Master Informatique de Grenoble – automne 2015
Recherche Opérationnelle, Combinatoire et Optimisation (ROCO)
Cours "programmation mathématique et applications industrielles"
Jérôme Malick

Feuille 3 – Décomposition de Benders

Exercice 1 – Décomposition par les variables en optimisation stochastique. Reprenons l'exemple vu au premier cours d'introduction. Nous devons prendre aujourd'hui une décision $x \in \mathbb{R}^n$, respectant certaines contraintes représentées par un polyèdre $X \subset \mathbb{R}^n$, dont le coût est double :

- un coût (déterministe) immédiat $c^\top x$ (avec $c \in \mathbb{R}^n$)
- un coût (aléatoire) futur $Q(x, \xi)$, dépendant des décisions présentes (x) et des évènements futurs inconnus (ξ), qui correspond aux coûts des décisions du futur $y \in \mathbb{R}^m$; plus précisement,

$$Q(x,\xi) = \left\{ \begin{array}{l} \min \quad q^\top y \\ \quad Tx + Wy = h(\xi) \\ \quad y \geqslant 0 \end{array} \right. \tag{1}$$

avec $q \in \mathbb{R}^m$, $T \in \mathbb{R}^{n \times p}$, $W \in \mathbb{R}^{m \times p}$ et $h(\xi) \in \mathbb{R}^p$. On suppose les matrices W et T sont telles que : ($\bullet$) l'ensemble réalisable de (1) est compact (pour tout ξ), ($\bullet$) le polyèdre d'équation $-W^\top \lambda \leqslant q$ est lui aussi compact, et ($\bullet$) $h(\xi)$ n'appartient jamais à l'image de T.

Nous voulons prendre la décision optimale qui minimise le coût total le plus probable (mathématiquement l'espérance du coût), ce qui nous amène au problème d'optimisation stochastique

$$\left\{ \begin{array}{l} \min \quad c^\top x + \mathbb{E}[Q(x,\xi)] \\ \quad x \in X \end{array} \right.$$

Nous faisons l'hypothèse que la variable aléatoire représentant le futur admet un nombre fini N de réalisations ξ_i ($i = 1, \ldots, N$) (les N scénarios du futur). À chaque scénario, sont associés une probabilité p_i et un vecteur $h_i = h(\xi_i)$. L'espérance s'écrit donc comme une somme et le problème d'optimisation stochastique s'écrit alors comme le problème déterministe suivant :

$$\left\{ \begin{array}{l} \min \quad f(x) = c^\top x + \displaystyle\sum_{i=1}^{N} p_i Q(x, \xi_i) \\ \quad x \in X. \end{array} \right. \tag{2}$$

En explicitant les $Q(x, \xi_i)$, on a la reformulation suivante du problème

$$\left\{ \begin{array}{l} \min \quad c^\top x + \displaystyle\sum_{i=1}^{N} p_i\, {q_i}^\top y_i \\ \quad Tx + Wy_i = h_i, \quad i = 1, \ldots, N \\ \quad y_i \geqslant 0, \quad i = 1, \ldots, N \\ \quad x \in X. \end{array} \right.$$

Remarquez que ce problème est linéaire en les variables $(x, y_1, \ldots, y_N) \in \mathbb{R}^n \times (\mathbb{R}^m)^N$. Mais il est aussi de très grande taille en pratique, et donc hors de porté des solvers classiques : en effet, en modélisant le futur, il est souhaitable de considérer un grand nombre de scénarios (N grand) pour garantir une représentation précise du processus stochastique sous-jacent. Cet exercice propose de résoudre ce problème (2) par une approche par décomposition basée sur la méthode des faisceaux.

a) Montrer que le problème dual de (1) s'écrit

$$\left\{ \begin{array}{l} \max \quad \lambda^\top (Tx - h(\xi)) \\ \quad -W^\top \lambda \leqslant q \\ \quad \lambda \in \mathbb{R}^p \end{array} \right. \tag{3}$$

Justifier que $Q(x, \xi)$ est égal à la valeur optimale de (3).

b) Montrer que, pour ξ fixé, la fonction $Q(\cdot, \xi)$ est convexe. En déduire que f (la fonction objectif de (2)) est convexe.

