Convex Chance-Constrained Programs with Wasserstein Ambiguity

Haoming Shen and **Ruiwei Jiang**
Department of Industrial and Operations Engineering
University of Michigan, Ann Arbor, MI 48109
Email: {hmshen, ruiwei}@umich.edu

Abstract

Chance constraints yield non-convex feasible regions in general. In particular, when the uncertain parameters are modeled by a Wasserstein ball, [Xie19] and [CKW18] showed that the distributionally robust (pessimistic) chance constraint admits a mixed-integer conic representation. This paper identifies sufficient conditions that lead to *convex* feasible regions of chance constraints with Wasserstein ambiguity. First, when uncertainty arises from the left-hand side of a pessimistic *individual* chance constraint, we derive a convex and conic representation if the Wasserstein ball is centered around a Gaussian distribution. Second, when uncertainty arises from the right-hand side of a pessimistic *joint* chance constraint, we show that the ensuing feasible region is convex if the Wasserstein ball is centered around a log-concave distribution (or, more generally, an α-concave distribution with $\alpha \geq -1$). In addition, we propose a block coordinate ascent algorithm for this class of chance constraints and prove its convergence to global optimum. Furthermore, we extend the convexity results and conic representation to *optimistic* chance constraints.

Keywords: Chance constraints; Convexity; Wasserstein ambiguity; Distributionally robust optimization; Distributionally optimistic optimization

1 Introduction

Many optimization models include safety principles taking the form

$$A(x)\xi \leq b(x),$$

where $x \in \mathbb{R}^n$ represents decision variables, $\xi \in \mathbb{R}^m$ represents model parameters, and $A(x) \in \mathbb{R}^{m\times n}$ and $b(x) \in \mathbb{R}^m$ are affine functions of x. When ξ is subject to uncertainty and follows a probability distribution $\mathbb{P}_{\text{true}}$, a convenient way of protecting these safety principles is to use chance constraint

$$\mathbb{P}_{\text{true}}\Big[A(x)\xi \leq b(x)\Big] \geq 1-\epsilon, \tag{\textbf{CC}}$$

where $1-\epsilon \in (0,1)$ represents a pre-specified risk threshold. (**CC**) requires to satisfy all safety principles with high probability (i.e., $1-\epsilon$ is usually close to one, e.g., 0.95). (**CC**) was first studied in the 1950s [CC59; CCS58; MW65; Pré70] and finds a wide range of applications in, e.g., power system [WGW11], vehicle routing [SG83], scheduling [DS16], portfolio management [Li95], and facility location [MG06]. We mention two examples.

Example 1. (Portfolio Management) Suppose that we manage a portfolio among n stocks and, for each $i \in [n] := \{1, \ldots, n\}$, x_i represents the amount of investment in stock i, which yields a random return ξ_i. Then, chance constraint

$$\mathbb{P}_{\text{true}}\left[x^{\mathsf{T}}\xi \geq \eta\right] \geq 1 - \epsilon \tag{\textbf{PTO}}$$

assures that we receive at least η dollar in return with high probability. Here, $m = 1$, and accordingly $A(x)$ in (**CC**) reduces to the row vector $-x^{\mathsf{T}}$ and $b(x)$ reduces to the scalar $-\eta$.

Example 2. (Production Planning) Suppose that we produce certain commodity at n facilities to serve m demand locations. If x_j denotes the production capacity of facility j and T_{ij} denotes the service coverage of facility j for location i (i.e., $T_{ij} = 1$ if facility j can serve location i and $T_{ij} = 0$ otherwise) for all $i \in [m]$ and $j \in [n]$, then chance constraint

$$\mathbb{P}_{\text{true}}\left[Tx \geq \xi\right] \geq 1 - \epsilon \tag{\textbf{PP}}$$

assures that the production capacities are able to satisfy the demands ξ at all locations. Here, $A(x)$ in (**CC**) equals the $m \times m$ identify matrix and $b(x)$ equals Tx.

In (**PP**), the random vector ξ is decoupled from the decision variables x because, in this example, $A(x)$ is independent of x. For such chance constraints with $A(x) \equiv A$, we follow the convention in the literature and refer to them as chance constraints with right-hand side (RHS) uncertainty. In contrast, ξ and x are multiplied in (**PTO**). To distinguish chance constraints in this form from those with RHS uncertainty, we call them chance constraints with left-hand side (LHS) uncertainty. In addition, we say a chance constraint is *individual* if $m = 1$ (such as in (**PTO**)) and *joint* if $m \geq 2$ (such as in (**PP**)).

Although (**CC**) provides an intuitive way to model uncertainty in safety principles, it produces a non-convex feasible region in general, giving rise to concerns of challenging computation. To this end, a stream of prior work proposed effective mixed-integer programming (MIP) approaches based on the notions of, e.g., sample average approximation [LA08; LAN08] and p-efficient points [Pré90; BR02], and derived valid inequalities to strengthen the ensuing MIP formulations (see, e.g., [Küç12; Lue14] and a recent survey [KJ21]). Another stream of prior work identified sufficient conditions for (**CC**) to produce a convex feasible region. For individual (**CC**), [PP63] derived a second-order conic (SOC) representation when ξ follows a Gaussian distribution, and [LLS01] and [CE06] further extended this result when ξ follows an elliptical log-concave distribution (see Definition 2). For joint (**CC**) with RHS uncertainty, [Pré13] (see his Theorem 10.2) proved the convexity of the ensuing feasible region when ξ follows a log-concave distribution, examples of which include Gaussian, exponential, beta (if both shape parameters are at least 1), uniform on convex support, etc. Furthermore, [SDR09] generalized this result to α-concave distributions (see Definition 2).

In most practical problems, the (true) distribution $\mathbb{P}_{\text{true}}$ of the random parameters ξ is unknown or ambiguous to the modeler, who often replaces $\mathbb{P}_{\text{true}}$ in (**CC**) with a crude estimate, denoted by $\mathbb{P}$. Candidates of $\mathbb{P}$ includes the empirical distribution based on past observations of ξ and Gaussian distribution, whose mean and covariance matrix can be estimated based on these past observations. Since $\mathbb{P}$ may not perfectly model the uncertainty of ξ, it is reasonable to take into account its neighborhood, or more formally, an ambiguity set $\mathcal{P}$ around $\mathbb{P}$. In this paper, we adopt a Wasserstein ambiguity set defined as

$$\mathcal{P} := \left\{ \mathbb{Q} \in \mathcal{P}_0 \colon d_W(\mathbb{Q}, \mathbb{P}) \leq \delta \right\},$$

where $\mathcal{P}_0$ is the set of all probability distributions, $\delta > 0$ is a pre-specified radius of $\mathcal{P}$, and $d_W(\cdot, \cdot)$ represents the Wasserstein distance (see, e.g., [MK18]). Specifically, the Wasserstein distance between two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined through

$$d_W(\mathbb{P}_1, \mathbb{P}_2) := \inf_{\mathbb{P}_0 \sim (\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{\mathbb{P}_0}\Big[\|X_1 - X_2\|\Big], \tag{1}$$

where X_1, X_2 are two random variables following distributions $\mathbb{P}_1$, $\mathbb{P}_2$ respectively, $\mathbb{P}_0$ is the coupling of $\mathbb{P}_1$ and $\mathbb{P}_2$, and $\|\cdot\|$ represents a norm. $d_W(\mathbb{P}_1, \mathbb{P}_2)$ can be interpreted as the minimum cost, with respect to $\|\cdot\|$, of transporting the probability masses of $\mathbb{P}_1$ to recover $\mathbb{P}_2$. Hence, the Wasserstein ambiguity set $\mathcal{P}$ is a ball (in the space of probability distributions) centered around $\mathbb{P}$, which for this reason is referred to as the reference distribution. Additionally, $\mathcal{P}$ may include the true distribution $\mathbb{P}_{\text{true}}$, i.e., $\mathbb{P}_{\text{true}} \in \mathcal{P}$, when the radius δ is large enough. As a result, the pessimistic counterpart

$$\inf_{\mathbb{Q} \in \mathcal{P}} \mathbb{Q}\left[A(x)\xi \leq b(x)\right] \geq 1 - \epsilon \tag{\textbf{P-CC}}$$

implies (**CC**) because it requires that (**CC**) holds with respect to all distributions in $\mathcal{P}$. In contrast, an optimistic modeler may be satisfied as long as there exists some distribution in $\mathcal{P}$, with respect to which (**CC**) holds. This gives rise to the following optimistic counterpart of (**CC**):

$$\sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{Q}\left[A(x)\xi \leq b(x)\right] \geq 1 - \epsilon. \tag{\textbf{O-CC}}$$

In the existing literature, (**P-CC**) is also known as distributionally robust chance constraint and, depending on the value of m and the ambiguity set $\mathcal{P}$, the feasible region of (**P-CC**) may be convex or non-convex. For individual (**P-CC**) (i.e., $m = 1$), convex representations have been derived when $\mathcal{P}$ is Chebyshev, i.e., when $\mathcal{P}$ consists of all distributions sharing the same mean and covariance matrix of ξ. Specifically, [EOO03; CE06] derived semidefinite and SOC representations of (**P-CC**) with a Chebyshev $\mathcal{P}$. With the same ambiguity set, [ZKR11] showed that (**P-CC**) is equivalent to its approximation based on conditional Value-at-Risk (CVaR) even when the safety principle becomes nonlinear in ξ. Additionally, [Han+15] and [LJM19] incorporated structural information (e.g., unimodality) into the Chebyshev $\mathcal{P}$ and derived semidefinite and SOC representations of (**P-CC**), respectively. For joint (**P-CC**) (i.e., $m \geq 2$), however, convexity results become scarce. [Han+17] characterized $\mathcal{P}$ by a conic support, the mean, and a positively homogeneous dispersion measure of ξ, and showed that (**P-CC**) with RHS uncertainty is conic representable. In addition, they showed that this result falls apart if one relaxes these conditions even in a mildest possible manner. More recently, [XA16] extended the convexity result when the safety principles depend on ξ nonlinearly and $\mathcal{P}$ is characterized by a single moment constraint of ξ. In this paper, we study (**P-CC**) and (**O-CC**) with $\mathcal{P}$ being a Wasserstein ambiguity set.

To the best of our knowledge, the convexity results for either (**P-CC**) or (**O-CC**) with Wasserstein ambiguity do not exist in the existing literature to date. This is not surprising because [XA20] showed that it is strongly NP-hard to optimize over the feasible region of (**P-CC**), if $\mathcal{P}$ is centered around an empirical distribution of ξ. In addition, for the same setting [Xie19; CKW18; JL20] derived mixed-integer conic representations for (**P-CC**), implying a non-convex feasible region. This paper seeks to revise the choice of the reference distribution $\mathbb{P}$, with regard to which (**P-CC**) and (**O-CC**) with Wasserstein ambiguity produce convex feasible regions. Our main results include:

1. For individual (**P-CC**) with LHS uncertainty, we derive a convex and SOC representation if (i) the reference distribution $\mathbb{P}$ of $\mathcal{P}$ is Gaussian with a positive definite covariance matrix $\Sigma \succ 0$ and (ii) the Wasserstein distance d_W in (1) is defined through the norm $\|\cdot\| := \|\Sigma^{-1/2}(\cdot)\|_2$. This result can be extended to a case where $\mathbb{P}$ is radial [CE06].

2. For joint (**P-CC**) with RHS uncertainty, we prove that the ensuing feasible region is convex if the reference distribution $\mathbb{P}$ is log-concave and d_W is defined through a general norm. More generally, this result holds when $\mathbb{P}$ is α-concave with $\alpha \geq -1$. In addition, we derive a block coordinate ascent algorithm for optimization models involving (**P-CC**) and prove its convergence to global optimum.
3. We extend the aforementioned convexity results for individual (**P-CC**) with LHS uncertainty and joint (**P-CC**) with RHS uncertainty to their optimistic counterparts (**O-CC**).

