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 par dualité Lagrangienne est de donner une minoration du problème initial $\min_K f$ (qu'on appellera maintenant problème primal) par un autre problème, en général, plus facile à résoudre. Tout l'enjeu en dualité Lagrangienne 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

$$\min_{x \in K} f(x) \tag{0.1}$$

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

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

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

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

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 minimisée 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 :

$$\min_{x \in K} f(x) \geq \sup_{(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l} \psi(\lambda, \mu). \tag{0.2}$$

La fonction $\mathcal{L}$ est appelée **fonction de Lagrange** ou **Lagrangien** associé au problème primal (0.1). La fonction ψ est appelée **fonction duale** et le problème qui cherche à la maximiser est

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

appelé 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) – c'est ce qu'on appellera plus tard, la **dualité faible**. 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 parlera dans ce cas de **dualité forte**. 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'interversion de "min" et de "sup" de la fonction de Lagrange qu'on va chercher à résoudre. On cherchera donc à identifier les situations pour lesquelles cette interversion est exacte. 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. On les appliquera par la suite au cas particulier de la fonction de Lagrange pour en déduire des liens entre problème primal et problème dual. En particulier, ces liens nous seront utiles à la fois pour déterminer des solutions au problème primal à l'aide du problème dual mais aussi pour construire des algorithmes et des critères d'arrêt.

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

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

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

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

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

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

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

Les trois propositions qui suivent vont établir des liens entre dualité forte, existence de points-selle, solutions du problème primal et solutions du problème dual.

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) $\varphi(x^*) \leq \psi(y^*)$
3) $\varphi(x^*) = \psi(y^*)$
4) $\varphi(x^*) = \psi(y^*) = \mathcal{L}(x^*, y^*)$

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

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

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

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

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

<u>3) ⇒ 4)</u> Par définition de φ et ψ, on a toujours

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

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

<u>4) ⇒ 1)</u> Par définition de φ et ψ, on a pour tout $x \in X$ et $y \in Y$,

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

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. 1) ⇒ 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.

5) ⇒ 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

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

1) ⇒ 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.

6) ⇒ 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 un 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-selle 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

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

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.

Définition 2.1 *La* ***fonction de Lagrange*** *ou* ***Lagrangien*** *associé(e) au problème d'optimisation sous contrainte* (2.1) *où K est défini dans* (2.2) *est défini(e) 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 $\mathcal{L}$ comme définie dans la section précédente.

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 :

$$\psi : (\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l \to \inf_{x\in U} \mathcal{L}(x, (\lambda, \mu)) \tag{2.3}$$

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.

Le problème duale est un problème d'optimisation convexe car c'est un problème de maximisation d'une fonction concave sur un ensemble convexe (qui est donc équivalent à un problème de minimisation d'une fonction convexe sur un convexe). Ceci est vrai tout le temps, sans aucune hypothèse ni sur K ni sur f comme on le montre dans le résultat ci-dessous.

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 d'abord le lien entre des conditions proches de KKT (points *i), ii) et iii)* ci-dessous) et l'existence d'un point-selle de la fonction de Lagrange (et donc de la dualité forte et de solution primale et duale). Puis, en utilisant le lien entre points-selles de la fonction de Lagrange et solution du problème (2.1) on établit un 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 $j = 1, \dots, l$.

Preuve. 1) ⇒ 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 $j = 1, \dots, l$."

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 $j = 1, \ldots, l$.

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 $\mathcal{L}$. ■

D'après le Théorème 3.1, il suffit de trouver un point $(x^*, (\lambda^*, \mu^*))$ satisfaisant *i), ii) et iii)* pour assurer l'existence d'un point-selle de $\mathcal{L}$ et donc la dualité forte, une solution x^* au problème primal et une solution (λ^*, μ^*) au problème dual. Dans le cas de l'OCD, ces trois conditions sont équivalentes aux conditions KKT car dans ce cadre $\mathcal{L}(\cdot, (\lambda^*, \mu^*))$ est convexe et donc dire que x^* est un point critique (comme dans les conditions KKT) est équivalent à dire que c'est un minimum (comme dans *i)*).

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 de cette solution (λ^*, μ^*) (on peut ici prendre n'importe quelle solution au problème duale), 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 $\mu_j^* h_j(x^*) = 0, \forall j = 1, \dots, l$.

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 9).

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 $g_1, \dots, g_r$ sont affines et $h_1, \dots, h_l$ 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 $j = 1, \dots, l$. 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 $j = 1, \dots, l$ donc la propriété 7) du Théorème 3.1 est satisfaite et donc par le Théorème 3.1, $(x^*, (\lambda^*, \mu^*))$ 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

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

où les $g_1, \dots, g_r : U \to \mathbb{R}$, définissant les contraintes d'égalité, sont des **fonctions affines** et les $h_1, \dots, h_l : U \to \mathbb{R}$, 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 $j = 1, \dots, l, h_j(x_0) < 0$. 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 $K = \{x \in U : g_i(x) = 0, \forall i = 1, \dots, r, h_j(x) \le 0, \forall j = 1, \dots, l\}$,*

$$x^* \in K \text{ et } \mu^* \geq 0 \tag{KKT1}$$

$$\mu_j^* h_j(x^*) = 0, \forall j = 1, \dots, l \tag{KKT2}$$

$$\nabla_x \mathcal{L}(x^*, (\lambda^*, \mu^*)) = 0 \tag{KKT3}$$

*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 $\mathcal{L}$,

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

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 $j = 1, \ldots, r$ 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.

<u>Conclusion 1</u> : *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.

<u>Conclusion 2</u> : *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 $(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT.

On voit que dans la première conclusion, on n'a pas besoin de qualifier la contrainte : il suffit de trouver une solution (λ^*, μ^*) au problème duale 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.

On peut tirer d'autres conclusions du Théorème 5.2. Par exemple, du point *1.* de ce théorème, on voit que si le duality gap est nul (c'est pas exemple le cas en (OCD) sous l'hypothèse que A est de rang r et la condition de Slater est vérifiée – voir Section **??**) et si (λ^*, μ^*) est une solution au problème dual alors toute solution x^* du problème primal doit aussi satisfaire que $(x^*, (\lambda^*, \mu^*))$ satisfait les conditions KKT. En particulier, si $L(\cdot, (\lambda^*, \mu^*))$ admet un unique minimum sur U alors soit ce minimum est dans K et $\mu_j^* h_j(x^*) = 0, j = 1, \ldots, l$ et alors le primal admet une unique solution donnée par x^* soit ce minimum n'est pas dans K on ne vérifie pas $\mu_j^* h_j(x^*) = 0, \forall j = 1, \ldots, l$ et dans ce cas, le primal n'a pas de solution.

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 l'hypothèse de qualification. 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 $f, h_1, \ldots, h_l$ sont bien convexes et différentiables.
2) On vérifie que la contrainte est qualifiée (Slater ou contraintes affines ou vérifiant Mangasarian-Fromovitch)
3) On résout le problème dual : on trouve une solution (λ^*, μ^*) au problème dual
4) on cherche les $x^* \in U$ tel que $(x^*, (\lambda^*, \mu^*))$ vérifie KKT (càd les x^* tels que $\nabla_x \mathcal{L}(x^*, (\lambda^*, \mu^*)) = 0$, $x^* \in K$ et $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$). On note par $E_2 \subset K$ cet ensemble.
5) Par 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. En effet, sous l'hypothèse de qualification, on sait que pour n'importe quelle solution (λ^*, μ^*) au problème dual alors $(x^*, (\lambda^*, \mu^*))$ vérifie KKT ssi x^* est solution au problème primal.

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 doit trouver au moins un point $(x^*, (\lambda^*, \mu^*))$ vérifiant KKT. C'est l'existence d'un tel point qui par la *Conclusion 1* permet de dire qu'on a dualité forte. Une fois qu'on a certifié la dualité forte (ici par l'existence d'un point vérifiant KKT), on peut appliquer le théorème 5.2 et donc avoir l'équivalence entre 'x^* est solution primale et (λ^*, μ^*) est solution duale' et '$(x^*, (\lambda^*, \mu^*))$ vérifiant KKT'. En particulier, il suffit de trouver une solution (λ^*, μ^*) au problème duale et de chercher tous les x^* tel que $(x^*, (\lambda^*, \mu^*))$ vérifie KKT pour avoir toutes les solutions au problème. Cependant, si on ne trouve pas de point $(x^*, (\lambda^*, \mu^*))$ vérifiant KKT cela ne signifie pas que le problème primal n'a pas de solution car il se peut que la contrainte ne soit pas qualifiée en un point x^* qui serait solution du problème. C'est cette situation que permet d'éviter l'hypothèse de qualification quand elle est vérifiée.

Méthodologie d'application de la dualité Lagrangienne en OCD sans qualification. 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 $x^* \in U$ tel que $(x^*, (\lambda^*, \mu^*))$ vérifie KKT (càd $\nabla_x \mathcal{L}(x^*, (\lambda^*, \mu^*)) = 0$, $x^* \in K$ et $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \dots, l$). On note par $E_2 \subset K$ cet ensemble.
5b) Par dualité Lagrangienne (ici la conclusion 1 de la section précédente), les points de E_2 sont les solutions au problème $\min_{x\in K} f(x)$ (il ne peut pas y en avoir d'autre car l'existence d'un point $(x^*, (\lambda^*, \mu^*))$ vérifiant KKT à permit de certifier la dualité forte et donc toute solution primale x^* est telle que $(x^*, (\lambda^*, \mu^*))$ vérifie KKT pour n'importe quelle solution (λ^*, μ^*) du duale).

Exemple : 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

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

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 $\mu \geq 0$,

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

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

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

où $(e_1, \ldots, e_l)$ est la base canonique de $\mathbb{R}^l$ et $J(\mu^*) = \{j \in \{1, \ldots, l\} : \mu_j^* = 0\}$ 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 $\big\langle -v, \mu^*\big\rangle = 0$

En effet, si $v \in N_{(\mathbb{R}_+^l)}(\mu^*)$ alors il existe $a_1, \ldots, a_l \geq 0$ 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

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

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 $\big\langle -v, \mu^*\big\rangle = 0$. On écrit $v = \sum_{j=1}^l v_j e_j$ où $v_j \leq 0$ car $-v \in (\mathbb{R}_+^l)$. Par ailleurs, $0 = \big\langle v, \mu^*\big\rangle = \sum_{j \notin J(\mu^*)} v_j \mu_j^*$ 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

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

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 Interprétation géométrique des dualités faible et forte et Théorème de Slater.

