DATA-DRIVEN ROBUST OPTIMIZATION

Moahammad Namakshenas and Mir Saman Pishvae*
School of Industrial Engineering,
Iran University of Science and Technology,
Tehran, Iran

ABSTRACT

Nowadays, data becomes a critical corporate asset. Integration of data in optimization methods has made a leap forward in recent years. The volume of available data has grown exponentially and enabled new insights and models in decision-making and especially the decision making under uncertainty. Recent advances in machine learning combined with polyhedral theory and conic programs pave new ways to discover intelligent and anti-conservative robust optimization models. This chapter introduces different robust models induced by three well-known data-driven uncertainty sets, distributional, clustering-oriented, and cutting hyperplanes uncertainty sets.

Keywords: data driven, robust optimization, uncertainty set, stochastic programming, distributionally robust, convex optimization

*Corresponding Author Email: pishvae@iust.ac.ir.
INTRODUCTION

Robust optimization is a tractable alternative to stochastic programming, particularly suited for the problems in which the probability distribution of parameters is unknown. In many real-world situations, a precise stochastic description of the uncertain events may not be available. With less structured information, such as the support of an uncertain parameter, one might describe the existing uncertainties by dedicating a set in which all realizations should lie, i.e., “uncertainty set.” The goal is to guarantee the feasibility of the underlying constraints for any possible realization while optimizing an objective protecting against the worst possible consequence. It is non-trivial to say that “uncertainty set” is the heart of robust optimization.

A key reason for its success has been its computational tractability. Computational experience suggests that with well-chosen sets, robust models yield tractable optimization problems whose solutions are reliable than other approaches. However, robust models constructed by poorly chosen sets may be overly-conservative or computationally intractable. The earliest attempts in formulating uncertainty sets date back to the 1970s, with the work of Soyster [1], in which uncertainty was limited to a column-wise structure or the box uncertainty set. Despite its computational convenience and guaranteed feasibility, the box uncertainty set tends to induce over-conservative decisions. Later, immense research effort has been made on devising more flexible robust models to ameliorate over-conservatism. Ellipsoidal uncertainty sets have been put forward independently by El Ghaoui et al. [2] and Ben-tal et al. [3] based on which the robust counterpart model simplifies to a conic quadratic problem in the presence of linear constraints. To further decrease the conservatism, intersections of basic uncertainty sets have been developed, including the ellipsoidal and polyhedral uncertainty sets [4]. Bertsimas and Sim [5] robustified linear programs using a polyhedral uncertainty set adjustable with the so-called budget, say budgeted uncertainty set. That attempt led to tractable models. Bertsimas and Ruiter [6] introduced a generalized polyhedral uncertainty set
in which the affine uncertainty coefficients could be formulated as the smallest convex set that includes \( U \).

On the other hand, the last decade witnessed an explosion in the availability of data. Massive amounts of data are now routinely collected in many industries and real-time mega-systems. The researches suggest that the full coverage of uncertain values usually leads to impractical and conservative results. A natural question, then, is how should robust optimization theories be tailored to this new paradigm? To reduce the conservatism and improve the robustness of solutions obtained from the set-induced optimization models, the notion of data is introduced to the uncertainty sets, which resulted in the “data-driven robust optimization.” Thus historical data or the uncertainty event should be introduced to correct the description of uncertainties and reformulate robust counterparts [7].

The distributional geometry of massive uncertain data can be effectively captured as a compact convex uncertainty set, which considerably reduces conservatism of robust optimization problems. Distributionally robust optimization (DRO) is an effective method to address the inexactness of probability distributions of uncertain parameters. In this approach, partial distributional information, such as support set and moment statistics are obtained from available empirical data [8]. In this way, DRO could effectively leverage data information and optimize the expected value without any presumption about the probability distribution of uncertainties, which stochastic programming typically relies on. On the other hand, DRO can avoid conservatism by incorporating partial stochastic information, which is disregarded by the classical robust optimization.

An alternative streamline of data-driven optimization is the statistical robust optimization, which utilizes data, hypothesis tests, and likelihood to construct the statistically interpretable uncertainty set at a high confidence level [9, 10, 11]. The resulted models avoid the conservatism of some prior robust methods by ruling out unrealistic distributions while maintaining robustness for any statistically likely outcomes.
Practically speaking, the underlying distribution of uncertainties may be intrinsically complicated and vary under different conditions. It is rather challenging to choose the type of uncertainty sets by prior knowledge when one is faced with high-dimensional uncertainties. The procedure even becomes harder by tuning the coefficients and further evaluating its divergence with the hypothesis testing. Developing uncertainty sets based on historical data can be viewed as an unsupervised learning problem [10]. By exploiting the statistical properties of clustering approaches, the data-driven uncertainty set is covered with the fraction of data. Shang et al. [10] reformulated the robust model by adjusting only one parameter. Controlling this parameter is tantamount to controlling the conservatism and excluding outliers. Another attempt is made by Zhang et al. [12] to remove unnecessary uncertain scenarios in the uncertainty sets by generating cutting planes.

Three important criteria play significant roles in prescribing an intelligent robust model for a problem contaminated by uncertain parameters: 1) conservatism-aversion, 2) reliability, and 3) computational tractability of the robust counterparts. Figure 1 illustrates the very essence of why a decision maker would prefer a “treated” uncertainty set. Figure 1.a addresses the first criterion, i.e., the real-world data increases the rate of risk.
acceptance of $U_1$; hence, it modifies it into a smaller and bounded ellipsoid, $U_2$. On the other side, Figure 1.b exemplifies the second criterion, the reliability. The worst case of the interval uncertainty set w.r.t positivity of uncertain parameters lies in $w_1 \in U_1$; on the other hand, the realized information suggests that the real worst-case should be in $w_2 \in U_2$. Considering the reliability could make the model more conservative than the prior counterpart. Therefore, the conservatism of the two models should be compared in the light of the realized worst-case.

