Conditions nécessaires et suffisantes d'optimalité en optimisation convexe et différentiable

Guillaume Lecué[1]

En **optimisation convexe différentiable (OCD)**, on s'intéresse aux 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

$$K = \left\{ x \in U : \begin{array}{c} g_1(x) = \cdots = g_r(x) = 0 \\ h_1(x) \leq 0, \ldots, h_l(x) \leq 0 \end{array} \right\} \tag{0.1}$$

où les $g_1, \ldots, g_r : U \to \mathbb{R}$, définissant les contraintes d'égalité, sont des **fonctions affines** et les $h_1, \ldots, h_l : U \to \mathbb{R}$, définissant les contraintes d'inégalité, sont des fonctions **convexes de classe $\mathcal{C}^1$**.

Étant donné que les contraintes d'égalité sont affines, on peut trouver une matrice $A \in \mathbb{R}^{r\times n}$ et un vecteur $b \in \mathbb{R}^r$ tels que

$$\{x \in U : g_1(x) = \cdots = g_r(x) = 0\} = \{x \in U : Ax = b\}.$$

La contrainte K en OCD s'écrit alors sous la forme

$$K = \{x \in U : Ax = b \text{ et } h_1(x) \leq 0, \ldots, h_l(x) \leq 0\} \tag{0.2}$$

où U est un convexe ouvert et $h_1, \ldots, h_l$ sont convexes et de classe $\mathcal{C}^1$.

Le but de cette section est de donner des conditions nécessaires et suffisantes d'optimalité pour les problèmes de la forme $\min_{x\in K} f(x)$ en OCD. Contrairement au cadre (non forcément convexe) de l'OD pour lequel on avait seulement des conditions nécessaires pour l'optimalité, ici, en OCD, on obtiendra des conditions nécessaires et suffisantes (CNS). Par exemple, en OCD sans contrainte, être un point critique est une CNS, en OCD avec contrainte (sous hypothèse de qualification) être un point vérifiant les conditions KKT est une CNS d'optimalité du problème et, finalement, le duality gap est nul en dualité Lagrangienne sous hypothèse de qualification et donc le problème dual va permettre de nous aider à résoudre le problème primal.

1 Rappels sur la notion de convexité

Définition 1.1 *Soit C un sous-ensemble de $\mathbb{R}^n$. On dit que C est **convexe** quand pour tout $x, y \in C$ et $0 \leq \alpha \leq 1$, on a $\alpha x + (1-\alpha)y \in C$. Soit C un ensemble convexe non vide de $\mathbb{R}^n$ et $f : C \to \mathbb{R}$. On dit que f **est une fonction convexe** quand pour tout $x, y \in C$ et $0 \leq \alpha \leq 1$ on a $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$.*

Géométriquement, les notions de convexité d'un ensemble et d'une fonction se visualisent assez bien : un ensemble C est convexe si pour tout couples de points $x, y \in C$ le segment $[x, y]$ est dans

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

C et pour une fonction f, dire qu'elle est convexe revient à dire que son graphe $\{(x, f(x)) : x \in \mathbb{R}^2\}$ est tel que si on prend deux points dans son graphe alors le segment $[(x, f(x)), (y, f(y))]$ se trouve au dessus de son graphe.

On commence par un résultat utile pour définir des notions telles que "convexe engendré par un ensemble" ou encore d'enveloppe convexe d'une fonction.

Proposition 1.2 *Soit $(C_i)_{i\in I}$ une famille d'ensembles convexes de $\mathbb{R}^n$. Alors $\cap_{i\in I} C_i$ est un ensemble convexe.*

Preuve. Si $x, y \in \cup_{i\in I} C_i$ et $0 \leq \alpha \leq 1$ alors, pour tout $i \in I$, comme $x, y \in C_i$ et que C_i est convexe, on a $\alpha x + (1-\alpha)y \in C_i$. Ceci étant vrai pour tout $i \in I$, on a bien $\alpha x + (1-\alpha)y \in \cap_{i\in I} C_i$ et donc $\cap_{i\in I} C_i$ est convexe. ■

Proposition 1.2 dit que la propriété de convexité d'un ensemble est stable par passage à l'intersection. Ce type de résultat est particulièrement intéressant quand on veut définir des notions de "plus petit convexe contenant tel ensemble" : si $A \subset \mathbb{R}^n$, on définit $\mathrm{conv}(A)$ comme étant le plus petit ensemble convexe contenant A. Cette définition a bien un sens grâce à la Proposition 1.2 vu que si $\mathcal{C} = \{C \subset \mathbb{R}^n : A \subset C \text{ et } C \text{ est convexe}\}$ alors $\mathcal{C}$ est une famille d'ensembles convexes et donc $\cap_{C\in\mathcal{C}} C$ est un ensemble convexe (d'après la Proposition 1.2). Par ailleurs, $\cap_{C\in\mathcal{C}} C$ est le plus petit ensemble convexe contenant A car si $A \subset C$ et C est convexe on a $C \in \mathcal{C}$ et donc $\cap_{C\in\mathcal{C}} C \subset C$. On a donc $\mathrm{conv}(A) = \cap_{C\in\mathcal{C}} C$.

On peut définir la notion de fonction convexe à partir de la notion d'ensemble convexe quand on introduit l'épigraphe de f.

Proposition 1.3 *Soit U un ensemble convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$. **L'épigraphe de** f est défini par $\mathrm{Epi}(f) = \{(x, y) \in U \times \mathbb{R} : f(x) \leq y\}$. Il y a équivalence entre :*

1) $\mathrm{Epi}(f)$ est un ensemble convexe

2) f est convexe.

Preuve. On montre *1)* implique *2)* : Soit $x, y \in U$ et $0 \leq \alpha \leq 1$. Comme l'épigraphe de f est convexe, on a $\alpha(x, f(x)) + (1-\alpha)(y, f(y)) \in \mathrm{Epi}(f)$. Ainsi, $(\alpha x + (1-\alpha)y, \alpha f(x) + (1-\alpha)f(y)) \in \mathrm{Epi}(f)$ et par définition de l'épigraphe, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$.

On montre *2)* implique *1)* : Soit $(x_1, y_1), (x_2, y_2) \in \mathrm{Epi}(f)$ et $0 \leq \alpha \leq 1$. On a

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) \leq \alpha y_1 + (1-\alpha)y_2$$

donc $(\alpha x_1 + (1-\alpha)x_2, \alpha y_1 + (1-\alpha)y_2) \in \mathrm{Epi}(f)$ et donc $\alpha(x_1, y_1) + (1-\alpha)(x_2, y_2) \in \mathrm{Epi}(f)$. On a donc bien que $\mathrm{Epi}(f)$ est convexe. ■

On donne finalement un exemple de fonction convexe qui nous sera utile en dualité Lagrangienne.

Proposition 1.4 *Soit $(L(\cdot, u))_{u\in U}$ une famille de fonction linéaires (càd pour tout $u \in U$, $x \to L(x, u)$ est une fonction linéaire). Alors $f : x \to \sup_{u\in U} L(x, u)$ est une fonction convexe.*

Preuve. Soit x, y et $\alpha \in [0, 1]$. On a pour tout $u \in U$, $L(\alpha x + (1-\alpha)y, u) = \alpha L(x, u) + (1-\alpha)L(y, u)$ et donc

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \sup_{u\in U} L(\alpha x + (1-\alpha)y, u) = \sup_{u\in U} \left(\alpha L(x, u) + (1-\alpha)L(y, u)\right) \\ &\leq \alpha \sup_{u\in U} L(x, u) + (1-\alpha) \sup_{u\in U} L(y, u) = \alpha f(x) + (1-\alpha)f(y). \end{aligned}$$

En fait, il y a une réciproque à ce résultat qui est que toute fonction convexe peut s'écrire comme le sup d'une famille de fonctions linéaires. Dans le cas différentiable, on le verra dans la suite où on montrera qu'une fonction est convexe si et seulement si elle est au-dessus de ses tangentes. On pourra alors prendre le sup de la famille des tangentes pour retrouver la fonction.

1.1 Fonctions convexes de $\mathbb{R}$

Pour les fonctions régulières, on peut lier la notion de convexité à des propriétés de ses dérivées première et seconde. C'est le but de cette section, de démontrer ces liens.

