Two-stage Stochastic Programming with Linearly Bi-parameterized Quadratic Recourse

Junyi Liu, Ying Cui, Jong-Shi Pang, and Suvrajeet Sen

Department of Industrial and Systems Engineering

University of Southern California

presented at

Mathematics Workshop on Electricity systems of the future

Cambridge University

Tuesday March 19, 2019, 15:45–16:30 AM

Stochastic Programming:

A mathematical tool for smart planning under uncertainty.

Stochastic Programming:
A mathematical tool for smart planning under uncertainty.

This workshop: Electricity systems of the future.
This talk: Stochastic programming of the future.

Stochastic Programming:
A mathematical tool for smart planning under uncertainty.

This workshop: Electricity systems of the future.
This talk: Stochastic programming of the future.

From linearity and convexity
to coupled nonconvexity and nondifferentiability

Stochastic Programming:
A mathematical tool for smart planning under uncertainty.

This workshop: Electricity systems of the future.
This talk: Stochastic programming of the future.

From linearity and convexity
to coupled nonconvexity and nondifferentiability

From globally optimal solutions
to sharp stationary solutions with significance.

Stochastic Programming:
A mathematical tool for smart planning under uncertainty.

This workshop: Electricity systems of the future.
This talk: Stochastic programming of the future.

From linearity and convexity
to coupled nonconvexity and nondifferentiability

From globally optimal solutions
to sharp stationary solutions with significance.

From smooth calculus
to nonsmooth analysis

Two-stage Stochastic Programming

Features:

- the decision must be made prior to the observation of the uncertainty
- allow for the option to adjust to the realization of the uncertainty

Two-stage Stochastic Programming

Features:

- the decision must be made prior to the observation of the uncertainty
- allow for the option to adjust to the realization of the uncertainty

Motivation: With exceptions, the bulk of computational two-stage SP is restricted to the setting of linear programming based recourse with a linearly parameterized right-hand side by the first-stage variable; the objective is not parameterized. **WHY?**

Focus: two-stage SP where both the linear cost vector and the right-hand constraint vector in the second-stage problem are simultaneously parameterized by the first-stage decision and the uncertainty and with a deterministic convex quadratic cost.

Two-stage SP with bi-parameterized recourse

$$\underset{x \in X}{\text{minimize}} \quad c(x) + \mathbf{E}_{\widetilde{\omega}}[\, h(x, \widetilde{\omega}) \,]$$

where the **bi-parameterized** recourse function

$$h(x, \omega) \; = \underset{y}{\text{minimum}} \quad d(x, \omega)^\top y + \tfrac{1}{2}\, y^\top Q y$$
$$\text{subject to} \quad Dy \geq b(x, \omega)$$

Motivation of the bi-parameterized recourse

joint pricing and shipment planning problem

Motivation of the bi-parameterized recourse

joint pricing and shipment planning problem

- **decision variables:**
  - p: price
  - x_i: production (generation) at plant i
  - y_i: (residual) production at plant i
  - $z_{i,j}$: shipment from plant i to demand point j for sale
- **random variable:** demand ω

Motivation of the bi-parameterized recourse

joint pricing and shipment planning problem

- **decision variables:**
  - p: price
  - x_i: production (generation) at plant i
  - y_i: (residual) production at plant i
  - $z_{i,j}$: shipment from plant i to demand point j for sale
- **random variable:** demand ω
- **the decision-making problem:**

$$\underset{x,\,p \geq 0}{\text{minimize}} \quad c_1 \sum_i x_i + \mathbf{E}_{\widetilde{\omega}\,|\,p}\left[\, h(x,p,\widetilde{\omega}\,)\,\right]$$

$$\text{where} \quad h(x,p,\omega\,) = \underset{y,\,z \geq 0}{\text{minimum}} \quad c_2 \sum_i y_i + \sum_{i,j} (s_{i,j} - p) z_{i,j}$$

$$\text{subject to} \quad \sum_i z_{i,j} \leq \omega_j, \quad \sum_j z_{i,j} \leq x_i + y_i$$

Another motivation

economic dispatch in renewable power systems planning

Another motivation

economic dispatch in renewable power systems planning

- **decision variables:**
  - power generation from slow-response generators x
  - power generation from fast-response generators y

Another motivation

economic dispatch in renewable power systems planning

- **decision variables:**
  - power generation from slow-response generators x
  - power generation from fast-response generators y
- **random variable:**
  - renewable power generation ω_1 and electricity demand ω_2

Another motivation

economic dispatch in renewable power systems planning

- **decision variables:**
  - power generation from slow-response generators x
  - power generation from fast-response generators y
- **random variable:**
  - renewable power generation ω_1 and electricity demand ω_2
- **parameters:**
  - cost functions of slow- and fast-response generators c_1 and c_2, respectively.

Another motivation

economic dispatch in renewable power systems planning

- **decision variables:**
  - power generation from slow-response generators x
  - power generation from fast-response generators y
- **random variable:**
  - renewable power generation ω_1 and electricity demand ω_2
- **parameters:**
  - cost functions of slow- and fast-response generators c_1 and c_2, respectively.
- **the decision-making problem:**

$$\underset{x \geq 0}{\text{minimize}} \quad c_1(x) + \mathbf{E}_{\widetilde{\omega}_1, \widetilde{\omega}_2}\left[\, h(x, \widetilde{\omega}_1, \widetilde{\omega}_2) \,\right]$$

$$\text{where} \quad h(x, \omega_1, \omega_2) = \begin{array}{ll} \underset{y \geq 0}{\text{minimum}} & c_2(\,[\,\omega_2 - x - \omega_1\,]_+\,)^\top y \\ \text{subject to} & y \geq \omega_2 - x - \omega_1 \end{array}$$

Decision-Dependent Uncertainty

A recent paper

L. Hellemo, P.I. Barton, and A. Tomasgard.
Decision-dependent probabilities in stochastic programs with recourse.
Computational Management Science 15 (2018) 369–395

$$\mathbf{E}_{\widetilde{\omega}}[\, h(x,\widetilde{\omega})\,] = \int_{\Omega} h(x,\widetilde{\omega}) \underbrace{p(x,\widetilde{\omega})}_{\text{decision-dependent density}} d\widetilde{\omega}.$$

Some examples: all decision dependent

Mixture of uncertainties, distorted distributions, scaled distributions.

