

Distributionally robust second-order stochastic dominance constrained optimization with Wasserstein distance

Yu Mei, Jia Liu, Zhiping Chen*

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, P. R. China,
Center for Optimization Technique and Quantitative Finance, Xi'an International Academy
for Mathematics and Mathematical Technology, Xi'an, 710049, P. R. China

Abstract

We consider a distributionally robust second-order stochastic dominance constrained optimization problem. We require the dominance constraints hold with respect to all probability distributions in a Wasserstein ball centered at the empirical distribution. We adopt the sample approximation approach to develop a linear programming formulation that provides a lower bound. We propose a novel split-and-dual decomposition framework which provides an upper bound. We establish quantitative convergency for both lower and upper approximations given some constraint qualification conditions. To efficiently solve the non-convex upper bound problem, we use a sequential convex approximation algorithm. Numerical evidences on a portfolio selection problem valid the convergency and effectiveness of the proposed two approximation methods.

Keywords:

stochastic dominance, distributionally robust optimization, Wasserstein distance, sequential convex approximation

MSC:

Primary, 90C15, 91B70; Secondary, 90C31, 90-08

1 Introduction

Stochastic dominance (SD), originated from economics, is popular in comparing random outcomes. In their pioneering work [8], Dentcheva and Ruszczyński studied the stochastic optimization problem with univariate SD constraints, where they developed the optimality conditions and duality theory. The commonly adopted univariate SD concepts in stochastic optimization are first-order SD (FSD) and second-order SD (SSD). Researchers have investigated the stochastic optimization problem with FSD constraints from different aspects, such as stability and sensitivity analysis [6], mixed-integer linear programming formulations [29], and linear programming relaxations [34]. The stochastic optimization problem with SSD constraints has been intensively studied in quite a few literature. For theoretical foundations, the stability and sensitivity analysis were presented in [7]. For solution methods, different linear programming formulations were derived in [8, 29] and the cutting plane methods were adopted in [14, 38, 39]. The stochastic programs with SD constraints induced by mixed-integer linear recourse were studied in [17] for FSD and in [16] for SSD. Stochastic optimization problems with multivariate extensions of SD constraints were considered in [20, 21, 33]. There is also a rich literature considering SD under dynamic settings, such as [13]. Applications of SD constraints in finance were investigated in [5, 9, 22].

A challenge of stochastic programming problems is the accessibility of the true probability distribution of the uncertain parameters. In some practical problems, the true probability distribution sometimes could not be completely observed. For this reason, distributionally robust optimization (DRO) models have been proposed to address the lack of complete information on the true probability distribution, where the expectations are taken under the worst-case probability distribution in a specific ambiguity set. There are mainly two types of ambiguity sets in the existing literature. The first type is the moment-based ambiguity sets, which is characterized by some moment inequalities [18, 44]. The second type is the distance-based ambiguity sets, which contain all probability distributions close to a nominal distribution measured by some probability metrics, such as Kullback-Leibler

*corresponding author, email: zchen@mail.xjtu.edu.cn

divergence [26], ϕ -divergence [24], and Wasserstein distance [2, 23, 26, 31, 40]. Esfahani and Kuhn [32] estimated a priori probability that the true distribution belongs to the Wasserstein ball and established finite sample and asymptotic guarantees for the distributionally robust solutions. With the duality theory, the DRO problem with Wasserstein ball can be reformulated as convex programs [15, 32, 43]. Such reformulations were then applied to chance-constrained DRO problems [2, 23, 40].

Incorporating the basic ideas of SD and distributional robustness, Dentcheva and Ruszczyński [10] first introduced the distributionally robust SD and established the optimality conditions of the stochastic optimization problem with distributionally robust SSD constraints. Since then, a stream of research has paid attention to stochastic optimization with distributionally robust SD constraints. Dupačová and Kopa [12] modeled the ambiguity of the distribution in FSD by a linear combination of a nominal distribution and a known contamination distribution with the combination parameter being in a parametric uncertainty set. Guo, Xu and Zhang [18] proposed a discrete approximation scheme for the moment-based ambiguity sets and approximately solved the resulting stochastic optimization problem with distributionally robust SSD constraints. Under a moment-based ambiguity set, Liesiö et al. [25] identified optimal portfolios robustly SSD dominating a given benchmark. Chen and Jiang [3] and Zhang et al. [42] studied stability of DRO problems with k th order SD constraints induced by full random recourse. The optimality conditions and duality theory of DRO problems with multivariate SD were discussed in [4, 19].

As is mentioned above, SD constrained optimization under distributional ambiguity is an important class of problems. While the distributionally robust SSD constrained optimization with Wasserstein ball has not been well studied in the existing literature. The main difficulties of solving such problems lie in three aspects.

- The semi-infiniteness induced from both the SSD and the distributionally robust counterpart are the main challenge.
- Distributionally robust SSD constraints are non-smooth such that gradient based methods fail to work here.
- Compared to moment-based ambiguity sets, the Wasserstein distance contains an extra optimization problem on computing the optimal transportation from the true distribution to the nominal distribution. Such an inner-level optimization problem leads a min-max-min structure and non-convexity of the distributionally robust SSD constraints.

Therefore, it is quite challenging for us to study the approximation schemes and algorithms for the distributionally robust SSD constrained optimization problem with Wasserstein ball. Thanks to the rapid development recently on the strong duality theory of DRO problems with Wasserstein ball [15, 32], we have a chance to show in this paper efficient approximation methods for such SSD constrained problem.

In detail, we first utilize the strong duality results for DRO problem with Wasserstein ball in [15] to derive a reformulation of distributionally robust SSD constraints. Then we adopt the sampling approach to approximate the infinitely many constraints by finitely many constraints and develop a linear programming formulation which is a lower bound approximation. We further analyze the quantitative convergency of the lower bound approximation. To overcome the ‘curse of dimensionality’ of the linear programming approximation, we propose a novel split-and-dual decomposition framework. We separate the support set of the parameter in distributionally robust SSD constraints into finite sub-intervals. For each sub-interval, we exchange the order of the supremum operator and the expectation operator to get an upper bound approximation. We prove that the optimal value of the upper bound approximation converges to that of the original problem as the number of sub-intervals goes to infinity and we quantitatively estimate the approximation error. As the derived upper bound approximation problem is non-convex, we apply the sequential convex approximation method to solve it.

This paper improves results in quite a few papers. Specifically, we extend the DRO with Wasserstein ball [2, 15, 23, 32, 40] to a more complicated case with infinitely many constraints induced by SSD. Compared with robust SD constrained optimization problems in [4, 18, 25, 42], we study Wasserstein ball rather than moment-based ambiguity sets. The main contributions of this paper include:

- We derive a lower bound approximation of the distributionally robust SSD constrained optimization with Wasserstein ball by the sample approximation approach, and establish the quantitative convergency of the approximation problem.
- We propose a novel split-and-dual decomposition framework, which provides an upper bound approximation of the problem. As far as we know, the upper bound approximations of SD constrained problems are seldom studied in existing literature. We prove the convergency of the approximation approach and quantitatively estimate the approximation error when the number of sub-intervals is sufficiently large.

While preparing this paper for publication, we became aware of an independent work by Peng and Delage [35] on distributionally robust SSD constraints with Wasserstein ball. Peng and Delage [35] formulated the distributionally robust SSD constrained problem as a multistage robust optimization problem, and proposed a tractable conservative approximation that exploits finite adaptability and a scenario-based lower bounding problem with nice numerical feasibility. We distinguish our work from Peng and Delage [35] in a different split-and-dual decomposition framework for the upper bound and detailed quantitative convergence analysis.

The rest of this paper is organized as follows. In section 2, we introduce the distributionally robust SSD constrained optimization problem. In section 3, we adopt the sampling approximation approach to obtain the lower bound approximation and establish the quantitative convergence. In section 4, we propose a split-and-dual decomposition framework to derive the upper bound approximation, whose optimal value can be obtained by solving a sequence of second-order cone programming problems. We also quantitatively estimate the approximation error when the number of sub-intervals is sufficiently large. Numerical evidences valid the convergence and effectiveness of the proposed approximation methods in Section 5. Section 6 concludes the paper.

2 Preliminaries

2.1 Distributionally robust second-order stochastic dominance

First we introduce some notations. Let \mathcal{U} be the set of all non-decreasing and concave utility functions $u : \mathbb{R} \rightarrow \mathbb{R}$. We use $(\cdot)_+ = \max\{\cdot, 0\}$ to denote the positive part function. Let $d(x, A) := \inf_{y \in A} \|x - y\|$ be the distance from a point x to a set A . Denote the deviation of a set A from another set B by $\mathbb{D}(A, B) = \sup_{x \in A} d(x, B)$ and the Hausdorff distance between A and B by $\mathbb{H}(A, B) = \max\{\mathbb{D}(A, B), \mathbb{D}(B, A)\}$. Let (Ω, \mathcal{F}) be a measurable space with \mathcal{F} being the Borel σ -algebra on Ω , and \mathcal{P} be the set of all probability measures on (Ω, \mathcal{F}) .

Before introducing the distributionally robust SSD, we recall the definition of classic SSD. Consider the random variables X and Y on a probability space (Ω, \mathcal{F}, P) with finite first order moments, here $P \in \mathcal{P}$ is the true distribution. We say that X stochastically dominates Y in the second order, denoted by $X \succeq_{(2)}^P Y$, if $\mathbb{E}_P[u(X)] \geq \mathbb{E}_P[u(Y)]$, $\forall u \in \mathcal{U}$. $X \succeq_{(2)}^P Y$ is equivalent to

$$\mathbb{E}_P[(\eta - X)_+ - (\eta - Y)_+] \leq 0, \forall \eta \in \mathbb{R}. \quad (1)$$

Let \mathcal{Y} be the set of all realizations of the random variable Y . It has been shown in [28, Proposition 1] that (1) is equivalent to

$$\mathbb{E}_P[(\eta - X)_+ - (\eta - Y)_+] \leq 0, \forall \eta \in \mathcal{Y}. \quad (2)$$

In some data-driven problems, it is difficult to obtain the complete information about the true probability measure P . To address this issue, Dentcheva and Ruszczyński [10] introduced distributionally robust SSD by considering an ambiguity set of probability measures instead of P .

Definition 1. X dominates Y robustly in the second order over a set of probability measures $\mathcal{Q} \subset \mathcal{P}$, denoted by $X \succeq_{(2)}^{\mathcal{Q}} Y$, if

$$\mathbb{E}_P[u(X)] \geq \mathbb{E}_P[u(Y)], \forall u \in \mathcal{U}, \forall P \in \mathcal{Q}.$$

In the rest of this paper, we investigate the following distributionally robust SSD constrained optimization problem

$$(P_{SSD}) \quad \begin{aligned} & \min_{z \in Z} f(z) \\ & \text{s.t.} \quad z^T \xi \succeq_{(2)}^{\mathcal{Q}} z_0^T \xi, \end{aligned}$$

where f is proper and continuous, ξ denotes the random vector, $Z \subset \mathbb{R}^n$ is a bounded polyhedral set, and $z_0 \in Z$ is a given benchmark. From (1), problem (P_{SSD}) can be rewritten as

$$\begin{aligned} & \min_{z \in Z} f(z) \\ & \text{s.t.} \quad \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \leq 0, \forall \eta \in \mathbb{R}, \forall P \in \mathcal{Q}. \end{aligned} \quad (3)$$

We can observe that the semi-infiniteness of constraints in problem (3) arises from $\eta \in \mathbb{R}$ and $P \in \mathcal{Q}$, induced from the SSD constraints and the distributionally robust ambiguity set, respectively. Moreover, the constraint

functions in problem (3) are non-smooth as $(\cdot)_+$ is involved. Therefore, problem (3), as well as problem (P_{SSD}) , is hard to solve. To reduce the difficulties in solving problem (P_{SSD}) , we firstly assume that the support set Ξ has a polyhedral structure. The polyhedral structure of Ξ , also assumed in [32, Corollary 5.1], contributes to applying the duality theory of second-order conic programming when deriving the upper bound approximation later in this paper.

