Multistage stochastic programming problems
Multiplier method

Welington de Oliveira

BAS Lecture 18, May 10, 2016, IMPA
Set YouTube resolution to 480p for best viewing
**T-SLP**

\[
\begin{aligned}
\min & \quad \mathbb{E} \left[ c_1^T x_1 + c_2^T x_2(\xi_2) + \cdots + c_T^T x_T(\xi_T) \right] \\
\text{s.t.} & \quad x_1 \in \mathcal{X}_1 \\
& \quad x_t(\xi_t) \in \mathcal{X}_t(x_{t-1}(\xi_{t-1}), \xi_t), \quad t = 2, \ldots, T \\
& \quad x_t(\xi_t) \in \mathcal{F}_t, \quad t = 1, \ldots, T
\end{aligned}
\]

**Equivalent formulation**

\[
\begin{aligned}
\min & \quad \mathbb{E} \left[ f_1(x_1) + f_2(x_2(\xi_2), \xi_2) + \cdots + f_T(x_T(\xi_T), \xi_T) \right] \\
\text{s.t.} & \quad x_t(\xi_t) \in \mathcal{F}_t, \quad t = 1, \ldots, T
\end{aligned}
\]

with

\[
f_t(x_t(\xi_t), \xi_t) := \begin{cases} 
  c_t^T x_t(\xi_t) & \text{if } x_t(\xi_t) \in \mathcal{X}_t(x_{t-1}(\xi_{t-1}), \xi_t) \\
  +\infty & \text{otherwise}
\end{cases}
\]
Finitely many scenarios

\[
\begin{align*}
\min \quad & \sum_{k=1}^{K} p_k \left[ f_1^k(x_1^k) + f_2^k(x_2^k) + \cdots + f_T^k(x_T^k) \right] \\
\text{s.t.} \quad & x_t^k \in \mathcal{F}_t, \quad t = 1, \ldots, T, \quad k = 1, \ldots, K
\end{align*}
\]

Useful spaces

- $\mathcal{X}$ be the vector space of all sequences $(x_1^k, \ldots, x_T^k), \ k = 1, \ldots, K$ (such space has dimension $(n_1 + \ldots + n_T)K$)

- $\mathcal{L}$ be the subspace of $\mathcal{X}$ defined by the nonanticipativity constraints (i.e., $x \in \mathcal{L}$ means that $x-\mathcal{F}$ measurable)

Compact formulation

\[
\min_{x \in \mathcal{X}} f(x) \quad \text{s.t.} \quad x \in \mathcal{L}
\]

where $f(x) := \sum_{k=1}^{K} p_k \sum_{t=1}^{T} f_t^k(x_t^k)$. 

Svan 2016
Dual decomposition

\[ x^k := (x^k_1, x^k_2, \ldots, x^k_T) \quad \text{AND} \quad x := (x^1, x^2, \ldots, x^K) \]

\[ f(x) := \sum_{k=1}^{K} p_k \sum_{t=1}^{T} f^k_t(x^k_t) \]

The problem

\[ \min_{x \in \mathcal{X}} f(x) \quad \text{s.t.} \quad x_t \triangleleft \mathcal{F}_t, \ t = 1 \ldots, T \]

is thus equivalent to

\[ \min_{x \in \mathcal{X}} f(x) \quad \text{s.t.} \quad Gx = 0 \]

Lagrangian function

\[ L(x, u) := f(x) + u^\top Gx \]

\[ = \sum_{k=1}^{K} p_k \sum_{t=1}^{T} f^k_t(x^k_t) + \sum_{k=1}^{K} u^\top G^k x^k \]

\[ = \sum_{k=1}^{K} \left[ p_k \sum_{t=1}^{T} f^k_t(x^k_t) + u^\top G^k x^k \right] \]
Dual decomposition

\[ D(u) := \inf_{x \in \mathcal{X}} L(x, u) = \sum_{k=1}^{K} D^k(u), \quad D^k(u) := \inf_{x^k_1} p_k \sum_{t=1}^{T} f^k_t(x^k_t) + u^\top G^k x^k \]

Given our assumptions on the T-SLP, the each subproblem is a LP!

\[
D^k(u) := \begin{cases} 
\min_{x^k_t} & p_k \sum_{t=1}^{T} (c^k_t)^\top x^k_t + u^\top G^k x^k \\
\text{s.t.} & A_1 x_1 = b_1 \\
& B^k_t x_{t-1} + A^k_t x_t = b^k_t, \quad t = 2, \ldots, T.
\end{cases}
\]

Computing \(D(u)\) for each given \(u\) amounts to solving \(K\) LPs.

