

# Programmation stochastique

## Méthode *L*-shaped

Fabian Bastin

IFT-6512 – Hiver 2010

## Lecture demandée : Section 3.2, Kall et Wallace

Résumons nos différentes formulations.

$$\min_{x \in \mathbb{R}_+^n \mid Ax=b} \left\{ c^T x + E_{\xi} \left[ \min_{y \in \mathbb{R}_+^p} \{ q(\omega)^T y \mid Wy = h(\omega) - T(\omega)x \} \right] \right\}$$

$$\min_{x \in \mathbb{R}_+^n \mid Ax=b} \left\{ c^T x + E_{\xi} [v(h(\omega) - T(\omega)x)] \right\}$$

$$\min_{x \in \mathbb{R}_+^n \mid Ax=b} \left\{ c^T x + E_{\xi} [Q(x, \xi)] \right\}$$

$$\min_{x \in \mathbb{R}_+^n} \left\{ c^T x + Q(x) \mid Ax = b \right\}$$

Nous avons étudié ses propriétés. Mais comment résoudre en pratique ?

- Ensemble de faisabilité de première étape :

$$K_1 = \{x \in \mathbb{R}_+^n \mid Ax = b\}.$$

- Ensemble de faisabilité de seconde étape :

$$K_2 = \{x \mid Q(x) < \infty\}.$$

Par conséquent nous pouvons réécrire (encore !) le problème comme suit :

$$\min_x \{c^T x + Q(x) \mid x \in K_1 \cap K_2\}.$$

Ensemble de faisabilité de seconde étape :

$$K_2 = \{x \mid Q(x) < \infty\}.$$

Que signifie  $Q(x) < \infty$  ?

$Q(x) = E_{\xi}[Q(x, \xi)]$ , avec

$$Q(x, \xi) = \min_{y \in \mathbb{R}_+^p} \{q(\xi)^T y \mid Wy = h(\omega) - T(\omega)x\}.$$

En fait, on devrait écrire  $\inf_{y \in \mathbb{R}_+^p} \{q(\xi)^T y \mid Wy = h(\xi) - T(\xi)x\}$  : nous n'avons aucune garantie que la fonction  $Q(x, \xi)$  soit bornée inférieurement sur son ensemble réalisable (défini par rapport à  $y$ ).

On peut donc avoir  $Q(x, \xi) = -\infty$ , pour  $y^*$  non défini.

**Convention** : si  $Q(x, \xi)$  n'admet aucun point réalisable, on notera  $Q(x, \xi) = +\infty$ . La notation de  $K_2$  donne ainsi l'ensemble des  $x$  tels que le problème de seconde étape est réalisable, mais rien de plus ! En particulier, on peut avoir  $Q(x) = -\infty$ .

Mais...  $|Q(x, \xi)| = \infty$  n'implique pas que  $|Q(x)| = \infty$ .  
(Pourquoi ?)

Rappel : nous travaillons avec un **recours fixe**.

Pour la simplicité du raisonnement, nous supposerons dans la suite que le recours est **relativement complet** (i.e. ?).

# Types de recours

Définissons le cône  $\text{pos}W$  par

$$\text{pos}W = \{t \mid t = Wy, y \geq 0\}.$$

Cône :  $\forall t \in \text{pos}W, \alpha > 0, \alpha t \in \text{pos}W$ .

Nous avons

$Wy = t, y \geq 0$  est réalisable ssi  $t \in \text{pos}W$ .

**Recours complet** :  $\text{pos}W = \mathbb{R}^m$ . Ceci implique

$$h(\omega) - T(\omega)x \in \text{pos}W \quad \forall \omega, x.$$

**Recours relativement complet** :

$h(\omega) - T(\omega)x \in \text{pos}W$  pour tout  $\omega$  et tout  $x \geq 0$  t.q.  $Ax = b$ .

Comment identifier le recours relativement complet ?

Utilisation du lemme de Farkas : condition nécessaire et suffisante pour la faisabilité d'un système d'équations linéaires.

