Multistage stochastic linear programming problems
Block separable recourse

Welington de Oliveira

BAS Lecture 16, May 3, 2016, IMPA
Nested Decomposition - convergence analysis

Block separable recourse
Set YouTube resolution to 480p for best viewing
Exercises

Second list of exercises is available!

Deadline: 02/06/2016
MINI COURSES > SCENARIO GENERATION AND SAMPLING METHODS

Güzin Bayraksan  
Ohio State University, USA

Tito Homem-de-Mello  
University Adolfo Ibáñez, Chile

From May 9th to May 13th, 2016
MINI COURSES > EQUILIBRIUM ROUTING UNDER UNCERTAINTY

Roberto Cominetti, University Adolfo Ibáñez, Chile

From May 16th to May 20th, 2016
MINI COURSES > STOCHASTIC CONVEX OPTIMIZATION METHODS IN MACHINE LEARNING

Mark Schmidt, University of British Columbia

From May 16th to May 20th, 2016
Nested formulation

\[
\begin{align*}
\min_{x_1 \geq 0} & \quad c_1^\top x_1 + \mathbb{E} \left[ \min_{x_2 \geq 0} & \quad c_2^\top x_2 + \mathbb{E} \left[ \cdots + \mathbb{E} \left[ \min_{x_T \geq 0} & \quad c_T^\top x_T \right] \right] \right] \\
A_1 x_1 &= b_1 \\
B_2 x_1 + A_2 x_2 &= b_2 \\
&\quad \cdots \\
&\quad \cdots \\
&\quad \cdots \\
&\quad \cdots
\end{align*}
\]

- Some elements of the data \( \xi = (c_t, B_t, A_t, b_t) \) depend on uncertainties.
Dynamic programming formulation

▶ Stage \( t = T \)

\[
Q_T(x_{T-1}, \xi_T) := \min_{B_T x_{T-1} + A_T x_T = b_T, x_T \geq 0} c_T^\top x_T
\]

▶ At stages \( t = 2, \ldots, T - 1 \)

\[
Q_t(x_{t-1}, \xi_t) := \min_{B_t x_{t-1} + A_t x_t = b_t, x_t \geq 0} c_t^\top x_t + Q_{t+1}(x_t, \xi_t)
\]

▶ Stage \( t = 1 \)

\[
\min_{A_1 x_1 = b_1, x_1 \geq 0} c_1^\top x_1 + Q_2(x_1, \xi_1)
\]

Recourse function

\[
Q_{t+1}(x_t, \xi_t) := \mathbb{E}_{|\xi_t} [Q_{t+1}(x_t, \xi_{t+1})]
\]
Dynamic programming formulation

Scenario tree

- Stage $t = T$

$$Q_T(x_{T-1}, \xi_{[T]}) := \min_{B_T x_{T-1}^a + A_T x_T = b_T} c_T^T x_T$$

- At stages $t = 2, \ldots, T - 1$

$$Q_t(x_{t-1}, \xi_{[t]}) := \min_{B_t x_{t-1}^a + A_t x_t = b_t} c_t^T x_t + \tilde{Q}_{t+1}(x_t, \xi_{[t]})$$

- Stage $t = 1$

$$\min_{A_1 x_1 = b_1} c_1^T x_1 + \tilde{Q}_2(x_2, \xi_{[1]})$$

Cutting-plane model

$$\tilde{Q}_{t+1}(x_t, \xi_{[t]}) := \sum_{j \in C_t} p^{(j)} \left[ Q_{t+1}(x_t, \xi_{[t+1]^j}) \right]$$
**CUTTING-PLANE APPROXIMATION**

▶ Stage \( t = T \)

\[
Q_T(x^k_{T-1}, \xi[T]) := \min_{B_T x^k_{T-1} + A_T x_T = b_T} c_T \top x_T \\
\text{s.t. } x_T \geq 0
\]

