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, where the true distribution of the uncertain parameters is ambiguous. The ambiguity set contains all probability distributions close to the empirical distribution under the Wasserstein distance. We adopt the sample approximation technique 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 prove that both lower and upper bound approximations are asymptotically tight when there are enough samples or pieces. We present quantitative error estimation for the upper bound under a specific constraint qualification condition. 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 efficiency and asymptotically tightness of the proposed two approximation methods.
Keywords:
stochastic dominance; distributionally robust optimization; Wasserstein distance; sequential convex approximation
MSC:
90C59,90C34

1 Introduction

Stochastic dominance (SD), originated from economics, is popular in comparing random outcomes. In their pioneering work [7], Dentcheva and Ruszczyński studied the stochastic optimization problem with univariate SD constraints, where they developed the necessary and sufficient conditions of optimality and duality theory for the problem. 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 the stability and sensitivity analysis [5], the integer and linear programming formulation [26], and linear programming relaxations [31]. The stochastic optimization problem with SSD constraints has been intensively studied in quite a few literature. For theoretical foundations, the stability and sensitivity analysis was presented in [6]. For the solution methods of SSD constrained stochastic optimization problem, different linear programming formulations were derived in [7, 26], and the idea of cutting plane is adopted in [12, 33, 34]. The stochastic programs with SD constraints induced by mixed-integer linear recourse were studied in [15] for FSD and in [14] for SSD. Stochastic optimization problems with the multivariate extensions of SD constraints were considered in [18, 19, 30]. Haskell et al. defined multivariate SD in [18] using multivariate utility functions and Noyan and Rudolf defined in [30] by using scalarization functions to transform a multivariate random vector into a univariate random variable. To solve multivariate SD constrained stochastic optimization problems, Haskell et al. developed primal-dual algorithms [18], while Noyan and Rudolf adopted a cut generation method [30]. There is also a rich literature considering SD under dynamic settings, such as time stochastic dominance [11]. Applications of optimization with SD constraints in finance were investigated in [4, 8, 20].

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

A challenge of stochastic programming problems is the accessibility of the true probability distribution of the uncertain parameters. In practice, the true probability distribution sometimes is unknown or imprecisely observed through some training dataset. For this reason, distributionally robust optimization models have been proposed to address the lack of perfect probability distribution information, 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. Moment-based ambiguity sets, whose resulting distributionally robust optimization problems have been widely studied [9, 39], and distance-based ambiguity sets, which contain all the probability distributions close to some nominal distribution measured by some probability metrics, such as Kullback-Leibler divergence [24], ϕ-divergence [22], and Wasserstein distance [1, 21, 24, 28, 35]. Mohajerin Esfahani and Kuhn estimated the priori probability that the true distribution belongs to the Wasserstein ambiguity set and established finite sample and asymptotic guarantees for the distributionally robust solutions in [29]. Using the duality theory, researchers reformulated distributionally robust optimization under the Wasserstein ambiguity set as convex programs [13, 29, 38]. Such reformulations were then applied to chance-constrained distributionally robust optimization problems [1, 35].

Incorporating the ideas of SD and distributional robustness, Dentcheva and Ruszczyński first introduced the distributionally robust SD in [9], where they established the optimality conditions of the stochastic optimization problem with distributionally robust SSD constraints. Since then, a stream of researches has paid attention to stochastic optimization with distributionally robust SD constraints. Considering the distributional uncertainty, Dupačová and Kopa in [10] exploited the contamination technique to derive a robust decision that is FSD efficient. Guo, Xu, and Zhang in [16] proposed a discrete approximation scheme for the moment-based ambiguity sets and approximately solved the resulting stochastic optimization problem with distributionally robust SSD constraints. Also, under a moment-based type ambiguity set, Liesiö, Xu, and Kuosmanen in [23] developed models that identify a portfolio that is robust SSD over a given benchmark. In addition to distributionally robust FSD and SSD, some researches also focused on robust kth ($k > 2$) order SD. The stability of distributionally robust optimization problem with kth order distributionally robust SD constraints induced by full random recourse was established [2, 37]. Besides, multivariate extensions of distributionally robust SD and optimality conditions and duality theory of resulting stochastic optimization problems were discussed in [3, 17].

As is mentioned above, SD constrained optimization under distributional ambiguity is an important issue in many practical applications such as financial decision making. However, to the best of our knowledge, distributionally robust SSD constrained optimization with Wasserstein distance has not been studied in the existing literature. The main difficulties of solving such problems lie in three aspects.

- The two levels of semi-infinite constraints with respect to the SSD constraints and the distributional robustness are the main challenge in solving distributionally robust SSD constrained optimization problems.
- Distributionally robust SSD constraints are non-smooth, which makes gradient-based methods fail to work here.
- Compared to moment-based ambiguity sets, the ambiguity set with Wasserstein distance contains an extra optimization problem on computing the optimal transportation between the true and reference distributions. Such an inner-level optimization problem leads a min-max 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 under Wasserstein distance. Thanks to the rapid development recently on the strong duality theory of distributionally robust optimization problems under Wasserstein distance [13, 29], we have a chance to show in this paper efficient approximation methods for the distributionally robust SSD constrained optimization problem under Wasserstein distance.

In detail, we first utilize the duality theory in [13] to derive a dual reformulation of distributionally robust SSD constraints. Then we adopt the sampling technique to approximate the infinitely many constraints by finitely many constraints and develop a linear programming formulation to obtain a lower bound approximation for the problem, which is asymptotically tight as the sample size goes to infinity. 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 and the expectation 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 also 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 distributionally robust optimization with Wasserstein distance [13, 35] to a more complicated case with infinitely many constraints induced by SSD. Compared with robust SD constrained optimization problems in [16, 23, 37], we study the ambiguity set with Wasserstein distance rather than moment-based ambiguity sets. The main contributions in this paper include:

- For the first time, we study the distributionally robust SSD constrained optimization with Wasserstein distance.
- We adopt the sample approximation technique to develop a linear programming formulation that provides a lower bound, and prove that the lower bound approximation is asymptotically tight when there exist enough sample points.
- We propose a novel split-and-dual decomposition framework, which provides an upper bound approximation of the problem. In the existing literature, the upper bounds of SD constrained problems are seldom studied. We prove the asymptotic tightness of the split-and-dual decomposition method and quantitatively estimate the approximation error when the number of sub-intervals goes to infinity.

The rest of this paper is organized as follows. In section 2, we recall distributionally robust SSD and specify the ambiguity set as a Wasserstein ball. In section 3, we elaborate on the distributionally robust SSD constrained optimization with Wasserstein distance in detail. We develop a linear programming formulation to obtain a lower approximation and solve a sequence of second-order cone programming problems for an upper approximation. Numerical evidences valid the efficiency and asymptotically tightness of the proposed approximation methods in Section 5. Section 6 concludes the paper.

2 Preliminaries

2.1 Distributionally robust second-order stochastic dominance

We first introduce some notations. Let $\mathcal{U}$ be the set of all non-decreasing and concave functions $u : \mathbb{R} \to \mathbb{R}$. We use $(\cdot)_+ = \max\{\cdot, 0\}$ to denote the positive part function. Let $(\Omega, \mathscr{F})$ be a measurable space with $\mathscr{F}$ being the Borel σ-algebra on Ω, and $\mathcal{M}$ be the set of all probability measures on $(\Omega, \mathscr{F})$.

Before introducing the distributionally robust second-order stochastic dominance, we recall the definition of classic second-order stochastic dominance. Consider the integrable random variables X and Y on a probability space $(\Omega, \mathscr{F}, P)$, here $P \in \mathcal{M}$ 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

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

Let $\mathcal{Y}$ be the set of all realizations of the random variable Y. It has been shown in [25, Proposition 1] that (1) is equivalent to

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

In practical applications, it is very difficult to know the full information about the true probability measure P. To address the lack of perfect information of P, Dentcheva and Ruszczyński introduced the distributionally robust second-order stochastic dominance in [9] 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 $\mathcal{Q} \subset \mathcal{M}$, denoted by $X \succeq_{(2)}^{\mathcal{Q}} Y$, if*

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

It is known from (1) that $X \succeq_{(2)}^{\mathcal{Q}} Y$ is equivalent to

$$\mathbb{E}_P[(\eta - X)_+ - (\eta - Y)_+] \leq 0,\ \forall \eta \in \mathbb{R},\ \forall P \in \mathcal{Q}. \tag{3}$$

In the rest of this paper, we investigate the following distributionally robust second-order stochastic dominance constrained optimization problem

$$(P_{SSD}) \qquad \min_{z \in Z} \quad f(z)$$

$$\text{s.t.} \quad z^T\xi \succeq_{(2)}^{\mathcal{Q}} z_0^T\xi,$$

where ξ denotes the random vector, Z is a bounded polyhedral set, and $z_0 \in Z$ is a given benchmark. From (3), problem (P_{SSD}) can be rewritten as

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

We can observe that problem (4) has two levels of semi-infinite constraints, $\eta \in \mathbb{R}$ and $P \in \mathcal{Q}$, induced from the SSD constraints and the distributionally robust ambiguity set, respectively. Moreover, the constraint functions in problem (4) are non-smooth as $(\cdot)_+$ is involved. Therefore, problem (4), 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 is bounded and has a polyhedral structure. The boundedness of Ξ helps us reduce the index set $\mathbb{R}$ into a compact set [25]. The polyhedral structure of Ξ, also assumed in [29, 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 bounded and polyhedral, i.e., $\Xi = \{\xi | C\xi \le d\}$, where $C \in \mathbb{R}^{l\times n}$, $d \in \mathbb{R}^l$.*

Keeping in mind the equivalence of (1) and (2), problem (4) can be formulated as

$$\begin{aligned} \min_{z\in Z} \quad & f(z) \\ \text{s.t.} \quad & \mathbb{E}_P[(\eta - z^T\xi)_+ - (\eta - z_0^T\xi)_+] \le 0, \ \forall \eta \in \mathcal{R}, \ \forall P \in \mathcal{Q}. \end{aligned} \tag{5}$$

where $\mathcal{R} := z_0^T\Xi$. By Assumption 1, $\mathcal{R}$ is a compact set, which is much easier to handle than the whole real line $\mathbb{R}$. We denote the smallest and largest numbers in $\mathcal{R}$ by $\mathcal{R}_{\mathbf{min}}$ and $\mathcal{R}_{\mathbf{max}}$, respectively, that is, $\mathcal{R} = [\mathcal{R}_{\mathbf{min}}, \mathcal{R}_{\mathbf{max}}]$.

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 distributionally robust optimization problems under Wasserstein ambiguity set [13, 29, 38].

To begin with, let $\mathcal{M}(\Xi)$ be the space of all probability measures Q supported on Ξ with $\mathbb{E}_Q[\|\xi\|] < \infty$. Now, we recall the definition of 1-Wasserstein distance.

Definition 2. *The* 1-Wasserstein distance $d\colon \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \to \mathbb{R}_+$ *is defined via*

$$d(Q_1, Q_2) := \inf \left\{ \int_{\Xi^2} \|\xi_1 - \xi_2\| \Pi(d\xi_1, d\xi_2) \colon \begin{array}{c} \Pi \textit{ is a joint distribution of } \xi_1 \textit{ and } \xi_2 \\ \textit{with marginals } Q_1 \textit{ and } Q_2, \textit{ respectively} \end{array} \right\}.$$

Given some observations $\{\widehat{\xi}_i\}_{i=1}^N$ of ξ, we define the data-driven Wasserstein ambiguity set $\mathcal{Q}$ as the set of all distributions close to the empirical distribution $\widehat{P}_N = \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{\xi}_i}$ under 1-Wasseretein metric, that is,

$$\mathcal{Q} := \{P \in \mathcal{M}(\Xi) : d(P, \widehat{P}_N) \le \epsilon\}, \tag{6}$$

where ϵ is a prespecified robust radius. In [29], Esfahani and Kuhn stated that with any prescribed $\beta \in (0, 1)$, by appropriately defining $\epsilon(\beta)$, the true distribution P belongs to $\mathcal{Q}$ with a confidence level $1 - \beta$. Then it is reasonable to consider the worst-case expectation under the ambiguity set $\mathcal{Q}$.

