

# Chapitre 3

## Problèmes de flots

Les problèmes de flots remontent aux origines de la recherche opérationnelle : ils comprennent comme cas particulier le problème de transport de masse, énoncé par Monge en 1781 et reformulé par Kantorovitch en termes de programmation linéaire vers 1942 ; leur théorie inclut les résultats fondamentaux sur les problèmes de couplage dans les graphes bipartis, dus à König, développée à partir de 1914, en relation avec des problèmes combinatoires apparaissant en algèbre linéaire (théorème de König-Frobenius). Un résultat central, le théorème de Ford-Fulkerson “flot-maximal=coupe-minimale”, démontré en 1956, permet de déterminer le flot maximal entre deux points donnés d’un réseau sachant que les flots sur les arcs sont majorés par des capacités attachées aux arcs. Les problèmes de flots à coût minimum constituent une classe générale permettant de modéliser et de résoudre de nombreux problèmes pratiques. Ils interviennent aussi comme outils dans le traitement de problèmes plus difficiles. La théorie des problèmes de flots offre un exemple remarquable de synthèse entre l’optimisation et la théorie des graphes, l’interprétation combinatoire des conditions d’optimalité et des résultats de dualité en programmation linéaire conduisant naturellement à des algorithmes polynomiaux très efficaces. Nous prenons ici le parti d’aller du général au particulier, en donnant un aperçu de la théorie des problèmes de flots à coût minimum, qui incluent les problèmes de plus court chemins comme cas particulier. Ces derniers peuvent être traités par des méthodes spéciales, notamment de programmation dynamique, que nous abordons au chapitre 4.

### 3.1 Flots dans un graphe : définitions, premières propriétés

Considérons un graphe orienté  $\mathcal{G} = (\mathcal{N}, \mathcal{A})$  : rappelons que  $\mathcal{N}$  est l’ensemble des *nœuds*, et  $\mathcal{A} \subset \mathcal{N} \times \mathcal{N}$  est l’ensemble des *arcs*. On munit chaque arc  $(i, j) \in \mathcal{A}$  d’une *capacité*  $u_{ij} \in \mathbb{R}_+ \cup \{+\infty\}$  et d’un *coût*  $c_{ij} \in \mathbb{R}$ . (Le “*u*” dans  $u_{ij}$  est pour “upper bound”.) On se donne aussi en chaque nœud du graphe un *flot entrant exogène*  $b_i \in \mathbb{R}$ ,

compté algébriquement (ce qui signifie que si  $b_i < 0$ , il s'agit d'un flot sortant en  $i$  d'une valeur  $-b_i$ ). On appelle *flot* (positif) une fonction  $x \in \mathbb{R}^{\mathcal{A}}$ ,  $(i, j) \mapsto x_{ij}$ , vérifiant la *loi des nœuds de Kirchhoff*, exprimant l'égalité des flots entrants et sortants en tous points :

$$b_i + \sum_{j \in \mathcal{N}, (j, i) \in \mathcal{A}} x_{ji} = \sum_{j \in \mathcal{N}, (i, j) \in \mathcal{A}} x_{ij}, \quad \forall i \in \mathcal{N} \quad (3.1)$$

ainsi que la contrainte de positivité

$$0 \leq x_{ij}, \quad \forall (i, j) \in \mathcal{A} . \quad (3.2)$$

Dans toute la suite, on supposera que le graphe n'a pas de boucles, c'est-à-dire pas d'arcs de la forme  $(i, i)$ , ce qui n'entraîne pas de perte de généralité (car les flots associés aux boucles disparaissent dans le bilan (3.1)).

On parle de *flot algébrique* lorsque l'on impose (3.1) mais pas (3.2). Dans le cas d'un graphe orienté, il est naturel de considérer des flots positifs plutôt que des flots algébriques, ce qui revient à dire que le flot sur un arc orienté de  $i$  à  $j$  circule en direction de  $i$  vers  $j$ . Dans la suite, les flots seront toujours positifs, sauf mention explicite du contraire.

Il est commode d'introduire le *flot net sortant* au nœud  $i$  :

$$f_i(x) := \sum_{j \in \mathcal{N}, (i, j) \in \mathcal{A}} x_{ij} - \sum_{j \in \mathcal{N}, (j, i) \in \mathcal{A}} x_{ji} ,$$

ce qui permet de résumer la loi des nœuds de Kirchhoff par :

$$b = f(x) .$$

Un flot tel que  $f(x) = 0$  est appelé *circulation*.

En pratique, un flot pourra représenter un fluide circulant en régime stationnaire sur un réseau (eau, gaz, courant électrique), mais aussi des biens (nombre de colis, des containers), si l'on considère des flots à valeurs entières.

En sommant les flots nets sortants  $f_i(x)$  sur l'ensemble des nœuds  $i$ , on obtient aussitôt :

**Observation 3.1** Une condition nécessaire pour l'existence d'un flot est que la somme des flots exogènes entrants soit nulle :

$$\sum_{i \in \mathcal{N}} b_i = 0 . \quad (3.3)$$

Nous supposerons toujours que la condition (3.3) est vérifiée.

Un flot est dit *admissible* s'il satisfait les contraintes de capacité :

$$x_{ij} \leq u_{ij}, \quad \forall (i, j) \in \mathcal{A} . \quad (3.4)$$

**Définition 3.2 (Problème de flot à coût minimum)** On appelle problème de *flot à coût minimum* le programme linéaire :

$$\min \sum_{(i,j) \in \mathcal{A}} c_{ij} x_{ij} \text{ sous les contraintes (3.1), (3.2), (3.4).} \quad (3.5)$$

Le problème de flot à coût minimum admet plusieurs sous-problèmes importants.

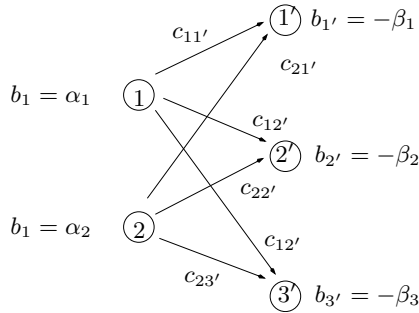
**Définition 3.3 (Problème de transport)** On appelle *problème de transport* un problème de flot à coût minimum pour lequel les capacités sont infinies, i.e.,  $u \equiv +\infty$ .

La terminologie “transport” est motivée par le problème fondamental suivant.

**Exemple 3.4 (Problème de transport de masse)** Reprenons à l’aide du formalisme des flots le problème de transport de masse, déjà mentionné dans l’exemple 1.9 du Chapitre 1. Dessinons pour cela un graphe orienté, avec  $n + m$  nœuds, les  $n$  premiers nœuds étant désignés par  $1, \dots, n$ , et les  $m$  suivants par  $1', \dots, m'$ . Pour tous  $1 \leq i \leq n$  et  $1 \leq j \leq m$ , dessinons un arc du nœud  $i$  au nœud  $j'$ . Si l’on veut être formel, on posera :  $\mathcal{N} := \{1, \dots, n, 1', \dots, m'\}$ ,  $\mathcal{A} = \{(i, j') \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ . Munissons l’arc  $(i, j')$  du coût  $c_{ij'}$  et de la capacité  $u_{ij'} = +\infty$ . Quant au vecteur  $b$  de flots exogènes, définissons le par :

$$b_i = \alpha_i, 1 \leq i \leq n, \quad b_{j'} = -\beta_j, 1 \leq j \leq m ,$$

où  $\alpha_i$  est la masse du déblai  $i$  et  $\beta_j$  la masse du remblai  $j$ . Par exemple, lorsque  $n = 2$  et  $m = 3$ , le graphe est :



En un nœud de type  $i$ , pour  $1 \leq i \leq n$ , la loi des nœuds de Kirchhoff,  $b_i = f_i(x)$ , donne

$$\alpha_i = \sum_{1 \leq j \leq m} x_{ij'} . \quad (3.6)$$

De même, un nœud de type  $j$ , pour  $1 \leq j' \leq n$ ,  $b_{j'} = f_{j'}(x)$ , donne

$$\sum_{1 \leq i \leq n} x_{ij'} = \beta_j . \quad (3.7)$$

On voit donc que l'ensemble des flots admissibles est exactement l'ensemble des plans de transport. Le problème de flot à coût minimum (3.5) consiste exactement à minimiser le coût d'un plan de transport.

**Remarque 3.5** Une variante du problème de transport de masse consiste à remplacer chaque contrainte (3.6) par

$$\alpha_i \geq \sum_{1 \leq j \leq m} x_{ij'} ,$$

ce qui modélise la situation où l'on s'autorise à consommer seulement une partie de la masse  $\alpha_i$  du déblai  $i$ . On conserve les contraintes (3.7). et l'on suppose maintenant que

$$\alpha_1 + \dots + \alpha_n \geq \beta_1 + \dots + \beta_m .$$

Ce problème se ramène à celui de l'exemple 3.4 : il suffit pour cela de rajouter un remblai imaginaire 0, de masse égale à la masse manquante,

$$\beta_0 := \alpha_1 + \dots + \alpha_n - \beta_1 - \dots - \beta_m ,$$

avec un coût de transport  $c_{i0'} = 0$ , pour tout  $1 \leq i \leq n$ .

**Remarque 3.6** Des contraintes inférieures et supérieures sur le flot

$$l_{ij} \leq x_{ij} \leq u_{ij}$$

peuvent se ramener au cas traité ici où  $l_{ij} = 0$ . Il suffit en effet d'effectuer le changement de variable  $x_{ij} = l_{ij} + x'_{ij}$ , où les nouvelles variables de flots  $x'_{ij}$  vérifient maintenant  $0 \leq x'_{ij} \leq u'_{ij}$  avec  $u'_{ij} = u_{ij} - l_{ij}$ , et satisfont une nouvelle loi des nœuds de Kirchhoff  $f(x') = b'$  avec  $b' = b - f(\ell)$ . Cette transformation ne fait que rajouter une constante au critère.

Un autre cas particulier fondamental du problème de flot à coût minimum est le *problème du flot maximal*, parfois appelé tout simplement *problème de flot*. Alors que dans les données du problème de transport, on avait des coûts, et pas de capacités, dans les données du problème du flot maximal, on aura des capacités, et pas de coûts. Pour définir le problème de flot maximal, on supposera que  $\mathcal{G}$  a deux nœuds distingués,  $s$  et  $p$ , appelés respectivement *source* et *puits*, tels que  $s$  n'a pas de prédécesseur (i.e.,  $\{i \in \mathcal{N} \mid (i, s) \in \mathcal{A}\} = \emptyset$ ), et  $p$  n'a pas de successeur (i.e.,  $\{i \in \mathcal{N} \mid (p, i) \in \mathcal{A}\} = \emptyset$ ). Soit  $v \in \mathbb{R}_+$ . On appelle *flot admissible de  $s$  à  $p$  de valeur  $v$*  un vecteur  $x = (x_{ij})_{(i,j) \in \mathcal{A}}$  tel que le flot net entrant en  $s$  soit égal à  $v$ ,  $f_s(x) = v$ , le flot net sortant en  $p$  soit égal à  $-v$ ,  $f_p(x) = -v$ , le flot net en tout autre nœud soit nul,  $f_i(x) = 0$  pour  $i \in \mathcal{N} \setminus \{s, p\}$ , les contraintes de positivité et de capacité étant satisfaites,  $0 \leq x_{ij} \leq u_{ij}$ . Le vecteur  $x$  est donc solution de (3.1),(3.2),(3.4), avec

$$b_i = \begin{cases} v & \text{si } i = s, \\ -v & \text{si } i = p, \\ 0 & \text{sinon.} \end{cases}$$

**Définition 3.7 (Problème du flot maximal)** Le problème du flot maximal consiste à trouver un flot admissible de  $s$  à  $p$  de valeur maximale.

**Remarque 3.8** Le problème du flot maximal de  $s$  à  $p$  est un cas particulier du problème de flot à coût minimal. Rajoutons en effet un arc de  $p$  à  $s$ , muni d'un coût  $-1$ , en munissant tous les autres arcs d'un coût nul, et donnons à ce nouvel arc une capacité  $+\infty$ , en conservant les autres capacités. Un flot  $x$  dans ce nouveau graphe tel que le flot net  $f_i(x)$  soit nul en tout nœud  $i$  correspond exactement à un flot admissible de  $s$  à  $p$  dans l'ancien graphe, et le coût du flot est l'opposé de sa valeur.

**Remarque 3.9** Dans l'énoncé du problème de flot maximal, l'hypothèse que le nœud source  $s$  n'a pas de prédécesseur, et que le nœud puits  $p$  n'a pas de successeur, n'est pas restrictive : si l'on cherche un flot d'une origine  $o$  à une destination  $d$  données, on peut toujours rajouter à un graphe un nœud source imaginaire  $s$  et un nœud puits imaginaire  $p$ , avec un arc de  $s$  à  $o$  de coût nul et de capacité infinie, et un arc de  $d$  à  $p$  de coût nul et de capacité infinie.

On peut décomposer un flot quelconque en termes de chemins. À chaque chemin de  $s$  à  $p$  et à chaque  $v \in \mathbb{R}_+$ , on associe un flot  $x$  de valeur  $v$ , tel que  $x_{ij} = v$  si  $(i, j)$  appartient au chemin, et  $x_{ij} = 0$  sinon. À chaque circuit du graphe on associe selon le même procédé un flot  $x$ , tel que  $x_{ij} = v$  si  $(i, j)$  appartient au circuit, et  $x_{ij} = 0$  sinon. Ces flots seront dits unitaires si  $v = 1$ . Rappelons qu'un chemin est *élémentaire* si tous ses nœuds sont distincts, et qu'un circuit est *élémentaire* si tous ses nœuds à l'exception du dernier sont distincts.

