GLOBALIZED DISTRIBUTIONALLY ROBUST OPTIMIZATION BASED ON SAMPLES∗

YUEYAO LI† AND WENXUN XING‡

Abstract. It is known that the set of perturbed data is key in robust optimization modelling. Distributionally robust optimization (DRO) is a methodology used for optimization problems affected by random parameters with uncertain probability distribution. In terms of the information of the perturbed data, it is essential to estimate an appropriate support set of the probability distribution in formulating DRO models. In this paper, we introduce two globalized distributionally robust optimization (GDRO) models which choose a core set based on data and a sample space containing the core set to balance the degree of robustness and conservatism at the same time. The degree of conservatism can be controlled by the expected distance of random parameters from the core set. Under some assumptions, we further reformulate several GDRO models into tractable semi-definite programs. In addition, numerical experiments are provided showing the relationship between the optimal objective values of the GDRO models and the size of the sample space and the core set.

Key words. distributionally robust optimization, sample space, semi-definite program, computationally solvable

MSC codes. 90C15, 90C22

1. Introduction. Distributionally robust optimization (DRO) model is generally formulated as

\begin{equation}
\min_{x \in X} \max_{P \in D} E_P[h(x, \xi)],
\end{equation}

where \( x \in \mathbb{R}^n \) is the decision vector, \( X \subseteq \mathbb{R}^n \) is a closed convex set of feasible solutions, \( h(x, \xi) \) is a convex function in \( x \) that depends on the random vector \( \xi \in \mathbb{R}^p \), \( E_P[\cdot] \) is the expectation taken with respect to \( \xi \) and \( D \) is the ambiguity set. The only information we know about \( \xi \) is a set of samples of it.

There are wide range of ambiguity sets constructed by different methods, such as matrix moment constraints [21], Wasserstein metric [14] and polynomial density functions [10]. Here we consider the ambiguity set of matrix moment constraints [11] as follows,

\begin{equation}
D = \left\{ P \in \mathcal{P} \mid \begin{array}{l}
E_P[|\xi - \mu_0|^T \Sigma_0^{-1}(E_P[\xi] - \mu_0)] \leq \gamma_1 \\
E_P[(\xi - \mu_0)(\xi - \mu_0)^T] \leq \gamma_2 \Sigma_0 \\
P(\xi \in \Xi) = 1
\end{array} \right\},
\end{equation}

where \( \mathcal{P} \) is the set of all probability measures on the measurable space \( (\Xi, \mathcal{B}(\Xi)) \), \( \Xi \subseteq \mathbb{R}^p \) is the sample space, \( \mathcal{B}(\Xi) \) is the Borel \( \sigma \)-algebra on \( \Xi \), \( \xi : \Xi \to \mathbb{R}^p \) is a vector of random variables in probability space \( (\Xi, \mathcal{B}(\Xi), P) \) satisfying \( \xi(\omega) = \omega \), \( \mu_0 \) is the sample mean, \( \Sigma_0 \) is the sample covariance matrix and \( \gamma_1 \) and \( \gamma_2 \) are estimated by statistical method to guarantee that the true mean and covariance matrix are contained with high probability. Details of determining \( \gamma_1 \) and \( \gamma_2 \) can be seen in Appendix B.

∗April 29th.

†Department of Mathematical Sciences, Tsinghua University, Beijing 100084, People’s Republic of China (liyueyao18@mails.tsinghua.edu.cn).

‡Department of Mathematical Sciences, Tsinghua University, Beijing 100084, People’s Republic of China (wx-ning@mail.tsinghua.edu.cn).
In this paper, we present a new globalized distributionally robust optimization (GDRO) framework, in which a core set based on given samples is fixed and the sample space of probability distribution is globalized. GDRO can be regarded as an extension of the traditional DRO. Here two kinds of GDRO models are provided. The first one called GDRO1 embodying the globalization in the objective function in (1.1) is

\[
\min_{x \in X} \max_{P \in D} E_P[h(x, \xi) - \min_{\xi' \in Y} \phi(\xi, \xi')],
\]

where the core set \(Y\) is a convex compact set containing core samples and \(\phi : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}\) is a closed, jointly convex and nonnegative function for which \(\phi(\xi, \xi) = 0\) for all \(\xi \in \mathbb{R}^p\). The ambiguity set \(D\) remains to be (1.2).

Besides, the set \(Y\) and \(\Xi\) are defined by

\[
Y = \{\mu_0 + A\zeta | \zeta \in Z_1\},
\]

\[
\Xi = \{\mu_0 + A\zeta | \zeta \in Z_2\},
\]

where \(Z_1\) is a given nonempty, convex and compact set with \(0 \in \text{ri}(Z_1)\) and \(Z_2\) is also convex and contains \(Z_1\).

We observe that

\[
E_P[h(x, \xi) - \min_{\xi' \in Y} \phi(\xi, \xi')]
= \int_{\Xi} (h(x, \xi) - \min_{\xi' \in Y} \phi(\xi, \xi'))dP
= \int_{Y} h(x, \xi)dP + \int_{\Xi \setminus Y} (h(x, \xi) - \min_{\xi' \in Y} \phi(\xi, \xi'))dP,
\]

where \(\min_{\xi' \in Y} \phi(\xi, \xi')\) can be regarded as a penalty item. If \(P(\Xi \setminus Y) = 0\), i.e. the support set of \(P\) is contained in \(Y\), the objective function is consistent with that in the classical DRO model (1.1). Otherwise, the value of the expectation over \(\Xi \setminus Y\) is subtracted by \(\int_{\Xi \setminus Y} \min_{\xi' \in Y} \phi(\xi, \xi')dP\). According to the construction of \(Y\), core samples \(\xi_1, \ldots, \xi_N\) are contained in \(Y\). Then we can believe that the probability \(P(\Xi \setminus Y)\) is extremely small. It is reasonable to reduce the influence of the expectation over \(\Xi \setminus Y\). Thus, the possible worst-case expectation is decreased and the degree of conservatism is reduced.

The second model called GDRO2 embodies the globalization in a constraint,

\[
\min_{x \in X} \max_{P \in D_1} E_P[h(x, \xi)],
\]

where the ambiguity set is defined as

\[
D_1 = \left\{ P \in \mathcal{P} | \begin{array}{l}
(E_P[\xi] - \mu_0)^T \Sigma_0^{-1} (E_P[\xi] - \mu_0) \leq \gamma_1 \\
E_P[(\xi - \mu_0)(\xi - \mu_0)^T] \leq \gamma_2 \Sigma_0 \\
E_P[\min_{\xi' \in Y} \phi(\xi, \xi')] \leq d_0 \\
P(\xi \in \Xi) = 1
\end{array} \right\}.
\]

All notations keep the same meanings as above. It is obvious that the performance of the probability distribution on \(\Xi \setminus Y\) is limited by the constraint

\[
E_P[\min_{\xi' \in Y} \phi(\xi, \xi')] \leq d_0.
\]
It forces that the probability of that $\xi$ is farther away from $Y$ is smaller.

The key idea of formulating the above two models is to realize the idea of globalization by introducing the distance function $\phi$. Besides, GDRO models are able to control the degree of conservatism by adjusting $\phi$. The term controlling the degree of conservatism $\min_{\xi' \in Y} \phi(\xi, \xi')$ appears in the objective function in GDRO1, while it is transferred to a constraint in the ambiguity set in GDRO2. Therefore, GDRO2 has one more parameter $d_0$ than GDRO1 to control conservatism.

The idea of globalization comes from the globalized robust optimization (GRO) introduced by Ben-Tal et al. [3]. Their motivation is to find a way to balance conservatism and robustness in robust optimization. Since the robust optimization was proposed, there were some articles [3, 4] that aimed to discuss the trade-off between conservatism and robustness. The robustness requires constraints to be feasible for all parameter values in the uncertainty set. The resulting optimal solutions are over-conservative in the sense that decision makers give up too much optimality of the nominal problem as [4] said. In extreme cases, there may be no feasible solution. In order to decrease the degree of conservatism and guarantee the existence of feasible solutions, we have to shrink the uncertainty set, resulting in the loss of robustness at the same time.

To deal with this, the GRO approach uses two uncertainty sets: an inner uncertainty set and an outer uncertainty set. Constraints in the GRO model are forced to be feasible for all parameter values in the inner uncertainty set but allowed to be infeasible for parameter values in the outer uncertainty set, where the violation is controlled by the distance of the parameter value and the inner uncertainty set. Compared with the RO approach, the GRO approach allows decision makers to consider the infection of more values of uncertain parameters even though the parameter values in the outer uncertainty set are less likely to occur in the real life. At the same time, the rise of robustness does not lead to much higher degree of conservatism.

The first published study of DRO was introduced by Scarf [20], but had not attracted much attention until robust optimization was popular. DRO handles the uncertainty by using some information like mean and variance of the samples under the assumption of not knowing the exact probability distribution. However, it is not so easy to estimate the support set of the probability distribution based on given samples. Generally, a set $\Phi$ known to contain the support set is used in their ambiguity sets, which is assumed to be a convex compact set, such as in [11, 18]. Under the assumption that $\Phi$ is bounded, the ambiguity set cannot contain the true probability distribution if its support set is the whole space $\mathbb{R}^p$. On the other side, it may lead to over-conservatism if $\Phi$ is directly taken as $\mathbb{R}^p$. In consideration of the similarity of the uncertainty set in RO and the support set in DRO, it is natural to introduce the idea of globalization to the DRO problem. That is why we propose such globalized distributionally robust optimization models.

In consideration of $\xi(\omega) = \omega$ for all $\omega \in \Xi$, the support set is contained in the sample space $\Xi$. Besides, there is no problem to let the sample space to be the same as the support set if the latter is known. Hence, we choose to use the term “sample space” which is more rigorous in expression, instead of “a set known to contain the support set”.

In GDRO models, we set a core set $Y$ containing core samples. In application, they can be eliminated if there are abnormal samples. In other words, the core set only include the key information which may depend on the preference of decision makers. As for the sample space, we hope it include as much information of samples as possible. Thus, we can balance the degree of robustness and conservatism by adjusting the core set $Y$, the sample space $\Xi$ and the distance function $\phi$. Besides, each of the above GDRO models with appropriate settings of $\Xi$, $Y$ and $\phi$ can be equivalently reformulated as a semi-definite program (SDP) under the assumption that $h(x, \xi)$ is piecewise linear in $\xi$. 

There are a lot of works concentrating on DRO under moment uncertainty. Delage and Ye [11] discuss the complexity of such DRO problem and reformulate the DRO model of a portfolio optimization problem to a SDP when the set $\Phi$ is an ellipsoid or the whole space. Liu et al. [18] consider a DRO model under moment uncertainty and obtain tractable reformulations when the set $\Phi$ is the whole space or a convex polyhedral set. Cheng et al. [8] provide an approximation method based on principal component analysis for solving DRO problems with moment-based ambiguity sets, the calculation time of which is reduced in the large-scale case compared with solving SDP reformulation directly. As for more related works, we refer to [19, 21, 22].

There are also some works similar to our ideas. Li and Kwon [17] present a penalty framework based on DRO for the portfolio problem. They assume that the investor has a particularly strong preference for a specific measure $Q$. Therefore, it is reasonable to subtract $kH(P|Q)$ from the objective function, where $H$ is the penalty function and $k$ is the penalty coefficient. In their moment-based framework, the single measure $Q$ is replaced by a set of first two moments $(\mu_c, \Sigma_c) = \{(\mu_i, \Sigma_i)|i \in C\}$ and the penalty item becomes $P_k(\mu, \Sigma|\mu_c, \Sigma_c)$. The SDP reformulation can be generated under appropriate choices of the confidence region $(\mu_c, \Sigma_c)$ and the penalty function. Besides, other globalized distributionally robust optimization model has appeared in [12]. Different from what we care about, they apply globalization to the ambiguity set. Their main idea is to construct two ambiguity sets: an inner ambiguity set of “normal range” of distributions and an outer ambiguity set of “physically possible” distributions. Expectation constraints in their model are forced to be feasible for all distributions in the inner ambiguity set but allowed to be infeasible for distributions in the outer ambiguity set, where the violation is controlled by the distance of the distribution and the inner ambiguity set.