**DISTRIBUTIONALLY ROBUST OPTIMIZATION**

The fundamental assumption of stochastic programming (SP) is that the underlying statistical distribution of uncertain parameters is either defined or approximated with some degree of accuracy. SP models generally seek reliable solutions avoiding conservatism; however, they often lose their computational attractiveness. Practically speaking, for example, a monte-carlo based procedure called sample average approximation (SAA) can be adopted as a standard approach for solving intractable SP models. On the other hand, while SAA enjoys the computational convenience and asymptomatic convergence guarantee, it is known to result in unstable solutions.

In contrast, a tractable alternative to SP is robust optimization (RO), where all puts the weight of uncertainty on the parameters’ support set. The computational advantages of RO induced models are out-of-question [5, 13]; however, RO does not exploit the distributional characteristics of an uncertain parameter, and they often produce conservative and unreliable solutions.

To fill this gap, distributionally robust optimization (DRO), which can be traced back to the 1950s [14], was progressed. In DRO, the uncertainty set of a parameter is so modified that it also encompasses the probability distribution or the statistical moments, aside from the support set, of the
uncertain parameter. It is not too peculiar that some researchers embrace DRO as, in fact, a marriage of SP and RO [15].

Consider the following SP problem:

\[
\min_{x \in X} E_F [g(x, \xi)]
\]  

where \( x \in \mathbb{R}^n \) is a decision variable, \( X \subseteq \mathbb{R}^m \) is a convex set, \( \xi \in \mathbb{R}^m \) is a random vector with distribution \( F \), and \( g(x, \xi) \) is a convex utility function in \( x \) for a given \( \xi \). In SP, it is assumed that distribution \( F \) is at hand or can be estimated through sample data. However, in many practical cases, it is cumbersome to reach or approximate a genuine distribution given limited data samples. DRO suggests that the support and the moments of the distribution \( F \) be incorporated into some uncertainty set \( U \). In other words, the uncertainty set \( U \) mimics the functional and structural properties of the distribution \( F \) in the simplest possible form. According to RO, the worst case outcome of the expected utility function (1) is derived as follows:

\[
\min_{x \in X} \max_{F \in U} E_F [g(x, \xi)]
\]  

The functionality of moment-based uncertainty sets is comparable to that of the classical ellipsoidal uncertainty sets. Suppose that the support, mean, and covariance of \( \xi \) is known explicitly. The mean of \( \xi \) can be considered in an ellipsoid set with the center \( \mu \), and the covariance matrix \( \Sigma \). Therefore, the distributional uncertainty set is described as \( U(S, \mu, \Sigma) \) accounting for the convex support \( S \), \( \mu \) in the interior of \( S \), and positive definiteness of \( \Sigma \). The description of \( U \) is as follows:

\[
U(S, \mu, \Sigma) = \left\{ F \left| \begin{array}{l}
\mathbf{P}(\xi \in S) = 1 \\
E_F[\xi] = \mu \\
E_F \left[ (\xi - \mu) (\xi - \mu)^T \right] \leq \Sigma
\end{array} \right\} \right.
\]
Delage and Ye [8] proved that the robust counterpart of the Problem (2) according to $U$ can be formulated as a semidefinite program (SDP).

**Theorem 1.** Given the uncertainty set $U$, if $g(x, \xi)$ is continuous and differentiable in $x$, the robust counterpart of the Problem (2) is as follows:

\[
\begin{align*}
\text{minimize} & \quad t + \mu^T p + \langle \Sigma + \mu^T \mu, Q \rangle \\
\text{s.t.} & \quad t + \xi^T p + \xi^T Q \xi \geq g(x, \xi), \quad \forall \xi \in S \\
& \quad t \in \mathbb{R}, \quad p \in \mathbb{R}^m, \quad Q \in \mathbb{R}^{m \times m} \\
& \quad Q \succeq 0, \quad x \in X
\end{align*}
\]  

(4a)

where $\langle \cdot, \cdot \rangle$ is the inner product defined by $\langle A, B \rangle = \sum_{i,j} A_{ij} B_{ij}$.

**Proof.** The inner maximization problem of (2) according to $U$ can be reformulated as the following problem:

\[
\begin{align*}
\text{maximize} & \quad \int_S g(x, \xi)dF(\xi) \\
\text{s.t.} & \quad \int_S dF(\xi) = 1 \\
& \quad \int_S \xi dF(\xi) = \mu \\
& \quad \int_S (\xi - \mu)(\xi - \mu)^T dF(\xi) \preceq \Sigma \\
& \quad Q \succeq 0, \quad x \in X
\end{align*}
\]  

(5)

The last constraint of (5) can be rearranged as follows:

\[
\int_S \xi \xi^T dF(\xi) \preceq \Sigma + \mu \mu^T
\]  

(6)

The resulted problem (5) can be considered as a conic linear program. Applying the lagrangian duality with the lagrangian multipliers $t, p, \text{ and } Q$ over (5) yields:
\[ L(t, p, Q, dF) = t + \mu^T p + \left\langle \Sigma + \mu^T \mu, Q \right\rangle + \int_S \left( g(x, \xi) - t - \xi^T p - \xi^T Q \xi \right) dF(\xi) \]  

(7)

Then the original Problem (2) becomes:

\[
\begin{align*}
\text{minimize} & \quad \max_{dF(\xi) \geq 0} L(t, p, Q, dF) = \\
\text{minimize} & \quad \left\{ t + \mu^T p + \left\langle \Sigma + \mu^T \mu, Q \right\rangle + \right. \\
& \left. \max_{dF(\xi) \geq 0} \int_S \left( g(x, \xi) - t - \xi^T p - \xi^T Q \xi \right) dF(\xi) \right\} \\
\end{align*}
\]

(8a)

(8b)

(8c)

which equals to

\[
\begin{align*}
\text{minimize} & \quad \left\{ t + \mu^T p + \left\langle \Sigma + \mu^T \mu, Q \right\rangle \right. \\
& \left. \text{s.t.} \quad g(x, \xi) - t - \xi^T p - \xi^T Q \xi \leq 0, \quad \xi \in S \right\} \\
\end{align*}
\]

(9)

The Problem (9) is completely dependent on the support \( S \). In what follows, we limit \( S \) into a polyhedral set.