Assumption 1. Ξ is polyhedral, i.e., $\Xi = \{\xi \in \mathbb{R}^n \mid C\xi \leq d\}$, where $C \in \mathbb{R}^{l \times n}$, $d \in \mathbb{R}^l$, and $z_0^T \Xi := \{z_0^T \xi \mid \xi \in \Xi\}$ is a compact set.

If Ξ is a polyhedron, then Assumption 1 holds automatically. Keeping in mind the equivalence of (1) and (2), problem (3) can be formulated as

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{s.t.} \quad & \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \leq 0, \quad \forall \eta \in \mathcal{R} := z_0^T \Xi, \quad \forall P \in \mathcal{Q}. \end{aligned} \quad (4)$$

We denote the smallest and largest numbers in \mathcal{R} by \mathcal{R}_{\min} and \mathcal{R}_{\max} , respectively, thus, $\mathcal{R} = [\mathcal{R}_{\min}, \mathcal{R}_{\max}]$.

2.2 Data-driven Wasserstein ambiguity set

In this section, we introduce the data-driven Wasserstein ambiguity set \mathcal{Q} and recall a fundamental duality result in DRO problems with Wasserstein ball [15, 32, 43].

Let $\mathcal{P}(\Xi)$ be the space of all probability measures P supported on Ξ with $\mathbb{E}_P[\|\xi\|] < \infty$. We consider 1-Wasserstein distance, also known as Kantorovich metric.

Definition 2. The Kantorovich metric $d_K: \mathcal{P}(\Xi) \times \mathcal{P}(\Xi) \rightarrow \mathbb{R}_+$ is defined via

$$d_K(P, Q) := \inf_{\pi} \left\{ \int_{\Xi^2} \|\xi_1 - \xi_2\| \pi(d\xi_1, d\xi_2) : \begin{array}{l} \pi \text{ is a joint distribution of } \xi_1 \text{ and } \xi_2 \\ \text{with marginals } P \text{ and } Q, \text{ respectively} \end{array} \right\}.$$

Kantorovich metric can be written in a pseudo metric form [1].

Proposition 1. Let \mathcal{G} be the set of all Lipschitz continuous functions $h: \Xi \rightarrow \mathbb{R}$ with modulus 1. Then

$$d_K(P, Q) := \sup_{h \in \mathcal{G}} \left| \int_{\Xi} h(\xi) P(d\xi) - \int_{\Xi} h(\xi) Q(d\xi) \right|.$$

Given $\mathcal{P}, \mathcal{Q} \subset \mathcal{P}(\Xi)$, define the deviation of \mathcal{P} from \mathcal{Q} by $\mathbb{D}_K(\mathcal{P}, \mathcal{Q}) := \sup_{P \in \mathcal{P}} \inf_{Q \in \mathcal{Q}} d_K(P, Q)$, and the Hausdorff distance between \mathcal{P} and \mathcal{Q} by $\mathbb{H}_K(\mathcal{P}, \mathcal{Q}) := \max\{\mathbb{D}_K(\mathcal{P}, \mathcal{Q}), \mathbb{D}_K(\mathcal{Q}, \mathcal{P})\}$.

Given N observations $\{\hat{\xi}_i\}_{i=1}^N$ of ξ , we define the data-driven Wasserstein ambiguity set \mathcal{Q} as a ball centered at the empirical distribution $\hat{P}_N = \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\xi}_i}$,

$$\mathcal{Q} := \{P \in \mathcal{P}(\Xi) : d_K(P, \hat{P}_N) \leq \epsilon\}, \quad (5)$$

where ϵ is a prespecified robust radius. Esfahani and Kuhn [32] proved that with any prescribed $\beta \in (0, 1)$, by appropriately defining $\epsilon(\beta)$, the true distribution P belongs to \mathcal{Q} with the confidence level $1 - \beta$.

Under some mild conditions, strong duality results of DRO problems with Wasserstein ball have been established in [32, Theorem 4.2], [15, Corollary 2] and [43, Proposition 2].

Lemma 1. If $\Psi(\xi)$ is proper, continuous, and for some $\zeta \in \Xi$, the growth rate $\kappa := \limsup_{\|\xi - \zeta\| \rightarrow \infty} \frac{\Psi(\xi) - \Psi(\zeta)}{\|\xi - \zeta\|} < \infty$, then the optimal values of

$$\sup_{P \in \mathcal{P}(\Xi)} \left\{ \int_{\Xi} \Psi(\xi) P(d\xi) : d_K(P, \hat{P}_N) \leq \epsilon \right\}$$

and

$$\min_{\lambda \geq 0} \left\{ \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [\Psi(\xi) - \lambda \|\xi - \hat{\xi}_i\|] \right\} \quad (6)$$

are equal. Moreover, the optimal solution set of (6) is nonempty and compact.

Proof. The strong duality was established in [15, Theorem 1]. We only need to prove the nonemptiness and compactness of the optimal solution set of (6). Let $f_D(\lambda) := \lambda\epsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [\Psi(\xi) - \lambda \|\xi - \widehat{\xi}_i\|]$. It is easy to see that f_D is lower semi-continuous and convex, and $f_D(\lambda) \geq \lambda\epsilon + \frac{1}{N} \sum_{i=1}^N \Psi(\widehat{\xi}_i), \forall \lambda \geq 0$. Since Ψ is proper, f_D is bounded from below on bounded sets and

$$\liminf_{\lambda \rightarrow \infty} \frac{f_D(\lambda)}{\lambda} \geq \liminf_{\lambda \rightarrow \infty} \frac{\lambda\epsilon + \frac{1}{N} \sum_{i=1}^N \Psi(\widehat{\xi}_i)}{\lambda} = \epsilon > 0.$$

This means that f_D is level-coercive [37, Definition 3.25].

Next, we prove that f_D is proper by showing $f_D(\kappa) < \infty$ at $\kappa := \limsup_{\|\xi - \zeta\| \rightarrow \infty} (\Psi(\xi) - \Psi(\zeta)) / \|\xi - \zeta\|$. Consider for each $i = 1, \dots, N$, the optimal value ϑ_i^* of problem $\sup_{\xi \in \Xi} f_i(\xi) := \Psi(\xi) - \kappa \|\xi - \widehat{\xi}_i\|$. Assume there exists a sequence $\{\xi_i^t\}_{t=1}^\infty \subset \Xi$ such that $f_i(\xi_i^t) \rightarrow \vartheta_i^*$. We then have two cases. Case 1: $\|\xi_i^t\| \rightarrow \infty$. As the growth rate κ does not depend on the choice of ζ [15, lemma 4], we can select $\xi = \xi_i^t$ and $\zeta = \widehat{\xi}_i$, and thus have $\vartheta_i^* = \lim_{t \rightarrow \infty} \Psi(\xi_i^t) - \kappa \|\xi_i^t - \widehat{\xi}_i\| \leq \Psi(\widehat{\xi}_i) < \infty$. Case 2: $\xi_i^t \rightarrow \xi_i^*$ with $\|\xi_i^*\| < \infty$. Since Ξ is closed, $\xi_i^* \in \Xi$. By continuity of Ψ , we have $\vartheta_i^* = \lim_{t \rightarrow \infty} \Psi(\xi_i^t) - \kappa \|\xi_i^t - \widehat{\xi}_i\| = \Psi(\xi_i^*) - \kappa \|\xi_i^* - \widehat{\xi}_i\| < \infty$. Therefore, $f_D(\kappa) = \kappa\epsilon + \frac{1}{N} \sum_{i=1}^N \vartheta_i^* < \infty$ and thus f_D is proper. By [37, Corollary 3.27], f_D is level-bounded. Therefore, it is known from [37, Theorem 1.9] that $\operatorname{argmin}_\lambda f_D(\lambda)$ is nonempty and compact. \square

2.3 Flowchart of the lower and upper bounds approximation schemes

Later on, we will derive for problem (P_{SSD}) a lower bound approximation in Section 3 and an upper bound approximation in Section 4. The relationship of formulations in intermediate steps of the two approximation schemes is illustrated in Figure 1.

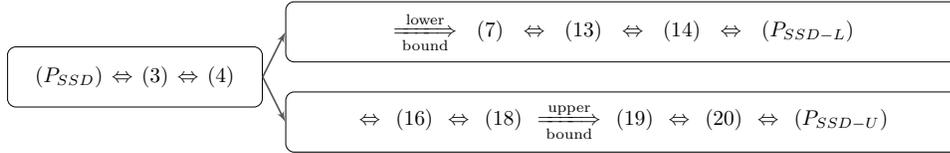


Figure 1: The flowchart of the two approximation schemes.

The key reformulation or approximation steps in the two approximation schemes can be summarized as follows:

- 1) Reformulations $(P_{SSD}) \Leftrightarrow (3)$ and $(3) \Leftrightarrow (4)$ are due to the definition of distributionally robust SSD.
- 2) Approximation $(4) \xrightarrow[\text{bound}]{\text{lower}} (7)$ comes from the finite sample approximation; Reformulation $(7) \Leftrightarrow (13)$ is due to the duality theory of DRO problems with Wasserstein ball from Lemma 1; Reformulations $(13) \Leftrightarrow (14)$ and $(14) \Leftrightarrow (P_{SSD-L})$ are obtained by adding auxiliary variables.
- 3) We propose a split-and-dual decomposition framework for the upper bound approximation. In detail, $(4) \Leftrightarrow (16)$ is a rewrite; We split the interval \mathcal{R} into sub-intervals in the reformulation $(16) \Leftrightarrow (18)$; We exchange the order of the expectation and supremum to derive the upper bound approximation $(18) \xrightarrow[\text{bound}]{\text{upper}} (19)$; Reformulations $(19) \Leftrightarrow (20)$ is due to the duality theory of DRO problems with Wasserstein ball from Lemma 1; Reformulation $(20) \Leftrightarrow (P_{SSD-U})$ is due to the strong duality of second-order cone programming.

3 Lower bound approximation of distributionally robust SSD constrained optimization

In order to tackle the semi-infiniteness arising from the constraints in problem (4), we consider the approximation of the sets \mathcal{Q} and \mathcal{R} . Let $\Xi_{\mathcal{N}} = \{\bar{\xi}_j\}_{j=1}^{\mathcal{N}}$ be a set of finite samples in Ξ and $\Gamma_{\mathcal{M}} = \{\eta_k\}_{k=1}^{\mathcal{M}}$ be a set of finite samples in $\mathcal{R} = [\mathcal{R}_{\min}, \mathcal{R}_{\max}]$, here \mathcal{N} and \mathcal{M} denote the sample sizes. We then approximate the ambiguity set \mathcal{Q} by the following Wasserstein ball:

$$\mathcal{Q}_{\mathcal{N}} := \{P \in \mathcal{P}(\Xi_{\mathcal{N}}) : d_K(P, \widehat{P}_{\mathcal{N}}) \leq \epsilon\}.$$

Different from \mathcal{Q} which covers both discrete and continuous probability measures supported on Ξ , $\mathcal{Q}_{\mathcal{N}}$ only contains discrete probability measures supported on $\Xi_{\mathcal{N}}$. It is easy to see $\mathcal{Q}_{\mathcal{N}} \subset \mathcal{Q}$ and $\Gamma_{\mathcal{M}} \subset \mathcal{R}$. Therefore, we have a lower bound approximation of problem (4):

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{s.t.} \quad & \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \leq 0, \quad \forall \eta \in \Gamma_{\mathcal{M}}, \quad \forall P \in \mathcal{Q}_{\mathcal{N}}. \end{aligned} \quad (7)$$

In subsection 3.1, we establish the quantitative convergency for problem (7) in terms of the feasible set, the optimal value, and the optimal solution set. Then in subsection 3.2, we show how problem (7) can be reformulated as a linear programming problem, and the computational efficiency for large sample sizes can be further improved by the cutting-plane method.

