13 Dualité

Les sages lui réservent toujours l'autre moitié du sens que mère nature sagement redoubla.

B. Gracián (1647). Oracle manuel et art de prudence.
Traduction de B. Pelegrín [182].

Trouver une solution ou plus généralement un point stationnaire d'un problème d'optimisation avec contraintes, ici appelé *problème primal*, consiste à déterminer un point optimal $\bar{x}$ et un multiplicateur associé $\bar{\lambda}$. Les méthodes primales de résolution se concentrent sur la détermination du point optimal $\bar{x}$ en construisant une suite $\{x_k\}$ qui converge vers lui. Dans ces méthodes, le multiplicateur est souvent obtenu comme sous-produit de l'algorithme, permettant en particulier de vérifier l'optimalité de $\bar{x}$. Dans les méthodes de dualité l'effort est porté sur la recherche du multiplicateur optimal : on génère une suite de multiplicateurs $\{\lambda_k\}$ convergeant vers $\bar{\lambda}$. Dans les bons cas, la solution primale $\bar{x}$ est obtenue comme sous-produit de l'algorithme.

Pour générer la suite $\{\lambda_k\}$ approchant le multiplicateur optimal, on a besoin de formuler un problème dont $\bar{\lambda}$ est solution, c'est le *problème dual*. Il s'agit d'un problème d'optimisation, différent du problème primal, qui fait apparaître en quelque sorte un autre aspect de celui-ci. Le problème dual n'est pas déterminé de manière unique et on verra une technique permettant d'en générer autant que l'on veut. Ceux-ci sont plus ou moins adaptés à un problème donné et plus ou moins aisés à résoudre. L'intérêt des méthodes de dualité apparaît lorsqu'on peut trouver un problème dual qui est plus simple à résoudre que le problème original et (cela paraît évident) qui a un rapport avec lui.

La notion de dualité est très générale et apparaît dans divers contextes. En optimisation, elle peut s'appliquer à la minimisation de fonctions définies sur un ensemble X non vide quelconque. Celui-ci peut être $\mathbb{R}^n$, un espace fonctionnel, un ensemble discret ou une partie d'un de ces espaces éventuellement définie par des relations fonctionnelles. Nous allons garder cette généralité aussi longtemps que possible.

Les fonctions considérées dans ce chapitre seront à valeurs dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Elles peuvent donc prendre les valeurs $-\infty$ et $+\infty$. L'intérêt de cette légère généralisation vient de ce que le problème d'optimisation

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

peut alors représenter un problème d'optimisation avec contrainte. Par exemple, le problème

$$\begin{cases} \min \tilde{f}(x) \\ x \in A, \end{cases}$$

où A est une partie de $\mathbb{R}^n$ et $\tilde{f}$ est à valeurs réelles, pourra s'écrire comme ci-dessus si l'on définit

$$f = \tilde{f} + \mathcal{I}_A,$$

où $\mathcal{I}_A$ est la fonction indicatrice de l'ensemble A. On peut donc traiter dans un même cadre les problèmes avec et sans contrainte. On rappelle que le *domaine* d'une fonction $f : X \to \overline{\mathbb{R}}$ est l'ensemble des points de X où elle ne prend pas la valeur $+\infty$. On le note

$$\operatorname{dom} f := \{x \in X : f(x) < +\infty\}.$$

Nous commencerons par introduire la dualité min-max (section 13.1). Pour l'utiliser, on écrit le critère du problème d'optimisation original, qualifié ici de *primal*, comme un supremum, si bien que ce problème primal s'écrit comme l'inf-sup d'une fonction de couplage φ. Par définition, le problème *dual* s'obtient alors en inversant l'infimum et le supremum. Il s'écrit donc comme le sup-inf de la même fonction φ. Cette notion de dualité est liée au concept de point-selle. Il y a beaucoup de manières d'écrire une fonction comme un supremum et toutes n'ont pas un intérêt équivalent, parce que le problème dual résultant n'est pas plus simple à résoudre que le problème primal ou n'apporte pas d'information pertinente. À la section 13.2, nous présentons l'approche souvent fructueuse qui consiste à "plonger" le problème original dans une famille de problèmes qui sont des perturbations de celui-ci. Cela donne une procédure pour écrire f comme un supremum et se ramener ainsi à la dualité min-max. Ces notions sont ensuite appliquées au cadre particulier de la dualité de Fenchel (section 13.3) et à la minimisation de fonctions sur $\mathbb{R}^n$ en présence de contraintes d'égalité et d'inégalité (section 13.4). Finalement, deux approches numériques de résolution par dualité sont présentées (section 13.5); toutes deux recherchent un point-selle de φ : la première approche le fait par l'intermédiaire de la fonction duale, la seconde travaille directement sur φ.

On peut lire ce chapitre de deux manières différentes. Dans la première, on se passe de la notion de fonction conjuguée et on fait l'impasse sur le lien entre dualité et perturbation de problème d'optimisation (section 13.2). Pour cela, on lit d'abord la section 13.1, puis directement la section 13.4 en considérant les lagrangiens des formules (13.20) et (13.23) comme des fonctions φ (celles de la section 13.1) données. Pour avoir une connaissance plus approfondie de la dualité, il faudra passer par la notion de fonction conjuguée, que l'on pourra aborder en seconde lecture. Ce concept permet d'établir les formules des lagrangiens classique (13.20) et augmenté (13.23) à partir de perturbations du problème d'optimisation original.

13.1 Dualité min-max

La notion de dualité introduite dans cette section est très générale. On suppose donnés deux ensembles X et Y *quelconques*, qui ne doivent donc pas être des espaces vectoriels.

13.1.1 Introduction d'un problème dual

Soit $f : X \to \overline{\mathbb{R}}$ une fonction. On considère le problème d'optimisation

$$(P) \qquad \inf_{x \in X} f(x),$$

que l'on appelle *problème primal* et dont les solutions sont appelées *solutions primales.*

De manière à introduire un problème dual de (P), on cherchera à représenter $f(x)$ comme un supremum :

$$f(x) = \sup_{y \in Y} \varphi(x, y), \tag{13.1}$$

où $\varphi : X \times Y \to \overline{\mathbb{R}}$ sera appelée la *fonction de couplage* entre variables x et y. Par exemple, si $f \in \overline{\mathrm{Conv}}(\mathbb{R}^n)$, elle est l'enveloppe supérieure de ses minorantes affines (proposition 3.28). Un autre exemple, plus pratique, est donné ci-dessous (l'exemple 13.1) et nous verrons à la section 13.2, une procédure permettant d'écrire f de cette manière. Lorsque f s'écrit comme ci-dessus, le problème primal devient

$$(P) \qquad \inf_{x \in X} \sup_{y \in Y} \varphi(x, y).$$

Insistons sur le fait que cette écriture veut dire que l'on cherche à minimiser la fonction $x \mapsto \sup_{y \in Y} \varphi(x, y)$.

On appelle *problème dual* de (P) relatif à φ le problème noté (D) et défini par

$$(D) \qquad \sup_{y \in Y} \inf_{x \in X} \varphi(x, y). \tag{13.2}$$

On a donc simplement inversé l'ordre dans lequel les extremums en $x \in X$ et en $y \in Y$ sont pris. Ses solutions sont appelées *solutions duales.* Le problème dual consiste donc à minimiser ce que l'on appelle la *fonction duale* δ associée à φ :

$$\delta : y \mapsto \delta(y) := -\inf_{x \in X} \varphi(x, y).$$

On gardera à l'esprit que, pour chaque $y \in Y$, il faut résoudre un problème de minimisation pour connaître la valeur $\delta(y)$ de la fonction duale !

Les problèmes

$$\inf_{x \in X} \varphi(x, y) \qquad \text{et} \qquad \sup_{y \in Y} \varphi(x, y)$$

sont appelé respectivement *problème interne primal* en $y \in Y$ et *problème interne dual* en $x \in X$. Dans certains contextes, ils portent parfois le nom de *problèmes de Lagrange.* On note respectivement

$$\bar{X}(y) \qquad \text{et} \qquad \bar{Y}(x)$$

leur ensemble de solutions.

On peut souvent représenter f comme en (13.1) au moyen de différentes fonctions φ. A chacune d'elles correspond un problème dual différent. Il n'y a donc pas unicité du problème dual.

Exemple 13.1 Considérons le problème d'optimisation avec contraintes d'égalité

$$\inf_{\substack{x \in X \\ c(x)=0}} f(x),$$

dans lequel $c : X \to \mathbb{R}^m$. Ce problème peut s'écrire sous la forme

$$\inf_{x \in X} \tilde{f}(x),$$

si on définit $\tilde{f} = f + \mathcal{I}_{X_c}$ où $X_c = \{x : c(x) = 0\}$. L'observation essentielle est que l'on peut écrire $\tilde{f}$ comme un supremum :

$$\tilde{f}(x) = \sup_{y \in \mathbb{R}^m} \Big(f(x) + y^\top c(x) \Big).$$

Si $X = \mathbb{R}^n$, la fonction de couplage φ utilisée ici est le lagrangien du problème. Le problème primal s'écrit

$$\inf_{x \in X} \sup_{y \in \mathbb{R}^m} \Big(f(x) + y^\top c(x) \Big)$$

et le problème dual est

$$\sup_{y \in \mathbb{R}^m} \inf_{x \in X} \Big(f(x) + y^\top c(x) \Big).$$

Nous reviendrons sur cette dualisation à la section 13.4.1. □

Les algorithmes de résolution de (P) par dualité cherchent à résoudre le problème dual (D). Ils génèrent une suite $\{y_k\} \subset Y$, l'itéré suivant y_{k+1} étant obtenu à partir de l'itéré courant y_k et des informations recueillies en résolvant le problème interne primal suivant (éventuellement de manière approchée)

$$\inf_{x \in X} \varphi(x, y_k). \tag{13.3}$$

Pour que cette approche soit efficace, il faut que les conditions suivantes soient remplies.

(D_1) En résolvant (D) on résout aussi (P). Il faut donc que les deux problèmes aient un lien entre eux, ce qui n'est pas manifeste pour l'instant. Ce que l'on peut en dire dans le cadre général ci-dessus est donné à la section 13.1.2. Le concept-clé est celui de point-selle.

(D_2) Il est plus facile de résoudre (D) que (P). Pour cela, il est essentiel que, quel que soit y_k fixé, le problème interne primal (13.3) ait (en général) une solution et soit beaucoup plus simple à résoudre que le problème original. En effet, ce problème doit être résolu un grand nombre de fois (pour chaque itéré y_k).

(D_3) On aura intérêt à ce que les solutions du problème interne primal $\inf_x \varphi(x, \bar{y})$ soient solutions du problème primal.

Comme tout dépend de la fonction φ, son choix sera guidé par les trois remarques ci-dessus. Dans l'exemple 13.1, on voit que le problème interne primal est un problème sans contrainte, ce qui n'est pas le cas du problème original. C'est un avantage.

13.1.2 Liens entre problèmes primal et dual, point-selle

La proposition suivante donne une relation entre les valeurs optimales primale et duale. Il s'agit d'un résultat très général puisqu'il ne fait aucune hypothèse sur X, Y et φ.