**Remark 1.** To preserve the convex optimization class of the Problem (4) under classical uncertainty sets, it is assumed that \( g(x, \xi) \) be a piecewise affine convex function in \( \xi \), i.e.,

\[ g(x, \xi) = \gamma_k^\xi(x)^T \xi + \gamma_k^0(x) \]

(10)

where \( \gamma_k(x) = \left( \gamma_1^k(x), \ldots, \gamma_m^k(x) \right) \) and \( \gamma_k^0(x) \) are affine in \( x \) for \( k = 1, \ldots, K \). Note that \( K \) is the number of terms in including the uncertain parameter.
Theorem 2. Given the support $S$ be a polyhedral set subject to $\text{Int}(S) \neq \emptyset$, i.e. $S = \{\xi | A\xi \leq b\} \neq \emptyset$ with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^{n}$, the problem (4) is reduced, according to Remark 1, to the following problem:

\[
\begin{align*}
&\text{minimize} & & t + \mu^T p + \left(\Sigma + \mu^T \mu\right) Q \\
&\text{s.t.} & & \left[\begin{array}{c}
t - \gamma^0_k(x) - \lambda_k^T b \\
p - \gamma_k(x) + A^T \lambda_k \\
Q
\end{array}\right] \succeq 0 \\
& & & k = 1, \ldots, K \\
& & & Q \succeq 0, \quad x \in X \\
& & & \lambda_k \in \mathbb{R}^n_+, \quad k = 1, \ldots, K
\end{align*}
\]

where $\lambda_k$ is the lagrangian multiplier.

Proof. Applying the worst-case approach to the realization of $\xi$ in constraint set (4a) [16], we obtain:

\[
\sup_{\xi \in \mathbb{R}^n, \ A\xi \leq b} \left\{\gamma^0_k(x) + \gamma_k(x)^T \xi - t - \xi^T p - \xi^T Q \xi \leq 0\right\}
\]

The latter is equivalent to the following:

\[
\exists \lambda_k \geq 0, \sup_{\xi \in \mathbb{R}^n} \left\{\gamma^0_k(x) + \gamma_k(x)^T \xi - t - \xi^T p - \xi^T Q \xi + \lambda_k (A\xi - b) \leq 0\right\}
\]

where $\lambda_k$ is the lagrangian multiplier. Now, we are able to transform the Inequality (13) into the form $v^T \psi \nu \geq 0$ (arranging the Inequality (13) in terms of $\xi$) where $\nu = \left(1 \quad \xi \right)^T$, the proof continues as follows:
\[
\exists \lambda_k \geq 0, \quad v^T \psi_k v \geq 0, \quad \forall k = 1, \ldots, K \quad \Rightarrow \quad (14a)
\]

\[
\exists \lambda_k \geq 0, \quad \psi_k \succeq 0, \quad \forall k = 1, \ldots, K \quad \Rightarrow \quad (14b)
\]

\(\psi_k\) is a PSD matrix in the following form:

\[
\psi_k = \begin{bmatrix}
t - \gamma_k^*(x) - \lambda_k^T b & \frac{(p - \gamma_k^*(x) + A^T \lambda_k)^T}{2} \\
\frac{(p - \gamma_k^*(x) + A^T \lambda_k)^T}{2} & Q
\end{bmatrix} \succeq 0
\quad (15)
\]

where \(\lambda_k \in \mathbb{R}^n, \quad \forall k = 1, \ldots, K\). □

**Example 1.** The standard portfolio optimization model is also known as the Markowitz’ portfolio model is formulated as follows:

\[
\begin{align*}
\text{maximize} & \quad x^T \tilde{r} \\
\text{s.t.} & \quad 1x = 1 \\
& \quad x \in \mathbb{R}^m
\end{align*}
\quad (16)
\]

where \(r\) is the return vector and \(x\) is the decision vector of investment in the set of securities \(I\). It is also assumed that the return vector \(r\) is uncertain with the support set \(S_r = \left\{ \xi^{(r)} \mid \sum_{i=1}^{m} a_i \xi_i^{(r)} \leq b \right\}\). According to Theorem 1, the convex utility function is adopted as follows:

\[
\begin{align*}
\text{maximize} & \quad g(x, \xi^{(r)}) = x^T \tilde{r} \\
\text{subject to} & \quad g(x, \xi^{(r)}) \\
& \quad (17)
\end{align*}
\]

Hence, DRO of (16) can be deduced as follows:

\[
\begin{align*}
\text{maximize} & \quad \min_{F^{r(\xi)}} \mathbb{E}_{F^{r(\xi)}} \left( g(x, \xi^{(r)}) \right) \\
\text{subject to} & \quad x \in \mathbb{R}^m
\end{align*}
\quad (18)
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 $\Sigma^{(r)}$.

First, the utility function should be defined in terms of a linear piecewise function, $\gamma^0(x) = 0$ and $\gamma(x) = x^T \xi^{(r)} = x^T \tilde{r}$. Note that we drop the index $K$, since it equals to one. Given $|I| = 3$, three random vectors with their corresponding means and covariance matrix are generated.

