Multistage stochastic programming problems
Dealing with nonantecipativity constraints

Welington de Oliveira

BAS Lecture 17, May 5, 2016, IMPA

Set YouTube resolution to 480p

for best viewing

T-SLP

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

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

Implementable policy

$$x_t : \Re^{d_1} \times \cdots \Re^{d_t} \to \Re^{n_t}$$

An implementable policy is said to be feasible if

$$x_t(\xi_{[t]}) \in \mathcal{X}_t(x_{t-1}(\xi_{[t-1]}), \xi_t), \quad t = 2, 3, \dots, T \quad w.p.\ 1.$$

T-SLP

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

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

Implementable policy

$$x_t : \Re^{d_1} \times \cdots \Re^{d_t} \to \Re^{n_t}$$

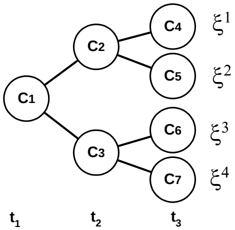
$x_t(\xi_t) \lhd \mathcal{F}_t$ means that the function $x_t(\xi_{[t]})$ is measurable with respect to the σ-algebra $\mathcal{F}_t$.

Example

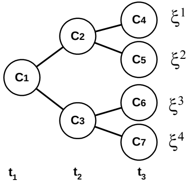
Consider the following 3-SLP:

$$\begin{cases} \min_{x,r} & \mathbb{E}\left[\sum_{t=1}^{3} \xi_t x_t\right] \\ s.t. & (x_t, r_t) \in \Re_+^2 \\ & r_t - r_{t-1} = x_t, \quad t = 2,3 \\ & r_1 = 0, r_3 = 5 \\ & (x_t, r_t) \lhd \mathcal{F}_t \end{cases}$$

with $\Xi := \{\xi^1, \xi^2, \xi^3, \xi^4\}$, and equiprobable scenarios ξ^i, $i = 1, \ldots, 4$ $(p_i = 1/4)$



Example



The nonantecipativity constraints can be made explicit:

$$\begin{cases} \min_{x,r} & \left\{ \begin{array}{l} (c_1x_1^1 + c_2x_2^1 + c_4x_3^1)/4 + (c_1x_1^2 + c_2x_2^2 + c_5x_3^2)/4 \quad + \\ \\ (c_1x_1^3 + c_3x_2^3 + c_6x_3^3)/4 + (c_1x_1^4 + c_3x_2^4 + c_7x_3^4)/4 \end{array} \right\} \\ \\ s.t. & (x_t^i, r_t^i) \in \Re_+^2, t = 1,2,3 \text{ and } i = 1,\dots,4 \\ & r_t^i - r_{t-1}^i = x_t^i, \; t = 2,3 \text{ and } i = 1,\dots,4 \\ & r_1^i = 0, r_3^i = 5, \; i = 1,\dots,4. \\ \\ & x_1^i = x_1^j, \; i,j = 1,\dots,4, \; x_2^1 = x_2^2, \; x_2^3 = x_2^4 \end{cases}$$

$(x_t^i := x(\xi_t^i))$.

T-SLP

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

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

Equivalent formulation

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

with

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

Finitely many scenarios

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

- Let's assume that the stochastic process is represented by a scenario tree composed of K scenarios ξ^k with associate probability p_k.
- We use the shorthand $f_t^k(x_t^k)$ for $f_t(x_t(\xi_{[t]}^k),\xi_t^k)$

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

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 is $\mathcal{F}$- measurable)
- Inner product: $\langle\langle \mathbf{x}, \mathbf{y} \rangle\rangle := \sum_{k=1}^{K} \sum_{t=1}^{T} p_k \langle x_t^k, y_t^k \rangle$

Compact formulation

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

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

Optimality conditions

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

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

Theorem

A policy $\bar{\mathbf{x}} \in \mathcal{L}$ is an optimal solution of the above problem iff there exists a multiplier $\bar{\lambda} \in \mathcal{L}^{\perp}$ such that

$$\bar{\mathbf{x}} \in \arg\min_{\mathbf{x} \in \mathfrak{X}} L(\mathbf{x}, \bar{\lambda}), \quad \textit{with} \quad L(\mathbf{x}, \lambda) := f(\mathbf{x}) + \langle\langle \lambda, \mathbf{x} \rangle\rangle$$

Dual problem

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

- Primal problem $\min_{\mathbf{x}\in\mathfrak{X}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{L}$
- Lagrangian function: $L(\mathbf{x}, \lambda) := f(\mathbf{x}) + \langle\langle \lambda, \mathbf{x} \rangle\rangle$
- Dual function $D(\lambda) = \inf_{x\in\mathfrak{X}} \ L(\mathbf{x}, \lambda)$

Dual problem

$$\max_{\lambda\in\mathcal{L}^\perp} D(\lambda)$$

Theorem

The primal and optimal values are equal unless both problems are infeasible. If their (common) optimal value is finite, then both problems have optimal solutions.

Dual problem

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

- Primal problem $\min_{\mathbf{x}\in\mathfrak{X}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{L}$
- Lagrangian function: $L(\mathbf{x}, \lambda) := f(\mathbf{x}) + \langle\langle \lambda, \mathbf{x} \rangle\rangle$
- Dual function $D(\lambda) = \inf_{x\in\mathfrak{X}} \; L(\mathbf{x}, \lambda)$

Dual problem

$$\max_{\lambda\in\mathcal{L}^\perp} D(\lambda)$$

Theorem

The primal and optimal values are equal unless both problems are infeasible. If their (common) optimal value is finite, then both problems have optimal solutions.

Proof: The result follows from the linear programming theory (remember that f_t are polyhedral!).