3.1 Quantitative analysis of the lower approximation

We denote the feasible sets of problem (4) and problem (7) by \mathcal{F} and $\mathcal{F}_{\mathcal{N}, \mathcal{M}}$, the optimal solution sets by \mathcal{S} and $\mathcal{S}_{\mathcal{N}, \mathcal{M}}$, and the optimal values by v and $v_{\mathcal{N}, \mathcal{M}}$, respectively. To establish the quantitative convergency for problem (7), we need the following Slater constraint qualification.

Assumption 2. *There exist a point $\bar{z} \in Z$ and a constant $\theta > 0$ such that*

$$\sup_{\eta \in \mathcal{R}} \sup_{P \in \mathcal{Q}} \mathbb{E}_P[(\eta - \bar{z}^T \xi)_+ - (\eta - z_0^T \xi)_+] < -\theta.$$

Let $L := \sup_{z \in Z} (\|z\| + \|z_0\|) < \infty$, since Z is compact. Denote the Hausdorff distance between Ξ and $\Xi_{\mathcal{N}}$ by $\alpha_{\mathcal{N}} := \sup_{\xi \in \Xi} \inf_{\xi' \in \Xi_{\mathcal{N}}} \|\xi - \xi'\|$, and the Hausdorff distance between \mathcal{R} and $\Gamma_{\mathcal{M}}$ by $\gamma_{\mathcal{M}} := \sup_{\eta \in \mathcal{R}} \inf_{\eta' \in \Gamma_{\mathcal{M}}} |\eta - \eta'|$.

Assumption 3. *$\lim_{\mathcal{N} \rightarrow \infty} \alpha_{\mathcal{N}} = 0$ and $\lim_{\mathcal{M} \rightarrow \infty} \gamma_{\mathcal{M}} = 0$.*

Theorem 1. *Given Assumptions 2 and 3, the following assertions hold.*

(i) *For any \mathcal{N} and \mathcal{M} ,*

$$\mathbb{H}(\mathcal{F}_{\mathcal{N}, \mathcal{M}}, \mathcal{F}) \leq \frac{2D_Z}{\theta} (L\alpha_{\mathcal{N}} + \gamma_{\mathcal{M}}),$$

where D_Z denotes the diameter of Z and θ is defined as in Assumption 2.

(ii) *$\lim_{\mathcal{N} \rightarrow \infty} v_{\mathcal{N}, \mathcal{M}} = v$ and $\limsup_{\mathcal{M} \rightarrow \infty} \mathcal{S}_{\mathcal{N}, \mathcal{M}} \subset \mathcal{S}$.*

(iii) *If, in addition, the objective function f is Lipschitz continuous with modulus L_f , then for any \mathcal{N} and \mathcal{M} ,*

$$|v_{\mathcal{N}, \mathcal{M}} - v| \leq \frac{2D_Z L_f}{\theta} (L\alpha_{\mathcal{N}} + \gamma_{\mathcal{M}}).$$

Moreover, if problem (4) satisfies the second-order growth condition at the optimal solution set \mathcal{S} , i.e., there exists a positive constant ρ such that

$$f(z) - v \geq \rho d(z, \mathcal{S})^2, \quad \forall z \in \mathcal{F},$$

then for sufficiently large \mathcal{N} and \mathcal{M} ,

$$\mathbb{D}(\mathcal{S}_{\mathcal{N}, \mathcal{M}}, \mathcal{S}) \leq \left(\sqrt{\frac{2D_Z}{\theta}} + \sqrt{\frac{4L_f D_Z}{\rho \theta}} \right) \sqrt{L\alpha_{\mathcal{N}} + \gamma_{\mathcal{M}}}. \quad (8)$$

Proof. (i) We write $\phi(\eta, z, \xi) = (\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+$, $w(z) = \sup_{\eta \in \mathcal{R}} \sup_{P \in \mathcal{Q}} \mathbb{E}_P[\phi(\eta, z, \xi)]$, and $w_{\mathcal{N}, \mathcal{M}}(z) = \sup_{\eta' \in \Gamma_{\mathcal{M}}} \sup_{P' \in \mathcal{Q}_{\mathcal{N}}} \mathbb{E}_{P'}[\phi(\eta', z, \xi)]$. Since $\mathcal{Q}_{\mathcal{N}} \subset \mathcal{Q}$ and $\Gamma_{\mathcal{M}} \subset \mathcal{R}$, then for any $z \in Z$,

$$w_{\mathcal{N}, \mathcal{M}}(z) \leq w(z). \quad (9)$$

One can observe that ϕ is uniformly Lipschitz continuous w.r.t. ξ with modulus L and also uniformly Lipschitz continuous w.r.t. η with modulus 2. Then by Proposition 1, we have

$$\begin{aligned} & |\mathbb{E}_P[\phi(\eta, z, \xi)] - \mathbb{E}_{P'}[\phi(\eta', z, \xi)]| \\ & \leq |\mathbb{E}_P[\phi(\eta, z, \xi)] - \mathbb{E}_{P'}[\phi(\eta, z, \xi)]| + |\mathbb{E}_{P'}[\phi(\eta, z, \xi)] - \mathbb{E}_{P'}[\phi(\eta', z, \xi)]| \end{aligned}$$

$$\leq Ld_K(P, P') + 2|\eta - \eta'|.$$

Therefore, for any $z \in Z$,

$$\begin{aligned} w(z) - w_{\mathcal{N}, \mathcal{M}}(z) &\leq \sup_{\eta \in \mathcal{R}} \inf_{\eta' \in \Gamma_{\mathcal{M}}} \sup_{P \in \mathcal{Q}} \inf_{P' \in \mathcal{Q}_{\mathcal{N}}} \left(\mathbb{E}_P[\phi(\eta, z, \xi)] - \mathbb{E}_{P'}[\phi(\eta', z, \xi)] \right) \\ &\leq \sup_{\eta \in \mathcal{R}} \inf_{\eta' \in \Gamma_{\mathcal{M}}} \sup_{P \in \mathcal{Q}} \inf_{P' \in \mathcal{Q}_{\mathcal{N}}} \left(Ld_K(P, P') + 2|\eta - \eta'| \right) \\ &= \sup_{P \in \mathcal{Q}} \inf_{P' \in \mathcal{Q}_{\mathcal{N}}} Ld_K(P, P') + \sup_{\eta \in \mathcal{R}} \inf_{\eta' \in \Gamma_{\mathcal{M}}} 2|\eta - \eta'| = L\mathbb{D}_K(\mathcal{Q}, \mathcal{Q}_{\mathcal{N}}) + 2\gamma_{\mathcal{M}}. \end{aligned} \quad (10)$$

By [1, Theorem 2], we have $\mathbb{H}_K(\mathcal{Q}, \mathcal{Q}_{\mathcal{N}}) \leq 2\alpha_{\mathcal{N}}$. Thus, combining inequalities (9) and (10) gives

$$\sup_{z \in Z} |w(z) - w_{\mathcal{N}, \mathcal{M}}(z)| \leq L\mathbb{H}_K(\mathcal{Q}, \mathcal{Q}_{\mathcal{N}}) + 2\gamma_{\mathcal{M}} \leq 2L\alpha_{\mathcal{N}} + 2\gamma_{\mathcal{M}}.$$

With Assumption 2, we can apply Robinson's error bound for convex inequality system [36] and obtain $d(z, \mathcal{F}) \leq \frac{D_Z}{\theta} [w(z)]_+$, $\forall z \in Z$. On one hand, for any $z \in \mathcal{F}_{\mathcal{N}, \mathcal{M}}$, we have

$$\begin{aligned} d(z, \mathcal{F}) &\leq \frac{D_Z}{\theta} [w(z)]_+ \leq \frac{D_Z}{\theta} (|w(z) - w_{\mathcal{N}, \mathcal{M}}(z)| + [w_{\mathcal{N}, \mathcal{M}}(z)]_+) \\ &= \frac{D_Z}{\theta} |w(z) - w_{\mathcal{N}, \mathcal{M}}(z)| \leq \frac{2D_Z}{\theta} (L\alpha_{\mathcal{N}} + \gamma_{\mathcal{M}}). \end{aligned}$$

This implies

$$\mathbb{D}(\mathcal{F}_{\mathcal{N}, \mathcal{M}}, \mathcal{F}) \leq \frac{2D_Z}{\theta} (L\alpha_{\mathcal{N}} + \gamma_{\mathcal{M}}). \quad (11)$$

Since $w_{\mathcal{N}, \mathcal{M}}(z) \leq w(z)$, $\forall z \in Z$, Assumption 2 implies $w_{\mathcal{N}, \mathcal{M}}(\bar{z}) < -\theta$. This means that the convex inequality constraint $w_{\mathcal{N}, \mathcal{M}}(z) \leq 0$ also satisfies the Slater constraint qualification. On the other hand, for any $z \in \mathcal{F}$, we have

$$d(z, \mathcal{F}_{\mathcal{N}, \mathcal{M}}) \leq \frac{D_Z}{\theta} [w_{\mathcal{N}, \mathcal{M}}(z)]_+ \leq \frac{D_Z}{\theta} |w_{\mathcal{N}, \mathcal{M}}(z) - w(z)| \leq \frac{2D_Z}{\theta} (L\alpha_{\mathcal{N}} + \gamma_{\mathcal{M}}),$$

which implies

$$\mathbb{D}(\mathcal{F}, \mathcal{F}_{\mathcal{N}, \mathcal{M}}) \leq \frac{2D_Z}{\theta} (L\alpha_{\mathcal{N}} + \gamma_{\mathcal{M}}). \quad (12)$$

Inequalities (11) and (12) mean conclusion (i).

(ii) Let $\tilde{f}(z) = f(z) + \delta_{\mathcal{F}}(z)$ and $\tilde{f}_{\mathcal{N}, \mathcal{M}}(z) = f(z) + \delta_{\mathcal{F}_{\mathcal{N}, \mathcal{M}}}(z)$. Since $\mathcal{F}_{\mathcal{N}, \mathcal{M}}$ converges to \mathcal{F} , then by [37, Proposition 7.4(f)], $\delta_{\mathcal{F}_{\mathcal{N}, \mathcal{M}}}$ epi-converges to $\delta_{\mathcal{F}}$ as $\mathcal{N} \rightarrow \infty, \mathcal{M} \rightarrow \infty$. As f is continuous and finite, we obtain by [37, Exercise 7.8] that $\tilde{f}_{\mathcal{N}, \mathcal{M}} = f + \delta_{\mathcal{F}_{\mathcal{N}, \mathcal{M}}}$ epi-converges to $\tilde{f} = f + \delta_{\mathcal{F}}$ when $\mathcal{N} \rightarrow \infty, \mathcal{M} \rightarrow \infty$. As $\mathcal{F}, \mathcal{F}_{\mathcal{N}, \mathcal{M}}$ are closed and f is continuous, $\tilde{f}_{\mathcal{N}, \mathcal{M}}$ and \tilde{f} are lower semi-continuous. Moreover, since $\tilde{f}_{\mathcal{N}, \mathcal{M}}$ and \tilde{f} are proper, it can then be deduced from [37, Theorem 7.33] that $v = \lim_{\mathcal{N} \rightarrow \infty} v_{\mathcal{N}, \mathcal{M}}$ and $\limsup_{\mathcal{M} \rightarrow \infty} \mathcal{S}_{\mathcal{N}, \mathcal{M}} \subset \mathcal{S}$.