c) On suppose qu'il existe un vecteur $y \in (\mathbb{R}_+^*)^m$ avec vérifiant les contraintes $Tx+Wy = h(\xi)$. Notons $\lambda(x, \xi)$ une solution optimale du problème. Donner l'expression d'un sous-gradient de f en $x \in \mathbb{R}^n$.

d) Écrire précisément ce que font les simulateurs des fonctions $Q_i(x) = Q(x, \xi_i)$ (pour $i = 1, \ldots, N$), ainsi que le simulateur de la fonction f.

e) Appliquons la méthode des faisceaux pour minimiser f sur le polyèdre X. On suppose qu'on est à l'itération k de cet algorithme ; écrire le modèle des plans sécants, et écrire précisément le sous-problème quadratique qui donne x_{k+1}.

On va maintenant exploiter le fait que l'objectif est une somme pour construire un modèle plus précis et une version plus efficace de l'algorithme des faisceaux.

f) Proposer un modèle de type « plans sécants » de la fonction f en exploitant le fait que c'est une somme.

g) Écrire précisément le sous-problème quadratique qui correspond à une itération d'un algorithme des faisceaux utilisant cette structure. Exliquer pourquoi on s'attend à ce que x_{k+1} calculé par cette méthode soit meilleur que pour la méthode des faisceaux standard. Expliquer aussi le coût numérique additionnel d'utiliser la deuxième méthode.

On suppose maintenant que N est si grand qu'il n'est pas possible de résoudre tous les N sous-problèmes linéaires (donnant les $Q(x, \xi_i)$ pour $i = 1, \ldots, N$) à chaque itération de la méthode des faisceaux. On suppose qu'on peut en résoudre uniquement un certain nombre (disons, pour fixer les idées, qu'on ne peut résoudre que les $N/10$ premiers).

h) Soient les deux sous-problèmes indexés par i et j. On suppose qu'on a résolu le problème j en x (c'est-à-dire qu'on connait $\lambda(x, \xi_j)$). Définissons l'application affine $\ell_{i,j}(z) = \lambda(x, \xi_j)^\top (Tz - h_i)$. Montrer que $\ell_{i,j}$ minore Q_i (c'est-à-dire qu'elle vérifie $Q_i(z) \geqslant \ell_{i,j}(z)$ pour tout z).

i) Soient $x \in \mathbb{R}^n$ et $i > N/10$ fixés ; on choisit l'application affine minorante $\ell_{i,I(i)}$ pour l'indice $I(i)$ tel que :

$$I(i) = \underset{j=1,\ldots,N/10}{\operatorname{argmax}} \left(\frac{Tx - h_i}{\|Tx - h_i\|} \right)^\top \frac{Tx - h_j}{\|Tx - h_j\|}$$

Expliquer ce choix.

On considère alors le « simulateur inexact » de Q_i qui donne, pour un x en entrée :

— si $i \leqslant N/10$, la valeur exacte de $Q_i(x)$ ainsi qu'un sous-gradient $g(x) \in \partial Q_i(x)$

— si $i > N/10$, la valeur $\lambda(x, \xi_{I(i)})^\top (Tx - h_i)$ et le vecteur $T^\top \lambda(x, \xi_{I(i)})$ donnés par $\ell_{i,I(i)}$.

j) Proposer un modèle de type « plans sécants » pour Q_i utilisant les simulateurs inexacts. (Faire un dessin.) En déduire un troisième modèle pour f qui exploite aussi le fait que ce soit une somme.

k) Écrire précisément le sous-problème quadratique qui correspond à une itération d'un algorithme de type faisceaux utilisant ce modèle. Exliquer pourquoi on s'attend à ce que x_{k+1} calculé par cette méthode soit moins bon que pour la méthode de la question **h)**. Expliquer l'intérêt d'utiliser cette troisième méthode.

Exercice 2 – Décomposition de Benders d'un problème de localisation.

Cet exercice porte sur un problème classique de localisation dans les réseaux : il s'agit de placer des "centres névralgiques d'échanges" ("hub" en anglais) qui collectent des marchandises de leur origines, les transfèrent entre eux, puis les distribuent à leur destinations finales. Le problème consiste à localiser les noeuds "hub" et à faire le transfert des marchandises, le tout de manière optimale.