Proposition 1.5 (Lemme des trois pentes) *Soit U un convexe de $\mathbb{R}$. Soit $x, y, z \in U$ tels que $x \leq y \leq z$ alors*

$$\frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(z) - \varphi(x)}{z - x} \leq \frac{\varphi(z) - \varphi(y)}{z - y}.$$

Preuve. Comme $y \in [x, z]$, il existe $0 \leq \alpha \leq 1$ tel que $y = \alpha x + (1 - \alpha)z$. On peut prendre

$$\alpha = \frac{y - z}{x - z}. \tag{1.1}$$

Par convexité de φ, on a

$$\varphi(y) = \varphi(\alpha x + (1 - \alpha)z) \leq \alpha\varphi(x) + (1 - \alpha)\varphi(z)$$

et donc

$$\varphi(y) - \varphi(z) \leq \alpha(\varphi(x) - \varphi(z)).$$

En remplaçant α par sa valeur (1.1) dans la dernière inégalité, on obtient

$$\frac{\varphi(z) - \varphi(x)}{z - x} \leq \frac{\varphi(z) - \varphi(y)}{z - y}.$$

De même, on peut écrire $y = tz + (1 - t)x$ où $t = (y - x)/(z - x)$. Par convexité de φ, on a

$$\varphi(y) \leq t\varphi(z) + (1 - t)\varphi(x)$$

donc $\varphi(y) - \varphi(x) \leq t(\varphi(z) - \varphi(x))$. Finalement, en utilisant que $t = (y - x)/(z - x)$ dans la dernière expression, on obtient que

$$\frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(z) - \varphi(x)}{z - x}.$$

■

Proposition 1.6 *Soit U un ouvert convexe de $\mathbb{R}$ et $\varphi : U \to \mathbb{R}$ une fonction dérivable. On a équivalence entre :*

1) φ' est croissante

2) φ est convexe.

Preuve. On montre que *1)* implique *2)* : Soit $x, z \in U$ tel que $x < z$ et $0 \leq \alpha \leq 1$. On note $y = \alpha x + (1 - \alpha)y$. On a

$$\varphi(y) - \varphi(x) = \int_x^y \varphi'(t)dt \leq \varphi'(y)(y - x) = \varphi'(y)(1 - t)(z - x)$$

et

$$\varphi(z) - \varphi(y) = \int_y^z \varphi'(t)dt \geq \varphi'(y)(z-y) = \varphi'(y)t(z-x).$$

On obtient alors

$$\frac{\varphi(y) - \varphi(x)}{(1-t)(z-x)} \leq \varphi'(y) \leq \frac{\varphi(z) - \varphi(y)}{t(z-x)}$$

donc

$$\frac{\varphi(y) - \varphi(x)}{1-t} \leq \frac{\varphi(z) - \varphi(y)}{t}$$

càd

$$\varphi(y) \leq t\varphi(x) + (1-t)\varphi(z).$$

On montre que *2)* implique *1)* : Soit $x, z \in U$ tels que $x \leq z$. Comme φ est convexe, elle vérifie le lemme des 3 pentes : pour tout $y \in [x, z]$,

$$\frac{\varphi(y) - \varphi(x)}{y - x} \overset{(a)}{\leq} \frac{\varphi(z) - \varphi(x)}{z - x} \overset{(b)}{\leq} \frac{\varphi(z) - \varphi(y)}{z - y}.$$

On fait tendre $y \to x$ dans *(a)* : comme

$$\varphi'(x) = \lim_{y \to x} \frac{\varphi(y) - \varphi(x)}{y - x},$$

on en déduit que

$$\varphi'(x) \leq \frac{\varphi(z) - \varphi(x)}{z - x}.$$

De même, on fait tendre $y \to z$ dans *(b)* pour obtenir

$$\frac{\varphi(z) - \varphi(x)}{z - x} \leq \varphi'(z).$$

On conclut que $\varphi'(x) \leq \varphi'(z)$. Ceci étant vrai pour tout $x \leq z$ dans U, on a donc bien que φ' est croissante sur U. ■

Corollaire 1.7 *Soit U un ouvert convexe de $\mathbb{R}^n$ et $\varphi : U \to \mathbb{R}$ une fonction de classe $\mathcal{C}^2$. Il y a équivalence entre :*

1) $\varphi''(x) \geq 0$ pour tout $x \in U$

2) φ est convexe.

Preuve. Pour une fonction de classe $\mathcal{C}^2$, on sait que φ' est croissante si et seulement si $\varphi''(x) \geq 0$ pour tout $x \in U$. On obtient donc le résultat en utilisant la Proposition 1.6. ■

1.2 Fonctions convexes de $\mathbb{R}^n$

On peut caractériser la convexité des fonctions régulières de la manière suivante.

Proposition 1.8 *Soit U un ensemble convexe de $\mathbb{R}^n$. Soit $f : U \to \mathbb{R}$. On suppose que f est de différentiable. Il y a équivalence entre :*

1) f est convexe

2) pour tout $x, y \in U$, $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$

3) pour tout $x, y \in U$, $\big\langle \nabla f(x) - \nabla f(y), x - y \big\rangle \geq 0$.
Si f est deux fois différentiable, il y a équivalence entre :
1) f est convexe
4) pour tout $x \in U, \nabla^2 f(x) \succeq 0$.

Preuve. On montre que *1)* implique *2)* : Pour tout $x, y \in \mathbb{R}^n$ et $0 < t \leq 1$, on a par convexité

$$f(x) - f(y) \geq \frac{f(y + t(x-y)) - f(y)}{t}$$

et donc quand $t \to 0$, on obtient $f(x) - f(y) \geq \big\langle \nabla f(y), x - y \big\rangle$.

On montre que *2)* implique *3)* : On a

$$\begin{aligned} f(y) - f(x) &\geq \big\langle \nabla f(x), y - x \big\rangle \\ f(x) - f(y) &\geq \big\langle \nabla f(y), x - y \big\rangle \end{aligned}$$

et donc en sommant les deux inégalités, on obtient le résultat.

On montre que *3)* implique *2)* : D'après la formule de Taylor-Lagrange à l'ordre 1 appliquée à l'application $g : t \in [0,1] \to f(x + t(y-x))$, il existe $t \in [0,1]$ tel que

$$g(1) = g(0) + (1-0)g'(t).$$

Comme $g(1) = f(y)$, $g(0) = f(x)$ et $g'(t) = \big\langle \nabla f(x + t(y-x)), y - x \big\rangle$, on obtient que

$$\begin{aligned} f(y) &= f(x) + \big\langle \nabla f(x + t(y-x)), y - x \big\rangle \\ &= f(x) + \big\langle \nabla f(x), y - x \big\rangle + \big\langle \nabla f(x + t(y-x)) - \nabla f(x), y - x \big\rangle \\ &\geq f(x) + \big\langle \nabla f(x), y - x \big\rangle \end{aligned}$$

car $\big\langle \nabla f(x + t(y-x)) - \nabla f(x), y - x \big\rangle \geq 0$.

On montre que *2)* implique *1)* : On note $x_t = x + t(y-x)$. On a

$$\begin{aligned} f(x) &\geq f(x_t) + \big\langle \nabla f(x_t), x - x_t \big\rangle \\ f(y) &\geq f(x_t) + \big\langle \nabla f(x_t), y - x_t \big\rangle \end{aligned}$$

et comme $x - x_t = -t(y-x)$ et $y - x_t = (1-t)(y-x)$, on obtient

$$\begin{aligned} f(x) &\geq f(x_t) - t\big\langle \nabla f(x_t), y - x \big\rangle \\ f(y) &\geq f(x_t) + (1-t)\big\langle \nabla f(x_t), y - x \big\rangle. \end{aligned}$$

En multipliant la première inégalité par $(1-t)$ et la deuxième par t et en sommant, on obtient

$$(1-t)f(x) + tf(y) \geq f(x_t) = f((1-t)x + ty).$$

On montre que *1)* implique *4)* : Comme *1)* implique *3)*, on a pour tout $0 < t \leq 1$,

$$\big\langle \nabla f(x + t(y-x)) - \nabla f(x), t(y-x) \big\rangle \geq 0$$

alors, en divisant par t et en passant à la limite quand $t \to 0$, on obtient

$$\big\langle \nabla^2 f(x)(y-x), y - x \big\rangle \geq 0.$$

Ceci étant vrai pour tout $x, y \in U$ et U étant ouvert, on a bien que $\nabla^2 f(x) \succeq 0$.