Mathematical Formulation

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \zeta(x) \triangleq \underbrace{\varphi(x)}_{\text{cvx, diff}} + \mathbf{E}_{\widetilde{\omega}}[\quad \underbrace{\psi(x, \widetilde{\omega})}_{\text{random recourse}} \quad] \\ \text{subject to} & x \in X, \text{ a compact convex set in } \mathbb{R}^{n_1} \end{array}$$

Mathematical Formulation

$$\underset{x}{\text{minimize}} \quad \zeta(x) \triangleq \underbrace{\varphi(x)}_{\text{cvx, diff}} + \mathbf{E}_{\widetilde{\omega}}\Big[\underbrace{\psi(x,\widetilde{\omega})}_{\text{random recourse}}\Big]$$

$$\text{subject to} \quad x \in X, \text{ a compact convex set in } \mathbb{R}^{n_1}$$

A Linearly Bi-parameterized Quadratic Recourse Function

$$\psi(x,\omega) \triangleq \underset{y}{\text{minimum}} \Big[\underbrace{f(\omega) + G(\omega)x}_{\text{linearly parameterized by } x}\Big]^\top y + \underbrace{\tfrac{1}{2}\, y^\top Q y}_{\text{Convex}}$$

$$\text{subject to } y \in \underbrace{Y(x,\omega) \triangleq \left\{ y \in \mathbb{R}^{n_2} \mid A(\omega)x + Dy \geq \xi(\omega) \right\}}_{\text{a linearly parameterized polyhedron}}$$

Mathematical Formulation

$$\underset{x}{\text{minimize}} \quad \zeta(x) \triangleq \underbrace{\varphi(x)}_{\text{cvx, diff}} + \mathbf{E}_{\widetilde{\omega}}\Big[\underbrace{\psi(x,\widetilde{\omega})}_{\text{random recourse}} \Big]$$

$$\text{subject to} \quad x \in X, \text{ a compact convex set in } \mathbb{R}^{n_1}$$

A Linearly Bi-parameterized Quadratic Recourse Function

$$\psi(x,\omega) \triangleq \underset{y}{\text{minimum}} \; \Big[\underbrace{f(\omega) + G(\omega)x}_{\text{linearly parameterized by } x} \Big]^\top y + \underbrace{\tfrac{1}{2}\, y^\top Q y}_{\text{Convex}}$$

$$\text{subject to } y \in \underbrace{Y(x,\omega) \triangleq \left\{ y \in \mathbb{R}^{n_2} \mid A(\omega)x + Dy \geq \xi(\omega) \right\}}_{\text{a linearly parameterized polyhedron}}$$

Assumptions

- the recourse function $\psi(x,\omega)$ is finite on an open set Ξ containing X and for any $\omega \in \Omega$;
- the random functions $G(\omega)$, $A(\omega)$, $f(\omega)$ and $\xi(\omega)$ are essentially bounded.

A first approach: Complete deterministic equivalent

Finite scenarios: $\Omega = \{\omega^s\}_{s=1}^S$ with respective probabilities $\{p_s\}_{s=1}^S$.

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & \varphi(x) + \sum_{s=1}^{S} p_s \left[f(\omega^s) + G(\omega^s)x \right]^\top y^s + \frac{1}{2} (y^s)^\top Q y^s \\ \text{subject to} & x \in X \\ \text{and} & A(\omega^s)\, x + D y^s \geq \xi(\omega^s), \quad s = 1, \cdots, S. \end{array}$$

A (potentially) large-scale nonconvex quadratic program due to the bilinear terms $x^\top y^s$.

Alternative approach, bypassing the auxiliary scenario-dependent variables $\{ y^s \}_{s=1}^S$?

Background: Difference-of-convex programming

$$\underset{x \in X}{\text{minimize}}\ g(x) - h(x) \quad \text{with} \quad g \text{ and } h \text{ both convex.}$$

Background: Difference-of-convex programming

$$\underset{x \in X}{\text{minimize}}\; g(x) - h(x) \quad \text{with} \quad g \text{ and } h \text{ both convex.}$$

An iterative algorithm:. Given $x^\nu \in X$, let $a^\nu \in \partial h(x^\nu)$ be available. Solve the convex subprogram for a given scalar $\gamma > 0$,

$$\underset{x \in X}{\text{minimize}}\; \underbrace{g(x) - \left[\underbrace{h(x^\nu) + (\, a^\nu \,)^T (\, x - x^\nu}_{\text{linearization of } h \text{ at } x^\nu}\right] + \underbrace{\frac{\gamma}{2} \,\|\, x - x^\nu \,\|^2}_{\text{a proximal term}}}_{\text{a convex function in } x},$$

obtaining $x^{\nu+1}$. $\square$

How to employ this algorithm to the bi-parameterized SP on hand?

Recall the problem on hand:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \varphi(x) + \mathbf{E}_{\widetilde{\omega}}[\, \psi(x, \widetilde{\omega}) \,] \\ \text{subject to} & x \in X, \quad \text{where} \\ \psi(x,\omega) \triangleq & \underset{y}{\text{minimum}}\ [\, f(\omega) + G(\omega)x \,]^\top y + \frac{1}{2}\, y^\top Q y \\ & \text{subject to } A(\omega)x + Dy \geq \xi(\omega). \end{array}$$

Motivating result: (Nouiehed, Pang, Razaviyayn 2018)
$\psi(\bullet, \omega)$ is a difference-of-convex (piecewise-linear quadratic) function.

Recall the problem on hand:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \varphi(x) + \mathbf{E}_{\widetilde{\omega}}[\,\psi(x,\widetilde{\omega})\,] \\ \text{subject to} & x \in X, \quad \text{where} \\ \psi(x,\omega) \triangleq & \underset{y}{\text{minimum}}\ [\,f(\omega) + G(\omega)x\,]^\top y + \frac{1}{2}\,y^\top Q y \\ & \text{subject to } A(\omega)x + Dy \geq \xi(\omega). \end{array}$$

Motivating result: (Nouiehed, Pang, Razaviyayn 2018)
$\psi(\bullet,\omega)$ is a difference-of-convex (piecewise-linear quadratic) function.

