Distributionally Robust Optimization with Decision-Dependent Ambiguity Set

Nilay Noyan

Sabancı University, Istanbul, Turkey

Joint work with

G. Rudolf, Koç University

M. Lejeune, George Washington University

Uncertainty in optimization

- Stochastic programming represents uncertain parameters by a random vector - a classical stochastic optimization:

$$\min_{\mathbf{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}}\left(G(\mathbf{x},\boldsymbol{\xi})\right)$$

- Classical assumptions in stochastic programming:
  - The probability distribution of the random parameter vector is independent of decisions - exogenously given $\longrightarrow$ relaxing it requires addressing endogenous uncertainty.
  - The "true" probability distribution of the random parameter vector is known $\longrightarrow$ relaxing it requires addressing distributional uncertainty.

Endogenous uncertainty

Sabancı Üniversitesi

- The underlying probability space may depend on the decisions:

$$\min_{\mathbf{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}(\mathbf{x})}\left(G(\mathbf{x},\boldsymbol{\xi}(\mathbf{x}))\right)$$

- Decisions can affect the likelihood of underlying random future events.
  - Example. Pre-disaster planning – strengthening/retrofitting transportation links can reduce failure probabilities in case of a disaster (Peeta et al., 2010).

- Decisions can affect the possible realizations of the random parameters.
  - Example. Machine scheduling - stochastic processing times can be compressed by control decisions (Shabtay and Steiner, 2007).

Endogenous uncertainty

Sabancı Üniversitesi

- Its use in stochastic programming remains a tough endeavor, and is far from being a well-resolved issue (Dupacova, 2006; Hellemo et al., 2018).

- Mainly two types of optimization problems (Goel and Grossmann, 2006):
  - decision-dependent information revelation
  - decision-dependent probabilities (literature is very sparse) $\longrightarrow$ our focus

- Stochastic programs with decision-dependent probability measures
  - Straightforward modeling approach expresses probabilities as non-linear functions of decision variables and leads to *highly non-linear models*.
  - A large part of the literature focuses on a particular stochastic *pre-disaster investment problem* (Peeta et al., 2010; Laumanns et al., 2014; Haus et al., 2017).
  - Existing algorithmic developments are mostly *specific to the problem structure*.

Distributional uncertainty

Sabancı Üniversitesi

- In practice, the "true" probability distribution of uncertain model parameters/data may *not be known*.
  - Access to limited information about the prob. distribution (e.g. samples).
  - Future might not be distributed like the past.
  - Solutions might be sensitive to the choice of the prob. distribution.

- Distributionally robust optimization (DRO) is an appreciated approach (e.g., Goh and Sim, 2010; Wiesemann et al., 2014, Jiang and Guan, 2015).
  - Considers a set of probability distributions (*ambiguity set*).
  - Determines decisions that provide hedging against the worst-case distribution by solving a minimax type problem.
  - An intermediate approach between *stochastic programming* and *traditional robust optimization.*

DRO - Choice of ambiguity set

- Moment-based versus statistical distance-based ambiguity sets
  - Exact moment-based sets typically do not contain the true distribution.
  - Conservative solutions: very different distributions can have the same lower moments and the use of higher moments can be impractical.

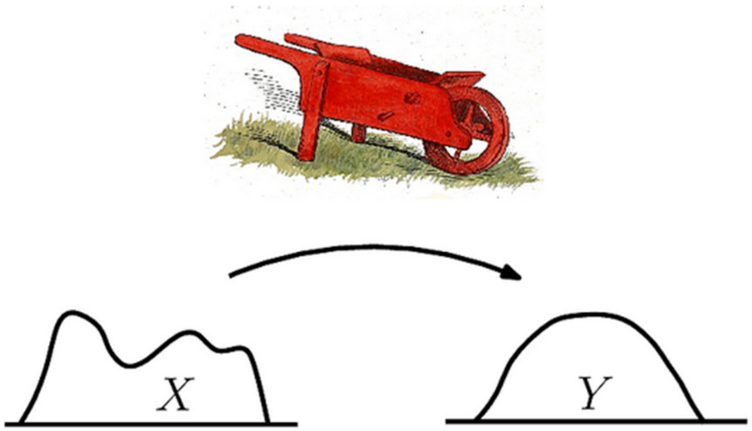
- Choice of statistical distance: (Bayraksan and Love, 2015; Rubner et al. 1998)

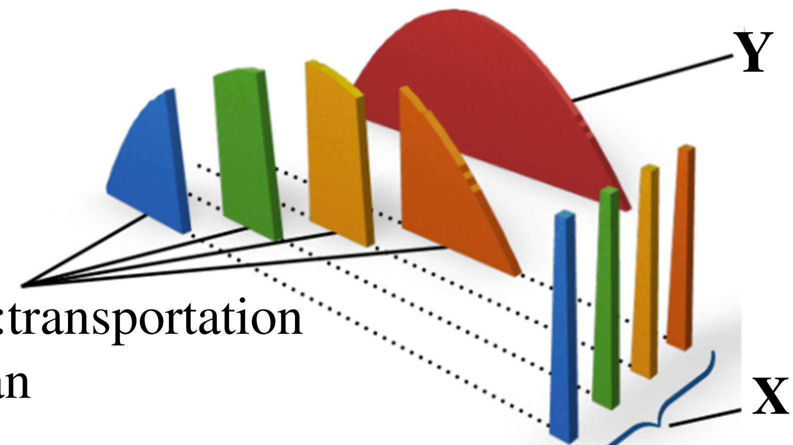
  Two of the more common ones: *Phi-divergence* versus *Earth Mover's Distances*

  - Divergence distances do not capture the metric structure of realization space.
  - In some cases, phi-divergences limit the support of the measures in the set.
  - Our particular focus - *Wasserstein distance* with the desirable properties:
    - Consistency, tractability, etc.

A general class of Earth Mover's Distances (EMDs)

Sabancı Universitesi





$$\Delta\left([\mathbb{P},\boldsymbol{\xi}_1],[\mathbb{Q},\boldsymbol{\xi}_2]\right)=\inf\left\{\int_{\Omega_1\times\Omega_2}\delta\left(\boldsymbol{\xi}(\omega_1),\boldsymbol{\xi}(\omega_2)\right)\,\mathbb{P}^*(\mathrm{d}\omega_1,\mathrm{d}\omega_2)\ :\ \begin{array}{c}\mathbb{P}^*\in\mathcal{P}(\Omega_1\times\Omega_2,\mathcal{A}_1\times\mathcal{A}_2),\\ \Pi_1(\mathbb{P}^*)=\mathbb{P},\ \Pi_2(\mathbb{P}^*)=\mathbb{Q}\end{array}\right\}$$

- In a pair $[\mathbb{P},\boldsymbol{\xi}]\in\mathcal{V}^m(\Omega,\mathcal{A})$, $\boldsymbol{\xi}:\Omega\to\mathbb{R}^m$ is a rand. var. on the prob. space $(\Omega,\mathcal{A},\mathbb{P})$
- δ: a measure of dissimilarity (or distance) between real vectors (*transportation cost*)
- For any two measurable spaces $(\Omega_1,\mathcal{A}_1)$ and $(\Omega_2,\mathcal{A}_2)$, the function δ induces an EMD
- Minimum-cost transportation plan

A general class of Earth Mover's Distances

Sabancı Üniversitesi

- Transportation problem – discrete case: $\delta^{ij} = \delta\left(\boldsymbol{\xi}(\omega^i), \boldsymbol{\xi}(\omega^j)\right)$ for $i, j \in [n]$

$$\min_{\boldsymbol{\gamma} \in \mathbb{R}_+^{n \times n}} \left\{ \sum_{i \in [n]} \sum_{j \in [n]} \delta^{ij} \gamma^{ij} \; : \; \sum_{j \in [n]} \gamma^{ij} = p^i \quad \forall\, i \in [n], \sum_{i \in [n]} \gamma^{ij} = q^j \quad \forall\, j \in [n] \right\}$$

