle que l'on pose

$$X = \{ \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) = 0, i \in [p] \quad \text{et} \quad h_j(\mathbf{x}) \leq 0, j \in [q] \}.$$

Théorème 3.11. *Supposons f et les h_j convexes et différentiables, et les g_i affines. Soit $\mathbf{x}^* \in X \cap \text{dom } f$. Si \mathbf{x}^* est une solution optimale du problème (Q) et si les contraintes sont qualifiées en \mathbf{x}^* , alors il existe p réels λ_i et q réels positifs ou nuls μ_j vérifiant*

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^p \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^q \mu_j \nabla h_j(\mathbf{x}^*) = \mathbf{0} \quad \text{et} \quad \mu_j h_j(\mathbf{x}^*) = 0 \quad \forall j \in [q].$$

Réciproquement, s'il existe p réels λ_i et q réels positifs μ_j vérifiant ces relations, alors \mathbf{x}^ est une solution optimale.*

Noter que pour la réciproque, la qualification des contraintes en \mathbf{x}^* n'est pas requise.

Démonstration du théorème 3.11. Le théorème 2.4 assure que si \mathbf{x}^* est une solution optimale, on a bien l'existence des λ_i et μ_j comme assuré par l'énoncé. Prouvons donc la réciproque et supposons l'existence de tels λ_i et μ_j . Prenons une direction admissible \mathbf{d} en \mathbf{x}^* . Il a été noté à la Section 2.2.1 qu'alors nécessairement $\mathbf{d} \in D'(\mathbf{x}^*)$. On a donc $\nabla g_i(\mathbf{x}^*) \cdot \mathbf{d} = 0$ pour $i \in [p]$ et $\nabla h_j(\mathbf{x}^*) \cdot \mathbf{d} \leq 0$ pour $j \in J(\mathbf{x}^*)$. D'autre part, comme $\sum_{j=1}^q \mu_j h_j(\mathbf{x}^*) = 0$, on a $\mu_j = 0$ si $j \notin J(\mathbf{x}^*)$. Par conséquent, l'égalité satisfaite par les λ_i et les μ_j fournit l'inégalité $\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0$, ce qui implique l'optimalité de \mathbf{x}^* grâce au théorème 3.8. \square

3.4 Dualité forte

Le théorème 3.11 a une conséquence intéressante sur le problème dual du problème (Q), i.e. le problème (D) donné en Section 2.3, car il donne une condition sous laquelle on a dualité forte.

Corollaire 3.12. *Supposons f et les h_j convexes et différentiables et les g_i affines. Supposons de plus que les contraintes de (Q) soient qualifiées partout et que le problème primal ait au moins une solution optimale. Alors le saut de dualité est nul : la valeur optimale du primal est égale à la valeur optimale du dual.*

Démonstration. Soit \mathbf{x}^* une solution optimale de (Q). Notons $\boldsymbol{\lambda}^* = (\lambda_i^*)$ et $\boldsymbol{\mu}^* = (\mu_j^*)$ les multiplicateurs de Lagrange dont l'existence est assurée par le théorème 3.11. Notons que $\mathbf{x} \mapsto \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est une fonction convexe en \mathbf{x} car f et les h_j sont convexes, les g_i sont affines et les μ_j^* sont positifs. L'équation

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^p \lambda_i^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^q \mu_j^* \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

assure alors que \mathbf{x}^* est un minimum global de $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ (théorème 3.8 appliqué à $\mathbf{x} \mapsto \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ et à $X = \mathbb{R}^n$). On a donc $\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. D'autre part, $\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*)$. On peut alors écrire la chaîne

$$\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \geq \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \geq \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*),$$

où la première inégalité est donnée par l'équation (2.2) et la seconde est évidente, et où l'égalité est celle qui vient d'être justifiée au début de la démonstration. Toutes les inégalités de cette chaîne sont donc des égalités et on a

$$f(\mathbf{x}^*) = \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}). \quad \square$$

Noter qu'alors les multiplicateurs de Lagrange $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ associés à la solution optimale \mathbf{x}^* du primal forment une solution optimale du dual car

$$v_D = \sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^p \times \mathbb{R}_+^q} \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*).$$

De plus, conformément à ce qui a été expliqué à la Section 2.3, le point $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est alors par définition un point-selle du lagrangien.

3.5 Algorithmes

3.5.1 Quelques remarques préliminaires

Les algorithmes qui vont suivre ne permettent pas en général de calculer la solution optimale d'un problème d'optimisation de manière exacte. Cela est dû à la nature des algorithmes, mais aussi au fait que les solutions optimales n'ont pas de raison d'être exprimables sous une forme close. Les algorithmes prendront donc systématiquement en données une précision ε . La situation sera différente dans la cas de l'optimisation linéaire ; voir Chapitre 4.

Les deux algorithmes qui vont suivre utilisent la projection sur un convexe fermé. La *projection* sur un convexe fermé non vide K , notée P_K , est définie comme suit : $P_K(\mathbf{x})$ est l'unique $\mathbf{z}^* \in K$ réalisant le minimum de $\|\mathbf{z} - \mathbf{x}\|_2$. Un tel \mathbf{z}^* existe et est bien unique car il réalise le minimum de $\mathbf{z} \mapsto \|\mathbf{z} - \mathbf{x}\|_2^2$, fonction fortement convexe, sur un convexe fermé et l'on peut donc appliquer la proposition 3.9. On a de plus le lemme suivant.

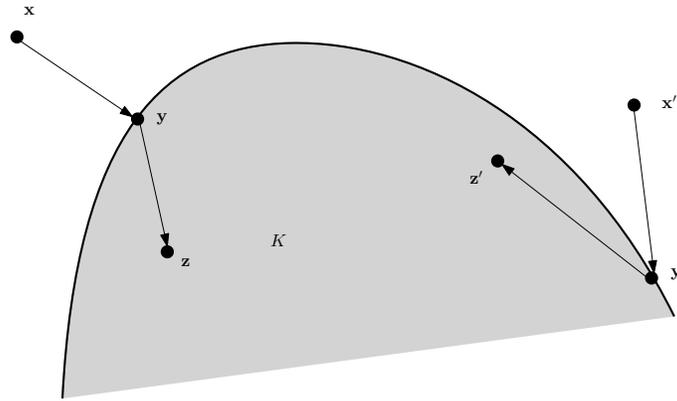


FIGURE 3.3 – Illustration du lemme 3.13 : le point \mathbf{y} est le projeté du point \mathbf{x} , alors que le point \mathbf{y}' n'est pas le projeté du point \mathbf{x}' .

Lemme 3.13. Soit K un convexe fermé. Pour $\mathbf{x} \in \mathbb{R}^n$ et $\mathbf{y} \in K$, on a $\mathbf{y} = P_K(\mathbf{x})$ si et seulement si pour tout point $\mathbf{z} \in K$, on a $(\mathbf{y} - \mathbf{x}) \cdot (\mathbf{z} - \mathbf{y}) \geq 0$.

Démonstration. Il suffit de noter que $\mathbf{z} - P_K(\mathbf{x})$ est une direction admissible pour le problème de minimisation dont la projection est solution et d'appliquer le théorème 3.8. \square

Exemple 3.3. Ce lemme est illustré sur la figure 3.3. Le point \mathbf{y} est bien le projeté du point \mathbf{x} sur K . On a $(\mathbf{y} - \mathbf{x}) \cdot (\mathbf{z} - \mathbf{y}) \geq 0$ pour tout $\mathbf{z} \in K$, en particulier pour celui indiqué. En revanche, le point \mathbf{y}' n'est pas le projeté du point \mathbf{x}' sur K . On a $(\mathbf{y}' - \mathbf{x}') \cdot (\mathbf{z}' - \mathbf{y}') < 0$.

Une conséquence immédiate du lemme 3.13 est le résultat suivant.

Lemme 3.14. Soit K un convexe fermé. Alors P_K est une application lipschitzienne de paramètre 1.

Démonstration. Il suffit d'écrire

$$\|P_K(\mathbf{x}) - P_K(\mathbf{y})\|_2^2 = (P_K(\mathbf{x}) - \mathbf{x}) \cdot (P_K(\mathbf{x}) - P_K(\mathbf{y})) + (\mathbf{x} - \mathbf{y}) \cdot (P_K(\mathbf{x}) - P_K(\mathbf{y})) + (\mathbf{y} - P_K(\mathbf{y})) \cdot (P_K(\mathbf{x}) - P_K(\mathbf{y})),$$

de noter que le lemme 3.13 permet de voir que les premier et troisième termes de cette égalité sont négatifs et d'utiliser le théorème de Cauchy-Schwartz. \square

3.5.2 Algorithme du gradient projeté

L'algorithme qui suit, l'*algorithme du gradient projeté*, est un grand classique de l'optimisation. Les idées le sous-tendant remontent à Cauchy [3]. Il s'applique à un convexe fermé X de \mathbb{R}^n .

Données : Une précision $\varepsilon > 0$, un vecteur $\mathbf{x}^0 \in X$ et un pas $\ell > 0$
Sortie : Une solution approchée

$k \leftarrow 0$;
repeat
 $\mathbf{x}^{k+1} \leftarrow P_X(\mathbf{x}^k - \ell \nabla f(\mathbf{x}^k))$;
 $k \leftarrow k + 1$;
until $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \leq \varepsilon$;
return \mathbf{x}^{k+1}

Algorithme 1 : Algorithme du gradient projeté

Quand X est l'espace \mathbb{R}^n tout entier, on parle alors simplement d'*algorithme du gradient*. Si P_X se calcule simplement, cet algorithme est d'implémentation aisée, et a un bon comportement en général. Malheureusement, à part lorsque X est vraiment particulier (boule ou pavé), la projection P_X est difficile à calculer, souvent d'une complexité équivalente au problème de départ.