Challenge: The dc decomposition is not practical for computation; because it involves a pointwise minimum of an exponential number of quadratic functions.

- **Goal:** Compute a "sharp stationary" solution.
- **Approach:** variable sample average approximation and dc approximation.

- **Goal:** Compute a "sharp stationary" solution.
- **Approach:** variable sample average approximation and dc approximation.
- **Three cases:**
  - Q is PD, leading to a relatively straightforward difference-of-convex (dc) decomposition
  - two special PSD cases, leading to a Clarke-regular recourse function
  - Q is generally PSD, introducing *a generalized critical point*

Concepts of Differentiability

Let $f : \Xi \to \mathbb{R}$ be a locally Lipschitz function defined on an open set $\Xi \subseteq \mathbb{R}^n$, Let $\bar{z} \in \Xi$ be given.

Concepts of Differentiability

Let $f : \Xi \to \mathbb{R}$ be a locally Lipschitz function defined on an open set $\Xi \subseteq \mathbb{R}^n$, Let $\bar{z} \in \Xi$ be given.

- The directional derivative of f at $\bar{z}$ along the direction $d \in \mathbb{R}^n$

$$f'(\bar{z}; d) \triangleq \lim_{\tau \downarrow 0} \frac{f(\bar{z} + \tau d) - f(\bar{z})}{\tau}$$

Concepts of Differentiability

Let $f : \Xi \to \mathbb{R}$ be a locally Lipschitz function defined on an open set $\Xi \subseteq \mathbb{R}^n$, Let $\bar{z} \in \Xi$ be given.

- The directional derivative of f at $\bar{z}$ along the direction $d \in \mathbb{R}^n$

$$f'(\bar{z}; d) \triangleq \lim_{\tau \downarrow 0} \frac{f(\bar{z} + \tau d) - f(\bar{z})}{\tau}$$

- The Clarke directional derivative of f at $\bar{z}$ along the direction $d \in \mathbb{R}^n$

$$f^0(\bar{z}; d) \triangleq \limsup_{\substack{z \to \bar{z} \\ \tau \downarrow 0}} \frac{f(z + \tau d) - f(z)}{\tau},$$

Concepts of Differentiability

Let $f : \Xi \to \mathbb{R}$ be a locally Lipschitz function defined on an open set $\Xi \subseteq \mathbb{R}^n$, Let $\bar{z} \in \Xi$ be given.

- The directional derivative of f at $\bar{z}$ along the direction $d \in \mathbb{R}^n$

$$f'(\bar{z}; d) \triangleq \lim_{\tau \downarrow 0} \frac{f(\bar{z} + \tau\, d) - f(\bar{z})}{\tau}$$

- The Clarke directional derivative of f at $\bar{z}$ along the direction $d \in \mathbb{R}^n$

$$f^0(\bar{z}; d) \triangleq \limsup_{\substack{z \to \bar{z} \\ \tau \downarrow 0}} \frac{f(z + \tau\, d) - f(z)}{\tau},$$

- if f is directional differentiable at $\bar{x}$, then $f^0(\bar{z}; d) \geq f'(\bar{z}; d), \forall d \in \mathbb{R}^n$.

Concepts of Differentiability

Let $f : \Xi \to \mathbb{R}$ be a locally Lipschitz function defined on an open set $\Xi \subseteq \mathbb{R}^n$, Let $\bar{z} \in \Xi$ be given.

- The directional derivative of f at $\bar{z}$ along the direction $d \in \mathbb{R}^n$

$$f'(\bar{z}; d) \triangleq \lim_{\tau \downarrow 0} \frac{f(\bar{z} + \tau d) - f(\bar{z})}{\tau}$$

- The Clarke directional derivative of f at $\bar{z}$ along the direction $d \in \mathbb{R}^n$

$$f^0(\bar{z}; d) \triangleq \limsup_{\substack{z \to \bar{z} \\ \tau \downarrow 0}} \frac{f(z + \tau d) - f(z)}{\tau},$$

- if f is directional differentiable at $\bar{x}$, then $f^0(\bar{z}; d) \geq f'(\bar{z}; d), \forall d \in \mathbb{R}^n$.

- We say the function f is Clarke regular at $\bar{z} \in Z$ if f is directionally differentiable at $\bar{z}$ and $f^0(\bar{z}; d) = f'(\bar{z}; d)$ for all $d \in \mathbb{R}^n$.

Concepts of Stationarity

A vector $\bar{x} \in X$ is said to be a xxx stationary point of f on a closed convex set X in the sense of

- d(irectional) if $f'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$;
- C(larke) if $f^0(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$.

Concepts of Stationarity

A vector $\bar{x} \in X$ is said to be a xxx stationary point of f on a closed convex set X in the sense of

- d(irectional) if $f'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$;
- C(larke) if $f^0(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$.

If $f = g - h$ is a dc function with g and h being convex,

- $\bar{x} \in X$ is a d-stationary point of f on X if $\partial h(\bar{x}) \subseteq \partial g(\bar{x}) + \mathcal{N}(\bar{x}; X)$;
- $\bar{x} \in X$ is a critical point of f on X if $0 \in \left[\, \partial g(\bar{x}) - \partial h(\bar{x}) \,\right] + \mathcal{N}(\bar{x}; X)$.

