

# Distributionally Robust Chance Constraints for Non-Linear Uncertainties

Wenzhuo Yang · Huan Xu

Received: date / Accepted: date

**Abstract** This paper investigates the computational aspects of distributionally robust chance constrained optimization problems. In contrast to previous research that mainly focused on the linear case (with a few exceptions discussed in detail below), we consider the case where the constraints can be non-linear to the decision variable, and in particular to the uncertain parameters. This formulation is of great interest as it can model non-linear uncertainties that are ubiquitous in applications. Our main result shows that distributionally robust chance constrained optimization is tractable, provided that the uncertainty is characterized by its mean and variance, and the constraint function is concave in the decision variables, and quasi-convex in the uncertain parameters. En route, we establish an equivalence relationship between distributionally robust chance constraint and the robust optimization framework that models uncertainty in a deterministic manner. This links two broadly applied paradigms in decision making under uncertainty and extends previous results of the same spirit in the linear case to more general cases. We then consider probabilistic envelope constraints, a generalization of distributionally robust chance constraints first proposed in Xu *et al.* [40] for the linear case. We extend this framework to the non-linear case, and derive sufficient conditions that guarantee its tractability. Finally, we investigate tractable approximations of joint probabilistic envelope constraints, and provide the conditions when these approximation formulations are tractable.

**Keywords** Robust optimization · Chance constraints · Probabilistic envelope constraints

---

Wenzhuo Yang  
Department of Mechanical Engineering, National University of Singapore  
E-mail: A0096049@nus.edu.sg

Huan Xu  
Department of Mechanical Engineering, National University of Singapore  
E-mail: mpexuh@nus.edu.sg

## 1 Introduction

Many optimization and decision making problems, when facing stochastic parameter uncertainty, can be tackled via the celebrated *chance constraint* paradigm. Here, a deterministic constraint is relaxed, and instead is required to hold with a certain probability (w.r.t. the uncertain parameter). That is, given a constraint  $f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha$  where  $\mathbf{x}$  denotes the decision variable,  $\alpha \in \mathbb{R}$  denotes the target value, and  $\boldsymbol{\delta}$ , the uncertain parameter, follows a distribution  $\mu$ , one solves:

$$\mathbb{P}_{\boldsymbol{\delta} \sim \mu}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha] \geq p, \quad (1)$$

for some value  $p \in (0, 1)$ . Chance constraints were first proposed by Charnes and Cooper [11], and since then there has been considerable work, e.g., Miller and Wagner [31], Prékopa [34], Delage and Mannor [16], and many others; we refer the reader to the textbook by Prékopa [35] and references therein for a thorough review.

While the chance constraint formulation is conceptually intuitive, it has two disadvantages that limit its practical applications. First, it is usually difficult to obtain enough samples to accurately estimate the distribution  $\mu$ . Second, optimization problems involving chance constraints are notoriously hard to solve, even when  $f(\cdot, \cdot)$  is bilinear (i.e., linear in either argument) and  $\mu$  is a uniform distribution (Nemirovski and Shapiro [32]). Indeed, the only known tractable case of the chance constraint formulation is when  $f(\cdot, \cdot)$  is bilinear and  $\mu$  follows a radial distribution (Calafiore and El Ghaoui [10]; Alizadeh and Goldfarb [1]).

A natural extension of the chance constraint paradigm that overcomes the above mentioned problems is the *distributionally robust chance constrained* (DRCC) approach (e.g., Calafiore and El Ghaoui [10], Erdogan and Iyengar [22], Delage and Ye [17], Zymler *et al.* [41]). In this paradigm, the distribution of the uncertain parameter is not precisely known, but instead, it is assumed to belong to a given set  $\mathfrak{P}$ . Constraint (1) is then replaced with the following constraint

$$\inf_{\mu \in \mathfrak{P}} \mathbb{P}_{\boldsymbol{\delta} \sim \mu}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha] \geq p. \quad (2)$$

In words, (2) requires that for all possible probability distributions of the stochastic uncertainty, the chance constraint must hold. Typically,  $\mathfrak{P}$  is characterized by the mean, the covariance, and sometimes the support of the distribution as well, all of which can be readily estimated from finite samples. The DRCC approach also brings in computational advantages, e.g., Cheung *et al.* [14] developed safe tractable approximations of chance constrained affinely perturbed linear matrix inequalities. A celebrated result by Calafiore and El Ghaoui [10] shows that when  $f(\cdot, \cdot)$  is bilinear and  $\mathfrak{P}$  is characterized by the mean and the variance, DRCC (2) can be converted into a *tractable* second order cone constraint.

Yet, most previous results on the tractability of DRCC are restricted to the case that  $f(\cdot, \cdot)$  is bilinear, whereas not much has been discussed when  $f(\cdot, \cdot)$  is non-linear. One exception that we are aware of is Zymler *et al.* [41], where they showed that DRCC is tractable when  $f(\mathbf{x}, \boldsymbol{\delta})$  is linear in the decision variable  $\mathbf{x}$  and *quadratic* or *piecewise linear* in the uncertainty  $\boldsymbol{\delta}$ . However, their method is built upon the S-lemma, and hence it is not clear how to extend the method to more general cases.

Another one is Cheng *et al.* [13] where they studied the knapsack problem with distributionally robust chance constraints when  $f(\mathbf{x}, \boldsymbol{\delta})$  is *piecewise linear* in the uncertainty  $\boldsymbol{\delta}$  and provided its equivalent formulation when the first and second moment and the support information of  $\boldsymbol{\delta}$  are known. To the best of our knowledge, the *general* non-linear case is largely untouched.

This paper is devoted to analyzing the tractability of DRCC (and its variants) under general – i.e., *non-linear* –  $f(\cdot, \cdot)$ . This problem is of interest, because in many applications the uncertainty is inherently non-linear, and cannot be modeled using a bilinear  $f(\cdot, \cdot)$ , e.g., [9, 28, 42]; see Section 2 for a more detailed discussion. In particular, we consider the following constraint

$$\inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha] \geq p, \quad (3)$$

where  $f(\mathbf{x}, \boldsymbol{\delta})$  is *concave* in  $\mathbf{x}$ , and *quasi-convex* in  $\boldsymbol{\delta}$ . Here, following the notations from Xu *et al.* [40], we use  $(0, \boldsymbol{\Sigma})$  to denote all distributions with mean zero and variance  $\boldsymbol{\Sigma}$ , and let  $\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})$  stand for  $\boldsymbol{\delta}$  follows some unknown distribution  $\mu$  that belongs to  $(0, \boldsymbol{\Sigma})$ . Notice that DRCC is a special case of *distributionally robust optimization* (e.g., [36, 21, 33, 17]) by setting the utility function to the indicator function. However, because the indicator function is neither convex nor concave in either argument, previous results on the tractability of DRO do not apply in our setup.

Our first contribution, presented in Section 3, establishes that Constraint (3), when  $f(\cdot, \cdot)$  is concave-quasiconvex, is tractable. En route, we derive an equivalence relationship between (3) and a robust optimization formulation using a deterministic uncertainty model (e.g., Ben-Tal *et al.* [6, 7, 5] Bertsimas and Sim [8]). This result thus links the two arguably most widely used approaches in optimization under uncertainty, and extends previous results of the same spirit for the linear case (e.g., Delage and Mannor [16], Shivaswamy *et al.* [38]).

Our second result, presented in Section 4, establishes the tractability of the probabilistic envelope model in the non-linear case. The probabilistic envelope model is proposed in Xu *et al.* [40], based on the following observation: the chance constraint (1) only guarantees that the given constraint will be satisfied with probability  $p$  or violated with the remaining  $(1 - p)$  probability, but no control is provided on the degree of violation. To overcome this, Xu *et al.* [40] proposed the probabilistic envelope constraint framework – essentially a set of infinite number of chance constraints at *all* levels of potential violation. That is, replace the single DRCC in (3) with the following

$$\inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha - s] \geq \mathfrak{B}(s), \quad \forall s \geq 0, \quad (4)$$

where  $\mathfrak{B}(s)$  is a given non-decreasing and right-continuous function of  $s$ . However, only the bilinear case has been investigated. In this paper, we extend the probabilistic envelope constraint to non-linear uncertainties. We prove that the optimization problem involving the probabilistic envelope constraint (4) is tractable when  $f(\cdot, \cdot)$  is concave-quasiconvex and  $\mathfrak{B}(s)$  satisfies some weak conditions. Similarly as for the (single) DRCC case, we establish a linkage between probabilistic envelope constraints and the comprehensive robust optimization framework using a deterministic uncertainty model (Ben-tal *et al.* [3, 2]).

It is worthwhile to note that the probabilistic envelope constraint is closely related to *stochastic dominance constraints* in the literature of stochastic programming (Dentcheva and Ruszczyński [18–20]); see Chapter 4 of the book by Shapiro *et al.* [37] for more details. A stochastic dominance constraint refers to a constraint of the form  $X \succeq_{(k)} Y$  where  $X$  and  $Y$  are random variables and  $\succeq_{(k)}$  stands for  $k$ -th order stochastic dominance. Thus, a probabilistic envelope constraint is indeed a first-order stochastic dominance constraint with the right hand side is a random variable whose cumulative distribution function is  $\mathfrak{B}(s)$ . However, most of the literature in optimization with stochastic dominance constraints does not address this specific case and instead focuses on the second (or higher) order constraints case, a case that preserves convexity and is more amenable to analysis. As we restrict our attention to this specific case, we choose to use the name “probabilistic envelope constraint”.

Finally, we extend our results in two ways, namely, more flexible uncertainty modeling and joint constraints. In Section 5, we provide tractability results for the case where the mean and variance themselves are unknown, and the case that the mean and the support of the distribution of the uncertain parameters are known. For more general uncertainty models where exact results appear difficult, we provide a conservative approximation scheme based on CVaR approximation of the chance constraints. In Section 6, we extend the probabilistic envelope constraint formulation to its joint chance constraint counterpart. This typically leads to a computationally challenging problem, and we adopt the CVaR approximation approach proposed by Zymler *et al.* [41], and show that the joint probabilistic envelope constraint can be approximated tractably under some technical conditions.

**Notation.** We use lower-case boldface letters to denote column vectors, upper-case boldface letters to denote matrices, and the transpose (superscript  $\top$ ) of the column vectors to denote row vectors. The all-ones vector is denoted by  $\mathbf{1}$ . The space of symmetric matrices of dimension  $n$  is denoted by  $\mathbb{S}^n$ . For any two matrices  $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n$ ,  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr}(\mathbf{X}\mathbf{Y})$  denotes the trace scalar product, and the relation  $\mathbf{X} \succeq \mathbf{Y}$  ( $\mathbf{X} \succ \mathbf{Y}$ ) implies that  $\mathbf{X} - \mathbf{Y}$  is positive semi-definite (positive definite). Random variables are always represented by  $\delta$ . Finally, we call an optimization problem *tractable* if it can be solved in polynomial time and call a set *tractable* if it is convex and a polynomial-time separation oracle can be constructed.

## 2 Formulation and Motivating Examples

We first propose the distributionally robust chance constraint, the probabilistic envelope constraint and the joint probabilistic envelope constraint discussed in this paper. For clarity, we repeat some of the definitions given in the introduction. Given a random variable  $\delta$  and a function  $f(\mathbf{x}, \delta)$ , a chance constraint places a lower-bound on the probability that the constraint reaches a certain target, which is defined as

$$\textbf{Distributionally Robust Chance Constraint: } \inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha] \geq p. \quad (5)$$

As discussed above, the distributionally robust chance constraint provides protection against noise by bounding the probability of failing to achieve a pre-defined target

$\alpha$ . It says nothing about what happens when, with probability at most  $(1 - p)$ , the target is not met. In particular, there is no control over the magnitude of violation of the constraint. To overcome this shortcoming, the probabilistic envelope constraint is proposed, which can enforce all levels of probabilistic guarantees. Given a non-decreasing function  $\mathfrak{B}(s)$ , the probabilistic envelope constraint can be written as

$$\textbf{Probabilistic Envelope Constraint: } \inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha - s] \geq \mathfrak{B}(s); \forall s \geq 0. \quad (6)$$

For example, if we want the probability of *large* constraint violation to decrease exponentially, then we can set  $\mathfrak{B}(s) = 1 - \gamma \exp(-\beta s)$ . Besides the individual probabilistic envelope constraint discussed above, we propose the following joint probabilistic envelope constraint (JPEC):

$$\textbf{JPEC: } \inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f_i(\mathbf{x}, \delta) \geq \alpha_i - s, \forall i = 1, \dots, m] \geq \mathfrak{B}(s); \forall s \geq 0. \quad (7)$$

Computationally, the joint envelope constraint is more complicated. A common method to simplify it is to decompose it into  $m$  individual envelope constraints by applying Bonferroni's inequality. However, since Bonferroni's inequality is not tight, this approximation method is usually overly conservative. In this paper, we use the worst-case CVaR method proposed by Zymler *et al.* [41] to give a tractable and tighter approximation for this joint envelope constraint.

Although the three types of constraints (5), (6), and (7) above can be general, they may not be tractable due to the non-convex feasible sets. To ensure tractability, we focus on the ‘‘concave-quasiconvex’’ case, i.e., the function  $f$  is concave w.r.t. the decision variable, and quasi-convex w.r.t. the uncertain parameters, see the following for a precise description:

**Assumption 1** *Let  $\mathbb{X}$  and  $\mathbb{Y}$  be two convex sets, and let  $f$  be a function mapping from  $\mathbb{X} \times \mathbb{Y}$  to  $\mathbb{R}$ ,*

1. *For each  $\mathbf{x} \in \mathbb{X}$ , the function  $f(\mathbf{x}, \cdot)$  is quasi-convex and continuous on  $\mathbb{Y}$ . For each  $\mathbf{y} \in \mathbb{Y}$ , the function  $f(\cdot, \mathbf{y})$  is concave on  $\mathbb{X}$ .*
2. *The uncertainty  $\delta$  is modeled as a random variable whose mean and variance are known but its distribution is unknown. Without loss of generality, we assume the mean is zero.*

Notice that Assumption 1 generalizes the case where  $f(\cdot, \cdot)$  is bilinear – a setup that previous literature mainly focused on – to the non-linear case. In particular, the uncertainty can be non-linear. Non-linearity of uncertainty arises naturally in a broad range of applications, as we demonstrate by the following examples.

### Example 1: Portfolio Optimization

Consider a stylized portfolio optimization problem, where an amount is to be allocated to  $n$  stocks and held for a time period  $T$ . Denote the price of the  $i$ th stock after

time  $T$  by  $S_i$ , and our goal is to maximize the Value at Risk (VaR) of the total return of the portfolio, which leads to the following formulation, for a fixed  $\gamma \in (0, 1)$

$$\begin{aligned} & \text{Maximize: } \mathbf{x} \geq 0, z \quad z \\ & \text{Subject to: } \mathbb{P} \left[ \sum_i S_i x_i \geq z \right] \geq 1 - \gamma; \mathbf{1}^\top \mathbf{x} = 1, \end{aligned}$$

where  $x_i$  is the allocation for the  $i$ th stock. It is well believed that the true drivers of the uncertainty in stock price is not the stock return  $S_i$  itself, but instead the compounded rates of return, i.e.,  $S_i = \exp(\delta_i)$  where  $\delta_i$  is the random variable to model and analyze. For example, the celebrated log-normal model, pioneered by Black and Scholes [9], models  $S_i$  as  $S_i = \exp((\mu_i - \sigma_i^2/2)T + \sqrt{T}\xi_i)$  where the vector  $\boldsymbol{\xi}$  is Normally distributed with mean 0 and covariance matrix  $\mathbf{Q}$ . This can be rewritten as  $S_i = \exp(\delta_i)$  where  $\boldsymbol{\delta} \sim \mathcal{N}((\mu_i - \sigma_i^2/2)T, T\mathbf{Q})$ .

One common criticism of the log-normal model is that it assumes  $\boldsymbol{\xi}$  to be Gaussian, whereas empirical evidence suggests that  $\boldsymbol{\xi}$  (and hence  $\boldsymbol{\delta}$ ) is fat-tailed (e.g., Jansen and de Vries [26], Cont [15], Kawas and Thiele [27]). Since the Gaussian assumption ignores the fat tails, it essentially leads the managers to take more risk than she is willing to accept. On the other hand, it remains controversial about what is the most appropriate fat-tail distribution to use in modeling returns [23, 29, 26, 15], and “this controversy has proven hard to resolve” as Jensen and de Vries stated [26]. In light of this, one possible approach is to not commit to any distribution, but instead only require that the first two moments match. This leads to the following problem:

$$\begin{aligned} & \text{Maximize: } \mathbf{x} \geq 0, z \quad z \\ & \text{Subject to: } \inf_{\boldsymbol{\delta} \sim ((\mu_i - \sigma_i^2/2)T, T\mathbf{Q})} \mathbb{P} \left[ \sum_i \exp(\delta_i) x_i \geq z \right] \geq 1 - \gamma; \mathbf{1}^\top \mathbf{x} = 1, \end{aligned} \tag{8}$$

Observe that this formulation satisfies Assumption 1, i.e., the constraint is linear to the decision variable and non-linearly convex to the uncertain parameters, and the decision variables are non-negative.