**Proposition 3.10 (Flots et chemins)** *Tout flot de  $s$  à  $p$  peut s'écrire comme combinaison linéaire positive de flots unitaires associés à des chemins élémentaires de  $s$  à  $p$  et de flots unitaires associés à des circuits élémentaires.*

Cette proposition se déduit du résultat suivant, grâce à la construction de la Remarque 3.8.

**Proposition 3.11** *Toute circulation dans un graphe est combinaison linéaire positive de flots unitaires associés à des circuits élémentaires.*

**Démonstration.** Dans le cas d'une circulation, la conservation des flots au nœud  $j$  s'écrit  $\sum_{(i,j) \in \mathcal{A}} x_{ij} = \sum_{(j,k) \in \mathcal{A}} x_{jk}$ . Donc, pour tout arc  $(i, j)$  tel que  $x_{(i,j)} > 0$ , on peut trouver un arc  $(j, k)$  tel que  $x_{(j,k)} > 0$ . Si une circulation  $x$  est non-nulle, on en déduit qu'il existe un chemin infini  $i_0, i_1, i_2, \dots$  tel que  $x_{i_0 i_1} > 0, x_{i_1 i_2} > 0, \dots$ . Comme le nombre de nœuds du graphe est fini, ce chemin passe deux fois par le même nœud. Autrement dit,  $i_{r+c} = i_r$  pour un certain  $r \geq 0$  et  $c \geq 1$ . Le chemin  $(i_r, \dots, i_{r+c})$  est donc un circuit, et ce circuit est même élémentaire si  $c$  est choisi minimal. En notant  $v$  le minimum des valeurs du flot sur ce circuit, et en définissant la circulation  $y$  valant 1 sur ce circuit, et 0 ailleurs, on voit que  $x' := x - vy$  est encore une circulation, et que par rapport à  $x$ ,  $x'$  s'annule sur au moins un arc supplémentaire. Par récurrence

sur le nombre d'arcs sur lesquels la circulation ne s'annule pas, on conclut que  $x$  est combinaison linéaire positive de flots unitaires associés à des circuits élémentaires. ■

**Remarque 3.12** La Proposition 3.10 fournit une bonne illustration du théorème de Minkowski-Weyl (Théorème 2.21). L'ensemble des flots de  $s$  à  $p$  de valeur unité est un polyèdre, dont les points extrémaux correspondent aux chemins élémentaires de  $s$  à  $p$ . Les directions extrémales du cône asymptotique de ce polyèdre correspondent aux circuits élémentaires.

Les problèmes de plus court chemin sont des cas particuliers de problèmes de flot à coût minimum. Considérons en effet le problème de calcul du chemin de coût minimum entre une origine  $o$  et une destination  $d$  d'un graphe orienté, les coûts (ou longueur) des arcs étant donnés par les  $c_{ij}$ . Cherchons pour cela un flot  $x$  de  $o$  à  $d$  de valeur 1 et de coût minimum, c'est-à-dire un vecteur  $x = (x_{ij})$  tel que  $x_{ij} \geq 0$  pour tout arc  $(i, j)$ , tel que le flot net sortant en tout nœud  $i$  distinct de  $o$  et  $d$  soit nul, soit  $f_i(x) = 0$  pour  $i \in \mathcal{N} \setminus \{o, d\}$ , et tel que  $f_o(x) = -f_d(x) = 1$ , minimisant le coût :

$$\sum_{(i,j) \in \mathcal{A}} c_{ij} x_{ij}$$

En utilisant la Proposition 3.10 (et éventuellement la Remarque 3.9), on voit que l'optimum, s'il existe, est atteint par un flot associé à un chemin de  $o$  à  $d$ .

Notons que la loi des nœuds de Kirchhoff (3.1) peut s'écrire  $Mx = b$ , où la matrice  $M \in \mathbb{R}^{\mathcal{N} \times \mathcal{A}}$ , appelée *matrice d'incidence nœuds-arcs* de  $\mathcal{G}$ , est définie par

$$M_{i,(j,k)} = \begin{cases} -1 & \text{si } i = k, \\ 1 & \text{si } i = j, \\ 0 & \text{sinon.} \end{cases}$$

Notons que la matrice d'incidence est bien définie, car le graphe est supposé sans boucles. L'ensemble des flots admissibles s'écrit donc

$$\{x \in \mathbb{R}^{\mathcal{A}} \mid Mx = b, 0 \leq x \leq u\} . \tag{3.8}$$

**Proposition 3.13** *La matrice d'incidence nœuds-arcs d'un graphe est totalement unimodulaire.*

**Démonstration.** C'est une conséquence immédiate de la Proposition 2.37. ■

Le résultat suivant est fondamental.

**Corollaire 3.14 (Optimalité des flots entiers)** *Si les flots entrants exogènes  $b_i$  sont entiers, et si les capacités  $u_{ij}$  sont entières ou infinies, alors, les points extrémaux de l'ensemble (3.8) des solutions admissibles d'un problème de flot à coût minimal sont entiers. En particulier, si l'ensemble des flots admissibles est borné et non-vide, il existe un flot entier optimal.*

**Démonstration.** C'est une conséquence immédiate du Corollaire 2.34 et de la Proposition 3.13. ■

**Exercice 3.15** Montrer que le problème de couverture (exercice 2.38) peut se modéliser par un problème de flot à coût minimum, et retrouver ainsi la conclusion de l'exercice 2.38.

**Exemple 3.16 (Problème d'affectation optimale)** Ce problème, objet de l'Exemple 1.6, a pour expression

$$\max_{\sigma \in \mathfrak{S}_n} \sum_{1 \leq i \leq n} s_{i\sigma(i)} . \tag{3.9}$$

Il s'agit d'une version avec contrainte d'intégrité du du problème de transport de masse (exemple 3.4), consistant à maximiser

$$\sum_{1 \leq i, j \leq n} s_{ij} x_{ij} \tag{3.10}$$

sur l'ensemble des  $x$  vérifiant :

$$\forall 1 \leq i \leq n, \quad 1 = \sum_{1 \leq k \leq n} x_{ik}, \tag{3.11a}$$

$$\forall 1 \leq j \leq n, \quad 1 = \sum_{1 \leq k \leq n} x_{kj}, \tag{3.11b}$$

$$\forall 1 \leq i, j \leq n, \quad 0 \leq x_{ij}. \tag{3.11c}$$

Ce problème de transport de masse est lui même un problème de flot à coût minimum. Il résulte alors du théorème d'optimalité des flots entiers (Corollaire 3.14) que l'optimum de ce problème est atteint pour un  $x$  entier, c'est-à-dire, pour un  $x$  représentant une matrice de permutation. Ce résultat a une interprétation plaisante en termes de mariage : si l'on envisage de marier  $n$  garçons et  $n$  filles,  $s_{ij}$  étant le bonheur du couple d'individus  $(i, j)$ , la valeur du problème d'affectation optimale représente la valeur d'un ensemble de  $n$  mariages monogames, alors qu'une solution non-entière  $x$  de (3.11) correspond à un ensemble de mariages polygames :  $x_{ij}$  désigne le pourcentage de temps que  $i$  passe avec  $j$ . Le théorème d'intégrité des flots démontre ainsi que la monogamie est optimale. Cette conclusion est due au caractère linéaire du critère. On pourrait considérer plus généralement un critère de la forme :

$$\sum_{1 \leq i, j \leq n} s_{ij} x_{ij}^\gamma ,$$

avec  $\gamma > 0$ . Supposons les coefficients  $s_{ij}$  positifs. Alors, le cas où  $\gamma < 1$  fournit un critère concave, qui modélise la lassitude. Dans ce cas, l'optimum n'est plus nécessairement atteint en un  $x$  entier. Par contre, si  $\gamma > 1$ , le critère est convexe, et d'après le Corollaire 2.34, le maximum du critère est encore atteint en un point entier.

## 3.2 Algorithmes de flots

Nous présentons maintenant les algorithmes de flots, qui vont nous permettre de traduire de manière efficace le résultat théorique d'intégrité des flots (Corollaire 3.14). Nous commençons par le problème du flot de valeur maximale (Définition 3.7), car l'algorithme est plus simple à présenter dans ce cas. Nous cherchons donc un flot de valeur maximale  $v$  d'une source  $s$  à un puits  $p$ , qui soit admissible, c'est-à-dire qui vérifie la contrainte de capacité (3.4).

### 3.2.1 Flots maximaux et coupes minimales

Nous allons tout d'abord majorer la valeur  $v$  d'un flot admissible. Appelons pour cela *coupe* du graphe  $\mathcal{G} = (\mathcal{N}, \mathcal{A})$  une partition  $\mathcal{N} = S \cup \bar{S}$  telle que  $S \neq \emptyset$  et  $S \neq \mathcal{N}$ . On dit que cette coupe *sépare*  $s$  de  $p$  si  $s \in S$  et  $p \in \bar{S}$ . Si  $I$  et  $J$  sont des sous-ensembles de  $\mathcal{N}$ , et si  $y$  est une fonction de  $\mathcal{A}$  dans  $\mathbb{R}_+ \cup \{+\infty\}$ , nous définissons :

$$y(I, J) = \sum_{i \in I, j \in J, (i,j) \in \mathcal{A}} y_{ij} .$$

En particulier,  $u(S, \bar{S}) = \sum_{i \in S, j \in \bar{S}, (i,j) \in \mathcal{A}} u_{ij}$  est appelée *capacité* de la coupe  $(S, \bar{S})$ .

**Lemme 3.17** *La valeur d'un flot admissible de  $s$  à  $p$  est majorée par la capacité de toute coupe séparant  $s$  de  $p$ .*

**Démonstration.** On va montrer que pour toute coupe  $(S, \bar{S})$  séparant  $s$  de  $p$  et pour tout flot  $x$  de valeur  $v$

$$x(S, \bar{S}) - x(\bar{S}, S) = v . \tag{3.12}$$

En effet, en sommant les lois des nœuds (3.1) pour tout  $i \in S$ , il vient

$$\sum_{i \in S} \left( \sum_{j \in \mathcal{N}, (i,j) \in \mathcal{A}} x_{ij} - \sum_{j \in \mathcal{N}, (j,i) \in \mathcal{A}} x_{ji} \right) = v , \tag{3.13}$$

Quand  $i$  et  $j$  appartiennent tous deux à  $S$ ,  $x_{ij}$  se simplifie dans (3.13), et quand  $i$  et  $j$  appartiennent tous deux à  $\bar{S}$ ,  $x_{ij}$  n'apparaît pas dans (3.13), ce qui nous permet de réécrire (3.13) sous la forme (3.12). Le lemme résulte aussitôt de (3.12), de  $x(S, \bar{S}) \leq u(S, \bar{S})$ , et de  $x(\bar{S}, S) \geq 0$ . ■

Le théorème fondamental de la théorie des flots, dû à Ford et Fulkerson [FF56], affirme que la majoration du Lemme 3.17 est optimale, i.e. que "max flow=min cut".

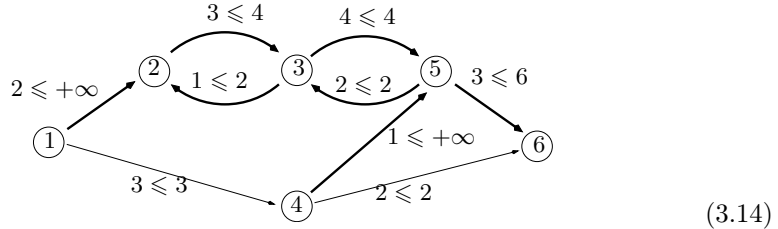
**Théorème 3.18 (Flot maximal=coupe minimale)** *Le supremum des valeurs des flots admissibles de  $s$  à  $p$  est égal à l'infimum des capacités des coupes séparant  $s$  de  $p$ .*



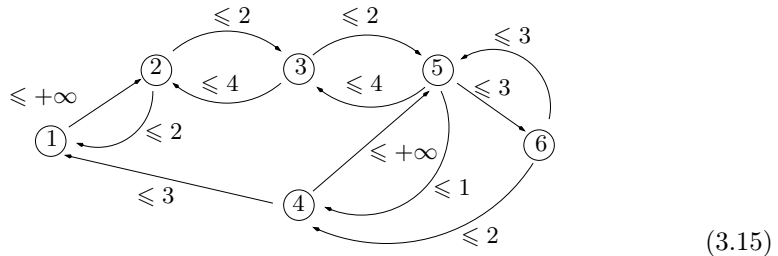
La preuve de ce théorème utilise la notion de *graphe résiduel* que nous introduisons maintenant. Afin de rendre plus simples les écritures algébriques, il sera commode de raisonner dans un graphe augmenté, toujours noté  $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ , tel que s'il y a un arc  $(i, j) \in \mathcal{A}$ , il y a aussi un arc en sens inverse,  $(j, i) \in \mathcal{A}$  : si l'arc  $(i, j)$ , mais non l'arc  $(j, i)$ , est dans le graphe initial, nous rajoutons l'arc  $(j, i)$  en le munissant d'une capacité  $u_{ji} = 0$ , ce qui ne change rien au problème. (Nous ne représenterons pas les arcs de capacité nulle sur les dessins.) Considérons donc un flot admissible  $x$ , et appelons, pour tout  $(i, j) \in \mathcal{A}$ , flot net de  $i$  à  $j$  la quantité  $x_{ij} - x_{ji}$ . Si l'on cherche à accroître localement le flot net, on peut augmenter  $x_{ij}$  de  $u_{ij} - x_{ij}$ , et symétriquement faire passer  $x_{ji}$  à 0. Aussi appelons-nous *capacité résiduelle*  $r_{ij}$  de l'arc  $(i, j)$  l'accroissement maximal possible du flot net de  $i$  à  $j$ , soit

$$r_{ij} = u_{ij} - x_{ij} + x_{ji} .$$

Le graphe  $\mathcal{G}_r(x) = (\mathcal{N}, \mathcal{A}_r)$ , où  $\mathcal{A}_r = \{(i, j) \in \mathcal{N}^2 \mid r_{ij} > 0\}$  est appelé *graphe résiduel*. La capacité d'une coupe  $(S, \bar{S})$  dans le graphe résiduel est par définition  $r(S, \bar{S})$ . Par exemple, considérons le graphe suivant muni d'un flot et de capacités :



On a représenté sur chaque arc les valeurs “ $x_{ij} \leq u_{ij}$ ”. Ici, la source est  $s = 1$ , le puits est  $p = 6$ , et la valeur du flot est 5. On a indiqué en traits gras dans (3.14) les arcs qui appartiennent au graphe résiduel. Le graphe résiduel lui même peut être représenté comme suit avec ses capacités résiduelles :



Le lemme suivant nous permet de vérifier l'optimalité d'un flot.