The remainder of this paper is organized as follows. In Section 2, we derive the computationally tractable reformulations of the two GDRO models under some assumptions. Section 3 presents some numerical results. Section 4 provides some conclusions.

**Notations.** $x$ is the decision vector and $X$ is the feasible region of $x$. $\Xi \subseteq \mathbb{R}^p$ is the sample space and $\mathcal{B}(\Xi)$ is the Borel $\sigma$-algebra on $\Xi$. $\mathcal{P}$ is the set of all probability measures on the measurable space $(\Xi, \mathcal{B}(\Xi))$. $\xi$ is a random vector satisfying $\xi(\omega) = \omega$ in probability space $(\Xi, \mathcal{B}(\Xi), P)$. $E_P[\cdot]$ represents the expectation with respect to $\xi$. $\mu_0$ is the sample mean and $\Sigma_0$ is the sample covariance matrix. $Y$ is the core set in GDRO models. $\mathbb{R}^n$ is the set of real $n$-dimensional vectors, $\mathbb{S}^n$ is the set of real $n \times n$ symmetric matrices and $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ are the set of positive semi-definite and positive definite matrices in $\mathbb{S}^n$, respectively. $e_p \in \mathbb{R}^p$ is a vector of 1 and $I_p \in \mathbb{R}^{p \times p}$ is the identity matrix.

2. Computationally tractable GDRO reformulations. Under some assumptions, corresponding tractable SDPs are reformulated considering two settings of the sample space of GDRO models. All proofs of the lemmas and theorems in this section are left in Appendix A.

2.1. Preparations. It is a common way to write the dual of the original model to reformulate the problem, which is also an essential part of deriving computationally tractable GDRO reformulations. Delage and Ye [11] have proved the following lemma.

**Lemma 2.1** (Delage and Ye [11]). Suppose $\gamma_1 > 0$, $\gamma_2 > 0$ and $\Sigma_0 \in \mathbb{S}^p_{++}$ in the ambiguity set (1.2) and that $h(x, \xi)$ is integrable for all $P \in \mathcal{D}$ for a fixed $x \in \mathbb{R}^n$. Then DRO (1.1) with the...
ambiguity set (1.2) can be reformulated as

\[
\min_{x,\Lambda,q,t} \quad t + \Lambda \cdot (\gamma_2 \Sigma_0 + \mu_0 \mu_0^T) + \sqrt{\gamma_1}||\Sigma_0^{1/2}(q + 2\Lambda \mu_0)||_2 + q^T \mu_0 \\
\text{s.t.} \quad h(x,\xi) - \xi^T \Lambda \xi - q^T \xi \leq t \quad \forall \xi \in \Xi, \\
\Lambda \in S^n_+, \quad q \in \mathbb{R}^p, \quad t \in \mathbb{R}, \\
x \in X \subseteq \mathbb{R}^n,
\]

(2.1)

where \(\Sigma_0^{1/2}\) is one decomposition of \(\Sigma_0\) such that \(\Sigma_0 = (\Sigma_0^{1/2})^T \Sigma_0^{1/2}\).

This lemma is key for the following SDP reformulations. (2.1) decouples the min/max operations in the objective and the expectation of random vector \(\xi\). Now it is a semi-infinite programming without uncertainty.

In order to simplify the models and derive computationally tractable counterparts, we make the following assumption.

**Assumption 2.2.** \(h(x,\xi)\) is a piecewise linear concave function in \(\xi\) as following,

\[
h(x,\xi) = \max_{k \in \{1,2,...,K\}} \{a_k(x)^T \xi + b_k(x)\},
\]

where \(a_k(x)\) and \(b_k(x)\) are linear in \(x\) for all \(k\).

It is easy to verify that \(h(x,\xi)\) satisfies the integrable condition in Lemma 2.1 under this assumption. Piecewise linear function is widely used as the objective function in the portfolio optimization [16, 7, 17] and inventory problems [2, 15]. Besides, some utility functions can be approximated accurately using piecewise linear functions in the portfolio optimization as stated in [11]. Especially, optimization problems with the worst case CVaR measure as the objective function can be regarded as DRO problems satisfying Assumption 2.2 if the loss function \(L(x,\xi)\) is linear in \(\xi\) and \(x\).

\[
\text{WC-CVaR}_\varepsilon(L(x,\xi)) = \inf_{\beta \in \mathbb{R}} \left( \beta + \frac{1}{\varepsilon} \sup_{P \in \mathcal{D}} E_P[(L(x,\xi) - \beta)^+] \right).
\]

Here \(h(x,\xi) = (L(x,\xi) - \beta)^+ = \max\{L(x,\xi) - \beta, 0\}\).

### 2.2. Globalization in objective function – GDRO1.

In this subsection, we recall GDRO1 (1.3) with the ambiguity set (1.2). For a fixed \(x \in \mathbb{R}^n\), assume that \(h(x,\xi) - \min_{\xi \in \gamma} \phi(\xi,\xi')\) is integrable for all \(P \in \mathcal{D}\). According to Lemma 2.1, GDRO1 (1.3) with the ambiguity set (1.2) and \(\gamma_1 > 0, \gamma_2 > 0, \) and \(\Sigma_0 \in S^n_{++}\) can be equivalently reformulated as

\[
\min_{x,\Lambda,q,t} \quad t + \Lambda \cdot (\gamma_2 \Sigma_0 + \mu_0 \mu_0^T) + \sqrt{\gamma_1}||\Sigma_0^{1/2}(q + 2\Lambda \mu_0)||_2 + q^T \mu_0 \\
\text{s.t.} \quad h(x,\xi) - \xi^T \Lambda \xi - q^T \xi - t \leq \min_{\xi \in \gamma} \phi(\xi,\xi') \quad \forall \xi \in \Xi, \\
\Lambda \in S^n_+, \quad q \in \mathbb{R}^p, \quad t \in \mathbb{R}, \\
x \in X \subseteq \mathbb{R}^n.
\]

Especially, if we have no idea about the estimation of the sample space, then it is reasonable to set

\[
Y = \{\xi \in \mathbb{R}^p | (\xi - \bar{\mu})^T \Sigma_0^{-1}(\xi - \bar{\mu}) \leq \tilde{\gamma}\},
\]

\(\Xi = \mathbb{R}^p\).
Theorem 2.3. Let $\phi(\xi, \xi') = \theta \|\xi - \xi'\|_2$ for a given $\theta \geq 0$ and $Y$ and $\Xi$ be defined as (2.3). Under Assumption 2.2, (2.2) is equivalent to the following SDP problem when $\gamma_1 > 0$, $\gamma_2 > 0$ and $\Sigma_0 \in S^n_{++}$,

$$\min_{x, \Lambda, q, t, v_k, z_k} t + \Lambda \cdot (\gamma_2 \Sigma_0 + \mu_0 \mu_0^T) + \sqrt{\gamma_1} \|\Sigma_0^{1/2} (q + 2\Lambda \mu_0)\|_2 + q^T \mu_0$$

s.t. \begin{align*}
\Lambda &\geq 0 \\
\frac{1}{2} (v_k + q - a_k(x))^T &\leq b_k(x) - \mu^T v_k - z_k \\
\sqrt{\gamma_1} \|\Sigma_0^{1/2} v_k\|_2 &\leq z_k, \\
\|v_k\|_2 &\leq \theta,
\end{align*}

(2.4)

However, the case that the sample space is known to be bounded may also arise in the real life. We may not be able to know the exact shape and size of the sample space. In this case, the idea of globalization also make sense by setting

$$\begin{align*}
Y &= \{\xi \in \mathbb{R}^p | (\xi - \bar{\mu})^T \Sigma_0^{-1} (\xi - \bar{\mu}) \leq \bar{\gamma}_1\}, \\
\Xi &= \{\xi \in \mathbb{R}^p | (\xi - \bar{\mu})^T \Sigma_0^{-1} (\xi - \bar{\mu}) \leq \bar{\gamma}_2\},
\end{align*}$$

(2.5)

where $0 < \bar{\gamma}_1 < \bar{\gamma}_2$.

Theorem 2.4. Let $\phi(\xi, \xi') = \theta \|\xi - \xi'\|_2$ for a given $\theta \geq 0$ and $Y$ and $\Xi$ be defined as (2.5). Under Assumption 2.2, (2.2) is equivalent to the following SDP problem when $\gamma_1 > 0$, $\gamma_2 > 0$, and $\Sigma_0 \in S^n_{++}$,

$$\min_{x, \Lambda, q, t, v_k, w_k, z_{1k}, z_{2k}} t + \Lambda \cdot (\gamma_2 \Sigma_0 + \mu_0 \mu_0^T) + \sqrt{\gamma_1} \|\Sigma_0^{1/2} (q + 2\Lambda \mu_0)\|_2 + q^T \mu_0$$

s.t. \begin{align*}
\Lambda &\geq 0 \\
\frac{1}{2} (v_k + w_k + q - a_k(x))^T &\leq b_k(x) - \bar{\mu}^T (v_k + w_k) - z_{1k} - z_{2k} \\
\sqrt{\gamma_1} \|\Sigma_0^{1/2} v_k\|_2 &\leq z_{1k}, \\
\sqrt{\gamma_2} \|\Sigma_0^{1/2} w_k\|_2 &\leq z_{2k}, \\
\|v_k\|_2 &\leq \theta, \\
v_k, w_k &\in \mathbb{R}^p, \ z_{1k}, z_{2k} \in \mathbb{R}, \ 1 \leq k \leq K, \\
\Lambda &\in S^n_{++}, \ q \in \mathbb{R}^p, \ t \in \mathbb{R}, \\
x &\in X \subseteq \mathbb{R}^n.
\end{align*}

(2.6)

It is obvious that we are able to control the degree of conservatism by adjusting the coefficient $\theta$ in the distance function $\phi(\xi, \xi') = \theta \|\xi - \xi'\|_2$. If $\theta = 0$, the two GDRDRO models degenerate into the classical DRO models. In (2.6), we observe that it forces $w_k = 0$ and $z_{2k} = 0$ for all $k$ when the parameter $\bar{\gamma}_2$ tends to be infinite, resulting in (2.6) to be the same as (2.4).
2.3. Globalization in ambiguity set – GDRO2. In this subsection, we recall GDRO2 (1.4) with the ambiguity set (1.5). Similarly to Lemma 2.1, under assumptions that $\gamma_1 > 0$, $\gamma_2 > 0$ and $\mathbf{\Sigma}_0 \in \mathbb{S}^p_+$, and $h(x, \xi)$ is integrable for all $P \in \mathcal{D}$, the GDRO (1.4) with the ambiguity set (1.5) for a fixed $x \in \mathbb{R}^n$ can be reformulated as

\[
\begin{align*}
\min_{x, A, q, t, r} & \quad t + \mathbf{\Lambda} \cdot (\gamma_2 \mathbf{\Sigma}_0 + \mu_0 \mu_0^T) + \frac{1}{\sqrt{T}} \| \mathbf{\Sigma}_0^{1/2} (q + 2\Lambda \mu_0) \|_2 + q^T \mu_0 + d_0 r \\
\text{s.t.} & \quad h(x, \xi) - \xi^T \mathbf{\Lambda} \xi - q^T \xi - t \leq r \min_{\xi' \in \mathcal{Y}} \phi(\xi, \xi') \quad \forall \xi \in \Xi,
\end{align*}
\]

(2.7)

where

\[
\mathbf{\Lambda} \in \mathbb{S}^p_+, \quad q \in \mathbb{R}^p, \quad t \in \mathbb{R}, \quad r \geq 0,
\]

\[x \in X \subseteq \mathbb{R}^n.
\]

Similarly, we have the following as follows for the case (2.3) and (2.5).