In portfolio optimization, options are another cause of non-linearity of the uncertainty (Kawas and Thiele [28], Zymler *et al.* [42]). Suppose for each stock, the investor is allowed to purchase an European call option at the price of  $c_i$  per unit, which gives her the right to buy a unit of stock  $i$  at time  $T$  with the strike price  $p_i$ . Thus, denote the stock return as  $S_i$ , the return of this option is  $\max(S_i - p_i, 0)$ , since the investor will execute the option if and only if  $S_i > p_i$ . The portfolio optimization problem is thus formulated as

$$\begin{aligned} & \text{Maximize: } \mathbf{x} \geq 0, \mathbf{y} \geq 0, z \quad z \\ & \text{Subject to: } \mathbb{P} \left[ \sum_i (S_i x_i + \max(S_i - p_i, 0) y_i) \geq z \right] \geq 1 - \gamma; \\ & \quad \sum_i (x_i + c_i y_i) = 1, \end{aligned}$$

where  $\mathbf{y}$  is the investment of the European call options. Notice that the constraints are non-linear, yet convex to  $S_i$ . Indeed, following the previous argument, we may

further model  $S_i = \exp(\delta_i)$ , and require that the first two moments of  $\delta$  are known. This makes the probabilistic constraint again satisfy Assumption 1.

### Example 2: Transportation Problem

Solving multi-stage optimization problems may also result in non-linearity of uncertainty and decision variables. We illustrate this using a transportation decision problem. Given a directed graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , and let  $\mathcal{S} \subset \mathcal{V}$  be the set of source nodes, and  $\mathcal{D} \subset \mathcal{V}$  be the set of destination nodes, with  $\mathcal{S} \cap \mathcal{D} = \emptyset$ . One can think of each node in  $\mathcal{S}$  as a supplier, and each node in  $\mathcal{D}$  as a consumer.

The decision to make contains two stages: in the first stage, the decision maker needs to decide the required flow of each source node and each destination node, i.e.,  $s(i)$  for  $i \in \mathcal{S}$  and  $d(j)$  for  $j \in \mathcal{D}$ . One can think of this as deciding how much amount of good to order from each supplier, and how much to sell to each client. Certain linear constraints on the required flow are imposed: for example, the total supply equals to the total demand, and they must be larger than a minimum demand  $L$ , i.e.,  $\sum_{i \in \mathcal{S}} s(i) = \sum_{j \in \mathcal{D}} d(j) \geq L$ .

In the second stage, after all the ordered goods are produced by the suppliers, the decision maker needs to decide how to transport these goods, i.e., the flow on the network from sources to destinations, by solving a minimum cost flow problem given  $s_i$  and  $d_j$ . This can be formulated as a linear program, where the decision variable  $f(u, v)$  is the flow from node  $u$  to node  $v$ :

$$\begin{aligned}
\text{Minimize:} \quad & \sum_{(u \rightarrow v) \in \mathcal{E}} \delta(u, v) f(u, v) \\
\text{Subject to:} \quad & \sum_{(u \rightarrow v) \in \mathcal{E}} f(u, v) - \sum_{(v \rightarrow u) \in \mathcal{E}} f(v, u) = 0 \quad \forall u \notin \mathcal{S} \cup \mathcal{D}; \\
& \sum_{(u \rightarrow v) \in \mathcal{E}} f(u, v) - \sum_{(v \rightarrow u) \in \mathcal{E}} f(v, u) = s(u) \quad \forall u \in \mathcal{S}; \\
& \sum_{(u \rightarrow v) \in \mathcal{E}} f(u, v) - \sum_{(v \rightarrow u) \in \mathcal{E}} f(v, u) = -d(u) \quad \forall u \in \mathcal{D}; \\
& f(u, v) \geq 0 \quad \forall (u \rightarrow v) \in \mathcal{E}; \\
& f(u, v) = 0 \quad \forall (u \rightarrow v) \notin \mathcal{E}.
\end{aligned}$$

Denote the optimal value by  $h(\mathbf{s}, \mathbf{d}, \boldsymbol{\delta})$ . Suppose  $\boldsymbol{\delta}$  represents uncertain parameters whose values are only revealed at stage two, then to ensure that the total transportation cost is low with high probability, the first stage decision can be formulated using DRCC:

$$\begin{aligned}
\text{Maximize: } & \mathbf{s} \geq 0, \mathbf{d} \geq 0, z \quad z \\
\text{Subject to: } & \inf_{\boldsymbol{\delta} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})} \mathbb{P}[-h(\mathbf{s}, \mathbf{d}, \boldsymbol{\delta}) \geq z] \geq 1 - \gamma; \\
& \sum_{i \in \mathcal{S}} s(i) = \sum_{j \in \mathcal{D}} d(j) \geq L.
\end{aligned}$$

It is easy to verify that  $-h(\mathbf{s}, \mathbf{d}, \boldsymbol{\delta})$  is non-linearly concave w.r.t. the decision variables  $(\mathbf{s}, \mathbf{d})$  and non-linearly convex w.r.t.  $\boldsymbol{\delta}$ . Thus, the above transportation problem satisfies Assumption 1.

### 3 The Chance Constraint Case

This section is devoted to the (individual) distributionally robust chance constraint case (5). Our main theorem shows that when function  $f(\mathbf{x}, \boldsymbol{\delta})$  satisfies Assumption 1, then a DRCC is equivalent to a robust optimization constraint. This bridges the two main approaches in optimization under uncertainty, namely, stochastic programming, and robust optimization. We then investigate the tractability of DRCC, providing sufficient conditions for the individual DRCC (5) to be tractable.

#### 3.1 Equivalence to Robust Optimization

In this subsection we show that DRCC is equivalent to robust optimization by analyzing the feasible set given by the constraint (5), which we denote by

$$S \triangleq \{\mathbf{x} \mid \inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha] \geq p\} = \{\mathbf{x} \mid \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) < \alpha] \leq 1 - p\}.$$

Our main tool to analyze  $S$  is the following result from Marshall and Olkin [30].

**Lemma 1** *Let  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_k)$  be a random vector with  $\mathbb{E}[\boldsymbol{\delta}] = 0$ ,  $\mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^\top] = \boldsymbol{\Sigma}$ , and  $T \subseteq \mathbb{R}^k$  be a closed convex set. Then we have*

$$\mathbb{P}[\boldsymbol{\delta} \in T] \leq \frac{1}{1 + \theta^2},$$

where  $\theta = \inf_{\mathbf{y} \in T} \sqrt{\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}}$ , and the equality can always be attained.

Notice that one technical difficulty that we face to apply Lemma 1 is that the set  $\{\boldsymbol{\delta} \mid f(\mathbf{x}, \boldsymbol{\delta}) < \alpha\}$  may not be closed. Hence we extend Lemma 1 to the case where  $T$  is not necessarily closed:

**Lemma 2** *Let  $T \subseteq \mathbb{R}^k$  be a convex set. Denote  $\theta = \inf_{\mathbf{y} \in T} \sqrt{\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}}$ . Then we have*

$$\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta} \in T] = \frac{1}{1 + \theta^2}.$$

*Proof* When  $T$  is empty, we have  $\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta} \in T] = 0$ . On the other hand,  $\theta = \inf_{\mathbf{y} \in T} \sqrt{\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}} = +\infty$  which implies  $1/(1 + \theta^2) = 0$ . Hence the lemma holds.

When  $T$  is non-empty,  $T$  has a non-empty relative interior. Let  $\mathbf{x}_0$  be a point in the relative interior of  $T$ . Let  $\bar{T}$  be the closure of  $T$ , and for  $0 \leq \lambda < 1$  define  $\underline{T}(\lambda)$  by

$$\underline{T}(\lambda) = \{\lambda(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0 \mid \mathbf{x} \in \bar{T}\}.$$



Thus, we have  $\underline{T}(\lambda)$  is closed, convex, and  $\underline{T}(\lambda) \subseteq T$ . Define

$$\bar{\theta} = \inf_{\mathbf{y} \in \bar{T}} \sqrt{\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}}, \quad \theta = \inf_{\mathbf{y} \in T} \sqrt{\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}}, \quad \underline{\theta}(\lambda) = \inf_{\mathbf{y} \in \underline{T}(\lambda)} \sqrt{\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}},$$

and hence  $\underline{\theta}(\lambda) \geq \theta \geq \bar{\theta}$ . On the other hand, for any  $\mathbf{x} \in \bar{T}$ , one can construct a sequence  $\mathbf{x}_i \rightarrow \mathbf{x}$  such that  $\mathbf{x}_i \in \underline{\theta}(\lambda_i)$  for some  $\{\lambda_i\}_{i=1}^\infty$ , by the definition of  $\underline{T}(\lambda)$ . Thus, since  $\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}$  is a continuous function of  $\mathbf{y}$ , we have  $\inf_{\lambda \in [0,1]} \underline{\theta}(\lambda) \leq \bar{\theta}$ , which implies  $\inf_{\lambda \in [0,1]} \underline{\theta}(\lambda) = \bar{\theta}$ . By Lemma 1, the following inequalities hold for  $0 \leq \lambda < 1$  since  $\underline{T}(\lambda)$  and  $\bar{T}$  are both closed convex sets:

$$\frac{1}{1 + \underline{\theta}(\lambda)^2} = \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta} \in \underline{T}(\lambda)] \leq \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta} \in T] \leq \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta} \in \bar{T}] = \frac{1}{1 + \bar{\theta}^2}.$$

Since  $\sup_{\lambda \in [0,1]} \frac{1}{1 + \underline{\theta}(\lambda)^2} = \frac{1}{1 + \bar{\theta}^2}$ , we have

$$\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta} \in T] = \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta} \in \bar{T}]$$

which establishes the lemma.  $\square$

Now we are ready to present the main result of this subsection.

**Theorem 1** *Suppose  $f(\mathbf{x}, \cdot)$  is quasi-convex for every  $\mathbf{x} \in \mathbb{X}$  and  $f(\cdot, \mathbf{y})$  is concave for every  $\mathbf{y} \in \mathbb{Y}$ , and let  $p \in (0, 1)$  and set  $r = p/(1 - p)$ , then the feasible set  $S$  of the DRCC (5) is convex and admits*

$$S = \{\mathbf{x} | \forall \mathbf{y} \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} < r \Rightarrow f(\mathbf{x}, \mathbf{y}) \geq \alpha\}.$$

If  $f(\mathbf{x}, \cdot)$  is further assumed to be continuous for every  $\mathbf{x} \in \mathbb{X}$ , then the distributionally robust chance constraint  $\inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha] \geq p$  is equivalent to

$$f(\mathbf{x}, \mathbf{y}) \geq \alpha, \quad \forall \mathbf{y} \in \Omega \triangleq \{\mathbf{y} | \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq r\}.$$

*Proof* Since  $f(\mathbf{x}, \cdot)$  is quasi-convex for each  $\mathbf{x} \in \mathbb{X}$ , the set  $T_{\mathbf{x}} \triangleq \{\mathbf{y} | f(\mathbf{x}, \mathbf{y}) < \alpha\}$  is convex for fixed  $\mathbf{x}$ . Then from Lemma 2, the feasible set of the constraint (5) satisfies

$$\begin{aligned} S &\triangleq \{\mathbf{x} | \inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha] \geq p\} = \{\mathbf{x} | \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) < \alpha] \leq 1 - p\} \\ &= \{\mathbf{x} | \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta} \in T_{\mathbf{x}}] \leq 1 - p\} \stackrel{(a)}{=} \{\mathbf{x} | \inf_{\mathbf{y} \in T_{\mathbf{x}}} \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \geq r\} \\ &= \{\mathbf{x} | \inf_{f(\mathbf{x}, \mathbf{y}) < \alpha} \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \geq r\} = \{\mathbf{x} | \forall \mathbf{y} \text{ such that } f(\mathbf{x}, \mathbf{y}) < \alpha \Rightarrow \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \geq r\} \\ &= \{\mathbf{x} | \forall \mathbf{y} \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} < r \Rightarrow f(\mathbf{x}, \mathbf{y}) \geq \alpha\}, \end{aligned}$$

where (a) holds by Lemma 2. Since  $f(\cdot, \mathbf{y})$  is concave for every  $\mathbf{y}$ , we know that  $S$  is convex, as the property is preserved under arbitrary intersection. Hence we proved the first part:  $S = \{\mathbf{x} | \forall \mathbf{y} \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} < r \Rightarrow f(\mathbf{x}, \mathbf{y}) \geq \alpha\}$ .

To show the second part, further notice that  $p \in (0, 1)$  implies  $r > 0$ . Thus we have

$$S = \{\mathbf{x} | f(\mathbf{x}, \mathbf{y}) \geq \alpha, \quad \forall \mathbf{y} \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq r\},$$

where the equality holds because for each  $\mathbf{x} \in \mathbb{X}$ ,  $f(\mathbf{x}, \mathbf{y})$  and  $\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}$  are both continuous in  $\mathbf{y}$  so that we can replace “<” by “ $\leq$ ” without effect on  $S$ .  $\square$

Thus the probabilistic uncertainty model is linked to the deterministic set based uncertainty model of robust optimization (e.g., Ben-Tal and Nemirovski [6, 7], Bertsimas and Sim [8]). This result is in the spirit of past work that has linked chance constraints to robust optimization in the linear case (e.g., Delage and Mannor [16], Shivaswamy *et al.* [38]).

Interestingly, based on the above theorem, we can establish an equivalence relationship between the distributionally robust chance constraint and the Worst Case Conditional Value at Risk (WCCVaR) in the convex case, which recovers a result first shown in [41] using a different proof.

**Corollary 1** *Suppose  $f(\mathbf{x}, \cdot)$  is convex and continuous for every  $\mathbf{x} \in \mathbb{X}$ , then for  $p \in (0, 1)$ ,*

$$\inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha] \geq p \Leftrightarrow \sup_{\delta \sim (0, \Sigma)} \text{CVaR}_{1-p}(-f(\mathbf{x}, \delta)) \leq -\alpha.$$

*Proof* See Appendix A.  $\square$

In the most general case, i.e.,  $f(\mathbf{x}, \delta)$  is quasi-convex, the equivalence shown in Corollary 1 does not hold. Consider a constraint with a random variable  $\delta$ :

$$\inf_{\delta \sim (0, \sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha] \geq 0.5.$$

We now construct a function  $f(\mathbf{x}, \delta)$  that is quasi-convex but not convex w.r.t.  $\delta$ . In particular, we construct  $f(\mathbf{x}, \delta)$  that is decreasing (hence quasi-convex) and *concave* w.r.t.  $\delta$ , such that the DRCC above holds but the constraint on the worse-case CVaR does not hold. For simplicity, denote  $-f(\mathbf{x}, \cdot)$  by  $L(\cdot)$  and let  $\alpha = -\sigma$ . Define  $L(\cdot)$  as follows:

$$L(x) = \begin{cases} \sigma, & x \leq \sqrt{\sigma}; \\ x^2, & x > \sqrt{\sigma}. \end{cases}$$

It can be easily shown that the constraint  $\inf_{\delta \sim (0, \sigma)} \mathbb{P}[L(\delta) \leq \sigma] \geq 0.5$  holds. Consider an uniform distribution over the interval  $[-\sqrt{3\sigma}, \sqrt{3\sigma}]$  which has mean 0 and variance  $\sigma$ . By simple computation, we can see that  $\text{CVaR}_{0.5}(L(\delta)) > \sigma$  w.r.t. this uniform distribution when  $\sigma = 1$ .

### 3.2 Tractability of Individual DRCC

In this subsection we investigate the tractability of DRCC. We first provide sufficient conditions for optimization problems involving chance constraint (5) with function  $f(\mathbf{x}, \delta)$  being tractable. We then show that for the special case where  $f(\mathbf{x}, \delta) = g(\delta)^\top \mathbf{x}$  and  $g(\delta)$  is linear or convex quadratic, we can convert (5) to an equivalent semi-definite constraint.