**Dual problem**

\[
\max_u D(u) \equiv \max_u \sum_{k=1}^{K} D^k(u)
\]

Given \(u^j\), let \(x^j \in \arg\min_{x \in \mathcal{X}} L(x, u^j)\). Then

\[ Gx^j \in \partial D(u^j). \]

**Exercise.** Define the Lagrangian function by \(L(x, u) = f(x) + \langle u, Gx \rangle\). Show how the resulting dual function \(D(u)\) can be decomposable, and specific how to compute a subgradient.

Inner product: \(\langle x, y \rangle := \sum_{k=1}^{K} \sum_{t=1}^{T} p_k \langle x^k_t, y^k_t \rangle\)
**Dual decomposition**

\[ D(u) := \inf_{x \in X} L(x, u) = \sum_{k=1}^{K} D^k(u), \quad D^k(u) := \inf_{x^k_t} \sum_{t=1}^{T} f^k_t(x^k_t) + u^\top G^k x^k \]

**Cutting-plane model**

\[ \tilde{D}_\ell(u) := \min_{j \leq \ell} \{ D(u^j) + \langle g^j, u - u^j \rangle \} \quad \text{with} \quad g^j := Gx^j. \]

Next iterate:

\[ u^{\ell+1} \in \arg \max_{z \in B(0, M)} \tilde{D}_\ell(u) \]
Primal recovering

\[ \tilde{v} := \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad Gx = 0 \]

After finitely many steps the cutting-plane algorithm finds a point \( \tilde{u} \) such that

\[ D(\tilde{u}) = \max_u D(u) \quad (= \max_{u \in B(0,M)} \tilde{D}_\ell(u)) \]

Since there is no optimality gap,

\[ \tilde{v} = D(\tilde{u}) \quad (= \max_{u \in B(0,M)} \tilde{D}_\ell(u)) \]

**Proposition**

Let \( \ell \) the iteration counter in which the optimal solution \( \tilde{u} \) is found by the algorithm. Suppose that \( \tilde{u} \in \text{int} B(0,M) \). Let \( \alpha_j \geq 0 \) Lagrange multiplies associate to the LP

\[ \max_{u \in B(0,M)} \tilde{D}_\ell(u) \equiv \left\{ \begin{array}{l} \max_{u,r} \quad r \\ \text{s.t.} \quad r \leq D(u^j) + \langle g^j, u - u^j \rangle, \quad \forall \ j \leq \ell \quad (\alpha_j) \end{array} \right. \]

Then \( \tilde{x} := \sum_{j=1}^{\ell} \alpha_j x^j \) is an optimal (primal) solution to the T-SLP.
**Dual decomposition**

- requires solving one LP per scenario (involving the whole horizon), at each iteration

- has a hard-to-evaluate dual function

- $u$ is in general a very large dimensional dual variable

- dual decomposition has slow convergence...

Alternatives to the (simple) dual decomposition:
**Dual decomposition**

- requires solving one LP per scenario (involving the whole horizon), at each iteration
- has a hard-to-evaluate dual function
- $u$ is in general a very large dimensional dual variable
- dual decomposition has slow convergence...

Alternatives to the (simple) dual decomposition:

- Regularized decomposition
- augmented Lagrangian method + nonlinear Jacobi method
- level decomposition with on-demand accuracy *(not studied yet for the multistage setting)*
- augmented Lagrangian method + (advanced) block-coordinate methods
The augmented Lagrangian method

Multiplier method

The high dimension of the dual vector $u$ suggests another approach: the augmented Lagrangian method.

Augmented Lagrangian

Let $\rho > 0$ be a penalty coefficient.

$$L_\rho(x, u) := L(x, u) + \frac{\rho}{2} \|Gx\|^2$$

$$= f(x) + u^\top Gx + \frac{\rho}{2} \|Gx\|^2$$

The Multiplier Method applied to the T-SLP works as follows.
Multiplier method

\[ L_\rho(x, u) = f(x) + u^\top Gx + \frac{\rho}{2} \|Gx\|^2 \]

- **Step 0: initialization.** Choose \( \text{tol} > 0, \ u^0 \) and set \( \ell = 0 \)

- **Step 1: primal update.** Compute

  \[ x^\ell = \arg \min_{x \in X} L_\rho(x, u^\ell) \]

- **Step 2: stopping test.** If \( \|Gx^\ell\| \leq \text{tol} \), stop.