## Lemme de Farkas

$\{x \mid Ax = b, x \geq 0\} \neq \emptyset$  si et seulement si

$$A^T u \geq 0 \text{ implique } b^T u \geq 0.$$

Recours relativement complet :  $z \in \text{pos}W$  ou

$\{y \mid Wy = z, y \geq 0\} \neq \emptyset$ .

Par le lemme de Farkas, cela équivaut à

$(W^T u \geq 0) \Rightarrow (z^T u \geq 0)$ .

# Implications du lemme de Farkas

Par conséquent,  $(-W^T u \geq 0) \Rightarrow (-z^T u \geq 0)$ , ou  
 $(W^T u \leq 0) \Rightarrow (z^T u \leq 0)$ , ou encore

$$z^T t \leq 0 \text{ quand } t \in \{u \mid W^T u \leq 0\}.$$

Or, par définition du produit scalaire,

$$\{u \mid W^T u \leq 0\} = \{u \mid y^T W^T u \leq 0 \forall y \geq 0\}$$

ou encore, par définition de  $\text{pos } W$ ,

$$\{u \mid W^T u \leq 0\} = \{u \mid u^T W y \leq 0 \forall y \geq 0\} = \{u \mid u^T z \leq 0 \forall z \in \text{pos } W\}.$$

Cette dernière expression définit le **cône polaire** de  $\text{pos } W$  :

$$\text{pol } \text{pos } W = \{u \mid u^T z \leq 0 \forall z \in \text{pos } W\}.$$

Le lemme de Farkas va nous être utile pour vérifier et **forcer** le recours relativement complet.

# Equivalent déterministe

Supposons è présent que le support de  $\xi$  est fini (le cas continu ou discret, infini, pourra s'y ramener pas échantillonnage) ; développons suivant les  $S$  scénarios.

Probabilité du scénario  $s$  :  $p_s$ .

$$\min_{x \in X, y_1, \dots, y_S} c^T x + p_1 q^T y_1 + p_2 q^T y_2 + \dots + p_S q^T y_S$$

t.q.

$$\begin{array}{rcl} Ax & & = b \\ T_1 x + W y_1 & & = h_1 \\ T_2 x & + W y_2 & = h_2 \\ \vdots & & \ddots \\ T_S x & & + W y_S = h_S \end{array}$$

$$x \in X, y_1 \in Y, y_2 \in Y, \dots, y_S \in Y.$$

# Propriétés de l'équivalent déterministe

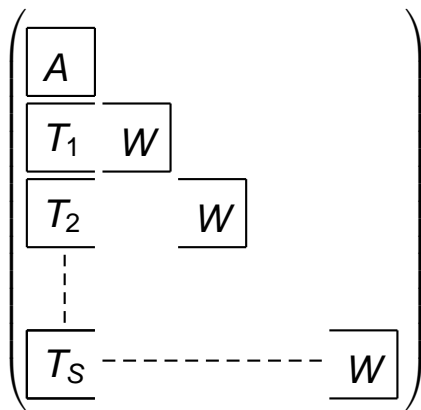
- $y_s \stackrel{def}{=} y(\omega_s)$  est l'action de recours à prendre si le scénario  $s$  a lieu.
- On obtient un programme linéaire, . . . de très grande taille :
  - $n + pS$  variables ;
  - $m1 + mS$  contraintes.
- Mais la matrice du programme linéaire a une structure (en escalier) très spéciale.
- **Idée** : utiliser des approches de décompositions.

**Principe général** : le terme non-linéaire de l'objectif, à savoir la fonction de recours, implique de calculer la solution de tous les programmes linéaires de recours de seconde étape.

Est-ce possible d'éviter des évaluations répétées des fonctions impliquées afin de calculer ces solutions ?

Idée : construire un programme maître en  $x$ , mais n'évaluer la fonction objectif exactement que comme sous-problème.

# Structure par blocs



Cette figure a donné le nom à la méthode : *L-shaped*.

# Structure par blocs (dual)

$$\begin{pmatrix} A^T & T_1^T & T_2^T & \dots & T_S^T \\ & W^T & & & \\ & & W^T & & \\ & & & \dots & \\ & & & & W^T \end{pmatrix}$$

## Section 1.6.4, Kall et Wallace

Supposons pour le moment que  $\Xi$  contient un seul élément.