**Lemme 3.19** *Soit  $x$  un flot admissible de valeur  $v$ . Les assertions suivantes sont équivalentes :*

1.  $x$  est de valeur maximale,

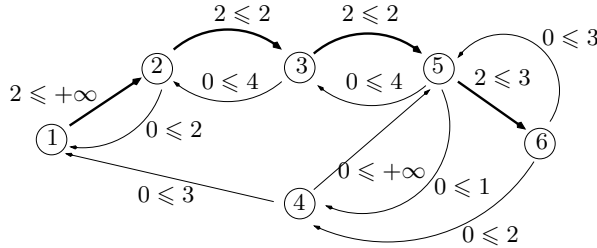
2. il n'existe pas de chemin de  $s$  à  $p$  dans le graphe résiduel  $\mathcal{G}_r(x)$ ,
3. il existe une coupe de capacité 0 séparant  $s$  de  $p$  dans le graphe résiduel  $\mathcal{G}_r(x)$ ,
4. il existe une coupe de capacité  $v$  séparant  $s$  de  $p$  dans le graphe original.

**Démonstration.** Montrons d'abord (non 2)  $\Rightarrow$  (non 1). Soit  $(\ell_0, \dots, \ell_k)$  un chemin de  $s$  à  $p$  dans  $\mathcal{G}_r(x)$ , que l'on peut toujours supposer élémentaire, et notons  $m = \min(r_{\ell_0\ell_1}, \dots, r_{\ell_{k-1}\ell_k})$  la capacité résiduelle minimale le long du chemin. Pour chaque arc  $(i, j)$  du chemin, par définition du graphe résiduel, il est possible d'accroître le flot net de  $i$  à  $j$  de  $m$  en décomposant  $m = \alpha_{ij} + \beta_{ji}$ , avec  $0 \leq \alpha_{ij} \leq u_{ij} - x_{ij}$  et  $0 \leq \beta_{ji} \leq x_{ji}$ , puis, d'une part, en accroissant le flot de  $\alpha_{ij}$  sur l'arc  $(i, j)$ , et d'autre part, en diminuant le flot de  $\beta_{ji}$  sur l'arc  $(j, i)$ . Formellement, définissons le flot  $x'$  :

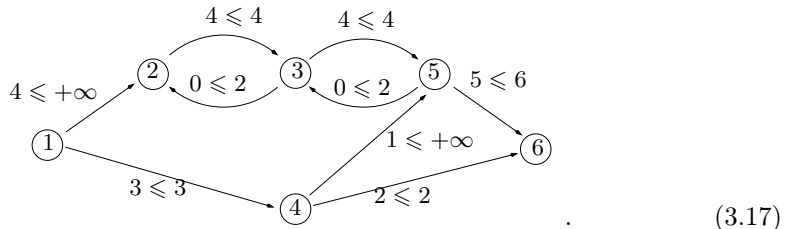
$$x'_{ij} = \begin{cases} x_{ij} + \alpha_{ij} & \text{si } (i, j) \text{ appartient au chemin } (\ell_0, \dots, \ell_k), \\ x_{ij} - \beta_{ij} & \text{si } (j, i) \text{ appartient au chemin } (\ell_0, \dots, \ell_k), \\ x_{ij} & \text{sinon.} \end{cases} \quad (3.16)$$

(Comme l'on a supposé le chemin élémentaire,  $(i, j)$  et  $(j, i)$  ne peuvent tous deux appartenir au chemin, ce qui garantit que les deux premiers cas de (3.16) sont bien exclusifs.) Le flot  $x'$  est encore admissible, et sa valeur est  $v + m$ , ce qui montre que  $x$  n'est pas maximal.

Illustrons cette partie de la preuve sur l'exemple (3.14). Dans ce cas, il y a un seul chemin élémentaire de  $s$  à  $p$  dans  $\mathcal{G}_r(x)$ , et sa capacité vaut  $m = 2$ . Ce chemin est représenté comme suit en gras, avec le flot dans le graphe résiduel qui lui est associé :



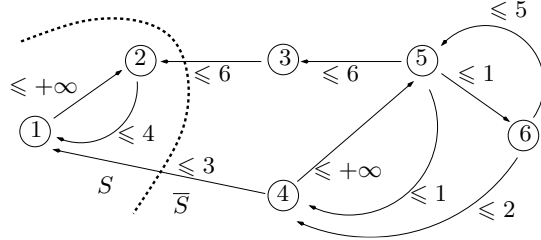
En revenant au graphe initial, partant du flot  $x$  de valeur 5, la construction (3.16) fournit le flot  $x'$  suivant de valeur 7 :



Revenons à la preuve, et montrons maintenant (2)  $\Rightarrow$  (3). Considérons l'ensemble  $S$  des nœuds  $i \in \mathcal{N}$  tel qu'il existe un chemin de  $s$  à  $i$  dans le graphe résiduel  $\mathcal{G}_r$ , et

soit  $\bar{S} = \mathcal{N} \setminus S$ . Par définition de  $S$ ,  $s \in S$ . L'assertion 2 entraîne que  $p \notin S$ . D'autre part, pour tout  $i \in S$  et  $j \in \bar{S}$  tels que  $(i, j) \in \mathcal{A}$ , nous avons nécessairement  $r_{ij} = 0$  : sinon, on aurait  $j \in S$ , une contradiction. Ainsi,  $r(S, \bar{S}) = \sum_{i \in S, j \in \bar{S}, (i, j) \in \mathcal{A}} r_{ij} = 0$ .

A titre d'illustration, cet argument appliqué au graphe résiduel du flot (3.17) fournit la coupe suivante :



Montrons ensuite (3) $\Rightarrow$ (4). Soit  $(S, \bar{S})$  une coupe de capacité nulle séparant  $s$  de  $p$  dans le graphe résiduel. Notons  $v$  la valeur de  $x$ . Par définition du graphe résiduel,  $u_{ij} = x_{ij}$  et  $x_{ji} = 0$  pour tout arc  $(i, j)$  du graphe résiduel tel que  $i \in S$  et  $j \in \bar{S}$ , ce qui nous permet de réécrire (3.12) sous la forme

$$\begin{aligned} v &= \sum_{i \in S, j \in \bar{S}, (i, j) \in \mathcal{A}} x_{ij} - \sum_{i \in S, j \in \bar{S}, (j, i) \in \mathcal{A}} x_{ji} \\ &= \sum_{i \in S, j \in \bar{S}, (i, j) \in \mathcal{A}} u_{ij} - \sum_{i \in S, j \in \bar{S}, (j, i) \in \mathcal{A}} 0 = u(S, \bar{S}), \end{aligned}$$

ce qui montre (4).

Enfin, (4) $\Rightarrow$ (1) résulte aussitôt du Lemme 3.17. ■

**Démonstration du Théorème 3.18.** Le calcul de flot maximum est un programme linéaire réalisable puisque le flot nul l'est. Si sa valeur est finie, il a une solution et on conclut avec l'implication (1) $\Rightarrow$ (4) du Lemme 3.19. Si elle vaut  $+\infty$ , on conclut avec le lemme 3.17. ■

### 3.2.2 Algorithme de Ford et Fulkerson

L'algorithme de Ford et Fulkerson pour calculer le flot admissible maximal dans un graphe consiste à appliquer la preuve du Lemme 3.19 : on part d'un flot admissible  $x^1$  quelconque, par exemple le flot nul, on construit ensuite le graphe résiduel  $\mathcal{G}_r(x^1)$  associé à  $x^1$ , puis l'on recherche un chemin de  $s$  à  $p$  dans  $\mathcal{G}_r(x^1)$ , (ce qui peut se faire en temps linéaire en le nombre d'arcs et de sommets de  $\mathcal{G}_r(x^1)$ ). Si l'on trouve un tel chemin, on construit un nouveau flot  $x^2$  de valeur strictement supérieure à la valeur de  $x^1$ , et l'on continue de la sorte. Nous avons donné dans la preuve du Lemme 3.19 une trace de cet algorithme.

Pour appliquer l'algorithme de Ford et Fulkerson en toute généralité, il faut d'abord vérifier que la valeur des flots admissibles de  $s$  à  $p$  est majorée ce qui peut se faire (toujours en temps linéaire) grâce à l'exercice suivant.

**Exercice 3.20** Montrer que la valeur des flots admissibles de  $s$  à  $p$  n'admet pas de majorant fini si, et seulement si, il existe un chemin de  $s$  à  $p$  formé d'arcs de capacité infinie.

Nous dirons qu'un flot  $x$  est *entier* si  $x_{ij} \in \mathcal{N}$ , pour tout  $(i, j) \in \mathcal{A}$ . Nous pouvons maintenant énoncer le résultat de convergence.

**Proposition 3.21** *Supposons que les capacités  $u_{ij}$  des arcs soient entières ou  $+\infty$ , supposons en outre qu'il n'existe pas de chemin de  $s$  à  $p$  dont tous les arcs ont des capacités infinies, et soit  $v^*$  la valeur maximale d'un flot admissible de  $s$  à  $p$ . Alors, l'algorithme de Ford et Fulkerson fournit un flot admissible maximal de  $s$  à  $p$ , qui est entier, en temps  $O(|\mathcal{A}|v^*)$ .*

**Démonstration.** Lorsque les données sont entières, que l'on part d'un flot initial entier, et que l'on applique (3.16) en choisissant des valeurs entières de  $\alpha_{ij}$  et  $\beta_{ij}$ , l'algorithme de Ford et Fulkerson produit des flots entiers dont la valeur s'accroît d'au moins une unité à chaque itération. Le nombre d'itérations est donc au plus  $v^*$ . Chaque itération prend un temps  $O(|\mathcal{A}|)$ . ■

**Remarque 3.22** La proposition 3.21 permet de déduire le résultat suivant : si  $u$  est à valeurs entières ou  $+\infty$ , et si la valeur des flots admissibles est majorée alors la valeur maximale des flots de  $s$  à  $p$  est atteinte par un flot entier. Nous avons donc retrouvé de manière algorithmique, dans un cas particulier, le théorème d'optimalité des flots entiers (Corollaire 3.14).

**Remarque 3.23 (Approche via la programmation linéaire)** La notion de coupe se retrouve à l'aide du théorème de dualité en programmation linéaire. Le problème du flot maximal admissible de  $s$  à  $p$  consiste en effet à maximiser la valeur  $v$  sur l'ensemble des couples  $(x, v)$  dans  $\mathbb{R}^{\mathcal{A}} \times \mathbb{R}$ , tels que  $x \geq 0$ , tel que les lois des nœuds  $f_i(x) = 0$  pour  $i \in \mathcal{N} \setminus \{s, p\}$ ,  $v = f_s(x)$ , et  $-v = f_p(x)$ , soient vérifiées, ainsi que  $x_{ij} \leq u_{ij}$ . Nous supposons toutes les capacités  $u_{ij}$  finies, afin d'alléger la notation. Notons  $y = (y_i) \in \mathbb{R}^{\mathcal{N}}$  le multiplicateur associé aux lois des nœuds,  $z = (z_{ij}) \in \mathbb{R}_+^{\mathcal{A}}$  le multiplicateur associé aux contraintes  $x_{ij} \leq u_{ij}$ , et introduisons le Lagrangien

$$L(x, v; y, z) = v + \sum_{i \in \mathcal{N} \setminus \{s, p\}} y_i f_i(x) + y_s(f_s(x) - v) + y_p(f_p(x) + v) + \sum_{(i, j) \in \mathcal{A}} z_{ij}(u_{ij} - x_{ij}),$$

de sorte que le problème du flot admissible maximum de  $s$  à  $p$  peut s'écrire :

$$\sup_{x \in \mathbb{R}_+^{\mathcal{A}}, v \in \mathbb{R}} \inf_{y \in \mathbb{R}^{\mathcal{N}}, z \in \mathbb{R}_+^{\mathcal{A}}} L(x, v; y, z).$$

Le problème dual s'obtient en commutant le sup et l'inf. En faisant apparaître les variables  $x_{ij}$  et  $v$  en facteur, on obtient la forme suivante du problème dual :

$$\inf_{y \in \mathbb{R}^{\mathcal{N}}, z \in \mathbb{R}_+^{\mathcal{A}}} \sup_{x \in \mathbb{R}_+^{\mathcal{A}}, v \in \mathbb{R}} \sum_{(i,j) \in \mathcal{A}} x_{ij}(y_i - z_{ij} - y_j) + v(1 - y_s + y_p) + \sum_{(i,j) \in \mathcal{A}} u_{ij}z_{ij}, \quad (3.18)$$

soit

$$(D) \quad \inf_{y \in \mathbb{R}^{\mathcal{N}}, z \in \mathbb{R}_+^{\mathcal{A}}} \sum_{(i,j) \in \mathcal{A}} u_{ij}z_{ij}; \quad y_i \leq z_{ij} + y_j, \quad \forall (i,j) \in \mathcal{A}, \quad 1 + y_p = y_s \quad .$$

Soit  $(y, z)$  une solution duale. Ajoutons la contrainte  $\pm y_i \leq M$ , où  $M \in \mathbb{N}$ ,  $M > \max_i |y_i|$ . Alors l'ensemble admissible dual ne contient pas de droite affine. La matrice du problème dual étant totalement unimodulaire (appliquer la proposition 2.37 à la transposée de cette matrice), le corollaire 2.34 assure l'existence d'une solution duale  $(y, z)$  entière. L'ensemble des solutions duales étant invariant par ajout d'une même constante à chaque composante de  $y$ , on peut supposer que  $y_p = 0$  et  $y_s = 1$ .