On considère l'ensemble

$$\mathcal{G} = \{(g_1(x), \cdots, g_r(x), h_1(x), \cdots, h_l(x), f(x)) : x \in U\} \subset \mathbb{R}^r \times \mathbb{R}^l \times \mathbb{R}.$$

La valeur optimale atteinte par le problème primal, notée $\varphi^* = \min(f(x) : x \in K)$, peut se retrouver à l'aide de $\mathcal{G}$ car

$$\varphi^* = \min\left(t : (u, v, t) \in \mathcal{G}, u = 0, v \leq 0\right).$$

On peut aussi retrouver la fonction duale à partir de $\mathcal{G}$ car pour tout $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$, on a

$$\psi(\lambda, \mu) = \inf\Big((\lambda, \mu, 1)^\top (u, v, t) : (u, v, t) \in \mathcal{G}\Big).$$

En particulier, pour tout $(u, v, t) \in \mathcal{G}$, on a

$$(\lambda, \mu, 1)^\top (u, v, t) - \psi(\lambda, \mu) \geq 0. \tag{7.1}$$

Ainsi, pour tout $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ tel que $\psi(\lambda, \mu)$ est fini, la ligne de niveau 0 de la fonction affine

$$(u, v, t) \in \mathbb{R}^r \times \mathbb{R}^l \times \mathbb{R} \to (\lambda, \mu, 1)^\top (u, v, t) - \psi(\lambda, \mu)$$

est un hyperplan supportant $\mathcal{G}$ (voir Figure 1).

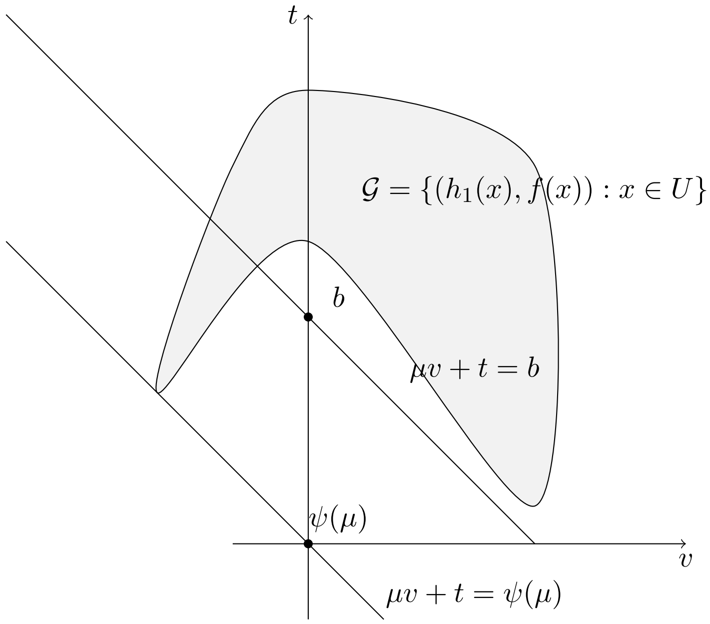

FIGURE 1 – Étant donné un $\mu \geq 0$ et $b \in \mathbb{R}$, la droite $\{(v,t) : \mu v + t = b\}$ a une pente négative donnée par $-\mu$ et un intercepte égal à b. Pour retrouver la fonction duale $\psi : \mu \geq 0 \to \inf_{x\in U}(f(x) + \mu h_1(x))$, on fait diminuer b (et donc la droite $\{(v,t) : \mu v + t = b\}$ est abaissée parallèlement à celle d'origine) jusqu'au point dernier point de contact avec $\mathcal{G}$. On peut ensuite voir la valeur de $\psi(\mu)$ comme étant l'intercept de cette dernière droite tangente à $\mathcal{G}$ et de pente $-\mu$.

On rappelle la définition d'un hyperplan supportant un ensemble.

Définition 7.1 *Soit $\mathcal{G}$ un ensemble d'un espace de Hilbert H. Pour $w \in H$ et $b \in \mathbb{R}$, on dit que $\{x \in H : \langle w, x \rangle = b\}$ est un* ***hyperplan supportant*** *$\mathcal{G}$ quand pour tout $x \in \mathcal{G}$, on a $\langle w, x \rangle \geq b$.*

Selon la Définition 7.1, $\{(u,v,t) : (\lambda,\mu,1)^\top(u,v,t) - \psi(\lambda,\mu) = 0\}$ est un hyperplan supportant $\mathcal{G}$. Par ailleurs, comme le vecteur normal $(\lambda,\mu,1)$ de cet hyperplan a un coefficient égal à 1 en dernière coordonnée, on dit que c'est un hyperplan supportant $\mathcal{G}$ non-vertical. Réciproquement, si $\{(u,v,t) : (\lambda,\mu,1)^\top(u,v,t) = b\}$ est un hyperplan supportant $\mathcal{G}$ non-vertical alors nécessairement $b \leq \psi(\lambda,\mu)$ étant donné la définition de $\psi(\lambda,\mu)$. Ainsi, $\{(u,v,t) : (\lambda,\mu,1)^\top(u,v,t) - \psi(\lambda,\mu) = 0\}$ pour $(\lambda,\mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ sont tous les hyperplans supportant $\mathcal{G}$ 'tangent' à $\mathcal{G}$ en dessous de $\mathcal{G}$ avec $\mu \geq 0$ (càd avec des coefficients positifs sur les coordonnées de v).

Par ailleurs, l'hyperplan d'équation $(\lambda,\mu,1)^\top(u,v,t) - \psi(\lambda,\mu) = 0$ coupe le dernier axe des coordonnées en $t = \psi(\lambda,\mu)$. On peut donc retrouver $\psi(\lambda,\mu)$ à partir cet hyperplan supportant $\mathcal{G}$ qui lui est aussi tangent.

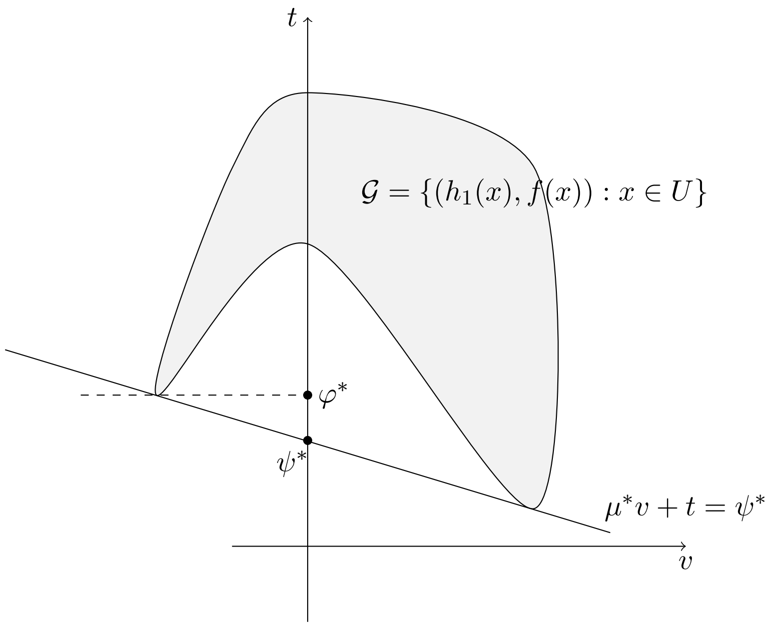

FIGURE 2 – Interprétation géométrique de la dualité faible pour le problème $\min_{x \in U}(f(x) : h_1(x) \leq 0)$. La fonction duale $\psi : \mu \geq 0 \to \inf_{x \in U}(f(x) + \mu h_1(x))$ se retrouve comme $\psi(\mu) = \inf(\mu v + t : (v,t) \in \mathcal{G})$ et peut se lire sur l'axe des 't' car $\mu \geq 0$ càd pour $v = 0$ dans l'hyperplan $\mu v + t = \psi(\mu)$. On retrouve la valeur optimale du problème primal $\varphi^* = \min(f(x) : h_1(x) \leq 0)$ comme $\varphi^* = \min(t : (t,v) \in \mathcal{G}, v \leq 0)$ et la valeur optimale du problème dual par $\psi^* = \max(\psi(\mu) : \mu \geq 0)$. On peut ainsi visualiser la dualité faible sur cette figure : $\psi^* \leq \varphi^*$ (et ici un duality gap $\varphi^* - \psi^* > 0$).

Application au Théorème de Slater. On considère un problème de type (OCD) où on rappelle que la contrainte est de la forme

$$K = \{x \in U : Ax = b \text{ et } h_1(x) \leq 0, \cdots, h_l(x) \leq 0\} \tag{7.2}$$

où $A \in \mathbb{R}^{r \times n}$ et $b \in \mathbb{R}^r$. On rappelle que la contrainte K vérifie la condition de Slater quand il existe $x_0 \in K$ tel que pour tout $j = 1, \ldots, l$, on a $h_j(x_0) < 0$.

Théorème 7.2 (Théorème de Slater) *Soit K un contrainte en (OCD) de la forme (7.2) où A est de rang r. On suppose que K vérifie la condition de Slater alors il y a dualité forte. De plus si $\inf(f(x) : x \in K)$ est fini alors le problème dual admet une solution.*

Preuve. La preuve du Théorème de Slater que nous donnons maintenant est basée sur un argument géométrique de séparation de deux ensembles convexes disjoints dans l'espace $\mathbb{R}^r \times \mathbb{R}^l \times \mathbb{R}$ où vit l'ensemble $\mathcal{G}$ qui a été introduit ci-dessus. Cet argument est similaire à celui que nous avons démontré concernant la séparation stricte entre un convexe fermé non vide et un point qui lui est extérieur. Ici on utilise le résultat suivant : si $\mathcal{A}$ et $\mathcal{B}$ sont deux ensembles convexes disjoints d'un espace de Hilbert H de dimension finie alors il existe $w \in H \backslash \{0\}$ et $b \in \mathbb{R}$ tels que a) pour tout $x \in \mathcal{A}$, $\langle w, x \rangle + b \leq 0$ et b) pour tout $x \in \mathcal{B}$, $\langle w, x \rangle + b \geq 0$. Ce résultat est aussi appelé Théorème de Hahn-Banach géométrique.

