Module A7

Dualité lagrangienne et conditions KKT

Sauf mention contraire, $\mathcal{X}$ est un espace vectoriel réel de dimension finie, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\| \cdot \|$ la norme euclidienne associée. Pour toutes les autres notations et définitions, le lecteur est invité à se reporter aux cours précédents.

Dans ce module, on s'intéresse au problème d'optimisation sous contraintes suivant :

$$\begin{array}{ll} \text{Minimiser} & J(x) \\ \text{sous les contraintes} & g_j(x) \le 0, \quad j = 1, \ldots, n \\ & h_i(x) = 0, \quad i = 1, \ldots, m \\ & x \in \mathcal{X} \end{array} \tag{$\mathcal{P}$}$$

où $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ est une fonction et les $g_j : \mathcal{X} \to \mathbb{R}$ et les $h_i : \mathcal{X} \to \mathbb{R}$ sont des fonctions différentiables.

1 Optimisation sous contraintes

1.1 Contraintes d'inégalité, contraintes d'égalité

Considérons le problème d'optimisation sous contraintes de la forme

$$\begin{array}{ll} \text{Minimiser} & J(x) \\ \text{sous les contraintes} & g_j(x) \le 0, \quad j = 1, \ldots, n \\ & h_i(x) = 0, \quad i = 1, \ldots, m \\ & x \in \mathcal{X} \end{array} \tag{$\mathcal{P}_c$}$$

Les contraintes de la forme $g_j(x) \le 0$ sont appelées *contraintes d'inégalités* et les contraintes de la forme $h_i(x) = 0$ *contraintes d'égalité*. Dans le contexte de l'optimisation sous contraintes, on dit que x^* est une solution du problème $(\mathcal{P}_c)$ si x^* est un point *admissible* de $(\mathcal{P}_c)$, c'est-à-dire que

$$x^* \in \mathcal{R} = \left\{ x \in \mathcal{X} \mid \forall\, j \in [\![1\,;\, n]\!]\,, g_j(x) \le 0 \text{ et } \forall\, i \in [\![1\,;\, m]\!]\,, h_i(x) = 0 \right\}$$

et que $$\forall\, x \in \mathcal{R}, \qquad J(x^*) \le J(x)$$

Autrement dit, x^* est un minimiseur de la fonction $J + \chi_{\mathcal{R}}$, l'ensemble $\mathcal{R}$ étant appelé *ensemble admissible* de $(\mathcal{P}_c)$.

Notons qu'en général, la convexité des fonctions définissant les contraintes ne suffit pas assurer la convexité de l'ensemble admissible et donc, du problème lui-même. Or, cette condition est essentielle pour appliquer la CNS d'optimalité du premier ordre. Aussi, pour se ramener à ce cas, il est courant de se restreindre à considérer un cas relativement simple de telles fonctions.

Proposition 1

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction **convexe**, $g_j : \mathcal{X} \to \mathbb{R}$ des fonctions convexes pour tout $1 \le j \le n$ et $h_i : \mathcal{X} \to \mathbb{R}$ des fonctions affines pour tout $1 \le i \le m$. Alors le problème $(\mathcal{P}_c)$ est un problème d'optimisation convexe.

On rappelle qu'un problème d'optimisation sous contraintes est convexe si sa fonction objectif et son ensemble admissible sont convexes.

Démonstration : La fonction J étant supposée convexe, il suffit de montrer que l'ensemble admissible $\mathcal{R}$ du problème $(\mathcal{P}_c)$ des points admissibles est convexe. Celui-ci s'écrit comme l'intersection de $I + J$ ensembles :

$$\mathcal{R} = \left(\bigcap_{j=1}^{n} \left\{ x \in \mathcal{X} \mid g_j(x) \le 0 \right\} \right) \cap \left(\bigcap_{i=1}^{m} \left\{ x \in \mathcal{X} \mid h_i(x) = 0 \right\} \right)$$

Chacun des ces ensembles sont convexes ; en effet, l'image réciproque d'un convexe par une fonction affine est convexe, tandis qu'il est aisé de vérifier que l'image réciproque par une fonction convexe de l'intervalle $]-\infty\,; 0]$ est également convexe (il suffit d'écrire la définition d'un convexe). Ainsi, $\mathcal{R}$ est convexe également. ■

Remarquons les contraintes d'égalité peuvent toujours s'écrire comme deux contraintes d'inégalité ; en effet, on a

$$h_i(x) = 0 \qquad \iff \qquad \begin{cases} h_i(x) \le 0 \\ -h_i(x) \le 0 \end{cases}$$

Aussi, désormais, on supposera que les problèmes sous contraintes d'inégalité et d'égalité ne comportent que des contraintes d'inégalité, et sont donc de la forme :

$$\begin{array}{ll} \text{Minimiser} & J(x) \\ \text{sous les contraintes} & g_j(x) \le 0, \quad j = 1, \ldots, n \\ & x \in \mathcal{X} \end{array} \tag{$\mathcal{P}_{ci}$}$$

Ce problème est convexe si les fonctions J et g_j sont convexes. Notons en particulier que, dans ce cas, les fonctions g_j correspondant aux contraintes d'égalité découplées sont nécessairement affines (puisque $-g_j$ est convexe aussi). Introduisons alors la définition suivante :

Définition 1 (Condition de Slater)

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** et $g_j : \mathcal{X} \to \mathbb{R}$ des fonctions **convexes** pour tout $1 \le j \le n$. Posons $\mathcal{J}$ l'ensemble des indices $1 \le j \le n$ correspondant à une contrainte g_j affine. On dit que les contraintes du problème $(\mathcal{P}_{ci})$ vérifient la condition de Slater s'il existe un vecteur $x_0 \in \mathcal{X}$ tel que

$$\forall\, 1 \le j \le n, \qquad \begin{cases} g_j(x_0) \le 0 & \text{si } j \in \mathcal{J} \\ g_j(x_0) < 0 & \text{si } j \notin \mathcal{J} \end{cases}$$

> Dans le vocabulaire de l'optimisation sous contraintes, on dit qu'une contrainte d'inégalité $g_i(x) \le 0$ est *active*, ou *saturée*, en un point x^0 si $g_i(x^0) = 0$. 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.

Remarquons que, dans la définition précédente, le vecteur x_0, s'il existe, est nécessairement un point admissible du problème. Aussi, pour vérifier que les contraintes satisfont la condition de Slater, il suffit de vérifier que les conditions introduites dans la définition sont vérifiées pour au moins un point admissible du problème.En particulier, lorsque les contraintes sont affines, la condition de Slater équivaut exactement à la réalisabilité du problème considéré. On voit donc que cette condition ne joue un rôle utile que dans le cas de contraintes non affines.

Comme on le verra par la suite, la condition de Slater est une condition suffisante pour pouvoir résoudre par dualité le problème $(\mathcal{P}_{ci})$. 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 (au sens de Slater).

Exemple

Contraintes non qualifiées. Considérons le problème suivant

$$\begin{array}{ll} \text{Minimiser} & J(x_1,x_2)=x_1 \\ \text{sous les contraintes} & x_2 \le 0 \\ & \min(x_1,0)^2 - x_2 \le 0 \\ & (x_1,x_2)\in\mathbb{R}^2 \end{array}$$

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

$$\mathcal{R} = \Big\{(x_1,0)\in\mathbb{R}^2 \mid x_1 \ge 0\Big\}$$

Or, pour tous ces points, la seconde contrainte s'annule ; les contraintes de ce problème ne sont donc pas qualifiées. Il est intéressant de noter que la qualification ou non des contraintes dépend de la manière dont on représente ces contraintes. En effet, on peut choisir la représentation suivante pour le problème considéré :

$$\begin{array}{ll} \text{Minimiser} & J(x_1,x_2)=x_1 \\ \text{sous les contraintes} & -x_1 \le 0 \\ & x_2 \le 0 \\ & -x_2 \le 0 \\ & (x_1,x_2)\in\mathbb{R}^2 \end{array}$$

Il est alors aisé de vérifier qu'il s'agit du même problème. Cependant, sous cette écriture, les contraintes sont qualifiées (car elles sont affines, et que le problème est réalisable).