Définissons  $(y', z') \in \mathbb{R}^{\mathcal{N}} \times \mathbb{R}_+^{\mathcal{A}}$  par  $y'_i := \max(0, \min(1, y_i))$ , pour tout  $i \in \mathbb{R}^{\mathcal{N}}$ , et pour tout  $ij \in \mathcal{A}$ ,  $z'_{ij} := (y'_i - y'_j)_+$ . Alors  $(y', z')$  est admissible, et  $z'_{ij} \leq z_{ij}$  pour tout  $ij \in \mathcal{A}$ , donc  $\sum_{(i,j) \in \mathcal{A}} u_{ij}z'_{ij} \leq \sum_{(i,j) \in \mathcal{A}} u_{ij}z_{ij}$ , ce qui prouve que  $(y', z')$  est solution duale. Nous avons construit une solution duale de composantes dans  $\{0, 1\}$ .

On retrouve les coupes séparant  $s$  de  $p$  à partir de cette formulation. En effet, soit  $(y, z)$  une solution duale de composantes dans  $\{0, 1\}$ . Posons  $S = \{i \in \mathcal{N} \mid y_i = 1\}$ , et  $\bar{S} = \mathcal{N} \setminus S$ . On voit qu'on a nécessairement  $s \in S$  et  $p \in \bar{S}$ , et que le plus petit vecteur  $z$  vérifiant  $y_i \leq z_{ij} + y_j$  n'est autre que le vecteur indicateur  $z$  de la coupe  $(S, \bar{S})$ , qui est tel que  $z_{ij} = 1$  si  $i \in S$ ,  $j \in \bar{S}$  et  $z_{ij} = 0$  sinon. En outre,  $\sum_{(i,j) \in \mathcal{A}} u_{ij}z_{ij}$  n'est autre que la capacité  $u(S, \bar{S})$  de la coupe  $(S, \bar{S})$ .

Nous retrouvons ainsi comme conséquence du théorème de dualité faible en programmation linéaire le fait que la valeur du flot maximal admissible de  $s$  à  $p$  est majorée par la capacité de toute coupe séparant  $s$  de  $p$ . En fait, le problème dual (D) admet une solution  $(y, z)$  telle que  $z$  soit la fonction indicatrice d'une coupe minimale. En effet, si  $(S, \bar{S})$  est une coupe de capacité minimale, à laquelle sont associés les vecteurs  $y, z$  définis ci-dessus, la valeur correspondante du critère du problème dual (D) est  $u(S, \bar{S})$ , qui est égal à la valeur maximale d'un flot  $x$ , d'après le théorème de Ford-Fulkerson, autrement dit, à la valeur du problème primal, ce qui montre que  $(y, z)$  est une solution optimale du problème dual.

**Exercice 3.24 (Choix de représentants et problèmes de flots)** Une ville a  $r$  citoyens  $M_1, \dots, M_r$ ,  $q$  clubs  $C_1, \dots, C_q$ , et  $n$  partis politiques  $P_1, \dots, P_n$ . (Rappelons qu'un club, ou un parti politique, est un ensemble de citoyens.) On suppose que chaque citoyen est membre d'au moins un club, et d'au plus un parti politique. Chaque club désigne l'un de ses membres pour le représenter au conseil municipal. Si un membre du conseil municipal représente plusieurs clubs, il a autant de voix que de clubs qui

l'ont désigné (par exemple s'il est désigné par deux clubs, son avis compte double). On voudrait savoir s'il est possible de choisir de tels représentants de sorte que pour tout  $1 \leq i \leq n$ , le nombre total des voix détenues par les membres du conseil municipal appartenant au parti  $P_i$  n'excède pas un quota  $u_i$  fixé a priori.

1. Montrer que le problème plus général suivant : “trouver le nombre maximal  $N \leq q$  de clubs qui peuvent choisir chacun un représentant de sorte que, pour tout  $1 \leq i \leq n$ , le nombre total de voix détenues par les membres du conseil municipal appartenant au parti  $P_i$  n'excède pas  $u_i$ ” est équivalent à un problème de flot maximal dans un certain graphe que l'on décrira. On pourra notamment expliciter le graphe pour l'exemple suivant :  $r = 7$ ,  $q = 4$ ,  $n = 3$ ,  $C_1 = \{M_1, M_3\}$ ,  $C_2 = \{M_1, M_2, M_3\}$ ,  $C_3 = \{M_2, M_3, M_4\}$ ,  $C_4 = \{M_2, M_4, M_5, M_6, M_7\}$ ,  $P_1 = \{M_1, M_2\}$ ,  $P_2 = \{M_3, M_4\}$ ,  $P_3 = \{M_5, M_6, M_7\}$ ,  $u_1 = 1$ ,  $u_2 = 1$ ,  $u_3 = 2$ .

2. Démontrer que pour l'exemple qui précède, on a  $N < 4$  (au moins un club ne peut choisir de représentant). Indication : on pourra utiliser le théorème max-flot=min-cut.

3. Peut-on modéliser par un problème de flot la variante dans laquelle deux clubs ne peuvent pas désigner le même représentant ?

**Exercice 3.25 (Un problème de coupe minimale apparaissant en traitement d'image).** On représente un dessin en noir et blanc avec  $n$  pixels par une application  $\{1, \dots, n\} \rightarrow \{\pm 1\}$  : la valeur d'un pixel est  $-1$  ou  $+1$  (pour blanc et noir).

Soit  $\alpha : \{1, \dots, n\} \rightarrow \{\pm 1\}, i \mapsto \alpha_i$  une image initiale, qui peut être dégradée. L'image restaurée  $\sigma : \{1, \dots, n\} \rightarrow \{\pm 1\}, i \mapsto \sigma_i$  est obtenue en minimisant une énergie du type :

$$J(\sigma) = \sum_{1 \leq i \leq n} -\alpha_i \sigma_i - \sum_{1 \leq i, j \leq n} \gamma_{ij} \sigma_i \sigma_j,$$

où les  $\gamma_{ij} \geq 0$  sont donnés. (En pratique,  $\gamma_{ij} > 0$  si  $i$  et  $j$  sont des pixels voisins, et  $\gamma_{ij} = 0$  sinon.)

En effet, dans ce problème de minimisation, le terme  $-\alpha_i \sigma_i$  incite la valeur reconstruite  $\sigma_i$  à coïncider avec la valeur initiale  $\alpha_i$ , alors que, lorsque  $\gamma_{ij} > 0$ , le terme  $\gamma_{ij} \sigma_i \sigma_j$ , incite les valeurs  $\sigma_i$  et  $\sigma_j$  à être identiques (régularisation).

Nous allons montrer que la minimisation de  $J$  se réduit au calcul d'un flot maximum dans un graphe.

1. Dire pourquoi il n'y a pas de perte de généralité à supposer que  $\gamma_{ij} = \gamma_{ji}$ , ce que l'on fera désormais.

2. À toute image  $\sigma$ , on associe la partition de  $\{1, \dots, n\}$  formée des deux ensembles :

$$T^+ = \{i \mid \sigma_i = 1\}, \quad T^- = \{i \mid \sigma_i = -1\}$$

On forme le graphe orienté complet dont les sommets sont  $1, \dots, n$  (complet signifie que pour chaque  $1 \leq i, j \leq n$ , l'arc orienté  $(i, j)$  est présent), l'arc  $(i, j)$  étant muni de la capacité  $\gamma_{ij}$ .

Montrer que

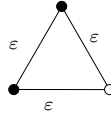
$$\sum_{1 \leq i, j \leq n} \gamma_{ij}(1 - \sigma_i \sigma_j) / 2 = 2C(T^+, T^-)$$

où  $C(T^+, T^-)$  désigne la capacité de la coupe  $(T^+, T^-)$  (avec la convention que la capacité est nulle si  $T^+$  ou  $T^-$  est vide).

3. Montrer qu'en rajoutant un nœud source et un nœud puits au graphe, reliés au nœuds  $1, \dots, n$  par des arcs de capacités bien choisies, on peut interpréter  $J(\sigma)$  comme la capacité d'une coupe, à des facteurs additifs et multiplicatifs près, et conclure que la minimisation de l'énergie  $J$  se ramène au calcul d'une coupe minimale séparant  $s$  de  $p$ .

Signalons que ce résultat n'a plus lieu si certains des  $\gamma_{ij}$  sont strictement négatifs. En particulier, lorsque tous les  $\gamma_{ij}$  sont négatifs, on obtient le problème "MAXCUT" (maximiser la capacité une coupe au lieu de la minimiser), qui est NP-dur. (Les problèmes de maximisation et de minimisation ne sont pas équivalents via un changement de signe, car les capacités des arcs sont toujours supposées positives ou nulles.)

4. Discuter la régularisation de la petite image suivante (3 pixels voisins dont deux noirs et un blanc) en fonction de la valeur de  $\gamma_{ij} \equiv \epsilon > 0, \forall 1 \leq i, j \leq 3$ .



### 3.2.3 Algorithme d'accroissement sur le chemin le plus court

L'algorithme d'accroissement du flot sur le chemin le plus court, dû à *Dinits* et à *Edmonds et Karp*, est une variante de l'algorithme de Ford et Fulkerson, qui permet de remplacer la borne pseudo-polynomiale de la proposition 3.21 par une véritable borne polynomiale. Ce raffinement consiste tout simplement à sélectionner à chaque étape de l'algorithme de Ford et Fulkerson, le plus court chemin de  $s$  à  $p$ , c'est-à-dire le chemin de  $s$  à  $p$  qui a le plus petit nombre d'arcs.

**Théorème 3.26** *L'algorithme d'accroissement du flot sur le chemin le plus court effectue au plus  $|\mathcal{N}||\mathcal{A}|$  accroissements.*

**Démonstration.** Notons  $d_x(i, j)$  la distance de  $i$  à  $j$  dans le graphe résiduel  $\mathcal{G}_r(x)$ . On pose  $d_x(i, j) = +\infty$  s'il n'existe pas de chemin de  $i$  à  $j$ . On prétend que si l'on passe de  $x$  à  $x'$  en effectuant l'accroissement (3.16) sur un chemin de longueur minimale, alors

$$d_x(s, p) \leq d_{x'}(s, p) . \tag{3.19}$$

Pour voir cela, introduisons, pour  $k \geq 0$ ,

$$\mathcal{N}_k = \{j \in \mathcal{N} \mid d_x(s, j) = k\} ,$$

# Chapitre 9

## Optimisation sous contraintes d'inégalités matricielles, ou “SDP”

### 9.1 Introduction : le problème SDP

Rappelons que le problème de programmation linéaire standard consiste à maximiser une forme linéaire sur l'intersection du cône positif de  $\mathbb{R}^n$  avec un nombre fini d'hyperplans, soit

$$\text{Min}_{x \in K} c \cdot x; \quad Ax = b \tag{9.1}$$

avec

$$K = \mathbb{R}_+^n := \{x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \dots, n\}$$

et  $c \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ .

Certains algorithmes de résolution de programmes linéaires, comme les très efficaces méthodes de points intérieurs présentées dans le chapitre 10, s'étendent à d'autres cônes, tels que le cône des matrices symétriques positives, que nous examinons maintenant.

**Le cône des matrices symétriques positives,  $\mathcal{S}_n^+$**  Nous considérons le cas où l'inconnue  $x$  n'est pas un vecteur mais est une matrice symétrique. Notons  $\mathcal{S}_n$  l'espace vectoriel des matrices symétriques

$$\mathcal{S}_n := \{X \in \mathbb{R}^{n \times n} \mid X = X^\top\} .$$

Un cône qui intervient souvent est le cône  $\mathcal{S}_n^+ \subset \mathcal{S}_n$  formé des matrices symétriques positives de dimension  $n$ , soit

$$\mathcal{S}_n^+ = \{X \in \mathcal{S}_n \mid X \succcurlyeq 0\} .$$



Nous désignons ici par  $\succcurlyeq$  l'ordre usuel des matrices symétriques, parfois appelé *ordre de Loewner*, qui n'est autre que l'ordre des formes quadratiques associées :

$$X \preccurlyeq Y \iff u^\top X u \leq u^\top Y u, \forall u \in \mathbb{R}^n .$$

Le cône  $\mathcal{S}_n^+$  est fermé : en effet, pour tout  $u$ , l'ensemble des  $X \in \mathcal{S}_n$  telles que  $u^\top X u \geq 0$  est la préimage du fermé  $[0, +\infty[$  par l'application continue  $X \mapsto u^\top X u$ , et  $\mathcal{S}_n^+$  est l'intersection de ces préimages lorsque  $u$  parcourt  $\mathbb{R}^n$ .