**Theorem 2.5.** Let $\phi(\xi, \xi^\prime) = \theta \| \xi - \xi^\prime \|_2$ for a given $\theta \geq 0$ and $Y$ and $\Xi$ be defined as (2.3). Under Assumption 2.2, (2.7) is equivalent to the following SDP problem when $\gamma_1 > 0$, $\gamma_2 > 0$ and $\mathbf{\Sigma}_0 \in \mathbb{S}^p_+$,

\[
\begin{align*}
\min_{x, A, q, t, v_k, z_k, r} & \quad t + \mathbf{\Lambda} \cdot (\gamma_2 \mathbf{\Sigma}_0 + \mu_0 \mu_0^T) + \sqrt{T} \| \mathbf{\Sigma}_0^{1/2} (q + 2\Lambda \mu_0) \|_2 + q^T \mu_0 + d_0 r \\
\text{s.t.} & \quad \mathbf{\Lambda} \in \mathbb{S}^p_+, \quad q \in \mathbb{R}^p, \quad t \in \mathbb{R}, \quad r \geq 0,
\end{align*}
\]

(2.8)

\[
\begin{align*}
\sqrt{T} \| \mathbf{\Sigma}_0^{1/2} v_k \|_2 & \leq z_k, \\
\| v_k \|_2 & \leq r \theta, \\
v_k & \in \mathbb{R}^p, \quad z_k \in \mathbb{R}, \quad 1 \leq k \leq K, \\
\mathbf{\Lambda} & \in \mathbb{S}^p_+, \quad q \in \mathbb{R}^p, \quad t \in \mathbb{R}, \quad r \geq 0,
\end{align*}
\]

\[x \in X \subseteq \mathbb{R}^n.
\]

**Theorem 2.6.** Let $\phi(\xi, \xi^\prime) = \theta \| \xi - \xi^\prime \|_2$ for a given $\theta \geq 0$ and $Y$ and $\Xi$ be defined as (2.5). Under Assumption 2.2, (2.7) is equivalent to the following SDP problem when $\gamma_1 > 0$, $\gamma_2 > 0$ and $\mathbf{\Sigma}_0 \in \mathbb{S}^p_+$,

\[
\begin{align*}
\min_{x, A, q, t, r, v_k, w_k, z_{1k}, z_{2k}} & \quad t + \mathbf{\Lambda} \cdot (\gamma_2 \mathbf{\Sigma}_0 + \mu_0 \mu_0^T) + \frac{1}{\sqrt{T}} \| \mathbf{\Sigma}_0^{1/2} (q + 2\Lambda \mu_0) \|_2 + q^T \mu_0 + d_0 r \\
\text{s.t.} & \quad \mathbf{\Lambda} \in \mathbb{S}^p_+, \quad q \in \mathbb{R}^p, \quad t \in \mathbb{R}, \quad r \geq 0,
\end{align*}
\]

(2.9)

\[
\begin{align*}
\sqrt{T} \| \mathbf{\Sigma}_0^{1/2} v_k \|_2 & \leq z_{1k}, \\
\sqrt{T} \| \mathbf{\Sigma}_0^{1/2} w_k \|_2 & \leq z_{2k}, \\
\| v_k \|_2 & \leq r \theta, \\
v_k, w_k & \in \mathbb{R}^p, \quad z_{1k}, z_{2k} \in \mathbb{R}, \quad 1 \leq k \leq K, \\
\mathbf{\Lambda} & \in \mathbb{S}^p_+, \quad q \in \mathbb{R}^p, \quad t \in \mathbb{R}, \quad r \geq 0,
\end{align*}
\]

\[x \in X \subseteq \mathbb{R}^n.
\]
For the two GDRO models with two settings (2.3) and (2.5) of \(Y\) and \(\Xi\), we get four SDP problems (2.4), (2.6), (2.8), and (2.9), which are polynomially solvable. Then numerical experiments can be adopted to evaluate the optimal objective values of the problems influenced by the control parameter \(\theta\) and the two settings. The GDRO models with setting (2.3) are referred as the unbounded case, while the GDRO models with setting (2.5) are referred as the bounded case in the following section.

3. Numerical results. The portfolio optimization problem is used in our numerical experiments which has been previously studied by Delage and Ye [11] in this section. Given \(n\) options, \(\xi\) is a random vector of returns and \(E_P[u(\xi^T x)]\) is the expected utility with respect to the probability distribution \(P\). Given the ambiguity set \(D\), the distributionally robust portfolio optimization model is written as

\[
\max_{x \in X} \min_{P \in D} E_P[u(\xi^T x)],
\]

where

\[
X = \{ x \in \mathbb{R}^n | \sum_{i=1}^{n} x_i = 1, x \geq 0 \}.
\]

Here the same utility function as the one in [11] is considered, which is a piecewise linear function.

\[
u(\xi^T x) = \min_{k \in \{1, 2, ..., K\}} \{a_k \xi^T x + b_k\}.
\]

Reformulate problem (3.1) as a standard form:

\[
\min_{x \in X} \max_{P \in D} E_P[\max_{k \in \{1, 2, ..., K\}} -a_k \xi^T x - b_k].
\]

In our globalized distributionally robust optimization models, we set \(\phi(\xi, \xi') = \theta \| \xi - \xi' \|_2\).

In the numerical experiments, all data are randomly generated without considering the real background of the portfolio optimization problem. We take \(n = p = 10, K = 5\) and \(a_k, b_k\) being data of independently uniform distribution on \((0, 10)\). In the following subsections of numerical experiments, 30 samples are generated for the uncertain \(\xi\) of each configuration according to a given distribution and a small perturbation \(N(\epsilon e_p, 0.5I_p)\), where \(\epsilon \in (-0.5, 0.5), e_p \in \mathbb{R}^p\) is a vector of 1 and \(I_p \in \mathbb{R}^{p \times p}\) is the identity matrix. The small perturbation added to the samples aims to make difference between the means and variances from the computation and given distribution.

The sample mean \(\mu_0\) and the sample covariance matrix \(\Sigma_0\) are computed based on the samples. Theoretical determination of \(\gamma_1\) and \(\gamma_2\) is given in Appendix B. Here \(\gamma_1\) and \(\gamma_2\) as defined by (B.2) with \(\alpha = 0.05\) are used and computed in all numerical experiments due to the small sample size. The other parameters may have different settings in different experiments. For convenience, notations in Table 1 are used to represent the models in the following numerical experiments.

The influence of the parameter \(\theta\) and the size of the core set or sample space on the optimal objective value is considered in subsections 3.1 to 3.3. Discussions on the determination of the core set, the sample space and the control parameters are presented in the subsection 3.4 based on the sample data.
Let $\bar{\theta}$ as the true probability distribution. Then 30 samples with the small perturbation are generated. It is obvious that the optimal objective value of GDRO models is decreased with the increase of $\theta$. In Figure 1a, $v(GDRO1.1, \gamma_1) = v(DRO1)$ and $v(GDRO1.2, \gamma_1, 100 \gamma_1) = v(DRO2, 100 \gamma_1)$.

### 3.1. Selection of $\theta$ and sample space.

The uniform distribution on the box $[0, 20]^p$ is chosen as the true probability distribution. Then 30 samples with the small perturbation are generated. Let $\mu = \mu_0$ and $\bar{\gamma} = \gamma_1$ in (2.3), $\mu = \mu_0$, $\bar{\gamma}_1 = \gamma_1$ and $\bar{\gamma}_2 = 100 \gamma_1$ in (2.5) and $d_0 = 0.5$ in (1.5). The variation of the optimal objective value of GDRO models with the parameter $\theta$ is shown in Figure 1.

### Table 1

**Models in numerical experiments**

| notation | model | sample space $\Xi$ | core set $Y$ | optimal objective value |
|----------|-------|--------------------|--------------|-------------------------|
| DRO1     | $\min_{x \in X} \max_{P \in D} E_P[h(x, \xi)]$ | $\mathbb{R}^p$ | -           | $v(DRO1)$               |
| DRO2     | $\min_{x \in X} \max_{P \in D} E_P[h(x, \xi)]$ | $\mathcal{E}(\gamma)$ | -           | $v(DRO2, \gamma)$      |
| GDRO1.1  | $\min_{x \in X} \max_{P \in D} E_P[h(x, \xi) - \min_{\xi \in Y} \phi(\xi, \xi')]$ | $\mathbb{R}^p$ | $\mathcal{E}(\gamma)$ | $v(GDRO1.1, \gamma)$ |
| GDRO1.2  | $\min_{x \in X} \max_{P \in D} E_P[h(x, \xi) - \min_{\xi \in Y} \phi(\xi, \xi')]$ | $\mathcal{E}(\gamma_2)$ | $\mathcal{E}(\gamma_1)$ | $v(GDRO1.2, \gamma_1, \gamma_2)$ |
| GDRO2.1  | $\min_{x \in X} \max_{P \in D_1} E_P[h(x, \xi)]$ | $\mathbb{R}^p$ | $\mathcal{E}(\gamma_2)$ | $\mathcal{E}(\gamma_1)$ | $v(GDRO2.1, \gamma)$ |
| GDRO2.2  | $\min_{x \in X} \max_{P \in D_2} E_P[h(x, \xi)]$ | $\mathcal{E}(\gamma_2)$ | $\mathcal{E}(\gamma_1)$ | $v(GDRO2.2, \gamma_1, \gamma_2)$ |

Note: $\mathcal{E}(\gamma) = \{\xi \in \mathbb{R}^p | (\xi - \tilde{\mu})^T \Sigma_0^{-1} (\xi - \tilde{\mu}) \leq \gamma\}$.

![Figure 1](image1.png)

**Fig. 1.** The optimal objective values of GDRO models via $\theta$.

In Figure 1, $v(DRO1)$, $v(DRO2, 100 \gamma_1)$ and $v(DRO2, \gamma_1)$ are optimal objective values of the DRO model with sample spaces $\mathbb{R}^p$, $\mathcal{E}(100 \gamma_1)$ and $\mathcal{E}(\gamma_1)$, respectively. Thus, it can be seen that the optimal objective value decreases obviously as the size of sample space decreases. The smaller the sample space is, the better the optimal objective value and the lower degree of conservatism are.

It is obvious that the optimal objective value of GDRO models is decreased with the increase of $\theta$. In Figure 1a, $v(GDRO1.1, \gamma_1) = v(DRO1)$ and $v(GDRO1.2, \gamma_1, 100 \gamma_1) = v(DRO2, 100 \gamma_1)$.
when \( \theta = 0 \). Actually, GDRO1.1 degenerates into DRO1 and GDRO1.2 degenerates into DRO2 with \( \gamma = 100 \gamma_1 \) at this time. In other words, the core set does not work in this situation. When \( \theta \) becomes large enough, \( v(GDRO1.1, \gamma_1) = v(GDRO1.2, \gamma_1, 100 \gamma_1) = v(DRO2, \gamma_1) \). The worst case solution\(^1\) of the inner maximization problem of GDRO1 (1.3) may be a probability measure \( P \) whose support set is contained in the core set \( Y = \mathcal{E}(\gamma_1) \). Hence, GDRO1.1 and GDRO1.2 will have the same performance as the model DRO2 with \( \gamma = \gamma_1 \) if \( \theta \) is large enough.