1.2 Lagrangien d'un problème sous contraintes d'inégalité

Pour tout $x\in\mathcal{X}$, posons

$$g(x) = \big(g_1(x),\dots,g_n(x)\big)$$

Définition 2 (Lagrangien pour un problème sous contraintes d'inégalité)

Soit $J:\mathcal{X}\mapsto\mathbb{R}$ et $g_j:\mathcal{X}\to\mathbb{R}$ des fonctions. Le *lagrangien* du problème $(\mathcal{P}_{ci})$ est la fonction $\mathcal{L}:\mathcal{X}\times(\mathbb{R}^+)^n\to\mathbb{R}$ définie par

$$\forall\,(x,\lambda)\in\mathcal{X}\times(\mathbb{R}^+)^n,\qquad \mathcal{L}(x;\lambda) = J(x) + \langle\lambda, g(x)\rangle$$

Les coefficients du vecteur λ sont appelés *multiplicateurs de* Lagrange.

On parle parfois également de *fonction lagrangienne*. Les anglo-saxons distinguent les multiplicateurs de Lagrange pour les contraintes d'égalité et les multiplicateurs de Kuhn–Tucker pour les contraintes d'inégalité, les premiers n'étant pas imposés positifs.

Remarque : Dans la littérature, les contraintes d'inégalité sont parfois écrites comme des contraintes de positivité, auquel cas il faut considérer l'opposé du second terme dans la définition du lagrangien (car λ est, par convention, toujours supposé à composantes positives).

La définition précédente va nous permettre de représenter la fonction objectif du problème $(\mathcal{P}_{ci})$ comme l'enveloppe supérieure d'une famille de fonctions, et pose de ce fait les bases de la dualité min-max que l'on développera dans la section suivante.

Pour ne plus se préoccuper du signe des multiplicateurs de Lagrange λ, il peut être utile de prolonger la définition du lagrangien en-dehors de l'ensemble $\mathcal{X}\times(\mathbb{R}^+)^n$, en posant par convention

$$\forall\,\lambda\notin(\mathbb{R}^+)^n,\qquad \mathcal{L}(x;\lambda) = -\infty$$

Proposition 2

Soit $J:\mathcal{X}\to\mathbb{R}$ une fonction **convexe** et $g_j:\mathcal{X}\to\mathbb{R}$ des fonctions **convexes** pour tout $1\le j\le n$. Alors le lagrangien du problème $(\mathcal{P}_{ci})$ est une fonction convexe-concave.

Démonstration : Soit $\lambda\in(\mathbb{R}^+)^n$. Puisque les multiplicateurs de Lagrange λ_j sont positifs, la fonction

$$x\mapsto\langle\lambda,g(x)\rangle = \sum_{j=1}^n \lambda_j\,g_j(x)$$

est une fonction convexe. La somme de deux fonctions convexes étant convexe, on en déduit que, pour tout $\lambda\in(\mathbb{R}^+)^n$, la fonction partielle $x\mapsto\mathcal{L}(x;\lambda)$ est une fonction convexe. De plus, quel que soit $x\in\mathcal{X}$, la fonction partielle $\lambda\mapsto-\mathcal{L}(x;\lambda)$ est la somme d'une fonction affine et de la fonction indicatrice convexe $\chi_{(\mathbb{R}^+)^n}$; elle est donc convexe. ■

2 Solutions d'un problème sous contraintes d'inégalités

2.1 Dualité lagrangienne

Dans ce paragraphe, nous allons montrer qu'il est possible d'utiliser le lagrangien pour définir une dualité min-max. En effet, remarquons que, pour tout $x\in\mathcal{X}$,

$$\sup_{\lambda\in(\mathbb{R}^+)^n}\mathcal{L}(x;\lambda) = \sup_{\lambda\in(\mathbb{R}^+)^n\times\mathbb{R}^m}\Big\{J(x)+\langle\lambda,g(x)\rangle\Big\} = \begin{cases} J(x) & \text{si } x\in\mathcal{R} \\ +\infty & \text{sinon}\end{cases}$$

où on rappelle que $\mathcal{R}$ désigne l'ensemble admissible du problème $(\mathcal{P}_{ci})$. La fonction objectif $J+\chi_{\mathcal{R}}$ du problème non contraint peut donc s'exprimer comme l'enveloppe supérieure d'une famille de fonctions, de sorte que le problème $(\mathcal{P}_{ci})$ s'écrit

$$\min_{x\in\mathcal{X}}\ \sup_{\lambda\in(\mathbb{R}^+)^n}\ \mathcal{L}(x;\lambda)$$

Le problème dual associé est le problème de maximisation

$$\max_{\lambda \in (\mathbb{R}^+)^n} \inf_{x \in \mathcal{X}} \Big\{ J(x) + \langle \lambda, g(x) \rangle \Big\} \tag{$\mathcal{D}$}$$

Réécrivons dans ce contexte un des résultats énoncés dans le corollaire 4 du module A6.

Proposition 3

Soit $J : \mathcal{X} \mapsto \mathbb{R}$ et $g_j : \mathcal{X} \to \mathbb{R}$ des fonctions, et notons $\mathcal{L}$ le lagrangien du problème $(\mathcal{P}_{ci})$. Soit $(x^*, \lambda^*) \in \mathcal{X} \times (\mathbb{R}^+)^n$ un point-selle de $\mathcal{L}$. Alors x^* est une solution du problème $(\mathcal{P}_{ci})$.

2.2 Points-selles du lagrangien

En vertu de la proposition **??**, il suffit de nous intéresser aux points-selles (s'ils existent) du lagrangien associé au problème $(\mathcal{P}_{ci})$. On suppose que J et les contraintes g_j sont des fonctions différentiables.

> On se place ici dans le cas différentiable (tant pour J que pour les contraintes), car il s'agit d'un cadre historique pour les conditions KKT que l'on étudiera en détails dans cette section. Cependant, comme on le verra dans la section suivante, il est tout-à-fait possible d'étendre la majorité des résultats établis ici dans un cadre plus large, où J est non différentiable mais les g_j continûment différentiables (ce qui est automatiquement le cas pour des contraintes affines, par exemple).