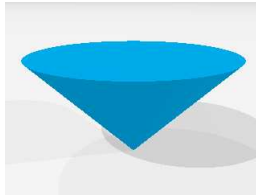
On peut aussi écrire la contrainte  $X \succcurlyeq 0$  sous la forme

$$\lambda_{\min}(X) \geq 0 ,$$

où  $\lambda_{\min}$  désigne la plus petite valeur propre d'une matrice symétrique. Appelons *cône de Lorentz* de dimension  $n$ , l'ensemble de  $\mathbb{R}^n$  défini par

$$x_n \geq \sqrt{x_1^2 + \dots + x_{n-1}^2} . \tag{9.2}$$

En dimension 3, le cône de Lorentz a la forme d'un "cornet de glace" :



Quand  $n = 2$ , on peut identifier l'ensemble des matrices symétriques positives au cône de Lorentz de dimension 3, en écrivant une matrice symétrique  $X$  comme combinaison linéaire des *matrices de Pauli*

$$X = tI + y_1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + y_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad t, y_1, y_2 \in \mathbb{R} , \tag{9.3}$$

et constater (Exercice) que la contrainte  $\lambda_{\min}(X) \geq 0$  est équivalente à

$$t \geq \sqrt{y_1^2 + y_2^2} .$$

Noter que pour  $n \geq 3$ , le cône  $\mathcal{S}_n^+$  ne coïncide plus avec un cône de Lorentz.

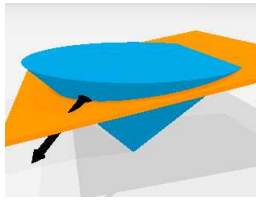
**Formulation primale du problème SDP** L'analogie du problème de programmation linéaire standard consiste à minimiser une forme linéaire sur l'intersection du cône  $\mathcal{S}_n^+$  avec un nombre fini d'hyperplans. Afin de l'expliciter, définissons, pour toutes matrices  $A$  et  $B$  de même taille, le produit scalaire suivant dit *de Frobenius*

$$\langle A, B \rangle = \text{trace}(AB^\top) = \sum_{i,j} A_{ij} B_{ij} .$$

Le problème linéaire sous contrainte de positivité matricielle, appelé problème SDP (pour “semi-definite program”<sup>1</sup>), s’écrit :

$$\text{Min}_{X \in \mathcal{S}_n^+} \langle C, X \rangle; \quad \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \quad (9.4)$$

pour des matrices  $C, A_1, \dots, A_m \in \mathcal{S}_n$  et des scalaires  $b_1, \dots, b_m \in \mathbb{R}$  donnés. Un exemple de problème SDP, pour  $n = 2$ , est illustré sur le dessin suivant : l’ensemble des points admissibles est l’ellipse obtenue comme intersection du cône  $\mathcal{S}_2^+$  avec un plan  $\langle A_1, X \rangle = b_1$ , l’opposé de la direction de  $C$  est représentée par une flèche, dont l’origine est le point de minimum de la fonction  $X \mapsto \langle C, X \rangle$  :



**Remarque 9.1 (SDP par blocs)** On peut aussi considérer un problème d’optimisation dont les variables sont plusieurs matrices positives  $X_1 \in \mathcal{S}_{n_1}^+, \dots, X_p \in \mathcal{S}_{n_p}^+$ , avec  $n_1, \dots, n_p \geq 1$ , soit

$$\text{Min}_{X_1 \in \mathcal{S}_{n_1}^+, \dots, X_p \in \mathcal{S}_{n_p}^+} \langle C_1, X_1 \rangle + \dots + \langle C_p, X_p \rangle; \quad \sum_{1 \leq j \leq p} \langle A_{ij}, X_j \rangle = b_i, \quad i = 1, \dots, m', \quad (9.5)$$

avec  $C_i \in \mathcal{S}_{n_i}$ ,  $A_{ij} \in \mathcal{S}_{n_j}$ , et  $b_i \in \mathbb{R}$ .

Le problème (9.5) se ramène à (9.4). Notons en effet  $X := \text{diag}(X_1, \dots, X_p)$  la matrice bloc diagonale formée des blocs  $X_1, \dots, X_p$ . La contrainte  $X \succcurlyeq 0$  est équivalente à  $X_i \succcurlyeq 0$ , pour  $1 \leq i \leq p$ . D’autre part, la contrainte “ $X$  bloc diagonale” peut s’écrire  $\langle X, E_{ij} \rangle = 0$ , si  $i$  et  $j$  appartiennent à des blocs différents, en notant  $E_{ij}$  la matrice avec un coefficient 1 en position  $(i, j)$ , et des 0 ailleurs. En tenant compte du caractère symétrique de  $X$ , on a  $\langle X, E_{ij} \rangle = 0$  si et seulement si  $\langle X, E_{ij} + E_{ji} \rangle = 0$ , cette dernière contrainte ayant la forme requise dans un SDP. En outre, les contraintes affines apparaissant dans (9.5) peuvent s’écrire comme des contraintes affines en la matrice  $X$ . Ceci permet d’écrire le problème (9.5) sous la forme (9.4).

Cet argument montre en particulier que le problème linéaire standard (9.1) est un cas particulier du problème SDP (9.4).

1. Mettons ici en garde le lecteur sur les pièges de traduction : en français, une matrice symétrique  $X$  est *positive* si  $\lambda_{\min}(X) \geq 0$  et *définie positive* si  $\lambda_{\min}(X) > 0$ , alors qu’en anglais,  $X$  est “positive” si  $\lambda_{\min}(X) > 0$  et “positive semi-definite” si  $\lambda_{\min}(X) \geq 0$ .

On retiendra de ce qui précède la règle suivante : plusieurs inégalités de type  $X_i \succcurlyeq 0$  peuvent se ramener à une seule inégalité de type  $X \succcurlyeq 0$  en considérant la matrice  $X$  bloc diagonale formée à partir des  $X_i$ .

Notons qu'en pratique, il est plus efficace de traiter des SDP "par blocs" (ce qui permet d'exploiter la structure creuse). Dans un souci de simplicité, nous limitons cependant l'exposé au problème standard (9.4). •

**Forme duale du problème SDP** Une autre formulation du problème SDP est la suivante :

$$\text{Max}_{y \in \mathbb{R}^m} b \cdot y; \quad y_1 A_1 + \dots + y_m A_m \preccurlyeq C \quad (9.6)$$

avec  $b \in \mathbb{R}^m$  et  $A_1, \dots, A_m, C \in \mathcal{S}_n$ . La contrainte

$$y_1 A_1 + \dots + y_p A_m \preccurlyeq C$$

satisfaite par tout point admissible  $y$  est qualifiée d'*inégalité linéaire matricielle*, ou "LMP".

Nous verrons dans la section 9.2 que (9.6) est le *problème dual* de (9.4), au sens de la dualité générale en optimisation convexe, décrite dans la section 2.5.4.

**Remarque 9.2** Malgré les apparences, le problème primal (9.4) et son dual (9.6) sont de même nature. En effet, notons  $\mathcal{E} = \{(k, l) \mid 1 \leq k \leq l \leq n\}$ , introduisons  $y = (y_{kl}) \in \mathbb{R}^{\mathcal{E}}$ , et posons  $F_{kk} = E_{kk}$  et  $F_{kl} = E_{kl} + E_{lk}$  pour  $k \neq l$ , (la matrice de base  $E_{kl}$  est définie comme dans la remarque 9.1) de sorte que toute matrice de  $\mathcal{S}_n$  peut s'écrire sous la forme

$$X = \sum_{(k,l) \in \mathcal{E}} y_{kl} F_{kl} .$$

Le problème (9.4) se réécrit comme suit :

$$\text{Min}_{y \in \mathbb{R}^{\mathcal{E}}} \sum_{k \leq l} \langle C, F_{kl} \rangle y_{kl}; \quad - \sum_{(k,l) \in \mathcal{E}} y_{kl} F_{kl} \preccurlyeq 0, \quad \sum_{(k,l) \in \mathcal{E}} y_{kl} \langle A_i, F_{kl} \rangle \leq \pm b_i, \quad i = 1, \dots, m.$$

En posant  $\mathcal{B} = (-\langle C, F_{kl} \rangle)_{(k,l) \in \mathcal{E}}$ ,  $\mathcal{D}_{kl} = \text{diag}(\langle A_1, F_{kl} \rangle, \dots, \langle A_m, F_{kl} \rangle)$ , ainsi que  $\mathcal{A}_{kl} = \text{diag}(-F_{kl}, \mathcal{D}_{kl}, -\mathcal{D}_{kl})$  et  $\mathcal{C} = \text{diag}(0, b_1, \dots, b_m, -b_1, \dots, -b_m)$ , on voit que (9.4) est équivalent à

$$\text{Max}_{y \in \mathbb{R}^{\mathcal{E}}} -\mathcal{B} \cdot y; \quad \sum_{(k,l) \in \mathcal{E}} y_{kl} \mathcal{A}_{kl} \preccurlyeq \mathcal{C} ,$$

qui est bien de la forme (9.6). En guise d'exercice, nous laissons le lecteur se convaincre, que, réciproquement, le problème (9.6) peut se ramener à un problème de la forme (9.4). •

**Exercice 9.3 (Complément de Schur)** Soit  $A \in \mathcal{S}_p$  une matrice définie positive, et

$$M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix},$$

avec  $C \in \mathcal{S}_q$  et  $B \in \mathbb{R}^{p \times q}$ . Montrer que

$$M \succ 0 \iff C - B^\top A^{-1} B \succ 0.$$

(La matrice  $C - B^\top A^{-1} B$  est appelée *complément de Schur*.)

Application. Soient  $F \in \mathbb{R}^{r \times n}$ ,  $g \in \mathbb{R}^n$  et  $c \in \mathbb{R}$ . Montrer que la contrainte convexe quadratique

$$x^\top F^\top F x \leq g^\top x + c \tag{9.7}$$

est équivalente à la contrainte ‘‘LMI’’ :

$$\begin{pmatrix} I & Fx \\ x^\top F^\top & g^\top x + c \end{pmatrix} \succ 0 \tag{9.8}$$

**Fonction barrière pour le problème SDP** Les méthodes de points intérieur pour le problème de programmation linéaire standard, qui font l’objet du chapitre 10, reposent sur l’emploi de la fonction strictement convexe, dite *potentiel logarithmique*,

$$\pi(x) = - \sum_{1 \leq i \leq n} \log x_i.$$

Les méthodes de points intérieurs, et notamment les résultats de complexité polynomiale, se généralisent au problème SDP (9.4), en prenant le potentiel :

$$\pi(X) = - \log \det X.$$

Pour minimiser une fonction convexe  $f(X)$  sous la contrainte  $X \succ 0$ , on considère, pour  $\mu > 0$ , la fonction

$$f_\mu(X) := f(X) + \mu \pi(X).$$

Comme  $\pi$  est strictement convexe sur l’ensemble des matrices symétriques définies positives, le point de minimum  $X_\mu$  de  $f_\mu$ , s’il existe est unique. L’idée est d’approcher  $X_\mu$  pour  $\mu > 0$  assez petit, ce qui fournit une approximation du point de minimum de  $f$  sur  $\mathcal{S}_n^+$ . Nous traiterons cette méthode en détail dans le chapitre 10, dans le cas plus simple où lequel le cône  $\mathcal{S}_n^+$  est remplacé par le cône standard  $\mathbb{R}_+^n$  de  $\mathbb{R}^n$ . Nous renvoyons le lecteur à [WSV00] pour le cas du cône  $\mathcal{S}_n^+$ , les preuves étant trop techniques pour être traitées ici. Limitons nous à vérifier la stricte convexité du potentiel  $-\log \det X$ .

**Proposition 9.4** Si  $X$  et  $Y$  sont des matrices définies positives, on a

$$\det \frac{1}{2}(X + Y) \geq \sqrt{\det(X) \det(Y)},$$

et l’égalité a lieu si et seulement si  $X = Y$ .

**Démonstration.** Comme une matrice symétrique diagonalise dans une base orthonormale, on peut écrire  $X = UDU^\top$  où  $U$  est une matrice orthogonale (i.e.  $UU^\top = U^\top U = I$ ) et  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$  est la matrice diagonale formée des valeurs propres de  $X$ . Posons  $\sqrt{D} := \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ . L'inégalité à prouver pour  $X$  et  $Y$  est équivalente à la même inégalité pour les matrices :

$$X' = \sqrt{D}^{-1}U^\top XU\sqrt{D}^{-1} = I \text{ et } Y' = \sqrt{D}^{-1}U^\top YU\sqrt{D}^{-1} .$$

On peut donc supposer que  $X = I$ . On peut écrire  $Y = VD'V^\top$  avec  $V$  orthogonale et  $D'$  diagonale. Quitte à considérer  $X' = V^\top XV = I$  et  $Y' = V^\top YV$  au lieu de  $X$  et  $Y$ , on peut supposer que  $X = I$  et que  $Y$  est diagonale, soit  $Y = \text{diag}(\mu_1, \dots, \mu_n)$ . Dans ce cas, l'inégalité demandée est conséquence de

$$\frac{1}{2}(1 + \mu_i) \geq \sqrt{\mu_i}$$

ce qui est vrai car la moyenne arithmétique majore la moyenne géométrique. Le cas d'égalité est laissé au lecteur. ■

**Corollaire 9.5** *La fonction  $X \mapsto -\log \det X$  est strictement convexe sur l'ensemble des matrices symétriques définies positives.* ■

## 9.2 Dualité SDP

### 9.2.1 Dualité faible

Associés au problème SDP sous forme primale (9.4) le lagrangien suivant :

$$L(X; Y, \lambda) = \langle C, X \rangle - \langle X, Y \rangle + \sum_{1 \leq i \leq m} \lambda_i (b_i - \langle A_i, X \rangle) ,$$

qui est fonction des variables  $X, Y \in \mathcal{S}_n$  et  $\lambda \in \mathbb{R}^m$ . Ce choix de lagrangien est justifié par la proposition :

**Proposition 9.6** *Le problème SDP sous forme primale (9.4) est équivalent au problème :*

$$\text{Min}_{X \in \mathcal{S}_n} \sup_{Y \in \mathcal{S}_n^+, \lambda \in \mathbb{R}^m} L(X; Y, \lambda) \tag{9.9}$$

Afin de démontrer cette proposition, rappelons la notion de cône polaire positif, introduite en section 2.5.5, noté dans ce chapitre par une étoile afin d'éviter la confusion avec l'ensemble des matrices positives :

$$K^* = \{y \in \mathbb{R}^k \mid \forall x \in K, x \cdot y \geq 0\} .$$

Le cône  $K = \mathbb{R}_+^n$  est un exemple de cône auto-polaire :  $(\mathbb{R}_+^n)^* = \mathbb{R}_+^n$ . Il en est de même du cône des matrices symétriques positives.

