OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0030-364X | EISSN 1526-5463 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000
© 0000 INFORMS

Optimization under Probabilistic Envelope Constraints

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

Constantine Caramanis
Department of Electrical and Computer Engineering, The University of Texas at Austin, USA, caramanis@mail.utexas.edu,

Shie Mannor
Department of Electrical Engineering, Technion, Israel, shie.mannor@ee.technion.ac.il,

Chance constraints are an important modeling tool in stochastic optimization, providing probabilistic guarantees that a solution "succeeds" in satisfying a given constraint. While they control the probability of "success," they provide no control whatsoever in the event of a "failure." That is, they do not distinguish between a slight over- or under-shoot of the bounds, and more catastrophic violation. In short, they do not capture the *magnitude* of violation of the bounds. This paper addresses precisely this topic, focusing on linear constraints and ellipsoidal (Gaussian-like) uncertainties. We show that the problem of requiring different probabilistic guarantees at each level of constraint violation can be reformulated as a semi-infinite optimization problem. We provide conditions that guarantee polynomial-time solvability of the resulting semi-infinite formulation. We show further that this resulting problem is what has been called a *comprehensive robust optimization* problem in the literature. As a byproduct, we provide tight probabilistic bounds for comprehensive robust optimization. Thus, analogously to the connection between chance constraints and robust optimization, we provide a broader connection between probabilistic envelope constraints and comprehensive robust optimization.

Subject classifications: Programming: Stochastic; Statistics: Nonparametric

1. Introduction

An important paradigm for handling stochastic parameter uncertainty in optimization, is the so-called chance constraint paradigm. Here, a deterministic constraint is relaxed, and required to hold with at least some specified probability. Thus, given a constraint $f(\mathbf{x}, \delta^r)$ where $\mathbf{x}$ denotes the decision variable, and δ^r the stochastic uncertainty (we add the superscript r to emphasize δ^r is a random variable), one solves:

$$\mathbb{P}(f(\mathbf{x}, \delta^r) \geq \alpha) \geq p, \tag{1}$$

for some value $p \in (0, 1)$ and target α. Chance constraints date at least as far back as (e.g., Charnes and Cooper 1959), and since then there has been considerable work, e.g., Miller and Wagner (1965),

Prékopa (1970), Delage and Mannor (2010) and many others; we refer the reader to the textbook Prékopa (1995) and reference therein for a thorough review.

The chance constraint formulation in (1) guarantees that the given constraint will be satisfied with probability p. With the remaining $(1-p)$ probability, the constraint is violated, and no control whatsoever is provided on the degree of violation. In many important practical applications, the decision-maker may not be indifferent to the degree of constraint violation (cf Payne et al. 1980, 1981, Chen and Sim 2009). The *example par excellence* is portfolio optimization. Here, the decision maker may enforce a chance constraint that with a certain confidence the portfolio achieves a target value. Yet the behavior when that target is not met is arguably equally important, as the investor is also interested in knowing and perhaps bounding how bad the return can be, in case the portfolio fails to achieve the targeted return. Neglecting the magnitude of constraint violation is particularly problematic when the uncertain parameter follows a heavy-tail distribution, as is often the case in financial applications.

One natural remedy to this shortcoming of chance constrains is to enforce different levels of probabilistic guarantees. Thus our investor might require the portfolio return to achieve target α_1 with probability at least 50%, target α_2 with probability at least 90%, and α_3 with probability at least 99%, hence providing a variety of hedging guarantees for when the primary target, α_1, is not achieved. This multiple-chance-constraint idea can be easily generalized to any finite number (say N) of levels of protection, resulting to a set of N chance constraints. However, the computational effort of solving such a problem increases (and in fact does so super-linearly) as N increases. The failure of this naïve implementation is not an indictment of the idea. What is missing is capturing structure between multiple chance constraints at different levels of protection. We tackle precisely this problem.

In this paper we propose what we call the *probabilistic envelope constraint* framework that generalizes chance constraints. Instead of requiring probabilistic guarantees for a single or even a *finite number* of target values (i.e., constraint violation), as in the multiple-chance-constraint setup, we enforce chance constraints at *all* levels of potential violation. Thus, the single chance constraint in (1) above, becomes the following infinite set of chance constraints:

$$\mathbb{P}(f(\mathbf{x},\delta^r) \geq \alpha - s) \geq \mathfrak{P}(s), \quad \forall s \geq 0,$$

where $\mathfrak{P}(s)$ is a given non-decreasing function of s. Thus, we guarantee that the entire tail behavior is bounded by an envelope function.

While the probabilistic envelope constraint framework can be general, all technical results derived in this paper focus on the following specialized setup. Detailed and precise definitions can be found in subsequent sections.

Condition 1 (General Setup).

- $f(\cdot)$ is linear and δ^r is additive, i.e., the constraint has a form

$$\mathbb{P}\left[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \geq b - s\right] \geq \mathfrak{P}(s), \quad \forall s \geq 0.$$

- $\boldsymbol{\delta}^r$ is a random variable with an ellipsoidal distribution (e.g., Gaussian); alternatively, the distribution of $\boldsymbol{\delta}^r$ is unknown, and only its mean and variance are known.

Unlike the naïve multi-chance-constraint formulation, we show that optimization under probabilistic envelope constraints is often computational friendly. This is because the (infinite) collection of chance constraints expressed, can be dealt with directly. Indeed, we show that under Condition 1, the envelope constraint can be converted into an easier to analyze *deterministic* semi-infinite program. We then give further conditions under which the resulting semi-infinite program can be solved in polynomial time. One sufficient condition (in addition to Condition 1) for tractability is as follows:

Condition 2 (Tractability).

- $\boldsymbol{\delta}^r$ has a log-concave distribution.
- $\mathfrak{P}(s) = 1 - \gamma \exp(-g(s))$ for an increasing, concave $g(\cdot)$

Thus, under these conditions, even though the computational cost of the multiple-chance-constraint formulation increases super-linearly with the number of levels of protection, the continuum limit results in a tractable problem.

Converting the probabilistic envelope program into a semi-infinite *deterministic* program provides an interesting link between problems with stochastic models of uncertainty, and deterministic problems with *set-based* uncertainty. The latter class of problems have attracted much attention in the last decade, under the banner of Robust Optimization (e.g., Ben-Tal and Nemirovski 1998, 1999, 2000, El Ghaoui et al. 1998, Bertsimas and Sim 2004, Ben-Tal et al. 2009, Bertsimas et al. 2011). In particular, we show that probabilistic envelope constraints can be transformed into semi-infinite constraints that in turn can be re-written as a *comprehensive robust optimization* problem (Ben-Tal et al. 2006, 2010). As we describe in further detail below, comprehensive robustness is an extension of robust optimization that provides different levels of deterministic protection

against different magnitudes of uncertainty. Thus, as a by-product of this connection, we present a probabilistic interpretation of comprehensive robust optimization.

Readers familiar with stochastic programming literature may recognize that the probabilistic chance constraint formulation is closely related to the *stochastic dominance constraints* (Dentcheva and Ruszczyński 2003, 2004a,b); see Chapter 4 of Shapiro et al. (2009) and reference therein for more details. 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. Indeed, a probabilistic chance constraint can be formulated as a first-order stochastic dominance constraint. However, most of the literature in optimization with stochastic dominance constraints focuses on the second (or higher) order constraints case, a case that preserves convexity and is more amenable to analysis.

It is worth pointing out that the probabilistic envelope program enforces the desired probabilistic requirements in the *design stage*. This is in contrast to the post-analysis approach (e.g., Paschalidis and Kang 2005), where one obtains a solution using alternative methods such as robust optimization or standard chance constraints, and then analyzes the tail probability of the constraint violation for the obtained solution. Recently, there has been some work on robust optimization that takes into account the probabilistic requirements on the solution, e.g., (Chen et al. 2007, 2008, Bertsimas and Brown 2009). These papers construct uncertainty sets such that the obtained solution is guaranteed to satisfy certain probabilistic requirements. In general, this approach seems to lead to *sufficient* but not *necessary* conditions for the desired probabilistic requirements. In contrast, we begin with the desired probabilistic envelope constraint, and subsequently show its equivalence to the semi-infinite deterministic formulation.

Finally, we comment on an alternative approach to capture constraints about the magnitude of the losses or gains: building stochastic optimization problems using risk measures. Considerable work has been done in pursuing this avenue, particularly for portfolio optimization (see, e.g., Artzner et al. 1999, Delbaen 2002, Novosyolov 2002, Ruszczynski and Shapiro 2006, Rockafellar and Uryasev 2000, Lüthi and Doege 2005, Föllmer and Schied 2002, El Ghaoui et al. 2003, Brown and Sim 2009, Ben-Tal et al. 2010). Most of work along this line of research is based on optimizing over an utility function or a mean risk functional. This allows for the modeling of refined risk preferences, beyond what simple chance constraints offer. However, they do not directly offer protection in the form of probability guarantees against losses exceeding some pre-specified level. Moreover, the decision maker has to articulate his/her utility function or determine the parameter of the risk functional. This can be subjective and unintuitive. In contrast, the probability of constraint

violation, or more generally of not meeting a target, is more intuitive, and often easy for the decision maker to set. Indeed, extensive empirical study shows that in daily decision making, risk is primarily considered by decision makers as failure to meet a pre-specified target (e.g., Lanzillotti 1958, Simon 1959, Mao 1970, Payne et al. 1980, 1981).

We remark that the term *envelope constraint* appeared as early as the 70's in the field of signal processing, see for example Evans et al. (1977a,b) and more recently Vu et al. (1997). Here, one seeks to design a filter such that its response to a specified input lies within a pre-defined envelope, consistent with the commonly used term *envelope function*, and thus, at a high level, the idea of performing within an envelope constraint is a common theme. Beyond that, the motivation, setup and technical details are, of course, quite different.

Organization

This paper is organized as follows. In Section 2 we propose the probabilistic envelope constraint framework, and present examples that motivate the formulation. In particular we point out the inadequacy of the traditional chance constraint setup in these settings. We then show in Section 3 that the probabilistic envelope program is equivalent to a deterministic semi-infinite program. This equivalence relationship has a nice interpretation as providing tight probabilistic bounds for comprehensive robust optimization formulations. The computational issue of solving the probabilistic envelope program and equivalently the comprehensive robust optimization is discussed in detail in Section 4. We then present two extensions of the proposed framework: in Section 5 we consider the case where we must satisfy probabilistic envelope constraints jointly for a group of constraints. Then, in Section 6, we discuss the distributionally robust approach to the probabilistic envelope framework: instead of assuming the precise distribution of the uncertainty is known, we take it to be known only approximately. This setting is particularly relevant for problems where our only knowledge of the distribution comes from estimates formed from a finite sample set. We report simulation results in Section 7, where we see the protection at all levels offered by envelope constraints in a portfolio optimization problem. All proofs, except a few succinct ones, are deferred to the appendix.

Notation

We use boldface letters to denote column vectors, and row vectors are represented using the transpose (superscript $\top$) of the column vectors. To distinguish between stochastic noise as used in the probabilistic formulation, and deterministic set-based uncertainty, as used in comprehensive robust

optimization, we use a superscript r for each random variable, as we have done in the introduction. As is standard, we use $\mathcal{N}(0,\Sigma)$ to denote the Gaussian distribution with mean zero and covariance matrix Σ. To lighten notation, given a positive definite matrix Σ, we write $\|\mathbf{x}\|_\Sigma$ to denote $\sqrt{\mathbf{x}^\top\Sigma\mathbf{x}}$. Finally, we call an optimization problem *tractable* if it can be solved in polynomial time.

2. Formulation and motivating examples

We first propose the probabilistic envelope constraint as a generalization of the chance constraint. For clarity, we repeat some of the definitions given in the introduction. Given a random variable δ^r representing the uncertain parameter, and a constraint function $f(\mathbf{x},\delta^r)$, a chance constraint places a lower-bound on the probability that f reaches a certain target. That is, for a fixed $\alpha\in\mathbb{R}$ and $p\in[0,1]$, we require

$$\textbf{Chance Constraint:}\ \ \mathbb{P}(f(\mathbf{x},\delta^r)\geq\alpha)\geq p.$$

As discussed above, the chance constraint provides protection against noise by bounding the probability of failing to achieve target α. 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 rectify this shortcoming, we propose a constraint called a *probabilistic envelope constraint*, which bounds the probability of failing to meet the target, α, at all levels. Given a non-decreasing function $\mathfrak{P}(s)$, the envelope constraint on f becomes:

$$\textbf{Probabilistic Envelope Constraint:}\ \ \mathbb{P}(f(\mathbf{x},\delta^r)\geq\alpha-s)\geq\mathfrak{P}(s);\qquad\forall s\geq 0.$$

One example of particular interest is when we require the probability of violation of the constraint by s to decay exponentially in s. Thus, this would give: $\mathfrak{P}(s)=1-\gamma\exp(-\alpha s)$. We call this an *exponentially decaying probabilistic envelope constraint*. We pay particular attention to such functions $\mathfrak{P}(s)$ in the sequel.

A chance constraint is a special case of a probabilistic envelope constraint, which we recover by setting $\mathfrak{P}(s)\equiv p$. In general, a probabilistic envelope constraint can be regarded as an infinite set of chance constraints, since for any fixed $s>0$, $\mathbb{P}(f(\mathbf{x},\delta^r)\geq\alpha-s)\geq\mathfrak{P}(s)$ is a chance constraint.

