Distributionally Robust Optimization with Principal Component Analysis

Jianqiang Cheng, Richard L.-Y. Chen, Habib Najm, Ali Pinar, Cosmin Safta, and Jean-Paul Watson

Systems and Industrial Engineering, University of Arizona
Sandia National Laboratories, Livermore, CA and Albuquerque, NM

DRO Workshop BIRS
March 5–9, 2018

Outlines

Distributionally robust optimization

$$(DRO) \quad \min_{\mathbf{x} \in X} \max_{F \in \mathcal{D}} \mathbb{E}_F[f(\mathbf{x}, \xi)]$$

- $f(\mathbf{x}, \xi)$ is a cost function in $\mathbf{x}$ that depends on a random vector ξ
- $\xi \in \mathcal{S} \subset \mathbb{R}^m$ with a distribution F
- $\mathcal{D}$ is an ambiguity set of F that encompasses the partial information on F.

Literature review

- **Moment-based** ambiguity sets
  - Ambiguity sets with first and second moments (see e.g., Delage and Ye '10)
  - Higher-order moment ambiguity sets (see e.g., Mehrotra and Papp '14)

Literature review

- **Moment-based** ambiguity sets
  - Ambiguity sets with first and second moments (see e.g., Delage and Ye '10)
  - Higher-order moment ambiguity sets (see e.g., Mehrotra and Papp '14)
- **Metric-based** ambiguity sets: Distance from reference (nominal) distribution (such as empirical distribution obtained from data):
  - Kullback-Leibler divergence (see e.g., Jiang and Guan '15)
  - Wasserstein Distance (see e.g., Gao and Kleywegt '16, Esfahani and Kuhn '15)

Literature review

- **Moment-based** ambiguity sets
  - Ambiguity sets with first and second moments (see e.g., Delage and Ye '10)
  - Higher-order moment ambiguity sets (see e.g., Mehrotra and Papp '14)
- **Metric-based** ambiguity sets: Distance from reference (nominal) distribution (such as empirical distribution obtained from data):
  - Kullback-Leibler divergence (see e.g., Jiang and Guan '15)
  - Wasserstein Distance (see e.g., Gao and Kleywegt '16, Esfahani and Kuhn '15)
- We consider the **moment-based** ambiguity sets.

Distributionally robust optimization

Assumption 1

$$\mathcal{D}(\mathcal{S}, \mu, \Sigma) = \left\{ F \;\middle|\; \begin{array}{l} \mathbb{P}(\xi \in \mathcal{S}) = 1 \\ \mathbb{E}_F[\xi] = \mu \\ \mathbb{E}_F[(\xi - \mu)(\xi - \mu)^T] \preceq \Sigma \end{array} \right\}$$

Remark: An extension to a **more general moment-based ambiguity set**

– For instance, the mean of ξ lies in an ellipsoid with the center μ is straightforward and is **omitted to simplify the introduction** of the proposed method.

Theorem (Delage and Ye '10)

Under Assumption 1, the target problem has the same optimal value as the following semi-infinite problem:

$$\begin{array}{rrcl} & f^* & := & \underset{\mathbf{x},s,\mathbf{q},\mathbf{Q}}{\text{minimize}}\; s+\mu^T\mathbf{q}+(\Sigma+\mu\mu^T)\bullet\mathbf{Q} \\ \textbf{(DRO-ORI)} & S.t. & & s+\xi^T\mathbf{q}+\xi^T\mathbf{Q}\xi \geq f(\mathbf{x},\xi)\;,\;\forall\xi\in\mathcal{S} \\ & & & \mathbf{Q}\succeq 0,\;\mathbf{x}\in X\;. \end{array}$$

- $s \in \mathbb{R}$, $\mathbf{q} \in \mathbb{R}^m$, $\mathbf{Q} \in \mathbb{R}^{m\times m}$: m is the size of ξ
- "$\bullet$" is the inner product defined by $A \bullet B = \sum_{i,j} A_{ij}B_{ij}$