**Lemme 9.7** On a  $\mathcal{S}_n^+ = (\mathcal{S}_n^+)^*$ .

Afin d'établir ce lemme, rappelons deux résultats d'algèbre linéaire, que nous énonçons sous une forme permettant de les réexploiter dans la suite.

**Lemme 9.8** Une matrice  $X \in \mathcal{S}_n$  peut s'écrire

$$X = \sum_{1 \leq k \leq n} \lambda_k v^k (v^k)^\top,$$

où les  $\lambda^k$  sont les valeurs propres de  $X$ , et où les  $v^k$  forment une famille orthonormale, telle que  $v^k$  soit vecteur propre de  $X$  pour la valeur propre  $\lambda^k$ .

**Démonstration.** Le théorème spectral affirme qu'une matrice symétrique diagonalise par un changement de base orthonormal, ce qui s'écrit  $X = UDU^\top$  avec  $D = \text{diag}(\lambda_1, \dots, \lambda_k)$ , et  $UU^\top = U^\top U = I$ . En prenant pour  $v^k$  la  $k$ -ième colonne de  $U$ , on a la décomposition annoncée. ■

Si  $X \in \mathcal{S}_n$ , nous notons  $\text{Im } X = \{Xu \mid u \in \mathbb{R}^n\}$  et  $\text{Ker } X = \{u \mid Xu = 0\}$ .

**Lemme 9.9** Toute matrice  $X \in \mathcal{S}_n^+$  de rang  $r$  peut s'écrire :

$$X = \sum_{1 \leq k \leq r} v^k (v^k)^\top, \tag{9.10}$$

où  $v^1, \dots, v^r$  forment une famille orthogonale de vecteurs non-nuls de  $\mathbb{R}^n$ . D'autre part, si  $X$  est de la forme (9.10), pour une famille quelconque de vecteurs  $v^1, \dots, v^r \in \mathbb{R}^n$ , et si  $V$  désigne l'espace vectoriel engendré par  $v^1, \dots, v^r$ , on a  $\text{Im } X = V$ , et  $\text{Ker } X = V^\perp$ . En particulier,  $\text{rg}(X) = \dim V \leq r$ .

**Démonstration.** Soit  $X \in \mathcal{S}_n^+$ . D'après le lemme 9.8, on peut écrire

$$X = \sum_{1 \leq k \leq n} \lambda_k u^k (u^k)^\top,$$

avec  $\lambda_k$  valeur propre de  $X$  et  $u^k$  vecteur propre associé. On a  $\lambda_k \geq 0$  car  $X \succcurlyeq 0$ . Quitte à réordonner les valeurs propres, on peut supposer  $\lambda_1, \dots, \lambda_r > 0$  et  $\lambda_{r+1} = \dots = \lambda_n = 0$ . En posant  $v^k = \sqrt{\lambda_k} u^k$ , on a bien une représentation de  $X$  sous la forme (9.10), les vecteurs non-nuls  $v^1, \dots, v^r$  formant une famille orthogonale.

Supposons maintenant  $X$  de la forme (9.10), et soit  $V$  l'espace vectoriel engendré par  $v^1, \dots, v^r$ . Tout vecteur  $y$  de l'image de  $X$  est de la forme  $y = Xz$ , donc  $y = \sum_{1 \leq k \leq r} v^k (v^k \cdot z) \in V$ , donc  $\text{Im } X \subset V$ . Tout vecteur  $y$  du noyau de  $X$  vérifie  $0 = y^\top Xy = \sum_{1 \leq k \leq r} (v^k \cdot y)^2$ , donc  $\text{Ker } X \subset V^\perp$ . Notons  $s$  le rang de  $X$  et  $t$  la dimension de  $V$ . De  $\text{Im } X \subset V$ , on déduit  $s \leq t$ . De  $\text{Ker } X \subset V^\perp$ , on déduit  $n - s \leq n - t$ . Donc  $s = t$ ,  $\text{Im } X = V$ , et  $\text{Ker } X = V^\perp$ . ■

**Démonstration du lemme 9.7.** Soit  $X \in (\mathcal{S}_n^+)^*$ . Alors, pour tout  $u \in \mathbb{R}^n$ ,  $u^\top X u = \langle X, uu^\top \rangle \geq 0$ , donc  $X \in \mathcal{S}_n^+$ . Ainsi,  $(\mathcal{S}_n^+)^* \subset \mathcal{S}_n^+$ . Soit  $X \in \mathcal{S}_n^+$ . D'après le lemme 9.9, tout  $Y \in \mathcal{S}_n^+$  peut s'écrire comme somme d'un nombre fini de matrices de la forme  $u^\top u$ , avec  $u \in \mathbb{R}^n$ . Comme  $\langle X, u^\top u \rangle = u^\top X u \geq 0$ , on déduit que  $\langle X, Y \rangle \geq 0$  pour tout  $Y \in \mathcal{S}_n^+$ , et donc  $\mathcal{S}_n^+ \subset (\mathcal{S}_n^+)^*$ . ■

**Corollaire 9.10** Pour toute matrice  $X \in \mathcal{S}_n$ , on a

$$\inf_{Y \in \mathcal{S}_n^+} \langle X, Y \rangle = \begin{cases} 0 & \text{si } X \in \mathcal{S}_n^+ \\ -\infty & \text{sinon.} \end{cases} \quad (9.11)$$

**Démonstration.** Si  $X \in \mathcal{S}_n^+$ , il résulte du lemme 9.7 que  $\langle X, Y \rangle \geq 0$  pour tout  $Y \in \mathcal{S}_n^+$ . Si  $X \in \mathcal{S}_n \setminus \mathcal{S}_n^+$ , il résulte à nouveau du lemme 9.7 qu'il existe  $Z \in \mathcal{S}_n^+$  telle que  $\langle X, Z \rangle < 0$ . Donc

$$\inf_{Y \in \mathcal{S}_n^+} \langle X, Y \rangle \leq \inf_{t>0} \langle X, tZ \rangle = \lim_{t \rightarrow \infty} t \langle X, Z \rangle = -\infty .$$

**Démonstration de la Proposition 9.6.** Cela résulte aussitôt du corollaire 9.10. ■

Le problème dual de (9.4) s'obtient formellement en commutant le sup et l'inf dans (9.9), soit

$$\begin{aligned} \text{Max}_{Y \in \mathcal{S}_n^+, \lambda \in \mathbb{R}^m} \inf_{X \in \mathcal{S}_n} L(X; Y, \lambda) &= \text{Max}_{Y \in \mathcal{S}_n^+, \lambda \in \mathbb{R}^m} \inf_{X \in \mathcal{S}_n} \lambda \cdot b + \langle C - Y - \sum_{1 \leq i \leq m} \lambda_i A_i, X \rangle \\ &= \text{Max}_{Y \in \mathcal{S}_n^+, \lambda \in \mathbb{R}^m} \lambda \cdot b; \quad C - Y - \sum_{1 \leq i \leq m} \lambda_i A_i = 0 \\ &= \text{Max}_{\lambda \in \mathbb{R}^m} \lambda \cdot b; \quad C \succcurlyeq \sum_{1 \leq i \leq m} \lambda_i A_i . \end{aligned}$$

On reconnaît la forme (9.6) du problème SDP, ce qui justifie le nom de "formulation duale" pour (9.6). Comme le sup des inf est inférieur à l'inf des sups (lemme 2.41), on a :

**Proposition 9.11 (Dualité faible SDP).** La valeur du problème primal (9.4) SDP majore celle du problème dual (9.6) associé. ■

## 9.2.2 Dualité forte

En programmation linéaire, primal et dual ont toujours même valeur, sauf dans le cas dégénéré où les ensembles admissibles du primal et du dual sont tous deux vides. En outre, un programme linéaire dont la valeur est finie a toujours une solution optimale. Il n'en est pas de même pour les SDP.

**Exercice 9.12 (Absence de solution)** Considérons le SDP sous forme primale

$$\text{Min}_{X \in \mathcal{S}_2^+} X_{11} ; \quad X_{12} = 1 . \quad (9.12)$$

Montrer que la valeur de ce problème est 0, mais qu'il n'existe pas de solution optimale. Expliciter le problème dual : montrer que sa valeur est également 0 et qu'il admet une solution optimale.

**Exercice 9.13 (Saut de dualité)** Considérons le SDP sous forme primale

$$\text{Min}_{X \in \mathcal{S}_2^+} -X_{12} - X_{22} ; \quad X_{11} + X_{22} + 2X_{12} = 0 . \quad (9.13)$$

Montrer que la valeur ce problème est 0 et que le problème dual n'admet aucun point admissible, de sorte que la valeur du dual est  $-\infty$ .

Afin d'énoncer un résultat de dualité forte, écrivons la contrainte du SDP dual (9.6) sous la forme

$$\sum_{1 \leq i \leq m} y_i A_i + Z = C; \quad Z \in \mathcal{S}_n^+ . \quad (9.14)$$

**Théorème 9.14 (Dualité forte pour les SDP)** *Supposons l'existence de  $\bar{y} \in \mathbb{R}^m$  ainsi que d'une matrice définie positive  $\bar{Z}$  solution de (9.14). Alors la valeur du primal (9.4) et la valeur du dual (9.6) coïncident, et si le primal est réalisable, il admet une solution optimale.*

**Démonstration.** Soit  $d$  la valeur du dual. Comme le point  $\bar{y}$  est admissible pour le dual, on a  $d > -\infty$ . En raison de l'inégalité de dualité faible, le seul cas à considérer est celui où  $d$  est finie. Il faut montrer que le système suivant a une solution :

$$\text{Il existe } X \in \mathcal{S}_n^+; \quad \langle C, X \rangle = d; \quad \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m. \quad (9.15)$$

Pour  $X \in \mathcal{S}_n$ , notons  $\mathcal{B}X := \langle C, X \rangle, \langle A_1, X \rangle, \dots, \langle A_m, X \rangle$ , et posons  $g := (d, b_1, \dots, b_m)^\top$  et  $\mathcal{K} := \mathcal{B}\mathcal{S}_n^+$ . La relation (9.15) équivaut à  $g \in \mathcal{K}$ . Si ce n'est pas le cas, appliquons le lemme 9.15 énoncé ci-dessous aux matrices  $B_1 = A_1, \dots, B_m = A_m, B_{m+1} = C$ . L'existence de  $\bar{y}, \bar{Z}$  solution de (9.14) avec  $\bar{Z}$  défini positif entraîne que l'hypothèse du lemme 9.15 est vérifiée. Le cône  $\mathcal{K}$  est donc fermé. Il est évidemment convexe. Comme  $g \notin \mathcal{K}$ , on peut séparer strictement  $g$  de  $\mathcal{K}$ , et ainsi trouver  $\alpha \in \mathbb{R}$  et  $\beta \in \mathbb{R}^m$  tels que

$$\alpha d + \beta \cdot b < 0$$

et

$$\langle \alpha C + \beta_1 A_1 + \dots + \beta_m A_m, X \rangle \geq 0, \quad \forall X \in \mathcal{S}_n^+ .$$

En vertu de  $(\mathcal{S}_n^+)^* = \mathcal{S}_n$ , cette dernière condition se réécrit

$$\alpha C + \beta_1 A_1 + \dots + \beta_m A_m \succ 0 .$$



Quitte à diviser  $\beta$  par  $|\alpha|$  si  $\alpha \neq 0$ , on peut supposer que  $\alpha \in \{0, \pm 1\}$ .

Si  $\alpha = 0$ , on a  $b \cdot \beta < 0$ , et comme pour tout scalaire  $t \geq 0$ ,  $y^t := \bar{y} - t\beta$  est un point admissible du dual, avec  $b \cdot y^t \rightarrow +\infty$  quand  $t \rightarrow +\infty$ , la valeur du dual est  $d = +\infty$ , une contradiction.

Si  $\alpha = 1$ ,  $y := -\beta$  est un point admissible du dual, de valeur  $b \cdot y > d$ , ce qui contredit le fait que  $d$  est la valeur du dual.

Reste à considérer le cas où  $\alpha = -1$ , de sorte que  $\beta \cdot b < d$ . Choisissons un point admissible du dual  $y$  tel que  $\beta \cdot b < y \cdot b < d$ . Observons que pour tout  $t \geq 0$ ,  $y^t := \bar{y} + t(y - \beta)$  est encore un point admissible du dual. On a  $b \cdot (y - \beta) > 0$ , et donc  $b \cdot y^t \rightarrow +\infty$  quand  $t \rightarrow +\infty$ , ce qui montre que  $d = +\infty$ , une contradiction. La conclusion s'ensuit. ■

**Lemme 9.15** Soient  $B_1, \dots, B_p \in \mathcal{S}_n$ , et soit  $\mathcal{B}$  l'opérateur linéaire de  $\mathcal{S}_n$  dans  $\mathbb{R}^p$  associant à  $X$  le vecteur de coordonnées  $\langle B_i, X \rangle$ ,  $i = 1, \dots, p$ . S'il existe un vecteur  $z$  tel que la matrice  $\sum_{1 \leq i \leq p} z_i B_i$  soit définie positive, alors, l'image de  $\mathcal{S}_n^+$  par  $\mathcal{B}$  est fermée.