Proposition 13.2 (dualité faible) *On a*

$$\boxed{\sup_{y\in Y}\inf_{x\in X}\varphi(x,y)\leq\inf_{x\in X}\sup_{y\in Y}\varphi(x,y).} \tag{13.4}$$

Démonstration. On a

$$\inf_{x\in X}\varphi(x,y')\leq\varphi(x',y'),\quad\forall x'\in X,\quad\forall y'\in Y.$$

En prenant le supremum en $y'\in Y$, on obtient

$$\sup_{y\in Y}\inf_{x\in X}\varphi(x,y)\leq\sup_{y\in Y}\varphi(x',y),\quad\forall x'\in X.$$

Le membre de gauche est indépendant de x', on peut donc prendre l'infimum en $x'\in X$ à droite et garder l'inégalité. Ceci conduit au résultat. □

Dans le cadre de la dualité min-max, on appelle *saut de dualité* l'écart positif entre les deux membres de (13.4) :

$$\text{Saut de dualité} := \inf_{x\in X}\sup_{y\in Y}\varphi(x,y)-\sup_{y\in Y}\inf_{x\in X}\varphi(x,y).$$

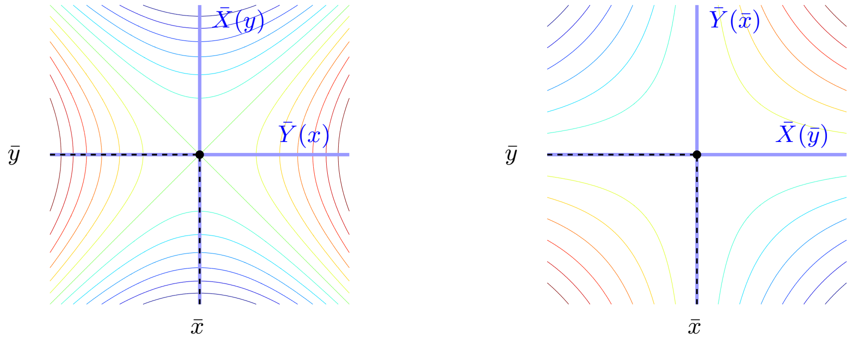
On dit qu'*il n'y a pas de saut de dualité* si ce dernier est nul. En général, lorsqu'il y a un saut de dualité, les solutions éventuelles des problèmes primal et dual n'ont pas de rapports entre elles. D'autre part, l'existence de solutions primale et duale et l'égalité en (13.4) sont étroitement liés à l'existence de point-selle de φ.

Définition 13.3 Soient X et Y deux ensembles et $\varphi : X\times Y\to\overline{\mathbb{R}}$. On dit que $(\bar{x},\bar{y})\in X\times Y$ est un *point-selle* de φ sur $X\times Y$ si on a

$$\varphi(\bar{x},y)\leq\varphi(\bar{x},\bar{y})\leq\varphi(x,\bar{y}),\quad\forall x\in X,\quad\forall y\in Y. \tag{13.5}$$

Donc $x\mapsto\varphi(x,\bar{y})$ atteint un minimum en $\bar{x}$ et $y\mapsto\varphi(\bar{x},y)$ atteint un maximum en $\bar{y}$.

Dans la définition du point-selle, on ne s'intéresse qu'aux valeurs prises par φ aux points (x,y) qui se trouvent sur les droites $\{\bar{x}\}\times Y$ et $X\times\{\bar{y}\}$. La définition est illustrée à la figure 13.1 dans le cas où $X=Y=\mathbb{R}$. On y a tracé les courbes de niveaux de deux fonctions φ. Le tracé de gauche correspond à la fonction $\varphi(x,y)=x^2-y^2$ dont le graphe se présente comme une selle, convexe suivant l'axe des abscisses et



$\varphi(x, y) = x^2 - y^2$ $\qquad$ $\varphi(x, y) = xy$

Fig. 13.1. Exemples de fonction φ avec point-selle $(\bar{x}, \bar{y}) = (0, 0)$

concave suivant l'axe des ordonnées : il y a un unique point-selle $(\bar{x}, \bar{y}) = (0, 0)$, $\bar{X}(y)$ représente les solutions de $\inf_x \varphi(x, y)$ et $\bar{Y}(x)$ représente les solutions de $\sup_y \varphi(x, y)$. Le tracé de droite est un peu moins trivial et correspond à la fonction $\varphi(x, y) = xy$: on a toujours le même unique point-selle $(\bar{x}, \bar{y}) = (0, 0)$, mais les ensembles de solutions ne sont plus des courbes puisque $\bar{X}(y)$ n'est non vide que si $y = \bar{y}$ (c'est l'ensemble $\mathbb{R}$) et $\bar{Y}(x)$ n'est non vide que si $x = \bar{x}$ (aussi l'ensemble $\mathbb{R}$).

La proposition suivante précise le lien annoncé entre la notion de point-selle et l'existence de solutions primale et duale.

Proposition 13.4 *Un couple de points $(\bar{x}, \bar{y}) \in X \times Y$ est point-selle de φ sur $X \times Y$ si et seulement si $\bar{x}$ est solution du problème primal, $\bar{y}$ est solution du problème dual et on a*

$$\sup_{y \in Y} \inf_{x \in X} \varphi(x, y) = \inf_{x \in X} \sup_{y \in Y} \varphi(x, y). \tag{13.6}$$

Dans ces conditions, la valeur en (13.6) *est* $\varphi(\bar{x}, \bar{y})$.

Démonstration. D'après l'inégalité de dualité faible (13.4) et la définition même des problèmes primal et dual, quel que soit le couple $(\bar{x}, \bar{y}) \in X \times Y$, on a

$$\inf_{x \in X} \varphi(x, \bar{y}) \leq \sup_{y \in Y} \inf_{x \in X} \varphi(x, y) \leq \inf_{x \in X} \sup_{y \in Y} \varphi(x, y) \leq \sup_{y \in Y} \varphi(\bar{x}, y). \tag{13.7}$$

Si $(\bar{x}, \bar{y}) \in X \times Y$ est un point-selle de φ, alors les membres à l'extrême gauche et à l'extrême droite dans (13.7) sont tous deux égaux à $\varphi(\bar{x}, \bar{y})$. On en déduit que l'on a égalité partout dans (13.7), c'est-à-dire que $\bar{x}$ est solution primale (par l'égalité de droite), que $\bar{y}$ est solution duale (par l'égalité de gauche) et qu'il n'y a pas de saut de dualité (par l'égalité du milieu).

Réciproquement, si $\bar{x}$ est solution du problème primal, si $\bar{y}$ est solution du problème dual et s'il n'y a pas de saut de dualité, on a égalité partout dans (13.7). Dès lors (les inégalités à gauche et à droite ci-dessous proviennent de la définition de l'infimum et du supremum)

$$\varphi(\bar{x},\bar{y}) \geq \inf_{x\in X} \varphi(x,\bar{y}) = \sup_{y\in Y} \varphi(\bar{x},y) \geq \varphi(\bar{x},\bar{y}).$$

On a donc égalité partout dans cette dernière relation. On en déduit (13.5). □

L'ensemble des points-selles d'une fonction $\varphi : X \times Y \to \overline{\mathbb{R}}$ a une structure très particulière, comme le montre le résultat suivant.

Corollaire 13.5 *L'ensemble des points-selles d'une fonction* $\varphi : X \times Y \to \overline{\mathbb{R}}$ *est un produit cartésien* $\bar{X} \times \bar{Y}$*, où* $\bar{X} \subset X$ *et* $\bar{Y} \subset Y$*. La fonction* φ *prend une valeur constante sur* $\bar{X} \times \bar{Y}$*, disons* $\bar{\varphi}$*, et on a*

$$\bar{X} = \bigcap_{y\in Y} \{x \in X : \varphi(x,y) \leq \bar{\varphi}\} \quad \textit{et} \quad \bar{Y} = \bigcap_{x\in X} \{y \in Y : \varphi(x,y) \geq \bar{\varphi}\}.$$

Démonstration. Soient (x_1, y_1) et (x_2, y_2) deux points-selles de φ. Alors, par la proposition 13.4, (x_1, y_2) est aussi un point-selle. En effet, x_1 est solution primale (car (x_1, y_1) est un point-selle), que y_2 est solution duale (car (x_2, y_2) est un point-selle) et qu'il n'y a pas de saut de dualité (car φ a un point-selle).

Par ailleurs, la valeur de φ est constante en tout point-selle. En effet, en utilisant le fait que (x_1, y_2) est un point-selle (inégalité (13.5)), on a

$$\varphi(x_1, y_1) \leq \varphi(x_1, y_2) \leq \varphi(x_2, y_2).$$

En permutant les indices, on trouve que $\varphi(x_1, y_1) = \varphi(x_2, y_2)$.

Soit $\bar{X}' = \cap_{y\in Y}\{x \in X : \varphi(x,y) \leq \bar{\varphi}\}$. Il s'agit de montrer que $\bar{X} = \bar{X}'$ (l'expression de $\bar{Y}$ s'obtient de la même manière). Si $x_1 \in \bar{X}$, alors $\varphi(x_1, y) \leq \bar{\varphi}$ pour tout $y \in Y$ (c'est l'inégalité de gauche dans (13.5)); donc $x_1 \in \bar{X}'$. Inversement, si $x_1 \in \bar{X}'$, on a $\sup_y \varphi(x_1, y) \leq \bar{\varphi} = \inf_x \sup_y \varphi(x,y)$ donc x_1 est solution primale et $x_1 \in \bar{X}$. □

Lorsque φ a un point-selle, on dispose d'une *approche duale* pour résoudre le problème (P), celle qui consiste à résoudre le problème dual : on cherche $\bar{y} \in Y$ solution de

$$\sup_{y\in Y} \left(\inf_{x\in X} \varphi(x,y) \right).$$

Considérons alors le problème interne primal en une solution duale $\bar{y}$:

$$\inf_{x\in X} \varphi(x,\bar{y}). \tag{13.8}$$

Le corollaire suivant montre que les solutions du problème primal sont solutions de ce problème.

Corollaire 13.6 *Si φ a un point-selle $(\bar{x}, \bar{y})$, alors*

$$\varnothing \neq \bar{X} \subset \bar{X}(\bar{y}) \qquad \textit{et} \qquad \varnothing \neq \bar{Y} \subset \bar{Y}(\bar{x}).$$

Démonstration. D'après la proposition 13.4, $\bar{x}$ est solution primale et $\bar{y}$ est solution duale; donc $\bar{X} \neq \varnothing$ et $\bar{Y} \neq \varnothing$. Par ailleurs, si $\hat{x}$ est solution primale, alors $(\hat{x}, \bar{y})$ est un point-selle (aussi par la proposition 13.4), et donc $\hat{x} \in \bar{X}(\bar{y})$. De même, si $\hat{y}$ est solution duale, alors $(\bar{x}, \hat{y})$ est un point-selle et donc $\hat{y} \in \bar{Y}(\bar{x})$. □

13.1.3 Existence de point-selle ▲

Nous concluons cette section par un résultat d'existence de point-selle. Commençons par une définition. Elle nous place d'emblée dans le cadre de l'analyse convexe, qui est celui du théorème 13.7 ci-dessous.

Théorème 13.7 *Supposons que X et Y soient des parties convexes fermées non vides d'espaces vectoriels normés et que φ soit convexe-concave sur $X \times Y$. On suppose également que*
(i) soit X est borné, soit il existe $y_0 \in Y$ tel que $\varphi(x, y_0) \to +\infty$ lorsque $\|x\| \to \infty$ avec $x \in X$;
(ii) soit Y est borné, soit il existe $x_0 \in X$ tel que $\varphi(x_0, y) \to -\infty$ lorsque $\|y\| \to \infty$ avec $y \in Y$.
Alors l'ensemble des points-selles de φ sur $X \times Y$ est un convexe compact non vide.

Démonstration. □

13.2 Dualité par perturbation

Dans cette section, nous décrivons une méthode permettant dans les bons cas d'écrire f comme le supremum d'une fonction, appelée *lagrangien*. On se ramène ainsi à la notion de dualité min-max introduite en section 13.1. Un intérêt de cette approche est de donner des conditions dans lesquelles il n'y a pas de saut de dualité, sans faire référence à la notion de point-selle comme à la section 13.1.2, mais en examinant l'allure de la fonction valeur associée aux perturbations.

L'idée est de plonger le *problème primal*

$$(P) \qquad \inf_{x \in X} f(x),$$

dans une famille de problèmes qui sont des perturbations de (P). On note $\mathrm{val}(P)$ la valeur de l'infimum dans (P). L'ensemble X dans (P) est toujours quelconque. Les perturbations de (P) sont introduites en se donnant une *fonction de perturbation*

$$F : X \times P \to \overline{\mathbb{R}} : (x, p) \mapsto F(x, p),$$

où P est un espace vectoriel de dimension finie, vérifiant

$$F(x, 0) = f(x), \quad \forall x \in X.$$

Cette condition montre que $x \mapsto F(x, p)$ est une perturbation de f. On considère alors la famille de problèmes

$$(P_p) \qquad \inf_{x \in X} F(x, p),$$

qui sont des perturbations de (P).

