Two-stage Stochastic Programs with Nonconvex Recourse

Ying Cui

Department of Industrial and Systems Engineering
University of Minnesota

Joint work with Hanyang Li (UMN)

Also acknowledge: Junyi Liu, Jong-Shi Pang and Suvrajeet Sen

Two-stage stochastic programs

stage-one decision x: here and now

↓

nature makes a random decision ω

↓

stage-two decision y_ω: wait and see

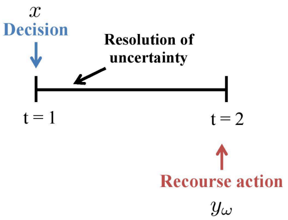


Figure taken from [Li and Grossmann, 2021]

Origin: George B. Dantzig (1955)

Monographs: [Shapiro, Dentcheva and Ruszczyński, 2009], [Birge and Louveaux, 2011] ...

An example: power system planning

Stage 1: Determine the capacity of each power plant and unit price

↓

The random demand and random production (from renewable energy) are observed

↓

Stage 2: Allocate the production of each power plant to different locations

Two-stage linear programs

A standard two-stage linear program (X polyhedral):

$$\underset{x \in X}{\text{minimize}} \ (c^0)^\top x + \mathbb{E}_{\omega \sim \mathbb{P}} \left[\psi(x;\omega) \right],$$

where $\psi(x;\omega)$ is the second-stage recourse:

$$\begin{aligned} \psi(x;\omega) \triangleq \underset{y}{\text{minimum}} \quad & c(\omega)^\top y \\ \text{subject to} \quad & T(\omega)x + W(\omega)y = h(\omega). \end{aligned}$$

x: stage-one decision (independent of ω)

y: stage-two decision (depends on ω)

Two-stage linear programs

A standard two-stage linear program (X polyhedral):

$$\underset{x \in X}{\text{minimize}} \ (c^0)^\top x + \mathbb{E}_{\omega \sim \mathbb{P}} [\, \psi(x;\omega) \,],$$

where $\psi(x;\omega)$ is the second-stage recourse:

$$\begin{aligned} \psi(x;\omega) \triangleq \ \underset{y}{\text{minimum}} & \quad c(\omega)^\top y \\ \text{subject to} & \quad T(\omega) x + W(\omega) y = h(\omega). \end{aligned}$$

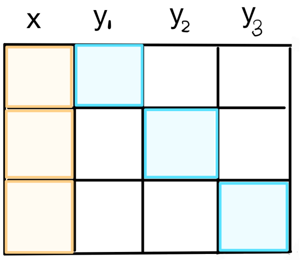
Given S realizations of ω, the problem becomes

$$\begin{aligned} \underset{x \in X, y^1, \cdots, y^S}{\text{minimize}} & \quad (c^0)^\top x + \frac{1}{S} \sum_{s=1}^{S} \left[\, (c^s)^\top y^s \,\right] \\ \text{subject to} & \quad T^s x + W^s y^s = h^s, \quad s = 1, \cdots, S. \end{aligned}$$

Problem size: dim(x) + S×dim(y) $\longrightarrow$ if S is large: decomposition!

Decomposition

Two-stage problems: **block angular** constraints



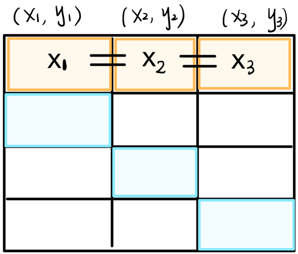
Benders decomposition

many variants: single cut vs. multicut

stochastic decomposition

regularized decomposition

...



Lagrangian relaxation

(dual decomposition)

Subgradient Method

Progressive Hedging

...