(iii) Let $z \in \mathcal{S}$ and $z_{\mathcal{N}, \mathcal{M}} \in \mathcal{S}_{\mathcal{N}, \mathcal{M}}$. By definition of $\mathbb{D}(\mathcal{F}_{\mathcal{N}, \mathcal{M}}, \mathcal{F})$, $d(z_{\mathcal{N}, \mathcal{M}}, \mathcal{F}) \leq \mathbb{D}(\mathcal{F}_{\mathcal{N}, \mathcal{M}}, \mathcal{F})$. Then there exists a $z' \in \mathcal{F}$ such that $\|z_{\mathcal{N}, \mathcal{M}} - z'\| \leq \mathbb{D}(\mathcal{F}_{\mathcal{N}, \mathcal{M}}, \mathcal{F})$. From the Lipschitz continuity of f , we have

$$\begin{aligned} v = f(z) &\leq f(z') \leq |f(z') - f(z_{\mathcal{N}, \mathcal{M}})| + f(z_{\mathcal{N}, \mathcal{M}}) \\ &\leq L_f \|z' - z_{\mathcal{N}, \mathcal{M}}\| + f(z_{\mathcal{N}, \mathcal{M}}) \leq L_f \mathbb{D}(\mathcal{F}_{\mathcal{N}, \mathcal{M}}, \mathcal{F}) + v_{\mathcal{N}, \mathcal{M}}. \end{aligned}$$

Exchanging the roles of $z_{\mathcal{N}, \mathcal{M}}$ and z , we obtain $v_{\mathcal{N}, \mathcal{M}} \leq L_f \mathbb{D}(\mathcal{F}, \mathcal{F}_{\mathcal{N}, \mathcal{M}}) + v$. Applying conclusion (i), for any \mathcal{N} and \mathcal{M} , we have

$$|v_{\mathcal{N}, \mathcal{M}} - v| \leq L_f \mathbb{H}(\mathcal{F}_{\mathcal{N}, \mathcal{M}}, \mathcal{F}) \leq \frac{2D_Z L_f}{\theta} (L\alpha_{\mathcal{N}} + \gamma_{\mathcal{M}}).$$

Now, we show (8). Let $z \in \mathcal{S}$ and $z_{\mathcal{N}, \mathcal{M}} \in \mathcal{S}_{\mathcal{N}, \mathcal{M}}$. Denote by $\Pi_A(x) \in \operatorname{arginf}_{s \in A} \|x - s\|$ the projection of a point x on a set A . If problem (4) satisfies the second order growth condition, then

$$\begin{aligned} f(z_{\mathcal{N}, \mathcal{M}}) - f(\Pi_{\mathcal{F}}(z_{\mathcal{N}, \mathcal{M}})) &= f(z_{\mathcal{N}, \mathcal{M}}) - f(z) - (f(\Pi_{\mathcal{F}}(z_{\mathcal{N}, \mathcal{M}})) - f(z)) \\ &\leq f(\Pi_{\mathcal{F}_{\mathcal{N}, \mathcal{M}}}(z)) - f(z) - \rho d(\Pi_{\mathcal{F}}(z_{\mathcal{N}, \mathcal{M}}), \mathcal{S})^2. \end{aligned}$$

Since f is Lipschitz continuous, we have

$$d(\Pi_{\mathcal{F}}(z_{\mathcal{N},\mathcal{M}}), \mathcal{S}) \leq \sqrt{(L_f/\rho)(\|\Pi_{\mathcal{F}_{\mathcal{N},\mathcal{M}}}(z) - z\| + \|\Pi_{\mathcal{F}}(z_{\mathcal{N},\mathcal{M}}) - z_{\mathcal{N},\mathcal{M}}\|)}.$$

By triangle inequality and the definition of projection Π , we get

$$\begin{aligned} d(z_{\mathcal{N},\mathcal{M}}, \mathcal{S}) &\leq \|z_{\mathcal{N},\mathcal{M}} - \Pi_{\mathcal{F}}(z_{\mathcal{N},\mathcal{M}})\| + d(\Pi_{\mathcal{F}}(z_{\mathcal{N},\mathcal{M}}), \mathcal{S}) \\ &\leq \|z_{\mathcal{N},\mathcal{M}} - \Pi_{\mathcal{F}}(z_{\mathcal{N},\mathcal{M}})\| + \sqrt{(L_f/\rho)(\|\Pi_{\mathcal{F}_{\mathcal{N},\mathcal{M}}}(z) - z\| + \|\Pi_{\mathcal{F}}(z_{\mathcal{N},\mathcal{M}}) - z_{\mathcal{N},\mathcal{M}}\|)} \\ &= d(z_{\mathcal{N},\mathcal{M}}, \mathcal{F}) + \sqrt{(L_f/\rho)(d(z, \mathcal{F}_{\mathcal{N},\mathcal{M}}) + d(z_{\mathcal{N},\mathcal{M}}, \mathcal{F}))} \\ &\leq \mathbb{H}(\mathcal{F}_{\mathcal{N},\mathcal{M}}, \mathcal{F}) + \sqrt{(2L_f/\rho)\mathbb{H}(\mathcal{F}_{\mathcal{N},\mathcal{M}}, \mathcal{F})} \leq (1 + \sqrt{2L_f/\rho})\sqrt{\mathbb{H}(\mathcal{F}_{\mathcal{N},\mathcal{M}}, \mathcal{F})}, \end{aligned}$$

where the last inequality holds as $\mathbb{H}(\mathcal{F}_{\mathcal{N},\mathcal{M}}, \mathcal{F}) \leq \sqrt{\mathbb{H}(\mathcal{F}_{\mathcal{N},\mathcal{M}}, \mathcal{F})}$ for sufficiently large \mathcal{N} and \mathcal{M} . Since $z_{\mathcal{N},\mathcal{M}}$ is arbitrarily chosen from $\mathcal{S}_{\mathcal{N},\mathcal{M}}$, then (8) follows from (i). \square

Theorem 1 establishes the quantitative convergency for problem (7) in the sense of the feasible set, the optimal value and the optimal solution set. For the case that the support set Ξ is finite, the lower bound approximation (7) is tight.

Corollary 1. *If Ξ is a finite set, $\Xi_{\mathcal{N}} = \Xi$, and $\Gamma_{\mathcal{M}} = \{z_0^T \xi \mid \xi \in \Xi\}$, the optimal values of problem (4) and problem (7) are equal.*

Proof. The conclusion follows from [8, Proposition 3.2]. \square

3.2 Tractability of the lower bound approximation problem (7)

Recall that $\Xi_{\mathcal{N}} = \{\bar{\xi}_j\}_{j=1}^{\mathcal{N}}$ and $\Gamma_{\mathcal{M}} = \{\eta_k\}_{k=1}^{\mathcal{M}}$, problem (7) can be rewritten as

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{s.t.} \quad & \sup_{P \in \mathcal{Q}_{\mathcal{N}}} \mathbb{E}_P[(\eta_k - z^T \xi)_+ - (\eta_k - z_0^T \xi)_+] \leq 0, \quad k = 1, \dots, \mathcal{M}. \end{aligned}$$

Then by Lemma 1, its optimal value is equal to that of

$$\begin{aligned} \min_{z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{M}}} \quad & f(z) \\ \text{s.t.} \quad & \lambda_k \epsilon - \frac{1}{N} \sum_{i=1}^N \min_{1 \leq j \leq \mathcal{N}} \left[\lambda_k \|\bar{\xi}_j - \hat{\xi}_i\| - (\eta_k - z^T \bar{\xi}_j)_+ + (\eta_k - z_0^T \bar{\xi}_j)_+ \right] \leq 0, \\ & k = 1, \dots, \mathcal{M}. \end{aligned} \quad (13)$$

By introducing auxiliary variables β_{ik} , $i = 1, \dots, N, k = 1, \dots, \mathcal{M}$, problem (13) can be reformulated as

$$\min_{z, \lambda, \beta} \quad f(z) \quad (14a)$$

$$\text{s.t.} \quad \lambda_k \epsilon - \frac{1}{N} \sum_{i=1}^N \beta_{ik} \leq 0, \quad k = 1, \dots, \mathcal{M}, \quad (14b)$$

$$\beta_{ik} \leq \lambda_k \|\bar{\xi}_j - \hat{\xi}_i\| - (\eta_k - z^T \bar{\xi}_j)_+ + (\eta_k - z_0^T \bar{\xi}_j)_+, \quad (14c)$$

$$i = 1, \dots, N, \quad j = 1, \dots, \mathcal{N}, \quad k = 1, \dots, \mathcal{M}, \quad (14d)$$

$$z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{M}}, \beta \in \mathbb{R}^{N \times \mathcal{M}}.$$

Problem (14) is equivalent to problem (7), and thus is a lower bound approximation of problem (P_{SSD}).

By introducing auxiliary variables s_{jk} , $j = 1, \dots, \mathcal{N}, k = 1, \dots, \mathcal{M}$, to handle $(\eta_k - z^T \bar{\xi}_j)_+$ (refer to [8] (3.10)-(3.12)), we have a linear programming reformulation of problem (14)

$$\min_{z, \lambda, \beta, s} \quad f(z)$$

$$\begin{aligned}
& \text{s.t.} \quad \lambda_k \epsilon - \sum_{i=1}^N \frac{1}{N} \beta_{ik} \leq 0, \quad k = 1, \dots, \mathcal{M}, \\
(P_{SSD-L}) \quad & \beta_{ik} + s_{jk} \leq \lambda_k \|\bar{\xi}_j - \hat{\xi}_i\| + (\eta_k - z_0^T \bar{\xi}_j)_+, \\
& \quad \quad \quad i = 1, \dots, N, \quad j = 1, \dots, \mathcal{N}, \quad k = 1, \dots, \mathcal{M}, \\
& s_{jk} \geq \eta_k - z^T \bar{\xi}_j, \quad j = 1, \dots, \mathcal{N}, \quad k = 1, \dots, \mathcal{M}, \\
& z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{M}}, \beta \in \mathbb{R}^{N \times \mathcal{M}}, s \in \mathbb{R}_+^{N \times \mathcal{M}}.
\end{aligned}$$

In fact, the dimension of s is $\mathcal{N} \times \mathcal{M}$ and the number of constraints in problem (P_{SSD-L}) is $\mathcal{M} + N \times \mathcal{N} \times \mathcal{M} + \mathcal{N} \times \mathcal{M}$. Thus the size of problem (P_{SSD-L}) increases rapidly with the increase of the sample sizes \mathcal{N} and \mathcal{M} .

Algorithm 1 Cutting-plane Method

Start from $\iota = 1$ and $\mathcal{J}_1^{\iota} = \mathcal{J}_2^{\iota} = \emptyset$.

while $\iota \geq 1$ **do**

Solve the approximate problem:

$$\begin{aligned}
& \min_{z, \lambda, \beta, s} f(z) \\
& \text{s.t.} \quad \beta_{ik} + s_{jk} \leq \lambda_k \|\bar{\xi}_j - \hat{\xi}_i\| + (\eta_k - z_0^T \bar{\xi}_j)_+, \quad i = 1, \dots, N, \quad j \in \mathcal{J}_1^{\iota}, \quad k \in \mathcal{J}_2^{\iota}, \\
& \quad \quad \quad s_{jk} \geq \eta_k - z^T \bar{\xi}_j, \quad j \in \mathcal{J}_1^{\iota}, \quad k \in \mathcal{J}_2^{\iota}, \\
& (14b), \quad z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{M}}, \beta \in \mathbb{R}^{N \times \mathcal{M}}, s \in \mathbb{R}_+^{N \times \mathcal{M}}.
\end{aligned} \tag{15}$$

Let $(z^{\iota}, \lambda^{\iota}, \beta^{\iota}, s^{\iota})$ denote the optimal solution of problem (15).