▶ At stages \( t = 2, \ldots, T - 1 \)

\[
Q_t(x^k_{t-1}, \xi[t]) := \begin{cases} 
\min_{x_t \geq 0, r_{t+1}} & c_t \top x_t + r_{t+1} \\
\text{s.t.} & B_t x^k_{t-1} + A_t x_t = b_t \\
& r_{t+1} \geq \alpha^j_{t+1} + \beta^j_{t+1} x_t & j = 1, \ldots, k
\end{cases}
\]

▶ Stage \( t = 1 \)

\[
\bar{z}^k := \begin{cases} 
\min_{x_1 \geq 0, r_2} & c_1 \top x_1 + r_2 \\
\text{s.t.} & A_1 x_1 = b_1 \\
& r_2 \geq \alpha^j_2 + \beta^j_2 x_1 & j = 1, \ldots, k
\end{cases}
\]
Computing cuts

- At stages $t = 2, \ldots, T - 1$

\[ Q_t(x_{t-1}^k, \xi_{[t]}) := \begin{cases} 
\min & c_t^\top x_t + r_{t+1} \\
\text{s.t.} & B_t x_{t-1}^k + A_t x_t = b_t \\
 & r_{t+1} \geq \alpha_{t+1}^j + \beta_{t+1}^j x_t \\
& j = 1, \ldots, k 
\end{cases} \]

- Cuts ($t = T$)

\[ \alpha_T^k := \mathbb{E}|\xi_{[T-1]}[b_T^\top \pi_T^k] \quad \text{and} \quad \beta_T^k := -\mathbb{E}|\xi_{[T-1]}[B_T^\top \pi_T^k] \]

- Cuts ($t = T - 1, \ldots, 2$)

\[ \alpha_t^k := \mathbb{E}|\xi_{[t-1]}[b_t^\top \pi_t^k + \sum_{j=1}^k \alpha_{t+1}^j \rho_t^j] \quad \text{and} \quad \beta_t^k := -\mathbb{E}|\xi_{[t-1]}[B_t^\top \pi_t^k] \]

\[ \tilde{Q}_{t+1}(x_t, \xi_{[t]}^i) = \sum_{j \in C_t} p^{(j)} \left[ Q_{t+1}(x_t, \xi_{[t+1]}^j) \right] \\
= \max_{j=1, \ldots, k} \left\{ \alpha_{t+1}^k + \beta_{t+1}^k x_t \right\} \]
**Algorithm - Nested Decomposition**

**stages** \( t = 2, \ldots, T - 1 \)

\[
Q_t(x_{t-1}^k, \xi_{[t]}) := \begin{cases} 
\min_{x_t \geq 0, r_{t+1}} & c_t \top x_t + r_{t+1} \\
\text{s.t.} & B_t x_{t-1}^k + A_t x_t = b_t \\
& r_{t+1} \geq \alpha_j^{t+1} + \beta_j^{t+1} x_t & j = 1, \ldots, k 
\end{cases} \quad (\pi_t, \rho_t)
\]

**Step 0: initialization.** Define \( k = 1 \) and add the constraint \( r_t = 0 \) in all LPs \( Q_t \), \( t = 2, \ldots, T - 1 \). Compute \( z^1 \) and let its solution be \( x_1^1 \).

**Step 1: forward.** For \( t = 2, \ldots, T \), solve the LP \( Q_t \) to obtain \( x_t^k := x_t^k(\xi_{[t]}) \). Define \( \bar{z}^k := \mathbb{E}[\sum_{t=1}^T c_t \top x_t^k] \).

**Step 2: backward.** Compute \( \alpha_T^k \) and \( \beta_T^k \). Set \( t = T \). Loop:

- While \( t > 2 \)
- \( t \leftarrow t - 1 \)
- solve the LP \( Q_t(x_{t-1}^k, \xi_{[t]}) \)
- Compute \( \alpha_t^k \) and \( \beta_t^k \)

Compute \( \bar{z}^k \) and let its solution be \( x_1^{k+1} \).