La variation de l'infimum de (P_p) avec p est représentée par la *fonction valeur* $v : P \to \overline{\mathbb{R}}$

$$v(p) = \inf_{x \in X} F(x, p).$$

Comme on a pris soin de prendre la perturbation p dans un espace vectoriel, on peut considérer la conjuguée v^* de v. On sait que cette opération ne sera utile que si v vérifie (3.22), c'est-à-dire

$$\operatorname{dom} v \neq \varnothing \quad \text{et} \quad v \text{ a une minorante affine.}$$

Dans ces conditions $v^* \in \overline{\mathrm{Conv}}(P)$ (proposition 3.27) et il y a un sens à minimiser v^* : d'une part, le caractère fermé de v^* a tendance à assurer l'existence d'une solution (par exemple si v^* tend vers $+\infty$ à l'infini) et, d'autre part, la convexité annule l'ambiguïté entre solution locale et globale. Ceci nous amène à considérer le problème

$$(D) \qquad -\inf_{\lambda \in P} v^*(\lambda) = \sup_{\lambda \in P} -v^*(\lambda).$$

Le signe "$-$" est introduit pour se ramener au cadre de la section 13.1. Ce problème est appelé *problème dual associé aux perturbations* et la fonction

$$\delta : \lambda \mapsto \delta(\lambda) = v^*(\lambda)$$

est appelée *fonction duale associée aux perturbations* : le problème dual consiste donc à minimiser la fonction duale. On note $\mathrm{val}(D)$ la valeur du supremum dans (D). On a clairement

$$\boxed{\mathrm{val}(D) = v^{**}(0)} \qquad \text{et} \qquad \boxed{\mathrm{sol}(D) = \partial v^{**}(0).} \tag{13.9}$$

La seconde identité vient du fait que $\bar{\lambda}$ est solution du problème dual si et seulement si $-v^*(\bar{\lambda}) = \sup_{\lambda \in P} -v^*(\lambda) = v^{**}(0)$, ce qui s'écrit aussi $0 = v^*(\bar{\lambda}) + v^{**}(0)$. D'après la proposition 3.37 (avec $f = v^{**}$ et $f^* = (v^{**})^* = v^*$), cette dernière égalité est équivalente au fait que $\bar{\lambda} \in \partial v^{**}(0)$.

Nous montrerons avec la proposition 13.8 que le problème primal a "quelques chances" d'être de la forme "inf sup" et que le problème dual est bien de la forme "sup inf". Auparavant, explicitons l'expression de $v^*(\lambda)$, en tenant compte de la forme particulière de v. On a

$$\begin{aligned}
v^*(\lambda) &= \sup_{p\in P}\Big(\langle\lambda,p\rangle - v(p)\Big)\\
&= \sup_{p\in P}\Big(\langle\lambda,p\rangle - \inf_{x\in X} F(x,p)\Big)\\
&= \sup_{p\in P}\sup_{x\in X}\Big(\langle\lambda,p\rangle - F(x,p)\Big) \qquad (13.10)\\
&= F^*(0,\lambda).
\end{aligned}$$

Dans le calcul précédent, en (13.10), le supremum en p dépend essentiellement de la perturbation que l'on se donne. Si cette perturbation est simple, il sera en général possible de donner une expression explicite de la valeur du supremum en p. Cela nous conduit à introduire la notion de *lagrangien associé aux perturbations*, qui est la fonction

$$\ell : X \times P \to \overline{\mathbb{R}},$$

définie par

$$\boxed{\ell(x,\lambda) = -\sup_{p\in P}\Big(\langle\lambda,p\rangle - F(x,p)\Big).} \qquad (13.11)$$

Ce lagrangien dépend des perturbations choisies et n'est donc pas nécessairement identique au lagrangien que nous avons introduit en (4.24) pour écrire les conditions d'optimalité. Nous verrons plus loin (à la section 13.4.1) que l'on peut choisir des perturbations telles que le lagrangien (13.11) redonne sur son domaine de définition le lagrangien (4.24).

Remarquons bien que, dans le couple (x,λ) dont dépend ℓ, x est une variable primale, appartenant à l'espace sur lequel est défini le problème (P), et λ est une variable duale, variant dans l'espace dual de l'espace des perturbations.

Proposition 13.8 (dualité faible) *On a*

$$\begin{aligned}
\mathrm{val}(D) = &\sup_{\lambda\in P}\inf_{x\in X}\ell(x,\lambda)\\
\le &\inf_{x\in X}\sup_{\lambda\in P}\ell(x,\lambda) \qquad (13.12)\\
\le &\,\mathrm{val}(P)
\end{aligned}$$

Démonstration. L'égalité $(13.12)_a$ vient de la définition de (D) et de ce que

$$-v^*(\lambda) = -\sup_{x\in X}(-\ell(x,\lambda)) = \inf_{x\in X}\ell(x,\lambda).$$

L'inégalité $(13.12)_b$ découle de la proposition 13.2. Enfin l'inégalité $(13.12)_c$ est obtenue comme suit

$$\sup_{\lambda\in P}\ell(x,\lambda) = \sup_{\lambda\in P}\inf_{p\in P}(F(x,p) - \langle\lambda,p\rangle) \le \sup_{\lambda\in P}F(x,0) = f(x).$$

$\square$

Définition 13.9 On appelle *saut de dualité*, la quantité positive $\mathrm{val}(P) - \mathrm{val}(D)$. On dit qu'*il n'y a pas de saut de dualité* si le saut de dualité est nul, c'est-à-dire si

$$\mathrm{val}(D) = \mathrm{val}(P).$$

Pour ne pas avoir de saut de dualité, il faudra d'abord avoir égalité en $(13.12)_c$, ce qui est le cas si l'application

$$p \mapsto F_x(p) := F(x,p)$$

vérifie

$$F_x^{**}(0) = F_x(0).$$

On a en effet $f(x) = F_x(0)$ et, d'après (13.11), $\ell(x,\lambda) = -F_x^*(\lambda)$ et donc

$$\sup_{\lambda \in P} \ell(x,\lambda) = \sup_{\lambda \in P}(-F_x^*(\lambda)) = F_x^{**}(0).$$

En particulier, on aura donc égalité en $(13.12)_c$ si la perturbation apportée au problème primal est telle que l'application $F_x \in \overline{\mathrm{Conv}}(P)$ ou encore que $F_x \in \mathrm{Conv}(P)$ et $\partial F_x(0) \neq \varnothing$ (corollaire (3.42)). Dans ce cas on est ramené à la notion de dualité introduite à la section 13.1 puisque $f(x)$ s'exprime comme un supremum.

Observons d'autre part que l'inégalité $\mathrm{val}(D) \leq \mathrm{val}(P)$ s'écrit aussi $v^{**}(0) \leq v(0)$. Il n'y aura donc pas de saut de dualité, si la fonction valeur et sa biconjuguée (qui la convexifie) coïncide en 0. Les propriétés démontrées au chapitre 3 conduisent aux résultats suivants.

Proposition 13.10 (existence de solution duale et saut de dualité)
1) *Si la fonction valeur* $v \in \mathrm{Conv}(P)$ *et* $0 \in \mathrm{dom}\, v$*, on a*
   a) $\mathrm{val}(P) = \mathrm{val}(D) \iff v$ *est s.c.i. en* 0*,*
   b) $\partial v(0) \neq \varnothing \iff \mathrm{val}(P) = \mathrm{val}(D)$ *et* $\mathrm{sol}(D) \neq \varnothing$*,*
   et dans ces cas $\mathrm{sol}(D) = \partial v(0)$*,*
   c) si $0 \in (\mathrm{dom}\, v)^{\ominus}$*, alors* $\mathrm{val}(P) = \mathrm{val}(D)$*,* $\mathrm{sol}(D) = \partial v(0)$ *est non vide et* v *est s.c.i. en* 0*,*
   d) si $0 \in (\mathrm{dom}\, v)^{\circ}$*, alors* $\mathrm{val}(P) = \mathrm{val}(D)$*,* $\mathrm{sol}(D) = \partial v(0)$ *est non vide et compact et* v *est lipschitzienne dans un voisinage de* 0*.*
2) *Si* $v \in \overline{\mathrm{Conv}}(P)$*, alors* $\mathrm{val}(P) = \mathrm{val}(D) \in \overline{\mathbb{R}}$*.*

DÉMONSTRATION. 1.*a*) C'est une récriture du corollaire 3.29.

1.*b*) Si $\partial v(0) \neq \varnothing$, alors $v(0) = v^{**}(0)$ (corollaire 3.42, point 2) et $\partial v(0) = \partial v^{**}(0)$ (corollaire 3.42, point 3). L'égalité $v(0) = v^{**}(0)$ montre qu'il n'y a pas de saut de dualité et le fait que $\partial v(0) \neq \varnothing$ implique alors que $\mathrm{sol}(D) \equiv \partial v^{**}(0)$ est non vide.

Réciproquement, $v(0) = v^{**}(0)$ (pas de saut de dualité) implique que $\partial v(0) = \partial v^{**}(0)$ (corollaire 3.42, point 3). Comme $\mathrm{sol}(D) \equiv \partial v^{**}(0)$ est non vide, il en est de même de $\partial v(0)$.

1.*c*) Si $0 \in (\mathrm{dom}\, v)^{\ominus}$, v est sous-différentiable en 0 (proposition 3.43). On applique alors les points 1.*a* et 1.*b*.

1.*d*) Si $0 \in (\mathrm{dom}\, v)^{\circ}$, alors v est lipschitzienne dans un voisinage de 0 (proposition 3.17) et $\partial v(0)$ est non vide et compact (propoistion 3.44).

2) Si $v \in \overline{\mathrm{Conv}}(P)$, $v^{**} = v$ (proposition 3.28), donc $v(0) = v^{**}(0)$, c'est-à-dire qu'il n'y a pas de saut de dualité (mais les valeurs optimales peuvent être infinies). $\square$

13.3 Dualité de Fenchel

On se donne deux espaces euclidiens E et F (produits scalaires tous deux notés $\langle \cdot, \cdot \rangle$), une fonction $f : E \to \mathbb{R} \cup \{+\infty\}$, une application linéaire $A : E \to F$ (on note A^* son adjointe) et une fonction $g : F \to \mathbb{R} \cup \{+\infty\}$. On considère le problème, dit *primal*,

$$(P_F) \qquad \inf_{x \in E} \Big(f(x) + g(Ax) \Big). \tag{13.13}$$

Introduisons un problème dual par perturbation de Ax. On considère le fonction valeur

$$v_F : F \to \overline{\mathbb{R}} : p \mapsto v_F(p) := \inf_{x \in E} \Big(f(x) + g(Ax + p) \Big). \tag{13.14}$$

Sa conjuguée s'écrit

$$\begin{aligned} v_F^*(\lambda) &= \sup_{x \in E} \sup_{p \in F} \Big(\langle \lambda, p \rangle - f(x) - g(Ax + p) \Big) \\ &= \sup_{x \in E} \Big(-\langle \lambda, Ax \rangle - f(x) + g^*(\lambda) \Big) \\ &= f^*(-A^*\lambda) + g^*(\lambda). \end{aligned}$$

Le problème dual consiste à maximiser $-v_F^*(-\lambda)$ (on fait un changement de variable $\lambda \curvearrowright -\lambda$) :

$$(D_F) \qquad \sup_{\lambda \in F} \Big(-f^*(A^*\lambda) - g^*(-\lambda) \Big). \tag{13.15}$$

Un résultat de dualité faible découle directement de la proposition 13.8.

Proposition 13.11 (dualité de Fenchel faible) *On a* $\mathrm{val}(D_F) \leq \mathrm{val}(P_F)$.

Pour obtenir un résultat de dualité forte, assurant en particulier l'absence de saut de dualité, $\mathrm{val}(P_L) = \mathrm{val}(D_F)$, nous allons analyser la régularité de la fonction valeur v_F et utiliser les résultats du chapitre 3. Calculons d'abord le domaine de v_F. Clairement, $p \in \mathrm{dom}\, v_F$ si et seulement si il existe un $x \in \mathrm{dom}\, f$ tel que $Ax + p \in \mathrm{dom}\, g$ ou encore $p \in \mathrm{dom}\, g - A(\mathrm{dom}\, f)$. Nous avons donc montré que

$$\mathrm{dom}\, v_F = \mathrm{dom}\, g - A(\mathrm{dom}\, f).$$

On comprend donc pourquoi l'ensemble à droite intevient dans le résultat suivant.