On remarque que pour $\varphi^* = \min_{x \in K} f(x)$ on a soit $\varphi^* = -\infty$ ou φ^* est fini. Si $\varphi^* = -\infty$ alors par dualité faible $\psi^* = \max(\psi(\lambda, \mu) : \mu \geq 0)$ vaut aussi $-\infty$ et donc $\psi^* = \varphi^*$, càd pas de duality gap. On peut maintenant supposer que φ^* est fini.

Les deux ensembles que nous allons faire intervenir dans la preuve du théorème de Slater peuvent être visualisés sur la Figure 3. Ils sont définis par

$$\mathcal{A} = \Big\{(u, v, t) \in \mathbb{R}^r \times \mathbb{R}^l \times \mathbb{R} : \exists x \in U, g_i(x) = u_i, h_j(x) \leq v_j, f(x) \leq t\Big\}$$

et

$$\mathcal{B} = \Big\{(0, 0, s) \in \mathbb{R}^r \times \mathbb{R}^l \times \mathbb{R} : s < \varphi^*\Big\}$$

où on rappelle que $\varphi^* = \min_{x\in K} f(x)$. L'ensemble $\mathcal{A}$ peut être vu comme un épigraphe de $\mathcal{G}$ sur la fonction objectif et les contraintes d'inégalité.

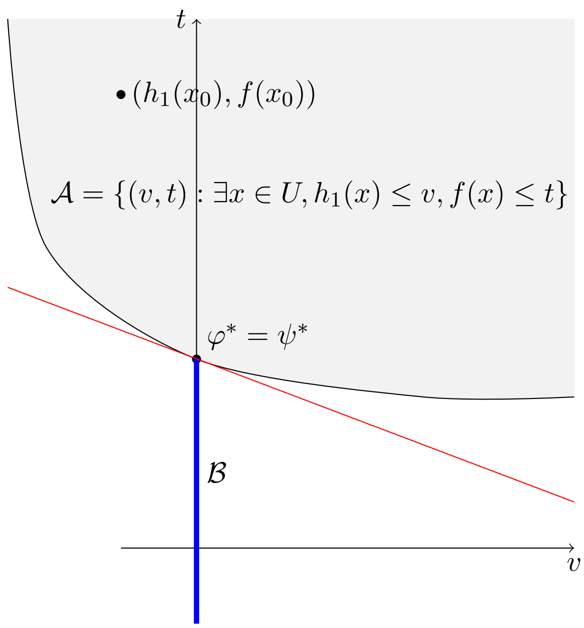

FIGURE 3 – Preuve du théorème de Slater via le théorème de Hahn-Banach géométrique. Les ensembles $\mathcal{A}$ et $\mathcal{B}$ sont séparés par l'hyperplan représenté par une ligne rouge. L'existence d'un point $x_0 \in K$ tel que $h_1(x_0) < 0$ (càd la condition de Slater) assure l'existence d'un hyperplan séparateur non verticale entre $\mathcal{A}$ et $\mathcal{B}$.

On montre que $\mathcal{A}$ et $\mathcal{B}$ sont **convexes** et **disjoints**. Il est clair que $\mathcal{B}$ est convexe. Pour $\mathcal{A}$, on prend (u_1, v_1, t_1) et (u_2, v_2, t_2) deux points de $\mathcal{A}$ et $0 \leq \alpha \leq 1$. Il existe $x_1, x_2 \in U$ tels que $G(x_1) = u_1, H(x_1) \leq v_1, f(x_1) \leq t_1$ et $G(x_2) = u_2, H(x_2) \leq v_2, f(x_2) \leq t_2$ où on note $G(x) = (g_i(x))_i$ et $H(x) = (h_j(x))_j$. Par convexité de $f, h_1, \cdots, h_l$ et affinité de $g_1, \ldots, g_r$, on a $G(\alpha x_1+(1-\alpha)x_2) = \alpha G(x_1)+(1-\alpha)G(x_2) = \alpha u_1+(1-\alpha)u_2$, $H(\alpha x_1+(1-\alpha)x_2) \leq \alpha v_1+(1-\alpha)v_2$ et $f(\alpha x_1+(1-\alpha)x_2) \leq \alpha t_1+(1-\alpha)t_2$. Finalement, par convexité de U, on a que $\alpha x_1+(1-\alpha)x_2 \in U$ et donc $\alpha(u_1, v_1, t_1) + (1 - \alpha)(u_2, v_2, t_2) \in \mathcal{A}$. Par ailleurs, $\mathcal{A}$ et $\mathcal{B}$ sont bien disjoints vu que $\inf(t : (0, 0, t) \in \mathcal{A}) = \inf(f(x) : g_i(x) = 0, h_j(x) \leq 0) = \varphi^*$ et $(0, 0, \varphi^*) > (0, 0, s)$ pour tout $(0, 0, s) \in \mathcal{B}$.

On peut donc appliquer le théorème de séparation (rappelé ci-dessus) aux ensembles convexes disjoints $\mathcal{A}$ et $\mathcal{B}$: il existe $(\tilde{\lambda}, \tilde{\mu}, \tilde{a}) \in \mathbb{R}^r \times \mathbb{R}^l \times \mathbb{R}\backslash\{(0, 0, 0)\}$ et $\alpha \in \mathbb{R}$ tel que

$$\forall (u, v, t) \in \mathcal{A}, \tilde{\lambda}^\top u + \tilde{\mu}^\top v + \tilde{a}t \geq \alpha \tag{7.3}$$

et pour tout $(u, v, t) \in \mathcal{B}$, $\tilde{\lambda}u + \tilde{\mu}v + \tilde{a}t \leq \alpha$ càd $\tilde{a}s \leq \alpha$ pour tout $s < \varphi^*$ et donc par passage à la limite quand $s \uparrow \varphi^*$, on a

$$\tilde{a}\varphi^* \leq \alpha. \tag{7.4}$$

De (7.3), on déduit que $\tilde{\mu} \geq 0$ et $\tilde{a} \geq 0$ sinon, $\tilde{\mu}v + \tilde{a}t$ ne pourrait pas être borné inférieurement sur $\mathcal{A}$ au vue de la définition de $\mathcal{A}$. On déduit de (7.3) et (7.4) que pour tout $x \in U$,

$$\tilde{\lambda}^\top (Ax - b) + \tilde{\mu}^\top H(x) + \tilde{a} f(x) \geq \alpha \geq \tilde{a}\varphi^*. \tag{7.5}$$

Si on peut prouver que $\tilde{a} > 0$ (pour l'instant, on sait seulement que $\tilde{a} \geq 0$) alors on en déduirait que pour tout $x \in U$,

$$\mathcal{L}(x, (\tilde{\lambda}/\tilde{a}, \tilde{\mu}/\tilde{a})) \geq \varphi^*.$$

Ceci étant vrai pour tout $x \in U$, on aurait aussi $\psi(\tilde{\lambda}/\tilde{a}, \tilde{\mu}/\tilde{a}) = \inf_{x \in U} \mathcal{L}(x, (\tilde{\lambda}/\tilde{a}, \tilde{\mu}/\tilde{a})) \geq \varphi^*$ et donc $\psi^* \geq \varphi^*$. On aurait donc dualité forte et $(\tilde{\lambda}/\tilde{a}, \tilde{\mu}/\tilde{a})$ serait solution du problème dual. Il reste donc à montrer que $\tilde{a} \neq 0$, càd que l'hyperplan séparant $\mathcal{A}$ et $\mathcal{B}$ n'est pas vertical. C'est ici que l'hypothèse de Slater intervient.

Si on avait $\tilde{a} = 0$ alors dans (7.5), on aurait $\tilde{\lambda}^\top (Ax - b) + \tilde{\mu}^\top H(x) \geq 0$. En particulier, au point x_0 de la condition de Slater, on aurait $0 \leq \tilde{\lambda}^\top (Ax_0 - b) + \tilde{\mu}^\top H(x_0) = \tilde{\mu}^\top H(x_0)$ et comme $H(x_0) < 0$ on aurait forcément $\tilde{\mu} = 0$. On a donc $\tilde{\mu} = 0$, $\tilde{a} = 0$ et comme $(\tilde{\lambda}, \tilde{\mu}, \tilde{a}) \neq 0$, on en déduit que $\tilde{\lambda} \neq 0$. Ainsi, dans (7.5), pour tout $x \in U$, $\tilde{\lambda}^\top (Ax - b) \geq 0$. Comme U est un ouvert et que $Ax_0 = b$, il existe $x \in U$ tel que $\tilde{\lambda}^\top (Ax - b) < 0$ à moins que $A^\top \tilde{\lambda} = 0$ (en effet, $k : x \in U \to \tilde{\lambda}^\top (Ax - b)$ est une fonction affine telle que $k(x_0) = 0$, alors soit $k \equiv 0$ qui est équivalent à $A^\top \tilde{\lambda} = 0$ ou k prend forcément des valeurs négatives et positives dans un voisinage de x_0). Or comme A est de rang maximal r, on a $\mathrm{Im}(A) = \mathbb{R}^r$ et donc $\ker(A^\top) = \mathrm{Im}(A)^\perp = \{0\}$. Ce qui contredit le fait que $\tilde{\lambda} \neq 0$ donc il existe bien un $x \in U$ tel que $\tilde{\lambda}^\top (Ax - b) < 0$. Ce qui contredit le fait que pour tout $x \in U$, $\tilde{\lambda}^\top (Ax - b) \geq 0$. On en déduit que $\tilde{a} \neq 0$ et le résultat. ■

Remarque 7.3 *1) On se sert de l'hypothèse de Slater uniquement pour montrer que l'hyperplan séparateur entre $\mathcal{A}$ et $\mathcal{B}$ est non vertical.*

2) On n'a pas besoin d'hypothèse de régularité sur les fonctions $f, h_1, \cdots, h_l$. Seule la convexité est ici utilisée.