Benders decomposition

In two-stage LP

$$\underset{x \in X}{\text{minimize}} \ (c^0)^\top x + \mathbb{E}_{\omega \sim \mathbb{P}} [\, \psi(x; \omega) \,],$$

where $\psi(x; \omega)$ is the second-stage recourse:

$$\begin{aligned} \psi(x;\omega) \triangleq \underset{y}{\text{minimum}} \quad & c(\omega)^\top y \\ \text{subject to} \quad & T(\omega) x + W(\omega) y = h(\omega). \end{aligned}$$

Benders decomposition

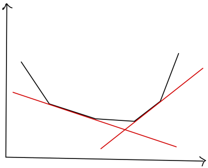
In two-stage LP

$$\underset{x \in X}{\text{minimize}} \;\; (c^0)^\top x + \mathbb{E}_{\omega \sim \mathbb{P}} \left[\psi(x;\omega) \right],$$

where $\psi(x;\omega)$ is the second-stage recourse:

$$\begin{aligned} \psi(x;\omega) \triangleq \; & \underset{y}{\text{minimum}} && c(\omega)^\top y \\ & \text{subject to} && T(\omega)x + W(\omega)y = h(\omega). \end{aligned}$$

$\psi(\bullet\,;\omega)$ is convex piecewise affine!



For each scenario ω, one can get a linear lower approximation of ψ at a given x

Nonconvex recourse

What if x appears in the objective function?

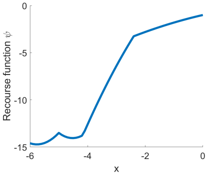
$$\underset{x \in X}{\text{minimize}} \ (c^0)^\top x + \mathbb{E}_{\omega \sim \mathbb{P}}\left[\, \psi(x;\omega) \,\right],$$

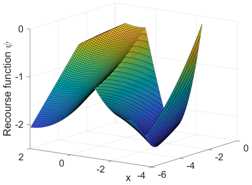
where $\psi(x;\omega)$ is the second-stage recourse:

$$\begin{array}{rcll} \psi(x;\omega) & = & \underset{y}{\text{minimum}} & x^\top D(\omega) y \\ & & \text{subject to} & T(\omega) x + W(\omega) y = h(\omega) \end{array}$$

Applications in decision-dependent uncertainty, stochastic interdiction problem... [Liu, Cui, Pang and Sen, 2020]

The function $\psi(\bullet;\omega)$ is no longer convex!





Nonconvex recourse

In general, we consider

$$\underset{x \in X}{\text{minimize}} \ \varphi(x) + \mathbb{E}_{\omega \sim \mathbb{P}}\left[\, \psi(x;\omega) \,\right],$$

where $\psi(x;\omega)$ is the second-stage recourse:

$$\begin{aligned} \psi(x;\omega) \triangleq \ & \underset{y}{\text{minimum}} & & f(x,y;\omega) \\ & \text{subject to} & & T(\omega)x + W(\omega)y = h(\omega), \ \ G(x,y;\omega) \leq 0. \end{aligned}$$

- φ convex on a closed convex set X
- $f(\bullet, \bullet; \omega)$ concave-convex
- $G(\bullet, \bullet; \omega)$ jointly convex
- all functions can be nonsmooth

Q: How to generalize the Benders decomposition to solve such problems?

Recourse functions: absences of Clarke Regularity

A locally Lipschitz continuous function f is said to be Clarke regular at $\bar{x}$ if it is directionally differentiable and

$$f'(\bar{x}; d) \triangleq \lim_{t \downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} = \limsup_{x \to \bar{x}, t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$

Recourse functions: absences of Clarke Regularity

A locally Lipschitz continuous function f is said to be Clarke regular at $\bar{x}$ if it is directionally differentiable and

$$f'(\bar{x}; d) \triangleq \lim_{t \downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} = \limsup_{x \to \bar{x}, t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$

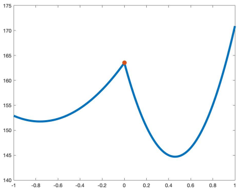
Examples: (weakly) convex functions, smooth functions, convex $\circ$ smooth functions...

Recourse functions: absences of Clarke Regularity

A locally Lipschitz continuous function f is said to be Clarke regular at $\bar{x}$ if it is directionally differentiable and

$$f'(\bar{x}; d) \triangleq \lim_{t \downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} = \limsup_{x \to \bar{x}, t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$

downward cusp



inward corner

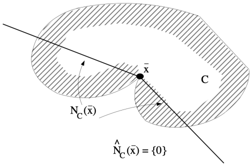


figure from [Rockafellar and Wets 09]

Absences of Clarke regularity

What can we do?