- Wasserstein-*p* metric: $W_p\left([\mathbb{P}_1, \boldsymbol{\xi}_1], [\mathbb{P}_2, \boldsymbol{\xi}_2]\right) = \Delta^p\left([\mathbb{P}_1, \boldsymbol{\xi}_1], [\mathbb{P}_2, \boldsymbol{\xi}_2]\right)^{\frac{1}{p}},$

  where Δ^p is the EMD induced by $\delta^p(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_p^p$

- Total variation distance (also a phi-divergence distance); the EMD induced by the discrete metric

$$\delta(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} 0 & \text{if } \mathbf{x}_1 = \mathbf{x}_2 \\ 1 & \text{if } \mathbf{x}_1 \neq \mathbf{x}_2. \end{cases}$$

DRO - Decision-dependent ambiguity set

Sabancı Üniversitesi

- Incorporate distributional uncertainty into decision problems via EMD balls centered on a *nominal random vector*

$$[\mathbb{P}, \boldsymbol{\xi}] \in \mathcal{V}^m(\Omega, \mathcal{A}) = \mathcal{P}(\Omega, \mathcal{A}) \times \mathcal{L}^m(\Omega, \mathcal{A})$$

- Continuous EMD ball: ambiguity *both in probability measure and realizations*

$$\mathcal{B}_{\delta,\kappa}\left([\mathbb{P}, \boldsymbol{\xi}]\right) = \{\boldsymbol{\zeta} \in \mathcal{L}^m([0,1], \mathcal{A}_B) \ : \ \Delta\left([\mathbb{P}, \boldsymbol{\xi}], [\mathbb{B}, \boldsymbol{\zeta}]\right) \leq \kappa\}$$

- Discrete EMD ball: the probability measure can change while the realization mapping $\boldsymbol{\xi}$ is fixed

$$\mathcal{B}^{\boldsymbol{\xi}}_{\delta,\kappa}(\mathbb{P}) = \{\mathbb{Q} \in \mathcal{P}(\Omega, \mathcal{A}) \ : \ \Delta\left([\mathbb{P}, \boldsymbol{\xi}], [\mathbb{Q}, \boldsymbol{\xi}]\right) \leq \kappa\}$$

DRO with decision-dependent ambiguity set

Sabancı Üniversitesi

Continuous EMD ball case:

$$\min_{\mathbf{x}\in\mathcal{X}} \sup_{\boldsymbol{\zeta}\in\mathcal{B}_{\delta,\kappa}([\mathbb{P}(\mathbf{x}),\boldsymbol{\xi}(\mathbf{x})])} \mathbb{E}_{\mathbb{B}}\left(G(\mathbf{x},\boldsymbol{\zeta})\right)$$

Discrete EMD ball case:

$$\min_{\mathbf{x}\in\mathcal{X}} \sup_{\mathbb{Q}\in\mathcal{B}^{\boldsymbol{\xi}(\mathbf{x})}_{\delta,\kappa}(\mathbb{P}(\mathbf{x}))} \mathbb{E}_{\mathbb{Q}}\left(G(\mathbf{x},\boldsymbol{\xi}(\mathbf{x}))\right)$$

- DRO with Wasserstein distance has been receiving increasing attention
  - See, e.g., Pflug and Wozabal, 2007; Zhao and Guan, 2015; Gao and Kleywegt, 2016; Esfahani and Kuhn, 2018; Luo and Mehrotra, 2017; Blanchet and Murthy, 2016.
- Using a decision-dependent ambiguity set: an almost untouched research area until recently
  - Zhang et al., 2016; Royset and Wets, 2017, Luo and Mehrotra, 2018.
- A very recent interest on a related concept in the context of robust optimization
  - Lappas and Gounaris, 2018, Nohadani and Sharma, 2018; using *decision-dependent uncertainty sets.*

Risk-averse variants

Continuous EMD ball case:

$$\min_{\mathbf{x}\in\mathcal{X}} \sup_{\boldsymbol{\zeta}\in\mathcal{B}_{\delta,\kappa}([\mathbb{P}(\mathbf{x}),\boldsymbol{\xi}(\mathbf{x})])} \rho\left(G(\mathbf{x},\boldsymbol{\zeta})\right)$$

Discrete EMD ball case:

$$\min_{\mathbf{x}\in\mathcal{X}} \sup_{\mathbb{Q}\in\mathcal{B}^{\boldsymbol{\xi}(\mathbf{x})}_{\delta,\kappa}(\mathbb{P}(\mathbf{x}))} \rho\left([\mathbb{Q}, G(\mathbf{x},\boldsymbol{\xi}(\mathbf{x}))]\right)$$

- Incorporating risk is crucial for *rarely occurring events* such as disasters.
- Law invariant coherent risk measures defined on a standard L_p space.
- Any such risk measure can be naturally extended to p-integrable random variables defined on an arbitrary probability space

$$\rho\left([\mathbb{P}, X]\right) = \rho(X) = \rho\left(F_X^{(-1)}\right)$$

- Our main focus: Conditional value-at-risk (Rockafellar and Uryasev, 2000).