- **Step 3: dual iterate.** Set \( u^{\ell+1} = u^\ell + \rho Gx^\ell \)

- **Step 4: loop.** Set \( \ell = \ell + 1 \) and go back to Step 1.

**Theorem**

Assume that the T-SLP dual problem \( \max_u D(u) \) has an optimal solution.
Then the sequence \( \{u^\ell\} \) generated by the Multiplier Method after finitely many steps finds an optimal solution.
**Multiplier method**

\[ L_\rho(x, u) = f(x) + u^\top Gx + \frac{\rho}{2} \|Gx\|^2 \]

▶ **Step 0: initialization.** Choose tol > 0, \( u^0 \) and set \( \ell = 0 \)

▶ **Step 1: primal update.** Compute

\[ x^\ell = \arg \min_{x \in X} L_\rho(x, u^\ell) \]

▶ **Step 2: stopping test.** If \( \|Gx^\ell\| \leq \text{tol} \), stop.

▶ **Step 3: dual iterate.** Set \( u^{\ell+1} = u^\ell + \rho Gx^\ell \)

▶ **Step 4: loop.** Set \( \ell = \ell + 1 \) and go back to Step 1.

**Theorem**

Assume that the T-SLP dual problem \( \max_u D(u) \) has an optimal solution. Then the sequence \( \{u^\ell\} \) generated by the Multiplier Method after finitely many steps finds an optimal solution.

We’ll rely on the Proximal Method theory to prove this theorem.
**Proximal point method**

Let $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and assume that $\text{dom} \varphi \neq \emptyset$. Consider the convex problem

$$\min_{y \in \mathbb{R}^n} \varphi(y). \quad (\star)$$

**Proximal Point Method**

The *proximal point method* defines iterates by the rule

$$y^{\ell+1} := \arg \min \varphi(y) + \frac{\rho}{2} \|y - y^\ell\|^2, \quad \ell = 0, 1, \ldots$$

**Theorem**

Suppose that the convex problem $(\star)$ has an optimal solution. Then the sequence $\{y^\ell\}$ generated by the proximal point method is convergent to an optimal solution of $(\star)$. Furthermore, if $\varphi$ is a polyhedral function then the convergence is finite.
Proximal point method

Let \( \varphi : \mathbb{R}^n \to \mathbb{R} \) be a convex function and assume that \( \text{dom} \varphi \neq \emptyset \). Consider the convex problem

\[
\min_{y \in \mathbb{R}^n} \varphi(y). \quad (\star)
\]

Proximal Point Method

The proximal point method defines iterates by the rule

\[
y^{\ell+1} := \arg \min \varphi(y) + \frac{\rho}{2} \|y - y^\ell\|^2, \quad \ell = 0, 1, \ldots
\]

Theorem

Suppose that the convex problem \((\star)\) has an optimal solution. Then the sequence \( \{y^\ell\} \) generated by the proximal point method is convergent to an optimal solution of \((\star)\).

Furthermore, if \( \varphi \) is a polyhedral function then the convergence is finite.

Coming back to the multiplier method’s analysis ...
Multiplier method

\[ L(x, u) = f(x) + u^\top Gx + \frac{\rho}{2} \|Gx\|^2 \]

Multiplier method
For \( \ell = 0, 1, \ldots \)

(A) Given \( u^\ell \), find \( x^\ell = \arg\min_{x \in X} L_\rho(x, u^\ell) \)

(B) set \( u^{\ell+1} = u^\ell + \rho Gx^\ell \)

Theorem
Assume that the T-SLP dual problem \( \max_u D(u) \) has an optimal solution. Then the sequence \( \{u^\ell\} \) generated by the Multiplier Method after finitely many steps finds an optimal solution.
**Multiplier method**

\[ L(x, u) = f(x) + u^\top Gx + \frac{\rho}{2} \|Gx\|^2 \]

**Multiplier method**

For \( \ell = 0, 1, \ldots \)

(A) Given \( u^\ell \), find \( x^\ell = \arg\min_{x \in \mathcal{X}} L_\rho(x, u^\ell) \)

(B) set \( u^{\ell+1} = u^\ell + \rho Gx^\ell \)

- The method’s main advantage is its simplicity
- Requires less iterations than the dual decomposition!