**Démonstration.** Soit  $X^k$  une suite de matrices de  $\mathcal{S}_n^+$  telle que  $\mathcal{B}X^k$  converge vers un vecteur  $b$ . On a  $z \cdot \mathcal{B}X^k = \sum_{1 \leq i \leq p} \langle z_i B_i, X^k \rangle = \langle F, X^k \rangle$ , où  $F := \sum_{1 \leq i \leq p} z_i B_i$  est définie positive. Notons  $\lambda_{kl}$ , pour  $l = 1, \dots, n$ , les valeurs propres de  $X^k$ , et soit  $(u^{kl})_{1 \leq k \leq n}$  la famille orthonormée de vecteurs propres correspondants, de sorte que, d'après le lemme 9.8,

$$X^k = \sum_{1 \leq l \leq n} \lambda_{kl} u^{kl} (u^{kl})^\top . \quad (9.16)$$

On a

$$\langle F, X^k \rangle = \sum_{1 \leq l \leq n} \lambda_{kl} (u^{kl})^\top F u^{kl} \geq \sum_{1 \leq l \leq n} \lambda_{kl} \lambda_{\min}(F) . \quad (9.17)$$

où  $\lambda_{\min}(F) > 0$  désigne la plus petite valeur propre de  $F$ , car  $u^\top F u \geq \lambda_{\min} \|u\|^2$ , pour tout vecteur  $u$ , et  $\|u^{kl}\| = 1$ . Comme  $X^k$  est positive, les  $\lambda_{kl}$  sont positifs. Comme  $\mathcal{B}X^k$  converge vers  $b$ ,  $\langle F, X^k \rangle = z \cdot \mathcal{B}X^k$  converge. On déduit de (9.17) que les valeurs propres  $\lambda_{kl}$  sont majorées. Les vecteurs  $u^{kl}$ , qui sont unitaires, sont bornés indépendamment de  $k$ . Il résulte de (9.16) que la suite  $X^k$  est bornée. Quitte à prendre une sous suite, on peut supposer que  $X^k$  converge vers une matrice  $X$ , qui appartient à  $\mathcal{S}_n^+$  car  $\mathcal{S}_n^+$  est fermé. Par continuité de  $\mathcal{B}$ , on a  $b = \mathcal{B}X$ , ce qui montre que  $\mathcal{B}(\mathcal{S}_n^+)$  est fermé. ■

**Remarque 9.16** Le théorème 9.14 est un cas particulier du corollaire 2.80. La formulation "primale" de ce dernier correspond à la formulation "duale" du théorème 9.14. On peut également appliquer le corollaire 2.81, qui sous une hypothèse primale assure l'absence de saut de dualité et l'existence de solutions duales. ●

**Exercice 9.17 (Réécriture basée sur le cône de Lorentz)** On va montrer comment certains problèmes non linéaires peuvent se réécrire comme des problèmes linéaires (avec critère et contraintes linéaires) et contraintes de cône de Lorentz.

1. *Contrainte de carré de norme.* Soient  $w \in \mathbb{R}^n$ ,  $\alpha$  et  $\beta$  scalaires. Montrer que

$$\{\alpha \geq 0, \beta \geq 0, \|w\|^2 \leq \alpha\beta\} \Leftrightarrow \alpha + \beta \geq \left\| \begin{pmatrix} 2w \\ \alpha - \beta \end{pmatrix} \right\|. \quad (9.18)$$

2. *Critères inverses.* Réécrire linéairement avec cône de Lorentz le problème

$$\begin{aligned} \text{Min}_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^p 1/(a_i \cdot x + b_i); \\ & a_i \cdot x + b_i > 0, \quad i = 1, \dots, p, \\ & c_i \cdot x + d_i \geq 0, \quad i = 1, \dots, q, \end{aligned} \quad (9.19)$$

avec les  $a_i, c_i$ , vecteurs de  $\mathbb{R}^n$ , et  $b, d$  vecteurs de  $\mathbb{R}^p$  et  $\mathbb{R}^q$ .

3. *Approximation uniforme en échelle logarithmique.* Réécrire linéairement avec cône de Lorentz le problème d'approximation uniforme, en échelle logarithmique :

$$\text{Min}_{x \in \mathbb{R}^n} \max |\log(a_i \cdot x) - \log(b_i)| \quad (9.20)$$

avec  $b_i > 0$  pour tout  $i$ , et la contrainte implicite  $a_i \cdot x > 0$  pour tout  $i$ .

4. *Puissances fractionnaires I.* Soient  $\ell$  un entier strictement positif. Réécrire linéairement avec cône de Lorentz la relation

$$x \in \mathbb{R}_+^{2\ell}; \quad t \in \mathbb{R}; \quad t \leq (x_1 x_2 \cdots x_{2\ell})^{1/2\ell}. \quad (9.21)$$

5. *Puissances fractionnaires II.* Réécrire linéairement avec cône de Lorentz la relation

$$x \in \mathbb{R}_+^n; \quad t \in \mathbb{R}_+; \quad t \leq x_1^{\pi_1} x_2^{\pi_2} \cdots x_n^{\pi_n}. \quad (9.22)$$

On suppose les exposant de la forme  $\pi_i = p_i/p$ , avec  $p_i$  entier strictement positifs et  $p$  entier,  $p \geq \sum_i p_i$ .

**Exercice 9.18 Maximisation de forme quadratique sous contrainte de boîte).** Soit  $A$  une matrice symétrique réelle, et soit

$$B_n = \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$$

l'hypercube unité. On s'intéresse au problème d'optimisation :

$$\text{Max } x^\top A x; \quad x \in B_n.$$

Montrer que la valeur de ce problème est majorée par la valeur du problème SDP

$$\text{Min} \sum_{1 \leq i \leq n} \mu_i; \quad \mu \in (\mathbb{R}_+)^n, \quad A - \text{diag}(\mu) \preceq 0.$$

## 9.3 Quelques problèmes combinatoires où interviennent des SDP

### 9.3.1 Problème de la coupe maximale

Reprenons le problème de la coupe maximale, décrit dans l'exemple 1.22. Étant donné un graphe non-orienté  $(\mathcal{V}, \mathcal{E})$ , chaque arête  $\{i, j\} \in \mathcal{E}$  étant munie d'une capacité  $c_{ij}$ , on cherche une bi-partition non-triviale :

$$\mathcal{V} = I^- \cup I^+, \quad I^- \cap I^+ = \emptyset, \quad I^- \neq \emptyset, \quad I^+ \neq \emptyset \quad (9.23)$$

maximisant

$$c(I^-, I^+) := \sum_{\{i,j\} \in \mathcal{E}, i \in I^-, j \in I^+} c_{ij} .$$

On dit que l'ensemble des arêtes reliant  $I^-$  et  $I^+$  est une *coupe*, et que  $c(I^-, I^+)$  est la capacité de la coupe.

**Remarque 9.19** Le cas particulier où  $c_{ij} \leq 0$ , pour tout  $\{i, j\} \in \mathcal{E}$ , est facile : on se ramène à *minimiser* la capacité d'une coupe dans un graphe dont les arêtes sont munies de capacités  $-c_{ij} \geq 0$ . On retrouve ainsi le problème de la coupe minimale (globale), qui, comme nous l'avons vu dans l'exercice 5.2, peut être résolu notamment au moyen d'algorithmes de flots. Nous renvoyons à nouveau à [CGK<sup>+</sup>96] pour un état de l'art. •

Nous nous intéressons ici au problème général de la coupe maximale (qui est NP-difficile). Pour l'aborder, associons à toute partition (9.23) le vecteur de signe  $x \in \{\pm 1\}^{\mathcal{E}}$ ,

$$x_i = \pm 1 \text{ si } i \in I^\pm .$$

Posons  $c_{ij} = 0$  si  $\{i, j\} \notin \mathcal{E}$ . La capacité de la coupe  $(I^-, I^+)$  s'écrit :

$$c(I^-, I^+) = \sum_{i < j} c_{ij} \frac{1 - x_i x_j}{2} = \frac{1}{4} \sum_{i,j} c_{ij} (1 - x_i x_j) \quad (9.24)$$

$$= \frac{1}{4} \left( \sum_{ij} c_{ij} x_i x_i - \sum_{ij} c_{ij} x_i x_j \right) = \frac{1}{4} x^\top L x \quad (9.25)$$

où  $L$  désigne la matrice telle que

$$L_{ii} = \sum_j c_{ij}, \quad L_{ij} = -c_{ij} .$$

La matrice  $-L$  est appelée *Laplacien* du graphe  $\mathcal{G}$  (relativement aux capacités  $c_{ij}$ ). Nous venons de montrer :

**Proposition 9.20** *Le problème de la coupe maximale est équivalent à :*

$$\max_{x \in \{\pm 1\}^V} x^\top Lx ,$$

où  $-L$  est le Laplacien associé au graphe. ■

Ceci suggère d'étudier plus généralement, pour toute matrice  $C \in \mathcal{S}^n$ , le problème :

$$\max_{x \in \{\pm 1\}^n} x^\top Cx . \tag{9.26}$$

La proposition suivante formule la généralisation (9.26) du problème de la coupe maximale comme un problème SDP, sous contrainte de rang. En oubliant la contrainte de rang, on obtient un majorant calculable par une méthode de points intérieurs :

**Proposition 9.21 Relaxation SDP du problème de la coupe maximale).**

*Toute matrice  $C \in \mathcal{S}_n$  vérifie*

$$\max_{x \in \{\pm 1\}^n} x^\top Cx = \max_{X \in \mathcal{S}_n^+} \langle C, X \rangle \quad X_{ii} = 1 \text{ pour tout } i, \text{ rg}(X) = 1 . \tag{9.27}$$

*En particulier, on a la majoration :*

$$\max_{x \in \{\pm 1\}^n} x^\top Cx \leq \sup_{X \in \mathcal{S}_n^+} \langle C, X \rangle \quad X_{ii} = 1 \text{ pour tout } i. \tag{9.28}$$

Le problème à droite de (9.28) est appelé *relaxation SDP* du problème (9.26). Afin d'établir cette proposition, énonçons le cas particulier du lemme 9.9 :

**Corollaire 9.22** *Les matrices de  $\mathcal{S}_n^+$  qui sont de rang 1 sont précisément les matrices de la forme  $xx^\top$ , où  $x$  est un vecteur non-nul de  $\mathbb{R}^n$ . ■*

**Démonstration de la Proposition 9.21.** Associons à tout  $x \in \{\pm 1\}^n$  la matrice positive de rang 1 :

$$X = xx^\top .$$

Observons que  $X_{ii} = 1$ , pour tout  $i$ . Comme

$$x^\top Cx = \langle C, X \rangle ,$$

on a aussitôt

$$\max_{x \in \{\pm 1\}^n} x^\top Cx \leq \max_{X \in \mathcal{S}_n^+} \langle C, X \rangle \quad X_{ii} = 1, \forall i, \text{ rg}(X) = 1 .$$

Réciproquement, si  $X$  est positive de rang 1, d'après le Corollaire 9.22, on peut écrire  $X = xx^\top$ , avec  $x \in \mathbb{R}^n$ . Si  $X_{ii} = 1$ , on a  $x_i^2 = 1$ , et donc  $x_i = \pm 1$ , ce qui démontre l'égalité (9.27). L'inégalité (9.28) en découle aussitôt. ■



On appelle *capacité de Lovasz* la quantité :

$$\vartheta(\mathcal{G}) := \inf_{A \in \mathcal{S}_n} \left\{ \lambda_{\max}(A); \quad A_{ii} = 1, \forall i \in \mathcal{V}, \quad A_{ij} = 1, \forall \{i, j\} \in \bar{\mathcal{E}} \right\}, \quad (9.30)$$

où  $\lambda_{\max}(A)$  désigne la plus grande valeur propre d'une matrice symétrique  $A$ . Le problème (9.30) est bien de type SDP. En introduisant une variable supplémentaire scalaire,  $t$ , on a en effet :

$$\vartheta(\mathcal{G}) = \inf_{A \in \mathcal{S}_n, t \in \mathbb{R}} \left\{ t; \quad tI - A \succcurlyeq 0, \quad A_{ii} = 1, \forall i \in \mathcal{V}, \quad A_{ij} = 1, \forall \{i, j\} \in \bar{\mathcal{E}} \right\}, \quad (9.31)$$

où  $I$  désigne la matrice identité. Ce problème est de la forme (9.6), la collection des variables  $y_k$  étant ici constituée des variables  $A_{ij}$ , pour  $\{i, j\} \in \mathcal{E}$ , ainsi que de la variable scalaire  $t$ .