On montre que *4)* implique *1)* : D'après la formule de Taylor-Lagrange à l'ordre 2 appliquée à l'application $t \in [0,1] \to f(x+t(y-x))$, il existe $t \in [0,1]$ tel que

$$g(1) = g(0) + (1-0)g'(0) + \frac{(1-0)^2}{2!} g''(t).$$

Comme $g(1) = f(y)$, $g(0) = f(x)$, $g'(t) = \langle \nabla f(x+t(y-x)), y-x \rangle$ et $g''(t) = \langle \nabla^2 f(x+t(y-x))(y-x), y-x \rangle$ on obtient que

$$f(y) = f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2}(y-x)^\top \nabla^2 f(x+t(y-x))(x-y)$$

et comme $\nabla^2 f(x+t(y-x)) \succeq 0$, *2)* est vérifiée et donc *1)* aussi. ∎

Les exemples classiques de fonctions convexes sont $f : x \to x^2$ ou encore $f : x \to |x|$. Néanmoins, la fonction $f : x \to x^2$ semble en quelque sorte "plus convexe" que $f : x \to |x|$ qui est en fait seulement linéaire par morceau et donc sans "courbure". Pour formaliser cette intuition, on introduit les notions de fonctions strictement et fortement convexes.

Définition 1.9 *Soit U un ensemble convexe de $\mathbb{R}^n$. Soit $f : U \to \mathbb{R}$. On dit que f est **strictement convexe** quand pour tout $x \neq y \in U$ et $0 < \alpha < 1$, on a*

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y).$$

*On dit que f est **fortement convexe** quand il existe $c > 0$ tel que pour tout $x, y \in U$ et $0 \leq \alpha \leq 1$, on a*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{c\alpha(1-\alpha)}{2} \|x-y\|_2^2 . \tag{1.2}$$

*On dit aussi que f est c-**convexe** ou c-**elliptique** lorsque (1.2) est satisfaite.*

La fonction $x \to x^2$ est fortement convexe alors que $x \to |x|$ ne l'est pas. La fonction $x > 0 :\to -\log(x)$ est strictement convexe mais pas fortement convexe. La stricte convexité implique l'unicité des solutions aux problèmes d'optimisation.

Proposition 1.10 *Soit U un ensemble convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une fonction strictement convexe. Soit K un sous-ensemble convexe de U. Le problème $\min_{x\in K} f(x)$ admet au plus une solution.*

Preuve. Soit x_1^* et x_2^* sont deux solutions du problème $\min_{x\in K} f(x)$. On pose $f^* = \min_{x\in K} f(x)$. Supposons que $x_1^* \neq x_2^*$. D'après la stricte convexité, on a pour tout $0 < \alpha < 1$, $f(\alpha x_1^* + (1-\alpha)x_2^*) < \alpha f(x_1^*) + (1-\alpha) f(x_2^*) = \alpha$. Or $\alpha x_1^* + (1-\alpha)x_2^* \in K$ donc $f(\alpha x_1^* + (1-\alpha)x_2^*) \geq \min_{x\in K} f(x) = f^*$. C'est une contradiction, donc $x_1^* = x_2^*$. ∎

La Proposition 1.10 n'implique pas l'existence d'une solution au problème $\min_{x\in K} f(x)$ seulement son unicité si elle existe. Néanmoins, si on suppose de plus que f est fortement convexe sur $\mathbb{R}^n$ et que K est fermé alors on a existence et unicité d'une solution au problème $\min_{x\in K} f(x)$.

Proposition 1.11 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction fortement convexe. Soit K un sous-ensemble convexe et fermé de $\mathbb{R}^n$. Le problème $\min_{x\in K} f(x)$ admet une unique solution.*

Preuve. Pour l'unicité, on applique la Proposition 1.10 vu qu'une fonction fortement convexe et aussi strictement convexe. Pour l'existence, on voit que f est coercive car fortement convexe et comme K est fermé, on conclut avec Weierstrass (voir la Proposition 1.5 du deuxième chapitre). Pour la coercivité de f, on peut par exemple utiliser la Proposition 1.13 suivante pour voir que pour tout $x \in \mathbb{R}^n$, on a $f(x) \geq f(0) + \langle \nabla f(0), x \rangle + (c/2) \|x\|_2^2$ et donc $f(x) \to +\infty$ quand $\|x\|_2 \to +\infty$. ∎

On peut caractériser la convexité forte de la manière suivante en montrant que même si on retranche $c/2$ fois la norme ℓ_2^n au carré à f alors elle reste convexe.

Proposition 1.12 *Soit U un convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$. Soit $c > 0$. Il y a équivalence entre :*
1) f est c-convexe
2) $x \in U \to f(x) - (c/2) \|x\|_2^2$ est convexe

Preuve. On montre que *1)* implique *2)* : On note $g : x \in U \to f(x) - (c/2) \|x\|_2^2$. On a pour tout $x, y \in U$ et $0 \le t \le 1$,

$$\begin{aligned}\|tx + (1-t)y\|_2^2 &= t^2 \|x\|_2^2 + (1-t)^2 \|y\|_2^2 + 2t(1-t)\langle x, y\rangle \\ &= t^2 \|x\|_2^2 + (1-t)^2 \|y\|_2^2 - t(1-t)\left(\|x-y\|_2^2 - \|x\|_2^2 - \|y\|_2^2\right) \\ &= t\|x\|_2^2 + (1-t)\|y\|_2^2 - t(1-t)\|x-y\|_2^2\end{aligned}$$

(au passage, on voit que $x \to \|x\|_2^2$ est 2-convexe) et donc

$$\begin{aligned}&g(tx + (1-t)y) - tg(x) - (1-t)g(y) \\ &\qquad = f(tx + (1-t)y) - tf(x) - (1-t)f(y) - \frac{ct(1-t)}{2}\|x-y\|_2^2 \le 0.\end{aligned} \tag{1.3}$$

On montre que *2)* implique *1)* : On note $g : x \in U \to f(x) - (c/2) \|x\|_2^2$. Le résultat s'obtient à partir de (1.3). ■

On peut caractériser la convexité forte de la manière suivante quand f a de la régularité supplémentaire.

Proposition 1.13 *Soit U un ouvert convexe et $f : U \to \mathbb{R}$ différentiable. Soit $c > 0$. Il y a équivalence entre :*
1) f est c-convexe
2) pour tout $x, y \in U$,

$$f(x) \ge f(y) + \langle \nabla f(y), x - y\rangle + (c/2)\|x-y\|_2^2$$

3) pour tout $x, y \in U$,

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge c\|x-y\|_2^2.$$

Si f est supposée deux fois différentiable, alors il y a équivalence entre les trois propriétés ci-dessus et
4) pour tout $x \in U$, $\nabla^2 f(x) \succeq cI_n$.

Preuve. On peut démontrer la Proposition 1.13 grâce à la Proposition 1.12 et la Proposition 1.8 appliquée à $g : x \in U \to f(x) - (c/2)\|x\|_2^2$. Néanmoins, on donne une preuve directe pour certaine implications ci-dessous.

On montre que *1)* implique *2)* : Pour tout $x, y \in U$ et $0 < \alpha \le 1$, on a

$$f(x) - f(y) - \frac{c(1-\alpha)}{2}\|x-y\|_2^2 \ge \frac{f(y + \alpha(x-y)) - f(y)}{\alpha}.$$

En faisant tendre $\alpha \to 0$, on obtient,

$$f(x) - f(y) - \frac{c}{2}\|x-y\|_2^2 \ge \langle \nabla f(y), x - y\rangle.$$

On montre que *2)* implique *3)* : pour tout $x, y \in U$, on a d'après *2)*,

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle + (c/2) \|x - y\|_2^2$$
$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle + (c/2) \|x - y\|_2^2 .$$

Alors en sommant ces deux inégalités, on obtient bien le résultat.