- Step (A) is difficult... This is the method’s main disadvantage.

\[
\min_{x \in \mathcal{X}} L_\rho(x, u^\ell) \equiv \min_{x \in \mathcal{X}} \sum_{k=1}^{K} \left[ p_k \sum_{t=1}^{T} f_t^k(x_t^k) + (u^\ell)^\top G^k x^k + \frac{\rho}{2} \sum_{k=1}^{K} \|G^k x_t^k\|^2 \right]
\]

\[
f_t(x_t(\xi_{[t]}), \xi_t) := \begin{cases} 
  c_t^\top x_t(\xi_{[t]}) & \text{if } x_t(\xi_{[t]}) \in \mathcal{X}_t(x_{t-1}(\xi_{[t-1]}), \xi_t) \\
  +\infty & \text{otherwise}
\end{cases}
\]

Subproblem is not decomposable...
The separable approximation

One possibility to overcome the nonseparability of the augmented Lagrangian is to apply an iterative nonlinear Jacobi method to

\[
\min_{x \in X} L_\rho(x, u^\ell) \equiv \min_{x \in X} \sum_{k=1}^K \left[ p_k \sum_{t=1}^T f^k_t(x^k_t) + (u^\ell)^\top G^k x^k + \frac{\rho}{2} \left\| \sum_{k=1}^K G^k x^k \right\|^2 \right]
\]

Let’s keep \( u^\ell \) fixed, and focus on the solution of the above subproblem. Let \( \tilde{x}^j \) be approximate solution at iteration \( j \)

\[
\tilde{x}^k := (\tilde{x}^k_1, \tilde{x}^k_2, \ldots, \tilde{x}^k_T) \quad \text{and} \quad \tilde{x}^j := (\tilde{x}^{1,j}, \tilde{x}^{2,j}, \ldots, \tilde{x}^{K,j})
\]

The approach solves, for each scenario \( k \), simplified problems:

\[
x^{k,j+1} := \arg \min_{x^k} p_k \sum_{t=1}^T f^k_t(x^k_t) + (u^\ell)^\top G^k x^k + \frac{\rho}{2} \left\| G^k x^k + \sum_{s \neq k} G^s \tilde{x}^{s,j} \right\|^2
\]

When \( x^{k,j+1} \) is computed for all scenario \( k = 1, \ldots, K \), the new iterate \( \tilde{x}^{j+1} \) is given as

\[
\tilde{x}^{k,j+1} := (1 - \tau)\tilde{x}^{k,j} + \tau x^{k,j+1} \quad \forall k = 1, \ldots, K, \quad \text{and given} \ \tau \in (0, 1).
\]
The separable approximation

\[ x_{k+1} := \arg \min p_k \sum_{t=1}^{T} f_t^k (x_t^k) + (u^\ell)^\top G^k x^k + \frac{\rho}{2} \|G^k x^k + \sum_{s \neq k} G_s \tilde{x}_{s,j} \|^2 \]  

The quadratic term of the augmented Lagrangian has then the form

\[ \|Gx\|^2 = \sum_{k=1}^{K} \|G^k x^k\|^2 = \sum_{t=1}^{T} \sum_{i \in \Omega_t} \sum_{k \in \mathcal{S}(i)} \|x_t^k - x_{t+1}^k\|^2 \]

- \( \Omega_t \) is the set of all nodes at stage \( t \)
- \( \mathcal{S}(i) \) is the set of all scenarios passing through node \( i \)

The minimization in (★) involves at most two simple quadratic terms for each subvector \( x_t^k, t = 1, \ldots, T \), relating it to the reference values at its neighbors:

\[ \tilde{x}_{t-1}^k \quad \text{and} \quad \tilde{x}_{t+1}^k. \]

This not only makes the subproblems easier to manipulate and solve, but it has a positive impact on the speed of convergence.
The separable approximation

\[
x^{k,j+1} := \arg \min_{p_k} \sum_{t=1}^{T} f_t^k(x_t^k) + (u^\ell)^\top G^k x^k + \frac{\rho}{2} \|G^k x^k + \sum_{s \neq k} G^s \tilde{x}^s,j\|_2^2
\]

Theorem
Assume that the feasible set of each scenario-based subproblem is bounded. Then

- For all \(k = 1, \ldots, K\) we have \(\lim_{j \to \infty} G^k (x^{k,j} - \tilde{x}^{k,j}) = 0\);
- Every accumulation point of the sequence \(\{\tilde{x}^j\}\) is a solution to

\[
\min_{x \in \mathcal{X}} L_\rho(x, u^\ell).
\]