IRAN UNIVERSITY of SCIENCE and TECHNOLOGY

UNIVERSITY OF TWENTE.

Data-driven Robust Optimzation

An Introduction to Developing Smart Uncertainty Sets

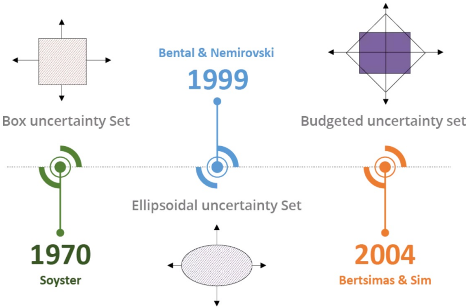
Mohammad Namakshenas

Research Fellow at CHOIR

March 18, 2019

M. Namakshenas and M. S. Pishvaee, Data Driven Robust Optimization, in Robust and Constrained Optimization: Methods and Applications, D. Clark, Ed. NOVA SCIENCE PUBLISHERS, INC., 2019, pp. 140.

A punchline on RO



- BOX (Interval) uncertainty space
- Ellipsoidal uncertainty space
- Polyhedral uncertainty space (Budgeted uncertainty space)

Motivations to DDRO

SIM et al.

DDRO is a Marriage of Robust Optimization (RO) and Stochastic Programming (SO).

Motivations to DDRO

SIM et al.

DDRO is a Marriage of Robust Optimization (RO) and Stochastic Programming (SO).

Reliability

Bring the RO to a real and reliable taste.

Motivations to DDRO

SIM et al.

DDRO is a Marriage of Robust Optimization (RO) and Stochastic Programming (SO).

Reliability

Bring the RO to a real and reliable taste.

Conservatism

To produce less conservative models.

Motivations to DDRO

SIM et al.

DDRO is a Marriage of Robust Optimization (RO) and Stochastic Programming (SO).

Reliability

Bring the RO to a real and reliable taste.

Conservatism

To produce less conservative models.

The Golden Key!

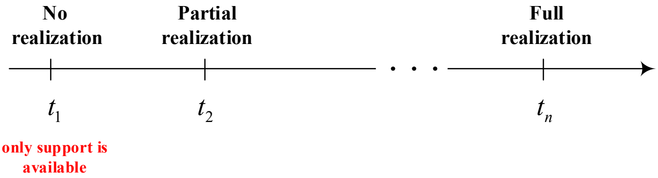
The uncertainty (ambiguity) set is the heart of RO.

Concept 1.

Motivations to DDRO

Concept 1.

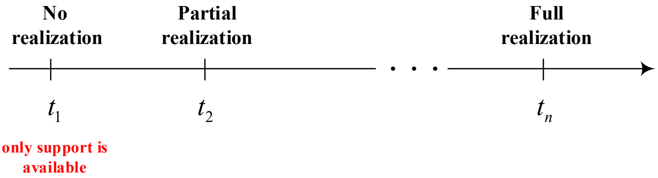
Concept 2.



Motivations to DDRO

Concept 1.

Concept 2.



Question. How is it possible to update an RO model in t_2?

Motivations to DDRO

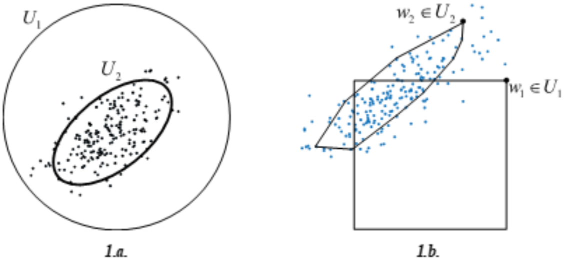


Figure: A conservative uncertainty set vs. a realized uncertainty set w.r.t a posteriori data, A priori worst case vs. realized worst-case w.r.t a posteriori data.

Distributionally Robust Optimization

Moment-based uncertainty sets

Moment-based uncertainty sets is comparable to that of the classical ellipsoidal uncertainty sets (we don't know the underlying distribution). Suppose that the support S, mean μ, and co-variance Σ of ξ is known explicitly. Hence, the moment-based uncertainty set $U(S, \mu, \Sigma)$ is defined such that the convexity of S, $\mu \in \text{Int}(S)$, and $\Sigma \in \mathcal{PSD}$ as follows:

$$U(S, \mu, \Sigma) = \begin{cases} P(\xi \in S) = 1, \\ \mathbb{E}_F[\xi] = \mu, \\ \mathbb{E}_F\left[(\xi - \mu)(\xi - \mu)^T\right] \preceq \Sigma. \end{cases}$$

Distributionally Robust Optimization

Expected Utility (Loss) Function

The goal is to minimize the worst case outcome of the expected utility function.