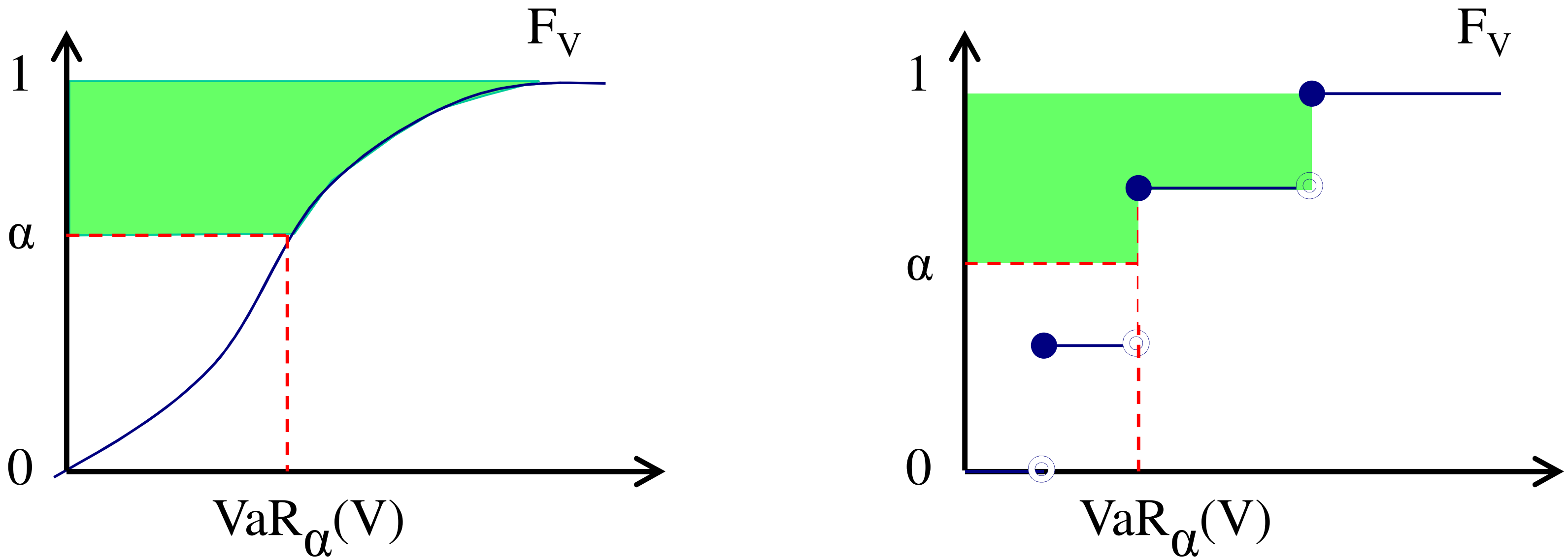
Theory of risk functionals

Sabancı Üniversitesi

- A risk functional ρ assigns to a random variable a scalar value, providing a direct way to define stochastic preference relations: $\rho(G(\mathbf{x}_1)) \leq \rho(G(\mathbf{x}_2))$
- Desirable properties of risk measures, such as law invariance and coherence, have been axiomized starting with the work of Artzner et al. (1999).
- Law invariance: Functionals that depend only on distributions of random vars.
- Coherence (smaller values of risk measures are preferred):
  - Monotonicity: $X \leq Y$ a.s. $\Rightarrow$ $\varrho(X) \leq \varrho(Y)$
  - Translation equivariance: $\varrho(X+\lambda) = \varrho(X) + \lambda$
  - Convexity: $\varrho(\lambda X + (1-\lambda)Y) \leq \lambda\varrho(X) + (1-\lambda)\varrho(Y)$ for $\lambda \in [0,1]$
  - Positive homogeneity: $\varrho(\lambda X) = \lambda\varrho(X)$ for $\lambda \geq 0$
- CVaR serves as a fundamental building block for other law invariant coherent risk measures (Kusuoka, 2001); supremum of convex combinations of CVaR at various confidence levels.

Conditional Value-at-Risk (CVaR)

Sabancı Universitesi



- Value-at-risk (α-quantile): $\mathrm{VaR}_{0.95}(V)$ is exceeded only with a small probability of at most 0.05.
- If unlucky (5% worst outcomes), the expected loss is $\mathrm{CVaR}_{0.95}(V)$ (shaded area).
- Alternative representations – Discrete case (v_i with prob p_i, $i \in [n]$):

$$\begin{aligned}\mathrm{CVaR}_\alpha(V) &= \min\left\{\eta + \frac{1}{1-\alpha}\mathbb{E}([V-\eta]_+), \eta \in \mathbb{R}\right\} \\ &= \max\left\{\frac{1}{1-\alpha}\sum_{i\in[n]} v_i\beta_i \;:\; \sum_{i\in[n]}\beta_i = 1-\alpha,\ 0 \le \beta_i \le p_i,\ \ \forall\, i \in [n]\right\} \\ &= \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_a(V)\,\mathrm{d}a \qquad \to \text{A weighted sum of the least favorable outcomes!}\end{aligned}$$

Formulations - Continuous EMD ball case

- Robustification of risk measures
  - Outcome mapping has a *bilinear structure:* $G(\mathbf{x},\boldsymbol{\zeta}) = \boldsymbol{\zeta}^\top \mathbf{v}(\mathbf{x}),\ \mathbf{v} : \mathcal{X} \to \mathbb{R}^m$
  - Law invariant convex risk measure $\rho : L_p \to \mathbb{R}$ is *well-behaved* with factor *C*.
  - Wasserstein-*p* ball of radius κ centered on a random vector $[\mathbb{B},\boldsymbol{\xi}]$
  - Key result of Pflug et al. (2012): $$\sup_{\boldsymbol{\zeta}\in\mathcal{B}_{\delta^p,\kappa^p}([\mathbb{B},\boldsymbol{\xi}])} \rho(\boldsymbol{\zeta}^\top\mathbf{v}) = \rho(\boldsymbol{\xi}^\top\mathbf{v}) + C\kappa\|\mathbf{v}\|_q$$

- Reformulation of the DRO problem under endogenous uncertainty:

$$\min_{\mathbf{x}\in\mathcal{X}} \rho\left(\boldsymbol{\xi}^\top(\mathbf{x})\mathbf{v}(\mathbf{x})\right) + C\kappa\|\mathbf{v}(\mathbf{x})\|_q$$

Formulations- Discrete EMD ball case

Sabancı Üniversitesi

Robustifying risk measures in finite spaces

- The closed-form in the continuous case is not valid.

  Example. Let $\boldsymbol{\xi}$ be a 2-dimensional random vector with possible realizations $(1,0)^\top$ and $(0,1)^\top$, and let $\mathbf{x} = (1,1)^\top$. For any probability distribution $\mathbb{Q}$:

$$\mathbb{E}_{\mathbb{Q}}(\mathbf{x}^\top \boldsymbol{\xi}) = 1 < 1 + \kappa \|\mathbf{x}\|_q$$

- Using LP duality, the supremum involved in robustification of certain risk measures can be replaced with an equivalent minimization.

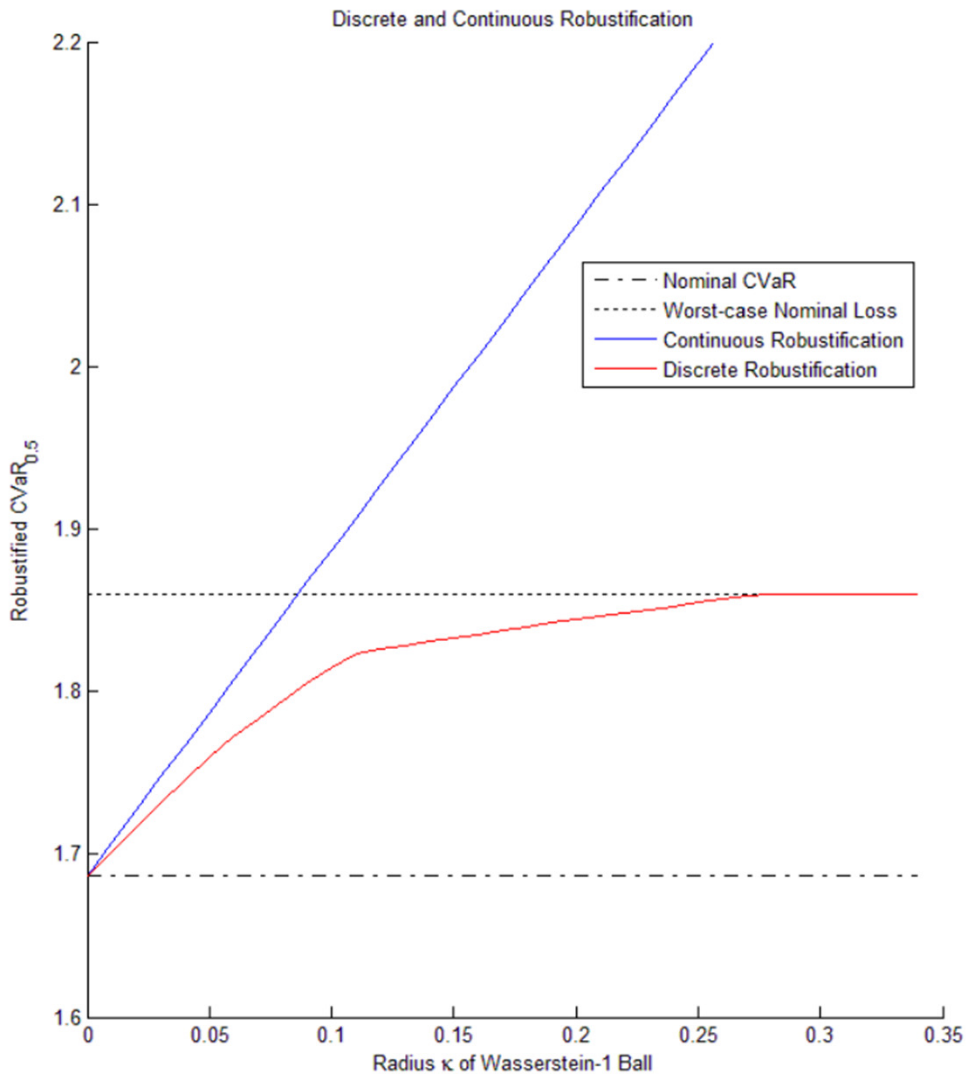
- The robustified CVaR value

$$\sup \left\{ \mathrm{CVaR}_\alpha ([\mathbb{Q}, Z]) \,:\, \mathbb{Q} \in \mathcal{B}^{\boldsymbol{\xi}}_{\delta,\kappa}(\mathbb{P}) \right\} = \begin{array}{ll} \min & \eta + \dfrac{1}{1-\alpha} \displaystyle\sum_{i \in [n]} p^i v^i + \dfrac{1}{1-\alpha} \kappa \tau \\ \text{s.t.} & v^i \geq z^j - \eta - \delta^{ij} \tau, \quad \forall i,j \in [n] \\ & \mathbf{v} \in \mathbb{R}^n_+, \ \tau \in \mathbb{R}_+. \end{array}$$