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

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

The resulted SD problem is as follows:

\[
\begin{aligned}
\text{minimize} & & t + 35.80p_1 + 69.50p_2 + 50.15p_3 + 8811.7q_{11} + 8482.8q_{12} + 8630.6q_{13} + 8482.8q_{21} + 9043.8q_{22} + 8700.7q_{23} + 8630.6q_{31} + 8700.7q_{32} + 8724.1q_{33} \\
\text{s.t.} & & \begin{bmatrix}
\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \lambda \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \\
2
\end{bmatrix}^T \begin{bmatrix}
q_{11} & q_{12} & q_{13} \\
q_{21} & q_{22} & q_{23} \\
q_{31} & q_{32} & q_{33}
\end{bmatrix} \geq 0 \\
1x &= 1 \\
Q &\succeq 0
\end{aligned}
\]
The polyhedral parameters $A$ and $b$ can be generated through the convex hull of the random vector $\xi^{(r)}$. Here, it is assumed that $A = (2 \hspace{1em} 1 \hspace{1em} 1)$ and $b = 100$ the result is $x_i^* = 0.5$, $x_2^* = x_3^* = 0.25$. It is suggested that sensitivity analysis can be helpful in determining reliable intervals for the parameters $A$ and $b$.

The DRO induced models in conjunction with other convex uncertainty sets (e.g., ellipsoidal uncertainty set) could be also challenging to solve. Moreover, they lose their practicality when the utility function cannot be divided into piecewise elements, which preserves convexity in $x$.

**CLUSTERING ORIENTED UNCERTAINTY SETS**

The distributional uncertainty sets are strongly dependent upon the distribution or the moment information of uncertain parameter. In real-world applications, the underlying distribution of uncertain parameter may be intrinsically complicated and the moment information is not at hand. Hopefully, the era of big data analytics paves promising ways to make decisions under uncertainty by exploiting the massive collection of the realized data. Shang et al. [10] proposed as an effective data-driven approach for robust optimization with the aid of support vector clustering (SVM). SVM can be adopted to estimate the support of uncertain parameters and to construct decent uncertainty sets from random data samples. Although the computational advantageous of the robust counterparts induced by clustering based uncertainty sets has not been thoroughly investigated, this class of methods allows the user to readily control the conservatism and they produce reliable solutions in comparison with models induced by the classical uncertainty sets.

Consider the following maximization problem with the uncertain parameter $\tilde{a}$.

$$\max_{x \in X} \hspace{1em} c^T x \hspace{1em} (19)$$
Data Driven Robust Optimization

\[ \max_{\tilde{a} \in U(\nu, D)} \tilde{a}^T x \leq b \]

where \( N \) samples \( D = \{ \xi_i^{(\tilde{a})}, \forall i = 1, \ldots, N \} \) are collected as the realization of the uncertain parameter \( \tilde{a} \) and the regularization parameter \( \nu \) is chosen based on the degree of conservatism. Note that the superscript of \( \xi_i^{(\tilde{a})} \) is dropped onwards for convenience. To characterize the uncertainty set \( U(\nu, D) \), it is required to determine which individual sample lies in the boundary of supporting vectors. The following supporting vector problem answers to the latter:

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j K(\xi_i, \xi_j) - \sum_{i=1}^{N} \alpha_i K(\tilde{\xi}_i, \xi_i) \\
\text{s.t.} & \quad \|\alpha\| = 1 \\
& \quad \alpha_i \leq 1/N\nu, \quad \forall i = 1, \ldots, N
\end{align*}
\]

(20)

where \( \{i \mid \alpha_i > 0, \forall i = 1, \ldots, N\} \) is the set of all supporting vectors; in this way, the outliers are also determined. \( K(\bullet, \bullet) \) is a kernel function and usually replaced with a tractable expression with a high accuracy regime [17].

**Proposition 1.** Given \( \alpha_i, \forall i = 1, \ldots, N \), the random vector with the condition \( P\{\tilde{a} = \xi_i\} = 1/N \), \( \forall i = 1, \ldots, N \) the kernel function \( K(u, v) = \sum_{k=1}^{N} \min(u_k, v_k) \geq 0 \), the uncertainty set \( U(\nu, D) \) is deduced as follows:

\[
U(\nu, D) = \left\{ \xi \mid \sum_{\forall i | \alpha_i > 0} \alpha_i \|Q(\xi - \hat{\xi}_i)\| \leq \sum_{\forall i' | \alpha_{i'} > 0} \alpha_{i'} \|Q(\xi_{i'} - \hat{\xi}_i)\|, \quad \forall i' | 0 < \alpha_{i'} < 1/N\nu \right\}
\]

(21)

where \( Q \) is the whitening matrix and directly calculated from the covariance matrix, \( Q = \Sigma^{-1/2} \).
**Proof.** The result can be readily deduced from expression (37) in [10]. □

**Theorem 3.** The uncertainty set $U(\nu, D)$ is bounded and non-empty with $0 \leq \nu < 1$.

**Proof.** The result is Direct from Theorem 1 in [10]. □

**Theorem 4.** The robust counterpart of the problem (19) w.r.t $0 \leq \nu < 1$ is as follows:

\[
\begin{align*}
&\text{maximize} \quad c^T x \\
&\text{s.t.} \quad \sum_{\forall i|\alpha_i > 0} (\mu_i - \lambda_i) Q \xi_i + \eta \beta \leq b \\
&\quad \sum_{\forall i|\alpha_i > 0} Q(\lambda_i - \mu_i) + x = 0 \tag{22} \\
&\quad \lambda_i + \mu_i = \eta \alpha_i 1, \quad \forall i | \alpha_i > 0 \\
&\quad \lambda_i, \mu_i \in \mathbb{R}^n \quad \forall i | \alpha_i > 0 \\
&\quad \eta \geq 0
\end{align*}
\]

where,

\[
\beta = \min_{\forall i | \xi_i \in [-\beta, \beta]} \sum_{\forall i|\alpha_i > 0} \alpha_i \|Q(\xi_i - \xi_i)\| \tag{23}
\]