Expected Utility (Loss) Function

The goal is to minimize the worst case outcome of the expected utility function.

$$\underset{x \in S}{\text{minimize}} \quad \max_{F \in U} \mathbb{E}_F \left[g(x, \xi) \right] \tag{1}$$

Distributionally Robust Optimization

Conjecture

The robust counterpart of the Problem according to U can be formulated as a semi-definite program (SDP).

Distributionally Robust Optimization

Conjecture

The robust counterpart of the Problem according to U can be formulated as a semi-definite program (SDP).

Theorem 1

Given the uncertainty set U, if is continuous and differentiable in x, the robust counterpart of the Problem (1) is as follows:

$$\begin{aligned}
\underset{x,t,p,Q}{\text{minimize}} \quad & t + \mu^T p + \langle \Sigma + \mu^T \mu, Q \rangle \\
\text{subject to} \quad & t + \xi^T p + \xi^T Q \xi \geq \ g(x, \xi), \forall \xi \in S, \\
& t \in \mathbb{R}, p \in \mathbb{R}^m, Q \in \mathbb{R}^{m \times m}, \\
& Q \succeq 0, x \in X.
\end{aligned} \tag{2}$$

Theorem 2

Given the support $S = \{\xi \,|\, A\xi \leq b\} \neq \emptyset$ be a polyhedral set with $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$, the problem (2) is reduced to the following problem:

Distributionally Robust Optimization

Theorem 2

Given the support $S = \{\xi \,|\, A\xi \leq b\} \neq \emptyset$ be a polyhedral set with $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$, the problem (2) is reduced to the following problem:

$$
\begin{aligned}
\underset{x,t,p,\lambda,Q}{\text{minimize}} \quad & t + \mu^T p + \langle \Sigma + \mu^T \mu, Q \rangle \\
\text{subject to} \quad & \begin{pmatrix} t - \gamma_k^0(x) - \lambda_k^T b & \dfrac{\left(p - \gamma_k(x) + A^T \lambda_k\right)^T}{2} \\ \dfrac{p - \gamma_k(x) + A^T \lambda_k}{2} & Q \end{pmatrix} \succeq 0, \forall k \in K, \\
& Q \succeq 0, \\
& \lambda_k \in \mathbb{R}_+^n, k \in K.
\end{aligned} \tag{3}
$$

Proof. Refer to the chapter book.

Example

The standard portfolio optimization model is also known as the Markowitz portfolio model is formulated as follows:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & x^T \widetilde{r} \\ \text{subject to} \quad & \mathbf{1} x = 1, \\ & x \in \mathbb{R}^m \end{aligned} \tag{4}$$

Setup

- Assume that the return vector r is uncertain with the support set $S_r = \{\xi^{(r)} | \sum_{i \in I} a_i {\xi_i}^{(r)} \leq b\}$, where the distribution $F^{(r)}$ of the random return vector belongs to some uncertainty set $U^{(r)}$ that encompasses the moment information $\mu^{(r)}$ and $\Sigma^{(r)}$.
- The utility function should be defined in terms of a linear piece-wise function, $\gamma^0(x) = 0$ and $\gamma(x)^T \xi^{(r)} = x^T \widetilde{r}$.

DRO of Problem 4 can be deduced as follows:

$$\underset{\{x \in \mathbb{R}^m | \mathbf{1}x = 1\}}{\text{maximize}} \quad \min_{F^{(r)} \in U^{(r)}} \mathbb{E}_{F^{(r)}} \left[g(x, \xi^{(r)}) = x^T \widetilde{r} \right] \qquad (5)$$

Sample Data.

$$\mu^{(r)} = \begin{pmatrix} 35.80 & 69.50 & 50.15 \end{pmatrix}$$

$$\Sigma^{(r)} = \begin{pmatrix} 184.80 & -144.10 & 3.66 \\ -144.10 & 416.89 & 73.81 \\ 3.66 & 73.81 & 97.18 \end{pmatrix}$$

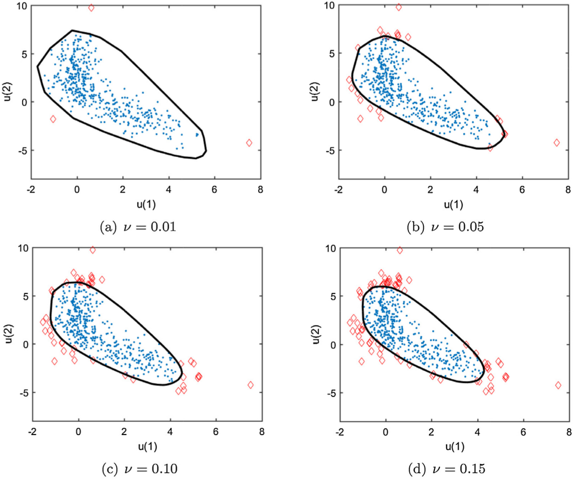
SDP Counerpart.