Le théorème suivant assure la convergence vers la solution optimale, sous certaines conditions. On y considère la suite infinie (\mathbf{x}^k) obtenue en oubliant le critère d'arrêt.

Théorème 3.15. *Supposons les g_i affines, les h_j convexes, f convexe différentiable et ∇f lipschitzienne de paramètre L . On a les résultats de convergence suivants.*

1. Si (Q) admet une solution optimale \mathbf{x}^* et $\ell < \frac{1}{L}$, alors on a l'inégalité

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) \leq \frac{1}{2\ell k} \|\mathbf{x}^0 - \mathbf{x}^*\|_2^2$$

pour tout k et on a donc en particulier convergence de la suite $(f(\mathbf{x}^k))$ vers la valeur optimale. Si cette solution optimale \mathbf{x}^* est unique, alors on a de plus la convergence de la suite (\mathbf{x}^k) vers \mathbf{x}^* .

2. Si f est fortement convexe de paramètre α et $\ell < \frac{2\alpha}{L^2}$, alors on a l'inégalité

$$\|\mathbf{x}^k - \mathbf{x}^*\|_2 \leq (1 - 2\ell\alpha + \ell^2 L^2)^{k/2} \|\mathbf{x}^0 - \mathbf{x}^*\|_2$$

pour tout k , où \mathbf{x}^* est l'unique solution optimale de (Q), et on a donc en particulier convergence de la suite (\mathbf{x}^k) vers \mathbf{x}^* .

Sous les conditions du point 2, lorsque $\ell < \frac{2\alpha}{L^2}$, on a une convergence rapide de l'algorithme, dite *linéaire* (car le nombre de décimales déterminées correctement croît asymptotiquement de manière linéaire).

Démonstration du théorème 3.15. Prouvons d'abord le point 1. Comme ∇f est lipschitzienne de paramètre L , on a pour tout \mathbf{x}, \mathbf{y} de \mathbb{R}^n la relation $f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2$. (Cela se prouve aisément en considérant $\int_0^1 (g'(t) - g'(0)) dt$, avec $g(t) = f((1-t)\mathbf{x} + t\mathbf{y})$.) En prenant $\mathbf{y} = \mathbf{x}^{k+1}$ and $\mathbf{x} = \mathbf{x}^k$, cette inégalité donne

$$f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k) \cdot (\mathbf{x}^{k+1} - \mathbf{x}^k) + \frac{L}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2. \quad (3.4)$$

En utilisant le fait que f est convexe, la proposition 3.3 appliquée en $\mathbf{x} = \mathbf{x}^*$ et $\mathbf{y} = \mathbf{x}^k$ permet de substituer $f(\mathbf{x}^k)$ dans l'inégalité et d'obtenir

$$f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^k) \cdot (\mathbf{x}^{k+1} - \mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2.$$

En utilisant le lemme 3.13 avec $\mathbf{x} = \mathbf{x}^k - \ell \nabla f(\mathbf{x}^k)$ et $\mathbf{z} = \mathbf{x}^*$, un calcul immédiat fournit

$$f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^*) + \frac{1}{2\ell} (\|\mathbf{x}^k - \mathbf{x}^*\|_2^2 - \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2^2) + \frac{1}{2} \left(L - \frac{1}{\ell} \right) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2.$$

Comme $\ell < \frac{1}{L}$, cette dernière inégalité donne

$$f(\mathbf{x}^{k+1}) - f(\mathbf{x}^*) \leq \frac{1}{2\ell} (\|\mathbf{x}^k - \mathbf{x}^*\|_2^2 - \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2^2).$$

En sommant sur les k premières itérations, on obtient $\sum_{i=1}^k (f(\mathbf{x}^i) - f(\mathbf{x}^*)) \leq \frac{1}{2\ell} \|\mathbf{x}^0 - \mathbf{x}^*\|_2^2$. L'inégalité (3.4) et le lemme 3.13 avec $\mathbf{x} = \mathbf{x}^k - \ell \nabla f(\mathbf{x}^k)$ et $\mathbf{z} = \mathbf{x}^k$ impliquent la décroissance de la suite $(f(\mathbf{x}^k))$. Par conséquent, $f(\mathbf{x}^k) - f(\mathbf{x}^*) \leq \frac{1}{k} \sum_{i=1}^k (f(\mathbf{x}^i) - f(\mathbf{x}^*))$.

Nous prouvons maintenant la convergence de la suite (\mathbf{x}^k) lorsque la solution optimale est unique. Cela conclura la preuve du point 1. Soit $\varepsilon > 0$. Notons r la valeur minimale prise par f sur la sphère de centre \mathbf{x}^* et de rayon ε . Cette valeur existe par compacité. Soit K tel que $f(\mathbf{x}^k) < r$ pour tout $k \geq K$. La convexité de f va nous permettre de voir que $f(\mathbf{y}) \geq r$ pour tout \mathbf{y} tel que $\|\mathbf{y} - \mathbf{x}^*\|_2 \geq \varepsilon$, ce qui impliquera que $\|\mathbf{x}^k - \mathbf{x}^*\|_2 < \varepsilon$ pour tout $k \geq K$. Considérons donc un tel \mathbf{y} . Posons $t = \frac{\varepsilon}{\|\mathbf{y} - \mathbf{x}^*\|_2}$. Par les hypothèses sur \mathbf{y} , on a $t \in [0, 1]$. Pour un tel t , on a $(1-t)\mathbf{x}^* + t\mathbf{y}$ sur la sphère de centre \mathbf{x}^* et de rayon ε . La convexité implique alors $(1-t)f(\mathbf{x}^*) + tf(\mathbf{y}) \geq r$, et en utilisant $f(\mathbf{x}^*) < r$, on obtient $f(\mathbf{y}) \geq r$, comme attendu.

Prouvons maintenant le point 2. Posons $\theta(\mathbf{x}) = \mathbf{x} - \ell \nabla f(\mathbf{x})$. On a

$$\|\theta(\mathbf{x}^k) - \theta(\mathbf{x}^*)\|_2^2 = \|\mathbf{x}^k - \mathbf{x}^*\|_2^2 - 2\ell(\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^*)) \cdot (\mathbf{x}^k - \mathbf{x}^*) + \ell^2 \|\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^*)\|_2^2.$$

Par conséquent, en utilisant la forte convexité de f (proposition 3.7) et le caractère lipschitzien de ∇f , on obtient $\|\theta(\mathbf{x}^k) - \theta(\mathbf{x}^*)\|_2 \leq \sqrt{1 - 2\ell\alpha + \ell^2 L^2} \|\mathbf{x}^k - \mathbf{x}^*\|_2$. Comme P_X est lipschitzienne de paramètre 1 (lemme 3.14), on en déduit $\|\mathbf{x}^{k+1} - P_X(\theta(\mathbf{x}^*))\|_2 \leq \sqrt{1 - 2\ell\alpha + \ell^2 L^2} \|\mathbf{x}^k - \mathbf{x}^*\|_2$. On termine la preuve en prouvant que \mathbf{x}^* est point fixe de $P_X \circ \theta$, une récurrence immédiate permettant alors de conclure. Le lemme 3.13 avec $\mathbf{x} = \mathbf{x}^* - \ell \nabla f(\mathbf{x}^*)$ et $\mathbf{z} = \mathbf{x}^*$ donne

$$\ell \nabla f(\mathbf{x}^*) \cdot (P_X(\theta(\mathbf{x}^*)) - \mathbf{x}^*) \leq -\|P_X(\theta(\mathbf{x}^*)) - \mathbf{x}^*\|_2^2.$$

Le théorème 3.8 montre qu'alors nécessairement $\|P_X(\theta(\mathbf{x}^*)) - \mathbf{x}^*\|_2 = 0$. □

3.5.3 Algorithme d'Uzawa

Nous présentons un autre algorithme qui s'applique à peu près sous les mêmes conditions que l'algorithme du gradient projeté, mais qui ne requiert pas de savoir projeter efficacement sur l'ensemble des solutions réalisables. C'est l'algorithme d'Uzawa, proposé en 1958 [11].

Données : Une précision $\varepsilon > 0$, un vecteur $\boldsymbol{\lambda}^0 \in \mathbb{R}^p$, un vecteur $\boldsymbol{\mu}^0 \in \mathbb{R}_+^q$ et un pas $\ell > 0$

Sortie : Une solution approchée

$k \leftarrow 0$;
repeat
 $\mathbf{x}^k \leftarrow \arg \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$;
 $\boldsymbol{\lambda}^{k+1} \leftarrow \boldsymbol{\lambda}^k + \ell \mathbf{g}(\mathbf{x}^k)$;
 $\boldsymbol{\mu}^{k+1} \leftarrow P_{\mathbb{R}_+^q}(\boldsymbol{\mu}^k + \ell \mathbf{h}(\mathbf{x}^k))$;
 $k \leftarrow k + 1$;
until $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \leq \varepsilon$;
return \mathbf{x}^{k+1}

Algorithme 2 : Algorithme d'Uzawa

On a utilisé la notation $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_p(\mathbf{x}))$ et $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), \dots, h_q(\mathbf{x}))$.

L'intuition générale est la suivante : l'algorithme d'Uzawa est l'algorithme du gradient projeté appliqué au problème dual. En notant d la fonction objectif du problème dual de (Q) (comme dans la section 2.3), on a $d(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) = \mathcal{L}(\mathbf{x}^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$, et $(\mathbf{g}(\mathbf{x}^k), \mathbf{h}(\mathbf{x}^k))$ est le gradient de d en $(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$. La projection est aisée pour le dual car l'ensemble de ses solutions réalisables est $\mathbb{R}^p \times \mathbb{R}_+^q$.