**Step 3: Stopping test.** If \( \bar{z}^k - \bar{z}^k \leq \epsilon \), stop. Otherwise set \( k \leftarrow k + 1 \) and go back to Step 1.
**Convergence analysis**

**Assumptions**

- The set of nodes $\Omega_t$ has finitely many elements, $t = 1, \ldots, T$

- The problem has recourse relatively complete (for simplicity, only)

- The feasible set, in each stage $t = 1, \ldots, T$, is compact

**Lemma**

$$\tilde{Q}_t^k(x_{t-1}, \xi_{[t-1]}) \leq Q_t(x_{t-1}, \xi_{[t-1]}) \quad \forall x_{t-1} \text{ and } \forall t = 2, \ldots, T$$

**Theorem**

The Nested Decomposition converges finitely to an optimal solution of the considered $T$-SLP.
If the T-SLP problem has block separable recourse, then a more efficient algorithm might be employed (this will, of course, depend on the application).

**Definition**
A T-SLP has block separable recourse if for all stage \( t = 1, \ldots, T \) and all \( \xi \), the decision vectors, \( x_t \), can be written as \( x_t = (w_t, y_t) \) where \( w_t \) represents aggregate level decisions and \( y_t \) represents detailed level (local) decisions. The constraints also follow these partitions:

- The stage \( t \) cost is \( c_t^\top x_t = c_t^w w_t + c_t^y y_t \)
- The matrices in the coupling constraint \( B_t x_{t-1} + A_t x_t = b_t \) are given by

\[
B_t = \begin{pmatrix} T_t & 0 \\ S_t & 0 \end{pmatrix} \quad A_t = \begin{pmatrix} W_t & 0 \\ 0 & D_t \end{pmatrix} \quad b_t = \begin{pmatrix} h_t \\ d_t \end{pmatrix}
\]
Block separable recourse

\[ x_t = (w_t, y_t) \quad c_t^\top x_t = c_t^w w_t + c_t^y y_t \]

In this manner

\[ B_t x_{t-1} + A_t x_t = b_t \quad \iff \quad \begin{cases} T_t w_{t-1} + W_t w_t = h_t \\ S_t w_{t-1} + D_t y_t = d_t \end{cases} \]

and the cost-to-go function

\[ Q_t(x_{t-1}, \xi_{[t]}) := \min_{B_t x_{t-1} + A_t x_t = b_t, \quad x_t \geq 0} c_t^\top x_t + Q_{t+1}(x_t, \xi_{[t]}) \]

becomes the sum of two quantities

\[ Q_t(x_{t-1}, \xi_{[t]}) = Q_t^w(w_{t-1}, \xi_{[t]}) + Q_t^y(w_{t-1}, \xi_{[t]}) \]

with

\[ Q_t^w(w_{t-1}, \xi_{[t]}) := \min_{T_t w_{t-1} + W_t w_t = h_t, \quad w_t \geq 0} c_t^w w_t + Q_{t+1}(w_t, \xi_{[t]}) \]

and

\[ Q_t^y(w_{t-1}, \xi_{[t]}) := \min_{S_t w_{t-1} + D_t y_t = d_t, \quad y_t \geq 0} c_t^y y_t \]
The great advantage of block separability is that we need not consider nesting among the detailed level decisions. In this way, the $w$ variables can all be pulled together into a first stage of aggregate level decisions.

$$\min_{x_1,w} c_1^\top x_1 + \mathbb{E}[c_2^w \top w_2 + \cdots + c_T^w \top w_T] + \mathbb{E} \left[ \sum_{t=2}^{T} Q^y_t(w_{t-1}, \xi[t]) \right]$$

s.t.