In addition, we summarize the main convexity results in the following table.

	(**P-CC**)	(**O-CC**)
LHS Uncertainty	Theorem 6	Theorem 11
RHS Uncertainty	Theorem 8	Theorem 12

The remainder of this paper is organized as follows. Section 2 reviews key definitions and results for α-concavity and log-concavity. Sections 3 and 4 study convexity and solution approach for (**P-CC**), respectively. Section 5 extends the convexity results to (**O-CC**). Section 6 demonstrates (**P-CC**) and (**O-CC**) through two numerical experiments.

Notation: We use $\mathcal{X}^p$ and $\mathcal{X}^o$ to denote the feasible region of (**P-CC**) and (**O-CC**), respectively. We denote the n-dimensional extended real system by $\overline{\mathbb{R}}^n$. For a given decision x, we denote by $\mathcal{S}(x)$ the event $\{\, \xi \colon A(x)\xi \leq b(x) \,\}$ and by $\mathcal{S}^c(x)$ its complement. For $a \in \mathbb{R}$, $(a)^+ := \max\{a, 0\}$ and $(a)^- := \min\{a, 0\}$. For a norm $\|\cdot\|$, $\|\cdot\|_*$ denotes its dual norm. $\|\cdot\|_2$ represents the 2-norm, i.e., for $a \in \mathbb{R}^n$, $\|a\|_2 = \sqrt{\sum_{i=1}^n a_i^2}$. I_n denotes the $n \times n$ identity matrix, $\mathbf{Leb}(\cdot)$ denotes the Lebesgue measure, and the indicator $\mathbb{1}\{\, x \in \Omega \,\}$ equals one if $x \in \Omega$ and zero if $x \notin \Omega$.

2 Preliminary results

We review definitions and properties frequently used in subsequent discussions.

Definition 1 (Definition 4.7 in [SDR09])**.** A nonnegative function f defined on a convex subset of $\mathbb{R}^n$ is said to be α-concave with $\alpha \in \overline{\mathbb{R}}$ if for all $x, y \in \mathbf{dom}\, f$ and $\theta \in [0, 1]$

$$f(\theta x + (1-\theta)y) \geq m_\alpha(f(x), f(y); \theta),$$

where $m_\alpha \colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ is defined as

$$m_\alpha(a, b; \theta) := 0 \quad \text{if } ab = 0,$$

and if $a > 0, b > 0, \theta \in [0, 1]$, then

$$m_\alpha(a, b; \theta) := \begin{cases} a^\theta b^{(1-\theta)} & \text{if } \alpha = 0, \\ \max\{a, b\} & \text{if } \alpha = +\infty, \\ \min\{a, b\} & \text{if } \alpha = -\infty, \\ (\theta a^\alpha + (1-\theta) b^\alpha)^{1/\alpha} & \text{otherwise.} \end{cases}$$

When $\alpha = 0$ or $\alpha = -\infty$, we say f is log-concave or quasi-concave, respectively.

Lemma 1 (Lemma 4.8 in [SDR09])**.** The mapping $\alpha \mapsto m_\alpha(a,b;\theta)$ is nondecreasing and continuous.

The monotonicity of m_α implies that if f is α-concave, then it is β-concave for all $\beta \le \alpha$. Under certain conditions, summation preserves α-concavity.

Theorem 1 (Theorem 4.19 in [SDR09])**.** If the function $f\colon \mathbb{R}^n \to \mathbb{R}_+$ is α-concave and the function $g\colon \mathbb{R}^n \to \mathbb{R}_+$ is β-concave, where $\alpha, \beta \ge 1$, then $f(x)+g(x)$ is $\min\{\alpha,\beta\}$-concave.

The Minkowski sum of two Borel measurable subsets $A, B \subset \mathbb{R}^n$ is Borel measurable. Let $\theta \in [0,1]$, then the convex combination of A, B is defined through

$$\theta A + (1-\theta)B := \{\, \theta x + (1-\theta)y \colon x \in A, y \in B \,\}.$$

Definition 2. A probability measure $\mathbb{P}$ defined on the Lebesgue subsets of a convex subset $\Omega \subseteq \mathbb{R}^n$ is said to be α-concave if for any Borel measurable sets $A, B \subseteq \Omega$ and for all $\theta \in [0,1]$,

$$\mathbb{P}(\theta A + (1-\theta)B) \ge m_\alpha(\mathbb{P}(A), \mathbb{P}(B); \theta).$$

For a random variable ξ supported on $\mathbb{R}^n$, we say it is α-concave if the probability measure induced by ξ is α-concave. In particular, ξ is log-concave if it induces a 0-concave distribution.

Next, we review the relationship between α-concave probability measures and their densities.

Theorem 2 (Theorem 4.15 in [SDR09])**.** Let Ω be a convex subset of $\mathbb{R}^n$ and s be the dimension of the smallest affine subspace $\mathcal{H}(\Omega)$ containing Ω. The probability measure $\mathbb{P}$ is α-concave with $\alpha \le 1/s$ if and only if its probability density function (PDF) with respect to the Lebesgue measure on $\mathcal{H}$ is α'-concave with

$$\alpha' := \begin{cases} \alpha/(1-s\alpha) & \text{if } \alpha \in (-\infty, 1/s), \\ -1/s & \text{if } \alpha = -\infty, \\ +\infty & \text{if } \alpha = 1/s. \end{cases}$$

Example 3. (Example 4.9 in [SDR09]) The PDF of an n-dimensional nondegenerate Gaussian is

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left[-\frac{1}{2}\|\Sigma^{-1/2}(x-\mu)\|_2\right],$$

where μ and Σ represent its mean and covariance, respectively. Since $\ln f$ is concave, f is a log-concave function and Gaussian random variables are log-concave.

Example 4 (Example 4.10 in [SDR09])**.** The PDF of a uniform distribution defined on a bounded convex subset $\Omega \subset \mathbb{R}^n$ is

$$f(x) = \frac{1}{\mathbf{vol}(\Omega)} \mathbb{1}\{\, x \in \Omega \,\},$$

where $\mathbf{vol}(\Omega)$ represents the volume of Ω. f is $+\infty$-concave on Ω. Therefore, n-dimensional uniform distributions over a bounded convex subset are $(1/n)$-concave.

Theorem 3 (Theorem 2 in [Gup80])**.** Let f_0, f_1 be two non-negative Borel-measurable functions on $\mathbb{R}^n$ with non-empty supports S_0 and S_1, respectively. Assume that f_0 and f_1 are integrable with respect to the Lebesgue measure on $\mathbb{R}^n$. Let $\theta \in (0,1)$ be a fixed number and f be a non-negative, measurable function on $\mathbb{R}^n$ such that

$$f(x) \ge m_\alpha[f_0(x_0), f_1(x_1); \theta],$$

whenever $x = (1-\theta)x_0 + \theta x_1$ with $x_0 \in S_0, x_1 \in S_1$; $-1/n \le \alpha \le +\infty$. Then

$$\int_{(1-\theta)S_0+\theta S_1} f(x)\,\mathrm{d}x \ge m_{\alpha_n^*}\left[\int_{S_0} f_0(x)\,\mathrm{d}x, \int_{S_1} f_1(x)\,\mathrm{d}x; \theta\right],$$

where

$$\alpha_n^* := \begin{cases} \alpha/(1+n\alpha) & \text{if } \alpha > -1/n, \\ 1/n & \text{if } \alpha = +\infty, \\ -\infty & \text{if } \alpha = -1/n. \end{cases}$$

Before presenting the next lemma, we review the (reverse) Minkowski's inequality.

Theorem 4 (Minkowski's Inequality; see Theorem 9 in Chapter 3 of [Bul13])**.** For $p > 1$ and $a_i, b_i \in \mathbb{R}_+$ for all $i \in [n]$, the following holds:

$$\left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{1/p} \le \left(\sum_{i=1}^{n} a_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} b_i^p\right)^{1/p}.$$

If $p < 1$ and $p \neq 0$, then the inequality holds with the inequality sign reversed.

Lemma 2. If the function $f\colon \mathbb{R}^n \to \mathbb{R}_+$ is an α-concave function with $\alpha \in \overline{\mathbb{R}}$ and $c \in \mathbb{R}_+$ is a constant, then $g(x) := f(x) - c$ is α-concave on $D := \{\, x \in \mathbb{R}^n \colon f(x) > c \,\}$.

Proof. See Appendix A. □

3 Pessimistic Chance Constraint

We first review the definitions of value-at-risk (VaR) and CVaR [RU99], as well as the CVaR reformulation of $\mathcal{X}^p$ derived by [Xie19]. Then, we derive a new reformulation of $\mathcal{X}^p$ for α-concave reference distribution $\mathbb{P}$. The new reformulation leads to convexity proofs for individual (**P-CC**) with LHS uncertainty and joint (**P-CC**) with RHS uncertainty in Sections 3.1 and 3.2, respectively.

Definition 3. Let X be a random variable, inducing probability distribution $\mathbb{P}_X$. The $(1-\epsilon)$-VaR of X is defined through

$$\mathrm{VaR}_{1-\epsilon}(X) := \inf\{\, x : \mathbb{P}_X[X \le x] \ge 1-\epsilon \,\},$$

and its $(1-\epsilon)$-CVaR is defined through

$$\mathrm{CVaR}_{1-\epsilon}(X) := \min_{\gamma}\left\{ \gamma + \frac{1}{\epsilon}\mathbb{E}\Big[\, (X-\gamma)^+ \,\Big] \right\}.$$

Theorem 5 (Adapted from Theorem 1 in [Xie19])**.** For $\delta > 0$, it holds that

$$\mathcal{X}^p = \left\{ x \in \mathbb{R}^n \colon \frac{\delta}{\epsilon} + \mathrm{CVaR}_{1-\epsilon}\Big(-\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right) \Big) \le 0 \right\}. \tag{2}$$

Here, random variable ζ follows the reference distribution $\mathbb{P}$ and $\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right)$ represents the distance from ζ to the "unsafe" set $\mathcal{S}^c(x)$ [CKW18],

$$\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right) := \inf_{\xi \in \Xi}\{\, \|\zeta - \xi\| \colon A(x)\xi \not\le b(x) \,\},$$

and Ξ is the support of ξ.

For all $x \in \mathcal{X}^p$, it holds that

$$a_i(x) = 0 \;\Rightarrow\; b_i(x) \geq 0 \quad \forall i \in [m],$$

where $a_i(x)^\top$ represents row i of matrix $A(x)$ and $b_i(x)$ represents entry i of vector $b(x)$, because otherwise $\mathbb{P}[A(x)\zeta \leq b(x)] = 0$ and $x \notin \mathcal{X}^p$. Assuming the above implication without loss of generality, we define function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$,

$$f(x, \zeta) := \min_{i \in [m] \setminus I(x)} \left\{ \frac{b_i(x) - a_i(x)^\top \zeta}{\|a_i(x)\|_*} \right\},$$

where $I(x) := \{i \in [m] : a_i(x) = 0\}$. Then, it follows from [Xie19; CKW18] that

$$\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right) = \Big(f(x, \zeta)\Big)^+.$$

In what follows, we derive new reformulations of $\mathcal{X}^p$ based on $f(x, \zeta)$. To this end, we need the following lemma to relate the the CVaR of $f(x, \zeta)$ to that of $-\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right)$ in (2).

Lemma 3. Let X be a random variable, then

$$\mathrm{CVaR}_{1-\epsilon}(X^-) = \mathbb{1}\left\{\, 0 \geq \mathrm{VaR}_{1-\epsilon}(X) \,\right\} \cdot \left[\mathrm{CVaR}_{1-\epsilon}(X) - \frac{1}{\epsilon}\mathbb{E}[X^+]\right].$$

Proof. See Appendix B. □

Combining Theorem 5 and Lemma 3 leads to the following reformulation of $\mathcal{X}^p$.