Théorème 13.12 (dualité de Fenchel forte) *On suppose que* $f \in \mathrm{Conv}(E)$, *que* $g \in \mathrm{Conv}(F)$, *que* $\mathrm{val}(P) \in \mathbb{R}$ *et que*

$$0 \in \Big(\mathrm{dom}\, g - A(\mathrm{dom}\, f)\Big)^{\circ}. \tag{13.16}$$

Alors $\mathrm{val}(P) = \mathrm{val}(D)$ *et* $\mathrm{sol}(D)$ *est non vide et compact.*

Démonstration. En tant que fonction marginale définie par (13.14), avec $f \in \mathrm{Conv}(E)$ et $g \in \mathrm{Conv}(F)$, v_F est une fonction convexe. Elle est aussi propre car $v_F(0) \in \mathbb{R}$ et, par (13.16), $0 \in (\mathrm{dom}\, v_F)^{\circ}$ (donc v_F ne peut pas prendre la valeur $-\infty$). Le résultat se déduit alors de la proposition 13.10, point 1.*d*. □

Ce résultat a de multiples applications. Nous en donnons deux à titre d'illustration. Le corollaire 13.13 étend des résultats des propositions 3.51 et 3.52 au cas de fonctions pouvant prendre la valeur $+\infty$. Le corollaire 13.14 étudie la dualité de Fenchel pour des problèmes avec contraintes linéaires d'égalité.

Corollaire 13.13 *Dans le cadre défini ci-dessus, avec* f *et* g *convexes, on a*

$$\partial(f + g \circ A)(x) \supset \partial f(x) + A^* \partial g(Ax),$$

avec égalité si la condition de stabilité (13.16) *a lieu.*

Démonstration. Soit $x \in \mathrm{dom}\, f$ tel que $\partial f(x) \neq \varnothing$ et $\partial g(Ax) \neq \varnothing$ (sinon, l'inclusion est triviale). Si $x^* \in \partial f(x)$ et $y^* \in \partial g(Ax)$, alors $x \in \mathrm{dom}\, f$, $Ax \in \mathrm{dom}\, g$ et quel que soit $d \in E$, on a

$$f'(x; d) \geq \langle x^*, d\rangle \quad \text{et} \quad g'(Ax; d) \geq \langle y^*, d\rangle.$$

Alors

$$\Big(f + g \circ A\Big)'(x; d) = f'(x; d) + g'(Ax; Ad) \geq \langle x^* + A^* y^*, d\rangle.$$

Donc $x^* + A^* y^* \in \partial(f + g \circ A)(x)$ et l'inclusion a lieu.

Soient à présent $\bar{x} \in \mathrm{dom}(f + g \circ A)$ et $\bar{x}^* \in \partial(f + g \circ A)(\bar{x})$, ce qui s'écrit aussi $0 \in \partial(f + g \circ A)(\bar{x}) - \bar{x}^*$. Alors $\bar{x}$ est solution du problème

$$\inf_{x \in E} \Big(\tilde{f}(x) + g(Ax)\Big),$$

où $\tilde{f}(x) := f(x) - \langle \bar{x}^*, x\rangle$. Comme $\tilde{f}^*(x^*) = f^*(x^* + \bar{x}^*)$, le dual de ce problème s'écrit

$$\sup_{\lambda \in F} \Big(-f^*(A^*\lambda + \bar{x}^*) - g^*(-\lambda)\Big).$$

Si la condition de stabilité (13.16) a lieu ($\mathrm{dom}\, \tilde{f} = \mathrm{dom}\, f$), les valeurs optimales des problèmes ci-dessus sont égales (d'après le théorème). Comme elles sont finies, le problème dual a une solution, disons $\bar{\lambda}$ et on a

$$f(\bar{x}) + f^*(A^*\bar{\lambda} + \bar{x}^*) + g(A\bar{x}) + g^*(-\bar{\lambda}) = \langle \bar{x}^*, \bar{x}\rangle.$$

En utilisant le fait que $f(\bar{x}) + f^*(A^*\bar{\lambda} + \bar{x}^*) \geq \langle A^*\bar{\lambda} + \bar{x}^*, \bar{x}\rangle$ et $g(A\bar{x}) + g^*(-\bar{\lambda}) \geq -\langle A^*\bar{\lambda}, \bar{x}\rangle$, on obtient

$$\langle \bar{x}^*, \bar{x}\rangle \leq f(\bar{x}) + f^*(A^*\bar{\lambda} + \bar{x}^*) + g(A\bar{x}) + g^*(-\bar{\lambda}).$$

On a donc des égalités dans les trois inégalités précédentes. Celles-ci impliquent que $A^*\bar{\lambda} + \bar{x}^* \in \partial f(\bar{x})$ et $-\bar{\lambda} \in \partial g(A\bar{x})$. Donc $\bar{x}^* \in \partial f(\bar{x}) + A^*\partial g(A\bar{x})$, ce qui conclut la démonstration. $\square$

Corollaire 13.14 *On suppose donnés $f : E \to \mathbb{R} \cup \{+\infty\}$, une application linéaire $A : E \to F$ et un élément $b \in F$. Alors*

$$\sup_{y\in F}\Big(\langle b, y\rangle - f^*(A^*y)\Big) \quad \leq \quad \inf_{\substack{x\in E\\ Ax=b}} f(x)$$

Si f et g sont convexes et si $b \in (A(\operatorname{dom} f))^\circ$, alors on a égalité ci-dessus et, lorsqu'il est fini, le supremum à gauche est atteint.

Démonstration. On considère le problème (13.13) avec $g = \mathcal{I}_{\{b\}}$, la fonction indicatrice du singleton $\{b\}$. On a $g^*(y^*) = \langle b, y^*\rangle$ et le dual de (13.13) s'écrit comme à gauche dans l'inégalité ci-dessus. Enfin la condition de stabilité (13.16) s'écrit $0 \in (b - A(\operatorname{dom} f))^\circ$, c'est-à-dire $b \in (A(\operatorname{dom} f))^\circ$. $\square$

13.4 Dualisation de contraintes fonctionnelles

Dans cette section, on considère le problème d'optimisation sous contraintes fonctionnelles suivant

$$(P) \qquad \begin{cases} \min f(x) \\ x \in X \\ c_i(x) = 0,\ \forall i \in E \\ c_i(x) \leq 0,\ \forall i \in I, \end{cases} \tag{13.17}$$

où X est toujours un ensemble quelconque et E et I sont des ensembles finis d'indices, contenant respectivement $m_E = |E|$ et $m_I = |I|$ éléments, formant une partition de $\{1, \ldots, m\}$. Donc $m = m_E + m_I$ et on note $c_E = \{c_i\}_{i\in E}$ et $c_I = \{c_i\}_{i\in I}$. La fonction f et les fonctions c_i, $i \in E \cup I$, sont définies sur X à valeurs dans $\mathbb{R}$.

Nous introduisons deux problèmes duaux de (13.17) en utilisant deux types de problèmes perturbés. Le lagrangien correspondant aux premières perturbations est très voisin du lagrangien de la remarque 4.24 servant à écrire les conditions d'optimalité. Cette approche est bien adaptée aux problèmes convexes. En revanche, lorsque le problème (13.17) n'est pas convexe, c'est le second type de perturbations qui s'avère le plus approprié. Celles-ci conduisent au lagrangien augmenté.

Les problèmes duaux introduits ci-dessous sont obtenus en perturbant les contraintes fonctionnelles de (13.17), sans toucher à la “contrainte” $x \in X$ considérée comme “contrainte dure”. On dit que l'on *dualise* les contraintes fonctionnelles, celles définies au moyen des fonctions c_E et c_I. Comme X est un ensemble quelconque, la démarche suivie englobe le cas où on ne dualiserait qu'une partie des contraintes fonctionnelles. En effet, X peut être un ensemble défini par d'autres contraintes d'égalité ou d'inégalité et peut donc prendre en compte les contraintes non dualisées. Cette remarque soulève la question du choix des contraintes à dualiser dans un cas concret. Sur ce sujet, ce sont les conditions (D_1)–(D_3) de la page 386 qui servent de guide. Comme les contraintes non dualisées (celles prisent en compte par X) se retrouvent dans les problèmes interne primal (13.3), on cherchera à dualiser les contraintes difficilement traitées par un algorithme de minimisation.

13.4.1 Le lagrangien classique

Pour $p_i \in \mathbb{R}$, $i \in E \cup I$, on note $p_E = \{p_i\}_{i\in E}$, $p_I = \{p_i\}_{i\in I}$, $p = (p_E, p_I)$ et on introduit le problème perturbé suivant :

$$\begin{cases} \min f(x) \\ x \in X \\ c_E(x) + p_E = 0 \\ c_I(x) + p_I \leq 0. \end{cases} \tag{13.18}$$

La fonction de perturbation F_0 est donc définie sur $X \times \mathbb{R}^m$ par

$$F_0(x,p) = \begin{cases} f(x) \text{ si } \ c_E(x) + p_E = 0 \ \text{ et } \ c_I(x) + p_I \leq 0 \\ +\infty \ \text{sinon.} \end{cases} \tag{13.19}$$

Le lagrangien associé à ces perturbations se calcule au moyen de la formule (13.11). Avec $\lambda = (\lambda_E, \lambda_I) \in \mathbb{R}^m$, on trouve

$$\begin{aligned} \ell_0(x,\lambda) &= -\sup_{p=(p_E,p_I)} \Big(\langle \lambda, p \rangle - F_0(x,p) \Big) \\ &= -\sup_{\substack{p=(p_E,p_I) \\ p_E=-c_E(x) \\ p_I \leq -c_I(x)}} \Big(\langle \lambda, p \rangle - f(x) \Big). \end{aligned}$$

Finalement

$$\boxed{\ell_0(x,\lambda) = \begin{cases} f(x) + \displaystyle\sum_{i\in E\cup I} \lambda_i c_i(x) \text{ si } \ \lambda_I \geq 0 \\ -\infty \qquad\qquad\qquad\qquad \text{sinon.} \end{cases}} \tag{13.20}$$

Pour λ tel que $\lambda_I \geq 0$, on retrouve le lagrangien de la remarque 4.24, servant à exprimer les conditions d'optimalité. Celui-ci correspond donc aux perturbations (13.18) du problème (13.17).

Le problème dual s'écrit

$$\sup_{\substack{\lambda=(\lambda_E,\lambda_I) \\ \lambda_I \geq 0}} \inf_{x\in X} \ell_0(x,\lambda) \tag{13.21}$$

On notera que les seules contraintes qui portent sur les variables duales sont des contraintes de positivité sur λ_I. Ces contraintes sont simples, surtout si on les compare aux contraintes (éventuellement non linéaires) d'égalité et d'inégalité du problème original (13.17).

Le problème primal (13.17) s'écrit aussi

$$\inf_{x \in X} \sup_{\substack{\lambda=(\lambda_E,\lambda_I)\\ \lambda_I \geq 0}} \ell_0(x,\lambda),$$

ce qui montre que l'on a égalité en $(13.12)_c$. On remarque en effet que cette quantité s'écrit

$$\begin{aligned}
&\inf_{x\in X} \sup_{\substack{\lambda\\ \lambda_I \geq 0}} \left(f(x) + \sum_{i \in E \cup I} \lambda_i c_i(x) \right)\\
&= \inf_{x \in X} \begin{cases} f(x) \text{ si } \ c_E(x) = 0 \ \text{ et } \ c_I(x) \leq 0 \\ +\infty \ \text{ sinon} \end{cases}\\
&= \operatorname{val}(P).
\end{aligned}$$

La proposition suivante analyse le lien entre les points-selles du lagrangien (13.20) et les solutions de (13.17). Schématiquement, les points-selles du lagrangien classique sont des solutions de (13.17) et la réciproque est vraie si le problème est convexe. Le point 2 de ce résultat apporte des précisions sur ce qui est affirmé à la proposition 12.14 : sous des hypothèses de convexité, non seulement $\bar{x}$ minimise le lagrangien $\ell_0(\cdot, \bar{\lambda})$, mais $(\bar{x}, \bar{\lambda})$ est un point-selle de ℓ_0.