Dual problem

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

- Primal problem $\min_{\mathbf{x}\in\mathfrak{X}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{L}$
- Lagrangian function: $L(\mathbf{x}, \lambda) := f(\mathbf{x}) + \langle\!\langle \lambda, \mathbf{x} \rangle\!\rangle$
- Dual function $D(\lambda) = \inf_{x\in\mathfrak{X}} \; L(\mathbf{x}, \lambda)$

Dual problem

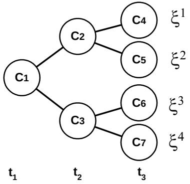
$$\max_{\lambda\in\mathcal{L}^\perp} D(\lambda) \quad \equiv \quad \max_{\lambda\in\mathfrak{X}} D(\lambda) \quad \text{s.t.} \quad \mathbb{E}_{|\xi_{[t]}}[\lambda_t] = 0 \quad t = 1, \ldots, T$$

Theorem

The primal and optimal values are equal unless both problems are infeasible. If their (common) optimal value is finite, then both problems have optimal solutions.

Proof: The result follows from the linear programming theory (remember that f_t are polyhedral!).

Dual decomposition



$$x_t \lhd \mathcal{F}_t \quad \equiv \quad \left\{ \begin{array}{l} x_1^1 - x_1^2 = 0,\ x_1^2 - x_1^3 = 0,\ x_1^3 - x_1^4 = 0 \\ x_2^1 - x_2^2 = 0 \ \text{ and } \ x_2^3 - x_2^4 = 0\,. \end{array} \right\}$$

$$x_t \lhd \mathcal{F}_t \quad \equiv \quad G\mathbf{x} = 0$$

$$G = \left[\begin{array}{cc|cc|cc|cc} I & & -I & & & & & \\ & & I & & -I & & & \\ & & & & I & & -I & \\ & I & & -I & & & & \\ & & & & & I & & -I \end{array} \right],$$

- I is the identity matrix of appropriate dimensions
- G is composed of K blocks $G = [G^1\, G^2\, \ldots\, G^k]$

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

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

Dual decomposition

Dual function

$$D(u) := \inf_{\mathbf{x}\in\mathfrak{X}} L(x,u) = \sum_{k=1}^{K} D^k(u)$$

where

$$\begin{aligned} D^k(u) &:= \inf_{x_t^k} p_k \textstyle\sum_{t=1}^T f_t^k(x_t^k) + u^\top G^k \mathbf{x} \\ &= -p_k \sup_{x_t^k} \{-(\tfrac{1}{p_k} u^\top G^k)\mathbf{x} - \textstyle\sum_{t=1}^T f_t^k(x_t^k)\} \\ &= -p_k (f^k)^*(-\tfrac{1}{p_k} u^\top G^k) \end{aligned}$$

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

- If $\bar{x}^k$ is a solution of the minimization problem, then $G^k\bar{x}^k \in \partial D^k(u)$ and thus

$$G\bar{\mathbf{x}} \in \partial D(u)$$

Dual decomposition

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

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

$$D^k(u) := \begin{cases} \min_{x_t^k} & p_k \sum_{t=1}^{T} (c_t^k)^\top x_t^k + u^\top G^k \mathbf{x} \\ \text{s.t.} & A_1 x_1 = b_1 \\ & B_t^k x_{t-1} + A_t^k x_t = b_t^k, \quad t = 2, \ldots, T \\ & x_t^k \geq 0\,. \end{cases}$$

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

Dual problem

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

Dual decomposition algorithm

- **Step 0: initialization**. Choose tol > 0, $M > 0$, $u^0 \in B(0, M)$ and call the oracle to compute $D(z^0)$ and $g^0 \in D(u^0)$. Set $f_0^{\text{low}} = D(u^0)$ and $\ell = 0$

- **Step 1: next iterate**. Compute

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

  and let $f_\ell^{\text{up}} = \check{D}_\ell(u^{\ell+1})$.

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

- **Step 3: oracle call**. Compute $D(u^{\ell+1})$ and $g^{\ell+1} \in D(u^{\ell+1})$ and set $f_{\ell+1}^{\text{low}} = \max\{f_\ell^{\text{low}}, D(u^{\ell+1})\}$.

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

Cutting-plane model

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

Theorem

Let tol ≥ 0 *be given and suppose that* M *is large enough such that*

$$B(0, M) \cap \arg\max D(u) \neq \emptyset\,.$$

Furthermore, assume that $B(0, M) \in \mathtt{dom}D$*. Then the Dual Decomposition Algorithm determines* $\Delta_k \leq \mathrm{tol}$ *in finitely many iterations. Furthermore, the point* $\bar{u}$ *yielding* $f_\ell^{\mathrm{low}} = D(\bar{u})$ *is a* tol*-solution to the problem.*

Proof: The a algorithm is a mere cutting-plane applied to a convex and polyhedral program. The result thus follows from the analysis of the cutting-plane method.

Convergence analysis

Theorem

Let tol ≥ 0 *be given and suppose that* M *is large enough such that*

$$B(0, M) \cap \arg\max D(u) \neq \emptyset\,.$$

Furthermore, assume that $B(0, M) \in \mathtt{dom} D$*. Then the Dual Decomposition Algorithm determines* $\Delta_k \leq \mathrm{tol}$ *in finitely many iterations. Furthermore, the point* $\bar{u}$ *yielding* $f_\ell^{\mathrm{low}} = D(\bar{u})$ *is a* tol*-solution to the problem.*

Proof: The a algorithm is a mere cutting-plane applied to a convex and polyhedral program. The result thus follows from the analysis of the cutting-plane method.

But $\bar{u}$ is a dual solution... We need a primal solution!

Primal recovering

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

Consider tol $= 0$ in the algorithm. After finitely many steps the 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.