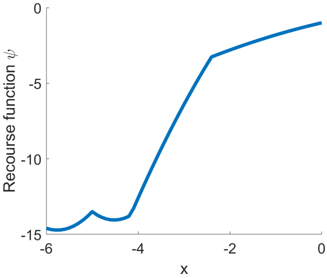
Back to the recourse function

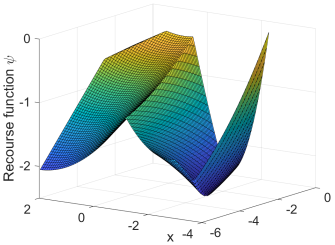
$$\begin{aligned}\psi(x;\omega) \triangleq \underset{y}{\text{minimum}} \quad & x^\top D(\omega) y \\ \text{subject to} \quad & T(\omega)x + W(\omega)y = h(\omega)\end{aligned}$$

What can we do?

Back to the recourse function

$$\psi(x;\omega) \triangleq \underset{y}{\text{minimum}} \quad x^\top D(\omega) y$$
$$\text{subject to} \quad T(\omega)x + W(\omega)y = h(\omega)$$





What can we do?

Back to the recourse function

$$\psi(x;\omega) \triangleq \underset{y}{\text{minimum}} \quad x^\top D(\omega) y \\ \text{subject to} \quad T(\omega)x + W(\omega)y = h(\omega)$$

Theoretically, this is a (piecewise linear-quadratic) difference-of-convex function
[Nouiehed, Pang and Razaviyayn, 2019]

However, its dc decomposition is tedious...

What can we do?

Recall the sensitivity analysis of linear programs. Let us fix $\bar{x}$.

Perturbed constraints: ψ is convex piecewise affine

$$\psi_{\mathrm{cvx}}(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & \bar{x}^\top D y \\ \text{subject to} & Tx + Wy = h \end{array}\right]$$

Perturbed objective: ψ concave piecewise affine

$$\psi_{\mathrm{cve}}(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & x^\top D y \\ \text{subject to} & T\bar{x} + Wy = h \end{array}\right]$$

What can we do?

Recall the sensitivity analysis of linear programs. Let us fix $\bar{x}$.

Perturbed constraints: ψ is convex piecewise affine

$$\psi_{\text{cvx}}(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & \bar{x}^\top D y \\ \text{subject to} & Tx + Wy = h \end{array} \right]$$

Perturbed objective: ψ concave piecewise affine

$$\psi_{\text{cve}}(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & x^\top D y \\ \text{subject to} & T\bar{x} + Wy = h \end{array} \right]$$

Joint perturbations:

$$\psi(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & x^\top D y \\ \text{subject to} & Tx + Wy = h \end{array} \right]$$

Variational analysis of optimal value functions can be found in [Bonnans and Shapiro, 2000]

Lifting: implicit convexity-concavity

$$\psi(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & x^\top D y \\ \text{subject to} & Tx + Wy = h \end{array} \right]$$

$$\Downarrow$$

$$\overline{\psi}(x, z) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & z^\top D y \\ \text{subject to} & Tx + Wy = h \end{array} \right]$$

Lifting: implicit convexity-concavity

$$\psi(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & x^\top D y \\ \text{subject to} & Tx + Wy = h \end{array} \right]$$

⇩

$$\overline{\psi}(x, z) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & z^\top D y \\ \text{subject to} & Tx + Wy = h \end{array} \right]$$

⇨ convex-concave

Lifting: implicit convexity-concavity

$$\psi(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & x^\top D y \\ \text{subject to} & Tx + Wy = h \end{array}\right]$$

implicitly convex-concave

$$\overline{\psi}(x, z) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & z^\top D y \\ \text{subject to} & Tx + Wy = h \end{array}\right]$$

convex-concave

Lifting: implicit convexity-concavity

$$\psi(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & x^\top D y \\ \text{subject to} & Tx + Wy = h \end{array}\right]$$

implicitly convex-concave

$$\overline{\psi}(x, z) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & z^\top D y \\ \text{subject to} & Tx + Wy = h \end{array}\right]$$

convex-concave

This implicitly convex-concave property holds for a broad class of optimal value functions [Cui and Pang, 2021]