On montre que *3)* implique *1)* : On note $g : x \in U \to f(x) - (c/2) \|x\|_2^2$. Pour tout $x, y \in U$, on a

$$\langle \nabla g(x) - \nabla g(y), x - y \rangle = \langle \nabla f(x) - \nabla f(y), x - y \rangle - c \|x - y\|_2^2 \geq 0$$

donc g est convexe et on conclut avec la Proposition 1.12.

Finalement, quand on suppose de plus que f est deux fois différentiable, comme la Hessienne de $g : x \in U \to f(x) - (c/2) \|x\|_2^2$ est $\nabla g(x) = \nabla f(x) - cI_n$ et que g est convexe si et seulement si $\nabla g(x) \succeq 0$ pour tout $x \in U$, on a bien l'équivalence entre *1)* et *4)* au vue de la Proposition 1.12.

■

Il est très utile d'étudier les propriétés de contraction de certaines fonctions quand on étudie la convergence d'algorithmes de points fixes comme nous le ferons plus tard. La forte convexité implique cette propriété de contraction pour une certaine fonction très utilisée par les algorithmes de descente. On rappelle qu'une fonction $F : U \to \mathbb{R}$ est une **contraction** quand elle est Lipschitzienne de constante de Lipschitz strictement plus petite que 1.

Proposition 1.14 *Soit U un ouvert convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une fonction c-convexe différentiable. On suppose que ∇f est C-Lipschitzien (i.e. $\|\nabla f(x) - \nabla f(y)\|_2 \leq C \|x - y\|_2$ pour tout $x, y \in U$). La fonction $F : x \in U \to x - \eta \nabla f(x)$ est Lipschitzienne de constante de Lipschitz $\sqrt{1 + \eta^2 C^2 - 2c\eta}$. En particulier, F est une contraction quand $0 < \eta \leq 2c/C^2$.*

Preuve. Comme le gradient est Lipschitzien et que f est c-convexe, d'après la Proposition 1.12, pour tout $x, y \in U$, on a

$$\|(x - \eta \nabla f(x)) - (y - \eta \nabla f(y))\|_2^2 = \|x - y\|_2^2 + \eta^2 \|\nabla f(x) - \nabla f(y)\| - 2\eta \langle x - y, \nabla f(x) - \nabla f(y) \rangle$$
$$\left(1 + \eta^2 C^2 - 2c\eta\right) \|x - y\|_2^2 .$$

■

Remarque 1.15 *Si $f : U \to \mathbb{R}$ une fonction c-convexe différentiable et ∇f est C-Lipschitzien alors d'après la Proposition 1.12, on a pour tout $x, y \in U$,*

$$c \|x - y\|_2^2 \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle \leq \|\nabla f(x) - \nabla f(y)\|_2 \|x - y\|_2 \leq C \|x - y\|_2^2$$

et donc $C \geq c$. En particulier, $1 + \eta^2 C^2 - 2c\eta \geq 0$ pour tout $\eta \in \mathbb{R}$.

2 Conditions nécessaires et suffisantes d'optimalité pour les problèmes d'optimisation convexes et différentiables

Dans cette section, on donne la "version OCD" du théorème d'Euler/Pénao/Kantorovitch. Contrairement au cas général (sans l'hypothèse de convexité), on a ici une condition nécessaire *et suffisante* d'optimalité. C'est un des avantages théoriques de l'optimisation convexe (on verra aussi des avantages pratiques à la convexité plus tard).

2.1 Problèmes d'optimisation sans contrainte en OCD

On commence ici avec les problèmes d'optimisation sans contrainte même si ce résultat peut être obtenu comme corollaire du Théorème d'Euler/Pénao/Kantorovitch de la section suivante.

Théorème 2.1 *Soit U un ouvert convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une fonction convexe et différentiable. Il y a équivalence entre :*
1) $x^* \in \mathrm{argmin}_{x\in U} f(x)$
2) $\nabla f(x^*) = 0$.

Preuve. On montre que *1)* implique *2)* : On a déjà vu cette implication dans le cas général (pas forcément convexe). On sait en effet que si x^* est solution du problème $\min_{x\in U} f(x)$ alors c'est un minimum local de f et donc c'est un point critique. On donne à nouveau la preuve ici.

Comme U est ouvert, il existe $\epsilon > 0$ tel que la boule $B_2(x^*, \epsilon)$ centrée en x^* et de rayon ϵ est dans U. Soit $d \in \mathbb{R}^n$. D'un côté, pour tout $0 < \lambda < \epsilon$, on a $f(x^* + \lambda d) \geq f(x^*)$ et d'un autre côté, quand $\lambda \to 0$, $f(x^* + \lambda d) = f(x^*) + \langle \nabla f(x^*), \lambda d\rangle + o(\lambda)$. On a donc quand $\lambda \to 0$, $\langle \nabla f(x^*), d\rangle + o(1) \geq 0$ donc $\langle \nabla f(x^*), d\rangle \geq 0$. Ceci étant vrai pour tout $d \in \mathbb{R}^n$, on en déduit que $\nabla f(x^*) = 0$.

On montre que *2)* implique *1)* : C'est cette réciproque qui est propre à la convexité de f. Comme f est convexe, d'après la Proposition 1.8, pour tout $y \in U$, on a $f(y) \geq f(x^*) + \langle \nabla f(x^*), y - x\rangle$. Ainsi, quand on a $\nabla f(x^*) = 0$, on a bien $f(y) \geq f(x^*)$ pour tout $y \in U$ et donc x^* est solution du problème $\min_{x\in U} f(x)$. ■

Sous hypothèse de convexité de f, le Théorème 2.1 assure bien que x^* est solution du problème $\min_{x\in U} f(x)$ si et seulement si x^* est un point critique de f. Pour trouver une solution du problème $\min_{x\in U} f(x)$, il suffit donc de chercher les points critiques de f. Néanmoins, le Théorème 2.1 n'assure pas l'existence ni l'unicité d'une solution, il se peut très bien que f n'atteigne aucun minimimum sur U ou que f atteigne son minimum en plusieurs points de U.

2.2 Problèmes d'optimisation avec contrainte en OCD

On donne ici la "version OCD" du théorème d'Euler/Pénao/Kantorovitch. On aura comme précédemment une CNS pour l'optimalité d'une solution au problème $\min_{x\in K} f(x)$.

Théorème 2.2 (Théorème d'Euler/Péano/Kantorovitch en OCD) *Soit U un ouvert convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une fonction convexe et différentiable. Soit $K \subset U$ un ensemble convexe. Il y a équivalence entre :*
1) $x^* \in \mathrm{argmin}_{x\in K} f(x)$
2) $-\nabla f(x^*) \in N_K(x^*)$
3) pour tout $y \in K$, $\langle \nabla f(x^*), y - x^*\rangle \geq 0$ *(condition du premier ordre, voir Figure 2.2).*

Preuve. L'équivalence entre *2)* et *3)* est immédiate vue que K est convexe et donc

$$N_K(x^*) = \{z \in \mathbb{R}^n : \langle z, y - x^*\rangle \leq 0, \forall y \in K\}.$$

On montre que *1)* implique *2)* : On connaît déjà cette implication dans le cas général. On démontre à nouveau ce résultat ici. Soit d un vecteur unitaire tangent à K en x^*. Soit $(x_n)_n$ une suite d'éléments de K convergeant vers x^* tel que $((x_n - x^*)/\|x_n - x^*\|_2)_n$ converge vers d. On a $f(x_n) = f(x^*) + \langle \nabla f(x^*), x_n - x^*\rangle + o(\|x_n - x^*\|_2)$ quand $n \to +\infty$. Par ailleurs, $f(x_n) \geq f(x^*)$ pour tout n, donc $\langle \nabla f(x^*), x_n - x^*\rangle + o(\|x_n - x^*\|_2) \geq 0$ quand $n \to 0$. On a donc $\langle \nabla f(x^*), (x_n - x^*)/\|x_n - x^*\|_2\rangle \to 0$ quand $n \to \infty$ càd $\langle \nabla f(x^*), d\rangle \geq 0$. Ceci étant vrai pour tout vecteur unitaire tangent à K en x^*, c'est aussi vrai pour le cône engendré par ces vecteurs, càd par $T_K(x^*)$. On a donc bien $-\nabla f(x^*) \in (T_K(x^*))^\circ = N_K(x^*)$.