**Theorem 2** *If function  $f(\mathbf{x}, \delta)$  satisfies Assumption 1, set  $\Psi \subseteq \mathbb{X}$  is tractable and  $p \in (0, 1)$ , then the following optimization problem*

$$\begin{aligned} & \text{Minimize: }_{\mathbf{x} \in \Psi} \mathbf{c}^\top \mathbf{x} \\ & \text{Subject to: } \inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha] \geq p \end{aligned} \quad (9)$$

can be solved in polynomial time, if (i) for any fixed  $\delta$  the super-gradient of  $f(\cdot, \delta)$  can be evaluated in polynomial time; and (ii) for any fixed  $\mathbf{x} \in \Psi$  the following optimization problems on  $\mathbf{y}$  can be solved in polynomial time,

$$\begin{aligned} & \text{Minimize:}_{\mathbf{y}} \quad f(\mathbf{x}, \mathbf{y}) \\ & \text{Subject to:} \quad \mathbf{y}^T \Sigma^{-1} \mathbf{y} \leq \frac{p}{1-p}. \end{aligned} \quad (10)$$

*Proof* By Theorem 1, the feasible set  $S$  of the constraint (5) is given by

$$S = \{\mathbf{x} | f(\mathbf{x}, \mathbf{y}) \geq \alpha, \forall \mathbf{y} \text{ such that } \mathbf{y}^T \Sigma^{-1} \mathbf{y} \leq \frac{p}{1-p}\}.$$

To establish the theorem, it suffices to construct a polynomial-time *separation oracle* for  $S$  (Grötschel et al. [25]). A “separation oracle” is a routine such that for  $\mathbf{x}^*$ , it can be verified in polynomial time that (a) whether  $\mathbf{x}^* \in S$  or not; and (b) if  $\mathbf{x}^* \notin S$ , a hyperplane that separates  $\mathbf{x}^*$  with  $S$ .

We now construct such a separation oracle. To verify the feasibility of  $\mathbf{x}^*$ , notice that  $\mathbf{x}^* \in S$  if and only if the optimal value of the optimization problem (10) is greater than or equal to  $\alpha$ , which can be verified by solving Problem (10) directly. By assumption, this can be done in polynomial time.

If  $\mathbf{x}^* \notin S$ , then by solving Problem (10), we can find in polynomial time  $\mathbf{y}^*$  such that  $f(\mathbf{x}^*, \mathbf{y}^*) < \alpha$ . Because  $f(\mathbf{x}, \mathbf{y})$  is concave in  $\mathbf{x}$  for each  $\mathbf{y} \in \mathbb{Y}$ , for any  $\mathbf{x} \in S$ , the following holds

$$f(\mathbf{x}^*, \mathbf{y}^*) + \nabla_{\mathbf{x}} f(\mathbf{x}^*, \mathbf{y}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq f(\mathbf{x}, \mathbf{y}^*) \geq \alpha.$$

Thus, the hyperplane separating  $\mathbf{x}^*$  from the feasible set  $S$  is the following

$$f(\mathbf{x}^*, \mathbf{y}^*) + \nabla_{\mathbf{x}} f(\mathbf{x}^*, \mathbf{y}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq \alpha,$$

which can be generated in polynomial time since the super-gradient of  $\mathbf{x}$  can be obtained in polynomial time.  $\square$

We now consider the special case that  $f(\mathbf{x}, \delta) = g(\delta)^\top \mathbf{x}$  and each component  $g_i(\delta)$  of  $g(\delta)$  is either quadratic convex or linear.

**Corollary 2** *If  $f(\mathbf{x}, \delta) = g(\delta)^\top \mathbf{x}$  and satisfies Assumption 1 and each component of  $g(\delta)$  is a convex quadratic or linear function, i.e., it has the form  $g_i(\delta) = \delta^\top \mathbf{G}_i \delta + \mathbf{p}_i^\top \delta + q_i$ , where  $\mathbf{p}_i \in \mathbb{R}^n$ ,  $q_i \in \mathbb{R}$  and  $\mathbf{G}_i \in \mathbb{S}^n$  is a symmetric semi-definite matrix ( $G_i$  is zero if  $g_i(\delta)$  is linear), then the following optimization problem*

$$\begin{aligned} & \text{Minimize:}_{\mathbf{x} \in \Psi} \quad c(\mathbf{x}) \\ & \text{Subject to:} \quad \inf_{\delta \sim (0, \Sigma)} \mathbb{P}[g(\delta)^\top \mathbf{x} \geq \alpha] \geq p \end{aligned} \quad (11)$$

where  $p \in (0, 1)$ , is equivalent to

$$\begin{aligned} & \text{Minimize:}_{\mathbf{x} \in \Psi, \beta \geq 0} \quad c(\mathbf{x}) \\ & \text{Subject to:} \quad \begin{pmatrix} \beta \Sigma^{-1} + G(\mathbf{x}) & \frac{1}{2} P(\mathbf{x}) \\ \frac{1}{2} P(\mathbf{x})^\top & Q(\mathbf{x}) - \frac{\beta p}{1-p} \end{pmatrix} \succeq 0, \end{aligned} \quad (12)$$

where  $G(\mathbf{x}) \triangleq \sum_{i=1}^n x_i \mathbf{G}_i$ ,  $P(\mathbf{x}) \triangleq \sum_{i=1}^n x_i \mathbf{p}_i$ , and  $Q(\mathbf{x}) \triangleq \sum_{i=1}^n x_i q_i - \alpha$ .

*Proof* See Appendix B.  $\square$

Notice that  $G(\mathbf{x})$ ,  $P(\mathbf{x})$  and  $Q(\mathbf{x})$  are all linear functions of  $\mathbf{x}$ , and hence the semi-definite constraint in Problem (12) is a linear matrix inequality. Compare to the result by Calafiore and El Ghaoui [10] which only considers the case where  $f(\cdot, \cdot)$  is bilinear, the result above holds when  $f(\mathbf{x}, \cdot)$  is convex quadratic. Zymler *et al.* [41] showed that DRCC is tractable when  $f(\mathbf{x}, \delta)$  is linear in  $\mathbf{x}$  and quadratic in  $\delta$ . However, their method is built upon S-lemma, and hence it is not clear how to extend the method to more general cases. Our formulation needs stronger conditions –  $f(\mathbf{x}, \cdot)$  is convex quadratic – than [41], but the equivalent formulation is simpler than [41].

#### 4 Probabilistic Envelope Constraint

Recall that the probabilistic envelope constraint refers to the following:

$$\inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha - s] \geq \mathfrak{B}(s); \forall s \geq 0. \quad (13)$$

Here,  $s$  represents allowed magnitude of constraint violation, and  $\mathfrak{B}(s)$  is the probabilistic guarantee associated with a constraint violation no more than  $s$ . Hence,  $\mathfrak{B}(s) \in (0, 1)$  for all  $s \geq 0$ , and is assumed to be non-decreasing without loss of generality.

When  $f(\mathbf{x}, \delta)$  is bilinear, the envelope constraint (13) is shown to be equivalent to a comprehensive robust constraint, and proved to be tractable under mild technical conditions in Xu *et al.* [40]. We consider in this section the tractability of (13) where  $f(\mathbf{x}, \delta)$  satisfies Assumption 1. For convenience of exposition, we rewrite (13) to an equivalent formulation as shown in the following lemma.

**Lemma 3** *If  $\mathfrak{B}(s) : \mathbb{R}^+ \mapsto (0, 1)$  is a non-decreasing function that is continuous from the right, then the probabilistic envelope constraint (13) is equivalent to*

$$\inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha - t(r)] \geq \frac{r}{1+r}; \forall r \geq 0. \quad (14)$$

Here  $t(r) \triangleq \mathfrak{B}^{-1}(\frac{r}{1+r})$  and  $\mathfrak{B}^{-1}(x)$  is defined as

$$\mathfrak{B}^{-1}(x) \triangleq \begin{cases} \inf\{y \geq 0 | \mathfrak{B}(y) \geq x\} & \text{if } \exists y \text{ such that } \mathfrak{B}(y) \geq x; \\ +\infty & \text{otherwise.} \end{cases}$$

Furthermore,  $t(\cdot)$  is non-decreasing,  $t(0) = 0$ ,  $\lim_{r \uparrow +\infty} t(r) = +\infty$ , and  $t(\cdot)$  is continuous at the neighborhood of 0.

*Proof* See Appendix C.  $\square$

Hence in the sequel, we analyze the probabilistic envelope constraint (14) instead of (13). The following theorem shows that a probabilistic envelope constraint is equivalent to a comprehensive robust constraint proposed in Ben-Tal *et al.* [3], Ben-Tal *et al.* [4] and Ben-Tal *et al.* [2]. This thus extends previous results for affine cases in Xu *et al.* [40] to general  $f(\cdot, \cdot)$  satisfying Assumption 1.

**Theorem 3** Suppose  $t : \mathbb{R}^+ \mapsto [0, +\infty)$  is non-decreasing,  $t(0) = 0$ ,  $\lim_{r \uparrow +\infty} t(r) = +\infty$  and continuous at the neighborhood of 0. Then if function  $f(\mathbf{x}, \boldsymbol{\delta})$  satisfies Assumption 1, the probabilistic envelope constraint

$$\inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha - t(r)] \geq \frac{r}{1+r}; \forall r \geq 0 \quad (15)$$

is equivalent to the comprehensive robust constraint

$$f(\mathbf{x}, \mathbf{y}) \geq \alpha - t(\|\mathbf{y}\|_{\boldsymbol{\Sigma}^{-1}}^2), \forall \mathbf{y} \in \mathbb{R}^n. \quad (16)$$

*Proof* Define the feasible set of (15) as  $S$ . For any fixed  $r \geq 0$ , we have

$$\begin{aligned} S(r) &= \left\{ \mathbf{x} \mid \inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha - t(r)] \geq \frac{r}{1+r} \right\} \\ &= \{ \mathbf{x} \mid \forall \mathbf{y} \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} < r \Rightarrow f(\mathbf{x}, \mathbf{y}) + t(r) \geq \alpha \}. \end{aligned}$$

by Lemma 3 and Theorem 1. Thus, we have

$$\begin{aligned} S &= \left\{ \mathbf{x} \mid \inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha - t(r)] \geq \frac{r}{1+r}; \forall r \geq 0 \right\} \\ &= \{ \mathbf{x} \mid \forall \mathbf{y} \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} < r \Rightarrow f(\mathbf{x}, \mathbf{y}) + t(r) \geq \alpha; \forall r \geq 0 \}. \end{aligned}$$

Notice that without loss of generality, we can neglect the case  $r = 0$  in the right hand side, as  $\{ \mathbf{y} \mid \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} < 0 \} = \emptyset$ . Thus we have

$$S = \{ \mathbf{x} \mid \forall \mathbf{y} \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq r \Rightarrow f(\mathbf{x}, \mathbf{y}) + t(r) \geq \alpha; \forall r \geq 0 \},$$

where in the last equality we use the fact that  $\forall \mathbf{x} \in \mathbb{X}$ ,  $f(\mathbf{x}, \mathbf{y})$  and  $\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}$  are both continuous in  $\mathbf{y}$ , we can replace “ $<$ ” by “ $\leq$ ” without effect on  $S$  as long as  $\{ \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} < r \}$  is non-empty. By continuity of  $t(r)$  at  $r = 0$ , we further have

$$S = \{ \mathbf{x} \mid \forall (\mathbf{y}, r) \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq r \Rightarrow f(\mathbf{x}, \mathbf{y}) + t(r) \geq \alpha \}.$$

The second equality holds because there exists no  $\mathbf{y}$  such that  $\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq r$  when  $r < 0$  so that the constraint  $r \geq 0$  can be removed. Hence the probabilistic envelope constraint is equivalent to

$$f(\mathbf{x}, \mathbf{y}) + t(r) \geq \alpha, \forall (\mathbf{y}, r) \text{ such that } \|\mathbf{y}\|_{\boldsymbol{\Sigma}^{-1}}^2 \leq r. \quad (17)$$

Notice that (17) is equivalent to constraint (16) by monotonicity of  $t(\cdot)$ .  $\square$

It is known that comprehensive robust optimization generalizes robust optimization (e.g., Ben-Tal *et al.* [3], Ben-Tal *et al.* [4] and Ben-Tal *et al.* [2]). Indeed, if  $t(\cdot)$  is taken to be an indicator function, i.e.,  $t(r) = 0$  for  $r \in [0, c]$  and  $+\infty$  for  $r > c$ , the formulation (16) recovers the standard robust optimization formulation with the ellipsoidal uncertainty set  $\boldsymbol{\Omega} = \{ \mathbf{y} \mid \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq c \}$ . On the other hand, while robust optimization guarantees that the constraint is not violated for any realization of the uncertain parameters in the set  $\boldsymbol{\Omega}$ , it makes no guarantees for realizations outside that set. In contrast, the comprehensive robust optimization formulation allows us to

choose different functions  $t(\cdot)$ , in order to provide different levels of protection for different parameter realizations, as opposed to the “all-or-nothing” view of standard robust optimization.

We now investigate the tractability of probabilistic envelope chance constraints. We first consider the general case where  $f(\mathbf{x}, \boldsymbol{\delta})$  is an arbitrary “concave-quasiconvex” function. The following theorem is essentially an envelope constraint counterpart of Theorem 2.

**Theorem 4** *If  $t(\cdot)$  satisfies the conditions in Theorem 3,  $f(\mathbf{x}, \boldsymbol{\delta})$  satisfies Assumption 1 and set  $\Psi \subseteq \mathbb{X}$  is tractable, then the optimization problem with a linear objective function and the probabilistic envelope constraint (13):*

$$\begin{aligned} & \text{Minimize: } \mathbf{x} \in \Psi \quad \mathbf{c}^\top \mathbf{x} \\ & \text{Subject to: } \quad \inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha - t(r)] \geq \frac{r}{1+r}; \quad \forall r \geq 0 \end{aligned} \quad (18)$$

can be solved in polynomial time if (1) one can provide the super-gradient of  $f(\mathbf{x}, \boldsymbol{\delta})$  at  $\mathbf{x}$  for fixed  $\boldsymbol{\delta}$  in polynomial time, and (2) for any fixed  $\mathbf{x}$  the following optimization problems can be solved in polynomial time:

$$\begin{aligned} & \text{Minimize: } \mathbf{y}, r \quad f(\mathbf{x}, \mathbf{y}) + t(r) \\ & \text{Subject to: } \quad \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq r. \end{aligned} \quad (19)$$

*Proof* By Theorem 3, the feasible set  $S$  can be rewritten as

$$S = \{\mathbf{x} | \forall (\mathbf{y}, r) \text{ such that } \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq r \Rightarrow f(\mathbf{x}, \mathbf{y}) + t(r) \geq \alpha\}.$$

Similar to the proof of Theorem 2, we construct a separation oracle to prove tractability. In order to verify the feasibility of a given  $\mathbf{x}^*$ , notice that  $\mathbf{x}^* \in S$  if and only if the optimal objective value of the optimization problem (19) is greater than or equal to  $\alpha$ , which can be verified by directly solving Problem (19). By assumption, this can be done in polynomial time.

If  $\mathbf{x}^* \notin S$ , then by solving Problem (19), we can find in polynomial time  $(\mathbf{y}^*, r^*)$  such that  $f(\mathbf{x}^*, \mathbf{y}^*) + t(r^*) < \alpha$ . Because  $f(\mathbf{x}, \mathbf{y})$  is concave in  $\mathbf{x}$  for each  $\mathbf{y} \in \mathbb{Y}$ , for any  $\mathbf{x} \in S$ , we have

$$f(\mathbf{x}^*, \mathbf{y}^*) + \nabla f(\mathbf{x}^*, \mathbf{y}^*)^\top (\mathbf{x} - \mathbf{x}^*) + t(r^*) \geq f(\mathbf{x}, \mathbf{y}^*) + t(r^*) \geq \alpha.$$

Hence the hyperplane separating  $\mathbf{x}^*$  from the feasible set  $S$  is the following:

$$f(\mathbf{x}^*, \mathbf{y}^*) + \nabla f(\mathbf{x}^*, \mathbf{y}^*)^\top (\mathbf{x} - \mathbf{x}^*) + t(r^*) \geq \alpha, \quad (20)$$

which can be generated in polynomial time since the super-gradient of  $\mathbf{x}$  can be obtained in polynomial time. This completes the proof.  $\square$

Our next result states that when  $f(\mathbf{x}, \boldsymbol{\delta}) = g(\boldsymbol{\delta})^\top \mathbf{x}$  and  $g_i(\boldsymbol{\delta})$  is quadratic, (14) can be converted to a semi-definite constraint.

