Dualité Lagrangienne

Guillaume Lecué[1]

Le but de cette section est de présenter une approche permettant de donner des informations supplémentaires sur les coefficients de Lagrange $\lambda_1, \ldots, \lambda_r, \mu_1, \ldots, \mu_l$ apparaissant dans le théorème de KKT grâce à des résultats de dualité.

L'idée de l'approche de relaxation Lagrangienne est de donner une minoration du problème initial (qu'on appellera problème primal) par un autre problème plus facile à résoudre. Tout l'enjeu est alors de montrer que cette minoration est exacte sous certaines conditions.

Pour fixer les idées, on donne maintenant la minoration du problème primal car elle fait apparaître la fonction de Lagrange dont l'étude est au centre de cette section. On considère alors le problème d'optimisation sous contrainte dans sa forme générale

$$\min_{x \in K} f(x)$$

où $f : U \to \mathbb{R}$, U est un ouvert de $\mathbb{R}^n$ et K est une contrainte de la forme

$$K = \left\{ x \in U : \begin{array}{c} g_1(x) = \cdots = g_r(x) = 0 \\ h_1(x) \leq 0, \ldots, h_l(x) \leq 0 \end{array} \right\}$$

pour $g_1, \ldots, g_r : U \to \mathbb{R}$ définissant les contraintes d'égalité et $h_1, \ldots, h_l : U \to \mathbb{R}$ définissant les contraintes d'inégalité. On note $G : x \in U \to (g_i(x))_{i=1}^r \in \mathbb{R}^r$ et $H : x \in U \to (h_j(x))_{j=1}^l \in \mathbb{R}^l$. Ainsi, la contrainte est $\{x \in U : G(x) = 0, H(x) \leq 0\}$ (où on utilise la définition qu'un vecteur v est plus petit qu'un vecteur u quand $v_i \leq u_i$ pour tout i).

On voit que pour tout $x \in K$ et tout $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ on a

$$f(x) \geq f(x) + \left\langle \lambda, G(x) \right\rangle + \left\langle \mu, H(x) \right\rangle := \mathcal{L}(x, (\lambda, \mu)).$$

En particulier, on a (pour tout $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$)

$$\min_{x \in K} f(x) \geq \min_{x \in U} \mathcal{L}(x, (\lambda, \mu)) := \psi(\lambda, \mu)$$

(on minimise bien $\mathcal{L}(\cdot, (\lambda, \mu))$ sur tout U alors que f n'est minimisé que sur K – on souhaite en effet retrouver le problème primal à gauche et un problème non contraint (donc sensé être plus facile à résoudre) à droite). Ce dernier résultat étant vrai pour tout $(\lambda, \mu)) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$, on peut prendre le maximum à droite :

$$\min_{x \in K} f(x) \geq \sup_{(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l} \psi(\lambda, \mu). \tag{0.1}$$

La fonction $\mathcal{L}$ est appelée **fonction de Lagrange** ou **Lagrangien** associé au problème primal. La fonction ψ est appelée **fonction duale** et le problème qui cherche à la maximiser est appelée

1. CREST, ENSAE. Bureau 3029, 5 avenue Henry Le Chatelier. 91 120 Palaiseau. Email : guillaume.lecue@ensae.fr.

le **problème duale**. L'objectif de cette section est d'identifier des situations où le problème dual permet de résoudre le problème primal.

L'analyse précédente montre que le problème dual peut toujours être utilisé pour minorer le problème primal (et ceci sans aucune hypothèse). Les situations qui nous intéressent sont celles pour lesquelles cette minoration est exacte, c'est-à-dire quand

$$\min_{x\in K} f(x) = \sup_{(\lambda,\mu)\in\mathbb{R}^r\times(\mathbb{R}_+)^l} \psi(\lambda,\mu).$$

On voit que cette égalité est en fait équivalente à avoir

$$\min_{x\in U} \sup_{(\lambda,\mu)\in\mathbb{R}^r\times(\mathbb{R}_+)^l} \mathcal{L}(x,(\lambda,\mu)) = \sup_{(\lambda,\mu)\in\mathbb{R}^r\times(\mathbb{R}_+)^l} \min_{x\in U} \mathcal{L}(x,(\lambda,\mu)).$$

C'est donc un problème d'intervertion de "min" et "sup" de la fonction de Lagrange qu'on va chercher à résoudre. Dans les situations (qu'on cherchera à identifier) pour lesquelles cette intervertion est exacte, on dira qu'on a **dualité forte** – l'inégalité (0.1), toujours vraie, étant appelée dualité faible. On s'intéresse donc aux propriétés de dualité forte de $\mathcal{L}$ et aux problèmes de mini-maximisation de $\mathcal{L}$. La section suivante étudie ces propriétés dans un cadre générale.

1 Problèmes de mini-maximisation et points selles

Dans cette section, on donne des résultats généraux sur les problèmes de mini-maximisation et leurs solutions pour des fonctions $\mathcal{L}$ à deux variables quelconque. On les appliquera par la suite à la fonction de Lagrange en cas particulier.

Définition 1.1 *Soient X et Y deux espaces. Soit $\mathcal{L}: X\times Y\to\mathbb{R}$ et $(x^*,y^*)\in X\times Y$. On dit que (x^*,y^*) est un **point-selle de $\mathcal{L}$ sur** $X\times Y$ quand pour tout $(x,y)\in X\times Y$, on a*

$$\mathcal{L}(x^*,y) \le \mathcal{L}(x^*,y^*) \le \mathcal{L}(x,y^*).$$

On peut par exemple voir que $(0,0)$ est un point-selle de $\mathcal{L}:(x,y)\to x^2-y^2+2$ car pour tout $(x,y)\in\mathbb{R}\times\mathbb{R}$, on a $\mathcal{L}(0,y)\le\mathcal{L}(0,0)\le\mathcal{L}(x,0)$.

Déterminer les points-selle de la fonction de Lagrange associée à un problème d'optimisation sous contraintes nous sera utile pour résoudre ce problème d'optimisation sous contraintes.

Dans cette section on va s'intéresser aux problèmes de minimaximisation de $\mathcal{L}$, càd aux problèmes de la forme

$$\sup_{y\in Y}\inf_{x\in X}\mathcal{L}(x,y) \text{ ou } \inf_{x\in X}\sup_{y\in Y}\mathcal{L}(x,y).$$

On va identifier des conditions assurant que les deux problèmes sont en fait identiques.

Définition 1.2 *Étant donnés X et Y deux espaces et $\mathcal{L}: X\times Y\to\mathbb{R}$. On appelle **fonction primale** la fonction $\varphi: x\in X\to\sup_{y\in Y}\mathcal{L}(x,y)$ (à valeurs $+\infty$ éventuellement). Le problème d'optimisation $\inf_{x\in X}\varphi(x)$ est appelé **problème primal**. De même, la **fonction duale** est définie par $\psi: y\in Y\to\inf_{x\in X}\mathcal{L}(x,y)$ (à valeurs $-\infty$ éventuellement). Le problème d'optimisation $\sup_{y\in Y}\psi(y)$ est appelé **problème dual**. Dans la fonction, $(x,y)\in X\times Y\to\mathcal{L}(x,y)$, la variable x est appelée **variable primale** et la variable y est appelée **variable duale**.*

Proposition 1.3 (dualité faible) *Étant donnés X et Y deux espaces et $\mathcal{L}: X\times Y\to\mathbb{R}$. On note φ la fonction primale et ψ la fonction duale. On a $\sup_{y\in Y}\psi(y)\le\inf_{x\in X}\varphi(x)$.*

Preuve. On a pour tout $x \in X$ et $y \in Y$, $\mathcal{L}(x,y) \le \sup_{y\in Y} \mathcal{L}(x,y)$ et donc $\inf_{x\in X} \mathcal{L}(x,y) \le \inf_{x\in X}\sup_{y\in Y} \mathcal{L}(x,y)$. Ceci étant vrai pour tout y, on obtient

$$\sup_{y\in Y}\inf_{x\in X} \mathcal{L}(x,y) \le \inf_{x\in X}\sup_{y\in Y} \mathcal{L}(x,y)$$

ce qui est, par définition de ψ et φ, équivalent à $\sup_{y\in Y}\psi(y) \le \inf_{x\in X}\varphi(x)$. ■

La Proposition (1.3) dit qu'on a toujours de la dualité faible :

$$\psi^* := \sup_{y\in Y}\inf_{x\in X} \mathcal{L}(x,y) \le \inf_{x\in X}\sup_{y\in Y} \mathcal{L}(x,y) =: \varphi^* \tag{1.1}$$

Autrement dit $\varphi^* - \psi^* \ge 0$ et la quantité $\varphi^* - \psi^*$ est appelée **duality gap**. On va identifier des situations où l'inégalité inverse a aussi lieu dans (1.1), càd quand le duality gap est nul.

Définition 1.4 *Le **duality gap** est $\varphi^* - \psi^* \ge 0$. Quand le duality gap est nul (càd quand $\varphi^* = \psi^*$), on dit qu'on a **dualité forte**.*

Proposition 1.5 *Étant donnés X et Y deux ensembles, $\mathcal{L} : X \times Y \to \mathbb{R}$ et $(x^*, y^*) \in X \times Y$. Il y a équivalence entre les points suivants :*
1) (x^*, y^*) est un point-selle de $\mathcal{L}$ sur $X \times Y$
2) $\varphi(x^*) \le \psi(y^*)$
3) $\varphi(x^*) = \psi(y^*)$
4) $\varphi(x^*) = \psi(y^*) = \mathcal{L}(x^*, y^*)$

Preuve. 1) $\Rightarrow$ 2) Comme (x^*, y^*) est un point-selle de $\mathcal{L}$, on a pour tout $(x,y) \in X \times Y$, $\mathcal{L}(x^*, y) \le \mathcal{L}(x^*, y^*) \le \mathcal{L}(x, y^*)$. Ainsi en passant au sup sur $y \in Y$ à gauche et à l'inf sur $x \in X$ à droite, on obtient

$$\varphi(x^*) = \sup_{y\in Y} \mathcal{L}(x^*, y) \le \mathcal{L}(x^*, y^*) \le \inf_{x\in X} \mathcal{L}(x, y^*) = \psi(y^*).$$

2) $\Rightarrow$ 3) Par définition de φ, on a pour tout $x \in X$, $\varphi(x) \ge \mathcal{L}(x,y)$ pour tout $y \in Y$. En particulier, pour y^*, on a $\varphi(x) \ge \mathcal{L}(x, y^*)$. Alors en passant à l'inf sur $x \in X$, on obtient

$$\varphi(x^*) \ge \inf_{x\in X} \varphi(x) \ge \inf_{x\in X} \mathcal{L}(x, y^*) = \psi(y^*).$$

Ainsi si on suppose que 2) est vraie, on a bien 3).

3) $\Rightarrow$ 4) Par définition de φ et ψ, on a toujours

$$\varphi(x^*) = \sup_{y\in Y} \mathcal{L}(x^*, y) \ge \mathcal{L}(x^*, y^*) \ge \inf_{x\in X} \mathcal{L}(x, y^*) = \psi(y^*).$$

Alors si on suppose que $\varphi(x^*) = \psi(y^*)$, on en déduit bien qu'il n'y a que des égalités dans la suite d'égalité ci-dessus et donc $\varphi(x^*) = \mathcal{L}(x^*, y^*) = \psi(y^*)$.