Under some mild conditions, a nice duality result of distributionally robust optimization problems under Wasserstein ambiguity set has been established in [29, Theorem 4.2], [13, Corollary 2] and [38, Proposition 2]. We adopt the version in [13] in the rest of this paper.

Lemma 1 [13]**.** *If Ξ is bounded and $\Psi(\xi)$ is upper semi-continuous, then the optimal values of*

$$\sup_{P\in\mathcal{M}(\Xi)} \left\{ \int_\Xi \Psi(\xi) P(d\xi) : d(P, \widehat{P}_N) \le \epsilon \right\}$$

and

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

are equal. Moreover, the optimal solution of the latter problem can always be obtained.

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.

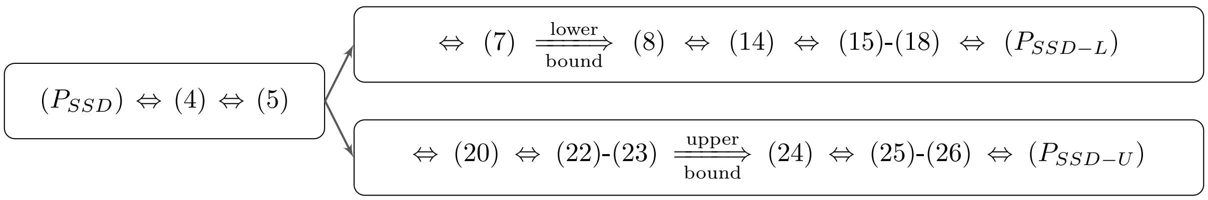

Figure 1: The relationship of formulations in intermediate steps of the two approximation schemes.

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

1) Reformulations $(P_{SSD}) \Leftrightarrow (4)$ and $(4) \Leftrightarrow (5)$ are due to the definition of robust SSD and Assumption 1.

2) Reformulations (5)$\Leftrightarrow$(7) is due to the duality theory of Wasserstein robust optimization from Lemma 1; Approximation (7) $\xrightarrow[\text{bound}]{\text{lower}}$ (8) comes from the finite discrete approximation; Reformulation (8) $\Leftrightarrow$ (14) is a simple rewritting; Reformulations (14) $\Leftrightarrow$ (15)-(18) and (15)-(18) $\Leftrightarrow$ (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, we exchange the order of two supremums equivalently in (5) $\Leftrightarrow$ (20); We split the interval $\mathcal{R}$ into sub-intervals in the reformulation (20) $\Leftrightarrow$ (22)-(23); We exchange the order of the expectation and supremum to derive the upper bound approximation (22)-(23) $\xrightarrow[\text{bound}]{\text{upper}}$ (24); Reformulations (24)$\Leftrightarrow$(25)-(26) is due to the duality theory of Wasserstein robust optimization from Lemma 1; Reformulation (25)-(26) $\Leftrightarrow$ (P_{SSD-U}) is due to the strong duality of second-order cone programming.

3 Lower bound approximation of distributionally robust SSD constrained optimization

We start with the lower bound approximation. Note that problem (5) can be written as

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

whose optimal value is equal to

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{s.t.} \quad & \sup_{\eta \in \mathcal{R}} \min_{\lambda \geq 0} \left\{ \lambda\epsilon - \frac{1}{N} \sum_{i=1}^{N} \inf_{\xi \in \Xi} \left[\lambda \|\xi - \widehat{\xi}_i\| - (\eta - z^T\xi)_+ + (\eta - z_0^T\xi)_+ \right] \right\} \leq 0 \end{aligned} \tag{7}$$

by Lemma 1.

The difficulty in solving problem (7) arises from two aspects: 1) taking infimum over $\xi \in \Xi$ where the function $\lambda\|\xi - \widehat{\xi}_i\| - (\eta - z^T\xi)_+ + (\eta - z_0^T\xi)_+$ is not convex w.r.t. ξ; 2) taking supremum over $\eta \in \mathcal{R}$ where the function $\min_{\lambda \geq 0} \left\{ \lambda\epsilon - \frac{1}{N} \sum_{i=1}^{N} \inf_{\xi \in \Xi} \left[\lambda\|\xi - \widehat{\xi}_i\| - (\eta - z^T\xi)_+ + (\eta - z_0^T\xi)_+ \right] \right\}$ is not convex w.r.t. η, either.

In order to tackle these difficulties, a natural approach is to approximate the sets Ξ and $\mathcal{R}$ by their finite subsets. Then the non-convex min-max problem reduces to a tractable approximation problem with limited enumeration. Let $\Xi_{\mathcal{N}_1} = \{\bar{\xi}_j\}_{j=1}^{\mathcal{N}_1}$ be the set of finite samples in Ξ, and $\Gamma_{\mathcal{N}_2} = \{\eta_k\}_{k=1}^{\mathcal{N}_2}$ be the set of finite samples in $\mathcal{R} = [\mathcal{R}_{\mathbf{min}}, \mathcal{R}_{\mathbf{max}}]$, where $\mathcal{N}_1$ and $\mathcal{N}_2$ denote the sample sizes. We then have an approximation of problem (7)

$$\begin{aligned} \min_{z\in Z} \quad & f(z) && (8)\\ \text{s.t.} \quad & \max_{1\le k\le \mathcal{N}_2} \min_{\lambda\ge 0} \left\{ \lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\le j\le \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T\bar{\xi}_j)_+ + (\eta_k - z_0^T\bar{\xi}_j)_+\right]\right\} \le 0. \end{aligned}$$

In subsection 3.1, we prove that problem (8) forms a lower bound approximation to problem (7). Besides, when the sample sizes $\mathcal{N}_1$ and $\mathcal{N}_2$ go to infinity, problem (8) converges to problem (7) in sense of the feasible solution set, the optimal solution set and the optimal value. In subsection 3.2, we show how problem (8) 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 Asymptotic tightness of the lower bound approximation (7) $\xRightarrow[\textbf{bound}]{\textbf{lower}}$ (8)

First, we have the following proposition.

Proposition 1. *Problem* (8) *is a lower bound approximation of problem* (7)*.*

Proof. Observe that for any $\eta \in \mathcal{R}$ and $\lambda \ge 0$, we have

$$\begin{aligned} &\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \inf_{\xi\in\Xi} \left[\lambda\|\xi - \widehat{\xi}_i\| - (\eta - z^T\xi)_+ + (\eta - z_0^T\xi)_+\right] \\ \ge\ &\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\le j\le \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta - z^T\bar{\xi}_j)_+ + (\eta - z_0^T\bar{\xi}_j)_+\right]. \end{aligned}$$

Taking minimization over $\lambda \ge 0$ on both sides, we obtain

$$\begin{aligned} &\min_{\lambda\ge 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \inf_{\xi\in\Xi} \left[\lambda\|\xi - \widehat{\xi}_i\| - (\eta - z^T\xi)_+ + (\eta - z_0^T\xi)_+\right]\right\} \\ \ge\ &\min_{\lambda\ge 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\le j\le \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta - z^T\bar{\xi}_j)_+ + (\eta - z_0^T\bar{\xi}_j)_+\right]\right\}. \end{aligned}$$

This implies that

$$\begin{aligned} &\sup_{\eta\in\mathcal{R}}\min_{\lambda\ge 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \inf_{\xi\in\Xi} \left[\lambda\|\xi - \widehat{\xi}_i\| - (\eta - z^T\xi)_+ + (\eta - z_0^T\xi)_+\right]\right\} \\ \ge\ &\sup_{\eta\in\mathcal{R}}\min_{\lambda\ge 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\le j\le \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta - z^T\bar{\xi}_j)_+ + (\eta - z_0^T\bar{\xi}_j)_+\right]\right\} \\ \ge\ &\max_{1\le k\le \mathcal{N}_2}\min_{\lambda\ge 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\le j\le \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T\bar{\xi}_j)_+ + (\eta_k - z_0^T\bar{\xi}_j)_+\right]\right\}. \end{aligned}$$

Therefore, the feasible solution set of problem (8) provides an outer approximation to that of problem (7). Since problem (7) and problem (8) are minimization problems, problem (8) is a lower bound approximation of problem (7). □

To establish the theoretical result that problem (8) converges to problem (7), we first need to show that

$$\max_{1\le k\le \mathcal{N}_2}\min_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le \mathcal{N}_1}\left[\lambda\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_k-z^T\bar{\xi}_j)_+ +(\eta_k-z_0^T\bar{\xi}_j)_+\right]\right\}$$

converges to

$$\sup_{\eta\in\mathcal{R}}\min_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\inf_{\xi\in\Xi}\left[\lambda\|\xi-\widehat{\xi}_i\|-(\eta-z^T\xi)_+ +(\eta-z_0^T\xi)_+\right]\right\},$$

when $\mathcal{N}_1$ and $\mathcal{N}_2$ go to infinity. To this end, we need the following assumption which ensures that the subsets $\Xi_{\mathcal{N}_1}$ and $\Gamma_{\mathcal{N}_2}$ cover the sets Ξ and $\mathcal{R}$ with small neighborhoods, and these subsets should be monotonously increasing with respect to $\mathcal{N}_1$ and $\mathcal{N}_2$ in sense of set inclusion.

Assumption 2. *The sample sets $\Xi_{\mathcal{N}_1}$ and $\Gamma_{\mathcal{N}_2}$ satisfy that:*
a) there exist positive radii $\Delta_1=o(\frac{1}{\mathcal{N}_1})$ and $\Delta_2=o(\frac{1}{\mathcal{N}_2})$, such that for each $\xi\in\Xi$, there exists at least one $\widetilde{\xi}\in\Xi_{\mathcal{N}_1}$ in ξ's Δ_1-neighborhood, and for each $\eta\in\mathcal{R}$, there exists at least one $\widetilde{\eta}\in\Gamma_{\mathcal{N}_2}$ in η's Δ_2-neighborhood;
b) if $\mathcal{N}_1^1\le\mathcal{N}_1^2$, then $\Xi_{\mathcal{N}_1^1}\subset\Xi_{\mathcal{N}_1^2}$;
c) if $\mathcal{N}_2^1\le\mathcal{N}_2^2$, then $\Gamma_{\mathcal{N}_2^1}\subset\Gamma_{\mathcal{N}_2^2}$.

Additionally, to guarantee the convergence, the robust radius ϵ should not be too small.

Assumption 3. $\epsilon>\max_{\xi_1,\xi_2\in\Xi}\|\xi_1-\xi_2\|$.

Under Assumption 1, Ξ is bounded. Thus $\max_{\xi_1,\xi_2\in\Xi}\|\xi_1-\xi_2\|$ is finite. Assumption 3 can then be satisfied for some finite positive number ϵ. Therefore, we have the following proposition.