Corollary 1. For $\delta > 0$, it holds that

$$\mathcal{X}^p = \left\{ x \in \mathbb{R}^n : \begin{array}{c} \dfrac{\delta}{\epsilon} + \mathrm{CVaR}_{1-\epsilon}\left(-f(x,\zeta)\right) \leq \dfrac{1}{\epsilon}\mathbb{E}\left[(-f(x,\zeta))^+\right] \\ 0 \geq \mathrm{VaR}_{1-\epsilon}\left(-f(x,\zeta)\right) \end{array} \right\} \begin{array}{r} (3) \\ (4) \end{array}$$

In this paper, we focus on cases in which $\mathbb{P}$ is α-concave. The next lemma shows that an α-concave $\mathbb{P}$ yields atomless $\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right)$ and $f(x, \zeta)$, which lead to a further reformulation of $\mathcal{X}^p$.

Lemma 4. If the reference distribution $\mathbb{P}$ is α-concave, then for all x, $\mathbb{P}\big[\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right) = y\big] = 0$ for all $y > 0$ and $\mathbb{P}\big[f(x, \zeta) = y\big] = 0$ for all $y \in \mathbb{R}$.

Proof. See Appendix C. □

Proposition 1. Suppose that $\mathbb{P}$ is α-concave. Then, for $\delta > 0$, it holds that

$$\mathcal{X}^p = \left\{ x \in \mathbb{R}^n : \begin{array}{c} \delta \leq \displaystyle\int_0^{\mathrm{VaR}_\epsilon(f(x,\zeta))} \Big(\mathbb{P}\left[f(x,\zeta) \geq t\right] - (1-\epsilon)\Big)\,\mathrm{d}t \\ \mathbb{P}\left[A(x)\zeta \leq b(x)\right] \geq 1-\epsilon \end{array} \right\} \begin{array}{r} (5) \\ (6) \end{array}$$

Proof. First, moving the CVaR term to the RHS of (3) yields

$$\begin{aligned} \delta &\leq \mathbb{E}\Big[f(x,\zeta) \cdot \mathbb{1}\left\{ -f(x,\zeta) \geq \mathrm{VaR}_{1-\epsilon}\left(-f(x,\zeta)\right) \right\}\Big] - \mathbb{E}\Big[f(x,\zeta) \cdot \mathbb{1}\left\{ -f(x,\zeta) \geq 0 \right\}\Big] \\ &= \mathbb{E}\Big[f(x,\zeta) \cdot \mathbb{1}\left\{ \mathrm{VaR}_{1-\epsilon}\left(-f(x,\zeta)\right) \leq -f(x,\zeta) \leq 0 \right\}\Big] \\ &= \mathbb{E}\Big[f(x,\zeta) \cdot \mathbb{1}\left\{ 0 \leq f(x,\zeta) \leq \mathrm{VaR}_\epsilon\left(f(x,\zeta)\right) \right\}\Big], \qquad (7) \end{aligned}$$

where the first equality is because $f(x,\zeta)$ is atomless and the second equality is because $\text{VaR}_{1-\epsilon}(-X) = -\text{VaR}_{\epsilon}(X)$. Now, we use the wedding lake representation of nonnegative integrable functions to further recast the RHS of (7) as

$$\begin{aligned}
&\mathbb{E}\Big[f(x,\zeta)\cdot \mathbb{1}\{\, 0 \le f(x,\zeta) \le \text{VaR}_{\epsilon}\left(f(x,\zeta)\right) \,\}\Big] \\
= & \int_{\Xi} f(x,\zeta)\cdot \mathbb{1}\{\, 0 \le f(x,\zeta) \le \text{VaR}_{\epsilon}\left(f(x,\zeta)\right) \,\}\, \mathrm{d}\mathbb{P}(\zeta) \\
= & \int_{\Xi}\int_{\mathbb{R}_+} \mathbb{1}\{\, t \le f(x,\zeta)\cdot \mathbb{1}\{\, 0 \le f(x,\zeta) \le \text{VaR}_{\epsilon}\left(f(x,\zeta)\right) \,\}\,\}\, \mathrm{d}t\ \mathrm{d}\mathbb{P}(\zeta) \\
= & \int_{\Xi}\int_{\mathbb{R}_+} \mathbb{1}\{\, t \le f(x,\zeta) \le \text{VaR}_{\epsilon}\left(f(x,\zeta)\right) \,\}\, \mathrm{d}t\ \mathrm{d}\mathbb{P}(\zeta) \\
= & \int_{\mathbb{R}_+} \mathbb{P}\Big[t \le f(x,\zeta) \le \text{VaR}_{\epsilon}\left(f(x,\zeta)\right)\Big]\, \mathrm{d}t, \qquad \text{(by the Tonelli's Theorem)} \\
= & \int_0^{\text{VaR}_{\epsilon}(f(x,\zeta))} \Big(\mathbb{P}\left[f(x,\zeta) \ge t\right] - (1-\epsilon)\Big)\, \mathrm{d}t,
\end{aligned}$$

where the first two equalities are by definitions of expectation and wedding lake representation, respectively. We justify the third equality by arguing that, for any $x \in \mathcal{X}^p$ and $\zeta \in \Xi$,

$$\mathbb{1}\{\, t \le f(x,\zeta)\cdot \mathbb{1}\{\, 0 \le f(x,\zeta) \le \text{VaR}_{\epsilon}\left(f(x,\zeta)\right) \,\}\,\} = \mathbb{1}\{\, t \le f(x,\zeta) \le \text{VaR}_{\epsilon}\left(f(x,\zeta)\right) \,\} \tag{8}$$

holds Lebesgue-almost everywhere for $t \in \mathbb{R}_+$. We discuss the following three cases:

(i) If ζ makes $f(x,\zeta) < 0$, then the LHS of (8) simplifies to $\mathbb{1}\{\, t \le 0 \,\}$, which coincides with the RHS.

(ii) If ζ makes $f(x,\zeta) \in [0, \text{VaR}_{\epsilon}(f(x,\zeta))]$, then the LHS of (8) simplifies to $\mathbb{1}\{\, t \le f(x,\zeta) \,\}$, coinciding with the RHS.

(iii) If ζ makes $f(x,\zeta) > \text{VaR}_{\epsilon}(f(x,\zeta))$, then the LHS and RHS of (8) simplify to $\mathbb{1}\{\, t \le 0 \,\}$ and 0, respectively, which differ only at $t = 0$ for $t \in \mathbb{R}_+$.

The last equality is because

$$\mathbb{P}\Big[t \le f(x,\zeta) \le \text{VaR}_{\epsilon}\left(f(x,\zeta)\right)\Big] = \mathbb{P}\Big[t \le f(x,\zeta)\Big] - \mathbb{P}\Big[t \ge \text{VaR}_{\epsilon}\left(f(x,\zeta)\right)\Big]$$

when $t \in [0, \text{VaR}_{\epsilon} f(x,\zeta)]$. This recasts (3) into (5).

Second, constraint (4) is equivalent to $\mathbb{P}\big[f(x,\zeta) \ge 0\big] \ge 1-\epsilon$ by definition of VaR, which can be further recast as

$$\mathbb{P}\Big[a_i(x)^{\mathsf{T}}\zeta \le b_i(x),\ \forall i \in [m] \setminus I(x)\Big] \ge 1-\epsilon$$

by definition of $f(x,\zeta)$. For all $x \in \mathcal{X}^p$ and $i \in [m]$, we assume without loss of generality that $b_i(x) \ge 0$ whenever $a_i(x) = 0$ (because otherwise $\mathbb{P}[A(x)\zeta \le b(x)] = 0$), and it holds that $a_i(x)^{\mathsf{T}}\zeta \le b_i(x)$ for all $i \in I(x)$. It follows that (4) is equivalent to (6), which completes the proof. □

Remark 1. We notice that constraint (6) is simply (**CC**) with respect to the reference distribution $\mathbb{P}$ of the Wasserstein ball $\mathcal{P}$. In addition, constraint (5) encodes a robust guarantee. Intuitively, the RHS of (5) evaluates the budget needed to shift the probability masses of $\mathbb{P}$ so that the corresponding (**CC**) can be violated. Constraint (5) makes sure that this budget is beyond the radius of $\mathcal{P}$, i.e., (**CC**) will not be violated as long as the shifted distribution lies within $\mathcal{P}$.

3.1 Individual (P-CC) with LHS Uncertainty

When $m = 1$, the feasible region of (**P-CC**) simplifies to

$$\mathcal{X}_{\mathrm{L}}^{p} := \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{Q} \in \mathcal{P}} \mathbb{Q}\left[a(x)^\mathsf{T} \xi \leq b(x) \right] \geq 1 - \epsilon \right\}.$$

We provide sufficient conditions, under which $\mathcal{X}_{\mathrm{L}}^{p}$ admits a convex and SOC representation.

Theorem 6. Suppose that the reference distribution $\mathbb{P}$ is Gaussian with mean μ and covariance matrix $\Sigma \succ 0$. In addition, suppose that the norm $\|\cdot\|$ defining $\mathcal{P}$ satisfies $\|\cdot\| = \|\Sigma^{-1/2}(\cdot)\|_2$. Then, for $\delta > 0$ and $\epsilon \in (0, \frac{1}{2}]$ it holds that

$$\mathcal{X}_{\mathrm{L}}^{p} = \left\{ x \in \mathbb{R}^n : a(x)^\mathsf{T} \mu + c_p \|\Sigma^{1/2} a(x)\|_2 \leq b(x) \right\}, \tag{9}$$

where

$$c_p := \inf_{\epsilon' \in (0, \epsilon)} \frac{\delta + \epsilon \mathrm{CVaR}_{1-\epsilon}(Y) - \epsilon' \mathrm{CVaR}_{1-\epsilon'}(Y)}{\epsilon - \epsilon'} \geq 0$$

and Y is a standard, 1-dimensional Gaussian random variable.

Proof. Pick any $x \in \mathcal{X}_{\mathrm{L}}^{p}$. If $a(x) = 0$, then $\mathbb{Q}[a(x)^\mathsf{T}\xi \leq b(x)] = 1$ when $b(x) \geq 0$ and $\mathbb{Q}[a(x)^\mathsf{T}\xi \leq b(x)] = 0$ otherwise. It follows that $\mathcal{X}_{\mathrm{L}}^{p} = \{ x \in \mathbb{R}^n : b(x) \geq 0 \}$, as desired. For the rest of the proof, we assume that $a(x) \neq 0$. By Corollary 1, x satisfies constraints (3)–(4). We recast (3) as

$$\begin{aligned}
(3) \iff & \frac{\delta}{\epsilon} + \mathrm{CVaR}_{1-\epsilon}\left(-f(x,\zeta)\right) \leq \frac{1}{\epsilon} \mathbb{E}\Big[- f(x,\zeta) \cdot \mathbb{1}\left\{ -f(x,\zeta) \geq 0 \right\} \Big] \\
\iff & \frac{\delta}{\epsilon} + \mathrm{CVaR}_{1-\epsilon}\left(-f(x,\zeta)\right) \leq \frac{1}{\epsilon} \sup_{t \in \mathbb{R}} \mathbb{E}\Big[- f(x,\zeta) \cdot \mathbb{1}\left\{ -f(x,\zeta) \geq t \right\} \Big] \\
\iff & \exists\, t \in \mathbb{R} \colon \delta + \epsilon \cdot \mathrm{CVaR}_{1-\epsilon}\left(-f(x,\zeta)\right) \leq \mathbb{E}\Big[- f(x,\zeta) \cdot \mathbb{1}\left\{ -f(x,\zeta) \geq t \right\} \Big] \\
\iff & \exists\, \epsilon_t > 0 \colon \delta + \epsilon \cdot \mathrm{CVaR}_{1-\epsilon}\left(-f(x,\zeta)\right) \leq \epsilon_t \cdot \mathrm{CVaR}_{1-\epsilon_t}\left(-f(x,\zeta)\right),
\end{aligned}$$