4) $\Rightarrow$ 1) Par définition de φ et ψ, on a pour tout $x \in X$ et $y \in Y$,

$$\mathcal{L}(x^*, y) \le \varphi(x^*) = \mathcal{L}(x^*, y^*) = \psi(y^*) \le \mathcal{L}(x, y^*).$$

Donc (x^*, y^*) est bien un point selle de $\mathcal{L}$ sur $X \times Y$. ■

Proposition 1.6 *Étant donnés X et Y deux ensembles, $\mathcal{L} : X \times Y \to \mathbb{R}$ et $(x^*, y^*) \in X \times Y$. Il y a équivalence entre les points suivants :*

1) (x^*, y^*) est un point-selle de $\mathcal{L}$ sur $X \times Y$

5) x^* atteint l'infimum de φ sur X (càd x^* est solution du problème primal), y^* atteint le supremum de ψ sur Y (càd y^* est solution du problème dual) et $\varphi(x^*) = \psi(y^*) = \mathcal{L}(x^*, y^*)$.

Preuve. $\underline{1) \Rightarrow 5)}$ D'après la Proposition 1.5, on sait que 1) implique 4) donc

$$\inf_{x \in X} \varphi(x) \leq \varphi(x^*) = \mathcal{L}(x^*, y^*) = \psi(y^*) \leq \sup_{y \in Y} \psi(y). \tag{1.2}$$

Par ailleurs, la Proposition 1.3 dit qu'on a toujours de la dualité faible càd

$$\sup_{y \in Y} \psi(y) \leq \inf_{x \in X} \varphi(x).$$

Ainsi, la dualité faible implique la suite d'égalités suivante dans (1.2)

$$\inf_{x \in X} \varphi(x) = \varphi(x^*) = \mathcal{L}(x^*, y^*) = \psi(y^*) = \sup_{y \in Y} \psi(y).$$

En particulier, $\inf_{x \in X} \varphi(x)$ est atteint en x^*, $\sup_{y \in Y} \psi(y)$ est atteint en y^* et $\varphi(x^*) = \mathcal{L}(x^*, y^*) = \psi(y^*)$. Donc 5) est vrai.

$\underline{5) \Rightarrow 1)}$ Comme 5) est plus fort que 4) dans la Proposition 1.5 et que 4) implique 1), on a bien le résultat. ■

Proposition 1.7 *Étant donnés X et Y deux espaces, $\mathcal{L} : X \times Y \to \mathbb{R}$ et $(x^*, y^*) \in X \times Y$. Il y a équivalence entre les points suivants :*

1) (x^*, y^*) est un point-selle de $\mathcal{L}$ sur $X \times Y$

6) x^* atteint l'infimum de φ sur X (càd x^* est solution du problème primal), y^* atteint le supremum de ψ sur Y (càd y^* est solution du problème dual) et le duality gap est nul.

Preuve. D'après la Proposition 1.3, on a toujours

$$\sup_{y \in Y} \psi(y) \leq \inf_{x \in X} \varphi(x).$$

Ainsi, dire que le duality gap est nul est équivalent à dire que

$$\inf_{x \in X} \varphi(x) \leq \sup_{y \in Y} \psi(y). \tag{1.3}$$

$\underline{1) \Rightarrow 6)}$ D'après la Proposition 1.6, 1) implique 5), il reste donc seulement à montrer (1.3). Par ailleurs, d'après 5), $\inf_{x \in X} \varphi(x)$ est atteint par x^* et $\sup_{y \in Y} \psi(y)$ est atteint par y^*. On a donc

$$\inf_{x \in X} \varphi(x) = \varphi(x^*) = \psi(y^*) = \sup_{y \in Y} \psi(y).$$

Donc (1.3) est satisfaite.

$\underline{6) \Rightarrow 1)}$ D'après 6), on a

$$\varphi(x^*) = \inf_{x \in X} \varphi(x) = \sup_{y \in Y} \psi(y) = \psi(y^*).$$

Donc 3) est vrai et donc d'après la Proposition 1.5, 1) est vrai. ■

La Proposition 1.7 permet donc d'assurer la dualité forte (càd le duality gap nul) quand il y a un point-selle. La réciproque est aussi vraie si de plus le problème primal et le problème dual ont une solution. Dans ce dernier cas, le couple formé par ces solutions est un point-selle.

2 Fonction de Lagrange

Dans cette section, on introduit une fonction associée à un problème d'optimisation sous contrainte, c'est la fonction de Lagrange qu'on a déjà vu en début de chapitre. Il sera intéressant de déterminer les points-selles de cette fonction pour résoudre le problème d'optimisation initial.

On considère un problème d'optimisation sous contrainte dans sa forme générale

$$\min_{x\in K} f(x) \tag{2.1}$$

où $f : U \to \mathbb{R}$, U est un ouvert de $\mathbb{R}^n$ et K est une contrainte de la forme

$$K = \left\{ x \in U : \begin{array}{c} g_1(x) = \cdots = g_r(x) = 0 \\ h_1(x) \leq 0, \dots, h_l(x) \leq 0 \end{array} \right\} \tag{2.2}$$

pour $g_1, \dots, g_r : U \to \mathbb{R}$ définissant les contraintes d'égalité et $h_1, \dots, h_l : U \to \mathbb{R}$ définissant les contraintes d'inégalité.

Définition 2.1 *La* ***fonction de Lagrange*** *ou* ***Lagrangien*** *associé au problème d'optimisation sous contrainte* (2.1) *où K est défini dans* (4.1) *est définie par*

$$\mathcal{L} : \left\{ \begin{array}{ccc} U \times (\mathbb{R}^r \times (\mathbb{R}_+)^l) & \to & \mathbb{R} \\ (x, (\lambda, \mu)) & \to & f(x) + \sum_{i=1}^r \lambda_i g_i(x) + \sum_{j=1}^l \mu_j h_j(x). \end{array} \right.$$

Les variables $\lambda \in \mathbb{R}^r$ et $\mu \in (\mathbb{R}_+)^l$ sont appelées les ***multiplicateurs de Lagrange****.*

Dans la suite, on étudie les points-selles de $\mathcal{L}$. On montrera, en particulier, les liens entre points-selles de $\mathcal{L}$ et les solutions au problème (2.1). On commence par identifier les fonctions primales et duales associées à $\mathcal{L}$.

La fonction primale associée à la fonction de Lagrange est donnée pour tout $x \in U$ par

$$\varphi(x) = \sup_{(\lambda,\mu)\in\mathbb{R}^r\times(\mathbb{R}_+)^l} \mathcal{L}(x, (\lambda, \mu)) = \left\{ \begin{array}{ll} f(x) & \text{if } x \in K \\ +\infty & \text{otherwise.} \end{array} \right.$$

La fonction primale φ de $\mathcal{L}$ coïncide donc avec f sur K et vaut $+\infty$ en dehors de K. En particulier, minimiser φ sur U est équivalent à minimiser f sur K et donc *le problème primal associé à la fonction de Lagrange est exactement le problème* (2.1) qu'on souhaite résoudre.

Pour la fonction duale associée à $\mathcal{L}$, on rappelle sa définition :

$$\psi(\lambda, \mu) = \inf_{x\in U} \mathcal{L}(x, (\lambda, \mu)) \tag{2.3}$$

et le problème dual associé est

$$\sup_{(\lambda,\mu)\in\mathbb{R}^r\times(\mathbb{R}_+)^l} \psi(\lambda, \mu). \tag{2.4}$$

Le problème dual est un problème d'optimisation sous-contrainte mais il a deux avantages : 1) la contrainte n'est faite que de l contraintes d'inégalité qu'on peut écrire de manière abrégée par "$\mu \geq 0$" ; c'est une contrainte "facile" car il est facile de projeter dessus (on reviendra sur ce point plus tard lorsqu'on parlera de l'algorithme de Uzawa) 2) la fonction duale est facile à maximiser car elle est concave.

Proposition 2.2 *La fonction duale ψ définie dans* (2.3) *est concave.*

Preuve. Pour tout $x \in U$, la fonction $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l \to \mathcal{L}(x, (\lambda, \mu))$ est une fonction affine. Ainsi, ψ est l'enveloppe inférieure d'une famille de fonction affine, elle est donc concave. En effet, pour tout $x \in U$, (λ_0, μ_0), (λ_1, μ_1) et $\alpha \in [0, 1]$, on a

$$\mathcal{L}(x, \alpha(\lambda_0, \mu_0) + (1 - \alpha)(\lambda_1, \mu_1)) = \alpha\mathcal{L}(x, (\lambda_0, \mu_0)) + (1 - \alpha)\mathcal{L}(x, (\lambda_1, \mu_1)).$$

On prend ensuite l'inf en $x \in U$ des deux côtés et comme l'inf d'une somme et plus grand que la somme des infs on conclut. ■

L'idée centrale de la dualité Lagrangienne est que le problème dual (2.4) va nous aider à résoudre le problème primal qui est ici exactement le problème d'optimisation (2.1) qu'on souhaite résoudre. Le lien entre les solutions du problème dual et les solutions du problème primal se fait grâce aux points-selles de la fonction de Lagrange quand le duality gap est nul. Identifier les situations où le duality gap de la fonction de Lagrange est nul est donc une question centrale en dualité Lagrangienne.

3 Points-selles et duality gap de la fonction de Lagrange

Dans cette section, on fait le lien entre points-selles de la fonction de Lagrange et solution du problème (2.1) et le lien entre solution du problème dual et solution du problème (2.1) sous l'hypothèse que le duality gap est nul.

Théorème 3.1 *Soit $(x^*, (\lambda^*, \mu^*)) \in U \times (\mathbb{R}^r \times (\mathbb{R}_+)^l)$. Il y a équivalence entre :*

1) $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$ ($\mathcal{L}$ étant la fonction de Lagrange de la Définition 2.1),

7) $(x^*, (\lambda^*, \mu^*))$ vérifie :

i) x^* minimise $x \in U \to \mathcal{L}(\cdot, (\lambda^*, \mu^*))$ sur U,

ii) $x^* \in K$,

iii) $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$.

Preuve. 1) $\Rightarrow$ 7) On suppose que $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$ alors pour tout $x \in U$ et $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$, on a

$$\mathcal{L}(x^*, (\lambda, \mu)) \overset{(a)}{\leq} \mathcal{L}(x^*, (\lambda^*, \mu^*)) \overset{(b)}{\leq} \mathcal{L}(x, (\lambda^*, \mu^*)).$$

On déduit de *(b)* que x^* minimise $x \in U \to \mathcal{L}(\cdot, (\lambda^*, \mu^*))$ sur U. On déduit de *(a)* que

$$\sup_{(\lambda,\mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l} \mathcal{L}(x^*, (\lambda, \mu)) \leq \mathcal{L}(x^*, (\lambda^*, \mu^*)) < +\infty$$

et donc $x^* \in K$, sinon on aurait $\sup_{(\lambda,\mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l} \mathcal{L}(x^*, (\lambda, \mu)) = +\infty$. Il ne reste plus qu'à démontrer la complementary condition : "$\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$."