Low-rank approximation

We introduce a linear combination of a lower-dimensional random vector $\xi_r \in \mathbb{R}^{m_1}$ $(m_1 \leq m)$ to approximate the ξ:

$$\xi \approx A_r \xi_r + \mu$$

- $A_r \in \mathbb{R}^{m \times m_1}$
- $$\mathcal{D}_r(\mathcal{S}_r, \mu_r, \Sigma_r) = \left\{ F_r \left| \begin{array}{l} \mathbb{P}(\xi_r \in \mathcal{S}_r) = 1 \\ \mathbb{E}_{F_r}[\xi_r] = 0 \\ \mathbb{E}_{F_r}[(\xi_r)(\xi_r)^T] \preceq \mathbf{I}_{m_1} \end{array} \right. \right\}.$$
- $\mathcal{S}_r := \{\xi_r \in \mathbb{R}^{m_1} : A_r \xi_r + \mu \in \mathcal{S}\}$
- $\mathbf{I}_{m_1}$ is an identity matrix of size m_1

PCA approximation

- $A_r\xi_r + \mu \in \mathcal{S}$ for any $\xi_r \in \mathcal{S}_r$ –Support
- $A_r\xi_r + \mu$ has the same mean as ξ –First Moment
- The covariance of $A_r\xi_r + \mu$ is $A_r\mathbb{E}_F[(\xi_r)(\xi_r)^T]A_r^T \preceq A_rA_r^T$ –Second Moment

PCA approximation

- $A_r\xi_r + \mu \in \mathcal{S}$ for any $\xi_r \in \mathcal{S}_r$ –Support
- $A_r\xi_r + \mu$ has the same mean as ξ –First Moment
- The covariance of $A_r\xi_r + \mu$ is $A_r\mathbb{E}_F[(\xi_r)(\xi_r)^T]A_r^T \preceq A_rA_r^T$ –Second Moment
- The **closer** $A_rA_r^T$ is to Σ; the **better** the approximation is.
- How to choose the **best** A_r?

PCA approximation

Eigendecomposition of Σ

$$\Sigma = U\Lambda U^T = U\Lambda^{\frac{1}{2}}(U\Lambda^{\frac{1}{2}})^T$$

- $U \in \mathbb{R}^{m\times m}$, $\Lambda \in \mathbb{R}^{m\times m}$ is a diagonal matrix of eigenvalues.
- $\Lambda^{\frac{1}{2}}$ replaces diagonal entries of Λ with their square roots.
- WLOG, the diagonal elements of Λ are arranged in **decreasing order**.

PCA approximation

Principal component analysis (PCA) as one of dimensionality reduction techniques:

- Employ a linear transformation to **project** the data to lower dimensional space
- Capture the **largest variance** (variability)
- $A_r = U_{m\times m_1}\Lambda_{m_1}^{\frac{1}{2}}$ which projects m-dimensional space to m_1-dimensional space.
  where $U_{m\times m_1} \in \mathbb{R}^{m\times m_1}$ is the $m \times m_1$ upper-left submatrix of U and $\Lambda_{m_1}^{\frac{1}{2}}$ is the $m_1 \times m_1$ upper-left submatrix of $\Lambda^{\frac{1}{2}}$.

Remark: m_1 is the number of principal components in PCA.

Distance functions

Least square error

$$\underset{A_r}{\text{minimize}} \quad \sum_i \sum_j ((A_r A_r^T)_{i,j} - \Sigma_{i,j})^2$$
$$S.t. \quad A_r \in \mathbb{R}^{m \times m_1}$$

Spectral norm

$$\underset{A_r}{\text{minimize}} \quad ||\Sigma - A_r A_r^T||$$
$$S.t. \quad A_r A_r^T \preceq \Sigma$$

where $||A|| = \sqrt{\rho(AA^T)}$ where A is a real square matrix and $\rho(A)$ is the largest eigenvalues of A.

Proposition

$A_r = U_{m\times m_1}\Lambda_{m_1}^{\frac{1}{2}}$ is an **optimal** solution of both the least square error and spectral norm problems.

