

# Chapitre 3

## Méthodes de plans sécants et de faisceaux

### 3.1 Rappels de dualité

On s'intéresse dans ce chapitre au problème suivant :

$$\inf_{x \in X \subset \mathbb{R}^n} f(x), \quad (3.1a)$$

$$\text{sous } c_j(x) = 0, \quad j \in \llbracket 1, m_E \rrbracket, \quad (3.1b)$$

$$c_j(x) \leq 0, \quad j \in \llbracket m_E + 1, m_E + m_I \rrbracket, \quad (3.1c)$$

où le critère  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est une fonction convexe, et où les fonctions  $c_j : \mathbb{R}^n \rightarrow \mathbb{R}$  représentent les contraintes du problème. On note  $m = m_E + m_I$ , et on dit qu'un point  $x \in X$  est *admissible* s'il vérifie les contraintes (3.1b) et (3.1c). Le problème (3.1) s'écrit aussi de manière plus compacte :

$$\begin{aligned} \inf_{x \in X \subset \mathbb{R}^n} f(x), \\ \text{sous } c_E(x) = 0, \\ c_I(x) \leq 0, \end{aligned}$$

avec  $c_E : \mathbb{R}^{m_E} \rightarrow \mathbb{R}$  et  $c_I : \mathbb{R}^{m_I} \rightarrow \mathbb{R}$ .

#### 3.1.1 Définition du problème dual

Soit  $\Lambda$  une partie convexe fermée de  $\mathbb{R}^m$ . On définit le Lagrangien  $L : X \times \Lambda \rightarrow \mathbb{R} \cup \{+\infty\}$ , de telle sorte que

$$\sup_{\lambda \in \Lambda} L(x, \lambda) = \begin{cases} f(x) & \text{si } x \text{ est admissible,} \\ +\infty & \text{sinon.} \end{cases}$$

Il est clair par cette définition que le problème (3.1) est équivalent au problème *primal* suivant :

$$\inf_{x \in X} \sup_{\lambda \in \Lambda} L(x, \lambda).$$

L'idée de la dualité est d'inverser dans ce problème les opérateurs inf et sup, et donc de considérer le problème *dual* :

$$\sup_{\lambda \in \Lambda} \inf_{x \in X} L(x, \lambda).$$

Ceci conduit donc à considérer la fonction duale :

$$\varphi : \lambda \mapsto \inf_{x \in X} L(x, \lambda), \quad (3.2)$$

le problème dual étant alors équivalent à :

$$\sup_{\lambda \in \Lambda} \varphi(\lambda). \quad (3.3)$$

Ce nouveau problème sera intéressant, d'une part si la fonction duale  $\varphi$  est facile à calculer, et d'autre part si le problème (3.3) est plus simple à résoudre que le problème primal (3.1).

### 3.1.2 Cas de la relaxation Lagrangienne

On choisit ici d'utiliser le Lagrangien classique, aussi appelé Lagrangien « simple » (plutôt que le Lagrangien « augmenté »). Le convexe fermé de  $\mathbb{R}^m$  sur lequel est défini le Lagrangien est :

$$\Lambda = \mathbb{R}^{m_E} \times \mathbb{R}_+^{m_I},$$

où  $\mathbb{R}_+^{m_I}$  désigne l'orthant positif de  $\mathbb{R}^{m_I}$ . Un élément  $\lambda \in \Lambda$  est donc de la forme  $\lambda = (\lambda_E, \lambda_I)$  avec  $\lambda_I \geq 0$ , et le Lagrangien a pour expression :

$$L(x, \lambda) = f(x) + \langle \lambda_E, c_E(x) \rangle + \langle \lambda_I, c_I(x) \rangle.$$

La « relaxation Lagrangienne » consiste à chercher des solutions du problème (3.3) afin de résoudre le problème (3.1).

### 3.1.3 Régularité de la fonctionnelle duale

Au vu de la définition (3.2), la fonction duale  $\varphi$  est construite comme un infimum de fonctions affines, et on en déduit alors que  $\varphi$  est une fonction concave.

On fait les hypothèses suivantes.

**Hypothèse 4.** La fonction  $\varphi$  vérifie les propriétés suivantes :

- $\forall \lambda \in \Lambda, \varphi(\lambda) > -\infty,$
- $\forall \lambda \in \Lambda, \exists \hat{x}_\lambda \in X$  tel que  $\varphi(\lambda) = L(\hat{x}_\lambda, \lambda),$

c'est-à-dire que le problème de minimisation dans (3.2) admet toujours au moins une solution.