**Proof.** Plugging the uncertainty set $U$ into the problem (18) and replacing $\bar{a}$ with the random vector $\xi_i$ with the condition $P \{ \bar{a} = \xi_i \} = 1/N, \forall i = 1, \ldots, N$, the Problem (19) becomes as follows:

\[
\begin{align*}
&\text{maximize} \quad c^T x \\
&\text{s.t.} \quad \max_{\xi} \left\{ \sum_{\forall i|\alpha_i > 0} \alpha_i \|Q(\xi_i - \xi_i)\| \leq \sum_{\forall i|\alpha_i > 0} \alpha_i \|Q(\xi_i - \xi_i)\|, \right. \\
&\quad \left. \forall i' | 0 < \alpha_i < 1/N \nu \right\} \leq b
\end{align*}
\]
Hence, the inner maximization Problem (24) is rearranged as follows:

$$\max_{\xi, z_i} \xi^T x$$

s.t. $\sum_{i \mid \alpha_i > 0} \alpha_i 1z_i \leq \beta$

$$Q\xi - z_i \leq Q\xi_i$$

$$-Q\xi - z_i \leq -Q\xi_i$$

$$\xi \in \mathbb{R}^M, z_i \in \mathbb{R}^M_+, \forall i \mid \alpha_i > 0$$

It is also immediate that $\beta$ is the infimum of all values of $\sum_{i \mid \alpha_i > 0} \alpha_i \|Q(\xi_i - \xi_i)\|$ over the set $\{ \forall i' \mid 0 < \alpha_i' < 1/N \nu \}$. Then, the Problem (24) is dualized by introducing the lagrangian multipliers $\mu_i$, $\lambda_i$, and $\eta$:

$$\min_{\lambda_i, \mu_i, \eta} \sum_{i \mid \alpha_i > 0} (\mu_i - \lambda_i)^T Q\xi_i + \eta \beta$$

s.t. $\sum_{i \mid \alpha_i > 0} Q(\lambda_i - \mu_i) + x = 0$

$$\lambda_i + \mu_i = \eta \alpha_i 1, \quad \forall i \mid \alpha_i > 0$$

$$\lambda_i, \mu_i \in \mathbb{R}^n_+, \forall i \mid \alpha_i > 0$$

$$\eta \geq 0$$

According to Theorem 4, the feasible region of the Problem (26) is bounded and non-empty w.r.t $0 \leq \nu < 1$. It is immediate that the primal and the dual have a unique optimal solution. Hence, the objective values of both coincide and that completes the proof. $\square$

In what follows, we sketch the steps required to build the robust counterpart according to the SVM method:
Step 1. Random vector for the uncertain parameter

Generate $\xi_i \in \mathbb{R}^{M \times N}$ based on the following probability criteria (Note that $x^*$ is the optimal decision vector w.r.t nominal values of the uncertain parameters):$$P\left\{ \tilde{a}^T x^* \leq \tilde{b} \right\} \geq 1 - \nu \quad \text{and} \quad P\{ \tilde{a} = \xi_i \} = 1/N, \quad i = 1, \ldots, N$$

Step 2. Whitening matrix of samples

$$\Sigma = \frac{1}{N-1} \left[ \sum_{i=1}^{N} \xi_i (\xi_i)^T - \left( \sum_{i=1}^{N} \xi_i \right) \left( \sum_{i=1}^{N} \xi_i \right)^T \right], \quad Q = (q_1, q_2) = \Sigma^{-\frac{1}{2}}$$

Step 3. Kernel matrix

$$\inf_{\gamma_k} \left\{ \gamma_k > \max_i q_k^T \xi_i - \min_i q_k^T \xi_i \right\} \left| K(\xi_i, \xi_j) \geq 0 \right\}, \quad \forall k$$

$$K(\xi_i, \xi_j) = \sum_{k=1}^{M} \gamma_k - \left\| Q(\tilde{\xi}_i - \tilde{\xi}_j) \right\|, \quad \forall i, j$$

Step 4. Set of support vector

$$\text{minimize} \quad \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j K(\xi_i, \xi_j) - \sum_{i=1}^{N} \alpha_i K(\xi_i, \xi_i)$$

s.t. \quad ||\alpha|| \leq 1

$$\alpha_i \leq 1/N \nu, \quad \forall i$$

Step 5. Robust counterpart
maximize \( c^T x \)

\[
\begin{align*}
\text{s.t.} \quad & \sum_{\forall i: \alpha_i > 0} (\mu_i - \lambda_i)^T Q \xi_i + \eta \beta \leq b \\
& \sum_{\forall i: \tilde{\alpha}_i > 0} Q (\lambda_i - \mu_i)^T + x = 0 \\
& \mu_i + \lambda_i = \eta \alpha_i \mathbf{1}, \quad \forall i: \alpha_i > 0 \\
& \left( \mu, \lambda \right) \in \mathbb{R}_{+}^{M \times N}, \quad x \in X, \quad \eta \geq 0 \\
& \beta = \min_{\forall i: \tilde{\alpha}_i < 1/N} \sum_{\forall i: \alpha_i > 0} \alpha_i \|Q(\xi_i - \xi_j)\|_1
\end{align*}
\]