Proposition 13.15 1) *Si* $(\bar{x}, \bar{\lambda})$ *est point-selle du lagrangien* (13.20) *sur* $X \times \mathbb{R}^m$*, alors* $\bar{x}$ *est solution de* (13.17)*. Si, de plus,* $X = \mathbb{R}^n$ *et* f *et* c *sont différentiables, alors* $\bar{\lambda}$ *est un multiplicateur de Lagrange tel que les conditions d'optimalité* (4.23) *soient vérifiées avec* $(x_*, \lambda_*) = (\bar{x}, \bar{\lambda})$.

2) *Réciproquement, supposons que* $X = \mathbb{R}^n$*, que le problème* (13.17) *soit convexe au sens de la définition* 4.17*, que* $\bar{x}$ *en soit une solution et que* f *et* c *soient différentiables en* $\bar{x}$*. Supposons également qu'il existe un multiplicateur* $\bar{\lambda}$ *tel que les conditions d'optimalité de Karush, Kuhn et Tucker* (4.23) *aient lieu. Alors,* $(\bar{x}, \bar{\lambda})$ *est un point-selle du lagrangien* (13.20)*.*

Démonstration. 1) Si $(\bar{x}, \bar{\lambda})$ est point-selle du lagrangien (13.20) sur $X \times \mathbb{R}^m$, $\bar{x}$ est solution du problème primal (proposition 13.4), qui n'est autre que (13.17).

Si, de plus, $X = \mathbb{R}^n$ et f et c sont différentiables, le fait que $\bar{x}$ minimise $\ell_0(\cdot, \bar{\lambda})$ sur $\mathbb{R}^n$ implique que $\nabla_x \ell_0(\bar{x}, \bar{\lambda}) = 0$. Il reste à montrer $\bar{\lambda}_I \geq 0$ et la complémentarité $\bar{\lambda}_I^\top c_I(\bar{x}) = 0$. Pour cela, on exploite l'inégalité

$$\ell_0(\bar{x}, \lambda) \leq \ell_0(\bar{x}, \bar{\lambda}), \quad \forall \lambda \in \mathbb{R}^m.$$

En prenant $\lambda = 0$, on a $f(\bar{x}) \leq \ell_0(\bar{x}, \bar{\lambda})$, ce qui montre que $\bar{\lambda}_I \geq 0$. Comme $c_E(\bar{x}) = 0$, on a donc

$$\lambda_I^\top c_I(\bar{x}) \leq \bar{\lambda}_I^\top c_I(\bar{x}), \quad \forall \lambda_I \in \mathbb{R}_+^{m_I}.$$

En y prenant $\lambda_I = 0$ et $\lambda_I = 2\bar{\lambda}_I$, on obtient $\bar{\lambda}_I^\top c_I(\bar{x}) = 0$.

2) La minimalité de $\ell_0(\cdot, \bar{\lambda})$ en $\bar{x}$ a été démontrée par la proposition 12.14. Il reste donc à montrer que $\ell_0(\bar{x}, \cdot)$ est maximisé en $\bar{\lambda}$. Soit $\lambda \in \mathbb{R}^m$. Comme $\bar{\lambda}_I \geq 0$, on a clairement $\ell_0(\bar{x}, \lambda) \leq \ell_0(\bar{x}, \bar{\lambda})$ si $\lambda_I \not\geq 0$. Si $\lambda_I \geq 0$, on a

$$\begin{aligned}
\ell_0(\bar{x}, \lambda) &= f(\bar{x}) + \sum_{i \in I} \lambda_i c_i(\bar{x}) \qquad [\text{car } c_E(\bar{x}) = 0] \\
&\leq f(\bar{x}) \qquad [\text{car } \lambda_i c_i(\bar{x}) \leq 0 \text{ pour } i \in I] \\
&= f(\bar{x}) + \sum_{i \in E \cup I} \bar{\lambda}_i c_i(\bar{x}) \qquad [\text{par complémentarité}] \\
&= \ell_0(\bar{x}, \bar{\lambda}).
\end{aligned}$$

□

D'après les propositions 13.4 et 13.15, on voit que, pour les problèmes *convexes*, il n'y a pas de saut de dualité et que l'on peut trouver $\bar{\lambda}$ en résolvant le problème dual (13.21). Sans convexité, il peut y avoir ou ne pas y avoir de saut de dualité comme le montre l'exemple simple suivant.

Exemple 13.16 Pour la fonction f définie par

$$f(x) = \begin{cases} 1 & \text{si } x \leq 0 \\ 1 - x & \text{si } 0 < x < 1 \\ x - 1 & \text{si } x \geq 1 \end{cases}$$

et $\alpha \in \mathbb{R}$, on considère le problème d'optimisation trivial

$$\min_{x = \alpha} f(x),$$

dont la solution est bien sûr $\bar{x} = \alpha$. La fonction valeur correspondant aux perturbations comme en (13.18) vaut $v(p) = f(\alpha - p)$ et sa biconjuguée est $v^{**}(p) = f^{**}(\alpha - p)$. Ces fonctions sont représentées à la figure 13.2. Il y a un saut de dualité ssi

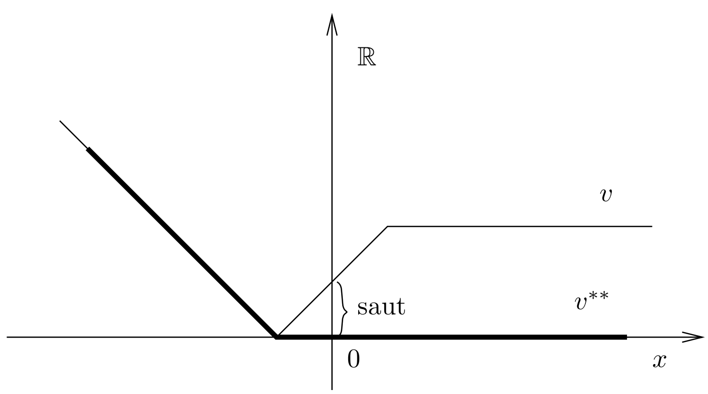


Fig. 13.2. Saut de dualité ($\alpha = 1/2$)

$v(0) \neq v^{**}(0)$, c'est-à-dire ici ssi $\alpha < 1$.

On peut aussi calculer la fonction duale

$$\delta(\lambda) = -\inf_{x\in\mathbb{R}} \ell_0(x,\lambda) = \begin{cases} \lambda(\alpha-1) & \text{si } -1 \leq \lambda \leq 0 \\ \infty & \text{sinon.} \end{cases}$$

On en déduit que les solutions duales sont

$$(\bar{\lambda}, \delta(\bar{\lambda})) = \begin{cases} (0,0) & \text{si } \alpha < 1 \\ ([-1,0],0) & \text{si } \alpha = 1 \\ (-1, 1-\alpha) & \text{si } \alpha > 1. \end{cases}$$

Ce n'est que pour $\alpha \geq 1$ que l'on a $\delta(\bar{\lambda}) = -f(\bar{x})$ $[= 1-\alpha]$ et donc que pour ces valeurs de α qu'il n'y a pas de saut de dualité. □

13.4.2 Le lagrangien augmenté

On considère à présent un autre type de perturbation du problème (13.17), bien adapté, comme on le verra, aux problèmes non convexes. Le but visé est de convexifier la fonction valeur dans un voisinage de 0. Nous retrouverons le lagrangien augmenté, déjà introduit à la section 12.4.

Soit $r \geq 0$. Pour $p = (p_E, p_I)$, on introduit le problème perturbé suivant

$$(P_p) \qquad \begin{cases} \min f(x) + \frac{r}{2}\|p\|_2^2 \\ x \in X \\ c_E(x) + p_E = 0 \\ c_I(x) + p_I \leq 0. \end{cases}$$

On a noté $\|\cdot\|_2$ la norme ℓ_2. Ce problème équivaut à minimiser en x la fonction définie sur $X \times \mathbb{R}^m$ par

$$F_r(x,p) = \begin{cases} f(x) + \frac{r}{2}\|p\|_2^2 & \text{si } \ c_E(x) + p_E = 0 \ \text{ et } \ c_I(x) + p_I \leq 0 \\ +\infty & \text{sinon.} \end{cases}$$

Observons que la fonction valeur correspondant à ces perturbations s'écrit

$$v_r(p) = v_0(p) + \frac{r}{2}\|p\|_2^2. \tag{13.22}$$

Ceci montre qu'en prenant r suffisamment grand on convexifie v_0 dans un voisinage de $p = 0$. C'est grâce à cela qu'on pourra éviter un saut de dualité, *localement*.

Calculons le lagrangien ℓ_r associé aux perturbations introduites. Pour $\lambda = (\lambda_E, \lambda_I)$, il s'écrit

$$\begin{aligned} \ell_r(x,\lambda) &= -\sup_{p=(p_E,p_I)} \Big(\langle\lambda,p\rangle - F_r(x,p)\Big) \\ &= f(x) - \sup_{\substack{p=(p_E,p_I) \\ p_E=-c_E(x) \\ p_I\leq -c_I(x)}} \Big(\langle\lambda,p\rangle - \frac{r}{2}\|p\|_2^2\Big). \end{aligned}$$

Le supremum en p_I est atteint pour $p_I = \min(\lambda_I/r, -c_I(x))$, si bien que l'on a

$$\boxed{\begin{aligned}\ell_r(x,\lambda) = f(x) &+ \sum_{i\in E}\left[\lambda_i c_i(x) + \frac{r}{2}c_i(x)^2\right] \\ &+ \sum_{i\in I}\left[\lambda_i \max\left(\frac{-\lambda_i}{r}, c_i(x)\right) + \frac{r}{2}\left(\max\left(\frac{-\lambda_i}{r}, c_i(x)\right)\right)^2\right].\end{aligned}} \tag{13.23}$$

On a donc retrouvé le lagrangien augmenté défini en (12.24). Remarquons que, contrairement au lagrangien classique, le lagrangien augmenté ne prend que des valeurs finies. Cette formule est compliquée, surtout du fait de la présence des contraintes d'inégalité. S'il n'y a que des contraintes d'égalité, elle devient

$$\ell_r(x,\lambda) = f(x) + \lambda^\top c(x) + \frac{r}{2}\|c(x)\|_2^2,$$

qui est le lagrangien classique avec pénalisation quadratique des contraintes.

Le problème dual s'écrit comme d'habitude

$$\sup_{\lambda=(\lambda_E,\lambda_I)} \inf_{x\in X} \ell_r(x,\lambda)$$

et le problème primal s'écrit

$$\inf_{x\in X} \sup_{\lambda=(\lambda_E,\lambda_I)} \ell_r(x,\lambda).$$

Pour ce dernier, on constate en effet que

$$\begin{aligned}&\sup_{\lambda=(\lambda_E,\lambda_I)} \ell_r(x,\lambda) \\ &= f(x) + \sum_{i\in E}\sup_{\lambda_i}\left[\lambda_i c_i(x) + \frac{r}{2}c_i(x)^2\right] \qquad (13.24)\\ &\quad + \sup_{\lambda_I\in\mathbb{R}^{m_I}}\left[-\sum_{\substack{i\in I\\ rc_i(x)+\lambda_i\leq 0}}\frac{(\lambda_i)^2}{2r} + \sum_{\substack{i\in I\\ rc_i(x)+\lambda_i>0}}\left(\lambda_i c_i(x) + \frac{r}{2}c_i(x)^2\right)\right].\end{aligned}$$

Le supremum en λ_i $(i \in E)$ du premier crochet vaut $+\infty$ si $c_i(x) \neq 0$ et vaut 0 si $c_i(x) = 0$. Le supremum en λ_i $(i \in I)$ du second crochet vaut $+\infty$ si $c_i(x) > 0$ et 0 si $c_i(x) \leq 0$ [il est atteint en tout $\lambda_i \geq 0$ si $c_i(x) = 0$ et en $\lambda_i = 0$ si $c_i(x) < 0$]. Dès lors, on retrouve le problème primal :

$$\inf_{x\in X} \sup_{\lambda=(\lambda_E,\lambda_I)} \ell_r(x,\lambda) = \inf_{x\in X}\begin{cases} f(x) \text{ si } \ c_E(x) = 0 \ \text{ et } \ c_I(x) \leq 0\\ +\infty \ \text{ sinon.}\end{cases}$$

On a donc, ici aussi, égalité en $(13.12)_c$.