Afin de résoudre le problème dual (3.3), on cherche à identifier des sous-gradients de la fonction  $\varphi$ . Les deux lemmes suivants montrent que les contraintes du problème permettent de trouver des sous-gradients de la fonction duale  $\varphi$ . On se reportera au §1.6.2 pour plus de détails mathématiques sur les propriétés de différentiabilité des fonctions marginales, ici utilisées dans un cadre à peine différent.

**Lemme 1.** Pour tout  $\lambda \in \Lambda$ , pour tout  $\hat{x}_\lambda$  solution du problème  $\inf_{x \in X} L(x, \lambda)$ , on a :

$$(c_E(\hat{x}_\lambda), c_I(\hat{x}_\lambda)) \in \partial\varphi(\lambda).$$

Avec un peu plus d'hypothèses, on peut même caractériser l'ensemble du sous-différentiel de la fonction duale.

**Lemme 2.** Si on suppose que l'ensemble  $X$  est compact, que les fonctions  $f$  et  $c_I$  sont fermées et que les fonctions  $c_E$  sont continues, alors on a :

$$\partial\varphi(\lambda) = \text{cl} \left( \text{conv} \left\{ (c_E(\hat{x}_\lambda), c_I(\hat{x}_\lambda)), \hat{x}_\lambda \text{ solution de } \min_{x \in X} L(x, \lambda) \right\} \right).$$

On décrit donc *tout* le sous-différentiel de la fonction duale à partir de l'image des solutions du problème (3.2). par les contraintes.

*Commentaires.*

1. Lorsque la solution  $\hat{x}_\lambda$  du problème (3.2) est unique, le sous-différentiel de la fonction duale  $\varphi$  se réduit à un singleton, et  $\varphi$  est donc *différentiable*. Ceci se produit par exemple lorsque l'ensemble  $X$  est convexe compact, la fonction  $f$  est strictement convexe, les fonctions  $c_E$  sont affines et les fonction  $c_I$  convexes, car le Lagrangien est alors *strictement* convexe en  $x$ .
2. L'hypothèse de compacité de  $X$  peut être remplacée par des conditions sur le critère et les contraintes, comme par exemple le critère  $f$  coercif et les contraintes  $c_E$  et  $c_I$  bornées inférieurement.
3. Utiliser le Lagrangien augmenté plutôt que le Lagrangien classique revient à considérer la régularisée de Moreau-Yosida de la fonction, cette régularisée étant différentiable. On se reportera au §2.2.1 pour les propriétés de la régularisée de Moreau-Yosida.

### 3.1.4 Liens entre les problèmes primal et dual

On rappelle la définition standard suivante.

**Définition 7.** On appelle *saut de dualité* l'écart entre les valeurs des problèmes primal et dual :

$$\inf_{x \in X} \sup_{\lambda \in \Lambda} L(x, \lambda) - \sup_{\lambda \in \Lambda} \inf_{x \in X} L(x, \lambda) .$$

Le *saut de dualité* est toujours positif ou nul : en effet, on a  $L(x, \lambda) \geq \inf_{x \in X} L(x, \lambda)$  et donc  $\sup_{\lambda \in \Lambda} L(x, \lambda) \geq \sup_{\lambda \in \Lambda} \inf_{x \in X} L(x, \lambda)$ . Comme le membre de droite ne dépend pas de  $x$ , prendre l'inf en  $x$  dans le membre de gauche ne change pas l'inégalité.

On dispose du théorème très général suivant, appelé théorème d'Everett, permettant d'interpréter une solution du problème dual comme une solution d'un problème primal « perturbé ».

**Théorème 21** (Everett). *Soit  $\lambda = (\lambda_E, \lambda_I) \in \Lambda$  fixé. On suppose que le problème  $\inf_{x \in X} L(x, \lambda)$  admet une solution  $\hat{x}_\lambda$ . Alors,  $\hat{x}_\lambda$  est aussi solution de :*

$$\inf_{x \in X \subset \mathbb{R}^n} f(x) , \tag{3.4a}$$

$$\text{sous } c_E(x) = c_E(\hat{x}_\lambda) , \tag{3.4b}$$

$$c_I(x) \leq c_I(\hat{x}_\lambda) . \tag{3.4c}$$

*Démonstration.* Soit  $x_\lambda^\sharp$  une solution du problème (3.4). Alors,

