Module A6
Dualité min-max
Dualité lagrangienne

Sauf mention contraire, $\mathcal{X}$ (resp. $\mathcal{Y}$) est un espace de Hilbert, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\|\cdot\|$ la norme qui découle du produit scalaire.

Dans ce module, on va introduire la notion de dualité en optimisation, qui consiste à remplacer un problème d'optimisation (c'est-à-dire de recherche d'un minimiseur) par un problème de recherche de point-selle. L'exemple de la dualité lagrangienne sera étudié sous ce prisme, de même en particulier que le théorème de Karush–Kuhn–Tucker.

1 Point-selle d'une fonction convexe-concave

1.1 Définition

Définition 1 (Point-selle)

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$. On appelle *point-selle* de $\mathcal{L}$ tout point $(\bar{x}; \bar{y})$ de $\operatorname{dom} \mathcal{L}$ vérifiant l'encadrement suivant :

$$\forall \big((\bar{x}, y), (x, \bar{y})\big) \in (\operatorname{dom} \mathcal{L})^2, \qquad \mathcal{L}(\bar{x}; y) \leq \mathcal{L}(\bar{x}; \bar{y}) \leq \mathcal{L}(x; \bar{y})$$

On parle parfois de *point-col*. En anglais, on parle de *saddle point*.

Exemple

Point-selle d'une fonction convexe-concave. Considérons la fonction

$$\mathcal{L} : \left\{ \begin{array}{ccc} \mathbb{R} \times \mathbb{R} & \to & \mathbb{R} \\ (x, y) & \mapsto & x^2 - y^2 \end{array} \right.$$

Cette fonction admet un point-selle $(0, 0)$, car

$$\forall (x, y) \in \mathbb{R} \times \mathbb{R}, \qquad -y^2 \leq 0 \leq x^2$$

Il s'agit par ailleurs de l'unique point-selle de $\mathcal{L}$, puisque tout point-selle $(\bar{x}, \bar{y})$ doit vérifier

$$0 = \bar{x}^2 - \bar{x}^2 \leq \bar{x}^2 - \bar{y}^2 \leq \bar{y}^2 - \bar{y}^2 = 0$$

impliquant que $\mathcal{L}(\bar{x}; \bar{y}) = 0$. Il s'ensuit que la définition du point-selle se lit

$$\forall y \in \mathbb{R}, \qquad \bar{x}^2 \leq y^2 \qquad \text{et} \qquad \forall x \in \mathbb{R}, \qquad \bar{y}^2 \leq x^2$$

En choisissant $x = y = 0$, il en découle que $(\bar{x}, \bar{y}) = (0, 0)$.

Dans ce cours, on va s'intéresser aux points-selles d'un cas de fonctions très particulières qui apparaîtront naturellement dans les problèmes d'optimisation.

Définition 2 (Fonction convexe-concave)

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$. On dit que $\mathcal{L}$ est *convexe-concave* si

- pour tout $y \in \mathcal{Y}$, la fonction partielle $x \mapsto \mathcal{L}(x; y)$ est convexe ou identiquement égale à $-\infty$;
- pour tout $x \in \mathcal{X}$, la fonction partielle $y \mapsto \mathcal{L}(x; y)$ est concave ou identiquement égale à $+\infty$.

On dit par ailleurs que $\mathcal{L}$ est *propre* si le domaine de $\mathcal{L}$ est non vide.

Exemple

Fonctions convexes-concaves additivement séparables. Voici un exemple simple de fonctions convexes-concaves :

$$\mathcal{L}(x; y) = f(x) - g(y)$$

construites à l'aide de $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ deux fonctions convexes. La fonction $\mathcal{L}$ est propre si et seulement si les fonctions f et g le sont.

1.2 Conditions d'existence

Commençons par introduire la notion de saut de dualité.

Définition 3 (Saut de dualité)

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$ une fonction. On définit le *saut de dualité* de $\mathcal{L}$ comme étant la quantité (éventuellement infinie) donnée par

$$\mathcal{G} = \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \mathcal{L}(x; y) - \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \mathcal{L}(x; y)$$

En anglais, on parle de *duality gap*. Le saut de dualité peut prendre une valeur infinie, mais lorsqu'il est fini, il est nécessairement positif, comme on va le démontrer dans la proposition suivante.

Proposition 1 (Dualité faible)

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$ une fonction. On suppose qu'il existe un couple $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$ tel que

$$\forall\, (x, y) \in \mathcal{X} \times \mathcal{Y}, \qquad \mathcal{L}(x; y^0) > -\infty \qquad \text{et} \qquad \mathcal{L}(x^0, y) < +\infty$$

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

Remarque : La condition portant sur l'existence du couple (x^0, y^0) se traduit, dans le cas où $\mathcal{L}$ est convexe-concave, par le fait que $\mathcal{L}$ est propre.

Démonstration : Soit $(x', y') \in \mathcal{X} \times \mathcal{Y}$. Puisque, par hypothèse, $\mathcal{L}(x'; y^0) > -\infty$, on en déduit par définition de la borne supérieure que

$$\forall\, y' \in \mathcal{Y}, \qquad \mathcal{L}(x'; y') \leq \sup_{y \in \mathcal{Y}} \mathcal{L}(x'; y) \in \mathbb{R} \cup \{+\infty\}$$

On démontre de la même façon que

$$\forall\, x' \in \mathcal{X}, \qquad \mathcal{L}(x'; y') \geq \inf_{x \in \mathcal{X}} \mathcal{L}(x; y') \in \mathbb{R} \cup \{-\infty\}$$

En particulier, on a (les deux termes ci-dessous étant possiblement infinis)

$$\forall\, (x', y') \in \mathcal{X} \times \mathcal{Y}, \qquad \inf_{x \in \mathcal{X}} \mathcal{L}(x; y') \leq \sup_{y \in \mathcal{Y}} \mathcal{L}(x'; y)$$

Puisque le membre de gauche ne dépend pas de x', ni celui de droite de y', on peut passer successivement à la borne supérieure, puis la borne inférieure :

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

ce qui est le résultat souhaité. ■

Le saut de dualité donne une information importante sur l'existence d'un point-selle d'une fonction, comme en témoigne le résultat suivant.

