Chapitre 5

Optimisation différentiable sous contraintes

Sommaire

Considérons un problème d'optimisation très général de la forme :

$$(P) \qquad \min f(x) \quad \text{s.c.} \quad x \in X, \tag{5.1}$$

où X est un sous-ensemble non vide de $\mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction que l'on suppose différentiable.

Les résultats d'existence de solutions que nous connaissons, ne sont d'aucune aide pour trouver une solution du problème (P). Ce qu'il nous faut, c'est une caractérisation analytique de l'optimalité, i.e. un ensemble d'équations ou d'inéquations qui pourront être résolues par les algorithmes.

Ce chapitre est consacré à l'écriture de conditions d'optimalité associées à différents problèmes d'optimisation. Tout d'abord nous donnerons une condition nécessaire d'optimalité du problème (5.1), puis nous montrerons comment s'écrit cette condition dans les cas suivants :

1. X ouvert de $\mathbb{R}^n$.

2. X convexe de $\mathbb{R}^n$.

3. X défini par des égalités et/ou des inégalités fonctionnelles :

$$X = \{x \in \mathbb{R}^n : h(x) = 0,\ g(x) \leqslant 0\}.$$

5.1 Cas général : condition d'optimalité géométrique

5.1.1 Le cône tangent ou comment se déplacer dans le domaine des contraintes

Une difficulté importante en optimisation sous contrainte consiste à savoir se déplacer dans l'ensemble des contraintes i.e. étant donnée une direction de recherche comment garantir que l'on reste dans l'ensemble X. Pour cela, on introduit la notion de direction admissible :

Définition 5.1: Direction admissible

Soit $x \in X$. Une direction $d \in \mathbb{R}^n$ sera dite admissible en x s'il existe $\eta > 0$ tel que $x + sd$ soit admissible quel que soit $s \in]0, \eta]$. On dit également que la direction d est rentrante dans X en x.

Dans le cas particulier où le domaine des contraintes est convexe, déterminer une direction d admissible en x revient à déterminer un point admissible y, différent de x : $d = y - x$ est alors une direction admissible.

Définition 5.2: Cône tangent

Soit $x \in X$. Le cône tangent (ou cône des directions admissibles) à X en x, noté $T_x(X)$, est l'ensemble des vecteurs $v \in \mathbb{R}^n$ tels qu'il existe une suite $(v_n)_{n\in\mathbb{N}}$ de limite v et une suite $(\varepsilon_n)_{n\in\mathbb{N}}$ de réels strictement positifs de limite nulle telles que :

$$x + \varepsilon_n v_n \in X.$$

Autrement dit, le cône admissible est l'ensemble des directions admissibles dans X au point x, ainsi que les limites de ces directions.

Le passage à la limite est essentiel, sous peine d'appauvrir radicalement le cône $T_x(X)$ et de le rendre ainsi inutilisable. On remarquera que $T_x(X)$ est un cône fermé, et que les seuls cas intéressants sont ceux où le point x est sur la frontière de X.

Proposition 5.3

Soit X un sous-ensemble d'intérieur non vide de $\mathbb{R}^n$ et $x \in X$.

(i) $T_x(X)$ est un cône fermé.

(ii) $x \notin \bar{X} \Rightarrow T_x(X) = \varnothing$.

(iii) $x \in \operatorname{int} X \Rightarrow T_x(X) = \mathbb{R}^n$.

Démonstration. La preuve du point *(i).* est laissée en exercice. Si $x \notin \bar{X}$, il existe des voisinages de x qui ne rencontrent pas X. Aucune direction n'est donc admissible ; le point *(ii).* est ainsi démontré. Intéressons nous à la troisième assertion : soit x à l'intérieur de X (en supposant X d'intérieur non vide). Par définition, il existe $r > 0$ tel que : $B(x, r) \subset X$. Pour toute direction $d \in \mathbb{R}^n$ non nulle et pour tout $s < r/\|d\|$, on a donc :

$$x + sd \in B(x, r) \subset X.$$

Toute direction d dans $\mathbb{R}^n$ est donc admissible au point x, d'où : $\mathbb{R}^n \subset T_x(X)$, ce qui implique : $T_x(X) = \mathbb{R}^n$. □

Exercice 5.1. *Soit $X = \{x \in \mathbb{R}^2 | x_1^2 \leqslant x_2 \leqslant 2x_1^2\}$. Dessiner X dans $\mathbb{R}^2$ et calculer l'ensemble des directions admissibles à X au point $(0,0)$ puis le cône tangent à X en $(0,0)$.*