On rappelle que, d'après la proposition 3 du module A6, si $(\bar{x}, \bar{\lambda}) \in \operatorname{dom} \mathcal{L}$ est un point-selle de $\mathcal{L}$, alors il vérifie les deux inclusions

$$0 \in \partial_x \mathcal{L}(\bar{x}; \bar{\lambda}) \qquad \text{et} \qquad 0 \in \partial_\lambda (-\mathcal{L})(\bar{x}; \bar{\lambda})$$

Puisque la fonction partielle $x \mapsto \mathcal{L}(x; \bar{\lambda})$ est définie et différentiable sur $\mathcal{X}$, la première inclusion devient

$$0 = \frac{\partial \mathcal{L}}{\partial x}(\bar{x}; \bar{\lambda}) = \nabla J(\bar{x}) + \sum_{j=1}^n \bar{\lambda}_j \, \nabla g_j(\bar{x})$$

Pour la seconde inclusion, on commence par noter que

$$\forall \, \lambda \in \mathbb{R}^n, \quad -\mathcal{L}(x; \lambda) = -J(x) + \sum_{j=1}^n \big(\underbrace{-\lambda_j \, g_j(x) - \chi_{\mathbb{R}^+}(\lambda_j)}_{= a_j(\lambda_j)} \big)$$

La somme définie donc une fonction séparable. Ainsi, d'après la proposition 15 du module A2, en notant que $J(x)$ est une constante par rapport à λ,

$$\partial_\lambda (-\mathcal{L})(x; \lambda) = \prod_{j=1}^n \Big\{ \partial a_j(\lambda_j) \Big\}$$

de sorte que la seconde inclusion vérifiée par les points-selles du lagrangien s'écrit

$$\forall \, j \in [\![1 \, ; \, n]\!], \qquad 0 \in \partial a_j(\lambda_j)$$

Pour tout $1 \leq j \leq n$, on peut vérifier que

$$\forall \, \lambda_j \in \mathbb{R}, \qquad \partial a_j(\lambda_j) = \begin{cases} \{ - g_j(x) \} & \text{si } \lambda_j > 0 \\ ] -\infty \, ; - g_j(x) \,] & \text{si } \lambda_j = 0 \\ \emptyset & \text{sinon.} \end{cases}$$

On a ainsi démontré que

Lemme 1

Soit $J : \mathcal{X} \to \mathbb{R}$ et $g_j : \mathcal{X} \to \mathbb{R}$ des fonctions différentiables. Soit $(\bar{x}, \bar{\lambda}) \in \mathcal{X} \times (\mathbb{R}^+)^n$ un point-selle du lagrangien $\mathcal{L}$ du problème $(\mathcal{P}_{ci})$. Alors

(a) $$\nabla J(\bar{x}) + \sum_{j=1}^n \bar{\lambda}_j \, \nabla g_j(\bar{x}) = 0$$

(b) $$\forall \, 1 \leq j \leq n, \qquad \begin{cases} g_j(\bar{x}) = 0 & \text{si } \bar{\lambda}_j > 0 \\ g_j(\bar{x}) \leq 0 & \text{si } \bar{\lambda}_j = 0 \end{cases}$$

Si le problème est convexe, alors on a également la réciproque (d'après la proposition 4 du module A6).

Notons que ce lemme ne garantit en aucune manière l'existence de tels points.

2.3 Conditions d'optimalité de Karush–Kuhn–Tucker

On va pouvoir à présent établir des résultats concernant les solutions d'un problème d'optimisation sous contraintes d'inégalité.

Proposition 4

Soit $J : \mathcal{X} \to \mathbb{R}$ et $g_j : \mathcal{X} \to \mathbb{R}$ des fonctions différentiables. Soit $(x^*, \lambda^*) \in \mathcal{X} \times (\mathbb{R}^+)^n$ vérifiant les deux conditions suivantes :

(a) $$\nabla J(x^*) + \sum_{j=1}^n \lambda_j^* \, \nabla g_j(x^*) = 0$$

(b) $$\forall \, 1 \leq j \leq n, \qquad \begin{cases} g_j(x^*) = 0 & \text{si } \lambda_j^* > 0 \\ g_j(x^*) \leq 0 & \text{si } \lambda_j^* = 0 \end{cases}$$

Alors x^* est un point critique de la fonction $J + \chi_{\mathcal{R}}$.

Remarque : Notons que, dans le cas convexe, il s'agit d'une conséquence de la proposition **??** et du lemme **??**.

Démonstration : Soit $(x^*, \lambda^*) \in \mathcal{X} \times (\mathbb{R}^+)^n$ vérifiant les deux conditions (a) et (b) ci-dessus. Alors d'après les calculs menés au paragraphe 2.2, on a

$$\frac{\partial \mathcal{L}}{\partial x}(x^*; \lambda^*) = 0 \qquad \text{et} \qquad 0 \in \partial_\lambda (-\mathcal{L})(x^*; \lambda^*)$$

Un développement limité de la fonction différentiable $x \mapsto \mathcal{L}(x; \lambda^*)$ assure d'une

part que
$$J(x) + \langle \lambda^*, g(x)\rangle = J(x^*) + \langle \lambda^*, g(x^*)\rangle + o(\|x - x^*\|)$$
D'autre part, la fonction $\lambda \mapsto -\mathcal{L}(x^*;\lambda)$ est convexe ; on a donc
$$\forall\,\lambda \in \mathbb{R}^n, \qquad -J(x^*) - \langle \lambda^*, g(x^*)\rangle \leq -J(x^*) - \langle \lambda, g(x^*)\rangle$$
En combinant les deux relations, on obtient
$$\forall\,\lambda \in \mathbb{R}^n, \qquad J(x) + \langle \lambda^*, g(x)\rangle \geq J(x^*) + \langle \lambda, g(x^*)\rangle + o(\|x - x^*\|)$$
Si $x \in \mathcal{R}$ et $\lambda = 0$, on a
$$J(x) \geq J(x) + \langle \lambda^*, g(x)\rangle \geq J(x^*) + o(\|x - x^*\|)$$
de sorte que $0 \in \hat{\partial}\tilde{J}(x^*) \subset \partial\tilde{J}(x^*)$ en définissant $\tilde{J} = J + \chi_{\mathcal{R}}$. ■

Autrement dit, si un tel couple (x^*, λ^*) existe, alors on peut en déduire un point critique de la fonction objectif $J + \chi_{\mathcal{R}}$ du problème non contraint, c'est-à-dire, si le problème est convexe, une solution du problème sous contraintes convexe $(\mathcal{P}_{ci})$. Une question importante est donc de savoir si de tels points existent. On va démontrer que, dans le cas convexe, c'est le cas lorsque les contraintes sont qualifiées dans le sens de Slater. Pour cela, on va commencer par établir le résultat suivant :

Proposition 5 (Inégalité d'Euler)

Soit $J : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ une fonction **convexe** et $\mathcal{R} \subset \operatorname{dom} J$ un **convexe** de $\mathcal{X}$. Soit $x^* \in \mathcal{R}$. On suppose que J est différentiable en x^*. Alors on a l'équivalence entre les affirmations suivantes :

(i) $$\forall\,x \in \mathcal{R}, \qquad J(x) \geq J(x^*)$$

(ii) $$\forall\,x \in \mathcal{R}, \qquad \langle \nabla J(x^*), x - x^*\rangle \geq 0$$

Remarque : La relation (ii) est une *inégalité variationnelle.*

Démonstration : Remarquons que la proposition (i) équivaut à dire que x^* est un minimiseur de la fonction convexe $J + \chi_{\mathcal{R}}$.