Le problème devient

$$\begin{aligned} \min_{x,y} \quad & c^T x + q^T y \\ \text{t.q.} \quad & Ax = b, \\ & Tx + Wy = h, \\ & x \geq 0, \quad y \geq 0. \end{aligned}$$

Le problème peut être reformulé comme

$$\begin{aligned} \min_x \quad & c^T x + f(x) \\ \text{t.q.} \quad & Ax = b, \quad x \geq 0, \end{aligned}$$

avec

$$f(x) \stackrel{\text{def}}{=} \min_y \{q^T y \mid Wy = h - Tx, y \geq 0\}.$$

# Reformulation du problème

De plus, nous supposons que le problème est réalisable et que  $\{x \mid Ax = b, x \geq 0\}$  est borné. Ceci implique que

$$\inf\{f(x) \mid Ax = b, Tx + Wy = h, x \geq 0\}$$

est fini et  $f(x)$  peut être uniformément bornée inférieurement sur  $K_1$ , disons par  $\theta$ .

Si on connaissait  $f(x)$  explicitement, on pourrait écrire

$$\begin{aligned} \min_x \quad & c^T x + f(x) \\ \text{t.q.} \quad & Ax = b, \\ & \theta - f(x) \geq 0, \\ & x \geq 0. \end{aligned}$$

Essayons de construire une **séquence de nouvelles contraintes linéaires (additionnelles)**, afin de définir un ensemble réalisable  $\mathcal{B}_1$  de vecteurs  $(x_1, \dots, x_n, \theta)^T$  dans  $\mathbb{R}^{(n+1)}$ , tel que, avec

$$\mathcal{B}_0 = \{(x^T, \theta^T) \mid Ax = b, x \geq 0, \theta \in \mathbb{R}\},$$

le problème

$$\min_{(x, \theta) \in \mathcal{B}_0 \cap \mathcal{B}_1} c^T x + \theta$$

fournisse une solution (de première étape) du problème de départ.

# Méthode de décomposition duale

**Etape 1** Soit  $\theta_0$  borne inférieure pour

$$\min_y \{q^T y \mid Ax = b, Tx + Wy = h, x \geq 0, y \geq 0\}.$$

Notons  $(\hat{x}, \hat{\theta})$  une solution de

$$\min_x \{c^T x + \theta \mid Ax = b, \theta \geq \theta_0, x \geq 0\}.$$

Posons également  $\mathcal{B}_1 := \{\mathbb{R}^n \times \{\theta\} \mid \theta \geq \theta_0\}$ .

**Etape 2** Evaluons

$$\begin{aligned} f(\hat{x}) &= \min_y \{q^T y \mid Wy = h - T\hat{x}, y \geq 0\} \\ &= \max_u \{(h - T\hat{x})^T u \mid W^T u \leq q\}. \text{ (dualité)} \end{aligned}$$

Nous devons distinguer deux cas.

## Méthode de décomposition duale (2)

**Etape 2 (a)** Si  $f(\hat{x}) = +\infty$ ,  $\hat{x}$  n'est pas réalisable avec toutes les contraintes de seconde étape : aucun  $y$  ne satisfait le système linéaire. Du lemme de Farkas,  $\exists \tilde{u}$  tel que

$$W^T \tilde{u} \leq 0 \text{ et } (h - T\hat{x})^T \tilde{u} > 0.$$

Au contraire, pour tout  $x$  réalisable, il existe un  $y \geq 0$  t.q.  $Wy = h - Tx$ . Comme dans la dérivation du cône polaire, nous avons

$$\tilde{u}^T (h - Tx) = \tilde{u}^T Wy \leq 0.$$

Dès lors

$$\tilde{u}^T h \leq \tilde{u}^T Tx,$$

et nous pouvons introduire la **coupe de faisabilité**

$$\tilde{u}^T (h - Tx) \leq 0.$$

Redéfinissons  $\mathcal{B}_1 := \mathcal{B}_1 \cap \{(x^T, \theta) \mid \tilde{u}^T (h - Tx) \leq 0\}$ .