**Corollary 3** Suppose  $t(\cdot)$  satisfies the conditions in Theorem 3 and is convex,  $f(\mathbf{x}, \boldsymbol{\delta}) = g(\boldsymbol{\delta})^\top \mathbf{x}$  satisfies Assumption 1 and  $\Psi \subseteq \mathbb{X}$  is tractable, then if each component  $g_i(\boldsymbol{\delta})$  of  $g(\boldsymbol{\delta})$  is linear or convex quadratic as in Corollary 2, the optimization problem (18) is equivalent to

$$\begin{aligned} & \text{Minimize: }_{\mathbf{x} \in \Psi, \beta \geq 0} \quad \mathbf{c}^\top \mathbf{x} \\ & \text{Subject to: } \quad \begin{pmatrix} \beta \boldsymbol{\Sigma}^{-1} + G(\mathbf{x}) & \frac{1}{2} P(\mathbf{x}) \\ \frac{1}{2} P(\mathbf{x})^\top & Q(\mathbf{x}) - t^*(\beta) \end{pmatrix} \succeq 0 \end{aligned} \quad (21)$$

where  $t^*(\beta)$  is the conjugate function of  $t(r)$ , i.e.,  $t^*(\beta) \triangleq \sup_{r \geq 0} (\beta r - t(r))$ ; and  $P(\cdot)$ ,  $G(\cdot)$ ,  $Q(\cdot)$  are defined as in Corollary 2. Furthermore, the optimization problem (18) with a linear objective function and the probabilistic envelope constraint can be solved in polynomial time if for any  $\beta \geq 0$  the following optimization problem on  $r$  can be solved in polynomial time:

$$\text{Minimize: }_{r \geq 0} t(r) - \beta r. \quad (22)$$

*Proof* See Appendix C.  $\square$

In particular, when  $t(r)$  is a convex function, the optimization problems (19) and (22) are both convex and can be solved efficiently.

## 5 Chance Constraints: Beyond Mean and Variance

Thus far we have studied the setup that models unknown parameters as following an ambiguous distribution with known mean and covariance. In this section we extend our results to some other models of uncertain parameters – this includes the case where the mean and the covariance themselves are unknown and can only be estimated from data; and the case where other information of the uncertain parameter (e.g., the support) may be available. Specifically, we first show that the chance constraint (5) and the probabilistic envelope constraint (6) with *uncertain* mean and covariance are still tractable. Then we deal with the case where the mean and support of the uncertain parameter are known. Finally, we apply *distributionally robust optimization* to make a conservative approximation for constraints (5) and (6) when additional information on the uncertain parameter is available.

### 5.1 Uncertain Mean and Covariance

We first study the uncertain mean and covariance case. This model of ambiguity was first proposed and studied in [17] for *distributionally robust optimization*, and was also investigated for linear chance constraints in [40]. We formulate the robust counterparts of the distributionally robust chance constraint (5) and the probabilistic envelope constraint (6) where the mean and covariance themselves are uncertain, and then show that optimization problems with these constraints are tractable under mild conditions. Based on Theorem 1 and Theorem 3, we can easily obtain the following

corollaries. Corollary 4 and Corollary 5 show that the DRCC and the probabilistic envelope constraint with unknown mean and covariance is equivalent to a set of (infinitely many) *deterministic* constraints. Note that the uncertainty sets  $\mathcal{U}$  and  $\mathcal{S}$  can be arbitrary. Corollary 6 shows the tractability of probabilistic envelope constraints.

**Corollary 4** *If function  $f(\mathbf{x}, \boldsymbol{\delta})$  satisfies Assumption 1, then for  $p \in (0, 1)$  the chance constraint*

$$\inf_{\boldsymbol{\delta} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\mu} \in \mathcal{U}, \boldsymbol{\Sigma} \in \mathcal{S}} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha] \geq p, \quad (23)$$

*is equivalent to the constraint  $f(\mathbf{x}, \mathbf{y} + \boldsymbol{\mu}) \geq \alpha$ ,  $\forall \mathbf{y} \in \mathbb{R}^n, \boldsymbol{\mu} \in \mathcal{U}$  and  $\boldsymbol{\Sigma} \in \mathcal{S}$  such that  $\begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{y} \\ \mathbf{y}^\top & \frac{p}{1-p} \end{pmatrix} \succeq 0$ , where  $\mathcal{U}$  and  $\mathcal{S}$  are the uncertainty sets of mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ , respectively.*

**Corollary 5** *Suppose  $t : \mathbb{R}^+ \mapsto [0, +\infty)$  is non-decreasing,  $t(0) = 0$ ,  $\lim_{r \uparrow +\infty} t(r) = +\infty$  and is continuous at the neighborhood of zero. Then if function  $f(\mathbf{x}, \boldsymbol{\delta})$  satisfies Assumption 1, the probabilistic envelope constraint*

$$\inf_{\boldsymbol{\delta} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\mu} \in \mathcal{U}, \boldsymbol{\Sigma} \in \mathcal{S}} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha - t(r)] \geq \frac{r}{1+r}; \forall r \geq 0, \quad (24)$$

*is equivalent to the constraint*

$$\inf_{\boldsymbol{\mu} \in \mathcal{U}} f(\mathbf{x}, \mathbf{y} + \boldsymbol{\mu}) \geq \alpha - t\left(\inf_{\boldsymbol{\Sigma} \in \mathcal{S}} \|\mathbf{y}\|_{\boldsymbol{\Sigma}^{-1}}^2\right), \forall \mathbf{y} \in \mathbb{R}^n, \quad (25)$$

*where  $\mathcal{U}$  and  $\mathcal{S}$  are the uncertainty sets of mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ , respectively.*

**Corollary 6** *Under the conditions of Corollary 5, an optimization problem with a linear objective function and the probabilistic envelope constraint (24) can be solved in polynomial time if one can provide the super-gradient of  $f(\mathbf{x}, \boldsymbol{\delta})$  at  $\mathbf{x}$  for fixed  $\boldsymbol{\delta}$  in polynomial time, and for any fixed  $\mathbf{x}$  the following optimization problem can be solved in polynomial time:*

$$\begin{aligned} & \text{Minimize: } f(\mathbf{x}, \mathbf{y} + \boldsymbol{\mu}) + t(r) \\ & \text{Subject to: } \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{y} \\ \mathbf{y}^\top & r \end{pmatrix} \succeq 0 \\ & \boldsymbol{\Sigma} \in \mathcal{S}, \boldsymbol{\mu} \in \mathcal{U}. \end{aligned} \quad (26)$$

*Proof* See Appendix D.  $\square$

From Corollary 6 we see that if  $t(\cdot)$  is convex, and  $\mathcal{U} \subseteq \mathbb{R}^n$  and  $\mathcal{S} \in \mathbb{S}_+^{n \times n}$  are both convex sets, then the optimization problem (26) is a SDP problem which can be solved efficiently. The tractability result of the chance constraint (23) is a special case of Corollary 6, namely,  $t(r) = 0$  and  $r = \frac{p}{1-p}$ .



## 5.2 Known Mean and Support

We now investigate the case where the mean and the support of the uncertain parameter  $\delta$  are known. We show that the corresponding robust chance constraint can be reformulated as a set of infinitely many deterministic constraints, and is tractable under mild technical conditions. Unfortunately, it seems that these results can not be easily extended to the probabilistic envelope constraint case, which is hence left for future research.

**Theorem 5** *Suppose the mean  $\mu$  and support  $\mathcal{S}$  of the uncertain parameter  $\delta$  are known and  $\mathcal{S}$  is a closed convex set. If  $f(\mathbf{x}, \cdot)$  is continuous and quasi-convex for every  $\mathbf{x} \in \mathbb{X}$ , then for  $p \in (0, 1]$ , the chance constraint*

$$\inf_{\delta \sim (\mu, \mathcal{S})} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha] \geq p, \quad (27)$$

is equivalent to

$$f(\mathbf{x}, \delta_1) \geq \alpha, \quad \forall \delta_1, \delta_2 \text{ such that } (1-p)\delta_1 + p\delta_2 - \mu = 0, \delta_1 \in \mathcal{S}, \delta_2 \in \mathcal{S}. \quad (28)$$

*Proof* See Appendix D.  $\square$

**Theorem 6** *If  $f(\mathbf{x}, \cdot)$  is quasi-convex and continuous for every  $\mathbf{x} \in \mathbb{X}$  and  $f(\cdot, \mathbf{y})$  is concave for every  $\mathbf{y} \in \mathbb{Y}$ , the mean  $\mu$  and support  $\mathcal{S}$  of the uncertain parameter  $\delta$  are known, and  $\mathcal{S}$  is a closed convex set, then for  $0 < p \leq 1$ , the optimization problem with a linear objective function and a chance constraint (27):*

$$\begin{aligned} \text{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{Subject to:} \quad & \inf_{\delta \sim (\mu, \mathcal{S})} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha] \geq p \end{aligned} \quad (29)$$

can be solved in polynomial time if (1) one can provide the super-gradient of  $f(\mathbf{x}, \mathbf{y})$  at  $\mathbf{x}$  for fixed  $\mathbf{y}$  in polynomial time, and (2) for any fixed  $\mathbf{x}$  the following optimization problems can be solved in polynomial time:

$$\begin{aligned} \text{Minimize:}_{\delta_1, \delta_2} \quad & f(\mathbf{x}, \delta_1) \\ \text{Subject to:} \quad & (1-p)\delta_1 + p\delta_2 - \mu = 0 \\ & \delta_1, \delta_2 \in \mathcal{S}. \end{aligned} \quad (30)$$

*Proof* From Theorem 5 we know that the chance constraint is satisfied if and only if the optimal value of (30) is greater than or equal to  $\alpha$ . Thus, the theorem can be proved following a similar argument as the proof of Corollary 6.  $\square$

## 5.3 Conservative Approximation

For general sets of ambiguous distributions, optimization problems involving chance constraints are notoriously hard to solve. Recall that CVaR provides a conservative approximation of chance constraints (Nemirovski *et al.* [32]), which allows us to apply DRO to approximately solve such problems. For completeness, we give the following lemma which is an extension of Nemirovski *et al.* [32]:

**Lemma 4** Suppose that  $\mathcal{D}$  is the ambiguity set of distributions of the uncertain parameter  $\delta$ , then the chance constraint

$$\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}[f(\mathbf{x}, \delta) \geq 0] \leq p, \quad (31)$$

can be conservatively approximated by

$$-tp + \gamma \leq 0, \quad \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[[f(\mathbf{x}, \delta) + t]_+] \leq \gamma, \quad (32)$$

where  $0 \leq p \leq 1$ ,  $t \in \mathbb{R}$  and  $\gamma \in \mathbb{R}$  are decision variables, and  $[x]_+ = \max\{x, 0\}$ . Here, by ‘‘conservative approximation’’ we mean that any solution that satisfies (32) also satisfies (31).

Wiesemann *et al.* [39] proposed a unified framework for modeling and solving distributionally robust optimization problems by introducing standardized ambiguity sets

$$\mathcal{D} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^m, \mathbb{R}^n) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\mathbf{A}\delta + \mathbf{B}\mu] = \mathbf{b} \\ \mathbb{P}[(\delta, \mu) \in \mathcal{C}_i] \in [p_i, \bar{p}_i], \forall i \in \mathcal{I} \end{array} \right\}, \quad (33)$$

where  $\mathbb{P}$  represents a joint probability distribution of the random vector  $\delta \in \mathbb{R}^m$  appearing in the constraint function  $f(\mathbf{x}, \delta)$  and some auxiliary random vector  $\mu \in \mathbb{R}^n$ , with  $\mathbf{A} \in \mathbb{R}^{k \times m}$ ,  $\mathbf{B} \in \mathbb{R}^{k \times n}$ ,  $\mathbf{b} \in \mathbb{R}^k$ ,  $\mathcal{I} = \{1, \dots, I\}$ ,  $p_i, \bar{p}_i \in [0, 1]$  and  $\mathcal{C}_i$  are the confidence sets.

Applying Theorem 1 and 5 in [39], the constraint  $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[[f(\mathbf{x}, \delta) + t]_+] \leq \gamma$  can be reformulated as a semi-infinite constraint system. For succinctness, we only present the conservative approximation of the chance constraints when  $p_i = \bar{p}_i = 1$  and  $|\mathcal{I}| = 1$  to illustrate our approach.

**Theorem 7** If the ambiguity set  $\mathcal{P}$  can be converted into

$$\mathcal{D} = \{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^m, \mathbb{R}^n) : \mathbb{E}_{\mathbb{P}}[\mathbf{A}\delta + \mathbf{B}\mu] = \mathbf{b}, \mathbb{P}[(\delta, \mu) \in \mathcal{C}] = 1 \}$$

by the lifting theorem (Theorem 5 in [39]) where  $g(\cdot)$  is a convex function, then the chance constraint  $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha] \geq p$  with  $p \in (0, 1)$  can be conservatively approximated by

$$(\mathbf{A}\delta + \mathbf{B}\mu)^\top \beta \geq \max \left[ -\lambda, \alpha - f(\mathbf{x}, \delta) + \frac{\mathbf{b}^\top \beta}{1-p} + \frac{p\lambda}{1-p} \right], \quad \forall (\delta, \mu) \in \mathcal{C} \quad (34)$$

where  $\beta, \lambda, \mathbf{x}$  are decision variables. Furthermore, the optimization problem with a linear objective function and the constraint (34) can be solved in polynomial time if (1) one can provide the super-gradient of  $f(\mathbf{x}, \mathbf{y})$  at  $\mathbf{x}$  for fixed  $\mathbf{y}$  in polynomial time, and (2) for any fixed  $(\mathbf{x}, \beta, \lambda)$  the following optimization problems

$$\begin{array}{ll} \text{Minimize: } \delta, \mu & (\mathbf{A}\delta + \mathbf{B}\mu)^\top \beta + \lambda \\ \text{Subject to:} & (\delta, \mu) \in \mathcal{C}, \end{array} \quad (35)$$

and

$$\begin{array}{ll} \text{Minimize: } \delta, \mu & (\mathbf{A}\delta + \mathbf{B}\mu - \frac{\mathbf{b}}{1-p})^\top \beta - \frac{p\lambda}{1-p} + f(\mathbf{x}, \delta) \\ \text{Subject to:} & (\delta, \mu) \in \mathcal{C}, \end{array} \quad (36)$$

can be solved in polynomial time.

*Proof* From Theorem 1 in [39] and Lemma 4, the conservative approximation formulation can be easily obtained. The proof of the tractability result is similar to that of Corollary 6, and hence omitted.  $\square$

We now extend this result to the probabilistic envelope constraint case.

**Theorem 8** *Suppose  $t : \mathbb{R}^+ \mapsto [0, +\infty)$  is convex, non-decreasing and continuous at the neighborhood of zero, and  $t(0) = 0$ ,  $\lim_{r \uparrow +\infty} t(r) = +\infty$ . If the ambiguity set  $\mathcal{D}$  satisfies the condition in Theorem 7, the probabilistic envelope constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}[f(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha - t(r)] \geq r/(1+r)$ ,  $\forall r \geq 0$  can be conservatively approximated by*

$$\begin{aligned} (\mathbf{A}\boldsymbol{\delta} + \mathbf{B}\boldsymbol{\mu})^\top \boldsymbol{\beta} \geq \\ \max[-\lambda, \alpha - f(\mathbf{x}, \boldsymbol{\delta}) - t(r) + (1+r)\mathbf{b}^\top \boldsymbol{\beta} + r\lambda], \quad \forall (\boldsymbol{\delta}, \boldsymbol{\mu}) \in \mathcal{C}, r \geq 0. \end{aligned} \quad (37)$$

Furthermore, the optimization problem with a linear objective function and this probabilistic envelope constraint can be solved in polynomial time if one can provide the super-gradient of  $f(\mathbf{x}, \mathbf{y})$  at  $\mathbf{x}$  for fixed  $\mathbf{y}$  in polynomial time, and for any fixed  $(\mathbf{x}, \boldsymbol{\beta}, \lambda)$  the following optimization problems:

$$\begin{aligned} \text{Minimize: }_{\boldsymbol{\delta}, \boldsymbol{\mu}, r} \quad & (\mathbf{A}\boldsymbol{\delta} + \mathbf{B}\boldsymbol{\mu})^\top \boldsymbol{\beta} + \lambda \\ \text{Subject to: } \quad & (\boldsymbol{\delta}, \boldsymbol{\mu}) \in \mathcal{C}, r \geq 0, \end{aligned} \quad (38)$$

and

$$\begin{aligned} \text{Minimize: }_{\boldsymbol{\delta}, \boldsymbol{\mu}, r} \quad & [\mathbf{A}\boldsymbol{\delta} + \mathbf{B}\boldsymbol{\mu} - (1+r)\mathbf{b}]^\top \boldsymbol{\beta} + f(\mathbf{x}, \boldsymbol{\delta}) + t(r) - r\lambda \\ \text{Subject to: } \quad & (\boldsymbol{\delta}, \boldsymbol{\mu}) \in \mathcal{C}, r \geq 0. \end{aligned} \quad (39)$$

can be solved in polynomial time.