$$
\begin{aligned}
\underset{x,t,p,\lambda,Q}{\text{minimize}} \quad & t + 35.80p_1 + 69.50p_2 + 50.15p_3 + 8811.7q_{11} + 8482.8q_{12} + \\
& 8630.6q_{13} + 8482.8q_{21} + 9043.8q_{22} + 8700.7q_{23} + 8630.6q_{31} + \\
& 8700.7q_{32} + 8724.1q_{33} \\
\text{subject to} \quad & \mathbf{1}x = 1, \\
& \begin{pmatrix} t - \lambda b & \dfrac{\left(\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \lambda \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}\right)^T}{2} \\ \dfrac{\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \lambda \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}}{2} & \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix} \end{pmatrix} \succeq 0, \\
& Q \succeq 0, x \in X
\end{aligned}
$$

CVX Package.

```
n=3;
b = 2;
mu =   [  35.8000   69.5000     50.1500  ];
covm = [  184.8000 -144.1053    3.6632
         -144.1053  416.8947    73.8158
          3.6632    73.8158     97.1868];
A = [ 2 ; 1 ; 1 ];
I = ones(1,n);
cvx_begin sdp

   variable t
   variable lambda
   variable x(n,1)
   variable p(n,1)
   variable Q(n,3) symmetric

   minimize(t + mu*p +  sum(sum(((covm + mu*mu' ).*Q))));

   #Constraints
    [t-lambda*b          (p-x+A*lambda)'/2 ;...
    (p-x+A*lambda)/2     Q          ] >= 0 ;
    I*x == 1;
    Q >= 0;
    p(1) >=0 ; p(2) >=0 ; p(3) >=0 ;
    x(1) >=0 ; x(2) >=0 ; x(3) >=0 ;
    lambda >= 0;

cvx_end
```

C. Shang, X. Huang, and F. You, Data-driven robust optimization based on kernel learning, Comput. Chem. Eng., vol. 106, pp. 464479, 2017.



(a) $\nu = 0.01$

(b) $\nu = 0.05$

(c) $\nu = 0.10$

(d) $\nu = 0.15$

Pros

- The robust counterpart is convex and linear w.r.t x (LP).
- The accuracy and outliers can be controlled by the parameter ν.
- This approach ensures that the final uncertainty set is closed and convex.

Pros

- The robust counterpart is convex and linear w.r.t x (LP).
- The accuracy and outliers can be controlled by the parameter ν.
- This approach ensures that the final uncertainty set is closed and convex.

Cons

- A lot of data preparation and pre-calculations are needed.
- The model size is intractable when the kernel matrix size is increased to a certain number.

Cutting Hyper-planes & RO

Definition 1

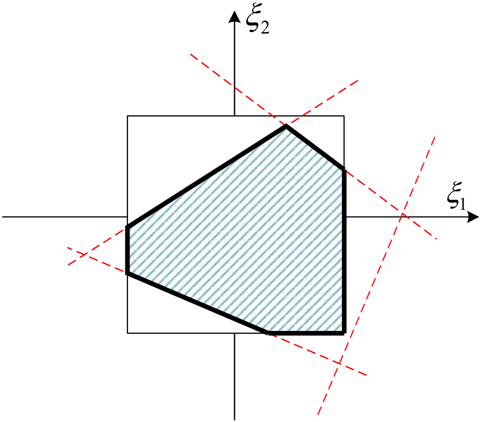
Given C number of cutting hyper-planes with the gradient vector $Q \in \mathbb{R}^{C\times|J_i|}$ and the intercept $d \in \mathbb{R}^C$, the hyper-plane $h(\xi)$ is defined over the random vector ξ as follows ($|J_i|$ is the number of the uncertain parameter in the ith constraint set):

$$h(\xi) = Q\xi_i + d \Leftrightarrow h_c(\xi) = q_{cj}\xi_{ij} + d_c, \quad \forall\ i, \forall\ j \in |J_i|, \forall\ c \in \{1, \ldots, C\} \tag{6}$$

Note. You already know the definition of ξ.

Cutting Hyper-planes & RO

A Punchline.