where the second equivalence is because $t = 0$ is a maximizer of $\sup_{t \in \mathbb{R}} \mathbb{E}\big[-f(x,\zeta) \cdot \mathbb{1}\{ -f(x,\zeta) \geq t \}\big]$, and the final equivalence is because $f(x,\zeta)$ is atomless and hence

$$\mathbb{E}\Big[- f(x,\zeta) \cdot \mathbb{1}\left\{ -f(x,\zeta) \geq t \right\} \Big] = \epsilon_t \cdot \mathrm{CVaR}_{1-\epsilon_t}\left(-f(x,\zeta)\right)$$

with $\epsilon_t := \mathbb{P}\left[-f(x,\zeta) \geq t\right]$. Also, since $\delta > 0$, it holds that $\epsilon_t < \epsilon$ because otherwise $\epsilon \cdot \mathrm{CVaR}_{1-\epsilon}\big(-f(x,\zeta)\big) \geq \epsilon_t \cdot \mathrm{CVaR}_{1-\epsilon_t}\big(-f(x,\zeta)\big)$ due to (4). Now, by definition

$$f(x,\zeta) = \frac{b(x) - a(x)^\mathsf{T}\zeta}{\|a(x)\|_*} = m(x) - \frac{a(x)^\mathsf{T}(\zeta - \mu)}{\|a(x)\|_*},$$

where we define $m(x) := \big(b(x) - a(x)^\mathsf{T}\mu\big)/\|a(x)\|_*$. Then, the normalized random variable $a(x)^\mathsf{T}(\zeta - \mu)/\|a(x)\|_*$ is standard Gaussian because $\mathbb{E}[\zeta] = \mu$ and

$$\mathbb{E}\left[\frac{a(x)^\mathsf{T}(\zeta - \mu)(\zeta - \mu)^\mathsf{T} a(x)}{\|a(x)\|_*^2} \right] = \frac{a(x)^\mathsf{T}\Sigma a(x)}{a(x)^\mathsf{T}\Sigma a(x)} = 1.$$

It follows from translation invariance and the symmetry of Y that $\mathrm{CVaR}_{1-\epsilon}\big(-f(x,\zeta)\big) = \mathrm{CVaR}_{1-\epsilon}(Y) - m(x)$ and

$$\begin{aligned}
(3) &\iff \exists\, \epsilon_t \in (0,\epsilon)\colon \delta + \epsilon \cdot \mathrm{CVaR}_{1-\epsilon}(Y) - \epsilon_t \cdot \mathrm{CVaR}_{1-\epsilon_t}(Y) \le m(x)(\epsilon - \epsilon_t) \\
&\iff a(x)^\mathsf{T}\mu + c_p \|\Sigma^{1/2} a(x)\|_2 \le b(x),
\end{aligned}$$

which proves that (3)–(4) imply (9). On the contrary, suppose that x satisfies (9). Then, by definition of CVaR we have

$$\begin{aligned}
\epsilon \mathrm{CVaR}_{1-\epsilon}(Y) - \epsilon' \mathrm{CVaR}_{1-\epsilon'}(Y) &= \mathbb{E}\Big[Y \cdot \mathbb{1}\{\, Y \in [\mathrm{VaR}_{1-\epsilon}(Y), \mathrm{VaR}_{1-\epsilon'}(Y)] \,\} \Big] \\
&= \int_{\mathrm{VaR}_{1-\epsilon}(Y)}^{\mathrm{VaR}_{1-\epsilon'}(Y)} y \cdot \mathrm{d}F_Y(y) \\
&= \int_{1-\epsilon}^{1-\epsilon'} \mathrm{VaR}_q(Y)\,\mathrm{d}q \ge (\epsilon - \epsilon')\mathrm{VaR}_{1-\epsilon}(Y),
\end{aligned}$$

where F_Y represents the CDF of Y and the final equality is by a change-of-variable $q = F_Y(y)$ (or equivalently, $y = \mathrm{VaR}_q(Y)$). It follows that $c_p \ge \mathrm{VaR}_{1-\epsilon}(Y)$. Hence, constraint (9) implies

$$\begin{aligned}
&\frac{b(x) - a(x)^\mathsf{T}\mu}{\|\Sigma^{1/2} a(x)\|_2} \ge \mathrm{VaR}_{1-\epsilon}(Y) \\
\iff &\frac{b(x) - a(x)^\mathsf{T}\mu}{\|\Sigma^{1/2} a(x)\|_2} \ge \mathrm{VaR}_{1-\epsilon}\left(\frac{a(x)^\mathsf{T}(\zeta - \mu)}{\|\Sigma^{1/2} a(x)\|_2} \right) \implies (4),
\end{aligned}$$

where the last implication is by translation invariance of VaR. Since x satisfies (4), the reformulation of (3) presented above show that x also satisfies (3), which completes the proof. □

Remark 2. We underscore that the value of c_p is independent of x and can be obtained offline through, e.g., a line search. In addition, in the proof of Theorem 6, we used the fact that the CVaR of the normalized random variable $a(x)^\mathsf{T}(\zeta - \mu)/\|a(x)\|_*$ is independent of x. As a result, with a similar proof, the representation in Theorem 6 can be extended to a case where $\mathbb{P}$ is radial [CE06], examples of which include Gaussian and uniform on an ellipsoidal support.

We close this subsection by mentioning the asymptotics of $\mathcal{X}_\mathrm{L}^p$ as the Wasserstein ball shrinks.

Remark 3. Since the Wasserstein ball shrinks to the reference distribution $\mathbb{P}$ as δ tends to zero, one would expect that $\mathcal{X}_\mathrm{L}^p$ asymptotically recovers the feasible region of (**CC**) with respect to $\mathbb{P} \sim \mathrm{Gaussian}(\mu, \Sigma)$, which reads $\big\{\, x \in \mathbb{R}^n \colon a(x)^\mathsf{T}\mu + \mathrm{VaR}_{1-\epsilon}(Y)\|\Sigma^{1/2}a(x)\|_2 \le b(x) \,\big\}$. This is indeed the case, since it can be shown that c_p decreases to $\mathrm{VaR}_{1-\epsilon}(Y)$ as δ tends to zero. See a proof in Appendix D.

3.2 Joint (P-CC) with RHS Uncertainty

For (**CC**) with RHS uncertainty, it is well celebrated that the ensuing feasible region is convex when ξ has an α-concave distribution (particularly, ξ is log-concave when $\alpha = 0$) [Pré13; SDR09].

Theorem 7 (Theorem 4.39 and Corollary 4.41 in [SDR09])**.** Let the function $h\colon \mathbb{R}^n \times \mathbb{R}^m$ be quasi-concave. If $\xi \in \mathbb{R}^m$ is a random vector that has an α-concave probability distribution, then $H(x) := \mathbb{P}[h(x,\xi) \ge 0]$ is α-concave on the set $\mathcal{D} := \{\, x \in \mathbb{R}^n \colon \exists\, \xi \text{ such that } h(x,\xi) \ge 0 \,\}$ and the following set is convex and closed:

$$\mathcal{X} := \{\, x \in \mathbb{R}^n \colon \mathbb{P}[h(x,\xi) \ge 0] \ge 1 - \epsilon \,\}.$$

In this subsection, we seek to extend this result to (**P-CC**).

Theorem 8. Suppose that the reference distribution $\mathbb{P}$ of $\mathcal{P}$ is α-concave with $\alpha \geq -1$. Then, for $\delta > 0$ the set

$$\mathcal{X}_{\mathrm{R}}^{p} := \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{Q} \in \mathcal{P}} \mathbb{Q}\left[A\xi \leq b(x)\right] \geq 1 - \epsilon \right\}$$

is convex and closed.

Before presenting a proof of Theorem 8, we present some useful lemmas. Without loss of generality, we assume that each row of matrix A, denoted by $a_i^\top$ for all $i \in [m]$, satisfies

(i) $a_i \neq 0$, because otherwise we can add a deterministic constraint $b_i(x) \geq 0$ to $\mathcal{X}_{\mathrm{R}}^{p}$ and eliminate inequality i from (**P-CC**);

(ii) $\|a_i\|_* = 1$, because otherwise we can divide both sides of inequality i by $\|a_i\|_*$ and set $a_i \leftarrow a_i/\|a_i\|_*$, $b_i(x) \leftarrow b_i(x)/\|a_i\|_*$.

Recall that for $\zeta \in \mathbb{R}^m$ the distance $\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right)$ to the unsafe set satisfies $\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right) = \big(f(x,\zeta)\big)^+$ with

$$f(x,\zeta) = \min_{i \in [m]} \left\{ b_i(x) - a_i^\top \zeta \right\}$$

and $f(x,\zeta)$ is jointly concave in (x,ζ).

Lemma 5. For all $\epsilon \in (0,1)$, if ζ has a α-concave distribution with $\alpha \geq -1$, then $\mathrm{VaR}_{1-\epsilon}\left(f(x,\zeta)\right)$ is concave in x on $\mathbb{R}^n$.

Proof. See Appendix E. □

Lemma 6. Suppose that $f(\cdot,\cdot)\colon \mathbb{R}^n \times \Xi \to \mathbb{R}$ is a continuous function, ζ follows an α-concave distribution $\mathbb{P}$, and $f(x,\zeta)$ is atomless for any $x \in \mathbb{R}^n$. Then,

$$\psi(x,t) := \mathbb{P}\left[f(x,\zeta) \geq t\right] - (1-\epsilon) \qquad \text{and} \qquad \phi(x,y) := \int_0^y \psi(x,t)\,\mathrm{d}t$$

are both continuous on $\mathbb{R}^n \times \mathbb{R}_+$.

Proof. See Appendix F. □

Now we are ready to prove Theorem 8.

Proof of Theorem 8. First, recall that by Proposition 1 we recast $\mathcal{X}_{\mathrm{R}}^{p}$ as constraints (5)–(6). For ease of exposition, we denote by $G(x)$ the RHS of (5).

Second, to show that $\mathcal{X}_{\mathrm{R}}^{p}$ is closed, it suffices to prove the closedness of the feasible region of (5) because that of (6) follows from Theorem 7. To this end, we notice that $\mathrm{VaR}_\epsilon\big(f(x,\zeta)\big)$ is continuous in x due to its concavity. Then, by Lemma 6 the mapping

$$x \mapsto \int_0^{\mathrm{VaR}_\epsilon(f(x,\zeta))} \mathbb{P}\Big[f(x,\zeta) \geq t\Big]\,\mathrm{d}t$$

is continuous. It follows that $G(x)$ is continuous and the feasible region of (5) is closed.