$$\begin{aligned} f(x_\lambda^\sharp) &= \inf_{x \in X} \sup_{\mu \in \Lambda} f(x) + \langle \mu_E, c_E(x) - c_E(\hat{x}_\lambda) \rangle + \langle \mu_I, c_I(x) - c_I(\hat{x}_\lambda) \rangle , \\ &\geq \sup_{\mu \in \Lambda} \inf_{x \in X} f(x) + \langle \mu_E, c_E(x) - c_E(\hat{x}_\lambda) \rangle + \langle \mu_I, c_I(x) - c_I(\hat{x}_\lambda) \rangle , \\ &\geq \inf_{x \in X} f(x) + \langle \lambda_E, c_E(x) - c_E(\hat{x}_\lambda) \rangle + \langle \lambda_I, c_I(x) - c_I(\hat{x}_\lambda) \rangle , \\ &= f(\hat{x}_\lambda) , \end{aligned}$$

car  $\hat{x}_\lambda$  est une solution de  $\inf_{x \in X} L(x, \lambda)$ . Comme  $\hat{x}_\lambda$  vérifie les contraintes du problème (3.4), on en déduit que  $\hat{x}_\lambda$  est aussi solution de ce dernier problème.  $\square$

*Commentaires.*

1. Le point  $\hat{x}_\lambda$ , qui minimise le Lagrangien  $L$ , est solution de la « version perturbée » du problème initial (3.1), au sens où le second membre des contraintes du problème perturbée (3.4) vaut  $(c_E(\hat{x}_\lambda), c_I(\hat{x}_\lambda))$  plutôt que  $(0, 0)$ .
2. Le théorème d'Everett est très général et ne nécessite aucune hypothèse sur le problème d'optimisation. En conséquence, les solutions du problème  $\inf_{x \in X} L(x, \lambda)$ , quand elles existent, ne sont pas nécessairement uniques, et pas nécessairement admissibles pour le problème (3.1).

**Corollaire 6.** *Si on suppose dans le théorème d'Everett que  $\hat{x}_\lambda$  vérifie les contraintes  $c_E(\hat{x}_\lambda) = 0$  et  $c_I(\hat{x}_\lambda) = 0$ , et donc les contraintes du problème (3.1), alors  $\hat{x}_\lambda$  est solution de (3.1).*

*Remarque 22.* La condition  $c_I(\hat{x}_\lambda) = 0$  du corollaire peut être affaiblie en  $\langle \lambda_I, c_I(\hat{x}_\lambda) \rangle = 0$ .  $\diamond$

Les deux lemmes suivants font le lien entre problème primal et problème dual.

**Lemme 3.** *Pour tout point  $x$  admissible pour le problème (3.1) et pour tout  $\lambda \in \Lambda$ , on a :*

$$f(x) \geq \varphi(\lambda) .$$

*Démonstration.* Comme  $x$  est admissible, on a  $x \in X$ ,  $c_E(x) = 0$  et  $c_I(x) \leq 0$ , ce qui implique que  $\langle \lambda, c(x) \rangle \leq 0$ . On en déduit :

$$\begin{aligned} f(x) &\geq f(x) + \langle \lambda_E, c_E(x) \rangle + \langle \lambda_I, c_I(x) \rangle , \\ &= L(x, \lambda) , \\ &\geq \inf_{y \in X} L(y, \lambda) , \\ &= \varphi(\lambda) , \end{aligned}$$

d'où le résultat.  $\square$

**Corollaire 7.** *Soit un point  $x$  admissible pour le problème (3.1) et soit  $\lambda \in \Lambda$  tels que :*

$$f(x) = \varphi(\lambda) .$$

*Alors, la condition des écarts complémentaires  $\langle \lambda, c(x) \rangle = 0$  est satisfaite, le point  $x$  est solution du problème :  $\inf_{y \in X} L(y, \lambda)$ , et le multiplicateur  $\lambda$  est solution du problème :  $\sup_{\mu \in \Lambda} \varphi(\mu)$ .*

*Démonstration.* C'est une conséquence directe de la preuve du lemme précédent, dans laquelle toutes les inégalités deviennent des égalités, et du fait que  $f(x) \geq \varphi(\lambda)$  pour tout  $\lambda \in \Lambda$  implique  $f(x) \geq \sup_{\mu \in \Lambda} \varphi(\mu)$ .  $\square$

On en déduit les propriétés suivantes.