Calculate

$$\delta^{\iota} := \max_{\substack{i \in \{1, \dots, N\}, j \in \{1, \dots, \mathcal{N}\}, \\ k \in \{1, \dots, \mathcal{M}\}}} \left\{ \beta_{ik}^{\iota} - \lambda_k^{\iota} \|\bar{\xi}_j - \hat{\xi}_i\| + (\eta_k - (z^{\iota})^T \bar{\xi}_j)_+ - (\eta_k - z_0^T \bar{\xi}_j)_+ \right\}.$$

if $\delta^{\iota} \leq 0$ **then**

Stop.

else

Determine

$$\begin{aligned}
& (i^{\iota}, j^{\iota}, k^{\iota}) \in \\
& \quad \quad \quad \operatorname{argmax}_{\substack{i \in \{1, \dots, N\}, j \in \{1, \dots, \mathcal{N}\}, \\ k \in \{1, \dots, \mathcal{M}\}}} \left\{ \beta_{ik}^{\iota} - \lambda_k^{\iota} \|\bar{\xi}_j - \hat{\xi}_i\| + (\eta_k - (z^{\iota})^T \bar{\xi}_j)_+ - (\eta_k - z_0^T \bar{\xi}_j)_+ \right\}.
\end{aligned}$$

Let $\mathcal{J}_1^{\iota+1} = \mathcal{J}_1^{\iota} \cup j^{\iota}$, $\mathcal{J}_2^{\iota+1} = \mathcal{J}_2^{\iota} \cup k^{\iota}$ and $\iota \leftarrow \iota + 1$.

end if

end while

In order to numerically solve problem (14) for large \mathcal{N} , \mathcal{M} , we propose a cutting-plane method, see Algorithm 1. At each iteration of the cutting-plane method, we solve problem (15), a relaxation of problem (14). After solving (15), we check whether all the constraints in (14c) are satisfied or not. If all the constraints in (14c) are satisfied, then the optimal solution we find for problem (15) is also optimal for problem (14). Otherwise, we add the violated constraint to the approximate problem (15) at the next iteration.

Proposition 2. *Algorithm 1 stops at the optimal value and optimal solution of problem (14) within finite steps.*

Proof. It is easy to see that $\mathcal{J}_1^{\iota} \subsetneq \mathcal{J}_1^{\iota+1}$ or $\mathcal{J}_2^{\iota} \subsetneq \mathcal{J}_2^{\iota+1}$. As the possible number of constraints that can be added is finite, Algorithm 1 must stop at the optimal value and optimal solution of problem (14) within finite steps. \square

To conclude this section, we develop a lower bound approximation (7) for the distributionally robust SSD constrained optimization problem (P_{SSD}) . Problem (7) can be reformulated as problem (14), which can be

easily solved using linear programming formulation (P_{SSD-L}) or by Algorithm 1.

4 Upper bound approximation of distributionally robust SSD constrained problem

Notice that problem (4) can be rewritten as

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{s.t.} \quad & \sup_{P \in \mathcal{Q}} \sup_{\eta \in \mathcal{R}} \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \leq 0. \end{aligned} \quad (16)$$

If we exchange the order of operators $\sup_{\eta \in \mathcal{R}}$ and \mathbb{E}_P in problem (16), we obtain an upper bound approximation for problem (16). However, such an upper bound approximation might be loose or even infeasible since the gap

$$\mathbb{E}_P \left[\sup_{\eta \in \mathcal{R}} \{(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+\} \right] - \sup_{\eta \in \mathcal{R}} \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \quad (17)$$

might be large. This is because we determine an η for all possible ξ 's in the latter supremum in (17), while we determine an η for each realization of ξ in the former supremum in (17). The larger the range \mathcal{R} of η , the larger the gap in (17). As an extreme case, when \mathcal{R} reduces to a singleton, the gap (17) becomes 0. This observation motivates us to divide \mathcal{R} into small sub-intervals, and exchange the order of the expectation operator and the supremum over each sub-interval, which provides an upper bound approximation of the sub-problem in the sub-interval. Summing all sub-problems in all sub-intervals, we obtain an improved upper bound approximation of problem (16). We name such a bounding method a split-and-dual framework.

In detail, we divide $\mathcal{R} = [\mathcal{R}_{\min}, \mathcal{R}_{\max}]$ into \mathcal{K} intervals with disjoint interiors, $[\underline{\eta}_k, \bar{\eta}_k]$, $k = 1, \dots, \mathcal{K}$, where the boundary points of the intervals are specified by $\underline{\eta}_k = \mathcal{R}_{\min} + (k-1) \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}$, $\bar{\eta}_k = \mathcal{R}_{\min} + k \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}$, $k = 1, \dots, \mathcal{K}$. Notice that problem (16) can also be reformulated as

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{s.t.} \quad & \max_{1 \leq k \leq \mathcal{K}} \sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \sup_{P \in \mathcal{Q}} \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \leq 0, \end{aligned}$$

or, equivalently,

$$\min_{z \in Z} \quad f(z) \quad (18a)$$

$$\text{s.t.} \quad \sup_{P \in \mathcal{Q}} \sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \leq 0, \quad k = 1, \dots, \mathcal{K}. \quad (18b)$$

Exchanging the order of operators $\sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]}$ and \mathbb{E}_P in (18b), we have the following approximation problem

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{s.t.} \quad & \sup_{P \in \mathcal{Q}} \mathbb{E}_P \left[\sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \{(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+\} \right] \leq 0, \quad k = 1, \dots, \mathcal{K}. \end{aligned} \quad (19)$$

The feasible set of problem (18) contains that of problem (19). Thus problem (19) provides an upper bound approximation for problem (18).

4.1 Quantitative analysis of the upper approximation

In what follows, we show that when the interval number \mathcal{K} goes to infinity, the optimal value of problem (19) converges to that of (18). To this end, we first prove the convergence from

$$g(z, \mathcal{K}) := \max_{1 \leq k \leq \mathcal{K}} \sup_{P \in \mathcal{Q}} \mathbb{E}_P \left[\sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \{(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+\} \right]$$

to

$$\begin{aligned} g(z) &:= \sup_{P \in \mathcal{Q}} \sup_{\eta \in \mathcal{R}} \mathbb{E}_P \left[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+ \right] \\ &= \max_{1 \leq k \leq \mathcal{K}} \sup_{P \in \mathcal{Q}} \sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \mathbb{E}_P \left[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+ \right]. \end{aligned}$$

As $\cup_{k=1, \dots, \mathcal{K}} [\underline{\eta}_k, \bar{\eta}_k] = \mathcal{R}$, the function $g(z)$ does not depend on the splitting of \mathcal{R} .

Proposition 3. *Given Assumption 1, for any positive integer \mathcal{K} , $g(\cdot, \mathcal{K})$ and $g(\cdot)$ are Lipschitz continuous with modulus $\mathcal{C} = \sup_{P \in \mathcal{Q}} \mathbb{E}_P[\|\xi\|] < \infty$.*

Proof. It follows from [11, page 164]. □

Next, we prove that $g(z, \mathcal{K})$ converges to $g(z)$ when \mathcal{K} goes to infinity.

Proposition 4. *Given Assumption 1, we have that*

$$g(z, \mathcal{K}) - g(z) \leq 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}},$$

and $\lim_{\mathcal{K} \rightarrow \infty} g(z, \mathcal{K}) = g(z)$, uniformly with respect to $z \in Z$.

Proof. Denote

$$\eta_k^*(\omega) \in \operatorname{argsup}_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \left\{ (\eta - z^T \xi(\omega))_+ - (\eta - z_0^T \xi(\omega))_+ \right\}, \quad \omega \in \Omega, \quad k = 1, \dots, \mathcal{K}$$

Denote

$$\eta_k^{**} \in \operatorname{argsup}_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \mathbb{E}_P \left[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+ \right], \quad k = 1, \dots, \mathcal{K}.$$

Notice that η_k^* is a random variable, while η_k^{**} is a real number. Since $\eta_k^*(\omega)$ and η_k^{**} take values in the same interval $[\underline{\eta}_k, \bar{\eta}_k]$, for any $\omega \in \Omega$, we have $|\eta_k^*(\omega) - \eta_k^{**}| \leq \bar{\eta}_k - \underline{\eta}_k = \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}$. Then we obtain

$$\begin{aligned} g(z, \mathcal{K}) - g(z) &\leq \max_{1 \leq k \leq \mathcal{K}} \sup_{P \in \mathcal{Q}} \left\{ \mathbb{E}_P \left[\sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \left\{ (\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+ \right\} \right] \right. \\ &\quad \left. - \sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \mathbb{E}_P \left[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+ \right] \right\} \\ &\leq \max_{1 \leq k \leq \mathcal{K}} \sup_{P \in \mathcal{Q}} \mathbb{E}_P \left| \left[(\eta_k^* - z^T \xi)_+ - (\eta_k^* - z_0^T \xi)_+ \right] - \left[(\eta_k^{**} - z^T \xi)_+ - (\eta_k^{**} - z_0^T \xi)_+ \right] \right| \\ &\leq 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}, \end{aligned}$$

where the last inequality is due to the Lipschitz continuity of the positive part function $(\cdot)_+$. Then the conclusion immediately follows. □

Proposition 4 shows that to control the approximation error of the constraint function, the interval number \mathcal{K} should be large enough when the range \mathcal{R} is large.

We denote the feasible sets of problem (18) and problem (19) by \mathcal{F} and $\mathcal{F}_{\mathcal{K}}$, the optimal solution sets by \mathcal{S} and $\mathcal{S}_{\mathcal{K}}$, and the optimal values by v and $v_{\mathcal{K}}$, respectively. It is clear that $\mathcal{F}_{\mathcal{K}} \subset \mathcal{F}$, $\forall \mathcal{K}$. To derive the convergence from $v_{\mathcal{K}}$ to v , as well as the quantitative approximation error estimation, we need some constraint qualification, e.g., Mangasarian Fromovitz constraint qualification (MFCQ) [30]. However, classical MFCQ works only in the differentiable case (e.g., [27]), while function $g(\cdot, \mathcal{K})$ here is non-smooth. We can find that $g(\cdot, \mathcal{K})$ and $g(\cdot)$ are continuous and convex, and thus subdifferentiable everywhere. Therefore, it is reasonable for us to extend MFCQ to the subdifferentiable case.

Definition 3. (ND-MFCQ) *Let $F(t) := \{x \in \mathbb{R}^n \mid g_j(x, t) \leq 0, j \in J\}$ with subdifferentiable g_j , here t is the parameter in the constraints. Given \bar{t} and $\bar{x} \in F(\bar{t})$, if there exist some vector θ and real constants $\sigma < 0$, $\alpha_1 > 0$, $\alpha_2 > 0$ such that*

$$\langle \varsigma, \theta \rangle \leq \sigma < 0, \quad \forall \varsigma \in \partial g_j(x, t), \forall x : \|x - \bar{x}\| \leq \alpha_1, \forall t : \|t - \bar{t}\| \leq \alpha_2, \forall j \in J_0(\bar{x}, \bar{t}),$$

where $J_0(\bar{x}, \bar{t}) := \{j \in J \mid g_j(\bar{x}, \bar{t}) = 0\}$, then we say that non-differentiable MFCQ (ND-MFCQ) holds at (\bar{x}, \bar{t}) with θ, σ, α_1 and α_2 ,

We notice that the MFCQ condition under nonsmooth cases has been discussed in some literature, such as [41, Page 14]. In fact, the ND-MFCQ in Definition 3 is more strict than that in [41]. We require ND-MFCQ hold uniformly in a neighborhood of (\bar{x}, \bar{t}) with the same vector θ , while [41] only requires ND-MFCQ hold at the point \bar{x} . ND-MFCQ is an extension of MFCQ [30], and is equivalent to MFCQ if the constraint functions are differentiable.