Third, to show that $\mathcal{X}_{\mathrm{R}}^p$ is convex, it suffices to prove the convexity of the feasible region of (5) because that of (6) follows from Theorem 7. To that end, by Theorem 7 and Lemma 2, ψ is α-concave in (x,t) on $\mathbf{dom}\,\psi := \{\,(x,t)\colon \psi(x,t) \geq 0\,\} = \{\,(x,t)\colon t \leq \mathrm{VaR}_\epsilon(f(x,\zeta))\,\}$, which is convex by Lemma 5. Then, for any $x_0, x_1 \in \mathcal{X}_{\mathrm{R}}^p$ and any $t_0 \in S_0 := [0, \mathrm{VaR}_\epsilon(f(x_0,\zeta))]$, $t_1 \in S_1 := [0, \mathrm{VaR}_\epsilon(f(x_1,\zeta))]$ it holds that

$$\psi(x_{1/2}, t_{1/2}) \geq m_\alpha\left[\psi(x_0,t_0), \psi(x_1,t_1); \frac{1}{2}\right],$$

where $x_{1/2} = (x_0+x_1)/2$ and $t_{1/2} = (t_0+t_1)/2$. It follows that

$$\begin{aligned} m_{\alpha_1^*}\left[\int_{S_0} \psi(x_0,t)\,\mathrm{d}t, \int_{S_1} \psi(x_1,t)\,\mathrm{d}t; \frac{1}{2}\right] &\leq \int_{\frac{1}{2}S_0+\frac{1}{2}S_1} \psi(x_{1/2},t)\,\mathrm{d}t \\ &\leq \int_{S_{1/2}} \psi(x_{1/2},t)\,\mathrm{d}t \end{aligned}$$

where the first inequality is due to Theorem 3 and $\alpha_1^* \geq -\infty$ is a function of α (see Theorem 3), and the second inequality is because $\frac{1}{2}S_0 + \frac{1}{2}S_1 \subseteq S_{1/2} := [0, \mathrm{VaR}_\epsilon(f(x_{1/2},\zeta))]$. In other words, we obtain that

$$m_{\alpha_1^*}\left[G(x_0), G(x_1); \frac{1}{2}\right] \leq G\left(x_{1/2}\right)$$

and $G(x)$ is midpoint α_1^*-concave, and particularly, midpoint quasi-concave. Then, its continuity implies that $G(x)$ is quasi-concave and constraint (5) yields a convex feasible region. This finishes the proof. $\square$

Remark 4. Although Theorem 8 is concerned with (**P-CC**) with linear inequalities, its proof extends to (**P-CC**) with nonlinear inequalities, as long as $f(x,\zeta)$ is jointly concave in (x,ζ) and $f(x,\zeta)$ is atomless for all x.

4 Solution Approach for Pessimistic Chance Constraint

We study an algorithm for solving optimization models involving (**P-CC**). In view that individual (**P-CC**) with LHS uncertainty admits a convex and SOC representation, which can be computed effectively by commercial solvers, we focus on a model with joint (**P-CC**) and RHS uncertainty: $\min_{x\in X}\{c^\mathsf{T}x : (\textbf{P-CC})\}$, where $c \in \mathbb{R}^n$ represents cost coefficients and $X \subseteq \mathbb{R}^n$ represents a set that is deterministic, compact, and convex. By Proposition 1, this model is equivalent to

$$\min_{x\in X} \quad c^\mathsf{T}x \tag{10a}$$

$$\text{s.t.} \quad \delta \leq \int_0^{\mathrm{VaR}_\epsilon(f(x,\zeta))} \Big(\mathbb{P}\big[f(x,\zeta) \geq t\big] - (1-\epsilon)\Big)\,\mathrm{d}t, \tag{10b}$$

$$\mathrm{VaR}_\epsilon\big(f(x,\zeta)\big) \geq 0, \tag{10c}$$

where $f(x,\zeta) = \min_{i\in[m]}\{b_i(x) - a_i^\mathsf{T}\zeta\}$. Here, constraint (10b) appears challenging because its RHS involves an integral with upper limit $\mathrm{VaR}_\epsilon(f(x,\zeta))$. To make the model computable, we define a new variable $y \geq 0$ to represent $\mathrm{VaR}_\epsilon(f(x,\zeta))$.

Proposition 2. For $y \geq 0$, define

$$\phi(x,y) := \int_0^y \Big(\mathbb{P}\Big[f(x,\zeta) \geq t\Big] - (1-\epsilon)\Big)\,\mathrm{d}t.$$

If $\mathbb{P}$ is α-concave with $\alpha \geq -1$, then $\phi(x,y)$ is α_1^*-concave on

$$\mathbf{dom}\,\phi := \{\, (x,y) \in X \times \mathbb{R}_+ \colon \mathbb{P}\left[f(x,\zeta) \geq y\right] \geq (1-\epsilon) \,\},$$

where α_1^* is defined in Theorem 3. In addition, $\mathbf{dom}\,\phi$ is closed and constraints (10b)–(10c) is equivalent to

$$\delta \leq \max_{y \geq 0} \phi(x,y). \tag{10d}$$

Proof. See Appendix G. □

By Proposition 2, formulation (10a)–(10c) is equivalent to $\min_{x \in X}\{c^\mathsf{T} x \,:\, \text{(10d)}\}$. To address the integral arising from the RHS of constraint (10d), we switch the objective function with the constraint to obtain

$$\rho(u) := \sup_{x \in X, y \geq 0} \left\{ \phi(x,y) \colon c^\mathsf{T} x \leq u \right\}, \tag{11}$$

where u represents a budget limit on the (original) objective function. We notice that if we can evaluate $\rho(u)$ efficiently, then an optimal solution to (10a)–(10c) can be readily obtained through a bisection search on u. That is, solving (10a)–(10c) reduces to evaluating $\rho(u)$. In addition, $\rho(u)$ may be interesting in its own right because it represents the largest Wasserstein radius δ that allows us to find a solution x that satisfies (**P-CC**) and incurs a cost no more than u. Hence, the graph of $\rho(u)$ depicts a risk envelope that interprets the trade-off between the robustness and the cost effectiveness of (**P-CC**). We demonstrate the risk envelope numerically in Section 6.2.

Evaluating $\rho(u)$ is equivalent to maximizing $\phi(x,y)$ over the intersection of $\{\, (x,y) \in X \times \mathbb{R}_+ : c^\mathsf{T} x \leq u \,\}$ and $\mathbf{dom}\,\phi$. Unfortunately, projecting onto $\mathbf{dom}\,\phi$ may be inefficient since it is the feasible region of (**CC**). To avoid projection, we propose a block coordinate ascent algorithm (Algorithm 1; see, e.g., [Aus76; LT93; Ber97; GS99; BT13; Bec15]). This algorithm iteratively maximizes over y with x fixed and then maximizes over x with y fixed. Here, for fixed x with $\mathbb{P}[A\zeta \leq b(x)] \geq 1-\epsilon$, i.e., when x satisfies (**CC**), the maximization over y admits a closed-form solution $y = \text{VaR}_\epsilon\big(f(x,\zeta)\big)$, that is,

$$\max_{y \geq 0} \phi(x,y) = \phi\left(x, \text{VaR}_\epsilon\Big(f(x,\zeta)\Big)\right)$$

because $\phi(x,y)$ is increasing in y on the interval $\big[0, \text{VaR}_\epsilon\big(f(x,\zeta)\big)\big]$ and it becomes decreasing in y when $y > \text{VaR}_\epsilon\big(f(x,\zeta)\big)$. On the other hand, for fixed y, we seek to maximize $\phi(x,y)$, which appears challenging as it is an integral. Fortunately, we can recast $\phi(x,y)$ as

$$\begin{aligned}
\phi(x,y) &= \int_0^y \mathbb{P}\left[f(x,\zeta) \geq t\right] \mathrm{d}t - y \cdot (1-\epsilon) \\
&= y \int_0^1 \mathbb{P}\left[f(x,\zeta) \geq sy\right] \mathrm{d}s - y \cdot (1-\epsilon) \\
&= y \int_{\Xi} \int_{\mathbb{R}} \mathbb{1}\left\{\, (\zeta,s) \colon f(x,\zeta) \geq sy \,\right\} \cdot \mathbb{1}_{[0,1]}(s)\, \mathrm{d}s\, \mathrm{d}\mathbb{P}(\zeta) - y \cdot (1-\epsilon), \\
&= y \int_{\Xi \times \mathbb{R}} \mathbb{1}\left\{\, (\zeta,s) \colon f(x,\zeta) \geq sy \,\right\} \mathrm{d}\widehat{\mathbb{P}}(\zeta,s) - y \cdot (1-\epsilon), \\
&= y \cdot \widehat{\mathbb{P}}\left[f(x,\zeta) \geq sy\right] - y \cdot (1-\epsilon),
\end{aligned}$$

where the third equality is by Tonelli's theorem and $\widehat{\mathbb{P}}$ represents the product measure of $\mathbb{P}$ and the uniform distribution on $[0,1]$. Since these two distributions are log-concave on Ξ and $[0,1]$, respectively, $\widehat{\mathbb{P}}$ is log-concave on $\Xi \times [0,1]$. As a result, the problem simplifies to the P-model of (**CC**) with respect to a log-concave distribution, which has been well studied in [Pré13; Nor93]. As a result, Algorithm 1 uses the existing solution approach as a building block and assumes that there exists an oracle, denoted by $\mathcal{O}_u(y, \varepsilon)$, which for given y and $\varepsilon > 0$ returns an ε-optimal solution $\widehat{x} \in \left\{ x \in X : c^\mathsf{T} x \le u \right\}$ such that

$$\widehat{\mathbb{P}}\left[f(\widehat{x}, \zeta) \ge sy\right] \ge \max_{x \in X:\ c^\mathsf{T} x \le u} \left\{ \widehat{\mathbb{P}}\left[f(x, \zeta) \ge sy\right] \right\} - \varepsilon.$$

We are now ready to present Algorithm 1.

Algorithm 1: Evaluation of $\rho(u)$

Inputs: budget u, risk level ϵ, oracle $\mathcal{O}_u$, a diminishing sequence $\{ \varepsilon_k \}_k$, and an x_1 such that $y_1 := \mathrm{VaR}_\epsilon\big(f(x_1, \zeta)\big) > 0$.

```
for k = 1, 2, ... do
    x_{k+1} ← O_u(y_k, ε_k);
    y_{k+1} ← VaR_ϵ(f(x_{k+1}, ζ));
    if stopping criterion is satisfied then
        return φ(x_{k+1}, y_{k+1}).
```

Algorithm 1 needs an starting point (x_1, y_1) such that $\mathrm{VaR}_\epsilon(f(x_1, \zeta)) > 0$. This can be obtained by solving a (**CC**) feasibility problem,

$$\min_{x \in X} \left\{ 0 \colon \mathbb{P}\big[f(x, \zeta) \ge \varepsilon_0\big] \ge 1 - \epsilon,\ c^\mathsf{T} x \le u \right\}, \tag{12}$$

where ε_0 is a small positive constant. If formulation (12) is infeasible for all $\varepsilon > 0$, then $\rho(u) = 0$ because $\mathrm{VaR}_\epsilon\big(f(x, \zeta)\big)$ always remains non-positive. Numerically, one can solve (12) for a sequence of diminishing ε_0's to find a valid starting point. We close this section by showing that Algorithm 1 achieves global optimum.

Theorem 9. Let $\{ (x_k, y_k) \}_k$ be an infinite sequence of iterates produced by Algorithm 1. Suppose that $\mathbb{P}$ is log-concave and, for all $k \ge 2$, x_k and y_k are ε_k-optimal, i.e.,

$$\max_x \phi(x, y_{k-1}) - \varepsilon_k \le \phi(x_k, y_{k-1}) \le \max_x \phi(x, y_{k-1}) \quad \text{and} \quad \Big| y_k - \mathrm{VaR}_\epsilon\big(f(x_k, \zeta)\big) \Big| \le \varepsilon_k$$

with $\lim_{k \to \infty} \varepsilon_k = 0$. Then, any limit point of $\{ (x_k, y_k) \}_k$ is a global optimal solution to (11).

Proof. The proof relies on preparatory Lemmas 8, 9, and 10, whose proofs are provided in Appendix H.

First, we define set $S := \mathbf{dom}\, \phi \cap \{(x, y) \in X \times \mathbb{R}_+ : c^\mathsf{T} x \le u\}$. Then, by compactness of X and closedness of $\mathbf{dom}\, \phi$ (see Proposition 2), S is compact. Since all iterates (x_k, y_k) lives in S (see Lemma 8), $\{(x_k, y_k)\}_k$ has a limit point $(x^*, y^*) \in S$.