3) Comme aucune hypothèse de régularité sur les fonction n'est utilisée, on n'a plus besoin que U soit ouvert. En fait, on peut remplacer l'ouvert convexe U dans le théorème de Slater par un ensemble convexe (non nécessairement ouvert) K_0 à condition de supposer l'existence d'un point de Slater x_0 dans l'intérieur relatif de K_0 – cet ensemble, noté $\mathrm{relint}(K_0)$, *est l'ensemble de tous les points de K_0 tel qu'il existe un voisinage ouvert V dans $\mathbb{R}^n$ de ce point pour lequel $V \cap \mathrm{aff}(K_0) \subset K_0$ où $\mathrm{aff}(K_0)$ est le plus petit espace affine contenant K_0 – tel que $g_i(x_0) = 0$ et $h_j(x_0) < 0$ pour les contraintes qu'on a choisies de dualiser. Cette remarque peut être utile quand on ne dualise pas toutes les contraintes. Par exemple, si $K = \{x \in U : g_1(x) = 0, h_1(x) \leq 0\}$, on peut poser $K_0 = \{x \in U : h_1(x) \leq 0\}$ et considérer la fonction duale $\psi(\lambda) = \inf_{x \in K_0}(f(x) + \lambda g_1(x))$. Cette fonction duale ne fait intervenir que la contrainte g_1 dans la fonction de Lagrange ; la contrainte h_1 est 'restée' dans la contrainte K_0. Le théorème de Slater dans ce cadre dit que s'il existe un $x_0 \in \mathrm{relint}(K_0)$ tel que $g_1(x_0) = 0$ alors $\max(\psi(\lambda) : \lambda \in \mathbb{R}) = \inf(f(x) : x \in K)$.*

Exemple : On considère une norme $\|\cdot\|$ sur $\mathbb{R}^n$ et sa norme duale $x \in \mathbb{R}^n \to \|x\|^* = \sup_{\|y\| \leq 1} \langle y, x \rangle$. On considère une matrice $A \in \mathbb{R}^{r \times n}$ de rang r et $b \in \mathbb{R}^r$. Le problème d'optimisation que nous considérons ici est

$$\min \left(\|x\| : Ax = b \right). \tag{7.6}$$

La fonction duale associée à (7.6) est $\psi : \lambda \in \mathbb{R}^r \to \inf_{x\in\mathbb{R}^n} \big(\|x\| + \big\langle \lambda, Ax - b\big\rangle\big)$. On a

$$\begin{aligned}\inf_{x\in\mathbb{R}^n}\Big(\|x\| + \big\langle A^\top\lambda, x\big\rangle\Big) &= \inf_{r\geq 0}\inf_{\|x\|=r}\Big(r + r\big\langle A^\top\lambda, x/\|x\|\big\rangle\Big) = \inf_{r\geq 0}\Big(r - r\left\|A^\top\lambda\right\|^*\Big)\\ &= \begin{cases} 0 & \text{si } \left\|A^\top\lambda\right\|^* \leq 1\\ -\infty & \text{si } \left\|A^\top\lambda\right\|^* > 1.\end{cases}\end{aligned}$$

On en déduit que pour tout $\lambda \in \mathbb{R}^r$,

$$\psi(\lambda) = \begin{cases} -\big\langle\lambda, b\big\rangle & \text{si } \left\|A^\top\lambda\right\|^* \leq 1\\ -\infty & \text{si } \left\|A^\top\lambda\right\|^* > 1.\end{cases}$$

Le problème dual associé à (7.6) est $\max_{\lambda\in\mathbb{R}^r}\psi(\lambda)$ qui est équivalent à

$$\max_{\lambda\in\mathbb{R}^r}\Big(-\big\langle\lambda, b\big\rangle : \left\|A^\top\lambda\right\|^* \leq 1\Big). \tag{7.7}$$

La contrainte du problème primal est $\{x \in \mathbb{R}^n : Ax = b\}$. Elle vérifie la condition de Slater si elle est non vide (il n'y a pas de contraintes d'inégalité ici). Or A est de rang r donc b est bien dans l'image de A et donc K est non vide. Alors la contrainte vérifie la condition de Slater et donc par le théorème de Slater il y a dualité forte :

$$\min\left(\|x\| : Ax = b\right) = \max_{\lambda\in\mathbb{R}^r}\Big(-\big\langle\lambda, b\big\rangle : \left\|A^\top\lambda\right\|^* \leq 1\Big).$$

Prenons un exemple où la norme $\|\cdot\|$ est différentiable : on choisit la norme Euclidéenne. Dans ce cas, sa norme duale est elle-même : $\|\cdot\| = \|\cdot\|^* = \|\cdot\|_2$. Par ailleurs, la contrainte vérifie la condition de Slater et est donc qualifiée. On a donc équivalence entre :

a) x^* est solution de (7.6) et λ^* est solution de (7.7)

b) (x^*, λ^*) vérifie les conditions de KKT.

Ici les conditions KKT sont :

KKT1 $Ax = b$

KKT2 il n'y a pas de contraintes d'égalité donc pas de complementary condition

KKT3 $\nabla_x\mathcal{L}(x^*, \lambda^*) = 0$ où $\mathcal{L} : (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^r \to \|x\|_2 + \big\langle\lambda, Ax - b\big\rangle$

On calcul le gradient de la fonction de Lagrange :

$$\nabla_x\mathcal{L}(x, \lambda) = \frac{x}{\|x\|_2} + A^\top\lambda$$

On voit que pour $x^* = A^\top(AA^\top)^{(-1)}b$ et $\lambda^* = -(AA^\top)^{(-1)}b/\|x^*\|$ on a $Ax^* = b$ (car $b \in \mathrm{Im}(A)$ donc $AA^\top(AA^\top)^{(-1)}b = b$) et

$$A^\top\lambda^* = \frac{-A^\top(AA^\top)^{(-1)}b}{\|x^*\|} = \frac{-x^*}{\|x^*\|}.$$

Donc (x^*, λ^*) vérifie les conditions KKT. Etant donné l'équivalence rappelée plus haut, on a que $x^* = A^\top(AA^\top)^{(-1)}b$ est solution du problème, on a aussi $\|x^*\|_2 = -\big\langle\lambda^*, b\big\rangle$ (càd pas de dualité gap). En fait, on reconnaît ici avec $x^* = A^\top(AA^\top)^{(-1)}b$ la projection de 0 sur l'espace affine $\{x : Ax = b\}$.

Conditions KKT en (OCD) via le théorème de Slater. On peut rétrouver le théorème de KKT en (OCD) grâce à la dualité forte donnée par le théorème de Slater et le lien entre solutions primale/duale et point-selle de la fonction de Lagrange. On considère un problème de type (OCD) où la contrainte est de la forme (7.2). On suppose que rang$(A) = r$. Ainsi par le théorème de Slater, il y a dualité forte.

On suppose maintenant que x^* est solution du problème primal et (λ^*, μ^*) est solution du probème dual. Par la Proposition 1.7, on sait que $(x^*, (\lambda^*, \mu^*))$ est un point-selle du Lagrangien : pour tout $(x, (\lambda, \mu)) \in U \times \mathbb{R}^r \times (\mathbb{R}_+)^l$, on a

$$\mathcal{L}(x^*, (\lambda, \mu)) \overset{(a)}{\leq} \mathcal{L}(x^*, (\lambda^*, \mu^*)) \overset{(b)}{\leq} \mathcal{L}(x, (\lambda^*, \mu^*)). \tag{7.8}$$

En particulier, on voit dans *(b)* que x^* est un minimum de $x \in U \to \mathcal{L}(x, (\lambda^*, \mu^*))$. Or dans le cadre de l'OCD, cette fonction est convexe et différentiable donc x^* est un point critique de cette foncion et donc $\nabla_x \mathcal{L}(x^*, (\lambda^*, \mu^*)) = 0$, on retrouve la condition KKT3 du théorème de KKT. Pour KKT1, *(a)* étant vraie pour tout $(\lambda, \mu)) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ et $\mathcal{L}(x^*, (\lambda^*, \mu^*)) < +\infty$, on voit forcément que $Ax^* = b$ et $h_j(x^*) \leq 0, j = 1, \ldots, l$. On a donc $x^* \in K$. Par ailleurs, $\mu^* \geq 0$, donc KKT1 est bien vraie. Finalement, pour KKT2, on a par *(a)* et $Ax^* = b$, que pour tout $\mu \geq 0$,

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

Ainsi $\sum_j \mu_j^* h_j(x^*) \geq 0$ et comme $\mu^* \geq 0$ et $h_j(x^*) \leq 0, j = 1, \ldots, l$, on en déduit que $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$, càd KKT2 est aussi satisfaite. On a donc bien montrer que $(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT.

Réciproquement, on suppose que $(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT. Par KKT3, x^* est un point critique de $x \in U \to \mathcal{L}(x, (\lambda^*, \mu^*))$ qui est une fonction convexe diffirentiable donc x^* est un minimum de cette fonction. Donc l'inégalité *(b)* dans (7.8) est satisfaite. Par KKT1 et KKT2, on a pour tout $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ que

$$\mathcal{L}(x^*, (\lambda, \mu)) = f(x^*) + \sum_j \mu_j h_j(x^*) \leq f(x^*) = \mathcal{L}(x^*, (\lambda^*, \mu^*))$$

et donc l'inégalité *(a)* dans (7.8) est satisfaite. On en déduit que $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$. Par la Proposition 1.7, on en déduit que x^* est solution du problème primal et (λ^*, μ^*) est solution du problème dual.

On retrouve donc bien le résultat suivant en (OCD).

Théorème 7.4 *On considère le cadre de l'OCD où $f, h_1, \ldots, h_l : U \to \mathbb{R}$ sont convexes et différentiables sur l'ouvert convexe U. Soit K un contrainte de la forme (7.2) où A est de rang r. On suppose que K vérifie la condition de Slater. Soit $x^* \in U$ et $(\lambda^*, \mu^*) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$. Il y a équivalence entre :*

1) x^* est solution du problème primal et (λ^*, μ^*) est solution du problème dual

2) $(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT.

Pour retrouver le théorème de KKT en OCD sous condition de Slater, il suffit d'utiliser le Théorème 7.4 et de montrer l'existence d'une solution au problème dual.

Théorème 7.5 *On considère le cadre de l'OCD où $f, h_1, \ldots, h_l : U \to \mathbb{R}$ sont convexes et différentiables sur l'ouvert convexe U. Soit K un contrainte de la forme (7.2) où A est de rang r. On suppose que K vérifie la condition de Slater. Soit $x^* \in U$. Il y a équivalence entre :*

1) x^* est solution du problème primal.