## Méthode de décomposition duale (3)

**Etape 2 (b)** Si  $f(\hat{x})$  est fini, pour  $\hat{x}$  donné, nous avons une solution optimale primale de base  $\hat{y}$  et une solution optimale duale basique  $\hat{u}$ , avec

$$f(\hat{x}) = (h - T\hat{x})^T \hat{u}.$$

Or, pour n'importe quel  $x$ ,

$$\begin{aligned} f(x) &= \sup_u \left\{ (h - Tx)^T u \mid W^T u \leq q \right\} \\ &\geq (h - Tx)^T \hat{u} \\ &= \hat{u}^T (h - Tx). \end{aligned}$$

$\theta$  servant à approcher  $f(x)$ , nous voudrions

$$\theta \geq \hat{u}^T (h - Tx).$$

# Méthode de décomposition duale (3)

$\theta \geq \hat{u}^T(h - Tx)$  est violée en  $(\hat{x}, \hat{\theta})^T$  ssi  
 $\hat{u}^T(h - T\hat{x}) > \hat{\theta}$ .

Dans ce cas, nous introduisons la coupe  
d'optimalité, rejetant la solution nonoptimale  
 $(\hat{x}^T, \hat{\theta})^T$  :

$$\theta \geq \hat{u}^T(h - Tx),$$

et nous redéfinissons

$$\mathcal{B}_1 := \mathcal{B}_1 \cap \{(x^T, \theta) \mid \theta \geq \hat{u}^T(h - Tx) \leq 0\}.$$

Sinon, si  $f(\hat{x}) \leq \hat{\theta}$ , stop :  $\hat{x}$  est une solution  
optimale de première étape. Autrement, étape 3.

**Etape 3** Résoudre le problème mis à jour

$$\min\{c^T x + \theta \mid (x^T, \theta) \in \mathcal{B}_0 \cap \mathcal{B}_1\}.$$

Recommencer l'étape 2 avec la solution  $(\hat{x}, \hat{\theta})$   
ainsi trouvée.

# Justification : détour par la programmation linéaire

Considérons le système  $Ax = b$ . Comme précédemment, nous supposons que  $\text{rang}(A) = m$ , et l'ensemble réalisable du problème linéaire standard

$$\mathcal{B} = \{x \mid Ax = b, x \geq 0\}.$$

**Solution réalisable de base** :  $\hat{x} \in \mathcal{B}$  (faisabilité), et soit  $I(\hat{x}) = \{i \mid \hat{x}_i > 0\}$ , l'ensemble  $\{A_i \mid i \in I(\hat{x})\}$  de colonnes de  $A$  est linéairement indépendant.

Les composantes  $\hat{x}_i$ ,  $i \in I(\hat{x})$ , sont l'unique solution du système

$$\sum_{i \in I(\hat{x})} A_i x_i = b.$$

## Théorème (Optimalité LP)

*Considérons le problème linéaire standard*

$$\begin{aligned} \min \quad & c^T x \\ \text{t.q.} \quad & Ax = b, \quad x \geq 0. \end{aligned}$$

*Ce problème peut être résolu si et seulement si*

$$B = \{x \mid Ax = b, x \geq 0\} \neq \emptyset, \text{ faisabilité}$$

*et*

$$c^T y \geq 0, \quad \forall y \in \mathcal{C} = \{y \mid Ay = 0, y \geq 0\}.$$

*Si ces deux conditions sont satisfaites, il y a au moins une solution réalisable de base qui est une solution optimale.*

Vérifions d'abord que la méthode est bien définie.

## Justification de l'étape 1

Nous avons supposé le problème résolvable. Par le théorème précédent (optimalité LP)

$$\{(x, y) \mid Ax = b, Tx + Wy = h, x \geq 0, y \geq 0\} \neq \emptyset,$$

$$\{v \mid Wv = 0, q^T v < 0, v \geq 0\} = \emptyset.$$

De plus,  $\{x \mid Ax = b, x \geq 0\}$  est fermé et borné. Dès lors  $\inf\{f(x) \mid Ax = b, x \geq 0\}$  est fini (fonction continue sur un ensemble non vide compact) et il existe une borne inférieure  $\theta_0$  pour  $f(x)$ .