- **Sens direct.** Soit $x \in \mathcal{R}$ et $t \in \,]\,0\,;1\,]$. On a
$$t\,x + (1-t)\,x^* = t\,(x - x^* + x^*) + (1-t)\,x^* = x^* + t\,(x - x^*)$$
Par convexité de $\mathcal{R}$, on en déduit que $t\,(x - x^*) + x^* \in \mathcal{R}$. Puisque J est différentiable en x^*, on a le développement limité suivant pour tout $t \in \,]\,0\,;1\,]$:
$$J\big(x^* + t\,(x - x^*)\big) - J(x^*) = t\big(\langle \nabla J(x^*), (x - x^*)\rangle + \|x - x^*\|\,\varepsilon(t\,\|x - x^*\|)\big)$$
avec $\varepsilon(t) \to 0$ si $t \to 0$. Par l'absurde, supposons que (ii) n'est pas vérifié pour $x - x^*$. Lorsque t tend avec 0, le terme entre parenthèses dans l'égalité précédente tend donc vers une quantité strictement négative. On en déduit l'existence d'un $t_0 \in \,]\,0\,;1\,]$ tel que
$$J\big(x^* + t\,(x - x^*)\big) - J(x^*) < 0$$
ce qui contredit l'optimalité de x^*.
- **Réciproque.** Contrairement au sens direct, la réciproque nécessite l'hypothèse de convexité sur J. Celle-ci permet en effet d'écrire que
$$\forall\,x \in \mathcal{R}, \qquad J(x) \geq J(x^*) + \langle \nabla J(x^*), x - x^*\rangle \geq J(x^*)$$

ce qui démontre (i). ■

Lemme 2 (Farkas–Minkowski)

Soit $n \in \mathbb{N}$. Pour tout $(a_j)_{1\leq j\leq n} \in \mathcal{X}^n$ et $b \in \mathcal{X}$, on a équivalence entre les deux relations suivantes :
(i) pour tout $x \in \mathcal{X}$, on a l'implication suivante :
$$\big(\langle a_j, x\rangle\big)_{1\leq j\leq n} \in (\mathbb{R}^+)^n \qquad \Longrightarrow \qquad \langle b, x\rangle \geq 0$$
(ii) il existe un vecteur $(\lambda_j)_{1\leq j\leq n} \in (\mathbb{R}^+)^n$ tel que
$$b = \sum_{j=1}^n \lambda_j\,a_j$$

Démonstration : Soit $n \in \mathbb{N}$.

- **Sens direct.** On va démontrer le sens direct par l'absurde. Autrement dit, on suppose qu'il existe $\{a_j\}_{1\leq j\leq n} \in \mathcal{X}^n$ et $b \in \mathcal{X}$ tels que, pour tout $x \in \mathcal{X}$,
$$\big(\langle a_j, x\rangle\big)_{1\leq j\leq n} \in (\mathbb{R}^+)^n \qquad \Longrightarrow \qquad \langle b, x\rangle \geq 0$$
mais tels que (ii) n'est pas vraie. Commençons par introduire le cône suivant :
$$\mathcal{K} = \left\{\sum_{j=1}^n \lambda_j\,a_j \mid \forall\,1 \leq j \leq n, \quad \lambda_j \geq 0\right\}$$
La proposition (ii) équivaut à $b \in \mathcal{K}$; on suppose donc que $b \notin \mathcal{K}$. Le lecteur peut vérifier que $\mathcal{K}$ est un convexe fermé non vide (il contient en particulier le vecteur nul). On peut donc définir la projection orthogonale de b sur $\mathcal{K}$ comme le point b^+ tel que
$$b^+ = \operatorname*{argmin}_{x\in\mathcal{K}} \left\{\frac{1}{2}\,\|x - b\|^2\right\} = \operatorname{prox}_{\chi_{\mathcal{R}}}(b)$$
Ce point vérifie l'inégalité proximale (proposition 8 du module A5)
$$\forall\,x \in \mathcal{K}, \qquad \langle x - b^+, b - b^+\rangle \leq 0$$
Puisque $b \notin \mathcal{K}$, on a $\|b^+ - b\|^2 > 0$, soit, après développement du carré,
$$\|b^+\|^2 - \langle b^+, b\rangle - \langle b^+, b\rangle + \|b\|^2 > 0 \qquad \Longleftrightarrow \qquad \langle b^+, b^+ - b\rangle > \langle b, b^+ - b\rangle$$
Posons $h = b^+ - b \neq 0$. D'après ce qui précède, il existe un réel α compris strictement entre $\langle b, h\rangle$ et $\langle b^+, h\rangle$. Revenons à l'inégalité proximale. Celle-ci se lit
$$\forall\,x \in \mathcal{K}, \qquad \langle x - b^+, h\rangle \geq 0$$
soit, puisque $\langle b^+, h\rangle$ est strictement supérieur à α,
$$\forall\,x \in \mathcal{K}, \qquad \langle x, h\rangle > \alpha > \langle b, h\rangle$$
En particulier, cela implique que α est strictement négatif (il suffit de remplacer x par 0 dans l'encadrement précédent). Par conséquent, on en déduit que $\langle b, h\rangle$ est également strictement négatif. D'après l'affirmation (i), il existe donc nécessairement un indice $1 \leq j_0 \leq n$ pour lequel
$$\langle a_{j_0}, h\rangle < 0$$

Démontrons que cette situation n'est pas possible. En effet, puisque $t\,a_{j_0} \in \mathcal{K}$ quel que soit $t > 0$, on a démontré que

$$\langle t\,a_{j_0}, h\rangle > \alpha \qquad \text{soit} \qquad \langle a_{j_0}, h\rangle > \frac{\alpha}{t}$$

Il suffit alors de faire tendre t vers $+\infty$ pour obtenir la contradiction souhaitée.

- **Réciproque.** Supposons qu'il existe des coefficients positifs λ_j tels que b soit une combinaison linéaire des a_j avec ces coefficients. Soit $x \in \mathcal{X}$. Si les $\langle a_j, x\rangle$ sont positifs pour tout $1 \leq j \leq n$, alors la quantité

$$\langle b, x\rangle = \sum_{j=1}^{n} \lambda_j\,\langle a_j, x\rangle$$

est positive, ce qui démontre le résultat souhaité. ■

On peut donc à présent conclure dans le cas convexe avec le célèbre théorème suivant :

Théorème 1 (Condition d'optimalité de Karush–Kuhn–Tucker)

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** et différentiable. On suppose que les fonctions g_j sont **convexes** et différentiables pour tout $1 \leq j \leq n$, et que les contraintes du problème $(\mathcal{P}_{ci})$ satisfont la condition de Slater. Alors on a l'équivalence entre les deux affirmations suivantes :

(i) $x^* \in \mathcal{X}$ est une solution du problème $(\mathcal{P}_c)$;

(ii) il existe $\lambda^* \in \mathbb{R}^n$ satisfaisant les conditions suivantes

(a) $$\nabla J(x^*) + \sum_{j=1}^{n} \lambda_j^*\,\nabla g_j(x^*) = 0$$

(b) $$\forall\, 1 \leq j \leq n, \quad g_j(x^*) \leq 0, \quad \lambda_j^* \geq 0 \quad \text{et} \quad \lambda_j^*\,g_j(x^*) = 0$$

> La dernière condition dans le théorème précédent se traduit par le fait que si le multiplicateur de Lagrange associée à une contrainte d'inégalité est non nul à l'optimum, alors la contrainte en question est active.

Démonstration : La réciproque est une conséquence de la proposition **??**. Aussi, on se contente de démontrer le sens direct. On suppose que x^* est une solution du problème $(\mathcal{P}_{ci})$. D'après l'inégalité d'Euler (proposition **??**), on a

$$\forall\, x \in \mathcal{R}, \qquad \langle \nabla J(x^*), x - x^*\rangle \geq 0$$

Aussi, pour tout $x' \in -x^* + \mathcal{R}$, on a (a). Soit x' tel que $x' + x^* \notin \mathcal{R}$, c'est-à-dire il existe j_0 tel que $g_{j_0}(x' + x^*) > 0$. Posons $\mathcal{J}$ l'ensemble des indices correspondant aux contraintes actives en x^* :

$$\mathcal{J} = \Big\{1 \leq j \leq n \mid g_j(x^*) = 0\Big\}$$

On suppose que $$\forall\, j \in \mathcal{J}, \qquad \langle \nabla g_j(x^*), x'\rangle \leq 0$$

Montrons que, dans ce cas-là, x' est une direction admissible, c'est-à-dire que $x' + x^*$ est réalisable. Les contraintes sont convexes, donc

$$g_j(x' + x^*) \geq \langle \nabla g_j(x^*), x'\rangle$$

On peut démontrer que soit $\langle \nabla J(x^*), x'\rangle \geq 0$, soit de tels x' n'existent pas. On applique le lemme de Farkas–Minkowski aux vecteurs

$$b = \nabla J(x^*) \qquad \text{et} \qquad \forall\, 1 \leq j \leq n, \quad \{a_j\}_{1 \leq j \leq n}$$

On obtient que l'égalité (a) est satisfaite, de même que la positivité des multiplicateurs de Lagrange. Puisque x^* est admissible, on a également les inégalités $g_j(x^*) \leq 0$. ■

On voit l'importance que revêt la condition de Slater, car elle sous-tend l'existence de points-selles pour le lagrangien. 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.