We now introduce optimization with probabilistic envelope constraints, which we call *probabilistic envelope programs*. While the original definition is general, in this paper we focus on using probabilistic envelope constraints in linear programs. Consider a linear program on $\mathbf{x}\in\mathbb{R}^n$

$$\begin{aligned}\text{Minimize:}\quad&\mathbf{c}^\top\mathbf{x}\\ \text{Subject to:}\quad&\mathbf{a}_i^\top\mathbf{x}\geq b_i;\quad i=1,\ldots,m.\end{aligned}$$

We can assume without loss of generality (introducing an additional variable, if necessary) that there is uncertainty only in the matrix, i.e., the $\{\mathbf{a}_i\}$. We consider an additive model for uncertainty, where the true (unknown) parameter is equal to the nominal value plus a random variable: $\hat{\mathbf{a}}_i = \mathbf{a}_i + \boldsymbol{\delta}_i^r$ for some random noise $\boldsymbol{\delta}_i^r$. As with chance constraints, the probabilistic envelope constraint we obtain depends on how much we know about the distribution of $\boldsymbol{\delta}_i^r$. We focus primarily on two cases: the setting where the distribution μ_i, of the $\{\boldsymbol{\delta}_i^r\}$ is known only through its first two moments, and the setting where it is known exactly (and completely). In the setting where only the mean and covariance are known (say 0 and Σ), then the probabilistic envelope constraint becomes a minimization over all distributions with that mean and variance:

$$\inf_{\boldsymbol{\delta}_i^r \sim (0,\Sigma)} \mathbb{P}\left[(\mathbf{a}_i + \boldsymbol{\delta}_i^r)^\top \mathbf{x} \geq b_i - s\right] \geq \mathfrak{P}_i(s); \quad \forall s \geq 0; \quad i = 1, \ldots, m.$$

If the distribution of $\boldsymbol{\delta}_i^r$ is known exactly, the envelope constraint becomes:

$$\mathbb{P}_{\boldsymbol{\delta}_i^r \sim \mu_i}\left[(\mathbf{a}_i + \boldsymbol{\delta}_i^r)^\top \mathbf{x} \geq b_i - s\right] \geq \mathfrak{P}_i(s); \quad \forall s \geq 0; \quad i = 1, \ldots, m.$$

We conclude this section by presenting two motivating examples where the probabilistic envelope constraint appears particularly useful.

2.1. Example 1: Portfolio Optimization

Consider a stylized portfolio optimization problem over n stocks. We model the unit return of each stock as $Z_i + c_i Z_0$. Random variable Z_i captures randomness due to the ith stock, and random variable Z_0 models the impact of the market. Thus we assume that the random variables $\{Z_i\}$ are independent across stocks, while random variable Z_0 is common across all stocks. Suppose we assume we know the distribution of these random variables, and moreover they are normal, so that $Z_i \sim \mathcal{N}(a_i, \sigma_i^2)$, and $Z_0 \sim \mathcal{N}(a_0, \sigma_0^2)$.

We would like our portfolio to meet a target return of T, with probability at least $1-\gamma$. Furthermore, we would like the probability of missing the target T by more than s, to decay exponentially in s. Thus we have:

$$\begin{aligned}
\text{Maximize:} \quad & \mathbb{E}[\sum_{i=1}^{n}(Z_i + c_i Z_0)x_i] \\
\text{Subject to:} \quad & \mathbb{P}(\sum_{i=1}^{n}(Z_i + c_i Z_0)x_i \geq T - s) \geq 1 - \gamma \exp(-\alpha s); \quad \forall s \geq 0; \\
& \sum_{i=1}^{n} x_i = 1; \\
& x_i \geq 0; \quad i = 1, \ldots, n.
\end{aligned}$$

By definition of Z_0 and Z_i, this formulation is equivalent to the following probabilistic envelope program:

$$\begin{aligned}
\text{Maximize:} \quad & \sum_{i=1}^{n}(a_i + c_i a_0)x_i \\
\text{Subject to:} \quad & \mathbb{P}\Big(\sum_{i=1}^{n}(a_i + c_i a_0 + \delta_i^r)x_i \geq T - s\Big) \geq 1 - \gamma\exp(-\alpha s); \quad \forall s \geq 0; \\
& \sum_{i=1}^{n} x_i = 1; \\
& x_i \geq 0; \quad i = 1, \ldots, n;
\end{aligned}$$

where $\boldsymbol{\delta}^r \triangleq (\delta_1^r, \ldots, \delta_n^r) \sim \mathcal{N}(0, \Sigma)$ is a random variable, and Σ is such that $\Sigma_{ii} = \sigma_i^2 + c_i^2\sigma_0^2$, and $\Sigma_{ij} = c_i c_j \sigma_0^2$.

2.2. Example 2: Robust Regression

The second example we consider is linear regression. Given a matrix A and observed vector $\mathbf{b}$, the nominal problem is to minimize $\|A\mathbf{x} - \mathbf{b}\|^2$. This can be rewritten as:

$$\begin{aligned}
\text{Minimize:} \quad & \sum_{i=1}^{n} \xi_i^2 \\
\text{Subject to:} \quad & |b_i - \mathbf{a}_i\mathbf{x}| \leq \xi_i, \quad i = 1, \ldots, m; \\
& \xi_i \geq 0, \quad i = 1, \ldots, m.
\end{aligned}$$

In many typical examples, the linear assumption is merely an approximation, either because of nonlinearity, or because of noise in our measurements of the matrix A. Let us assume that the regression matrix A is uncertain. In particular, suppose each row in fact equals $\mathbf{a}_i + \boldsymbol{\delta}_i^r$, for Gaussian noise $\boldsymbol{\delta}_i^r \sim \mathcal{N}(0, \Sigma_i)$. In the standard formulation of regression, large fluctuations of the uncertain parameter could produce very skewed fits, because of the squared loss. Thus large fluctuations could jeopardize the entire fit. One approach to combat this is to require a relaxed fitting condition, with probabilistic fitting, but controlling the probability of "large" constraint violation. This is precisely the setting for probabilistic envelope constraints. Using this framework, we have the following problem for any given fixed s^*.

$$\begin{aligned}
\text{Minimize:} \quad & \sum_{i=1}^{n} \xi_i^2 \\
\text{Subject to:} \quad & |b_i - \mathbf{a}_i^\top\mathbf{x}| \leq \xi_i, \quad i = 1, \ldots, m; \\
& \mathbb{P}\big(|b_i - (\mathbf{a}_i + \boldsymbol{\delta}_i^r)^\top\mathbf{x}| \leq \xi_i + s\big) \geq 1 - \gamma\exp(-\alpha s), \quad \forall s \geq s^*. \quad i = 1, \ldots, m; \\
& \xi_i \geq 0, \quad i = 1, \ldots, m.
\end{aligned} \tag{2}$$

Note that while the constraint $\mathbb{P}\big(|b_i - (\mathbf{a}_i + \boldsymbol{\delta}_i^r)^\top\mathbf{x}| \leq \xi_i + s\big) \geq 1 - \gamma\exp(-\alpha s)$ is a probabilistic envelope constraint, due to the non-linearity of $|b_i - (\mathbf{a}_i + \boldsymbol{\delta}_i^r)^\top\mathbf{x}|$, Problem (2) is not a probabilistic

envelope program. After establishing the basic computational complexity results for probabilistic envelope programs, we revisit this problem in Section 5, and show that it can be approximated by a tractable probabilistic envelope program.

3. Deterministic re-formulation of probabilistic envelope constraints

In this section we present the main result of the paper: the probabilistic envelope problem can be re-formulated as an equivalent deterministic semi-infinite program. Moreover, we show that this re-formulated problem is what is known as a *comprehensive robust optimization* proposed in Ben-Tal et al. (2006) (also called a *global robust counterpart* in Ben-Tal et al. 2009). This formulation has been well-studied and we refer to the original papers on the topic (Ben-Tal et al. 2006, Ben-Tal et al. 2009, Ben-Tal et al. 2010). Thus the probabilistic uncertainty model is linked to the deterministic set-based uncertainty model of (comprehensive) robust optimization. This result is in the spirit of past work that has linked (standard) chance constraints to (standard) robust optimization (e.g., Shivaswamy et al. 2006, Delage and Mannor 2010).

We begin with a brief introduction of the comprehensive robust optimization formulation. We then derive tight probabilistic bounds for comprehensive robust constraints. These results lead to our main theorems, which provide a deterministic re-formulation of the probabilistic envelope problem.

3.1. Comprehensive robust constraints

Comprehensive Robust Optimization (e.g., Ben-Tal et al. 2006, 2010) aims to relax the robust optimization formulation, and provide different levels of protection against different levels of noise. For a linear program, the corresponding comprehensive robust optimization formulation takes the form:

$$\begin{aligned} \text{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{Subject to:} \quad & (\mathbf{a}_i + \boldsymbol{\delta}_i)^\top \mathbf{x} \geq b_i - f_i(\boldsymbol{\delta}_i), \quad \forall \boldsymbol{\delta}_i \in \mathbb{R}^n;\ i = 1, \ldots, m, \end{aligned}$$

where each $f_i(\cdot)$ is a non-negative *penalty function*. Note that this is deterministic: the constraints must be satisfied for every $\boldsymbol{\delta}$, as these are no longer stochastic. Comprehensive robust optimization generalizes robust linear optimization (e.g., Ben-Tal and Nemirovski 1998, 1999, Bertsimas and Sim 2004) — indeed, if $f_i(\cdot)$ is taken to be the indicator function of a set Ω_i, i.e., $f_i(\delta_i) = 0$ for $\delta_i \in \Omega_i$ and $+\infty$ otherwise, then the formulation above recovers the standard robust optimization formulation:

$$\begin{aligned} \text{Minimize:} \ & \mathbf{c}^\top \mathbf{x} \\ \text{Subject to:} \ & (\mathbf{a}_i + \boldsymbol{\delta}_i)^\top \mathbf{x} \geq b_i, \qquad \forall\, \boldsymbol{\delta}_i \in \Omega_i. \end{aligned}$$

The robust optimization formulation guarantees that the constraint will not be violated for any realization of the uncertain parameters in the set Ω_i, but makes no guarantees for realizations outside that set. The comprehensive robust optimization formulation allows us to choose different functions $f_i(\cdot)$, in order to provide different levels of protection for different parameter realizations, as opposed to the "all-or-nothing" view of standard robust optimization. Thus, given the comprehensive robust formulation above, for a large parameter deviation, the constraint $(\mathbf{a}_i + \boldsymbol{\delta}_i)^\top \mathbf{x} \geq b$ only needs to be approximately satisfied, i.e., a gap of $f_i(\boldsymbol{\delta}_i)$ is allowed.

3.2. Probabilistic bounds of comprehensive robust constraints

While robust optimization has seen remarkable success as a tractable optimization tool for providing probabilistic protection to optimization solutions (e.g., Shivaswamy et al. 2006, Delage and Mannor 2010), there has been no successful effort to date, to develop the probabilistic side of the story of comprehensive robust optimization. This subsection seeks to develop such a link. More specifically we derive tight bounds on the probability that the solution to a comprehensive robust optimization violates a given constraint with a magnitude of at least s.

As we discuss in Section 2, throughout the paper we consider two different noise models. In the first, we assume we only know the mean and variance of the noise, and want to bound the worst-case probability among all distributions with that given mean and variance. In the second model, we assume we have complete (and perfect) information about the noise distribution. More specifically, we consider ellipsoidal noise, i.e., noise of the form $H\boldsymbol{\zeta}^r$, where $\boldsymbol{\zeta}^r$ is a spherically symmetric random variable with mean 0 and variance I. Gaussian noise is a special case of such a noise model. In both cases, the bounds obtained are *tight* in the sense that if the comprehensive robust constraint is not satisfied, then there exists a value of s, such that the corresponding probabilistic bound at level s will be violated.

THEOREM 1 **(Mean-Variance Model)**. *Let $t : \mathbb{R}^+ \mapsto [0, +\infty]$ be a non-decreasing function such that $t(0) = 0$ and $\lim_{r\uparrow+\infty} t(r) = +\infty$. Then the constraint*

$$(\mathbf{a} + \boldsymbol{\delta})^\top \mathbf{x} \geq b - t(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n, \tag{3}$$

is equivalent to

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a} + \boldsymbol{\delta}^r)^\top \mathbf{x} \geq b - s\big] \geq 1 - \frac{1}{\big(t^{-1}(s)\big)^2 + 1},\ \forall s \geq 0. \tag{4}$$

The infimum here is taken over all random variables $\boldsymbol{\delta}^r$ with mean zero and covariance matrix Σ, and $t^{-1}(s) \triangleq \sup\{r | t(r) \leq s\}$.

Next we consider the case where the distribution of the deviation is known. As a representative (and important) example, we consider the Gaussian case.

THEOREM 2 **(Gaussian Model)**. *Let $t : \mathbb{R}^+ \mapsto [0, +\infty]$ be a non-decreasing function such that $t(0) = 0$ and $\lim_{r\uparrow+\infty} t(r) = +\infty$. If $\boldsymbol{\delta}^r \sim \mathcal{N}(0, \Sigma)$, then the constraint*

$$(\mathbf{a}+\boldsymbol{\delta})^\top \mathbf{x} \geq b - t(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n, \tag{5}$$

is equivalent to

$$\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \geq b - s\big] \geq \Phi\big(t^{-1}(s)\big), \forall s \geq 0. \tag{6}$$

Here $t^{-1}(s) \triangleq \sup\{r | t(r) \leq s\}$.

It is straightforward to extend Theorem 2 to noise that follows an ellipsoidal distribution. Recall that such a random variable can be represented as a spherical random variable under a linear transformation, i.e., $H\boldsymbol{\zeta}^r$ where H is a matrix and $\boldsymbol{\zeta}^r$ is spherical.

THEOREM 3 **(Ellipsoidal Model)**. *Let $t : \mathbb{R}^+ \mapsto [0, +\infty]$ be a non-decreasing function such that $t(0) = 0$ and $\lim_{r\uparrow+\infty} t(r) = +\infty$. Let $\boldsymbol{\zeta}^r \in \mathbb{R}^n$ be a spherical random variable with mean 0 and variance I_n, and $H \in \mathbb{R}^{n\times n}$ be a full rank matrix. Let $\Sigma = H^\top H$, and let $\Psi(\cdot)$ be the cumulative distribution function of the one-dimensional marginal of ζ. Then the constraint*

$$(\mathbf{a}+\boldsymbol{\delta})^\top \mathbf{x} \geq b - t(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n,$$

is equivalent to

$$\mathbb{P}\big[(\mathbf{a}+H\boldsymbol{\zeta}^r)^\top \mathbf{x} \geq b - s\big] \geq \Psi\big(t^{-1}(s)\big), \forall s \geq 0.$$

3.3. Re-formulation of probabilistic envelope constraints

One may notice that the probabilistic bounds in Theorem 1 are indeed probabilistic envelope constraints. This indeed implies our main result that probabilistic envelope constraints can be re-formulated as an *deterministic* semi-infinite program – a comprehensive robust optimization problem, thus linking two widely-used models in treating uncertainty: the deterministic uncertainty model used in (comprehensive) robust optimization, and the probabilistic uncertainty model.