Le théorème suivant assure la convergence l'algorithme vers la valeur optimale, ou même vers la solution optimale, sous certaines conditions. De manière similaire au cas du gradient projeté, on y considère la suite infinie $((\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k))$ obtenue en oubliant le critère d'arrêt. Noter que les conditions assurent que l'on a dualité forte, et donc que valeurs optimales des problèmes primal et dual coïncident. La preuve est omise car elle utilise des résultats un tout petit peu plus avancés de convexité que ceux présentés dans ce cours.

Théorème 3.16. *Supposons les g_i affines et les h_j convexes, f convexe différentiable et $\mathbf{x} \mapsto (\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}))$ lipschitzienne de paramètre L . Supposons également les contraintes qualifiées en tout point de X . On a les résultats de convergence suivants.*

1. Si (Q) admet une solution optimale et $\ell < \frac{1}{L}$, alors on a l'inégalité

$$d(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) - d(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) \leq \frac{1}{2\ell k} (\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|_2^2 + \|\boldsymbol{\mu}^0 - \boldsymbol{\mu}^*\|_2^2)$$

pour tout k , où $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est solution optimale du dual de (Q), et on a donc en particulier convergence de la suite $(d(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k))$ vers la valeur optimale de (Q).

2. Si f est fortement convexe de paramètre α et $\ell < \frac{2\alpha}{L^2}$, alors on a convergence de la suite (\mathbf{x}^k) vers l'unique solution optimale de (Q).

Une grande variété d’algorithmes

Cette section sur les algorithmes ne constitue qu’une brève initiation aux questions algorithmiques du monde de l’optimisation convexe. C’est en réalité un domaine gigantesque, où la recherche est encore très active, en particulier suite aux succès de l’optimisation convexe en apprentissage automatique (ou “machine learning”).

Par exemple, lorsqu’on parle d’algorithme du gradient, on fait en général référence à toute une famille d’algorithmes, et l’algorithme 1 de la Section 3.5.2 n’en est en fait que la version la plus épurée. Une amélioration simple consiste à autoriser le pas ℓ à varier à chaque itération. En choisissant bien ce pas, on obtient l’*algorithme du gradient à pas optimal*, qui reste très simple à programmer et dont la vitesse de convergence peut être substantiellement supérieure à la version donnée ici ; voir par exemple la Section 9.3 du livre de Boyd et Vandenberghe [1].

D’autres algorithmes peuvent être particulièrement efficaces dans le cas de l’optimisation convexe avec contraintes (cas où, rappelons-le, l’algorithme du gradient projeté est d’une efficacité limitée), comme l’algorithme de Frank–Wolfe [2, Chapitre 3] ou les méthodes de points intérieurs [1, Chapitre 11].

Optimisation linéaire

4.1 Motivation

L'*optimisation linéaire* consiste à résoudre des problèmes de la forme

$$\begin{aligned} \text{Min} \quad & \mathbf{c} \cdot \mathbf{x} \\ \text{s.c.} \quad & A\mathbf{x} \leq \mathbf{b}, \end{aligned} \tag{4.1}$$

où A est une matrice à m lignes et n colonnes et où \mathbf{b} et \mathbf{c} sont deux vecteurs respectivement dans \mathbb{R}^m et dans \mathbb{R}^n . C'est donc un cas particulier de l'optimisation convexe. Les problèmes d'optimisation linéaire apparaissent naturellement dans de nombreux contextes et possèdent, tout comme l'optimisation convexe, des propriétés remarquables. On peut se demander pourquoi s'intéresser à ce cas particulier de l'optimisation convexe. Il y a de nombreuses façons de répondre à cette question. Comme l'optimisation linéaire est un cas particulier de l'optimisation convexe, on peut s'attendre à des résultats plus forts. C'est vrai, mais il y a une autre raison : la plupart des résultats du Chapitre 3, et en particulier ceux relatifs à la convergence des algorithmes, ne sont valables que si la fonction objectif est fortement convexe. Or, dans le cas de l'optimisation linéaire, la fonction objectif n'est même pas strictement convexe. Il faut cependant noter qu'il existe des algorithmes puissants pour l'optimisation convexe couvrant de nombreux cas, dont ceux de l'optimisation linéaire, comme par exemple les *méthodes de points intérieurs* évoquées à la fin de la Section 3.5, mais ils dépassent le cadre de ce cours. Nous allons voir l'algorithme le plus fréquemment utilisée en pratique, l'algorithme du simplexe, et dont le principe est relativement simple. Parmi les ouvrages récents sur l'optimisation linéaire, on peut signaler celui publié en 2007 par Matoušek et Gärtner [9] qui, tout en étant court, couvre la plupart des résultats théoriques et présente des applications variées.

4.2 Formulations équivalentes et hypothèses pour le chapitre

Le problème (4.1) est sous *forme inéquationnelle*. Il existe d'autres formes possibles, comme la *forme canonique* :

$$\begin{aligned} \text{Min} \quad & \mathbf{c} \cdot \mathbf{x} \\ \text{s.c.} \quad & A\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Une des formes les plus utilisées est la *forme standard* :

$$\begin{aligned} \text{Min} \quad & \mathbf{c} \cdot \mathbf{x} \\ \text{s.c.} \quad & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{PL}$$

Quitte à changer les dimensions des problèmes et les définitions de A , \mathbf{b} et \mathbf{c} , tout problème sous l'une des trois formes (inéquationnelle, canonique, standard) peut être réécrit sous l'une quelconque des deux autres, voir Exercices. Par exemple, un problème sous forme canonique peut être mis sous forme standard en écrivant

$$\begin{aligned} \text{Min} \quad & \mathbf{c} \cdot \mathbf{x} \\ \text{s.c.} \quad & A\mathbf{x} + \mathbf{y} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \\ & \mathbf{y} \geq \mathbf{0}. \end{aligned}$$

La matrice (A, I) joue alors le rôle de la matrice A dans (PL).

Dans la suite de ce chapitre, nous allons nous focaliser sur la forme standard et on notera m le nombre de lignes de A et n son nombre de colonnes. Si A n'est pas de rang m , alors soit le système $A\mathbf{x} = \mathbf{b}$ n'a pas de solution, et donc alors le problème (PL) n'a pas de solution réalisable (et sa valeur optimale est $+\infty$), soit ce système possède des lignes redondantes que l'on peut supprimer sans changer l'ensemble des solutions réalisables. On peut donc faire l'hypothèse suivante, qui sera donc supposée vérifiée pour le reste du chapitre.

La matrice A est de rang m .

Noter que cela implique alors en particulier que $m \leq n$.

4.3 Existence et caractérisation d'une solution optimale

4.3.1 Bases

Pour un sous-ensemble $K \subseteq [n]$, on notera A_K la matrice réduite aux colonnes $(A_i)_{i \in K}$. De même, si \mathbf{x} est un vecteur de \mathbb{R}^n , on notera \mathbf{x}_K le vecteur réduit aux composantes $(x_i)_{i \in K}$.

Un sous-ensemble $B \subseteq [n]$ est une *base* du problème (PL) si A_B est inversible. Noter que cette notion n'est pas *stricto sensu* celle de base en algèbre linéaire. Au sens de l'algèbre linéaire, $(A_i)_{i \in B}$ est une base de \mathbb{R}^m . Un tel B est une *base réalisable* si on a de plus $A_B^{-1}\mathbf{b} \geq \mathbf{0}$. L'intérêt de la notion de base est le suivant. Soit B une base. Définissons le vecteur \mathbf{y} de \mathbb{R}^n :

$$\mathbf{y}_B = A_B^{-1}\mathbf{b} \quad \text{et} \quad \mathbf{y}_N = \mathbf{0},$$

où $N = [n] \setminus B$. Alors \mathbf{y} satisfait automatiquement les contraintes d'égalité du problème (PL) car on a $A\mathbf{y} = \mathbf{b}$. Le vecteur \mathbf{y} est alors une *solution basique* (associée à B). Si B est une base réalisable, on a de plus $\mathbf{y} \geq \mathbf{0}$. Un tel \mathbf{y} est alors réalisable et constitue une *solution basique réalisable* (associée à B).

4.3.2 Condition d'existence

Si le problème (PL) a au moins une solution réalisable, il possède nécessairement une base réalisable. Cette remarque, pas complètement triviale, peut être utile. On a cependant bien plus, comme le montre le théorème suivant (et dont la preuve permet aussi de montrer la remarque précédente).

Théorème 4.1. *Si le problème (PL) a au moins une solution réalisable et si la fonction objectif est bornée inférieurement sur l'ensemble des solutions réalisables, alors il existe au moins une solution optimale.*

S'il existe une solution optimale, alors il en existe au moins une qui soit basique réalisable.

La base associée à une solution basique réalisable optimale est appelée *base optimale*.

Le théorème 4.1 est spécifique à l'optimisation linéaire. Le premier point déjà n'est pas satisfait en général pour l'optimisation convexe. Par exemple, si l'on souhaite minimiser y sur l'ensemble $\{(x, y) \in (\mathbb{R}_+ \setminus \{0\})^2 : y \geq 1/x\}$, on se trouve confronté à un problème d'optimisation convexe réalisable, avec une fonction objectif bornée inférieurement par 0, et pourtant ce problème n'a pas de solution optimale finie. Ce théorème n'est pas non plus une conséquence du théorème 2.1 car $\mathbf{x} \mapsto \mathbf{c} \cdot \mathbf{x}$ n'est pas coercive.

Démonstration du théorème 4.1. Supposons donc que le problème (PL) a au moins une solution réalisable et que sa fonction objectif est bornée inférieurement sur l'ensemble des solutions réalisables. Nous allons prouver que si \mathbf{x} est une solution réalisable de (PL), alors il existe une solution basique réalisable $\tilde{\mathbf{x}}$ telle que $\mathbf{c} \cdot \tilde{\mathbf{x}} \leq \mathbf{c} \cdot \mathbf{x}$. Comme le nombre de bases réalisables est fini, cela implique le théorème dans sa totalité.