Comme $x^* \in K$, on peut réécrire *(a)* comme : pour tout $\mu \geq 0$,

$$f(x^*) + \sum_{j=1}^{l} \mu_j h_j(x^*) \leq f(x^*) + \sum_{j=1}^{l} \mu_j^* h_j(x^*).$$

Comme $h_j(x^*) \leq 0$ et $\mu^* \geq 0$, on a

$$0 = \sup_{\mu \geq 0} \sum_{j=1}^{l} \mu_j h_j(x^*) \leq \sum_{j=1}^{l} \mu_j^* h_j(x^*) \leq 0$$

et donc $\sum_{j=1}^{l} \mu_j^* h_j(x^*) = 0$, mais comme chaque terme de cette somme est négatif (car $\mu_j^* \geq 0$ et $h_j(x^*) \leq 0$ vu que $x^* \in K$), chaque terme doit être nul càd $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, r$.

7) $\Rightarrow$ 1) Par hypothèse, x^* minimise $x \in U \to \mathcal{L}(\cdot, (\lambda^*, \mu^*))$ sur U donc pour tout $x \in U$,

$$\mathcal{L}(x^*, (\lambda^*, \mu^*)) \leq \mathcal{L}(x, (\lambda^*, \mu^*)).$$

Il reste donc à montrer que pour tout $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$, on a

$$\mathcal{L}(x^*, (\lambda, \mu)) \leq \mathcal{L}(x^*, (\lambda^*, \mu^*)).$$

Comme $x^* \in K$, on a $g_i(x^*) = 0$ pour tout $i = 1, \ldots, r$ et comme $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$, on a

$$\mathcal{L}(x^*, (\lambda^*, \mu^*)) = f(x^*) + \sum_{i=1}^{r} \lambda_i^* g_i(x^*) + \sum_{j=1}^{l} \mu_j^* h_j(x^*) = f(x^*).$$

Par ailleurs, si $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ on a $\lambda_i g_i(x^*) = 0$ car $x^* \in K$ et $\mu_j h_j(x^*) \leq 0$ car $\mu_j \geq 0$ et $h_j(x^*) \leq 0$ donc

$$\mathcal{L}(x^*, (\lambda, \mu)) = f(x^*) + \sum_{i=1}^{r} \lambda_i g_i(x^*) + \sum_{j=1}^{r} \mu_j h_j(x^*) \leq f(x^*) = \mathcal{L}(x^*, (\lambda^*, \mu^*)).$$

On obtient donc bien que $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$. ∎

On fait finalement le lien entre solutions du problème dual et solution du problème primal sous l'hypothèse que le duality gap de la fonction de Lagrange est nul.

Théorème 3.2 *On suppose que le duality gap de la fonction de Lagrange est nul. Si (λ^*, μ^*) est une solution du problème dual (2.4) alors il y a équivalence entre :*

8) x^* est solution du problème (2.1)

7) $(x^*, (\lambda^*, \mu^*))$ vérifie :

i) x^* minimise $x \in U \to \mathcal{L}(\cdot, (\lambda^*, \mu^*))$ sur U,

ii) $x^* \in K$,

iii) $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$.

Preuve. D'après le Théorème 3.1, il y a équivalence entre *1)* et *7)*. De plus, d'après la Proposition 1.7, il y a équivalence entre *1)* et *6)* : "x^* atteint l'infimum de φ sur X, y^* atteint le supremum de ψ sur Y et le duality gap est nul" qui, appliqué au cas particulier de la fonction de Lagrange, s'énonce sous la forme : "x^* est solution du problème (2.1), (λ^*, μ^*) est solution du problème dual et le duality gap est nul". Ainsi, si on suppose que (λ^*, μ^*) est solution du problème dual et le duality gap est nul, on a bien équivalence entre 7) et 8). ∎

En conséquence, identifier les situations où le duality gap de la fonction de Lagrange est nulle permet de résoudre le problème initial d'optimisation à partir du problème dual. En effet, dans ce cas (càd quand le duality gap est nul), on peut d'abord résoudre le problème dual (qui a des contraintes "faciles" et une fonction objectif concave à maximiser) : on obtient une solution (λ^*, μ^*). A l'aide d'une solution (λ^*, μ^*), on peut identifier toutes les solutions du problème primal (c'est le problème qui nous intéresse), en résolvant le problème d'optimisation sans contrainte :"minimiser $x \in U \to \mathcal{L}(x, (\lambda^*, \mu^*))$ sur U" et en ne gardant que le ou les solution(s) x^* de ce problème pour laquelle ou lesquelles on a $x^* \in K$ et $\mu_j^* h_j(x^*) = 0, \forall j = 1, \ldots, l$.

4 Qualification de contrainte et dualité Lagrangienne

Dans la section précédente, on a vu que sous hypothèse que le duality gap est nul, on peut retrouver toutes les solution du problème primal grâce au problème dual et en solutionnant un problème d'optimisation sans contrainte. Néanmoins, vérifier que le duality gap est nul peut être difficile (voir cependant les théorèmes généraux à ce sujet en Section 8).

On peut simplifier cette approche en supposant la qualification de la contrainte. On peut en effet faire un lien entre la qualification de contrainte et l'existence d'un point-selle de la fonction de Lagrange (et donc d'un duality gap nul). On peut par exemple, montrer le théorème suivant grâce au théorème de KKT.

Théorème 4.1 (Théorème du point-selle) *Soit $x^* \in U$. On a :*

1) *S'il existe $(\lambda^*, \mu^*) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ tel que $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$ alors x^* est solution de (2.1).*
2) *On suppose que U est un ouvert convexe et que les fonctions définissant les contraintes sont telles que $g_1, \ldots, g_r$ sont affines et $h_1, \ldots, h_l$ sont convexes. Si x^* est solution de (2.1) et que K est qualifiée en x^* alors il existe $(\lambda^*, \mu^*) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ tel que $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$.*

Preuve. 1) $\Rightarrow$ 2) On sait d'après la Proposition 1.6 que si $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$ alors x^* est solution du problème primal, donc du problèmes (2.1) dans le cas de la fonction de Lagrange (et aussi que (λ^*, μ^*) est solution du problème dual et le duality gap est nul).

2) $\Rightarrow$ 1) D'après le Théorème de KKT, comme K est qualifiée en x^* et que x^* est solution de (2.1), on sait qu'il existe $(\lambda^*, \mu^*) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ tel que $\nabla_x \mathcal{L}(x^*, (\lambda^*, \mu^*)) = 0$ et $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$. Or sous les hypothèses que les g_i sont affines et les h_j sont convexes la fonction $x \in U \to \mathcal{L}(x, (\lambda^*, \mu^*)$ est convexe et différentiable. Alors ses minima sur U sont exactement ses points critiques. Or x^* est un point critique de cette fonction (par KKT), donc x^* minimise $x \in U \to \mathcal{L}(x, (\lambda^*, \mu^*)$ sur U. Par ailleurs, $x^* \in K$ et, par KKT, on a $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$ donc la propriété 7) du Théorème 3.2 est satisfaite et donc par le Théorème 3.2, $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$. ■

Le théorème du point-selle ne s'applique réellement bien (càd, sous hypothèse de qualification, on a une CNS "x^* est solution de (2.1)" si et seulement si "il existe $(\lambda^*, \mu^*) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ tel que $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$") que pour les problèmes d'**optimisation convexe différentiable (OCD)** qui sont des problèmes d'optimisation sous contrainte de la forme $\min_{x \in K} f(x)$ où $f : U \to \mathbb{R}$, U est un **ouvert convexe** de $\mathbb{R}^n$ et K est une contrainte de la forme

$$K = \left\{ x \in U : \begin{array}{c} g_1(x) = \cdots = g_r(x) = 0 \\ h_1(x) \le 0, \ldots, h_l(x) \le 0 \end{array} \right\} \tag{4.1}$$

où les $g_1, \ldots, g_r : U \to \mathbb{R}$, définissant les contraintes d'égalité, sont des **fonctions affines** et les $h_1, \ldots, h_l : U \to \mathbb{R}$, définissant les contraintes d'inégalité, sont des fonctions **convexes différentiables**.

Pour les problèmes d'OCD, la contrainte de qualification la plus utilisée est la contrainte de Slater : on dit que K satisfait la condition de Slater quand il existe $x_0 \in K$ tel que pour tout $j = 1, \ldots, l, h_j(x_0) < 0$. On a vu que cette condition implique que la contrainte K est qualifiée (en tout point de K).

Remarque 4.2 *On sait, d'après la Proposition 1.3, que si $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$ alors nécessairement (λ^*, μ^*) est solution du problème dual. On peut alors préciser dans le théorème du point-selle dans les deux points 1) et 2) que (λ^*, μ^*) est une solution du problème dual.*

5 Conditions KKT et dualité Lagrangienne

On considère de nouveau le problème d'optimisation différentiable sous contraintes (2.1). Quand le duality gap est nul, on peut faire le lien entre l'optimalité pour les deux problèmes primal et dual et les conditions KKT. On rappelle ici les conditions KKT.

Définition 5.1 *Soit $f, (g_i)_{i=1,\dots,r}, (h_j)_{j=1,\dots,l} : U \mapsto \mathbb{R}$ des fonctions différentiables sur un ouvert U de $\mathbb{R}^n$. On dit que $(x^*, (\lambda^*, \mu^*))$ satisfait les* ***conditions KKT*** *quand, pour $K = \{x \in U : g_i(x) = 0, \forall i = 1, \dots, r, h_j(x) \leq 0, \forall j = 1, \dots, l\}$,*

$$x^* \in K \text{ et } \mu^* \geq 0 \tag{KKT1}$$
$$\mu_j^* h_j(x^*) = 0, \forall j = 1, \dots, l \tag{KKT2}$$
$$\nabla_x \mathcal{L}(x^*, (\lambda^*, \mu^*)) = 0 \tag{KKT3}$$

La condition (KKT2) est appelée ***complementary slackness condition****.*

La première condition (KKT1) dit simplement que les éléments x^* et (λ^*, μ^*) sont dans leur ensemble de contraintes respectifs (pour le primal et le dual). La condition (KKT2) signifie que soit $h_j(x^*) = 0$, càd la j-ième contrainte du problème primal est saturée, soit la j-iéme contrainte du problème dual est saturée, càd $\mu_j^* = 0$. Finalement, la troisième condition (KKT3) signifie que x^* est un point critique de $\mathcal{L}(\cdot, \lambda^*, \mu^*)$.

Théorème 5.2 *On considère le problème d'optimisation sous contraintes (2.1) où $f, (g_i)_{i=1}^r, (h_j)_{j=1}^l$ sont différentiables sur U. On a les résultats suivant :*

1. *si le duality gap de $\mathcal{L}$ est nul, si x^* est une solution primale et si (λ^*, μ^*) est une solution duale, alors $(x^*, (\lambda^*, \mu^*))$ satisfait les conditions KKT.*
2. *si on suppose que $(g_i)_{i=1}^r$ sont affines et que f et $(h_j)_{j=1}^l$ sont convexes alors les conditions KKT sont suffisantes : si $(x^*, (\lambda^*, \mu^*))$ satisfait KKT alors x^* est solution du problème primal, (λ^*, μ^*) est solution du problème duale et le duality gap est nul.*

Preuve. . On montre *1.*. C'est une conséquence directe de la Proposition 1.7 et du Théorème 3.1. On a équivalence entre

6) x^* minimise φ sur U (ce qui équivaut à dire que x^* est solution primale), (λ^*, μ^*) maximise ψ sur $\mathbb{R}^r \times (\mathbb{R}_+)^l$ (ce qui équivaut à dire que (λ^*, μ^*) est solution duale), le duality gap est nul,

1) $(x^*, (\lambda^*, \mu^*))$ est un point-selle de L,

7) i) x^* minimise $\mathcal{L}(\cdot, (\lambda^*, \mu^*))$ sur U
   ii) $x^* \in K$
   iii) $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \dots, r$.