As before, the first model considers a random deviation with mean zero and variance Σ while the specific distribution is unknown, and the second one considers a random variable with known distribution (e.g., Gaussian).

THEOREM 4 **(Mean-Variance Model)**. *If $\mathfrak{P}:\mathbb{R}^+\mapsto[0,1)$ is a non-decreasing function that is continuous from the right, then the probabilistic envelope constraint*

$$\inf_{\boldsymbol{\delta}^r\sim(0,\Sigma)}\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top\mathbf{x}\geq b-s\big]\geq\mathfrak{P}(s),\ \forall s\geq 0,$$

is equivalent to the comprehensive robust constraint

$$(\mathbf{a}+\boldsymbol{\delta})^\top\mathbf{x}\geq b-\mathfrak{P}^{-1}\left(\frac{\|\boldsymbol{\delta}\|^2_{\Sigma^{-1}}}{1+\|\boldsymbol{\delta}\|^2_{\Sigma^{-1}}}\right),\ \forall\boldsymbol{\delta}\in\mathbb{R}^n, \tag{7}$$

where $\mathfrak{P}^{-1}(x)\triangleq\inf\{y\geq 0|\mathfrak{P}(y)\geq x\}$.

We also obtain an equivalence of the probabilistic envelope constraint to comprehensive robust optimization when the distribution of the uncertainty is known exactly. Note that the range of $\mathfrak{P}(\cdot)$ is $[0.5,\,1)$ because the nominal constraint $\mathbf{a}^\top\mathbf{x}\geq b$ implies that for any $s>0$, $\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top\mathbf{x}\geq b-s\big]$ is at least 0.5 since the random variable is symmetric.

THEOREM 5 **(Gaussian Model)**. *If $\mathfrak{P}:\mathbb{R}^+\mapsto[0.5,1)$ is a non-decreasing function that is continuous from right, and $\boldsymbol{\delta}^r\sim\mathcal{N}(0,\Sigma)$, then the probabilistic constraint*

$$\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top\mathbf{x}\geq b-s\big]\geq\mathfrak{P}(s),\ \forall s\geq 0,$$

is equivalent to

$$(\mathbf{a}+\boldsymbol{\delta})^\top\mathbf{x}\geq b-\mathfrak{P}^{-1}\left(\Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}})\right),\ \forall\boldsymbol{\delta}\in\mathbb{R}^n,$$

where $\mathfrak{P}^{-1}(x)\triangleq\inf\{y\geq 0|\mathfrak{P}(y)\geq x\}$ and $\Phi(\cdot)$ is the cumulative distribution function of $\mathcal{N}(0,1)$.

THEOREM 6 **(Ellipsoidal Model)**. *Let $\boldsymbol{\zeta}^r\in\mathbb{R}^n$ be a spherical random variable, with mean 0 and variance I_n, and $H\in\mathbb{R}^{n\times n}$ be a full rank matrix. Let $\Sigma=H^\top H$, and denote the cumulative distribution function of the one-dimensional marginal of $\boldsymbol{\zeta}^r$ by $\Psi(\cdot)$, which is strictly increasing. If $\mathfrak{P}:\mathbb{R}^+\mapsto[0.5,1)$ is a non-decreasing function that is continuous from the right, then the constraint*

$$\mathbb{P}\big[(\mathbf{a}+H\boldsymbol{\zeta}^r)^\top\mathbf{x}\geq b-s\big]\geq\mathfrak{P}(s),\ \forall s\geq 0,$$

is equivalent to

$$(\mathbf{a}+\boldsymbol{\delta})^\top\mathbf{x}\geq b-\mathfrak{P}^{-1}\left(\Psi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}})\right),\ \forall\boldsymbol{\delta}\in\mathbb{R}^n,$$

where $\mathfrak{P}^{-1}(x)\triangleq\inf\{y\geq 0|\mathfrak{P}(y)\geq x\}$.

As an example to illustrate the re-formulation, consider the exponential decay in the motivating example: for $\boldsymbol{\delta}^r\sim\mathcal{N}(0,\Sigma)$ and $\gamma<1/2$,

$$\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top\mathbf{x}\geq b_i-s\big]\geq 1-\gamma\exp(-\alpha s);\quad\forall s\geq 0. \tag{8}$$

Thus, $\mathfrak{P}(s) = 1 - \gamma\exp(-\alpha s)$, and hence $\mathfrak{P}^{-1}(x) = \max[-\frac{1}{\alpha}\log(1-x) + \frac{\log\gamma}{\alpha}, 0]$. Therefore, $\mathfrak{P}^{-1}(\Phi(t)) = \max\big[-\frac{1}{\alpha}\log(1-\Phi(t)) + \frac{\log\gamma}{\alpha}, 0\big]$. Theorem 5 asserts that Constraint (8) is equivalent to the infinite collection of deterministic constraints

$$(\mathbf{a}+\boldsymbol{\delta})^\top\mathbf{x} \geq b + \min\Big[\frac{1}{\alpha}\log(1-\Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}})) - \frac{\log\gamma}{\alpha}, 0\Big],\ \forall\boldsymbol{\delta}\in\mathbb{R}^n,$$

which is further equivalent to

$$(\mathbf{a}+\boldsymbol{\delta})^\top\mathbf{x} \geq b + \frac{1}{\alpha}\log(1-\Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}})) - \frac{\log\gamma}{\alpha},\ \forall\boldsymbol{\delta} \text{ such that } \Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}) \geq 1-\gamma. \tag{9}$$

In Section 4 we further show that the feasible set to Constraint (9) has a polynomial time "separation Oracle", and hence tractable.

4. Computational Tractability

A question of immediate interest is the computational tractability of the probabilistic envelope constraints. Given the equivalence in the previous section of probabilistic envelope constraints and comprehensive robust optimization, we investigate the tractability of the latter. A comprehensive robust optimization problem is tractable exactly when finding the most adversarial disturbance can be done efficiently. We show that this amounts to minimizing a function of a scalar variable, and it can be minimized efficiently whenever the penalty function of the comprehensive robust optimization is convex. Based on these results, we conclude this section by showing that the exponential decay probabilistic envelope constraint leads to a tractable optimization problem.

Theorem 7. *The following comprehensive robust optimization problem*

$$\begin{aligned} &\textit{Minimize:} && \mathbf{c}^\top\mathbf{x} \\ &\textit{Subject to:} && (\mathbf{a}_i+\boldsymbol{\delta}_i)^\top\mathbf{x} \geq b_i - t_i(\|\boldsymbol{\delta}_i\|_{\Sigma_i^{-1}}), \qquad \forall\boldsymbol{\delta}_i\in\mathbb{R}^n;\ i=1,\ldots,m. \end{aligned} \tag{10}$$

can be solved in polynomial time if for each i, and any $\beta \geq 0$, the following optimization on y can be solved in polynomial time:

$$\textit{Minimize:}\quad t_i(y) - \beta y,\ \textit{subject to: } y\in\mathbb{R}^+. \tag{11}$$

Remark 1. Notice that when $t_i(\cdot)$ is convex, then the function $t_i(y) - \beta y$ is a convex function of a *scalar variable* y. Therefore, this function is *unimodal*, and in particular, can be solved in polynomial time using line-search and bisection. Similarly, if $t_i(\cdot)$ is concave, then the function $t_i(y) - \beta y$ is a concave function of *scalar variable* y. Therefore, the minimum is attained at one of the two extremes, i.e., $y = 0$ or $y \uparrow \infty$.

Remark 2. Note that en route to proving Theorem 7, we establish that for fixed $\mathbf{x}_0$, the most adversarial noise, i.e., $\boldsymbol{\delta}_i$ such that constraint $(\mathbf{a}_i+\boldsymbol{\delta}_i)^\top\mathbf{x}_0 \geq b_i - f_i(\boldsymbol{\delta}_i)$ has a largest violation gap, is achieved at

$$\boldsymbol{\delta}_i^* = -y^*\Sigma_i\mathbf{x}_0/\sqrt{\mathbf{x}_0^\top\Sigma_i\mathbf{x}_0},$$

where y^* is the optimal solution of

$$\text{Minimize: } t_i(y) - y\sqrt{\mathbf{x}_0^\top\Sigma_i\mathbf{x}_0}, \text{ Subject to: } y\in\mathbb{R}^+.$$

4.1. Exponentially decaying probabilistic envelope constraints

As an example, we next investigate explicitly the tractability of exponentially decaying probabilistic envelope constraints. The next lemma shows that for an ellipsoidal random variable with log-concave density, the resulting penalty function of the exponential decaying probabilistic envelope constraints is convex.

Lemma 1. *Let $\boldsymbol{\zeta}^r\in\mathbb{R}^n$ be a spherical random variable with mean 0 and variance I_n, and having a log-concave density function. Denote the cumulative distribution function of the one-dimensional marginal of ζ by $\Psi(\cdot)$, which is strictly increasing. If $\mathfrak{P}(s) = 1-\gamma\exp(-\alpha s)$ for $\alpha>0$, then $\mathfrak{P}^{-1}(\Psi(s))$ is convex.*

Proof: Note that since $\boldsymbol{\zeta}^r$ has a log-concave density, the density function of its one-dimensional margin is also log-concave. This further implies that the cumulative distribution function of its marginal, $\Psi(\cdot)$, is log-concave (e.g., Boyd and Vandenberghe 2004). Furthermore, algebraic manipulation yields that $\mathfrak{P}^{-1}(t) = \max[-\frac{1}{\alpha}\log(1-t)+\frac{\log\gamma}{\alpha},0]$. Thus, we have

$$\mathfrak{P}^{-1}(\Psi(s)) = \max[-\frac{1}{\alpha}\log(1-\Psi(s))+\frac{\log\gamma}{\alpha},0] = \max[-\frac{1}{\alpha}\log(\Psi(-s))+\frac{\log\gamma}{\alpha},0].$$

The last equality holds since $\boldsymbol{\zeta}^r$ being spherical implies that $\Psi(\cdot)$ is symmetric. Thus, log concavity of $\Psi(\cdot)$ implies that $-\frac{1}{\alpha}\log(\Psi(-s))$ is convex for any $\alpha>0$. The lemma thus holds since maximization preserves convexity. Q.E.D.

Theorem 8. *For $i=1,\ldots,m$, let $\boldsymbol{\zeta}_i^r\in\mathbb{R}^n$ be a spherical random variable with mean 0 and variance I_n, and having a log-concave density function. Let $H_i\in\mathbb{R}^{n\times n}$ be a full rank matrix. Let $\Sigma_i = H_i^\top H_i$, and denote the cumulative distribution function of the one-dimensional marginal of $\boldsymbol{\zeta}_i^r$ by $\Psi_i(\cdot)$, which is strictly increasing. Further let $\alpha_i>0$ and $0<\gamma_i\leq 0.5$. The following optimization problem*

$$\begin{aligned} &\textit{Minimize:} \quad \mathbf{c}^\top\mathbf{x} \\ &\textit{Subject to:} \quad \mathbb{P}\big[(\mathbf{a}_i+H_i\boldsymbol{\zeta}_i^r)^\top\mathbf{x}\geq b_i-s\big]\geq 1-\gamma_i\exp(-\alpha_i s), \quad \forall s\geq 0, \quad i=1,\ldots,m; \end{aligned} \tag{12}$$

can be solved in polynomial time.

Proof: Let $\mathfrak{P}(s) = 1 - \gamma \exp(-\alpha s)$. By Theorem 6, we have that Problem (12) is equivalent to

$$\begin{aligned} \text{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{Subject to:} \quad & (\mathbf{a}_i + \boldsymbol{\delta}_i)^\top \mathbf{x} \geq b_i - \mathfrak{P}^{-1}(\Psi(\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})), \quad \forall \boldsymbol{\delta}_i \in \mathbb{R}^n, \quad i = 1, \ldots, m. \end{aligned} \tag{13}$$

Note that $\mathfrak{P}^{-1}(\Psi(\cdot))$ is convex due to Lemma 1. By Theorem 7 together with Remark 1 this implies that Problem (13), and equivalently Problem (12) can be solved in polynomial time. Q.E.D.

Remark 3. Indeed, if $\mathfrak{P}(s) = 1 - \gamma \exp(-f(s))$ for an increasing, concave $f(\cdot)$, then we have $\mathfrak{P}^{-1}(s) = \max[f^{-1}(\log \gamma - \log(\Psi(s))), 0]$, which is again convex. The resulting optimization problem is therefore tractable.

As a special case of Theorem 8, we see that the exponentially decaying probabilistic envelope constraints with Gaussian noise are tractable.

Corollary 1. *For $i = 1, \ldots, m$, let $\boldsymbol{\delta}_i^r \in \mathbb{R}^n \sim \mathcal{N}(0, \Sigma_i)$. Further let $\alpha_i > 0$ and $0 < \gamma_i \leq 0.5$. The following optimization problem*

$$\begin{aligned} \textit{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \textit{Subject to:} \quad & \mathbb{P}\big[(\mathbf{a}_i + \boldsymbol{\delta}_i^r)^\top \mathbf{x} \geq b_i - s\big] \geq 1 - \gamma_i \exp(-\alpha_i s), \quad \forall s \geq 0, \quad i = 1, \ldots, m; \end{aligned}$$

can be solved in polynomial time.

Next we investigate exponentially decaying probabilistic envelope constraints for noise described by the mean-variance model. The next theorem shows that it is impossible for the mean-variance model to satisfy an exponentially-decaying probabilistic envelope for the entire tail. Intuitively, this is due to the fact that without further assumption on the distribution, fixing the mean and variance of a random variable only guarantees a power-law tail decay (by, e.g., the Markov inequality). On the other hand, it is possible to require an exponential decay on a bounded interval for the mean-variance model. Indeed, such a formulation leads to tractable problems. As before, the proof is deferred to the appendix.