Proposition 2 (Existence d'un point-selle)

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$ une fonction. Soit $(\bar{x}, \bar{y}) \in \operatorname{dom} \mathcal{L}$. Alors $(\bar{x}, \bar{y})$ est un point-selle de $\mathcal{L}$ si et seulement si les trois conditions suivantes sont vérifiées :

(i) la fonction $x \mapsto \sup_{y \in \mathcal{Y}} \mathcal{L}(x; y)$ atteint son minimum en $\bar{x}$;

(ii) la fonction $y \mapsto \inf_{x \in \mathcal{X}} \mathcal{L}(x; y)$ atteint son maximum en $\bar{y}$;

(iii) (*dualité forte*) son saut de dualité $\mathcal{G}$ est nul.

Dans ce cas, le point $(\bar{x}, \bar{y})$ vérifie

$$\mathcal{L}(\bar{x}; \bar{y}) = \max_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \mathcal{L}(x; y) = \min_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \mathcal{L}(x; y)$$

Démonstration : Démontrons séparément les deux sens de la proposition.

- **Sens direct.** On suppose que $\mathcal{L}$ admet un point-selle $(\bar{x}, \bar{y})$. Par définition, on a alors que

$$\forall\, (x, y) \in \mathcal{X} \times \mathcal{Y}, \qquad \mathcal{L}(\bar{x}; y) \leq \mathcal{L}(\bar{x}; \bar{y}) \leq \mathcal{L}(x; \bar{y})$$

En passant à la borne supérieure sur y dans le première inégalité, et à la borne inférieure dans la seconde, on obtient que

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

Puisque, par ailleurs,

$$\inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \mathcal{L}(\bar{x}; y) \leq \sup_{y \in \mathcal{Y}} \mathcal{L}(\bar{x}; y) \qquad \text{et} \qquad \inf_{x \in \mathcal{X}} \mathcal{L}(x; \bar{y}) \leq \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \mathcal{L}(x; \bar{y})$$

on obtient l'inégalité suivante :

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

Or, on a démontré plus haut que

$$\inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \mathcal{L}(\bar{x}; y) \geq \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \mathcal{L}(x; \bar{y})$$

On en déduit l'égalité dans cette relation. Celle-ci implique en particulier que toutes les inégalités écrites ci-dessus sont également des égalités ; on a alors notamment les deux égalités

$$\inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \mathcal{L}(\bar{x};y) = \sup_{y\in\mathcal{Y}} \mathcal{L}(\bar{x};y) \qquad \text{et} \qquad \inf_{x\in\mathcal{X}} \mathcal{L}(x;\bar{y}) = \sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \mathcal{L}(x;\bar{y})$$

ce qui implique, puisque ces deux quantités sont égales et sont finies, que les fonctions

$$x \mapsto \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y) \qquad \text{et} \qquad y \mapsto \inf_{x\in\mathcal{X}} \mathcal{L}(x;y)$$

atteignent respectivement leur minimum et leur maximum (la première en $\bar{x}$ et la seconde en $\bar{y}$), ce qui permet d'écrire

$$\inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y) = \min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y) \quad \text{et} \quad \sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \mathcal{L}(x;y) = \max_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \mathcal{L}(x;y)$$

- **Réciproque.** Supposons que le saut de dualité est nul et qu'il existe deux points $\bar{x} \in \mathcal{X}$ et $\bar{y} \in \mathcal{Y}$ tels que

$$\sup_{y\in\mathcal{Y}} \mathcal{L}(\bar{x};y) = \min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y) \qquad \text{et} \qquad \inf_{x\in\mathcal{X}} \mathcal{L}(x;\bar{y}) = \max_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \mathcal{L}(x;y)$$

Alors, par définition des bornes inférieure et supérieure, on a

$$\mathcal{L}(\bar{x};\bar{y}) \leq \sup_{y\in\mathcal{Y}} \mathcal{L}(\bar{x};y) = \min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y) = \max_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \mathcal{L}(x;y) = \inf_{x\in\mathcal{X}} \mathcal{L}(x;\bar{y})$$

puis $$\forall\, x \in \mathcal{X}, \qquad \mathcal{L}(\bar{x};\bar{y}) \leq \inf_{x\in\mathcal{X}} \mathcal{L}(x;\bar{y}) \leq \mathcal{L}(x;\bar{y})$$

On démontre de la même façon que,

$$\forall\, y \in \mathcal{Y}, \qquad \mathcal{L}(\bar{x};y) \leq \mathcal{L}(\bar{x};\bar{y})$$

ce qui assure que $(\bar{x},\bar{y})$ est un point-selle de $\mathcal{L}$. ■

REMARQUE : Puisque

$$\forall\, (\bar{x},\bar{y}) \in \mathcal{X}\times\mathcal{Y}, \qquad \inf_{x\in\mathcal{X}} \mathcal{L}(x;\bar{y}) \leq \mathcal{L}(\bar{x};\bar{y}) \leq \sup_{y\in\mathcal{Y}} \mathcal{L}(\bar{x};y)$$

on en déduit que, si $\mathcal{L}(\bar{x};\bar{y})$ est fini, alors $\bar{x}$ est un minimiseur de $x \mapsto \mathcal{L}(x;\bar{y})$ et $\bar{y}$ est un maximiseur de $y \mapsto \mathcal{L}(\bar{x};y)$. Or, si les conditions (i) et (ii) de la proposition précédente sont vérifiées, alors $\mathcal{L}(\bar{x};\bar{y})$ est finie car les deux membres extrémaux de l'encadrement plus haut sont finis. Réciproquement, si $\bar{x}$ est un minimiseur de $x \mapsto \mathcal{L}(x;\bar{y})$ et si $\bar{y}$ est un maximiseur de $y \mapsto \mathcal{L}(\bar{x};y)$, alors

$$\forall\, (x,y) \in \mathcal{X}\times\mathcal{Y}, \qquad \mathcal{L}(x;\bar{y}) \leq \mathcal{L}(\bar{x};\bar{y}) \leq L(\bar{x};y)$$