Comme *1.* est une réécriture de *6)*, on a *1.* implique *7)*. Dans le cas différentiable, *7)* implique les conditions KKT vu que si x^* minimise $\mathcal{L}(\cdot, (\lambda^*, \mu^*))$ sur U alors nécessairement x^* est un point critique de $\mathcal{L}(\cdot, (\lambda^*, \mu^*))$ càd $\nabla_x \mathcal{L}(x^*, (\lambda^*, \mu^*)) = 0$. Par ailleurs, d'après *7)*, $x^* \in K$, $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \dots, r$ et par définition de (λ^*, μ^*) dans *1.*, on a $\mu_j^* \geq 0$. Donc $(x^*, (\lambda^*, \mu^*))$ satisfait KKT.

On démontre le deuxième point *2.* sous l'hypothèse supplémentaire que f et $(h_j)_j$ sont convexes et $(g_i)_i$ sont affines. Soit $(x^*, (\lambda^*, \mu^*))$ satisfaisant les conditions KKT. On a x^* et (λ^*, μ^*) sont dans les ensembles de contraintes du problème primal et dual respectivement. Par ailleurs, comme

$\mu^* \geq 0$, $\mathcal{L}(\cdot, \lambda^*, \mu^*) = f(\cdot) + \sum_{i=1}^r \lambda_i^* g_i(\cdot) + \sum_{j=1}^l \mu_j^* h_j(\cdot)$ est convexe et comme $\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0$, x^* est un minimum de $\mathcal{L}(\cdot, \lambda^*, \mu^*)$ sur U (en optimisation convexe différentiable les points critiques sont les extrema). Ainsi la condition *i)* de la condition 7) est satisfaite. Par ailleurs, la condition (KKT1) assure que $x^* \in K$ donc *ii)* est aussi satisfaite. Finalement, la condition (KKT3) implique la condition *iii)* de la condition 7). Donc 7) est satisfaite et comme elle est équivalente à 6), on obtient bien le résultat. ■

Dans les deux conclusions suivantes, on combine les résultats du Théorème 4.1 et du Théorème 5.2.

Conclusion 1 : Pour les problèmes d'optimisation convexe différentiables (OCD) (où les contraintes d'égalités $(g_i)_{i=1}^r$ sont affines et $f, (h_j)_{j=1}^l$ sont convexes et différentiables) les conditions KKT sont nécessaires et suffisantes : il y a équivalence entre

1. *le duality gap est nul, x^* est solution primal et (λ^*, μ^*) est solution duale*
2. *$(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT.*

Dans ce cas, pour trouver une solution au problème primale, il suffit de trouver une solution duale (λ^*, μ^*) et de trouver un point x^* tel que $(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT. La réciproque n'est vraie que si on sait de plus que le duality gap est nul.

Conclusion 2 : Pour un problème d'optimisation convexe différentiables (OCD) (où les contraintes d'égalités $(g_i)_{i=1}^r$ sont affines et $f, (h_j)_{j=1}^l$ sont convexes et différentiables) ayant une contrainte qualifiée (par exemple en vérifiant la condition de Slater ou pour des contraintes affines), il y a équivalence entre

1. *x^* est solution primal et (λ^*, μ^*) est solution duale*
2. *$(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT.*

Dans ce cas, pour trouver toutes les solutions au problème primale, il suffit de trouver une solution duale (λ^*, μ^*) et de déterminer l'ensemble des points x^* tels que (x^*, λ^*, μ^*) vérifie les conditions KKT.

On voit que dans la première conclusion, on n'a pas besoin de qualifier la contrainte : il suffit de résoudre le problème duale (λ^*, μ^*) et de chercher un x^* tel que $(x^*, (\lambda^*, \mu^*))$ vérifie KKT. Dans ce cas, on sait que x^* est une solution du problème primal. Néanmoins, si on veut avoir toutes les solution du problème primal, on a besoin d'une réciproque à cette approche ; c'est la deuxième conclusion qui permet de le faire. Mais dans ce cas, on a besoin de l'hypothèse de qualification.

6 Méthodologie et exemple d'application de la dualité Lagrangienne

Les deux conclusions de la section précédente sont propices à une méthodologie d'application de la dualité Lagrangienne. Cette méthodologie est aussi présentée dans le chapitres sur l'OCD car c'est dans ce cadre que la dualité Lagrangienne s'applique le mieux (c'est aussi dans ce cadre qu'ont été énoncées les deux conclusions précédentes). On donne maintenant la méthodologie d'application de la deuxième conclusion.

Méthodologie d'application de la dualité Lagrangienne en OCD. Sous hypothèse de qualification de la contrainte (par Slater ou contraintes affines par exemple), on peut appliquer la méthode suivante pour trouver toutes les solutions d'un problème d'OCD :

0) On vérifie que les fonctions $f, h_1, \dots, h_l$ sont bien convexes et différentiables.
2) On vérifie que la contrainte est qualifiée (Slater ou contraintes linéaires en général)
3) On résout le problème dual : on trouve les solutions (λ^*, μ^*) du problème dual
4) on cherche les $x^* \in U$ qui minimise $x \to \mathcal{L}(x, (\lambda^*, \mu^*))$ sur U tels que $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \dots, l$ et ceci pour toutes les solutions du problème dual. On note par $E_2 \subset K$ cet ensemble.
5) Pa dualité Lagrangienne (ici la conclusion 2 de la section précédente), l'ensemble des solutions au problème $\min_{x \in K} f(x)$ est donné exactement par E_2.

On peut aussi énoncer une méthodologie d'application pour la première conclusion. Dans ce cas, on n'a pas besoin de vérifier la qualification mais on n'a seulement un ou des solutions au problème dual sans avoir de certitude de les avoir toutes :

Méthodologie d'application de la dualité Lagrangienne en OCD. Sans hypothèse de qualification de la contrainte, on peut appliquer la méthode suivante pour trouver une ou des solutions d'un problème d'OCD :

0) On vérifie que les fonctions $f, h_1, \dots, h_l$ sont bien convexes et différentiables.
3) On résout le problème dual : on trouve une solution (λ^*, μ^*) du problème dual
4) on cherche des $x^* \in U$ qui minimise $x \to \mathcal{L}(x, (\lambda^*, \mu^*))$ sur U tels que $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \dots, l$. On note par $E_2 \subset K$ cet ensemble.
5b) Par de dualité Lagrangienne (ici la conclusion 1 de la section précédente), les points de E_2 sont solutions au problème $\min_{x \in K} f(x)$ (mais il peut y en avoir d'autres).

Example : On considère $M \in \mathbb{R}^{n \times n}$ telle que $M \succ 0$ (càd M est symétrique définie positive), $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{l \times n}$ et $b \in \mathbb{R}^l$. On souhaite trouver les solutions au problème

$$\min_{x \in \mathbb{R}^n} \left(\frac{1}{2} \langle Mx, x \rangle + \langle q, x \rangle : Ax + b \leq 0 \right) \tag{6.1}$$

où $v \leq 0$ pour $v = (v_j)_{j=1}^l \in \mathbb{R}^l$ quand $v_j \leq 0$ pour tout $j = 1, \dots, l$.

1. Donner la fonction duale de (6.1) et formuler le problème dual (D)
2. Caractériser les solutions de (D) par un système d'inéquations et d'équations linéaires et quadratiques
3. donner une condition suffisante pour que (D) n'admette qu'une seule solution
4. si μ^* est solution de (D), donner les solutions de (6.1) en fonction de λ^*.

1. La fonction de Lagrange associée à (6.1) est

$$\mathcal{L} : \left\{ \begin{array}{ccl} \mathbb{R}^n \times (\mathbb{R}_+)^l & \to & \mathbb{R} \\ (x, \mu) & \to & f(x) + \langle \mu, H(x) \rangle = \frac{1}{2} \langle Mx, x \rangle + \langle q, x \rangle + \langle \mu, Ax + b \rangle \end{array} \right.$$

pour $f(x) = \langle Mx, x \rangle + \langle q, x \rangle$ et $H(x) = Ax + b$ est à valeur vectorielle dans $\mathbb{R}^r$. C'est-à-dire pour tout $x \in \mathbb{R}^n$ et $\mu \geq 0$, on a

$$\mathcal{L}(x, \mu) = \frac{1}{2} \langle Mx, x \rangle + \langle q + A^\top \mu, x \rangle + \langle \mu, b \rangle.$$

La fonction duale est donnée par

$$\psi : \left\{ \begin{array}{ccc} (\mathbb{R}_+)^l & \to & \mathbb{R} \\ \mu & \to & \inf_{x\in\mathbb{R}^n} \left(\frac{1}{2}\langle Mx, x\rangle + \langle q + A^\top \mu, x\rangle + \langle \mu, b\rangle\right) \end{array} \right.$$

Or pour tout $\mu \geq 0$, $\mathcal{L}(\cdot, \mu) : x \to \frac{1}{2}\langle Mx, x\rangle + \langle q + A^\top \mu, x\rangle + \langle \mu, b\rangle$ est convexe (car sa Hessienne est $M \succ 0$). Donc x^*_μ minimise $\mathcal{L}(\cdot, \mu)$ sur $\mathbb{R}^n$ si et seulement si x^*_μ est un point critique de $\mathcal{L}(\cdot, \mu)$ càd $\nabla_x \mathcal{L}(x^*_\mu, \mu) = 0$. Mais comme $\nabla_x \mathcal{L}(x, \mu) = Mx + q + A^\top \mu$, on a nécessairement que

$$x^*_\mu = -M^{-1}(q + A^\top \mu). \tag{6.2}$$

On a donc pour tout $\mu \geq 0$,

$$\begin{aligned} \psi(\mu) = \mathcal{L}(x^*_\mu, \mu) &= \frac{1}{2}\langle q + A^\top \mu, M^{-1}(q + A^\top \mu)\rangle - \langle q + A^\top \mu, M^{-1}(q + A^\top \mu)\rangle + \langle \mu, b\rangle \\ &= \frac{-1}{2}\langle q + A^\top \mu, M^{-1}(q + A^\top \mu)\rangle + \langle \mu, b\rangle \\ &= -\frac{1}{2}\langle \mu, AM^{-1}A^\top \mu\rangle + \langle b - AM^{-1}q, \mu\rangle - \frac{1}{2}\langle q, M^{-1}q\rangle. \end{aligned}$$

On retrouve bien le fait que la fonction duale ψ est concave (sa Hessienne est $-AM^{-1}A \preceq 0$). Le problème dual est

$$\max_{\mu\geq 0} \psi(\mu). \tag{6.3}$$

2) Comme ψ est concave les solutions de (6.3) sont donnés par la condition d'Euler/Pénao/Kantorovitch en OCD : $\mu^* \geq 0$ est solution de (6.3) si et seulement si $\nabla\psi(\mu^*) \in N_{(\mathbb{R}^l_+)}(\mu^*)$ où $N_{(\mathbb{R}^l_+)}(\mu^*)$ est le cône normal à la contrainte $(\mathbb{R}^l_+)$ en μ^*. On a

$$N_{(\mathbb{R}^l_+)}(\mu^*) = \left\{ \sum_{j\in J(\mu^*)} a_j(-e_j) : a_1, \dots, a_l \geq 0 \right\} \tag{6.4}$$

où $(e_1, \dots, e_l)$ est la base canonique de $\mathbb{R}^l$ et $J(\mu^*) = \{j \in \{1, \dots, l\} : \mu^*_j = 0\}$ et

$$\nabla\psi(\mu^*) = -AM^{-1}A^\top \mu^* + (b - AM^{-1}q).$$

On peut caractériser les vecteurs qui appartiennent au cône normal (6.4) de la manière suivante : il y a équivalence entre

a) $v \in N_{(\mathbb{R}^l_+)}(\mu^*)$

b) $-v \in (\mathbb{R}^l_+)$ et $\langle -v, \mu^*\rangle = 0$

En effet, si $v \in N_{(\mathbb{R}^l_+)}(\mu^*)$ alors il existe $a_1, \dots, a_l \geq 0$ tels que $v = \sum_{j\in J(\mu^*)} a_j(-e_j)$. En particulier, $-v = \sum_{j\in J(\mu^*)} a_j e_j \in (\mathbb{R}^l_+)$ et

$$\langle -v, \mu^*\rangle = \sum_{j\in J(\mu^*)} a_j \mu^*_j = 0$$

car $\mu^*_j = 0$ pour tout $j \in J(\mu^*)$. Réciproquement, on prends $v \in \mathbb{R}^l$ tel que $-v \in (\mathbb{R}^l_+)$ et $\langle -v, \mu^*\rangle = 0$. On écrit $v = \sum_{j=1}^l v_j e_j$ où $v_j \leq 0$ car $-v \in (\mathbb{R}^l_+)$. Par ailleurs, $0 = \langle v, \mu^*\rangle = \sum_{j\notin J(\mu^*)} v_j \mu^*_j$ et comme les termes de cette somme sont tous négatifs car $v_j \leq 0$ et $\mu^*_j < 0$, on a donc forcément $v_j\mu_j^* = 0$ pour tout $j \notin J(\mu^*)$ et comme $\mu_j^* < 0$, on a $v_j = 0$ pour tout $\mu_j^* < 0$. On a donc bien $v \in N_{(\mathbb{R}^l_+)}(\mu^*)$.

En utilisant la caractérisation précédente des éléments de $N_{(\mathbb{R}^l_+)}(\mu^*)$, on obtient que $\nabla\psi(\mu^*) \in N_{(\mathbb{R}^l_+)}(\mu^*)$ si et seulement si

$$AM^{-1}A^\top\mu^* - (b - AM^{-1}q) \geq 0 \text{ et } \left\langle AM^{-1}A^\top\mu^* - (b - AM^{-1}q), \mu^*\right\rangle = 0. \tag{6.5}$$

Les solution du problème dual (6.3) sont les solutions du système d'inéquations et d'équations quadratiques et linéaires donné en (6.5).

3. Pour qu'une fonction concave admette un unique maximum il suffit que sa Hessienne soit strictement négative càd ici que $AM^{-1}A^\top \succ 0$. Dans ce cas, le système (6.5) admet une unique solution.

4. La fonction objective f est convexe et différentiable car c'est un polynôme et sa Hessienne est $M \succ 0$. La contrainte est faite uniquement de l contraintes d'inégalités affines, elle est donc qualifiée. Ainsi, par dualité Lagrangienne, si $\mu^* \geq 0$ est une solution du problème dual (6.3), on a équivalence entre :

1) x^* est solution du problème (6.1)

2) x^* minimise $\mathcal{L}(\cdot, \mu^*)$ sur $\mathbb{R}^n$ et $\mu_j^*(Ax^* + b)_j = 0$ pour tout $j = 1, \ldots, l$.

On a déjà vu en (6.2) que l'unique solution au problème de minimisation de $\mathcal{L}(\cdot, \mu^*)$ sur $\mathbb{R}^n$ est donnée par $x^*_{\mu^*} = -M^{-1}(q + A^\top\mu^*)$. De plus, on a pour tout $j = 1, \ldots, l$,

$$\mu_j^*(Ax^*_{\mu^*} + b)_j = -\mu_j^*(AM^{-1}A^\top\mu^* - (b - AM^{-1}q))_j$$

qui est bien nul d'après (6.5) vu que $(AM^{-1}A^\top\mu^* - (b - AM^{-1}q))_j \geq 0$, $\mu_j \geq 0$ et $\sum_j \mu_j(AM^{-1}A^\top\mu^* - (b - AM^{-1}q))_j = 0$. Donc pour toute solution μ^* au problème dual (càd vérifiant (6.5)), $x^*_{\mu^*}$ est solution du problème (6.3) et toute solution de ce problème s'écrit sous la forme $x^*_{\mu^*}$ où μ^* est solution au problème dual.

7 Annexe : Dualité Lagrangienne et conditions KKT en programmation linéaire

Dans cette section, on applique les idées des sections précédentes (dualité Lagrangienne et conditions KKT) au cas particulier des problèmes d'optimisation en Programmation Linéaire sous forme standard

$$\min_{x \in \mathbb{R}^n} \left\langle c, x\right\rangle \text{ tel que } Ax = b, x \geq 0 \tag{PPL}$$

où $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ et $b \in \mathbb{R}^m$. On rappelle que $x \geq 0$ signifie que $x_j \geq 0$ pour tout $j \in \{1, \ldots, n\}$ quand $x = (x_j)_j$.

On suppose que l'ensemble de contrainte $K = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ est non vide (sinon le problème (PPL) n'est pas défini). Ainsi la condition de qualification QC-A, quand toutes les contraintes sont localement affines et continues, est bien vérifiée ici. Par ailleurs, la fonction objectif $x \mapsto \left\langle x, c\right\rangle$ est convexe de classe $\mathcal{C}^1$ et les contraintes "$Ax = b, x \geq 0$" sont affines. On peut donc appliquer les principaux résultats des sections précédentes, à savoir les Théorème 4.1 et Théorème 5.2.

La <u>fonction de Lagrange</u> est ici

$$\mathcal{L}(x, \lambda, \mu) = \left\langle c, x\right\rangle - \left\langle \mu, x\right\rangle + \left\langle \lambda, b - Ax\right\rangle = \left\langle c - \mu - A^\top\lambda, x\right\rangle + \left\langle \lambda, b\right\rangle$$

définie pour tout $x \in \mathbb{R}^n, \mu \in (\mathbb{R}_+)^n$ et $\lambda \in \mathbb{R}^m$. La fonction duale est définie pour tout $\mu \in (\mathbb{R}_+)^n$ et $\lambda \in \mathbb{R}^m$ par

$$\psi(\lambda, \mu) = \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu, \lambda) = \left\{ \begin{array}{cl} -\infty & \text{quand } A^\top \lambda + \mu \neq c \\ \langle \lambda, b \rangle & \text{quand } A^\top \lambda + \mu = c. \end{array} \right.$$

Le problème dual $\max_{\lambda,\mu \geq 0} \psi(\lambda, \mu,)$ est donc ici

$$\max_{\lambda,\mu \geq 0} \langle b, \lambda \rangle \text{ tel que } A^\top \lambda + \mu = c, \mu \geq 0 \tag{DPL}$$

C'est aussi un problème de programmation linéaire.

On rappelle que la condition de qualification QC-A (quand toutes les contraintes sont localement affines et continues) est satisfaites ici car K est supposé non-vide. On a donc de la dualité forte (cf. Théorème 4.1) et d'après les conditions KKT (cf. Théorème 5.2) il y a équivalence entre :

1. x^* est solution de (PPL) et (μ^*, λ^*) est solution de (DPL)
2. (x^*, μ^*, λ^*) vérifie les conditions de KKT :

$$\begin{array}{ll} (KKT1): & Ax^* = b, x^* \geq 0 \text{ et } \mu^* \geq 0 \\ (KKT2): & \mu_i^* x_i^* = 0, \forall i = 1, \ldots, n \\ (KKT3): & 0 = \nabla_x \mathcal{L}(x^*, \mu^*, \lambda^*) = -A^\top \lambda^* - \mu^* + c \end{array}$$

On peut vérifier que le dual du problème dual (DPL) est ici le problème primal (PPL). En effet, le lagrangien de (DPL) est

$$\tilde{L}(\mu, \lambda, x, y) = -\langle b, \lambda \rangle + \langle x, A^\top \lambda + \mu - c \rangle - \langle y, \mu \rangle = \langle Ax - b, \lambda \rangle + \langle x - y, \mu \rangle - \langle x, c \rangle$$

où $(x, y) \in \mathbb{R}^n \times (\mathbb{R}_+)^m$ est ici le multiplicateur de Lagrange. Puis, la fonction duale de Lagrange est ici définie pour tout $y \geq 0, x$ par

$$\tilde{D}(x, y) = \min_{\mu \in \mathbb{R}^n, \lambda \in \mathbb{R}^m} \tilde{L}(\mu, \lambda, x, y) = \left\{ \begin{array}{cl} -\infty & \text{quand } Ax \neq b \text{ ou } x \neq y \\ -\langle c, x \rangle & \text{quand } Ax = b \text{ et } x = y. \end{array} \right.$$

Le problème duale $\max_{y \geq 0, x} \tilde{D}(x, y)$ est alors

$$\max_{x,y} \Big(-\langle c, x \rangle : Ax = b \text{ et } x = y \text{ et } y \geq 0 \Big)$$

qui est bien équivalent è (PPL).

On peut donc de nouveau appliquer la dualité Lagrangienne et obtenir la caractérisation des solutions aux problèmes primal et dual par les conditions KKT qui sont identiques à celles obtenues précédemment.

Théorème 7.1 *Si le problème primal* (PPL) *a une solution (resp. si le problème dual* (DPL) *a une solution) alors c'est aussi le cas pour le problème dual (resp. le problème primal) et les fonctions objectifs sont égales en leurs valeurs optimales. Si un des problèmes n'est pas borné alors l'ensemble des contraintes de l'autre est vide. Par ailleurs, on a l'équivalence entre :*

1. *x^* est solution de* (PPL) *et (λ^*, μ^*) est solution de* (DPL)
2. *$(x^*, (\lambda^*, \mu^*))$ vérifie les conditions de KKT :*

$$Ax^* = b, x^* \geq 0 \textit{ et } \mu^* \geq 0; \quad \mu_i^* x_i^* = 0, \forall i = 1, \ldots, n; \quad A^\top \lambda^* + \mu^* = c.$$

3. *x^* et (λ^*, μ^*) sont faisables (pour leur problème respectif) et $\langle \mu^*, x^* \rangle = 0$.*

8 Annexe : Théorème minmax de Von Neuman - Sion

On a vu dans la section précédente qu'identifier des situations où le duality gap de la fonction de Lagrange est nul est important pour faire le lien entre problème dual et problème primal. On va donc donner ici deux théorèmes assurant un duality gap nul.

Théorème 8.1 (Théorèmes minimax de von Neumann-Sion) *Soit $X \subset \mathbb{R}^n$ et $Y \subset \mathbb{R}^m$ deux ensembles compacts et convexes. Soit $\mathcal{L} : X \times Y \to \mathbb{R}$ une fonction continue et* ***convexe-concave*** *càd :*

a) pour tout $y \in Y, x \in X \to \mathcal{L}(x, y)$ est convexe

b) pour tout $x \in X, y \in Y \to \mathcal{L}(x, y)$ est concave.

Alors,

$$\min_{x\in X} \max_{y\in Y} \mathcal{L}(x, y) = \max_{y\in Y} \min_{x\in X} \mathcal{L}(x, y).$$

Le théorème minmax de von Neuman a été généralisé à plusieures reprises. Une de ces généralisation est donnée ci-dessous. Pour cela, on rappelle les définitions suivantes.

Définition 8.2 *Soit X un espace topologique et $f : X \to \mathbb{R} \cup \{-\infty, +\infty\}$. Soit $x_0 \in X$. On dit que f est* ***semi-continue inférieurement en*** *x_0 quand, si $f(x_0) \neq -\infty$ alors pour tout $\epsilon > 0$ il existe un voisinage ouvert U de x_0 tel que pour tout $x \in U$, $f(x) \leq f(x_0) + \epsilon$ et si $f(x_0) = -\infty$ alors $f(x) \to -\infty$ quand $x \to x_0$.*

De même, on dit que f est ***semi-continue supérieurement*** *en x_0 quand, si $f(x_0) \neq +\infty$, pour tout $\epsilon > 0$, il existe un voisinage ouvert U de x_0 tel que pour tout $x \in U$, on a $f(x) \geq f(x_0) - \epsilon$ et si $f(x_0) = +\infty$ alors $f(x) \to f(x_0)$ quand $x \to x_0$.*

Une fonction continue en x_0 est à la fois semi-continue inférieurement et supérieurement en x_0. La réciproque est vraie : si f est semi-continue inférieurement et supérieurement en x_0 alors f est continue en x_0. Les deux type de semi-continuité sont nécessaires pour avoir la continuité (coir Figure 1).

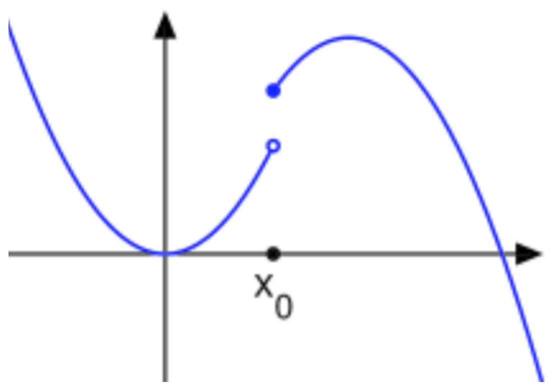

Figure 1 – Fonction semi-continue inférieurement en x_0.

Définition 8.3 *Soit C un ensemble convexe d'un espace linéaire et $f : C \to \mathbb{R}$. On dit que f est* ***quasi-convexe*** *quand pour tout $x, y \in C$ et $0 \leq \alpha \leq 1$, on a $f(\alpha x + (1-\alpha)y) \leq \max(f(x), f(y))$. De même, on dit que f est* ***quasi-concave*** *quand pour tout $x, y \in C$ et $0 \leq \alpha \leq 1$, on a $f(\alpha x + (1 - \alpha)y) \geq \max(f(x), f(y))$*

On peut vérifier qu'une fonction est quasi-convexe si et seulement si tout ces ensembles de niveaux sont convexes, càd pour tout $t \in \mathbb{R}, \{x \in C : f(x) \leq t\}$ est convexe. Une fonction convexe est quasi-convexe mais la réciproque est fausse (voir Figure 2). On peux aussi voir que la quasi-convexité n'implique pas la continuité.

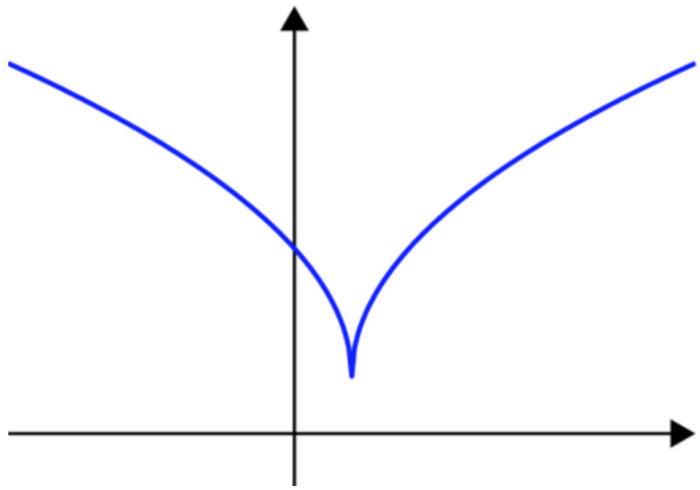

FIGURE 2 – Fonction quasi-convexe mais pas convexe.

Théorème 8.4 *Soit X un ensemble convexe et compact d'un espace linéaire topologique et Y un ensemble convexe d'un espace linéaire. Soit $\mathcal{L} : X \times Y \to \mathbb{R}$. On suppose que*

a) pour tout $x \in X$, $\mathcal{L}(x, \cdot)$ est semi-continue supérieurement et quasi-concave

b) pour tout $y \in Y$, $\mathcal{L}(\cdot, y)$ est semi-continue inférieurement et quasi-convexe.

On a alors

$$\min_{x \in X} \max_{y \in Y} \mathcal{L}(x, y) = \max_{y \in Y} \min_{x \in X} \mathcal{L}(x, y).$$

9 Annexe : Théorème de Danskin et propriétés de différentiation de la fonction duale

La fonction primale est de la forme $\varphi : x \in X \to \sup_{y \in Y} \mathcal{L}(x, y)$, de même la fonctions duale peut s'écrire sous la forme $\psi : y \in Y \to -\sup_{x \in X}(-\mathcal{L}(x, y))$ ou encore la transformée de Fenchel d'une fonction $f : C \to \mathbb{R}$ où $C \subset \mathbb{R}^d$ s'écrie $f^* : y \in \mathbb{R}^d \to \sup_{x \in C}(\langle x, y \rangle - f(x))$. Il y a beaucoup de fonction qui peuvent s'obtenir comme le supremum d'une famille de fonction. Il peut être utile de savoir calculer leur gradient ou sous-gradient par exemple pour construire des algorithmes. Dans ce cas on peut utiliser un Théorème de Danskin ou une extension comme celle de Bertsekas. Avant celà on rappelle la définition de la sous-différentielle et d'un sous-gradient d'une fonction convexe.

Définition 9.1 *Soit C un ensemble convexe de $\mathbb{R}^n$ et $f : C \to \mathbb{R}$ une fonction convexe. Soit $x \in C$. On dit que $g \in \mathbb{R}^n$ est un* ***sous-gradient de** f **en** x* *quand pour tout $z \in C$, on a*

$$f(z) \geq f(x) + \langle g, z - x \rangle.$$

L'ensemble des sous-gradients de f en x est appelé la ***sous-différentielle de** f **en** x* *et est noté $\partial^- f(x)$.*

Si $h : C \to \mathbb{R}$ est une fonction concave et $x \in C$, on dit que g est un ***sur-gradient de** h **en*** *x que $-g$ est un sous-gradient de $-f$ en x et l'ensemble des sur-gradients de h est appelé* ***sur-différentielle*** *et est noté $\partial^+ h(x)$.*

La notion de sous-différentielle généralise celle de différentielle pour les fonctions convexes car dans le cas d'une fonction f convexe différentiable en x on a $\partial^- f(x) = \{\nabla f(x)\}$. On peut voir que la sous-différentielle (resp. sur-différentielle) est un ensemble convexe et fermé. On peut aussi énoncer un résultat du type Euler/Péano/Kantorovitch dans le cas convexe mais non forcément différentiable : soit f et K convexes, on a $x^* \in \operatorname{argmin}_{x\in K} f(x)$ si et seulement si $0 \in \partial^- f(x^*) + N_K(x^*)$ (on retrouve la condition $-\nabla f(x^*) \in N_K(x^*)$ dans le cas différentiable car $\partial^- f(x^*) = \{\nabla f(x^*)\}$ dans ce cas).

Dans le résultat suivant, on caractérise les propriétés de différentiation d'une fonction définie comme le sup d'une famille de fonctions.

Théorème 9.2 (Théorème de Danskin) *Soit U un ouvert convexe de $\mathbb{R}^n$ et S un ensemble compact de $\mathbb{R}^m$. Soit $\mathcal{L} : U \times S \to \mathbb{R}$. On suppose que :*

a) $\mathcal{L}$ est continue sur $U \times S$

b) pour tout $z \in S$, $\mathcal{L}(\cdot, z)$ est convexe.

On pose $f : x \in U \to \max_{z\in S} \mathcal{L}(x, z)$. Pour tout $x \in U$, on pose $Z_x = \{z \in Z : \max_{z\in S} \mathcal{L}(x, z) = \mathcal{L}(x, z)\}$. On a :

1) f est convexe

2) pour tout $x \in U$, pour tout $v \in \mathbb{R}^n$ f admet une dérivée directionnelle en x dans la direction v donnée par $\partial_v f(x) = \max_{z\in Z_x} \partial_v \mathcal{L}(\cdot, y)_{|x}$ où $\partial_v \mathcal{L}(\cdot, y)_{|x}$ est la dérivée directionnelle de $\mathcal{L}(\cdot, y)$ en x dans la direction v.

3) soit $x \in U$. Si Z_x est un singleton $\{z_x\}$ alors f est différentiable en x et dans ce cas le gradient de f est $\nabla f(x) = \nabla \mathcal{L}(\cdot, z_x)_{|x}$, où $\nabla \mathcal{L}(\cdot, z_x)_{|x}$ est le gradient de $\mathcal{L}(\cdot, z_x)$ en x.

4) si $\mathcal{L}(\cdot, z)$ est différentiable sur U pour tout $z \in Z$ et si $\nabla_x \mathcal{L}(x, \cdot)$ est continue pour tout $x \in U$ alors le sous-gardient de f en x est donné par

$$\partial^- f(x) = \operatorname{conv}\left(\nabla_x \mathcal{L}(x, z) : z \in Z_x\right).$$

Il existe un extension du Théorème de Danskin donnée par Bertsekas. En ce qui nous concerne nous nous intéresserons principalement aux propriétés de différentiation de la fonction duale (et de la transformée de Fenchel dans un autre chapitre). Dans le cas de la fonction duale ψ, le théorème de Danskin ne peut pas s'appliquer directement car l'ensemble de la variable dual n'est pas compact : $\mathcal{L} : (x, (\lambda, \mu)) \in U \times (\mathbb{R}^r \times (\mathbb{R}_+)^l) \to \mathbb{R}$ et donc $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ qui n'est pas compact. On peut néanmoins, obtenir des résultats intéressants sur le sous-gradient de ψ et de dualité forte quand ψ est différentiable.

On va obtenir ces résultats dans un cadre plus générale que les contraintes de type inégalités "$g_i(x) = 0$" et d'inégalité "$h_j(x) \leq 0$" qu'on appelle les **contraintes coniques généralisées**. On peut en effet remarquer que l'ensemble des contraintes

$$K = \{x \in U : g_i(x) = 0, i = 1, \ldots, r \text{ et } h_j(x) \leq 0, j = 1, \ldots, l\}$$

peut s'écrire sous la forme $K = \{x \in U : F(x) \leq 0\}$ où $F : x \in U \to \mathbb{R}^{2r+l}$ est une fonction vectorielle donnée par $F(x) = (g_1(x), \cdots, g_r(x), -g_1(x), \ldots, -g_r(x), h_1(x), \ldots, h_l(x))$. Par ailleurs, dire que $F(x) \leq 0$ peut se récrire de manière équivalente en $F(x) \in C$ où C est le cône $C = \{u \in \mathbb{R}^{2r+l} : u \leq 0\}$. On va alors généraliser la forme des contraintes vues jusqu'ici en regardant les contraintes de la forme

$$K = \{x \in U : F(x) \in C\} \text{ où } C \text{ est un cône de } \mathbb{R}^m$$

et m est un entier quelconque. On cherche donc les solutions aux problèmes de la forme

$$\min_{x\in U: F(x)\in C} f(x) \tag{9.1}$$

où C est un cône de $\mathbb{R}^m$. C'est dans ce cadre des problèmes d'optimisation avec contraintes coniques généralisées qu'on va s'intéresser aux propriétés de différentiation de la fonction duale. On retrouvera les résultats dans le cadre classique des contraintes d'inégalité et d'égalité en prenant $C = (\mathbb{R}_-)^{2r+l}$.

Dans le cadre de l'optimisation avec contraintes coniques généralisées, la variable duale prends ses valeurs dans le cône dual de C. On retrouve bien la contrainte duale $\mu \geq 0$ car le cône dual de $C = (\mathbb{R}_-)^{2r+l}$ est $C^\circ = \{\mu : \langle \mu, v\rangle \leq 0, \forall v \in C\} = (\mathbb{R}_+)^{2r+l}$. On définit alors la fonction de Lagrange par

$$\mathcal{L} : \left\{ \begin{array}{ccc} U \times C^\circ & \to & \mathbb{R} \\ (x,\mu) & \to & f(x) + \langle \mu, F(x)\rangle \end{array} \right.$$

et la fonction duale par

$$\psi : \left\{ \begin{array}{ccc} C^\circ & \to & \mathbb{R} \\ \mu & \to & \inf_{x\in U} \big(f(x) + \langle \mu, F(x)\rangle\big) \end{array} \right.$$

C'est bien le cône dual C° de C qui apparaît ici comme espace de la variable duale μ car c'est en faisant ce choix qu'on récupère le problème d'origine (9.1) comme problème primal de $\mathcal{L}$. En effet, on a pour tout $x \in U$,

$$\varphi(x) = \sup_{\mu\in C^\circ} \mathcal{L}(x,\mu) = \left\{ \begin{array}{cc} f(x) & \text{si } F(x) \in C \\ +\infty & \text{sinon} \end{array} \right.$$

et donc le problème primal $\inf_{x\in U} \varphi(x)$ est exactement le problème (9.1).

On a toujours de la dualité faible sans aucune hypothèse (voir Proposition 1.3) :

$$\sup_{\mu\in C^\circ} \inf_{x\in U} \mathcal{L}(x,\mu) \leq \inf_{x\in U} \sup_{\mu\in C^\circ} \mathcal{L}(x,\mu) \tag{9.2}$$

on s'intéresse aux cas où il y a dualité forte càd où il y a égalité dans l'expression du dessus car dans ce cas, la Proposition 1.6 dit que $\mathcal{L}$ a un point-selle et donc dans ce cas on peut appliquer les théorèmes de dualité Lagrangienne, par exemple celui de la conclusion 1, pour déduire des solutions du problème dual, les solutions du problème primal (et donc, de manière équivalente, de (9.1)).

On va montrer ici que si ψ est différentiable en un $\mu^* \in C^\circ$ solution du problème dual alors le duality gap est nul et donc on a dualité forte. On donne d'abord une caractérisation générale de la sur-différentielle de ψ.

Proposition 9.3 *Pour tout $\mu \in C^\circ$, on note $U_\mu = \{x \in U : \mathcal{L}(x,\mu) = \min_{x\in U} \mathcal{L}(x,\mu)\}$. On a*

$$\overline{\text{conv}}\{F(x) : x \in U_\mu\} \subset \partial^+\psi(\mu)$$

où $\overline{\text{conv}}$ désigne l'enveloppe convexe fermée.

Preuve. On suppose que $U_\mu \neq \emptyset$. Soit $x_\mu \in U_\mu$. On a pour tout $\mu_0 \in C^\circ$,

$$\begin{aligned} \psi(\mu_0) &= \inf_{x\in U} \mathcal{L}(x,\mu_0) \leq \mathcal{L}(x_\mu,\mu_0) = f(x_\mu) + \langle \mu_0, F(x_\mu)\rangle \\ &= f(x_\mu) + \langle \mu, F(x_\mu)\rangle + \langle \mu_0 - \mu, F(x_\mu)\rangle = \mathcal{L}(x_\mu,\mu) + \langle \mu_0 - \mu, F(x_\mu)\rangle \\ &= \psi(\mu) + \langle \mu_0 - \mu, F(x_\mu)\rangle \end{aligned}$$

donc $F(x_\mu)$ est un sur-gradient de ψ en μ. On obtient le résultat vu que $\partial^+\psi(\mu)$ est convexe et fermé. ■

Définition 9.4 *On dit que la fonction duale vérifie la* ***filling property en*** $\mu \in C^\circ$ *quand* $\overline{\mathrm{conv}}\{F(x) : x \in U_\mu\} = \partial^+\psi(\mu)$.

10 Annexe : conic duality, relaxation Lagrangienne et dualité forte sous condition de Slater

Dans cette section, on montre un théorème de dualité forte pour la fonction de Lagrange associée à un problème d'optimisation généralisé de la forme

$$\min_{x\in X : F(x)\in C} f(x) \tag{10.1}$$

où X est un ensemble convexe (d'un espace linéaire), $f : X \to \mathbb{R}$, $F : x \in X \to (f_i(x))_{i=1}^m$ et C est un **cône propre** (càd C est un cône convexe fermé d'intérieur non vide et ne contenant pas de droite (càd si $x, -x \in C$ alors $x = 0$)) de $\mathbb{R}^m$. On suppose que $f, f_1, \dots, f_m$ sont convexes.

On ne fait aucune hypothèse sur X autre que la convexité. En particulier, X n'est pas forcément ouvert ou fermé, il peut donc contenir d'autres contraintes que celles données par "$F(x) \in C$". On voit ici que la contrainte générale "$x \in X$ et $F(x) \in C$" est donc décomposée en deux contraintes : une contrainte "facile"(i.e. sur laquelle on peut minimiser facilement) donnée ici par $x \in X$ et une contrainte "difficile" donnée ici par $F(x) \in C$. L'idée de ce qu'on appelle la **relaxation Lagrangienne** est de remplacer la contrainte difficile $F(x) \in C$ par un terme de régularisation "$\langle \mu, F(x)\rangle$" dans la fonction objectif où μ est un paramètre de régularisation qu'on va regarder comme un variable de Lagrange. Dans ce cas, la fonction de Lagrange est donnée par

$$\mathcal{L} : \left\{ \begin{array}{ccc} X \times C^\circ & \to & \mathbb{R} \\ (x, \mu) & \to & f(x) + \langle \mu, F(x)\rangle \end{array} \right.$$

et la fonction duale par

$$\psi : \left\{ \begin{array}{ccc} C^\circ & \to & \mathbb{R} \\ \mu & \to & \inf_{x\in X} \big(f(x) + \langle \mu, F(x)\rangle\big) \end{array} \right.$$

C'est le cône dual $C^\circ = \{z \in \mathbb{R}^m : \langle v, z\rangle \le 0, \forall v \in C\}$ de C qui apparaît ici comme espace de la variable duale μ de $\mathcal{L}$ et ψ.

En effet, la méthode de relaxation Lagrangienne est une méthode par décomposition de la contrainte (on a décomposé la contrainte en deux partie "$x \in X$" et "$F(x) \in C$") qui est aussi une méthode de minoration : pour tout $x \in X$ tel que $F(x) \in C$ on a pour tout $\mu \in C^\circ$,

$$f(x) + \langle \mu, F(x)\rangle \le f(x)$$

et donc

$$\sup_{\mu\in C^\circ} \inf_{x\in X} \mathcal{L}(x, \mu) \le \inf_{x\in X : F(x)\in C} f(x). \tag{10.2}$$

Autrement dit, pour tout $\mu \in C^\circ$, la fonction duale prise en μ : $\psi(\mu)$ est un minorant du problème primal. Ainsi en maximisant la fonction duale ψ sur C°, on aura un plus grand minorant et donc une meilleure évaluation du problème primal. Idéalement, on aimerait que les deux quantités $\sup_{\mu\in C^\circ} \psi(\mu)$ et $\min_{x\in X:F(x)\in C} f(x)$ coïncident – c'est ce qu'on appellera dans la suite la dualité forte – et on aimerait aussi pouvoir trouver une solution au problème primal $\min_{x\in X:F(x)\in C} f(x)$ à partir d'une solution du problème duale $\sup_{\mu\in C^\circ} \psi(\mu)$.

On s'intéresse premièrement à la dualité forte. On observe que la propriété de minoration donnée en (10.2) permet de retrouver la propriété de dualité faible des fonctions de Lagrange même pour cette approche de relaxation Lagrangienne :

$$\sup_{\mu\in C^\circ}\inf_{x\in X}\mathcal{L}(x,\mu)\leq \inf_{x\in X:F(x)\in C} f(x)=\inf_{x\in X}\sup_{\mu\in C^\circ}\mathcal{L}(x,\mu).$$

On veut ici identifier les situations où on a égalité i.e. où on a de la dualité forte pour la fonction de Lagrange :

$$\inf_{x\in X}\sup_{\mu\in C^\circ}\mathcal{L}(x,\mu)=\sup_{\mu\in C^\circ}\inf_{x\in X}\mathcal{L}(x,\mu).$$

En OCD, une condition suffisante de qualification est la condition de Slater : "*il existe $x_0\in X$ tel que $F(x_0)\in \mathring{C}$*". Où $\mathring{C}$ est l'intérieur de C. On a déjà vu grâce au théorème de dualité Lagrangienne dans le cas où X est un ouvert, $f, f_1,\cdots,f_m$ sont différentiables et pour $C=\{z\in\mathbb{R}^m: z\leq 0\}$ (tel que $F(x)\in C$ est équivalent à $f_j(x)\leq 0$ pour tout $j=1,\ldots,m$), qu'on a de la dualité forte sous cette condition de Slater (aussi appelée "faisabilité stricte"). On généralise ce résultat ici et on en donne une preuve simple.

Théorème 10.1 *On suppose que*

1) $f,f_1,\ldots,f_m: X\to\mathbb{R}$ sont convexes où X est un ensemble convexe et C est un cône propre,

2) $p^*=\inf_{x\in X:F(x)\in C} f(x)$, est finie

3) la condition de Slater est vérifiée : il existe $x_0\in X$ tel que $F(x_0)\in\mathring{C}$.

Alors, il y a dualité forte et l'ensemble $C_0=\{\mu_0\in C:\psi(\mu_0)=\sup_{\mu\in C}\psi(\mu)\}$ est non vide et borné.

Preuve. Comme C est un cône propre on peut définir une inégalité généralisée : $y\succeq_C x$ si et seulement si $y-x\in C$. On vérifie que $\succeq_C$

1. préserve l'addition : si $y\succeq_C x$ et $u\succeq_C v$ alors $y+u\succeq_C x+v$
2. est transitive : si $y\succeq_C x$ et $x\succeq_C z$ alors $y\succeq_C z$
3. est positivement homogène : si $y\succeq_C x$ et $\lambda\geq 0$ alors $\lambda y\succeq_C \lambda x$
4. est réflexive : $x\succeq_C x$
5. est anti-symétrique : si $y\succeq_C x$ et $y\succeq_C x$ alors $x=y$

On considère l'ensemble

$$T=\left\{(u,w)\in\mathbb{R}^m\times\mathbb{R}:(F(x),f(x))\succeq_{C\times\mathbb{R}_-}(u,w),\forall x\in X\right\}$$

autrement dit

$$T=\left\{(u,w)\in\mathbb{R}^m\times\mathbb{R}:F(x)-u\in C \text{ et } f(x)\leq w,\forall x\in X\right\}.$$

L'ensemble T est (i) convexe car si $(u_0,w_0),(u_1,w_1)\in T$ et $0\leq t\leq 1$ alors pour tout $x\in X$

$$(F(x),f(x))=t(F(x),f(x))+(1-t)(F(x),f(x))\succeq_{C\times\mathbb{R}_-}t(u_0,w_0)+(1-t)(u_1,w_1).$$

(autrement dit on a pour tout $x\in X$, $f(x)=tf(x)+(1-t)f(x)\leq tw_0+(1-t)w_1$ et $F(x)-u_0\in C$ et $F(x)-u_1\in C$ alors $F(x)-(tu_0+(1-t)u_1)=t(F(x)-u_0)+(1-t)(F(x)-u_1)\in tC+(1-t)C=C$, car C est convexe). On a donc bien $t(u_0,w_0)+(1-t)(u_1,w_1)\in T$.

De plus, l'ensemble T est tel que (ii) si $(u_1,w_1)\in T$ et $(u_1,w_1)\succeq(u_0,w_0)$ alors $(u_0,w_0)\in T$. En effet, pour tout $x\in X,(F(x),f(x))\succeq(u_1,w_1)\succeq(u_0,w_0)$ et donc $(u_0,w_0)\in T$ (autrement dit

$w_0 \geq w_1$ et $u_1 - u_0 \in C$ alors pour tout $x \in X$, $w_0 \geq w_1 \geq f(x)$ et $F(x) - u_0 = 2[(1/2)(F(x) - u_0)] = 2[(1/2)(u_1 - u_0) + (1/2)(F(x) - u_1)] \in C$ car C est cône convexe).

On prouve que $(0, p^*) \notin \mathring{T}$. On raisonne par l'absurde. Si $(0, p^*) \in \mathring{T}$ alors il existe $\epsilon > 0$ tel que $(0, p^* - \epsilon) \in T$ ce qui contredit clairement l'optimalité de p^*. On a donc soit $p^* \in \partial T$ soit $p^* \notin T$. Par le théorème de l'hyperplan séparateur (on utilise ici la convexité de T), il existe $(\lambda, \lambda_0) \in \mathbb{R}^m \times \mathbb{R}$ tel que $(\lambda, \lambda_0) \neq 0$ et pour tout $(u, w) \in T$, on a

$$\langle \lambda, u \rangle + \lambda_0 w \geq \lambda_0 p^*. \tag{10.3}$$

En particulier, on a $-\lambda \in C^\circ$ et $\lambda_0 \geq 0$ sinon, il existerait $\mu \in C$ tel que $\langle -\lambda, \mu \rangle > 0$ ou $\lambda_0 < 0$. Si λ_0

une coordonnée λ_{i_0} de (λ, λ_0) négative et comme T vérifié la propriété (ii), on peut prendre la i_0-ième coordonnée de (u, w) aussi grande qu'on veut alors le produit scalaire $\langle (\lambda, \lambda_0), (u, w) \rangle$ peut être rendu aussi proche de $-\infty$ qu'on veut, ce qui contredit (10.3) car p^* est fini. On considère maintenant deux cas : (a) "$\lambda_0 = 0$" et (b) "$\lambda_0 > 0$".

(a) "$\lambda_0 = 0$" : D'après (10.3) et $\lambda \neq 0$, on a

$$\inf_{(u,w) \in T} \langle \lambda, u \rangle = 0 \tag{10.4}$$

Par ailleurs, Comme $(\lambda, \lambda_0) \neq 0$ et $(\lambda, \lambda_0) \geq 0$ on a nécessairement $\lambda \neq 0$

$$\inf_{(u,w) \in T} \langle \lambda, u \rangle =$$

■

Bidual relaxation Lagrangienne et la relaxation de rang pour le problème de MAX-CUT.

Au premier chapitre, on a vu que le problème du MAX-CUT d'un graphe non-orienté et non pondéré peut se réécrire sous la forme

$$\min_{x \in \mathbb{R}^n} \left(x^\top A x : x_i^2 = 1, i = 1, \ldots, n \right) \tag{10.5}$$

où A est la matrice d'adjacence du graphe – on suppose ici le graphe non orienté et donc A est symétrique. Les contraintes "difficiles" sont ici toutes les $x_i^2 = 1, i = 1, \ldots, n$ car elles imposent aux x_i d'être discrètes (ici $x_i^2 = 1$ ssi $x_i \in \{-1, 1\}$). Ce sont ces contraintes qui font de MAX-CUT un problème combinatoire. On va alors dualiser ces contraintes.

La fonction de Lagrange obtenue par relaxation Lagrangienne des contraintes difficiles associée au problème du MAX-CUT est

$$\mathcal{L} : (x, u) \in \mathbb{R}^n \times \mathbb{R}^n \to x^\top A x - \sum_{i=1}^n u_i (x_i^2 - 1) = x^\top (A - \mathrm{Diag}(u)) x + \langle e, u \rangle$$

où $\mathrm{Diag}(u)$ est la matrice diagonale de taille $n \times n$ dont les éléments diagonaux sont donnés par les coordonnées de u et $e = (1)_1^n$. La fonction duale est

$$\psi : u \in \mathbb{R}^n \to \min_{x \in \mathbb{R}^n} \mathcal{L}(x, u) = \begin{cases} \langle e, u \rangle & \text{if } A - \mathrm{Diag}(u) \succeq 0 \\ -\infty & \text{sinon.} \end{cases}$$

Le problème dual est donc le problème

$$\max_{u \in \mathbb{R}^n} \psi(u) = \max_{u \in \mathbb{R}^n} \left(\langle e, u \rangle : A - \mathrm{Diag}(u) \succeq 0 \right). \tag{10.6}$$

En posant $F(u) = A - \mathrm{Diag}(u)$ et en notant $\mathcal{S}_+^n$ le cône des matrices symétriques semi-définie positives, on voit que le problème dual est un problème de sous-contrainte conique. On peut alors dualiser ce problème comme nous l'avons vu plus haut. Ici, (10.6) est déjà un problème dual (c'est le dual de MAX-CUT), on parle alors de **problème bidual**. La fonction dual de (10.6) est

$$\mathcal{L}' : (u, X) \in \mathbb{R}^n \times (\mathcal{S}_+^n)^\circ \to \big\langle e, u \big\rangle - \big\langle X, A - \mathrm{Diag}(u) \big\rangle = \big\langle e + \mathrm{Diag}(X), u \big\rangle - \big\langle X, A \big\rangle$$

où $\mathrm{Diag}(X)$ est la matrice diagonale de taille $n \times n$ dont les éléments diagonaux sont donnés par les éléments diagonaux de X et $(\mathcal{S}_+^n)^\circ$ est le cône dual de $\mathcal{S}_+^n$. On peut montrer que c'est le cône des matrices symétriques semi-définies négatives. On obtient alors que la fonction duale de (10.6) est

$$\psi' : X \preceq 0 \to \max_{u \in \mathbb{R}^n} \mathcal{L}'(u, X) = \left\{ \begin{array}{cc} +\infty & \text{si } e + \mathrm{Diag}(X) \neq 0 \\ -\big\langle X, A \big\rangle & \text{sinon.} \end{array} \right.$$

En remplaçant X par $-X$, on voit que le problème dual est

$$\min_{X \succeq 0} \psi'(-X) = \min_{X \succeq 0} \big(\big\langle X, A \big\rangle : X_{ii} = 1, i = 1, \ldots, n \big) \tag{10.7}$$

Il est intéressant de remarquer que le problème bidual de MAX-CUT obtenu en (10.7) peut aussi être obtenu en faisant une relaxation convexe de l'ensemble des contraintes. En effet, une autre manière d'écrire le problème initial (10.5) est d'introduire le produit scalaire matriciel $\big\langle A, B \big\rangle = \sum A_{ij} B_{ij}$. On voit alors que $x^\top A x = \big\langle x x^\top, A \big\rangle$ où $x x^\top$ est une matrice symétrique de rang 1. On a alors

$$\min_{x \in \mathbb{R}^n} \Big(x^\top A x : x_i^2 = 1, i = 1, \ldots, n \Big) = \min \Big(\big\langle X, A \big\rangle : X_{ii} = 1, X = X^\top, \mathrm{rang}(X) = 1 \Big). \tag{10.8}$$

En utilisant cette dernière formulation et en remarquant que l'enveloppe convexe de $\{X \in \mathbb{R}^{n \times n} : X_{ii} = 1, X = X^\top, \mathrm{rang}(X) = 1\}$ est incluse dans $\{X \in \mathbb{R}^{n \times n} : X_{ii} = 1, X \succeq 0\}$, on voit qu'on obtient le problème bidual (10.7) à partir de (10.8) en prenant une relaxation convexe de son ensemble de contraintes. La contrainte de rang : "rang$(X) = 1$" est à l'origine des difficultés computationelle de MAX-CUT. L'approche qu'on utilise ici est de tout simplement enlever cette contrainte. En utilisant cette dernière remarque et la dualité faible (de la première dualisation), on obtient l'encadrement suivant du problème de MAX-CUT :

$$\max_{u \in \mathbb{R}^n} \big(\big\langle e, u \big\rangle : A - \mathrm{Diag}(u) \succeq 0 \big) \leq \min_{X \succeq 0} \big(\big\langle X, A \big\rangle : X_{ii} = 1 \big) \leq \min_{x \in \mathbb{R}^n} \Big(x^\top A x : x_i^2 = 1 \Big).$$

11 Annexe : dualité forte pour les problèmes SDP