2) il existe (λ^*, μ^*) une solution du problème dual tel que $(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT.

Preuve. L'implication *2)* implique *1)* est immédiate à partir du Théorème 7.4. Il reste à montrer l'implication *1)* implique *2)*. On suppose donc que x^* est solution du problème primal. D'après le Théorème 7.4, il suffit de démontrer l'existence d'une solution au problème duale pour conclure. On a $\inf(f(x) : x \in K) = f(x^*)$ qui est donc fini alors d'après le théorème de Slater, le problème duale admet une solution. ■

Remarque 7.6 *En OCD, '$(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT' est équivalent à dire que '$(x^*, (\lambda^*, \mu^*))$ est un point-selle de la fonction de Lagrange' (voir Théoreme 3.1). Ainsi si 2) est vraie alors on a dualité forte et l'existence d'une solution x^* au primal et d'une solution (λ^*, μ^*) au dual. Pour le sens 2) implique 1) on peut donc se passer de la condition de Slater. On peut même se passer de l'hypothèse que (λ^*, μ^*) est une solution du problème dual vu que si $(x^*, (\lambda^*, \mu^*))$ vérifie les conditions KKT alors (λ^*, μ^*) est solution du problème dual.*

Le théorème de Slater permet donc de retrouver le théorème de KKT en OCD sous la condition de qualification de Slater. En particulier, cette preuve ne fait pas intervenir d'éléments de géométrie différentielle comme la preuve de KKT que nous avons donné dans les chapitres précédent. Elle est principalement basé sur un argument de convexité qui est le théorème de Hahn-Banach géométrique (que nous avons aussi utilisé pour la preuve de KKT des chapitres précédent). Néanmoins, le théorème de Slater ne permet pas de retrouver le théorème de KKT dans le cas de l'OD ni celui dans le cas de l'OCD sous l'hypothèse générale de qualification.

Il existe plusieurs preuves du théorème de KKT selon le point de vue qu'on adopte. Celle présentée aux chapitres d'avant utilise le point de vue de la géométrie différentielle et propose une hypothèse de qualification qui apparaît naturellement. La preuve qu'on donne ici via le théorème de Slater est une preuve de l'analyse convexe. On peut aussi dans ce cadre proposer une hypothèse de qualification de contrainte ; ici cette hypothèse de qualification est de dire qu'il y a dualité forte. Ainsi, même dans ce sens, la condition de Slater est une hypothèse de qualification (et la preuve en est donnée par le théorème de Slater). Il est intéressant de voir quels sont les liens entre ces deux hypothèse de qualifications. Par exemple, au Corollaire 4.2 du chapitre sur l'optimisation convexe, on a vu que si K est qualifiée en x^* alors il y a dualité forte. Ainsi la qualification au sens de la géométrie différentielle en un point solution primale implique la 'qualification au sens de l'analyse convexe'.

8 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

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

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 $j \in \{1, \dots, n\}$ 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 $x \mapsto \langle x, c\rangle$ 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 fonction de Lagrange est ici

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

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

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

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 :

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

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

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

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 8.1 *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 :*

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

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

9 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 9.1 (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 9.2 *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 4).

Définition 9.3 *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 $f(\alpha x + (1-\alpha)y) \geq \max(f(x), f(y))$*

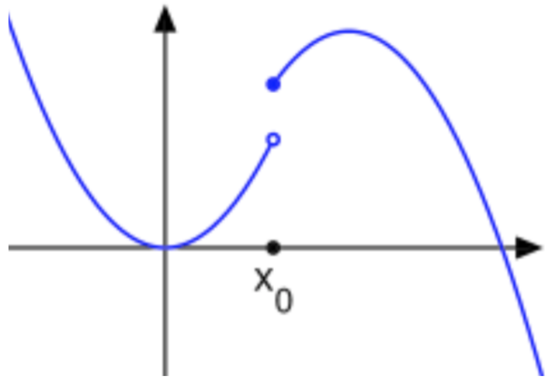

FIGURE 4 – Fonction semi-continue inférieurement en x_0.

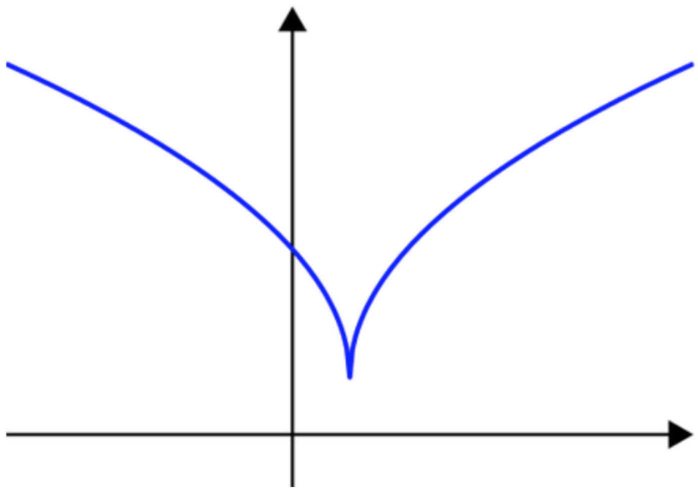

FIGURE 5 – Fonction quasi-convexe mais pas convexe.

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 5). On peux aussi voir que la quasi-convexité n'implique pas la continuité.

Théorème 9.4 *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

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

10 Annexe : dualité forte en conic LP sous condition de Slater

Dans cette section, on considère des problèmes de minimisation d'une fonction linéaire sous une contrainte qui est l'intersection d'un cône et d'un espace affine. Le problème typique qu'on va considérer est le problème *SDP* pour semi-definite programming. On rappelle cette classe de problème qu'on rencontre souvent en optimisation.

Définition 10.1 *Soit $C \in \mathbb{R}^{n\times n}$, $A_1 \dots, A_m \in \mathbb{R}^{n\times n}$ et $b \in \mathbb{R}^m$. On note $\mathcal{A} : X \in \mathbb{R}^{n\times n} \to (\langle A_i, X\rangle)_{i=1}^m$. Un **problème SDP** est de la forme*

$$\min\big(\langle C, X\rangle : X \succeq 0, \mathcal{A}(X) = b\big)$$

où on rappelle que $X \succeq 0$ signifie que X est positive (càd $\langle v, Xv\rangle \geq 0$ pour tout $v \in \mathbb{R}^n$).

La fonction objective d'un problème SDP est bien linéaire et sa contrainte est bien l'intersection d'un cône (ici le cône des matrices positives – si $X \succeq 0$ alors $\lambda X \succeq 0$ pour tout $\lambda \geq 0$) et d'un espace affine, l'ensemble des X tel que $\mathcal{A}(X) = b$. On va étudier les problème SDP en détail dans la suite. En particulier, on cherchera à identifier des conditions impliquant la dualité forte pour ce type de problème. Pour l'instant, on considère une classe plus grande de problèmes qu'on appelle les problèmes **conic LP** pour problème de programmation linéaire conique.

Définition 10.2 *Soit E un espace Euclidien (i.e. un espace vectoriel de dimension finie) muni d'un produit scalaire noté $\langle\cdot,\cdot\rangle$. Soit $d \in E$, $a_1, \dots, a_m \in E$, $b \in \mathbb{R}^n$ et $C \subset E$ un cône convexe et fermé de E. On note $\mathcal{A} : x \in E \to (\langle a_i, x\rangle)_{i=1}^m$. Un problème de type **conic LP** est un problème d'optimisation de la forme*

$$\min_{x\in E}\big(\langle c, x\rangle : x \in C, \mathcal{A}(x) = b\big)\,. \tag{10.1}$$

Les problèmes de type *conic LP* sont donc bien tous de la forme 'minimum d'une fonction linéaire sous une contrainte qui est l'intersection d'un cône et d'un espace affine'. Les problèmes SDP en donnent des exemples où le cône est ici le cône des matrices SDP (semi-definite positive).

L'objectif de cette section est d'étudier les problèmes de type conic LP sous l'angle de la dualité Lagrangienne et de l'analyse convexe. On va donc obtenir leur problème dual et introduire la condition de Slater dans ce cadre qui va nous permette d'assurer la dualité forte pour ces problèmes. Un fois la dualité forte établie, on pourra utiliser le problème dual pour 'certifier' une solution du problème primal. Cette approche qu'on appelle **certification duale** est une approche classique en relaxation convexe car elle permet de montrer qu'une solution d'un problème obtenu par relaxation convexe est bien aussi solution du problème d'origine grâce à la construction d'un certificat dual qui 'certifie' que la solution du problème d'origine est bien solution du problème relâché (on peut même montrer l'unicité dans certain cas). On utilise cette approche par exemple pour des problème de détection de communautés ou de synchronisation ou de coupes optimales dans des graphes.

Problème dual et dualité faible en conic LP. On commence par donner le problème dual associé à un conic LP. On va devoir dualiser les contrainte de ce type de problème. Les contraintes d'égalité sont gérées classiquement mais pour la contrainte de cône '$x \in C$' on doit introduire un ensemble pour sa variable duale associée. Cet ensemble est donné par le **cône dual** :

$$C^\circ = \{z \in E : \langle z, x\rangle \leq 0, \forall x \in C\}. \tag{10.2}$$

Par exemple, le cône dual des matrices $X \succeq 0$ est le cône des matrices négatives càd telle que $Z \preceq 0$ (autrement dit $-Z \succeq 0$.)