Lifting: implicit convexity-concavity

$$\psi(x) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & x^\top D y \\ \text{subject to} & Tx + Wy = h \end{array}\right]$$

implicitly convex-concave

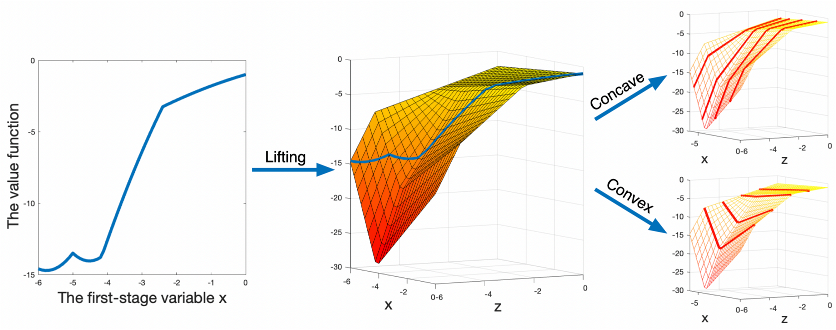
$$\overline{\psi}(x,z) \triangleq \left[\begin{array}{ll} \underset{y}{\text{minimum}} & z^\top D y \\ \text{subject to} & Tx + Wy = h \end{array}\right]$$

convex-concave

Note: A difference-of-convex function is explicitly convex-concave:

$$\underbrace{f(x)}_{\text{convex}} - \underbrace{g(x)}_{\text{convex}} \dashrightarrow \underbrace{f(x) - g(z)}_{\text{convex-concave}}$$

Lifting: implicit convexity-concavity



Implicitly convex-concave in $\mathbb{R}$ | convex-concave in $\mathbb{R}^2$ | convex/concave in $\mathbb{R}$

Implicitly convex-concave: surrogations

Moreau envelope (f may not be convex)

$$e_\gamma^{\mathrm{ori}} f(x) \triangleq \inf_y \left\{ f(y) + \frac{1}{2\gamma}\|x-y\|^2 \right\}.$$

Implicitly convex-concave: surrogations

Moreau envelope (f may not be convex)

$$e_\gamma^{\mathrm{ori}} f(x) \triangleq \inf_y \left\{ f(y) + \frac{1}{2\gamma}\|x-y\|^2 \right\}.$$

can always be decomposed into difference-of-convex functions

$$e_\gamma^{\mathrm{ori}} f(x) = \frac{1}{2\gamma}\|x\|^2 - \underbrace{\sup_y \left\{ -f(y) - \frac{1}{2\gamma}\|y\|^2 + \frac{1}{\gamma} y^\top x \right\}}_{\text{convex in } x \text{ even if } f \text{ is nonconvex}}$$

Implicitly convex-concave: surrogations

Moreau envelope (f may not be convex)

$$e_\gamma^{\text{ori}} f(x) \triangleq \inf_y \left\{ f(y) + \frac{1}{2\gamma}\|x-y\|^2 \right\}.$$

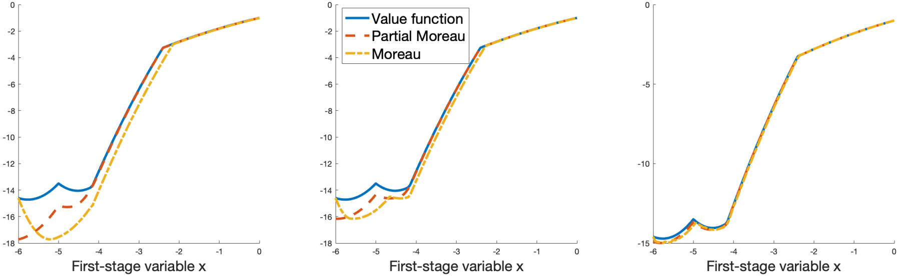
can always be decomposed into difference-of-convex functions

$$e_\gamma^{\text{ori}} f(x) = \frac{1}{2\gamma}\|x\|^2 - \underbrace{\sup_y \left\{ -f(y) - \frac{1}{2\gamma}\|y\|^2 + \frac{1}{\gamma} y^\top x \right\}}_{\text{not easy to compute if } f \text{ is not Clarke-regular!}}$$