PCA approximation for DRO

Then we have the PCA approximation:

$$\underset{\mathbf{x}\in X}{\text{minimize}}\ \underset{F_r\in\mathcal{D}_r}{\text{maximize}}\ \mathbb{E}_{F_r} f(\mathbf{x}, U_{m\times m_1}\Lambda_{m_1}^{\frac{1}{2}}\xi_r + \mu)$$

where

$$\mathcal{D}_r(\mathcal{S}_r, \mu_r, \Sigma_r) = \left\{ F \left| \begin{array}{l} \mathbb{P}(\xi_r \in \mathcal{S}_r) = 1 \\ \mathbb{E}_F[\xi_r] = 0 \\ \mathbb{E}_F[(\xi_r)(\xi_r)^T] \preceq \mathbf{I}_{m_1} \end{array} \right. \right\}.$$

Theorem: Main results of PCA approximation

The PCA approximation has the same optimal value as the following semi-infinite problem:

$$
\begin{aligned}
f^*(m_1) \quad &:= \quad \underset{\mathbf{x},s,\mathbf{q}_r,\mathbf{Q}_r}{\text{minimize}} \quad s + \mathbf{I}_{m_1} \bullet \mathbf{Q}_r \\
(\textbf{DRO-PCA}) \quad S.t. \quad & s + \xi_r^T \mathbf{q} + \xi_r^T \mathbf{Q}_r \xi_r \geq f(x, U_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \mu), \ \forall \xi_r \in \mathcal{S}_r \\
& \mathbf{Q}_r \succeq 0, \ \mathbf{x} \in X
\end{aligned}
$$

where $s \in \mathbb{R}$, $\mathbf{q}_r \in \mathbb{R}^{m_1}$ and $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$.

- DRO-PCA is a **relaxation** problem of the original problem and $f^*(m_1)$ is a lower bound, i.e., $f^*(m_1) \leq f^*$
- $f^*(m_1)$ is a **nondecreasing** function of m_1, i.e., $f^*(m_1) \leq f^*(m_2)$ if $m_2 \geq m_1$.
- If $m_1 = m$, then problem DRO-PCA has the **same optimal value** as problem DRO-ORI. Thus, $f^*(m) = f^*$.

Comparison

$$
\begin{array}{lll}
f^*(m_1) & := & \underset{\mathbf{x},s,\mathbf{q}_r,\mathbf{Q}_r}{\text{minimize}} \quad s + \mathbf{I}_{m_1} \bullet \mathbf{Q}_r \\
\textbf{(DRO-PCA)} & S.t. & s + \xi_r^T \mathbf{q} + \xi_r^T \mathbf{Q}_r \xi_r \geq f(x, U_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}} \xi_r + \mu), \; \forall \xi_r \in \mathcal{S}_r \\
 & & \mathbf{Q}_r \succeq 0, \; \mathbf{x} \in X
\end{array}
$$

$s \in \mathbb{R}$, $\mathbf{q}_r \in \mathbb{R}^{m_1}$ and $\mathbf{Q}_r \in \mathbb{R}^{m_1 \times m_1}$ $\rightarrow 1 + m_1 + m_1^2$

$$
\begin{array}{lll}
f^* & := & \underset{\mathbf{x},s,\mathbf{q},\mathbf{Q}}{\text{minimize}} \; s + \mu^T \mathbf{q} + (\Sigma + \mu\mu^T) \bullet \mathbf{Q} \\
\textbf{(DRO-ORI)} & S.t. & s + \xi^T \mathbf{q} + \xi^T \mathbf{Q} \xi \geq f(\mathbf{x}, \xi) \,, \; \forall \xi \in \mathcal{S} \\
 & & \mathbf{Q} \succeq 0, \; \mathbf{x} \in X \,.
\end{array}
$$

$s \in \mathbb{R}$, $\mathbf{q} \in \mathbb{R}^m$ and $\mathbf{Q} \in \mathbb{R}^{m \times m}$ $\rightarrow 1 + m + m^2$

- **DRO-PCA is easier to solve than DRO-ORI.**

Piecewise linear $f(\mathbf{x}, \xi)$ and polyhedra $\mathcal{S}$

- Support is polyhedral: $\mathcal{S} = \{\xi | A\xi \leq b\}$ with $A \in \mathbb{R}^{n_1 \times m}$ and $b \in \mathbb{R}^{n_1}$
- $f(\mathbf{x}, \xi)$ is a convex piecewise linear function in ξ: $f(\mathbf{x}, \xi) = \max_{k=1}^{K}(y_k^0(\mathbf{x}) + \mathbf{y}_k(\mathbf{x})^T \xi)$
  - $\mathbf{y}_k(\mathbf{x}) = [y_k^1(\mathbf{x}), \ldots, y_k^m(\mathbf{x})]^T$ and $y_k^0(\mathbf{x})$ are affine in $\mathbf{x}$