Theorem 9. *There is no non-zero $\mathbf{x}$ that satisfies*

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a} + \boldsymbol{\delta}^r)^\top \mathbf{x} \geq b - s\big] \geq 1 - \gamma \exp(-\alpha s), \ \forall s \geq 0. \tag{14}$$

On the other hand, for fixed $s^-, s^+ \geq 0$, the following constraint

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a} + \boldsymbol{\delta}^r)^\top \mathbf{x} \geq b - s\big] \geq 1 - \gamma \exp(-\alpha s), \ \forall s \in [s^-, s^+], \tag{15}$$

leads to a tractable optimization problem.

Original Function	Conjugate function
$t_i(x) = \begin{cases} 0 & x \le c, \\ \alpha(x-c) & x > c. \end{cases}$	$t_i^*(y) = \begin{cases} cy & y \le \alpha, \\ +\infty & y > \alpha. \end{cases}$ $= \max(I_\alpha,\, cy)$
$t_i(x) = \begin{cases} \alpha x & x \le c, \\ +\infty & x > c. \end{cases}$	$t_i^*(y) = \begin{cases} 0 & y \le \alpha, \\ c(y-\alpha) & y > \alpha. \end{cases}$ $= \max(0,\, c(y-\alpha))$
$t_i(x) = \begin{cases} 0 & x \le c_1, \\ \alpha(x-c_1) & c_1 < x \le c_2, \\ +\infty & x > c_2. \end{cases}$	$t_i^*(y) = \begin{cases} c_1 y & y \le \alpha, \\ c_2(y-\alpha) + \alpha c_1 & y > \alpha. \end{cases}$ $= \max(c_1 y,\, c_2 y + \alpha(c_1 - c_2))$
$t_i(x) = \begin{cases} 0 & x \le c, \\ \alpha(x-c)^2 & x > c. \end{cases}$	$t_i^*(y) = y^2/4\alpha + cy.$
$t_i(x) = \begin{cases} \alpha x^2 & x \le c, \\ +\infty & x > c. \end{cases}$	$t_i^*(y) = \begin{cases} y^2/4\alpha & y \le 2\alpha c, \\ cy - \alpha c^2 & y > 2\alpha c. \end{cases}$ $= \inf_{\lambda \ge 0} \Big((y-\lambda)^2/4\alpha + c\lambda \Big)$
$t_i(x) = \begin{cases} 0 & x \le c_1, \\ \alpha(x-c_1)^2 & c_1 < x \le c_2, \\ +\infty & x > c_2. \end{cases}$	$t_i^*(y) = \begin{cases} y^2/4\alpha + yc_1 & y \le 2\alpha(c_2 - c_1), \\ c_2 y - \alpha(c_2-c_1)^2 & y > 2\alpha(c_2 - c_1). \end{cases}$ $= \max\Big(c_1 y,\, \inf_{\lambda_1,\lambda_2 \ge 0} [\frac{(y+\lambda_1-\lambda_2)^2}{4\alpha} + c_1 y + (c_2 - c_1)\lambda_2] \Big)$

Table 1 Piecewise-defined functions and their conjugates.

4.2. Some examples of "easy" penalty functions

Thus far we have discussed polynomial time solvability. However, for large scale problems much stronger complexity requirements may be needed. In this section, we list some penalty functions $t_i(\cdot)$, such that the respective comprehensive robust optimization problem can be solved easily. For ease of notation and presentation, we consider the computational issues only for comprehensive robust optimization. Given the equivalence we prove above, tractability results for the corresponding probabilistic envelope programs is immediately implied.

The comprehensive robust optimization

$$\begin{aligned} \text{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{Subject to:} \quad & (\mathbf{a}_i + \boldsymbol{\delta}_i)^\top \mathbf{x} \ge b_i - t_i(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}), \quad \forall \boldsymbol{\delta}_i \in \mathbb{R}^n; \quad i = 1, \ldots, m; \end{aligned} \tag{16}$$

can be reduced to the following optimization formulation problem with finite number of constraints

$$\begin{aligned} \text{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{Subject to:} \quad & \mathbf{a}_i^\top \mathbf{x} - t_i^*(\|\mathbf{x}\|_\Sigma) \ge b; \quad i = 1, \ldots, m, \end{aligned}$$

where $t_i^*(y) = \sup_{x \ge 0} \big[xy - t(x) \big]$ is the conjugate function of $t(\cdot)$. Thus, a comprehensive robust optimization (16) can be easily solved if constraints of the form $t_i^*(x) \le \alpha$ lead to "simple" optimization problems.

We remark that all conjugate functions $t^*(\cdot)$ in Table 1 can be written as $t^*(x) = \max_{1,2}(s_1(x), s_2(x))$, where $s_i = \inf_{\boldsymbol{\lambda} \in \mathcal{S}_i} q_i(\lambda, x)$ for some "simple" functions q_i and polytope $\mathcal{S}_i$.

Here by "simple" we mean the function is a quadratic function, or a linear function, or an indicator function. Hence the constraint $t_i^*(x) \le \alpha$ is equivalent to

$$\begin{aligned} & q_1(x, \boldsymbol{\lambda}_1) \le \alpha; \\ & \boldsymbol{\lambda}_1 \in \mathcal{S}_1; \\ & q_2(x, \boldsymbol{\lambda}_2) \le \alpha; \\ & \boldsymbol{\lambda}_2 \in \mathcal{S}_2. \end{aligned}$$

Since a "simple" function leads to a second order cone constraint, the resulting formulation is a Second Order Cone Program, where medium to large scale problems can be solved using a standard solver in reasonable time.

5. Group probabilistic envelope constraints of correlated noise

In this section we extend the probabilistic envelope constraint to the case where some random variables $\boldsymbol{\delta}_i^r$ are correlated across different constraints. In contrast to independent noise where we bound *individually* the probability of each constraint being satisfied, in the correlated noise case, we bound the probability that a group of constraints are satisfied *simultaneously*. To be more specific, let $\boldsymbol{\delta}_1^r, \ldots, \boldsymbol{\delta}_m^r$ be random variables that follow a joint distribution. Let $\mathcal{I}_1, \ldots, \mathcal{I}_p$ be (not necessarily disjoint) subsets of $[1:m]$. The group probabilistic envelope constraint program considers the following problem:

$$\begin{aligned} \text{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{Subject to:} \quad & \mathbb{P}\big[\forall i \in \mathcal{I}_t : (\mathbf{a}_i + \boldsymbol{\delta}_i^r)^\top \mathbf{x} \ge b_i - s\big] \ge \mathfrak{P}_t(s),\ \forall s \ge 0; \quad t = 1, \ldots, p. \end{aligned}$$

While addressing general correlation seems hard, we investigate two special cases, where we can either exactly or approximately solve the group probabilistic envelope constraint program.

5.1. Identical noise

The first case we investigate is where identical noise $\boldsymbol{\delta}_t^r$ affects each constraint within a group $\mathcal{I}_t$. Then, the group probabilistic envelope constrained program reduces to the following:

$$\begin{aligned} \text{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{Subject to:} \quad & \mathbb{P}\big[\forall i \in \mathcal{I}_t : (\mathbf{a}_i + \boldsymbol{\delta}_t^r)^\top \mathbf{x} \ge b_i - s\big] \ge \mathfrak{P}_t(s),\ \forall s \ge 0; \quad t = 1, \ldots, p. \end{aligned} \tag{17}$$

THEOREM 10. *Problem (17) is equivalent to*

$$\begin{aligned} \textit{Minimize:} \quad & \mathbf{c}^\top \mathbf{x} \\ \textit{Subject to:} \quad & \mathbb{P}\big[(\mathbf{a}_i + \boldsymbol{\delta}_t^r)^\top \mathbf{x} \ge b_i - s\big] \ge \mathfrak{P}_t(s),\ \forall s \ge 0; \quad \forall i \in \mathcal{I}_t; \quad t = 1, \ldots, p. \end{aligned}$$

Theorem 10 follows from the following lemma.

LEMMA 2. *Fix* $\mathbf{x}$, $\mathbf{a}_i$, c_i *and let* $\boldsymbol{\delta}^r$ *be a random variable, then*

$$\mathbb{P}\big(\forall i \in \mathcal{I} : (\mathbf{a}_i + \boldsymbol{\delta}^r)^\top \mathbf{x} \geq c_i\big) = \inf_{i \in \mathcal{I}} \mathbb{P}\big((\mathbf{a}_i + \boldsymbol{\delta}^r)^\top \mathbf{x} \geq c_i\big).$$

Proof: Note that

$$\begin{aligned}\mathbb{P}\big(\forall i \in \mathcal{I} : (\mathbf{a}_i + \boldsymbol{\delta}^r)^\top \mathbf{x} \geq c_i\big) &= \mathbb{P}\big(\inf_{i \in \mathcal{I}}[(\mathbf{a}_i + \boldsymbol{\delta}^r)^\top \mathbf{x} - c_i] \geq 0\big) \\ = \mathbb{P}\big(\boldsymbol{\delta}^{r\top} \mathbf{x} \geq \sup_{i \in \mathcal{I}}[c_i - \mathbf{a}_i^\top \mathbf{x}]\big) &= \inf_{i \in \mathcal{I}} \mathbb{P}\big((\mathbf{a}_i + \boldsymbol{\delta}^r)^\top \mathbf{x} \geq c_i\big),\end{aligned}$$

where in the last equality we use the fact that $\mathbf{a}_i$, c_i and $\mathbf{x}$ are fixed. Q.E.D.

Theorem 10 states that in the identical noise case, we can decompose a group probabilistic envelope constraint into individual probabilistic envelope constraints. Thus, Theorems 4 to 6 immediately imply the following corollaries.

COROLLARY 2 **(Mean-Variance Model)**. *For* $t = 1, \ldots, p$, *if* $\mathfrak{P}_t : \mathbb{R}^+ \mapsto [0, 1)$ *is non-decreasing and continuous from the right, then the following problem*

$$\begin{aligned}&\textit{Minimize:} && \mathbf{c}^\top \mathbf{x} \\ &\textit{Subject to:} && \inf_{\boldsymbol{\delta}_t^r \sim (0, \Sigma_t)} \mathbb{P}\big[\forall i \in \mathcal{I}_t : (\mathbf{a}_i + \boldsymbol{\delta}_t^r)^\top \mathbf{x} \geq b_i - s\big] \geq \mathfrak{P}_t(s),\ \forall s \geq 0; \quad t = 1, \ldots, p;\end{aligned}$$

is equivalent to

$$\begin{aligned}&\textit{Minimize:} && \mathbf{c}^\top \mathbf{x} \\ &\textit{Subject to:} && (\mathbf{a}_i + \boldsymbol{\delta})^\top \mathbf{x} \geq b_i - \mathfrak{P}_t^{-1}\left(\frac{\|\boldsymbol{\delta}\|^2_{\Sigma_t^{-1}}}{1 + \|\boldsymbol{\delta}\|^2_{\Sigma_t^{-1}}}\right),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n; \quad \forall i \in \mathcal{I}_t; \quad t = 1, \ldots, p\end{aligned}$$

where $\mathfrak{P}_t^{-1}(x) \triangleq \inf\{y \geq 0 | \mathfrak{P}_t(y) \geq x\}$.

COROLLARY 3 **(Gaussian Model)**. *For* $t = 1, \ldots, p$, *if* $\mathfrak{P}_t : \mathbb{R}^+ \mapsto [0, 1)$ *is non-decreasing and continuous from right, and* $\boldsymbol{\delta}_t^r \sim \mathcal{N}(0, \Sigma)_t$, *then the following problem*

$$\begin{aligned}&\textit{Minimize:} && \mathbf{c}^\top \mathbf{x} \\ &\textit{Subject to:} && \mathbb{P}\big[\forall i \in \mathcal{I}_t : (\mathbf{a}_i + \boldsymbol{\delta}_t^r)^\top \mathbf{x} \geq b_i - s\big] \geq \mathfrak{P}_t(s),\ \forall s \geq 0; \quad t = 1, \ldots, p;\end{aligned}$$

is equivalent to

$$\begin{aligned}&\textit{Minimize:} && \mathbf{c}^\top \mathbf{x} \\ &\textit{Subject to:} && (\mathbf{a}_i + \boldsymbol{\delta})^\top \mathbf{x} \geq b_i - \mathfrak{P}_t^{-1}\left(\Phi(\|\boldsymbol{\delta}\|_{\Sigma_t^{-1}})\right),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n; \quad \forall i \in \mathcal{I}_t; \quad t = 1, \ldots, p\end{aligned}$$

where $\mathfrak{P}_t^{-1}(x) \triangleq \inf\{y \geq 0 | \mathfrak{P}_t(y) \geq x\}$.

5.2. Two-sided condition

The second case we investigate is the "two-sided condition", i.e., the nominal constraint has the form

$$b - c \le \mathbf{a}^\top \mathbf{x} \le b + c.$$

This constraint is motivated by the linear regression setup, where we observe a set of input-output pairs $(b_i, \mathbf{a}_i)$ for $i = 1, \ldots, m$ and seek a solution $\mathbf{x}$ such that $b_i \approx \mathbf{a}_i^\top \mathbf{x}$ for all i.

The probabilistic envelope constraints for the two-sided condition thus have the form

$$\mathbb{P}(b - s \le (\mathbf{a} + \boldsymbol{\delta}^r)^\top \mathbf{x} \le b + s) \ge \mathfrak{P}(s); \quad \forall s \ge 0. \tag{18}$$

While Constraint (18) seems to be hard to solve exactly, we can approximate it with the following comprehensive robust constraint:

$$\boldsymbol{\delta}^\top \mathbf{x} \ge -\mathfrak{P}^{-1}\left[2(\Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}) - 1/2)\right], \ \forall \boldsymbol{\delta} \in \mathbb{R}^n,$$

under the condition that $\boldsymbol{\delta}^r \sim \mathcal{N}(0, \Sigma)$, and that $\mathfrak{P}(s) > 0$ only when $s \gg c$. The latter condition essentially means that the decision maker is only concerned with bounding the probability of "large" deviations when noise exists.

This approximation is justified by the following theorem. Note that by definition, it is easy to see that $\mathfrak{P}^{-1}$ is continuous from the left. Hence, Theorem 11 implies that when $\epsilon \downarrow 0$, then the inner-set and the outer-set converge to the feasible set of (19). That is, the approximation is asymptotically exact.