Il suffit alors de passer aux bornes supérieure et inférieure pour retrouver les conditions (i) et (ii). Autrement dit, ces deux conditions équivalent à

(i) la fonction $x \mapsto \mathcal{L}(x;\bar{y})$ atteint son minimum en $\bar{x}$;

(ii) la fonction $y \mapsto \mathcal{L}(\bar{x};y)$ atteint son maximum en $\bar{y}$.

L'annulation du saut de dualité ne permet pas à elle seule d'assurer l'existence d'un point-selle ; les deux conditions (i) et (ii) dans la proposition précédente sont essentielles, comme le montre le contre-exemple suivant.

CONTRE-EXEMPLE

Saut de dualité nul. Considérons la fonction convexe-concave

$$\mathcal{L} : \left\{ \begin{array}{ccc} \mathbb{R}\times\mathbb{R} & \to & \mathbb{R} \\ (x,y) & \mapsto & \exp(x) - y^2 \end{array} \right.$$

Calculons son saut de dualité : on a d'une part

$$\inf_{x\in\mathbb{R}} \sup_{y\in\mathbb{R}} \left\{ \exp(x) - y^2 \right\} = \inf_{x\in\mathbb{R}} \left\{ \exp(x) + \sup_{y\in\mathbb{R}} -y^2 \right\} = \inf_{x\in\mathbb{R}} \exp(x) = 0$$

et d'autre part

$$\sup_{y\in\mathbb{R}} \inf_{x\in\mathbb{R}} \left\{ \exp(x) - y^2 \right\} = \sup_{y\in\mathbb{R}} \left\{ \inf_{x\in\mathbb{R}} \exp(x) - y^2 \right\} = \sup_{y\in\mathbb{R}} \left\{ -y^2 \right\} = 0$$

Le lecteur peut vérifier que son saut de dualité est nul. Or, comme on le démontrera plus loin dans ce paragraphe (proposition 4), les points-selles d'une fonction différentiable convexe-concave sont exactement ses points critiques, c'est-à-dire ici les points annulant le gradient

$$\nabla \mathcal{L}(x;y) = \begin{pmatrix} \exp(x) \\ -2\,y \end{pmatrix}$$

qui, dans ce cas précis, n'admet pas de zéro. Cela tient au fait que la première condition dans la proposition précédente n'est pas satisfaite.

Voici une autre propriété des points-selles, qui repose sur le calcul sous-différentiel.

Proposition 3

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$ une fonction. On suppose que $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}$ est un point-selle de $\mathcal{L}$. Alors on a

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

Démonstration : Remarquons à nouveau que $\mathcal{L}$ admet un point-selle $(\bar{x}, \bar{y})$ si et seulement si $\bar{x}$ est un minimiseur de la fonction partielle $x \mapsto \mathcal{L}(x;\bar{y})$ et $\bar{y}$ est un minimiseur de la fonction partielle $y \mapsto -\mathcal{L}(\bar{x};y)$. La règle de Fermat s'écrit dans ce cas

$$0 \in \partial\big(x \mapsto \mathcal{L}(x;\bar{y})\big)(\bar{x}) \qquad \text{et} \qquad 0 \in \partial\big(y \mapsto -\mathcal{L}(\bar{x};y)\big)(\bar{y})$$

On obtient le résultat désiré en reconnaissant les sous-différentiels partiels de $\mathcal{L}$. ■

1.3 Caractérisation des points-selles

Dans le cas d'une fonction convexe-concave, la proposition précédente permet de caractériser les points-selles :

Proposition 4 (Caractérisation d'un point-selle)

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$ une fonction **convexe-concave** propre. Alors on a l'équivalence entre deux énoncés suivants :

(i) le point $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}$ est un point-selle de $\mathcal{L}$;

(ii) le point $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}$ vérifie

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

et le saut de dualité de $\mathcal{L}$ est nul.

REMARQUE : Ces deux sous-différentiels partiels sont défini comme les sous-différentiels des deux fonctions convexes $x \mapsto \mathcal{L}(x;\bar{y})$ et $y \mapsto -\mathcal{L}(\bar{x};y)$. Rappelons qu'en général,

$$\partial_y(-\mathcal{L})(x;y) \neq -\partial_y \mathcal{L}(x;y)$$

(il suffit de reprendre l'exemple de la fonction $y \mapsto -|y|$).

DÉMONSTRATION : Le sens direct a déjà été démontré, la réciproque provient de la réciproque dans la règle de FERMAT dans le cas convexe. ■

Corollaire 1

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ une fonction. On suppose qu'elle admet la décomposition suivante :

$$\forall\,(x,y) \in \mathcal{X} \times \mathcal{Y}, \qquad \mathcal{L}(x;y) = f(x) + h(x,y)$$

avec $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ une fonction continûment différentiable. On suppose que $(\bar{x},\bar{y}) \in \mathcal{X} \times \mathcal{Y}$ est un point-selle de $\mathcal{L}$. Alors on a

$$0 \in \partial\mathcal{L}(\bar{x};\bar{y})$$

Si, de plus, $\mathcal{L}$ est **convexe-concave**, alors la réciproque est vraie.