Corollary: simplification of two reformulations

DRO-ORI

$$f^* = \underset{\mathbf{x}, s, \mathbf{q}, \lambda, \mathbf{Q}}{\text{minimize}} \quad s + \mu^T \mathbf{q} + (\Sigma + \mu\mu^T) \bullet \mathbf{Q}$$

$$S.t. \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \lambda_k^T b & \frac{(\mathbf{q} - \mathbf{y}_k(\mathbf{x}) + A^T \lambda_k)^T}{2} \\ \frac{(\mathbf{q} - \mathbf{y}_k(\mathbf{x}) + A^T \lambda_k)}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \forall k \in \{1, \ldots, K\}$$

$$\mathbf{Q} \succeq 0, \ \lambda \in \mathbb{R}_+^{n_1}, \ \mathbf{x} \in X.$$

DRO-PCA

$$f^*(m_1) = \underset{\mathbf{x}, s, \mathbf{q}_r, \lambda, \mathbf{Q}_r}{\text{minimize}} \quad s + \mathbf{I}_{m_1} \bullet \mathbf{Q}_r$$

$$S.t. \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \lambda_k^T b - \mathbf{y}_k(\mathbf{x})^T \mu + \lambda_k^T A \mu & \frac{(\mathbf{q}_r + (U_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}})^T (A^T \lambda_k - \mathbf{y}_k(\mathbf{x})))^T}{2} \\ \frac{\mathbf{q}_r + (U_{m \times m_1} \Lambda_{m_1}^{\frac{1}{2}})^T (A^T \lambda_k - \mathbf{y}_k(\mathbf{x}))}{2} & \mathbf{Q}_r \end{bmatrix} \succeq 0$$

$$\forall k \in \{1, 2, ..., K\}$$

$$\mathbf{Q}_r \succeq 0, \ \lambda \in \mathbb{R}_+^{n_1}, \ \mathbf{x} \in X.$$

Corollary: simplification of two reformulations

DRO-ORI

$$f^* = \underset{\mathbf{x},s,\mathbf{q},\lambda,\mathbf{Q}}{\text{minimize}} \quad s + \mu^T\mathbf{q} + (\Sigma + \mu\mu^T) \bullet \mathbf{Q}$$

$$S.t. \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \lambda_k^T b & \frac{(\mathbf{q}-\mathbf{y}_k(\mathbf{x})+A^T\lambda_k)^T}{2} \\ \frac{(\mathbf{q}-\mathbf{y}_k(\mathbf{x})+A^T\lambda_k)}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \forall k \in \{1,\dots,K\}$$

$$\mathbf{Q} \succeq 0,\ \lambda \in \mathbb{R}_+^{n_1},\ \mathbf{x} \in X.$$ Dimension of LMI: $m+1$

DRO-PCA

$$f^*(m_1) = \underset{\mathbf{x},s,\mathbf{q}_r,\lambda,\mathbf{Q}_r}{\text{minimize}} \quad s + \mathbf{I}_{m_1} \bullet \mathbf{Q}_r$$

$$S.t. \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \lambda_k^T b - \mathbf{y}_k(\mathbf{x})^T\mu + \lambda_k^T A\mu & \frac{(\mathbf{q}_r+(U_{m\times m_1}\Lambda_{m_1}^{\frac{1}{2}})^T(A^T\lambda_k-\mathbf{y}_k(\mathbf{x})))^T}{2} \\ \frac{\mathbf{q}_r+(U_{m\times m_1}\Lambda_{m_1}^{\frac{1}{2}})^T(A^T\lambda_k-\mathbf{y}_k(\mathbf{x}))}{2} & \mathbf{Q}_r \end{bmatrix} \succeq 0$$

$$\forall k \in \{1, 2, ..., K\}$$

$$\mathbf{Q}_r \succeq 0,\ \lambda \in \mathbb{R}_+^{n_1},\ \mathbf{x} \in X.$$ Dimension of LMI: $m_1 + 1$

Quality of PCA Approximation

Proposition

When $\mathcal{S}$ is polyhedral and $f(x,\xi)$ is convex piecewise linear, then

$$0 \leq f^*(m) - f^*(m_1) \leq \sum_{k=1}^{K} \sqrt{\sum_{i=m_1+1}^{m} \Lambda_{i,i}(\mathbf{y}_k(\mathbf{x}^*)^T U_i)^2},$$

where $\mathbf{x}^*$ is an optimal solution of the PCA approximation

Remark: $f^*(m) = f^*$. The **smaller** $\Lambda_{i,i}, i = m_1 + 1, \ldots, m$ is, the **better** the PCA approximation is.