THEOREM 11. *Let $\boldsymbol{\delta}^r \sim \mathcal{N}(0, \Sigma)$. Suppose $\mathfrak{P} : \mathbb{R}^+ \mapsto [0, 1)$ is a non-decreasing function that is continuous from the right. Let $s^* = \min\{s | \mathfrak{P}(s) > 0\}$, and let $\epsilon = c/s^*$. Then the feasible set of*

$$\begin{aligned} &\mathbb{P}(b - s \le (\mathbf{a} + \boldsymbol{\delta})^\top \mathbf{x} \le b + s) \ge \mathfrak{P}(s); \quad \forall s \ge 0, \\ &b - c \le \mathbf{a}^\top \mathbf{x} \le b + c \end{aligned} \tag{19}$$

is bounded from the inside by that of

$$\begin{aligned} &\boldsymbol{\delta}^\top \mathbf{x} \ge -\mathfrak{P}^{-1}\left[(2 - \epsilon)(\Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}) - 1/2)\right], \ \forall \boldsymbol{\delta} \in \mathbb{R}^n, \\ &b - c \le \mathbf{a}^\top \mathbf{x} \le b + c; \end{aligned} \tag{20}$$

and bounded from the outside by that of

$$\begin{aligned} &\boldsymbol{\delta}^\top \mathbf{x} \ge -\mathfrak{P}^{-1}\left[2(\Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}) - 1/2)\right], \ \forall \boldsymbol{\delta} \in \mathbb{R}^n, \\ &b - c \le \mathbf{a}^\top \mathbf{x} \le b + c. \end{aligned} \tag{21}$$

As before, $\mathfrak{P}^{-1}(x) \triangleq \inf\{y \ge 0 | \mathfrak{P}(y) \ge x\}$.

Remark 4. A close inspection of the proof of Theorem 11 shows that it only depends on the fact that the density function of the marginal of the distribution is decreasing in $[0,+\infty)$. Thus, it is straightforward to generalize Theorem 11 to the ellipsoidal random variable case.

As an example of the approximation, recall the linear regression with probabilistic envelope constraint (2):

$$\begin{aligned}
\text{Minimize:}\quad & \sum_{i=1}^{n} \xi_i^2 \\
\text{Subject to:}\quad & |b_i - \mathbf{a}_i^\top \mathbf{x}| \le \xi_i, \quad i=1,\dots,m; \\
& \mathbb{P}\big(|b_i - (\mathbf{a}_i + \boldsymbol{\delta}_i^r)^\top \mathbf{x}| \le \xi_i + s\big) \ge 1 - \gamma \exp(-\alpha s), \quad \forall s \ge s^*. \quad i=1,\dots,m; \\
& \xi_i \ge 0, \quad i=1,\dots,m.
\end{aligned}$$

This can thus be approximated by the following program (the algebraic details are deferred to the appendix), which can be solved in polynomial time due to our results in Section 4.

$$\begin{aligned}
\text{Minimize:}\quad & \sum_{i=1}^{n} \xi_i^2 \\
\text{Subject to:}\quad & |b_i - \mathbf{a}_i^\top \mathbf{x}| \le \xi_i, \quad i=1,\dots,m; \\
& \boldsymbol{\delta}_i^\top \mathbf{x} \ge \frac{1}{\alpha}\log(2\Phi(-\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})) - \frac{\log\gamma}{\alpha}; \\
& \qquad \forall \|\boldsymbol{\delta}_i\|_{\Sigma^{-1}} \ge \Phi^{-1}\left(1 - \frac{\gamma\exp(-\alpha s^*)}{2}\right); \quad i=1,\dots,m; \\
& \xi_i \ge 0, \quad i=1,\dots,m.
\end{aligned}$$

6. Approximate Partial Distributional Information

In the previous sections we assume that either perfect information of the distribution, or at least the mean and variance of the noise, is available (i.e., explicitly given). In this section, we consider a more practical setup, where any probabilistic information of the noise is approximate. This setting is most relevant in the sample-driven regime, i.e., the setting where our only knowledge of the noise distribution comes through a finite set of observed noise samples. Given these, we must estimate the distribution (or alternatively the mean and the variance) of the noise. In such a case, it is not realistic to hope for a precise estimate of the distribution parameters (or mean/variance). Instead, we adapt the approach proposed in Delage and Ye (2010) where an interval estimate of mean and variance (or respective parameters of the Gaussian distribution) are given, and provide worst-case (w.r.t. mean–variance or distribution parameters) probabilistic bounds.

Theorem 12. *Suppose the random variable δ^r has unknown mean $\tilde{\mathbf{c}}$ and variance $\tilde{\Sigma}$, such that $\tilde{c}_i \in [-\epsilon_i, +\epsilon_i]$ and $\tilde{\Sigma} \preceq \Sigma^*$. Define $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ as*

$$T(\boldsymbol{\delta})_i = \begin{cases} 0 & \text{if } -\epsilon_i \le \delta_i \le \epsilon_i; \\ \delta_i - \epsilon_i & \text{if } \delta_i > \epsilon_i; \\ \delta_i + \epsilon_i & \text{if } \delta_i < -\epsilon_i. \end{cases}$$

Then the constraint

$$(\mathbf{a}+\boldsymbol{\delta})^\top \mathbf{x} \geq b - t(\|T(\boldsymbol{\delta})\|_{(\Sigma^*)^{-1}}),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n, \tag{22}$$

is equivalent to

$$\inf_{\mathbf{c}:c_i\in[-\epsilon_i,\epsilon_i],\Sigma\preceq\Sigma^*} \inf_{\boldsymbol{\delta}^r\sim(c,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \geq b-s\big] \geq 1 - \frac{1}{\big(t^{-1}(s)\big)^2+1},\ \forall s \geq 0.$$

Using a similar argument, we can derive a probabilistic bound when the parameters of the Gaussian random variable are not precisely known.

THEOREM 13. *The constraint*

$$(\mathbf{a}+\boldsymbol{\delta})^\top \mathbf{x} \geq b - t(\|T(\boldsymbol{\delta})\|_{(\Sigma^*)^{-1}}),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n,$$

is equivalent to

$$\inf_{\mathbf{c}:c_i\in[-\epsilon_i,\epsilon_i],\Sigma\preceq\Sigma^*} Pr_{\delta^r\sim\mathcal{N}(c,\Sigma)}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \geq b-s\big] \geq \Phi\big(t^{-1}(s)\big),\ \forall s \geq 0.$$

Theorems 12 and 13 imply the following corollaries that show that one can use the comprehensive robust optimization paradigm to enforce probabilistic requirements.

COROLLARY 4. *If $\mathfrak{P}:\mathbb{R}^+ \mapsto [0,1)$ is a non-decreasing function that is continuous from the right, then the probabilistic constraint*

$$\inf_{\mathbf{c}:c_i\in[-\epsilon_i,\epsilon_i];\Sigma\preceq\Sigma^*} \inf_{\boldsymbol{\delta}^r\sim(0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \geq b-s\big] \geq \mathfrak{P}(s),\ \forall s \geq 0,$$

is equivalent to

$$(\mathbf{a}+\boldsymbol{\delta})^\top \mathbf{x} \geq b - \mathfrak{P}^{-1}\left(\frac{\|T(\boldsymbol{\delta})\|^2_{(\Sigma^*)^{-1}}}{1+\|T(\boldsymbol{\delta})\|^2_{(\Sigma^*)^{-1}}}\right),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n,$$

where $\mathfrak{P}^{-1}(x) \triangleq \inf\{y \geq 0 | \mathfrak{P}(y) \geq x\}$.

COROLLARY 5. *If $\mathfrak{P}:\mathbb{R}^+ \mapsto [0.5,1)$ is a non-decreasing function that is continuous from the right, then the probabilistic constraint*

$$\inf_{\mathbf{c}:c_i\in[-\epsilon_i,\epsilon_i];\Sigma\preceq\Sigma^*} \mathrm{Pr}_{\boldsymbol{\delta}^r\sim\mathcal{N}(0,\Sigma)}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \geq b-s\big] \geq \mathfrak{P}(s),\ \forall s \geq 0,$$

is equivalent to

$$(\mathbf{a}+\boldsymbol{\delta})^\top \mathbf{x} \geq b - \mathfrak{P}^{-1}\left(\Phi(\|T(\boldsymbol{\delta})\|_{(\Sigma^*)^{-1}})\right),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n,$$

where $\mathfrak{P}^{-1}(x) \triangleq \inf\{y \geq 0 | \mathfrak{P}(y) \geq x\}$.

Before concluding this section, we remark that the confidence interval of the mean $[-\epsilon_i, \epsilon_i]$ can be easily estimated from i.i.d. samples using, for example, Hoeffding's bound. To derive the confidence interval of the variance from empirical observations, (i.e., Σ^*), we refer the readers to Lemma 3 of Delage and Ye (2010).

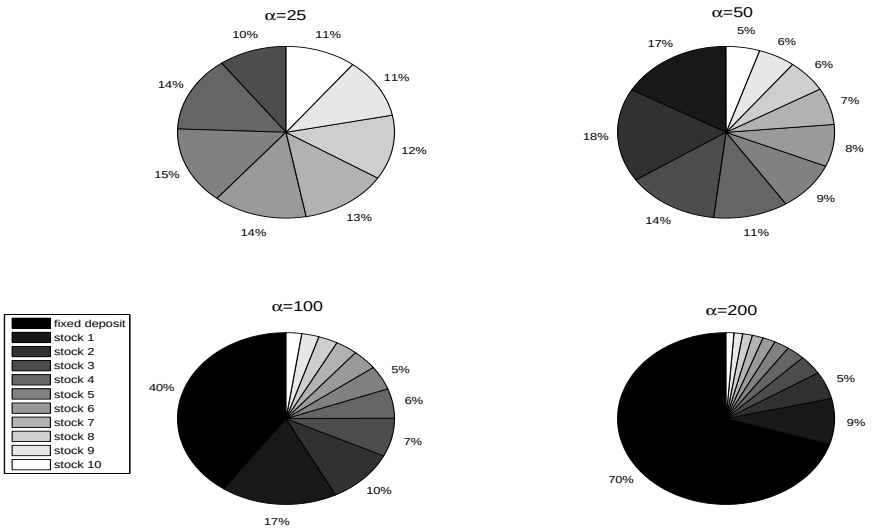

Figure 1 The portfolio allocations for different decay rates $\alpha = 25, 50, 100, 200$ of the envelope constraints.

7. Simulations

In this section we illustrate the proposed approach using a synthetic portfolio optimization example in the spirit of the motivating example discussed in Section 2.1. We consider allocating an investment among 11 assets: 10 stocks and a fixed deposit. The return of the fixed deposit is fixed as 1, while the return of stock i follows the equation

$$r_i = Z_i + Z_0,$$

where we set $Z_i \sim \mathcal{N}(1 + 0.01i, (0.03i)^2)$. Therefore, the larger the mean return of a stock, the larger the return's variance. Thus, stock 10 is the most risky stock, and stock 1 is the most conservative except for the fixed deposit. In addition, the market effect is set as $Z_0 \sim \mathcal{N}(0, (0.01)^2)$. Our goal is to maximize the expected return subject to the exponentially decaying probabilistic envelope $1 - \gamma \exp(-\alpha s)$. Setting $T = 1$ and $\gamma = 0.2$ requires a guarantee of no losses with at least 80% probability. We choose four values of $\alpha = 25$, 50, 100, 200, giving different rates of decay for the probability the constraint is violated at level s for each s. Figure 1 shows the resulting portfolio allocations. Not surprisingly, larger α corresponds to a more risk averse attitude towards large constraint violation (i.e., significant loss), and consequently, the resulting portfolio is more conservative and tends to invest more heavily in the fixed deposit. We also observe that all portfolio allocations are well diversified, each investing a non-negligible fraction in at least 8 out of the 11 assets. Table 2 reports the mean return of these portfolios. We also compute the probability that

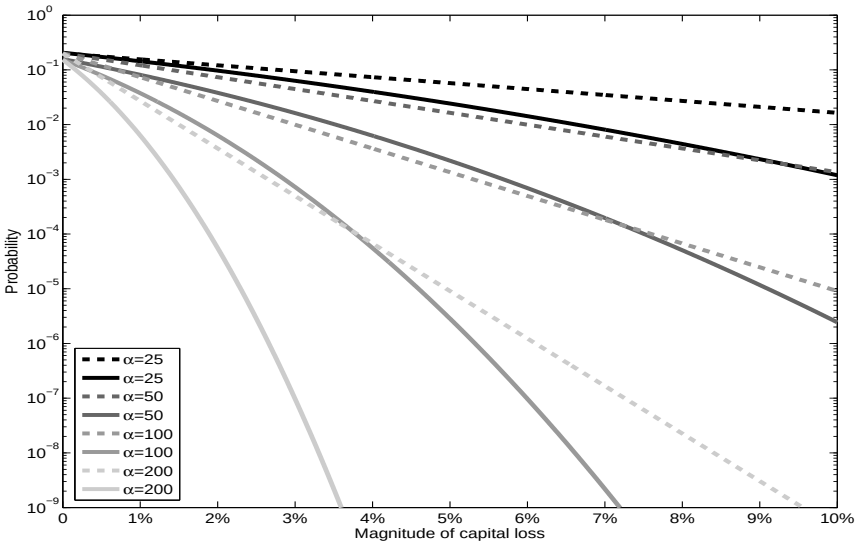

Figure 2 Probability that a portfolio suffers a given loss. Here, the dashed lines are probabilistic envelope constraints imposed, and the solid lines are the true probabilities of resulting portfolios suffer a given amount of loss.

each portfolio suffers given level of loss, as shown in Figure 2. As expected, each portfolio satisfies its respective envelope constraint. Moreover, as an indication that the construction is not conservative, we observe that portfolio allocations designed with less conservative constraints violate the more conservative envelope constraints.

	$\alpha = 25$	$\alpha = 50$	$\alpha = 100$	$\alpha = 200$
Mean return	1.0640	1.0428	1.0220	1.0110

Table 2 Mean return of the allocations for different decay rates $\alpha = 25, 50, 100, 200$ of the envelope constraints.