In this paper, we only have one constraint and thus $J = \{1\}$. Our decision variable z corresponds to x and our parameter $\frac{1}{\mathcal{K}}$ corresponds to t in Definition 3. To arrive at the convergence result, we also require the following assumption.

Assumption 4. *The optimal solution set of problem (19) with $\mathcal{K} = 1$, denoted by \mathcal{S}_1 , is nonempty.*

From (18b) and the constraints in (19), we have that $\mathcal{S} \supset \mathcal{S}_{\mathcal{K}} \supset \mathcal{S}_1$ for any \mathcal{K} . Therefore, if \mathcal{S}_1 is nonempty, then \mathcal{S} and $\mathcal{S}_{\mathcal{K}}, \forall \mathcal{K}$, are also nonempty.

Theorem 2. *Given Assumptions 1 and 4. For some $z^* \in \mathcal{S}$, assume that ND-MFCQ holds at $(z^*, 0)$ with θ, σ, α_1 , and α_2 as is defined in Definition 3. If the objective function f is Lipschitz continuous with modulus L_f , then for $\mathcal{K} \geq \max \left\{ \frac{1}{\alpha_2}, \frac{2}{|\sigma|} \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\alpha_1} \|\theta\|, -2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{g(z^*)} \left(C \frac{\|\theta\|}{|\sigma|} + 1 \right) \right\}$, we have that*

$$|v_{\mathcal{K}} - v| \leq L_f \frac{2\|\theta\|}{|\sigma|} \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}},$$

and $\lim_{\mathcal{K} \rightarrow \infty} v_{\mathcal{K}} = v$.

Proof. For any \mathcal{K} , let $z_{\mathcal{K}} = z^* - \frac{2}{\sigma} \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} \theta$. Since $z^* \in \mathcal{S}$, then obviously $g(z^*) \leq 0$.

Firstly, we claim that $z_{\mathcal{K}} \in \mathcal{F}_{\mathcal{K}}$. To prove this, we examine two cases on whether the constraint $g(z) \leq 0$ is active at z^* .

- Case 1: $g(z^*) = 0$. By Proposition 4, we have

$$\begin{aligned} g(z_{\mathcal{K}}, \mathcal{K}) &= g(z_{\mathcal{K}}, \mathcal{K}) - g(z^*, \mathcal{K}) + g(z^*, \mathcal{K}) - g(z^*) + g(z^*) \\ &= g(z_{\mathcal{K}}, \mathcal{K}) - g(z^*, \mathcal{K}) + g(z^*, \mathcal{K}) - g(z^*) \leq g(z_{\mathcal{K}}, \mathcal{K}) - g(z^*, \mathcal{K}) + 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}. \end{aligned}$$

By the extended mean-value theorem [37, Theorem 10.48], for some $\tau \in (0, 1)$ and the corresponding point $z_{\mathcal{K}}^{\tau} = (1 - \tau)z_{\mathcal{K}} + \tau z^*$, there is a vector $\varsigma \in \partial g(z_{\mathcal{K}}^{\tau}, \mathcal{K})$ satisfying

$$g(z_{\mathcal{K}}, \mathcal{K}) \leq g(z_{\mathcal{K}}, \mathcal{K}) - g(z^*, \mathcal{K}) + 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} = \langle \varsigma, z_{\mathcal{K}} - z^* \rangle + 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}.$$

For any $\mathcal{K} \geq \frac{2}{|\sigma|} \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\alpha_1} \|\theta\|$, $z_{\mathcal{K}}^{\tau}$ is in the α_1 -neighborhood of z^* , which can be seen from

$$\|z_{\mathcal{K}}^{\tau} - z^*\| = \|(1 - \tau)(z_{\mathcal{K}} - z^*)\| \leq \|z_{\mathcal{K}} - z^*\| = \frac{2}{|\sigma|} \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} \|\theta\| \leq \alpha_1.$$

Therefore, by ND-MFCQ at $(z^*, 0)$, we have $\langle \varsigma, \theta \rangle \leq \sigma$ for $\varsigma \in \partial g(z_{\mathcal{K}}^{\tau}, \mathcal{K})$ and $\mathcal{K} \geq \max \left\{ \frac{2}{|\sigma|} \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\alpha_1} \|\theta\|, \frac{1}{\alpha_2} \right\}$. Then we have

$$\begin{aligned} g(z_{\mathcal{K}}, \mathcal{K}) &\leq \langle \varsigma, z_{\mathcal{K}} - z^* \rangle + 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} = -\frac{2}{\sigma} \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} \langle \varsigma, \theta \rangle + 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} \\ &\leq -\frac{2}{\sigma} \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} \sigma + 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} = 0. \end{aligned}$$

Then $z_{\mathcal{K}} \in \mathcal{F}_{\mathcal{K}}$.

- Case 2: $g(z^*) < 0$. Let $\delta := -g(z^*) > 0$. If $\mathcal{K} \geq 2 \frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\delta} \left(C \frac{\|\theta\|}{|\sigma|} + 1 \right)$, then we obtain from Propositions 3 and 4 that

$$|g(z_{\mathcal{K}}, \mathcal{K}) - g(z^*)| \leq |g(z_{\mathcal{K}}, \mathcal{K}) - g(z^*, \mathcal{K})| + |g(z^*, \mathcal{K}) - g(z^*)|$$

$$\leq \mathcal{C}\|z_{\mathcal{K}} - z^*\| + 2\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} = 2\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} \left(\mathcal{C}\frac{\|\theta\|}{|\sigma|} + 1 \right) \leq \delta.$$

This indicates that $g(z_{\mathcal{K}}, \mathcal{K}) \leq g(z^*) + \delta = 0$. Thus $z_{\mathcal{K}} \in \mathcal{F}_{\mathcal{K}}$ under this case.

Next, we estimate the approximation error $|v_{\mathcal{K}} - v|$ of the optimal values. Since $\mathcal{F}_{\mathcal{K}} \subset \mathcal{F}$, then $v_{\mathcal{K}} \geq v$. By Assumption 4, we can choose $z_{\mathcal{K}}^* \in \mathcal{S}_{\mathcal{K}}$. Then we have

$$\begin{aligned} 0 &\leq v_{\mathcal{K}} - v = f(z_{\mathcal{K}}^*) - f(z^*) \leq f(z_{\mathcal{K}}) - f(z^*) \\ &\leq L_f\|z_{\mathcal{K}} - z^*\| = L_f\frac{2\|\theta\|}{|\sigma|}\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}. \end{aligned}$$

The conclusions follow immediately. \square

Proposition 2 quantitatively estimates the approximation error between the optimal values of problem (19) and problem (18).

4.2 Reformulation of problem (19)

By applying Lemma 1 to each supremum problem w.r.t. P for $k = 1, \dots, \mathcal{K}$, we have a reformulation of problem (19)

$$\min_{z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{K}}} f(z) \tag{20a}$$

$$\begin{aligned} \text{s.t.} \quad & \lambda_k \epsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} \left\{ \right. \\ & \left. \sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]} \{(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+\} - \lambda_k \|\xi - \widehat{\xi}_i\| \right\} \leq 0, \quad k = 1, \dots, \mathcal{K}. \end{aligned} \tag{20b}$$

To simplify the notation, we write (20b) as

$$\lambda_k \epsilon + \frac{1}{N} \sum_{i=1}^N V_S^{ik} \leq 0, \quad k = 1, \dots, \mathcal{K}, \tag{21}$$

where

$$V_S^{ik} := \sup_{(\xi, \eta) \in \Xi \times [\underline{\eta}_k, \bar{\eta}_k]} (\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+ - \lambda_k \|\xi - \widehat{\xi}_i\|, \quad i = 1, \dots, N, \quad k = 1, \dots, \mathcal{K}.$$

In what follows, we derive a reformulation for V_S^{ik} . According to Assumption 1, V_S^{ik} is equivalent to

$$\begin{aligned} & \sup_{\xi, \eta, s, m} (\eta - z^T \xi)_+ - s - \lambda_k m \\ \text{s.t.} \quad & s \geq \eta - z_0^T \xi, \quad C\xi \leq d, \\ & \eta \geq \underline{\eta}_k, \eta \leq \bar{\eta}_k, \quad \|\xi - \widehat{\xi}_i\| \leq m, \\ & \xi \in \mathbb{R}^n, \eta \in \mathbb{R}, s \in \mathbb{R}_+, m \in \mathbb{R}. \end{aligned} \tag{22}$$

Problem (22) is a non-convex optimization problem with a piecewise linear objective function with two pieces. Examining the two pieces of the objective function separately, we can split problem (22) into two convex sub-problems:

$$\begin{aligned} V_{S1}^{ik} = & \sup_{\xi, \eta, s, m} \eta - z^T \xi - s - \lambda_k m & V_{S2}^{ik} = & \sup_{\xi, \eta, s, m} -s - \lambda_k m \\ \text{s.t.} \quad & s \geq \eta - z_0^T \xi, & \text{s.t.} \quad & s \geq \eta - z_0^T \xi, \\ & \eta - z^T \xi \geq 0, & & \eta - z^T \xi \leq 0, \\ (P_{SSD-1}^{ik}) & C\xi \leq d, & (P_{SSD-2}^{ik}) & C\xi \leq d, \\ & s \geq 0, & & s \geq 0, \\ & \eta \geq \underline{\eta}_k, & & \eta \geq \underline{\eta}_k, \end{aligned}$$

4.3 Sequential convex approximation for (P_{SSD-U})

We observe that bilinear terms $\mu_2^{ik}z$ and $\tilde{\mu}_2^{ik}z$ in problem (P_{SSD-U}) make it difficult to solve problem (P_{SSD-U}) directly. We apply a sequential convex approximation method to solve problem (P_{SSD-U}) , see Algorithm 2. The idea is to separate coupling variables. At each iteration, we fix z and optimize w.r.t. $\mu, \tilde{\mu}$; then fix $\mu, \tilde{\mu}$ and optimize w.r.t. z . The sequential convex approximation method finally generates a sequence of decisions whose objective values converge to an upper bound of the optimal value of (P_{SSD-U}) .

Algorithm 2 Sequential convex approximation

Start from $z^\iota \in Z, \iota = 1$.
while $\iota \geq 1$ **do**
 Solve problem (P_{SSD-U}) with an additional constraint $z = z^\iota$. Denote the optimal $\mu, \tilde{\mu}$ by $\mu^\iota, \tilde{\mu}^\iota$, respectively.
 Solve problem (P_{SSD-U}) with additional constraints $\mu = \mu^\iota, \tilde{\mu} = \tilde{\mu}^\iota$. Denote the optimal z by $z^{\iota+1}$.
 if $z^{\iota+1} = z^\iota$ **then**
 Break.
 else
 $\iota \leftarrow \iota + 1$.
 end if
end while

Proposition 5. *Suppose that the optimal value of problem (P_{SSD}) is finite. Given a starting point z^1 . Algorithm 2 generates a sequence of decisions whose objective values converge to an upper bound of the optimal value of problem (P_{SSD-U}) .*

Proof. Denote the feasible set of problem (P_{SSD-U}) by \mathcal{F}_U . We write all the decision variables excluding $z, \mu, \tilde{\mu}$ by y . We can thus write problem (P_{SSD-U}) in a compact form $\min\{f(z) \mid (z, \mu, \tilde{\mu}, y) \in \mathcal{F}_U\}$.