Second, we show that (x^*, y^*) is a first-order local optimal solution to (11), which implies its global optimality due to the log-concavity of $\phi(x, y)$. To this end, let $\Delta := (d_x, d_y)$ be an arbitrary tangent direction of S at (x^*, y^*). Then, by definition there exists a sequence $\{(x_\ell, y_\ell)\}_\ell$ in S converging to (x^*, y^*) and $t_\ell \searrow 0$ such that

$$\Delta = \lim_{\ell\to\infty} \frac{(x_\ell, y_\ell) - (x^*, y^*)}{t_\ell}.$$

Then, we examine the directional derivative of $\phi(x, y)$ along direction Δ to obtain

$$\begin{aligned}
\phi'(x^*, y^*; \Delta) &= \phi'\left(x^*, y^*; \lim_{\ell\to\infty} \frac{1}{t_\ell}\Big[(x_\ell, y_\ell) - (x^*, y^*)\Big]\right) \\
&= \lim_{\ell\to\infty} \phi'\left(x^*, y^*; \frac{1}{t_\ell}\Big[(x_\ell, y_\ell) - (x^*, y^*)\Big]\right) \\
&= \lim_{\ell\to\infty} \frac{1}{t_\ell}\phi'\Big(x^*, y^*; (x_\ell, y_\ell) - (x^*, y^*)\Big) \le 0,
\end{aligned}$$

where the second and third equalities follow from the continuity and positive homogeneity of $\phi'(x^*, y^*; \Delta)$ in Δ, respectively (see Lemma 10), and the inequality follows from Lemma 9 because $(x^*, y^*) + (x_\ell, y_\ell) - (x^*, y^*) = (x_\ell, y_\ell) \in S$. This completes the proof. □

Remark 5. If $\ln\phi(x, y)$ has a Lipschitz continuous gradient with respect to x, then it follows from Theorem 3.7 in [Bec15] that Algorithm 1 admits an $O(1/k)$ rate of convergence because all iterates $\{(x_k, y_k)\}_k$ live in $\mathbf{dom}\,\phi$, on which $\ln\phi(x, y)$ is jointly concave in (x, y).

5 Optimistic Chance Constraint

This section extends the convexity results for (**P-CC**) in Section 3 to (**O-CC**). We first present a CVaR reformulation for $\mathcal{X}^o$ by adapting Theorem 1 in [Xie19]. Then, we study individual (**O-CC**) with LHS uncertainty and joint (**O-CC**) with RHS uncertainty in Sections 5.1 and 5.2, respectively.

Theorem 10. For $\delta > 0$ it holds that

$$\mathcal{X}^o = \left\{ x \in \mathbb{R}^n \colon \mathrm{CVaR}_\epsilon\Big(-\mathbf{d}\left(\zeta, \mathcal{S}(x)\right)\Big) + \frac{\delta}{1-\epsilon} \ge 0 \right\},$$

where the CVaR is with respect to the reference distribution $\mathbb{P}$ and $\mathbf{d}\left(\zeta, \mathcal{S}(x)\right)$ is the distance from $\zeta \in \mathbb{R}^m$ to the safe set $\mathcal{S}(x)$,

$$\mathbf{d}\left(\zeta, \mathcal{S}(x)\right) := \inf_{\xi\in\Xi} \left\{ \|\zeta - \xi\| \colon A(x)\xi \le b(x) \right\}.$$

Proof. See Appendix I. □

5.1 Individual (O-CC) with LHS Uncertainty

When $m = 1$, the feasible region of (**O-CC**) simplifies to

$$\mathcal{X}^o_{\mathrm{L}} := \left\{ x \in \mathbb{R}^n \colon \sup_{\mathbb{Q}\in\mathcal{P}} \mathbb{Q}\Big[a(x)^\mathsf{T}\xi \le b(x)\Big] \ge 1-\epsilon \right\}.$$

We show that, under the same sufficient conditions as for individual (**P-CC**), $\mathcal{X}^o_{\mathrm{L}}$ admits a SOC representation.

Theorem 11. Suppose that the reference distribution $\mathbb{P}$ is Gaussian with mean μ and covariance matrix $\Sigma \succ 0$. In addition, suppose that the norm $\|\cdot\|$ defining $\mathcal{P}$ satisfies $\|\cdot\| = \|\Sigma^{-1/2}(\cdot)\|_2$. Then, for $\delta > 0$ it holds that

$$\mathcal{X}^o_{\mathrm{L}} = \left\{\, x \in \mathbb{R}^n \colon a(x)^\mathsf{T}\mu + c_o\|\Sigma^{1/2}a(x)\|_2 \leq b(x) \,\right\},$$

$$\text{where} \quad c_o := \sup_{\epsilon' \in (\epsilon,1)} \frac{-\delta + (1-\epsilon')\mathrm{CVaR}_{\epsilon'}(Y) - (1-\epsilon)\mathrm{CVaR}_{\epsilon}(Y)}{\epsilon' - \epsilon}$$

and Y is a standard, 1-dimensional Gaussian random variable.

Proof. Pick any $x \in \mathcal{X}^o_{\mathrm{L}}$. If $a(x) = 0$, then $\mathbb{Q}\big[a(x)^\mathsf{T}\xi \leq b(x)\big] = 1$ if $b(x) \geq 0$ and $\mathbb{Q}\big[a(x)^\mathsf{T}\xi \leq b(x)\big] = 0$ otherwise. Hence, in this case $\mathcal{X}^o_{\mathrm{L}} = \{\, x \in \mathbb{R}^n \colon b(x) \geq 0 \,\}$, as desired. For the rest of the proof, we assume that $a(x) \neq 0$. Then, for any $\zeta \in \mathbb{R}^m$ the distance to the safe set is $\mathbf{d}\,(\zeta, \mathcal{S}(x)) = (f(x,\zeta))^+$ [CKW18] with

$$f(x,\zeta) = \frac{a(x)^\mathsf{T}\zeta - b(x)}{\|a(x)\|_*}.$$

By Theorem 10, (**O-CC**) is equivalent to

$$\frac{\delta}{1-\epsilon} + \mathrm{CVaR}_\epsilon\Big[\big(-f(x,\zeta)\big)^-\Big] \geq 0,$$

which by Lemma 3 is further equivalent to

$$\frac{\delta}{1-\epsilon} + \mathbb{1}\left\{\, 0 \geq \mathrm{VaR}_\epsilon\big(-f(x,\zeta)\big) \,\right\} \cdot \left(\mathrm{CVaR}_\epsilon\big(-f(x,\zeta)\big) - \frac{1}{1-\epsilon}\mathbb{E}\Big[\big(-f(x,\zeta)\big)^+\Big]\right) \geq 0.$$

It follows that $\mathcal{X}^o_{\mathrm{L}} = \mathcal{X}^o_{\mathrm{L},1} \cup \mathcal{X}^o_{\mathrm{L},2}$ with

$$\mathcal{X}^o_{\mathrm{L},1} := \left\{\, x \in \mathbb{R}^n \colon 0 < \mathrm{VaR}_\epsilon\Big(-f(x,\zeta)\Big) \,\right\},$$

$$\mathcal{X}^o_{\mathrm{L},2} := \left\{ x \in \mathbb{R}^n \colon \begin{array}{l} 0 \geq \mathrm{VaR}_\epsilon\Big(-f(x,\zeta)\Big) \\ \dfrac{\delta}{1-\epsilon} + \mathrm{CVaR}_\epsilon\Big(-f(x,\zeta)\Big) \geq \dfrac{1}{1-\epsilon}\mathbb{E}\Big[\big(-f(x,\zeta)\big)^+\Big]. \end{array} \right\}$$

First, since $f(x,\zeta)$ is Gaussian, $\mathcal{X}^o_{\mathrm{L},1} = \{\, x \in \mathbb{R}^n \colon m(x) < \mathrm{VaR}_\epsilon(Y) \,\}$, where $m(x)$ is defined through

$$m(x) := \frac{a(x)^\mathsf{T}\mu - b(x)}{\|a(x)\|_*}.$$

Second, we address $\mathcal{X}^o_{\mathrm{L},2}$. Since $f(x,\zeta)$ is Gaussian, the first constraint in $\mathcal{X}^o_{\mathrm{L},2}$ simplifies to $m(x) \geq \mathrm{VaR}_\epsilon(Y)$. For the second constraint, we recast the expectation of $\big(-f(x,\zeta)\big)^+$ in the form of CVaR:

$$\begin{aligned}
\mathbb{E}\left[\big(-f(x,\zeta)\big)^+\right] &= \mathbb{E}\Big[-f(x,\zeta) \cdot \mathbb{1}\left\{\, -f(x,\zeta) \geq 0 \,\right\}\Big] \\
&= \max_{t \,\geq\, \mathrm{VaR}_\epsilon\left(-f(x,\zeta)\right)} \mathbb{E}\Big[-f(x,\zeta) \cdot \mathbb{1}\left\{\, -f(x,\zeta) \geq t \,\right\}\Big] \\
&= \max_{\epsilon_t \in [\epsilon,1]} (1-\epsilon_t) \cdot \mathrm{CVaR}_{\epsilon_t}\Big(-f(x,\zeta)\Big),
\end{aligned}$$

where the second equality is because (i) the expectation in this equality attains its maximum at $t = 0$ and (ii) $\mathrm{VaR}_\epsilon\big(-f(x,\zeta)\big) \le 0$ by definition of $\mathcal{X}^o_{\mathrm{L},2}$, and the final equality uses a change-of-variable $\epsilon_t := \mathbb{P}\big[-f(x,\zeta) < t\big]$ (or equivalently, $t = \mathrm{VaR}_{\epsilon_t}(-f(x,\zeta))$). In addition, we notice that $(1-\epsilon_t)\cdot\mathrm{CVaR}_{\epsilon_t}\big(-f(x,\zeta)\big)$ decreases as ϵ_t converges to 1. As a result, we can drop the candidate solution $\epsilon_t = 1$ in the final equality. It follows that

$$\mathcal{X}^o_{\mathrm{L},2} = \left\{ x \in \mathbb{R}^n : \begin{array}{l} m(x) \ge \mathrm{VaR}_\epsilon(Y) \\ \delta + (1-\epsilon)\cdot\mathrm{CVaR}_\epsilon\Big(-f(x,\zeta)\Big) \ge (1-\epsilon_t)\cdot\mathrm{CVaR}_{\epsilon_t}\Big(-f(x,\zeta)\Big), \quad \forall \epsilon_t \in (\epsilon,1) \end{array} \right\}.$$

Above, we further drop the case $\epsilon_t = \epsilon$ in the second constraint because, in that case, this constraint always holds. To finish the reformulation, we rewrite

$$-f(x,\zeta) = -m(x) - \frac{a(x)^\intercal(\zeta-\mu)}{\|a(x)\|_*}$$

and notice that $\big(a(x)^\intercal(\zeta-\mu)\big)/\|a(x)\|_*$ is standard Gaussian. It follows from the translation invariance of CVaR and the symmetry of Y that

$$\mathrm{CVaR}_\epsilon\Big(-f(x,\zeta)\Big) = -m(x) + \mathrm{CVaR}_\epsilon(Y),$$

which implies

$$\begin{aligned} & \delta + (1-\epsilon)\cdot\mathrm{CVaR}_\epsilon\Big(-f(x,\zeta)\Big) \ge (1-\epsilon_t)\cdot\mathrm{CVaR}_{\epsilon_t}\Big(-f(x,\zeta)\Big) \quad \forall \epsilon_t \in (\epsilon,1) \\ \iff & \delta + (1-\epsilon)\cdot\Big(-m(x) + \mathrm{CVaR}_\epsilon(Y)\Big) \ge (1-\epsilon_t)\cdot\Big(-m(x) + \mathrm{CVaR}_{\epsilon_t}(Y)\Big) \quad \forall \epsilon_t \in (\epsilon,1) \\ \iff & -m(x) \ge c_o. \end{aligned}$$