8. Conclusion

Standard chance constraints ignore the magnitude of constraint violation altogether, controlling only the probability of violation. In many applications, including many finance applications, controlling this magnitude of violation can be paramount. We proposed a new class of probabilistic constraints we call probabilistic envelope constraints, which bound the probability that a constraint is violated by a certain gap s for all values of $s \geq 0$. We investigated linear programs with probabilistic envelope constraints under Gaussian uncertainties and their generalization, log-concave uncertainties, and show that this problem is equivalent to a semi-infinite program, known as comprehensive robust optimization in the literature. As a byproduct, we provided tight probabilistic bounds on comprehensive robust optimization. We further considered the tractability of

probabilistic envelope constraints, and showed that under mild technical conditions the resulting optimization problem can be solved in polynomial time. Extensions to the correlated uncertainty case and the approximate distribution based case were also provided.

Appendix A: Proofs of results in Section 3

This appendix is devoted to establishing results in Section 3. Our main goal is to prove Theorems 4 to 6. This is done in two steps: we first prove Theorems 1 to 3, thus establishing tight probabilistic bounds for comprehensive robust optimization. We then show that these bounds imply Theorems 4 to 6.

A.1. Proof of Theorem 1 to Theorem 3

Proof of Theorem 1: We first show that for any fixed $c \in \mathbb{R}$, $d > 0$, the following two inequalities are equivalent.

$$\sup_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \le c\big] \le \frac{1}{d^2+1} \tag{23}$$

$$\mathbf{a}^\top \mathbf{x} \ge c + d\|\mathbf{x}\|_\Sigma, \tag{24}$$

i.e., if $\mathbf{x}$ satisfies one inequality, then it also satisfies the other one. To see that, by Marshall and Olkin (1960), we have

$$\sup_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \le c\big] = (1+\alpha^2)^{-1},$$

where

$$\alpha = \inf_{\mathbf{n}|\mathbf{n}^\top \mathbf{x} \le c - \mathbf{a}^\top \mathbf{x}} \sqrt{\mathbf{n}^\top \Sigma^{-1} \mathbf{n}} = \begin{cases} \frac{\mathbf{a}^\top \mathbf{x} - c}{\|\mathbf{x}\|_\Sigma} & \text{if } \mathbf{a}^\top \mathbf{x} - c \ge 0; \\ 0 & \text{otherwise.} \end{cases}$$

Thus, if $\mathbf{x}$ satisfies Inequality (24), then $\mathbf{a}^\top \mathbf{x} - c \ge d\|\mathbf{x}\|_\Sigma$, which implies that $\alpha \ge d$, and hence $(1+\alpha^2)^{-1} \le (1+d^2)^{-1}$, which implies Inequality (23) holds. Conversely, if $\mathbf{x}$ does not satisfy Inequality (24), then $\mathbf{a}^\top \mathbf{x} - c < d\|\mathbf{x}\|_\Sigma$ which implies $a < d$, and hence Inequality (23) does not hold. Therefore, Inequalities (23) and (24) are equivalent.

Note that (24) is equivalent to

$$(\mathbf{a}+\boldsymbol{\delta})^\top \mathbf{x} \ge c; \quad \forall \|\boldsymbol{\delta}\|_{\Sigma^{-1}} \le d. \tag{25}$$

Hence for any $c \in \mathbb{R}$, $d > 0$, Inequalities (23) and (25) are equivalent.

We next show that Inequality (4) is equivalent to

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - t(r)\big] \ge 1 - \frac{1}{r^2+1}, \qquad \forall r \ge 0. \tag{26}$$

If Inequality (4) does not hold, then there exists s^* and $\epsilon > 0$ small enough, such that

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - s^*\big] < 1 - \frac{1}{\big(t^{-1}(s^*) - \epsilon\big)^2 + 1}.$$

Let $r^* = t^{-1}(s^*) - \epsilon$, by definition $t(r^*) \le s^*$, hence

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - t(r^*)\big] \le \inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - s^*\big] < 1 - \frac{1}{(r^*)^2+1},$$

hence Inequality (26) does not hold.

On the other-hand, if Inequality (26) does not hold, then there exists r^* such that

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - t(r^*)\big] < 1 - \frac{1}{(r^*)^2+1}.$$

Let $s^* = t(r^*)$, then $t^{-1}(s^*) \ge r^*$. We have

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - s^*\big] < 1 - \frac{1}{(r^*)^2+1} \le 1 - \frac{1}{(t^{-1}(s^*))^2+1},$$

hence Inequality (4) does not hold. Thus, we conclude that (4) and (26) is equivalent.

Finally, by the equivalence of (23) and (24), Constraint (26) is equivalent to:

$$(\mathbf{a}+\boldsymbol{\delta})^\top \mathbf{x} \ge b - t(r);\ \forall \|\boldsymbol{\delta}\|_{\Sigma^{-1}} \le r,\ \forall r \ge 0.$$

Since $t_i(\cdot)$ is nondecreasing, this is further equivalent to (3). Q.E.D.

Notice that the first building block of the proof is the equivalence relationship of a worst-case chance constraint (23) (with only the first and second moment information), and a deterministic constraint (23). We remark that in a recent paper, Zymler et al. (2011) has extended this equivalence relationship to the non-linear case. It would be interesting to see whether this makes it possible to analyze probabilistic envelope constraints for the non-linear optimization problem.

Proof of Theorem 2: For fixed $k \ge 1/2$ and constant l, the following constraints are equivalent:

$$\begin{aligned} & \mathbb{P}(\mathbf{x}^\top \boldsymbol{\delta}_i^r \ge l) \ge k \\ \iff\ & l \le \Phi^{-1}(k)\big(\mathbf{x}^\top \Sigma \mathbf{x}\big)^{1/2} \\ \iff\ & l \le \mathbf{x}^\top \boldsymbol{\delta},\ \forall \|\boldsymbol{\delta}\|_{\Sigma^{-1}} \le \Phi^{-1}(k). \end{aligned} \tag{27}$$

Next we show that (6) is equivalent to

$$\mathbb{P}\Big(\boldsymbol{\delta}^{r\top} \mathbf{x} \ge (b - \mathbf{a}^\top \mathbf{x}) - t(r)\Big) \ge \Phi(r),\ \forall r \ge 0, \tag{28}$$

If (6) does not hold, then there exists $s^* \ge 0$ and $\epsilon > 0$ small enough such that

$$\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - s^*\big) < \Phi\big(t^{-1}(s^*) - \epsilon\big).$$

Let $r^* = t^{-1}(s^*) - \epsilon$, then by definition $t(r^*) \le s^*$. Hence

$$\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - t(r^*)\big) \le \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - s^*\big) < \Phi\big(r^*\big),$$

i.e., Inequality (28) does not hold.

On the other hand, suppose Inequality (28) does not hold. Thus, there exists r^* such that

$$\mathbb{P}\Big(\boldsymbol{\delta}^{r\top}\mathbf{x} \geq (b - \mathbf{a}^\top\mathbf{x}) - t(r^*)\Big) < \Phi(r^*).$$

Let $s^* = t(r^*)$. By definition $t^{-1}(s^*) \geq r^*$. Thus

$$\mathbb{P}\Big(\boldsymbol{\delta}^{r\top}\mathbf{x} \geq (b - \mathbf{a}^\top\mathbf{x}) - s^*\Big) < \Phi(r^*) \leq \Phi(t^{-1}(s^*)),$$

i.e., (6) does not hold. Hence we conclude that (6) and (28) are equivalent.

By (27), Constraint (28) is equivalent to $\forall\gamma \geq 0$,

$$(\mathbf{a} + \boldsymbol{\delta}^r)^\top\mathbf{x} \geq b - t(\gamma), \quad \forall\|\boldsymbol{\delta}\|_{\Sigma^{-1}} \leq \Phi^{-1}\big(\Phi(\gamma)\big) = \gamma.$$

Since $t(\cdot)$ is nondecreasing, this is equivalent to (5). Q.E.D.

The proof of Theorem 3 is identical to the proof of Theorem 2, and hence omitted.

A.2. Proof of Theorem 4 to Theorem 6

We first prove Theorem 4, the mean-variance model.

Proof of Theorem 4: Since $\lim_{s\to+\infty}\mathfrak{P}(s)$ may not converge to 1, we need to pay special attention to the case where $\inf\{y \geq 0|\mathfrak{P}(y) \geq x\} = \emptyset$. To this end, let

$$\mathfrak{S} = \{s|\exists y \geq 0 : \mathfrak{P}(y) \geq s\}.$$

Note that $\mathfrak{P}(\cdot)$ is non-decreasing and continuous from right, we have for $x \in \mathfrak{S}$,

$$\mathfrak{P}^{-1}(x) = \min\{y \geq 0|\mathfrak{P}(y) \geq x\}.$$

And $\mathfrak{P}^{-1}(x) = +\infty$ for $x \notin \mathfrak{S}$. To simplify notation, let $g(x) \triangleq x^2/(1+x^2)$ and $v(x) \triangleq \mathfrak{P}^{-1}(g(x))$. Note that $g(x)$ is a strictly increasing function in $\mathbb{R}^+$ onto $[0,1)$, hence $g^{-1}(y)$ is uniquely defined for any $y \in [0,1)$. By definition we have when $g(x) \in \mathfrak{S}$,

$$v(x) = \mathfrak{P}^{-1}(g(x)) = \inf\{y \geq 0|\mathfrak{P}(y) \geq g(x)\} = \inf\{y \geq 0|g^{-1}(\mathfrak{P}(y)) \geq x\} = \min\{y \geq 0|g^{-1}(\mathfrak{P}(y)) \geq x\};$$

and $v(x) = +\infty$ when $g(x) \notin \mathfrak{S}$.

Note that Constraint (7) can be rewritten as

$$(\mathbf{a} + \boldsymbol{\delta})^\top\mathbf{x} \geq b - v(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}),\ \forall\boldsymbol{\delta} \in \mathbb{R}^n,$$

which by Theorem 1 is equivalent to

$$\inf_{\boldsymbol{\delta}^r\sim(0,\Sigma)} \mathbb{P}\big[(\mathbf{a} + \boldsymbol{\delta}^r)^\top\mathbf{x} \geq b - s\big] \geq g(v^{-1}(s)),\ \forall s \geq 0, \tag{29}$$

where $v^{-1}(s) = \sup\{x|v(x) \le s\}$.

Furthermore, we have the following

$$v^{-1}(s) = \sup\{x|v(x) \le s\} = \max\Big(\sup\{x|g(x) \in \mathfrak{S},\, v(x) \le s\}, \sup\{x|g(x) \notin \mathfrak{S},\, v(x) \le s\}\Big).$$

Note that $g(x) \notin \mathfrak{S}$ implies $v(x) = +\infty$, hence $\{x|g(x) \notin \mathfrak{S},\, v(x) \le s\} = \emptyset$, i.e., the second term equals $-\infty$. The first term equals

$$\begin{aligned}
&\sup\big\{x\,\big|\min\{y \ge 0|g^{-1}(\mathfrak{P}(y)) \ge x\} \le s\big\} \\
&\quad = \sup\big\{x|\exists y \in [0,s] : g^{-1}(\mathfrak{P}(y)) \ge x\big\} \\
&\quad = \sup\big\{x|g^{-1}(\mathfrak{P}(s)) \ge x\big\} \\
&\quad = g^{-1}(\mathfrak{P}(s)).
\end{aligned}$$

Thus, we have that

$$v^{-1}(s) = g^{-1}(\mathfrak{P}(s)).$$

Substitute it into (29) and note that $g(g^{-1}(z)) = z$ we conclude that Constraint (7) is equivalent to

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a} + \boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - s\big] \ge \mathfrak{P}(s),\ \forall s \ge 0,$$

which establishes the Theorem. Q.E.D.

If we let $g(x) \triangleq \Phi(x)$ (respectively $\Psi(x)$), then the proof of Theorem 5 (and of Theorem 6) are identical to the proof of Theorem 4. Hence we omit the details.

Appendix B: Proofs of results in Section 4

In this appendix we provide proofs to Theorem 7 and Theorem 9.

Proof of Theorem 7: Note that Problem (10) is a minimization of a linear objective over a convex set (denoted by $\mathcal{C}$), since there are infinitely many linear constraints. Grötschel et al. (1988) shows that a sufficient condition for such problem to be solved in polynomial time is the existence of a polynomial-time separation Oracle of $\mathcal{C}$, which is a sub-routine such that given a candidate solution $\mathbf{x}$, in polynomial time it either correctly reports that $\mathbf{x} \in \mathcal{C}$; or outputs an $\mathbf{h}$ and α such that $\mathbf{h}^\top \mathbf{x} < \alpha$, whereas for any $\mathbf{z} \in \mathcal{C}$, $\mathbf{h}^\top \mathbf{z} \ge \alpha$.

Thus, to complete the proof, we show the existence of a separation oracle. Note that it suffices to show that for each $i = 1, \ldots, m$, the following convex set has a separation oracle:

$$\mathcal{C}_i \triangleq \{\mathbf{x}|(\mathbf{a}_i + \boldsymbol{\delta}_i)^\top \mathbf{x} \ge b_i - t_i(\|\boldsymbol{\delta}_i\|_{\Sigma_i^{-1}}),\ \forall \boldsymbol{\delta}_i \in \mathbb{R}^n\}.$$

Fix a candidate solution $\mathbf{x}_0$, we have

$$\min_{\boldsymbol{\delta}_i \in \mathbb{R}^n} \big[\boldsymbol{\delta}_i^\top \mathbf{x}_0 + t_i(\|\boldsymbol{\delta}_i\|_{\Sigma_i^{-1}})\big] = \min_{y \ge 0} \min_{\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}} = y} \big[\boldsymbol{\delta}_i^\top \mathbf{x}_0 + t_i(\|\boldsymbol{\delta}_i\|_{\Sigma_i^{-1}})\big] = \min_{y \ge 0} \big[-y\sqrt{\mathbf{x}_0^\top \Sigma_i \mathbf{x}_0} + t_i(y)\big],$$

where the last equality is achieved at $\boldsymbol{\delta}_i = -y\Sigma_i \mathbf{x}_0 / \sqrt{\mathbf{x}_0^\top \Sigma_i \mathbf{x}_0}$.