Exercice 5.2. *Soit $X = \{x \in \mathbb{R}^2 | x_1^2 \leqslant x_2^2\}$. Dessiner X dans $\mathbb{R}^2$ et calculer $T_{(0,0)}(X)$ et $T_{(1,1)}(X)$*

5.1.2 Condition nécessaire d'optimalité géométrique

L'écriture des conditions d'optimalité en présence de contraintes est basée sur l'intuition qu'il est impossible de "descendre" à partir d'un minimum. Considérons le problème général :

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{sous la contrainte :} \quad x \in X.$$

La condition nécessaire d'optimalité du premier ordre énoncée ci-après met en évidence l'utilité de la notion de cône tangent :

Théorème 5.4

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une application différentiable et $x^\star$ un minimum local du problème (P). Alors :

$$\forall v \in T_{x^\star}(X), \ \langle \nabla f(x^\star), v \rangle \geqslant 0. \tag{5.2}$$

Démonstration. Soit $v \in T_{x^\star}(X)$. Par définition, il existe une suite $(v_n)_{n\in\mathbb{N}}$ d'éléments de $\mathbb{R}^n$ de limite v et une suite $(\varepsilon_n)_{n\in\mathbb{N}}$ de réels strictement positifs de limite nulle telles que :

$$x_n = x^\star + \varepsilon_n v_n \in X.$$

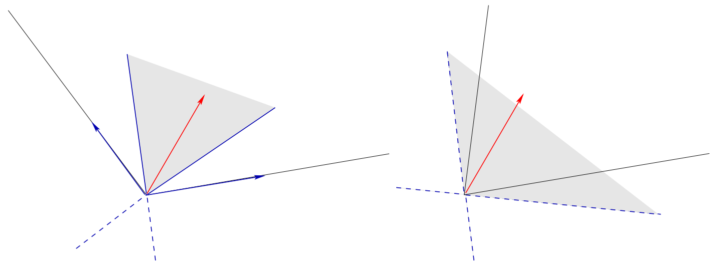

FIGURE 5.1 – Interprétation géométrique de la condition nécessaire d'optimalité (5.2)

De plus, $x^\star$ est un minimum local de f sur X, donc il existe $r > 0$ tel que :

$$\forall y \in B(x^\star, r) \cap X, \quad f(y) \geqslant f(x^\star),$$

et par construction, la suite $(x_n)_{n\in\mathbb{N}}$ est une suite d'éléments de X qui converge vers $x^\star$. Donc il existe $N \in \mathbb{N}$ tel que :

$$\forall n \geqslant N, \quad x_n = x^\star + \varepsilon_n v_n \in X \cap B(x^\star, r),$$

D'où : $f(x^\star + \varepsilon_n v_n) \geqslant f(x^\star)$, pour $n \geqslant N$. D'autre part, pour n assez grand, on a également :

$$f(x^\star + \varepsilon_n v_n) = f(x^\star) + \varepsilon_n \nabla f(x^\star)^\top v_n + \|\varepsilon_n v_n\| \epsilon(n), \quad \text{avec :} \lim_{n\to+\infty} \epsilon(n) = 0$$

D'où : $\varepsilon_n \nabla f(x^\star)^\top v_n + \|\varepsilon_n v_n\| \epsilon(n) \geqslant 0$, soit :

$$\nabla f(x^\star)^\top v_n + \|v_n\| \epsilon(n) \geqslant 0.$$

En passant à la limite quand n tend vers $+\infty$, on obtient le résultat attendu. □

5.1.3 Cas où l'ensemble des contraintes est convexe

Dans le cas où le domaine X des contraintes est convexe, la condition d'optimalité se simplifie dans le sens où elle ne fait plus intervenir le cône tangent mais uniquement l'ensemble des directions admissibles. Bien que ce dernier soit plus petit que le cône tangent, il n'y a pas de perte d'information lorsque f est différentiable en $x^\star$, que son adhérence est le cône tangent à X en $x^\star$ et que $v \mapsto \langle \nabla f(x^\star), v \rangle$ est continue.

Corollaire 5.5: CN et CS en présence de convexité

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une application différentiable et X un sous-ensemble convexe de $\mathbb{R}^n$. Si $x^\star$ est un minimum local de (P), alors :

$$\forall x \in X, \langle \nabla f(x^\star), x - x^\star \rangle \geqslant 0. \tag{5.3}$$

Si f est convexe sur le convexe X, alors la condition (5.3) est suffisante pour que $x^\star$ soit un point de minimum global de f sur X.