Figure 1b has the same trend as Figure 1a. In Figure 1b, it is also the same as that in Figure 1a when \( \theta = 0 \). However, for the model GDRO2.1 and GDRO2.2, the optimal objective value will be closed to \( v(DRO2, \gamma_1) \) but will never reach it with \( \theta \) increasing.

The decision maker can control the degree of conservatism by choosing the sample space and the value of \( \theta \). For the optimal objective values of GDRO models, the supremum (i.e., the least conservative objective value) is \( v(DRO2, \gamma_1) \) while the infimum (i.e., the most conservative objective value) is \( v(DRO1) \) (\( v(DRO2, 100 \gamma_1) \)) for unbounded case (bounded case) in the range where the degree of conservatism can be adjusted. Besides, the gap between the supremum and the infimum determines the size of the adjustment range of \( \theta \). It also can be seen that \( v(GDRO1.1, \gamma_1) \) and \( v(GDRO1.2, \gamma_1, 100 \gamma_1) \) reach a same value with a relatively smaller \( \theta \) for the later. Therefore, the adjustment range of \( \theta \) for the bounded case is smaller than that for the unbounded case.

Actually, the determination of \( \theta \) in GDRO models depends on the optimal value range of the objective function and the preference of decision makers. For this instance, the optimal objective value of GDRO1.1 reaches \( v(DRO2, \gamma_1) \) at \( \theta = 4.4 \) and that of GDRO1.2 reaches \( v(DRO2, \gamma_1) \) at \( \theta = 1.4 \). For general problems, we can solve the following problem to get an appropriate value of \( \theta \) for GDRO1. The same also works for GDRO2 by replacing \( h(x, \xi) - \min_{\xi \in Y} \phi(\xi, \xi') \) with \( h(x, \xi) \) and the ambiguity set \( D (1.2) \) with \( D_1 (1.5) \).

\[
(3.3) \quad \min \theta \quad \text{s.t.} \quad \min_{x \in X} \max_{P \in D} E_P[h(x, \xi) - \min_{\xi \in Y} \phi(\xi, \xi')] \leq v + \varepsilon,
\]

where \( v = v(DRO2, \gamma) \) and \( \varepsilon \) depends on the preference of decision makers. Under our assumptions, (3.3) is also equivalently reformulated as a SDP model which is computationally tractable.

### 3.2. Selection of core set \( Y \) under sample space \( \mathbb{R}^p \)

In this subsection, the optimal objective values of GDRO1.1 and GDRO2.1 are used to show the influence of the radius of the core set when the sample space is \( \mathbb{R}^p \). The optimal objective value of DRO2 is selected as a lower bound for that of GDRO1.1 and GDRO2.1. Each set of 30 samples is got from each of the following three distributions with the small perturbation \( N(\epsilon e_p, 0.5 I_p) \), the normal \( N(10\epsilon e_p, 100 I_p) \), uniform on box \([0, 20]^p\) and uniform on ellipsoid \( \{\xi | (\xi - 10 \epsilon e_p)^T \Sigma^{-1} (\xi - 10 \epsilon e_p) \leq 200\} \), respectively, where \( \Sigma \) is a positive definite matrix.

Let \( \bar{\mu} = \mu_0 \) and \( \bar{\gamma} \) be 0.25, 0.5, 0.75, 1, 3, 25, 49, 100, 225, 400, 625 times of \( \gamma_1 \), respectively. We repeat the experiments 100 times and compute the average and the variance of optimal objective values. Tables 2 to 4 show the experiment results of DRO2, GDRO1.1 with \( \theta = 4 \), and GDRO2.1 with \( d_0 = 0.5, \theta = 4 \). Besides, the average of \( \theta \) solved by (3.3) with \( \varepsilon = 0.01 \) is presented in Table 5. The optimal objective value of DRO1 is not influenced by the radius because its sample space is the whole space. Hence, the average and variance of \( v(DRO1) \) are recorded as a note under every table.

\(^1\)The worst case solution of the inner maximization problem of (1.3) is \( \arg \max_{P \in D} E_P[h(x, \xi) - \min_{\xi \in Y} \phi(\xi, \xi')] \).
Table 2

| $\gamma/\bar{\gamma}/\bar{\gamma}$ | 0.25 | 0.5 | 0.75 | 1 | 9 | 25 | 49 | 100 | 225 | 400 | 625 |
|-----------------------------------|------|-----|------|---|---|-----|-----|-----|-----|-----|-----|
| Mean                             |      |     |      |   |   |     |     |     |     |     |     |
| DRO2                             | -14.3856 | -14.0043 | -13.7411 | -13.5306 | -13.4130 | -9.7738 | -9.4538 | -9.4538 | -9.4538 | -9.4538 | -9.4538 |
| GDRO1.1                          | -14.3607 | -13.9972 | -13.7383 | -13.5304 | -13.4130 | -9.7738 | -9.4538 | -9.4538 | -9.4538 | -9.4538 | -9.4538 |
| GDRO2.1                          | -13.9004 | -13.5552 | -13.3076 | -13.1463 | -13.1422 | -9.7738 | -9.4538 | -9.4538 | -9.4538 | -9.4538 | -9.4538 |
| Var                              |      |     |      |   |   |     |     |     |     |     |     |
| DRO2                             | 0.2360 | 0.2333 | 0.2385 | 0.2465 | 0.2465 | 0.7192 | 4.9440 | 3.7592 | 2.7055 | 2.6813 | 2.6813 |
| GDRO1.1                          | 0.2184 | 0.2265 | 0.2353 | 0.2462 | 0.2461 | 0.7192 | 4.9440 | 3.7592 | 2.7055 | 2.6813 | 2.6813 |
| GDRO2.1                          | 0.2126 | 0.2186 | 0.2280 | 0.2394 | 0.2392 | 0.6867 | 4.5862 | 3.4638 | 2.6867 | 2.6813 | 2.6813 |

Note. mean($v(DRO1)$) = −9.4538, var($v(DRO1)$) = 2.7084.

Table 3

| $\gamma/\bar{\gamma}/\bar{\gamma}$ | 0.25 | 0.5 | 0.75 | 1 | 9 | 25 | 49 | 100 | 225 | 400 | 625 |
|-----------------------------------|------|-----|------|---|---|-----|-----|-----|-----|-----|-----|
| Mean                             |      |     |      |   |   |     |     |     |     |     |     |
| DRO2                             | -14.3539 | -13.9644 | -13.6961 | -13.4818 | -13.4818 | -13.2725 | -11.3096 | -9.4861 | -9.2181 | -9.2159 | -9.2159 |
| GDRO1.1                          | -14.3283 | -13.9559 | -13.6922 | -13.4813 | -13.4812 | -13.2721 | -11.3096 | -9.4861 | -9.2181 | -9.2159 | -9.2159 |
| GDRO2.1                          | -13.8785 | -13.5271 | -13.2750 | -13.1087 | -13.1050 | -12.9009 | -11.0399 | -9.4026 | -9.2161 | -9.2159 | -9.2159 |
| Var                              |      |     |      |   |   |     |     |     |     |     |     |
| DRO2                             | 0.2193 | 0.2065 | 0.2140 | 0.2270 | 0.2270 | 1.4371 | 5.3689 | 3.8619 | 2.9346 | 2.9309 | 2.9309 |
| GDRO1.1                          | 0.1807 | 0.1940 | 0.2066 | 0.2209 | 0.2208 | 1.4358 | 5.3689 | 3.8619 | 2.9346 | 2.9309 | 2.9309 |
| GDRO2.1                          | 0.1817 | 0.1861 | 0.1961 | 0.2098 | 0.2094 | 1.3661 | 5.0032 | 3.6102 | 2.9322 | 2.9309 | 2.9309 |

Note. mean($v(DRO1)$) = −9.2159, var($v(DRO1)$) = 2.9605.

In Tables 2 to 4, the means of optimal objective values are all have the same non-decreased trend as $\bar{\gamma}$ increases, which means the larger radius of the core set is, the more conservative GDRO models are. The same trend under three different distributions is trivial because both DRO and GDRO models are not relative to the true distribution. Therefore, we can use GDRO1.1 or GDRO2.1 if it is uncertain that the sample space of the true probability distribution is bounded or not.

It is obvious that GDRO1.1 and GDRO2.1 both give smaller optimal objective values than DRO1. Hence, GDRO models can reduce the degree of conservatism compared with DRO1. Meanwhile, GDRO1.1 and GDRO2.1 provide smaller variances of the optimal objective values than DRO2 in most cases. Thus, GDRO models are less affected by the set of samples than DRO models.

In addition, $\theta$ solved by (3.3) in Table 5 is non-increased and tends to 0 as $\bar{\gamma}$ increases. Therefore, we can control the degree of conservatism in GDRO models by adjusting $\theta$ only when the radius of $Y$ is not extremely large. Besides, the variances of the optimal objective values when $\gamma/\gamma_1 \geq 25$ in Tables 2 and 3, $\gamma/\gamma_1 \geq 225$ in Table 4) are obviously larger than those when $\gamma/\gamma_1$ is small for all three models. Thus, GDRO models with smaller core set is more robust with respect to samples. In summary, we should only pay attention to the core data while constructing the core set $Y$.

3.3. Selection of core set under bounded sample space. In this subsection, the 30 samples are got from the uniform distribution on the box $[0, 20]^p$ with the small perturbation $N(\epsilon p, 0.5I_p)$ in the numerical experiments. According to the numerical results in subsection 3.1, we still fix $d_0 = 0.5$ in (1.5) and choose $\theta = 0.5$ here.

In Table 6, we fix the size $\gamma_1$ of the core set $E(\gamma_1)$ and change the size $\gamma_2$ of the sample space $E(\gamma_2)$ in GDRO1.2 and GDRO2.2. Besides, the radius of the sample space $E(\gamma)$ of DRO2 is also
Table 4
Uniform distribution on ellipsoid \{(\xi - 10e_p)^T \Sigma^{-1} (\xi - 10e_p) \leq 200\}

| γ/γ₁ | 0.25 | 0.5 | 0.75 | 1 | 9 | 25 | 49 | 100 | 225 | 400 | 625 |
|-------|------|-----|------|---|---|----|----|-----|-----|-----|-----|
| Mean  |     |     |      |   |   |    |    |     |     |     |     |
| DRO2  | -14.4224 | -14.2647 | -14.1542 | -14.0650 | -14.0650 | -14.0650 | -14.0599 | -13.7621 | -13.4019 | -13.3045 |
| GDRO1.1 | -14.4158 | -14.2625 | -14.1532 | -14.0649 | -14.0649 | -14.0648 | -14.0596 | -13.7620 | -13.4019 | -13.3045 |
| GDRO2.1 | -14.0284 | -13.8918 | -13.7908 | -13.7441 | -13.7369 | -13.7290 | -13.7201 | -13.7001 | -13.4563 | -13.2882 | -13.2844 |
| Var   |     |     |      |   |   |    |    |     |     |     |     |
| DRO2  | 0.0669 | 0.0664 | 0.0677 | 0.0695 | 0.0695 | 0.0695 | 0.0744 | 0.2595 | 0.2391 | 0.1957 |
| GDRO1.1 | 0.0652 | 0.0660 | 0.0677 | 0.0696 | 0.0696 | 0.0696 | 0.0745 | 0.2595 | 0.2391 | 0.1957 |
| GDRO2.1 | 0.0647 | 0.0675 | 0.0702 | 0.0729 | 0.0731 | 0.0732 | 0.0735 | 0.0780 | 0.1975 | 0.1791 | 0.1764 |

Note. mean(\(v(DRO_1)\)) = -13.2844, var(\(v(DRO_1)\)) = 0.1782.

