

Programmation stochastique

Modèles avec recours

Fabian Bastin

IFT-6512 – Hiver 2011

Rappels (?) : convexité

Les techniques de programmations varient fortement pour les cas convexes et non-convexes. Il importe de donc de savoir de quoi on parle !

- Un ensemble A est convexe si pour tous points x, y dans A , le segment joignant x à y est entièrement à l'intérieur ou sur la frontière de A :

$$\forall x, y \in A, \alpha x + (1 - \alpha)y \in A, \forall 0 \leq \alpha \leq 1.$$

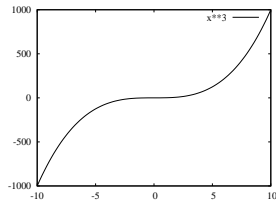
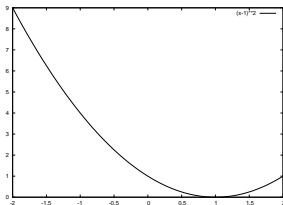
- Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est convexe si pour deux points quelconques x, y , le graphe de f est sous le segment de droite joignant $(x, f(x))$ à $(y, f(y))$ dans \mathbb{R}^{n+1} .
De manière équivalente, cela revient à exiger que

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y), \forall 0 \leq \alpha \leq 1.$$

- Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est concave si pour deux points quelconques x, y , le graphe de f est au-dessus le segment de droite joignant $(x, f(x))$ à $(y, f(y))$ dans \mathbb{R}^{n+1} .
De manière équivalente, cela revient à exiger que

$$f(\alpha x + (1 - \alpha)y) \geq \alpha f(x) + (1 - \alpha)f(y), \quad \forall 0 \leq \alpha \leq 1.$$

- Une fonction nonconvexe est une fonction que n'est ni convexe, ni concave.



Fonction et ensemble convexe ?

Pourquoi parler d'ensemble et de fonction convexes ?
Comment ces notions sont-elles reliées ?

L'**épigraphe** de f est défini comme

$$\text{epi}(f) = \{(x, y) \mid x \in \text{Dom}(f) \text{ et } f(x) \leq y\}.$$

f est convexe si l'épigraphe de f est un ensemble convexe.

Référence conseillée : Nocedal et Wright, Numerical Optimization, Springer, 1999.

Considérons le problème

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

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $S \subset \mathbb{R}^n$. Nous sommes à la recherche d'un point qui minimise f , ... mais c'est-à-dire ?

Minimum global

Un point x^* est un minimum global de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dans X si $f(x^*) \leq f(x) \forall x \in S$.

Un minimum global peut être difficile à trouver, en particulier quand la fonction objectif f est nonconvexe. Dans, on cherchera souvent plutôt un minimum local, c'est-à-dire atteignant la plus petite valeur de f dans un certain voisinage.

Minimum local

Un point x^* est un minimum local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dans X s'il existe un voisinage \mathcal{V} de x^* tel que $f(x^*) \leq f(x)$ pour tout $x \in \mathcal{V} \cap X$.

Un point qui satisfait la définition précédente est parfois appelé un minimum local faible. Cette terminologie le distingue d'un minimum local strict, formellement défini comme suit.

Minimum local strict Un point x^* est un minimum local strict (aussi appelé un minimum local fort) s'il existe un voisinage \mathcal{V} de x^* tel que $f(x^*) < f(x)$ pour tout $x \in \mathcal{V}$ avec $x \neq x^*$.

Un point x^* est un **minimum local isolé** s'il existe un voisinage \mathcal{N} de x^* tel que x^* est le seul minimum local dans \mathcal{N} .
Des minima locaux strict ne sont pas toujours isolés, mais tous les minima locaux isolés sont strict.

Exemple

Considérons la fonction suivante (Nocedal et Wright, page 14) :

$$f(x) = x^4 \cos \frac{1}{x} + 2x^4.$$

$f(x) \in C^2$ et a un minimum local strict en $x^* = 0$. On peut toutefois montrer qu'il existe d'autres minima locaux stricts dans n'importe quel voisinage de 0, autre que 0 lui-même. L'identification d'un minimum devient alors particulièrement délicate.

Choisir un programme convexe ou nonconvexe ?

Avantage de la convexité : tout minimum local est global !

Dès lors il suffit de chercher un minimum local.

La même chose prévaut en programmation non-convexe, car il est très difficile de trouver un minimum global, ou même de vérifier qu'un minimum local est global.

Par conséquent, la convexité est une propriété souvent désirée, mais pas forcément réaliste.

Recherche de solutions : vérification de conditions d'optimalité.
Nous y reviendrons.