Proposition 2. *Under Assumptions 1, 2 and 3, we have*

$$\begin{aligned}&\max_{\eta\in\mathcal{R}}\min_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\inf_{\xi\in\Xi}\left[\lambda\|\xi-\widehat{\xi}_i\|-(\eta-z^T\xi)_+ +(\eta-z_0^T\xi)_+\right]\right\}\\ =&\lim_{\substack{\mathcal{N}_1\to\infty\\ \mathcal{N}_2\to\infty}}\max_{1\le k\le \mathcal{N}_2}\min_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le \mathcal{N}_1}\left[\lambda\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_k-z^T\bar{\xi}_j)_+ +(\eta_k-z_0^T\bar{\xi}_j)_+\right]\right\}.\end{aligned}\tag{9}$$

Proof. Let η^* be optimal to the left-hand side problem of equation (9). Let $\tilde{k}\in\{1,\cdots,\mathcal{N}_2\}$ be the index such that $\eta_{\tilde{k}}\in\Gamma_{\mathcal{N}_2}$ is in η^*'s Δ_2-neighborhood. Denote

$$\widetilde{\lambda}\in\operatorname*{argmin}_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le \mathcal{N}_1}\left[\lambda\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right]\right\}.\tag{10}$$

Then we have

$$\begin{aligned}0\le\Delta:=&\max_{\eta\in\mathcal{R}}\min_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\inf_{\xi\in\Xi}\left[\lambda\|\xi-\widehat{\xi}_i\|-(\eta-z^T\xi)_+ +(\eta-z_0^T\xi)_+\right]\right\}\\ &-\max_{1\le k\le \mathcal{N}_2}\min_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le \mathcal{N}_1}\left[\lambda\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_k-z^T\bar{\xi}_j)_+ +(\eta_k-z_0^T\bar{\xi}_j)_+\right]\right\}\\ \le&\min_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\inf_{\xi\in\Xi}\left[\lambda\|\xi-\widehat{\xi}_i\|-(\eta^*-z^T\xi)_+ +(\eta^*-z_0^T\xi)_+\right]\right\}\\ &-\min_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le \mathcal{N}_1}\left[\lambda\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right]\right\}\\ \le&\widetilde{\lambda}\epsilon-\frac{1}{N}\sum_{i=1}^{N}\inf_{\xi\in\Xi}\left[\widetilde{\lambda}\|\xi-\widehat{\xi}_i\|-(\eta^*-z^T\xi)_+ +(\eta^*-z_0^T\xi)_+\right]\\ &-\left\{\widetilde{\lambda}\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le \mathcal{N}_1}\left[\widetilde{\lambda}\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right]\right\}\end{aligned}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le \mathcal{N}_1}\left[\widetilde{\lambda}\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right]$$
$$-\frac{1}{N}\sum_{i=1}^{N}\inf_{\xi\in\Xi}\left[\widetilde{\lambda}\|\xi-\widehat{\xi}_i\|-(\eta^*-z^T\xi)_+ +(\eta^*-z_0^T\xi)_+\right]. \tag{11}$$

Then we denote

$$\xi^*\in\operatorname*{argmin}_{\xi\in\Xi}\left[\widetilde{\lambda}\|\xi-\widehat{\xi}_i\|-(\eta^*-z^T\xi)_+ +(\eta^*-z_0^T\xi)_+\right],$$

the infimum can be obtained because Ξ is compact (Assumption 1) and $\widetilde{\lambda}\|\xi-\widehat{\xi}_i\|-(\eta^*-z^T\xi)_+ +(\eta^*-z_0^T\xi)_+$ is continuous with respect to ξ. Let $\tilde{j}\in\{1,\cdots,\mathcal{N}_1\}$ be the index such that $\bar{\xi}_{\tilde{j}}\in\Xi_{\mathcal{N}_1}$ is in ξ^*'s Δ_1-neighborhood. Then by (11)

$$\Delta\le\frac{1}{N}\sum_{i=1}^{N}\left\{\widetilde{\lambda}\|\bar{\xi}_{\tilde{j}}-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_{\tilde{j}})_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_{\tilde{j}})_+ -\widetilde{\lambda}\|\xi^*-\widehat{\xi}_i\|+(\eta^*-z^T\xi^*)_+-(\eta^*-z_0^T\xi^*)_+\right\}$$
$$\le(\widetilde{\lambda}+\|z\|+\|z_0\|)\|\bar{\xi}_{\tilde{j}}-\xi^*\|+2|\eta_{\tilde{k}}-\eta^*|\le(\widetilde{\lambda}+\|z\|+\|z_0\|)\Delta_1+2\Delta_2. \tag{12}$$

Letting $\mathcal{N}_1,\ \mathcal{N}_2\to\infty$, it is then known from Assumption 2 a) that $\Delta_1,\ \Delta_2\to 0$. Furthermore, $\widetilde{\lambda}$ is bounded (see Lemma 2), $z\in Z$ is also bounded, and these bounds are independent of $\mathcal{N}_1$ and $\mathcal{N}_2$. Then $\lim_{\substack{\mathcal{N}_1\to\infty\\ \mathcal{N}_2\to\infty}}\Delta=0$. Therefore, we obtain (9). □

Remark 1. *From our proof, one can notice that the conclusion in Proposition 2 still holds if the condition in Assumption 2 a) is* $\Delta_1=O(\frac{1}{\mathcal{N}_1})$ *or* $\Delta_2=O(\frac{1}{\mathcal{N}_2})$.

In fact, $\widetilde{\lambda}$ defined in (10) depends on $\mathcal{N}_1$ and $\mathcal{N}_2$. It is obvious from (10) that $\widetilde{\lambda}$ depends on $\mathcal{N}_1$. In addition, the choice of $\eta_{\tilde{k}}$ in (10) depends on $\mathcal{N}_2$. We write $\widetilde{\lambda}$ explicitly as a function of $\mathcal{N}_1$ and $\mathcal{N}_2$ as

$$\widetilde{\lambda}(\mathcal{N}_1,\mathcal{N}_2):=\widetilde{\lambda}\in\operatorname*{argmin}_{\lambda\ge 0}\left\{\lambda\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le\mathcal{N}_1}\left[\lambda\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right]\right\}.$$

Lemma 2. *Under Assumptions 1, 2 b) and 3, for any positive integers* $\mathcal{N}_1$ *and* $\mathcal{N}_2$*, we have* $\widetilde{\lambda}(1,1)=\widetilde{\lambda}(1,\mathcal{N}_2)\ge\widetilde{\lambda}(\mathcal{N}_1,\mathcal{N}_2)$.

Proof. Firstly, we show that $\widetilde{\lambda}(\mathcal{N}_1^1,\mathcal{N}_2)\ge\widetilde{\lambda}(\mathcal{N}_1^2,\mathcal{N}_2)$, if $\mathcal{N}_1^1\le\mathcal{N}_1^2$. Assume on the contrary that $\widetilde{\lambda}(\mathcal{N}_1^1,\mathcal{N}_2)<\widetilde{\lambda}(\mathcal{N}_1^2,\mathcal{N}_2)$. From Assumption 2 b) and the optimality of $\widetilde{\lambda}(\mathcal{N}_1^2,\mathcal{N}_2)$ under $\Xi_{\mathcal{N}_1^2}=\{\bar{\xi}_j\}_{j=1}^{\mathcal{N}_1^2}$ and $\Gamma_{\mathcal{N}_2}=\{\eta_k\}_{k=1}^{\mathcal{N}_2}$, we immediately obtain

$$\widetilde{\lambda}(\mathcal{N}_1^2,\mathcal{N}_2)\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le\mathcal{N}_1^2}\left[\widetilde{\lambda}(\mathcal{N}_1^2,\mathcal{N}_2)\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right]$$
$$\le\widetilde{\lambda}(\mathcal{N}_1^1,\mathcal{N}_2)\epsilon-\frac{1}{N}\sum_{i=1}^{N}\min_{1\le j\le\mathcal{N}_1^2}\left[\widetilde{\lambda}(\mathcal{N}_1^1,\mathcal{N}_2)\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right].$$

Then we have

$$\left(\widetilde{\lambda}(\mathcal{N}_1^2,\mathcal{N}_2)-\widetilde{\lambda}(\mathcal{N}_1^1,\mathcal{N}_2)\right)\epsilon$$
$$\le\frac{1}{N}\sum_{i=1}^{N}\left\{\min_{1\le j\le\mathcal{N}_1^2}\left[\widetilde{\lambda}(\mathcal{N}_1^2,\mathcal{N}_2)\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right]\right.$$
$$\left.-\min_{1\le j\le\mathcal{N}_1^2}\left[\widetilde{\lambda}(\mathcal{N}_1^1,\mathcal{N}_2)\|\bar{\xi}_j-\widehat{\xi}_i\|-(\eta_{\tilde{k}}-z^T\bar{\xi}_j)_+ +(\eta_{\tilde{k}}-z_0^T\bar{\xi}_j)_+\right]\right\} \tag{13}$$
$$\le\frac{1}{N}\sum_{i=1}^{N}\max_{1\le j\le\mathcal{N}_1^2}\left[(\widetilde{\lambda}(\mathcal{N}_1^2,\mathcal{N}_2)-\widetilde{\lambda}(\mathcal{N}_1^1,\mathcal{N}_2))\|\bar{\xi}_j-\widehat{\xi}_i\|\right]$$

$$< \left(\widetilde{\lambda}(\mathcal{N}_1^2, \mathcal{N}_2) - \widetilde{\lambda}(\mathcal{N}_1^1, \mathcal{N}_2)\right)\epsilon,$$

where the last inequality is implied by Assumption 3. Then (13) indicates a contradiction. Consequently, $\widetilde{\lambda}(\mathcal{N}_1^1, \mathcal{N}_2) < \widetilde{\lambda}(\mathcal{N}_1^2, \mathcal{N}_2)$ cannot hold. Hence, $\widetilde{\lambda}(1, \mathcal{N}_2) \geq \widetilde{\lambda}(\mathcal{N}_1, \mathcal{N}_2)$ for any positive integers $\mathcal{N}_1$ and $\mathcal{N}_2$.

Now we investigate

$$\widetilde{\lambda}(1, \mathcal{N}_2) \in \operatorname*{argmin}_{\lambda \geq 0} \left\{ \lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \left[\lambda\|\bar{\xi}_1 - \widehat{\xi}_i\| - (\eta_{\tilde{k}} - z^T\bar{\xi}_1)_+ + (\eta_{\tilde{k}} - z_0^T\bar{\xi}_1)_+\right] \right\}.$$

It is easy to observe that the optimal solution $\widetilde{\lambda}(1, \mathcal{N}_2)$ does not change with respect to the choice of $\eta_{\tilde{k}}$. Hence, $\widetilde{\lambda}(1, \mathcal{N}_2) = \widetilde{\lambda}(1, 1)$ for any positive integer $\mathcal{N}_2$. Therefore, $\widetilde{\lambda}(\mathcal{N}_1, \mathcal{N}_2)$ is bounded by $\widetilde{\lambda}(1, 1)$ for any positive integers $\mathcal{N}_1$ and $\mathcal{N}_2$. □

Now, it is time to establish the theoretical result that problem (8) converges to problem (7) when $\mathcal{N}_1$ and $\mathcal{N}_2$ tend to infinity. We denote the feasible solution sets of problem (7) and problem (8) by $\mathcal{F}$ and $\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$, the optimal solution sets by $\mathcal{S}$ and $\mathcal{S}_{\mathcal{N}_1,\mathcal{N}_2}$, and the optimal values by v and $v_{\mathcal{N}_1,\mathcal{N}_2}$, respectively.

Theorem 1. *Given Assumptions 2 and 3, we have* $\mathcal{F} = \lim_{\substack{\mathcal{N}_1\to\infty \\ \mathcal{N}_2\to\infty}} \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$, $\limsup_{\substack{\mathcal{N}_1\to\infty \\ \mathcal{N}_2\to\infty}} \mathcal{S}_{\mathcal{N}_1,\mathcal{N}_2} \subset \mathcal{S}$, $v = \lim_{\substack{\mathcal{N}_1\to\infty \\ \mathcal{N}_2\to\infty}} v_{\mathcal{N}_1,\mathcal{N}_2}$.

Proof. Firstly, we claim that $\mathcal{F}_{\mathcal{N}_1+1,\mathcal{N}_2+1} \subset \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$. To see this, we use the same idea as that is the proof of Proposition 1. From Assumption 2 b), we know that for any $\eta \in \mathcal{R}$ and $\lambda \geq 0$ it holds that

$$\begin{aligned}
&\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\leq j\leq \mathcal{N}_1+1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta - z^T\bar{\xi}_j)_+ + (\eta - z_0^T\bar{\xi}_j)_+\right] \\
\geq\ &\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\leq j\leq \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta - z^T\bar{\xi}_j)_+ + (\eta - z_0^T\bar{\xi}_j)_+\right].
\end{aligned}$$

Taking minimization over $\lambda \geq 0$ on both sides, we obtain

$$\begin{aligned}
&\min_{\lambda\geq 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\leq j\leq \mathcal{N}_1+1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta - z^T\bar{\xi}_j)_+ + (\eta - z_0^T\bar{\xi}_j)_+\right]\right\} \\
\geq\ &\min_{\lambda\geq 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\leq j\leq \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta - z^T\bar{\xi}_j)_+ + (\eta - z_0^T\bar{\xi}_j)_+\right]\right\}.
\end{aligned}$$

Then taking maximization over $k \in \{1, \cdots, \mathcal{N}_2\}$, we have

$$\begin{aligned}
&\max_{1\leq k\leq \mathcal{N}_2+1}\min_{\lambda\geq 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\leq j\leq \mathcal{N}_1+1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T\bar{\xi}_j)_+ + (\eta_k - z_0^T\bar{\xi}_j)_+\right]\right\} \\
\geq\ &\max_{1\leq k\leq \mathcal{N}_2+1}\min_{\lambda\geq 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\leq j\leq \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T\bar{\xi}_j)_+ + (\eta_k - z_0^T\bar{\xi}_j)_+\right]\right\} \\
\geq\ &\max_{1\leq k\leq \mathcal{N}_2}\min_{\lambda\geq 0}\left\{\lambda\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\leq j\leq \mathcal{N}_1} \left[\lambda\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T\bar{\xi}_j)_+ + (\eta_k - z_0^T\bar{\xi}_j)_+\right]\right\},
\end{aligned}$$

where the last inequality is due to Assumption 2 c). Therefore, $\mathcal{F}_{\mathcal{N}_1+1,\mathcal{N}_2+1} \subset \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$.