Computational setup

- DRO **Conditional Value-At-Risk (CVaR)**
- A **Risk-Averse Production-Transportation** application
- All problems are solved using **Mosek** with their default parameters on a computer equipped with a Quad-core Intel Core i7 @ 2.2 GHz processor and 16 GB RAM.

DRO for Conditional Value-At-Risk(CVaR)

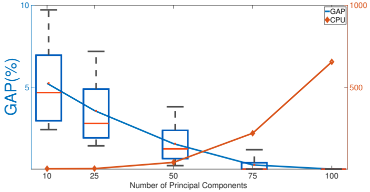
DRO $\text{CVaR}_{1-\alpha}$ of a cost function $\mathbf{x}^T\xi$ can be formulated as the following optimization problem (Rockafellar and Uryasev 02'):

$$\underset{\mathbf{x}\in X, t\in\mathbb{R}}{\text{minimize}} \ \underset{F\in\mathcal{D}}{\text{maximize}} \ t + \frac{1}{\alpha}\mathbb{E}_F[x^T\xi - t]^+$$

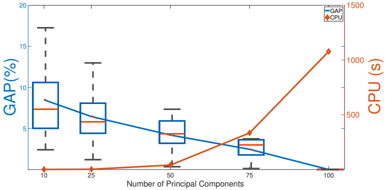
- where $\alpha \in (0,1)$ is a risk tolerance level
- function $[\cdot]^+ := \max\{0,\cdot\}$.
- $X = \{\mathbf{x} \in \mathbb{R}^n_+ | \sum_{i=1}^n x_i = 1\}$.

Numerical Study Setup

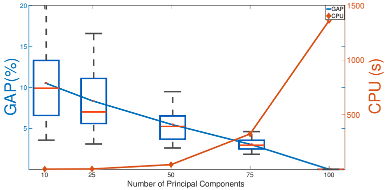
- $n = 200$ and $\alpha = 0.05$.
- Support $\mathcal{S} \in \{[-2\sigma, 2\sigma], [-3\sigma, 3\sigma], [-4\sigma, 4\sigma]\}$
- $\mu \sim \mathcal{U}[5, 10]$
- Σ is generated randomly using MATLAB function "gallery('randcorr',n)"
- Numbers of principal components $m_1 \in \{200, 150, 100, 50, 20\}$.



(a) Support=$[-2\sigma, 2\sigma]$



(b) Support=$[-3\sigma, 3\sigma]$



(c) Support=$[-4\sigma, 4\sigma]$

Randomly generated Σ

CVAR m=200 Support	Orig. time (secs)	PCA ($m_1 = 200$) time (secs)	Gap (%)	Gap2 (%)	PCA ($m_1 = 150$) time (secs)	Gap (%)	Gap2 (%)	PCA ($m_1 = 100$) time (secs)	Gap (%)	Gap2 (%)	PCA ($m_1 = 50$) time (secs)	Gap (%)	Gap2 (%)	PCA ($m_1 = 20$) time (secs)	Gap (%)	Gap2 (%)
$[-2\sigma, 2\sigma]$	1019.5	654.5	0.00	0.00	219.4	0.26	8.37	41.1	1.55	9.10	3.1	3.57	12.93	2.0	5.24	18.45
$[-3\sigma, 3\sigma]$	1290.9	1078.3	0.00	0.00	334.2	2.46	7.40	40.8	4.20	9.93	2.7	6.45	14.85	1.1	8.49	19.50
$[-4\sigma, 4\sigma]$	1309.2	1362.0	0.00	0.00	324.1	3.06	7.42	42.9	5.49	10.19	3.1	8.37	14.18	1.7	10.56	19.13

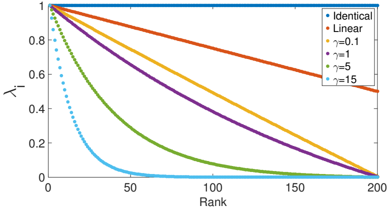
Table: Average results of PCA method for ten instances.