Exercice 5.3. *Démontrer le corollaire précédent.*

5.2 Cas différentiable sans contraintes

Supposons maintenant que le domaine X des contraintes est un ouvert de $\mathbb{R}^n$. Par conséquent : $int\ X = X$, d'où :

$$T_x(X) = \mathbb{R}^n.$$

Proposition 5.6: Condition nécessaire d'optimalité du 1er ordre

Soit $f : \mathbb{R}^n \to \mathbb{R}$ supposée différentiable et X un **<u>ouvert</u>** de $\mathbb{R}^n$. Si $x^\star$ est un point de minimum local de f sur X alors :

$$\nabla f(x^*) = 0.$$

Nous retrouvons ainsi la condition nécessaire d'optimalité du premier ordre pour l'optimisation sans contrainte. Cela nous rappelle qu'un problème d'optimisation sur un ouvert, doit être traité comme un problème d'optimisation sans contrainte et les contraintes vérifiées *a posteriori*.

Exercice 5.4. *Les problèmes* $\min_{x\in\mathbb{R}} x^2$ *s.t.* $|x| < 1$ *et* $\min_{x\in\mathbb{R}} x^2$ *s.t.* $x > 1$ *ont-ils des solutions ?*

Rappelons (sans démonstration) les conditions nécessaires du premier et du second ordre pour des problèmes d'optimisation différentiable sans contrainte :

- Conditions nécessaires d'optimalité locale
  Si $x^\star \in \mathbb{R}^n$ réalise un minimum local (resp. maximum local) de f, alors :
  $\nabla f(x^\star) = 0$ (CN d'optimalité du 1^{er} ordre)
  $H[f](x^\star)$ est semidéfinie positive (CN d'optimalité du 2^{nd} ordre)
  (resp. $H[f](x^\star)$ est semidéfinie négative)
- Condition suffisante d'optimalité locale
  Soit X un ouvert de $\mathbb{R}^n$ et $x^\star \in X$. Si :
  $\nabla f(x^\star) = 0$ et $H[f](x^\star)$ symétrique, définie positive (resp. définie négative)
  Alors $x^\star$ est un point de minimum local (resp. maximum local) de f sur X.
- Condition Suffisante d'optimalité globale
  Supposons $\nabla f(x^\star) = 0$.
  (i) Si f est convexe, alors $x^\star$ est un point de minimum global de f.
  (ii) Si f est strictement convexe, alors $x^\star$ est l'unique point de minimum global de f.

5.3 Cas différentiable avec contraintes fonctionnelles

Nous nous intéressons maintenant plus particulièrement à la résolution de problèmes d'optimisation dont le domaine des contraintes X est défini par des égalités et/ou des inégalités :

$$X = \{x \in \mathbb{R}^n : h_i(x) = 0,\ i = 1, \dots, p,\ g_j(x) \leqslant 0,\ j = 1, \dots, q\}$$

noté également :

$$X = \{x \in \mathbb{R}^n : h(x) = 0,\ g(x) \leqslant 0\}$$

où les fonctions $g : \mathbb{R}^n \to \mathbb{R}^q$ et $h : \mathbb{R}^n \to \mathbb{R}^p$ sont supposées différentiables sur $\mathbb{R}^n$. On rappelle que par continuité des applications g et h, l'ensemble X est un fermé de $\mathbb{R}^n$, non nécessairement borné (penser au cas $h \equiv 0$ et $g(x, y) = -x$).

A partir de maintenant, nous travaillerons toujours à partir d'un problème d'optimisation écrit sous *forme standard*, à savoir :

$$(P) \left| \begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ \text{s.c.} & h_i(x) = 0, \quad i = 1, \dots, p \\ & g_j(x) \leqslant 0, \quad j = 1, \dots, q. \end{array} \right.$$

Proposition 5.7

Lorsque h est affine, g convexe et la fonction objectif f est convexe sur X, alors le problème (P) est convexe.

Exercice 5.5. *Donner la forme standard du problème* $\max_{(x,y)\in\mathbb{R}^2} f(x,y)$ *s.t. :* $x^2+y^2 \geqslant 1$ *et* $x+y=5$. *Ce problème est-il convexe ?*

L'essentiel du travail dans ce chapitre pourrait se résumer à trouver une expression plus "pratique" de la condition (5.2), et donc du cône tangent associé à un ensemble d'égalités et/ou d'inégalités fonctionnelles.

Définition 5.8: Condition de Slater (CS de qualification convexe)

Soient f et $(g_j)_{1\leqslant j\leqslant q}$ des fonctions convexes, $(h_i)_{1\leqslant i\leqslant p}$ des fonctions affines. On dit que les contraintes du problème (P) vérifient la condition de Slater, s'il existe x_0 tel que

$$\begin{aligned} \forall i = 1,\ldots,p \quad & h_i(x_0) = 0 \\ \forall j = 1,\ldots,q \quad & g_j(x_0) < 0 \end{aligned}$$

Définition 5.9: Contrainte active

On dit qu'une contrainte en inégalité g est active ou saturée au point $x^\star \in \mathbb{R}^n$ si $g(x^\star) = 0$.