Exemple

Contraintes non qualifiées. Considérons à nouveau le problème

$$\begin{array}{ll} \text{Minimiser} & J(x_1, x_2) = x_1 \\ \text{sous les contraintes} & x_2 \leq 0 \\ & \min(x_1, 0)^2 - x_2 \leq 0 \\ & (x_1, x_2) \in \mathbb{R}^2 \end{array}$$

On a vu que les contraintes de ce problème ne sont pas qualifiées. L'ensemble des points admissibles de ce problème étant les points $(x_1, 0) \in \mathbb{R}^2$ tels que $x_1 \geq 0$, il est aisé de conclure que l'unique solution de ce problème convexe est le point $(x_1^*, x_2^*) = (0, 0)$. Introduisons le lagrangien de ce problème :

$$\forall\,(x, \lambda) \in \mathbb{R}^2 \times (\mathbb{R}^+)^2, \quad \mathcal{L}(x_1, x_2; \lambda_1, \lambda_2) = x_1 + \lambda_1\,x_2 + \lambda_2\,(\min(x_1, 0)^2 - x_2)$$

Les points-selles de ce lagrangien sont exactement les points $(\bar{x}, \bar{\lambda})$ vérifiant

$$\frac{\partial \mathcal{L}}{\partial x}(\bar{x}; \bar{\lambda}) = \begin{pmatrix} 1 + 2\,\bar{\lambda}_2\,\min(\bar{x}_1, 0) \\ \bar{\lambda}_1 - \bar{\lambda}_2 \end{pmatrix} = 0$$

et les deux ensembles de conditions

$$\begin{cases} \bar{x}_2 \leq 0 \\ \bar{\lambda}_1 \geq 0 \\ \bar{\lambda}_1\,\bar{x}_2 = 0 \end{cases} \qquad \text{et} \qquad \begin{cases} \min(\bar{x}_1, 0)^2 - \bar{x}_2 \leq 0 \\ \bar{\lambda}_2 \geq 0 \\ \bar{\lambda}_2\big(\min(\bar{x}_1, 0)^2 - \bar{x}_2\big) = 0 \end{cases}$$

Puisque $\bar{\lambda}_2$ est positif, on a en particulier

$$1 + 2\,\bar{\lambda}_2\,\min(\bar{x}_1, 0) = 0 \qquad \iff \qquad \bar{x}_1 = -\frac{1}{2\,\bar{\lambda}_2} < 0$$

avec $\bar{\lambda}_2 > 0$. Ainsi, la contrainte correspondante doit être saturée, c'est-à-dire que $\bar{x}_2 = 1/(4\,\bar{\lambda}_2)^2 > 0$, ce qui n'est pas possible. Par conséquent, on en déduit que $\mathcal{L}$ ne possède aucun point-selle. En revanche, en choisissant une autre représentation du problème,

$$\begin{array}{ll} \text{Minimiser} & J(x_1, x_2) = x_1 \\ \text{sous les contraintes} & -x_1 \leq 0 \\ & \;\; x_2 \leq 0 \\ & -x_2 \leq 0 \\ & (x_1, x_2) \in \mathbb{R}^2 \end{array}$$

où les contraintes sont qualifiées, on change de lagrangien, qui vaut cette fois

$$\forall\,(x, \lambda) \in \mathbb{R}^2 \times (\mathbb{R}^+)^3, \quad \widetilde{\mathcal{L}}(x_1, x_2; \lambda_1, \lambda_2, \lambda_3) = x_1 - \lambda_1\, x_1 + \lambda_2\, x_2 - \lambda_3\, x_2$$

On pourra vérifier que les points-selles de $\widetilde{\mathcal{L}}$ sont les points

$$\left\{(0, 0, 1, \lambda_2, \lambda_3) \in \mathbb{R}^2 \times \mathbb{R}^3 \mid \lambda_2 = \lambda_3 \geq 0\right\}$$

On voit donc que la condition de qualification est suffisante pour assurer l'existence d'un point-selle pour le lagrangien. Néanmoins, elle n'est pas nécessaire. En effet, si la fonction objectif J admet un point critique x^* admissible ($0 \in \partial J(x^*)$), alors le point $(x^*, 0)$ satisfait de manière triviale la condition (a), tandis que les conditions (b) sont automatiquement vérifiées pour ce même point.

> Notons qu'une conséquence des conditions KKT est qu'à l'optimum x^*, on peut exprimer le gradient de la fonction objectif en fonction des gradients de contraintes actives :
>
> $$\nabla J(x^*) = -\sum_{\substack{j=1 \\ g_j(x^*)=0}}^{n} \lambda_j^*\, \nabla g_j(x^*)$$

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

- si $(\bar{x}, \bar{\lambda}) \in \mathcal{X} \times (\mathbb{R}^+)^n$ est un point-selle du lagrangien, alors $\bar{x}$ est une solution du problème $(\mathcal{P}_{ci})$ et il vérifie les conditions KKT ;
- si $(\bar{x}, \bar{\lambda}) \in \mathcal{X} \times (\mathbb{R}^+)^n$ vérifie les conditions KKT, alors $\bar{x}$ est un point critique de $J + \chi_{\mathcal{R}}$.

En pratique, cela implique que l'on peut essayer de trouver les points $(\bar{x}, \bar{\lambda})$ 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 $J + \chi_{\mathcal{R}}$; Il faut donc vérifier par d'autres moyens que $\bar{x}$ est bien une solution du problème sous contraintes.

2.4 Application : minimisation sur le simplexe

Intéressons-nous à la fonction suivante

$$f(x) = \langle b, x\rangle + \chi_{\Sigma_n}(x) = \sum_{i=1}^{n} b_{(i)} x_{(i)} + \chi_{\Sigma_n}(x)$$

où $b \in (\mathbb{R}^+)^n$ et le simplexe Σ_n défini par

$$\begin{aligned} \Sigma_n &= \left\{ x = (x_{(i)})_{1 \leq i \leq n} \in \mathbb{R}^n \mid \forall\, i \in [\![\, 1\,;\, n \,]\!]\,, x_{(i)} \geq 0 \text{ et } \sum_{i=1}^{n} x_{(i)} = 1 \right\} \\ &= \left\{ x \in (\mathbb{R}^+)^n \mid \|x\|_1 = 1 \right\} \end{aligned}$$

Cette fonction apparaît dans des problèmes de répartition de masse ou de votes. Un exemple simple est le suivant : une entreprise souhaite s'approvisionner en matières premières. Il a n fournisseurs, qui proposent des tarifs différents $b_{(i)} \geq 0$. La question est la suivante : à quel(s) fournisseur(s) doit-il faire appel ? Si l'on note $x_{(i)}$ la proportion des matières premières achetées auprès du fournisseur i, alors ce problème s'écrit de manière suivante :

$$\min_{x \in \mathbb{R}^n} f(x)$$

On peut vérifier qu'une solution de ce problème est

$$x^*_{(i)} = \begin{cases} 1 & \text{si } i = i_0 \\ 0 & \text{sinon} \end{cases} \qquad \text{avec} \qquad b_{(i_0)} = \min_{1 \leq i \leq n} b_{(i)}$$

On notera que, s'il existe plusieurs i_0 possibles, on en choisit un seul (de sorte que la somme des $x_{(i)}$ vaille 1). Autrement dit, il suffit de choisir le fournisseur le moins cher.