Finally, we notice that $\mathcal{X}^o_{\mathrm{L},2} = \{\, x \in \mathbb{R}^n : c_o \le -m(x) \le -\mathrm{VaR}_\epsilon(Y) \,\}$ is non-empty because for all $\epsilon' \in (\epsilon, 1)$

$$\begin{aligned} & \frac{(1-\epsilon')\mathrm{CVaR}_{\epsilon'}(Y) - (1-\epsilon)\mathrm{CVaR}_\epsilon(Y)}{\epsilon'-\epsilon} \\ = & \frac{\mathbb{E}\Big[Y\cdot\mathbb{1}\{\, Y \ge \mathrm{VaR}_{\epsilon'}(Y) \,\}\Big] - \mathbb{E}\Big[Y\cdot\mathbb{1}\{\, Y \ge \mathrm{VaR}_\epsilon(Y) \,\}\Big]}{\epsilon'-\epsilon} \\ = & -\frac{\mathbb{E}\Big[Y\cdot\mathbb{1}\Big\{\, Y \in \Big[\mathrm{VaR}_\epsilon(Y), \mathrm{VaR}_{\epsilon'}(Y)\Big) \,\Big\}\Big]}{\epsilon'-\epsilon} \\ = & -\mathbb{E}\Big[Y \mid Y \in \Big[\mathrm{VaR}_\epsilon(Y), \mathrm{VaR}_{\epsilon'}(Y)\Big)\Big] \le -\mathrm{VaR}_\epsilon(Y), \end{aligned}$$

which implies that $c_o \le -\mathrm{VaR}_\epsilon(Y)$. Concatenating $\mathcal{X}^o_{\mathrm{L},1}$ and $\mathcal{X}^o_{\mathrm{L},2}$ yields $\mathcal{X}^o_{\mathrm{L}} = \{\, x \in \mathbb{R}^n : -m(x) \ge c_o \,\}$ and completes the proof. □

Remark 6. Like in (**P-CC**), the representation in Theorem 11 can be generalized to the case that $\mathbb{P}$ is radial. In addition, we underscore that the value of c_o is independent of x and can be obtained offline through, e.g., a line search. However, depending on the risk threshold ϵ and Wasserstein radius δ, c_o may take a negative value, rendering $\mathcal{X}^o_{\mathrm{L}}$ *non-convex*. In Figure 1, we depict the "watershed" of (ϵ, δ) combinations that produce a $c_o = 0$ (see the solid curve), while any combination lying below the watershed leads to a convex and SOC $\mathcal{X}^o_{\mathrm{L}}$. From this figure, we observe that, for any ϵ, $\mathcal{X}^o_{\mathrm{L}}$ is convex and SOC as long as δ is small enough.

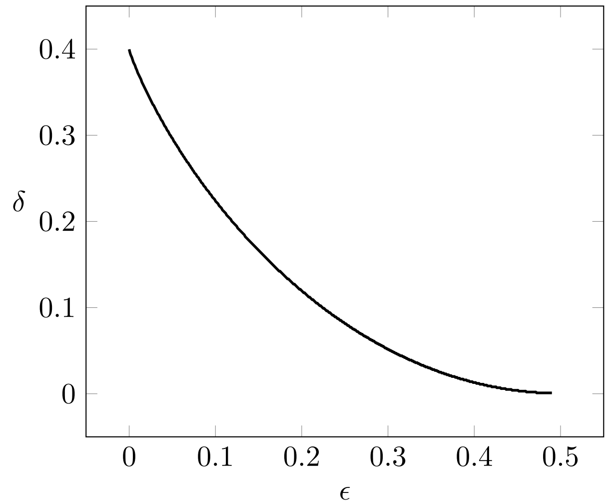


Figure 1: Combinations of the risk threshold ϵ and the Wasserstein radius δ that produce a $c_o = 0$; any (ϵ, δ) combination under the curve leads to a convex and SOC $\mathcal{X}_{\mathrm{L}}^{o}$

Remark 7. Once again, Like in (**P-CC**), it can be shown that c_o increases to $\mathrm{VaR}_{1-\epsilon}(Y)$ as δ tends to zero. Therefore, $\mathcal{X}_{\mathrm{L}}^{o}$ asymptotically recovers the feasible region of (**CC**) with respect to $\mathbb{P} \sim \mathrm{Gaussian}(\mu, \Sigma)$.

5.2 Joint (O-CC) with RHS uncertainty

When $m \geq 2$ and ξ arises from the RHS, we recall the formulation of (**O-CC**):

$$\mathcal{X}_{\mathrm{R}}^{o} := \left\{ x \in \mathbb{R}^n \colon \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{Q}\Big[A\xi \leq b(x) \Big] \geq 1 - \epsilon \right\}.$$

As a preparation, we show that the distance $\mathbf{d}\left(\zeta, \mathcal{S}(x)\right)$ from $\zeta \in \mathbb{R}^m$ to the safe set $\mathcal{S}(x)$ is convex.

Lemma 7. $\mathbf{d}\left(\zeta, \mathcal{S}(x)\right) \equiv \min_{\xi \in \Xi} \Big\{ \|\xi - \zeta\| : A\xi \leq b(x) \Big\}$ is jointly convex in (ζ, x) on $\Xi \times \mathbb{R}^n$.

Proof. See Appendix J. □

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

Theorem 12. Suppose that the reference distribution $\mathbb{P}$ of $\mathcal{P}$ is α-concave with $0 \leq \alpha \leq 1/m$. Then, $\mathcal{X}_{\mathrm{R}}^{o}$ is convex and closed for $\delta > 0$.

Proof. By Theorem 10, (**O-CC**) admits the following reformulations:

$$\begin{aligned}
& \mathrm{CVaR}_{\epsilon}\left\{ -\mathbf{d}\left(\zeta, \mathcal{S}(x)\right) \right\} \geq -\frac{\delta}{1-\epsilon} \\
\iff & \inf_{\gamma \in \mathbb{R}} \left\{ \gamma + \frac{1}{1-\epsilon} \mathbb{E}_{\mathbb{P}} \left\{ \Big[-\mathbf{d}\left(\zeta, \mathcal{S}(x)\right) - \gamma \Big]^{+} \right\} \right\} \geq -\frac{\delta}{1-\epsilon} \\
\iff & \gamma + \frac{1}{1-\epsilon} \mathbb{E}_{\mathbb{P}} \left\{ \Big[-\mathbf{d}\left(\zeta, \mathcal{S}(x)\right) - \gamma \Big]^{+} \right\} \geq -\frac{\delta}{1-\epsilon} \qquad \forall \gamma \in \mathbb{R}.
\end{aligned}$$

In what follows, we prove that the LHS of the last reformulation is log-concave in x for any fixed γ. Since log-concave functions are quasi-concave and continuous (see Lemma 2.4 in [Nor93]), the convexity and closedness of $\mathcal{X}_{\mathrm{R}}^{o}$ follows from their preservation under intersection. To this end, we notice that

$$\mathbb{E}_{\mathbb{P}}\Big[\phi(\zeta, x)\Big] = \int_{\Xi} \phi(x,\zeta) \cdot f_{\zeta}(\zeta)\,\mathrm{d}\zeta,$$

where $\phi(x,\zeta) := \big[-\mathbf{d}\,(\zeta, \mathcal{S}(x)) - \gamma\big]^{+}$ and f_{ζ} represents the probability density function of ζ. It suffices to show that $\phi(x,\zeta)\cdot f_{\zeta}(\zeta)$ is jointly log-concave in (x,ζ) because log-concavity preserves under marginalization (see Theorem 3.3 in [SW14]). In view that log-concavity also preserves under multiplication, we complete the proof by showing that $f_{\zeta}(\zeta)$ is log-concave in ζ and $\phi(x,\zeta)$ is jointly log-concave in (x,ζ).

1. Since $\mathbb{P}$ is α-concave, its density function f_{ζ} is α'-concave by Theorem 2, where

$$\alpha' = \begin{cases} \frac{\alpha}{1-m\alpha} & \text{if } \alpha \in [0, 1/m) \\ +\infty & \text{if } \alpha = 1/m \end{cases}$$

   and $\alpha' \geq 0$. Hence, f_{ζ} is log-concave by Lemma 1.

2. For any pair of $(x_1,\zeta_1), (x_2,\zeta_2) \in \mathbb{R}^n \times \Xi$ and any $\theta \in [0,1]$, define $(x_\theta, \zeta_\theta) := \theta(x_1,\zeta_1) + (1-\theta)(x_2,\zeta_2)$. Then, it holds that

$$\begin{aligned}\phi(x_\theta,\zeta_\theta) = \Big(-\mathbf{d}\,(\zeta_\theta, \mathcal{S}(x_\theta)) - \gamma\Big)^{+} &\geq \Big(m_1\big(-\mathbf{d}\,(\zeta_1,\mathcal{S}(x_1)) - \gamma, -\mathbf{d}\,(\zeta_2,\mathcal{S}(x_2)) - \gamma; \theta\big)\Big)^{+} \\ &\geq m_0\Big(\phi(x_1,\zeta_1), \phi(x_2,\zeta_2); \theta\Big),\end{aligned}$$

   where the first inequality is because $\mathbf{d}\,(\zeta, \mathcal{S}(x))$ is jointly convex in (x,ζ). To see the second inequality, we discuss the following two cases.

   (i) If either $\phi(x_1,\zeta_1)$ or $\phi(x_2,\zeta_2)$ equals zero, then $m_0\big(\phi(x_1,\zeta_1), \phi(x_2,\zeta_2);\theta\big)$ equals zero by definition.

   (ii) If both $\phi(x_1,\zeta_1)$ and $\phi(x_2,\zeta_2)$ are strictly positive, then

$$\begin{aligned}\Big(m_1\big(-\mathbf{d}\,(\zeta_1,\mathcal{S}(x_1)) - \gamma, -\mathbf{d}\,(\zeta_2,\mathcal{S}(x_2)) - \gamma;\theta\big)\Big)^{+} &= m_1\Big(\phi(x_1,\zeta_1), \phi(x_2,\zeta_2);\theta\Big) \\ &\geq m_0\Big(\phi(x_1,\zeta_1), \phi(x_2,\zeta_2);\theta\Big),\end{aligned}$$

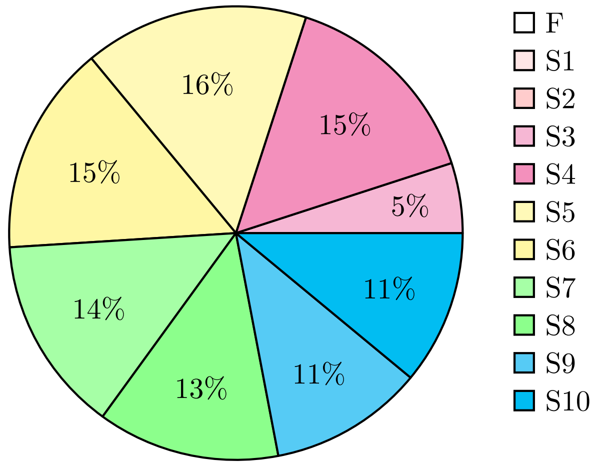
   where the inequality follows from Lemma 1.