Firstly, observe that each problem we solve in Algorithm 2 has an additional constraint compared with problem (P_{SSD-U}) . Therefore, $f(z^\iota), \iota = 1, \dots$, are upper bounds to the optimal value of problem (P_{SSD-U}) .

Next, the sequence $\{f(z^\iota)\}$ has a finite lower bound, the optimal value of problem (P_{SSD}) . Thus in order to show the convergence of $\{f(z^\iota)\}$, it is sufficient to prove that $\{f(z^\iota)\}$ is nonincreasing. From Algorithm 2, there exists y' such that $(\mu^\iota, \tilde{\mu}^\iota, y') = \operatorname{argmin}_{\mu, \tilde{\mu}, y} \{f(z^\iota) \mid (z^\iota, \mu, \tilde{\mu}, y) \in \mathcal{F}_U\}$. It follows immediately that $(z^\iota, \mu, \tilde{\mu}, y') \in \mathcal{F}_U$. Also there exists y'' such that $(z^{\iota+1}, y'') = \operatorname{argmin}_{z, y} \{f(z) \mid (z, \mu^\iota, \tilde{\mu}^\iota, y) \in \mathcal{F}_U\}$. Since $(z^\iota, \mu^\iota, \tilde{\mu}^\iota, y') \in \mathcal{F}_U$, we have $f(z^{\iota+1}) \leq f(z^\iota)$. \square

Here it is necessary to point out that any element in the sequence of optimal values generated by Algorithm 2 is an upper bound of the optimal value of problem (P_{SSD-U}) . Each problem we solve in Algorithm 2 is a second-order cone programming and thus is computationally tractable.

To conclude this section, we divide \mathcal{R} into sub-intervals and exchange the order of the expectation operator and the supremum over each sub-interval to derive an upper bound approximation (P_{SSD-U}) for the distributionally robust SSD constrained optimization problem (P_{SSD}) . We prove the convergence of the optimal value of the upper bound approximation problem and quantitatively estimate the approximation error. To cope with bilinear terms in problem (P_{SSD-U}) , we apply the sequential convex approximation method, Algorithm 2, to obtain an upper bound of the optimal value of problem (P_{SSD-U}) .

5 Numerical experiments

In this section, we present the results of numerical experiments to illustrate the validity and practicality of our lower and upper bound approximation methods for model (P_{SSD}) . The numerical experiments are carried out by calling the Gurobi solver in CVX package in MATLAB R2016a on a Dell G7 laptop with Windows 10 operating system, Intel Core i7 8750H CPU 2.21 GHz and 16 GB RAM.

5.1 Case study: an illustrative numerical example

We begin with a simple numerical example and examine the validation of the proposed lower and upper bound approximations. Consider the following problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \|z\|_2 \\ \text{s.t.} \quad & \mathbb{E}_P[(\eta - z^T \xi)_+] \leq \mathbb{E}_P[(\eta - z_0^T \xi)_+], \quad \forall \eta \in \mathbb{R}, \quad \forall P \in \mathcal{Q}, \\ & z \in \mathbb{R}_2^+, \|z\|_1 \leq 1. \end{aligned} \tag{25}$$

where $z_0 = (1, 0)^T$ and $\mathcal{Q} = \{P \in \mathcal{P}(\Xi) : d_K(P, \hat{P}_N) \leq \epsilon\}$ is defined as that in (5). Here $\hat{P}_N = \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\xi}_i}$ is the empirical distribution. The support set is supposed to be $\Xi = \{(\xi_1, \xi_2)^T \mid \xi_1 \in [0, 250], \xi_2 \in [0, 500]\}$. We set $\epsilon = 10^{-5}$, $N = 10$ and the observed sample set $\{\hat{\xi}_i\}_{i=1}^{10}$ consists of $(0, 0)^T$, $(250, 0)^T$, $(0, 500)^T$, $(100, 100)^T$, $(200, 200)^T$, $(100, 0)^T$, $(200, 0)^T$, $(0, 100)^T$, $(0, 200)^T$, $(200, 500)^T$.

Table 1: The optimal values and the optimal solutions of the lower and upper bound approximations to problem (25).

lower bound approximation ((P_{SSD-L}) or Algorithm 1)				upper bound approximation (Algorithm 2)			Gap
\mathcal{N}	\mathcal{M}	Optimal value	Optimal solution	\mathcal{K}	Optimal value	Optimal solution	
100	100	0.2922	$(0.4229, 0.4027)^T$	10	0.4097	$(0.8010, 0.1564)^T$	40.2122%
200	200	0.2964	$(0.4266, 0.4077)^T$	11	0.3044	$(0.4590, 0.4002)^T$	2.6991%
300	300	0.3014	$(0.4423, 0.4014)^T$	12	0.3025	$(0.4653, 0.3868)^T$	0.3650%

We get the lower bound approximation by solving the linear programming formulation (P_{SSD-L}) or Algorithm 1 and obtain the upper bound approximation by Algorithm 2. The optimal values and the optimal solutions are shown in Table 1. We also calculate the relative gaps of the optimal values of the lower and upper bound approximations (i.e., $\text{Gap} = |\frac{\text{upper} - \text{lower}}{\text{lower}}|$). From Table 1, we can see that the relative gap between the optimal values of the lower and upper bound approximations decreases quickly to 0 with the increase of sample sizes \mathcal{N} , \mathcal{M} and the interval number \mathcal{K} , which verifies the validation of the proposed approximation methods.

5.2 Case study: a practical portfolio selection problem

We consider a financial application of model (P_{SSD}) to the portfolio selection problem with distributionally robust SSD constraints:

$$\begin{aligned} \min_{z \in Z} \quad & \mathbb{E}_{\hat{P}_N}[-z^T \xi] \\ \text{s.t.} \quad & \mathbb{E}_P[(\eta - z^T \xi)_+] \leq \mathbb{E}_P[(\eta - z_0^T \xi)_+], \quad \forall \eta \in \mathbb{R}, \quad \forall P \in \mathcal{Q}, \end{aligned} \tag{26}$$

where $Z = \{z \in \mathbb{R}^n \mid z \geq 0, \sum_{i=1}^n z_i = 1\}$. Problem (26) is inspired by [18, Example 4.2]. The difference between problem (26) and that in [18, Example 4.2] lies in the construction method of the ambiguity set. In [18, Example 4.2], the ambiguity set \mathcal{Q} is determined by first two order moment information, while in problem (26), \mathcal{Q} is a Wasserstein ball.

We use the same historical annual return rate data of eight risky assets as that in [8, Table 8.1] (with a total of $N = 22$ years). We choose the equally weighted portfolio as the benchmark portfolio z_0 . We select z_0 as the starting point z^1 in Algorithm 2.

In what follows, we show the numerical results emphatically illustrating from the following aspects: the convergence of the lower and upper bound approximations with respect to the sample sizes and the interval number, the price of introducing distributional robustness in SSD constraints, and the influence of the robust radius.

5.2.1 Convergence of lower and upper bounds

Firstly, we demonstrate the convergence of the lower bound approximation with respect to the sample sizes \mathcal{N} , \mathcal{M} and the decreasing trend of the upper bound approximation when the interval number \mathcal{K} increases. We fix the robust radius $\epsilon = 10^{-4}$.

For the lower bound approximation, we consider the cases with the sample sizes being $\mathcal{N} = \mathcal{M} = 40, 60, 80, 100, 120$. To make fair comparison later in Section 5.2.2 with the portfolio optimization problem with classic (non-robust)

SSD constraints, we let $\Xi_{\mathcal{N}}$ contain all the historical annual return rates $\{\widehat{\xi}_i\}_{i=1}^N$ from [8, Table 8.1] and $\Gamma_{\mathcal{M}}$ contain $\{z_0^T \widehat{\xi}_i\}_{i=1}^N$. For the upper bound approximation, we consider the cases with the interval number being $\mathcal{K} = 1, 2, 4, 8, 12$, respectively. Figure 2 shows the convergence trend of the optimal values of the lower and upper bound approximations for problem (26).

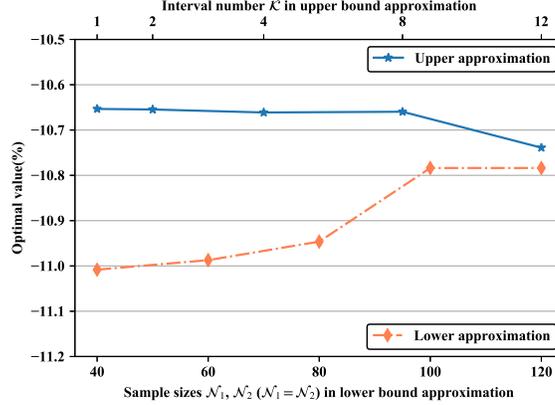


Figure 2: Optimal values of the lower bound approximation with respect to \mathcal{N}, \mathcal{M} and that of the upper bound approximation with respect to \mathcal{K} .

From Figure 2, we can observe that the lower bound monotonously increases with the increase of the sample sizes \mathcal{N}, \mathcal{M} and the upper bound decreases with the increase of the interval number \mathcal{K} . The gap between the lower and upper approaches 0. These observations verify the quantitative convergency established in Theorem 1 and Theorem 2. To see more details, we present in Table 2 the optimal values and the optimal solutions obtained from the lower and upper bound approximations.

Table 2: Optimal values and the optimal solutions of the lower bound approximation with respect to \mathcal{N}, \mathcal{M} , and those of the upper bound approximation with respect to \mathcal{K} .

lower bound approximation ((P_{SSD-L}) or Algorithm 1)			upper bound approximation (Algorithm 2)		Gap
\mathcal{N}	\mathcal{M}	Optimal value(%) Optimal solution	\mathcal{K}	Optimal value(%) Optimal solution	
40	40	-11.0082 (0.000,0.000,0.068,0.188,0.000,0.391,0.231,0.122)	1	-10.6534 (0.125,0.125,0.125,0.125,0.125,0.125,0.125,0.125)	3.2231%
60	60	-10.9872 (0.000,0.038,0.000,0.269,0.000,0.354,0.213,0.126)	2	-10.6543 (0.125,0.124,0.124,0.127,0.125,0.126,0.125,0.125)	3.0299%
80	80	-10.9463 (0.000,0.006,0.094,0.138,0.036,0.389,0.215,0.123)	4	-10.6546 (0.124,0.124,0.123,0.127,0.124,0.127,0.125,0.125)	2.6648%
100	100	-10.7838 (0.000,0.018,0.168,0.000,0.131,0.384,0.172,0.126)	8	-10.6551 (0.124,0.124,0.123,0.128,0.124,0.127,0.125,0.125)	1.1935%
120	120	-10.7838 (0.000,0.018,0.168,0.000,0.131,0.384,0.172,0.126)	12	-10.7389 (0.075,0.067,0.005,0.274,0.087,0.238,0.125,0.129)	0.4164%

From Table 2, we can see the changing trend of the optimal portfolios of the lower and upper bound approximations. Especially for the upper bound approximation, the optimal portfolio under $\mathcal{K} = 1$ is the equally weighted portfolio, while the optimal portfolio under $\mathcal{K} = 12$ is quite different from the equally weighted portfolio and approaches the optimal portfolios obtained from the lower bound approximation. We observe from Table 2 that the lower and upper bounds we finally obtain are not equal. This happens because when implementing the upper bound approximation we can only have finite sub-intervals and thus a gap is induced whenever we exchange the order of operators $\sup_{\eta \in [\underline{\eta}_k, \bar{\eta}_k]}$ and \mathbb{E}_P . We calculate the relative gap between the upper bound with $\mathcal{K} = 12$ and the lower bound with $\mathcal{N} = 120, \mathcal{M} = 120$, which is only $|\frac{10.7838 - 10.7389}{-10.7838}| = 0.4164\%$. This is quite satisfactory for real applications.