Imaginons maintenant que, pour commercer avec un fournisseur, il faut mettre en place un contrat qui pourrait être coûteux (en temps ou en énergie par exemple). On suppose par ailleurs que les prix peuvent fluctuer (tous les ans par exemple), et qu'un nouveau contrat doit être signé à chaque changement de fournisseur. On voit donc que choisir le fournisseur le moins cher peut s'avérer une mauvaise stratégie : si le fournisseur le moins cher change, il faudra signer n nouveaux contrats. On peut donc préférer répartir davantage les fournisseurs, et moduler cette répartition d'année en année. Une manière de procéder est, par exemple, de choisir une répartition initiale arbitraire, associée à un certain coût, est de faire baisser ce coût, sans pourtant atteindre le minimum. Puisque la fonction f n'est pas différentiable, on ne peut pas utiliser des méthodes de type gradient. En revanche, f est convexe, et, si l'on est capable de calculer

$$\operatorname{prox}_f(x^0) = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2}\,\|x - x^0\|^2 \right\} = \operatorname*{argmin}_{x \in \Sigma} \left\{ \langle b, x\rangle + \frac{1}{2}\,\|x - x^0\|^2 \right\}$$

on pourrait utiliser l'algorithme du point proximal par exemple (module B3). Intéressons-nous donc au problème d'optimisation sous contraintes

$$\text{Minimiser } \langle b, x\rangle \text{ sous contraintes } x \in \Sigma_n$$

On suppose que $x^0 - b$ est tel que

$$x^0_{(1)} - b_{(1)} \geq x^0_{(2)} - b_{(2)} \geq \cdots \geq x^0_{(n)} - b_{(n)}$$

Si ce n'est pas le cas, on détermine une permutation permettant d'obtenir cette relation, que l'on inversera pour obtenir le vecteur optimal x^*. Le lagrangien du problème considéré est donné par

$$\mathcal{L}(x; \lambda_1, \lambda_2, \lambda_3) = \langle b, x\rangle + \frac{1}{2}\,\|x - x^0\|^2 - \sum_{i=1}^{n} \left(\lambda_1\right)_{(i)} x_{(i)} + \lambda_2 \left(\sum_{i=1}^{n} x_{(i)} - 1\right) + \lambda_3 \left(1 - \sum_{i=1}^{n} x_{(i)}\right)$$

On a en particulier

$$\forall\, i \in [\![\, 1\,;\, n \,]\!]\,, \qquad \frac{\partial \mathcal{L}}{\partial x_{(i)}}(x; \lambda_1, \lambda_2, \lambda_3) = b_{(i)} + \frac{1}{2}\left(x_{(i)} - x^0_{(i)}\right) - (\lambda_1)_{(i)} + \lambda_2 - \lambda_3$$

et les conditions KKT s'écrivent

$$\begin{cases} \forall\, i \in [\![1\,;\, n]\!], & b_{(i)} + x^*_{(i)} - x^0_{(i)} - (\lambda^*_1)_{(i)} + \lambda^*_2 - \lambda^*_3 = 0 \\ \forall\, i \in [\![1\,;\, n]\!], & x^*_{(i)} \geq 0, \quad (\lambda_1)_{(i)} \geq 0 \quad \text{et} \quad (\lambda^*_1)_{(i)} x^*_{(i)} = 0 \\ & \displaystyle\sum_{i=1}^n x^*_{(i)} \leq 1, \quad \lambda_2 \geq 0 \quad \text{et} \quad \lambda^*_2 \left(\sum_{i=1}^n x^*_{(i)} - 1 \right) = 0 \\ & \displaystyle\sum_{i=1}^n x^*_{(i)} \geq 1, \quad \lambda^*_3 \geq 0 \quad \text{et} \quad \lambda^*_3 \left(\sum_{i=1}^n x^*_{(i)} - 1 \right) = 0 \end{cases}$$

Les deux dernières relations impliquent que

$$\sum_{i=1}^n x^*_{(i)} = 1, \quad \lambda^*_2 \geq 0 \quad \text{et} \quad \lambda^*_3 \geq 0$$

On voit donc qu'il y a aucune contrainte sur la différence $\lambda^*_2 - \lambda^*_3$, de sorte que l'on peut la remplacer par un multiplicateur de Lagrange $\mu^* \in \mathbb{R}$ non contraint à être positif.

> De manière générale, on peut traiter directement les contraintes d'égalité à l'aide d'un multiplicateur de Lagrange réel (non nécessairement positif).

Soit $i \in [\![1\,;\, n]\!]$; deux possibilités se présentent :

- soit $x^*_{(i)} \neq 0$; alors $(\lambda_1)^*_{(i)} = 0$ et la première relation des conditions KKT implique que $x^*_{(i)} = x^0_{(i)} - b_{(i)} - \mu^* > 0$;
- soit $x^*_{(i)} = 0$; alors la première relation des conditions KKT devient $(\lambda^*_1)_{(i)} = b_{(i)} - x^0_{(i)} + \mu^* \geq 0$.

On voit donc finalement que

$$x^*_{(i)} = 0 \qquad \iff \qquad b_{(i)} - x^0_{(i)} + \mu^* \geq 0$$

Puisque la somme des $x^*_{(i)}$ est contrainte à valoir 1, il n'est pas possible que tous les $x^*_{(i)}$ soient nuls. Ainsi, l'hypothèse sur l'ordre des coefficients de $b - x^0$ implique qu'il existe un indice $i_0 \in [\![1\,;\, n]\!]$ tel que

$$b_{(1)} - x^0_{(1)} + \mu^* \geq \cdots \geq b_{(i_0-1)} - x^0_{(i_0-1)} + \mu^* \geq 0 > b_{(i_0)} - x^0_{(i_0)} + \mu^* \geq \cdots \geq b_{(n)} - x^0_{(n)} + \mu^*$$

(si $i_0 = 1$, alors aucun $x^*_{(i)}$ n'est nul). Remarquons que, pour tout $i < i_0$, on a $x^*_{(i)} = 0$. On a donc

$$1 = \sum_{i=1}^n x^*_{(i)} = \sum_{i=i_0}^n x^*_{(i)} = \sum_{i=i_0}^n \left(x^0_{(i)} - b_{(i)} - \mu \right)$$

soit
$$\mu^* = \frac{1}{n - i_0 + 1} \left(\sum_{i=i_0}^n \left(x^0_{(i)} - b_{(i)} \right) - 1 \right)$$

Ainsi, pour déterminer μ^*, il suffit de vérifier pour tout $k \in [\![1\,;\, n]\!]$ si le réel μ_k

$$\mu_k = \frac{1}{n - k + 1} \left(\sum_{i=k}^n \left(x^0_{(i)} - b_{(i)} \right) - 1 \right)$$

est tel que
$$b_{(k-1)} - x^0_{(k-1)} \geq -\mu_k > b_{(k)} - x^0_{(k)}$$

Si c'est le cas, alors $\mu^* = \mu_k$ et

$$\forall\, 1 \leq i < k, \quad x^*_{(i)} = 0 \qquad \text{et} \qquad \forall\, 1 \leq k \leq n, \quad x^*_{(i)} = x^0_{(i)} - b_{(i)} - \mu^*$$

3 Lagrangien augmenté

3.1 Problème d'optimisation sous contraintes d'égalités

Plaçons-nous à présent dans le cas des problèmes d'optimisation sous contraintes d'égalités :

$$(\mathcal{P}_{ce}) \qquad \begin{array}{ll} \text{Minimiser} & J(x) \\ \text{sous les contraintes} & h_i(x) = 0, \quad i = 1, \ldots, m \\ & x \in \mathcal{X} \end{array}$$

où $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ est une fonction et les $h_i : \mathcal{X} \to \mathbb{R}$ des fonctions continûment différentiables. On notera que l'on abandonne ici l'hypothèse de différentiabilité pour la fonction J, mais que l'on renforce celles sur les contraintes.