DÉMONSTRATION : Il suffit de rappeler que (cf. proposition 19 du module **A2 : Sous-différentiabilité**)

$$\forall\,(x,y) \in \operatorname{dom}\mathcal{L}, \qquad \partial\mathcal{L}(x;y) = \partial_x\mathcal{L}(x;y) \times \partial_z\mathcal{L}(x;y)$$

Par ailleurs, on a $\partial_y\mathcal{L}(x;y) = \{\nabla\mathcal{L}(x;y)\}$, de sorte que, dans ce cas précis,

$$\partial_y(-\mathcal{L})(x;y) = -\partial_y\mathcal{L}(x;y)$$

On peut alors appliquer les propositions 3 et 4. ■

En réalité, on peut démontrer que les points-selles d'une fonction convexe-concave différentiable sont exactement ses points critiques. Attention : dans le cas d'une fonction non différentiable, la caractérisation des points-selles d'une fonction convexe-concave $\mathcal{L}$ n'en fait pas nécessairement des points critiques de $\mathcal{L}$, qui sont généralement distincts des points critiques partiels.

CONTRE-EXEMPLE

Points-selles et points critiques d'une fonction non différentiable. Considérons la fonction convexe-concave suivante :

$$\mathcal{L} : \left\{ \begin{array}{ccc} \mathbb{R} \times \mathbb{R} & \to & \mathbb{R} \\ (x,y) & \mapsto & -|y| \end{array} \right.$$

Il est aisé de vérifier que ses points-selles sont les points

$$\left\{ (\bar{x},0) \in \mathbb{R}^2 \mid \bar{x} \in \mathbb{R} \right\}$$

Intéressons-nous donc à ses points critiques. Commençons par remarquer que, pour tout $(x,y) \in \mathbb{R} \times \mathbb{R}^*$, la fonction $\mathcal{L}$ est différentiable au voisinage de (x,y). Ainsi, on a

$$\partial\mathcal{L}(x;y) = \hat{\partial}\mathcal{L}(x;y) = \left\{ \nabla\mathcal{L}(x;y) \right\} = \begin{cases} \{(0,\ \ 1)\} & \text{si } y < 0 \\ \{(0,-1)\} & \text{si } y > 0 \end{cases}$$

On en déduit donc qu'aucun de ces points n'est un point critique de $\mathcal{L}$. Étudions à présent les autres points, c'est-à-dire les points-selles de $\mathcal{L}$. D'après ce qui précède, on a par passage à la limite

$$\Big\{(0,1),(0,-1)\Big\} \in \partial\mathcal{L}(x;0)$$

En réalité, on peut démontrer que $(0,0) \in \partial\mathcal{L}(x;0)$ si et seulement s'il existe une suite $((x_k,0))_{k\in\mathbb{N}}$ de $\mathbb{R}^2$ telle que $(x_k)_{k\in\mathbb{N}}$ tend vers 0 et que

$$\forall\, k \in \mathbb{N}, \exists (p_k,q_k) \in \mathbb{R}^2, \quad (p_k,q_k) \in \hat{\partial}\mathcal{L}(x_k,0) \qquad \text{avec} \qquad \lim_{k\to+\infty} (p_k,q_k) = (0,0)$$

Or, les (p_k,q_k) doivent vérifier par définition

$$-|y| \geq p_k\,(x_k - x) + o(|x-x_k|) + o(|y|)$$

En faisant tendre x vers x_k, cette inégalité se lit

$$-|y| \geq |y|\,\varepsilon(|y|)$$

avec $\varepsilon(t) \to 0$ lorsque $t \to 0$. Si $y \neq 0$, alors on peut simplifier par $|y|$, puis faire tendre y vers 0. On aboutit alors à une contradiction. On démontre de la sorte que les points-selles de $\mathcal{L}$ ne sont pas des points critiques de $\mathcal{L}$.

2 Dualité min-max en optimisation convexe

2.1 Problème dual

Dans cette section, on s'intéresse au problème de minimisation suivant

$$\min_{x\in\mathcal{X}} J(x) \tag{$\mathcal{P}$}$$

où $J : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ est une fonction. Le principe de la dualité min-max est le suivant : on cherche à représenter la fonction objectif J du problème $(\mathcal{P})$ comme le supremum d'une famille de fonctions $\mathcal{L}(\cdot;y)$ paramétrée par $y \in \mathcal{Y}$:

$$\forall\, x \in \mathcal{X}, \qquad J(x) = \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y)$$

de sorte que le problème de minimisation s'écrive comme un problème de minimisation-maximisation :

$$\min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y)$$

La fonction $\mathcal{L}$ est, dans le cas général, appelée *fonction de couplage*. On montrera un peu plus loin qu'il existe un lien entre les minimiseurs de J et les points-selles de la fonction de couplage $\mathcal{L}$ sous certaines conditions, si bien que le problème ci-dessus peut être vu comme un problème de recherche de point-selle.

La réécriture du problème d'optimisation $(\mathcal{P})$ comme un problème de minimisation-maximisation conduit naturellement à considérer le problème miroir, à savoir le problème de maximisation-minimisation suivant :

$$\max_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \mathcal{L}(x;y) \tag{$\mathcal{D}$}$$

où l'on a simplement inversé l'ordre des deux optimisations. Ce problème peut être interprété comme la maximisation de la fonction

$$E : y \mapsto \inf_{x\in\mathcal{X}} \mathcal{L}(x;y)$$

On parle alors pour le problème $(\mathcal{D})$ de *problème dual*, dont la fonction objectif est également qualifiée de *duale* ; par opposition, le problème $(\mathcal{P})$ et sa fonction objectif J sont qualifiés de *primaux*. On verra dans la suite sous quelles conditions les problèmes primaux et duaux peuvent être liés, et comment on peut exploiter ce lien pour résoudre le problème primal.

Notons que, pour un problème primal donné, il est possible de définir plusieurs problèmes duaux associés à différentes fonctions de couplage. Ainsi, dans la suite, lorsque l'on parlera *du* problème dual, on sous-entendra toujours que la fonction de couplage est connue.

2.2 Convexité des problèmes primal et dual

Les propriétés de convexité partielle de la fonction de couplage peuvent entraîner la convexité des problèmes primal et dual.