**Example 2.** Consider the following maximization problem:

\[
\begin{align*}
\text{minimize} \quad & 2x_1 + 3x_2 \\
\text{s.t.} \quad & a_1 x_1 + a_2 x_2 \leq 5 \\
& x_i \geq 0, \quad \forall i \in \{1, 2\} \\
& \hat{\mu}_1 = 2, \quad \hat{\mu}_2 = 1 \\
& \hat{\Sigma} = \text{cov}(a_1, a_2) = \begin{pmatrix} 1 & 1.5 \\ 1.5 & 3 \end{pmatrix}
\end{align*}
\]

Five \((N=5)\) sample vectors are generated based on a bivariate normal distribution w.r.t \(\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}\), and \(\nu = 0.25\).

\[
\xi = \begin{pmatrix} 0.65 & 5.03 & 2.72 & 1.93 & 2.71 \\ 0.79 & 7.44 & 5.37 & 4.12 & 5.29 \end{pmatrix}, \quad Q = \begin{pmatrix} 2.16 & -1.34 \\ -1.34 & 0.92 \end{pmatrix}
\]

Note that \(Q\) is calculated from diagonalization of \(\Sigma\). We now set the infimum of \((\gamma_1, \gamma_2)\) as \((2.23, 1.43)\). The kernel matrix \(K\) is now calculated as
\[
K(\xi_i, \xi_j) = 2.23 + 1.43 - \begin{pmatrix} 2.16 & -1.34 \\ -1.34 & 0.92 \end{pmatrix}(\xi_i - \xi_j)
\]

\[
\begin{pmatrix} 
3.66 & 2.90 & 0.55 & 0.62 & 0.69 \\
2.90 & 3.66 & 0.24 & 0.31 & 0.38 \\
0.55 & 0.24 & 3.66 & 3.55 & 3.52 \\
0.62 & 0.31 & 3.55 & 3.66 & 3.53 \\
0.69 & 0.38 & 3.52 & 3.53 & 3.66 \\
\end{pmatrix}
\]

We can simply check that \( K \geq 0 \). Set of all supporting vectors and the boundary of supporting vectors are adopted as \( \{1, 2, 3, 4\} \) and \( \{1, 2, 4\} \), respectively. Finally, the LP robust counterpart is constructed as follows:

minimize \( x_1 + 3x_2 \)

s.t.

\[
\sum_{i|\xi_i > 0} \left( \begin{array}{c} \mu_{i1} - \lambda_{i1} \\ \mu_{i2} - \lambda_{i2} \end{array} \right) \left( \begin{array}{cc} q_{i1} & q_{i2} \\ q_{21} & q_{22} \end{array} \right) \left( \begin{array}{c} \xi_{i1} \\ \xi_{i2} \end{array} \right) + 1.75 \eta \leq 5
\]

\[
\sum_{i|\lambda_{i} > 0} \left( \begin{array}{c} q_{i1} \\ q_{21} \end{array} \right) \left( \begin{array}{cc} \mu_{i1} - \lambda_{i1} \\ \mu_{i2} - \lambda_{i2} \end{array} \right) + \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\] \hspace{1cm} (28)

\[
\begin{pmatrix} \mu_{i1} - \lambda_{i1} \\ \mu_{i2} - \lambda_{i2} \end{pmatrix} = \begin{pmatrix} \eta \alpha_i \\ \eta \alpha_i \end{pmatrix}, \quad \forall i| \alpha_i > 0
\]

**Cutting Uncertainty Space**

Zhang et al. [12] incorporated cutting planes into uncertainty sets to limit the support of an uncertain parameter after realization. Without the loss of computational attractiveness, the robust formulations induced by the cutting sets can reduce the conservatism and improve the robustness.
**Remark 2.** Given $C$ number of cutting hyperplanes with the gradient vector $Q \in \mathbb{R}^{C\times d}$ and the intercept $d \in \mathbb{R}^C$, the hyperplane $h(\xi)$ is defined over the random vector $\xi$ as follows:

$$h_i(\xi) = Q\xi_i + d \Rightarrow h_i(\xi) = q_{ij}\xi_j + d_c,$$

$$\forall i, \forall j \in |J_i|, \forall c \in \{1, \ldots, C\}$$

where $|J_i|$ is the number of the uncertain parameter in the $i$th constraint.

**Remark 3.** Consider the uncertain parameter $\hat{a}$ modeled as a symmetric and bounded and takes value in $[\hat{a} - \bar{a}, \hat{a} + \bar{a}]$; hence, we define the random vector variable $\xi = (\hat{a} - \bar{a})/\hat{a}$, which follows an unknown but symmetric distribution in $[-1, +1]$. We draw the worst-case situation on $\hat{a}$ w.r.t $U(h)$ in the following LP problem:

$$\begin{align*}
\max_{x \in \mathcal{X}} & \quad c^T x \\
\text{s.t.} & \quad \bar{a} x + \max_{\xi \in U(h)} \{ \hat{a} \xi \} \leq b_i, \quad \forall i
\end{align*}$$

(29)

**Remark 4.** The inner maximization of the Problem (29) is not bounded; intuitively, it is observed that the feasible region constructed by $U(h)$ cannot guarantee the boundedness. To address this issue, without disturbing its computational attractiveness, it is preferable to combine $U(h)$ with a classical uncertainty set\(^1\), e.g., interval uncertainty sets.

$$U_\infty(h) = \left\{ \xi \mid \|\xi\|_\infty \leq 1, \quad h(\xi) + d \geq 0 \right\}$$

(30)

\(^1\) The classical uncertainty sets are well-known for convexity, boundedness, and solidity.
Theorem 5. The robust counterpart of the problem (29) w.r.t $U_{\infty}(h)$ is as follows:

\[
\begin{align*}
\text{maximize} & \quad c^T x \\
\text{s.t.} & \quad \hat{a}_i x + \|\hat{a}_i x + Q^T \tau\|_\infty + d \tau \leq b_i, \quad \forall i \\
& \quad x \in X, \tau \in \mathbb{R}^c
\end{align*}
\]  