On montre que *2)* implique *1)* : Comme K est convexe, on sait que $-\nabla f(x^*) \in N_K(x^*)$ si et seulement si pour tout $y \in K$, on a $\left\langle -\nabla f(x^*), y - x^* \right\rangle \leq 0$. Par ailleurs, f est convexe donc pour tout $y \in K$, $f(y) \geq f(x^*) + \left\langle \nabla f(x^*), y - x^* \right\rangle$. On en déduit que $f(y) \geq f(x^*)$.

∎

Remarque 2.3 *On peut déduire immédiatement le Théorème 2.1 du Théorème 2.2 vu que si $K = U$ alors comme U est ouvert, on a $N_K(x^*) = \{0\}$ et donc dire que $-\nabla f(x^*) \in N_K(x^*)$ est dans ce cas équivalent à dire que $\nabla f(x^*) = 0$.*

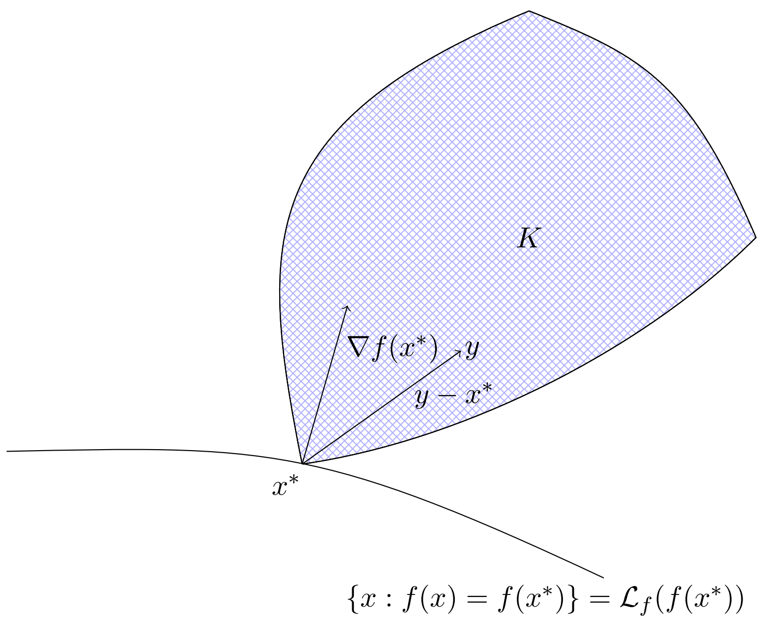

FIGURE 1 – Interprétation géométrique de la condition du premier ordre du théorème d'Euler/Péano/Kantorovitch.

2.3 Exemple : projection sur un convexe fermé

Un exemple d'application classique et très utile du théorème de EPK dans le cas convexe est celui de la projection sur un ensemble convexe fermé. Pour cela, on considère un ensemble $K \subset \mathbb{R}^n$ convexe, fermé et non vide. Pour tout $x \in \mathbb{R}^n$ la fonction $f : z \in \mathbb{R}^n \to (1/2) \|z - x\|_2^2$ est fortement convexe (car sa Hessienne est I_n qui vérifie bien la condition *4)* de la Proposition 1.13) et K est convexe fermé non vide donc $\min_{z \in K} f(z)$ admet une unique solution (voir Proposition 1.11). On peut donc définir la fonction

$$\mathrm{proj}_K : \left\{ \begin{array}{ccc} \mathbb{R}^n & \longrightarrow & K \\ x & \longrightarrow & \mathrm{argmin}_{z \in K} \|z - x\|_2 \, . \end{array} \right.$$

qui est appelé **opérateur de projection sur** K.

On remarque que minimiser $f : z \in \mathbb{R}^n \to (1/2) \|z - x\|_2^2$ ou $z \in \mathbb{R}^n \to \|z - x\|_2$ est exactement la même chose. Néanmoins, il est plus facile de travailler avec f car c'est une fonction quadratique. Pour tout $x \in \mathbb{R}^n$, le problème $\min_{z \in K} f(z)$ entre dans le domaine d'application du

Théorème d'EPK dans la cas convexe rappelé dans le Théorème 2.2 car f est convexe différentiable et K est convexe. On a donc équivalence entre :

1) $x^* = \text{proj}_K(x)$
2) pour tout $y \in K$, $\langle y - x^*, x - x^* \rangle \leq 0$.

On peut donc caractériser la projection sur un ensemble convexe par la condition du premier ordre qui est ici donnée en *2)*.

Exemple : projection sur un espace affine. On s'intéresse ici à l'opérateur de projection précédent dans le cas particulier où $K = \{y \in \mathbb{R}^n : Ay = b\}$ où $A \in \mathbb{R}^{m \times n}$ et $b \in \mathbb{R}^m$. On suppose que $K \neq \emptyset$, ce qui équivaut à dire que $b \in \text{Im}(A)$.

On vérifie que K est bien un convexe fermé. Comme $x^* \in K$, la condition du premier ordre s'écrit ici pour tout $h \in \ker(A)$, $\langle h, x - x^* \rangle = 0$ et donc $x - x^* \in (\ker(A))^\perp$. Par ailleurs, on a $(\ker(A))^\perp = \text{Im}(A^\top)$ (en effet, $z \in \ker(A)$ si et seulement si $Az = 0$ ssi $\forall y \in \mathbb{R}^n, \langle Az, y \rangle = 0$ ssi $\forall y \in \mathbb{R}^n, \langle z, A^\top y \rangle = 0$ ssi $z \in (\text{Im}(A^\top))^\perp$). On a donc $x^* = \text{proj}_K(x)$ ssi $x - x^* \in \text{Im}(A^\top)$ càd $x - x^* = A^\top \lambda$ pour un certain $\lambda \in \mathbb{R}^m$.

Pour déterminer un λ tel que $x^* = x - A^\top \lambda$, on utilise la contrainte que $x^* \in K$ càd $Ax^* = b$ donc $A(x - A^\top \lambda) = b$ càd $AA^\top \lambda = Ax - b$. On aimerait pouvoir inverser $AA^\top$ pour pouvoir déterminer un tel λ. Hors, en général $AA^\top$ n'est pas inversible (en fait $\ker(AA^\top) = \ker(A)$), il y a potentiellement une infinité de solution (en λ) à l'équation $AA^\top \lambda = Ax - b$. On va donc en déterminer une et montrer ensuite que pour cette solution on a bien su caractériser la projection sur K de x.

Pour cela, on considère l'**inverse généralisée** de $AA^\top$, notée $(AA^\top)^{(-1)}$. On rappelle sa définition : comme $AA^\top$ est une matrice symétrique positive, on peut l'écrire sous la forme $AA^\top = PDP^\top$ où $P \in O(n)$ est une matrice orthogonale et $D = \text{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0)$ où $\sigma_1 \geq \cdots, \geq \sigma_r > 0$. On pose $(AA^\top)^{(-1)} = PD^{(-1)}P^\top$ où $D^{(-1)} = \text{diag}(\sigma_1^{-1}, \cdots, \sigma_r^{-1}, 0, \cdots, 0))$. L'inverse généralisée permet en quelque sorte d'inverser $AA^\top$ uniquement là où elle est inversible càd sur $\text{Im}(AA^\top) = \text{Im}(A)$. On a en effet, pour tout $z \in \text{Im}(A), AA^\top(AA^\top)^{(-1)}z = z$. Pour voir cela, on montre que si $p_1, \cdots, p_r$ sont les vecteur propres de $AA^\top$ associés aux valeurs propres $\sigma_1, \ldots, \sigma_r$ formant les r premiers vecteurs colonne de $AA^\top$ alors $\text{Im}(A) = \text{vect}(p_1, \ldots, p_r)$ (car $AA^\top p_j = \sigma_j p_j$ donc $p_j \in \text{Im}(A)$ et comme $\text{rg}(A) = \text{rg}(AA^\top) = r$, on a bien trouvé une base orthonormale de $\text{Im}(A)$ avec $p_1, \ldots, p_r$). Si $z \in \text{Im}(A)$, on peut alors trouver $a_1, \ldots a_r$ des réels tels que $z = \sum_{j=1}^r a_j p_j$. On a donc

$$(AA^\top)(AA^\top)^{(-1)}z = PDP^\top(PD^{(-1)})P^\top z = PI_rP^\top \sum_{j=1}^r a_j p_j = \sum_{j=1}^r a_j p_j = z$$

où on a utiliser le fait que $p_1, \ldots, p_r$ sont les r premiers vecteurs colonnes de P et qu'ils sont orthonormaux.