Proposition 5

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$. Si la fonction partielle $x \mapsto \mathcal{L}(x; y)$ est **convexe** pour tout $y \in \mathcal{Y}$, alors

$$x \mapsto \sup_{y \in \mathcal{Y}} \mathcal{L}(x; y)$$

est une fonction convexe.

DÉMONSTRATION : La fonction considérée dans la proposition peut être vue comme l'exemple supérieure de la famille de fonctions $x \mapsto \mathcal{L}(x; y)$ convexe pour tout $y \in \mathcal{Y}$. Elle est donc convexe en vertu de la proposition 11 du module **A1 : Éléments de topologie**. ■

On peut de même démontrer le résultat analogue suivant :

Corollaire 2

Soit $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$. Si la fonction partielle $y \mapsto \mathcal{L}(x; y)$ est concave pour tout $x \in \mathcal{X}$, alors

$$y \mapsto \inf_{x \in \mathcal{X}} \mathcal{L}(x; y)$$

est une fonction concave.

On peut remarquer que

$$\sup_{y \in \mathcal{Y}} E(y) = -\inf_{y \in \mathcal{Y}} -E(y) \qquad \text{et} \qquad \arg\max_{y \in \mathcal{Y}} E(y) = \arg\min_{y \in \mathcal{Y}} -E(y)$$

On dit alors que le problème dual est **convexe** lorsque le problème de minimisation équivalent l'est, autrement dit lorsque E est concave.

Ainsi, même si le problème primal n'est pas convexe, on peut lui associer un problème dual convexe. Par ailleurs, si J n'est pas convexe, alors il n'est pas possible de le représenter à l'aide d'une fonction de couplage convexe-concave (par exemple). En revanche, si $\mathcal{L}$ est une fonction convexe-concave, alors les problèmes primal et dual sont des problèmes de minimisation convexe.

2.3 Lien entre minimiseurs de J et points-selles de $\mathcal{L}$

Dans ce paragraphe, on va s'intéresser au lien qui peut exister entre les minimiseurs de J (en supposant qu'ils existent) et les points-selles de la fonction de couplage $\mathcal{L}$.

On suppose qu'il existe une fonction de couplage $\mathcal{L}$ telle que

$$\forall\, x \in \mathcal{X}, \qquad J(x) = \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y)$$

La proposition 2 donne alors une condition nécessaire et suffisante d'optimalité, sous des hypothèses d'existence de minimiseurs et de maximiseurs pour les problèmes primal et dual, respectivement. On peut traduire ce résultat dans le cadre de l'optimisation :

Corollaire 3

Soit $J : \mathcal{X} \mapsto \mathbb{R} \cup \{+\infty\}$ une fonction s.c.i. On suppose qu'il existe une fonction s.c.i. $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty; -\infty\}$ telle que

$$\forall\, x \in \mathcal{X}, \qquad J(x) = \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y)$$

Soit $(x^*, y^*) \in \operatorname{dom} \mathcal{L}$. Alors (x^*, y^*) est un point-selle de $\mathcal{L}$ si et seulement si les trois conditions suivantes sont vérifiées :

(i) le problème primal $(\mathcal{P})$ admet comme solution le point x^* ;

(ii) le problème dual $(\mathcal{D})$ admet comme solution le point y^* ;

(iii) les problèmes primal et dual admettent la même valeur optimale, c'est-à-dire que

$$\min_{x\in\mathcal{X}} J(x) = \max_{y\in\mathcal{Y}} E(y)$$

Démonstration : Il suffit de rappeler que

$$\forall\, x \in \mathcal{X}, \qquad \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y) = J(x) \qquad \text{et} \qquad \forall\, y \in \mathcal{Y}, \qquad \inf_{x\in\mathcal{X}} \mathcal{L}(x;y) = E(y) \;\blacksquare$$

On dit alors que x^* est une *solution primale* et y^* une *solution duale*. Les points $x \in \mathcal{X}$ (resp. $y \in \mathcal{Y}$) sont parfois appelés *variables primales* (resp. *variables duales*). Dans la dualité min-max, la dualité forte se traduit donc par le fait que problèmes primal et dual ont même valeur optimale.

Le corollaire 3 est un résultat central en optimisation, car il implique que, si les problèmes primal et dual ont même valeur optimale, alors la résolution du problème primal est équivalent à la recherche d'un point-selle.

> Si la fonction de couplage $\mathcal{L}$ admet un point-selle, alors le problème dual peut être un problème de minimisation équivalent au problème $(\mathcal{P})$. Il arrive que, pour certains problèmes d'optimisation, le problème dual présente des propriétés plus intéressantes que le problème primal, le rendant plus facile à résoudre. On peut alors être amené à résoudre préférentiellement le problème dual au lieu du problème primal. On parle alors de *résolution par dualité*. Si le problème dual est un problème convexe, tous les résultats de cette partie du cours sont valables, de même que les méthodes de résolution présentées.

3 Exemple : dualité lagrangienne

3.1 Lagrangien d'un problème d'optimisation sous contraintes

Dans cette section, on considère le problème sous contraintes suivant

$$\min_{\substack{x \in \mathcal{X} \\ g(x) \leq 0 \\ h(x) = 0}} J(x) \tag{$\mathcal{P}_{\mathrm{c}}$}$$

où $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ est la fonction objectif, $g : \mathcal{X} \to (\mathbb{R} \cup \{+\infty\})^n$ encode les contraintes d'inégalité (avec $n \in \mathbb{N}^*$) et $h : \mathcal{X} \to (\mathbb{R} \cup \{+\infty\})^m$ encode les contraintes d'égalité (avec $m \in \mathbb{N}^*$). La notation $g(x) \leq 0$ remplace la notation $g(x) \in (\left]-\infty\,;0\right])^n$. Quitte à considérer des restrictions adaptées, on suppose que J, g et h ont même domaine.