Soit donc \mathbf{x} une solution réalisable arbitraire et soit $\tilde{\mathbf{x}}$ une solution réalisable avec le plus de zéros possible telle que $\mathbf{c} \cdot \tilde{\mathbf{x}} \leq \mathbf{c} \cdot \mathbf{x}$. Notons K l'ensemble des indices i tels que $\tilde{x}_i > 0$. Si les colonnes de A_K sont linéairement indépendantes, alors $\tilde{\mathbf{x}}$ est une solution basique réalisable : il suffit de choisir $B \supseteq K$ tel que A_B soit inversible (ce qui existe par un argument élémentaire d'algèbre linéaire), et $\tilde{\mathbf{x}}$ sera bien une solution basique réalisable associée à B .

Supposons donc que A_K a des colonnes dépendantes. Nous allons montrer que l'on aboutit alors à une contradiction. Il existe $\mathbf{y} \in \mathbb{R}^K$ tel que $A_K \mathbf{y} = \mathbf{0}$ et $\mathbf{y} \neq \mathbf{0}$. Définissons alors $\mathbf{z} \in \mathbb{R}^n$ par $z_i = y_i$ si $i \in K$ et $z_i = 0$ sinon. Le vecteur $\tilde{\mathbf{x}} + t\mathbf{z}$ est réalisable, i.e. $A(\tilde{\mathbf{x}} + t\mathbf{z}) = \mathbf{b}$ et $\tilde{\mathbf{x}} + t\mathbf{z} \geq \mathbf{0}$, précisément pour t dans l'intervalle $J = [\max_{i \in I_+} (-\tilde{x}_i/z_i), \min_{i \in I_-} (-\tilde{x}_i/z_i)]$, où

$$I_+ = \{i : z_i > 0\} \quad \text{et} \quad I_- = \{i : z_i < 0\}.$$

Si I_+ (resp. I_-) est vide, la borne inférieure (resp. supérieure) de J devient comme attendu $-\infty$ (resp. $+\infty$).

Si $\mathbf{c} \cdot \mathbf{z}$ est différent de zéro, on considère la borne de J qui est du signe opposé à $\mathbf{c} \cdot \mathbf{z}$. Cette borne est finie, sinon on pourrait faire tendre la fonction objectif vers $-\infty$ le long du rayon $\tilde{\mathbf{x}} + t\mathbf{z}$. Si $\mathbf{c} \cdot \mathbf{z}$ est égal à zéro, on considère arbitrairement une borne de J finie (il en existe forcément une car I_+ et I_- ne peuvent être tous deux vides). Lorsque t atteint cette borne, $\tilde{\mathbf{x}} + t\mathbf{z}$ a une composante nulle de plus que $\tilde{\mathbf{x}}$ tout en fournissant une solution au moins aussi bonne. Cela contredit la minimalité de son support. A_K a donc nécessairement ses colonnes linéairement indépendantes. \square

Le théorème 4.1 permet de voir qu'il existe, du moins en principe, un algorithme qui calcule en temps fini une solution optimale exacte d'un problème d'optimisation linéaire lorsqu'une telle solution existe. En effet, le nombre de bases est borné par $\binom{n}{m}$, qui est une quantité finie. On pourrait donc passer en revue tous les sous-ensembles B de cardinalité m de $[n]$, tester pour chacun de ces B si A_B est inversible et si oui, si l'on a bien $A_B^{-1} \mathbf{b} \geq \mathbf{0}$, et

enfin, garder, parmi ces B , celui qui maximise la quantité $\mathbf{c}_B \cdot (A_B^{-1} \mathbf{b})$. Cette approche n'est cependant pas envisageable en pratique, excepté peut-être sur de petits problèmes, car $\binom{n}{m}$ est en général très grand (par exemple, de l'ordre de 4^m , si $m = n/2$).

4.4 Algorithme du simplexe

4.4.1 Préliminaires

Il existe un algorithme particulièrement efficace permettant la résolution des problèmes d'optimisation linéaire : l'algorithme du simplexe. Il a été proposé par Dantzig en 1947, mais publié seulement quelques années plus tard [4]. Du fait de l'importance de l'optimisation linéaire, cet algorithme est l'un des plus utilisés dans l'industrie, toutes catégories d'applications confondues. Il existe de nombreux logiciels – gratuits ou payants – dans lesquels il est implémenté de manière très efficace. Un ingénieur ou un chercheur devant résoudre un problème d'optimisation linéaire n'aura donc pas à le reprogrammer. Signalons aussi que les méthodes de points intérieurs, déjà signalées, disposent de même de versions efficaces déjà programmées.

Il est donc inutile d'entrer dans les détails techniques de la description de l'algorithme du simplexe et nous allons simplement donner les principes sur lesquels il s'appuie. La principale propriété sous-tendant l'algorithme du simplexe est la proposition suivante.

Étant donné une base B et l'ensemble $N = [n] \setminus B$, on pose $\mathbf{r} = \mathbf{c}_N - A_N^T (A_B^{-1})^T \mathbf{c}_B$. Ce vecteur à $n - m$ composantes est le vecteur des *coûts réduits* relatifs à B .

Proposition 4.2. *Soit B est une base réalisable. Si $\mathbf{r} \geq \mathbf{0}$, alors B est une base optimale. Si k est tel que $r_k < 0$, alors une seule des deux possibilités suivantes se réalise :*

1. *Il existe une base réalisable $B' \subseteq B \cup \{k\}$, avec $B' \neq B$. Dans ce cas, la solution basique associée à B' est au moins aussi bonne que celle associée à B .*
2. *Il n'existe pas de base réalisable $B' \subseteq B \cup \{k\}$, avec $B' \neq B$. Dans ce cas, en notant $\tilde{\mathbf{x}}$ la solution basique associée à B , il existe $\mathbf{z} \in \mathbb{R}^n$ avec $z_k > 0$ tel que l'on a $\mathbf{c} \cdot \mathbf{z} = r_k z_k$ et $\mathbf{x}(t) = \tilde{\mathbf{x}} + t\mathbf{z}$ réalisable pour tout $t \in \mathbb{R}_+$. En particulier, la valeur de (PL) est $-\infty$.*

Dans le cas 2, l'équation $\mathbf{x}(t)$, pour $t \in \mathbb{R}_+$, décrit un rayon infini le long duquel le critère peut être rendu arbitrairement petit.

Démonstration de la proposition 4.2. Soit $\tilde{\mathbf{x}}$ la solution basique associée à B . Prenons \mathbf{x} réalisable quelconque. On a $\mathbf{c} \cdot \mathbf{x} = \mathbf{c}_B \cdot \mathbf{x}_B + \mathbf{c}_N \cdot \mathbf{x}_N$. Comme $\mathbf{x}_B = A_B^{-1} \mathbf{b} - A_B^{-1} A_N \mathbf{x}_N$ et $\mathbf{c} \cdot \tilde{\mathbf{x}} = \mathbf{c}_B^T A_B^{-1} \mathbf{b}$, on a $\mathbf{c} \cdot \mathbf{x} = \mathbf{c} \cdot \tilde{\mathbf{x}} + \mathbf{r} \cdot \mathbf{x}_N$. Donc si $\mathbf{r} \geq \mathbf{0}$, on a $\mathbf{c} \cdot \mathbf{x} \geq \mathbf{c} \cdot \tilde{\mathbf{x}}$, ce qui montre alors que $\tilde{\mathbf{x}}$ est solution optimale.

Supposons maintenant qu'il existe $k \in N$ tel que $r_k < 0$. Considérons la matrice $A_{B \cup \{k\}}$. Comme $k \notin B$, les colonnes de cette matrice sont linéairement dépendantes. Il existe donc $\mathbf{z} \in \mathbb{R}^n$ dont le support est précisément $B \cup \{k\}$ et tel que $A\mathbf{z} = \mathbf{0}$ et $\mathbf{z} \neq \mathbf{0}$. Quitte à prendre l'opposé de \mathbf{z} si nécessaire, on peut supposer que z_k est strictement positif. Le vecteur $\tilde{\mathbf{x}} + t\mathbf{z}$ est réalisable précisément pour $t \in [0, \min_{j \in J} (-\tilde{x}_j/z_j)]$, où $J = \{j \in B : z_j < 0\}$. Notons que $\mathbf{c} \cdot \mathbf{z} = r_k z_k$ (cela se voit en utilisant $\mathbf{z}_B = A_B^{-1} A_N \mathbf{z}_N$). Donc $\mathbf{c} \cdot \mathbf{z} < 0$.

Si J est non vide, $B' = B \cup \{k\} \setminus \{j^*\}$, où $j^* = \arg \min_{j \in J} (-\tilde{x}_j/z_j)$, est une base réalisable, car $z_{j^*} \neq 0$ (la colonne A_{j^*} peut s'écrire comme combinaison linéaire des colonnes de $A_{B'}$). La solution basique réalisable

associée à B' est $\tilde{\mathbf{x}} - \frac{\tilde{x}_{j^*}}{z_{j^*}} \mathbf{z}$. Comme $\mathbf{c} \cdot \mathbf{z} < 0$, cette nouvelle solution basique réalisable est au moins aussi bonne que $\tilde{\mathbf{x}}$. On a donc prouvé le cas 1.