Next, we show $\mathcal{F} = \lim_{\substack{\mathcal{N}_1\to\infty \\ \mathcal{N}_2\to\infty}} \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$. Since $\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2} \supset \mathcal{F}_{\mathcal{N}_1+1,\mathcal{N}_2+1}$, we have by [32, Exercise 4.3] that

$$\lim_{\substack{\mathcal{N}_1\to\infty \\ \mathcal{N}_2\to\infty}} \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2} = \bigcap_{\substack{\mathcal{N}_1=1 \\ \mathcal{N}_2=1}}^{\infty} \mathrm{cl}\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}.$$

Therefore, to obtain $\mathcal{F} = \lim_{\substack{\mathcal{N}_1\to\infty\\ \mathcal{N}_2\to\infty}} \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$, it is sufficient to prove that

$$\mathcal{F} = \bigcap_{\substack{\mathcal{N}_1=1\\ \mathcal{N}_2=1}}^{\infty} \mathrm{cl}\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}.$$

On one hand, $\mathcal{F} \subset \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}, \forall \mathcal{N}_1, \mathcal{N}_2$, which obviously leads to $\mathcal{F} \subset \cap_{\mathcal{N}_1=1,\mathcal{N}_2=1}^{\infty} \mathrm{cl}\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$. On the other hand, for any $z \in \cap_{\mathcal{N}_1=1,\mathcal{N}_2=1}^{\infty} \mathrm{cl}\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$, we must have $z \in \mathrm{cl}\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}, \forall \mathcal{N}_1, \mathcal{N}_2$. Thus $z \in \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}, \forall \mathcal{N}_1, \mathcal{N}_2$, since $\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$ is closed. This means that z satisfies the constraint in problem (8). Taking $\mathcal{N}_1 \to \infty, \mathcal{N}_2 \to \infty$, then by Proposition 2, z satisfies the constraint in problem (7). That is, $z \in \mathcal{F}$.

Finally, we verify that $v = \lim_{\substack{\mathcal{N}_1\to\infty\\ \mathcal{N}_2\to\infty}} v_{\mathcal{N}_1,\mathcal{N}_2}$ and $\limsup_{\substack{\mathcal{N}_1\to\infty\\ \mathcal{N}_2\to\infty}} \mathcal{S}_{\mathcal{N}_1,\mathcal{N}_2} \subset \mathcal{S}$. Let $\bar{f}(z) = f(z) + \delta_{\mathcal{F}}(z)$ and $\bar{f}_{\mathcal{N}_1,\mathcal{N}_2}(z) = f(z) + \delta_{\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}}(z)$, where $\delta_A(z) = 0$ if $z \in A$, otherwise $\delta_A(z) = +\infty$. Since $\mathcal{F} = \lim_{\substack{\mathcal{N}_1\to\infty\\ \mathcal{N}_2\to\infty}} \mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$, then by [32, Proposition 7.4(f)], $\delta_{\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}}$ epi-converges to $\delta_{\mathcal{F}}$ as $\mathcal{N}_1 \to \infty, \mathcal{N}_2 \to \infty$. As f is continuous and finite, we obtain by [32, Exercise 7.8] that $\bar{f}_{\mathcal{N}_1,\mathcal{N}_2} = f + \delta_{\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}}$ epi-converges to $\bar{f} = f + \delta_{\mathcal{F}}$ when $\mathcal{N}_1 \to \infty, \mathcal{N}_2 \to \infty$. As $\mathcal{F}$, $\mathcal{F}_{\mathcal{N}_1,\mathcal{N}_2}$ are closed and f is continuous, $\bar{f}_{\mathcal{N}_1,\mathcal{N}_2}$ and $\bar{f}$ are lower semi-continuous. Moreover, since $\bar{f}_{\mathcal{N}_1,\mathcal{N}_2}$ and $\bar{f}$ are proper, it can then be deduced from [32, Theorem 7.33] that $v = \lim_{\substack{\mathcal{N}_1\to\infty\\ \mathcal{N}_2\to\infty}} v_{\mathcal{N}_1,\mathcal{N}_2}$ and $\limsup_{\substack{\mathcal{N}_1\to\infty\\ \mathcal{N}_2\to\infty}} \mathcal{S}_{\mathcal{N}_1,\mathcal{N}_2} \subset \mathcal{S}$. □

Theorem 1 states that the lower bound approximation problem (8) is asymptotically tight. Specifically, the optimal value and the feasible solution set of problem (8) converge to those of problem (7), and the outer limit of the optimal solution set of problem (8) is included in the optimal solution set of problem (7).

For the case that the support set Ξ is finite, the lower bound approximation is tight.

Proposition 3. *When $\Xi_{\mathcal{N}_1} = \Xi$ and $\Gamma_{\mathcal{N}_2} = \{z_0^T\xi |\ \xi \in \Xi\}$, the optimal values of problem* (7) *and problem* (8) *are equal.*

Proof. The conclusion follows from [7, Proposition 3.2]. □

3.2 Tractability of the lower bound approximation problem (8)

In what follows, we derive a tractable equivalent formulation of problem (8).

Notice that problem (8) can be rewritten as

$$\begin{aligned}
\min_{z,\lambda}\quad & f(z) \\
\text{s.t.}\quad & \lambda_k\epsilon - \frac{1}{N}\sum_{i=1}^{N} \min_{1\le j\le \mathcal{N}_1} \left[\lambda_k\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T\bar{\xi}_j)_+ + (\eta_k - z_0^T\bar{\xi}_j)_+\right] \le 0,\ k = 1,\cdots,\mathcal{N}_2, \qquad (14)\\
& z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{N}_2}.
\end{aligned}$$

By introducing auxiliary variables $\beta_{ik},\ i = 1,\cdots,N, k = 1,\cdots,\mathcal{N}_2$ (which can be written for simplicity as a matrix $\beta \in \mathbb{R}^{N\times\mathcal{N}_2}$), problem (14) can be reformulated as

$$\min_{z,\lambda,\beta}\quad f(z) \tag{15}$$

$$\text{s.t.}\quad \lambda_k\epsilon - \frac{1}{N}\sum_{i=1}^{N}\beta_{ik} \le 0,\ k = 1,\cdots,\mathcal{N}_2, \tag{16}$$

$$\beta_{ik} \le \lambda_k\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T\bar{\xi}_j)_+ + (\eta_k - z_0^T\bar{\xi}_j)_+,\ i = 1,\cdots,N,\ j = 1,\cdots,\mathcal{N}_1,\ k = 1,\cdots,\mathcal{N}_2, \tag{17}$$

$$z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{N}_2}, \beta \in \mathbb{R}^{N\times\mathcal{N}_2}. \tag{18}$$

In fact, the feasible solution sets of problem (14) and problem (15)-(18) are equivalent in the following sense. On one hand, given any feasible solution (z,λ) of problem (14), let

$$\beta_{ik} = \min_{1\le j\le\mathcal{N}_1}\left[\lambda_k\|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T\bar{\xi}_j)_+ + (\eta_k - z_0^T\bar{\xi}_j)_+\right],\ i = 1,\cdots,N,\ k = 1,\cdots,\mathcal{N}_2.$$

Then (z,λ,β) is feasible for problem (15)-(18). On the other hand, given any feasible solution (z,λ,β) of problem (15)-(18), we can verify that for any $k = 1, \cdots, \mathcal{N}_2$, it holds that

$$\lambda_k \epsilon \leq \frac{1}{N} \sum_{i=1}^N \beta_{ik} \leq \frac{1}{N} \sum_{i=1}^N \min_{1 \leq j \leq \mathcal{N}_1} \left[\lambda_k \|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z^T \bar{\xi}_j)_+ + (\eta_k - z_0^T \bar{\xi}_j)_+ \right],$$

where the inequalities are from (16) and (17). Therefore, (z, λ) is also feasible for problem (14).

Further introducing auxiliary variables $s_{jk},\ j = 1, \cdots, \mathcal{N}_1, k = 1, \cdots, \mathcal{N}_2$, to handle $(\eta_k - z^T \bar{\xi}_j)_+$ (refer to [7] (3.10)-(3.12)), we have a reformulation of problem (15)-(18) as follows

$$(P_{SSD-L}) \quad \begin{aligned} \min_{z,\lambda,\beta,s} \quad & f(z) \\ \text{s.t.} \quad & \lambda_k \epsilon - \sum_{i=1}^N \frac{1}{N} \beta_{ik} \leq 0,\ k = 1, \cdots, \mathcal{N}_2, \\ & \beta_{ik} + s_{jk} \leq \lambda_k \|\bar{\xi}_j - \widehat{\xi}_i\| + (\eta_k - z_0^T \bar{\xi}_j)_+,\ i = 1, \cdots, N,\ j = 1, \cdots, \mathcal{N}_1,\ k = 1, \cdots, \mathcal{N}_2, \\ & s_{jk} \geq \eta_k - z^T \bar{\xi}_j,\ j = 1, \cdots, \mathcal{N}_1,\ k = 1, \cdots, \mathcal{N}_2, \\ & z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{N}_2}, \beta \in \mathbb{R}^{N \times \mathcal{N}_2}, s \in \mathbb{R}_+^{\mathcal{N}_1 \times \mathcal{N}_2}. \end{aligned}$$

Problem (P_{SSD-L}) is equivalent to problem (8), and thus is a lower bound approximation of problem (7) and problem (P_{SSD}). As a linear programming problem, (P_{SSD-L}) can be solved directly using many off-the-shelf optimization softwares. However, if $\mathcal{N}_1$ and $\mathcal{N}_2$ are large, solving problem (P_{SSD-L}) may be time-consuming. In fact, the dimension of s is $\mathcal{N}_1 \times \mathcal{N}_2$ and the number of constraints in problem (P_{SSD-L}) is $\mathcal{N}_2 + N \times \mathcal{N}_1 \times \mathcal{N}_2 + \mathcal{N}_1 \times \mathcal{N}_2$. They increase rapidly with the increase of the sample sizes $\mathcal{N}_1$ and $\mathcal{N}_2$.

In order to numerically solve problem (15)-(18) for large $\mathcal{N}_1, \mathcal{N}_2$, we propose a cutting-plane method, see Algorithm 1. At each iteration of the cutting-plane method, we solve problem (19), a relaxation of problem (P_{SSD-L}). Therefore, the approximate problem (19) provides a lower bound approximation for problem (15)-(18) at each iteration. After solving problem (19), we check whether all the constraints in (17) are satisfied or not. If all the constraints in (17) hold, then the optimal solution we find for problem (19) is also optimal for problem (15)-(18). Otherwise, if constraint (17) is violated for some index (j^l, k^l), then we add the violated constraint to the approximate problem (19) at the next iteration.