**Corollaire 8.**

1. *Si la valeur du problème dual (3.3) est  $+\infty$ , alors le problème primal (3.1) n'admet pas de solution.*
2. *Si la valeur du problème primal (3.1) est  $-\infty$ , alors le problème dual (3.3) n'admet pas de solution.*
3. *S'il existe un point  $x^\sharp$  admissible pour le problème (3.1) et un multiplicateur  $\lambda^\sharp \in \Lambda$  tels que  $f(x) = \varphi(\lambda)$ , alors  $x^\sharp$  est solution du problème primal et  $\lambda^\sharp$  est solution du problème dual.*

*Démonstration.* Ces résultats sont une conséquence directe des lemmes précédents.  $\square$

*Remarque 23.* Même si le saut de dualité est différent de zéro, les résultats précédents montrent que la résolution du problème dual fournit toujours une borne inférieure pour la valeur du problème primal. Ce résultat est souvent utilisé, notamment en optimisation combinatoire.  $\diamond$

**Exercice 4.** Soit  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$  et  $c \in \mathbb{R}^n$ . On considère le problème linéaire suivant :

$$\inf_{x \in \mathbb{R}^n} c^\top x \quad \text{sous} \quad Ax = b, \quad x \geq 0.$$

1. Montrer que le problème dual de ce problème s'écrit :

$$\sup_{(y,s) \in \mathbb{R}^m \times \mathbb{R}^n} b^\top y \quad \text{sous} \quad A^\top y + s = c, \quad s \geq 0.$$

2. Montrer que l'on obtient le même problème dual qu'à la question 1 en ne dualisant que les contraintes d'égalité du problème initial.

En conclusion, résoudre des problèmes d'optimisation sous contraintes conduit dans le cadre de la dualité à maximiser la fonction duale, qui n'est pas toujours différentiable. La dualité donne donc un cadre d'application naturel pour les algorithmes de l'optimisation sous-différentiable.

## 3.2 Méthode de plans sécants

On considère dans cette section une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , convexe propre. On rappelle qu'un élément  $r$  du sous-différentiel de la fonction  $f$  au point  $x$  vérifie par définition :

$$f(y) \geq f(x) + \langle r, y - x \rangle \quad \forall y \in \mathbb{R}^n,$$

et donc que  $y \mapsto f(x) + \langle r, y - x \rangle$  est une minorante affine exacte au point  $x$  de la fonction  $f$ . Une telle minorante affine sera appelée *coupe* de la fonction  $f$  au point  $x$  et est caractérisée par le couple  $(x, r)$ .

### 3.2.1 Approximation tangentielle

Soit  $\{x^i\}_{i \in \llbracket 0, p \rrbracket}$  un sous-ensemble discret de points de  $\mathbb{R}^n$ , tel que la fonction  $f$  soit sous-différentiable en chacun des points  $x^i$ . Pour tout  $i \in \llbracket 0, p \rrbracket$ , on suppose connus la valeur  $f(x^i)$  ainsi qu'un élément  $r^i$  dans le sous-différentiel  $\partial f(x^i)$ .

**Définition 8.** On appelle *approximation tangentielle* de  $f$  basée sur les points  $\{x^i\}_{i \in \llbracket 0, p \rrbracket}$  la fonction  $\widehat{f}_p$  définie par :

$$\widehat{f}_p(x) = \max_{i \in \llbracket 0, p \rrbracket} \left\{ f(x^i) + \langle r^i, x - x^i \rangle \right\}.$$

On note  $\mathcal{C}_p = \{(x^0, r^0), \dots, (x^p, r^p)\}$  l'ensemble des coupes constituant cette approximation.

Une approximation tangentielle  $\widehat{f}_p$  de  $f$  est donc l'enveloppe supérieure de  $p$  coupes de la fonction  $f$ . Elle vérifie les propriétés suivantes.

**Proposition 13.**

1. Soit  $p \in \mathbb{N}$ . La fonction  $\widehat{f}_p$  est convexe, vérifie pour tout  $x \in \mathbb{R}^n$  :

$$\widehat{f}_p(x) \leq f(x) \quad \text{et donc} \quad \inf_{x \in \mathbb{R}^n} \widehat{f}_p(x) \leq \inf_{x \in \mathbb{R}^n} f(x),$$

et on a :

$$\widehat{f}_p(x^i) = f(x^i) \quad \forall i \in \llbracket 0, p \rrbracket.$$

2. Soient  $p_1$  et  $p_2$  tels que  $0 \leq p_1 \leq p_2$ , avec  $\mathcal{C}_{p_1} \subset \mathcal{C}_{p_2}$ . Alors, pour tout  $x \in \mathbb{R}^n$ , on a :

$$\widehat{f}_{p_1}(x) \leq \widehat{f}_{p_2}(x) \quad \text{et donc} \quad \inf_{x \in \mathbb{R}^n} \widehat{f}_{p_1}(x) \leq \inf_{x \in \mathbb{R}^n} \widehat{f}_{p_2}(x).$$