*Proof* From Theorem 7, the probabilistic envelope constraint can be conservatively approximated by

$$\begin{aligned} 0 \leq \min_{r \geq 0} \max_{\boldsymbol{\beta}, \lambda} \min_{(\boldsymbol{\delta}, \boldsymbol{\mu}) \in \mathcal{C}} (\mathbf{A}\boldsymbol{\delta} + \mathbf{B}\boldsymbol{\mu})^\top \boldsymbol{\beta} - \\ \max[-\lambda, \alpha - f(\mathbf{x}, \boldsymbol{\delta}) - t(r) + (1+r)\mathbf{b}^\top \boldsymbol{\beta} + r\lambda], \end{aligned} \quad (40)$$

Furthermore, by switching ‘‘min’’ and ‘‘max’’, this can be conservatively approximated by (37). Then following a similar proof as that of Corollary 6, we obtain the tractability result to complete the proof.  $\square$

## 6 Joint Chance Constraint

In this section we investigate the case of joint probabilistic envelope constraint (7) which can be reformulated as (from Lemma 3)

$$\inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f_i(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha_i - t(r), \forall i = 1, \dots, m] \geq \frac{r}{1+r}; \quad \forall r \geq 0, \quad (41)$$

where  $t(r) = \mathfrak{B}^{-1}(r/(1+r))$ . The optimization problem with the constraint (41) is usually intractable (e.g., Nemirovski and Shapiro [32]; Zymler *et al.* [41]), even when  $f(\mathbf{x}, \boldsymbol{\delta})$  is a bi-linear function, and approximation schemes are often used to tackle them. The most straightforward method to approximate the constraints (41) is to decompose them into several individual probabilistic envelope constraints using Bonferroni's inequality (see below for details). A notable advantage of the Bonferroni approximation is that it is easy to implement and requires no assumptions on the function  $f_i(\mathbf{x}, \boldsymbol{\delta})$ .

However, the Bonferroni approximation can be overly conservative. Zymler *et al.* [41] proposed a tighter approximation method called *worst-case CVaR approximation* that outperforms other methods including the Bonferroni approximation (e.g. Nemirovski and Shapiro [32] and Chen *et al.* [12]). In the rest of the section, we extend both the Bonferroni approximation and worst-case CVaR methods to JPEC. We also investigate the tractability of the two approximation schemes for  $f_i(\mathbf{x}, \boldsymbol{\delta})$  satisfying Assumption 1.

### 6.1 The Bonferroni Approximation

The Bonferroni approximation for the joint probabilistic envelope constraint (41) can be easily derived from Bonferroni's inequality. From Theorem 2 and Theorem 4, we know that the optimization problem with a set of probabilistic envelope constraints generated by the Bonferroni approximation method is tractable, under mild technical conditions. More specifically we have the following theorem:

**Theorem 9** *Let  $t : \mathbb{R}^+ \mapsto [0, +\infty)$  be a non-decreasing function such that  $t(0) = 0$  and  $\lim_{r \uparrow +\infty} t(r) = +\infty$ , and  $\boldsymbol{\epsilon}$  be a constant vector such that  $\sum_{i=1}^m \epsilon_i = 1$  and  $\boldsymbol{\epsilon} \geq 0$ . The Bonferroni approximation of the joint probabilistic envelope constraint (41) which has the form*

$$\inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[f_i(\mathbf{x}, \boldsymbol{\delta}) \geq \alpha_i - t(r)] \geq 1 - \frac{\epsilon_i}{(1+r)}; \forall r \geq 0, \forall i = 1, \dots, m. \quad (42)$$

*is tractable if for each  $i$ , (1) one can provide the super-gradient of  $f_i(\mathbf{x}, \boldsymbol{\delta})$  at  $\mathbf{x}$  for fixed  $\boldsymbol{\delta}$  in polynomial time, and (2) for any fixed  $\mathbf{x}$  the following optimization problem can be solved in polynomial time:*

$$\begin{aligned} & \text{Minimize:}_{\mathbf{y}, r} \quad f_i(\mathbf{x}, \mathbf{y}) + t(r) \\ & \text{Subject to:} \quad \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \leq \frac{r+1}{\epsilon_i} - 1. \end{aligned} \quad (43)$$

*Proof* Let  $r' = (1+r)/\epsilon_i - 1$ , then we have  $r'/(1+r') = 1 - \epsilon_i/(1+r)$ . Let  $t'(r') \triangleq t(r)$ , then we apply Theorem 4 to complete the proof.  $\square$

### 6.2 The Worst-case CVaR Approximation

Zymler *et al.* [41] developed a new approximation scheme for robust joint chance constraints termed *Worst-case CVaR approximation*. In this subsection we extend the

worst-case CVaR approximation to JPEC (41). In contrast to the rest of the paper, we focus on the linear-quadratic uncertainty case, namely,  $f(\mathbf{x}, \boldsymbol{\delta})$  is linear in  $\mathbf{x}$  for any fixed  $\boldsymbol{\delta}$  and quadratic (possibly non-convex) in  $\boldsymbol{\delta}$  for each  $\mathbf{x} \in \mathbb{X}$ . Then (41) can be rewritten respectively as:

$$\inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta}^\top Q_i(\mathbf{x})\boldsymbol{\delta} + y_i(\mathbf{x})^\top \boldsymbol{\delta} + y_i^0(\mathbf{x}) + t(r) \leq 0, \forall i = 1, \dots, m] \geq \frac{r}{1+r}; \forall r \geq 0; \quad (44)$$

where  $Q_i(\mathbf{x})$ ,  $y_i^0(\mathbf{x})$  and  $y_i(\mathbf{x})$  are all linear functions for  $i = 1, \dots, m$ . Zymler *et al.* [41] provided the Worst-case CVaR approximation for the following robust joint chance constraint

$$\inf_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[\boldsymbol{\delta}^\top Q_i(\mathbf{x})\boldsymbol{\delta} + y_i(\mathbf{x})^\top \boldsymbol{\delta} + y_i^0(\mathbf{x}) \leq 0, \forall i = 1, \dots, m] \geq p. \quad (45)$$

**Theorem 10** [41] Let  $\mathcal{A} \triangleq \{\boldsymbol{\alpha} \in \mathbb{R}^m | \boldsymbol{\alpha} > 0\}$ . For any fixed  $\mathbf{x}$  and  $\boldsymbol{\alpha} \in \mathcal{A}$ , the feasible set of the worst-case CVaR approximation for the constraint (45) is

$$Z^{JCC}(\boldsymbol{\alpha}) = \left\{ \mathbf{x} \in \mathbb{R}^n : \begin{array}{l} \exists(\beta, \mathbf{M}) \in \mathbb{R} \times \mathbb{S}^{k+1}, \\ \beta + \frac{1}{1-p} \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \mathbf{M} \succeq 0 \\ \mathbf{M} - \begin{pmatrix} \alpha_i Q_i(\mathbf{x}) & \frac{1}{2} \alpha_i y_i(\mathbf{x}) \\ \frac{1}{2} \alpha_i y_i(\mathbf{x})^\top & \alpha_i y_i^0(\mathbf{x}) - \beta \end{pmatrix} \succeq 0 \\ \forall i = 1, \dots, m \end{array} \right\}, \quad (46)$$

where  $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\Sigma}, 1)$ .

Indeed, Zymler *et al.* [41] showed that the approximation quality of the worst-case CVaR is controlled by the parameter  $\boldsymbol{\alpha}$  and that the approximation becomes exact if  $\boldsymbol{\alpha}$  is chosen optimally. Notice that  $Z^{JCC}(\boldsymbol{\alpha})$  contains semi-definite constraints, and hence provides a tractable approximation to robust joint chance constraint. We now extend this methodology to the joint probabilistic envelope constraints (44). From Theorem 10, the feasible set of the constraint (44) can be approximated as

$$Z^P(\boldsymbol{\alpha}) = \left\{ \mathbf{x} \in \mathbb{R}^n : \begin{array}{l} \text{For any } r \geq 0 \text{ we have} \\ \exists(\beta, \mathbf{M}) \in \mathbb{R} \times \mathbb{S}^{k+1}, \\ \beta + (r+1) \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \mathbf{M} \succeq 0 \\ \mathbf{M} - \begin{pmatrix} \alpha_i Q_i(\mathbf{x}) & \frac{1}{2} \alpha_i y_i(\mathbf{x}) \\ \frac{1}{2} \alpha_i y_i(\mathbf{x})^\top & \alpha_i (y_i^0(\mathbf{x}) - t(r)) - \beta \end{pmatrix} \succeq 0 \\ \forall i = 1, \dots, m \end{array} \right\}. \quad (47)$$

Notice that in contrast to (46), (47) is defined through *uncountably many* sets of constraints, and hence we need the following theorem to establish the tractability of the set  $Z^P$ .

**Theorem 11** Fix  $\alpha \in \mathcal{A}$ . The optimization problem with a linear objective function and the feasible set  $Z^P(\alpha)$  in (47) can be solved in polynomial time if for any fixed  $\mathbf{x}$ , the following optimization problem can be solved in polynomial time:

$$\begin{aligned} \min_{\mathbf{Y}_i \succeq 0, r \geq 0} \quad & -\operatorname{tr}\left(\sum_{i=1}^m \alpha_i \mathbf{Y}_i \mathbf{B}_i\right) + t(r) \sum_{i=1}^m \alpha_i \operatorname{tr}(\mathbf{Y}_i \mathbf{E}) \\ \text{s.t.} \quad & \sum_{i=0}^m \mathbf{Y}_i = (r+1)\mathbf{\Omega}, \operatorname{tr}(\mathbf{E} \sum_{i=1}^m \mathbf{Y}_i) = 1, \end{aligned} \quad (48)$$

where  $\mathbf{B}_i = \begin{pmatrix} Q_i(\mathbf{x}) & \frac{1}{2}y_i(\mathbf{x}) \\ \frac{1}{2}y_i(\mathbf{x})^\top & y_i^0(\mathbf{x}) \end{pmatrix}$  and  $\mathbf{E} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ .

*Proof* See Appendix E.  $\square$

Interestingly, Theorem 11 provides a tractability result for *individual* probabilistic envelope constraint.

**Corollary 7** If each component  $f_i(\cdot)$  of  $f(\cdot)$  is quadratic (and possibly non-convex), the optimization problem with a linear objective function and the probabilistic envelope constraint (6) can be solved in polynomial time if for any fixed  $\mathbf{x}$ , the following optimization problem can be solved in polynomial time:

$$\begin{aligned} \min_{\mathbf{Y} \succeq 0, r \geq 0} \quad & -\operatorname{tr}(\mathbf{YB}) + t(r) \\ \text{s.t.} \quad & \mathbf{Y} = (r+1)\mathbf{\Omega}, \operatorname{tr}(\mathbf{EY}) = 1, \end{aligned} \quad (49)$$

where  $\mathbf{B} = \begin{pmatrix} Q(\mathbf{x}) & \frac{1}{2}y(\mathbf{x}) \\ \frac{1}{2}y(\mathbf{x})^\top & y^0(\mathbf{x}) \end{pmatrix}$  and  $\mathbf{E} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ .

*Proof* When  $m = 1$ ,  $\alpha$  can be chosen as  $\alpha = 1$  without effect on the optimal solution of (48). Then (48) can be simplified as (49).  $\square$

Notice that Corollary 7 does not require that  $f_i(\cdot)$  is a convex quadratic function, and hence, subject to the price of a more complex formulation, is more general than Corollary 3 that investigates the probabilistic envelope constraint under convex quadratic uncertainty.

## 7 Simulations

In this section we illustrate two proposed approaches – chance constraint (5) and probabilistic envelope constraint (6) using the synthetic transportation problem discussed in Section 2.

We consider the transportation problem where the graph  $\mathcal{G}$  is a bi-parti graph between sources and destinations, i.e.,  $\mathcal{V} = \mathcal{S} \cup \mathcal{D}$  and  $\mathcal{E} = \{(s \rightarrow d) | s \in \mathcal{S}, d \in$

$\mathcal{D}$ }. Let  $m = |\mathcal{S}|$  and  $n = |\mathcal{D}|$ , then the unit cost  $\boldsymbol{\delta}$  is an  $m \times n$  matrix, and the transportation problem can be rewritten as

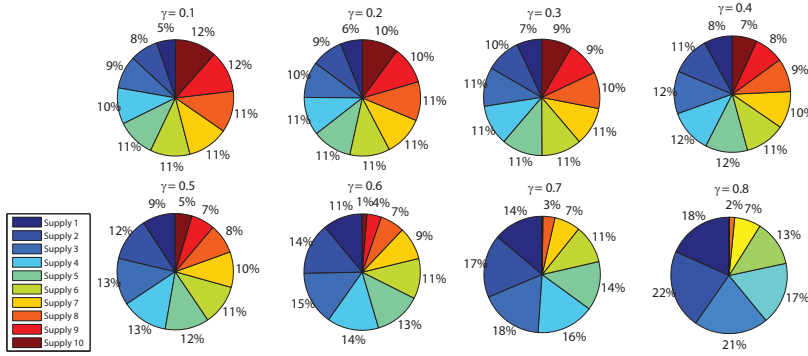
$$\begin{aligned} & \text{Maximize: } \mathbf{s} \geq 0, \mathbf{d} \geq 0 \quad z \\ & \text{Subject to: } \inf_{\boldsymbol{\delta} \sim \mathcal{P}} \mathbb{P}[-h(\mathbf{s}, \mathbf{d}, \boldsymbol{\delta}) \geq z] \geq 1 - \gamma; \\ & \quad \mathbf{1}_m^\top \mathbf{s} = \mathbf{1}_n^\top \mathbf{d} \geq L, \end{aligned}$$

where  $\boldsymbol{\delta} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{1}_m$  and  $\mathbf{1}_n$  are the all-one vectors with dimension  $m$  and  $n$  respectively. The function  $h(\mathbf{s}, \mathbf{d}, \boldsymbol{\delta})$  is defined by

$$\begin{aligned} h(\mathbf{s}, \mathbf{d}, \boldsymbol{\delta}) = & \text{Minimize: } \mathbf{F} \in \mathbb{R}^{m \times n} \quad \text{tr}\langle \boldsymbol{\delta}, \mathbf{F} \rangle \\ & \text{Subject to: } \mathbf{F}^\top \mathbf{1}_m = \mathbf{d}, \mathbf{F} \mathbf{1}_n = \mathbf{s}, \mathbf{F} \geq 0. \end{aligned}$$

By Theorem 2, one can solve this transportation problem by MATLAB and CVX [24]. We consider the case where there are 10 suppliers and 3 consumers, and the least demand  $L = 80$ . The mean  $M_{ij}$  and the variance  $\Sigma_{ij}$  of the transportation cost  $\delta_{ij}$  are set to  $100 + 0.1\sqrt{3(i-1) + j}$  and  $5/\sqrt{3(i-1) + j}$ , respectively. Then the transportation costs related to suppliers and consumers with lower serial numbers have smaller means but larger variances, i.e., lower mean cost but more risky.

Our first goal is to minimize the total cost to some fixed confidence parameter  $\gamma$ . Figure 1 shows the resulting allocations for different  $\gamma$ . As expected, small  $\gamma$  leads to



**Fig. 1** The transportation problem: the resulting allocations for different guarantees  $\gamma = 0.1 - 0.8$ .

more conservative allocations which tend to select supplies with higher mean costs and smaller variances, while large  $\gamma$  leads to less conservative allocations which select suppliers with lower mean costs and larger variances.