The Positive Definite case

dc decomposition:

$$\psi(x,\omega) = \psi_1(x,\omega) - \psi_2(x,\omega)$$

where, using duality,

$$\begin{aligned}\psi_1(x,\omega) \quad &\triangleq \max_{\lambda \geq 0} \; -\tfrac{1}{2}\,\lambda^\top D Q^{-1} D^\top \lambda \\ &\qquad +\lambda^\top \left[\, \xi(\omega) - A(\omega)x + DQ^{-1}\left(\, f(\omega) + G(\omega)x \,\right) \right] \\ \psi_2(x,\omega) \quad &\triangleq \tfrac{1}{2}\,\left[\, f(\omega) + G(\omega)x \,\right]^\top Q^{-1} \left[\, f(\omega) + G(\omega)x \,\right],\end{aligned}$$

The Positive Definite case

dc decomposition:

$$\psi(x,\omega) = \psi_1(x,\omega) - \psi_2(x,\omega)$$

where, using duality,

$$\begin{aligned}
\psi_1(x,\omega) &\triangleq \max_{\lambda \geq 0} \; -\tfrac{1}{2}\,\lambda^\top D Q^{-1} D^\top \lambda \\
&\qquad\qquad + \lambda^\top \left[\, \xi(\omega) - A(\omega)x + DQ^{-1}\left(\, f(\omega) + G(\omega)x \,\right) \right] \\
\psi_2(x,\omega) &\triangleq \tfrac{1}{2}\left[\, f(\omega) + G(\omega)x \,\right]^\top Q^{-1} \left[\, f(\omega) + G(\omega)x \,\right],
\end{aligned}$$

Two-stage SP = a stochastic dc program

$$\underset{x}{\text{minimize}} \quad \varphi(x) + \mathbf{E}_{\widetilde{\omega}}\Big[\underbrace{\psi_1(x,\widetilde{\omega})}_{\text{nondifferentiable}} \Big] - \mathbf{E}_{\widetilde{\omega}}\Big[\underbrace{\psi_2(x,\widetilde{\omega})}_{\text{quadratic}} \Big]$$

Combined Sequential SAA-DCA algorithm

- **the SAA discretization:**

$$\mathbf{E}_{\widetilde{\omega}}[\,\psi(x,\widetilde{\omega})\,] \approx \frac{1}{N}\sum_{i=1}^{N}\psi(x,\omega^i) = \frac{1}{N}\sum_{i=1}^{N}\Big[\,\psi_1(x,\omega^i) - \psi_2(x,\omega^i)\,\Big]$$

Combined Sequential SAA-DCA algorithm

- **the SAA discretization:**

$$\mathbf{E}_{\widetilde{\omega}}[\,\psi(x,\widetilde{\omega})\,] \approx \frac{1}{N}\sum_{i=1}^{N}\psi(x,\omega^i) = \frac{1}{N}\sum_{i=1}^{N}\Big[\,\psi_1(x,\omega^i) - \psi_2(x,\omega^i)\,\Big]$$

- **the linearized approximation**

$$\psi_2(x,\omega) \approx \widehat{\psi}_2(x,\omega;\, x') \triangleq \psi_2(x',\omega) + \nabla\psi_2(x',\omega)^\top(x - x')$$

Combined Sequential SAA-DCA algorithm

- **the SAA discretization:**

$$\mathbf{E}_{\widetilde{\omega}}[\,\psi(x,\widetilde{\omega})\,] \approx \frac{1}{N}\sum_{i=1}^{N}\psi(x,\omega^i) = \frac{1}{N}\sum_{i=1}^{N}\left[\,\psi_1(x,\omega^i) - \psi_2(x,\omega^i)\,\right]$$

- **the linearized approximation**

$$\psi_2(x,\omega) \approx \widehat{\psi}_2(x,\omega;\,x') \triangleq \psi_2(x',\omega) + \nabla\psi_2(x',\omega)^\top(x-x')$$

Algorithm

1. Initialization: Let $\{L_\nu\}_{\nu=0}^{\infty}$ be a sequence of positive integers. Let $\gamma > 0$ be a given scalar. Let an initial feasible vector $\widetilde{x}^{\,0} \in X$ be given.
2. At iteration ν, generate iid samples $\{\omega^{\nu,i}\}_{i=1}^{L_\nu}$ and let $\widetilde{x}^{\,\nu+1}$ be

$$\underset{x\in X}{\text{argmin}}\;\underbrace{\varphi(x) + \frac{1}{L_\nu}\sum_{i=1}^{L_\nu}\left[\,\psi_1(x,\omega^{\nu,i}) - \widehat{\psi}_2(x,\omega^{\nu,i};\widetilde{x}^{\,\nu})\,\right]}_{\text{convex in } x} + \frac{1}{2\,\gamma}\|\,x-\widetilde{x}^{\,\nu}\,\|^2.$$

Convergence in the PD case

Theorem: Suppose $\sum_{\nu=1}^{\infty} \frac{1}{L_\nu^{1/2}} < \infty$, then every accumulation point of the sequence $\{\widetilde{x}^\nu\}$ produced by the Algorithm, one of which must exist, is a d-stationary point of the two-stage SP problem, almost surely.

Convergence in the PD case

Theorem: Suppose $\sum_{\nu=1}^{\infty} \frac{1}{L_\nu^{1/2}} < \infty$, then every accumulation point of the sequence $\{\widetilde{x}^\nu\}$ produced by the Algorithm, one of which must exist, is a d-stationary point of the two-stage SP problem, almost surely.

Note

- the samples among iterations are independent
- the accumulated error of SAA is finite with a controllable growth of the sample size
- each convex subproblem needs to be solved by numerical algorithms

PSD: Special Case 1

Assumption: the vector f and matrix G in the recourse function are deterministic and satisfy $f + GX \subseteq \text{Range}\, Q$.

PSD: Special Case 1

Assumption: the vector f and matrix G in the recourse function are deterministic and satisfy $f + GX \subseteq \text{Range}\, Q$.

In this case, let $Qz = f + Gx$,

$$\begin{array}{rcll} \psi(x,\omega) & = & \underset{y}{\text{minimum}} & (\, f + Gx\,)^\top y + \frac{1}{2}\, y^\top Qy \\ & & \text{subject to} & A(\omega)x + Dy \, \geq \, \xi(\omega) \end{array}$$

$$= \underbrace{\left[\begin{array}{ll} \underset{w}{\text{minimum}} & \frac{1}{2}\, w^\top Qw \\ \text{subject to} & \left[\, A(\omega)x - Dz\,\right] + Dw \, \geq \, \xi(\omega) \end{array}\right]}_{\text{denoted by } \theta(x,z,\omega), \text{ a convex function}} - \frac{1}{2}\, z^\top Qz.$$

PSD: Special Case 1

Assumption: the vector f and matrix G in the recourse function are deterministic and satisfy $f + GX \subseteq \text{Range}\, Q$.