*Démonstration.* La preuve est une conséquence immédiate des définitions de l'approximation tangentielle et du sous-différentiel.  $\square$

*Commentaires.*

1. L'approximation tangentielle permet d'approcher « par en-dessous » une fonction  $f$  convexe à l'aide d'une fonction linéaire par morceaux.
2. Plus on augmente le nombre  $p$  de coupes constituant l'approximation tangentielle, et plus la fonction  $\widehat{f}_p$  s'approche de la fonction  $f$ .

La figure suivante illustre la notion d'approximation tangentielle pour une fonction  $V$ .

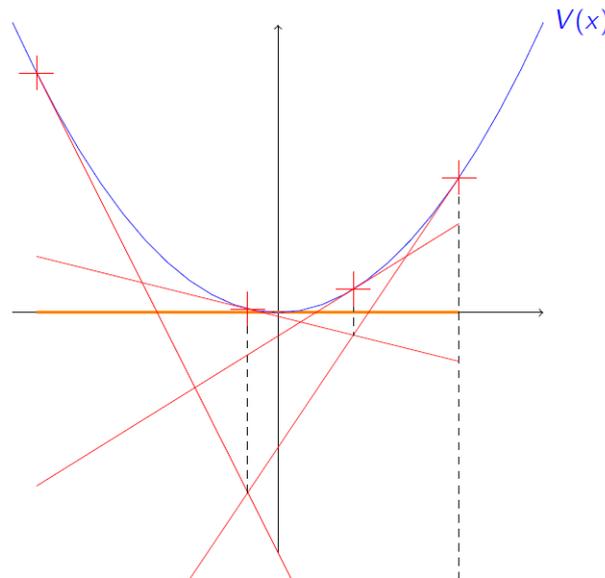


FIGURE 3.1 – Exemple d'approximation tangentielle

En pratique, on calcule l'infimum d'une approximation tangentielle de la manière suivante.

**Proposition 14.** L'infimum de la fonction  $\widehat{f}_p$  sur un ensemble  $X \subset \mathbb{R}^n$  s'obtient en calculant la valeur optimale  $\eta^\sharp$  du problème d'optimisation suivant :

$$\eta^\sharp = \inf_{(x, \eta) \in X \times \mathbb{R}} \eta, \\ \text{sous } \eta \geq f(x^i) + \langle r^i, x - x^i \rangle \quad \forall i \in \llbracket 0, p \rrbracket.$$

*Démonstration.* On a :

$$\begin{aligned}\eta^\sharp &= \inf_{x \in X} \left( \inf_{\eta \in \mathbb{R}} \eta \text{ sous } \eta \geq f(x^i) + \langle r^i, x - x^i \rangle, \forall i \in \llbracket 0, p \rrbracket \right) \\ &= \inf_{x \in X} \left( \max_{i \in \llbracket 0, p \rrbracket} \{f(x^i) + \langle r^i, x - x^i \rangle\} \right) \\ &= \inf_{x \in X} \widehat{f}_p(x),\end{aligned}$$

ce qui achève la preuve.  $\square$

Lorsque l'ensemble  $X$  est un polyèdre, le problème d'optimisation ci-dessus est un programme linéaire, qui peut donc être résolu par les algorithmes très efficaces de la programmation linéaire.

### 3.2.2 Algorithme de plans sécants

On s'intéresse au problème de minimiser une fonction  $f$  sur un sous-ensemble  $X \subset \mathbb{R}^n$  :

$$\min_{x \in X} f(x).$$

Le principe de la méthode des plans sécants est de construire de manière itérative une suite d'approximations tangentielles de plus en plus riche de la fonction  $f$  que l'on veut minimiser.

L'algorithme qui en découle est appelé *algorithme des plans sécants*, ou encore algorithme de Kelley [Kelley (1960)]. À chaque itération de l'algorithme, on dispose d'un point en lequel sont calculés la valeur de  $f$  et un élément du sous-différentiel de  $f$ , ce qui fournit une nouvelle coupe et donc permet d'enrichir l'approximation tangentielle existante. Sa minimisation donnera le nouveau point pour l'itération suivante.