Table 5
\(\theta\) solved by (3.3)

| \(\gamma/\gamma_1\) | 0.25 | 0.5 | 0.75 | 1 | 9 | 25 | 49 | 100 | 225 | 400 | 625 |
|---------------------|------|-----|------|---|---|----|----|-----|-----|-----|-----|
| Normal              | 3.9144 | 3.5688 | 3.4250 | 3.0183 | 3.0422 | 3.0010 | 2.2386 | 0.7087 | 0.0309 | 9.05E-10 | 1.06E-09 |
| Box                 | 3.8231 | 3.4642 | 3.3186 | 2.9223 | 2.9460 | 2.9046 | 2.0963 | 0.6029 | 0.0113 | 8.87E-10 | 1.17E-09 |
| Ellipsoid           | 3.6027 | 3.3241 | 3.2136 | 2.8214 | 2.8517 | 2.8846 | 2.9208 | 2.9753 | 2.4270 | 0.9938 | 0.2171 |

changed accordingly. Let \(\bar{\mu} = \mu_0\), \(\bar{\gamma}_1 = \gamma_1\) and \(\gamma_2\) be 1, 9, 25, 49, 100, 225, 400, 625 times of \(\gamma_1\), respectively. We repeat the experiments 100 times and compute the average and the variance of optimal objective values for each configuration.

Obviously, the optimal objective value increases with the increase of the radius of the sample space for all three models in Table 6. However, GDRO models provide smaller optimal values and hence reduce the degree of conservatism significantly, compared with the DRO model with the same sample space. Besides, GDRO models also provide smaller variances, meaning more robust with respect to different sets of samples than the DRO model.

When the radius of the sample space is small (\(\bar{\gamma}_2/\gamma_1 \leq 25\)), there is little difference between the objective values of GDRO1.2 and GDRO2.2. The difference between GDRO1.2 and GDRO2.2 becomes more obvious with the increase of the radius. Generally, GDRO2.2 provides smaller optimal values and smaller variances and shows better performance than GDRO1.2. Probably the reason is that GDRO2.2 is influenced by \(d_0\), which also controls the degree of conservatism and robustness in GDRO2.2.

In Table 7, we fix the size of the sample space (\(\mathbb{R}^p\) for GDRO1.1 and GDRO2.1, and \(\mathcal{E}(\bar{\gamma}_2)\) for GDRO1.2 and GDRO2.2) and change the radius (\(\sqrt{\gamma}\) for GDRO1.1 and GDRO2.1, and \(\sqrt{\gamma_1}\) for GDRO1.2 and GDRO2.2) of the core set in all four GDRO models. Besides, the radius of the sample space \(\mathcal{E}(\gamma)\) of DRO2 is also changed accordingly. Let \(\bar{\mu} = \mu_0\), \(\bar{\gamma}_2 = 225\gamma_1\) and \(\gamma_1\) be 0.25, 0.5, 0.75, 1, 9, 25, 49, 100 times of \(\gamma_1\), respectively. We repeat the experiments 100 times and compute the average and the variance of optimal objective values.

Obviously, the optimal objective value increases with the increase of the radius of the core set for all four GDRO models in Table 7. The DRO model is the least conservative because the sample space in the DRO model is the same as the core set in GDRO models. It can be seen that the bounded GDRO models (GDRO1.2 and GDRO2.2) are less conservative and more robust with respect to samples than the unbounded cases (GDRO1.1 and GDRO2.1), which shows that the sample space still play an important role in modelling. The variances of optimal objective values
of GDRO models are worse than those of DRO2 because the radius of the sample space of GDRO models is larger than that of DRO2. Similar to that in subsection 3.2, GDRO models with smaller core set is also more robust with respect to samples in the bounded case.

Therefore, a smaller core set and sample space are preferred in order to achieve a lower degree of conservatism.

3.4. Data-based GDROs. Obviously, all sample data are bounded in real world. If a bounded sample space is selected in the DRO modelling, the probability distribution with unbounded sample space is excluded, such as the normal distribution. Therefore, GDRO is an efficient tool which not only includes the probability distribution with the whole space as the sample space, but also controls the degree of conservatism by constructing the core set and adjusting the parameter $\theta$ or $d_0$.

When an unbounded sample space is considered for our GDRO modelling, the determination of the core set of samples is needed first. The minimal ellipsoid containing all samples with the variance matrix of the samples is used in our modelling. The following example shows our procedure of handling the portfolio optimization model (3.1). Here the normal distribution $N(10e_p, \frac{100}{3}I_p)$ is chosen as the true probability distribution and 30 samples are generated accordingly. For each sample, a small perturbation $N(\epsilon e_p, 0.5I_p)$ is added to the sample where $\epsilon \in (-0.5, 0.5)$. The mean $\mu_0$ and the covariance matrix $\Sigma_0$ of the samples are thus computed. Then the smallest $\gamma$ satisfying that all samples are included in the ellipsoid $\{\xi \in \mathbb{R}^p | (\xi - \mu_0)^T \Sigma_0^{-1} (\xi - \mu_0) \leq \gamma\}$ is computed, denoted as $\tilde{\gamma}$. If $\theta = 4$ and $d_0 = 0.5$ are fixed, then solve DRO1, DRO2, GDRO1.1 and GDRO2.1.

| $\gamma / \gamma_1: \gamma_2 / \gamma_1$ | 0.25 | 0.5 | 0.75 | 1 | 9 | 25 | 49 | 100 |
|----------------------------------------|------|-----|------|---|---|-----|-----|-----|
| Mean DRO2                              | -14.3210 | -13.9231 | -13.6490 | -13.4306 | -13.4306 | -13.2118 | -11.0446 | -9.2762 |
| GDRO1.1                                | -10.5824 | -10.5352 | -10.5159 | -10.4996 | -10.2376 | -9.9418 | -9.6131 | -9.1724 |
| GDRO2.1                                | -11.6007 | -11.2977 | -11.1412 | -11.1330 | -10.9953 | -10.6388 | -9.2100 | -9.0303 |
| GDRO1.2                                | -10.5919 | -10.5568 | -10.5364 | -10.5192 | -10.2476 | -9.9466 | -9.6154 | -9.1730 |
| GDRO2.2                                | -12.2025 | -11.8725 | -11.7011 | -11.6830 | -11.3672 | -10.7681 | -9.2130 | -9.0318 |
| Var DRO2                               | 0.2723 | 0.2723 | 0.3633 | 0.6044 | 0.5106 | 0.4000 | 0.3495 | 0.3276 |
| GDRO1.1                                | 0.1920 | 0.1869 | 0.2013 | 0.2013 | 0.8199 | 5.4549 | 3.7923 | 3.7923 |
| GDRO2.1                                | 1.5599 | 1.5212 | 1.5425 | 1.5608 | 1.8819 | 2.3088 | 2.8588 | 3.2493 |
| GDRO1.2                                | 0.2188 | 0.2498 | 0.2974 | 0.2978 | 0.3032 | 0.8085 | 2.9614 | 2.8832 |
| GDRO2.2                                | 1.6007 | 1.6028 | 1.6211 | 1.6370 | 1.9276 | 2.3347 | 2.8726 | 3.2532 |
| GDRO2.2                                | 0.2608 | 0.2922 | 0.3448 | 0.3466 | 0.3804 | 0.9548 | 2.9771 | 2.8928 |

*Table 6: Changing the radius of the sample space*

*Table 7: Changing the radius of the core set*
The means and variances of optimal objective values are shown in Table 8.

|                  | \(v(DRO1)\) | \(v(DRO2, \bar{\gamma})\) | \(v(DRO2, 25\bar{\gamma})\) | \(v(GDRO(1, \bar{\gamma})\) | \(v(GDRO(1, 25\bar{\gamma})\) | \(v(GDRO(2, \bar{\gamma})\) | \(v(GDRO(2, 25\bar{\gamma})\) |
|------------------|-------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| **Mean**         | -9.8150     | -13.4145                 | -8.9822                  | -13.4145                 | -8.9822                  | -13.0301                 | -8.9816                  |
| **Var**          | 3.8781      | 0.2573                   | 3.8838                   | 0.2573                   | 3.8838                   | 0.2542                   | 3.8786                   |

Due to the big gap between \(v(DRO1)\) and \(v(DRO2, \bar{\gamma})\), GDRO models make sense in application. Now there is a large alternative range for \(\theta\) to control the degree of conservatism when the core set is set \(E(\bar{\gamma}) = \{\xi \in \mathbb{R}^p | (\xi - \mu_0)^T\Sigma_0^{-1}(\xi - \mu_0) \leq \bar{\gamma}\}\). Here \(\theta = 4\) in GDRO models is fixed for our convenience of repeating the experiments. Actually, decision makers can solve (3.3) to get \(\theta\) in line with their own risk preference. If the core set is selected as \(E(25\bar{\gamma})\), then GDRO models are useless as the gap between \(v(DRO1)\) and \(v(DRO2, 25\bar{\gamma})\) is too small.

Suppose that it is known the sample space is bounded. Here the uniform distribution on the box \([0, 20]^p\) is chosen as the true probability distribution and 30 samples with the small perturbation as above are generated accordingly. Then the smallest \(\gamma\) satisfying that all samples are included in the ellipsoid \(\{\xi \in \mathbb{R}^p | (\xi - \mu_0)^T\Sigma_0^{-1}(\xi - \mu_0) \leq \gamma\}\) is computed, denoted as \(\gamma\). Let \(\theta = 0.5\) and \(\varrho_0 = 0.5\). The means and variances of optimal values are shown in Table 9.

|                  | \(v(DRO1)\) | \(v(DRO2, \bar{\gamma})\) | \(v(DRO2, 4\bar{\gamma})\) | \(v(DRO2, 25\bar{\gamma})\) | \(v(GDRO(1, 2, \bar{\gamma})\) | \(v(GDRO(1, 2, 4\bar{\gamma})\) | \(v(GDRO(1, 2, 25\bar{\gamma})\) | \(v(GDRO(2, 2, \bar{\gamma})\) | \(v(GDRO(2, 2, 4\bar{\gamma})\) | \(v(GDRO(2, 2, 25\bar{\gamma})\) |
|------------------|-------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| **Mean**         | -9.2564     | -13.5379                 | -11.0322                 | -9.2564                  | -12.0367                 | -10.3411                 | -8.9816                  | -13.0301                 | -12.5736                 | -11.2220                 |
| **Var**          | 2.8434      | 0.1745                   | 5.7760                   | 2.8434                   | 3.6655                   | 1.9666                   | 3.8781                   | 3.8838                   | 3.8786                   | 3.8786                   |

We observe that \(v(DRO2, 25\bar{\gamma})\) is extremely similar to \(v(DRO1)\). It is why we add the case \(4\bar{\gamma}\) in Table 9. For the two sample spaces of different sizes, the gap between \(v(GDRO(1, \bar{\gamma})\) and \(v(GDRO(1, 2, \bar{\gamma})\) or \(v(GDRO(2, \bar{\gamma})\) and \(v(GDRO(2, 2, \bar{\gamma})\) is different. Therefore, it is essential to estimate an appropriate sample space for the bounded case. If a bounded sample space cannot be estimated very well, the unbounded or a large bounded sample space like \(E(25\bar{\gamma})\) with the GDRO modelling is suggested in application. In fact, there is little difference between the results of bounded and unbounded cases if the bounded sample space is large enough.