Un programme linéaire aléatoire

Considérons le programme linéaire suivant, paramétrisé par le vecteur aléatoire ω :

$$\begin{aligned} \min_x \quad & c^T x \\ \text{t.q.} \quad & Ax = b \\ & T(\omega)x = h(\omega) \\ & x \in X, \end{aligned}$$

avec $X = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$. Exemple :

$$\begin{aligned} \min_x \quad & x_1 + x_2 \\ \text{t.q.} \quad & \omega_1 x_1 + x_2 \geq 7 \\ & \omega_2 x_1 + x_2 \geq 4 \\ & x_1, x_2 \geq 0. \end{aligned}$$

Evacuer l'aléatoire ?

Une approche populaire est de chercher à inférer des valeurs raisonnables pour ω_1 et ω_2 . Comment ?

Propositions :

- non biaisé : choisir les valeurs moyennes de chaque variable aléatoire ;
- pessimiste : choisir les valeurs de pire cas pour ω ;
- optimiste : choisir les valeurs de meilleur cas pour ω .

Chaque approche donnera une solution optimale différente !

A nouveau, nous devons traiter des problèmes de décisions où la décision x doit être faite avant que la réalisation de ω soit connue.

- Nous devons, cependant, connaître la distribution de ω sur Ω .
- Dans des modèles avec recours, les contraintes aléatoires sont des contraintes “douces”. Elles peuvent être violées, mais le coût des violations influencera le choix de x .
- En fait, un programme linéaire de seconde étape est introduit ; il décrit comment les contraintes aléatoires violées sont gérées.

Le nouveau problème. . .

Dans le cas le plus simple, nous pouvons simplement pénaliser la déviation des contraintes par des vecteurs de coefficients de pénalité q_+ et q_- .

$$\begin{aligned} \min \quad & c^T x + q_+^T s(\omega) + q_-^T t(\omega) \\ \text{t.q.} \quad & Ax = b, \\ & T(\omega)x + s(\omega) - t(\omega) = h(\omega), \\ & x \in X. \end{aligned}$$

Mais sous cette forme, il est toujours impossible de résoudre le problème !

Le nouveau problème d'optimisation

Un problème raisonnable, et soluble, est alors

$$\min c^T x + E_{\omega}[q_+^T s(\omega) + q_-^T t(\omega)]$$

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

$$T(\omega)x + s(\omega) - t(\omega) = h(\omega), \quad \forall \omega \in \Omega$$

$$x \in X.$$

- En général, nous pouvons réagir de manière adaptée (voire optimale) : nous disposons d'un recours pour “corriger” la première décision une fois l'inconnue levée.
- Une structure de recours en programmation linéaire est fournie par 3 éléments :
 - un ensemble $Y \subset \mathbb{R}^p$ qui décrit l'ensemble réalisable des actions de recours, par exemple $Y = \{y \in \mathbb{R}^p \mid y \geq 0\}$;
 - q : un vecteur de coûts de recours ;
 - W : une matrice $m \times p$, appelée la matrice de recours.

Les précédentes considérations nous conduisent à formuler le programme suivant :

$$\begin{aligned} \min \quad & c^T x + E_{\omega}[q^T y] \\ \text{t.q.} \quad & Ax = b, \\ & T(\omega)x + Wy(\omega) = h(\omega), \quad \forall \omega \in \Omega \\ & x \in X, \\ & y \in Y, \quad \forall \omega. \end{aligned}$$

On pourrait avoir W variant suivant la réalisation ω . Si comme dans la formulation, W est unique, on parlera de **recours fixe** : le recours ne change pas avec le scénario.

Mais comment décider de y ?

$$\min_{x \in X | Ax=b} \left\{ c^T + E_{\omega} \left[\min_{y \in Y} q^T y | Wy = h(\omega) - T(\omega)x \right] \right\}.$$

- **Fonction de seconde étape**, ou **fonction de recours** (pénalité) $v : \mathbb{R}^m \rightarrow \mathbb{R}$:

$$v(z) \stackrel{\text{def}}{=} \min_{y \in Y} \{ q^T y | Wy = z \};$$

cette fonction décrit les coûts correspondants à n'importe quel vecteur z de "déviations des contraintes aléatoires $T(\omega)x = f(\omega)$."

- **Fonction de valeur espérée**, ou recours de minimum d'espérance $Q : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$Q(x) = E_{\omega} [v(h(\omega) - T(\omega)x)].$$

Elle décrit le coût attendu (espéré) du recours, pour n'importe quelle politique $x \in \mathbb{R}^n$.

Le problème (linéaire) PS à 2 étapes

En utilisant ces définitions, nous pouvons réécrire notre problème de programmation stochastique (PS) avec recours en termes du vecteur x seulement :

$$\min_{x \in X} \{c^T x + Q(x) \mid Ax = b\}.$$

C'est un problème de programmation (nonlinéaire) dans \mathbb{R}^n . La facilité de résolution d'un tel problème dépend des propriétés de $Q(x)$.