> Le fait que les contraintes d'inégalité soient écrites sous la forme de contraintes de négativité et que les contraintes d'égalité soient écrites sous forme de contraintes de nullité est essentiel pour la suite, notamment pour l'écriture des conditions d'optimalité. On dit que les contraintes sont écrites sous forme *normale*. Pour plus d'informations, le lecteur est invité à consulter le module **B5 : Théorème de Karush–Kuhn–Tucker** (3MA261).

Définition 4 (Lagrangien d'un problème sous contraintes d'inégalité et d'égalité)

Soit $J : \mathcal{X} \mapsto \mathbb{R} \cup \{+\infty\}$, $g : \mathcal{X} \to (\mathbb{R} \cup \{+\infty\})^n$ et $h : \mathcal{X} \to (\mathbb{R} \cup \{+\infty\})^m$ trois applications de même domaine. On considère le problème d'optimisation sous contraintes

$$\min_{\substack{x \in \mathcal{X} \\ g(x) \leq 0 \\ h(x) = 0}} J(x) \tag{$\mathcal{P}_{\mathrm{c}}$}$$

Le *lagrangien* du problème $(\mathcal{P}_{\mathrm{c}})$ est la fonction $\mathcal{L} : \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ définie sur $\mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m$ par

$$\mathcal{L}(x;\lambda,\mu) = \begin{cases} J(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle & \text{si } (x;\lambda,\mu) \in \operatorname{dom} J \times (\mathbb{R}^+)^n \times \mathbb{R}^m \\ +\infty & \text{si } x \notin \operatorname{dom} J \\ -\infty & \text{si } x \in \operatorname{dom} J \text{ et } \lambda \notin (\mathbb{R}^+)^n \end{cases}$$

Les coefficients des vecteurs λ et μ sont appelés *multiplicateurs de Lagrange*.

Montrons que le lagrangien du problème $(\mathcal{P}_{\mathrm{c}})$ permet de représenter la fonction objectif du problème $(\mathcal{P}_{\mathrm{c}})$. Insistons sur le fait que, de par la présence des contraintes, la fonction objectif à considérer **n'est pas** la fonction J, mais la fonction

$$J + \chi_{\mathcal{R}} \qquad \text{avec} \qquad \mathcal{R} = \Big\{ x \in \operatorname{dom} J \mid g(x) \leq 0 \text{ et } h(x) = 0 \Big\}$$

Intéressons-nous à la quantité suivante :

$$\sup_{(\lambda,\mu) \in \mathbb{R}^n \times \mathbb{R}^m} \mathcal{L}(x;\lambda,\mu) = \sup \left\{ \sup_{\substack{\lambda \in (\mathbb{R}^+)^n \\ \mu \in \mathbb{R}^m}} \mathcal{L}(x;\lambda,\mu), \sup_{\substack{\lambda \notin (\mathbb{R}^+)^n \\ \mu \in \mathbb{R}^m}} \mathcal{L}(x;\lambda,\mu) \right\}$$

Soit $x \in \mathcal{R}$. Commençons par la première borne supérieure. Par définition de $\mathcal{R}$, on a $g(x) \leq 0$ et $h(x) = 0$, de sorte que

$$\sup_{\substack{\lambda \in (\mathbb{R}^+)^n \\ \mu \in \mathbb{R}^m}} \mathcal{L}(x; \lambda, \mu) = J(x) + \sup_{\lambda \in (\mathbb{R}^+)^n} \langle \lambda, g(x) \rangle = J(x) + \sum_{j=1}^n \sup_{\lambda_i \geq 0} \lambda_j \, g_j(x)$$

Les produits $\lambda_j \, g_j(x)$ étant négatifs, ils sont majorés par 0, qui est atteint pour $\lambda_j = 0$. Ainsi,

$$\sup_{\substack{\lambda \in (\mathbb{R}^+)^n \\ \mu \in \mathbb{R}^m}} \mathcal{L}(x; \lambda, \mu) = J(x) + 0 = J(x)$$

Calculons maintenant la seconde borne supérieure. Par définition du lagrangien lorsque les multiplicateurs de Lagrange λ_j ne sont pas tous positifs, on a

$$\sup_{\substack{\lambda \notin (\mathbb{R}^+)^n \\ \mu \in \mathbb{R}^m}} \mathcal{L}(x; \lambda, \mu) = -\infty$$

Finalement, puisque $J(x) > -\infty$, on a montré que

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

Considérons maintenant le cas où $x \notin \mathcal{R}$. Si $x \notin \operatorname{dom} J$, alors par définition du lagrangien, $\mathcal{L}(x; \lambda, \mu) = +\infty$ quels que soient λ et μ. Si $x \in \operatorname{dom} J$, alors la première borne supérieure (sur les $\lambda \geq 0$) vaut $+\infty$, car il existe au moins un $i \in [\![1\,; n]\!]$ tel que $g_j(x) > 0$. La seconde borne supérieure vaut quant à elle $-\infty$ par définition du lagrangien. Ainsi, on a montré cette fois que

$$\sup_{(\lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^m} \mathcal{L}(x; \lambda, \mu) = +\infty$$

3.2 Points-selles du lagrangien

Puisque le lagrangien représente le problème $(\mathcal{P}_\mathrm{c})$ (au sens de la dualité), on peut s'intéresser à ses points-selles et au problème dual associé. *On suppose à partir de maintenant que les applications g et h sont continûment différentiables sur leur domaine.*

On rappelle que, d'après la proposition 2, si $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \operatorname{dom} \mathcal{L}$ est un point-selle de $\mathcal{L}$, alors il vérifie les deux inclusions