**Algorithme 4.** Soit  $x^0 \in X$  et  $\epsilon > 0$ . On pose  $k = 0$

1. Calculer  $f(x^k)$  et  $r^k \in \partial f(x^k)$ .
2. Effectuer la minimisation de la fonction  $\widehat{f}_k$ , ce qui, par la proposition 14, s'écrit :

$$\begin{aligned}\min_{(x, \eta) \in X \times \mathbb{R}} \eta, \\ \text{sous } \eta \geq f(x^i) + \langle r^i, x - x^i \rangle \quad \forall i \in \llbracket 0, k \rrbracket,\end{aligned}$$

et dont la solution est notée  $(x^{k+1}, \eta^{k+1})$ . Par définition, on a :  $\widehat{f}_k(x^{k+1}) = \eta^{k+1}$ .

3. Tester la convergence.
  - Si  $f(x^{k+1}) - \widehat{f}_k(x^{k+1}) < \epsilon$ , STOP.
  - Sinon,  $k \leftarrow k + 1$  et retourner en 1.

*Remarque 24.* L'algorithme permet de construire une suite  $\{x^k\}_{k \in \mathbb{N}}$  d'éléments de  $X$ . Par la proposition 13, on sait que la suite  $\{\widehat{f}_k(x^{k+1})\}_{k \in \mathbb{N}}$  est croissante et bornée supérieurement. Par contre, la suite  $\{f(x^k)\}_{k \in \mathbb{N}}$  n'est pas nécessairement décroissante.  $\diamond$

### 3.2.3 Convergence

Le théorème suivant garantit la convergence asymptotique de la méthode des plans sécants.

**Théorème 22.** *On suppose que l'ensemble  $X$  est compact, que la fonction  $f$ , en plus d'être convexe propre, est Lipschitzienne et atteint son minimum sur  $X$  en un point  $x^\sharp$ . Alors,*

$$\forall \epsilon > 0, \exists k \in \mathbb{N}, \text{ tel que } f(x^k) \leq f(x^\sharp) + \epsilon.$$

*Démonstration.* Voir [Bonnans et al (2006), Theorem 9.6]. □

*Commentaires.*

1. Ce théorème ne donne aucune indication de *vitesse* de convergence. En fait, on peut construire des exemples où il faut plus de  $10^9$  itérations pour atteindre une précision de l'ordre de  $10^{-1}$  : la méthode des plans sécants peut converger de manière très lente ! On consultera [Nesterov (2004), Chapter 3] pour plus de détails sur l'algorithme des plans sécants.
2. Au cours des itérations, on accumule les coupes de la fonction  $f$  pour construire les approximations tangentielles. Si le nombre d'itérations devient grand, cette accumulation peut engendrer des problèmes de taille mémoire dans l'étape de minimisation de l'algorithme.

On va alors introduire une méthode permettant de conserver un nombre borné d'approximations tangentielles sans perdre les propriétés de convergence.

## 3.3 Méthode des faisceaux

### 3.3.1 Régularisé de Moreau-Yosida

On rappelle ici quelques résultats obtenus au §2.2 du chapitre 2. On suppose que la fonction  $f$  est convexe, s.c.i., propre et sous-différentiable. Soit  $c$  un coefficient strictement positif. La régularisée de Moreau-Yosida  $f_c$  est définie par :

$$f_c(x) = \min_{y \in \mathbb{R}^n} f(y) + \frac{1}{2c} \|y - x\|^2.$$

L'unique minimiseur de  $f_c$  au point  $x$  est le point proximal :

$$p_c(x) = \arg \min_{y \in \mathbb{R}^n} f(y) + \frac{1}{2c} \|y - x\|^2,$$

Le théorème suivant, dont les conclusions ont été démontrées au théorème 13 (§2.2), rappelle les propriétés de différentiabilité de la régularisée de Moreau-Yosida.

**Théorème 23.** *Soit  $f$  une fonction convexe, s.c.i., propre et sous-différentiable, et soit  $c > 0$ . Alors, la régularisée de Moreau-Yosida  $f_c$  de  $f$  est convexe, différentiable à gradient Lipschitzien, et le gradient de  $f_c$  est donné par la relation :*

$$\nabla f_c(x) = \frac{1}{c} (x - p_c(x)).$$

Ce théorème permet donc d'affirmer que, même si la fonction  $f$  n'est pas différentiable, sa régularisée de Moreau-Yosida l'est, et que son gradient est de plus Lipschitzien : une propriété de la plus haute importance pour l'analyse de convergence des algorithmes du premier ordre.

Le théorème suivant, dont les conclusions ont été démontrées au théorème 15 (§2.2), montre l'intérêt de la régularisée de Moreau-Yosida pour l'optimisation.

**Théorème 24.** Soit  $f$  une fonction convexe, s.c.i., propre et sous-différentiable, et soit  $c > 0$ . Alors, la régularisée de Moreau-Yosida  $f_c$  de  $f$  est telle que :

$$\inf_{x \in \mathbb{R}^n} f_c(x) = \inf_{x \in \mathbb{R}^n} f(x).$$

De plus, il y a équivalence entre les quatre propositions suivantes :

- (i)  $x^\sharp$  est un minimiseur de la fonction  $f$ ,
- (ii)  $x^\sharp$  est un minimiseur de la fonction  $f_c$ ,
- (iii)  $\nabla f_c(x^\sharp) = 0$ ,
- (iv)  $f(x^\sharp) = f_c(x^\sharp)$ .

**Exercice 5.** On pourra à titre d'exercice redémontrer les équivalences dans le théorème ci-dessus.

**Exemple 2.** La régularisée de Moreau-Yosida de la fonction « valeur absolue »  $f(x) = |x|$  est donnée par :

$$f_c(x) = \begin{cases} -x - \frac{c}{2} & \text{si } x < -c \\ \frac{1}{2c} \|x\|^2 & \text{si } x \in [-c, c] \\ x - \frac{c}{2} & \text{si } x > c \end{cases}.$$

que l'on a représentée sur la figure 3.2 (pour  $c = 0.5$ ) :

△

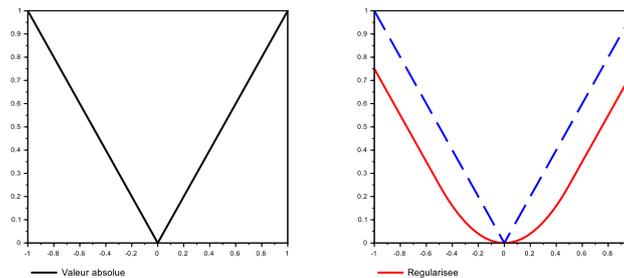


FIGURE 3.2 – Régularisation de la fonction valeur absolue

Ainsi, on a remplacé la résolution d'un problème convexe non différentiable par celle d'un problème différentiable ayant la même valeur et les mêmes solutions que le problème initial. L'idée serait donc d'appliquer à  $f_c$  les algorithmes standards de l'optimisation différentiable (gradient, gradient conjugué...). Cependant, le simple calcul de la fonction  $f_c$  est a priori aussi compliqué que la minimisation de la fonction  $f$ . L'utilisation directe de la régularisée de Moreau-Yosida n'est donc a priori pas directement opérationnelle.

**Une idée plus intéressante est de ne pas régulariser la fonction  $f$ , mais plutôt son approximation tangentielle.** Le calcul de la régularisée de Moreau-Yosida d'une approximation tangentielle correspond à résoudre un problème linéaire quadratique, et on dispose de méthodes performantes pour ce type de problème. Cette idée est la base de l'algorithme des faisceaux que l'on présente maintenant.

### 3.3.2 Algorithme des faisceaux

L'algorithme des faisceaux consiste à construire itérativement une suite d'approximations tangentielles de plus en plus précises, en calculant à chaque itération un nouveau point en utilisant la régularisée de l'approximation tangentielle courante.

Dans cet algorithme, on a deux types d'itérations.

- Itération *interne* (indice  $i$ ) : elle sert à améliorer la description de l'approximation tangentielle de  $f$  en utilisant sa régularisée, jusqu'à obtenir un point qui fait suffisamment décroître la valeur de la fonction.
  - Itération *externe* (indice  $k$ ) : elle change le point courant de l'algorithme afin de faire décroître la fonction, et construit l'approximation autour de ce nouveau point.
- Plus précisément, l'algorithme des faisceaux est le suivant.

**Algorithme 5.** Soit  $x^0 \in X$ ,  $y^0 \in X$ ,  $c > 0$ ,  $\epsilon > 0$  et  $\epsilon' > 0$ . On pose  $i = 0$  et  $k = 0$ .

1. Calculer  $f(y^i)$  et  $r^i \in \partial f(y^i)$  et résoudre

$$\min_{(y,\eta) \in X \times \mathbb{R}} \eta + \frac{1}{2c} \|x^k - y\|^2, \\ \text{sous } \eta \geq f(y^j) + \langle r^j, y - y^j \rangle \quad \forall j \in \llbracket 0, i \rrbracket,$$

dont la solution est notée  $(y^{i+1}, \eta^{i+1})$ . Calculer  $\widehat{f}_i(y^{i+1})$ .<sup>1</sup>

2. Tester la convergence.

- Si  $f(y^{i+1}) - \widehat{f}_i(y^{i+1}) \leq \frac{1}{2c} \|x^k - y^{i+1}\|^2$ ,
  - a) si  $\|x^k - y^{i+1}\| \leq \epsilon$ , *STOP*,
  - b) sinon, poser  $x^{k+1} = y^{i+1}$ , faire  $k \leftarrow k + 1$  et retourner à l'étape 1.<sup>2</sup>
- Sinon,
  - c) si  $f(y^{i+1}) - \widehat{f}_i(y^{i+1}) \leq \epsilon'$ , *STOP*,
  - d) sinon, faire  $i \leftarrow i + 1$  et retourner à l'étape 1.

*Commentaires.*

- L'étape 1 consiste à calculer la régularisée de Moreau-Yosida de l'approximation tangentielle  $\widehat{f}_i$  au point  $x^k$ , et donc à résoudre le problème

$$\min_{y \in X} \widehat{f}_i(y) + \frac{1}{2c} \|x^k - y\|^2.$$

- Le cas 2.a) correspond au cas où le min de l'approximation tangentielle est suffisamment proche du min de  $f$ .
- Dans le cas 2.b), on considère que la fonction  $f$  a suffisamment décréu et on effectue une itération externe (changement de point  $x^k$ ).
- Le cas 2.c) correspond au cas où le min de l'approximation tangentielle est suffisamment proche du min de  $f$ .
- Dans le cas 2.d), on considère que la fonction  $f$  n'a pas assez décréu : on affine son approximation tangentielle en lui ajoutant une nouvelle coupe (minorante affine en  $y^{i+1}$ ) et on effectue une itération interne.

### 3.3.3 Convergence

On donne ici un exemple de résultat de convergence obtenu par l'algorithme des faisceaux.

**Théorème 25.** Soit  $\epsilon > 0$  et  $\epsilon' > 0$ . Alors l'algorithme des faisceaux converge en un nombre fini d'itérations. De plus,

---

1. On rappelle que la fonction  $\widehat{f}_i$  est définie par :  $\widehat{f}_i(y) = \max_{j \in \llbracket 0, i \rrbracket} f(y^j) + \langle r^j, y - y^j \rangle$ .  
 2.  $y^{i+1}$  est alors appelé un « pas sérieux ».

— si l'algorithme s'arrête à l'étape 2.a) :  $\|x^k - y^{i+1}\| \leq \epsilon$ , alors

$$\|\nabla f_c(x^k)\| \leq \frac{2\epsilon}{c},$$

— si l'algorithme s'arrête à l'étape 2.c) :  $f(y^{i+1}) - \widehat{f}_{k+1}(y^{i+1}) \leq \epsilon'$ , alors

$$\|\nabla f_c(x^k)\| \leq 2\sqrt{\frac{2\epsilon'}{c}}.$$

*Démonstration.* Voir [Bonnans et al (2006), §10.3]. □

*Commentaires.*

1. L'algorithme consiste donc à construire un modèle qui approche la fonction, avec la recherche d'un centre dit « de stabilité » autour duquel on fait décroître la fonction tout en restant proche de ce centre. Si on ne décroît pas suffisamment, on conserve le même centre, mais on cherche un nouveau pas de descente  $y^{i+1}$ . Sinon, le pas  $y^{i+1}$  est considéré comme « sérieux » et on y déplace le centre de stabilité.
2. Le test d'arrêt de l'algorithme des faisceaux est de la même qualité que pour un algorithme de type gradient, alors que l'on traite un cas non différentiable.
3. Pour éviter les problèmes informatiques de taille mémoire et de temps de calcul lors de la minimisation de la régularisée de l'approximation tangentielle, liés à l'accumulation des coupes (contraintes) que l'on ajoute au modèle, on peut ajouter à l'algorithme une étape de sélection/compression des coupes.

La littérature sur les méthodes de faisceaux est fournie. On consultera par exemple l'ouvrage [Bonnans et al (2006)] pour approfondir le sujet. Pour les questions nouvelles et les perspectives sur ces méthodes, on pourra se reporter à l'article [Sagastizabal, 2018].