In this case, let $Qz = f + Gx$,

$$\begin{aligned}
\psi(x,\omega) \;&=\; \underset{y}{\text{minimum}} \quad (\, f + Gx\,)^\top y + \tfrac{1}{2}\, y^\top Q y \\
&\qquad \text{subject to} \quad A(\omega)x + Dy \;\geq\; \xi(\omega) \\
&=\; \underbrace{\left[\begin{array}{ll} \underset{w}{\text{minimum}} & \tfrac{1}{2}\, w^\top Q w \\ \text{subject to} & \big[\, A(\omega)x - Dz \,\big] + Dw \;\geq\; \xi(\omega) \end{array}\right]}_{\text{denoted by } \theta(x,z,\omega),\text{ a convex function}} - \tfrac{1}{2}\, z^\top Q z.
\end{aligned}$$

Two-stage SP $\Leftrightarrow$ a dc SP with a differentiable concave term.

$$\begin{array}{ll}
\underset{x,\,z}{\text{minimize}} & \varphi(x) + \mathbf{E}_{\widetilde{\omega}}\left[\,\theta(x,z,\widetilde{\omega})\,\right] - \tfrac{1}{2}\, z^T Q z \\
\text{subject to} & \underbrace{x \in X \quad \text{and} \quad f + Gx \;=\; Qz}_{\text{lifted feasible set of pairs } (x,z)}\;.
\end{array}$$

Sequential SAA-DCA can be applied.

PSD: Special Case 2

With the eigenvalue decomposition

$$Q = \begin{bmatrix} P_+ & P_0 \end{bmatrix} \begin{bmatrix} \Lambda_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P_+ & P_0 \end{bmatrix}^\top$$

let $\begin{bmatrix} f_+(\omega) & G_+(\omega) & D_+^T \\ f_0(\omega) & G_0(\omega) & D_0^T \end{bmatrix} \triangleq \begin{bmatrix} P_+^T \\ P_0^T \end{bmatrix} \begin{bmatrix} f(\omega) & G(\omega) & D^T \end{bmatrix}$

PSD: Special Case 2

With the eigenvalue decomposition

$$Q = \begin{bmatrix} P_+ & P_0 \end{bmatrix} \begin{bmatrix} \Lambda_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P_+ & P_0 \end{bmatrix}^\top$$

$$\text{let} \quad \begin{bmatrix} f_+(\omega) & G_+(\omega) & D_+^T \\ f_0(\omega) & G_0(\omega) & D_0^T \end{bmatrix} \triangleq \begin{bmatrix} P_+^T \\ P_0^T \end{bmatrix} \begin{bmatrix} f(\omega) & G(\omega) & D^T \end{bmatrix}$$

The recourse function $\psi(x,\omega)$ is equal to

$$- \left[\begin{array}{l} \underset{u \geq 0}{\text{minimum}} \ \frac{1}{2} u^\top \underbrace{D_+ \Lambda_+^{-1} D_+^\top}_{PSD\ \ matrix} u + \\ u^\top \underbrace{\left[A(\omega)x - \xi(\omega) - D_+ \Lambda_+^{-1} (f_+(\omega) + G_+(\omega)x) \right]}_{linearly\ \ parameterized\ \ by\ \ x} \\ \text{subject to} \quad f_0(\omega) + G_0(\omega)x - D_0^\top u = 0 \end{array} \right]$$

$$- \frac{1}{2} \underbrace{\left[f_+(\omega) + G_+(\omega)x \right]^\top \Lambda_+^{-1} \left[f_+(\omega) + G_+(\omega)x \right]}_{a\ \ quadratic\ \ function}$$

A range condition

Suppose $G_0(\omega) = 0$; or equivalently, $\left[\text{ Range } G(\omega) \subseteq \text{ Range } Q\right]$. Then $\psi(x, \omega)$ is equal to

$$\left[\begin{array}{l} \underset{u}{\text{minimum}} \ \frac{1}{2}\, u^\top \underbrace{D_+ \Lambda_+^{-1} D_+^\top}_{PSD\ \ matrix} u + \\ u^\top \underbrace{\left[A(\omega)x - \xi(\omega) - D_+ \Lambda_+^{-1}(f_+(\omega) + G_+(\omega)x) \right]}_{linearly\ \ parameterized\ \ by\ \ x} \\ \text{subject to} \quad \underbrace{f_0(\omega) - D_0^\top u \ = \ 0; \quad u \ \geq \ 0}_{no\ \ first-stage\ \ variable\ \ x} \end{array}\right]$$

$$-\frac{1}{2} \underbrace{\left[f_+(\omega) + G_+(\omega)x \right]^\top \Lambda_+^{-1} \left[f_+(\omega) + G_+(\omega)x \right]}_{a\ \ quadratic\ \ function}$$

Hence, two-stage SP $\Leftrightarrow$ a dc SP with a differentiable concave term.

The general PSD case

Challenge: Explicit computationally usable dc decomposition disappears in the general psd case.

The general PSD case

Challenge: Explicit computationally usable dc decomposition disappears in the general psd case.

Questions:

- What kind of a stationary solution can one aim to compute in the absence of dc decomposition?
- What sort of an algorithm can one design to approximate such a solution?

Convex-Concave condition

For a function $g : X \subseteq \mathbb{R}^n \to \mathbb{R}$, there exists a bivariate function $h : X \times X \to \mathbb{R}$ such that

- $g(x) = h(x, x)$ for any $x \in X$
- $h(\bullet, x)$ is convex and $h(x, \bullet)$ is concave.

Convex-Concave condition

For a function $g : X \subseteq \mathbb{R}^n \to \mathbb{R}$, there exists a bivariate function $h : X \times X \to \mathbb{R}$ such that

- $g(x) = h(x, x)$ for any $x \in X$
- $h(\bullet, x)$ is convex and $h(x, \bullet)$ is concave.