La condition de Slater se traduit donc de la manière suivante : il existe un point x_0 pour lequel toutes les contraintes non-affines sont inactives.

S'il n'y a que des contraintes d'égalité, la condition de Slater n'est pas restrictive : elle est équivalente à la condition d'admissibilité. Cette condition ne joue un rôle utile que dans le cas de contraintes d'inégalités.

Comme on le verra par la suite, la condition de Slater est une condition suffisante pour pouvoir résoudre par dualité le problème (P). Il s'agit d'un exemple de *qualification de contraintes*, parmi d'autres.

Dans l'exemple qui suit, on illustre un cas de contraintes non qualifiées. Considérons le problème suivant

$$\begin{aligned} \min \quad & f(x_1,x_2) \\ \text{sous les contraintes } & x_2 \leqslant 0 \\ & \min(x_1,0) - x_2 \leqslant 0 \\ & (x_1,x_2) \in \mathbb{R}^2 \end{aligned}$$

L'ensemble des points admissibles de ce problème sont les points

$$\mathcal{X} = \left\{(x_1,0) \in \mathbb{R}^2, \quad x_1 \geqslant 0\right\}.$$

Or pour tous ces points la seconde contrainte s'annule ; les contraintes ne sont pas qualifiées (au sens de Slater). Il est intéressant de noter que la qualification ou non des contraintes

dépend de la manière dont on représente les contraintes. En effet on peut choisir la représentation suivante :

$$\begin{aligned} \min \quad & f(x_1,x_2) \\ \text{sous les contraintes } & -x_1 \leqslant 0 \\ & x_2 = 0 \\ & (x_1,x_2) \in \mathbb{R}^2 \end{aligned}$$

Et sous cette forme, les contraintes vérifient la condition de Slater.

5.3.1 Intuition géométrique : cas d'une contrainte d'inégalité

Notons qu'une contrainte d'égalité $h_i = 0$ peut toujours se réécrire comme deux contraintes d'inégalités :

$$h_i(x) = 0 \Longleftrightarrow \begin{cases} h_i(x) & \leqslant 0 \\ -h_i(x) \leqslant 0. \end{cases}$$

Nous allons donc essayer de comprendre ce qui se passe géométriquement dans le cas de la minimisation d'une fonction convexe sous contrainte d'inégalité :

$$\min_{g(x)\leqslant 0} f(x)$$

où f et g sont des fonctions convexes de $\mathbb{R}^n$ dans $\mathbb{R}$. La condition d'optimalité sous contrainte est la suivante.

Théorème 5.10

Si f et g sont des fonctions différentiables de $\mathbb{R}^n$ dans $\mathbb{R}$ et si $x^\star$ est un minimum local de f sur l'ensemble $X := \{x|g(x) \leqslant 0\}$, si on suppose de plus que soit $\nabla g(x^\star) \neq 0$ soit g est non-active en $x^\star$. Alors il existe $\gamma \geqslant 0$ tel que

$$\nabla f(x^\star) + \gamma \nabla g(x^\star) = 0 \qquad \text{et} \qquad \gamma g(x^\star) = 0.$$

De plus si f et g sont convexes alors la dernière equation est suffisante pour assurer que $x^\star$ est un minimum local (et même global) de f sur l'ensemble X.

On ne démontre pas ce résulat, on démontrera un résultat plus général dans la suite. En revanche on peut essayer de comprendre géometriquement ce qu'il se passe, cf. Figure 5.2. On remarque que l'on a besoin d'une hypothèse qui est "$\nabla g(x^\star) \neq 0$" quand g est active en $x^\star$, cette hypothèse s'appelle "qualification des contraintes".

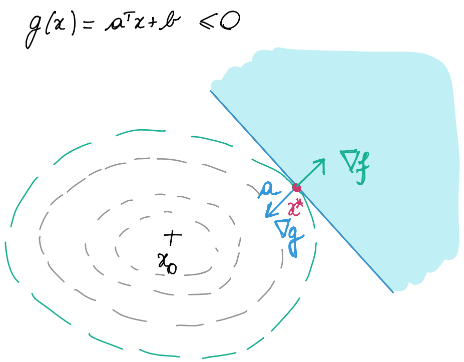

FIGURE 5.2 – Dans cet exemple, le point x_0 est un point de minimum sans contrainte et le minimum global sous contrainte est $x^\star$.