La fonction de Lagrange associée à un problème (10.1) est donnée par

$$\mathcal{L} : \left\{\begin{array}{ccc} E \times \mathbb{R}^m \times C^\circ & \longrightarrow & \mathbb{R} \\ (x, (\lambda, \mu)) & \longrightarrow & \langle c, x\rangle + \langle \lambda, b - \mathcal{A}(x)\rangle + \langle \mu, x\rangle. \end{array}\right.$$

La fonction duale est donnée par

$$\psi : \left\{ \begin{array}{ccc} C^\circ & \longrightarrow & \mathbb{R} \\ \mu & \longrightarrow & \inf_{x\in E}\left(\langle c,x\rangle + \langle \lambda, b-\mathcal{A}(x)\rangle + \langle \mu, x\rangle\right) \end{array}\right. = \left\{ \begin{array}{cc} \langle \lambda, b\rangle & \text{si } c-\mathcal{A}^\top(\lambda)+\mu = 0 \\ -\infty & \text{sinon} \end{array}\right.$$

où on note $\mathcal{A}^\top : \mathbb{R}^m \to E$ l'opérateur adjoint de $\mathcal{A}$ défini par $\langle \mathcal{A}^\top(\lambda), x\rangle = \langle \lambda, \mathcal{A}(x)\rangle$ pour tout $\lambda \in \mathbb{R}^m$ et $x \in E$. On vérifie que $\mathcal{A}^\top(\lambda) = \sum_i \lambda_i a_i$ pour tout $\lambda \in \mathbb{R}^m$. Le problème dual du problème conic LP (10.1) est alors

$$\max_{(\lambda,\mu)\in\mathbb{R}^n\times E}\left(\langle \lambda, b\rangle : \mathcal{A}^\top(\lambda) - \mu = c, \mu \in C^\circ\right). \tag{10.3}$$

On retrouve bien le problème (10.1) comme problème primal de $\mathcal{L}$ vu que pour tout $x \in E$,

$$\max\left(\mathcal{L}(x,(\lambda,\mu)) : (\lambda,\mu)\in\mathbb{R}^n\times C^\circ\right) = \left\{ \begin{array}{cc} +\infty & \text{si } \mathcal{A}(x)\neq b \text{ ou } x\notin C \\ \langle c,x\rangle & \text{sinon.} \end{array}\right.$$

Pour la dualité faible, on a toujours

$$\max_{(\lambda,\mu)\in\mathbb{R}^n\times C^\circ}\min_{x\in E}\mathcal{L}(x,(\lambda,\mu)) \leq \min_{x\in E}\max_{(\lambda,\mu)\in\mathbb{R}^n\times C^\circ}\mathcal{L}(x,(\lambda,\mu))$$

et donc

$$\max_{(\lambda,\mu)\in\mathbb{R}^n\times C^\circ}\psi(\lambda,\mu) \leq \min_{x\in K}\langle c,x\rangle \tag{10.4}$$

où $K = \{x \in E : \mathcal{A}(x) = b, x \in C\}$ est la contrainte primale. L'inégalité (10.4) exprime la dualité faible en conic LP. Elle est (comme d'habitude) toujours vraie sans aucune hypothèse. Notre objectif est d'identifier les situations où cette inégalité est une égalité, càd quand nous avons dualité forte. La dualité forte nous sera ensuite utile pour certifier des solutions du primal à l'aide d'une solution du dual. Ici au vue de la fonction duale, on aura dualité forte si on trouve un $\lambda^* \in \mathbb{R}^m$ et un $\mu^* \in C^\circ$ tels que $\mathcal{A}^\top(\lambda^*) + \mu^* = c$ et un x^* tel que $\mathcal{A}(x^*) = b$ et $x^* \in C$ qui satisfont

$$\langle \lambda^*, b\rangle = \langle c, x^*\rangle. \tag{10.5}$$

C'est l'objectif du paragraphe suivant que de construire un tel couple $(x^*, (\lambda^*, \mu^*))$ sous la condition de Slater.

Dualité forte en conic LP sous condition de Slater. Pour assurer la dualité forte on va utiliser une condition similaire à celle de Slater adaptée au type de contrainte conique que nous considérons en conic LP. On note $\text{int}(C)$ l'intérieur du cône C de même par $\text{int}(C^\circ)$ l'intérieur du cône C°. Le but de cette section est de démontrer le théorème suivant.

Théorème 10.3 *Soit E un espace Euclidien muni d'un produit scalaire noté $\langle \cdot, \cdot\rangle$. Soit $c \in E$, $a_1, \ldots, a_m \in E$, $b \in \mathbb{R}^n$ et $C \subset E$ un cône convexe et fermé de E. On note $\mathcal{A} : x \in E \to (\langle a_i, x\rangle)_{i=1}^m$. On considère le problème de **conic LP***

$$\min_{x\in E}\left(\langle c,x\rangle : x\in C, \mathcal{A}(x) = b\right) \tag{10.6}$$

et son problème dual

$$\max_{(\lambda,\mu)\in\mathbb{R}^n\times E}\left(\langle \lambda, b\rangle : \mathcal{A}^\top(\lambda) - \mu = c, \mu \in C^\circ\right). \tag{10.7}$$

On suppose qu'il existe $x_0 \in \text{int}(C)$ et $\mu_0 \in \text{int}(C^\circ)$. Alors le problème primal (10.6) admet une solution, le problème dual (10.7) aussi et il y a dualité forte. On a donc l'existence de solutions primales x^* et duales (λ^*, μ^*) et toutes ces solutions satisfont $\langle c, x^*\rangle = \langle \lambda^*, b\rangle$.

Le reste de ce paragraphe est dédié à donner une preuve au Théorème 10.3. Cette preuve (comme toutes les preuve de dualité forte) utilise le théorème de séparation de convexes disjoints. Ici ce théorème est présent au travers d'un **théorème de l'alternative**. Ces théorèmes sont souvent utilisé en optimisation le plus connu d'entre eux porte le nom de Lemme de Farkas. Le Lemme de Farkas est un résultat qui se démontre simplement mais qui est très important en mathématique car c'est un théorème d'existence. Il dit que si on a des points $a_1, \ldots, a_k$ dans $\mathbb{R}^n$ et un autre point $a \in \mathbb{R}^n$. Il y a deux alternatives possibles :

1) soit a est dans l'enveloppe convexe des k points $a_1, \ldots, a_k$: càd il existe $0 \leq \lambda_1, \ldots, \lambda_k \leq 1$ tel que $\sum_i \lambda_i = 1$ et $a = \sum_i \lambda_i a_i$.
2) soit a n'est pas dans l'enveloppe convexe des k points $a_1, \ldots, a_k$ et alors d'après le théorème de séparation stricte de convexes disjoints, il existe $w \in \mathbb{R}^n$ tel que $\langle a_k, w \rangle < \langle a, w \rangle$.

Ce type de théorème propose alors deux alternatives. On peut en effet réécrire le Lemme de Farkas de la manière suivante : le système $\mathcal{A}^\top(\lambda) = x$ (où $\mathcal{A}^\top(\lambda) = \sum_i \lambda_i a_i$) admet une solution telle que $0 \leq \lambda_i \leq 1$ et $\sum_i \lambda_i = 1$ si et seulement si il n'existe pas de $w \in \mathbb{R}^n$ tel que $\langle a_k - a, w \rangle < 0$.

On va mettre en place un théorème de l'alternative pour prouver le Théorème 10.3.

Théorème 10.4 *Soit E un espace Euclidien muni d'un produit scalaire noté $\langle \cdot, \cdot \rangle$. Soit $a_1, \ldots, a_m \in E$, $b \in \mathbb{R}^n$ et $C \subset E$ un cône convexe et fermé de E. On note $\mathcal{A} : x \in E \to (\langle a_i, x \rangle)_{i=1}^m$. On suppose qu'il existe un $\lambda_0 \in \mathbb{R}^m$ tel que $\mathcal{A}^\top(\lambda_0) \in \mathrm{int}(C^\circ)$. Alors le système $\mathcal{A}(x) = b$ a une solution dans C si et seulement si le système $\mathcal{A}^\top(\lambda) \in C^\circ$ et $\langle \lambda, b \rangle = 1$ n'a pas de solution.*

Pour démontrer le Théorème 10.4, on donne deux lemmas. Le premier porte sur l'intérieur d'un cône dual.

Lemme 10.5 *On a $\mathrm{int}(C^\circ) = \{z \in E : \langle x, z \rangle < 0, \forall x \in C \backslash \{0\}\}$.*

Le second lemme porte sur l'image de C par $\mathcal{A}$. C'est pour démontrer le résultat de fermeture dans ce lemme qu'on a besoin que l'intérieur du cône C° soit non vide.

Lemme 10.6 *On suppose qu'il existe un $\lambda_0 \in \mathbb{R}^m$ tel que $\mathcal{A}^\top(\lambda_0) \in \mathrm{int}(C^\circ)$. L'image de C par $\mathcal{A}$, notée $\mathcal{A}C$, est un cône non vide fermé et convexe.*

Preuve du Théorème 10.4. On démontre que les deux systèmes du Théorème 10.4 ne peuvent pas avoir de solution simultanément. On suppose qu'il existe $x \in C$ et $\lambda \in \mathbb{R}^m$ tels que

$$\mathcal{A}(x) = b, -\mathcal{A}^\top(\lambda) \in C^\circ \text{ et } \langle \lambda, b \rangle = 1. \tag{10.8}$$

Comme $x \in C$ et $\mathcal{A}^\top(\lambda) \in C^\circ$ on a

$$0 \leq \langle -\mathcal{A}^\top(\lambda), x \rangle = -\langle \lambda, \mathcal{A}(x) \rangle = -\langle \lambda, b \rangle = -1.$$

Ce qui est une contradiction et donc les deux systèmes du Théorème 10.4 ne peuvent pas avoir de solution simultanément.