The recourse function $\psi(\bullet, \omega)$ satisfies the convex-concave condition

Specifically, we define **the lifted recourse function $\bar{\psi}(x, z, \omega)$**

$$\begin{array}{ll} \underset{y}{\text{minimum}} & \left[\, f(\omega) + G(\omega) z \,\right]^\top y + \frac{1}{2}\, y^\top Q y \\ \text{subject to} & y \in Y(x, \omega) \triangleq \{\, y \in \mathbb{R}^{n_2} : A(\omega) x + D y \geq \xi(\omega) \,\}, \end{array}$$

we have

- $\psi(x, \omega) = \bar{\psi}(x, x, \omega)$
- $\bar{\psi}(\bullet, z, \omega)$ is convex, $\bar{\psi}(x, \bullet, \omega)$ is concave.

Directional Derivative of the Recourse Function

In the second-stage problem, let $M(x, \omega)$ denote the sets of optimal primal solutions; $\Lambda(x, \omega)$ denote the optimal dual solutions

Directional Derivative of the Recourse Function

In the second-stage problem, let $M(x,\omega)$ denote the sets of optimal primal solutions; $\Lambda(x,\omega)$ denote the optimal dual solutions

- the partial directional derivatives of the lifted recourse function $\bar{\psi}$ are

$$
\begin{aligned}
\bar{\psi}(\bullet,\bar{x},\omega)'(\bar{x};\, d) &= \max_{\lambda\in\Lambda(\bar{x},\omega)} \left[-A(\omega)^\top\lambda\right]^\top d \\
\bar{\psi}(\bar{x},\bullet,\omega)'(\bar{x};\, d) &= \min_{y\in M(\bar{x},\omega)} \left[\, G(\omega)^\top y\,\right]^\top d
\end{aligned}
$$

Directional Derivative of the Recourse Function

In the second-stage problem, let $M(x, \omega)$ denote the sets of optimal primal solutions; $\Lambda(x, \omega)$ denote the optimal dual solutions

- the partial directional derivatives of the lifted recourse function $\bar{\psi}$ are

$$\begin{aligned} \bar{\psi}(\bullet, \bar{x}, \omega)'(\bar{x};\, d) &= \max_{\lambda \in \Lambda(\bar{x}, \omega)} \left[-A(\omega)^\top \lambda \right]^\top d \\ \bar{\psi}(\bar{x}, \bullet, \omega)'(\bar{x};\, d) &= \min_{y \in M(\bar{x}, \omega)} \left[G(\omega)^\top y \right]^\top d \end{aligned}$$

- the directional derivative of the recourse function ψ is

$$\begin{aligned} \psi(\bullet, \omega)'(\bar{x}; d) &= \bar{\psi}(\bar{x}, \bullet, \omega)'(\bar{x};\, d) + \bar{\psi}(\bullet, \bar{x}, \omega)'(\bar{x};\, d) \\ &= \min_{y \in M(\bar{x}, \omega)} \max_{\lambda \in \Lambda(\bar{x}, \omega)} \nabla_x \underbrace{L(x, \omega; y, \lambda)}_{\text{Lagrangian}}{}^\top d \end{aligned}$$

Directional Derivative of the Recourse Function

In the second-stage problem, let $M(x, \omega)$ denote the sets of optimal primal solutions; $\Lambda(x, \omega)$ denote the optimal dual solutions

- the partial directional derivatives of the lifted recourse function $\bar{\psi}$ are

$$\bar{\psi}(\bullet, \bar{x}, \omega)'(\bar{x};\, d) = \max_{\lambda \in \Lambda(\bar{x},\omega)} \left[-A(\omega)^\top \lambda \right]^\top d$$

$$\bar{\psi}(\bar{x}, \bullet, \omega)'(\bar{x};\, d) = \min_{y \in M(\bar{x},\omega)} \left[G(\omega)^\top y \right]^\top d$$

- the directional derivative of the recourse function ψ is

$$\begin{aligned} \psi(\bullet, \omega)'(\bar{x}; d) &= \bar{\psi}(\bar{x}, \bullet, \omega)'(\bar{x};\, d) + \bar{\psi}(\bullet, \bar{x}, \omega)'(\bar{x};\, d) \\ &= \min_{y \in M(\bar{x},\omega)} \max_{\lambda \in \Lambda(\bar{x},\omega)} \nabla_x \underbrace{L(x, \omega; y, \lambda)}_{\text{Lagrangian}}{}^\top d \end{aligned}$$

Note: it establishes the equality between the total directional derivative of the bivariate function $\bar{\psi}(\bullet, \bullet, \omega)$ and the sum of the partial directional derivatives of the two partial functions $\bar{\psi}(\bullet, z, \omega)$ and $\bar{\psi}(x, \bullet, \omega)$.

Directional Derivative of the Recourse Function

In the second-stage problem, let $M(x, \omega)$ denote the sets of optimal primal solutions; $\Lambda(x, \omega)$ denote the optimal dual solutions

- the partial directional derivatives of the lifted recourse function $\bar{\psi}$ are

$$\begin{aligned}
\bar{\psi}(\bullet, \bar{x}, \omega)'(\bar{x};\, d) &= \max_{\lambda \in \Lambda(\bar{x},\omega)} \left[-A(\omega)^\top \lambda \right]^\top d \\
\bar{\psi}(\bar{x}, \bullet, \omega)'(\bar{x};\, d) &= \min_{y \in M(\bar{x},\omega)} \left[G(\omega)^\top y \right]^\top d
\end{aligned}$$

- the directional derivative of the recourse function ψ is

$$\begin{aligned}
\psi(\bullet, \omega)'(\bar{x}; d) &= \bar{\psi}(\bar{x}, \bullet, \omega)'(\bar{x};\, d) + \bar{\psi}(\bullet, \bar{x}, \omega)'(\bar{x};\, d) \\
&= \min_{y \in M(\bar{x},\omega)} \; \max_{\lambda \in \Lambda(\bar{x},\omega)} \nabla_x \underbrace{L(x, \omega; y, \lambda)}_{\text{Lagrangian}}{}^\top d
\end{aligned}$$

Note: it establishes the equality between the total directional derivative of the bivariate function $\bar{\psi}(\bullet, \bullet, \omega)$ and the sum of the partial directional derivatives of the two partial functions $\bar{\psi}(\bullet, z, \omega)$ and $\bar{\psi}(x, \bullet, \omega)$.