5.4 Introduction à la dualité lagrangienne

On définit le *Lagrangien* (ou fonction lagrangienne) associé à (P) et défini par :

$$\begin{aligned} L : \mathbb{R}^n \times \mathbb{R}^p \times (\mathbb{R}_+)^q &\to \mathbb{R} \\ (x, \lambda, \mu) &\mapsto L(x; \lambda, \mu) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x) + \sum_{j=1}^{q} \mu_j g_j(x), \\ &\qquad\qquad\quad = f(x) + \langle \lambda, h(x) \rangle_{\mathbb{R}^p} + \langle \mu, g(x) \rangle_{\mathbb{R}^q}. \end{aligned}$$

où l'on note λ et μ les vecteurs de coordonnées respectives $(\lambda_1, \ldots, \lambda_p)$ et $(\mu_1, \ldots, \mu_q)$.

5.4.1 Problème primal - Problème dual

Commençons par observer que le problème primal (P) peut se réécrire à l'aide du Lagrangien de la façon suivante :

$$\inf_{x \in X} f(x) = \inf_{x \in \mathbb{R}^n} \left(\sup_{(\lambda,\mu) \in \mathbb{R}^p \times (\mathbb{R}_+)^q} L(x; \lambda, \mu) \right) \qquad \text{PROBLÈME PRIMAL}$$

On définit alors :

- la fonction primale $\bar{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ associée au problème (P) :

$$\bar{f}(x) = \sup_{(\lambda,\mu) \in \mathbb{R}^p \times (\mathbb{R}_+)^q} L(x; \lambda, \mu) = \begin{cases} f(x) & \text{si } x \in X \\ +\infty & \text{sinon.} \end{cases}$$

- le *domaine admissible primal* :

$$\left\{ x \in \mathbb{R}^n : \bar{f}(x) < +\infty \right\} = X$$

sur lequel la fonction primale est à valeurs finies.

Le problème dual du problème (P) est obtenu en échangeant l'inf et le sup :

$$\sup_{(\lambda,\mu) \in \mathbb{R}^p \times (\mathbb{R}_+)^q} \left(\inf_{x \in \mathbb{R}^n} L(x; \lambda, \mu) \right) = \sup_{(\lambda,\mu) \in X^\star} f^\star(\lambda, \mu) \qquad \text{PROBLÈME DUAL}$$

où l'on définit :

- la fonction duale du problème (P) :

$$f^\star(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x; \lambda, \mu).$$

- le *domaine admissible dual* :

$$X^\star = \left\{ (\lambda, \mu) \in \mathbb{R}^p \times (\mathbb{R}_+)^q : \inf_{x \in \mathbb{R}^n} L(x; \lambda, \mu) > -\infty \right\},$$

sur lequel la fonction duale est à valeurs finies.

La question de la dualité est savoir si l'on change les solutions du problème primal en échangeant l'inf et le sup. D'un point de vue théorique, le problème dual est beaucoup plus simple à résoudre car il s'agit, sous sa forme standard, d'un problème d'optimisation convexe. En effet :

Proposition 5.11

La fonction duale $f^\star$ est toujours concave et le domaine admissible dual $X^\star$ convexe.

Exercice 5.6. *Le démontrer !*

5.4.2 Liens entre problème primal et problème dual

Principe de dualité faible.

$$\sup_{(\lambda,\mu) \in \mathbb{R}^p \times (\mathbb{R}_+)^q} \left(\inf_{x \in \mathbb{R}^n} L(x; \lambda, \mu) \right) \leqslant \inf_{x \in \mathbb{R}^n} \left(\sup_{(\lambda,\mu) \in \mathbb{R}^p \times (\mathbb{R}_+)^q} L(x; \lambda, \mu) \right).$$

Le principe de dualité faible s'obtient directement. Dans le cadre de la dualité Lagrangienne, on appelle *saut de dualité* la différence positive entre les deux quantités précédentes, à savoir :

$$\inf_{x\in\mathbb{R}^n}\left(\sup_{(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q} L(x;\lambda,\mu)\right) - \sup_{(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q}\left(\inf_{x\in\mathbb{R}^n} L(x;\lambda,\mu)\right).$$

Si le saut de dualité est non nul, cela signifie que les solutions des problèmes primal et dual n'ont *a priori* rien à voir entre elles. En l'absence de saut de dualité, les problèmes sont équivalents (même ensemble de solutions) et l'existence de solutions primale et duale dans ce cas est étroitement liée à l'existence de points selle du Lagrangien.