La proposition suivante donne des conditions pour ne pas avoir de saut de dualité localement. Pour les problèmes non convexes sur $\mathbb{R}^n$, on peut donc avoir des résultats de dualité locaux via le lagrangien augmenté. Ce résultat renforce celui du théorème 12.17 en affirmant que, non seulement $\ell_r(\cdot, \bar{\lambda})$ est minimisée localement en $\bar{x}$, mais que $\ell_r(\bar{x}, \cdot)$ est maximisée en $\bar{\lambda}$.

Proposition 13.17 *On considère le problème* (13.17) *avec* $X = \mathbb{R}^n$ *et on suppose que* f *et* $c_{E\cup I^0(\bar{x})}$ *sont deux fois dérivables en un minimum local* $\bar{x}$ *de ce problème. On suppose également que les conditions de* (*KKT*) *ont lieu en* $(x_*, \lambda_*) = (\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m$ *et que la condition suffisante d'optimalité du second ordre semi-forte* (4.41) *a lieu pour un certain multiplicateur optimal* $\bar{\lambda}$*. Alors, il existe un réel* $\bar{r} > 0$ *et un voisinage* V *de* $\bar{x}$*, tels que pour tout* $r \geq \bar{r}$*, tout* $x \in V \setminus \{\bar{x}\}$ *et tout* $\lambda \in \mathbb{R}^m$*, on ait*

$$\ell_r(\bar{x}, \lambda) \leq \ell_r(\bar{x}, \bar{\lambda}) < \ell_r(x, \bar{\lambda}). \tag{13.25}$$

DÉMONSTRATION. La minimalité stricte de $\ell_r(\cdot, \lambda)$ en $\bar{x}$ sur V, pour r au-delà d'un certain seuil $\bar{r} > 0$, a été établie au théorème 12.17. Il reste donc à montrer la maximalité de $\ell_r(\bar{x}, \cdot)$ en $\bar{\lambda}$. Ici le seuil $\bar{r} > 0$ n'est pas nécessaire : un $r > 0$ suffit.

Soient $\lambda \in \mathbb{R}^m$ et $r > 0$. D'après (13.23), on a

$$\ell_r(\bar{x}, \lambda) = f(\bar{x}) + \sum_{\substack{i\in I \\ rc_i(\bar{x})+\lambda_i>0}} \left(\lambda_i c_i(\bar{x}) + \frac{r}{2} c_i(\bar{x})^2\right) - \sum_{\substack{i\in I \\ rc_i(\bar{x})+\lambda_i\leq 0}} \frac{(\lambda_i)^2}{2r}.$$

Le dernier terme est clairement négatif. Quant au second, en utilisant $rc_i(\bar{x}) + \lambda_i > 0$ et $c_i(\bar{x}) \leq 0$, on voit que chaque terme de la somme est $\leq -rc_i(\bar{x})^2 + \frac{r}{2} c_i(\bar{x})^2 \leq 0$. Donc

$$\ell_r(\bar{x}, \lambda) \leq f(\bar{x}).$$

On conclut en observant que $f(\bar{x}) = \ell_r(\bar{x}, \bar{\lambda})$. En effet, il suffit d'utiliser dans l'expression de $\ell_r(\bar{x}, \bar{\lambda})$ ci-dessus le fait que $\bar{\lambda}_i c_i(\bar{x}) = 0$, que $rc_i(\bar{x}) + \bar{\lambda}_i > 0$ implique que $\bar{\lambda}_i > 0$ et donc $c_i(\bar{x}) = 0$, et que $rc_i(\bar{x}) + \bar{\lambda}_i \leq 0$ implique que $\bar{\lambda}_i = 0$. □

Le lagrangien augmenté a également un intérêt pour les *problèmes convexes*, car il a un effet régularisant sur la fonction duale. Les *fonctions duales* associées au lagrangien classique (13.20) et au lagrangien augmenté (13.23) sont définies en $\lambda \in \mathbb{R}^m$ respectivement par

$$\delta_0(\lambda) = -\inf_{x\in X} \ell_0(x, \lambda) \qquad \text{et} \qquad \delta_r(\lambda) = -\inf_{x\in X} \ell_r(x, \lambda).$$

Le résultat suivant montre que δ_r (qui est convexe) est la régularisée de Moreau-Yosida de δ_0 sur $\mathbb{R}^m$, pour la norme $\frac{1}{r^{1/2}}\|\cdot\|_2$ (cette notion de régularisation a été introduite à la section 3.6.2).

Proposition 13.18 *On suppose que le problème* (P_{EI}) *est convexe au sens de la définition* 4.17, *que* f *et* c *sont à valeurs réelles et que* $r > 0$. *Alors* δ_r *vérifie les propriétés suivantes :*

1) δ_r *est convexe fermée,*
2) δ_r *est la régularisée de Moreau-Yosida de* δ_0, *c'est-à-dire*

$$\delta_r(\lambda) = \inf_{\mu \in \mathbb{R}^m} \Big(\delta_0(\mu) + \frac{1}{2r} \|\mu - \lambda\|_2^2 \Big), \tag{13.26}$$

3) *si* $\delta_0 \not\equiv +\infty$, *alors* δ_r *prend des valeurs finies* (*elle est donc dans* $\overline{\mathrm{Conv}}(\mathbb{R}^m)$) *et est* $C^{1,1}$ *sur* $\mathbb{R}^m$; *si, de plus,* $\delta_r(\lambda) = -\ell_r(x, \lambda)$ *pour un certain* $x \in \mathbb{R}^n$, *le gradient de* δ_r *est donné par*

$$\nabla \delta_r(\lambda) = -\nabla_\lambda \ell_r(x, \lambda) = \begin{pmatrix} -c_E(x) \\ \min\left(\frac{1}{r}\lambda_I, -c_I(x)\right) \end{pmatrix}. \tag{13.27}$$

Démonstration. 1) Comme enveloppe supérieure (en x et $s \geq 0$) des fonctions affines $\lambda \mapsto -\ell_r(x, s, \lambda)$ (le critère dans (12.22)), δ_r est convexe et fermée (proposition 3.7).

2) On sait que $\delta_r(\lambda) = v_r^*(\lambda)$. La fonction valeur v_r étant reliée à v_0 par la formule (13.22) :

$$v_r(\lambda) = v_0(\lambda) + r\chi(\lambda),$$

où χ est la fonction auto-conjuguée $p \mapsto \frac{1}{2}\|p\|_2^2$. On aura donc un lien entre $\delta_r(\lambda)$ et $\delta_0(\lambda) = v_0^*(\lambda)$ en calculant v_r^* comme conjuguée d'une somme de fonctions (proposition 3.36).

La fonction F_0 définie par (13.19) est convexe et donc aussi v_0 qui en est la fonction marginale (proposition 3.8). D'autre part $v_0 \not\equiv +\infty$ (quel que soit $x \in \mathbb{R}^n$, $v_0(-c(x)) \leq f(x) < +\infty$). Si v_0 prend la valeur $-\infty$, il en est de même de v_r (pour la même valeur de l'argument), si bien que $\delta_0 \equiv +\infty$ et $\delta_r \equiv +\infty$: la formule (13.26) est vérifiée. Autrement $v_0 \in \mathrm{Conv}(\mathbb{R}^m)$ et en appliquant la proposition 3.36, on trouve également la formule (13.26) :

$$\begin{aligned} \delta_r(\lambda) &= v_r^*(\lambda) \\ &= \Big(v_0^* \uplus (r\chi)^* \Big)(\lambda) \\ &= \inf \left\{ \delta_0(\mu) + \frac{1}{2r}\|\nu\|_2^2 : \mu + \nu = \lambda \right\}. \end{aligned}$$

3) Par hypothèse, $\delta_0 \not\equiv +\infty$. D'autre part δ_0 ne prend pas la valeur $-\infty$ (du fait de la minimisation de ℓ_0 en x). Donc δ_0 est propre. Comme supremum de fonctions affines, δ_0 est alors dans $\overline{\mathrm{Conv}}(\mathbb{R}^m)$. Pour tout λ, le problème de minimisation dans (13.26) a alors une solution unique dans $\operatorname{dom} \delta_0$, si bien que δ_r est à valeurs réelles. Comme régularisée de Moreau-Yosida, δ_r est dans $\overline{\mathrm{Conv}}(\mathbb{R}^m)$ et est $C^{1,1}$ (proposition 3.55).

Enfin, si $\delta_r(\lambda) = -\ell_r(x, \lambda)$ pour un certain $x \in \mathbb{R}^n$, on a quel que soit $\lambda' \in \mathbb{R}^m$:

$$\begin{aligned}\delta_r(\lambda') &\leq -\ell_r(x,\lambda') \qquad [\text{définition de } \delta_r]\\ &\leq -\ell_r(x,\lambda) - \nabla_\lambda \ell_r(x,\lambda)^\top(\lambda'-\lambda) \qquad [\text{concavité de } \ell_r(x,\cdot)]\\ &= \delta_r(\lambda) - \nabla_\lambda \ell_r(x,\lambda)^\top(\lambda'-\lambda).\end{aligned}$$

On en déduit que $\nabla\delta_r(\lambda) = -\nabla_\lambda \ell_r(x,\lambda)$. □

13.5 Méthodes numériques ▲

On s'intéresse dans cette section à des algorithmes de résolution du problème dual

$$\sup_{y\in Y} \inf_{x\in X} \varphi(x,y),$$

où $\varphi : X \times Y \to \overline{\mathbb{R}}$. Deux méthodes sont décrites. La première minimise directement la fonction duale, si bien qu'il faut résoudre complètement les problèmes de Lagrange pour chaque valeur prise par la variable duale. Il faut donc que le problème soit bien adapté à cette approche : la fonction φ doit être facile à minimiser en $x \in X$. La seconde est une version simplifiée de la première, dans laquelle le problème de Lagrange est résolu de manière très grossière, puisque l'on ne fait qu'un seul pas de minimisation.

Nous présentons ces algorithmes dans le cadre de la *relaxation lagrangienne*, c'est-à-dire lorsque la fonction φ ci-dessus est le lagrangien d'un problème d'optimisation avec contraintes.

13.5.1 Minimisation de la fonction duale

Nous allons introduire un algorithme minimisant la fonction duale lorsqu'il s'agit de résoudre un problème d'optimisation sous la forme suivante :

$$\begin{cases}\min\ f(x)\\ c(x) \leq 0\\ x \in X,\end{cases} \tag{13.28}$$

où $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^m$ et X est une partie de $\mathbb{R}^n$ permettant d'exprimer les contraintes aisées à prendre en compte (typiquement des contraintes de borne). On note $\ell(x,\lambda) = f(x) + \lambda^\top c(x)$ le lagrangien qui "dualise" la contrainte d'inégalité. Le problème dual s'écrit

$$\inf_{\lambda\geq 0}\ \delta(\lambda), \tag{13.29}$$

où $\delta : \mathbb{R}^m_+ \to \mathbb{R}$ est la *fonction duale*, qui est définie par

$$\delta(\lambda) := -\inf_{x\in X} \ell(x,\lambda). \tag{13.30}$$

On rappelle que le problème de minimisation dans (13.30) s'appelle le *problème de Lagrange*.

Même en l'absence de convexité de f et c, la fonction duale est concave. Ceci provient de l'inf dans (13.30) et de la structure du lagrangien, qui est linéaire par rapport au multiplicateur λ. La fonction duale n'est pas différentiable en général; mais lorsque le problème de Lagrange a (au moins) une solution, on dispose d'un sous gradient de δ, dont on peut faire usage dans des algorithmes adaptés à la minimisation de fonctions non différentiables (par exemple la *méthode de faisceaux*, décrite dans [209; 1993; chapitres XIV et XV]).

Proposition 13.19 *Si $\delta \not\equiv +\infty$, alors $\delta \in \mathrm{C}\overline{\mathrm{onv}}(\mathbb{R}^n)$. De plus, si $\bar{x}_\lambda$ est une solution du problème de Lagrange* (13.30)*, alors $-c(\bar{x}_\lambda) \in \partial\delta(\lambda)$.*

Démonstration. Supposons que $\delta \not\equiv +\infty$. Alors δ est propre car elle ne prend pas la valeur $+\infty$. D'autre part, comme enveloppe supérieure (de la famille paramétrée par $x \in X$) de fonctions linéaires (donc convexes et fermées) $\lambda \mapsto -\ell(x,\lambda)$, $-\delta$ est convexe et fermée.