Theorem 2. *Algorithm 1 stops at the optimal value and optimal solution of problem* (15)-(18) *within finite steps.*

Proof. We claim that $\mathcal{J}_1^l \subsetneqq \mathcal{J}_1^{l+1}$ or $\mathcal{J}_2^l \subsetneqq \mathcal{J}_2^{l+1}$. In fact, since $(z^l, \lambda^l, \beta^l, s^l)$ is an optimal solution of problem (19), we immediately have that

$$\beta_{ik}^l + s_{jk}^l - \lambda_k^l \|\bar{\xi}_j - \widehat{\xi}_i\| - (\eta_k - z_0^T \bar{\xi}_j)_+ \leq 0, i = 1, \cdots, N, j \in \mathcal{J}_1^l, k \in \mathcal{J}_2^l,$$
$$s_{jk}^l \geq \eta_k - (z^l)^T \bar{\xi}_j,\ s_{jk}^l \geq 0, j \in \mathcal{J}_1^l, k \in \mathcal{J}_2^l.$$

This implies that

$$\begin{aligned} & \beta_{ik}^l - \lambda_k^l \|\bar{\xi}_j - \widehat{\xi}_i\| + (\eta_k - (z^l)^T \bar{\xi}_j)_+ - (\eta_k - z_0^T \bar{\xi}_j)_+ \\ \leq\ & \beta_{ik}^l - \lambda_k^l \|\bar{\xi}_j - \widehat{\xi}_i\| + s_{jk}^l - (\eta_k - z_0^T \bar{\xi}_j)_+ \leq 0, i = 1, \cdots, N, j \in \mathcal{J}_1^l, k \in \mathcal{J}_2^l. \end{aligned}$$

On the other hand, (i^l, j^l, k^l) is chosen such that for $(i, j, k) = (i^l, j^l, k^l)$, it holds that

$$\beta_{ik}^l - \lambda_k^l \|\bar{\xi}_j - \widehat{\xi}_i\| + (\eta_k - (z^l)^T \bar{\xi}_j)_+ - (\eta_k - z_0^T \bar{\xi}_j)_+ > 0.$$

Therefore, $j^l \notin \mathcal{J}_1^l$ or $k^l \notin \mathcal{J}_2^l$. But $j^l \in \mathcal{J}_1^{l+1}$ and $k^l \in \mathcal{J}_2^{l+1}$. This tells us that $\mathcal{J}_1^l \subsetneqq \mathcal{J}_1^{l+1}$ or $\mathcal{J}_2^l \subsetneqq \mathcal{J}_2^{l+1}$. 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 (15)-(18) within finite steps. □

To conclude this section, we develop an asymptotically tight lower bound approximation problem (8) for the distributionally robust second-order stochastic dominance constrained optimization problem (P_{SSD}). Problem (8) can be reformulated as problem (15)-(18), which can be easily solved using linear programming formulation (P_{SSD-L}) or by Algorithm 1.

Algorithm 1 Cutting-plane Method

Start from $l = 1$ and $\mathcal{J}_1^1 = \mathcal{J}_2^1 = \emptyset$.
while $l \geq 1$ **do**
Solve the approximate problem:

$$\begin{aligned} \min_{z,\lambda,\beta,s} \quad & f(z) \\ \text{s.t.} \quad & \lambda_k \epsilon - \sum_{i=1}^{N} \frac{1}{N} \beta_{ik} \leq 0, k = 1, \cdots, \mathcal{N}_2, \\ & \beta_{ik} + s_{jk} \leq \lambda_k \|\bar{\xi}_j - \widehat{\xi}_i\| + (\eta_k - z_0^T \bar{\xi}_j)_+, i = 1, \cdots, N, j \in \mathcal{J}_1^l, k \in \mathcal{J}_2^l, \\ & s_{jk} \geq \eta_k - z^T \bar{\xi}_j, j \in \mathcal{J}_1^l, k \in \mathcal{J}_2^l, \\ & z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{N}_2}, \beta \in \mathbb{R}^{N \times \mathcal{N}_2}, s \in \mathbb{R}_+^{\mathcal{N}_1 \times \mathcal{N}_2}. \end{aligned} \tag{19}$$

Let $(z^l, \lambda^l, \beta^l, s^l)$ denote the optimal solution of problem (19).
Calculate

$$\delta^l := \max_{i \in \{1,\cdots,N\}, j \in \{1,\cdots,\mathcal{N}_1\}, k \in \{1,\cdots,\mathcal{N}_2\}} \left\{ \beta_{ik}^l - \lambda_k^l \|\bar{\xi}_j - \widehat{\xi}_i\| + (\eta_k - (z^l)^T \bar{\xi}_j)_+ - (\eta_k - z_0^T \bar{\xi}_j)_+ \right\}.$$

if $\delta^l \leq 0$ **then**
Stop.
else
Determine

$$(i^l, j^l, k^l) \in \operatorname*{argmax}_{i \in \{1,\cdots,N\}, j \in \{1,\cdots,\mathcal{N}_1\}, k \in \{1,\cdots,\mathcal{N}_2\}} \left\{ \beta_{ik}^l - \lambda_k^l \|\bar{\xi}_j - \widehat{\xi}_i\| + (\eta_k - (z^l)^T \bar{\xi}_j)_+ - (\eta_k - z_0^T \bar{\xi}_j)_+ \right\}.$$

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

4 Upper bound approximation of distributionally robust SSD constrained problem

We derive an upper bound approximation for problem (5) in this section.

Notice that problem (5) can be rewritten as

$$\begin{aligned} \min_{z \in Z} \quad & f(z) \\ \text{s.t.} \quad & \sup_{P \in \mathcal{Q}} \sup_{\eta \in \mathcal{R}} \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \leq 0. \end{aligned} \tag{20}$$

If we exchange the order of operators $\sup_{\eta \in \mathcal{R}}$ and $\mathbb{E}_P$ in problem (20), we obtain an upper bound approximation for problem (20). However, such an upper bound approximation might be loose since for each $P \in \mathcal{Q}$, the gap

$$\mathbb{E}_P \left\{ \sup_{\eta \in \mathcal{R}} [(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \right\} - \sup_{\eta \in \mathcal{R}} \mathbb{E}_P[(\eta - z^T \xi)_+ - (\eta - z_0^T \xi)_+] \tag{21}$$

might be large. This is because we determine an η for all possible ξ's in the latter supremum in (21), while we determine an η for each realization of ξ in the former supremum in (21). Therefore, the larger the range $\mathcal{R}$ of η, the larger the gap in (21). As an extreme case, when $\mathcal{R}$ reduces to a singleton set, the gap in (21) 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 taking supremum over the sub-interval. Summing all sub-problems in all sub-intervals, we would obtain a better upper bound approximation of problem (20). We name such a bounding method a split-and-dual technique.

In detail, we divide $\mathcal{R} = [\mathcal{R}_{\mathbf{min}}, \mathcal{R}_{\mathbf{max}}]$ into $\mathcal{K}$ intervals with disjoint interiors, $[\underline{\eta}_k, \bar{\eta}_k]$, $k = 1, \cdots, \mathcal{K}$, where the boundary points of the intervals are specified by $\underline{\eta}_k = \mathcal{R}_{\mathbf{min}} + (k-1)\frac{\mathcal{R}_{\mathbf{max}} - \mathcal{R}_{\mathbf{min}}}{\mathcal{K}}$, $\bar{\eta}_k = \mathcal{R}_{\mathbf{min}} + k\frac{\mathcal{R}_{\mathbf{max}} - \mathcal{R}_{\mathbf{min}}}{\mathcal{K}}$,

$k = 1, \cdots, \mathcal{K}$. Notice that problem (20) can also be reformulated as

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

or, equivalently,

$$\min_{z\in Z} \ f(z) \tag{22}$$

$$\text{s.t.} \ \sup_{P\in\mathcal{Q}} \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} \mathbb{E}_P[(\eta - z^T\xi)_+ - (\eta - z_0^T\xi)_+] \le 0, \ k = 1, \cdots, \mathcal{K}. \tag{23}$$

Exchanging the order of operators $\sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]}$ and $\mathbb{E}_P$ in (23), we have the following approximation problem

$$\begin{aligned} \min_{z\in Z} \ & f(z) \\ \text{s.t.} \ & \sup_{P\in\mathcal{Q}} \mathbb{E}_P\Big\{ \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} [(\eta - z^T\xi)_+ - (\eta - z_0^T\xi)_+]\Big\} \le 0, \ k = 1, \cdots, \mathcal{K}. \end{aligned} \tag{24}$$

The feasible solution set of problem (22)-(23) contains the feasible solution set of (24). Thus problem (24) provides an upper bound approximation for problem (22)-(23).

By applying Lemma 1 to each supremum problem with respect to P for $k = 1, \ldots, \mathcal{K}$, we have an equivalent formulation of problem (24)

$$\min_{z\in Z,\lambda\in\mathbb{R}_+^{\mathcal{K}}} \ f(z) \tag{25}$$

$$\text{s.t.} \ \lambda_k\epsilon + \frac{1}{N}\sum_{i=1}^{N} \sup_{\xi\in\Xi}\Big\{ \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} [(\eta - z^T\xi)_+ - (\eta - z_0^T\xi)_+] - \lambda_k\|\xi - \widehat{\xi}_i\|\Big\} \le 0, \ k = 1, \cdots, \mathcal{K}. \tag{26}$$

To simplify the notation, we write (26) as

$$\lambda_k\epsilon + \frac{1}{N}\sum_{i=1}^{N} V_S^{ik} \le 0, \ k = 1, \cdots, \mathcal{K}, \tag{27}$$

where

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

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

$$\begin{aligned} \sup_{\xi,\eta,s,m} \quad & (\eta - z^T\xi)_+ - s - \lambda_k m \\ \text{s.t.} \quad & s \ge \eta - z_0^T\xi, \ C\xi \le d, \\ & \eta \ge \underline{\eta}_k, \eta \le \bar{\eta}_k, \ \|\xi - \widehat{\xi}_i\| \le m, \\ & \xi \in \mathbb{R}^n, \eta \in \mathbb{R}, s \ge 0, m \in \mathbb{R}. \end{aligned} \tag{28}$$

Problem (28) 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 (28) into two convex optimization problems:

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

$$
\begin{aligned}
& s \geq 0, && s \geq 0, \\
& \eta \geq \underline{\eta}_k, && \eta \geq \underline{\eta}_k, \\
& \eta \leq \bar{\eta}_k, && \eta \leq \bar{\eta}_k, \\
& \|\xi - \widehat{\xi}_i\| \leq m. && \|\xi - \widehat{\xi}_i\| \leq m.
\end{aligned}
$$

And we have that

$$V_S^{ik} = \max\{V_{S1}^{ik}, V_{S2}^{ik}\}. \tag{29}$$

First, we derive the dual problem of problem (P_{SSD-1}^{ik}). Using conic duality theory, we introduce the dual variables $(\mu_1, \mu_2, \nu, \mu_3, \mu_4, \mu_5, \delta, \kappa)$ such that $\mu_1 \geq 0,\ \mu_2 \geq 0,\ \nu \geq \mathbf{0},\ \mu_3 \geq 0,\ \mu_4 \geq 0,\ \mu_5 \geq 0$ and $\kappa \geq \|\delta\|$ for the seven constraints, respectively, and we obtain the standard formulation of the dual problem of (P_{SSD-1}^{ik}) as follows:

$$
\begin{aligned}
\inf_{\mu,\nu,\kappa,\delta} \quad & d^T\nu - \widehat{\xi}_i^T\delta - \mu_4\underline{\eta}_k + \mu_5\bar{\eta}_k \\
\text{s.t.} \quad & -z + \mu_1 z_0 - \mu_2 z - C^T\nu + \delta = 0, && (30) \\
& 1 - \mu_1 + \mu_2 + \mu_4 - \mu_5 = 0, \\
& -1 + \mu_1 + \mu_3 = 0, && (31) \\
& -\lambda_k + \kappa = 0, && (32) \\
& \kappa \geq \|\delta\|, \\
& \mu \in \mathbb{R}_+^5,\ \nu \in \mathbb{R}_+^l,\ \kappa \in \mathbb{R},\ \delta \in \mathbb{R}^n.
\end{aligned}
$$

Eliminating δ, μ_3, μ_5 and κ using (30), (31) and (32), we can reformulate the dual problem as

$$
(D_{SSD-1}^{ik}) \qquad
\begin{aligned}
\tilde{V}_{S1}^{ik} = \inf_{\mu,\nu} \quad & d^T\nu - \widehat{\xi}_i^T(z - \mu_1 z_0 + \mu_2 z + C^T\nu) - \mu_4\underline{\eta}_k + (1 - \mu_1 + \mu_2 + \mu_4)\bar{\eta}_k \\
\text{s.t.} \quad & 1 - \mu_1 + \mu_2 + \mu_4 \geq 0, \\
& \|z - \mu_1 z_0 + \mu_2 z + C^T\nu\| \leq \lambda_k, \\
& 1 \geq \mu_1 \geq 0,\ \mu_2 \geq 0,\ \mu_4 \geq 0,\ \nu \in \mathbb{R}_+^l.
\end{aligned}
$$