The robust counterpart of the below problem is needed!

$$
\begin{aligned}
\underset{x \in X}{\text{maximize}} \quad & c^T x \\
\text{subject to} \quad & \bar{a}_i x + \max_{\xi_i \in U_\infty(h)} \{\hat{a}_i \xi_i x\} \leq b_i, \qquad \forall i, \qquad (7) \\
& U_\infty(h) = \{\xi \,|\, \|\xi\|_\infty \leq 1,\ h(\xi) + d \geq 0\}
\end{aligned}
$$

Theorem 3

The robust counterpart of the problem (7) w.r.t $U_\infty(h)$ is as follows:

$$\begin{aligned} \underset{x \in X, \tau \in \mathbb{R}_+}{\text{maximize}} \quad & c^T x \\ \text{subject to} \quad & \bar{a}_i x + \left\| \hat{a}_i x + Q^T \tau \right\|_1 + d^T \tau \le b_i, \; \forall i. \end{aligned} \tag{8}$$

Proof. On the board.

Theorem 4

The robust counterpart of the problem (7) w.r.t $U_2(h)$ is an SOCP as follows:

$$\begin{aligned} \underset{x \in X, \tau \in \mathbb{R}_+}{\text{maximize}} \quad & c^T x \\ \text{subject to} \quad & \bar{a}_i x + \lambda + d^T \leq b_i, \quad \forall i, \\ & \left\| \hat{a}_i x + Q^T \tau \right\|_2^2 \leq \lambda^2. \end{aligned} \tag{9}$$

Proof. Refer to the chapter book.

Home Exercise.

Find the robust counterpart of the problem (7) w.r.t $U_1(h)$.

Note. So Eaaaaaasy!.

Example

Given the uncertain parameters $\widetilde{a}_1 = 2 \pm \xi_1$ and $\widetilde{a}_2 = 1 \pm 0.5\xi_2$, 2000 samples are generated according to a bi-variate normal distribution w.r.t $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\Sigma}$.

$$\begin{aligned} \underset{x \in \mathbb{R}^2_+}{\text{maximize}} \quad & 2x_1 + 3x_2 \\ \text{subject to} \quad & a_1 x_1 + a_2 x_2 \leq 5. \end{aligned} \tag{10}$$

$$\hat{\mu}_1 = 2, \hat{\mu}_2 = 1$$

$$\hat{\Sigma} = cov(a_1, a_2) = \begin{pmatrix} 1 & 1.5 \\ 1.5 & 3 \end{pmatrix}$$

Example Continued.

We produce $C = 4$ cutting planes with $\epsilon = 0.05$ as follows:

$$Q = \begin{pmatrix} -12 & 3 \\ -12 & 4 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$$

$$d = \begin{pmatrix} -13 \\ 2 \\ 1 \\ -1 \end{pmatrix}$$

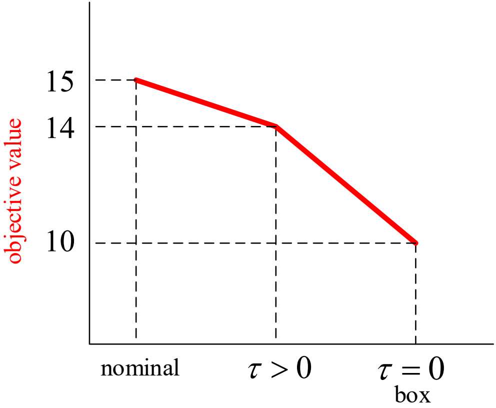
Cutting Hyper-planes & RO

The robust counterpart.

$$
\begin{aligned}
\underset{x \in \mathbb{R}^2_+, \tau \in \mathbb{R}^4_+}{\text{maximize}} \quad & 2x_1 + 3x_2 \\
\text{subject to} \quad & \\
2x_1 + x_2 + |x_1 - 12\tau_1 - 12\tau_2 + 2\tau_3 + 3\tau_4| + & \\
& |0.5x_2 + 3.5\tau_1 + 3.5\tau_2 + 0.5\tau_3 + 0.5\tau_4| + \\
& - 13\tau_1 + 1.5\tau_2 + \tau_3 - \tau_4 \leq 5
\end{aligned}
$$

Note. How does the model look like given $\tau = 0$?

Comparison of objectives.



Cutting Hyper-planes & RO

Pros

- The robust counterpart is convex w.r.t three uncertainty sets.
- The maximum and outliers can be controlled during generating hyper-planes (Q and d).
- This approach ensures that the final uncertainty set is closed and convex.

Cutting Hyper-planes & RO

Pros

- The robust counterpart is convex w.r.t three uncertainty sets.
- The maximum and outliers can be controlled during generating hyper-planes (Q and d).
- This approach ensures that the final uncertainty set is closed and convex.

Cons

- There is no deterministic approach to find Q and d.
- For 2×2 you need to create 4 cutting planes.

¿ More Questions ?