Implicitly convex-concave: surrogations

Partial Moreau envelope for an implicitly convex-concave function ψ:

$$e_\gamma \psi(z) \triangleq \inf_x \left\{ \overline{\psi}(x, z) + \frac{1}{2\gamma} \|x - z\|^2 \right\},$$



Implicitly convex-concave: surrogations

Partial Moreau envelope for an implicitly convex-concave function ψ:

$$e_\gamma \psi(z) \triangleq \inf_x \left\{ \overline{\psi}(x,z) + \frac{1}{2\gamma} \|x - z\|^2 \right\},$$

Difference-of-convex decomposition:

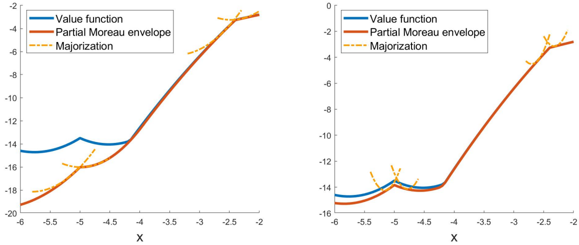
$$e_\gamma \psi(z) = \underbrace{\frac{1}{2\gamma}\|z\|^2}_{\text{strongly convex}} \underbrace{- \sup_x \left\{ -\overline{\psi}(x,z) - \frac{1}{2\gamma}\|x\|^2 + \frac{1}{\gamma} z^\top x \right\}}_{\text{denoted as } g_\gamma(z), \text{ convex}}.$$

The evaluation of the function value and subgradient of g_γ at x can be done by solving a convex problem

Implicitly convex-concave: surrogations

$$e_\gamma \psi(z) = \underbrace{\frac{1}{2\gamma}\|z\|^2}_{\text{strongly convex}} - \underbrace{\sup_x \left\{ -\overline{\psi}(x,z) - \frac{1}{2\gamma}\|x\|^2 + \frac{1}{\gamma} z^\top x \right\}}_{\text{denoted as } g_\gamma(z), \text{ convex}}$$

$$\leq \underbrace{\frac{1}{2\gamma}\|z\|^2 - \text{linearization of } g_\gamma \text{ at any point}}_{\text{strongly convex}}$$



Left: $\gamma = 0.15$

Right: $\gamma = 0.02$

A decomposition algorithm: fixed samples

Fixed S scenarios $\{(D^s, T^s, W^s, h^s)\}_{s=1}^{S}$

$$\underset{x \in X}{\text{minimize}} \quad \text{first-stage obj} \; + \; \frac{1}{S} \sum_{s=1}^{n} \psi_s(x),$$

where the s-th recourse is

$$\begin{aligned} \psi_s(x) \triangleq \underset{y}{\text{minimum}} \quad & x^\top D^s y \\ \text{subject to} \quad & T^s x + W^s y = h^s \end{aligned}$$

A decomposition algorithm: fixed samples

Fixed S scenarios $\{(D^s, T^s, W^s, h^s)\}_{s=1}^S$

$$\underset{x \in X}{\text{minimize}} \quad \text{first-stage obj} \; + \; \frac{1}{S}\sum_{s=1}^{n} \psi_s(x),$$

where the s-th recourse is

$$\psi_s(x) \triangleq \underset{y}{\text{minimum}} \quad x^\top D^s y \qquad \text{subject to} \quad T^s x + W^s y = h^s$$

Master problem:

$$x^{k+1} = \underset{x \in X}{\text{argmin}} \left[\text{first-stage obj} \; + \; \frac{1}{S}\sum_{s=1}^{n} \widehat{e}_\gamma \psi_s(x; x^k) \right]$$

Subproblem: for each scenario s, solve y^s of the recourse problem at $x = x^{k+1}$ to get the next surrogation $\widehat{e}_\gamma \psi_s(x; x^k)$
(decomposable over different scenarios)

Need an outer loop to update $\gamma \downarrow 0$

A decomposition algorithm: fixed samples

Theorem: Under technical conditions,

(a) any accumulation point is a (properly-defined) stationary point;

(b) if $\sum_{k \geq 0} \gamma_k < +\infty$, then the objective value sequence converges.