Similarly, the dual problem of problem (P_{SSD-2}^{i}) is

$$
\begin{aligned}
\inf_{\mu,\nu,\kappa,\delta} \quad & d^T\nu - \widehat{\xi}_i^T\delta - \mu_4\underline{\eta}_k + \mu_5\bar{\eta}_k \\
\text{s.t.} \quad & \mu_1 z_0 + \mu_2 z - C^T\nu + \delta = 0, \\
& -\mu_1 - \mu_2 + \mu_4 - \mu_5 = 0, \\
& -1 + \mu_1 + \mu_3 = 0, \\
& -\lambda_k + \kappa = 0, \\
& \kappa \geq \|\delta\|, \\
& \mu \in \mathbb{R}_+^5,\ \nu \in \mathbb{R}_+^l,\ \kappa \in \mathbb{R},\ \delta \in \mathbb{R}^n,
\end{aligned}
$$

or, equivalently,

$$
(D_{SSD-2}^{ik}) \qquad
\begin{aligned}
\tilde{V}_{S2}^{ik} = \inf_{\mu,\nu} \quad & d^T\nu - \widehat{\xi}_i^T(-\mu_1 z_0 - \mu_2 z + C^T\nu) - \mu_4\underline{\eta}_k + (-\mu_1 - \mu_2 + \mu_4)\bar{\eta}_k \\
\text{s.t.} \quad & -\mu_1 - \mu_2 + \mu_4 \geq 0, \\
& \|-\mu_1 z_0 - \mu_2 z + C^T\nu\| \leq \lambda_k, \\
& 1 \geq \mu_1 \geq 0,\ \mu_2 \geq 0,\ \mu_4 \geq 0,\ \nu \in \mathbb{R}_+^l.
\end{aligned}
$$

We observe that the infimum problems (D_{SSD-1}^{ik}) and (D_{SSD-2}^{ik}) can be reached by the corresponding minimization problems over the closed feasible solution sets, given that the optimal values are finite. By equation (29) and the duality theory, we have that $V_S^{ik} \leq \max\{\tilde{V}_{S1}^{ik}, \tilde{V}_{S2}^{ik}\}$, $i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}$.

Assumption 4. *For any $i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}$, problems (P_{SSD-1}^{ik}) and (P_{SSD-2}^{ik}) are strictly feasible.*

Under Assumption 4, the strong duality condition holds. Thus, the duality gap between V_{S1}^{ik} (resp. V_{S2}^{ik}) and $\tilde{V}_{S1}^{ik}$ (resp. $\tilde{V}_{S2}^{ik}$) is zero, and $V_S^{ik} = \max\{\tilde{V}_{S1}^{ik}, \tilde{V}_{S2}^{ik}\}$, $i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}$.

Introducing auxiliary variables $V^{ik},\ i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{N}_2$, we claim that constraints

$$\lambda_k \epsilon + \frac{1}{N}\sum_{i=1}^{N} V^{ik} \le 0, k = 1, \cdots, \mathcal{K} \tag{33}$$

$$V^{ik} \ge \tilde{V}_{S1}^{ik},\ i = 1, \cdots, N, k = 1, \cdots, \mathcal{K}, \tag{34}$$

$$V^{ik} \ge \tilde{V}_{S2}^{ik},\ i = 1, \cdots, N, k = 1, \cdots, \mathcal{K}. \tag{35}$$

are equivalent to the constraints in (27). To prove the assertion, first, if there exist $V^{ik},\ i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}$, such that constraints (33)-(35) hold, then

$$\lambda_k \epsilon + \frac{1}{N}\sum_{i=1}^{N} V_S^{ik} \le \lambda_k \epsilon + \frac{1}{N}\sum_{i=1}^{N} \max\{\tilde{V}_{S1}^{ik}, \tilde{V}_{S2}^{ik}\} \le \lambda_k \epsilon + \frac{1}{N}\sum_{i=1}^{N} V^{ik} \le 0,\ k = 1, \cdots, \mathcal{K},$$

and thus, the constraints in (27) hold. On the other hand, if $V_S^{ik},\ i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}$ satisfy the constraints in (27), let $V^{ik} = \max\{\tilde{V}_{S1}^{ik}, \tilde{V}_{S2}^{ik}\},\ i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}$. Then by strong duality condition between V_{S1}^{ik} (resp. V_{S2}^{ik}) and $\tilde{V}_{S1}^{ik}$ (resp. $\tilde{V}_{S2}^{ik}$), constraints (33)-(35) hold.

Taking the formulations (D_{SSD-1}^{ik}) and (D_{SSD-2}^{ik}) of $\tilde{V}_{S1}^{ik}, \tilde{V}_{S2}^{ik}$ into constraints (34)-(35) provides an upper bound approximation for problem (P_{SSD}).

Theorem 3. *Given Assumption 4, the optimal value of the following optimization problem*

$$
(P_{SSD-U}) \quad
\begin{aligned}
\min \quad & f(z) \\
\text{s.t.} \quad & \lambda_k \epsilon + \frac{1}{N}\sum_{i=1}^{N} V^{ik} \le 0,\ k = 1, \cdots, \mathcal{K}, \\
& \mu_1^{ik} \le 1,\ \tilde{\mu}_1^{ik} \le 1,\ 1 - \mu_1^{ik} + \mu_2^{ik} + \mu_3^{ik} \ge 0,\ -\tilde{\mu}_1^{ik} - \tilde{\mu}_2^{ik} + \tilde{\mu}_3^{ik} \ge 0, \\
& \qquad i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}, \\
& V^{ik} \ge d^T \nu^{ik} - \hat{\xi}_i^T (z - \mu_1^{ik} z_0 + \mu_2^{ik} z + C^T \nu^{ik}) - \mu_3^{ik} \underline{\eta}_k + (1 - \mu_1^{ik} + \mu_2^{ik} + \mu_3^{ik}) \bar{\eta}_k, \\
& \qquad i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}, \\
& \|z - \mu_1^{ik} z_0 + \mu_2^{ik} z + C^T \nu^{ik}\| \le \lambda_k,\ \| - \tilde{\mu}_1^{ik} z_0 - \tilde{\mu}_2^{ik} z + C^T \tilde{\nu}^{ik}\| \le \lambda_k, \\
& \qquad i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}, \\
& V^{ik} \ge d^T \tilde{\nu}^{ik} - \hat{\xi}_i^T (-\tilde{\mu}_1^{ik} z_0 - \tilde{\mu}_2^{ik} z + C^T \tilde{\nu}^{ik}) - \tilde{\mu}_3^{ik} \underline{\eta}_k + (-\tilde{\mu}_1^{ik} - \tilde{\mu}_2^{ik} + \tilde{\mu}_3^{ik}) \bar{\eta}_k, \\
& \qquad i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}, \\
& z \in Z, \lambda \in \mathbb{R}_+^{\mathcal{K}}, \mu^{ik} \in \mathbb{R}_+^3, \nu^{ik} \in \mathbb{R}_+^l, \tilde{\mu}^{ik} \in \mathbb{R}_+^3, \tilde{\nu}^{ik} \in \mathbb{R}_+^l, V^{ik} \in \mathbb{R}, \\
& \qquad i = 1, \cdots, N,\ k = 1, \cdots, \mathcal{K}.
\end{aligned}
$$

is an upper bound to that of problem (P_{SSD}).

Proof. From what we have demonstrated, problem (P_{SSD-U}) provides a reformulation for problem (24). Moreover, we have shown that problem (24) is an upper bound approximation for problem (22)-(23). The latter problem is a reformulation of problem (P_{SSD}), as illustrated in Figure 1. Therefore, problem (P_{SSD-U}) provides an upper bound approximation for problem (P_{SSD}). □

4.1 Asymptotic tightness of upper bound approximation (22)-(23) $\xRightarrow[\textbf{bound}]{\textbf{upper}}$ (24)

In what follows, we show that when the interval number $\mathcal{K}$ goes to infinity, the optimal value of problem (P_{SSD-U}) converges to that of problem (22)-(23). Since problem (P_{SSD-U}) is a reformulation of problem (24), as illustrated in Figure 1, we next prove that the optimal value of problem (24) converges to that of problem

(22)-(23). To this end, we first prove the asymptotic convergence of

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

to

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

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

We notice that $g(\cdot,\mathcal{K})$ and $g(\cdot)$ are Lipschitz continuous with respect to z.

Proposition 4. *Under Assumption 1, $g(\cdot,\mathcal{K})$ and $g(\cdot)$ are Lipschitz continuous with a concretely Lipschitz constant $\mathcal{C} = \max_{\xi\in\Xi}\|\xi\| < \infty$.*

Proof. We first prove the conclusion for $g(\cdot,\mathcal{K})$. We have from the Lipschitz continuity of the positive part function $(\cdot)_+$ that

$$\begin{aligned}
&|g(z_1,\mathcal{K}) - g(z_2,\mathcal{K})| \\
=&\Bigg|\max_{1\le k\le \mathcal{K}} \sup_{P\in\mathcal{Q}} \mathbb{E}_P\Big\{ \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} \big[(\eta - z_1^T\xi)_+ - (\eta - z_0^T\xi)_+\big] \Big\} \\
&- \max_{1\le k\le \mathcal{K}} \sup_{P\in\mathcal{Q}} \mathbb{E}_P\Big\{ \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} \big[(\eta - z_2^T\xi)_+ - (\eta - z_0^T\xi)_+\big] \Big\}\Bigg| \\
\le& \max_{1\le k\le \mathcal{K}} \sup_{P\in\mathcal{Q}} \Bigg|\mathbb{E}_P\Big\{ \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} \big[(\eta - z_1^T\xi)_+ - (\eta - z_0^T\xi)_+\big] - \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} \big[(\eta - z_2^T\xi)_+ - (\eta - z_0^T\xi)_+\big] \Big\}\Bigg| \\
\le& \max_{1\le k\le \mathcal{K}} \sup_{P\in\mathcal{Q}} \mathbb{E}_P\Big\{ \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} \Big| \big[(\eta - z_1^T\xi)_+ - (\eta - z_0^T\xi)_+\big] - \big[(\eta - z_2^T\xi)_+ - (\eta - z_0^T\xi)_+\big] \Big|\Big\} \\
\le& \max_{1\le k\le \mathcal{K}} \sup_{P\in\mathcal{Q}} \mathbb{E}_P\Big\{ \sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]} \Big|z_1^T\xi - z_2^T\xi\Big|\Big\} \\
\le& \max_{\xi\in\Xi}\|\xi\|\cdot\|z_1 - z_2\|.
\end{aligned}$$

The proof of the conclusion for $g(\cdot)$ is quite similar and thus is omitted. □

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

Proposition 5. *We have that*

$$\lim_{\mathcal{K}\to\infty} g(z,\mathcal{K}) = g(z),$$

and the convergence is uniform with respect to any $z \in Z$.

Proof. Denote

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

Denote

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

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

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

$$
\begin{aligned}
&= \max_{1\le k\le \mathcal{K}} \sup_{P\in\mathcal{Q}} \left\{ \mathbb{E}_P\left[(\eta_k^* - z^T\xi)_+ - (\eta_k^* - z_0^T\xi)_+\right] - \mathbb{E}_P\left[(\eta_k^{**} - z^T\xi)_+ - (\eta_k^{**} - z_0^T\xi)_+\right] \right\} \\
&\le \max_{1\le k\le \mathcal{K}} \sup_{P\in\mathcal{Q}} \mathbb{E}_P \Big| \left[(\eta_k^* - z^T\xi)_+ - (\eta_k^* - z_0^T\xi)_+\right] - \left[(\eta_k^{**} - z^T\xi)_+ - (\eta_k^{**} - z_0^T\xi)_+\right] \Big| \\
&\le 2\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}},
\end{aligned} \tag{36}
$$

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

4.2 Quantitative error analysis

We denote the feasible solution sets of problem (22)-(23) and problem (24) 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 of $v_\mathcal{K}$ to v, as well as the quantitative approximation error estimation, we need some constraint qualification, e.g., Mangasarian Fromovitz constraint qualification (MFCQ) [27]. Although function $g(\cdot, \mathcal{K})$ is non-smooth, it is 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 $F(t) = \{x \in \mathbb{R}^n | g_j(x,t) \le 0, j \in J\}$ with subdifferentiable g_j, here t is a parameter in the optimization. If there exist some vector θ, real number $\sigma < 0$, real number $\alpha_1 > 0$, and real number $\alpha_2 > 0$ such that*

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

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

We have noticed that the MFCQ condition under nonsmooth case has been discussed in many literature, such as [36, Page 14]. However, the non-differentiable MFCQ defined in Definition 3 is more strict than that in [36]. Firstly, Definition 3 defines non-differentiable MFCQ under a parameterized case, while [36] defines it under a case without parameters. Most importly, Definition 3 requires the non-differentiable MFCQ holds in a neighborhood of $(\bar{x},\bar{t})$ with a unified vector θ, while [36] only requires the non-differentiable MFCQ holds at a point $\bar{x}$.