Définition 5.12: Point-selle du Lagrangien

Soit $\bar{x}\in\mathbb{R}^n$ et $(\bar{\lambda},\bar{\mu})\in\mathbb{R}^p\times(\mathbb{R}_+)^q$. Le point $(\bar{x},\bar{\lambda},\bar{\mu})$ est un point-selle du Lagrangien si :

$$\begin{cases} \forall(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q, & L(\bar{x};\lambda,\mu) \leqslant L(\bar{x};\bar{\lambda},\bar{\mu}) \\ \forall x\in\mathbb{R}^n, & L(x;\bar{\lambda},\bar{\mu}) \geqslant L(\bar{x};\bar{\lambda},\bar{\mu}) \end{cases}$$

Autrement dit, si :

$$f(\bar{x}) = \sup_{(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q} L(\bar{x};\lambda,\mu) = L(\bar{x};\bar{\lambda},\bar{\mu}) = \inf_{x\in\mathbb{R}^n} L(x;\bar{\lambda},\bar{\mu}) = f^\star(\bar{\lambda},\bar{\mu}).$$

En combinant les deux inégalités de la définition précédente, on a que $(\bar{x},\bar{\lambda},\bar{\mu})$ est un point-selle du Lagrangien, si pour tout $(x,\lambda,\mu)\in\mathbb{R}^n\times\mathbb{R}^p\times\mathbb{R}^q_+$,

$$L(\bar{x},\lambda,\mu) \leqslant L(\bar{x},\bar{\lambda},\bar{\mu}) \leqslant L(x,\bar{\lambda},\bar{\mu}).$$

Théorème 5.13: Théorème de dualité

Le point $(\bar{x},\bar{\lambda},\bar{\mu})$ est un point-selle du Lagrangien si et seulement si :

(i) $\bar{x}$ est solution du problème primal.

(ii) $(\bar{\lambda},\bar{\mu})$ est solution du problème dual.

(iii) le saut de dualité est nul c'est-à-dire :

$$\sup_{(\lambda,\mu)\in X^\star}\left(\inf_{x\in\mathbb{R}^n} L(x;\lambda,\mu)\right) = \inf_{x\in X}\left(\sup_{(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q} L(x;\lambda,\mu)\right).$$

Dans ce cas, on a en plus :

$$\begin{aligned} f(\bar{x}) &= \inf_{x\in X} f(x) = \inf_{x\in\mathbb{R}^n} L(x;\bar{\lambda},\bar{\mu}) \\ &= \sup_{(\lambda,\mu)\in X^\star}\left(\inf_{x\in\mathbb{R}^n} L(x;\lambda,\mu)\right) = \inf_{x\in X}\left(\sup_{(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q} L(x;\lambda,\mu)\right). \end{aligned}$$

Démonstration. L'implication $(\Rightarrow)$ ne pose pas de problème. Démontrons l'implication $(\Leftarrow)$: soit $(\bar{x},\bar{\lambda},\bar{\mu})$ un point-selle du Lagrangien. Vérifions tout d'abord que $\bar{x}$ est un point admissible pour le problème (P), i.e. que : $\bar{x}\in X$.

Par définition du point selle $(\bar{x},\bar{\lambda},\bar{\mu})$, on a pour tout $(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q$:

$$L(\bar{x};\bar{\lambda},\bar{\mu}) \geqslant L(\bar{x};\lambda,\mu),$$

soit :

$$\forall(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q,\ (\lambda-\bar{\lambda})^\top h(\bar{x}) + (\mu-\bar{\mu})^\top g(\bar{x}) \leqslant 0,$$

ce qui implique :

$$(*)\quad \begin{cases} \forall\lambda, & (\lambda-\bar{\lambda})^\top h(\bar{x}) \leqslant 0 \\ \forall\mu\geqslant 0, & (\mu-\bar{\mu})^\top g(\bar{x}) \leqslant 0 \end{cases} \Rightarrow h(\bar{x}) = 0 \quad \text{et} \quad g(\bar{x}) \leqslant 0.$$

Donc le point $\bar{x}$ est un point admissible du problème (P) et :

$$L(\bar{x};\bar{\lambda},\bar{\mu}) = f(\bar{x}) + \mu^\top g(\bar{x}) \leqslant f(\bar{x}).$$

Appliquées en $\mu = 0$, les relations $(*)$ s'écrivent : $\bar{\mu}^\top g(\bar{x}) \geqslant 0$. Or $\bar{\mu}\geqslant 0$ et $g(\bar{x})\leqslant 0$, d'où : $\bar{\mu}^\top g(\bar{x}) \leqslant 0$, et par suite :

$$\mu^\top g(\bar{x}) = 0.$$

On en déduit alors :

$$L(\bar{x};\bar{\lambda},\bar{\mu}) = f(\bar{x}). \tag{5.4}$$

Démontrons maintenant que le point $\bar{x}$ est bien solution du problème (P). Revenant à la définition d'un point selle, on a alors :

$$\forall x \in X,\ f(x) \geqslant f(x) + \underbrace{\bar{\mu}^\top g(x)}_{\leqslant 0} = L(x;\bar{\lambda},\bar{\mu}) \geqslant L(\bar{x};\bar{\lambda},\bar{\mu}) = f(\bar{x}).$$

