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

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

- $\mathcal{X}_1 := \{x_1 \in \Re^{n_1} : A_1 x_1 = b_1,\ x_1 \geq 0\}$
- $\mathcal{X}_t(x_{t-1}, \xi_t) := \{x_t \in \Re^{n_t} : B_t x_{t-1} + A_t x_t = b_t,\ x_t \geq 0\}$

Equivalent formulation

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

with

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

Finitely many scenarios

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

Useful spaces

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

Compact formulation

$$\min_{\mathbf{x} \in \mathfrak{X}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{L}$$

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

Dual decomposition

$$x^k := (x_1^k, x_2^k, \ldots, x_T^k) \quad \text{AND} \quad \mathbf{x} := (x^1, x^2, \ldots, x^K)$$

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

The problem

$$\min_{\mathbf{x} \in \mathfrak{X}} f(\mathbf{x}) \quad \text{s.t.} \quad x_t \lhd \mathcal{F}_t,\ t = 1 \ldots, T$$

is thus equivalent to

$$\min_{\mathbf{x} \in \mathfrak{X}} f(\mathbf{x}) \quad \text{s.t.} \quad G\mathbf{x} = 0$$

Lagrangian function

$$\begin{aligned} L(\mathbf{x}, u) &:= f(\mathbf{x}) + u^\top G\mathbf{x} \\ &= \textstyle\sum_{k=1}^{K} p_k \sum_{t=1}^{T} f_t^k(x_t^k) + \sum_{k=1}^{K} u^\top G^k x^k \\ &= \textstyle\sum_{k=1}^{K} \left[p_k \sum_{t=1}^{T} f_t^k(x_t^k) + u^\top G^k x^k \right] \end{aligned}$$

Dual decomposition

$$D(u) := \inf_{\mathbf{x}\in\mathfrak{X}} L(\mathbf{x}, u) = \sum_{k=1}^{K} D^k(u), \quad D^k(u) := \inf_{x_t^k} p_k \sum_{t=1}^{T} f_t^k(x_t^k) + u^\top G^k x^k$$

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

$$D^k(u) := \begin{cases} \min_{x_t^k} & p_k \sum_{t=1}^{T} (c_t^k)^\top x_t^k + u^\top G^k x^k \\ \text{s.t.} & A_1 x_1 = b_1 \\ & B_t^k x_{t-1} + A_t^k x_t = b_t^k, \quad t = 2, \ldots, T\,. \end{cases}$$

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

Dual problem

$$\max_u \; D(u) \quad \equiv \quad \max_u \sum_{k=1}^{K} D^k(u)$$

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

$$G\mathbf{x}^j \in \partial D(u^j)\,.$$

Exercise. Define the Lagrangian function by $L(\mathbf{x}, u) = f(\mathbf{x}) + \langle\langle u, G\mathbf{x}\rangle\rangle$. Show how the resulting dual function $D(u)$ can be decomposable, and specific how to compute a subgradient.
Inner product: $\langle\langle \mathbf{x}, \mathbf{y}\rangle\rangle := \sum_{k=1}^{k} \sum_{t=1}^{T} p_k \langle x_t^k, y_t^k\rangle$

Dual decomposition

$$D(u) := \inf_{\mathbf{x}\in\mathfrak{X}} L(\mathbf{x}, u) = \sum_{k=1}^{K} D^k(u), \quad D^k(u) := \inf_{x_t^k} p_k \sum_{t=1}^{T} f_t^k(x_t^k) + u^\top G^k x^k$$

Cutting-plane model

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

Next iterate:

$$u^{\ell+1} \in \arg\max_{z\in B(0,M)} \check{D}_\ell(u)$$

Primal recovering

$$\bar{v} := \min_{\mathbf{x}\in\mathfrak{X}} f(\mathbf{x}) \quad \text{s.t.} \quad G\mathbf{x} = 0$$

After finitely many steps the cutting-plane algorithm finds a point $\bar{u}$ such that

$$D(\bar{u}) = \max_u D(u) \qquad (= \max_{u\in B(0,M)} \check{D}_\ell(u))$$

Since there is no optimality gap,

$$\bar{v} = D(\bar{u}) \qquad (= \max_{u\in B(0,M)} \check{D}_\ell(u))$$

Propositon

Let ℓ the iteration counter in which the optimal solution $\bar{u}$ is found by the algorithm. Suppose that $\bar{u} \in \mathtt{int}B(0,M)$. Let $\alpha_j \geq 0$ Lagrange multipliers associate to the LP

$$\max_{u\in B(0,M)} \check{D}_\ell(u) \quad \equiv \quad \begin{cases} \max_{u,r} & r \\ s.t. & r \leq D(u^j) + \langle g^j, u - u^j\rangle, \quad \forall\, j \leq \ell \quad (\alpha_j) \end{cases}$$

Then $\check{\mathbf{x}} := \sum_{j=1}^{\ell} \alpha_j \mathbf{x}^j$ *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 high dimension of the dual vector u suggests another approach: the augmented Lagrangian method.