(31)

Proof. The inner maximization Problem of (29) w.r.t $U_{\infty}(h)$ is deduced as follows:

\[
\begin{align*}
\text{maximize} & \quad \hat{a}_i \xi, x \\
\text{s.t.} & \quad Q \xi + d \geq 0 \\
& \quad P_{\infty} \xi + p_{\infty} \geq 0 \\
& \quad \xi \in \mathbb{R}^{j_i}
\end{align*}
\]  

(32)

where $P_{\infty} = \left[ I_{|j_i|}^1, 0_{|j_i|, 1} \right]$, $p_{\infty} = \left[ 0_{|j_i|, 1}, 1 \right]$, and $K_{\infty} = \left\{ x \in \mathbb{R}^{1|j_i|} : \|x\|_\infty \leq t \right\}$.

Hence, we define the dual variable $y_i \in \left[ z_i, w_i \right] \in \mathbb{R}^{1|j_i|}$ and the dual problem of (32) as

\[
\begin{align*}
\text{minimize} & \quad p^\tau y_i + d^\tau \tau \\
\text{s.t.} & \quad P_{\infty}^\tau y_i - Q^\tau \tau = \hat{a}_i x, \quad \forall i \\
& \quad y_i \in K_{\infty}^* = K_i \\
& \quad \tau \in \mathbb{R}^c
\end{align*}
\]  

(33)
which is simplified as

\[
\begin{align*}
\min_{\tau, z, w_i} & \quad w_i + d^T \tau \\
\text{s.t.} & \quad z_i = \tilde{a}_i x + Q^T \tau, \quad \forall i \\
& \quad \|z_i\| \leq w_i, \quad \forall i \\
& \quad \tau \in \mathbb{R}_+.
\end{align*}
\]

then the explicit form of (34) can be formulated as

\[
\min_{\tau} \left\{ \|\tilde{a}_i x + Q^T \tau\| + d^T \tau \right\}. \quad (35)
\]

Note that the latter expression is completely dependent on $\tau$. It is immediate that the Primal model (30) and the Dual model (33) are simultaneously bounded and non-empty; hence, their objective coincides and that completes the proof.

**Theorem 5.** The robust counterpart of the Problem (29) w.r.t $U_2(h)^2$ is an SOCP\(^2\) as follows:

\[
\begin{align*}
\max_{x, \tau} & \quad c^T x \\
\text{s.t.} & \quad \tilde{a}_i x + \lambda + d \tau \leq b_i, \quad \forall i \\
& \quad \|\tilde{a}_i x + Q^T \tau\|^2 \leq \lambda^2 \\
& \quad x \in X, \tau \in \mathbb{R}_+^C.
\end{align*}
\]

**Proof.** The proof is the same as $U_\infty(h)$ in Theorem 4 except the dual variable should be enclosed in a second-order cone, $y_i \in K_2 = K_2$. \(\Box\)

\(^2\) Combination of ellipsoidal and cutting hyperplane uncertainty sets.

\(^3\) Second-order conic programming.
Remark 5. The cutting planes’ parameters $Q$ and $d$ can be approximated through the following minimization problem:

$$
\begin{align*}
\text{minimize} \quad & \max_{\eta_m, \eta'_{m}} \eta_{m} \\
\text{s.t.} \quad & \sum_{j} q'_{ij} s_{j}^{(m)} + d'_{c} \geq \eta_{m} (1 - \varepsilon), \quad \forall \, m, \forall \, c \\
& \sqrt{\sum_{j} q'^{2}_{ij}} \prod_{j} q'_{ij} / \prod_{j} q_{(c+1)j} \leq 0, \quad \forall \, c = \{1, \cdots, C - 1\} \\
& |d'_{c} - d'_{c+2}| \geq \frac{\sum \hat{a}_{ij}}{|J_i|}, \quad \forall \, c = \{1, \cdots, C - 2\}
\end{align*}
$$

where $s_{j}^{(m)}, j = 1, \cdots, |J|$ is the $m$th generated sample. The parameters $d' = [d'_1, \cdots, d'_C]$ and $Q' = [q'_{ij}]$, converting the support $\bar{a} \in [\bar{a} - \hat{a}, \bar{a} + \hat{a}] \rightarrow \xi \in [-1, +1]$, are mapped into $d$ and $Q$ as

$$
d = Q' \hat{a}_{i} \, + d', \quad \forall \, i
$$

$$
Q = Q' \ast (\hat{a}_{i} \, 1)^T, \quad \forall \, i
$$

where $1 \in \mathbb{R}^C$ and “•” is the component-wise product of two consistent matrices. Note that any increase in $\varepsilon$ gives rise to exclude more samples.