In summary, the decision maker can determine the core set \(Y\) and the sample space \(\Xi\) firstly, then compute the optimal values of DRO with \(Y\) and \(\Xi\) as the sample space, respectively. If the gap between the two values is extremely small, there is no need to use the GDRO models. Conversely, GDRO models are good choice in controlling the degree of conservatism.

4. Conclusions. In this paper, we introduce the globalized distributionally robust optimization and derive the corresponding tractable GDRO models under the assumption that the objective function is piecewise linear. Compared with DRO, the GDRO approach allows wider range of the sample space and controls the degree of conservatism at the same time. Besides, the GDRO approach is more flexible in adjustment of conservatism by adjusting the parameter \(\theta\). In application, the decision maker can determine the core set \(Y\) and the sample space \(\Xi\) firstly. If there is no much priori information of sample space, the unbounded case \(\Xi = \mathbb{R}^p\) is suggested. Then the optimal objective values of DRO with \(Y\) and \(\Xi\) as the sample space can be computed respectively. If the gap between the two values is extremely small, DRO may be enough to get a satisfying result. If the gap
is large, the decision maker can determine an appropriate $\theta$ by (3.3) and solve the corresponding GDRO model.

Appendix A. Proofs.

A.1. Proof of Lemma 2.1.

Proof. We consider the inner maximization problem of (1.1) with the ambiguity set (1.2) and rewrite it as

$$\max_{P \in C(\mathcal{P})} E_P[h(x, \xi)]$$

s.t. $E_P[(\xi - \mu_0)(\xi - \mu_0)^T] \preceq \gamma_2 \Sigma_0,$

$$E_P \left( \frac{-\Sigma_0}{\gamma_1} \left( \mu_0 - \xi \right)^T - \gamma_1 \right) \preceq 0,$$

$$E_P[I_{\xi \in \Xi}] = 1,$$

where $C(\mathcal{P})$ is the cone generated by $\mathcal{P}$ and $I_{\xi \in \Xi}$ is the indicator function of the set $\Xi$, defined as

$$I_{\xi \in \Xi}(\xi) = \begin{cases} 1 & \text{if } \xi \in \Xi, \\ 0 & \text{otherwise.} \end{cases}$$

For the given $\Lambda_1 \in S^p_+$, $[A_2 \beta] \in S^{p+1}_+$ and $t \in \mathbb{R}$, the Lagrangian function of (A.1) is

$$L \left( \Lambda_1, \begin{bmatrix} A_2 \\ \beta^T \\ s \end{bmatrix}, t, P \right)$$

$$= E_P[h(x, \xi)] + \Lambda_1 \cdot (\gamma_2 \Sigma_0 - E_P[(\xi - \mu_0)(\xi - \mu_0)^T])$$

$$+ \begin{bmatrix} A_2 \\ \beta^T \\ s \end{bmatrix} \cdot E_P \left( \frac{\Sigma_0}{\gamma_1} \left( \mu_0 - \xi \right)^T \right) + t(1 - E_P[I_{\xi \in \Xi}])$$

$$= \Lambda_1 \cdot (\gamma_2 \Sigma_0 - \mu_0 \mu_0^T) + \Lambda_2 \cdot \Sigma_0 + \gamma_1 s - 2\mu_0^T \beta + t$$

$$+ E_P[h(x, \xi)] - \xi^T \Lambda_1 \xi + 2\xi^T \Lambda_1 \mu_0 + 2\beta^T \xi - tI_{\xi \in \Xi}].$$

And

$$\max_{P \in C(\mathcal{P})} L \left( \Lambda_1, \begin{bmatrix} A_2 \\ \beta^T \\ s \end{bmatrix}, t, P \right)$$

$$= \Lambda_1 \cdot (\gamma_2 \Sigma_0 - \mu_0 \mu_0^T) + \Lambda_2 \cdot \Sigma_0 + \gamma_1 s - 2\mu_0^T \beta + t$$

$$+ \max_{P \in C(\mathcal{P})} E_P[h(x, \xi)] - \xi^T \Lambda_1 \xi + 2\xi^T \Lambda_1 \mu_0 + 2\beta^T \xi - tI_{\xi \in \Xi}]$$

$$= \begin{cases} \Lambda_1 \cdot (\gamma_2 \Sigma_0 - \mu_0 \mu_0^T) + \Lambda_2 \cdot \Sigma_0 + \gamma_1 s - 2\mu_0^T \beta + t \\ \infty & \text{otherwise,} \end{cases}$$

where the constraint

$$E_P[h(x, \xi)] - \xi^T \Lambda_1 \xi + 2\xi^T \Lambda_1 \mu_0 + 2\beta^T \xi - tI_{\xi \in \Xi}] \leq 0 \quad \forall P \in C(\mathcal{P}),$$

$$\iff E_P[h(x, \xi)] - \xi^T \Lambda_1 \xi + 2\xi^T \Lambda_1 \mu_0 + 2\beta^T \xi - tI_{\xi \in \Xi}] \leq 0 \quad \forall P \in \mathcal{P},$$

$$\iff h(x, \xi) - \xi^T \Lambda_1 \xi + 2\xi^T \Lambda_1 \mu_0 + 2\beta^T \xi \leq t \quad \forall \xi \in \Xi.$$
Therefore, the Lagrangian dual of (A.1) is written as

\[
\min_{\Lambda_1, \Lambda_2, \beta, s, t} \quad \Lambda_1 \cdot (\gamma_2 \Sigma_0 - \mu_0 \mu_0^T) + \Lambda_2 \cdot \Sigma_0 + \gamma_1 s - 2\mu_0^T \beta + t
\]
\[\text{s.t.} \quad h(x, \xi) - \xi^T \Lambda_1 \xi + 2\xi^T \Lambda_1 \mu_0 + 2\beta^T \xi \leq t \quad \forall \xi \in \Xi,
\]
\[
\Lambda_1 \in \mathbb{S}_+^p, \quad \begin{bmatrix} \Lambda_2 & \beta \\ \beta^T & s \end{bmatrix} \in \mathbb{S}_+^{p+1}, \quad t \in \mathbb{R}.
\]  

(A.2)

We can simplify the dual problem by solving analytically for the variables (A, \beta, s) with fixed (\Lambda_1, t). Denote (A^*_1, \beta^*, s^*) as the optimal solution of (A.2) with fixed (\Lambda_1, t).

- If \( s^* > 0 \), then \( A^*_1 \geq \frac{1}{s^*} \beta^* \beta^T \) by \( \begin{bmatrix} A^*_2 & \beta^* \\ \beta^T & s^* \end{bmatrix} \in \mathbb{S}_+^{p+1} \) according to Schur’s complement [6].

The objective function gets its minimum at \( A^*_1 = \frac{1}{s^*} \beta^* \beta^T \). Then the objective function becomes

\[
t + \Lambda_1 \cdot (\gamma_2 \Sigma_0 - \mu_0 \mu_0^T) + \frac{1}{s^*} \beta^T \Sigma_0 \beta^* + \gamma_1 s^* - 2\mu_0^T \beta^*.
\]

Since \( \frac{1}{s^*} \beta^T \Sigma_0 \beta^* + \gamma_1 s^* \geq 2\sqrt{\gamma_1 \beta^T \Sigma_0 \beta^*} \) and the equality holds if and only if \( s^* = \sqrt{\beta^T \Sigma_0 \beta^*/\gamma_1} \), the objective function becomes

\[
t + \Lambda_1 \cdot (\gamma_2 \Sigma_0 - \mu_0 \mu_0^T) + \sqrt{\gamma_1}||\Sigma_0^{1/2}(q + 2\Lambda_1 \mu_0)||_2 + q^T \mu_0
\]

with \( q = -2(\beta^* + \Lambda_1 \mu_0) \).

- If \( s^* = 0 \), then \( \beta^* = 0 \) to keep \( \begin{bmatrix} A^*_2 & \beta^* \\ \beta^T & s^* \end{bmatrix} \in \mathbb{S}_+^{p+1} \).

The objective function gets its minimum at \( A^*_1 = 0 \) due to \( \Sigma_0 \geq 0 \).

Let \( q = -2(\beta^* + \Lambda_1 \mu_0) \). The objective function is written as

\[
t + \Lambda_1 \cdot (\gamma_2 \Sigma_0 - \mu_0 \mu_0^T)
\]

\[
= t + \Lambda_1 \cdot (\gamma_2 \Sigma_0 + \mu_0 \mu_0^T) + \sqrt{\gamma_1}||\Sigma_0^{1/2}(q + 2\Lambda_1 \mu_0)||_2 + q^T \mu_0.
\]

Thus, the dual of (A.1) can be equivalently rewritten as

\[
\min_{\Lambda, q, t} \quad t + \Lambda \cdot (\gamma_2 \Sigma_0 + \mu_0 \mu_0^T) + \sqrt{\gamma_1}||\Sigma_0^{1/2}(q + 2\Lambda \mu_0)||_2 + q^T \mu_0
\]
\[\text{s.t.} \quad h(x, \xi) - \xi^T \Lambda \xi - q^T \xi \leq t \quad \forall \xi \in \Xi,
\]
\[
\Lambda \in \mathbb{S}_+^p, \quad q \in \mathbb{R}^p, \quad t \in \mathbb{R}.
\]  

(A.3)

According to Proposition 2.8 in Shapiro[1], we can conclude that there is no dual gap if \( \gamma_1, \gamma_2 > 0, \Sigma_0 > 0 \) and \( h(x, \xi) \) is integrable for all \( P \in D \). Besides, the optimal solution set of the problem (A.3) is nonempty. \hfill \Box

A.2. Proofs of Theorems 2.3 to 2.6. We first introduce the following theorem given in [3], which is essential to our computationally tractable GDRO models.

**Theorem A.1** (Ben-Tal et al. [3]). Let \( f(\cdot, x) \) be a closed proper concave function in \( \mathbb{R}^m \) for all \( x \in \mathbb{R}^n \) and \( \phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R} \) be a closed, jointly convex and nonnegative function for which \( \phi(\xi, \xi) = 0 \) for all \( \xi \in \mathbb{R}^m \). Let the set \( Z_1 \subset \mathbb{R}^p \) be nonempty, convex and compact with \( 0 \in r_i(Z_1) \) and \( Z_2 \) be a closed convex set such that \( Z_1 \subset Z_2 \); and let \( U_1 \) and \( U_2 \) be defined as follows for fixed \( \xi_0 \) and \( A \in \mathbb{R}^{m \times L} \) such that \( \xi_0 \in r_i(U_1) \).
(A.4) \[ U_i = \{ \xi = \xi_0 + A\zeta | \zeta \in Z_i \}, \quad i = 1, 2. \]

Then \( x \) satisfies

(A.5) \[ f(\xi, x) \leq \min_{\xi \in U_1} \phi(\xi, \xi') \quad \forall \xi \in U_2, \]

if and only if there exist \( v, w \in \mathbb{R}^m \) that satisfy the single inequality

(A.6) \[ \xi_0^T (v + w) + \delta^*(A^T v | Z_1) + \delta^*(A^T w | Z_2) - f_*(v + w, x) + \phi^**(v, -v) \leq 0. \]

**Notation.** For any function \( g : \mathbb{R}^n \to \mathbb{R} \), let \( \text{dom}(g) = \{ x \in \mathbb{R}^n | g(x) < \infty \} \). The convex conjugate of \( g \) is defined as

\[ g^*(y) = \sup_{x \in \text{dom}(g)} \{ y^T x - g(x) \}. \]

The concave conjugate function of \( g \) is defined as

\[ g_*(y) = \inf_{x \in \text{dom}(-g)} \{ y^T x - g(x) \}. \]

For a function \( f(\cdot, \cdot) \) with two variables, let \( f^*(\cdot, \cdot) \) and \( f_*(\cdot, \cdot) \) denote the convex and concave conjugate function with respect to the first variable, respectively, while \( f^{**}(\cdot, \cdot) \) and \( f_{**}(\cdot, \cdot) \) denote the convex and concave conjugate function with respect to both variables, respectively.