In this example, the algorithm takes about 40 seconds on a desktop PC with Intel i7 3.4GHz CPU and 8G memory. The computational time for solving the transportation problems of different numbers of suppliers is reported in Table 1. For a large-scale problem, i.e. the number of suppliers is 1000, our algorithm finds the result in about 30 minutes. From the table, it appears that the computation time scales roughly linearly with respect to the number of suppliers. Note that one can use commercial

Number of suppliers	10	50	100	200	500	1000
Running time (min)	0.88	3.33	4.01	6.21	15.16	34.78

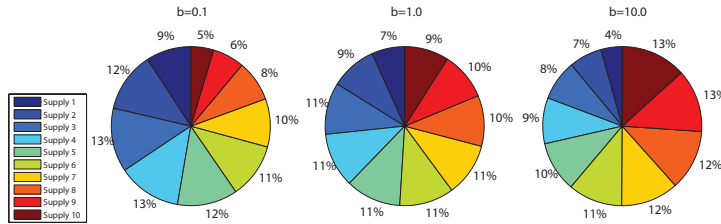
**Table 1** The running time for solving the transportation problem with different numbers of suppliers.

solvers such as CPLEX instead of CVX to implement this algorithm, which is typically more computationally efficient.

Using the same notations, the transportation problem with probabilistic envelope constraints can be formulated as

$$\begin{aligned}
& \text{Maximize:}_{\mathbf{s}, \mathbf{d}} \quad z \\
& \text{Subject to:} \quad \inf_{\boldsymbol{\delta} \sim \mathcal{P}} \mathbb{P}[-f(\mathbf{s}, \mathbf{d}, \boldsymbol{\delta}) \geq z - s] \geq \mathfrak{B}(s); \\
& \quad \quad \quad \mathbf{1}_m^\top \mathbf{s} = \mathbf{1}_n^\top \mathbf{d} \geq D; \\
& \quad \quad \quad \mathbf{s}, \mathbf{d} \geq 0.
\end{aligned}$$

Our second goal is to minimize the total cost subject to a decaying probabilistic envelope  $\mathfrak{B}(s) = 1 - 1/(1 + b\sqrt{s + a/b^2})$  which implies  $t(r) = \max\{(r^2 - a)/b^2, 0\}$  by Lemma 3. We choose  $a = 1$  and  $b = 0.1, 1.0, 10.0$ , giving different rates of decay for the probability the constraint is violated at level  $s$  for each  $s$ . Based on Theorem 4, we can easily solve this problem. Figure 2 shows the resulting allocations. Clearly,



**Fig. 2** The transportation problem: the resulting allocations for decay rates  $b = 0.1, 1.0$  and  $10.0$ .

larger  $b$  corresponds to a more risk averse attitude towards large constraint violation so that the resulting allocation is more conservative and tends to choose suppliers with larger mean costs and smaller variances.

## 8 Conclusion

The distributionally robust chance constraint formulation has been extensively studied. Yet, most previous work focused on the *linear* constraint function case. In this paper, motivated by applications where uncertainty is inherently *non-linear*, we investigate the computational aspects of distributionally robust chance constrained optimization problems for the general function case. We show that the distributionally robust chance constrained optimization is tractable, provided that the uncertainty



is characterized by its mean and variance, and the constraint function is concave-quasiconvex. This significantly expands the range of decision problems that can be modeled and solved efficiently via the DRCC framework. Along the way, we establish a relationship between the DRCC framework and robust optimization model, which links the stochastic model and the deterministic model of uncertainty. We then consider probabilistic envelope constraints, a generalization of distributionally robust chance constraint first proposed in Xu *et al.* [40], and extend this framework to the non-linear case, and obtain conditions that guarantee its tractability. Finally, we discuss two extensions of our approach, provide approximation schemes for JPEC, and establish conditions to ensure these approximation formulations are tractable.

**Acknowledgements** We thank the associate editor and two anonymous reviewers for their constructive comments which result in significant improvement of the paper. This work is supported by the Ministry of Education of Singapore through AcRF Tier Two grant R-265-000-443-112.

## Appendices

### A Proof of Corollary 1

For clarity, we denote  $-f(\mathbf{x}, \boldsymbol{\delta})$  and  $-\alpha$  by  $L_{\mathbf{x}}(\boldsymbol{\delta})$  and  $\beta$ , respectively. Since  $f(\mathbf{x}, \boldsymbol{\delta})$  is convex w.r.t.  $\boldsymbol{\delta}$  for fixed  $\mathbf{x}$ ,  $L_{\mathbf{x}}(\boldsymbol{\delta})$  is a concave function. Then the equivalence to establish can be rewritten as

$$\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[L_{\mathbf{x}}(\boldsymbol{\delta}) > \beta] \leq 1 - p \Leftrightarrow \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \text{CVaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) \leq \beta.$$

It is well known that  $\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \text{CVaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) \leq \beta \Rightarrow \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[L_{\mathbf{x}}(\boldsymbol{\delta}) > \beta] \leq 1 - p$ . Besides,  $\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \mathbb{P}[L_{\mathbf{x}}(\boldsymbol{\delta}) > \beta] > 1 - p \Leftrightarrow \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \text{VaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) > \beta$ , hence we only need to show that  $\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \text{CVaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) > \beta \Rightarrow \sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \text{VaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) > \beta$ .

Since  $\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \text{CVaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) > \beta$ , then there exists a probability distribution  $\mathbb{P}$  with zero mean, covariance  $\boldsymbol{\Sigma}$ , and  $\text{CVaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) > \beta$  when  $\boldsymbol{\delta} \sim \mathbb{P}$ . Decompose  $\mathbb{P} = \mu_1 + \mu_2$  where the measure  $\mu_1$  constitutes a probability of  $p$  and the measure  $\mu_2$  constitutes a probability of  $1 - p$ , and that  $L_{\mathbf{x}}(y_1) \leq L_{\mathbf{x}}(y_2)$  for any  $y_1$  and  $y_2$  that belong to the support of  $\mu_1$  and  $\mu_2$  respectively. By the CVaR constraint, we have  $(\int_{\boldsymbol{\delta}} L_{\mathbf{x}}(\boldsymbol{\delta}) d\mu_2) / (1 - p) > \beta$ .

We now construct a new probability  $\bar{\mathbb{P}}$  as follows: let  $\mu_2'$  be a measure that put a probability mass of  $1 - p$  on  $\int_{\boldsymbol{\delta}} \boldsymbol{\delta} d\mu_2 / (1 - p)$ , i.e., the conditional mean of  $\mu_2$ , and let  $\bar{\mathbb{P}} = \mu_1 + \mu_2'$ . Observe that  $\bar{\mathbb{P}}$  is a probability measure whose mean is the same as that of  $\mathbb{P}$ . Moreover, notice that  $\mu_2 / (1 - p)$  is a probability measure, by concavity of  $L_{\mathbf{x}}(\cdot)$  we have that

$$L_{\mathbf{x}}\left(\int_{\boldsymbol{\delta}} \boldsymbol{\delta} d\mu_2 / (1 - p)\right) \geq \left(\int_{\boldsymbol{\delta}} L_{\mathbf{x}}(\boldsymbol{\delta}) d\mu_2\right) / (1 - p) > \beta,$$

which implies that  $\text{VaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) > \beta$  for  $\boldsymbol{\delta} \sim \bar{\mathbb{P}}$ .

We now show that this also implies that  $\sup_{\boldsymbol{\delta} \sim (0, \boldsymbol{\Sigma})} \text{VaR}_{1-p}(L_{\mathbf{x}}(\boldsymbol{\delta})) > \beta$ . Denote the covariance w.r.t  $\bar{\mathbb{P}}$  by  $\bar{\boldsymbol{\Sigma}}$  and recall that both  $\mathbb{P}$  and  $\bar{\mathbb{P}}$  are zero mean, then

$$\begin{aligned} \boldsymbol{\Sigma} - \bar{\boldsymbol{\Sigma}} &= \int_{\boldsymbol{\delta}} \boldsymbol{\delta} \boldsymbol{\delta}^{\top} d\mathbb{P} - \int_{\boldsymbol{\delta}} \boldsymbol{\delta} \boldsymbol{\delta}^{\top} d\bar{\mathbb{P}} \\ &= \int_{\boldsymbol{\delta}} \boldsymbol{\delta} \boldsymbol{\delta}^{\top} d\mu_1 + \int_{\boldsymbol{\delta}} \boldsymbol{\delta} \boldsymbol{\delta}^{\top} d\mu_2 - \int_{\boldsymbol{\delta}} \boldsymbol{\delta} \boldsymbol{\delta}^{\top} d\mu_1 - \int_{\boldsymbol{\delta}} \boldsymbol{\delta} \boldsymbol{\delta}^{\top} d\mu_2' \\ &= \int_{\boldsymbol{\delta}} \boldsymbol{\delta} \boldsymbol{\delta}^{\top} d\mu_2 - (1 - p) \left[ \frac{\int_{\boldsymbol{\delta}} \boldsymbol{\delta} d\mu_2}{1 - p} \right] \left[ \frac{\int_{\boldsymbol{\delta}} \boldsymbol{\delta} d\mu_2}{1 - p} \right]^{\top} \\ &= \int_{\boldsymbol{\delta}} \left\{ \boldsymbol{\delta} - \left[ \frac{\int_{\boldsymbol{\delta}} \boldsymbol{\delta} d\mu_2}{1 - p} \right] \right\} \left\{ \boldsymbol{\delta} - \left[ \frac{\int_{\boldsymbol{\delta}} \boldsymbol{\delta} d\mu_2}{1 - p} \right] \right\}^{\top} d\mu_2 \succeq 0, \end{aligned}$$

where the third equality is due to the definition of  $\mu'_2$ . Note that from the construction of  $\bar{\mathbb{P}}$ , we have  $\sup_{\delta \sim (0, \bar{\Sigma})} \mathbb{P}[L_{\mathbf{x}}(\delta) > \beta] > 1 - p$ . Denote the set  $\{\delta | L_{\mathbf{x}}(\delta) > \beta\}$  by  $T_{\mathbf{x}}$ . First, we consider the case where  $\bar{\Sigma}$  is full rank. From Lemma 2, we have  $\inf_{\mathbf{y} \in T_{\mathbf{x}}} \mathbf{y}^\top \bar{\Sigma}^{-1} \mathbf{y} < r \triangleq p/(1-p)$ . Since  $\bar{\Sigma} \preceq \Sigma$  and  $\bar{\Sigma}$  is full rank,  $\inf_{\mathbf{y} \in T_{\mathbf{x}}} \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \leq \inf_{\mathbf{y} \in T_{\mathbf{x}}} \mathbf{y}^\top \bar{\Sigma}^{-1} \mathbf{y} < r$ , which implies that  $\sup_{\delta \sim (0, \Sigma)} \mathbb{P}[L_{\mathbf{x}}(\delta) > \beta] > 1 - p$ , which establishes the theorem.

The case where  $\bar{\Sigma}$  is not full rank requires additional work, as Lemma 2 or Theorem 1 can not be applied directly. Consider the spectral decomposition  $\bar{\Sigma} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top$  and denote the pseudo inverse of  $\bar{\Sigma}$  by  $\bar{\Sigma}^+$ . Suppose that the top  $d$  diagonal entries of  $\mathbf{\Lambda}$  are non-zero. Let  $\mathbf{Q}_d$  be the submatrix of  $\mathbf{Q}$  by selecting the first  $d$  columns of  $\mathbf{Q}$  and  $\mathbf{\Lambda}_d$  be the top  $d \times d$  submatrix of  $\mathbf{\Lambda}$ . Denote the column space of  $\bar{\Sigma}$  by  $\mathcal{C}$ , and let  $\mathcal{Q} \triangleq \{\mathbf{z} | \mathbf{z} = \mathbf{Q}_d^\top \delta, \forall \delta \in T_{\mathbf{x}} \cap \mathcal{C}\}$ . Since there is no uncertainty in  $\mathcal{C}^\perp$  w.r.t  $\bar{\mathbb{P}}$ ,

$$\sup_{\mathbf{z} \sim (0, \mathbf{\Lambda}_d)} \mathbb{P}[\mathbf{z} \in \mathcal{Q}] = \sup_{\delta \sim (0, \bar{\Sigma})} \mathbb{P}[\delta \in T_{\mathbf{x}} \cap \mathcal{C}] = \sup_{\delta \sim (0, \bar{\Sigma})} \mathbb{P}[\delta \in T_{\mathbf{x}}] > 1 - p.$$

From Lemma 2, we have  $\inf_{\mathbf{z} \in \mathcal{Q}} \mathbf{z}^\top \mathbf{\Lambda}_d^{-1} \mathbf{z} < r$ . In other words, there exists  $\mathbf{z} \in \mathcal{Q}$  such that  $\mathbf{z}^\top \mathbf{\Lambda}_d^{-1} \mathbf{z} < r$ , which implies that  $\mathbf{y}^\top \bar{\Sigma}^+ \mathbf{y} < r$  for  $\mathbf{y} \triangleq \mathbf{Q}_d \mathbf{z}$ . From the Schur complement, since  $\Sigma \succeq \bar{\Sigma} \succeq 0$ ,  $(\mathbf{I} - \bar{\Sigma} \bar{\Sigma}^+) \mathbf{y} = 0$  and  $r - \mathbf{y}^\top \bar{\Sigma}^+ \mathbf{y} = r - \mathbf{z}^\top \mathbf{\Lambda}_d^{-1} \mathbf{z} > 0$ , we have  $\begin{pmatrix} \Sigma & \mathbf{y} \\ \mathbf{y}^\top & r \end{pmatrix} \succeq \begin{pmatrix} \bar{\Sigma} & \mathbf{y} \\ \mathbf{y}^\top & r \end{pmatrix} \succeq 0$ . Hence  $\inf_{\mathbf{y} \in T_{\mathbf{x}}} \mathbf{y}^\top \Sigma^{-1} \mathbf{y} < r$ , which implies that  $\sup_{\delta \sim (0, \Sigma)} \mathbb{P}[L_{\mathbf{x}}(\delta) > \beta] > 1 - p$ .  $\square$

## B Proofs of Results in Section 3.2

### B.1 Proof of Corollary 2:

By Theorem 1, the feasible set  $S = \{\mathbf{x} | \mathbf{x}^\top g(\mathbf{y}) \geq \alpha, \forall \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \leq r\}$  where  $r = p/(1-p)$ . Hence, determining whether  $\mathbf{x} \in S$  is equivalent to determining whether the inner optimization problem  $\min_{\{\mathbf{y}^\top \Sigma^{-1} \mathbf{y} \leq r\}} \mathbf{x}^\top g(\mathbf{y}) - \alpha \geq 0$ . Rewrite the left hand side as an optimization problem on  $\mathbf{y}$ :

$$\begin{aligned} \text{Minimize: } & \mathbf{y}^\top G(\mathbf{x}) \mathbf{y} + P(\mathbf{x})^\top \mathbf{y} + Q(\mathbf{x}) \\ \text{Subject to: } & \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \leq r, \end{aligned} \quad (50)$$

by substituting  $g_i(\delta) = \delta^\top G_i \delta + \mathbf{p}_i^\top \delta + q_i$ . To prove Corollary 2, we need the following two results.

**Lemma 5** Fix  $\mathbf{x}$ . The optimal value of the optimization problem (50) equals that of the following SDP:

$$\text{Maximize: } \beta \geq 0, t, \quad \text{subject to: } \begin{pmatrix} \beta \Sigma^{-1} + G(\mathbf{x}) & \frac{1}{2} P(\mathbf{x}) \\ \frac{1}{2} P(\mathbf{x})^\top & Q(\mathbf{x}) - t - \beta r \end{pmatrix} \succeq 0. \quad (51)$$

*Proof* The dual problem of (50) is:  $\max_{\beta \geq 0} \min_{\mathbf{y}} \mathbf{y}^\top G(\mathbf{x}) \mathbf{y} + P(\mathbf{x})^\top \mathbf{y} + Q(\mathbf{x}) + \beta \mathbf{y}^\top \Sigma^{-1} \mathbf{y} - \beta r$ . By taking minimum over  $\mathbf{y}$  and the Schur complement, this can be reformulated as the SDP (51). Notice that there exists  $\mathbf{y}$  such that  $\mathbf{y}^\top \Sigma^{-1} \mathbf{y} < r$  since  $r > 0$ , hence Slater's condition is satisfied for (50), and the strong duality holds.  $\square$

Thus,  $\mathbf{x} \in S$  if and only if the optimal value of problem (50) is greater than or equal to 0. This means we can convert the constraint in  $S$  into a feasibility problem as follows:

**Lemma 6** Under the conditions of Corollary 2, and let  $r = p/(1-p)$ , we have the constraint

$$\inf_{\delta \sim (0, \Sigma)} \mathbb{P}[g(\delta)^\top \mathbf{x} \geq \alpha] \geq p, \quad (52)$$

is equivalent to the following problem

$$\text{Exist: } \beta \geq 0, \quad \text{s.t.: } \begin{pmatrix} \beta \Sigma^{-1} + G(\mathbf{x}) & \frac{1}{2} P(\mathbf{x}) \\ \frac{1}{2} P(\mathbf{x})^\top & Q(\mathbf{x}) - \beta r \end{pmatrix} \succeq 0. \quad (53)$$

*Proof* 1. Equation (52)  $\Rightarrow$  Equation (53): When Inequality (52) holds, the optimal value  $t$  of (50) must be greater than or equal to 0. So from Equation (51), we have

$$\begin{pmatrix} \beta \Sigma^{-1} + G(\mathbf{x}) & \frac{1}{2}P(\mathbf{x}) \\ \frac{1}{2}P(\mathbf{x})^T & Q(\mathbf{x}) - \beta r \end{pmatrix} \succeq \begin{pmatrix} \beta \Sigma^{-1} + G(\mathbf{x}) & \frac{1}{2}P(\mathbf{x}) \\ \frac{1}{2}P(\mathbf{x})^T & Q(\mathbf{x}) - t - \beta r \end{pmatrix} \succeq 0. \quad (54)$$

2. Equation (53)  $\Rightarrow$  Equation (52): Since the feasibility problem is solvable,  $t = 0$  must be a feasible solution of (51), which implies Inequality (52).  $\square$

Lemma 6 immediately implies Corollary 2.  $\square$

## C Proofs of Results in Section 4

### C.1 Proof of Lemma 3:

We now show that the constraints (13) and (14) are equivalent.