Therefore, $\mathbf{x}_0 \in \mathcal{C}_i$ if and only if

$$\min_{y \geq 0} \big[-y\sqrt{\mathbf{x}_0^\top \Sigma_i \mathbf{x}_0} + t_i(y) \big] \geq b_i - \mathbf{a}_i^\top \mathbf{x}_0.$$

In addition, if $\mathbf{x}_0 \notin \mathcal{C}$, then take $\boldsymbol{\delta}_i^* = -y^* \Sigma_i \mathbf{x}_0 / \sqrt{\mathbf{x}_0^\top \Sigma_i \mathbf{x}_0}$ where $y^* = \arg\min \big[-y\sqrt{\mathbf{x}_0^\top \Sigma_i \mathbf{x}_0} + t_i(\gamma) \big]$, we have that

$$\boldsymbol{\delta}_i^{*\top} \mathbf{x}_0 + t_i(\|\boldsymbol{\delta}_i^*\|_{\Sigma_i^{-1}}) < b_i - \mathbf{a}_i^\top \mathbf{x}_0,$$

whereas for any $\mathbf{z} \in \mathcal{C}_i$, we have

$$\boldsymbol{\delta}_i^{*\top} \mathbf{z} + t_i(\|\boldsymbol{\delta}_i^*\|_{\Sigma_i^{-1}}) \geq b_i - \mathbf{a}_i^\top \mathbf{z}.$$

That is, the following hyperplane separates $\mathbf{x}_0$ and $\mathcal{C}_i$,

$$(\boldsymbol{\delta}_i^* + \mathbf{a}_i)^\top \mathbf{x} = b_i - t_i(\|\boldsymbol{\delta}_i^*\|_{\Sigma_i^{-1}}).$$

Thus, since in polynomial time we can solve

$$\text{minimize: } t_i(y) - \beta y, \text{ subject to: } y \in \mathbb{R}^+;$$

we have a polynomial-time separation Oracle, which implies that Problem (10) is tractable. Q.E.D.

Proof of Theorem 9: We establish the first claim. Since $\mathfrak{P}(s) = 1 - \gamma \exp(-\alpha s)$, we have $\mathfrak{P}^{-1}(t) = \max[-\frac{1}{\alpha}\log(1-t) + \frac{\log\gamma}{\alpha}, 0]$. Thus, the penalty function is

$$t(s) \triangleq \mathfrak{P}^{-1}\Big(\frac{s^2}{1+s^2}\Big) = \max\Big[\frac{1}{\alpha}\log(1+s^2) + \frac{\log\gamma}{\alpha}, 0\Big].$$

It is easy to check, $\frac{1}{\alpha}\log(1+s^2) + \frac{\log\gamma}{\alpha}$ is convex on $[0,1]$, and concave on $[1,\infty)$.

Let $\mathcal{C}$ be the feasible set of Constraint (14), then by Theorem 4 we have

$$\mathcal{C} \triangleq \{\mathbf{x} | (\mathbf{a} + \boldsymbol{\delta})^\top \mathbf{x} \geq b - t(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}),\ \forall \boldsymbol{\delta} \in \mathbb{R}^n\}.$$

From the proof of Theorem 7, a candidate solution $\mathbf{x}_0 \in \mathcal{C}$ if and only if

$$\min_{y \geq 0} \big[-y\sqrt{\mathbf{x}_0^\top \Sigma \mathbf{x}_0} + t(y) \big] \geq b_i - \mathbf{a}^\top \mathbf{x}_0.$$

Note that Σ is full rank and positive semi-definite, which implies that $\sqrt{\mathbf{x}_0^\top \Sigma \mathbf{x}_0} > 0$ for any non-zero $\mathbf{x}_0$. Hence, substituting $t(s) = \frac{1}{\alpha}\log(1+s^2) + \frac{\log\gamma}{\alpha}$ we have that for for any non-zero $\mathbf{x}_0$,

$$\lim_{y \to +\infty} \big[-y\sqrt{\mathbf{x}_0^\top \Sigma \mathbf{x}_0} + t(y) \big] = -\infty,$$

which establishes the first claim.

We turn to establish the second claim. To apply Theorem 4 requires a probabilistic envelope function on $\mathbb{R}^+$ instead of $[s^-, s^+]$. Therefore, we let

$$\hat{\mathfrak{P}}(s) \triangleq \begin{cases} 0 & \text{if } s < s^-; \\ 1-\gamma\exp(-\alpha s) & \text{if } s^- \le s \le s^+; \\ 1-\gamma\exp(-\alpha s^+) & \text{if } s > s^+; \end{cases}$$

which is non-decreasing and continuous from the right. In addition, observe that Constraint (15) is equivalent to

$$\inf_{\boldsymbol{\delta}^r \sim (0,\Sigma)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top \mathbf{x} \ge b - s\big] \ge \hat{\mathfrak{P}}(s),\ \forall s \ge 0.$$

By definition, we have that

$$\hat{\mathfrak{P}}^{-1}(x) = \inf\{y \ge 0 | \hat{\mathfrak{P}}(y) \ge x\} = \begin{cases} 0 & \text{if } x = 0; \\ s^- & \text{if } 0 < x < 1-\gamma\exp(-\alpha s^-); \\ \max[-\frac{1}{\alpha}\log(1-x) + \frac{\log\gamma}{\alpha}, 0] & \text{if } 1-\gamma\exp(-\alpha s^-) \le x \le 1-\gamma\exp(-\alpha s^+); \\ +\infty & \text{if } 1-\gamma\exp(-\alpha s^+) < x. \end{cases}$$

Note that when $x \ge 1-\gamma\exp(-\alpha s^-)$, we have $-\frac{1}{\alpha}\log(1-x) + \frac{\log\gamma}{\alpha} \ge s^- \ge 0$. This leads to

$$\begin{aligned} \hat{t}(s) &= \hat{\mathfrak{P}}^{-1}\Big(\frac{s^2}{1+s^2}\Big) \\ &= \begin{cases} 0 & \text{if } s = 0; \\ s^- & \text{if } 0 < s < \sqrt{\frac{1}{\gamma\exp(-\alpha s^-)} - 1}; \\ \frac{1}{\alpha}\log(1+s^2) + \frac{\log\gamma}{\alpha} & \text{if } \sqrt{\frac{1}{\gamma\exp(-\alpha s^-)} - 1} \le s \le \sqrt{\frac{1}{\gamma\exp(-\alpha s^+)} - 1}; \\ +\infty & \text{if } \sqrt{\frac{1}{\gamma\exp(-\alpha s^+)} - 1} \le s. \end{cases} \end{aligned}$$

Theorem 7 asserts that Constraint (15) leads to a tractable optimization problem if for any $\beta \ge 0$, minimize:$_{s\ge 0}\,[\hat{t}(s) - \beta s]$ is tractable. By the definition of $\hat{t}(s)$, it thus suffices to show that the following is tractable:

$$\begin{aligned} \text{Minimize:} \quad & \frac{1}{\alpha}\log(1+s^2) + \frac{\log\gamma}{\alpha} - \beta s \\ \text{Subject to:} \quad & \sqrt{\frac{1}{\gamma\exp(-\alpha s^-)} - 1} \le s \le \sqrt{\frac{1}{\gamma\exp(-\alpha s^+)} - 1}. \end{aligned} \tag{30}$$

Recall that $\frac{1}{\alpha}\log(1+s^2) + \frac{\log\gamma}{\alpha}$ is convex on $[0,1]$ and concave on $[1,+\infty)$. Thus, due to convexity, by line search in polynomial time we can solve the scalar-variable optimization problem

$$\begin{aligned} \text{Minimize:} \quad & \frac{1}{\alpha}\log(1+s^2) + \frac{\log\gamma}{\alpha} - \beta s \\ \text{Subject to:} \quad & \sqrt{\frac{1}{\gamma\exp(-\alpha s^-)} - 1} \le s \le \sqrt{\frac{1}{\gamma\exp(-\alpha s^+)} - 1} \\ & 0 \le s \le 1. \end{aligned}$$

Due to concavity, we can solve (by checking the two extreme points of the feasible set)

$$\begin{aligned} \text{Minimize:} \quad & \frac{1}{\alpha}\log(1+s^2)+\frac{\log\gamma}{\alpha}-\beta s \\ \text{Subject to:} \quad & \sqrt{\frac{1}{\gamma\exp(-\alpha s^-)}-1}\le s\le\sqrt{\frac{1}{\gamma\exp(-\alpha s^+)}-1} \\ & s\ge 1. \end{aligned}$$

Hence in polynomial time we can solve (30), which implies the second claim. Q.E.D.

Appendix C: Proof of results in Section 5

In this appendix we prove Theorem 11. We start with a simple lemma.

Lemma 3. *Let $c_2>0$ and $\epsilon=|c_1|/c2<1$, then*

$$(1-\epsilon/2)[\Phi(c_2)-\Phi(-c_2)]\le\Phi(c_1+c_2)-\Phi(c_1-c_2)\le\Phi(c_2)-\Phi(-c_2).$$

Proof: Note that by symmetry of the density function of Gaussian random variable, we have that

$$\Phi(c_1+c_2)-\Phi(c_1-c_2)=\Phi(-c_1+c_2)-\Phi(-c_1-c_2).$$

Thus, without loss of generality we assume that $c_1\ge 0$. By definition of the cumulative distribution function, we have that

$$\begin{aligned} &\Phi(c_2)-\Phi(-c_2)-[\Phi(c_1+c_2)-\Phi(c_1-c_2)] \\ &\quad=\int_{-c_2}^{c_1-c_2}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,dx-\int_{c_2}^{c_1+c_2}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,dx \\ &\quad\ge c_1\frac{1}{\sqrt{2\pi}}\exp(-c_2^2/2)-c_1\frac{1}{\sqrt{2\pi}}\exp(-c_2^2/2)=0, \end{aligned}$$

where, the inequality holds due to the fact that the pdf of a Gaussian random variable is decreasing when positive, and increasing when negative. Hence we proved the right hand side of the lemma.

Again using the monotonicity of the pdf of Gaussian distribution, we have

$$\Phi(c_1-c_2)-\Phi(-c_2)=\int_{-c_2}^{c_1-c_2}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,dx\le\frac{c_1}{c_2}\int_{-c_2}^{0}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,dx=\frac{\epsilon(\Phi(c_2)-\Phi(-c_2))}{2}.$$

Thus,

$$\begin{aligned} &\Phi(c_1+c_2)-\Phi(c_1-c_2) \\ &\quad\ge\Phi(c_2)-\Phi(c_1-c_2) \\ &\quad=\Phi(c_2)-\Phi(-c_2)-[\Phi(c_1-c_2)-\Phi(-c_2)] \\ &\quad\ge(1-\epsilon/2)(\Phi(c_2)-\Phi(-c_2)), \end{aligned}$$

which establishes the left hand side of the lemma. Q.E.D.

Now we turn to prove Theorem 11.

Proof of Theorem 11: Let $\hat{\mathfrak{P}}(s) = (\mathfrak{P}(s)+1)/2$, which is also non-decreasing and continuous from right then we have

$$\mathfrak{P}^{-1}\left[2(\Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}}) - 1/2)\right] = \hat{\mathfrak{P}}^{-1}\left[\Phi(\|\boldsymbol{\delta}\|_{\Sigma^{-1}})\right],$$

which by Theorem 5 implies that Constraint (21) is equivalent to

$$\begin{aligned}
&\mathbb{P}\big[\boldsymbol{\delta}^{r\top}\mathbf{x} \geq -s\big) \geq \hat{\mathfrak{P}}(s) = (\mathfrak{P}(s)+1)/2; \quad \forall s \geq 0;\\
&b - c \leq \mathbf{a}^\top \mathbf{x} \leq b + c.
\end{aligned}$$

Since $\boldsymbol{\delta}^r \sim \mathcal{N}(0, \Sigma)$, Constraint (21) is thus equivalent to

$$\begin{aligned}
&\Phi\left(\frac{s}{\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}}\right) \geq \mathfrak{P}(s)/2 + 1/2; \quad \forall s \geq 0;\\
&b - c \leq \mathbf{a}^\top \mathbf{x} \leq b + c.
\end{aligned} \tag{31}$$

Similarly, Constraint (20) is equivalent to

$$\begin{aligned}
&\Phi\left(\frac{s}{\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}}\right) \geq \mathfrak{P}(s)/(2-\epsilon) + 1/2; \quad \forall s \geq 0;\\
&b - c \leq \mathbf{a}^\top \mathbf{x} \leq b + c.
\end{aligned} \tag{32}$$

Fix a $\mathbf{x}$ and a $s > 0$, then

$$\begin{aligned}
\mathbb{P}(b - s \leq (\mathbf{a} + \boldsymbol{\delta}^r)^\top \mathbf{x} \leq b + s) =& \mathbb{P}(b - \mathbf{a}^\top \mathbf{x} - s \leq \boldsymbol{\delta}^{r\top}\mathbf{x} \leq b - \mathbf{a}^\top \mathbf{x} + s)\\
=& \mathbb{P}(b - \mathbf{a}^\top \mathbf{x} - s \leq z \leq b - \mathbf{a}^\top \mathbf{x} + s),
\end{aligned}$$

where z is a random variable follows $\mathcal{N}(0, \sqrt{\mathbf{x}^\top \Sigma \mathbf{x}})$. Let $\hat{z} \sim \mathcal{N}(0,1)$, the right hand side equals

$$\begin{aligned}
&\mathbb{P}((b - \mathbf{a}^\top \mathbf{x} - s)/\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}} \leq \hat{z} \leq (b - \mathbf{a}^\top \mathbf{x} + s)/\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}})\\
&\quad = \Phi((b - \mathbf{a}^\top \mathbf{x} + s)/\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}) - \Phi((b - \mathbf{a}^\top \mathbf{x} - s)/\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}).
\end{aligned}$$

Note that for $\mathbf{x}$ satisfying that $|b - \mathbf{a}\mathbf{x}| \leq c$, we have that

$$(1-\epsilon/2)(2\Phi\left(\frac{s}{\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}}\right) - 1) \overset{(a)}{\leq} \Phi((b - \mathbf{a}^\top \mathbf{x} + s)/\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}) - \Phi((b - \mathbf{a}^\top \mathbf{x} - s)/\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}) \overset{(b)}{\leq} 2\Phi\left(\frac{s}{\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}}\right) - 1.$$

From (a), if $\Phi\left(\frac{s}{\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}}\right) \geq \mathfrak{P}(s)/(2-\epsilon) + 1/2$, then

$$\Phi((b - \mathbf{a}^\top \mathbf{x} + s)/\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}) - \Phi((b - \mathbf{a}^\top \mathbf{x} - s)/\sqrt{\mathbf{x}^\top \Sigma \mathbf{x}}) \geq \mathfrak{P}(s).$$

This implies that Constraint (32), and equivalently Constraint (20) bounds the original constraint from the inside. Similarly, (b) implies that Constraint (21) bounds the original constraint from the outside. Q.E.D.