Generalized Critical Point

Definition: Let $S \subseteq \mathbb{R}^n$ be a convex set and a function $g : S \to \mathbb{R}$ satisfies the convex-concave condition with the associated bivariate convex-concave function h. We say that $\bar{x} \in S$ is a *generalized critical point* of g on S if

$$0 \in \partial_x h(\bar{x}, \bar{x}) - \partial_z(-h(\bar{x}, \bar{x})) + \mathcal{N}(\bar{x}, S),$$

where $\partial_x h(\bar{x}, \bar{x})$ is the subgradient of the convex function $h(\bullet, \bar{x})$ at $\bar{x}$, and similarly for $\partial_z(-h)(\bar{x}, \bar{x})$.

Generalized Critical Point

Definition: Let $S \subseteq \mathbb{R}^n$ be a convex set and a function $g : S \to \mathbb{R}$ satisfies the convex-concave condition with the associated bivariate convex-concave function h. We say that $\bar{x} \in S$ is a *generalized critical point* of g on S if

$$0 \in \partial_x h(\bar{x}, \bar{x}) - \partial_z(-h(\bar{x}, \bar{x})) + \mathcal{N}(\bar{x}, S),$$

where $\partial_x h(\bar{x}, \bar{x})$ is the subgradient of the convex function $h(\bullet, \bar{x})$ at $\bar{x}$, and similarly for $\partial_z(-h)(\bar{x}, \bar{x})$.

For the two-stage SP in the general psd case

We say that $\bar{x} \in X$ is a *generalized critical point* if

$$0 \in \nabla\varphi(\bar{x}) + \partial_x \mathbf{E}_{\widetilde{\omega}} \left[\bar{\psi}(\bar{x}, \bar{x}, \widetilde{\omega}) \right] - \partial_z \mathbf{E}_{\widetilde{\omega}} \left[-\bar{\psi}(\bar{x}, \bar{x}, \widetilde{\omega}) \right] + \mathcal{N}(\bar{x}, X),$$

Equivalently, there exists $\bar{v} \in \partial_z \mathbf{E}_{\widetilde{\omega}} \left[-\bar{\psi}(\bar{x}, \bar{x}, \widetilde{\omega}) \right]$

$$\bar{x} \in \underset{x \in X}{\text{minimize}} \underbrace{\varphi(x) + \mathbf{E}_{\widetilde{\omega}} \left[\bar{\psi}(x, \bar{x}, \widetilde{\omega}) \right] - \bar{v}^\top (x - \bar{x})}_{\text{convex function in x}}.$$

Relation to the C-stationarity, d-stationarity

$S_{\rm gc}$ denote the sets of generalized critical points

$S_{\rm C}$ denote the set of Clarke stationary points

$S_{\rm d}$ denote the set of directional stationary points.

Relation to the C-stationarity, d-stationarity

$S_{\rm gc}$ denote the sets of generalized critical points

$S_{\rm C}$ denote the set of Clarke stationary points

$S_{\rm d}$ denote the set of directional stationary points.

Proposition: For the two-stage stochastic program (**P**), under the relatively complete assumption,

(a) $S_{\rm d} \subseteq S_{\rm C} \subseteq S_{\rm gc}$;

Relation to the C-stationarity, d-stationarity

$S_{\rm gc}$ denote the sets of generalized critical points

$S_{\rm C}$ denote the set of Clarke stationary points

$S_{\rm d}$ denote the set of directional stationary points.

Proposition: For the two-stage stochastic program (**P**), under the relatively complete assumption,

(a) $S_{\rm d} \subseteq S_{\rm C} \subseteq S_{\rm gc}$;

(b) if for almost every $\omega \in \Omega$, Range $G(\omega) \subseteq$ Range Q, then for almost all $\omega \in \Omega$, the function $\psi(\bullet, \omega)$ is Clarke regular on X;

Relation to the C-stationarity, d-stationarity

$S_{\rm gc}$ denote the sets of generalized critical points

$S_{\rm C}$ denote the set of Clarke stationary points

$S_{\rm d}$ denote the set of directional stationary points.

Proposition: For the two-stage stochastic program (**P**), under the relatively complete assumption,

(a) $S_{\rm d} \subseteq S_{\rm C} \subseteq S_{\rm gc}$;

(b) if for almost every $\omega \in \Omega$, Range $G(\omega) \subseteq$ Range Q, then for almost all $\omega \in \Omega$, the function $\psi(\bullet, \omega)$ is Clarke regular on X;

(c) if in addition to the range assumption in (b), the random functions $f(\omega)$, $\xi(\omega)$, $G(\omega)$ and $A(\omega)$ are essentially bounded, then $S_{\rm d} = S_{\rm C} = S_{\rm gc}$.

Regularization

Regularized recourse function $\psi_\alpha(x,\omega)$

$$\begin{array}{ll} \underset{y}{\text{minimum}} & \left[\, f(\omega) + G(\omega)x \,\right]^\top y + \frac{1}{2}\, y^\top \left[\, Q + \alpha\, \mathbb{I} \,\right] y \\ \text{subject to} & y \in Y(x,\omega) \triangleq \left\{\, y \in \mathbb{R}^{n_2} \mid A(\omega)x + Dy \geq \xi(\omega) \right\}, \end{array}$$

dc decomposition: $\psi_\alpha(x,\omega) = \psi_{\alpha,1}(x,\omega) - \psi_{\alpha,2}(x,\omega)$.

Regularization

Regularized recourse function $\psi_\alpha(x,\omega)$

$$\begin{array}{ll} \underset{y}{\text{minimum}} & \left[\, f(\omega) + G(\omega)x \,\right]^\top y + \frac{1}{2}\, y^\top \left[\, Q + \alpha\, \mathbb{I} \,\right] y \\ \text{subject to} & y \,\in\, Y(x,\omega) \,\triangleq\, \left\{\, y \,\in\, \mathbb{R}^{n_2} \,\mid\, A(\omega)x + Dy \,\geq\, \xi(\omega) \right\}, \end{array}$$

dc decomposition: $\psi_\alpha(x,\omega) \,=\, \psi_{\alpha,1}(x,\omega) - \psi_{\alpha,2}(x,\omega)$.