ND-MFCQ defined in Definition 3 is an extension of the classic MFCQ proposed in [27], and it is equivalent to the classic MFCQ if the constraint functions are differentiable.

Proposition 6. *When g_j, $j \in J$ are differentiable, ND-MFCQ is equivalent to classic MFCQ.*

In this paper, we only have one constraint and thus $J = \{1\}$. Our decision variable z corresponds to x in Definition 3 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 5. *a) The objective function $f(z)$ is continuous and differentiable, and its gradients are bounded by $\mathcal{C}_f = \max_{z\in Z} \|\nabla f(z)\|_2 < \infty$;*

b) $\mathcal{S}_1$, the optimal solution set for problem (24) with $\mathcal{K} = 1$, is nonempty.

Observing the constraints in (23) and (24), we know 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 $\mathcal{S}_\mathcal{K}$ are nonempty for any $\mathcal{K}$.

Theorem 4. *Given Assumptions 1 and 5. 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. Then, for $\mathcal{K} \ge \max\left\{\frac{1}{\alpha_2}, \frac{2}{|\sigma|}\frac{\mathcal{R}_{max}-\mathcal{R}_{min}}{\alpha_1}\|\theta\|, -2\frac{\mathcal{R}_{max}-\mathcal{R}_{min}}{g(z^*)}\left(\mathcal{C}\frac{\|\theta\|}{|\sigma|}+1\right)\right\}$, we have that,*

$$
|v_\mathcal{K} - v| \le \mathcal{C}_f \frac{2}{|\sigma|} \frac{\mathcal{R}_{max} - \mathcal{R}_{min}}{\mathcal{K}}.
$$

Thus, $\lim_{\mathcal{K}\to\infty} v_\mathcal{K} = v$.

Proof. For all $\mathcal{K}$, 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 $g(z^*) \le 0$.

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

• Case 1: $g(z^*) = 0$. Then by the inequality (36) in Proposition 5, we immediately have

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

From the extended mean-value theorem [32, 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 exists a vector $v \in \partial g(z_{\mathcal{K}}^{\tau}, \mathcal{K})$ satisfying

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

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

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

Therefore, by the ND-MFCQ assumption, we have for all $\mathcal{K} \ge \max\left\{\frac{2}{|\sigma|}\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\alpha_1}\|\theta\|, \frac{1}{\alpha_2}\right\}$ that

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

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

• Case 2: $g(z^*) < 0$. Let $\delta := -g(z^*) > 0$. If $\mathcal{K} \ge 2\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\delta}\left(\mathcal{C}\frac{\|\theta\|}{|\sigma|} + 1\right)$, then we obtain from Propositions 4 and 5 that

$$\begin{aligned} |g(z_{\mathcal{K}}, \mathcal{K}) - g(z^*)| &\le |g(z_{\mathcal{K}}, \mathcal{K}) - g(z^*, \mathcal{K})| + |g(z^*, \mathcal{K}) - g(z^*)| \\ &\le \mathcal{C}\|z_{\mathcal{K}} - z^*\| + 2\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}} \\ &= 2\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}\left(\mathcal{C}\frac{\|\theta\|}{|\sigma|} + 1\right) \\ &\le \delta. \end{aligned}$$

This indicates that

$$g(z_{\mathcal{K}}, \mathcal{K}) \le g(z^*) + \delta = 0.$$

We also have $z_{\mathcal{K}} \in \mathcal{F}_{\mathcal{K}}$ under this case.

Next, we estimate the approximation error $|v_{\mathcal{K}} - v|$ of the optimal values. By Assumption 5 b), we can choose $z_{\mathcal{K}}^* \in \mathcal{S}_{\mathcal{K}}$. From the mean-value theorem, for some $\varsigma \in (0,1)$ and the corresponding point $z_{\mathcal{K}}^{\varsigma} = (1-\varsigma)z_{\mathcal{K}} + \tau z^*$, we have

$$\begin{aligned} v_{\mathcal{K}} - v = f(z_{\mathcal{K}}^*) - f(z^*) &\le f(z_{\mathcal{K}}) - f(z^*) = \langle \nabla f(z^{\varsigma}), z_{\mathcal{K}} - z^* \rangle \\ &\le \mathcal{C}_f\|z_{\mathcal{K}} - z^*\| = \mathcal{C}_f\frac{2}{|\sigma|}\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}. \end{aligned}$$

Since $\mathcal{F}_{\mathcal{K}} \subset \mathcal{F}$, then $v_{\mathcal{K}} \ge v$. Therefore, it holds that

$$|v_{\mathcal{K}} - v| \le \mathcal{C}_f\frac{2}{|\sigma|}\frac{\mathcal{R}_{\max} - \mathcal{R}_{\min}}{\mathcal{K}}.$$

□

Proposition 4 quantitatively estimates the approximation error between the optimal value of problem (24) and that of problem (22)-(23).

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

Observing that bilinear terms $\mu_2^{ik}z$ and $\tilde{\mu}_2^{ik}z$ in problem (P_{SSD-U}) makes 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 with respect to $\mu, \tilde{\mu}$; then fix $\mu, \tilde{\mu}$ and optimize with respect to 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 problem (P_{SSD-U}).

Algorithm 2 Sequential convex approximation

Start from $z^1 \in Z_0$, $k = 1$.
while $k \geq 1$ **do**
  Solve problem (P_{SSD-U}) with an additional constraint $z = z^k$. Denote the optimal $\mu, \tilde{\mu}$ by $\mu^k, \tilde{\mu}^k$, respectively.
  Solve problem (P_{SSD-U}) with additional constraints $\mu = \mu^k, \tilde{\mu} = \tilde{\mu}^k$. Denote the optimal z by z^{k+1}.
  if $z^{k+1} = z^k$ **then**
    Break.
  else
    $k \leftarrow k + 1$.
  end if
end while

Proposition 7. *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 solution 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 $\min\{f(z)|\ (z, \mu, \tilde{\mu}, y) \in \mathcal{F}_U\}$.

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

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

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 second-order stochastic dominance 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 to the distributionally robust stochastic dominance constrained model (P_{SSD}).

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:

$$\min \quad \frac{1}{2}\|z\|_2$$

$$\text{s.t.} \quad \mathbb{E}_P[(\eta - z^T\xi)_+] \le \mathbb{E}_P[(\eta - z_0^T\xi)_+], \ \forall \eta \in \mathbb{R}, \ \forall P \in \mathcal{Q}, \qquad (37)$$
$$z \in \mathbb{R}_2^+, \|z\|_1 \le 10.$$

where $z_0 = (1,0)^T$ and $\mathcal{Q} = \{P \in \mathcal{M}(\Xi) : d(P, \widehat{P}_N) \le \epsilon\}$ is defined as that in (6). Here $\widehat{P}_N = \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{\xi}_i}$ is the empirical distribution. The support set is supposed to be $\Xi = \{(\xi_1,\xi_2)| \ \xi_1 \in [0,250], \xi_2 \in [0,500]\}$. We set $\epsilon = 10^{-5}, N = 10$ and the observed sample set $\{\widehat{\xi}_i\}_{i=1}^{10}$ 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 (37).

lower bound approximation (by solving (P_{SSD-L}))				upper bound approximation (by Algorithm 2)			Gap
$\mathcal{N}_1$	$\mathcal{N}_2$	Optimal value	Optimal solution	$\mathcal{K}$	Optimal value	Optimal solution	
100	100	0.292	$(0.423, 0.403)^T$	10	0.410	$(0.801, 0.156)^T$	40.205%
200	200	0.296	$(0.427, 0.408)^T$	11	0.304	$(0.459, 0.400)^T$	2.679%
300	300	0.301	$(0.442, 0.401)^T$	12	0.303	$(0.465, 0.387)^T$	0.381%

We get the lower bound approximation by solving the linear programming formulation (P_{SSD-L}) 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., 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 $\mathcal{N}_1, \mathcal{N}_2$ and the interval number $\mathcal{K}$, which verifies the validation of the proposed approximation methods.

5.2 Case study: a practical portfolio selection problem

Next, we consider a financial application of model (P_{SSD}) to the portfolio selection problem with distributionally robust second-order stochastic dominance constraints:

$$\min_{z\in Z} \quad \mathbb{E}_{\widehat{P}_N}[-z^T\xi] \qquad (38)$$
$$\text{s.t.} \quad \mathbb{E}_P[(\eta - z^T\xi)_+] \le \mathbb{E}_P[(\eta - z_0^T\xi)_+], \ \forall \eta \in \mathbb{R}, \ \forall P \in \mathcal{Q}, \qquad (39)$$

where $Z = \{z \in \mathbb{R}^n | \ z \ge 0, \sum_{i=1}^n z_i = 1\}$. Problem (38)-(39) is inspired by [16, Example 4.2]. The difference between problem (38)-(39) and that in [16, Example 4.2] lies in the construction method of the ambiguity set. In [16, Example 4.2], the ambiguity set $\mathcal{Q}$ is determined by first two order moment information, while in problem (38)-(39), $\mathcal{Q}$ is defined through Wasserstein distance.

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. We select eight risky assets to constitute the stock pool, which are U.S. three-month treasury bills, U.S. long-term government bonds, S&P 500, Willshire 5000, NASDAQ, Lehmann Brothers corporate bond index, EAFE foreign stock index, and gold. We use the same historical annual return rate data as that in [7, Table 8.1] (with a total of 22 years). We choose the equally weighted portfolio as the benchmark portfolio z_0. The support set is defined as $\Xi = \{x = (x_1, \cdots, x_n)| a_i \le x_i \le b_i, i = 1, \cdots, n\}$, where a_i and b_i denote the smallest and largest historical annual return rates of the ith asset, respectively. In Algorithm 2, for deriving an upper bound approximation, the choice of the starting point z^1 is an essential issue. We already know that 1) the benchmark portfolio z_0 is definitely feasible for problem (P_{SSD}); 2) problem (P_{SSD-U}) is an upper bound approximation to problem (P_{SSD}) from Theorem 3. Therefore, z_0 is probably feasible for problem (P_{SSD-U}) and we thus choose z_0 as the starting point z^1.

The lower bound can be obtained by solving a linear programming problem or a sequence of small-scale linear programming problems using Algorithm 1. When deriving the lower bound approximation, we should approximate sets Ξ and $\mathcal{R} = z_0^T\Xi$ by randomly selected samples $\Xi_{\mathcal{N}_1}$ and $\Xi_{\mathcal{N}_2}$, respectively. To obtain a better lower bound approximation, we can repeat the sampling-optimizing process for multiple times and choose the largest optimal value as the final lower bound. For deriving the upper bound approximation, we solve a sequence of second order conic programming problems. When Algorithm 2 stops, we take the optimal value of the last step as the upper bound.

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 size, the impact of the interval number on the optimal value of the upper bound approximation, the impact of robust second-order stochastic dominance constraints defined by Wasserstein distance, as well as the influence of the robust radius.