Randomly generated Σ

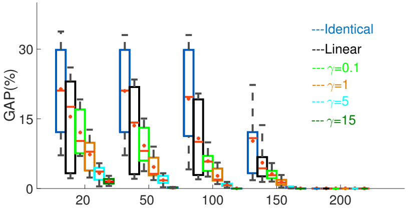
Size	Orig. time (h)	PCA ($m_1 = 300$) time (h)	PCA ($m_1 = 300$) Gap (%)	PCA ($m_1 = 225$) time (h)	PCA ($m_1 = 225$) Gap (%)	PCA ($m_1 = 150$) time (h)	PCA ($m_1 = 150$) Gap (%)	PCA ($m_1 = 75$) time (h)	PCA ($m_1 = 75$) Gap (%)	PCA ($m_1 = 30$) time (h)	PCA ($m_1 = 30$) Gap (%)
$m = 300$	9.416	8.605	0.00	0.867	1.56	0.088	3.71	0.004	5.89	0.000	7.55

Table: Average results of the PCA approximation on a 300-dimensional problem with Support=$[-3\sigma, 3\sigma]$.

Specially structured Σ



(d) Eigenvalue generating functions $\frac{e^{-0.01*i*\gamma+1.01*\gamma}-1}{e^{\gamma}-1}, i = 1, \ldots, 200$



(e) Performance of PCA approximation

Average results of PCA method with structured Σ

Slope	Orig. time (secs)	PCA ($m_1 = 200$) time (secs)	PCA ($m_1 = 200$) Gap (%)	PCA ($m_1 = 150$) time (secs)	PCA ($m_1 = 150$) Gap (%)	PCA ($m_1 = 100$) time (secs)	PCA ($m_1 = 100$) Gap (%)	PCA ($m_1 = 50$) time (secs)	PCA ($m_1 = 50$) Gap (%)	PCA ($m_1 = 20$) time (secs)	PCA ($m_1 = 20$) Gap (%)
Identical	1234.2	1036.9	0.00	148.8	10.24	21.6	19.29	2.2	21.02	2.0	21.40
Linear	1344.8	1326.5	0.00	296.8	5.58	41.7	10.82	3.0	13.58	2.0	15.47
0.1	1401.0	1561.2	0.00	337.9	3.12	42.4	5.88	3.1	9.24	2.0	12.06
1	1643.7	1800.1	0.00	340.0	1.38	51.1	2.70	2.7	4.62	1.0	7.31
5	1731.4	1560.0	0.00	346.5	0.26	45.4	0.75	2.8	1.83	1.0	3.26
15	1503.1	1624.7	0.00	325.2	0.00	42.3	0.01	2.6	0.21	1.1	1.59

Deterministic production-transportation problem (Bertsimas et al '10)

$$
\begin{aligned}
\underset{\mathbf{x},\mathbf{y}}{\text{minimize}} \quad & \sum_{i=1}^{m} c_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{n} \xi_{ij} y_{ij} \\
\text{subject to} \quad & \sum_{i=1}^{m} y_{ij} = d_j,\ j = 1,\ldots,n \\
& \sum_{j=1}^{n} y_{ij} = x_i,\ i = 1,\ldots,m \qquad (11) \\
& 0 \le x_i \le 1,\ i = 1,\ldots,m \\
& y_{ij} \ge 0,\ i = 1,\ldots,m, j = 1,\ldots,n
\end{aligned}
$$

Two-stage risk averse production-transportation problem (Bertsimas et al '10)

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \sum_{i=1}^{m} c_i x_i + \underset{F \in \mathcal{D}}{\text{maximize}}\, \mathbb{E}_F[\mathcal{U}(\mathcal{Q}(\mathbf{x}, \xi))] \\ \text{subject to} \quad & 0 \le x_i \le 1,\ i = 1, \ldots, m \end{aligned} \tag{12}$$

$$\begin{aligned} \mathcal{Q}(\mathbf{x}, \xi) = \underset{\mathbf{y} \ge 0}{\text{minimize}} \quad & \sum_{i=1}^{m} \sum_{j=1}^{n} \xi_{ij} y_{ij} \\ \text{subject to} \quad & \sum_{i=1}^{m} y_{ij} = d_j,\ j = 1, \ldots, n \\ & \sum_{j=1}^{n} y_{ij} = x_i,\ i = 1, \ldots, m \end{aligned}$$