The indicator function on the set \( S \) is defined as

\[ \delta(x | S) = \begin{cases} 0 & \text{if } x \in S, \\ \infty & \text{otherwise}. \end{cases} \]

**Proof of Theorem 2.3.** Note that

\[ h(x, \xi) = \max_k \{ a_k(x)^T \xi + b_k(x) \}, \quad k \in \{1, 2, \ldots, K\}. \]

So the constraint

\[ h(x, \xi) - \xi^T A\xi - q^T \xi - t \leq \min_{\xi' \in Y} \phi(\xi, \xi') \quad \forall \xi \in \mathbb{R}^p \]

is equivalent to

\[ a_k(x)^T \xi + b_k(x) - \xi^T A\xi - q^T \xi - t \leq \min_{\xi' \in Y} \phi(\xi, \xi') \quad \forall \xi \in \mathbb{R}^p \quad \forall k. \]

Let

\begin{align*}
  m &= p, \\
  Z_1 &= \{ \xi \mid \| \xi \|_2 \leq \sqrt{\bar{\gamma}} \}, \\
  Z_2 &= \mathbb{R}^p, \\
  A &= \Sigma_0^{1/2}, \\
  \xi_0 &= \bar{\mu}, \\
  U_1 &= \{ \xi | (\xi - \bar{\mu})^T \Sigma_0^{-1} (\xi - \bar{\mu}) \leq \bar{\gamma} \} = Y, \\
  U_2 &= \mathbb{R}^p, \\
  f_k(\xi, (x, A, q, t)) &= a_k(x)^T \xi + b_k(x) - \xi^T A\xi - q^T \xi - t, \\
  \phi(\xi, \xi') &= \theta \| \xi - \xi' \|_2
\end{align*}
in Theorem A.1. Then
\[ \delta^*(A^Tw|Z_1) = \sup_{\|\zeta\|_2 \leq \sqrt{\gamma}} \zeta^T A^Tv = \sqrt{\gamma}\|A^Tv\|_2 = \sqrt{\gamma}\|\Sigma_0^{1/2}v\|_2, \]
\[ \delta^*(A^Tw|Z_2) = \sup_{\zeta \in \mathbb{R}^p} \zeta^T A^Tw = \sup_{y \in \mathbb{R}^p} y^T w = \begin{cases} 0 & \text{if } w = 0, \\ \infty & \text{otherwise}, \end{cases} \]
\[ \phi^{**}(v, -v) = \sup_{s, t} \{(s - t)^T v - \theta s - t\|s\|_2\} \]
\[ = \sup_{s} s^T v - \theta s\|s\|_2 \]
\[ = \begin{cases} 0 & \text{if } \|v\|_2 \leq \theta, \\ \infty & \text{otherwise}, \end{cases} \]
\[ f_k,*(y, (x, \Lambda, q, t)) = \inf_{\xi \in \mathbb{R}^p} \{y^T \xi - f_k(\xi, (x, \Lambda, q, t))\} \]
\[ = \inf_{\xi \in \mathbb{R}^p} \{\xi^T \Lambda \xi + (y + q - a_k(x))^T \xi + t - b_k(x)\}. \]
According to Theorem A.1, we know that
\[ a_k(x)^T \xi + b_k(x) - \xi^T \Lambda \xi - q^T \xi - t \leq \min_{\xi \in \mathcal{Y}} \phi(\xi, \xi') \quad \forall \xi \in \mathbb{R}^p \]
is equivalent to
\[ \begin{cases} \bar{\mu}^Tv_k + \sqrt{\gamma}\|\Sigma_0^{1/2}v_k\|_2 - \inf_{\xi \in \mathbb{R}^p} \{\xi^T \Lambda \xi + (v_k + q - a_k(x))^T \xi + t - b_k(x)\} \leq 0, \\ \|v_k\| \leq \theta, \end{cases} \]
where
\[ \bar{\mu}^Tv_k + \sqrt{\gamma}\|\Sigma_0^{1/2}v_k\|_2 - \inf_{\xi \in \mathbb{R}^p} \{\xi^T \Lambda \xi + (v_k + q - a_k(x))^T \xi + t - b_k(x)\} \leq 0 \]
\[ \iff \bar{\mu}^Tv_k + \sqrt{\gamma}\|\Sigma_0^{1/2}v_k\|_2 \leq \xi^T \Lambda \xi + (v_k + q - a_k(x))^T \xi + t - b_k(x) \quad \forall \xi \in \mathbb{R}^p \]
\[ \iff \xi^T \Lambda \xi + (v_k + q - a_k(x))^T \xi + t - b_k(x) - \bar{\mu}^Tv_k - \sqrt{\gamma}\|\Sigma_0^{1/2}v_k\|_2 \geq 0 \quad \forall \xi \in \mathbb{R}^p \]
\[ \iff \left[ \xi^T, 1 \right] \begin{bmatrix} \Lambda & \frac{1}{2}(v_k + q - a_k(x)) \\ \frac{1}{2}(v_k + q - a_k(x))^T & t - b_k(x) - \bar{\mu}^Tv_k - \sqrt{\gamma}\|\Sigma_0^{1/2}v_k\|_2 \end{bmatrix} \begin{bmatrix} \xi \\ 1 \end{bmatrix} \geq 0 \quad \forall \xi \in \mathbb{R}^p \]
\[ \iff \left[ \Lambda & \frac{1}{2}(v_k + q - a_k(x)) \\ \frac{1}{2}(v_k + q - a_k(x))^T & t - b_k(x) - \bar{\mu}^Tv_k - \sqrt{\gamma}\|\Sigma_0^{1/2}v_k\|_2 \end{bmatrix} \geq 0. \]
Then the result follows.

Proof of Theorem 2.4. Note that
\[ h(x, \xi) = \max_k \{a_k(x)^T \xi + b_k(x)\}, \quad k \in \{1, 2, \ldots, K\}. \]
So the constraint
\[ h(x, \xi) - \xi^T \Lambda \xi - q^T \xi - t \leq \min_{\xi' \in Y} \phi(\xi, \xi') \quad \forall \xi \in \Xi \]
is equivalent to
\[ a_k(x)^T \xi + b_k(x) - \xi^T \Lambda \xi - q^T \xi - t \leq \min_{\xi' \in Y} \phi(\xi, \xi') \quad \forall \xi \in \Xi \forall k. \]

Let
\[
\begin{align*}
    m &= p, \\
    Z_1 &= \{ \xi \| \xi \|_2 \leq \sqrt{\gamma_1} \}, \quad Z_2 = \{ \xi \| \xi \|_2 \leq \sqrt{\gamma_2} \}, \\
    A &= \Sigma_0^{1/2}, \quad \xi_0 = \bar{\mu}, \\
    U_1 &= \{ \xi (\xi - \bar{\mu})^T \Sigma_0^{-1} (\xi - \bar{\mu}) \leq \gamma_1 \} = Y, \\
    U_2 &= \{ \xi (\xi - \bar{\mu})^T \Sigma_0^{-1} (\xi - \bar{\mu}) \leq \gamma_2 \} = \Xi, \\
    f_k(\xi, (x, \Lambda, q, t)) &= a_k(x)^T \xi + b_k(x) - \xi^T \Lambda \xi - q^T \xi - t, \\
    \phi(\xi, \xi') &= \theta \| \xi - \xi' \|_2
\end{align*}
\]
in Theorem A.1.

Then
\[
\delta^*(A^T v | Z_1) = \sup_{\| \xi \|_2 \leq \sqrt{\gamma_1}} \xi^T A^T v = \sqrt{\gamma_1} \| A^T v \|_2 = \sqrt{\gamma_1} \| \Sigma_0^{1/2} v \|_2,
\]
\[
\delta^*(A^T w | Z_2) = \sup_{\| \xi \|_2 \leq \sqrt{\gamma_2}} \xi^T A^T w = \sqrt{\gamma_2} \| A^T w \|_2 = \sqrt{\gamma_2} \| \Sigma_0^{1/2} w \|_2,
\]
\[
\phi^*(v, -v) = \sup_{s, t} \{(s - t)^T v - \theta \| s - t \|_2 \} = \sup_s \{s^T v - \theta \| s \|_2\}
\]
\[
= \left\{ \begin{array}{ll} 
0 & \text{if } \| v \|_2 \leq \theta, \\
\infty & \text{otherwise},
\end{array} \right.
\]
\[
f_{k, \ast}(y, (x, \Lambda, q, t)) = \inf_{\xi \in \mathbb{R}^p} \{ y^T \xi - f_k(\xi, (x, \Lambda, q, t)) \} = \inf_{\xi \in \mathbb{R}^p} \{ \xi^T \Lambda \xi + (y + q - a_k(x))^T \xi + t - b_k(x) \}.
\]

According to Theorem A.1, we know that
\[ a_k(x)^T \xi + b_k(x) - \xi^T \Lambda \xi - q^T \xi - t \leq \min_{\xi' \in Y_1} \phi(\xi, \xi') \quad \forall \xi \in Y_2 \]
is equivalent to
\[
\begin{cases}
\bar{\mu}^T (v_k + w_k) + \sqrt{\gamma_1} \| \Sigma_0^{1/2} v_k \|_2 + \sqrt{\gamma_2} \| \Sigma_0^{1/2} w_k \|_2 - f_{k, \ast}(v_k + w_k, (x, \Lambda, q, t)) \leq 0, \\
\| v_k \|_2 \leq \theta,
\end{cases}
\]
Then the result follows.

where

\[ \mu^T(v_k + w_k) + \sqrt{\gamma_1} \| \Sigma_0^{1/2} v_k \|_2 + \sqrt{\gamma_2} \| \Sigma_0^{1/2} w_k \|_2 - f_{k,}\ast(v_k + w_k, (x, \Lambda, q, t)) \leq 0 \]

\[ \iff \mu^T(v_k + w_k) + \sqrt{\gamma_1} \| \Sigma_0^{1/2} v_k \|_2 + \sqrt{\gamma_2} \| \Sigma_0^{1/2} w_k \|_2 \leq \xi^T \Lambda \xi + (v_k + w_k + q - a_k(x))^T \xi + t - b_k(x) \quad \forall \xi \in \mathbb{R}^p \]

\[ \iff \begin{bmatrix} \Lambda & 1/2 (v_k + w_k + q - a_k(x)) \\ 1/2 (v_k + w_k + q - a_k(x))^T & -b_k(x) - \mu^T(v_k + w_k) \end{bmatrix} \succeq 0 \]

\[ \iff \begin{bmatrix} \Lambda & 1/2 (v_k + w_k + q - a_k(x)) \\ 1/2 (v_k + w_k + q - a_k(x))^T & -b_k(x) - \mu^T(v_k + w_k) - z_{1k} - z_{2k} \end{bmatrix} \succeq 0, \]

\[ \sqrt{\gamma_1} \| \Sigma_0^{1/2} v_k \|_2 \leq z_{1k}, \quad \sqrt{\gamma_2} \| \Sigma_0^{1/2} w_k \|_2 \leq z_{2k}. \]

Then the result follows.

As for Theorems 2.5 and 2.6, the proofs are similar to that of Theorems 2.3 and 2.4, respectively. So we omit them here.

**Appendix B. Determine the Value of \( \gamma_1 \) and \( \gamma_2 \).**

**B.1. Boundedness Assumption.** Delage and Ye [11] show the confidence region for the mean and covariance matrix under the boundedness assumption.