Si J est vide, le cas 2 se produit : le vecteur $\tilde{\mathbf{x}} + t\mathbf{z}$ est réalisable précisément pour tout $t \in [0, +\infty[$, et l'on a bien $\mathbf{c} \cdot \mathbf{z} < 0$. \square

4.4.2 Description de l'algorithme

L'algorithme de simplexe s'écrit alors, en supposant que l'on dispose d'une première base réalisable B_0 :

```

Données : Une base réalisable  $B_0$ 
Sortie : La valeur optimale et si elle est finie, une base optimale

 $B \leftarrow B_0$ ;
 $N \leftarrow [n] \setminus B$ ;
 $Unbounded \leftarrow \text{false}$ ;
 $\mathbf{r} \leftarrow \mathbf{c}_N - A_N(A_B^{-1})^T \mathbf{c}_B$ ;
while  $Unbounded = \text{false}$  et  $\mathbf{r} \not\geq \mathbf{0}$  do
    Choisir  $k$  tel que  $r_k < 0$ ;
    if Il existe une base réalisable  $B' \subseteq B \cup \{k\}$ , avec  $B' \neq B$  then
         $B \leftarrow B'$ ;
         $N \leftarrow [n] \setminus B$ ;
         $\mathbf{r} \leftarrow \mathbf{c}_N - A_N^T(A_B^{-1})^T \mathbf{c}_B$ ;
    else
         $Unbounded = \text{true}$ ;
    end
end
if  $Unbounded \leftarrow \text{true}$  then
    return  $-\infty$ 
else
    return  $(\mathbf{c}_B^T A_B^{-1} \mathbf{b}, B)$ 
end

```

Algorithme 3 : Algorithme du simplexe

En général, il existe plusieurs k tels que $r_k < 0$. La façon de le choisir s'appelle la *règle de pivot* de l'algorithme. Les performances de l'algorithme sont fortement liées au choix de cette règle. La mise à jour de \mathbf{r} à chaque itération peut se faire aisément et rapidement par un simple pivot de Gauss, connaissant la valeur de \mathbf{r} à l'itération précédente. Il n'est pas nécessaire de procéder à chaque fois à une inversion de la matrice A_B , heureusement. Nous ne détaillons pas les calculs ici.

Avant d'expliquer comment trouver une première base réalisable, il faut justifier que cet algorithme termine. En fait, la proposition 4.2 ne permet pas de l'assurer. En effet, pour que cela fût le cas, il faudrait que dans le cas 1, la solution basique associée à B' soit *strictement* meilleure que celle associée à B , ce qui n'est pas nécessairement le cas. Il peut alors arriver que l'algorithme se mette à *cycler* sur une suite de bases réalisables donnant toute la même

valeur au critère. Le théorème suivant, établi en 1955 par Dantzig, Orden et Wolfe [5], indique qu'il existe des règles de pivot qui permettent à l'algorithme de ne pas cycler. On omet la preuve car elle est difficile et dépasse le cadre de ce cours.

Théorème 4.3. *Il existe des règles de pivot pour l'algorithme du simplexe qui assure que ce dernier se termine toujours en un nombre fini d'itérations.*

En particulier, si la valeur optimale est finie, l'algorithme parvient à trouver une base réalisable B telle que le vecteur des coûts réduits qui lui est associé a toutes ses composantes positives, qui est bien une base optimale d'après la proposition 4.2.

Il est également intéressant de noter que dans le cas de l'optimisation linéaire on est toujours capable de calculer la solution optimale exacte, ce qui diffère de ce que l'on a pu voir au Chapitre 3 pour l'optimisation convexe en général. En effet, il existe toujours une solution optimale entièrement déterminée par un ensemble de cardinalité finie : la base optimale.

4.4.3 Trouver une première base réalisable

Pour trouver une base réalisable au problème (PL), une façon standard de procéder consiste à considérer le problème linéaire suivant :

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^m z_i \\ \text{s.c.} \quad & A\mathbf{x} + \mathbf{z} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \\ & \mathbf{z} \geq \mathbf{0}, \end{aligned} \tag{PL'}$$

en supposant sans perte de généralité que $\mathbf{b} \geq \mathbf{0}$. Il est aisé de vérifier que (PL') a pour valeur optimale 0 si et seulement si (PL) admet une solution réalisable. Et si (PL) admet une solution réalisable, alors toute base optimale de (PL') donne une base réalisable de (PL). Comme on dispose d'une base réalisable évidente pour (PL') – celle formée par les variables z_i – on peut appliquer l'algorithme du simplexe sur ce problème pour tester si (PL) est réalisable et s'il l'est, pour en déterminer une première base réalisable.

4.5 Interprétation géométrique

L'optimisation linéaire a une interprétation géométrique assez agréable. On peut d'emblée noter que l'ensemble des solutions réalisables d'un problème linéaire forme un polyèdre. Rappelons qu'un *polyèdre* de \mathbb{R}^n est l'intersection d'un nombre fini de demi-espaces de \mathbb{R}^n . En fait, l'optimisation linéaire coïncide précisément avec l'optimisation d'un critère linéaire sur un polyèdre. Un exemple de polyèdre est donné sur le figure 4.1.

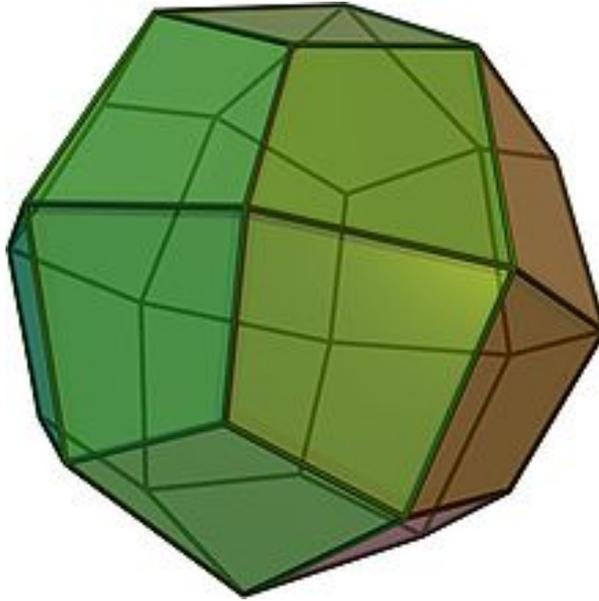


FIGURE 4.1 – Un polyèdre de dimension 3. (Source : [Wikimedia Commons](#), [Creative Commons Attribution-Share Alike 3.0 Unported](#))

On a en fait bien plus, comme le prouvent les deux résultats suivants, dont les preuves sont laissées en exercice. Un point \mathbf{x} d'un polyèdre P est un *sommet* s'il existe un hyperplan H tel que P soit entièrement contenu dans un des deux demi-espaces fermés délimités par H et tel que \mathbf{x} soit l'unique élément de $H \cap P$.

***Théorème 4.4.** *Considérons un problème linéaire écrit sous forme standard. Alors \mathbf{x} est une solution basique réalisable si et seulement si \mathbf{x} est un sommet du polyèdre formé par l'ensemble des solutions réalisables.*

Un segment $[\mathbf{x}, \mathbf{y}]$ est une *arête* s'il existe un hyperplan H tel que P soit entièrement contenu dans un des deux demi-espaces fermés délimités par H et tel que $[\mathbf{x}, \mathbf{y}] = H \cap P$. Deux sommets d'un polyèdre sont adjacents s'ils sont reliés par une arête.

***Proposition 4.5.** *Si les deux solutions basiques réalisables associées à deux bases visitées successivement par l'algorithme du simplexe sont distinctes, alors elles forment deux sommets adjacents du polyèdre des solutions réalisables.*

L'algorithme du simplexe parcourt donc une partie des sommets du polyèdre des solutions réalisables en passant d'un sommet à l'autre à l'aide des arêtes.

Compléments sur l'optimisation linéaire

5.1 Dualité en optimisation linéaire

5.1.1 Dualité forte

Dans le cas de l'optimisation linéaire, le principal fait à retenir concernant la dualité est que l'on a toujours dualité forte. Reprenons le problème (PL),

$$\begin{aligned} \text{Min} \quad & \mathbf{c} \cdot \mathbf{x} \\ \text{s.c.} \quad & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{PL}$$

Son dual est défini comme étant (voir Section 2.3)

$$\begin{aligned} \text{Max} \quad & \inf_{\mathbf{x} \in \mathbb{R}^n} \mathbf{c} \cdot \mathbf{x} + \boldsymbol{\lambda} \cdot (\mathbf{b} - A\mathbf{x}) - \boldsymbol{\mu} \cdot \mathbf{x} \\ \text{s.c.} \quad & \boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}_+^n, \end{aligned}$$

qui se réécrit

$$\begin{aligned} \text{Max} \quad & \mathbf{b} \cdot \boldsymbol{\lambda} + \inf_{\mathbf{x} \in \mathbb{R}^n} (\mathbf{c} - A^T \boldsymbol{\lambda} - \boldsymbol{\mu}) \cdot \mathbf{x} \\ \text{s.c.} \quad & \boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}_+^n. \end{aligned}$$

Comme

$$\inf_{\mathbf{x} \in \mathbb{R}^n} (\mathbf{c} - A^T \boldsymbol{\lambda} - \boldsymbol{\mu}) \cdot \mathbf{x} = \begin{cases} -\infty & \text{si } \mathbf{c} - A^T \boldsymbol{\lambda} - \boldsymbol{\mu} \neq \mathbf{0} \\ 0 & \text{sinon,} \end{cases}$$

on obtient que le dual de (PL) s'écrit

$$\begin{aligned} \text{Max} \quad & \mathbf{b} \cdot \boldsymbol{\lambda} \\ \text{s.c.} \quad & \mathbf{c} - A^T \boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0} \\ & \boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}_+^n, \end{aligned}$$

qui peut se réécrire

$$\begin{aligned} \text{Max} \quad & \mathbf{b} \cdot \mathbf{y} \\ \text{s.c.} \quad & A^T \mathbf{y} \leq \mathbf{c}. \end{aligned} \tag{DL}$$

En refaisant l'exercice d'écrire le dual de (DL), on peut d'ailleurs noter que l'on obtient un problème linéaire équivalent au primal de départ (PL). C'est une propriété partagée par tout problème linéaire, quelle que soit la forme sous laquelle il est écrit : le dual du dual est équivalent au primal.

On a le théorème suivant.

Théorème 5.1. *En optimisation linéaire, si au moins l'un des deux problèmes, du primal ou du dual, est réalisable, alors on a dualité forte.*

Ce théorème signifie que si le problème primal et le problème dual sont tous deux réalisables, alors la valeur optimale du primal est finie et égale à la valeur optimale du dual. Il signifie aussi que si l'un des deux problèmes n'est pas réalisable, alors l'autre est soit non borné, soit non réalisable. Par exemple, si le dual n'est pas réalisable mais que le primal l'est, alors la valeur optimale du primal est nécessairement $-\infty$.

La preuve de ce théorème est la conséquence du lemme de Farkas (lemme 5.2 ci-dessous). Noter que ce théorème peut être vu comme un cas particulier du corollaire 3.12, au moins lorsque le primal et le dual admettent tous deux des solutions réalisables, mais comme ce dernier est prouvé entre autre à l'aide du théorème 2.4, qui lui-même est une conséquence du lemme de Farkas, il est intéressant de voir que l'on peut procéder à la preuve directement à partir du lemme de Farkas. En exercice, on verra que le théorème 5.1 peut également être vu comme une conséquence de la terminaison de l'algorithme du simplexe, donc une conséquence du théorème 4.3.

5.1.2 Lemme de Farkas et preuve de la dualité forte

L'un des plus anciens résultats d'optimisation linéaire est le lemme suivant, dû à Farkas [6]. Une illustration en est donnée sur la figure 5.1.

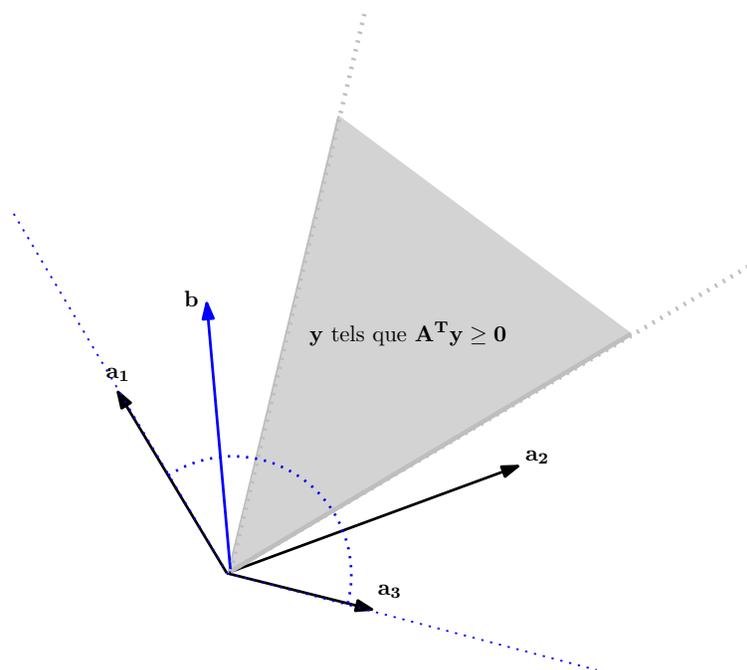


FIGURE 5.1 – Illustration du lemme de Farkas (lemme 5.2) : dans la zone bleue, tout vecteur s'écrit comme combinaison linéaire à coefficients positifs des vecteurs \mathbf{a}_j formant les colonnes de A ; la zone grisée correspond aux vecteurs \mathbf{y} tels que $A^T \mathbf{y} \geq 0$.

***Lemme 5.2** (Lemme de Farkas). *Soit A une matrice réelle et soit \mathbf{b} un vecteur. Il existe $\mathbf{x} \geq \mathbf{0}$ vérifiant $A\mathbf{x} = \mathbf{b}$ si et seulement si $\mathbf{y} \cdot \mathbf{b} \geq 0$ pour tout \mathbf{y} tel que $\mathbf{y}^T A \geq \mathbf{0}^T$.*

Démonstration. Notons m le nombre de lignes de A et n son nombre de colonnes. S'il existe $\mathbf{x} \in \mathbb{R}_+^n$ vérifiant $A\mathbf{x} = \mathbf{b}$, alors pour tout $\mathbf{y} \in \mathbb{R}^m$ tel que $\mathbf{y}^T A \geq \mathbf{0}$, on a $\mathbf{y} \cdot \mathbf{b} = \mathbf{y}^T A\mathbf{x}$ et donc $\mathbf{y} \cdot \mathbf{b} \geq 0$.

Réciproquement, montrons par récurrence sur n que si A est une matrice à n colonnes est si $\mathbf{y} \cdot \mathbf{b} \geq 0$ pour tout \mathbf{y} tel que $\mathbf{y}^T A \geq \mathbf{0}^T$, alors il existe $\mathbf{x} \in \mathbb{R}_+^n$ vérifiant $A\mathbf{x} = \mathbf{b}$.

Considérons le cas $n = 1$. La matrice A est alors un vecteur \mathbf{a} tel que pour tout vecteur \mathbf{y} , si $\mathbf{y} \cdot \mathbf{a} \geq 0$, alors $\mathbf{y} \cdot \mathbf{b} \geq 0$. Si $\mathbf{a} = \mathbf{0}$, alors cela implique que $\mathbf{b} = \mathbf{0}$ et l'on a le résultat. Si $\mathbf{a} \neq \mathbf{0}$, il est alors aisé de voir que cela implique que \mathbf{a} et \mathbf{b} sont normaux à un même hyperplan. \mathbf{a} et \mathbf{b} sont donc colinéaires, ce qui implique à nouveau le résultat.

Considérons maintenant le cas $n > 1$. Notons \mathbf{a}^i la i ème colonne de A . S'il existe x_1, \dots, x_{n-1} de \mathbb{R}_+ tels que $\sum_{i=1}^{n-1} x_i \mathbf{a}^i = \mathbf{b}$, on a l'égalité souhaitée en posant $\mathbf{x} = (x_1, \dots, x_{n-1}, 0)$. On peut donc supposer que pour tout $(n-1)$ -uplet (x_1, \dots, x_{n-1}) de réels positifs, on a $\sum_{i=1}^{n-1} x_i \mathbf{a}^i \neq \mathbf{b}$. D'après l'hypothèse de récurrence, il existe un vecteur $\mathbf{v} \in \mathbb{R}^m$ tel que $\mathbf{v} \cdot \mathbf{a}^i \geq 0$ pour tout $i = 1, \dots, n-1$ et tel que $\mathbf{v} \cdot \mathbf{b} < 0$. On a donc nécessairement $\mathbf{v} \cdot \mathbf{a}^n < 0$. Considérons l'hyperplan $H = \{\mathbf{z} \in \mathbb{R}^m : \mathbf{v} \cdot \mathbf{z} = 0\}$. Définissons \mathbf{b}' le projeté de \mathbf{b} sur H et pour tout $i \in [n-1]$ le projeté \mathbf{a}'^i de \mathbf{a}^i sur H selon la direction \mathbf{a}^n .

Supposons d'abord qu'il n'existe pas de $(n-1)$ -uplet $(x'_1, \dots, x'_{n-1}) \in \mathbb{R}_+^{n-1}$ tel que $\sum_{i=1}^{n-1} x'_i \mathbf{a}'^i = \mathbf{b}'$. D'après l'hypothèse de récurrence, cela implique qu'il existe alors un vecteur $\mathbf{v}' \in H$ tel que $\mathbf{v}' \cdot \mathbf{a}'^i \geq 0$ tandis que $\mathbf{v}' \cdot \mathbf{b}' < 0$. Notons

$$\mu = \frac{\mathbf{v}' \cdot \mathbf{a}^n}{\mathbf{v}' \cdot \mathbf{a}^n}$$

et posons $\mathbf{y} = \mathbf{v}' - \mu \mathbf{v}$. On a alors $\mathbf{y}^T A \geq \mathbf{0}$ et $\mathbf{y} \cdot \mathbf{b} < 0$, ce qui est en contradiction avec la supposition sur A et \mathbf{b} .

On a donc un $(n-1)$ -uplet $(x'_1, \dots, x'_{n-1}) \in \mathbb{R}_+^{n-1}$ tel que $\sum_{i=1}^{n-1} x'_i \mathbf{a}'^i = \mathbf{b}'$. Posons

$$\lambda_i = \frac{\mathbf{v} \cdot \mathbf{a}^i}{\mathbf{v} \cdot \mathbf{a}^n}$$

pour $i = 1, \dots, n-1$. Soit alors \mathbf{x} défini par $x_i = x'_i$ si $i \in [n-1]$ et par $x_n = \frac{\mathbf{v} \cdot \mathbf{b}}{\mathbf{v} \cdot \mathbf{a}^n} - \sum_{i=1}^{n-1} \lambda_i x'_i$. Comme $\mathbf{v} \cdot \mathbf{a}^i \geq 0$ pour $i \in [n-1]$ et $\mathbf{v} \cdot \mathbf{a}^n < 0$, on a $\lambda_i < 0$ pour tout $i \in [n-1]$. Par conséquent, un tel \mathbf{x} a toutes ses composantes positives et satisfait $A\mathbf{x} = \mathbf{b}$. \square

Démonstration du théorème 5.1. Si le primal et le dual sont non réalisables, il n'y a rien à prouver.