Piecewise linear convex nondecreasing disutility function (Bertsimas et al '10)

The definition of disutility function $\mathcal{U}(\cdot)$ is given as follows:

$$\mathcal{U}(\mathcal{Q}(\mathbf{x},\xi)) = \max_{k\in\{1,2,\dots,K\}} a_k \mathcal{Q}(\mathbf{x},\xi) + b_k, \tag{13}$$

with nonnegative coefficients, i.e., $a_k \geq 0$ for all k.

(m, n)	Orig. time (secs)	PCA (100%) time (secs)	PCA (100%) Gap (%)	PCA (75%) time (secs)	PCA (75%) Gap (%)	PCA (50%) time (secs)	PCA (50%) Gap (%)	PCA (25%) time (secs)	PCA (25%) Gap (%)	PCA (10%) time (secs)	PCA (10%) Gap (%)
(5, 20)	91.4	88.2	0.00	27.4	0.25	7.7	0.57	2.2	0.93	1.7	0.94
(8, 25)	2574.5	2392.1	0.00	609.6	0.06	99.0	0.11	9.2	0.12	2.5	0.12
(10, 30)	–	–	–	4888.2	1.07*	705.2	1.44*	42.7	1.76*	5.3	2.35*

Table: "–" indicates no solution found and "*" indicates an upper bound for the relative gap rather than the actual gap.

Summary

- We **propose a PCA approximation** method for DRO problems with moment-based ambiguity sets.

Summary

- We **propose a PCA approximation** method for DRO problems with moment-based ambiguity sets.
- We show that the PCA approximation is a **relaxation and quantify** the impact of the number of principal components on solution quality.

Summary

- We **propose a PCA approximation** method for DRO problems with moment-based ambiguity sets.
- We show that the PCA approximation is a **relaxation and quantify** the impact of the number of principal components on solution quality.
- The proposed approximation method provides decision makers more **flexibility** to deal with uncertainty, allowing for direct control of the **trade-offs** between solution quality and runtime.

Summary

- We **propose a PCA approximation** method for DRO problems with moment-based ambiguity sets.
- We show that the PCA approximation is a **relaxation and quantify** the impact of the number of principal components on solution quality.
- The proposed approximation method provides decision makers more **flexibility** to deal with uncertainty, allowing for direct control of the **trade-offs** between solution quality and runtime.
- One future research direction is to apply **more general matrix** decomposition other than eigen-decomposition in PCA.

Reference I

[1] Bertsimas, Dimitris, Xuan Vinh Doan, Karthik Natarajan, and Chung-Piaw Teo. "Models for minimax stochastic linear optimization problems with risk aversion." Mathematics of Operations Research 35, no. 3 (2010): 580-602.

[2] Delage, Erick, and Yinyu Ye. "Distributionally robust optimization under moment uncertainty with application to data-driven problems." Operations research 58.3 (2010): 595-612.

[3] Esfahani, Peyman Mohajerin, and Daniel Kuhn. "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations." Mathematical Programming (2015): 1-52.

[4] Gao, Rui, and Anton J. Kleywegt. "Distributionally Robust Stochastic Optimization with Wasserstein Distance." arXiv preprint arXiv:1604.02199 (2016).

[5] Jiang, Ruiwei, and Yongpei Guan. "Data-driven chance constrained stochastic program." Mathematical Programming (2015).

[6] Mehrotra, Sanjay, and Dvid Papp. "A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization." SIAM Journal on Optimization 24.4 (2014): 1670-1697.

[7] Rockafellar, R. Tyrrell, and Stanislav Uryasev. "Optimization of conditional value-at-risk." Journal of risk 2 (2000): 21-42.

[8] Van Parys, Bart PG, Paul J. Goulart, and Manfred Morari. "Distributionally robust expectation inequalities for structured distributions." Optimization-online.

Thank you for your attention!