Soit à présent $\bar{x}_\lambda$ une solution du problème de Lagrange (13.30). Pour tout $\mu \in \mathbb{R}^m$, on a

$$\delta(\mu) \geq -\ell(\bar{x}_\lambda, \mu) = \delta(\lambda) - (\mu - \lambda)^\top c(\bar{x}_\lambda).$$

Donc $-c(\bar{x}_\lambda) \in \partial\delta(\lambda)$. □

Voici à présent des conditions assurant la différentiabilité de la fonction duale. Dans ce cas et lorsqu'une solution $\bar{x}_\lambda$ du problème de Lagrange existe, d'après le résultat précédent, on doit avoir $\nabla\delta(\lambda) = -c(\bar{x}_\lambda)$.

Proposition 13.20 *Supposons que le problème de Lagrange* (13.30) *ait une solution unique, notée $\bar{x}_\lambda$, pour tout λ voisin d'un certain λ_0 et que l'application $\lambda \mapsto c(\bar{x}_\lambda)$ soit continue en λ_0; alors δ est Fréchet-différentiable en λ_0 et $\nabla\delta(\lambda_0) = -c(\bar{x}_{\lambda_0})$. Si de plus, $\lambda \mapsto c(\bar{x}_\lambda)$ est continue dans un voisinage de λ_0, alors δ est C^1 dans un voisinage de λ_0.*

Démonstration. Comme dans la démonstration de la proposition 13.19, pour λ et $\lambda' \in \operatorname{dom}\delta$ et pour une solution $\bar{x}_\lambda$ du problème de Lagrange en λ, on a $-\delta(\lambda')+\delta(\lambda) \leq (\lambda' - \lambda)^\top c(\bar{x}_\lambda)$. En inversant le rôle de λ et λ', on obtient

$$(\lambda' - \lambda)^\top c(\bar{x}_{\lambda'}) \leq -\delta(\lambda') + \delta(\lambda) \leq (\lambda' - \lambda)^\top c(\bar{x}_\lambda).$$

Fixons à présent $\lambda = \lambda_0$ et prenons $\lambda' = \lambda_0 + \mu$ avec μ petit, de telle sorte que $\bar{x}_{\lambda_0+\mu}$ existe et que $c(\bar{x}_{\lambda_0+\mu})$ dépende continûment de μ voisin de zéro. Alors

$$\mu^\top[c(\bar{x}_{\lambda_0+\mu}) - c(\bar{x}_{\lambda_0})] \leq -\delta(\lambda_0 + \mu) + \delta(\lambda_0) - \mu^\top c(\bar{x}_{\lambda_0}) \leq 0.$$

Dès lors

$$-\delta(\lambda_0 + \mu) + \delta(\lambda_0) - c(\bar{x}_{\lambda_0})^\top \mu = o(\|\mu\|), \quad \text{lorsque } \mu \to 0.$$

Ceci montre que δ est différentiable en λ_0 et que $\nabla\delta(\lambda_0) = -c(\bar{x}(\lambda_0))$. □

Il n'est pas surprenant que le gradient $\nabla\delta(\lambda)$ soit donné par $-c(\bar{x}_\lambda)$, si on prend le point de vue suivant. Supposons en effet des conditions un peu plus régulières que celles de la proposition ci-dessus : la solution $\bar{x}_\lambda \equiv \bar{x}(\lambda)$ du problème de Lagrange dans (13.30) est une fonction différentiable de λ et ℓ est différentiable. Alors $\delta(\lambda) = -\ell(\bar{x}(\lambda), \lambda)$ et on a

$$\delta'(\lambda) \cdot \mu = -\ell'_x(\bar{x}(\lambda), \lambda) \cdot (\bar{x}'(\lambda) \cdot \mu) - \ell'_\lambda(\bar{x}(\lambda), \lambda) \cdot \mu = -\ell'_\lambda(\bar{x}(\lambda), \lambda) \cdot \mu,$$

puisque $\ell'_x(\bar{x}(\lambda), \lambda) = 0$ par optimalité de $\bar{x}_\lambda$. On retrouve donc $\nabla\delta(\lambda) = -c(\bar{x}_\lambda)$.

On peut à présent énoncer l'algorithme, qui peut être vu comme une méthode du type gradient avec projection (section 11.1) appliquée au problème dual. Le gradient avec projection permet de prendre en compte la contrainte de positivité sur le multiplicateur dans (13.29).

Algorithme 13.21 (minimisation de la fonction duale)

1. Initialisation : choix de $\lambda_1 \geq 0$;
2. Pour $k = 1, 2, \ldots$ faire :
   2.1. Trouver x_k solution du problème $\min_{x \in X} \ell(x, \lambda_k)$;
   2.2. Si x_k est satisfaisant, on s'arrête;
   2.3. $\lambda_{k+1} = (\lambda_k + \alpha_k c(x_k))^+$, où $\alpha_k > 0$ est un pas bien choisi.

On retrouve le déplacement du gradient avec projection à l'étape 2.3. On passe pour l'instant sous silence l'étape délicate du choix du pas $\alpha_k > 0$.

Nous allons montrer la convergence de cet algorithme dans le cas des problèmes (fortement) convexes, avec de petits pas α_k. En ce qui concerne la convergence, des contraintes d'égalité affines peuvent être prise en compte par c (on les remplace par deux contraintes d'inégalité opposées).

Proposition 13.22 *On considère le problème* (13.28) *dans lequel on suppose que* $X = \mathbb{R}^n$*, que* $f : \mathbb{R}^n \to \mathbb{R}$ *est différentiable et fortement convexe de module* $\kappa > 0$ *et que* $c : \mathbb{R}^n \to \mathbb{R}^m$ *est lipschitzienne de constante* $L > 0$ *et convexe. On suppose également que le lagrangien* ℓ *a un point-selle* $(\bar{x}, \bar{\lambda})$ *sur* $\mathbb{R}^n \times \mathbb{R}^m_+$*. On considère l'algorithme* 13.21 *dans lequel* α_k *pris dans un compact de* $]0, \frac{2\kappa}{L^2}[$*. Alors l'algorithme génère une suite* $\{x_k\}$ *qui converge vers* $\bar{x}$*.*

Démonstration. Comme f est fortement convexe et l'ensemble admissible du problème (13.28) est convexe, ce dernier a une solution unique, qui n'est autre que $\bar{x}$. En effet $\bar{x}$ est solution du problème primal associé au lagrangien ℓ – proposition 13.4 – qui est le problème (13.28).

L'optimalité de $\bar{x}$ donne

$$\nabla f(\bar{x})^\top (x - \bar{x}) + \bar{\lambda}^\top \Big(c'(\bar{x}) \cdot (x - \bar{x}) \Big) = 0, \quad \forall x \in \mathbb{R}^n.$$

En utilisant la convexité de c et $\bar{\lambda} \geq 0$, on obtient

$$\nabla f(\bar{x})^\top (x - \bar{x}) + \bar{\lambda}^\top \Big(c(x) - c(\bar{x}) \Big) \geq 0, \quad \forall x \in \mathbb{R}^n.$$

De même pour l'optimalité en x_k :

$$\nabla f(x_k)^\top(x - x_k) + \lambda_k^\top\Big(c(x) - c(x_k)\Big) \geq 0, \quad \forall x \in \mathbb{R}^n.$$

En prenant $x = x_k$ dans l'avant-dernière inégalité, $x = \bar{x}$ dans la dernière et en sommant, on obtient

$$(\nabla f(x_k) - \nabla f(\bar{x}))^\top(x_k - \bar{x}) \leq -(\lambda_k - \bar{\lambda})^\top\Big(c(x_k) - c(\bar{x})\Big).$$

On utilise alors la forte convexité de f et on note $\mu_k := \lambda_k - \bar{\lambda}$, ce qui conduit à

$$\kappa\|x_k - \bar{x}\|^2 \leq -\mu_k^\top\Big(c(x_k) - c(\bar{x})\Big). \tag{13.31}$$

Il reste à utiliser la récurrence sur λ_k. On a $\nabla_\lambda \ell(\bar{x}, \lambda) = c(\bar{x})$ et on se rappelle que $\ell(\bar{x}, \cdot)$ a un maximum en $\bar{\lambda}$ sur $\mathbb{R}^m_+$. Dès lors, pour tout $\lambda \in \mathbb{R}^m_+$:

$$0 \leq -\alpha_k c(\bar{x})^\top(\lambda - \bar{\lambda}) = \Big(\bar{\lambda} - (\bar{\lambda} + \alpha_k c(\bar{x}))\Big)^\top(\lambda - \bar{\lambda}),$$

ce qui s'écrit aussi

$$\bar{\lambda} = P_{\mathbb{R}^m_+}(\bar{\lambda} + \alpha_k c(\bar{x})).$$

D'autre part, la récurrence sur λ_k donne

$$\lambda_{k+1} = P_{\mathbb{R}^m_+}(\lambda_k + \alpha_k c(x_k)).$$

On utilise alors le fait que le projecteur $P_{\mathbb{R}^m_+}$ est 1-lipschitzien, la propriété de Lipschitz de c et (13.31) :

$$\begin{aligned}
\|\mu_{k+1}\| &\leq \|\mu_k + \alpha_k(c(x_k) - c(\bar{x}))\| \\
\|\mu_{k+1}\|^2 &\leq \|\mu_k\|^2 + 2\alpha_k\mu_k^\top(c(x_k) - c(\bar{x})) + \alpha_k^2\|c(x_k) - c(\bar{x})\|^2 \\
&\leq \|\mu_k\|^2 + (\alpha_k L^2 - 2\kappa)\alpha_k\|x_k - \bar{x}\|^2 \\
&\leq \|\mu_k\|^2 - \epsilon^2 L^2\|x_k - \bar{x}\|^2,
\end{aligned}$$

si $\alpha_k \in [\epsilon, \frac{2\kappa}{L^2} - \epsilon]$, avec $\epsilon > 0$ petit. Dès lors $\{\|\mu_k\|\}$ est décroissante, donc converge. On en déduit que $x_k \to \bar{x}$. □

Le résultat précédent n'assure pas la convergence de la suite des multiplicateurs $\{\lambda_k\}$, alors que l'algorithme 13.21 s'intéresse principalement à cette suite ! En fait, sous les hypothèses de la proposition, il n'y a pas unicité du multiplicateur optimal $\bar{\lambda}$. Par contre, si l'on suppose que les gradients des contraintes actives sont linéairement indépendants, alors on peut montrer la convergence de λ_k vers l'unique multiplicateur optimal $\bar{\lambda}$ (voir l'exercice 13.5).

13.5.2 Minimisation de la fonction duale régularisée

On considère ici le problème avec contraintes d'égalité

$$\begin{cases} \min \ f(x) \\ c(x) = 0 \\ x \in X, \end{cases} \tag{13.32}$$

où $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^m$ et X est une partie de $\mathbb{R}^n$ permettant d'exprimer les contraintes aisées à prendre en compte (typiquement des contraintes de borne). On étudie ici le cas où l'approche duale se fait via le lagrangien augmenté associé, à savoir

$$\ell_r(x, \lambda) = f(x) + \lambda^\top c(x) + \frac{r}{2}\|c(x)\|^2.$$

La fonction duale δ_r est définie sur tout $\mathbb{R}^m$ par

$$\delta_r(\lambda) = -\inf_{x \in X} \ \ell_r(x, \lambda).$$

L'approche duale consiste à minimiser δ_r. Dans la version la plus simple, on prend

$$\lambda_+ = \lambda + rc(x_\lambda), \tag{13.33}$$

où x_λ est un minimiseur de $\ell_r(\cdot, \lambda)$. Cette approche est identique à la *méthode des multiplicateurs* introduite à la section 12.4.4. Notre objectif ici est d'en donner une interprétation en utilisant la dualité et l'opérateur proximal (section 3.6.1).