Robustification: continuous vs. discrete balls

Sabancı Üniversitesi

- A simple illustrative portfolio optimization with three equally weighted assets



- Nominal distribution:
  - Ten equally likely scenarios
  - Randomly generated losses
- Robustified $\mathrm{CVaR}_{0.5}$ of portfolio loss
  - Ambiguity set: Wasserstein-1 ball
  - Varying radius κ
- Continuous ball
  - Loss realizations are ambiguous
- Discrete ball
  - Loss realizations are fixed
  - Only probabilities are ambiguous

Formulations - Discrete EMD ball case

Sabancı Üniversitesi

- For ρ=CVaR$_\alpha$, minimax DRO problem as a conventional minimization:

$$
\begin{aligned}
\min \quad & \eta + \frac{1}{1-\alpha}\sum_{i\in[n]} p^i(\mathbf{x})v^i + \frac{1}{1-\alpha}\kappa\tau \\
\text{s.t.} \quad & v^i \geq G(\mathbf{x}, \boldsymbol{\xi}^j(\mathbf{x})) - \eta - \delta^{ij}\tau, && \forall i,j \in [n] \\
& \delta^{ij} = \delta\left(\boldsymbol{\xi}^i(\mathbf{x}), \boldsymbol{\xi}^j(\mathbf{x})\right), && \forall i,j \in [n] \\
& \mathbf{v} \in \mathbb{R}^n_+,\ \tau \in \mathbb{R}_+,\ \mathbf{x} \in \mathcal{X}.
\end{aligned}
$$

- Analogous, although more complex, formulations can be obtained for a *general class of coherent risk measures*
  - the family of risk measures with finite Kusuoka representations.

- Provide an overview of various settings leading to tractable formulations.

Tractable formulations - Discrete EMD ball

Sabancı Üniversitesi

$$
\begin{array}{lll}
\min & \eta + \dfrac{1}{1-\alpha}\displaystyle\sum_{i\in[n]} p^i(\mathbf{x})v^i + \dfrac{1}{1-\alpha}\kappa\tau & \\
\text{s.t.} & v^i \geq G^i - \eta & \forall i \in [n] \\
& v^i \geq \displaystyle\max_{j\in[n]} G^j - \eta - \tau & \forall i \in [n] \\
& \mathbf{v} \in \mathbb{R}^n_+,\ \tau \in \mathbb{R}_+,\ \mathbf{x} \in \mathcal{X}. &
\end{array}
$$

- Nominal realizations are decision-independent, and decision-dependent outcomes and scenario probabilities can be expressed via linear constraints
  - Quadratic program with linear constraints

- Both nominal realizations and outcomes are decision-independent
  - Using the discrete metric $\delta(\boldsymbol{\xi}^1, \boldsymbol{\xi}^2) = \begin{cases} 0 & \text{if } \boldsymbol{\xi}^1 = \boldsymbol{\xi}^2 \\ 1 & \text{if } \boldsymbol{\xi}^1 \neq \boldsymbol{\xi}^2 \end{cases}$
  - This metric allows to use *total variation distance-based balls* as ambiguity sets.
  - Still contains highly non-trivial instances of practical interest; pre-disaster planning (for strengthening a transportation network) and stochastic interdiction problems.

Tractable formulations - Discrete ball case

Sabancı Üniversitesi

$$
\begin{aligned}
\min \quad & \eta + \frac{1}{1-\alpha}\sum_{i\in[n]} p^i(\mathbf{x})v^i + \frac{1}{1-\alpha}\kappa\tau \\
s.t. \quad & v^i \geq G(\mathbf{x},\boldsymbol{\xi}^j(\mathbf{x})) - \eta - \sum_{k\in[m]} \nu_k^{ij}\,\tau, && \forall\, i\in[n],\ j\in[n] \\
& \nu_k^{ij} \leq \xi_k^i(\mathbf{x}) - \xi_k^j(\mathbf{x}) + M\lambda_k^{ij}, && \forall\, i\in[n],\ j\in[n],\ k\in[m] \quad (1) \\
& \nu_k^{ij} \leq -\xi_k^i(\mathbf{x}) + \xi_k^j(\mathbf{x}) + M(1-\lambda_k^{ij}), && \forall\, i\in[n],\ j\in[n],\ k\in[m] \quad (2) \\
& \boldsymbol{\lambda}\in\{0,1\}^{n\times n\times m}, \quad \boldsymbol{\nu}\in\mathbb{R}_+^{n\times n\times m}, \\
& \mathbf{v}\in\mathbb{R}_+^n,\ \tau\in\mathbb{R}_+,\ \mathbf{x}\in\mathcal{X},
\end{aligned}
$$

- Nominal realizations are decision-dependent, and the decision-dependent outcomes and scenario probabilities can be expressed via linear constraints
- Using the Wasserstein-1 metric: $\delta^{ij} = \left\|\boldsymbol{\xi}^i(\mathbf{x}) - \boldsymbol{\xi}^j(\mathbf{x}))\right\|_1 = \sum_{k\in[m]}\left|\xi_k^i(\mathbf{x}) - \xi_k^j(\mathbf{x})\right|$
  - Mixed-binary quadratic program with quadratic constraints
  - Make use of *comonotone structure* in the data to reduce the constraints of type (1)-(2), along with the corresponding binary and auxiliary variables.

Stochastic pre-disaster investment planning

Sabancı Üniversitesi

- Consider a transportation network where the links are subject to random failures in the event of a disaster.
  - each link is either operational or non-operational
  - the binary random variable: $\xi_l = 1$ (if link l survives) and $\xi_l = 0$ if it fails.

- Select the links to be strengthened to *reduce their failure probabilities*.
  - No strengthening: $x_l = 0$ and σ_l^0 : link survival prob.
  - Strengthening (with cost c_l): $x_l = 1$ and σ_l^1 : link survival prob.

- Decision-dependent probabilities:

$$[\mathbb{P}(\mathbf{x})](\{\xi_l = \xi_l^i\}) = \begin{cases} (1 - x_l)\sigma_l^0 + x_l\sigma_l^1 & \xi_l^i = 1 \\ (1 - x_l)(1 - \sigma_l^0) + x_l(1 - \sigma_l^1) & \xi_l^i = 0 \end{cases}$$

Stochastic pre-disaster investment planning

Sabancı Üniversitesi

- Improve post-disaster connectivity
  - Random outcome: weighted sum of shortest-path distances between a number of O-D pairs.