On pose alors $x^* = x - A^\top(AA^\top)^{(-1)}(Ax - b)$. On a bien $x - x^* \in \text{Im}(A^\top)$, donc la condition du premier ordre est vérifiée et, comme $Ax - b \in \text{Im}(A)$, on a $(AA^\top)(AA^\top)^{(-1)}(Ax - b) = Ax - b$ et donc $Ax^* = Ax - Ax + b = b$ càd $x^* \in K$. Alors, en utilisant la réciproque du théorème d'Euler/Péano/Kantorovitch, on a $x^* = \text{proj}_K(x)$. On a donc pour tout $x \in \mathbb{R}^n$,

$$\text{proj}_K(x) = x - A^\top(AA^\top)^{(-1)}(Ax - b).$$

3 Conditions KKT en optimisation convexe différentiable

Pour les problèmes d'OCD la contrainte est de la forme

$$K = \{x \in U : Ax = b \text{ et } h_1(x) \leq 0, \ldots, h_l(x) \leq 0\} \tag{3.1}$$

où U est un convexe ouvert, $A \in \mathbb{R}^{r\times n}, b \in \mathbb{R}^r$ et $h_1,\ldots,h_l$ sont convexes et de classe $\mathcal{C}^1$. L'hypothèse de qualification de K en un point $x \in K$ s'écrit donc ici : K est qualifiée en x si et seulement si

$$T_K(x) = \big\{v \in \mathbb{R}^n : Av = 0, \big\langle \nabla h_j(x), v\big\rangle \leq 0, \forall j \in J(x)\big\} \tag{3.2}$$

où $J(x) = \{j \in \{1,\ldots,l\} : h_j(v) = 0\}$.

Pour les problèmes d'OCD, une des conditions suffisantes de qualification la plus utilisée est la condition de Slater : pour une contrainte K de la forme (3.1) on dit que K satisfait la condition de Slater quand il existe $x_0 \in K$ tel que pour tout $j = 1,\ldots,l, h_j(x_0) < 0$. On a vu dans les chapitres précédents que cette condition implique que la contrainte K est qualifiée.

Sous hypothèse de qualification de K en x (càd quand (3.2) est vérifiée), le cône normal à K en x est donné par la forme général des cône tangents à un cône polyédral (vue dans les chapitres précédents), càd ici

$$N_K(x) = \left\{A^\top\lambda + \sum_{j\in J(x)} \mu_j \nabla h_j(x) : \lambda \in \mathbb{R}^r, \mu_j \geq 0, \forall j \in J(x)\right\}. \tag{3.3}$$

On a donc tous les éléments pour appliquer le théorème d'Euler/Péano/Kantorovitch (voir Théorème 2.2) dans le cas d'une contrainte de la forme (3.1) et donc obtenir une "version OCD" du théorème de KKT faisant, cette fois-ci, apparaître une CNS pour l'optimalité d'une solution au problème $\min_{x\in K} f(x)$.

Théorème 3.1 (KKT en OCD) *Soit $f, (h_j)_{j=1}^l : U \to \mathbb{R}$ des fonctions différentiables convexes définies sur un ouvert convexe U de $\mathbb{R}^n$. On considère la contrainte K définie dans (3.1). Soit $x^* \in K$. On suppose que K est qualifiée en x^* on a alors équivalence entre :*

1) $x^* \in \operatorname{argmin}_{x\in K} f(x)$

2) Il existe $\lambda^* \in \mathbb{R}^r$ et $\mu^* \in \mathbb{R}^l$ tels que :

i) $\nabla f(x^*) + A^\top\lambda^* + \sum_{j=1}^l \mu_j^* \nabla h_j(x^*) = 0$

ii) $\mu_j^* \geq 0$ *pour tout* $j = 1,\ldots,l$

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

Preuve. D'après le théorème de Euler/Péano/Kantorovitch, on a *1)* est équivalent à $-\nabla f(x^*) \in N_K(x^*)$. Par ailleurs, sous l'hypothèse de qualification de K en x^*, $N_K(x^*)$ est donnée dans (3.3). Ce qui conclut la preuve. ■

Le théorème de KKT en OCD fournit donc une CNS pour les solutions du problème $\min_{x\in K} f(x)$. On a donc obtenu en plus la réciproque de KKT dans le cadre OD grâce à l'hypothèse supplémentaire de convexité de f, $h_j, j = 1,\ldots,l$ et les g_i sont affines. Par contre, le théorème de KKT ne dit rien sur les coefficients $\lambda^* \in \mathbb{R}^r$ et $\mu^* \in \mathbb{R}^l$. On peut obtenir de l'information supplémentaire sur ces coefficients grâce à la dualité Lagrangienne et en particulier grâce au problème dual.

4 Dualité Lagrangienne en OCD

On commence par quelques rappels sur la dualité Lagrangienne. Dans la cadre de l'OCD pour une contrainte de la forme (3.1), la fonction de Lagrange s'écrit

$$\mathcal{L} : \left\{\begin{array}{ccc} U \times (\mathbb{R}^r \times (\mathbb{R}_+)^l) & \longrightarrow & \mathbb{R} \\ (x, (\lambda,\mu)) & \longrightarrow & f(x) + \big\langle Ax - b, \lambda\big\rangle + \sum_{j=1}^l \mu_j h_j(x). \end{array}\right. \tag{4.1}$$

Les fonctions primale φ et duale ψ sont définies pour tout $(x,(\lambda,\mu)) \in U \times (\mathbb{R}^r \times (\mathbb{R}_+)^l)$ par

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

Comme d'habitude, on retrouve le problème d'origine $\min_{x\in K} f(x)$ comme problème primal $\inf_{x\in U} \varphi(x)$ vu que

$$\varphi(x) = \begin{cases} +\infty & \text{si } x \notin K \\ f(x) & \text{si } x \in K. \end{cases}$$

Un des objectifs de la dualité Lagrangienne est de préciser les valeurs prises par les coefficients λ^* et μ^* apparaissant dans le théorème KKT (voir Théorème 3.1) grâce au problème dual

$$\sup_{(\lambda,\mu)\in\mathbb{R}^r\times(\mathbb{R}_+)^l} \psi(\lambda,\mu). \tag{4.2}$$

Ce sera en effet le cas sous la condition que le duality gap est nul ou d'existence d'un point-selle pour $\mathcal{L}$. Pour cela, on donne une application du théorème du point-selle vu dans les chapitres précédents et du théorème de KKT.

Théorème 4.1 (Théorème de dualité Lagrangienne en OCD) *On se place dans le cadre de l'OCD pour une contrainte de la forme (3.1) où $f, (h_j)_{j=1}^l$ sont convexes et différentiables. Soit $x^* \in K$, on suppose que K est qualifiée en x^*. Il y a équivalence entre :*

1) $x^* \in \operatorname{argmin}_{x\in K} f(x)$

2) il existe une solution $(\lambda^*,\mu^*) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$ du problème dual (4.2), telle que x^* minimise $\mathcal{L}(\cdot,(\lambda^*,\mu^*))$ sur U et pour tout $j = 1,\dots,l, \mu_j^* h_j(x^*) = 0$.

Preuve. On sait qu'il y a équivalence entre

3) $(x^*,(\lambda^*,\mu^*))$ est un point selle de $\mathcal{L}$

4) x^* est solution du problème primal, (λ^*,μ^*) est solution du problème dual et le duality gap est nul.