1. On sait que δ_r est différentiable. Son gradient s'écrit $\nabla\delta_r(\lambda) = -c(x_\lambda)$, si bien que (13.33) peut déjà être vu comme une méthode de gradient sur δ_r avec le pas $r > 0$.

2. La proposition 13.18 permet d'avoir un autre point de vue. La fonction δ_r étant la régularisée de Moreau-Yosida de δ_0, son gradient en λ pour le produit scalaire $\langle u, v\rangle = \frac{1}{r}(u^\top v)$ s'écrit aussi $\lambda_p - \lambda$, où λ_p est le point proximal de λ (proposition 3.55). Le gradient pour le produit scalaire euclidien s'écrit donc $\frac{1}{r}(\lambda_p - \lambda)$. D'après (13.27), ce dernier est aussi $c(x_\lambda)$, on voit que

   $$\lambda_+ = \text{le point proximal de } \lambda.$$

   Il s'agit donc de l'*algorithme proximal* sur δ_0. Ce point de vue implique que la méthode des multiplicateurs est un algorithme monotone sur δ_0 et de l'inégalité

   $$\delta_r(\lambda) = \delta_0(\lambda_p) + \frac{1}{2r}\|\lambda - \lambda_p\|^2 \le \delta_0(\lambda),$$

   on voit qu'à chaque itération on contrôle la croissance de δ_0 par

   $$\delta_0(\lambda_+) \le \delta_0(\lambda) - \frac{r}{2}\|c(x_\lambda)\|^2.$$

   Si δ_0 est bornée inférieurement, cette inégalité implique la convergence de $c(x_\lambda) \to 0$. Elle donne aussi une estimation de la vitesse de décroissance de $\|c(x_\lambda)\|$ vers zéro, qui est en $O(r^{-1/2})$, d'où l'intérêt de prendre r grand. Une grande valeur de r rend par ailleurs la minimisation de ℓ_r souvent plus difficile, ce qui nécessite de trouver un compromis.

13.5.3 L'algorithme d'Arrow-Hurwicz

Algorithme 13.23 (Arrow-Hurwicz)

1. Initialisation : choix de $\lambda_1 \in \Lambda$ et de $x_1 \in X$;
2. Pour $k = 1, 2, \ldots$ faire :
   2.1. $x_{k+1} = P_X\left(x_k - \alpha_k^1 \nabla_x \ell(x_k, \lambda_k)\right)$;
   2.2. $\lambda_{k+1} = P_\Lambda\left(\lambda_k + \alpha_k^2 \nabla_\lambda \ell(x_{k+1}, \lambda_k)\right)$;
   2.3. Si (x_{k+1}, λ_{k+1}) est satisfaisant on s'arrête.

Cet algorithme est très semblable à l'algorithme 13.21 mais on ne fait plus qu'un seul pas de minimisation en x. Dans le cas convexe, on montre que l'on a convergence de la méthode si α_k^1 et α_k^2 sont pris assez petits.

Notes

Sur la dualité en général, on pourra consulter l'ouvrage d'Hiriart-Urruty et Lemaréchal [209; chapitre XII] (problèmes convexes). La technique de dualisation par perturbation a été introduite par Rockafellar [319; 1969] et [320; 1970]. La théorie peut aussi se développer en dimension infinie: voir Laurent [243; 1972] et Ekeland et Temam [129; 1974]. La dualité de Fenchel a débuté avec le travail de Fenchel [132, 133] et a été ensuite étendue par Rockafellar [320, 325].

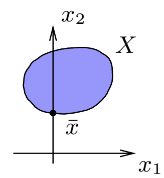
Nous l'avons déjà mentionné au chapitre 12 : le lagrangien augmenté (13.23), identique à (12.24), a été proposé par Rockafellar [321, 322; 1971-1973]. Nous avons suivi ici l'approche par perturbation de [322]. La proposition 13.17 est en partie due à Arrow, Gould et Howe [16; 1973] qui ont montré que, sous les conditions énoncées, $(\bar{x}, \bar{\lambda})$ est point-selle de $\ell_r(x, \lambda)$ sur $B(\bar{x}, \varepsilon) \times (\mathbb{R}^{m_E} \times \mathbb{R}_+^{m_I})$; la restriction $\lambda_I \geq 0$ n'est pas nécessaire comme l'avait déjà observé Rockafellar [322] dans le cas convexe. Le point de vue proximal de la proposition 13.18 est repris de [322].

Les algorithmes fondés sur la dualité n'ont été ici qu'esquissés. Pour leur utilisation dans la résolution de problèmes quadratiques, on pourra consulter la revue de Lin et Pang [250; 1987]. Pour des problèmes plus généraux, la non-différentiabilité de la fonction duale, lorsqu'elle est présente, complique évidemment beaucoup cette approche et il faut alors recourir à des algorithmes adaptés. La *méthode de faisceaux* en est un exemple; elle est étudiée en détail dans [209; 1993; chapitres XIV et XV]. L'utilisation de la dualité lagrangienne pour résoudre des problèmes variés est passée en revue par Lemaréchal [247]. Une autre possibilité est d'utiliser une fonction de couplage φ autre que le lagrangien, de manière à rendre la fonction duale différentiable. On l'a vu, le lagrangien augmenté a cette propriété (voir aussi [323]). D'autres fonctions partageant cette propriété et ayant divers avantages sont étudiées dans le livre de Gol'shteĭn et Tretyakov [175; section 2.5] et dans les articles d'Auslender, Ben-Tiba et Teboulle [23] et d'Auslender et Teboulle [21].

Exercices

13.1. *Dualisation d'un problème dans* $\mathbb{R}^2$ [40]. Soit X une partie quelconque de $\mathbb{R}^2$. On considère le problème :

$$(P) \quad \begin{cases} \inf\, x_2 \\ x \in X \\ x_1 = 0. \end{cases}$$



1) En dualisant la contrainte "$x_1 = 0$" par le lagrangien classique, montrez que le problème dual consiste à trouver la droite de $\mathbb{R}^2$ qui soit en-dessous de X et qui rencontre l'axe des ordonnées le plus haut possible.
2) Montrez que si X est un convexe fermé, si $\{x_2 \in \mathbb{R}\colon (0, x_2) \in X\}$ est borné inférieurement et si X a un point d'abscisse < 0 et un point d'abscisse > 0, alors le lagrangien a un point-selle.
3) En choisissant des ensembles X particuliers, montrez que l'on peut rencontrer les situations suivantes.
   (a) Il y a un saut de dualité et les solutions du problème interne $\inf_{x\in X} (\bar{\lambda} x_1 + x_2)$, où $\bar{\lambda}$ est une solution duale, ne sont pas solutions de (P).
   (b) Le lagrangien a un point-selle, mais certaines solutions du problème interne $\inf_{x\in X} (\bar{\lambda} x_1 + x_2)$, où $\bar{\lambda}$ est une solution duale, ne sont pas solutions de (P).
   (c) Tout point de $\mathbb{R}$ est solution duale.
   (d) Le problème primal a une (ou n'a pas de) solution, la valeur optimale primale est finie, le problème dual n'a pas de solution et il n'y a pas de saut de dualité.
4) Dualisez la contrainte "$x_1 = 0$" par le lagrangien augmenté et donnez une interprétation géométrique du problème dual. Donnez un exemple d'ensemble X pour lequel il n'y a pas de saut de dualité avec cette dualisation par le lagrangien augmenté, alors qu'il y en aurait un avec la dualisation par le lagrangien classique.

Remarque. Cet exercice décrit bien ce qui peut se passer lors de la dualisation de contraintes d'égalité. C'est alors l'épigraphe de la fonction valeur primale qui tient lieu d'ensemble X.

13.2. *Dualisation de problèmes.* Écrire un dual des problèmes suivants.

1) Le problème d'optimisation linéaire $\min\{c^\top x : x \in \mathbb{R}^n,\ Ax = b,\ x \geq 0\}$, où $c \in \mathbb{R}^n$, A est une matrice $m \times n$ et $b \in \mathbb{R}^m$.
2) Un dual du dual du problème linéaire défini au numéro 1.
3) Le problème généralisant le problème linéaire défini au numéro 1 : on se donne deux espaces euclidiens E et F, $c \in E$, un cône non vide K de E, $b \in F$, une application linéaire $A : E \to F$ et on considère le problème $\min\{\langle c, x\rangle : x \in \mathbb{R}^n,\ Ax = b,\ x \in K\}$. On dualisera soit la contrainte d'égalité, soit les deux contraintes (on suppose alors que K est aussi convexe et fermé), pour retrouver le même problème dual.
4) Le problème d'optimisation quadratique $\min\{c^\top x + \frac{1}{2} x^\top Q x : x \in \mathbb{R}^n,\ Ax = b,\ x \geq 0\}$, où $c \in \mathbb{R}^n$, Q est une matrice d'ordre n symétrique définie positive, A est une matrice $m \times n$ et $b \in \mathbb{R}^m$.

13.3. *Dualité en optimisation conique* [316]. Soit $\mathbb{R}^3_\triangledown := \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 \leq x_3^2,\ x_3 \geq 0\}$ le cône auto-dual déjà considérer à l'exercice 2.20. On considère le problème conique et son dual

$$(P) \quad \begin{cases} \min \ c^\top x \\ Ax = b \\ x \in \mathbb{R}^3_\triangledown \end{cases} \qquad \text{et} \quad (D) \quad \begin{cases} \min \ b^\top y \\ A^\top y + s = c \\ s \in \mathbb{R}^3_\triangledown \end{cases}$$

1) Montrez que $\inf \{x_2 + x_3 : x_1 = 1,\ x \in \mathbb{R}^3_\triangledown\}$ n'a pas de solution, que son dual en a une et qu'il n'y a pas de saut de dualité.

13.4. *Théorème d'Everett.* On considère le problème

$$(P_{EI}) \quad \begin{cases} \min \ f(x) \\ c_E(x) = 0 \\ c_I(x) \le 0, \end{cases}$$

où les fonctions $f : \mathbb{R}^n \to \mathbb{R}$, $c_E : \mathbb{R}^n \to \mathbb{R}^{m_E}$ et $c_I : \mathbb{R}^n \to \mathbb{R}^{m_I}$ n'ont pas de propriété particulière (de différentiabilité ou de convexité). On note $c(x) := (c_E(x), c_I(x)) \in \mathbb{R}^m$ ($m := m_E + m_I$), $(x, \lambda) \mapsto \ell(x, \lambda) = f(x) + \lambda^\top c(x)$ le lagrangien,

$$\lambda \mapsto \delta(\lambda) := \inf_{x \in \mathbb{R}^n} \ell(x, \lambda) \tag{13.34}$$

la fonction duale et $\sup\{\delta(\lambda) : \lambda \in \mathbb{R}^m,\ \lambda_I \ge 0\}$ le problème dual. Le problème à droite dans (13.34) est appelé problème de Lagrange. Démontrez les affirmations suivantes.

1) Si $\lambda_I \ge 0$ et si x_λ est solution du problème de Lagrange, alors x_λ est solution du problème (P_{EI}) perturbé suivant

$$\begin{cases} \min \ f(x) \\ c_E(x) = c_E(x_\lambda) \\ c_i(x) \le c_i(x_\lambda), \quad \forall i \in I \text{ tel que } \lambda_i > 0. \end{cases} \tag{13.35}$$

2) Si $\bar{x}$ est solution du problème de Lagrange avec un $\lambda = \bar{\lambda}$ vérifiant $\bar{\lambda}_I \ge 0$ et si $\bar{x}$ est admissible pour (P_{EI}) et vérifie $\bar{\lambda}_I^\top c_I(\bar{x}) = 0$, alors $\bar{x}$ est solution de (P_{EI}).

13.5. *Convergence des multiplicateurs dans l'algorithme* 13.21. On reprend le cadre défini par la proposition 13.22 en supposant en plus que les gradients des contraintes actives en $\bar{x}$ sont linéairement indépendants. Alors il y a un unique point-selle $(\bar{x}, \bar{\lambda})$ du lagrangien et la suite $\{\lambda_k\}$ générée par l'algorithme 13.21 converge vers $\bar{\lambda}$.

Extrait de "*Éléments d'Optimisation Différentiable : Théorie et Algorithmes*", J. Ch. Gilbert (`Jean-Charles.Gilbert@inria.fr`), à paraître. Version du 6 février 2007.