D'où le point $\bar{x}$ est solution du problème (P) i.e. :

$$f(\bar{x}) = \inf_{x\in X} f(x) = \inf_{x\in\mathbb{R}^n} \sup_{(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q} L(x;\lambda,\mu) \tag{5.5}$$

On démontre de façon similaire que $(\bar{\lambda},\bar{\mu})$ est solution du problème dual. En effet,

$$\forall(\lambda,\mu) \in \mathbb{R}^p \times (\mathbb{R}_+)^q,\ f^*(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} L(x;\lambda,\mu) \leqslant L(\bar{x};\lambda,\mu) \leqslant L(\bar{x};\bar{\lambda},\bar{\mu}) = f^*(\bar{\lambda},\bar{\mu}),$$

d'où :

$$f^*(\bar{\lambda},\bar{\mu}) = \sup_{\mathbb{R}^p\times(\mathbb{R}_+)^q} f^*(\lambda,\mu) = \sup_{(\lambda,\mu)\in\mathbb{R}^p\times(\mathbb{R}_+)^q} \inf_{x\in\mathbb{R}^n} L(\bar{x};\lambda,\mu) \tag{5.6}$$

D'après les égalités (5.4), (5.5) et (5.6), le saut de dualité est bien nul. □

En pratique le théorème de dualité ne nous dit pas comment calculer les points selle du Lagrangien (puisqu'il faudrait déjà connaitre les solutions optimales des problèmes primal et dual !).

Théorème 5.14: Théorème de dualité forte

On suppose le problème convexe. Soient $(\lambda^\star,\mu^\star)$ solutions du problème dual. Si $x^\star$ est un point solution de (P) et si la condition de Slater est vérifiée, alors le saut de dualité est nul, i.e.

$$D(\lambda^\star,\mu^\star) = L(x^\star,\lambda^\star,\mu^\star) = f(x^\star)$$

L'intérêt du problème dual est double : il est concave et les contraintes sont simples à prendre en compte (contraintes de positivité dans le cas général, pas de contrainte dans le cas de contraintes d'égalité). Pour le résoudre, on peut par exemple utiliser :

- une méthode de gradient dans le cas de contraintes d'égalité : c'est l'algorithme d'Uzawa.
- une méthode de gradient projeté dans le cas de contraintes d'inégalité.
- des méthodes de type Newton.

Dans le cas convexe, nous savons comment caractériser les points-selles du Lagrangien, et donc les solutions primales et duales : ce sont les conditions de Karush-Kuhn-Tucker (KKT).

Théorème 5.15: Condition d'optimalité de Karush-Kunh-Tucker (KKT)

Supposons le problème convexe, et la condition de Slater vérifiée. $x^\star$, $\lambda^\star$, $\mu^\star$ sont solutions des problèmes primal et dual, si et seulement si ils satisfont les conditions de Karush-Kuhn-Tucker (KKT), i.e.

$$\begin{cases} h(x^\star) = 0 \quad g(x^\star) \leqslant 0 & (1) \\ \forall j = 1,\dots,q, \quad \mu_j^\star \geqslant 0 & (2) \\ \forall j = 1,\dots,q, \quad \mu_j^\star g_j(x^\star) = 0 & (3) \\ \nabla f(x^\star) + \sum_{i=1}^p \lambda_i^\star \nabla h_i(x^\star) + \sum_{j=1}^q \mu_j^\star \nabla g_j^\star(x^\star) = 0 & (4) \end{cases}$$

Dans les conditions KKT, on peut reconnaître :

(1) contraintes primales

(2) contraintes duales

(3) conditions de complémentarités

(4) gradient du Lagrangien par rapport à x nul, i.e.

$$\nabla_x L(x^\star,\lambda^\star,\mu^\star) = 0.$$

On peut remarquer que les conditions (3) et (4) proviennent du fait que $(x^\star, \lambda^\star, \mu^\star)$ est un point-selle du Lagrangien, pour lequel on peut écrire les conditions d'optimalité du premier ordre (règle de Fermat) en x (ce qui donne (4)), en μ (et la condition de positivité donne (3)).

Démonstration. Tout d'abord, supposons que $x^\star \in \mathbb{R}^n$ et $(\lambda^\star,\mu^\star) \in \mathbb{R}^p \times \mathbb{R}_+^q$ sont solutions respectives des problèmes primal et dual.