> Les calculs menés dans cette section permettent entre autres de généraliser les résultats de la section précédente dans le cas d'un problème non différentiable.

De la même manière que pour les problèmes d'optimisation sous contraintes d'inégalité, on peut définir un lagrangien, en posant au préalable

$$\forall\, x \in \mathcal{X}, \qquad h(x) = \big(h_1(x), \ldots, h_m(x) \big)$$

Définition 3 (Lagrangien pour un problème sous contraintes d'égalité)

Soit $J : \mathcal{X} \mapsto \mathbb{R}$ et $g_j : \mathcal{X} \to \mathbb{R}$ des fonctions. Le *lagrangien* du problème $(\mathcal{P}_{ce})$ est la fonction $\mathcal{L} : \mathcal{X} \times \mathbb{R}^m \to \mathbb{R}$ définie par

$$\forall\, (x, \lambda) \in \mathcal{X} \times \mathbb{R}^m, \qquad \mathcal{L}(x; \mu) = J(x) + \langle \mu, h(x) \rangle$$

On voit en particulier que, dans ce cas, les multiplicateurs de Lagrange ne sont plus contraintes à être positifs. On peut vérifier que

$$\forall\, x \in \mathcal{R}, \qquad J(x) = \sup_{\lambda \in \mathbb{R}^m} \mathcal{L}(x; \mu)$$

où $\mathcal{R}$ désigne à nouveau l'ensemble réalisable du problème, c'est-à-dire ici les points annulant toutes les contraintes h_i. Commençons par noter (la preuve étant laissée au lecteur) que, si (x^*, μ^*) est un point-selle de $\mathcal{L}$, alors x^* est une solution de $(\mathcal{P}_{ce})$, c'est-à-dire un minimiseur de $J + \chi_{\mathbb{R}}$.

Caractérisons à l'aide du calcul sous-différentiel les points-selles de $\mathcal{L}$: puisque les fonctions h_i sont supposées continûment différentiables, on a

$$\partial_x \mathcal{L}(x; \mu) = \partial J(x) + \sum_{i=1}^m \mu_i \, \nabla h_i(x)$$

tandis que, puisque $\mu \mapsto \mathcal{L}(x; \mu)$ est affine, donc différentiable, on a

$$\partial_\mu(-\mathcal{L})(x; \mu) = -h(x) = -\partial_\mu \mathcal{L}(x; \mu)$$

Par ailleurs, toujours grâce au caractère continûment différentiables des contraintes, la fonction

$$(x, \mu) \mapsto \langle \mu, h(x) \rangle$$

est continûment différentiable. En particulier, il en découle que

$$\partial \mathcal{L}(x;\mu) = \partial_x \mathcal{L}(x;\mu) \times \partial_\mu \mathcal{L}(x;\mu)$$

On en déduit que

$$\big(0 \in \partial_x \mathcal{L}(x;\mu) \qquad \text{et} \qquad 0 \in \partial_\mu(-\mathcal{L})(x;\mu)\big) \qquad \Longleftrightarrow \qquad 0 \in \partial \mathcal{L}(x;\mu)$$

Autrement dit, tout point-selle (x^*, μ^*) de $\mathcal{L}$ est un point critique de $\mathcal{L}$, et vérifie donc

$$\begin{cases} 0 \in \partial J(x^*) + \displaystyle\sum_{i=1}^m \mu_i^* \, \nabla h_i(x^*) \\ h(x^*) = 0 \end{cases}$$

(la réciproque étant vraie si J est convexe, car $\mathcal{L}$ est dans ce cas convexe-concave).

Intéressons-nous au problème dual. La fonction duale associée à $\mathcal{P}_{ce}$ s'écrit

$$E(\mu) = \inf_{x \in \mathcal{X}} \left\{ J(x) + \sum_{i=1}^m \mu_i \, h_i(x) \right\}$$

Puisque $-E$ est l'enveloppe supérieure d'une famille de fonctions affines, il s'agit d'une fonction convexe (proposition 10 du module A1). Ainsi, le problème dual peut s'écrire comme un problème de minimisation convexe :

$$\min_{\mu \in \mathbb{R}^m} -E(\mu) \qquad \text{avec} \qquad -E(\mu) = \sup_{x \in \mathcal{X}} \Big\{ -J(x) - \langle \mu, h(x) \rangle \Big\}$$

Dans le module A5, on a vu qu'il est possible de remplacer le problème de la minimisation d'une fonction convexe par la minimisation de sa régularisation de Moreau, les deux problèmes ayant mêmes minimiseurs. Dans le cas du problème dual, cela revient à considérer le problème

$$\min_{\mu \in \mathbb{R}^m} {}^{\tau}(-E)(\mu) \tag{$\mathcal{D}_{\text{M-Y}}$}$$

pour un $\tau > 0$ donné, où ${}^{\tau}(-E)(\mu)$ est défini pour tout $\mu \in \mathbb{R}^m$ par

$$\min_{\mu' \in \mathbb{R}^m} \left\{ -E(\mu') + \frac{1}{2\,\tau} \|\mu' - \mu\|^2 \right\} = \min_{\mu' \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} \left\{ -J(x) - \langle \mu', h(x) \rangle + \frac{1}{2\,\tau} \|\mu' - \mu\|^2 \right\}$$

Soit $\mu \in \mathbb{R}^m$. Considérons la fonction de couplage

$$\widetilde{\mathcal{L}}(\mu', x) \mapsto -J(x) - \langle \mu', h(x) \rangle + \frac{1}{2\,\tau} \|\mu' - \mu\|^2$$

Commençons par remarquer que, pour tout $(x, \mu') \in \mathcal{X} \times \mathbb{R}^m$,

$$-\langle \mu', h(x) \rangle + \frac{1}{2\,\tau} \|\mu' - \mu\|^2 = \frac{1}{2\,\tau} \|\mu' - \mu - \tau\, h(x)\|^2 - \langle \mu, h(x) \rangle - \frac{\tau}{2} \|h(x)\|^2$$

de sorte que, pour tout $(x, \mu') \in \mathcal{X} \times \mathbb{R}^m$,

$$\widetilde{\mathcal{L}}(\mu', x) = -\mathcal{L}(x;\mu) + \frac{1}{2\,\tau} \|\mu' - \mu - \tau\, h(x)\|^2 - \frac{\tau}{2} \|h(x)\|^2$$

Par dualité faible, on a

$${}^{\tau}(-E)(\mu) \geq \sup_{x \in \mathcal{X}} \inf_{\mu' \in \mathbb{R}^m} \widetilde{\mathcal{L}}(\mu'; x) = \sup_{x \in \mathcal{X}} \left\{ -\mathcal{L}(x;\mu) - \frac{\tau}{2} \|h(x)\|^2 \right\}$$