$Q(x)$ est

- linéaire ?
- convexe ?
- continu ?
- différentiable ?

$$\min_{x, y(\omega)} E_{\omega}[c^T x + q^T y(\omega)]$$

tel que $Ax = b$

contraintes de 1^{ère} étape

$$T(\omega)x + Wy(\omega) = h(\omega), \forall \omega \in \Omega$$

contraintes de 2^e étape

$$x \in X, y(\omega) \in Y.$$

Considérons le cas (discret) où $\Omega = \{\omega_1, \omega_2, \dots, \omega_S\} \subset \mathbb{R}^r$.

$$P(\omega = \omega_s) = p_s, \quad s = 1, 2, \dots, S$$

$$T_s = T(\omega), \quad h_s = h(\omega)$$

Développons suivant les S scénarios.

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

t.q.

$$Ax = b$$

$$T_1 x + W y_1 = h_1$$

$$T_2 x + W y_2 = h_2$$

\vdots

\ddots

$$T_S x + W y_S = h_S$$

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

Equivalent déterministe (2)

- $y_s = y(\omega_s)$ est l'action de recours à prendre si le scénario ω_s se produit.
- Avantage : c'est un programme linéaire.
- Inconvénient : c'est programme linéaire de (très) grande dimension :
 - $n + pS$ variables ;
 - $m_1 + mS$ constraints.
- Avantage : la matrice du programme linéaire a une structure très spéciale (en escalier).
Pourra-t-on l'exploiter ?

Grande taille, . . . mais encore ?

Supposons que nous avons r variables aléatoires ($\Omega \subset \mathbb{R}^r$).

- Considérons le problème suivant (tiré de Linderoth). Une compagnie Télécom veut étendre son réseau de manière à répondre à une demande inconnue (aléatoire).
- Il y a 86 demandes inconnues. Chaque demande est indépendante et peut prendre une valeur parmi un ensemble de 7 valeurs. Par conséquent

$$S = |\Omega| = 7^{86} \approx 4.77 \times 10^{72}.$$

. . . nombre de particules subatomiques dans l'univers !

- Pire encore. . .
Si Ω n'est pas fini, mais contient un ensemble infini d'éléments ? C'est en particulier le cas avec des variables aléatoires continues. Notre "équivalent déterministe" aurait un ∞ -té de variables et de contraintes !
- Nous pouvons alors résoudre un problème approché, obtenu par échantillonnage.

Revenons sur notre problème jouet

$$\min_x x_1 + x_2$$

$$\text{t.q. } \omega_1 x_1 + x_2 \geq 7$$

$$\omega_2 x_1 + x_2 \geq 4$$

$$x_1, x_2 \geq 0,$$

où $\omega_1 \sim U[1, 4]$, $\omega_2 \sim U[1/3, 1]$.

Comment construire l'équivalent déterministe ?

Exemple : une formulation par recours

Supposons pour un moment que Ω est fini, avec un nombre de scénarios S .

$$\min_x x_1 + x_2 + \sum_{s \in S} p_s \lambda (y_{1s} + y_{2s})$$

$$\text{t.q. } \omega_{1s} x_1 + x_2 + y_{1s} \geq 7$$

$$\omega_{2s} x_1 + x_2 + y_{1s} \geq 4$$

$$x_1, x_2 \geq 0,$$

$$y_{1s}, y_{2s} \geq 0.$$

Une des difficultés est dès lors la construction de la formulation déterministe. Que choisir pour λ ici ?

Comment choisir les scénarios ? On pourra procéder par échantillonnage Monte-Carlo, avec $p_s = 1/N, \forall s$. Nous y reviendrons.

Exemple : une formulation par recours (suite)

De manière plus générale, nous pouvons construire le programme

$$\begin{aligned} \min_x \quad & x_1 + x_2 + E_\omega[Q(x)] \\ \text{t.q.} \quad & x_1, x_2 \geq 0, \end{aligned}$$

et

$$\begin{aligned} Q(x) = \min_y \quad & q_1 y_1 + q_2 y_2 \\ \text{t.q.} \quad & \omega_1 x_1 + x_2 + y_1 \geq 7, \\ & \omega_2 x_1 + x_2 + y_2 \geq 4. \end{aligned}$$

Un autre exemple : le vendeur de journaux

Source : Birge et Louveaux, Section 1.1.

- Un vendeur de journaux doit décider combien de journaux acheter afin de maximiser son profit. Il ne sait pas au début de la journée combien de journaux il pourra vendre (la demande).
- Chaque journal coûte c , et peut être vendu à un prix q .
- Le vendeur peut retourner chaque journal invendu à la fin de la journée à un prix r .
- Connaissant la distribution de probabilité $F(t) = P(\omega \leq t)$, combien de journaux le vendeur doit-il acheter pour maximiser ses profits ?