- Underlying risk-neutral stochastic program (Peeta et al. 2010):

$$\min_{\mathbf{x}\in\{0,1\}^L} \left\{ \sum_{i\in[n]} p^i(\mathbf{x}) \sum_{k\in[K]} w_k Q_k(\boldsymbol{\xi}^i) \ : \ \sum_{l\in[L]} c_l x_l \leq \hat{B} \right\}$$

- Solve a *shortest path* problem for each O-D pair and scenario

- Key challenge: expressing the decision-dependent scenario probabilities
- A straightforward approach results in highly non-linear functions of decision variables (under independence assumption):

$$p^i(\mathbf{x}) = \prod_{\ell\in[L]\,:\,\xi^i_\ell=1} \left[(1-x_\ell)\sigma^0_\ell + x_\ell\sigma^1_\ell\right] \prod_{\ell\in[L]\,:\,\xi^i_\ell=0} \left[(1-x_\ell)(1-\sigma^0_\ell) + x_\ell(1-\sigma^1_\ell)\right].$$

Stochastic pre-disaster investment planning

Sabancı Üniversitesi

- Benefit from an efficient characterization of decision-dependent scenario probabilities via a set of linear constraints (Laumanns et al. 2014)

- Our proposed risk-neutral or CVaR-based DRO-extension:

$$\min_{\mathbf{x}\in\mathcal{X}} \sup_{\mathbb{Q}\in\mathcal{B}^{\boldsymbol{\xi}}_{\delta,\kappa}(\mathbb{P}(\mathbf{x}))} \rho\left(\sum_{k\in[K]} w_k Q_k(\boldsymbol{\xi})\right)$$

- A natural choice of ambiguity set – total variation distance-based EMD ball using the discrete metric:

$$\delta(\boldsymbol{\xi}^1,\boldsymbol{\xi}^2) = \begin{cases} 0 & \text{if } \boldsymbol{\xi}^1 = \boldsymbol{\xi}^2 \\ 1 & \text{if } \boldsymbol{\xi}^1 \neq \boldsymbol{\xi}^2 \end{cases}$$

Stochastic pre-disaster investment planning

Sabancı Üniversitesi

- Reformulation: mixed-binary quadratic prog. with linear constraints

$$
\begin{aligned}
\min \quad & \sum_{i\in[n]} \pi_L^i v^i + \kappa\tau \\
\text{s.t.} \quad & v^i \geq G^i, \quad \forall i \in [n] \\
& v^i \geq \max_{j\in[n]} G^j - \tau, \qquad \forall i \in [n] \\
& \pi_\ell^i \leq \frac{\sigma_\ell^1}{\sigma_\ell^0}\pi_{\ell-1}^i + 1 - x_\ell, \qquad \forall\, \ell \in [L],\ i \in [n] \ :\ \xi_\ell^i = 1 \\
& \pi_\ell^i \leq \frac{1-\sigma_\ell^1}{1-\sigma_\ell^0}\pi_{\ell-1}^i + 1 - x_\ell, \qquad \forall\, \ell \in [L],\ i \in [n] \ :\ \xi_\ell^i = 0 \\
& \pi_\ell^i \leq \pi_{\ell-1}^i + x_\ell, \qquad \forall\, \ell \in [L],\ i \in [n] \\
& \sum_{i\in[n]} \pi_\ell^i = 1, \qquad \ell \in [L] \\
& \sum_{\ell\in[L]} c_\ell x_\ell \leq \hat{B}, \\
& \mathbf{v} \in \mathbb{R}^n, \quad \tau \in \mathbb{R}_+, \quad \mathbf{x} \in \{0,1\}^L, \quad \pi \in \mathbb{R}_+^{L\times n}
\end{aligned}
$$

Recursive distribution shaping $\pi_L^i = p^i(\mathbf{x})$

- Realizations $G^i = \sum_{k\in[K]} w_k Q_k(\boldsymbol{\xi}^i),\ i \in [n]$; Baseline Probs.: $\pi_0^i = \prod_{\ell:\xi_\ell^i=0}(1-\sigma_\ell^0)\prod_{\ell:\xi_\ell^i=1}\sigma_\ell^0$

Robustification in finite spaces

Sabancı Universitesi

- Robustified expectation $\mathbb{E}^{\kappa}(Z) = \sup \left\{ \mathbb{E}_{\mathbb{Q}}(Z) \ : \ \mathbb{Q} \in \mathcal{B}^{\boldsymbol{\xi}}_{\delta,\kappa}(\mathbb{P}) \right\}$

$$\begin{aligned} \min \quad & \sum_{i\in[n]} p^i v^i + \kappa\tau \\ \text{s.t.} \quad & v^i \geq z^i, & \forall i \in [n] \\ & v^i \geq \sup(Z) - \tau, & \forall i \in [n] \\ & \tau \geq 0. \end{aligned}$$

- For the total variation distance

$\mathbb{E}^{\kappa}(Z) = \kappa \sup(Z) + (1-\kappa)\,\mathrm{CVaR}_{\kappa}(Z)$ (Jiang and Guan, Rahimian et al., 2018)

$$\begin{aligned} \min \quad & \kappa \sup(Z) + (1-\kappa)\left(\eta - \frac{1}{1-\kappa}\sum_{i\in[n]} p^i \hat{v}^i\right) \\ \text{s.t.} \quad & \hat{v}^i \geq z^i - \eta, & \forall i \in [n] \\ & \hat{v}^i \geq 0, & \forall i \in [n] \\ & \eta \leq \sup(Z). \end{aligned}$$

- The change of variables $\eta = \sup(Z) - \tau, \ \ \hat{v}^i = v^i + \tau - \sup(Z)$ for $i \in [n]$

Robustification in finite spaces

Sabancı Universitesi

- Robustified expectation $\mathbb{E}^{\kappa}(Z) = \sup\left\{\mathbb{E}_{\mathbb{Q}}(Z) \ : \ \mathbb{Q} \in \mathcal{B}^{\boldsymbol{\xi}}_{\delta,\kappa}(\mathbb{P})\right\}$

$$
\begin{aligned}
\min \quad & \sum_{i\in[n]} p^i v^i + \kappa\tau \\
\text{s.t.} \quad & v^i \geq z^i, && \forall i \in [n] \\
& v^i \geq \sup(Z) - \tau, && \forall i \in [n] \\
& \tau \geq 0.
\end{aligned}
$$

- Optimum can be attained when $\tau = \sup(Z) - \mathrm{VaR}_{\kappa}(Z)$
- $\mathrm{VaR}_{\kappa}(Z) = z^j$ for at least one j∈ [n]: $v^i = \sum_{j\in[n]} a^{ij}\beta^j$ with $a^{ij} := \max\{z^i, z^j\}$.

$$
\begin{aligned}
\min \quad & \sum_{i\in[n]} p^i \sum_{j\in[n]} a^{ij}\beta^j + \kappa(\sup(Z) - \sum_{j\in[n]} z^j\beta^j) \\
\text{s.t.} \quad & \sum_{j\in[n]} \beta^j = 1, \quad \boldsymbol{\beta} \in \{0,1\}^n.
\end{aligned}
$$

Stochastic pre-disaster investment planning

Sabancı Üniversitesi

- Reformulation: mixed-binary quadratic prog. with linear constraints

$$\begin{aligned}
\min \quad & \sum_{i\in[n]} \pi_L^i v^i + \kappa\tau \\
\text{s.t.} \quad & v^i \geq G^i, \quad \forall i \in [n] \\
& v^i \geq \max_{j\in[n]} G^j - \tau, \qquad \forall i \in [n] \\
& \text{Distribution shaping constraints} \\
& \mathbf{v} \in \mathbb{R}^n, \quad \tau \in \mathbb{R}_+, \quad \mathbf{x} \in \mathcal{X}
\end{aligned}$$

$$\pi_L^i = p^i(\mathbf{x}) \longleftarrow \text{Distribution shaping constraints}$$

- Towards an MIP formulation:

$$\begin{aligned}
\min_{\mathbf{x}\in X} \quad & \sum_{i\in[n]} \pi_L^i \sum_{j\in[n]} a^{ij}\beta^j + \kappa(\max_{j\in[n]} G^j - \sum_{j\in[n]} G^j\beta^j) \\
\text{s.t.} \quad & \text{Distribution shaping constraints} \\
& \sum_{j\in[n]} \beta^j = 1, \quad \boldsymbol{\beta} \in \{0,1\}^n.
\end{aligned}$$

- McCormick envelopes and reformulation-linearization technique (Sherali and Adams, 1994); convex hull of (Gupte, et al. 2017)

$$P := \{(\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\pi}_L) \in \mathbb{R}_+^{n\times n} \times \{0,1\}^n \times C \mid \mathbf{z} = \boldsymbol{\pi}_L \boldsymbol{\beta}^\top, \quad \sum_{j\in[n]} \beta_j = 1\}$$

Stochastic pre-disaster investment planning

Sabancı Üniversitesi

- Considering all the network configurations, the number of scenarios is impractically large: 2^L.