Supposons le primal réalisable de valeur optimale finie, avec comme solution optimale \mathbf{x}^* . On a alors $(\mathbf{x}^*)^T M \geq \mathbf{r}^T$, avec

$$M = (A^T, -A^T, I) \quad \text{and} \quad \mathbf{r}^T = (\mathbf{b}^T, -\mathbf{b}^T, \mathbf{0}^T).$$

Notons J l'ensemble des j tels que $(\mathbf{x}^*)^T M_j = \mathbf{r}_j$. Si \mathbf{z} est tel que $\mathbf{z}^T M_j \geq \mathbf{0}$, on a $\mathbf{x}^* + t\mathbf{z}$ solution réalisable du primal pour certains t positifs, et donc par optimalité de \mathbf{x}^* , on a $\mathbf{c} \cdot \mathbf{z} \geq 0$. D'après le lemme de Farkas (lemme 5.2) appliqué à la matrice M_j et au vecteur \mathbf{c} , on sait qu'il existe alors $\boldsymbol{\mu}' \in \mathbb{R}_+^J$ tel que $M_j \boldsymbol{\mu}' = \mathbf{c}$. Définissons alors μ_j comme étant égal à μ'_j pour $j \in J$ et à 0 sinon. On a alors $M\boldsymbol{\mu} = \mathbf{c}$ avec $\boldsymbol{\mu} \geq \mathbf{0}$. En écrivant le vecteur $\boldsymbol{\mu}$ sous la forme $(\mathbf{y}_1, \mathbf{y}_2, \mathbf{s})$, on a donc $A^T \mathbf{y}_1 - A^T \mathbf{y}_2 + \mathbf{s} = \mathbf{c}$, avec les vecteurs $\mathbf{y}_1, \mathbf{y}_2$ et \mathbf{s} tous trois à composantes positives. Le vecteur $\mathbf{y} = \mathbf{y}_1 - \mathbf{y}_2$ satisfait donc $A^T \mathbf{y} \leq \mathbf{c}$. C'est une solution réalisable du dual. On a $\mathbf{b} \cdot \mathbf{y} = \mathbf{r} \cdot \boldsymbol{\mu} = \mathbf{r}_J \cdot \boldsymbol{\mu}_J = (\mathbf{x}^*)^T M_J \boldsymbol{\mu}_J = (\mathbf{x}^*)^T \mathbf{c}$. C'est donc une solution réalisable donnant une valeur au critère du dual égale à la valeur optimale du primal. Par dualité faible, on voit donc que \mathbf{y} est solution optimale du dual. Primal et dual ont donc la même valeur optimale.

Supposons le dual réalisable de valeur optimale finie, avec comme solution optimale \mathbf{y}^* . On a $(\mathbf{y}^*)^T A \leq \mathbf{c}$. Notons I les lignes de A pour lesquelles on a égalité. De manière semblable à ce qui vient d'être fait, on voit que si $\mathbf{z}^T A_I \geq \mathbf{0}$, alors par optimalité de \mathbf{y}^* , on doit avoir $\mathbf{b} \cdot \mathbf{z} \geq 0$. D'après le lemme de Farkas, il existe donc $\mathbf{x}' \geq \mathbf{0}$ tel que $A_I \mathbf{x}' = \mathbf{b}$. Complétant \mathbf{x}' avec des 0, on obtient donc une solution réalisable \mathbf{x} du primal tel que $x_i \neq 0$ si $i \notin I$. On a $\mathbf{c} \cdot \mathbf{x} = \mathbf{c}_I \cdot \mathbf{x}_I = (\mathbf{y}^*)^T A_I \mathbf{x}_I = \mathbf{y}^* \cdot \mathbf{b}$. C'est donc une solution réalisable donnant

une valeur au critère du primal égale à la valeur optimale du dual. Par dualité faible, on voit donc que \mathbf{x} est solution optimale du primal. Primal et dual ont donc la même valeur optimale.

Enfin, pour les deux cas restant, la dualité faible permet de conclure directement : si la valeur optimale du primal est $-\infty$, le dual ne peut avoir de solution réalisable, et si la valeur optimale du dual est $+\infty$, le primal ne peut avoir de solution réalisable. \square

Application 5.1 (Équilibres de Nash dans les jeux à somme nulle). Soit A une matrice réelle à m lignes et n colonnes. Notons $\Delta^k = \{\mathbf{u} \in \mathbb{R}_+^k : \sum_{i=1}^k u_i = 1\}$. On peut voir cet ensemble comme étant celui des distributions de probabilité sur $[k]$. Il existe toujours $\mathbf{x} \in \Delta^m$ et $\mathbf{y} \in \Delta^n$ tels que

$$\mathbf{u}^T A \mathbf{y} \leq \mathbf{x}^T A \mathbf{y} \leq \mathbf{x}^T A \mathbf{v} \quad \text{pour tout } \mathbf{u} \in \Delta^m \text{ et tout } \mathbf{v} \in \Delta^n. \quad (5.1)$$

C'est un théorème de von Neumann [12], dont nous allons donner la preuve à partir du théorème 5.1. Le couple (\mathbf{x}, \mathbf{y}) peut s'interpréter comme un équilibre de Nash en stratégies mixtes dans le jeu à somme nulle dont la matrice des paiements est donnée par A , voir l'encadré suivant.

Quitte à ajouter un même réel strictement positif à toutes les entrées de la matrice A , on peut supposer sans perte de généralité que A a toutes ses entrées strictement positives. Considérons les deux problèmes linéaires duaux l'un de l'autre

$$\begin{array}{ll} \text{Min} & \sum_{j=1}^n v_j \\ \text{s.c.} & A \mathbf{v} \geq \mathbf{1} \\ & \mathbf{v} \geq \mathbf{0} \end{array} \quad \text{et} \quad \begin{array}{ll} \text{Max} & \sum_{i=1}^m u_i \\ \text{s.c.} & A^T \mathbf{u} \leq \mathbf{1} \\ & \mathbf{u} \geq \mathbf{0}. \end{array}$$

Le problème en \mathbf{u} est réalisable ($\mathbf{u} = \mathbf{0}$ est solution) et borné car A a toutes ses entrées strictement positives. Il admet donc une solution optimale \mathbf{u}^* d'après le théorème 4.1. Le théorème 5.1 assure alors l'existence d'une solution optimale \mathbf{v}^* au problème en \mathbf{v} . Posons $t = \sum_{i=1}^m u_i^* = \sum_{j=1}^n v_j^*$, valeur optimale de ces deux problèmes linéaires. On montre aisément que $t > 0$. Posons alors

$$\mathbf{x} = \frac{1}{t} \mathbf{u}^* \quad \text{et} \quad \mathbf{y} = \frac{1}{t} \mathbf{v}^*.$$

On a alors bien $\mathbf{x} \in \Delta^m$ et $\mathbf{y} \in \Delta^n$. On a aussi $\mathbf{u}^T A \mathbf{y} = \frac{1}{t} \mathbf{u}^T A \mathbf{v}^* \geq \frac{1}{t}$ pour tout $\mathbf{u} \in \Delta^m$ et $\mathbf{x}^T A \mathbf{v} = \frac{1}{t} \mathbf{u}^{*T} A \mathbf{v} \leq \frac{1}{t}$ pour tout $\mathbf{v} \in \Delta^n$. La conclusion suit immédiatement.

Une conséquence intéressante de cette preuve est que l'algorithme du simplexe permet de calculer une solution de (5.1), i.e. un équilibre de Nash en stratégies mixtes dans les jeux à somme nulle.

Les jeux à somme nulle

Soit $A = (a_{\ell c})$ une matrice réelle à m lignes et n colonnes. Considérons deux joueurs, “ligne” et “colonne”, jouant au jeu suivant. “ligne” choisit une ligne ℓ de A et “colonne” choisit simultanément une colonne c de A . “ligne” gagne alors $a_{\ell c}$ euros et “colonne” gagne $-a_{\ell c}$ euros. C’est un *jeu à somme nulle*, un joueur gagnant exactement ce que l’autre perd.

Un équilibre de Nash en *stratégies pures* est choix d’une ligne ℓ^* pour “ligne” et d’une colonne c^* pour “colonne” qui soit pour chacun d’eux le meilleur choix possible, i.e.

$$a_{\ell c^*} \leq a_{\ell^* c^*} \leq a_{\ell^* c} \quad \text{pour tout } \ell \in [m] \text{ et tout } c \in [n].$$

Un tel équilibre en stratégies pures n’existe pas forcément. En revanche, il existe forcément un équilibre de Nash en *stratégies mixtes*, qui est le choix d’une distribution de probabilité sur les lignes pour “ligne” et sur les colonnes pour “colonne” conduisant au meilleur choix pour chacun d’eux relativement à l’espérance de gain. Mathématiquement, un équilibre de Nash en stratégies mixtes est un couple (\mathbf{x}, \mathbf{y}) satisfaisant la propriété (5.1).

L’étude de ces jeux, et plus généralement la théorie des jeux dont les fondements remontent aux années 1940, permet de comprendre et d’analyser les situations d’interaction stratégique entre des agents. L’un des pères fondateurs de cette discipline est John Nash, qui a en particulier donné son nom aux équilibres mentionnés ci-dessus.

5.2 Matrices totalement unimodulaires

Une matrice A est *totalement unimodulaire* si chacune de ses sous-matrices carrées est de déterminant $+1$, -1 ou 0 . En particulier, c’est une matrice dont tous les coefficients sont dans $\{+1, -1, 0\}$. Le théorème suivant a de nombreuses applications en optimisation combinatoire¹.

Théorème 5.3. *Considérons le problème (PL). Supposons A totalement unimodulaire et \mathbf{b} à coefficients entiers.*

Si (PL) admet au moins une solution réalisable, il admet une solution réalisable ayant toutes ses coordonnées entières.

Si (PL) admet au moins une solution optimale, il admet une solution optimale ayant toutes ses coordonnées entières.

Démonstration. Nous allons montrer que toute solution basique réalisable est à coordonnées entières. Ce sera suffisant pour conclure car l’existence d’une solution réalisable implique l’existence d’une solution basique réalisable (voir le début de la section 4.3.2), et l’existence d’une solution optimale implique l’existence d’une solution basique optimale (théorème 4.1).