Le vendeur de journaux (2)

- Selon nos définitions de recours, le vendeur de journaux souhaiterait résoudre le problème d'optimisation suivant :

$$\max_{x \geq 0} -cx + Q(x),$$

- $Q(x)$ est le montant espéré des ventes si le vendeur de journaux achète x journaux :

$$Q(x)E_{\omega}[Q(x, \omega)].$$

- Ici $Q(x, \omega)$ est le montant d'argent qu'obtient le vendeur s'il achète x journaux et que la demande est ω .

Le vendeur de journaux (3)

- Comme précédemment, nous pourrions construire un problème linéaire équivalent (présenté dans Birge et Louveaux).
- Peut-on simplifier ? Raisonement proposé par Linderoth. Il est facile de voir que

$$Q(x, \omega) = \begin{cases} qx & \text{si } x \leq \omega, \\ q\omega + r(x - \omega) & \text{si } x \geq \omega. \end{cases}$$

De là,

$$\begin{aligned} Q(x) &= E_{\omega}[Q(x, \omega)] = \int_{-\infty}^{\infty} Q(x, \omega) dF(\omega) \\ &= \int_{-\infty}^x (q\omega + r(x - \omega)) dF(\omega) + \int_x^{\infty} qx dF(\omega). \end{aligned}$$

Dès lors, nous avons

$$\begin{aligned} Q(x) &= (q - r) \int_{-\infty}^x \omega dF(\omega) + rx \int_{-\infty}^x dF(\omega) + qx \int_x^{\infty} dF(\omega) \\ &= (q - r) \int_{-\infty}^x \omega dF(\omega) + rxF(x) + qx(1 - F(x)). \end{aligned}$$

Quid de

$$\int_{-\infty}^{\infty} g(\omega) dF(\omega)?$$

Dans le cadre de variables continues, pour lesquelles F est continue et dérivable, il suit que $dF(\omega) = f(\omega)d\omega$.

Pour une loi discrète, nous pouvons réécrire F comme

$$F(x) = \sum_{j=1}^k p(x_j),$$

avec k tel que $x_k \leq x$ et $\forall j > k, x_j > x$.

Dès lors

$$\int_{-\infty}^{\infty} g(x) dF(x) = \sum_j g(x_j) p(x_j).$$

Nous pouvons aussi combiner les deux formes si F est une combinaison de variables discrètes et continues.

Intégration par parties

Supposons que F satisfait la propriété $\lim_{t \rightarrow -\infty} tF(t) = 0$.
Nous pouvons alors intégrer par parties pour obtenir que

$$\begin{aligned}\int_{-\infty}^x \omega dF(\omega) &= \omega F(\omega) \Big|_{-\infty}^x - \int_{-\infty}^x F(\omega) d(\omega) \\ &= xF(x) - \int_{-\infty}^x F(\omega) d(\omega).\end{aligned}$$

Bref,

$$Q(x) = qx - (q - r) \int_{-\infty}^x F(\omega) d\omega.$$

Rappelons le problème initial...

$$\max_{x \geq 0} -cx + Q(x),$$

Il nous “reste” donc à résoudre ce problème. Pour ce faire, nous utiliserons les conditions d’optimalité associées.

Solution de la deuxième étape

Les conditions KKT sont ici très simples...

Il suffit d'annuler le gradient de l'objectif, en notant que

$$Q(x) = q - (q - r)F(x),$$

aussi, il faut résoudre

$$-c + q - (q - r)F(x) = 0.$$

Le vendeur de journaux (suite)

La solution x^* est dès lors

$$x^* = F^{-1} \left(\frac{q - c}{q - r} \right).$$

Un exemple : $c = 0.15$, $q = 0.25$, $r = 0.02$, $\omega \sim N(650, 80)$.

Alors

$$x^* = N^{-1}(-0.1/0.23) \approx 636.86.$$

Autre interprétation, plus intuitive : supposons que le vendeur a acheté t journaux. Quel est le revenu marginal espéré s'il achète un journal supplémentaire ? D'un point de vue économique, nous souhaiterions que ce revenu marginal soit égal à 0. . . comme explicité par les conditions KKT !

Dénotons le revenu marginal espéré par MR ; on a

$$\begin{aligned}MR(t) &= -c + qP[\omega \geq t] + rP[\omega \leq t] \\ &= -c + q(1 - F(t)) + rF(t).\end{aligned}$$

En annulant cette expression, on obtient

$$MR(t) = 0 \text{ ssi } F(t) = \frac{q - c}{q - r},$$

et donc on retrouve la solution précédente.