1. (13)  $\Rightarrow$  (14): Since  $\lim_{y \rightarrow +\infty} \mathfrak{B}(y)$  may not converge to 1, we define  $\mathfrak{B}^{-1}(x) = +\infty$  when  $\{y \geq 0 | \mathfrak{B}(y) \geq x\} = \emptyset$ . Then if  $r/(1+r)$  is not in the range of  $\mathfrak{B}(s)$ , we have  $t(r) = +\infty$  so that the constraint (14) is always satisfied. Otherwise, suppose that  $y^* = t(r) = \inf\{y \geq 0 | \mathfrak{B}(y) \geq r/(1+r)\}$ , then we have

$$\inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha - t(r)] = \inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha - y^*] \geq \mathfrak{B}(y^*) \geq \frac{r}{1+r}.$$

2. (14)  $\Rightarrow$  (13): Since  $\mathfrak{B}(y^*) \in [0, 1)$  for any  $y^* \geq 0$ , there exists  $r^*$  such that  $\mathfrak{B}(y^*) = \frac{r^*}{1+r^*}$ . From the definition of  $t(r)$ , we have  $y^* \geq t(r^*) = \inf\{y \geq 0 | \mathfrak{B}(y) \geq r^*/(1+r^*)\}$ . Hence the following inequality holds

$$\inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha - y^*] \geq \inf_{\delta \sim (0, \Sigma)} \mathbb{P}[f(\mathbf{x}, \delta) \geq \alpha - t(r^*)] \geq \frac{r^*}{1+r^*} = \mathfrak{B}(y^*).$$

Furthermore,  $t(\cdot)$  is non-decreasing since both  $r/(1+r)$  and  $\mathfrak{B}(\cdot)$  are non-decreasing. By definition of  $t$ ,  $\mathfrak{B}(0) \geq 0$  leads to  $t(0) = 0$ ; and  $\mathfrak{B}(s) < 1$  for all  $s > 0$  leads to  $\mathfrak{B}^{-1}(1) = +\infty$  and hence  $\lim_{r \uparrow +\infty} t(r) = +\infty$ . Also,  $\mathfrak{B}(0) > 0$  implies for some  $\epsilon > 0$ ,  $\mathfrak{B}(0) \geq \epsilon$ , and hence  $t(\epsilon) = 0$ . Thus,  $t(\cdot)$  is continuous at a neighborhood of 0.  $\square$

### C.2 Proof of Corollary 3:

The feasible set  $S = \{\mathbf{x} | \inf_{\delta \sim (0, \Sigma)} \mathbb{P}[g(\delta)^\top \mathbf{x} \geq \alpha - t(r)] \geq \frac{r}{1+r}; \forall r \geq 0\}$  admits

$$\begin{aligned} S &\stackrel{(a)}{=} \{\mathbf{x} | \forall (\mathbf{y}, r) \text{ such that } \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \leq r \Rightarrow g(\mathbf{y})^\top \mathbf{x} \geq \alpha - t(r)\} \\ &= \{\mathbf{x} | \min_{\{\mathbf{y}, r | \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \leq r\}} g(\mathbf{y})^\top \mathbf{x} + t(r) - \alpha \geq 0\}, \end{aligned}$$

where (a) holds by Theorem 3. As each component  $g_i(\mathbf{y})$  of  $g(\mathbf{y})$  is linear or quadratic, i.e.,  $g_i(\mathbf{y}) = \mathbf{y}^\top \mathbf{G}_i \mathbf{y} + \mathbf{p}_i^\top \mathbf{y} + q_i$ , for fixed  $\mathbf{x}$  the inner optimization problem  $\min_{\{\mathbf{y}, r | \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \leq r\}} g(\mathbf{y})^\top \mathbf{x} + t(r) - \alpha$  can be rewritten as:

$$\begin{aligned} &\text{Minimize: }_{r \geq 0, \mathbf{y}} \quad \mathbf{y}^\top G(\mathbf{x}) \mathbf{y} + P(\mathbf{x})^\top \mathbf{y} + Q(\mathbf{x}) + t(r) \\ &\text{subject to: } \quad \mathbf{y}^\top \Sigma^{-1} \mathbf{y} - r \leq 0, \end{aligned} \quad (55)$$

where  $G(\mathbf{x}) \triangleq \sum_{i=1}^n x_i \mathbf{G}_i$ ,  $P(\mathbf{x}) \triangleq \sum_{i=1}^n x_i \mathbf{p}_i$ , and  $Q(\mathbf{x}) \triangleq \sum_{i=1}^n x_i q_i - \alpha$ . Thus, in order to analyze  $S$ , we need to analyze the optimization problem (55). We have the following lemma:

**Lemma 7** For any fixed  $\mathbf{x}$ , the optimal value of problem (55) is equivalent to that of the following:

$$\text{Maximize: } \beta \geq 0, \eta, \quad \text{subject to: } \begin{pmatrix} \beta \Sigma^{-1} + G(\mathbf{x}) & \frac{1}{2} P(\mathbf{x}) \\ \frac{1}{2} P(\mathbf{x})^\top & Q(\mathbf{x}) - t^*(\beta) - \eta \end{pmatrix} \succeq 0, \quad (56)$$

where  $t^*(x)$  is the conjugate function of  $t(r)$  defined as  $t^*(x) = \sup_{r \geq 0} (xr - t(r))$ .

*Proof* By the Schur complement and the strong duality of problem (55) (Slater's condition holds by picking  $r = 1$  and  $\mathbf{y} = 0$ ), one can easily obtain this lemma.  $\square$

From Lemma 7, the constraint  $\inf_{\delta \sim (0, \Sigma)} \mathbb{P}[g(\delta)^\top \mathbf{x} \geq \alpha - t(r)] \geq \frac{r}{1+r}, \forall r \geq 0$ , is equivalent to a constraint that the optimal value of the optimization problem (56) is non-negative. Thus,  $\mathbf{x}$  belongs to the feasible set of the envelope constraint if and only if  $\alpha = 0$  is a feasible solution of (56), for the same  $\mathbf{x}$ . This means we can remove the  $-\alpha$  term from (56). That is, when each component  $g_i(\cdot)$  of  $g(\cdot)$  is linear or quadratic, the envelope constraint is equivalent to the following feasibility problem:

$$\text{exist: } \beta \geq 0, \quad \text{s.t.: } \begin{pmatrix} \beta \Sigma^{-1} + G(\mathbf{x}) & \frac{1}{2} P(\mathbf{x}) \\ \frac{1}{2} P(\mathbf{x})^\top & Q(\mathbf{x}) - t^*(\beta) \end{pmatrix} \succeq 0. \quad (57)$$

Hence the optimization problem (18) is equivalent to (21), which proves the first part of the Theorem.

To prove the second part of the Theorem, it suffices to show that Problem (21) can be solved in polynomial time. We show this by constructing a polynomial time separation oracle. For any  $(\beta, \mathbf{x})$ , if the optimization problem (22) can be solved in polynomial time, which implies  $t^*(\beta)$  can be computed in polynomial time, then it can be verified in polynomial time whether the constraint in (21) is satisfied or not, and hence the feasibility of  $(\beta, \mathbf{x})$  can be determined in polynomial time. Moreover, if  $(\beta_0, \mathbf{x}_0)$  is infeasible and let  $r_0$  be the optimal solution of the problem (22) (by assumption  $r_0$  can be found in polynomial time), then we have

$$\begin{pmatrix} \beta_0 \Sigma^{-1} + G(\mathbf{x}_0) & \frac{1}{2} P(\mathbf{x}_0) \\ \frac{1}{2} P(\mathbf{x}_0)^\top & Q(\mathbf{x}_0) + t(r_0) - \beta_0 r_0 \end{pmatrix} \not\succeq 0,$$

and we can find in polynomial time (e.g., by SVD) a vector  $(\mathbf{y}_0^\top, 1)$  such that

$$\begin{aligned} & (\mathbf{y}_0^\top, 1) \begin{pmatrix} \beta_0 \Sigma^{-1} + G(\mathbf{x}_0) & \frac{1}{2} P(\mathbf{x}_0) \\ \frac{1}{2} P(\mathbf{x}_0)^\top & Q(\mathbf{x}_0) + t(r_0) - \beta_0 r_0 \end{pmatrix} \begin{pmatrix} \mathbf{y}_0 \\ 1 \end{pmatrix} \\ & = (\mathbf{y}_0^\top \Sigma^{-1} \mathbf{y}_0 - r_0) \beta_0 + \mathbf{y}_0^\top G(\mathbf{x}_0) \mathbf{y}_0 + P(\mathbf{x}_0)^\top \mathbf{y}_0 + Q(\mathbf{x}_0) + t(r_0) < 0. \end{aligned}$$

Notice that for any feasible solution  $(\beta, \mathbf{x})$ , we must have

$$(\mathbf{y}_0^\top \Sigma^{-1} \mathbf{y}_0 - r_0) \beta + \mathbf{y}_0^\top G(\mathbf{x}) \mathbf{y}_0 + P(\mathbf{x})^\top \mathbf{y}_0 + Q(\mathbf{x}) + t(r_0) \geq 0.$$

Hence we have a separating hyperplane.  $\square$

## D Proofs of Results in Section 5

### D.1 Proof of Corollary 6:

As before, we construct a separation oracle to prove tractability. In order to verify the feasibility of a given  $\mathbf{x}^*$ , from Corollary 5 we know that  $\mathbf{x}^*$  is feasible if and only if the optimal value of the optimization problem (26) is greater than or equal to  $\alpha$ , which can be verified by directly solving Problem (26). By assumption, this can be done in polynomial time.

If  $\mathbf{x}^*$  is not feasible, then we can find in polynomial time  $(\mathbf{y}^*, r^*, \boldsymbol{\mu}^*, \Sigma^*)$  such that  $f(\mathbf{x}^*, \mathbf{y}^* + \boldsymbol{\mu}^*) + t(r^*) < \alpha$ . Because  $f(\mathbf{x}, \mathbf{y} + \boldsymbol{\mu})$  is concave in  $\mathbf{x}$  for fixed  $\mathbf{y}$  and  $\boldsymbol{\mu}$ , for any feasible  $\mathbf{x}$ , we have

$$f(\mathbf{x}^*, \mathbf{y}^* + \boldsymbol{\mu}^*) + \nabla f(\mathbf{x}^*, \mathbf{y}^* + \boldsymbol{\mu}^*)^\top (\mathbf{x} - \mathbf{x}^*) + t(r^*) \geq f(\mathbf{x}, \mathbf{y}^* + \boldsymbol{\mu}^*) + t(r^*) \geq \alpha.$$

Hence the hyperplane separating  $\mathbf{x}^*$  from the feasible set is the following:

$$f(\mathbf{x}^*, \mathbf{y}^* + \boldsymbol{\mu}^*) + \nabla f(\mathbf{x}^*, \mathbf{y}^* + \boldsymbol{\mu}^*)^\top (\mathbf{x} - \mathbf{x}^*) + t(r^*) \geq \alpha, \quad (58)$$

which can be generated in polynomial time since the super-gradient of  $\mathbf{x}$  can be obtained in polynomial time.  $\square$

## D.2 Proof of Theorem 5:

If  $f(\mathbf{x}, \delta) \geq \alpha$  for all  $\delta \in \mathcal{S}$ , the constraints (27) and (28) are satisfied, so we only need to consider the case where there exists  $\delta \in \mathcal{S}$  such that  $f(\mathbf{x}, \delta) < \alpha$ . Note that (27) is equivalent to  $\sup_{\delta \sim (\mu, \mathcal{S})} \mathbb{P}[f(\mathbf{x}, \delta) < \alpha] \leq 1 - p$ , then we can apply the following lemma:

**Lemma 8** *If the conditions in Theorem 5 hold and  $\{\delta : f(\mathbf{x}, \delta) < \alpha\}$  is nonempty, then*

$$\sup_{\delta \sim (\mu, \mathcal{S})} \mathbb{P}[f(\mathbf{x}, \delta) < \alpha] \leq 1 - p \quad (59)$$

is equivalent to

$$1 - p \geq \begin{cases} \sup_{\theta, \delta_1, \delta_2} \theta \\ \text{such that } \theta \delta_1 + (1 - \theta) \delta_2 = \mu, \\ 0 \leq \theta \leq 1, \\ f(\mathbf{x}, \delta_1) < \alpha, \\ \delta_1, \delta_2 \in \mathcal{S}. \end{cases} \quad (60)$$

*Proof* Since  $\mu \in \mathcal{S}$  and  $\{\delta : f(\mathbf{x}, \delta) < \alpha\}$  is not empty, the optimization problem in (60) is always feasible. To show the equivalence of (59) and (60), one needs to prove that the optimal objective value  $\theta^*$  of the optimization problem in (60) equals  $\zeta = \sup_{\delta \sim (\mu, \mathcal{S})} \mathbb{P}[f(\mathbf{x}, \delta) < \alpha]$ .

The first step is to show  $\theta^* \leq \zeta$ : Since  $f(\mathbf{x}, \cdot)$  is continuous for fixed  $\mathbf{x} \in \mathbb{X}$  and  $\mathcal{S}$  is a closed convex set, for any  $\epsilon > 0$  there exists a feasible solution  $(\delta'_1, \delta'_2, \theta')$  such that  $|\theta' - \theta^*| < \epsilon$ . Construct a probability distribution  $\mathbb{P}'(\mathbf{x})$  such that  $\delta = \delta'_1$  with probability  $\theta'$  and  $\delta = \delta'_2$  with probability  $1 - \theta'$ , then we have  $\mathbb{P}' \in (\mu, \mathcal{S})$ . By construction we have  $\theta' \leq \mathbb{P}'[f(\mathbf{x}, \delta) < \alpha] \leq \zeta$  where the second inequality holds from  $\mathbb{P}' \in (\mu, \mathcal{S})$ . Thus we have  $\theta^* \leq \zeta$  as  $\epsilon$  can be arbitrarily small.