On suppose maintenant qu'il n'existe pas de solution $x \in C$ au système $\mathcal{A}(x) = b$. Montrons que le deuxième système a bien une solution, càd qu'il existe $\lambda \in \mathbb{R}^m$ tel que $\mathcal{A}^\top(\lambda) \in C^\circ$ et $\langle \lambda, b \rangle = 1$. Par hypothèse $b \notin \mathcal{A}C$ et comme $\mathcal{A}C$ est un convexe fermé non vide, il existe (d'après le théorème de séparation des convexes fermés non vides disjoints) un $\tilde{\lambda} \in \mathbb{R}^m$ tel que pour tout $x \in C$, $\langle \tilde{\lambda}, \mathcal{A}(x) \rangle < \langle \tilde{\lambda}, b \rangle$. En particulier, $0 \in C$ donc $\alpha := \langle \lambda, b \rangle > 0$. On a ainsi pour $\lambda = \tilde{\lambda}/\alpha$ que $\langle \lambda, b \rangle = 1$ et pour tout $x \in C$, $\langle \lambda, \mathcal{A}(x) \rangle < 1$ càd $\langle \mathcal{A}^\top(\lambda), x \rangle < 1$ pour tout $x \in C$ et donc $\mathcal{A}^\top(\lambda) \in C^\circ$. En effet, pour tout $x \in C$ et $\gamma > 0$, on a $\gamma x \in C$ alors $\langle \mathcal{A}^\top(\lambda), \gamma x) \rangle < 1$ et donc $\langle \mathcal{A}^\top(\lambda), x) \rangle < 1/\gamma$. Ceci étant vrai pour tout $\gamma > 0$, en prenant $\gamma \to +\infty$, on a $\langle \mathcal{A}^\top(\lambda), x) \rangle \leq 0$. Ceci étant vrai pour tout $x \in C$, on en déduit bien que $\mathcal{A}^\top(\lambda) \in C^\circ$. ■

Preuve du Théorème 10.3. On prouve le Théorème 10.3 grâce au théorème de l'alternative Théorème 10.4. Pour démontrer le Théorème 10.3, il suffit de montrer que le système

$$(S) : \left\{ \begin{array}{c} \langle \lambda, b \rangle - \langle c, x \rangle = 0 \\ \mathcal{A}(x) = b, x \in C \\ \mathcal{A}^\top(\lambda) - \mu = c, \mu \in C^\circ \end{array} \right.$$

admet une solution.

Prouver le Théorème 10.3 revient donc a prouver l'existence d'une solution à un système sous contraintes coniques et affines. On est donc dans le cadre d'appliquation du Théorème de l'alternative Théorème 10.4. Pour pouvoir appliquer ce résultat, on doit d'abord trouver le problème dual qui lui est associé. On réécrit le système (S) sous la forme demandée au Théorème 10.4. On considère l'opérateur

$$\mathcal{A}_0 : (x, \lambda, \mu) \in E \times \mathbb{R}^m \times E \to (\langle \lambda, b \rangle - \langle c, x \rangle, \mathcal{A}(x), \mathcal{A}^\top(\lambda) - \mu) \in \mathbb{R} \times \mathbb{R}^m \times E,$$

$b_0 = (0, b, c)$ et le cône convexe fermé non-vide $C_0 = C \times \mathbb{R}^m \times C^\circ$. On voit que (S) est équivalent à

$$\mathcal{A}_0(x, \lambda, \mu) = b_0 \text{ et } (x, \lambda, \mu) \in C_0. \tag{10.9}$$

Le problème alternatif associé à (10.9) est de trouver une solution y au système

$$(AS) : \left\{ \begin{array}{l} \mathcal{A}_0^\top y \in C_0^\circ \\ \langle b_0, y \rangle = 1. \end{array} \right.$$

On doit alors déterminer l'adjoint de $\mathcal{A}_0$ et le cône dual de C_0. On voit que

$$\mathcal{A}_0 = \left(\begin{array}{c|c|c} -c^\top & b^\top & 0 \\ \hline \mathcal{A} & 0 & 0 \\ \hline 0 & \mathcal{A}^\top & -I_E \end{array} \right) \text{ et donc } \mathcal{A}_0^\top = \left(\begin{array}{c|c|c} -c & \mathcal{A}^\top & 0 \\ \hline b & 0 & \mathcal{A} \\ \hline 0 & 0 & -I_E \end{array} \right)$$

où I_E est la matrice identité définie sur E. On précise ici les dimensions des opérateurs : $\mathcal{A}_0 : E \times \mathbb{R}^m \times E \to \mathbb{R} \times \mathbb{R}^m \times E$ et donc $\mathcal{A}_0^\top : \mathbb{R} \times \mathbb{R}^m \times E \to E \times \mathbb{R}^m \times E$. Pour le cône dual de C_0 on a $C_0^\circ = C^\circ \times \{0_m\} \times C$ où on note 0_m le 0 de $\mathbb{R}^m$ ($\{0_m\}$ étant le cône dual de $\mathbb{R}^m$). On peut maintenant réécrire le système (AS) de manière plus explicite : (AS) est équivalent à trouver $(\tau, \lambda', y') \in \mathbb{R} \times \mathbb{R}^m \times E$ tel que

$$(AS) : \left\{ \begin{array}{l} -\tau c + \mathcal{A}^\top(\lambda') \in C^\circ \\ \tau b + \mathcal{A}(y') = 0 \\ -y' \in C \\ \langle b, \lambda' \rangle + \langle c, y' \rangle = 1 \end{array} \right.$$

Pour appliquer le Théorème 10.4 aux système duaux (S) et (AS) on doit vérifier que l'intérieur du cône C_0° contient bien un élément de l'image de $\mathcal{A}_0^\top$, càd l'existence d'un point $(\tau, \lambda', y') \in \mathbb{R} \times \mathbb{R}^m \times E$ tel que $\mathcal{A}_0^\top(\tau, \lambda', y') \in \mathrm{int}(C_0^\circ)$. Ici l'intérieur (en fait, on devrait parler d'intérieur relatif càd de l'intérieur de C_0° relativement au plus petit espace affine le contenant, càd à $E \times \{0_m\} \times E$, de telle sorte que la deuxième composante $\{0_m\}$ ne fasse pas que C_0° soit d'intérieur vide ; on passe ce détail un peu sous silence pour éviter cette complication mais logiquement, on devrait écrire 'relint' – pour relatif intérieur – au lieu de 'int" depuis le début). On a alors que $\mathrm{int}(C_0^\circ) = \mathrm{int}(C^\circ) \times \{0_m\} \times \mathrm{int}(C)$ et donc on veut prouver l'existence d'un point $(\tau, \lambda', y') \in \mathbb{R} \times \mathbb{R}^m \times E$ tel que $-c\tau + \mathcal{A}^\top(\lambda') \in \mathrm{int}(C^\circ)$ et $y' \in \mathrm{int}(C)$. On peut alors prendre $\tau = 0$, $\lambda' = \mu_0$ et $y' = x_0$

pour les points x_0 et μ_0 dont on a supposé l'existence dans les hypothèses du Théorème 10.3. On peut alors bien appliquer Théorème 10.4 : (S) a une solution si et seulement si (AS) n'a pas de solution. Il reste alors à montrer que (AS) n'a pas de solution.

On raisonne par l'absurde : on suppose que (AS) a une solution. On note $(\tau, \lambda', y') \in \mathbb{R}\times\mathbb{R}^m\times E$ une telle solution au système (AS). On remarque que $\langle -\tau c + \mathcal{A}^\top(\lambda'), y'\rangle \leq 0$ car $-\tau c + \mathcal{A}^\top(\lambda') \in C^\circ$ et $y' \in C$. Mais comme $\mathcal{A}^\top(y') = -\tau b$ et que $\langle b, \lambda'\rangle + \langle c, y'\rangle = 1$ on a

$$0 \geq \langle -\tau c + \mathcal{A}^\top(\lambda'), y'\rangle = -\tau(\langle b, \lambda'\rangle + \langle c, y'\rangle) = -\tau.$$

On en déduit alors que $\tau \geq 0$. On considère deux cas : soit $\tau = 0$ soit $\tau > 0$.

Si $\tau = 0$: Dans ce cas $(\lambda', y') \in \mathbb{R}^m \times E$ sont tels que $\mathcal{A}^\top(\lambda') \in C^\circ$, $\mathcal{A}(y') = 0$, $-y' \in C$ et $\langle b, \lambda'\rangle + \langle c, y'\rangle = 1$. Soit $\bar{x}$ un point faisable pour le primal. On a pour tout $\alpha \geq 0$, que $\alpha(-y') + \bar{x} \in C$ et $\mathcal{A}(\alpha(-y') + \bar{x}) = b$ (car $\mathcal{A}(y') = 0$ et $\mathcal{A}(\bar{x}) = b$) donc $\alpha(-y') + \bar{x}$ est faisable pour le primal. La valeur de la fonction objective primale en ce point vaut

$$\langle c, \alpha(-y') + \bar{x}\rangle = -\alpha\langle c, y'\rangle + \langle c, \bar{x}\rangle$$

cette valeur est minorée par la valeur optimale du problème dual qui est finie car on a supposé l'existence d'un point faisable pour le dual. On a donc forcément que $\langle c, y'\rangle \leq 0$ sinon en faisant tendre $\alpha \to +\infty$ au-dessus on obtiendrait une contradiction. De même on prouve que $\langle b, \lambda'\rangle \leq 0$. Soit $(\bar{\lambda}, \bar{\mu})$ un point faisable pour le problème dual. Pour tout $\alpha > 0$, on a $\mathcal{A}^\top(\alpha\lambda' + \bar{\lambda}) - (\bar{\mu} + \mathcal{A}^\top(\alpha\lambda')) = c$ et $\bar{\mu} + \mathcal{A}^\top(\alpha\lambda') \in C^\circ$ car μ et $\mathcal{A}^\top(\lambda') \in C^\circ$. Alors $(\alpha\lambda' + \bar{\lambda}, \bar{\mu} + \mathcal{A}^\top(\alpha\lambda'))$ est faisable pour le problème dual. La valeur de la fonction objective du problème dual en ce point vaut

$$\langle b, \alpha\lambda' + \bar{\lambda}\rangle = \alpha\langle b, \lambda'\rangle + \langle b, \bar{\lambda}\rangle.$$

Cette valeur est majorée par la valeur optimale du problème primal par dualité faible et cette valeur est finie par l'existence d'un point faisable pour le problème primale. Ainsi on doit avoir $\langle b, \lambda'\rangle \leq 0$ sinon en faisant tendre $\alpha \to +\infty$, on aurait une contradiction. On en déduit que $\langle b, \lambda'\rangle \leq 0$ et $\langle c, y'\rangle \leq 0$ or $\langle b, \lambda'\rangle + \langle c, y'\rangle = 1$ ce qui constitue une contradiction.

Si $\tau > 0$: Alors $\lambda^* = \lambda'/\tau$ et $y^* = -y'/\tau$ sont tels que

$$\begin{cases} \mathcal{A}^\top(\lambda^*) - c \in C^\circ \\ \mathcal{A}(y^*) = b \\ y^* \in C \\ \langle b, \lambda^*\rangle - \langle c, y^*\rangle = 1/\tau \end{cases}$$

Si on pose $\mu^* = \mathcal{A}^\top(\lambda^*) - c$ alors y^* est faisable pour le primal et (λ^*, μ^*) est faisable pour le dual. En particulier, par dualité on a $\langle b, \lambda^*\rangle - \langle c, y^*\rangle \leq 0$ ce qui est une contradiction avec la dernière équation de (AS) qui impose $\langle b, \lambda^*\rangle - \langle c, y^*\rangle = 1/\tau$.