5.2.1 Convergence of lower and upper bounds

Firstly, we examine the effectiveness of the proposed lower and upper bound approximations. We also demonstrate the convergence of the lower bound approximation with respect to the sample sizes $\mathcal{N}_1$, $\mathcal{N}_2$ 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 start from the case with $\mathcal{N}_1 = 40$ and $\mathcal{N}_2 = 40$, that is, both the approximate sets $\Xi_{\mathcal{N}_1}$ and $\Xi_{\mathcal{N}_2}$ have 40 samples. To make fair comparison later in Section 5.2.2 with the portfolio optimization problem with non-robust second-order stochastic dominance constraints, we let $\Xi_{\mathcal{N}_1}$ include all the historical annual return rates $\{\widehat{\xi}_i\}_{i=1}^{22}$ from [7, Table 8.1] and other 18 randomly generated samples from Ξ; similarly, $\Xi_{\mathcal{N}_2}$ consists of all the samples in $\{z_0^T\widehat{\xi}_i\}_{i=1}^{22}$ and other 18 randomly generated samples from $z_0^T\Xi$. After solving problem (P_{SSD-L}), we obtain an optimal value and an optimal solution. Then we repeat the sampling-optimizing tests for 10 times, and adopt the largest optimal value as the lower bound under this case. Then we generate other 20 samples from Ξ and add them into the set $\Xi_{\mathcal{N}_1}$, and also generate 20 samples from $z_0^T\Xi$ and add them into the set $\Xi_{\mathcal{N}_2}$, which corresponds to the case with $\mathcal{N}_1 = 60$ and $\mathcal{N}_2 = 60$. Repeat the above testing procedure. The process stops when $\mathcal{N}_1$ and $\mathcal{N}_2$ are large enough such that problem (P_{SSD-L}) cannot be solved on MATLAB due to huge time consumption.

For the upper bound approximation, we consider the cases with the interval number being $\mathcal{K} = 1, 2, 4, 8, 12$, respectively. For each case, we obtain an optimal value and an optimal solution by Algorithm 2. Figure 2 shows the convergence trend of the optimal values of the lower and upper bound approximations for problem (38)-(39).

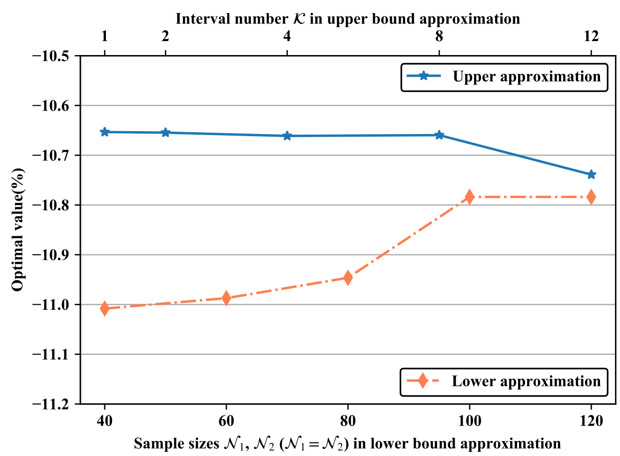

Figure 2: The optimal values of the lower bound approximation with respect to $\mathcal{N}_1, \mathcal{N}_2$ and that of the upper bound approximation with respect to $\mathcal{K}$.

From Figure 2, we can observe that the lower bound approximation monotonuously increases with the increase of the sample sizes $\mathcal{N}_1$ and $\mathcal{N}_2$. Besides, the upper bound approximation decreases with the increase of the interval number $\mathcal{K}$. The gap between the lower and upper bound approximations approaches to 0. To see more details, we present in Table 2 the optimal values and the optimal solutions obtained from the lower and upper bound approximations.

From Table 2 we can see that the optimal value obtained from the lower bound approximation is monotonously increasing with the increase of the sample sizes $\mathcal{N}_1, \mathcal{N}_2$. This verifies the asymptotic convergence result established in Theorem 1. We can also see from Table 2 that the optimal value obtained from the upper bound approximation decreases if the number of sub-intervals increases. This verifies the asymptotic convergence result in Theorem 4 and supports our split-and-dual technique for deriving the upper bound. From Table 2, we can also see the changing trend of the optimal portfolios of the lower and upper bound approximations. Especially

Table 2: The optimal values and the optimal solutions of the lower bound approximation with respect to $\mathcal{N}_1, \mathcal{N}_2$, and those of the upper bound approximation with respect to $\mathcal{K}$.

lower bound approximation (by solving (P_{SSD-L}))			upper bound approximation (by Algorithm 2)		Gap
$\mathcal{N}_1$	$\mathcal{N}_2$	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)	3.22%
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)	3.04%
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)	2.66%
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)	1.19%
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)	0.42%

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 is theoretically correct because the upper bound approximation is not tight. In fact, from problem (22)-(23) to problem (24), we exchange the order of operators $\sup_{\eta\in[\underline{\eta}_k,\bar{\eta}_k]}$ and $\mathbb{E}_P$, and such a transformation is not an equivalent reformulation. Therefore, a gap is induced here. To evaluate the difference between the lower and upper bound approximations, we calculate the relative gap between the upper bound with $\mathcal{K} = 12$ and the lower bound with $\mathcal{N}_1 = 120, \mathcal{N}_2 = 120$, which is only $|\frac{10.7838-10.7389}{-10.7838}| = 0.42\%$. This is quite satisfactory for real applications. Therefore, when considering distributionally robust second-order stochastic dominance constrained problems, we can efficiently approximate the original problem by the lower bound approximation (P_{SSD-L}) or the upper bound approximation (Algorithm 2), both optimal values are close to the true optimal value.

5.2.2 Price of distributional robustness

To examine the price of introducing distributional robustness, we compare the numerical results of robust stochastic dominance constrained portfolio optimization problem with those of classic stochastic dominance constrained portfolio optimization problem. Specifically, the latter model reads

$$\min \left\{ \mathbb{E}_{\widehat{P}_N}[-z^T\xi] |\ z \in Z, z^T\xi \succeq_{\widehat{P}_N} z_0^T\xi \right\},$$

which is equivalent to

$$\begin{aligned} \min_{z\in Z} \quad & \mathbb{E}_{\widehat{P}_N}[-z^T\xi] \\ \text{s.t.} \quad & \mathbb{E}_{\widehat{P}_N}[(\eta - z^T\xi)_+] \le \mathbb{E}_{\widehat{P}_N}[(\eta - z_0^T\xi)_+],\ \forall \eta \in \mathbb{R}. \end{aligned} \tag{40}$$

Here the expectations are taken under the empirical distribution $\widehat{P}_N$.

Table 3 reports the comparative results. In Table 3, for the distributionally robust stochastic dominance constrained problem, denoted by 'RSD', we present the optimal expected return rate (absolute value of the optimal value) and the optimal portfolio of the lower bound approximation obtained by solving (P_{SSD-L}) under $\epsilon = 10^{-4}, \mathcal{N}_1 = 120, \mathcal{N}_2 = 120$, and those of the upper bound approximation obtained by Algorithm 2 under $\mathcal{K} = 12$; for the classic stochastic dominance constrained problem, denoted by 'SD', we present the optimal expected return rate and the optimal portfolio. Table 3 also exhibits the benchmark portfolio and its expected return rate.

From Table 3, we can see that both the lower and upper bound approximations to problem (38)-(39) with distributionally robust stochastic dominance constraints derive a smaller optimal expected return rate than problem (40) with classic stochastic dominance constraints. Therefore, the optimal expected return rate of problem (38)-(39) must be smaller than that of problem (40). As we expected, considering the distributionally robust ambiguity in stochastic dominance 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

Table 3: The optimal expected return rates and the optimal portfolios for the lower and upper bound approximations to problem (38)-(39), and those of problem (40).

Portfolio optimization problem		Expected return rate(%)
RSD (38)-(39)	lower bound approximation	10.7838
	upper bound approximation	10.7389
SD (40)		11.0082
Benchmark		10.6534

than that of the benchmark portfolio, which means that model (38)-(39) 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 (38)-(39) with distributionally robust second-order stochastic dominance constraints. To obtain the lower bound approximation, we solve linear programming (P_{SSD-L}) with $\mathcal{N}_1 = 120$ and $\mathcal{N}_2 = 120$, while for the upper bound approximation, we adopt Algorithm 2 under $\mathcal{K} = 12$. The optimal expected return rates 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 radius.

Robust radius ϵ	Optimal values (%) lower bound approximation	upper bound approximation	Gap
10^{-5}	-10.8775	-10.8268	0.466%
10^{-4}	-10.7838	-10.7389	0.416%
10^{-3}	-10.7836	-10.6536	1.206%
10^{-2}	-10.7823	-10.6535	1.194%
0.1	-10.7689	-10.6534	1.072%
0.5	-10.6885	-10.6534	0.328%
1	-10.6534	-10.6534	0%

We can clearly see from Table 4 that both the optimal value of the lower and upper bound approximations to problem (38)-(39) are monotonously increasing with the increase of the robust radius, which implies that the optimal value for problem (38)-(39) also increase as the robust radius increases. This is theoretically natural because a problem with a larger robust radius has a smaller feasible solution set and thus has a larger optimal value. Table 4 also tells us that choosing a proper robust radius is a crucial issue in distributionally robust stochastic dominance constrained problems. We notice that for robust radius $\epsilon \geq 0.1$, the optimal portfolio obtained from the upper bound approximation problem coincides with the benchmark portfolio (the benchmark portfolio is always feasible for problem (38)-(39) and provides a trivial upper bound, -10.6534, for the optimal value of problem (38)-(39)), which means that the upper bound approximation does not provide additionally useful information for portfolio selection. Fortunately, when robust radius $\epsilon \leq 10^{-2}$, both the lower and upper bound approximations derive optimal portfolios different from the benchmark portfolio and thus are useful for portfolio selection.

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 sampling technique 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 prove that the lower bound approximation is asymptotically tight. 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 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. 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] Zhi Chen, Daniel Kuhn, and Wolfram Wiesemann. Data-driven chance constrained programs over Wasserstein balls. *Available at arXiv*, 2018.

[2] Zhiping Chen and Jie Jiang. Stability analysis of optimization problems with kth order stochastic and distributionally robust dominance constraints induced by full random recourse. *SIAM Journal on Optimization*, 28(2):1396–1419, 2018.

[3] Zhiping Chen, Yu Mei, and Jia Liu. Multivariate robust second-order stochastic dominance and resulting risk-averse optimization. *Optimization*, 68(9):1719–1747, 2019.

[4] Giorgio Consigli, Vittorio Moriggia, and Sebastiano Vitali. Long-term individual financial planning under stochastic dominance constraints. *Annals of Operations Research*, 292:973–1000, 2020.

[5] 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.

[6] Darinka Dentcheva and Werner Römisch. Stability and sensitivity of stochastic dominance constrained optimization models. *SIAM Journal on optimization*, 23(3):1672–1688, 2013.

[7] Darinka Dentcheva and Andrzej Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566, 2003.

[8] Darinka Dentcheva and Andrzej Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance*, 30:433–451, 2006.

[9] Darinka Dentcheva and Andrzej Ruszczyński. Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming*, 123(1):85–100, 2010.

[10] Jitka Dupačová and Miloš Kopa. Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research*, 234:434–441, 2014.

[11] 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.

[12] Csaba I Fábián, Gautam Mitra, and Diana Roman. Processing second-order stochastic dominance models using cutting-plane representations. *Mathematical Programming*, 130(1):33–57, 2011.

[13] Rui Gao and Anton J. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *Available at arXiv*, 2016.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] Jian Hu and Gevorg Stepanyan. Optimization with reference-based robust preference constraints. *SIAM Journal on Optimization*, 27(4):2230–2257, 2017.

[21] Ran Ji and Miguel Lejeune. Data-driven distributionally robust chance-constrained optimization with Wasserstein metric. *Available at SSRN 3201356*, 2020.

[22] Ruiwei Jiang and Yongpei Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, 158:291–327, 2016.

[23] 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.

[24] Jia Liu, Abdel Lisser, and Zhiping Chen. Distributionally robust chance constrained geometric optimization. *Available at Optimization Online*, 2019.

[25] 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.

[26] James Luedtke. New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19(3):1433–1450, 2008.

[27] 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.

[28] 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*, 2020.

[29] 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.

[30] Nilay Noyan and Gábor Rudolf. Optimization with stochastic preferences based on a general class of scalarization functions. *Operations Research*, 66(2):463–486, 2018.

[31] 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.

[32] R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*. Springer, Berlin, 2009.

[33] 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.

[34] 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.

[35] Weijun Xie. On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming*, 2019.

[36] Jane J. Ye and Daoli Zhu. Optimality conditions for bilevel programming problems. *Optimization*, 33(1):9–27, 1995.

[37] 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):124564, 2021.

[38] Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.

[39] 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.