Appendix D: Approximation of the robust linear regression problem

In this appendix we provide the detailed derivation of the approximation of the robust linear regression. Recall that as we discussed in Section 5.2, when $\min\{s|\mathfrak{P}(s)>0\}$ is large, the constraint

$$\mathbb{P}\big(|b_i-(\mathbf{a}_i+\boldsymbol{\delta}_i^r)^\top\mathbf{x}|\le\xi_i+s\big)\ge\mathfrak{P}(s),\quad\forall s\ge 0,$$

can be approximated by

$$\boldsymbol{\delta}^\top\mathbf{x}\ge-\mathfrak{P}^{-1}\left[2(\Phi(\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})-1/2)\right];\quad\forall\boldsymbol{\delta}_i\in\mathbb{R}^n. \tag{33}$$

Notice that in Problem (2), we have

$$\mathfrak{P}(s)=\begin{cases}0 & \text{if } s<s^*;\\ 1-\gamma\exp(-\alpha s) & \text{if } s\ge s^*.\end{cases}$$

Thus, by definition we have

$$\mathfrak{P}^{-1}(s)=\begin{cases}0 & \text{if } x=0;\\ s^* & \text{if } 0<x<1-\gamma\exp(-\alpha s^*);\\ -\frac{1}{\alpha}\log(1-x)+\frac{\log\gamma}{\alpha} & \text{if } 1-\gamma\exp(-\alpha s^*)\le x.\end{cases}$$

Therefore, Constraint (33) is equivalent to

$$\begin{array}{ll}(a)\ \boldsymbol{\delta}^\top\mathbf{x}\ge 0 & \text{if } 2(\Phi(\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})-1/2)=0;\\ (b)\ \boldsymbol{\delta}^\top\mathbf{x}\ge -s^* & \text{if } 0<2(\Phi(\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})-1/2)<1-\gamma\exp(-\alpha s^*);\\ (c)\ \boldsymbol{\delta}^\top\mathbf{x}\ge\frac{1}{\alpha}\log(1-2(\Phi(\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})-1/2))-\frac{\log\gamma}{\alpha} & \text{if } 1-\gamma\exp(-\alpha s^*)\le 2(\Phi(\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})-1/2).\end{array}$$

Note that Constraint (a) is superfluous, and Constraint (b) is implied by

$$\boldsymbol{\delta}^\top\mathbf{x}\ge\frac{1}{\alpha}\log(1-2(\Phi(\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})-1/2))-\frac{\log\gamma}{\alpha}\quad\text{if } 1-\gamma\exp(-\alpha s^*)=2(\Phi(\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}})-1/2).$$

Thus, Constraint (33) reduced to (c), which can be further simplified to

$$\boldsymbol{\delta}_i^\top\mathbf{x}\ge\frac{1}{\alpha}\log(2\Phi(-\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}}))-\frac{\log\gamma}{\alpha};\quad\forall\|\boldsymbol{\delta}_i\|_{\Sigma^{-1}}\ge\Phi^{-1}\left(1-\frac{\gamma\exp(-\alpha s^*)}{2}\right),$$

which leads to the approximation formulation.

Appendix E: Proofs to results in Section 6

In this appendix we prove Theorem 12.

Proof of Theorem 12: Fix x and s and let c^* be such that $c_i^*=-\text{sign}(x_i)\epsilon_i$. We note that

$$\begin{aligned}\inf_{\mathbf{c}:c_i\in[-\epsilon_i,\epsilon_i],\Sigma\preceq\Sigma^*}\ \inf_{\boldsymbol{\delta}^r\sim(c,\Sigma)}\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top\mathbf{x}\ge b-s\big]&=\inf_{\boldsymbol{\delta}^r\sim(c^*,\Sigma^*)}\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^r)^\top\mathbf{x}\ge b-s\big]\\&=\inf_{\boldsymbol{\delta}^{r'}\sim(0,\Sigma^*)}\mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^{r'})^\top\mathbf{x}\ge b-s+\sum_{i=1}^n\epsilon_i|x_i|\big].\end{aligned}$$

Thus, Constraint (23) is equivalent to

$$\inf_{\boldsymbol{\delta}^{r'}\sim(0,\Sigma^*)} \mathbb{P}\big[(\mathbf{a}+\boldsymbol{\delta}^{r'})^\top\mathbf{x} \geq b - s + \sum_{i=1}^n \epsilon_i|x_i|\big] \geq 1 - \frac{1}{\big(t^{-1}(s)\big)^2+1},\ \forall s \geq 0.$$

By Theorem 1, this is equivalent to

$$(\mathbf{a}+\boldsymbol{\delta})^\top\mathbf{x} \geq b - t(\|\boldsymbol{\delta}\|_{(\Sigma^*)^{-1}}) + \sum_{i=1}^n \epsilon_i|x_i|,\ \forall \boldsymbol{\delta}\in\mathbb{R}^n. \tag{34}$$

Therefore, it suffices to show that (22) and (34) are equivalent to establish the theorem. We prove this by showing that if $\mathbf{x}$ violates one constraint, then it must violate the other.

Suppose $\mathbf{x}$ violates (34), i.e., for some $\boldsymbol{\delta}'$ the following holds

$$(\mathbf{a}+\boldsymbol{\delta}')^\top\mathbf{x} < b - t(\|\boldsymbol{\delta}'\|_{(\Sigma^*)^{-1}}) + \sum_{i=1}^n \epsilon_i|x_i|.$$

Let $\boldsymbol{\delta}$ be such that $\delta_i = \delta_i' - \epsilon_i \text{sign}(x_i)$, then we have

$$(\mathbf{a}+\boldsymbol{\delta})^\top\mathbf{x} = (\mathbf{a}+\boldsymbol{\delta}')^\top\mathbf{x} - \sum_{i=1}^n \epsilon_i|x_i| < b - t(\|\boldsymbol{\delta}'\|_{(\Sigma^*)^{-1}}) \leq b - t(\|T(\boldsymbol{\delta})\|_{(\Sigma^*)^{-1}}),$$

where the last inequality follows from the definition of $T(\cdot)$. Thus, $\mathbf{x}$ violates Constraint (22).

Conversely, suppose now that $\mathbf{x}$ violates Constraint (22), i.e., for some $\boldsymbol{\delta}'$ the following holds

$$(\mathbf{a}+\boldsymbol{\delta}')^\top\mathbf{x} < b - t(\|T(\boldsymbol{\delta}')\|_{(\Sigma^*)^{-1}}).$$

We have

$$(\mathbf{a}+T(\boldsymbol{\delta}'))^\top\mathbf{x} < b - t(\|T(\boldsymbol{\delta}')\|_{(\Sigma^*)^{-1}}) + (T(\boldsymbol{\delta}')-\boldsymbol{\delta}')^\top\mathbf{x} \leq b - t(\|T(\boldsymbol{\delta}')\|_{(\Sigma^*)^{-1}}) + \sum_{i=1}^n \epsilon_i|x_i|,$$

where the last inequality follows from the definition of $T(\cdot)$. Letting $\delta = T(\boldsymbol{\delta}')$, $\mathbf{x}$ violates (34), which completes the proof. Q.E.D.

Acknowledgments

We thank the AE and the referees for their comments which lead to improvements of this paper. We thank the Area Editor for pointing out the relationship of the proposed model to stochastic dominance constraints. We are grateful to Prof. Aharon Ben-Tal for carefully reading an early version of the manuscript and providing insightful comments and suggestions. The research of H. Xu was partially supported by NUS startup grant R-265-000-384-133. The research of C. Caramanis was partially supported by NSF grants EFRI-0735905, EECS-1056028, and DTRA grant HDTRA 1-08-0029. The research of S. Mannor was partially supported by the Israel Science Foundation (contract 890015).

References

Artzner, P., F. Delbaen, J. Eber, D. Heath. 1999. Coherent measures of risk. *Mathematical Finance* **9**(3) 203–228.

Ben-Tal, A., D. Bertsimas, D. Brown. 2010. A soft robust model for optimization under ambiguity. *Operations Research* **58**(4-Part-2) 1220–1234.

Ben-Tal, A., S. Boyd, A. Nemirovski. 2006. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming, Series B* **107**(1-2) 63–89.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. *Robust Optimization*. Princeton University Press.

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* **23**(4) 769–805.

Ben-Tal, A., A. Nemirovski. 1999. Robust solutions of uncertain linear programs. *Operations Research Letters* **25**(1) 1–13.

Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming, Series A* **88**(3) 411–424.

Bertsimas, D., D. B. Brown. 2009. Constructing uncertainty sets for robust linear optimization. *Operations Research* **57**(6) 1483–1495.

Bertsimas, D., D. B. Brown, C. Caramanis. 2011. Theory and applications of robust optimization. *SIAM Review* **53**(3) 464–501.

Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Brown, D.B., M. Sim. 2009. Satisficing measures for analysis of risky positions. *Management Science* **55**(1) 71–84.

Charnes, A., W. W. Cooper. 1959. Chance constrained programming. *Management Science* **6**(1) 73–79.

Chen, W., M. Sim. 2009. Goal driven optimization. *Operations Research* **57**(2) 342–357.

Chen, X., M. Sim, P. Sun. 2007. A robust optimization perspective on stochastic programming. *Operations Research* **55**(6) 1058–1107.

Chen, X., M. Sim, P. Sun, J. Zhang. 2008. A linear-decision based approximation approach to stochastic programming. *Operations Research* **56**(2) 344–357.

Delage, E., S. Mannor. 2010. Percentile optimization for Markov decision processes with parameter uncertainty. *Operations Research* **58**(1) 203–213.

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with applications to data-driven problems. *Operations Research* **58**(3) 596–612.

Delbaen, F. 2002. Coherent risk measures on general probability spaces. K. Sandmann, P. J. Schönbucher, eds., *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*. Springer, 1–38.

Dentcheva, D., A. Ruszczyński. 2003. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization* **14**(2) 548–566.

Dentcheva, D., A. Ruszczyński. 2004a. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming, Series A* **99**(2) 329–350.

Dentcheva, D., A. Ruszczyński. 2004b. Semi-infinite probabilistic optimization: First order stochastic dominance constraints. *Optimization* **53**(5-6) 433–451.

El Ghaoui, L., M. Oks, F. Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research* **51**(4) 543–556.

El Ghaoui, L., F. Oustry, H. Lebret. 1998. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization* **9**(1) 33–52.

Evans, R., A. Cantoni, T. Fortmann. 1977a. Envelope-constrained filters – ii: Adaptive structures. *IEEE Transactions on Information Theory* **23**(4) 435–433.

Evans, R., T. Fortmann, A. Cantoni. 1977b. Envelope-constrained filters – i: Theory and applications. *IEEE Transactions on Information Theory* **23**(4) 421–434.

Föllmer, H., A. Schied. 2002. *Stochastic finance: An introduction in discrete time*. Berlin: Walter de Gruyter.

Grötschel, M., L. Lovász, A. Schrijver. 1988. *Geometric Algorithms and Combinatorial Optimization*. Springer, Heidelberg.

Lanzillotti, R. F. 1958. Pricing objectives in large companies. *American Economic Review* **48**(5) 921–940.

Lüthi, H.-J., J. Doege. 2005. Convex risk measures for portfolio optimization and concepts of flexibility. *Mathematical Programming Series B* **104**(2-3) 541–559.

Mao, J. 1970. Survey of capital budgeting: Theory and practice. *Journal of Finance* **25**(2) 349–360.

Marshall, A. W., I. Olkin. 1960. Multivariate Chebyshev inequalities. *Annuals of Mathematical Statistics* **31**(4) 1001–1014.

Miller, B., H. Wagner. 1965. Chance constrained programming with joint constraints. *Operations Research* **13**(6) 930–945.

Novosyolov, A. 2002. Lectures on risk theory. Available online at http://risktheory.narod.ru/download.htm.

Paschalidis, I., S.-C. Kang. 2005. Robust linear optimization: On the benefits of distributional information and applications in inventory control. *Proceedings of the 44th IEEE Conference on Decision and Control*. 4416–4421.

Payne, J. W., D. J. Laughhunn, R. Crum. 1980. Translation of gambles and aspiration level effects in risky choice behavior. *Management Science* **26**(10) 1039–1060.

Payne, J. W., D. J. Laughhunn, R. Crum. 1981. Further tests of aspiration level effects in risky choice behavior. *Management Science* **27**(8) 953–958.

Prékopa. 1970. On probabilistic constrained programming. *In Proceedings of the Princeton Symposium on Mathematical Programming*. 113–138.

Prékopa, A. 1995. *Stochastic Programming*. Kluwer.

Rockafellar, R.T., S.P. Uryasev. 2000. Optimization of conditional value-at-risk. *The Journal of Risk* **2**(3) 21–41.

Ruszczynski, A., A. Shapiro. 2006. Optimization of convex risk functions. *Mathematics of Operations Research* **31**(3) 433–452.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia.

Shivaswamy, P. K., C. Bhattacharyya, A. J. Smola. 2006. Second order cone programming approaches for handling missing and uncertain data. *Journal of Machine Learning Research* **7**(Jul) 1283–1314.

Simon, H. A. 1959. Theories of decision-making in economics and behavioral science. *American Economic Review* **49**(3) 253–283.

Vu, B.-N., A. Cantoni, K. L. Teo. 1997. Envelope constrained filter with linear interpolator. *IEEE Transactions on Signal Processing* **45**(6) 1405–1414.

Zymler, S., D. Kuhn, B. Rusterm. 2011. Distributionally robust joint chance constraints with second-order moment information. Forthcoming in *Mathematical Programming*.