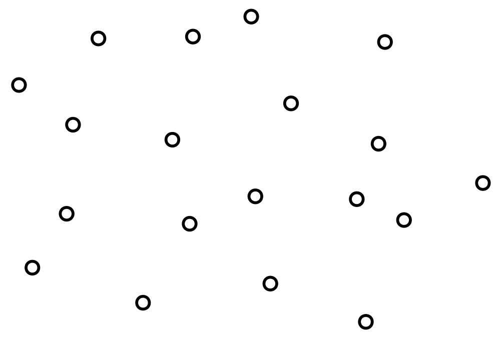
Soit un réseau représenté par un graphe complet (N, A), où la distance entre le noeud i et le noeud j est notée d_{ij}. Voici les données du problème, illustrées par les figures ci-contre. On a $n = |N|$ noeuds, et tout noeud i peut être un hub. On a K articles à distribuer sur ce réseau : pour l'article k, on dispose d'une quantité W_k à acheminer d'un noeud d'origine $o(k) \in N$ à un noeud de destination $d(k) \in N$. Notons χ, τ, δ les coûts nominaux de collecte, de transfert, et de distribution sur le réseau ; ainsi

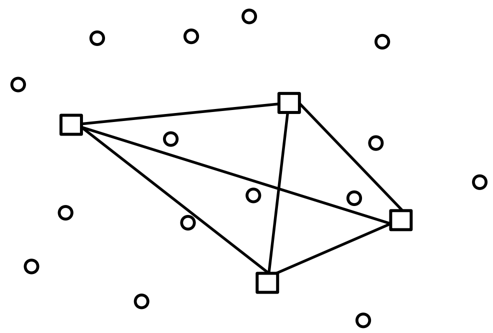
— $\chi d_{o(k)i}$ est le coût de collecte du noeud $o(k)$ vers le hub localié en i,

— τd_{ij} est le coût de transfert entre le hub en i et le hub en j,

— $\delta d_{jd(k)}$ est le coût de distribution du hub en j vers le noeud $o(k)$.

Finalement, notons $f = (f_1, \ldots, f_n) \in \mathbb{R}^n$ avec f_i le coût d'ouverture d'un hub au noeud i.

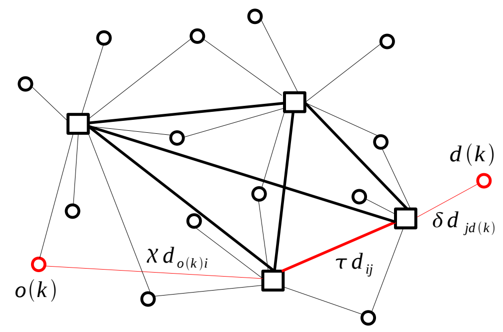




a) Modéliser le problème décrit ci-dessus comme un PLNE avec deux types de variables :

— les variables de localisation : $y_i \in \{0, 1\}$ qui vaut 1 si et seulement si un hub est localisé au noeud i,

— les variables de chemin : $x_{ijk} \in [0, 1]$ qui représente la fraction de la quantité de l'article k qui transite par l'arc entre les hubs i et j.



b) Montrer que le problème s'écrit

$$\text{(P)} \quad \min_{y \in \{0,1\}^n} f^\top y + \sum_{k=1}^{K} v_k(y)$$

où les $v_k(y)$ sont les valeurs optimales de K sous-problèmes indépendants (donc parallélisables). Quelle est la nature de ces K sous-problèmes ? Est-ce que ce sont des problèmes faciles ?

c) Ajouter à (P) une contrainte exprimant qu'il existe au moins un hub dans le réseau. Montrer qu'elle implique qu'on n'aura pas besoin de "coupes de faisabilité".

On admet qu'on est dans le cadre du cours : les fonctions v_k sont convexes et le calcul de $v_k(y)$ pour un y donné fournit aussi un sous-gradient $g(y) \in \partial v_k(y)$. On peut donc résoudre ce problème par l'algorithme de Benders (stabilisé).

d) Supposons être à l'itération ℓ de l'algorithme de Benders stabilisé. Écrire le problème donnant l'itéré suivant $y_{\ell+1}$. Quelle est la nature de ce problème ? Est-ce un problème facile ? Quelles sont les modifications à faire sur ce problème d'une itération à la suivante ?