Example 3. Given the problem and the uncertain parameters in Example 2 and $\tilde{a}_1 = 2 \pm \xi_1$, $\tilde{a}_2 = 1 \pm 0.5 \xi_2$, 2000 samples are generated according to a bivariate normal distribution w.r.t $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\Sigma}$. We produce $C = 4$ cutting planes with $\varepsilon = 0.05$ as follows:
Data Driven Robust Optimization

\[
Q = \begin{pmatrix} -12 & 3.5 \\ -12 & 3.5 \\ 2 & 0.5 \\ 3 & 0.5 \end{pmatrix}, \quad d = \begin{pmatrix} -13 \\ 1.5 \\ 1 \\ -1 \end{pmatrix}
\]

The robust counterpart is deduced as

\[
\begin{align*}
\text{maximize} & \quad 3x_1 + 2x_2 \\
\text{s.t.} & \quad 2x_1 + x_2 + \left| x_1 - 12\tau_1 - 12\tau_2 + 2\tau_3 + 3\tau_4 \right| + \\
& \quad \left| 0.5x_2 + 3.5\tau_1 + 3.5\tau_2 + 0.5\tau_3 + 0.5\tau_4 \right| \\
& \quad -13\tau_1 + 1.5\tau_2 + \tau_3 - \tau_4 \leq 5 \\
& \quad x \in \mathbb{R}^p_+, \quad \tau \in \mathbb{R}^c_+
\end{align*}
\]

REFERENCES


Biographical Sketch

Name: Mohammad Namakshenas

Affiliation: Iran University of Science and Technology
Education: Industrial Engineering
Business Address: School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran

Research and Professional Experience:

1. Professional
   1.1. Project management
   1.2. Dashboard designing and reporting using QlikView and SQL
   1.3. Studied and developed Software sizing, and Software metrics and Cost estimation of agile projects Guidelines; implementing ISO software sizing standard
   1.4. Agile software development project management
   1.5. Implemented document center for online documents in the Microsoft Sharepoint server 2013 and prepared a comprehensive guideline
   1.6. Feasibility studies of implementing blockchain technologies in open supply chains and logistic networks
   1.7. Oversaw documentation of comprehensive business process guidelines
   1.8. Run mind-mapping sessions and scenario analysis
   1.9. Data Mining in the aerospace industry

2. Academic
   2.1. Teaching Assistant
      2.1.1. Taught in part advanced robust optimization theories and solution methods
      2.1.2. Instructed computational optimization, IBM Ilog Cplex, and Gurobi
2.1.3. Taught advanced integer optimization techniques including constraint programming
2.1.4. Taught integer programming concepts in scheduling
2.1.5. Taught and instructed data-driven optimization with Python and Gurobi
2.1.6. Taught in part convex optimization and semi-definite programming
2.1.7. Instructed CVX optimization package on convex optimization problems
2.1.8. Lecturer
2.1.9. Taught system analysis, project Control, scheduling theories courses
2.1.10. Taught and instructed Matlab Toolboxes

2.2. Running Workshops
2.2.1. Taught and instructed advanced OPL language using IBM Ilog Cplex Optimization Studio at Bu Ali Sina University, Hamedan, IR
2.2.2. Taught and instructed OPL language using IBM Ilog Cplex Optimization Studio at Elmo Farhang and Shahed University, Tehran, IR
2.2.3. Taught, instructed, and oversaw advanced OPL language using IBM Ilog Cplex Optimization Studio, Gurobi optimization package, and advanced object-oriented optimization using Python platform at Iran University of Science and Technology, Tehran, IR

2.3. Editorial Assistant

Professional Appointments: --
Honors:
1. Ranked 3rd out of 24 peers in M.Sc.
2. Awarded 8160 EUR excellence scholarship for a short term study by the Ministry of Science, and Research and Technology

Publications from the Last 3 Years:

Moahammad Namakshenas and Mir Saman Pishvae


Name: Mir Saman Pishvae

Affiliation: Iran University of Science and Technology  
(Associate Prof.)

Education: Industrial Engineering (PhD)

Business Address: School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran

Research and Professional Experience:

1. Professional
   1.2. Manager of strategic and operational planning department, Iran Ministry of Health and Medical Education (2010-2011).
   1.5. Faculty member (assistant Prof.), Department of School of Industrial Engineering, Iran university of Science and Technology (from March 2013 by now)
   1.6. Deputy of socio-cultural affairs, School of Industrial Engineering, Iran university of Science and Technology (from Sep. 2013 by now).
   1.7. The reengineering and strategic change project of Nuclear Medicine Center of Shariati Hospital, Master analyst, 2009.
   1.8. The design of AVAPESHK (Medical Device Manufacturer Company) company distribution system, Project manager, 2010.
   1.9. The Strategic planning project of AVAPESHK (Medical Device Manufacturer Company), Project manager, 2010.
   1.10. The Strategic planning project of Elm-Gostaran-Alavi charity organization, Project manager, 2011.
1.11. Industrial Management Institute Consultants Network project (shams.imi.ir), Main manager, 2012.
1.12. The Strategic planning and business model design project of DANA (A large venture capital), Project manager, 2013.
1.16. The strategic planning, business model design and logistics network planning for Tashkhis networked Lab., 2018

2. Academics
2.1. Supply chain and logistics management:
   2.1.1. Introduction to supply chain management
   2.1.2. Supply and procurement management
   2.1.3. International logistics
   2.1.4. Reverse logistics management
2.2. System analysis, design and optimization:
   2.3. Robust optimization
   2.4. System analysis
   2.5. System dynamics
   2.6. Quantitative methods in Industrial Engineering
   2.7. Operations research in healthcare
2.8. Innovation management:
   2.8.1. Innovation theories, models, processes and policies
   2.8.2. Open innovation
   2.8.3. Crowdsourcing
2.9. Strategic management:
   2.9.1. Introduction to strategic management
   2.9.2. Corporate strategy
   2.9.3. Strategy implementation and change management
Professional Appointments: --

Honors:
Among 1 percent distinguished researchers in the world (2018)

Publications from the Last 3 Years:


42. Babazadeh R., Razmi J., Pishvaee M.S., (2016). Sustainable cultivation location optimization of the Jatropha curcas L. under


50. Kohneh J.N., Teymoury E., Pishvae M.S., (2016). Blood products supply chain design considering disaster circumstances (Case study:


59. Fazli-Khalaf M., Chaharsooghi S.K., Pishvae M.S., (2015). A reliable closed-loop supply chain network design under uncertainty:


