

Programmation stochastique

Méthode *L*-shaped - Partie 4

Fabian Bastin

IFT-6512 – Hiver 2011

Extensions de la méthode L -shaped

- Une idée clé de la méthode L -shaped est de sous-estimer $Q(\mathbf{x})$ par une variable auxiliaire θ .
- Nous pouvons construire cette sous-estimation en utilisant l'inégalité du sous-gradient.
- $Q(\mathbf{x}) = \sum_{s \in S} p_s Q(\mathbf{x}, \xi_s)$.
- Version simple-coupe : la contrainte $\theta \geq Q(\mathbf{x})$ donnait

$$\theta \geq \sum_s p_s (\hat{\pi}_s)^T (h(\xi_s) - T(\xi_s)\mathbf{x}) = \alpha + \beta^T \mathbf{x},$$

ou

$$-\beta^T \mathbf{x} + \theta \geq \alpha.$$

- Autrement dit, tous les programmes de seconde étape sont optimisés pour obtenir les multiplicateurs simplex optimaux. Ces multiplicateurs sont agrégés pour générer une coupe.

- Pourquoi ne construire qu'une coupe ? **Idée** : approximer **chaque** $Q(x, \xi_s)$ par une variable auxiliaire θ_s , $s \in S$.
- Nous introduisons a présent les coupes

$$\theta_s \geq p_s(\hat{\pi}_s)^T (h(\xi_s) - T(\xi_s)x) = \alpha_s + \beta_s^T x, \quad s = 1 \in S.$$

Autrement dit, nous gardons $|S|$ coupes au lieu de les agréger en une seule. S'il n'est pas possible de construire ces $|S|$ coupes, la solution est optimale.

- Quelle méthode marchera le mieux ? La version multi-coupes offre une information plus détaillée, aussi en général, on s'attend à avoir moins d'itérations majeures, mais comme plus de coupes sont ajoutées, le programme de première étape est plus coûteux à résoudre. En règle pratique (non formelle !), l'approche multi-coupes est plus efficace quand le nombre de réalisations S n'est pas significativement plus grand que le nombre de contraintes de première étape.

Rappels : sous-gradients

Rappelons qu'en raison du **principe de dualité**,

$$Q(\hat{x}, \xi_s) = \max\{\pi^T (h(\xi_s) - T(\xi_s)x) \mid \pi^T W \leq q(\xi_s)^T\}.$$

Et donc, il existe un $\hat{\pi}$ t.q. $Q(\hat{x}, \xi_s) = \hat{\pi}^T h(\xi_s) - \hat{\pi}^T T(\xi_s)\hat{x}$.
Par rapport aux variables duales,

$$Q(x, \xi_s) \geq \hat{\pi}^T h(\xi_s) - \hat{\pi}^T T(\xi_s)x, \quad \forall x \in K_2(\xi_s)$$

et donc

$$Q(x, \xi_s) \geq Q(\hat{x}, \xi_s) - \hat{\pi}^T T(\xi_s)(x - \hat{x}), \quad \forall x \in K_2(\xi_s).$$

En d'autres termes, $-T(\xi_s)^T \hat{\pi} \in \partial Q(\hat{x}, \xi_s)$.

De même (passage à l'espérance), $-\sum_j p_j T(\xi_j)^T \hat{\pi}_s \in \mathcal{Q}(x)$.

- Partitionnons les scénarios en C "clusters" $\mathcal{S}_1, \mathcal{S}_1, \dots, \mathcal{S}_C$, et définissons

$$Q_{[\mathcal{S}_k]}(\mathbf{x}) = \sum_{s \in \mathcal{S}_k} p_s Q(\mathbf{x}, \xi_s).$$

Dès lors

$$Q(\mathbf{x}) = \sum_{k=1}^C Q_{[\mathcal{S}_k]}(\mathbf{x})$$
$$- \sum_{s \in \mathcal{S}_k} p_s (T(\xi_s))^T \hat{\pi}_s \in \partial Q_{[\mathcal{S}_k]}(\mathbf{x}).$$

- Nous pouvons donc faire la même chose : approximer $Q_{[\mathcal{S}_k]}(\mathbf{x})$ avec l'inégalité du sous-gradient.

Etape 0 Soit θ_k^0 une borne inférieure pour $Q_{[S_k]}(\mathbf{x})$. Posons

$$\mathcal{B}_0 = \{\mathbb{R}_n^+ \times \{\theta_1, \theta_2, \dots, \theta_C\} \mid \mathbf{A}\mathbf{x} = \mathbf{b}\},$$

$$\mathcal{B}_1 = \{\mathbb{R}_n^+ \times \{\theta_1, \theta_2, \dots, \theta_C\} \mid \theta_k \geq \theta_k^0\}.$$

Poser $\nu = 0$.

Etape 1 Incrémenter ν et résoudre le **problème maître**

$$\min \mathbf{c}^T \mathbf{x} + \sum_{k=1}^C \theta_k \text{ t.q. } (\mathbf{x}, \theta_1, \theta_2, \dots, \theta_C) \in \mathcal{B}_0 \cap \mathcal{B}_1.$$

Soit $(\mathbf{x}^\nu, \theta_1^\nu, \theta_2^\nu, \dots, \theta_C^\nu)$ la solution obtenue.

- Evaluer $Q(x^\nu) = \sum_{s \in S} p_s Q(x^\nu, \xi_s)$.
- Si $Q(x^\nu) = \infty$, il y a un certain ξ_k tel que $Q(x^\nu, \xi_k) = \infty$.
Nous ajoutons une **coupe de faisabilité** :

$$B_1 := B_1 \cap \{(x, \theta) \mid \sigma^T (h(\xi_k) - T(\xi_k)x) \leq \theta\}.$$

Cette inégalité n'a pas de terme en θ_s ; c'est la même inégalité que dans la méthode *L-shaped* classique.
Retour à l'étape 1.

- Si $Q(x^\nu) < \infty$, les S LPs correspondant aux S scénarios peuvent être résolus, avec des multiplicateurs π_s^ν et

$$u_k := - \sum_{s \in S_k} p_s (T(\xi_s))^T \hat{\pi}_s \in \partial Q_{[S_k]}(x).$$

Si $Q_{[S_k]}(x) \leq \theta_k$, $k = 1, \dots, C$, arrêt : x^ν est optimal.

Pour chaque k tel que $Q_{[S_k]}(\mathbf{x}) > \theta_k$, ajouter une coupe d'optimalité :

$$B_1 := B_1 \cap \{(\mathbf{x}, \theta_1, \theta_2, \dots, \theta_C) \mid \theta_k \geq Q_{[S_k]}(\mathbf{x}^\nu) + \mathbf{u}_k^T(\mathbf{x} - \mathbf{x}^\nu)\}.$$

Ok, très bien, . . . mais que se passe-t-il si on part avec la solution optimale \mathbf{x}^* ?

La méthode *L-shaped* souffre de problèmes de "stabilité",

- spécialement au cours des premières itérations quand nous n'avons pas un modèle "assez bon" de $Q(x)$;
- spécialement quand nous partons d'un bon estimé de la solution.

Il convient de **régulariser** la méthode. Plusieurs techniques existent.

Quelques références sur la stabilisation de la décomposition de Benders :

- Marsten, Hogan, Blankenship (1975) ;
- Lemaréchal (1975, ...) ;
- Kiwiel (1983, ...) ;
- Neame, Boland, and Ralph (1998).

Régularisation de la méthode L -Shaped

- **Pénalisation de la longueur de pas** : A. Ruszczyński, *A Regularized Decomposition for Minimizing a Sum of Polyhedral Functions*, *Mathematical Programming* 35, pp. 309–333, 1986.
- **Approche en $\|\cdot\|_\infty$** : J. T. Linderoth and S. J. Wright, *Implementing a Decomposition Algorithm for Stochastic Programming on a Computational Grid*, *Computational Optimization and Applications* 24, pp. 207–250, 2003.

Pénalisation de la longueur du pas. A l'itération k , on résout le programme

$$\min_x c^T x + \sum_{j \in C} \theta_j + \frac{1}{2\rho} \|x - x_k\|^2.$$

- Large ρ : comme le L -shaped.
- Petit ρ : on ne bouge pas trop.

On parle de **méthode de décomposition régularisée**.

Le principe est aussi de pénaliser la longueur du pas. A l'itération k , nous imposons $\|x_{x})k\|_k \leq \Delta_k$, où $\|\cdot\|_k$ désigne une norme usuelle (traditionnellement 2 ou ∞).

- Grand Δ_k : comme le L -shaped.
- Petit Δ_k : on ne bouge pas trop.

On parlera de **régularisation** de la méthode.

Dans les deux cas, on voit apparaître des concepts de **programmation nonlinéaire**.

Soit $m(x)$ le "modèle" utilisé comme problème de 1^{ère} étape.

- **L-shaped multi-coupes**

$$m(x) = \min \left\{ \mathbf{c}^T \mathbf{x} + \sum_{k=1}^C \theta_k \mid (\mathbf{x}, \theta_1, \theta_2, \dots, \theta_C) \in \mathcal{B}_0 \cap \mathcal{B}_1 \right\}.$$

- **Région de confiance**

$$m(x) = \min \left\{ \mathbf{c}^T \mathbf{x} + \sum_{k=1}^C \theta_k \mid (\mathbf{x}, \theta_1, \theta_2, \dots, \theta_C) \in \mathcal{B}_0 \cap \mathcal{B}_1, \right. \\ \left. \|\mathbf{x} - \mathbf{a}'\|_l \leq \Delta_l \right\}.$$

- **Décomposition régularisée**

$$m(x) = \min \left\{ \mathbf{c}^T \mathbf{x} + \sum_{k=1}^C \theta_k + (1/2\rho_l) \|\mathbf{x} - \mathbf{a}'\|_2^2 \right. \\ \left. \text{t.q. } (\mathbf{x}, \theta_1, \theta_2, \dots, \theta_C) \in \mathcal{B}_0 \cap \mathcal{B}_1, \right\}.$$

Méthode de décomposition régularisée

Référence : Birge et Louveaux, Section 6.1.

Etape 0 Posons $\rho = 1$, $r = \nu = 0$, $s(k) = 0$ pour tout $k = 1, \dots, S$. Sélectionnons une solution réalisable a^1 .

Etape 1 Posons $\nu := \nu + 1$, et résolvons le programme maître régularisé

$$\min c^T x + \sum_{k=1}^K \theta_k + \frac{1}{2\rho} \|x - a^\nu\|^2,$$

$$\text{t.q. } Ax = b,$$

$$D_\ell^T x \geq d_\ell, \ell = 1, \dots, r,$$

$$-\beta_{\ell(k)}^T x + \theta_k \geq \alpha_{\ell(k)}, \ell(k) = 1, \dots, s(k),$$

$$x \geq 0.$$

Méthode de décomposition régularisée (2)

- Etape 1** (suite) Soit (x^ν, θ^ν) une solution optimale de ce programme, où $(\theta^\nu)^T = (\theta_1^\nu, \dots, \theta_K^\nu)^T$.
Si $s(k) = 0$ pour un certain k , θ_k^ν est ignoré dans le calcul. Si $c^T x^\nu + e^T \theta^\nu = c^T a^\nu + Q(a^\nu)$, arrêt : a^ν est optimal.
- Etape 2** Si une coupe de faisabilité est générée, poser $a^{\nu+1} = a^\nu$ (pas non-réalisable nul), et retour à l'étape 1.
- Etape 3** Pour $k = 1, \dots, S$, résoudre le sous-problème linéaire

$$\min_y q_k^T y$$

$$\text{t.q. } Wy = h_k - T_k x^\nu,$$

$$y \geq 0.$$

Calculer $Q_k(x^\nu)$.

Etape 3 (suite) Si

$$\theta_k^\nu < \rho_k (\pi_k^\nu)^T (h_k - T_k x^\nu),$$

ajouter une coupe d'optimalité de la forme

$$-\beta_{\ell(k)}^T x + \theta_k \geq \alpha_{\ell(k)}.$$

Poser $s(k) = s(k) + 1$.

Etape 4 Si pour tout k ,

$$\theta_k^\nu \geq \rho_k (\pi_k^\nu)^T (h_k - T_k x^\nu),$$

poser $a^{\nu+1} := x^\nu$ (pas sérieux exact). Retour à l'étape 1.

Etape 5 Si

$$c^T x^\nu + Q(x^\nu) \leq c^T a^\nu + Q(a^\nu),$$

poser $a^{\nu+1} := x^\nu$ (pas sérieux approximatif).

Retour à l'étape 1.

Sinon, poser $a^{\nu+1} := a^\nu$. Retour à l'étape 1.

Extension immédiate : permettre à ρ de varier.

L'implémentation du L -shaped régularisé pose un certain de nombre de questions pratiques. Une implémentation efficace est proposée dans Ruszczyński (1986).

Birge et Louveaux, Section 5.3, Exercice 1.

Supposons $n_1 = 1$ (dimension de x), $m_1 = 0$ (aucune contrainte de première étape), $m_2 = 3$ (trois contraintes de seconde étape), $n_2 = 6$ (dimension de y), et

$$W = \begin{pmatrix} 1 & -1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Supposons de plus $S = 2$, et que les 2 réalisations de ξ sont équiprobables. Ces réalisations sont

$$\begin{aligned}\xi^1 &= (q^1, h^1, T^1)^T, \\ \xi^2 &= (q^2, h^2, T^2)^T.\end{aligned}$$

Exemple (suite)

Supposons

$$q^1 = (1, 0, 0, 0, 0, 0)^T, \quad q^2 = (3/2, 0, 2/7, 1, 0, 0)^T,$$

$$h^1 = (-1, 2, 7)^T, \quad h^2 = (0, 2, 7)^T,$$

$$T^1 = T^2 = (1, 0, 0)^T.$$

Pour la première valeur ξ^1 , nous avons donc le programme de seconde étape

$$\min_y y_1$$

$$\text{t.q. } y_1 - y_2 - y_3 - y_4 = -1 - x,$$

$$y_2 + y_5 = 2,$$

$$y_3 + y_6 = 7,$$

$$y \geq 0.$$

Exemple (suite)

Il est facile de voir que pour ξ^1 ,

$$Q_1(x) = \begin{cases} -x - 1 & \text{si } x \leq -1, \\ 0 & \text{si } x \geq -1. \end{cases}$$

Pour la seconde valeur de $Q(x, \xi)$, on peut montrer que

$$Q_2(x) = \begin{cases} -1.5x & \text{si } x \leq 0, \\ 0 & \text{si } 0 \leq x \leq 2, \\ \frac{2}{7}(x - 2) & \text{si } 2 \leq x \leq 9, \\ x - 7 & \text{si } x \geq 9. \end{cases}$$

Supposons également que $-20 \leq x \leq 20$ et $c = 0$.

Quelles propriétés (attendues) retrouve-t-on pour $Q_1(x)$ et $Q_2(x)$?

Que peut-on dire du recours ?

Exemple : L-shaped

Partons de $x = -2$. Nous obtenons la séquence d'itérés suivantes

Iteration 1 $x^1 = -2$, θ^1 est omis, nous ajoutons la coupe (type ?)

$$\theta \geq -0.5 - 1.25x.$$

Iteration 2 $x^2 = 20$, $\theta^2 = -25.5$, nouvelle coupe :

$$\theta \geq 0.5x - 3.5.$$

Iteration 3 $x^3 = 12/7$, $\theta^3 = -37/14$, nouvelle coupe :

$$\theta \geq 0.$$

Iteration 4 $x^4 \in [-2/5, 7]$, θ^4 . Si x^4 est choisi comme n'importe quelle valeur dans $[0, 2]$, arrêt.

Exemple : L-shaped multi-coupes

L'approche multi-coupes donnerait la séquence suivante.

Iteration 1 $x^1 = -2$, θ^1 et θ^2 sont omis ; nous ajoutons les coupes $\theta_1 \geq -0.5 - 0.5x$, $\theta_2 \geq -3/4x$.

Iteration 2 $x^2 = 20$, $\theta_1^2 = -10.5$, $\theta_2^2 = -15$; nouvelles coupes : $\theta_1 \geq 0$, $\theta_2 \geq 0.5x - 3.5$.

Iteration 3 $x^3 = 2.8$, $\theta_1^3 = 0$, $\theta_2^3 = -2.1$; nouvelle coupe : $\theta_2^3 \geq 1/7(x - 2)$.

Iteration 4 $x^4 = 0.32$, $\theta_1^4 = 0$, $\theta_2^4 = -0.24$; nouvelle coupe : $\theta_2 \geq 0$.

Iteration 5 $x^5 = 0$, $\theta_1^5 = \theta_2^5 = 0$. Arrêt.

Cette exemple illustre que la méthode multi-coupes peut parfois prendre plus d'itérations que l'approche simple coupe.

Exemple : L-shaped regularisé

Prenons $a^1 = -0.5$ comme point de départ. Ce point correspond à la solution du problème obtenu en remplaçant ξ par $E[\xi]$. En effet, le programme de seconde étape devient

$$\begin{aligned} \min \quad & \frac{5}{4}y_1 + \frac{1}{7}y_3 + \frac{1}{2}y_4 \\ \text{t.q.} \quad & y_1 - y_2 - y_3 - y_4 = -\frac{1}{2} - x, \\ & y_2 + y_5 = 2, \\ & y_3 + y_6 = 7, \\ & y \geq 0. \end{aligned}$$

Question : comment ai-je formé ce programme ?

Exercice : vérifier que $a^1 = -0.5$ est la valeur optimale du programme (utilisez les conditions KKT).

Exemple : L -shaped régularisé (2)

Pour $a^1 = -0.5$, la solution optimale de seconde étape est $y = (0.5, 0, 0, 0, 0, 0, 0)^T$, $Q_1(a^1) = 0$, $Q_2(a^2) = 3/4$, aussi $Q(a^1) = 3/8$. L'approche régularisée conduit aux itérations suivantes.

Itération 1 Les coupes $\theta_1 \geq 0$, $\theta_2 \geq -3/4x$ sont ajoutées.
 $a^2 := a^1$.

Itération 2 Le programme maître régularisé est

$$\begin{aligned} \min \theta_1 + \theta_2 + \frac{1}{2}(x + 0.5)^2 \\ \text{t.q. } \theta_1 \geq 0, \theta_2 \geq -\frac{3}{4}x, \end{aligned}$$

avec pour solution $x^2 = 0.25$, $\theta_1 = 0$, $\theta_2 = -3/16$. Une nouvelle coupe $\theta_2 \geq 0$ est ajoutée. Comme $Q(0.25) = 0 < Q(a^1)$, $a^3 := 0.25$ (pas approximatif sérieux).

Exemple : L-shaped régularisé (3)

Itération 3 Le programme maître régularisé est

$$\begin{aligned} \min \theta_1 + \theta_2 + \frac{1}{2}(x - 0.25)^2 \\ \text{t.q. } \theta_1 \geq 0, \theta_2 \geq -\frac{3}{4}x, \theta_2 \geq 0, \end{aligned}$$

avec la solution $x^3 = 0.25$, $\theta_1 = 0$, $\theta_2 = 0$. Comme $\theta^\nu = Q(a^\nu)$, une solution est trouvée.

Théorème

Si le problème original a une solution, alors l'algorithme s'arrête après un nombre fini d'itérations. Sinon, il génère une séquence de points réalisables $\{a^\nu\}$ telle que $Q(a^\nu) \rightarrow -\infty$, as $\nu \rightarrow \infty$.

Démonstration.

See Birge et Louveaux, section 6.2. □