- For computational tractability: utilize *scenario bundling* techniques.

- Laumanns et al. (2014) and Haus et al. (2017) propose very effective scenario bundling approaches.
  - For example, 2^{30} scenarios is replaced by 223 bundles for 5 O-D pairs.

- In the DRO setting, bundling raises an important issue:
  - An EMD ball around the reduced version of the original distribution is not equivalent to considering the reduced versions of the distributions in the EMD ball around the original distribution.
  - We proved that for our choice of the discrete metric these two ambiguity sets are the same.

Stochastic single-machine scheduling

- L jobs with stochastic processing times;
  - machine breakdowns, inconsistency of the worker performance, changes in tool quality, variable setup times, etc.

- Find a non-preemptive job processing sequence before uncertain processing times are realized.

- Sequencing decision variables (linear ordering formulation):

$$\theta_{kl} = \begin{cases} 1, & \text{if task } k \text{ precedes task } l \\ 0, & \text{otherwise} \end{cases}$$

- The set $\mathcal{T}$ of feasible scheduling decisions:

$$\begin{aligned}
&\theta_{ll} = 1, && \forall l \in [L] \\
&\theta_{kl} + \theta_{lk} = 1, && \forall k,\ l \in [L] \ : \quad k < l \\
&\theta_{kl} + \theta_{lh} + \theta_{hk} \leq 2, && \forall k,\ l,\ h \in [L] \ : \ k < l < h \\
&\boldsymbol{\theta} \in \{0,1\}^{L \times L}
\end{aligned}$$

Controllable processing times

Sabancı Üniversitesi

- Processing times are stochastic and can be *affected by control decisions*.
- $\xi_l(\mathbf{u})$: random processing time of job $l \in$ [L] given control decision $\mathbf{u} \in \mathcal{U}$
- A variety of schemes can be used to control processing times (e.g., Shabtay and Steiner, 2007)

Control with discrete resources: a set of T control options for every job

- Set of feasible control decisions: $\mathcal{U} \subset \left\{\mathbf{u} \in \{0,1\}^{T\times L} : \sum_{t\in[T]} u_{tl} = 1 \ \forall l \in [L]\right\}$
- Option t for job l leads to a random processing time of $\hat{\xi}_{tl}$

$$\xi_l(\mathbf{u}) = \sum_{t\in[T]} \hat{\xi}_{tl} u_{tl}, \ \ l \in [L]$$

$$\xi_l(\mathbf{u}) = \hat{\xi}_l (1 - \sum_{t\in[T]} \hat{a}_{tl} u_{tl}) = \hat{\xi}_l \sum_{t\in[T]} a_{tl} u_{tl}, \ \ l \in [L]$$

Comonotonicity: $\xi_l^i(\mathbf{u}) \geq \xi_l^j(\mathbf{u})$ or $\xi_l^i(\mathbf{u}) \leq \xi_l^j(\mathbf{u})$ holds for all $\mathbf{u} \in \mathcal{U}$

Stochastic single-machine scheduling

Sabancı Üniversitesi

- Random outcome of interest: **total weighted completion time**

$$\sum_{l\in[L]} w_l \sum_{k\in[L]} \xi_k(\mathbf{u})\theta_{kl} = \sum_{k\in[L]} \sum_{l\in[L]} \xi_k(\mathbf{u})\theta_{kl} w_l = \boldsymbol{\xi}^\top(\mathbf{u})\Theta\mathbf{w}$$

- The risk-averse version of our stochastic scheduling problem:

$$\min_{(\boldsymbol{\theta},\mathbf{u})\in\mathcal{T}\times\mathcal{U}} \quad h(\mathbf{u}) + \rho\left(\boldsymbol{\xi}^\top(\mathbf{u})\Theta\mathbf{w}\right)$$

- The robustified risk-averse scheduling problem – discrete ball

$$\min_{(\boldsymbol{\theta},\mathbf{u})\in\mathcal{T}\times\mathcal{U}} h(\mathbf{u}) + \sup_{\mathbb{Q}\in\mathcal{B}_{\delta,\kappa}^{\boldsymbol{\xi}(\mathbf{u})}(\mathbb{P})} \rho\left(\boldsymbol{\xi}^\top(\mathbf{u})\Theta\mathbf{w}\right)$$

Stochastic single-machine scheduling

Sabancı Üniversitesi

- Reformulation (mixed-integer quadratic program):

$$
\begin{aligned}
\min \quad & h(\mathbf{u}) + \eta + \frac{1}{1-\alpha} \sum_{i \in [n]} p^i v^i + \frac{1}{1-\alpha} \kappa \tau \\
\text{s.t.} \quad & v^i \geq \sum_{l \in [L]} \sum_{k \in [L]} \sum_{t \in [T]} w_l \hat{\xi}^j_{tk} z_{tkl} - \eta - \sum_{l \in [L]} \nu^{ij}_l \tau, && \forall i, j \in [n] \\
& z_{tkl} \leq u_{tk}, && \forall t \in [T],\ k, l \in [L] \\
& z_{tkl} \leq \theta_{kl}, && \forall t \in [T],\ k, l \in [L] \\
& z_{tkl} \geq u_{tk} + \theta_{kl} - 1, && \forall t \in [T],\ k, l \in [L] \\
& \nu^{ij}_l \leq \xi^i_l(\mathbf{u}) - \xi^j_l(\mathbf{u}) + M \lambda^{ij}_l, && \forall i, j \in [n],\ l \in [L] \\
& \nu^{ij}_l \leq -\xi^i_l(\mathbf{u}) + \xi^j_l(\mathbf{u}) + M(1 - \lambda^{ij}_l), && \forall i, j \in [n],\ l \in [L] \\
& \boldsymbol{\lambda} \in \{0,1\}^{n \times n \times L}, \quad \boldsymbol{\nu} \in \mathbb{R}^{n \times n \times L}_+, \\
& (\boldsymbol{\theta}, \mathbf{u}) \in \mathcal{T} \times \mathcal{U}, \quad \mathbf{v} \in \mathbb{R}^n_+, \quad \tau \geq 0, \quad \mathbf{z} \in [0,1]^{T \times L \times L}.
\end{aligned}
$$

- Enhanced MIP formulations: Variable and constraint elimination, McCormick envelopes, and reformulation-linearization technique.

Computational performance

Sabancı Üniversitesi

$$\xi_l(\mathbf{u}) = \hat{\xi}_l \sum_{t \in [T]} a_{tl} u_{tl}, \;\; l \in [L]$$