Ce résultat est vrai sans aucune hypothèse. Par ailleurs, sous hypothèse de qualification de K en x^*, le Théorème de KKT en OCD (voir Théorème 3.1) donne l'équivalence entre :

1) $x^* \in \operatorname{argmin}_{x\in K} f(x)$

6) Il existe $\lambda^* \in \mathbb{R}^r$ et $\mu^* \in \mathbb{R}^l$ tels que :

   i) $\nabla f(x^*) + A^\top \lambda^* + \sum_{j=1}^l \mu_j^* \nabla h_j(x^*) = 0$ (càd x^* est un point critique de $x \in U \to \mathcal{L}(x,(\lambda^*,\mu^*))$)

   ii) $\mu_j^* \geq 0$ pour tout $j = 1,\dots,l$

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

Comme f et $(h_j)_j$ sont convexes et différentiables, la fonction $x \in U \to \mathcal{L}(x,(\lambda,\mu))$ est convexe et différentiable et ceci pour tout $\lambda \in \mathbb{R}^r$ et $\mu \in (\mathbb{R}_+)^l$. On a donc d'après le Théorème 2.1 l'équivalence entre

i) $\nabla f(x^*) + A^\top \lambda^* + \sum_{j=1}^l \mu_j^* \nabla h_j(x^*) = 0$ (càd x^* est un point critique de $x \in U \to \mathcal{L}(x,(\lambda^*,\mu^*))$)

iv) x^* atteint le minimum de $\mathcal{L}(\cdot,(\lambda^*,\mu^*))$ sur U.

On peut alors remplacer *i)* par *iv)* dans *6)*. On a donc équivalence entre

1) $x^* \in \operatorname{argmin}_{x\in K} f(x)$

7) Il existe $\lambda^* \in \mathbb{R}^r$ et $\mu^* \in \mathbb{R}^l$ tels que :

   iv) x^* atteint le minimum de $\mathcal{L}(\cdot,(\lambda^*,\mu^*))$ sur U.

   ii) $\mu_j^* \geq 0$ pour tout $j = 1,\dots,l$

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

Par ailleurs, montrons que si *7)* est vrai alors $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$. En effet, comme *iv)* est vrai, on a pour tout $x \in U$,

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

De plus, comme $x^* \in K$, on a $Ax^* = b$ et $h_j(x^*) \leq 0, j = 1, \ldots, l$ donc pour tout $(\lambda, \mu) \in \mathbb{R}^r \times (\mathbb{R}_+)^l$,

$$\mathcal{L}(x^*, (\lambda, \mu)) = f(x^*) + \sum_{j=1}^{l} \mu_j h_j(x^*) \leq f(x^*) = \mathcal{L}(x^*, (\lambda^*, \mu^*))$$

car $\sum_{j=1}^{l} \mu_j^* h_j(x^*) = 0$ d'après *iii)* et $Ax^* = b$ donc $f(x^*) = \mathcal{L}(x^*, (\lambda^*, \mu^*))$ est bien vérifiée. On a donc bien que $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$ càd *3)* est vraie.

Finalement, comme *3)* implique *4)*, on a bien que (λ^*, μ^*) est solution du problème dual. Par ailleurs, on a aussi montré que *1)* et *7)* sont équivalents donc d'après *iv)*, x^* atteint le minimum de $\mathcal{L}(\cdot, (\lambda^*, \mu^*))$ sur U et d'après *iii)*, $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$. On a donc bien que *7)* implique *2)* ; et donc *1)* implique *2)*.

Par ailleurs, l'implication "*2)* implique *7)*" étant triviale et qu'on a equivalence entre *7)* et *1)*, on a bien équivalence entre *2)* implique *1)* et donc équivalence entre *1)* et *2)*. ■

Corollaire 4.2 *Sous les mêmes hypothèses que le Théorème 4.1. Si $x^* \in K$ est tel que K est qualifié en x^* et si x^* est solution de $\min_{x \in K} f(x)$ alors*

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

En particulier, on a dualité forte.

Preuve. On a vu dans la preuve du Théorème 4.1 que si x^* est solution du problème primal alors il existe (λ^*, μ^*) solution du problème dual telle que $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$. Par ailleurs, on sait d'après la Proposition 1.4 du chapitre sur la dualité Lagrangienne que si $(x^*, (\lambda^*, \mu^*))$ est un point-selle de $\mathcal{L}$ alors pour φ la fonction primal on a $\varphi(x^*) = \mathcal{L}(x^*, (\lambda^*, \mu^*)) = \psi((\lambda^*, \mu^*))$. Comme ici $x^* \in K$, on a $\varphi(x^*) = f(x^*)$. On en décuit le résultat. ■

Remarque 4.3 *En OCD, si K est qualifiée en un point solution du problème primal alors d'après Corollaire 4.2, il y a dualité forte de la fonction de Lagrange. Or d'après le Théorème 3.2 du chapitre sur la dualité Lagrangienne, on sait que pour toute solution (λ^*, μ^*) au problème dual, il y a équivalence entre :*

8) x^* est solution du problème primal

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 $j = 1, \ldots, l$.*

En particulier, si K est qualifiée et que le problème primal admet une solution (alors, en particulier K est qualifiée en une solution du problème primal et donc le dualité gap est nul), il suffit de trouver les solutions (λ^*, μ^*) au problème dual et leq x^* minimisant $x \in U \to \mathcal{L}(\cdot, (\lambda^*, \mu^*))$ sur U tel que $x^* \in K$ et $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$, pour avoir toutes les solutions au problème primal. Néanmoins, on ne peut pas se contenter de trouver une seule solution (λ^*, μ^*) au problème dual et de chercher à minimiser $x \in U \to \mathcal{L}(\cdot, (\lambda^*, \mu^*))$ et de prendre uniquement les solutions dans K. En effet, on doit aussi s'assurer que $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$. On pourrait *imaginer qu'une solution x^* au problème primal soit exclue à cause de cette dernière condition qui ne serait pas satisfaite pour un certain couple (λ^*, μ^*) mais satisfaite pour un autre couple (λ_0^*, μ_0^*).*

Remarque 4.4 *Si on utilise un algorithme itératif $(x_k, (\lambda_k, \mu_k))_k$ de recherche de point-selle de $\mathcal{L}$ alors on peut utiliser le Corollaire 4.2 pour justifier le critère d'arrêt :*

$$\left|f(x_k) - \psi(\lambda_k, \mu_k)\right| \leq \epsilon$$

pour un certain ϵ (par exemple $\epsilon = 10^{-4}$) fixé par avance.

5 Méthodologie d'application de KKT et de la dualité Lagrangienne en OCD

De même que pour les problèmes en OD, on peut décrire une méthodologie d'application des Théorèmes de KKT et de dualité Lagrangienne en OCD. Les principales différences ici est qu'on peut utiliser le problème dual pour caractériser les coefficients de Lagrange et qu'on a en OCD des CNS. En particulier, on n'a pas besoin de vérifier l'existence de solution quand la contrainte est qualifiée, vu qu'un point vérifiant les conditions KKT est solution en OCD.

Méthodologie pour la résolution d'un problème d'optimisation convexe différentiable. En OCD, on peut soit utiliser la méthodologie associée au Théorème de KKT comme en OD soit on utilise une méthodologie associée à la dualité Lagrangienne. On rappelle les deux méthodes ici, en commençant par la méthode associée au **théorème de KKT** :

0) On vérifie que les fonctions $f, h_1, \ldots, h_l$ sont bien convexes et différentiables.
1) Preuve de l'existence d'une solution (par exemple, par Weiertsrass ou coercivité)
2) On cherche les points de K où la contrainte n'y est pas qualifiée (en général la condition de Slater est celle qu'on utilisera en OCD). On note par $E_1 \subset K$ cet ensemble.
3a) On cherche les points de K vérifiant les 3 conditions KKT. On note par $E_2 \subset K$ cet ensemble.
4) Par le théorème de KKT, si $\min_{x \in K} f(x)$ a une solution alors nécessairement elle est dans $E_1 \cup E_2$.
5) On évalue f en tous les points de $E_1 \cup E_2$ pour identifier celui ou ceux qui minimisent f.