**Assumption B.1.** There exists an \( R \geq 0 \) such that

\[ P \left( (\xi - \mu)^T \Sigma^{-1} (\xi - \mu) \leq R^2 \right) = 1, \]

where \( \mu \) is the mean of \( \xi \) and \( \Sigma \) is the covariance matrix of \( \xi \).

**Theorem B.2 (Delage and Ye[11]).** Let \( \{\xi_i\}_{i=1}^M \) be a set of \( M \) samples generated independently and randomly according to the distribution of \( \xi \). If \( \xi \) satisfies Assumption B.1 and \( M \) satisfies the following inequality

\[ M > R^4 \left( \sqrt{1 - p/R^2} + \sqrt{\ln(2/\delta)} \right)^2, \]

then with the probability greater than \( 1 - \delta \) over the choice of \( \{\xi_i\}_{i=1}^M \), the following set of constraints are met:

\[ (\mu_0 - \mu)^T \Sigma^{-1} (\mu_0 - \mu) \leq \beta(\delta/2), \]

\[ \Sigma \leq \frac{1}{1 - \alpha(\delta/4) - \beta(\delta/2)} \Sigma_0, \]

\[ \Sigma \geq \frac{1}{1 + \alpha(\delta/4)} \Sigma_0, \]
where \( \mu_0 \) is the sample mean, \( \Sigma_0 \) is the sample covariance matrix and \( \alpha(\delta/4) \) and \( \beta(\delta/2) \) are as follows:

\[
\begin{align*}
\alpha(\delta/4) &= (R^2/\sqrt{M})(\sqrt{1 - p/R^2} + \sqrt{\ln(4/\delta)}), \\
\beta(\delta/2) &= (R^2/M)(2 + \sqrt{2\ln(2/\delta)})^2.
\end{align*}
\]

Moreover, they show that if the conditions in Theorem B.2 are satisfied, the ambiguity set (1.2) with \( \gamma_1 \) and \( \gamma_2 \) defined as follows contains the true distribution of \( \xi \) with probability greater than \( 1 - \delta \) over the choice of \( \{\xi_i\}_{i=1}^M \).

\[
\begin{align*}
\gamma_1 &= \frac{\beta(\delta/2)}{1 - \alpha(\delta/4) - \beta(\delta/2)}, \\
\gamma_2 &= \frac{1 + \beta(\delta/2)}{1 - \alpha(\delta/4) - \beta(\delta/2)}.
\end{align*}
\]  

(B.1)

B.2. Normal Assumption. We consider an unbounded case that the true distribution is normal distribution.

Lemma B.3 (see [5] section 8.3). Let \( \{\xi_i\}_{i=1}^M \) be i.i.d \( N_p(\mu, \Sigma) \). By Hotelling-\( T^2 \) test, we have exactly

\[
P\left( (\mu_0 - \mu)^T\Sigma_0^{-1}(\mu_0 - \mu) \leq \frac{p}{M-p} F_\alpha(p, M-p) \right) = 1 - \alpha,
\]
where \( F_\alpha(p, M-p) \) is \( \alpha \) quantile of the distribution \( F(p, M-p) \).

Lemma B.4 (see [5] Proposition 7.1). For \( \{\xi_i\}_{i=1}^M \) i.i.d \( N_p(\mu, \Sigma) \),

\[
M\Sigma_0 \sim W_p(M-1, \Sigma),
\]

where \( W_p(M-1, \Sigma) \) is Wishart distribution.

The joint distribution of eigenvalues and the probability density function of the maximum and minimum eigenvalues of Wishart matrix all have closed form [13]. However, it’s still difficult to compute the confidence region of the maximum and minimum eigenvalues. Hence, we use the approximation of \( P(\lambda \leq \lambda_{\text{min}}(W)) \).

Lemma B.5 (see [9] section 4). For \( W \sim W_p(m, I) \),

\[
P(\lambda_{\text{max}}(W) < b) \to F_\beta\left( \frac{b - \mu_{mp}}{\sigma_{mp}} \right),
\]

\[
P(\lambda_{\text{min}}(W) > a) \to F_\beta\left( \frac{a - \mu^-_{mp}}{\sigma^-_{mp}} \right),
\]

where

\[
\mu_{mp} = (\sqrt{m} + \sqrt{p})^2, \sigma_{mp} = \sqrt{\mu_{mp}}\left( \frac{1}{\sqrt{p}} + \frac{1}{\sqrt{m}} \right)^{1/2},
\]

\[
\mu^-_{mp} = (\sqrt{m} - \sqrt{p})^2, \sigma^-_{mp} = \sqrt{\mu^-_{mp}}\left( \frac{1}{\sqrt{p}} - \frac{1}{\sqrt{m}} \right)^{1/2}.
\]

Besides, \( F_\beta(\cdot) \) is the distribution function of Tracy-Widom distribution \( TW_\beta \) and \( \beta = 1 \) for real matrices and \( \beta = 2 \) for complex matrices.
Corollary B.6. Let $\{\xi_i\}_{i=1}^M$ be i.i.d $N_p(\mu, \Sigma)$ and

$$\gamma_1 = \frac{p}{M - p} F_a(p, M - p),$$

(B.2)

$$\gamma_2 = \gamma_1 + \frac{M}{a}.$$ Then the constraint

(B.3) $$(E_P[\xi] - \mu_0)^T \Sigma_0^{-1} (E_P[\xi] - \mu_0) \leq \gamma_1$$

is met with the probability $1 - \alpha$. And the constraint

(B.4) $$E_P[(\xi - \mu_0)(\xi - \mu_0)^T] \preceq \gamma_2 \Sigma_0$$

is met with the probability $F_\beta(-\frac{a - \mu_{M-1,p}}{\sigma_{M-1,p}})$ approximately.

Proof. For the first part of the corollary, the proof follows from Lemma B.3 directly. We only need to proof the remainder of the corollary. According to Lemma B.4,

$$M \Sigma_0 \sim W_p(M - 1, \Sigma).$$

Hence,

$$W = M \Sigma_0^{-\frac{1}{2}} \Sigma_0^{-\frac{1}{2}} \sim W_p(M - 1, \Sigma).$$

According to Lemma B.5,

$$P(\lambda_{\min}(W) > a) \to F_\beta(-\frac{a - \mu_{M-1,p}}{\sigma_{M-1,p}}).$$

Hence,

$$P(\Sigma \preceq \frac{M}{a} \Sigma_0) = P(a \Sigma \preceq W) = P(\lambda_{\min}(W) \geq a) \approx F_\beta(-\frac{a - \mu_{M-1,p}}{\sigma_{M-1,p}}).$$

In addition,

$$\frac{M}{a} \Sigma_0 \preceq \Sigma = E_P[\xi \xi^T] - E_P[\xi] E_P[\xi]^T$$

$$\geq E_P[(\xi - \mu_0)(\xi - \mu_0)^T] - \gamma_1 \Sigma_0.$$ Therefore, (B.4) is met with the probability greater than $F_\beta(-\frac{a - \mu_{M-1,p}}{\sigma_{M-1,p}})$ approximately. \qed

Let $a = \mu_{M-1,p} - \sigma_{M-1,p}$ and $\beta = 1$, then $F_1(1) \approx 95\%$.

REFERENCES

[1] S. Alexander, On duality theory of conic linear problems, in Semi-Infinite Programming: Recent Advances, G. M. A. and L. M. A., eds., Springer US, Boston, MA, 2001, pp. 135–165.
[2] A. Ardestani-Jaafari and E. Delage, Robust optimization of sums of piecewise linear functions with application to inventory problems, Oper. Res., 64 (2016), pp. 474–494, https://doi.org/10.1287/opre.2016.1483.

[3] A. Ben-Tal, R. Brekelmans, D. den Hertog, and J.-P. Vial, Globalized robust optimization for nonlinear uncertain inequalities, INFORMS J. Comput., 29 (2017), pp. 350–366, https://doi.org/10.1287/ijsco.2016.0735.

[4] D. Bertsimas and M. Sim, The price of robustness, Oper. Res., 52 (2004), pp. 35–53, https://doi.org/10.1287/opre.1030.0065.

[5] M. Bilodeau and D. Brenner, Theory of Multivariate Statistics, Springer, New York, 1999.

[6] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, New York, 2004.

[7] X. Cai, K.-L. Teo, X. Yang, and X. Y. Zhou, Portfolio optimization under a minmax rule, Manage. Sci., 46 (2000), pp. 957–972, https://doi.org/10.1287/mnsc.46.7.957.12039.

[8] J. Cheng, R. L.-Y. Chen, H. N. Najm, A. Pinar, C. Safta, and J.-P. Waston, Distributionally robust optimization with principal component analysis, SIAM J. Optim., 28 (2018), pp. 1817–1841, https://doi.org/10.1137/16M1075910.

[9] M. Chiani, On the probability that all eigenvalues of gaussian, wishart, and double wishart random matrices lie within an interval, IEEE Trans. Inf. Theory, 63 (2017), pp. 4521–4531, https://doi.org/10.1109/TIT.2017.2694846.

[10] E. de Klerk, D. Kuhn, and K. Postek, Distributionally robust optimization with polynomial densities: theory, models and algorithms, Math. Program., 181 (2020), pp. 265–296, https://doi.org/10.1007/s10107-019-01429-5.

[11] E. Delage and Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, Oper. Res., 58 (2010), pp. 595–612, https://doi.org/10.1287/opre.1090.0741.

[12] K. Ding, N. Huang, and L. Wang, Globalized distributionally robust optimization problems under the moment-based framework, Aug. 2020, https://arxiv.org/abs/2008.08256.

[13] A. Edelman, Eigenvalues and Condition Numbers of Random Matrices, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[14] P. M. Esfahani and D. Kuhn, Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations, Math. Program., 171 (2018), pp. 115–166, https://doi.org/10.1007/s10107-017-1172-1.

[15] G. A. Hanasusanto, D. Kuhn, S. W. Wallace, and S. Zymler, Distributionally robust multi-item newsvendor problems with multimodal demand distributions, Math. Program., 152 (2015), pp. 1–32, https://doi.org/10.1007/s10107-014-0776-y.

[16] H. Kosino, Piecewise linear risk-function and portfolio optimization, J. Oper. Res. Soc. Jpn., 33 (1990), pp. 139–156, https://doi.org/10.1580/jorsj.33.139.

[17] J. Y. Li and R. H. Kwong, Portfolio selection under model uncertainty: a penalized moment-based optimization approach, J. Glob. Optim., 56 (2013), pp. 131–164, https://doi.org/10.1007/s10898-012-9969-1.

[18] Q. Liu, J. Wu, X. Xiao, and L. Zhang, A note on distributionally robust optimization under moment uncertainty, J. Numer. Math., 26 (2018), pp. 141–150, https://doi.org/10.1515/jnma-2017-0020.

[19] S. Lotfi and S. A. Zenios, Robust $\emph{var}$ and $\emph{cvar}$ optimization under joint ambiguity in distributions, means, and covariances, Eur. J. Oper. Res., 269 (2018), pp. 556–576, https://doi.org/10.1016/j.euro.2017.04.1163.

[20] H. E. Scarf, A min-max solution of an inventory problem, Econometrica, 25 (1957), p. 352.

[21] H. Xu, Y. Liu, and H. Sun, Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods, Math. Program., 169 (2018), pp. 489–529, https://doi.org/10.1007/s10107-017-1143-6.

[22] S. Zymler, D. Kuhn, and B. Rustem, Distributionally robust joint chance constraints with second-order moment information, Math. Program., 137 (2013), pp. 167–198, https://doi.org/10.1007/s10107-011-0494-7.