On en déduit donc que (AS) n'a pas de solution et par le Théorème 10.4 que (S) a une solution. Ce qui prouve le résultat sur l'existence d'une solution x^* au primal et (λ^*, μ^*) au dual tels que $\langle \lambda^*, b\rangle - \langle c, x^*\rangle = 0$. Il y a donc dualité forte. Ensuite, on peut appliquer de la première section sous dualité forte disant que si x^* est solution primale et (λ^*, μ^*) est solution duale alors forcément $(x^*, (\lambda^*, \mu^*))$ est un point-selle de la fonction de Lagrange et donc les valeurs des fonctions objectives coïncident en ces points càd $\langle \lambda^*, b\rangle = \langle c, x^*\rangle$. ■

Dualité forte pour les problèmes SDP. On considère dans cette section les problèmes d'optimisation de type SDP comme introduit dans la Définition 10.1. C'est un exemple de problème de type conic LP. Le but de cette section est d'obtenir un résultat de dualité forte pour les problèmes SDP et de montrer à quoi ils peuvent servir. On commence par identifier le problème dual d'un problème SDP.

On considère un problème de type SDP : soit $C \in \mathbb{R}^{n\times n}$, $A_1 \dots, A_m \in \mathbb{R}^{n\times n}$ et $b \in \mathbb{R}^m$. On note $\mathcal{A} : X \in \mathbb{R}^{n\times n} \to (\langle A_i, X\rangle)_{i=1}^m$. Un problème SDP est de la forme

$$\min\left(\langle C, X\rangle : X \succeq 0, \mathcal{A}(X) = b\right) \tag{10.10}$$

où on rappelle que $X \succeq 0$ signifie que X est positive. L'ensemble des matrices symétriques positives est un cône convexe fermé non vide. On retrouve donc un problème de type conic LP. Le problème dual associé est alors

$$\max\left(\langle \lambda, b\rangle : C - \mathcal{A}^\top(\lambda) \succeq 0\right) \tag{10.11}$$

où $\mathcal{A}^\top : \lambda \in \mathbb{R}^m \to \sum_i \lambda_i A_i$. La dualité faible dit que pour tout $\lambda \in \mathbb{R}^m$ tel que $C - \mathcal{A}^\top(\lambda) \succeq 0$ et pour tout $X \succeq 0$ tel que $\mathcal{A}(X) = b$, on a $\langle C, X\rangle \geq \langle \lambda, b\rangle$. Dans le théorème suivant, on identifie les situation où on peut avoir égalité $\langle C, X^*\rangle \geq \langle \lambda^*, b\rangle$ pour X^* solution primale et λ^* solution duale.

Théorème 10.7 *On suppose qu'il existe $X_0 \succ 0$ tel que $\mathcal{A}(X_0) = b$ et un $\lambda_0 \in \mathbb{R}^m$ tel que $C - \mathcal{A}^\top(\lambda_0) \succ 0$. Alors le problème SDP primal (10.10) admet une solution, le problème SDP dual (10.11) admet une solution, il y a dualité forte et pour toute solution X^* du primal et toute solution du dual λ^*, on a $\langle C, X^*\rangle = \langle b, \lambda^*\rangle$. Réciproquement, si X est faisable pour le primal (10.10) et λ est faisable pour le dual (10.11) et sont tels que $\langle C, X\rangle = \langle b, \lambda\rangle$ alors X est solution du primal et λ^* est solution du duale.*

Preuve. La première assertion est conséquence directe du résultat général de dualité forte sur les conic LP donné au Théorème 10.3. Pour la réciproque, on utilise la Proposition 1.7 du premier chapitre. ∎

Une application classique du Théorème 10.7 est la construction d'un **certificat dual**. Dans cette application, on n'utilise que la réciproque et donc seulement la Proposition 1.7 du premier chapitre. L'idée est qu'on a un problème initial (P_0) qu'on ne sait pas résoudre en général (par exemple, il est NP-hard). On 'convexifie' ce problème par relaxation convexe qui nous donne un autre problème, noté (P). Parfois cette relaxation donne un problème de type SDP. Dans ce cas, on peut considérer le problème dual (D) associé à ce SDP. Le but est de montrer que l'ensemble des solutions du problème initial (P_0) est aussi l'ensemble des solutions du problème relaché (P), on pourra ainsi résoudre (P_0) grâce à (P). Pour ce faire, on peut prendre une solution du problème initial, généralement c'est un unique vecteur x_0 et montrer que $X_0 = x_0 x_0^\top$ est l'unique solution du problème relaché (P). Pour cela, on cherche à construire un élément λ_0 faisable pour le dual, tel que $\langle C, X_0\rangle = \langle b, \lambda_0\rangle$. Ainsi, on sait que X_0 est solution de (P) : on dit que λ_0 a certifié X_0 comme étant solution de (P). C'est le principe du **dual certificate**. Toute la difficulté technique est dans la construction de λ_0 cependant, on fait en général de la 'retro engineering' en partant du fait que λ_0 doit à la fois être faisable pour le dual et aussi vérifier $\langle C, X_0\rangle = \langle b, \lambda_0\rangle$. On peut aussi construire λ_0 en cherchant à résoudre le problème dual vu qu'on est sûr que si λ_0 est faisable pour le dual et que $\langle C, X_0\rangle = \langle b, \lambda_0\rangle$ où X_0 est sensé être solution du primal alors nécessairement λ_0 est solution du dual.

Finalement, on peut voir que le problème dual (10.11) peut en fait se réécrire comme un problème de type conic LP (en changeant de variable $\lambda \to -\lambda$) :

$$\min_{(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^{n\times n}} \Big(\big\langle \lambda, b\big\rangle : \mu - \mathcal{A}^\top(\lambda) = C, \mu \succeq 0\Big). \tag{10.12}$$

C'est bien un problème de type conic LP en la variable $(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^{n\times n}$ où le cône convexe fermé et non vide est l'ensemble $\mathbb{R}^m\times\mathcal{S}_n^+$ où $\mathcal{S}_n^+$ est l'ensemble des matrices symétrique positives de $\mathbb{R}^{n\times n}$ et les contraintes d'égalité sont $C = \mu - \mathcal{A}^\top(\lambda) := \mathcal{A}_0(\lambda,\mu)$ où $\mathcal{A}_0 : (\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^{n\times n} \to \mu - \mathcal{A}^\top(\lambda)\in\mathbb{R}^{n\times n}$ est un opérateur qui s'écrit sous forme matricielle

$$\mathcal{A}_0 = \left(\begin{array}{c|c} -\mathcal{A}^\top & I_{n\times n}\end{array}\right)$$

où $I_{n\times n} : A\in\mathbb{R}^{n\times n}\to A$ est l'opérateur identité sur $\mathbb{R}^{n\times n}$ (ce n'est pas la matrice identité de $\mathbb{R}^{n\times n}$). On peut alors donner directement le problème dual associé à (10.12) :

$$\max\left(\big\langle Y, C\big\rangle : b + \mathcal{A}(Y) = 0, -Y \succeq 0\right). \tag{10.13}$$

En particulier, ce problème dual (du problème dual (10.11)) est équivalent au problème primal (10.10) (en effectuant le changement de variable $X = -Y$). Autrement dit le bidual Lagrangien est égal au primal pour les problème SDP. On a déjà vu ce phénomène pour les problèmes de Linear Programming. Ce n'est pas toujours le cas comme nous pouvons le voir dans ce qui suit pour le problème du MAX-CUT.

Problème bidual Lagrangien et 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

$$\min_{x\in\mathbb{R}^n}\Big(x^\top A x : x_i^2 = 1, i = 1,\dots,n\Big) \tag{10.14}$$

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 $x_i^2 = 1, i = 1,\dots,n$ 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

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

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

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

Le problème dual est donc le problème

$$\max_{u\in\mathbb{R}^n}\psi(u) = \max_{u\in\mathbb{R}^n}\left(\big\langle e,u\big\rangle : A - \mathrm{Diag}(u)\succeq 0\right). \tag{10.15}$$

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, (10.15) est déjà un problème dual (c'est le dual de MAX-CUT), on parle alors de **problème bidual**. La fonction dual de (10.15) est

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

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 (10.15) est

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

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

$$\min_{X \succeq 0} \psi'(-X) = \min_{X \succeq 0} \left(\langle X, A \rangle : X_{ii} = 1, i = 1, \dots, n \right) \tag{10.16}$$

Il est intéressant de remarquer que le problème bidual de MAX-CUT obtenu en (10.16) peut aussi être obtenu en faisant une relaxation convexe de l'ensemble des contraintes. En effet, une autre manière d'écrire le problème initial (10.14) est d'introduire le produit scalaire matriciel $\langle A, B \rangle = \sum A_{ij} B_{ij}$. On voit alors que $x^\top A x = \langle x x^\top, A \rangle$ où $x x^\top$ est une matrice symétrique de rang 1. On a alors

$$\min_{x \in \mathbb{R}^n} \left(x^\top A x : x_i^2 = 1, i = 1, \dots, n \right) = \min \left(\langle X, A \rangle : X_{ii} = 1, X = X^\top, \mathrm{rang}(X) = 1 \right). \tag{10.17}$$

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 (10.16) à partir de (10.17) en prenant une relaxation convexe de son ensemble de contraintes. La contrainte de rang : "$\mathrm{rang}(X) = 1$" 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 :

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

11 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 11.1 *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 $x^* \in \mathrm{argmin}_{x \in K} f(x)$ 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 11.2 (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

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

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

$$\min_{x \in U : F(x) \in C} f(x) \tag{11.1}$$

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 (11.1) 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 (11.1).

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

$$\sup_{\mu \in C^\circ} \inf_{x \in U} \mathcal{L}(x, \mu) \leq \inf_{x \in U} \sup_{\mu \in C^\circ} \mathcal{L}(x, \mu) \tag{11.2}$$

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 (11.1)).

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 11.3 *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*

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

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

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

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

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