Ainsi, en passant à la borne inférieure sur $\mu \in \mathbb{R}^m$, on obtient que

$$\min_{\mu \in \mathbb{R}^m} -E(\mu) = \min_{\mu \in \mathbb{R}^m} {}^{\tau}(-E)(\mu) \geq \inf_{\mu \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} \left\{ -\mathcal{L}(x;\mu) - \frac{\tau}{2} \|h(x)\|^2 \right\}$$

Toujours par dualité faible, on écrit

$$\min_{\mu \in \mathbb{R}^m} -E(\mu) \geq \sup_{x \in \mathcal{X}} \inf_{\mu \in \mathbb{R}^m} \left\{ -\mathcal{L}(x;\mu) - \frac{\tau}{2} \|h(x)\|^2 \right\} = \sup_{x \in \mathcal{X}} \left\{ -\sup_{\mu \in \mathbb{R}^m} \mathcal{L}(x;\mu) - \frac{\tau}{2} \|h(x)\|^2 \right\}$$

On reconnaît l'expression de $J(x) + \chi_{\mathcal{R}}(x)$ dans le membre de droite. Il s'ensuit que

$$\min_{\mu \in \mathbb{R}^m} -E(\mu) \geq \sup_{x \in \mathcal{X}} \inf_{\mu \in \mathbb{R}^m} \left\{ -\mathcal{L}(x;\mu) - \frac{\tau}{2} \|h(x)\|^2 \right\} = -\min_{x \in \mathcal{R}} J(x)$$

Or, par dualité forte, on sait que le minimum de $-E$, qui vaut le maximum de E, est également égal au minimum de J sur $\mathcal{R}$. Par conséquent, toutes les inégalités écrites au-dessus sont des égalités, et, en introduisant la définition suivante :

Définition 4 (Lagrangien augmenté)

Soit $\tau > 0$. Le *lagrangien augmenté* du problème $(\mathcal{P}_{ce})$ de paramètre τ est la fonction $\mathcal{L}_\tau : \mathcal{X} \times \mathbb{R}^m \to \mathbb{R}$ définie par

$$\forall\, (x, \mu) \in \mathcal{X} \times \mathbb{R}^m, \qquad \mathcal{L}_\tau(x;\mu) = \mathcal{L}(x;\mu) + \frac{\tau}{2} \|h(x)\|^2$$

on a démontré que

$$\min_{\mu \in \mathbb{R}^m} {}^{\tau}(-E)(\mu) = \inf_{\mu \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} \mathcal{L}_\tau(x;\mu) = \sup_{x \in \mathcal{X}} \inf_{\mu \in \mathbb{R}^m} \mathcal{L}_\tau(x;\mu)$$

Le lagrangien augmenté n'a donc pas de saut de dualité lorsque le lagrangien n'en a pas.

Remarque : Pour le distinguer du lagrangien augmenté, on qualifie parfois le lagrangien d'*ordinaire* ou de *classique*.

Intéressons-nous aux points-selles du lagrangien augmenté.

Proposition 6

Le lagrangien et le lagrangien augmenté d'un même problème $(\mathcal{P}_{ce})$ ont les mêmes points-selles.

Démonstration : Démontrons séparément les deux sens de l'équivalence.

- **Sens direct.** Soit $(\bar{x}, \bar{\mu}) \in \mathcal{X} \times (\mathbb{R}^+)^m$ un point-selle du lagrangien, défini par
$$\forall\, (x, \mu) \in \mathcal{X} \times \mathbb{R}^m, \qquad \mathcal{L}(\bar{x};\mu) \leq \mathcal{L}(\bar{x};\bar{\mu}) \leq \mathcal{L}(x, \bar{\mu})$$
Puisque $\bar{x}$ est une solution et, d'après le lemme **??**, tel que $h(\bar{x}) = 0$. Puisque
$$\forall\, x \in \mathcal{X}, \qquad 0 \leq \frac{1}{2\,\tau} \|h(x)\|^2$$

on en déduit que, pour tout $(x, \mu) \in \mathcal{X} \times \mathbb{R}^m$,

$$\mathcal{L}(\bar{x}; \mu) + \frac{1}{2\,\tau} \left\| h(\bar{x}) \right\|^2 \leq \mathcal{L}(\bar{x}; \bar{\mu}) + \frac{1}{2\,\tau} \left\| h(\bar{x}) \right\|^2 \leq \mathcal{L}(x, \bar{\mu}) + \frac{1}{2\,\tau} \left\| h(x) \right\|^2$$

Autrement dit, $(\bar{x}, \bar{\mu})$ est un point-selle de $\mathcal{L}_\tau$.

- **Sens réciproque.** Soit $(\bar{x}, \bar{\mu})$ est un point-selle de $\mathcal{L}_\tau$. D'après la remarque précédant la proposition courante, $\bar{\mu}$ est une solution duale de $(\mathcal{P}_e)$. Démontrons que $\bar{x}$ est une solution primale de $(\mathcal{P}_e)$. Commençons par montrer que c'est un point admissible pour le problème primal, c'est-à-dire que $h(\bar{x})$. En effet, par définition du point-selle, la fonction partielle

$$\mu \mapsto \mathcal{L}_\tau(\bar{x}; \mu) = J(\bar{x}) + \langle \mu, h(\bar{x}) \rangle + \frac{1}{2\,\tau} \left\| h(\bar{x}) \right\|^2$$

admet un maximum sur $\mathbb{R}^m$. Or, il est aisé de s'assurer que cette fonction n'est majorée que si $h(\bar{x}) = 0$. Par ailleurs, le point $\bar{x}$ est un minimiseur de la fonction partielle

$$x \mapsto \mathcal{L}_\tau(x; \bar{\mu}) = J(x) + \langle \bar{\mu}, h(x) \rangle + \frac{1}{2\,\tau} \left\| h(x) \right\|^2$$

qui est égale à la fonction partielle $x \mapsto \mathcal{L}(x, \bar{\mu})$ sur $\mathcal{R}$ l'ensemble admissible du problème primal. Or, la fonction primale vaut, sur cet ensemble,

$$J(x) = \sup_{\mu \in \mathbb{R}^m} \mathcal{L}(x; \mu) = \mathcal{L}(x; \bar{\mu}) = \mathcal{L}_\tau(x; \bar{\mu}) = \sup_{\mu \in \mathbb{R}^m} \mathcal{L}_\tau(x; \mu)$$

ce qui démontre que $\bar{x}$ est une solution primale, achevant la preuve. ■

Cette proposition assure donc que, pour trouver une solution d'un problème d'optimisation sous contraintes d'égalité, on peut, de manière équivalente, rechercher les points-selles de son lagrangien augmenté (tout comme on peut rechercher les points-selle de son lagrangien ordinaire). Suivant la méthode utilisée pour trouver de tels points, on verra que substituer le lagrangien augmenté au lagrangien ordinaire peut avoir des avantages, notamment en termes de stabilité (dus à l'ajout du terme quadratique).

Pour aller plus loin

Conditions de Karush–Kuhn–Tucker. Les conditions KKT sont classiquement utilisées en optimisation lisse sous contraintes lisses, à condition d'être capable de résoudre le système de conditions obtenus. Notons que ce système peut être ardu à résoudre.

Lagrangien et lagrangien augmenté. Le lagrangien, et surtout le lagrangien augmenté dans le cas de contraintes d'égalité, sont fréquemment utilisés pour gérer les contraintes s'écrivant de manière fonctionnelle, notamment lorsqu'elles sont affines, qu'elles soient d'inégalité ou d'égalité. Ainsi, des méthodes d'optimisation classiques s'attachent à trouver les points-selles de ces fonctions. Parmi elles, on peut citer la méthode des multiplicateurs de Lagrange, l'ADMM (méthode des directions alternées), que l'on verra en détails dans le module B5.