Augmented Lagrangian

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

$$\begin{aligned} L_\rho(\mathbf{x}, u) &:= L(\mathbf{x}, u) + \frac{\rho}{2} \|G\mathbf{x}\|^2 \\ &= f(\mathbf{x}) + u^\top G\mathbf{x} + \frac{\rho}{2} \|G\mathbf{x}\|^2 \end{aligned}$$

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

Multiplier method

$$L_\rho(\mathbf{x}, u) = f(\mathbf{x}) + u^\top G\mathbf{x} + \frac{\rho}{2}\|G\mathbf{x}\|^2$$

- **Step 0: initialization**. Choose tol > 0, u^0 and set $\ell = 0$
- **Step 1: primal update**. Compute
$$\mathbf{x}^\ell = \arg\min_{\mathbf{x}\in\mathfrak{X}} L_\rho(\mathbf{x}, u^\ell)$$
- **Step 2: stopping test**. If $\|G\mathbf{x}^\ell\| \leq$ tol, stop.
- **Step 3: dual iterate**. Set $u^{\ell+1} = u^\ell + \rho G\mathbf{x}^\ell$
- **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

$$L_\rho(\mathbf{x}, u) = f(\mathbf{x}) + u^\top G\mathbf{x} + \frac{\rho}{2}\|G\mathbf{x}\|^2$$

- **Step 0: initialization**. Choose tol > 0, u^0 and set $\ell = 0$
- **Step 1: primal update**. Compute
$$\mathbf{x}^\ell = \arg\min_{\mathbf{x}\in\mathfrak{X}} L_\rho(\mathbf{x}, u^\ell)$$
- **Step 2: stopping test**. If $\|G\mathbf{x}^\ell\| \leq$ tol, stop.
- **Step 3: dual iterate**. Set $u^{\ell+1} = u^\ell + \rho G\mathbf{x}^\ell$
- **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 $\varphi : \Re^n \to \bar{\Re}$ be a convex function and assume that $\mathtt{dom}\,\varphi \neq \emptyset$. Consider the convex problem

$$\min_{y \in \Re^n} \varphi(y)\,. \qquad (\bigstar)$$

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 $(\bigstar)$ *has an optimal solution. Then the sequence* $\{y^{\ell}\}$ *generated by the proximal point method is convergent to an optimal solution of* $(\bigstar)$*.*
Furthermore, if φ *is a polyhedral function then the convergence is finite.*

Proximal point method

Let $\varphi : \Re^n \to \bar{\Re}$ be a convex function and assume that $\mathtt{dom}\,\varphi \neq \emptyset$. Consider the convex problem

$$\min_{y \in \Re^n} \varphi(y)\,. \qquad (\bigstar)$$

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 $(\bigstar)$ *has an optimal solution. Then the sequence* $\{y^{\ell}\}$ *generated by the proximal point method is convergent to an optimal solution of* $(\bigstar)$*.*
Furthermore, if φ *is a polyhedral function then the convergence is finite.*

Coming back to the multiplier method's analysis ...

Multiplier method

$$L(\mathbf{x}, u) = f(\mathbf{x}) + u^\top G\mathbf{x} + \frac{\rho}{2}\|G\mathbf{x}\|^2$$

Multiplier method

For $\ell = 0, 1, \ldots$

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

(B) set $u^{\ell+1} = u^\ell + \rho G\mathbf{x}^\ell$

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

$$L(\mathbf{x}, u) = f(\mathbf{x}) + u^\top G\mathbf{x} + \frac{\rho}{2}\|G\mathbf{x}\|^2$$

Multiplier method

For $\ell = 0, 1, \ldots$

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

(B) set $u^{\ell+1} = u^\ell + \rho G\mathbf{x}^\ell$

- 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.

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

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

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

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

Let's keep u^ℓ fixed, and focus on the solution of the above subproblem. Let $\tilde{\mathbf{x}}^j$ be approximate solution at iteration j

$$\tilde{x}^k := (\tilde{x}_1^k, \tilde{x}_2^k, \ldots, \tilde{x}_T^k) \quad \text{and} \quad \tilde{\mathbf{x}}^j := (\tilde{x}^{1,j}, \tilde{x}^{2,j}, \ldots, \tilde{x}^{K,j})$$

The approach solves, for each scenario k, simplified problems:

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

When $x^{k,j+1}$ is computed for all scenario $k = 1, \ldots, K$, the new iterate $\tilde{\mathbf{x}}^{j+1}$ 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

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

The quadratic term of the augmented Lagrangian has then the form

$$\|G\mathbf{x}\|^2 = \sum_{k=1}^{K} \|G^k x^k\|^2 = \sum_{t=1}^{T} \sum_{\iota \in \Omega_t} \sum_{k \in \mathcal{S}^{(\iota)}} \|x_t^k - x_t^{k+1}\|^2$$

- Ω_t is the set of all nodes at stage t
- $\mathcal{S}^{(\iota)}$ is the set of all scenarios passing through node ι

The minimization in $(\bigstar)$ 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:

$$\tilde{x}_t^{k-1} \quad \text{and} \quad \tilde{x}_t^{k+1}\,.$$

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

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

Theorem

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

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

$$\min_{\mathbf{x}\in\mathfrak{X}} L_\rho(\mathbf{x}, u^\ell)\,.$$

The proof is given in Theorem 24 of the book:
Stochastic Programming. Handbook in Operations Research and Management Science, edited by Ruszczynski, A. and Shapiro, A., Elsevier, 2003.