Pour la méthodologie associée à la **dualité Lagrangienne** dans le cas où la contrainte est qualifiée, on a peut appliquer la méthode suivante :

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 linéaires en général)
3b) On résout le problème dual : on trouve les solutions (λ^*, μ^*) au problème dual et on cherche les $x^* \in U$ qui minimise $x \to \mathcal{L}(x, (\lambda^*, \mu^*))$ sur U tels $x^* \in K$ et $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \ldots, l$ pour chaque solution (λ^*, μ^*) du problème dual. On note par $E_2 \subset K$ cet ensemble.
4) Par le théorème de dualité Lagrangienne, l'ensemble des solutions au problème $\min_{x \in K} f(x)$ est donné exactement par E_2.

On minimise $x \to \mathcal{L}(x, (\lambda^*, \mu^*))$ pour chaque solution (λ^*, μ^*) au problème dual quand on veut avoir toute les solutions au problème primal. Si on ne cherche qu'une solution au problème primal, on peut se contenter de trouver un couple (λ^*, μ^*) solution au problème dual, puis chercher une solution x^* minimisant $x \to \mathcal{L}(x, (\lambda^*, \mu^*))$ tel que $x^* \in K$ et $\mu_j^* h_j(x^*) = 0$ pour tout $j = 1, \dots, l$ pour avoir une solution.

Exemple : On considère $s \in \mathbb{R}^n \backslash \{0\}$ et $c \in \mathbb{R}^n$. On veut résoudre le problème

$$\min_{x \in \mathbb{R}^n} \left(\frac{1}{2} \|x\|_2^2 - \langle c, x \rangle : \langle s, x \rangle \leq 0 \right). \tag{5.1}$$

1) Déterminer la fonction duale associée à (5.1)
2) Résoudre le problème dual (distinguer plusieurs cas selon le signe de $\langle c, s \rangle$)
3) Utiliser un argument de dualité pour résoudre (5.1).

1) La fonction de Lagrange est donnée par

$$\mathcal{L} : \left\{ \begin{array}{ccc} \mathbb{R}^n \times (\mathbb{R}_+) & \longrightarrow & \mathbb{R} \\ (x, \mu) & \longrightarrow & \frac{1}{2} \|x\|_2^2 - \langle c, x \rangle + \mu \langle s, x \rangle. \end{array} \right.$$

La fonction duale est définie pour tout $\mu \geq 0$ par $\psi(\mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu)$. Pour tout $\mu \geq 0$, $x \to \mathcal{L}(x, \mu)$ est convexe (car $\nabla_x^2 \mathcal{L}(\cdot, \mu) = I_d$) donc $x_\mu^* \in \mathbb{R}^n$ minimise $\mathcal{L}(\cdot, \mu)$ sur $\mathbb{R}^n$ si et seulement si $\nabla_x \mathcal{L}(x_\mu^*, \mu) = 0$ càd $x_\mu^* - c - \mu s = 0$. Donc $x_\mu^* = c - \mu s$ est l'unique solution de $\min_{x \in \mathbb{R}^n} \mathcal{L}(x, \mu)$. On a alors

$$\psi : \mu \geq 0 \to \mathcal{L}(x_\mu^*, \mu) = (1/2) \|c - \mu\|_2^2 + \langle \mu s - c, c - \mu s \rangle = -(1/2) \|c - \mu s\|_2^2 .$$

2) Le problème duale est $\max_{\mu \geq 0} \psi(\mu)$ qui s'écrit ici $\max_{\mu \geq 0} [-(1/2) \|c - \mu s\|_2^2]$. On cherche donc une solution au problème $\min_{\mu \geq 0} g(\mu)$ où $g : \mu \in \mathbb{R} \to \|c - \mu s\|_2^2 = \mu^2 \|s\|_2^2 - 2\mu \langle c, s \rangle + \|c\|_2^2$. Comme g est une fonction convexe et différentiable sur $\mathbb{R}$, g atteint son minimum sur $\mathbb{R}$ en μ^* si et seulement si $g'(\mu^*) = 0$ càd $2\mu^* \|s\|_2^2 - 2\langle c, s \rangle = 0$ càd $\mu^* = \langle c, s \rangle / \|s\|_2^2$. On a donc

$$\mu^{**} = \underset{\mu \geq 0}{\operatorname{argmin}}\, g(\mu) = \left\{ \begin{array}{cl} 0 & \text{si } \langle c, s \rangle \leq 0 \\ \langle c, s \rangle / \|s\|_2^2 & \text{si } \langle c, s \rangle \geq 0. \end{array} \right. \tag{5.2}$$

On a aussi

$$\max_{\mu \geq 0} \psi(\mu) = \psi(\mu^{**}) = \left\{ \begin{array}{cl} -\|c\|_2^2 / 2 & \text{si } \langle c, s \rangle \leq 0 \\ \frac{-1}{2} \left(\|c\|_2^2 - \frac{\langle c, s \rangle^2}{\|s\|_2^2} \right) & \text{si } \langle c, s \rangle \geq 0. \end{array} \right.$$

3) Comme la contrainte est faite d'une seule contrainte d'inégalité affine, elle est qualifiée. On peut alors appliquer le théorème de dualité Lagrangienne (Théorème 4.1) : il y a équivalence entre

1) x^* est solution du problème (5.1)
2) il existe μ^{**} solution du problème dual tel que x^* minimise $\mathcal{L}(\cdot, \mu^{**})$ sur $\mathbb{R}^n$, $x^* \in K$ et $\mu^{**} \langle s, x^* \rangle = 0$.

Il n'y a qu'une seule solution au problème dual donnée par μ^{**} dans (5.2). Il suffit donc de chercher les x^* qui minimise $x \in \mathbb{R}^n \to \mathcal{L}(x, \mu^{**})$ tels que $\langle x^*, s \rangle \leq 0$ (càd $x^* \in K$) et $\mu^{**} \langle s, x^* \rangle = 0$.

Pour le calcul de la fonction duale, on a identifié les solutions au problème $\min_{x\in\mathbb{R}^n} \mathcal{L}(x,\mu^{**})$ qui sont données par $x^* := x^*_{\mu^{**}} = c - \mu^{**}s$ càd $x^* = c$ quand $\langle c, s\rangle \leq 0$ et $x^* = c - \langle c,s\rangle (s/\left\|s\right\|_2^2)$ quand $\langle c,s\rangle \geq 0$. Par ailleurs, $\langle s, x^*_{\mu^{**}}\rangle = \langle s, c\rangle \leq 0$ quand $\langle s,c\rangle \leq 0$ et

$$\langle s, x^*_{\mu^{**}}\rangle = \langle s, c - \langle c,s\rangle (s/\left\|s\right\|_2^2)\rangle = 0$$

quand $\langle s,c\rangle \leq 0$. On a donc bien $x^* \in K$. De plus $\mu^{**}\langle s, x^*\rangle = 0$ car $\mu^{**} = 0$ quand $\langle c,s\rangle \leq 0$ et $\langle s, x^*\rangle = 0$ quand $\langle s,c\rangle \geq 0$. On a donc bien x^* minimise $\mathcal{L}(\cdot,\mu^{**})$ sur $\mathbb{R}^n$, $x^* \in K$ et $\mu^{**}\langle s, x^*\rangle = 0$.

Ainsi par le Théorème de dualité Lagrangienne en OCD, on a montré que (5.1) admet une unique solution donnée par $x_{\mu^{**}} = c - \mu^{**}s$ où

$$\mu^{**} = \left\{ \begin{array}{cl} 0 & \text{si } \langle c,s\rangle \leq 0 \\ \langle c,s\rangle / \left\|s\right\|_2^2 & \text{si } \langle c,s\rangle \geq 0. \end{array} \right.$$

càd, l'unique solution au problème (5.1) est donnée par

$$x^* = x_{\mu^{**}} = \left\{ \begin{array}{cl} c & \text{si } \langle c,s\rangle \leq 0 \\ c - \langle c, \frac{s}{\|s\|_2}\rangle \left(\frac{s}{\|s\|_2}\right) & \text{si } \langle c,s\rangle \geq 0 \end{array} \right.$$

càd $x^* = c$ quand $\langle c,s\rangle \leq 0$ et x^* est le projeté orthogonal de c sur $\mathrm{vect}(s)^\perp$ quand $\langle c,s\rangle \geq 0$.