A decomposition algorithm: fixed samples

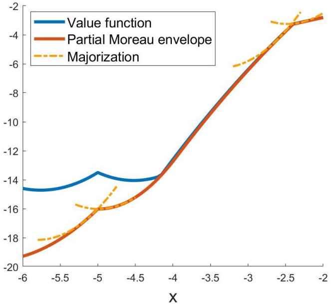
Theorem: Under technical conditions,

(a) any accumulation point is a (properly-defined) stationary point;

(b) if $\sum_{k \geq 0} \gamma_k < +\infty$, then the objective value sequence converges.

A technical note:

the surrogation is neither an upper bound nor a lower bound of the original recourse



Decomposition with sampling

For each step, we can also sample a batch of scenarios $\{(D^s, T^s, W^s, h^s)\}_{s \in S_k}$

Master problem:

$$x^{k+1} = \underset{x \in X}{\operatorname{argmin}} \left[\text{first-stage obj} \;+\; \frac{1}{|S_k|} \sum_{s \in S_k} \widehat{e}_\gamma \psi_s(x; x^k) \right]$$

Subproblem: for each sampled scenario $s \in S_k$, solve y^s at $x = x^{k+1}$ to get $\widehat{e}_\gamma \psi_s(x; x^k)$

Gradually add samples to the master problem

Decomposition with sampling

For each step, we can also sample a batch of scenarios $\{(D^s, T^s, W^s, h^s)\}_{s \in S_k}$

Master problem:

$$x^{k+1} = \underset{x \in X}{\text{argmin}} \left[\text{first-stage obj} \;\; + \;\; \frac{1}{|S_k|} \sum_{s \in S_k} \widehat{e}_\gamma \psi_s(x; x^k) \right]$$

Subproblem: for each sampled scenario $s \in S_k$, solve y^s at $x = x^{k+1}$ to get $\widehat{e}_\gamma \psi_s(x; x^k)$

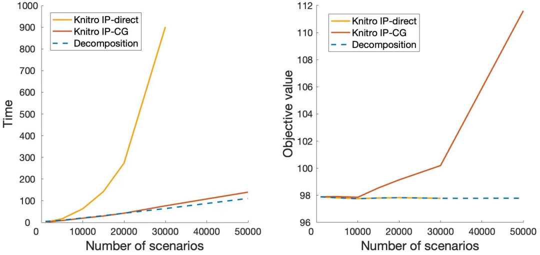
Gradually add samples to the master problem

Sample-size requirement:

$$\sum_{\nu=1}^{\infty} \frac{S_{k+1} - S_k}{S_{k+1} \, (S_k)^\eta} < \infty \quad \text{for some } \eta \in (0, 1/2).$$

Numerical experiments

compared with the general purpose nonlinear programming solver Knitro



1st stage: $x \in \mathbb{R}^{10}$

22 inequality constraints

2nd stage: $y^s \in \mathbb{R}^{40}$

93 inequality constraints

Algorithms stop if

KKT residual $\leq 10^{-4}$

Sizes of the deterministic equivalent problems:

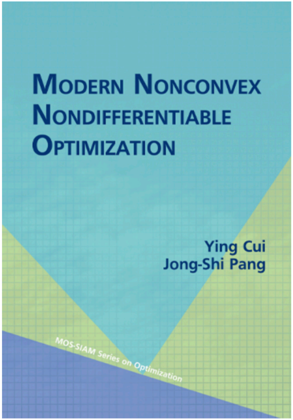
# of scenarios	10^4	5×10^4
# of variables (1st+2nd stages)	$400,010$	$2,000,010$
# of constraints	$850,030$	$4,250,030$

Advertisement

Our new monograph (2021)

Modern Nonconvex Nondifferentiable Optimization

Available at SIAM Bookstore



Thank You!