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

Intéressons-nous tout d'abord à la première inclusion. Par définition du domaine de $\mathcal{L}$, on a $\lambda \geq 0$. Puisque g et h sont continûment différentiables, on a la décomposition suivante :

$$\partial_x \mathcal{L}(\bar{x}; \bar{\lambda}, \bar{\mu}) = \partial J(\bar{x}) + \{\langle \bar{\lambda}, \nabla g(\bar{x}) \rangle + \langle \bar{\mu}, \nabla h(\bar{x}) \rangle\}$$

Pour la seconde inclusion, on commence par noter que

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

La somme définie donc une fonction séparable en les coefficients de λ et de μ. Ainsi, d'après la proposition 17 du module **A2 : Sous-différentiabilité**, en notant que $J(x)$ est une constante par rapport à (λ, μ), on obtient

$$\partial_{(\lambda,\mu)}(-\mathcal{L})(x; \lambda, \mu) = \prod_{j=1}^n \partial a_j(\lambda_j) \times \prod_{i=1}^m \Big\{ h_i(x) \Big\}$$

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

$$\forall\, j \in [\![1\,;\, n]\!]\,, \qquad 0 \in \partial a_j(\lambda_j) \qquad \text{et} \qquad \forall\, i \in [\![1\,;\, m]\!]\,, \qquad 0 = h_i(x)$$

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

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

On a ainsi démontré que

Proposition 6

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction sous-différentiable, $g : \mathcal{X} \to (\mathbb{R} \cup \{+\infty\})^n$ et $h : \mathcal{X} \to (\mathbb{R} \cup \{+\infty\})^m$ deux applications continûment différentiables sur leur domaine. Soit $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \operatorname{dom} J \times (\mathbb{R}^+)^n \times \mathbb{R}^m$ un point-selle du lagrangien $\mathcal{L}$ du problème $(\mathcal{P}_c)$. Alors

(a)
$$0 \in \partial J(\bar{x}) + \sum_{j=1}^{n} \left\{\bar{\lambda}_j\, \nabla g_j(\bar{x})\right\} + \sum_{i=1}^{m} \left\{\bar{\mu}_i\, \nabla h_i(\bar{x})\right\}$$

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

(c)
$$\forall\, 1 \leq i \leq m, \qquad h_i(x) = 0$$

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

Notons que ce lemme ne garantit en aucune manière l'existence de tels points. Pour étudier l'existence, il faut s'intéresser au problème dual associé. La fonction duale associée à $\mathcal{P}_c$ s'écrit

$$E(\lambda, \mu) = \inf_{x \in \mathcal{X}} \mathcal{L}(x; \lambda, \mu) = \begin{cases} \displaystyle\inf_{x \in \operatorname{dom} J} \Big\{ J(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle \Big\} & \text{si } \lambda \in (\mathbb{R}^+)^n \\ -\infty & \text{si } \lambda \notin (\mathbb{R}^+)^n \end{cases}$$

Puisque $-E$ est l'enveloppe supérieure d'une famille de fonctions affines, il s'agit d'une fonction convexe (proposition 11 du module **A1 : Éléments de topologie**). Ainsi, **quelles que soient les hypothèses du problème primal**, le problème dual est toujours convexe. Notons que s'il y a dualité forte (c'est-à-dire si le saut de dualité de $\mathcal{L}$ est nul), alors E admet nécessairement un maximum. La question de l'existence de points-selles se réduit donc ici à savoir s'il y a dualité forte (puisqu'on suppose que le problème primal admet une solution). C'est l'objet des *conditions de qualification.*

> Il existe une multitude de conditions de qualification, qui assurent donc la dualité forte, adaptées au problème étudié. On peut en citer deux :
> - condition de Slater : la fonction objectif J est de domaine ouvert et convexe, les composantes de g sont convexes, h est affine, et il existe un point admissible $x^0 \in \mathcal{R}$ et $j \in [\![1\,;\, n]\!]$ tels que $g_j(x) < 0$;
> - condition d'indépendance linéaire : en tout point $x \in \mathcal{R}$, les vecteurs $\nabla g_j(x)$ et $\nabla h_i(x)$ tels que $g_j(x) = 0$ forment une famille libre.
>
> Notons que les conditions de qualification ne sont que des conditions suffisantes. Pour plus d'informations, le lecteur est invité à consulter le module **B5 : Théorème de Karush–Kuhn–Tucker** (3MA261).

Ces considérations nous permettent d'énoncer le célèbre théorème de Karush–Kuhn–Tucker dans le cas différentiable :

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

Soit $\mathcal{U} \subset \mathcal{X}$ un ouvert non vide. Soit $J : \mathcal{U} \to \mathbb{R}$ une fonction différentiable. On suppose que les applications $g : \mathcal{U} \to \mathbb{R}^n$ et $h : \mathcal{U} \to \mathbb{R}^m$ sont continûment différentiables et que le problème $(\mathcal{P}_c)$ satisfait une condition de qualification. On suppose que le problème $(\mathcal{P}_c)$ admet une solution x^*. Alors il existe $(\lambda^*, \mu^*) \in (\mathbb{R}^+)^n \times \mathbb{R}^m$ satisfaisant les conditions suivantes

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

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

(c)
$$\forall\, 1 \leq i \leq m, \qquad h_i(x^*) = 0$$

Démonstration : Par hypothèse, le problème primal $(\mathcal{P}_c)$ admet une solution x^*. La condition de qualification assure que le problème dual admet aussi une solution (λ^*, μ^*) et que l'hypothèse de dualité forte est satisfaite. La proposition 2 permet donc de conclure à l'existence d'un point-selle (x^*, λ^*, μ^*) pour le lagrangien $\mathcal{L}$. Il suffit alors d'appliquer la proposition 6. ■

Notons que la réciproque est assurée lorsque le problème est de plus convexe (c'est-à-dire lorsque J et les composantes de g sont convexes et h affine).

3.3 Lagrangien augmenté

Dans ce paragraphe, on ne considère que des problèmes sous contraintes d'égalité :

$$\min_{\substack{x \in \mathcal{X} \\ h(x)=0}} J(x) \tag{$\mathcal{P}_{\text{ce}}$}$$

en conservant toutefois les mêmes hypothèses pour J et h.