Par conséquent, le problème

$$\min\{c^T x + \theta \mid Ax = b, \theta \geq \theta_0, x \geq 0\}$$

peut être résolu.

## Justification (partielle) de l'étape 2

$$f(\hat{x}) = +\infty$$

Condition pour que l'algorithme se termine : ce cas ne peut se produire qu'un nombre fini de fois.

Rappelons que

$$\begin{aligned} f(\hat{x}) &= \min\{q^T y \mid Wy = h - T\hat{x}, y \geq 0\} \\ &= \max\{(h - T\hat{x})^T u \mid W^T u \leq q\}. \end{aligned}$$

$f(\hat{x}) = +\infty$  signifie que le problème ne peut être résolu. Dès lors,  $\hat{x}$  n'est pas réalisable avec les contraintes de seconde étape. Du lemme de Farkas, on tire

$$\{u \mid W^T u \leq 0, (h - T\hat{x})^T u > 0\} \neq \emptyset.$$

En d'autres termes, le problème dual est réalisable mais non borné ( $W^T u \leq q$  est satisfait pour  $u$  assez grand).

## Justification de l'étape 2 (2)

Les propriétés de l'algorithme du simplexe permettent d'établir qu'après un nombre fini d'ajouts de coupes de faisabilité, nous devons avoir

$$(h - Tx)^T u \leq 0 \quad \forall u \in \{u \mid W^T u \leq 0\}.$$

Autrement dit, le problème de recours est réalisable.

$f(\hat{x})$  fini

La méthode du simplexe fournit des solutions de base optimales réalisables, primales et duales, que nous noterons respectivement  $\hat{y}$  et  $\hat{u}$ .

Supposons que nous avons déjà eu  $\tilde{u} = \hat{u}$  dans une étape précédente. Par construction des coupes d'optimalité,

$$\hat{\theta} \geq \hat{u}^T (h - T\hat{x}).$$

## Justification de l'étape 2 (3)

Soit  $u^1, \dots, u^k$  les solutions de base réalisables dans  $\{u \mid W^T u \leq q\}$  utilisées pour construire les coupes d'optimalité.

Au cours de l'étape 3, pour n'importe quel  $x$ , nous minimisons  $\theta$  par rapport à  $\mathcal{B}_1$ , ce qui implique que

$$\hat{\theta} = \max_{1 \leq i \leq k} (h - T\hat{x})^T u^i.$$

**Règle d'arrêt** :  $f(\hat{x}) \leq \hat{\theta}$ . Soit  $\{u^1, \dots, u^k, \dots, u^r\}$  l'ensemble de toutes les solutions de base réalisables de  $\{u \mid W^T u \leq q\}$ . Il suit que

$$\begin{aligned} \hat{\theta} &= \max_{1 \leq i \leq k} (h - T\hat{x})^T u^i \leq \max_{1 \leq i \leq r} (h - T\hat{x})^T u^i \\ &= f(\hat{x}) \leq \hat{\theta}. \end{aligned}$$

L'algorithme s'arrête donc pour  $\hat{\theta}$  optimal.

## Proposition

Si le programme stochastique à deux étapes est soluble et  $\{x \mid Ax = b, x \geq 0\}$  est borné, le méthode de décomposition duale fournit une solution optimale après un nombre fini d'itérations.

Si  $\Xi$  est fini mais n'est pas réduit à un singleton ?

Nous pouvons introduire des coupes de faisabilité et d'optimalité pour toutes les fonctions de recours

$$f_i(x) = \min\{q^{(iT)}y^i \mid Wy^i = h^i - T^ix, y^i \geq 0\},$$

$i = 1, \dots, K$ . On obtient une méthode de décomposition duale *multi – coupes*.

Alternativement, combiner les coupes isolées correspondant aux blocs particuliers  $i = 1, \dots, S$  avec leur probabilités respectives.