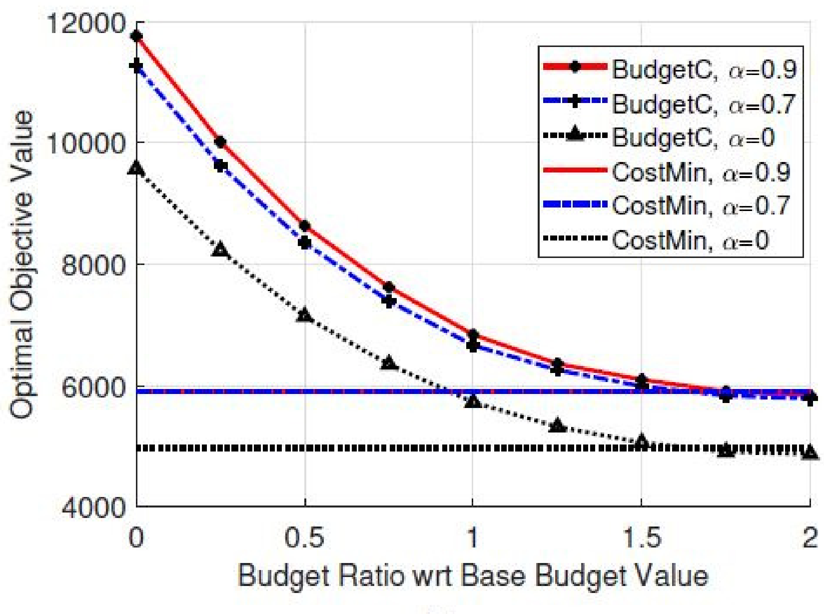
		Time [ROG (%)]			
		Cost Minimizing		Budget-Constrained	
L	n	CCM	CCM-RLT	CCM	CCM-RLT
10	50	8.4	7.4	8.9	7.4
	100	52.9	46.6	61.6	23.5
	150	271.3	137.0	226.2	175.5
	200	567.2	539.9	615.3	359.4
	250	1210.0	775.6	1626.2	451.8
	300	1002.6	1023.4	1251.2	2077.5
	400	1991.6	3155.5	1963.7	1742.2
	500	5160.7	†6046.3[0.13]	4496.1	†5783.11[0.45]
15	50	113.1	43.0	172.2	34.5
	100	603.3	190.0	627.8	155.3
	150	2513.8	555.4	2460.8	376.1
	200	5240.7	1221.9	5891.9	510.6
	250	†6737.8[3.29]	2659.1	††[4.47]	2135.5
	300	††[9.54]	3581.7	††[12.92]	2524.2
	400	–	††[1.34]	–	5013.1
	500	–	††[1.89]	–	††[10.91]
20	50	2318.6	62.4	3357.8	71.6
	100	††[7.89]	462.6	††[18.87]	523.0
	150	††[16.87]	1326.7	††[29.79]	1044.5
	200	–	2768.2	–	2697.5
	250	–	5411.8	–	5035.7
	300	–	††[1]	–	6057.3
	400	–	††[4.71]	–	††[3.23]

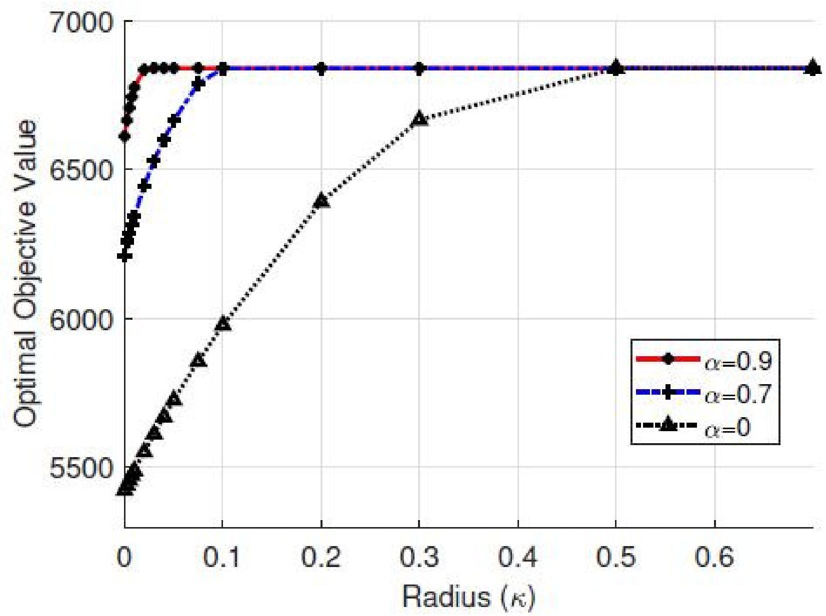
†: Each dagger sign indicates one instance hitting the time limit with an integer feasible solution.

Numerical Analysis

Sabancı Üniversitesi

- Optimal objective function value (robustified $CVaR_\alpha$ of TWCT) for varying radius and budget (L = 15 jobs and n = 100 scenarios)

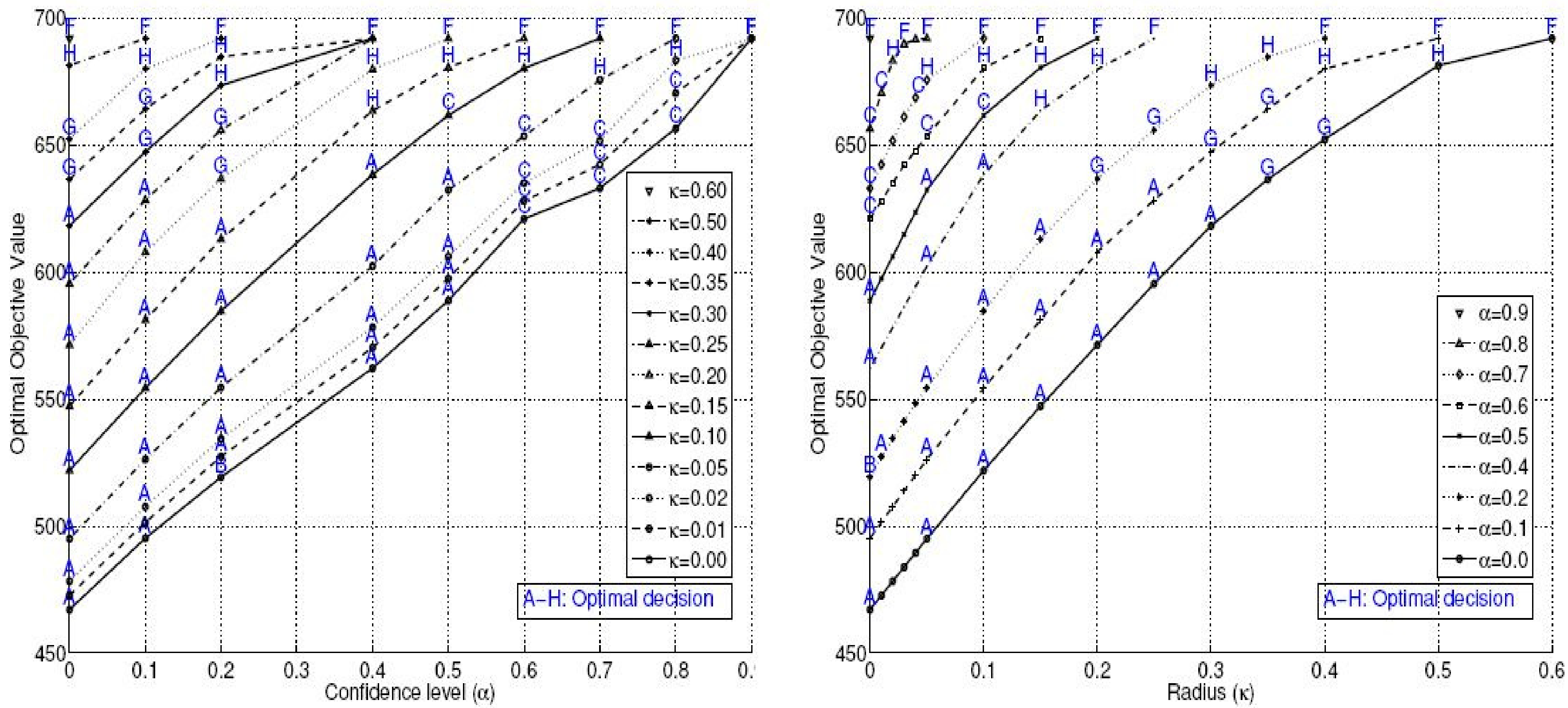




Optimal objective function values and solutions for a small illustrative example

Sabancı Üniversitesi

- Solution G is only optimal for high values of κ and low values of α, while, conversely, solution C is only optimal for lower κ and higher α values.
- Can express a range of risk-averse preferences that would not be possible to capture by either a "*purely robust*" or a "*purely CVaR-based*" approach.



Future avenues of research

Sabancı Üniversitesi

- Investigate meaningful and tractable characterizations of decision-dependent nominal parameter realizations and/or scenario probabilities for practical applications.