$$A_1 x_1 = b_1$$

$$T_t w_{t-1} + W_t w_t = h_t, \quad t = 2, \ldots, T \quad a.s.$$  

$$x_1, w \geq 0$$

with

$$Q^y_t(w_{t-1}, \xi[t]) := \min_{S_t w_{t-1} + D_t y_t = d_t} \min_{y_t \geq 0} c_t^y \top y_t$$
Block separable recourse

With finitely many scenarios

$$\min_{z \in Z} \, \bar{c}^\top z + Q(z)$$

with $Z$ a polyhedral set, $z$ containing all the node decisions $w_t^i$ and $x_1$ and

$$Q(z) = \sum_{t=2}^{T} \sum_{i \in \Omega_t} p^{(i)} Q^y_t (z, \xi^i)$$

$$Q^y_t (z, \xi^i) = \min_{S_t^i w_{t-1} + D_t^i y_t = d_t} \, c_t^{y,i} \top y_t$$

This is a convex programming problem and a subgradient of $Q$ is computable!
**Block separable recourse**

**Cutting-plane method**

**The problem**

\[
\min_{z \in \mathbb{Z}} f(z), \quad \text{with} \quad f(z) = \bar{c}^\top z + Q(z)
\]

**Oracle**

\[
z^\ell \xrightarrow{\text{oracle}} \begin{cases} \quad f(z^\ell) = \bar{c}^\top z^\ell + Q(z^\ell) \\ g^\ell \in \partial f(z^\ell) \end{cases}
\]

**Cutting-plane model**

\[
\tilde{f}_\ell(z) := \max_{j=1,\ldots,\ell} \{ f(z^j) + \langle g^j, x - x^j \rangle \}
\]

**Next iterate**

\[
z^{\ell+1} \in \arg \min_{z \in \mathbb{Z}} \tilde{f}_\ell(z)
\]
**Cutting-plane algorithm**

- **Step 0: initialization.** Choose $\text{tol} > 0$, $z^0 \in Z$ and call the oracle to compute $f(z^0)$ and $g^0 \in f(z^0)$. Set $f^{\text{up}}_0 = f(z^0)$ and $\ell = 0$.

- **Step 1: next iterate.** Compute
  \[
  z^{\ell+1} \in \arg \min_{z \in Z} \tilde{f}_\ell(z)
  \]
  and let $f^{\text{low}}_\ell = \tilde{f}_\ell(z^{\ell+1})$.

- **Step 2: stopping test.** Define $\Delta_\ell = f^{\text{up}}_\ell - f^{\text{low}}_\ell$. If $\Delta_\ell \leq \text{tol}$, stop.

- **Step 3: oracle call.** Compute $f(z^{\ell+1})$ and $g^{\ell+1} \in f(z^{\ell+1})$ and set
  \[
  f^{\text{up}}_{\ell+1} = \min\{f^{\text{up}}_\ell, f(z^{\ell+1})\}.
  \]

- **Step 4: loop.** Set $\ell = \ell + 1$ and go back to Step 1.
CONVERGENCE ANALYSIS

THEOREM

Let $\text{tol} > 0$ be given and suppose that $Z$ is compact. Then the cutting-plane algorithm determines $\Delta_{\ell} \leq \text{tol}$ in finitely many iterations. Furthermore, the point $\bar{z}$ yielding $f_{\ell}^{\text{up}} = f(\bar{z})$ is a $\text{tol}$-solution to the block separable T-SLP.
Theorem

Let \( \text{tol} > 0 \) be given and suppose that \( Z \) is compact. Then the cutting-plane algorithm determines \( \Delta_\ell \leq \text{tol} \) in finitely many iterations. Furthermore, the point \( \bar{z} \) yielding \( f^\text{up}_\ell = f(\bar{z}) \) is a \( \text{tol} \)-solution to the block separable T-SLP.

In fact, the result also holds if:

- \( \text{tol} = 0 \) (finite convergence)

- \( z \) is a mixed-integer variable (mixed-integer stochastic linear programming)!