Considérons donc une solution basique réalisable et notons B la base à laquelle cette solution basique réalisable est associée. On a donc $\tilde{\mathbf{x}}_B = A_B^{-1}\mathbf{b}$. En écrivant $A_B^{-1} = \frac{1}{\det A_B} \text{com}(A_B)^T$, on voit que la totale unimodularité de A implique que A_B^{-1} est à coefficients dans $\{+1, -1, 0\}$. La conclusion suit immédiatement. \square

1. L’optimisation combinatoire est la branche de l’optimisation qui s’intéresse au problème (P) du Chapitre 1 quand X est fini, ou du moins dénombrable. L’exemple 1.4 est un exemple typique de problème d’optimisation combinatoire.

Sous les conditions du théorème, la preuve précédente montre que toute solution basique réalisable est à coordonnées entières et donc que la solution optimale retournée par l'algorithme du simplexe est à coordonnées entières. On dispose donc d'une méthode pratique pour calculer une telle solution.

Il existe des algorithmes efficaces pour tester si une matrice A est totalement unimodulaire, mais ils dépassent largement le cadre de ce cours. En revanche, certaines conditions suffisantes à ce qu'une matrice soit totalement unimodulaire sont faciles à énoncer. Par exemple, si A est totalement unimodulaire, alors les matrices suivantes le sont également :

- A^T et $-A$
- $(A, \pm A)$
- (A, \mathbf{e}) où \mathbf{e} est un vecteur possédant exactement une composante non nulle égale à $+1$ ou -1
- toute matrice obtenue en multipliant par -1 toutes les entrées d'une ligne ou d'une colonne de A .

L'existence d'une solution optimale entière n'est pas conditionnée par la forme sous laquelle est exprimée le problème d'optimisation linéaire, comme l'indique par exemple le corollaire suivant.

Corollaire 5.4. *Considérons le problème linéaire*

$$\begin{array}{ll} \text{Min} & \mathbf{c} \cdot \mathbf{x} \\ \text{s.c.} & A\mathbf{x} \leq \mathbf{b} \\ & A\mathbf{x} \geq \mathbf{d}, \end{array}$$

où \mathbf{b} et \mathbf{d} sont des vecteurs à coefficients entiers, possiblement égaux à $\pm\infty$, et où A est totalement unimodulaire.

Si ce problème admet au moins une solution réalisable, il admet une solution réalisable ayant toutes ses coordonnées entières.

Si ce problème admet au moins une solution optimale, il admet une solution optimale ayant toutes ses coordonnées entières.

De même, l'algorithme du simplexe calculera une telle solution entière.

Démonstration du corollaire 5.4. Le problème peut se réécrire sous forme standard (voir Section 4.2), avec une nouvelle matrice A , mais qui reste totalement unimodulaire, comme on peut le voir en appliquant de manière répétée les constructions maintenant la totale unimodularité énoncées ci-dessus. \square

On a aussi la proposition suivante, établie par Poincaré en 1900 [10].

Proposition 5.5. *Soit A une matrice à coefficients dans $\{+1, -1, 0\}$. Si toute colonne contient au plus un $+1$ et au plus un -1 , alors A est totalement unimodulaire.*

Démonstration. Soit A une matrice carrée $n \times n$ à coefficients dans $\{+1, -1, 0\}$ et ayant toute colonne avec au plus un $+1$ et au plus un -1 . Nous allons montrer qu'alors $\det A \in \{+1, -1, 0\}$. La conclusion suivra alors automatiquement.

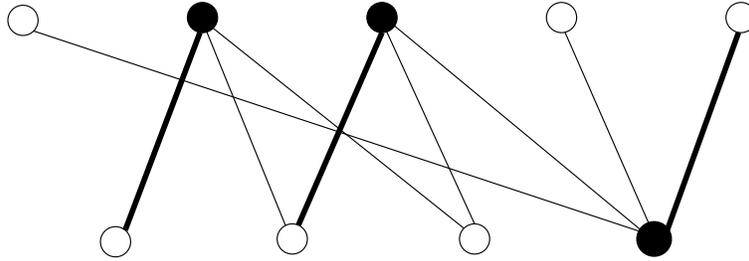


FIGURE 5.2 – Sur ce graphe biparti G , on a exhibé une couverture et un couplage, tous deux de cardinalité 3. La cardinalité d'un couplage étant toujours plus petite que celle d'une couverture, la couverture et le couplage exhibés sont tous deux optimaux, et l'on a bien $\nu(G) = \tau(G)$, comme énoncé par le théorème de König.

Procédons par récurrence sur n . Si $n = 1$, l'énoncé est évident. Prenons donc maintenant $n > 1$. Si l'une des colonnes possède au plus un élément non nul, la récurrence donne immédiatement le résultat. On peut donc supposer que A possède exactement un $+1$ et un -1 par colonne. La somme de toutes les lignes donne le vecteur $\mathbf{0}$, et donc $\det A = 0$. Ce qui achève la récurrence. \square

Application 5.2 (Théorème de König [7]). Un *graphe biparti* G est formé de deux ensembles finis disjoints U et V (dont les éléments sont les *sommets*) et d'un ensemble $E \subseteq U \times V$ (dont les éléments sont les *arêtes*). L'arête $e = (u, v) \in E$ est *incidente* à u et à v . On appelle *couplage* une partie M de E telle que tout sommet de G est incident à au plus un élément de M , et *couverture* un ensemble $C \subseteq U \cup V$ tel que toute arête est incidente à au moins un élément de C . Il est aisé de voir que, pour de tels ensembles, on a toujours $|M| \leq |C|$, mais on a bien plus. En notant $\nu(G)$ la cardinalité maximale d'un couplage de G , et $\tau(G)$ la cardinalité minimale d'une couverture de G , le théorème de König assure que l'on a $\nu(G) = \tau(G)$. Une illustration est donnée sur la figure 5.2. C'est une conséquence de la proposition 5.5 et des théorèmes 5.3 et 5.1, comme nous allons le voir maintenant.

Notons $A = (a_{v,e})$ la matrice suivante. Les lignes sont indicées par les sommets de G et les colonnes par ses arêtes. On pose

$$a_{v,e} = \begin{cases} +1 & \text{si } e \text{ est incidente à } v, \\ 0 & \text{sinon.} \end{cases}$$

A est totalement unimodulaire d'après la proposition 5.5 (il suffit de multiplier par -1 les lignes indicées par un élément de U pour s'en convaincre). Considérons maintenant le problème linéaire

$$\begin{aligned} \text{Max} \quad & \sum_{e \in E} x_e \\ \text{s.c.} \quad & A\mathbf{x} \leq \mathbf{1} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Si l'on impose en plus que ses solutions soient entières, alors ce problème linéaire modélise exactement le problème du couplage de cardinalité maximale. Comme la matrice A est totalement unimodulaire, la valeur optimale de ce problème linéaire est donc $\nu(G)$, sans avoir à

imposer que les solutions soient entières. Cela se voit en appliquant le corollaire 5.4 à

$$A' = \begin{pmatrix} A \\ I_E \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{1} \\ +\infty \end{pmatrix} \quad \text{et} \quad \mathbf{d} = \begin{pmatrix} -\infty \\ \mathbf{0} \end{pmatrix},$$

où I_E est la matrice identité dont les lignes et les colonnes sont indicées par les arêtes de G . Le dual de ce problème linéaire est

$$\begin{aligned} \text{Min} \quad & \sum_{v \in V} y_v \\ \text{s.c.} \quad & A^T \mathbf{y} \geq \mathbf{1} \\ & \mathbf{y} \geq \mathbf{0}. \end{aligned}$$

Si l'on impose en plus que ses solutions soient entières, alors ce problème linéaire modélise exactement le problème de la couverture de cardinalité minimale. A nouveau, la total unimodularité de A implique que sa valeur optimale est $\tau(G)$, sans avoir à imposer que les solutions soient entières. Le théorème 5.1 montre alors que $\nu(G) = \tau(G)$.

Bibliographie

- [1] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. [15](#), [26](#)
- [2] S. Bubeck, *Convex Optimization : Algorithms and Complexity*, now publishers inc., 2015. [26](#)
- [3] A. Cauchy, *Méthode générale pour la résolution des systèmes d'équations simultanées*, Comptes rendus hebdomadaires des séances de l'Académie des Sciences (1847). [22](#)
- [4] G. B. Dantzig, *Maximization of a linear function of variables subject to linear inequalities*, Activity Analysis of Production and Allocation (Tj. C. Koopmans, ed.), Wiley, 1951. [30](#)
- [5] G. B. Dantzig, A. Orden, and P. Wolfe, *The generalized simplex method for minimizing a linear form under linear inequality constraints*, Pacific Journal of Mathematics (1955). [32](#)
- [6] J. Farkas, *Vectortan és az egyszerű inaequatiók tana*, Kolozsvár (1900). [36](#)
- [7] D. König, *Über trennende Knotenpunkte in Graphen (nebst Anwendungen auf Determinanten und Matrizen)*, Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae (Szeged), Sectio Scientiarum Mathematicarum (1933). [41](#)
- [8] H. W. Kuhn and A. W. Tucker, *Nonlinear programming*, Proceedings of 2nd Berkeley Symposium, 1951. [8](#)
- [9] J. Matoušek and B. Gärtner, *Understanding and using linear programming*, Springer, 2007. [27](#)
- [10] H. Poincaré, *Second complément à l'analysis situs*, Proceedings of the London Mathematical Society (1900). [40](#)
- [11] H. Uzawa, *Iterative methods for concave programming*, Studies in linear and nonlinear programming (K. J. Arrow, L. Hurwicz, and H. Uzawa, eds.), Stanford University Press, 1958. [24](#)
- [12] J. Von Neumann, *Zur Theorie der Gesellschaftspiele*, Mathematischen Annalen (1928). [38](#)