The second step is to prove  $\theta^* \geq \zeta$ : Consider any probability distribution  $\mathbb{P} \in (\mu, \mathcal{S})$ , and define  $\bar{\theta} = \mathbb{P}[f(\mathbf{x}, \delta) < \alpha]$ ,  $\bar{\delta}_1 = \mathbb{E}_{\mathbb{P}}[\delta | f(\mathbf{x}, \delta) < \alpha]$  and  $\bar{\delta}_2 = \mathbb{E}_{\mathbb{P}}[\delta | f(\mathbf{x}, \delta) \geq \alpha]$ . We then have  $\bar{\delta}_1 \bar{\theta} + \bar{\delta}_2 (1 - \bar{\theta}) = \mu$ ,  $f(\mathbf{x}, \bar{\delta}_1) < \alpha$  and  $\bar{\delta}_1, \bar{\delta}_2 \in \mathcal{S}$ , or equivalently  $(\bar{\delta}_1, \bar{\delta}_2, \bar{\theta})$  is a feasible solution of the optimization problem in (60). Thus, we must have  $\mathbb{P}[f(\mathbf{x}, \delta) < \alpha] = \bar{\theta} \leq \theta^*$ , which implies that  $\theta^* \geq \zeta = \sup_{\delta \sim (\mu, \mathcal{S})} \mathbb{P}[f(\mathbf{x}, \delta) < \alpha]$ . Therefore, (59) is equivalent to (60).  $\square$

From the equivalence shown in Lemma 8, we consider the following feasibility problem parameterized by  $\theta \in [0, 1]$ , denoted  $\mathcal{F}_\theta$ :

$$\begin{aligned} &\text{exist: } \delta_1, \delta_2 \\ &\text{such that: } \theta \delta_1 + (1 - \theta) \delta_2 = \mu, \\ & \quad f(\mathbf{x}, \delta_1) < \alpha, \\ & \quad \delta_1, \delta_2 \in \mathcal{S}. \end{aligned}$$

Then we have that for any  $0 \leq \theta_1 \leq \theta_2 \leq 1$ ,  $\mathcal{F}_{\theta_2}$  being feasible implies  $\mathcal{F}_{\theta_1}$  being feasible. To see this, let  $(\delta_1^*, \delta_2^*)$  be a feasible solution to  $\mathcal{F}_{\theta_2}$ . Hence we have  $\theta_2 \delta_1^* + (1 - \theta_2) \delta_2^* = \mu$ . Let  $\delta'_2$  be such that

$$\mu - \delta'_2 = (\mu - \delta_2^*) \times \frac{(1 - \theta_2) \theta_1}{(1 - \theta_1) \theta_2}.$$

Since  $\theta_2 \geq \theta_1$ , we have that  $\delta'_2$  is on the line segment between  $\mu$  and  $\delta_2^*$ , and hence belongs to  $\mathcal{S}$  by its convexity. Furthermore, it is easy to check that  $(\delta_1^*, \delta'_2)$  is feasible to  $\mathcal{F}_{\theta_1}$ .

Thus, constraint (60) (and equivalently the chance constraint (27)) is equivalent to  $\mathcal{F}_{1-p+\epsilon}$  infeasible for all  $\epsilon > 0$ , i.e.,

$$\delta_2 = -\frac{1-p+\epsilon}{p-\epsilon}(\delta_1 - \mu) + \mu \notin \mathcal{S}, \quad \forall f(\mathbf{x}, \delta_1) < \alpha \text{ and } \delta_1 \in \mathcal{S}.$$

This further implies  $\mathcal{F}_{1-p}$  is infeasible, i.e.,

$$\delta_2 = -\frac{1-p}{p}(\delta_1 - \mu) + \mu \notin \mathcal{S}, \quad \forall f(\mathbf{x}, \delta_1) < \alpha \text{ and } \delta_1 \in \mathcal{S}. \quad (61)$$

To see this, we only need to show that  $\mathcal{F}_{1-p}$  being feasible implies that  $\mathcal{F}_{1-p+\varepsilon}$  is feasible for some  $\varepsilon > 0$ . Suppose that there exists  $\delta_1^* \in \mathcal{S}$  such that  $\delta_2^* = -\frac{1-p}{p}(\delta_1^* - \mu) + \mu \in \mathcal{S}$  and  $f(\mathbf{x}, \delta_1^*) < \alpha$ . By continuity of  $f(\mathbf{x}, \cdot)$ , we have that for a sufficiently small  $\eta > 0$ ,  $f(\mathbf{x}, \delta_1') < \alpha$  where  $\delta_1' \triangleq (1-\eta)\delta_1^* + \eta\mu$ . Note that  $\delta_1' \in \mathcal{S}$  and there exists  $\varepsilon > 0$  such that  $-\frac{1-p+\varepsilon}{p-\varepsilon}(\delta_1' - \mu) + \mu \in \mathcal{S}$ , which implies that  $\mathcal{F}_{1-p+\varepsilon}$  is feasible.

Finally, the constraint (61) can be rewritten as

$$0 < \min_{\delta_1, \delta_2} \|(1-p)\delta_1 + p\delta_2 - \mu\|_2 \text{ s.t. } f(\mathbf{x}, \delta_1) < \alpha, \delta_1 \in \mathcal{S}, \delta_2 \in \mathcal{S}, \quad (62)$$

which is equivalent to (28). Therefore, the theorem follows.  $\square$

## E Proofs of Results in Section 6

### E.1 Proof of Theorem 11:

The constraints in  $Z^P$  (47) requires that for any  $r \geq 0$ , we can find  $\beta$  and  $\mathbf{M}$  to satisfy  $\beta + (r+1)\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0$  and the other  $(m+1)$  semi-definite constraints. This is equivalent to requiring that the following optimization problem has an optimal value less than or equal to 0 (notice that for any  $r \geq 0$ , finding  $\beta$  and  $\mathbf{M}$  to satisfy the  $(m+1)$  semi-definite constraints itself is trivial):

$$\begin{aligned} \max_{r \geq 0} \min_{\mathbf{M} \succeq 0, \beta} & \beta + (r+1)\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} & \mathbf{M} - \begin{pmatrix} \alpha_i Q_i(\mathbf{x}) & \frac{1}{2} \alpha_i y_i(\mathbf{x}) \\ \frac{1}{2} \alpha_i y_i(\mathbf{x})^\top & \alpha_i (y_i^0(\mathbf{x}) - t(r)) - \beta \end{pmatrix} \succeq 0 \quad \forall i = 1, \dots, m. \end{aligned} \quad (63)$$

We analyze this requirement using duality. In order to find the dual problem of (63), it is more convenient for us to analyze the following problem:

$$\begin{aligned} \min_{r \geq 0} \max_{\mathbf{M} \succeq 0, \beta} & -\beta - (r+1)\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} & \mathbf{M} - \begin{pmatrix} \alpha_i Q_i(\mathbf{x}) & \frac{1}{2} \alpha_i y_i(\mathbf{x}) \\ \frac{1}{2} \alpha_i y_i(\mathbf{x})^\top & \alpha_i (y_i^0(\mathbf{x}) - t(r)) - \beta \end{pmatrix} \succeq 0 \quad \forall i = 1, \dots, m. \end{aligned} \quad (64)$$

Consider the dual problem, the ‘‘max’’ part in (64) is equivalent to

$$L(r) = \min_{\lambda_i \geq 0} \max_{\beta, \mathbf{M}} -\beta - (r+1)\langle \boldsymbol{\Omega}, \mathbf{M} \rangle + \sum_{i=1}^m \lambda_i \lambda_{\min}(\mathbf{M} - \mathbf{S}_i + \beta \mathbf{E}) + \lambda_0 \lambda_{\min}(\mathbf{M}), \quad (65)$$

where the function  $\lambda_{\min}(\mathbf{X})$  denotes minimum eigenvalue of matrix  $\mathbf{X}$ , and  $\mathbf{S}_i \triangleq \alpha_i \mathbf{B}_i - \alpha_i t(r) \mathbf{E}$ . Further note that the function  $\lambda_{\min}(\mathbf{X})$  is equivalent to the following optimization problem:  $\min_{\mathbf{Y} \succeq 0, \text{tr}(\mathbf{Y})=1} \text{tr}(\mathbf{Y}\mathbf{X})$ . Thus (65) is equivalent to

$$L(r) = \min_{\lambda_i \geq 0} \max_{\beta, \mathbf{M}} \min_{\{\mathbf{Y}_i | \text{tr}(\mathbf{Y}_i)=1, \mathbf{Y}_i \succeq 0\}} -\beta - (r+1)\langle \boldsymbol{\Omega}, \mathbf{M} \rangle + \text{tr}(\lambda_0 \mathbf{Y}_0 \mathbf{M}) + \sum_{i=1}^m \text{tr}(\lambda_i \mathbf{Y}_i (\mathbf{M} - \mathbf{S}_i + \beta \mathbf{E})).$$

Notice that for any fixed  $\lambda$ , the objective function is continuous, convex w.r.t.  $(\mathbf{Y}_i)_{i=0}^m$  and concave w.r.t.  $(\beta, \mathbf{M})$ . Moreover, the feasible set of  $(\mathbf{Y}_i)_{i=0}^m$  is compact and does not depend on  $(\beta, \mathbf{M})$ . Hence Sion’s minimax theorem applies, and we have

$$L(r) = \min_{\lambda_i \geq 0} \min_{\{\mathbf{Y}_i | \text{tr}(\mathbf{Y}_i)=\lambda_i, \mathbf{Y}_i \succeq 0\}} \max_{\beta, \mathbf{M}} -\sum_{i=1}^m \text{tr}(\mathbf{Y}_i \mathbf{S}_i) + \langle \mathbf{M}, \sum_{i=0}^m \mathbf{Y}_i - (r+1)\boldsymbol{\Omega} \rangle + \beta \langle \mathbf{E}, \sum_{i=1}^m \mathbf{Y}_i - 1 \rangle.$$

Taking maximum over  $\beta$  and  $M$ , we have that  $L(r)$  is equivalent to the following optimization problem with variables  $\mathbf{Y}_i$  and  $\lambda_i$ :

$$L(r) = \min_{\lambda_i \geq 0} \min_{\{\mathbf{Y}_i | \text{tr}(\mathbf{Y}_i) = \lambda_i, \mathbf{Y}_i \succeq 0\}} - \sum_{i=1}^m \text{tr}(\alpha_i \mathbf{Y}_i (\mathbf{B}_i - t(r) \mathbf{E}))$$

$$s.t. \quad \sum_{i=0}^m \mathbf{Y}_i = (r+1) \mathbf{\Omega}, \quad \text{tr}(\mathbf{E} \sum_{i=1}^m \mathbf{Y}_i) = 1.$$

By taking minimum over  $\lambda_i$ ,  $\min_{r \geq 0} L(r)$  can be further reformulated as (48). Hence from the analysis above, we know that (64) is equivalent to (48). To complete the proof, we construct a separation oracle of  $Z^P$  based on (48). Given  $\mathbf{x}$ , if the optimization problem (48) can be solved in polynomial time, then it can be verified whether  $\mathbf{x} \in Z^P$  or not in polynomial time since  $\mathbf{x}$  is feasible if and only if the optimal value of (48) is greater than or equal to 0. Furthermore, if  $\mathbf{x} \notin Z^P$ , let the optimal solution of (48) be  $(r^0, \{\mathbf{Y}_i^0\})$ , then we have  $-\sum_{i=1}^m \text{tr}(\alpha_i \mathbf{Y}_i^0 \mathbf{B}_i) + t(r^0) \sum_{i=1}^m \alpha_i \text{tr}(\mathbf{Y}_i^0 \mathbf{E}) < 0$  since  $\mathbf{x} \notin Z^P$ . On the other hand, for any  $\mathbf{x} \in Z^P$ , the following inequality must be satisfied

$$-\sum_{i=1}^m \text{tr}(\alpha_i \mathbf{Y}_i^0 \mathbf{B}_i) + t(r^0) \sum_{i=1}^m \alpha_i \text{tr}(\mathbf{Y}_i^0 \mathbf{E}) \geq 0,$$

which implies that a separating hyperplane can be generated in polynomial time.  $\square$

## References

1. Alizadeh, F., Goldfarb, D.: Second-order cone programming. *Mathematical Programming* **95**(1), 3–51 (2003)
2. Ben-Tal, A., Bertsimas, D., Brown, D.: A soft robust model for optimization under ambiguity. *Operations Research* **58**, 1220–1234 (2010)
3. Ben-Tal, A., Boyd, S., Nemirovski, A.: Extending scope of robust optimization: comprehensive robust counterparts of uncertain problems. *Mathematical Programming, Series B* **107**, 63–89 (2006)
4. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: *Robust Optimization*. Princeton University Press (2009)
5. Ben-Tal, A., den Hertog, D., Vial, J.P.: Deriving robust counterparts of nonlinear uncertain inequalities. *CentER Discussion Paper Series No. 2012-053* (2012)
6. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. *Mathematics of Operations Research* **23**, 769–805 (1998)
7. Ben-Tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. *Operations Research Letters* **25**(1), 1–13 (1999)
8. Bertsimas, D., Sim, M.: The price of robustness. *Operations Research* **52**(1), 35–53 (2004)
9. Black, F., Scholes, M.: The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637–659 (1973)
10. Calafiore, G., Ghaoui, L.E.: Distributionally robust chance constrained linear programs with applications. *Journal of Optimization Theory and Applications* **130**(1), 1–22 (2006)
11. Charnes, A., Cooper, W.W.: Chance constrained programming. *Management Science* **6**, 73–79 (1959)
12. Chen, W., Sim, M., Sun, J., Teo, C.: From CVaR to uncertainty set: implications in joint chance constrained optimization. *Operations Research* **58**(2), 470–485 (2010)
13. Cheng, J., Delage, E., Lissner, A.: Distributionally robust stochastic knapsack problem. working draft (2013)
14. Cheung, S.S., So, A.M., Wang, K.: Linear matrix inequalities with stochastically dependent perturbations and applications to chance-constrained semidefinite optimization. *SIAM Journal on Optimization* **22**(4), 1394C–1430 (2012)
15. Cont, R.: Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* **1**(2), 223–236 (2001)
16. Delage, E., Mannor, S.: Percentile optimization for Markov decision processes with parameter uncertainty. *Operations Research* **58**(1), 203–213 (2010)
17. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with applications to data-driven problems. *Operations Research* **58**(1), 595–612 (2010)

18. Dentcheva, D., Ruszczyński, A.: Optimization with stochastic dominance constraints. *SIAM Journal on Optimization* **14**(2), 548–566 (2003)
19. Dentcheva, D., Ruszczyński, A.: Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming, Series A* **99**(2), 329–350 (2004)
20. Dentcheva, D., Ruszczyński, A.: Semi-infinite probabilistic optimization: first order stochastic dominance constraints. *Optimization* **53**(5-6), 433–451 (2004)
21. Dupacová, J.: The minimax approach to stochastic programming and an illustrative application. *Stochastics* **20**, 73–88 (1987)
22. Erdogvan, E., Iyengar, G.: Ambiguous chance constrained problems and robust optimization. *Mathematical Programming, Ser. B* **107**, 37–61 (2006)
23. Fama, E.: The behavior of stock prices. *Journal of Business* **38**, 34–105 (1965)
24. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 1.21. <http://cvxr.com/cvx> (2011)
25. Grötschel, M., Lovász, L., Schrijver, A.: Geometric algorithms and combinatorial optimization, vol. 2. Springer (1988)
26. Jansen, D., deVries, C.: On the frequency of large stock returns: Putting booms and busts into perspective. *Review of Economics and Statistics* **73**(1), 18–24 (1991)
27. Kawas, B., Thiele, A.: A log-robust optimization approach to portfolio management. *OR Spectrum* **33**, 207–233 (2011)
28. Kawas, B., Thiele, A.: Short sales in log-robust portfolio management. *European Journal of Operations Research* **215**, 651–661 (2011)
29. Kon, S.: Models of stock returns - a comparison. *Journal of Finance* **39**, 147–65 (1984)
30. Marshall, A.W., Olkin, I.: Multivariate Chebyshev inequalities. *Annals of Mathematical Statistics* **31**(4), 1001–1014 (1960)
31. Miller, B., Wagner, H.: Chance-constrained programming with joint constraints. *Operations Research* **13**, 930–945 (1965)
32. Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. *SIAM Journal on Optimization* **17**(4), 969–996 (2006)
33. Popescu, I.: Robust mean-covariance solutions for stochastic optimization. *Operations Research* **55**(1), 98–112 (2007)
34. Prékopa, A.: On probabilistic constrained programming. In *Proceedings of the Princeton Symposium on Mathematical Programming* pp. 113–138 (1970)
35. Prékopa, A.: *Stochastic programming*. Kluwer (1995)
36. Scarf, H.: A min-max solution of an inventory problem. In: *Studies in Mathematical Theory of Inventory and Production*, pp. 201–209. Stanford University Press (1958)
37. Shapiro, A., Dentcheva, D., Ruszczyński, A.: *Lectures on stochastic programming: modeling and theory*. SIAM, Philadelphia (2009)
38. Shivaswamy, P., Bhattacharyya, C., Smola, A.: Second order cone programming approaches for handling missing and uncertain data. *Journal of Machine Learning Research* **7**, 1283–314 (2006)
39. Wiesemann, W., Kuhn, D., Sim, M.: *Distributionally robust convex optimization* (2013)
40. Xu, H., Caramanis, C., Mannor, S.: Optimization under probabilistic envelope constraints. *Operations Research* **60**, 682–700 (2012)
41. Zymmler, S., Kuhn, D., Rustem, B.: Distributionally robust joint chance constraints with second-order moment information. *Forthcoming, Mathematical Programming* (2011)
42. Zymmler, S., Kuhn, D., Rustem, B.: Worst-case value at risk of nonlinear portfolios. *Forthcoming, Management Science* (2012)