> Il est toujours possible de convertir un problème d'optimisation sous contraintes d'inégalité en un problème équivalent d'optimisation sous contraintes d'égalité, en rajoutant des *variables d'écart*.

On rappelle que le problème dual associé s'écrit

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

Dans le module **A5 : Opérateur proximal de Moreau**, 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

$$\min_{\mu' \in \mathbb{R}^m} \left\{ -E(\mu') + \frac{1}{2\,\tau} \|\mu' - \mu\|_2^2 \right\} = \min_{\mu' \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} \left\{ -J(x) - \langle \mu', h(x) \rangle + \frac{1}{2\,\tau} \|\mu' - \mu\|_2^2 \right\}$$

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

$$\widetilde{\mathcal{L}}(\mu', x) \mapsto -J(x) - \langle \mu', h(x) \rangle + \frac{1}{2\,\tau} \|\mu' - \mu\|_2^2$$

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

$$-\langle \mu', h(x) \rangle + \frac{1}{2\,\tau} \|\mu' - \mu\|_2^2 = \frac{1}{2\,\tau} \|\mu' - \mu - \tau\, h(x)\|_2^2 - \langle \mu, h(x) \rangle - \frac{\tau}{2} \|h(x)\|_2^2$$

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

$$\widetilde{\mathcal{L}}(\mu', x) = -\mathcal{L}(x; \mu) + \frac{1}{2\,\tau} \|\mu' - \mu - \tau\, h(x)\|_2^2 - \frac{\tau}{2} \|h(x)\|_2^2$$

Par dualité faible, on a

$${}^{\tau}(-E)(\mu) \geq \sup_{x \in \mathcal{X}} \inf_{\mu' \in \mathbb{R}^m} \widetilde{\mathcal{L}}(\mu'; x) = \sup_{x \in \mathcal{X}} \left\{ -\mathcal{L}(x; \mu) - \frac{\tau}{2} \|h(x)\|_2^2 \right\}$$

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

$$\min_{\mu \in \mathbb{R}^m} -E(\mu) = \min_{\mu \in \mathbb{R}^m} {}^{\tau}(-E)(\mu) \geq \inf_{\mu \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} \left\{ -\mathcal{L}(x; \mu) - \frac{\tau}{2} \|h(x)\|_2^2 \right\}$$

Toujours par dualité faible, on écrit

$$\min_{\mu \in \mathbb{R}^m} -E(\mu) \geq \sup_{x \in \mathcal{X}} \inf_{\mu \in \mathbb{R}^m} \left\{ -\mathcal{L}(x; \mu) - \frac{\tau}{2} \|h(x)\|_2^2 \right\} = \sup_{x \in \mathcal{X}} \left\{ -\sup_{\mu \in \mathbb{R}^m} \mathcal{L}(x; \mu) - \frac{\tau}{2} \|h(x)\|_2^2 \right\}$$

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

$$\min_{\mu \in \mathbb{R}^m} -E(\mu) \geq \sup_{x \in \mathcal{X}} \inf_{\mu \in \mathbb{R}^m} \left\{ -\mathcal{L}(x; \mu) - \frac{\tau}{2} \|h(x)\|_2^2 \right\} = -\min_{x \in \mathcal{R}} J(x)$$

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 5 (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

$$\forall\, (x, \mu) \in \mathcal{X} \times \mathbb{R}^m, \qquad \mathcal{L}_\tau(x; \mu) = \mathcal{L}(x; \mu) + \frac{\tau}{2} \|h(x)\|_2^2$$

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 7

Le lagrangien et le lagrangien augmenté d'un même problème $(\mathcal{P}_{\text{ce}})$ 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 du problème associé (lemme 6), on a en particulier $h(\bar{x}) = 0$. Puisque

$$\forall\, x \in \mathcal{X}, \qquad 0 \leq \frac{1}{2\,\tau}\, \|h(x)\|_2^2$$

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

$$\mathcal{L}(\bar{x}; \mu) + \frac{1}{2\,\tau}\, \|h(\bar{x})\|_2^2 \leq \mathcal{L}(\bar{x}; \bar{\mu}) + \frac{1}{2\,\tau}\, \|h(\bar{x})\|_2^2 \leq \mathcal{L}(x, \bar{\mu}) + \frac{1}{2\,\tau}\, \|h(x)\|_2^2$$

  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

$$\mu \mapsto \mathcal{L}_\tau(\bar{x}; \mu) = J(\bar{x}) + \langle \mu, h(\bar{x}) \rangle + \frac{1}{2\,\tau}\, \|h(\bar{x})\|_2^2$$

  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

$$x \mapsto \mathcal{L}_\tau(x; \bar{\mu}) = J(x) + \langle \bar{\mu}, h(x) \rangle + \frac{1}{2\,\tau}\, \|h(x)\|_2^2$$

  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

Dualité de Fenchel. Pour les problèmes convexes non différentiables ou sous contraintes non différentiables, on peut considérer d'autres fonctions de couplage, construite à partir de la conjuguée convexe dite de Fenchel. On étudiera en détails les propriétés de cette conjuguée dans le module **A8 : Dualité de Fenchel et conjuguée convexe**.

Méthodes d'éclatements primaux. La dualité min-max intervient dans de nombreux algorithmes en optimisation convexe. Elle permet de transformer un problème en un problème équivalent (module **B4 : Éclatement primal d'opérateurs**) ou d'interpréter les itérations d'un algorithme sur le problème primal en des itérations d'un autre algorithme sur un problème dual associé (module **B5 : Éclatement primal-dual**).

Méthodes d'éclatements primaux-duaux. La recherche d'un point-selle est au cœur de certains algorithmes, connus sous le nom de *méthodes primales-duales* (module **B5 : Éclatement primal-dual**).