(1) La faisabilité primale vient du fait que $x^\star$ est solution du problème primal ;

(2) La faisabilité duale vient du fait que $(\lambda^\star,\mu^\star) \in \mathbb{R}^p \times \mathbb{R}_+^q$ sont solutions du problème dual ;

(3) On peut écrire

$$\begin{aligned} f^\star(\lambda^\star,\mu^\star) &:= \inf_{x\in\mathbb{R}^d} L(x,\lambda^\star,\mu^\star) && \text{(définition)} \\ &\leqslant L(x^\star,\lambda^\star,\mu^\star) \\ &= f(x^\star) + \sum_{i=1}^p \lambda_i h_i(x^\star) + \sum_{j=1}^q \mu_j g_j(x^\star) \\ &\leqslant f(x^\star) && (\mu_j g_j(x^\star) \leqslant 0 \text{ et } h_i(x^\star) = 0). \end{aligned}$$

Mais par la condition de Slater, on a la dualité forte, i.e. $D(\lambda^\star,\mu^\star) = f(x^\star)$, donc les deux dernières inégalités sont en fait des égalités. On en déduit que $x^\star$ est minimiseur de $L(\cdot,\lambda^\star,\mu^\star)$ et que $\sum_{j=1}^q \mu_j g_j(x^\star) = 0$, d 'où les conditions de complémentarité.

(4) Du fait que $x^\star$ est minimiseur de $L(\cdot, \lambda^\star, \mu^\star)$ (qui est une fonction convexe), la condition d'optimalité du premier ordre (règle de Fermat) nous donne la dernière condition.

D'autre part, si $x^\star \in \mathbb{R}^n$ et $(\lambda^\star, \mu^\star) \in \mathbb{R}^p \times \mathbb{R}_+^q$ satisfont les conditions KKT, alors

1. (1) et (3) impliquent que $f(x^\star) = L(x^\star, \lambda^\star, \mu^\star)$;
2. Par (2), (4), la convexité de $L(\cdot, \lambda^\star, \mu^\star)$ et la condition d'optimalité du premier ordre, on a $f^\star(\lambda^\star, \mu^\star) = \inf_{x \in \mathbb{R}^n} L(x, \lambda^\star, \mu^\star) = L(x^\star, \lambda^\star, \mu^\star)$.

Donc $f(x^\star) = f^\star(\lambda^\star, \mu^\star)$. De plus, par dualité faible,

$$f(x^\star) \geqslant \inf_{x \in \mathcal{X}} f(x) \geqslant \sup_{(\lambda,\mu) \in \mathbb{R}^p \times \mathbb{R}_+^q} f^\star(\lambda, \mu) \geqslant f^\star(\lambda^\star, \mu^\star) = f(x^\star),$$

donc toutes les inégalités sont des égalités et, $x^\star$ et $(\lambda^\star, \mu^\star)$ sont solutions des problèmes primal et dual. □

Qualification des contraintes ? Dans le cas convexe, la condition de qualification des contraintes de Slater est là pour assurer que de tels points-selles existent. Attention de manière générale, si les contraintes ne sont pas qualifiées (au sens de Slater ou autre), alors le lagrangien peut ne pas posséder de points-selles. En revanche, cela n'implique pas que le problème sous contraintes associé ne possède pas de solutions. En effet, on a vu dans un exemple plus haut que, suivant la manière dont sont écrites les contraintes, celles-ci peuvent être qualifiées ou non. Or, la manière d'écrire des contraintes ne change pas l'ensemble des solutions. On peut donc considérer qu'exprimer l'ensemble admissible à l'aide de contraintes qualifiées correspond à définir un bon lagrangien, c'est-à-dire exploitable pour la résolution du problème.

Au delà du cas convexe Pour terminer, revenons au cas général. Dans le cas où le problème n'est pas convexe, on a les résultats suivants :

— si $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p \times (\mathbb{R}^+)^q$ est un point-selle du lagrangien, alors $\bar{x}$ est une solution du problème primal et il vérifie les conditions KKT ;

— si $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p \times (\mathbb{R}^+)^q$ vérifie les conditions KKT, alors $\bar{x}$ est un point critique de $J + \chi_{\mathcal{X}}$.

En pratique, cela implique que l'on peut essayer de trouver les points $(\bar{x}, \bar{\lambda}, \bar{\mu})$ qui satisfont les conditions KKT, mais sans garantie d'existence de tels points. S'il en existe, alors $\bar{x}$ est un point critique de $J + \chi_{\mathcal{X}}$; il faut donc vérifier par d'autres moyens que $\bar{x}$ est bien une solution du problème sous contraintes.