- While *scenario bundling* is a very effective method of reducing problem sizes, most EMDs are not compatible with this approach.
  - The total variation metric is a notable exception.
  - Other class of outcome-based scenario distances, which give rise to EMDs that can be used in conjunction with bundling?

- For problems of practical interest where bundling methods are not applicable, one might instead consider *sampling methods* to reduce the number of scenarios.
  - Appropriate sampling approaches?

Robustified risk measures in finite spaces

Sabancı Üniversitesi

$$\rho^{\kappa}(Z) = \sup\left\{\rho\left([\mathbb{Q}, Z]\right) \;:\; \mathbb{Q} \in \mathcal{B}^{\boldsymbol{\xi}}_{\delta,\kappa}(\mathbb{P})\right\} \quad \text{for } Z \in \mathcal{L}^1(\Omega, 2^{\Omega}).$$

- Replacing *the usual ordering* with a *parametric family of relations*, and introducing a corresponding "penalty term".

- **Definition.** The relation $\succeq_{\tau}$: $\quad V \succeq_{\tau} Z \quad \Leftrightarrow \quad v^i \geq z^j - \delta^{ij}\tau \quad \forall\, i, j \in [n].$

- **Robustified expectation:** $\quad \mathbb{E}^{\kappa}(Z) = \inf\left\{\mathbb{E}_{\mathbb{P}}(V) + \kappa\tau \;:\; \tau \geq 0,\; V \succeq_{\tau} Z\right\}$

- **Robustified CVaR:**

$$\mathrm{CVaR}^{\kappa}_{\alpha}(Z) = \inf\left\{\eta + \mathbb{E}_{\mathbb{P}}\left(\frac{1}{1-\alpha}S\right) + \frac{1}{1-\alpha}\kappa\tau \;:\; \eta \in \mathbb{R},\; \tau \geq 0,\; S \succeq_{\tau} [Z-\eta]_+\right\}$$

Robustified risk measures in finite spaces

Sabancı Üniversitesi

- CVaR serves as a fundamental building block for other law invariant coherent risk measures (Kusuoka, 2001)

- **Robustified mixed CVaR:**

$$\rho_{\{\mu\}}(Z) = \int_0^1 \mathrm{CVaR}_\alpha(Z)\,\mu(\mathrm{d}\alpha) = \sum_{\alpha \in \mathrm{supp}(\mu)} \mu(\{\alpha\})\,\mathrm{CVaR}_\alpha(Z) = \mathbb{E}_\mu\left(\mathrm{CVaR}_A([\mathbb{P}, Z])\right)$$

$$\begin{aligned} \rho^{\kappa}_{\{\mu\}}(Z) = \inf \Big\{ & \mathbb{E}_\mu(H) + \mathbb{E}_{\mathbb{P}}(S) + \kappa\tau \ : \\ & H \in \mathbb{R}^{[0,1)},\ \tau \geq 0,\ S \succeq_\tau \mathbb{E}_\mu\left(\frac{1}{1-A}[Z-H]_+\right)\Big\}. \end{aligned}$$

- **Robustified finitely representable risk measures:**

$$\rho_{\mathcal{M}}([\mathbb{P}, Z]) = \sup_{\mu \in \mathcal{M}} \rho_{\{\mu\}}([\mathbb{P}, Z])$$

$$\rho^{\kappa}_{\mathcal{M}}(Z) = \inf \left\{ R \in \mathbb{R} \ : \ H \in \mathbb{R}^{[0,1)},\ \boldsymbol{\tau} \in \mathbb{R}^{\mathcal{M}}_+, \ \begin{array}{l} S_\mu \succeq_{\tau_\mu} \mathbb{E}_\mu\left(\frac{1}{1-A}[Z-H]_+\right), \\ R \geq \mathbb{E}_\mu(H) + \mathbb{E}_{\mathbb{P}}(S_\mu) + \kappa\tau_\mu \end{array} \quad \forall \mu \in \mathcal{M} \right\}$$

Controllable processing times

Sabancı Universitesi

- Processing times are stochastic and can be *affected by control decisions*.

- $\xi_l(\mathbf{u}) \in \mathcal{L}^1(\Omega, \mathcal{A})$: random processing time of job l given decision $\mathbf{u} \in \mathcal{U}$
  - $\mathcal{U}$ **:** set of feasible control decisions
  - The mapping $\boldsymbol{\xi} : \mathcal{U} \to \mathcal{L}^L(\Omega, \mathcal{A})$ for an arbitrary prob. space $(\Omega, \mathcal{A}, \mathbb{P})$

- A wide variety of schemes can be used to control processing times
  - **Linearly compressible processing times** (e.g., Shabtay and Steiner, 2007)

$$\xi_l(\mathbf{u}) = \hat{\xi}_l - a_l u_l; \quad \text{a special case } \xi_l(\mathbf{u}) = \hat{\xi}_l(1 - u_l)$$

$$\mathcal{U} \subset \left\{ \mathbf{u} \in \mathbb{R}^L \ : \ 0 \leq u_l \leq \operatorname{ess\,inf} \frac{\hat{\xi}_l}{a_l} \quad \forall l \in [L] \right\}.$$

  - **Control with discrete resources** (later)

Computational performance

. . Sabancı . Üniversitesi

L	n	Time [ROG (%)] NCM	PCM	PCM-RLT	PCM-RLT CPLEX mipgap=2%
10	50	3036.7	151.6	149.5	138.8[0.84]
	100	††[13.87]	1361.3	4050.3	1618.8[0.93]
	150	††[38.88]	††[0.78]	††[3.14]	–
	200	–	††[23.15]	††[10.41]	–
	250	–	††[10.10]	††[4.03]	–
15	50	††[0.59]	1375.8	398.5	40.9[1.23]
	100	††[25.25]	††[4.14]	††[0.42]	1403.4[1.9]
	150	††[76.41]	††[22.66]	††[2.27]	†4538.4[2.48]
	200	–	–	††[2.60]	–
20	50	††[18.9]	4570.5	1405.3	105.5[0.68]
	100	††[65.81]	††[23.54]	††[1.28]	2598.2[1.71]
	150	–	††[43.5]	††[2.21]	†5140.6[2.28]
25	50	††[42.27]	††[19.04]	1670.8	161.7[1.95]
	100	††[79.21]	††[43.98]	††[0.52]	1156.6[1.11]
	150	–	–	††[2.69]	–
30	50	††[57.81]	††[36.44]	4501.7	234[0.54]
	100	–	–	††[1.37]	3220.5[1.78]

†: Each dagger sign indicates one instance hitting the time limit with an integer feasible solution.

Computational performance

Sabancı Üniversitesi

Impact of modeling parameters on performance of CCM-RLT

		Time [ROG (%)]				
		Cost Minimizing (α)		Budget-Constrained (BR, α)		
L	n	**0.7**	**0.9**	**(1, 0.7)**	**(1, 0.9)**	**(0.3, 0.9)**
15	50	37.8	43.0	31.7	34.5	17.3
	100	179.1	190.0	100.4	155.3	79.9
	150	526.0	555.4	367.8	376.1	274.4
	200	902.2	1221.9	814.6	510.6	540.5
	250	3155.5	2659.1	1909.5	2135.5	1132.5
	300	2921.2	3581.7	2640.6	2524.2	2815.5
	400	†† [0.95]	†† [1.35]	† 5469.5[1.78]	5013.1	3457.6
20	50	80.5	62.4	62.1	71.6	49.4
	100	367.7	462.6	337.3	523.0	260.6
	150	972.1	1326.7	833.2	1044.5	592.2
	200	3146.0	2768.2	1742.4	2697.5	1226.6
	250	5459.7	5411.8	3172.8	5035.7	2408.3
	300	4820.1	†† [1]	† 5973.3[1.28]	6057.3	† 5080.345[0.53]

†: Each dagger sign indicates one instance hitting the time limit with an integer feasible solution.