Sequential Regularization: For each positive scalar α_ν, each x^ν is a d-stationary solution of the dc stochastic program:

$$\underset{x \in X}{\text{minimize}}\ \zeta_{\alpha_\nu}(x) \,\triangleq\, \varphi(x) + \mathbf{E}_{\widetilde{\omega}}\left[\, \psi_{\alpha_\nu}(x,\widetilde{\omega}) \,\right];$$

Regularization

Regularized recourse function $\psi_\alpha(x,\omega)$

$$\begin{array}{ll} \underset{y}{\text{minimum}} & \left[f(\omega) + G(\omega)x \right]^\top y + \frac{1}{2} y^\top \left[Q + \alpha \mathbb{I} \right] y \\ \text{subject to} & y \in Y(x,\omega) \triangleq \left\{ y \in \mathbb{R}^{n_2} \mid A(\omega)x + Dy \geq \xi(\omega) \right\}, \end{array}$$

dc decomposition: $\psi_\alpha(x,\omega) = \psi_{\alpha,1}(x,\omega) - \psi_{\alpha,2}(x,\omega)$.

Sequential Regularization: For each positive scalar α_ν, each x^ν is a d-stationary solution of the dc stochastic program:

$$\underset{x \in X}{\text{minimize}}\ \zeta_{\alpha_\nu}(x) \triangleq \varphi(x) + \mathbf{E}_{\widetilde{\omega}} \left[\psi_{\alpha_\nu}(x,\widetilde{\omega}) \right];$$

Assumption: For almost all $\omega \in \Omega$ and all $x \in X$, $[\, Dv \geq 0, Qv = 0 \,] \Rightarrow v = 0$ (for the boundedness of the second-stage primal optimal solution)

Theorem: If the nonincreasing sequence $\{\alpha_\nu\}$ converges to 0, then every accumulation point of the sequence $\{x^\nu\}$

Simultaneous Regularization and Linearization

For each positive scalar α_ν, each $\widehat{x}^{\,\nu+1}$ is the unique minimizer of the strongly convex SP:

$$\underset{x \in X}{\text{minimize}} \quad \varphi(x) + \mathbf{E}_{\widetilde{\omega}}\left[\,\psi_{\alpha_\nu,1}(x,\widetilde{\omega})\,\right] - \mathbf{E}_{\widetilde{\omega}}\left[\,\widehat{\psi}_{\alpha_\nu,2}(x,\widetilde{\omega};\widehat{x}^{\,\nu})\,\right] + \frac{1}{2\,\gamma}\,\|\,x - \widehat{x}^{\,\nu}\,\|^2.$$

Theorem: If $\{\alpha_\nu\}$ is a nonincreasing sequence of positive scalars satisfying $\alpha_\nu = \Omega(\nu^{-1/2})$, then every accumulation point of the sequence $\{\widehat{x}^{\,\nu}\}$, at least one of which must exist, is a generalized critical point of the two-stage SP. □

Sampling, Regularization and Linearization

For each positive scalar α_ν, each $\widetilde{x}^{\,\nu+1}$ is the unique minimizer of the discretized strongly convex (sampled) program:

$$\underset{x \in X}{\text{minimize}} \quad \varphi(x) + \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \left[\psi_{\alpha_\nu,1}(x, \omega^{\nu,i}) - \widehat{\psi}_{\alpha_\nu,2}(x, \omega^{\nu,i}; \widetilde{x}^{\,\nu}) \right] + \frac{1}{2\gamma} \| x - \widetilde{x}^{\,\nu} \|^2.$$

Assumption: For almost $\omega \in \Omega$,

$$[\, D^\top \lambda = 0 \text{ and } \lambda \geq 0 \,] \Rightarrow A(\omega)^\top \lambda = 0.$$

• implied by either Slater ($\exists\, \widehat{y}$ such that $D\widehat{y} > 0$), or Range $A(\omega) \subseteq$ Range D (contrast to Range $G(\omega) \subseteq$ Range Q);

• ensures boundedness of $\partial \mathbf{E}_{\widetilde{\omega}} \left[\psi_{\alpha,1}(x, \widetilde{\omega}) \right]$ for $\alpha > 0$.

Theorem: If the sequences of parameters $\{\alpha_\nu\}$ and $\{L_\nu\}$ satisfy that $\alpha_\nu = \Omega(\nu^{-1/2})$ and $\sum_\nu \frac{1}{\alpha_\nu^2 L_\nu^{1/2}} < \infty$, then any accumulation point of the sequence $\{\widetilde{x}^\nu\}$, at least one of which must exist, is a generalized critical point of the two-stage SP with probability 1. $\square$

Theorem: If the sequences of parameters $\{\alpha_\nu\}$ and $\{L_\nu\}$ satisfy that $\alpha_\nu = \Omega(\nu^{-1/2})$ and $\sum_\nu \frac{1}{\alpha_\nu^2 L_\nu^{1/2}} < \infty$, then any accumulation point of the sequence $\{\widetilde{x}^\nu\}$, at least one of which must exist, is a generalized critical point of the two-stage SP with probability 1. □

Note: the slower α_ν diminishes, the slower the sample size Ł$_\nu$ increases.
e.g. $\alpha_\nu = \frac{1}{\log\, \nu}$, and $L_\nu = \nu^{2+2\delta} \log^4 \nu$ for $\delta > 0$.

In Summary

For the two-stage SP with a linearly bi-parameterized quadratic recourse, we develop algorithmic schemes to compute stationary points using **sampling, convexification, and regularization**, resulting in implementable algorithms that need to be refined and enhanced for practical application.

PD case	SAA + DCA	d stationary point w.p.1
special PSD case	SAA +DCA	d stationary point w.p.1
general PSD case	regularization + SAA + DCA	generalized critical w.p.1