5.2.2 Price of distributional robustness

To examine the price of introducing distributional robustness, we compare the numerical result of problem (26) with that of classic SSD constrained portfolio optimization problem

$$\min \left\{ \mathbb{E}_{\widehat{P}_N} [-z^T \xi] \mid z \in Z, z^T \xi \succeq_{(2)} \widehat{P}_N z_0^T \xi \right\}. \quad (27)$$

Table 3 reports the comparative results of the optimal expected return rates, which are absolute values of the optimal values of problems (26) and (27) since they are minimization problems.

Table 3: Optimal expected return rates (absolute value of the optimal value) of the lower and upper bound approximations to problem (26), optimal expected return rate of problem (27), and expected return rate of z_0 .

Portfolio optimization problem	Expected return rate(%)
(26) lower bound approximation ($\mathcal{N} = \mathcal{M} = 120$)	10.7838
upper bound approximation ($\mathcal{K} = 12$)	10.7389
(27)	11.0082
Benchmark	10.6534

From Table 3, we can see that both the lower and upper bound approximations to problem (26) with distributionally robust SSD constraints derive a smaller optimal expected return rate than problem (27) with classic SSD constraints. Therefore, the optimal expected return rate of problem (26) must be smaller than that of problem (27). As we expected, considering the distributionally robust ambiguity in SSD constraints induces a more conservative solution. It can also be seen from Table 3 that the expected return rates of the lower and upper bound approximations are larger than that of the benchmark portfolio, which means that model (26) derives a portfolio better than the benchmark portfolio in sense of the expected return rate. These numerical results demonstrate that introducing distributional robustness brings in conservation without loss of stochastic dominance.

5.2.3 Influence of the robust radius

Finally, we briefly examine the impact of robust radius on the lower and upper bound approximations to the portfolio optimization problem (26). The optimal values of the lower and upper bound approximations under different robust radii are shown in Table 4.

Table 4: Optimal values of the lower and upper bound approximations, and their relative gaps with respect to different robust radii.

Robust radius ϵ	Optimal values (%)		Gap
	lower bound approximation	upper bound approximation	
10^{-5}	-10.8775	-10.8268	0.4661%
10^{-4}	-10.7838	-10.7389	0.4164%
10^{-3}	-10.7836	-10.6536	1.2055%
10^{-2}	-10.7823	-10.6535	1.1946%
0.1	-10.7689	-10.6534	1.0725%
0.5	-10.6885	-10.6534	0.3284%
1	-10.6534	-10.6534	0%

We can see from Table 4 that, as is expected both the optimal values of the lower and upper bound approximations of problem (26) are monotonously increasing, which implies that the optimal value of problem (26) increases as the robust radius increases. Table 4 also tells us that choosing a proper robust radius is a crucial issue in distributionally robust SSD constrained problems. For robust radius $\epsilon \geq 0.1$, the upper bound coincides with $\mathbb{E}_{\hat{\mathcal{P}}_N}[-z_0^T \xi]$, while for $\epsilon \leq 10^{-2}$ both the lower and upper bound approximations derive optimal portfolios better than the benchmark portfolio.

6 Conclusion

We consider a distributionally robust SSD constrained optimization problem, where the true distribution of the uncertain parameters is ambiguous. The ambiguity set contains those probability distributions close to the empirical distribution under the Wasserstein distance.

We propose two approximation methods to obtain bounds on the optimal value of the original problem. We adopt the sample approximation approach to develop a linear programming formulation to obtain a lower bound approximation for the problem. The lower bound approximation can be easily solved by using linear programming formulation or by the cutting-plane method. Moreover, we establish the quantitative convergency for the lower bound approximation problem. We also develop an upper bound approximation and quantitatively estimate the approximation error between the optimal value of the upper bound approximation and that of the

original problem. We propose a novel split-and-dual decomposition framework to reformulate distributionally robust SSD constraints. The upper bound approximation problem can be solved by a sequence of second-order cone programming problems. We carry out numerical experiments on a portfolio optimization problem to illustrate our lower and upper bound approximation methods.

One of future research topics would be modifying the design of cutting-planes to solve the lower bound approximation problem more efficiently. While for the upper bound approximation, it is interesting to investigate the critical number of intervals for further enhancing the practicality of the approximation scheme. Besides, finding efficient approximation and solution methods for distributionally robust multivariate robust SSD constrained optimization is also a promising topic.

Funding

This research was supported by the National Natural Science Foundation of China under grant numbers 11991023, 11991020, 11735011, and 11901449.

References

- [1] Yannan Chen, Hailin Sun, and Huifu Xu. Decomposition and discrete approximation methods for solving two-stage distributionally robust optimization problems. *Computational Optimization and Applications*, 78:205–238, 2021.
- [2] Zhi Chen, Daniel Kuhn, and Wolfram Wiesemann. Data-driven chance constrained programs over Wasserstein balls. *Available at arXiv*, 2018.
- [3] Zhiping Chen and Jie Jiang. Stability analysis of optimization problems with k th order stochastic and distributionally robust dominance constraints induced by full random recourse. *SIAM Journal on Optimization*, 28(2):1396–1419, 2018.
- [4] Zhiping Chen, Yu Mei, and Jia Liu. Multivariate robust second-order stochastic dominance and resulting risk-averse optimization. *Optimization*, 68(9):1719–1747, 2019.
- [5] Giorgio Consigli, Vittorio Moriggia, and Sebastiano Vitali. Long-term individual financial planning under stochastic dominance constraints. *Annals of Operations Research*, 292:973–1000, 2020.
- [6] Darinka Dentcheva, René Henrion, and Andrzej Ruszczyński. Stability and sensitivity of optimization problems with first order stochastic dominance constraints. *SIAM Journal on Optimization*, 18(1):322–337, 2007.
- [7] Darinka Dentcheva and Werner Römisch. Stability and sensitivity of stochastic dominance constrained optimization models. *SIAM Journal on optimization*, 23(3):1672–1688, 2013.
- [8] Darinka Dentcheva and Andrzej Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566, 2003.
- [9] Darinka Dentcheva and Andrzej Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance*, 30:433–451, 2006.
- [10] Darinka Dentcheva and Andrzej Ruszczyński. Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming*, 123(1):85–100, 2010.
- [11] Darinka Dentcheva, Andrzej Ruszczyński, and Alexander Shapiro. *Lectures on Stochastic Programming: Modeling and Theory*. Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia, 2009.
- [12] Jitka Dupačová and Miloš Kopa. Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research*, 234:434–441, 2014.
- [13] Laureano F. Escudero, María Araceli Garín, María Merino, and Gloria Pérez. On time stochastic dominance induced by mixed integer-linear recourse in multistage stochastic programs. *European Journal of Operational Research*, 249:164–176, 2016.

- [14] Csaba I Fábrián, Gautam Mitra, and Diana Roman. Processing second-order stochastic dominance models using cutting-plane representations. *Mathematical Programming*, 130(1):33–57, 2011.
- [15] Rui Gao and Anton J. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *Available at arXiv*, 2016.
- [16] Ralf Gollmer, Uwe Gotzes, and Rüdiger Schultz. A note on second-order stochastic dominance constraints induced by mixed-integer linear recourse. *Mathematical Programming*, 126(1):179–190, 2011.
- [17] Ralf Gollmer, Frederike Neise, and Rüdiger Schultz. Stochastic programs with first-order dominance constraints induced by mixed-integer linear recourse. *SIAM Journal on Optimization*, 19(2):552–571, 2008.
- [18] Shaoyan Guo, Huifu Xu, and Liwei Zhang. Probability approximation schemes for stochastic programs with distributionally robust second-order dominance constraints. *Optimization Methods and Software*, 32(4):770–789, 2017.
- [19] William B. Haskell, Lunce Fu, and Maged Dessouky. Ambiguity in risk preferences in robust stochastic optimization. *European Journal of Operational Research*, 254:214–225, 2016.
- [20] William B. Haskell, J. George Shanthikumar, and Z. Max Shen. Primal-dual algorithms for optimization with stochastic dominance. *SIAM Journal on Optimization*, 27(1):34–66, 2017.
- [21] Tito Homem-de-Mello and Sanjay Mehrotra. A cutting-surface method for uncertain linear programs with polyhedral stochastic dominance constraints. *SIAM Journal on Optimization*, 20(3):1250–1273, 2009.
- [22] Jian Hu and Gevorg Stepanyan. Optimization with reference-based robust preference constraints. *SIAM Journal on Optimization*, 27(4):2230–2257, 2017.
- [23] Ran Ji and Miguel Lejeune. Data-driven distributionally robust chance-constrained optimization with Wasserstein metric. *Journal of Global Optimization*, 79:779–811, 2021.
- [24] Ruiwei Jiang and Yongpei Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, 158:291–327, 2016.
- [25] Juuso Liesiö, Peng Xu, and Timo Kuosmanen. Portfolio diversification based on stochastic dominance under incomplete probability information. *European Journal of Operational Research*, 286:755–768, 2020.
- [26] Jia Liu, Abdel Lisser, and Zhiping Chen. Distributionally robust chance constrained geometric optimization. *Available at Optimization Online*, 2019.
- [27] Jia Liu, Shen Peng, Abdel Lisser, and Zhiping Chen. Rectangular chance constrained geometric optimization. *Optimization and Engineering*, 21:537–566, 2020.
- [28] Yongchao Liu, Hailin Sun, and Huifu Xu. An approximation scheme for stochastic programs with second order dominance constraints. *Numerical Algebra, Control & Optimization*, 6(4):473–490, 2016.
- [29] James Luedtke. New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19(3):1433–1450, 2008.
- [30] Olvi L. Mangasarian and Stan Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17:37–47, 1967.
- [31] Yu Mei, Zhiping Chen, Bingbing Ji, Zhujia Xu, and Jia Liu. Data-driven stochastic programming with distributionally robust constraints under Wasserstein distance: asymptotic properties. *Journal of the Operations Research Society of China*, 9:525–542, 2021.
- [32] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171:115–166, 2018.
- [33] Nilay Noyan and Gábor Rudolf. Optimization with stochastic preferences based on a general class of scalarization functions. *Operations Research*, 66(2):463–486, 2018.

- [34] Nilay Noyan, Gábor Rudolf, and Andrzej Ruszczyński. Relaxations of linear programming problems with first order stochastic dominance constraints. *Operations Research Letters*, 34(6):653–659, 2006.
- [35] Chun Peng and Erick Delage. Data-driven optimization with distributionally robust second-order stochastic dominance constraints. *Available at Optimization Online*, 2020.
- [36] Stephen M. Robinson. An application of error bounds for convex programming in a linear space. *SIAM Journal on Control*, 13(2):271–273, 1975.
- [37] R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*. Springer, Berlin, 2009.
- [38] Gábor Rudolf and Andrzej Ruszczyński. Optimization problems with second order stochastic dominance constraints: duality, compact formulations, and cut generation methods. *SIAM Journal on Optimization*, 19(3):1326–1343, 2008.
- [39] Hailin Sun, Huifu Xu, Rudabeh Meskarian, and Yong Wang. Exact penalization, level function method, and modified cutting-plane method for stochastic programs with second order stochastic dominance constraints. *SIAM Journal on Optimization*, 23(1):602–631, 2013.
- [40] Weijun Xie. On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming*, 186:115–155, 2021.
- [41] Jane J. Ye and Daoli Zhu. Optimality conditions for bilevel programming problems. *Optimization*, 33(1):9–27, 1995.
- [42] Sainan Zhang, Shaoyan Guo, Liwei Zhang, and Hongwei Zhang. On distributionally robust optimization problems with k -th order stochastic dominance constraints induced by full random quadratic recourse. *Journal of Mathematical Analysis and Applications*, 493(2):1–24, 2021.
- [43] Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.
- [44] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.