**Théorème 9.25** *Pour tout graphe non-orienté  $\mathcal{G}$ , on a*

$$\alpha(\mathcal{G}) \leq \vartheta(\mathcal{G}). \quad (9.32)$$

Afin de démontrer ce résultat, faisons l'observation élémentaire :

**Lemme 9.26** *Si  $A \in \mathcal{S}_p$  est une sous-matrice principale d'une matrice  $B \in \mathcal{S}_n$ , on a*

$$\lambda_{\max}(A) \leq \lambda_{\max}(B).$$

**Démonstration.** Pour toute matrice symétrique  $C$ , on a

$$\lambda_{\max}(C) = \sup_{u \neq 0} \frac{u^\top C u}{\|u\|^2}.$$

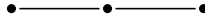
On peut supposer que la matrice  $A$  est formée des  $p$  premières lignes et colonnes de  $B$ . On a

$$\begin{aligned} \lambda_{\max}(A) &= \sup_{v \in \mathbb{R}^p \setminus \{0\}} \frac{v^\top A v}{\|v\|^2} = \sup_{v \in \mathbb{R}^p \setminus \{0\}} \frac{\begin{pmatrix} v \\ 0 \end{pmatrix}^\top C \begin{pmatrix} v \\ 0 \end{pmatrix}}{\|v\|^2} \\ &\leq \sup_{u \in \mathbb{R}^n \setminus \{0\}} \frac{u^\top C u}{\|u\|^2} = \lambda_{\max}(C). \end{aligned}$$

■

**Démonstration du Théorème 9.25.** Si  $\mathcal{S}$  est un ensemble indépendant, alors, pour toute matrice  $A$  vérifiant les contraintes (9.30), la sous-matrice  $B$  de  $A$  obtenue en se restreignant aux lignes et colonnes de  $\mathcal{S}$  a toutes ses composantes égales à 1. Les valeurs propres de  $B$  sont  $|\mathcal{S}|$  (de multiplicité 1), et 0 (de multiplicité  $|\mathcal{S}| - 1$ ). Donc  $\lambda_{\max}(B) = \max(0, |\mathcal{S}|) = |\mathcal{S}|$ . D'après le lemme 9.26,  $\lambda_{\max}(B) \leq \lambda_{\max}(A)$ , et donc  $|\mathcal{S}| \leq \lambda_{\max}(A)$ . Il en résulte que  $\alpha(\mathcal{G}) \leq \vartheta(\mathcal{G})$ . ■

**Exemple 9.27** Soit  $\mathcal{G}$  la chaîne à trois sommets :



Sa capacité de Lovasz est donnée par :

$$\vartheta(\mathcal{G}) = \inf_{x_{12}, x_{23} \in \mathbb{R}} \lambda_{\max}(A(x_{12}, x_{23})) \quad \text{avec } A(x_{12}, x_{23}) := \begin{pmatrix} 1 & x_{12} & 1 \\ x_{12} & 1 & x_{23} \\ 1 & x_{23} & 1 \end{pmatrix}$$

Cette chaîne admet un ensemble indépendant de cardinal 2 (formé des extrémités de la chaîne). D'après le théorème 9.25, on a  $\vartheta(\mathcal{G}) \geq 2$ . En prenant  $x_{12} = x_{23} = 0$ , on obtient une matrice  $A(x)$  telle que  $\lambda_{\max}(A(x)) = 2$ . Donc  $\vartheta(\mathcal{G}) = 2$ .

### 9.3.3 Nombre chromatique d'un graphe

Soit  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  un graphe non-orienté. Le *nombre chromatique* de  $\mathcal{G}$ , noté  $\chi(\mathcal{G})$ , est le nombre minimal de couleurs nécessaires pour colorier tous les sommets de  $\mathcal{G}$ , de telle sorte que deux sommets adjacents n'aient pas la même couleur.

Le problème de calcul du nombre chromatique peut être vu comme une version très purifiée du problème d'*allocation de fréquence* : une couleur représente une fréquence, les sommets représentent des émetteurs, la contrainte de coloriage exprime que deux émetteurs voisins ont des fréquences différentes (en pratique on demande plutôt que des émetteurs voisins aient des fréquences suffisamment éloignées).

Pour colorier une clique, il faut autant de couleurs que la clique a de sommets, et donc, pour tout graphe  $\mathcal{G}$ ,

$$\omega(\mathcal{G}) \leq \chi(\mathcal{G}) . \tag{9.33}$$

Nous allons améliorer cette minoration de  $\chi(\mathcal{G})$ .

**Théorème 9.28** *Pour tout graphe non-orienté  $\mathcal{G}$ , on a*

$$\vartheta(\mathcal{G}) \leq \chi(\bar{\mathcal{G}}) . \tag{9.34}$$

Avant de démontrer ce théorème, énonçons le corollaire obtenu à l'aide du théorème 9.25 et de l'identité (9.29).

**Corollaire 9.29 (du sandwich)** *Pour tout graphe non-orienté  $\mathcal{G}$ , on a*

$$\omega(\mathcal{G}) \leq \vartheta(\bar{\mathcal{G}}) \leq \chi(\mathcal{G}) . \tag{9.35}$$

L'intérêt de ces inégalités est que  $\vartheta(\bar{\mathcal{G}})$  est facilement calculable, alors que  $\omega(\mathcal{G})$  et  $\chi(\mathcal{G})$  ne le sont probablement pas (calculer  $\omega(\mathcal{G})$  et calculer  $\chi(\mathcal{G})$  sont des problèmes NP-durs).

Afin de montrer le théorème 9.28, donnons une formulation duale de la capacité  $\vartheta$  de Lovasz.

**Proposition 9.30** *On a*

$$\vartheta(\mathcal{G}) = \max_{X \in \mathcal{S}_n^+} \left\{ \sum_{i,j \in \mathcal{V}} X_{ij}; \quad X_{ij} = 0, \text{ pour tout } i, j \in \mathcal{E}, \quad \sum_{i \in \mathcal{V}} X_{ii} = 1 \right\} \quad (9.36)$$

**Démonstration.** Soit  $E_{ij}$  la matrice de base définie dans la remarque 9.1. Nous posons  $F_{ij} = E_{ij} + E_{ji}$  pour  $j \neq i$ . En faisant apparaître les degrés de liberté  $y_{kl} := X_{kl}$ , avec  $(k, l) \in \mathcal{E}$ , dans la formulation (9.31), on obtient :

$$-\vartheta(\mathcal{G}) = \max_{y \in \mathbb{R}^{\mathcal{E}}, t \in \mathbb{R}} \left\{ -t; \quad \sum_{(i,j) \in \bar{\mathcal{E}}} F_{ij} + \sum_{i \in \mathcal{V}} E_{ii} + \sum_{(i,j) \in \mathcal{E}} y_{ij} F_{ij} - tI \preceq 0 \right\}. \quad (9.37)$$

Ce problème est le problème dual du problème de minimisation ci-dessous :

$$\text{Min}_{X \in \mathcal{S}_n^+} - \sum_{(i,j) \in \bar{\mathcal{E}}} X_{ij} - \sum_{i \in \mathcal{V}} X_{ii}; \quad X_{ij} = 0 \text{ pour tout } (i, j) \in \mathcal{E}, \quad \sum_{i \in \mathcal{V}} X_{ii} = 1. \quad (9.38)$$

Prenons un  $y \in \mathbb{R}^{\mathcal{E}}$  quelconque. Pour  $t$  assez grand,  $(y, t)$  est un point admissible de (9.37) tel que la variable d'écart  $Z$  solution de

$$\sum_{(i,j) \in \mathcal{E}} y_{ij} F_{ij} - tI + Z = - \sum_{(i,j) \in \bar{\mathcal{E}}} F_{ij},$$

soit définie positive. En raison du théorème de dualité forte 9.14, les problèmes (9.37) et (9.38) ont même valeur. L'égalité (9.36) en résulte. ■

**Démonstration du théorème 9.28.** Notons  $n$  le nombre de sommets. Soit  $X$  un point admissible de la formulation (9.36). D'après le Lemme 9.9, on peut écrire  $X = VV^T$ , avec  $V \in \mathbb{R}^{n \times r}$ , et en notant  $z^i$  la ligne  $i$  de  $V$ , il vient  $X_{ij} = z^i \cdot z^j$ . La valeur du critère vaut donc

$$\sum_{ij} X_{ij} = \|z^1 + \dots + z^n\|^2. \quad (9.39)$$

Considérons maintenant une coloration de  $\bar{\mathcal{G}}$  avec  $k$  couleurs, c'est-à-dire, une application  $\gamma$  de  $\mathcal{V}$  dans  $\{1, \dots, k\}$  telle que

$$(i, j) \in \bar{\mathcal{E}} \implies \gamma(i) \neq \gamma(j).$$

Si les sommets  $i$  et  $j$  sont de la même couleur (i.e.,  $\gamma(i) = \gamma(j)$ ), alors  $i$  et  $j$  ne peuvent être reliés par une arête de  $\bar{\mathcal{E}}$ , donc  $(i, j) \in \mathcal{E}$ , donc  $z^i \cdot z^j = X_{ij} = 0$ . On a montré :

$$\gamma(i) = \gamma(j) \implies z^i \cdot z^j = 0. \quad (9.40)$$

D'autre part,

$$\sum_{i \in \mathcal{V}} X_{ii} = \sum_{i \in \mathcal{V}} \|z^i\|^2 = 1.$$



Pour toute couleur  $c \in \{1, \dots, k\}$ , considérons la somme des vecteurs associés à des sommets de couleur  $c$  :

$$w^c = \sum_{i \in \mathcal{V}, \gamma(i)=c} z^i .$$

En utilisant (9.40), on obtient

$$\sum_{1 \leq c \leq k} \|w^c\|^2 = \sum_{i \in \mathcal{V}} \|z^i\|^2 = 1 .$$

Comme  $\sum_{1 \leq c \leq k} w^c = \sum_{i \in \mathcal{V}} z^i$ , en utilisant (9.39) et l'inégalité triangulaire, on obtient :

$$\sum_{ij} X_{ij} = \left\| \sum_{1 \leq c \leq k} w^c \right\|^2 \leq \left( \sum_{1 \leq c \leq k} \|w^c\| \right)^2 .$$

Soit  $t_c := \|w^c\|$ . On a  $\sum_{1 \leq c \leq k} t_c^2 = 1$ , et  $\sum_{ij} X_{ij} \leq (\sum_{1 \leq c \leq k} t_c)^2$ . Un simple exercice d'optimisation montre que

$$\sqrt{k} = \max \left\{ \sum_{1 \leq c \leq k} t_c ; \quad \sum_{1 \leq c \leq k} t_c^2 = 1 \right\}$$

(en termes géométriques, la boule  $\ell_2$  de rayon 1 est incluse dans la boule  $\ell_1$  de rayon  $\sqrt{k}$ ). On a donc  $\sum_{ij} X_{ij} \leq k$ . Ceci étant vrai pour tout point admissible  $X$  de la formulation (9.36), on a  $\vartheta(\mathcal{G}) \leq k$ . Cette inégalité étant vraie pour toute coloration  $\gamma$  de  $\bar{\mathcal{G}}$  en  $k$  couleurs, on a  $\vartheta(\mathcal{G}) \leq \chi(\bar{\mathcal{G}})$ . ■

**Exercice 9.31 (Coupe maximale)** On considère la généralisation du problème de la coupe maximale :

$$\max_{x \in \{\pm 1\}^n} x^\top C x , \tag{9.41}$$

avec  $C = (C_{ij})$ , et sa relaxation SDP

$$\max_{X \in \mathcal{S}_n^+, X_{ii}=1, \forall i} \langle C, X \rangle , \tag{9.42}$$

1. Montrer que cette relaxation SDP peut s'écrire :

$$\text{Max} \sum_{ij} C_{ij} v^i \cdot v^j ; \quad \|v^i\|^2 = 1, \forall i, v^1, \dots, v^n \in \mathbb{R}^n .$$

Si  $v$  désigne un point admissible de ce problème, une manière de fabriquer une coupe est de prendre un vecteur  $h$  aléatoire, tiré selon une distribution Gaussienne standard, et de poser :

$$x_i = \text{sgn } v^i \cdot h .$$

La coupe associée est formée des arcs reliant des sommets de signe opposé. La valeur de cette coupe est

$$H := \sum_{i,j} C_{ij} (\text{sgn } v^i \cdot h) (\text{sgn } v^j \cdot h) .$$

2. Montrer que

$$\Pr(\text{sgn}(v^i \cdot h) \neq \text{sgn}(v^j \cdot h)) = \frac{1}{\pi} \arccos(v^i \cdot v^j) .$$

(On admettra que la projection d'un vecteur aléatoire Gaussien standard sur un plan est encore un vecteur aléatoire Gaussien standard).

3. En déduire que l'espérance de  $H$  vaut :

$$E(H) = \frac{2}{\pi} \sum_{i,j} C_{ij} \arcsin(v^i \cdot v^j) .$$

4. On définit  $\arcsin(X) := (\arcsin(X_{ij}))$ . Montrer que (9.41) équivaut à

$$\max_{X \in \mathcal{S}_n^+, X_{ii}=1, \forall i} \frac{2}{\pi} \langle C, \arcsin X \rangle , \tag{9.43}$$

5. Montrer que si  $A$  et  $B$  sont symétriques positives, alors, le produit coefficient par coefficient de  $A$  est  $B$  est encore une matrice positive. En notant que

$$\arcsin(X) = X + \frac{1}{2} \frac{X^{\odot 3}}{3} + \frac{1 \times 3}{2 \times 4} \frac{X^{\odot 5}}{5} + \dots$$

où  $X^{\odot k} := (X_{ij}^k)$ , déduire de la question précédente que si  $C \in \mathcal{S}_n^+$ , la valeur de (9.41) est minorée par  $2/\pi$  fois la valeur de (9.42). Ce corollaire est dû à Nesterov.

## 9.4 Optimisation polynomiale et problèmes de moments

Cette section étudie les problèmes d'optimisation à données polynomiales. On montre comment relaxer le problème par des contraintes de positivité des matrices de moments. Cette relaxation s'interprète comme un certain type de décomposition du critère en somme pondérée de carrés de polynômes.

### 9.4.1 Optimisation polynomiale

Appliquons maintenant l'optimisation SDP à la résolution de problèmes de calcul du minimum global de problèmes de la forme

$$\text{Min}_{x \in \mathbb{R}^n} f(x); \quad g_i(x) \geq 0, \quad i = 1, \dots, r, \tag{P}$$