□

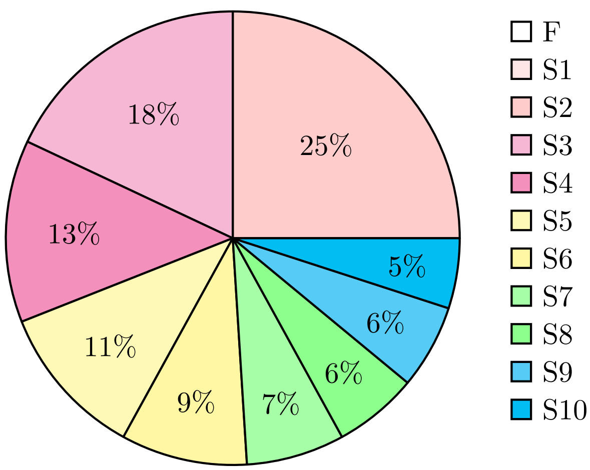
Remark 8. Although Theorem 12 is concerned with (**O-CC**) with linear inequalities, its proof extends to (**O-CC**) with nonlinear inequalities, as long as $\mathbf{d}\,(\zeta, \mathcal{S}(x))$ remains jointly convex in (x,ζ).

6 Numerical Experiments

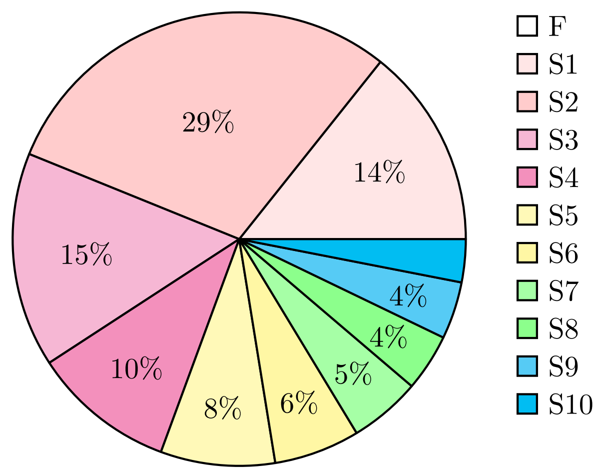
We demonstrate the theoretical results through two numerical experiments: a (**PTO**) model using individual (**P-CC**) or (**O-CC**) in Section 6.1 and a (**PP**) model using joint (**P-CC**) in Section 6.2.



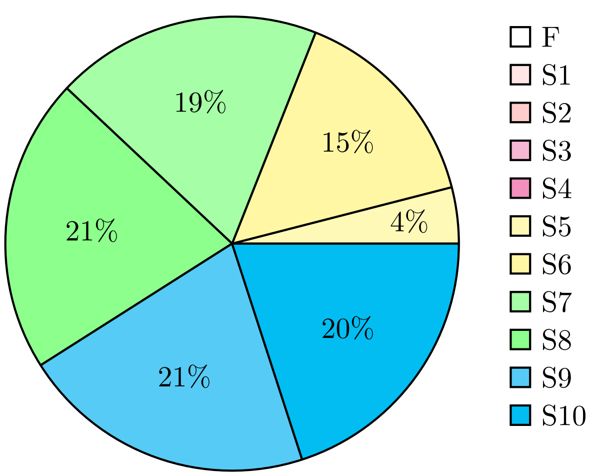
(a) (**CC**) Optimal Portfolio; F = Fixed Return, $S1$ = Least Profitable/Risky, $S10$ = Most Profitable/Risky



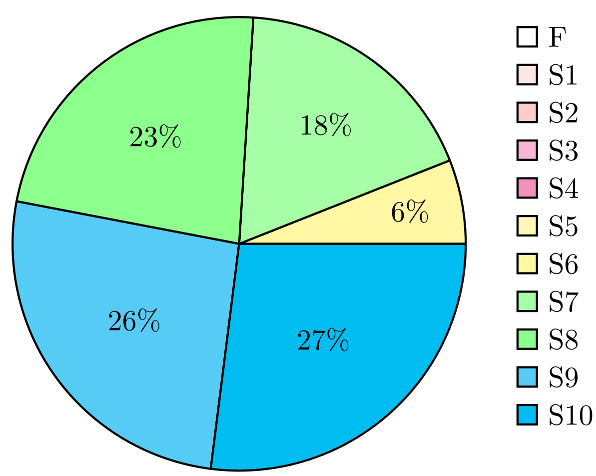
(b) (**P-CC**) with $\delta = 0.005$



(c) (**P-CC**) with $\delta = 0.01$



(d) (**O-CC**) with $\delta = 0.005$



(e) (**O-CC**) with $\delta = 0.01$

Figure 2: Optimal Portfolios of (**CC**), (**P-CC**), and (**O-CC**) Models

6.1 Portfolio Management

We consider a (**PTO**) model that seeks to maximize the expected return of a portfolio, while achieving a certain target return with high probability (see Example 1 and Section 7 in [XCM12]). Specifically, we consider 11 investments consisting of a fixed deposit, denoted by F, with a deterministic rate of return 1 and 10 stocks, denoted by $S1$–$S10$, with random rate of returns R_i and

$$R_i := R_{0,i} + r \qquad \forall i \in [10],$$

where $R_{0,i} \sim \mathcal{N}\big(1 + 0.01i, (0.03i)^2\big)$ represents the randomness of stock i and $r \sim \mathcal{N}\big(0, (0.01)^2\big)$ denotes the market effect on all stocks. Hence, $S1$ is the least profitable/risky and $S10$ is the most profitable/risky. We consider the following models based on (**P-CC**) and (**O-CC**), respectively:

$$\max_{\substack{x \in \mathbb{R}^n_+,\\ \mathbf{1}^\mathsf{T} x = 1}} \left\{ \mathbb{E}\left[\sum_{i=1}^n R_i x_i\right] \text{ s.t. } \inf_{\mathbb{Q} \in \mathcal{P}} \mathbb{Q}\left(\sum_{i=1}^n R_i x_i \geq \eta\right) \geq 1 - \epsilon \right\},$$

$$\max_{\substack{x \in \mathbb{R}^n_+,\\ \mathbf{1}^\mathsf{T} x = 1}} \left\{ \mathbb{E}\left[\sum_{i=1}^n R_i x_i\right] \text{ s.t. } \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{Q}\left(\sum_{i=1}^n R_i x_i \geq \eta\right) \geq 1 - \epsilon \right\},$$

where $\eta = 1.0$ represents the target return and ϵ is set to be 0.15. We characterize the Wasserstein ball $\mathcal{P}$ so that the assumptions of Theorems 6 and 11 are satisfied and the two models above admit SOC representations. As a benchmark, we also consider a (**CC**) model by setting $\delta = 0$ in either model above.

The optimal portfolios produced by the (**CC**), (**P-CC**), and (**O-CC**) models are displayed in Figure 2. Comparing the optimal portfolio of (**CC**) in Figure 2a with those of (**P-CC**) in Figures 2b–2c, we observe that a pessimistic investor decreases her investment in the more profitable/risky stocks. In contrast, an optimistic investor focuses on the more profitable/risky stocks (see Figures 2d–2e).

6.2 Production Planning

We consider a (**PP**) model that seeks to procure production capacity so that all demands can be satisfied with high probability and a minimal procurement cost (see Example 2). Specifically, we consider the following formulation with (**P-CC**):

$$\begin{aligned} \min \ & c^\mathsf{T} x, \\ \text{s.t. } & \inf_{\mathbb{Q} \in \mathcal{P}} \mathbb{Q}\left[Tx \geq \xi\right] \geq 1 - \epsilon, \\ & 0 \leq x_i \leq U, \forall i \in [n], \end{aligned}$$

where c represents the procurement costs, U represents a homogeneous upper bound of production capacity for all facilities, and the reference distribution $\mathbb{P}$ of the Wasserstein ball $\mathcal{P}$ is assumed to be pairwise independent and Gaussian. To apply Algorithm 1, we switch the objective function with (**P-CC**) to obtain

$$\begin{aligned} \rho(u) = \max_{x \in \mathbb{R}^n_+, y \in \mathbb{R}_+} \ & \phi(x, y) \equiv \int_0^y \left(\mathbb{P}\left[\min_{t \in [m]} (T_i x - \zeta_i) \geq t\right] - (1 - \epsilon) \right) \mathrm{d}t \\ \text{s.t. } & c^\mathsf{T} x \leq u, \\ & 0 \leq x_i \leq U, \quad \forall i \in [n], \end{aligned}$$

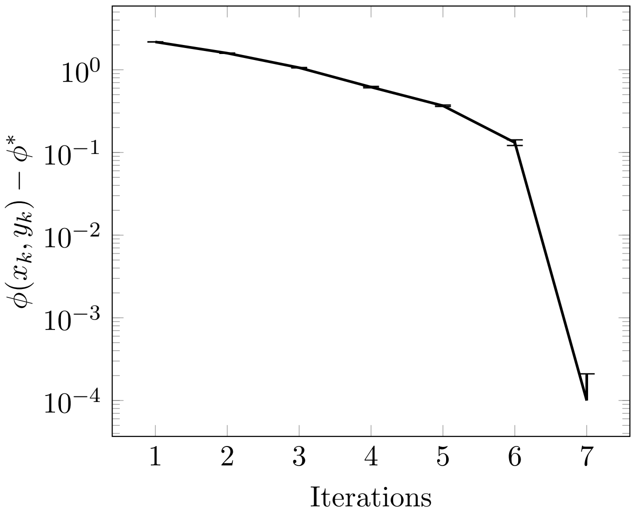
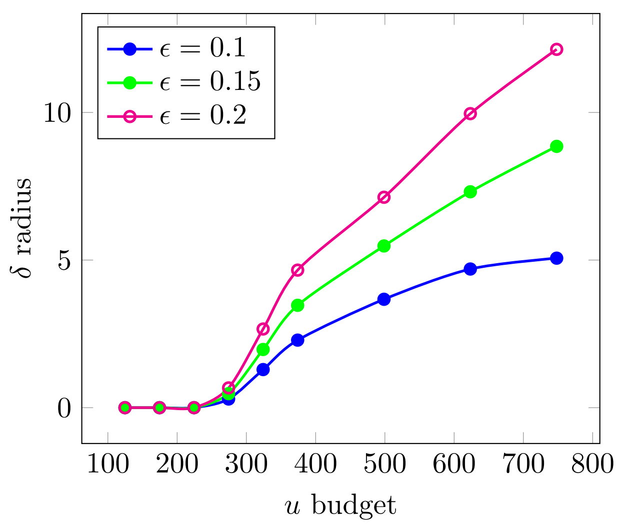


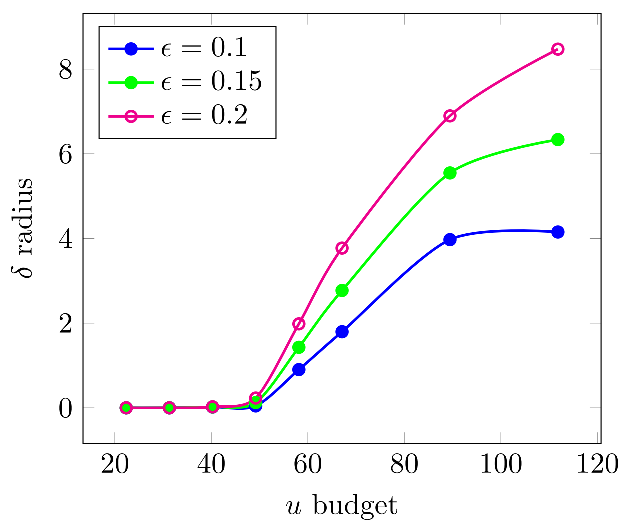
Figure 3: Convergence of Algorithm 1 on Production Planning Instances; solid line = average of the difference $\phi(x_k, y_k) - \phi^*$ across five runs, error bar = standard deviation of the difference

where we adjust the procurement budget u and apply the algorithm with various u to obtain a risk envelope. In addition, when applying Algorithm 1, we employ the stochastic approach described in [Nor93] to be the oracle $\mathcal{O}_u(y_k, \varepsilon_k)$ in Step 5 and terminate the algorithm whenever the change in y_k becomes sufficiently small, specifically, when $|y_k - y_{k+1}| \leq 10^{-6}$.

We demonstrate the convergence of Algorithm 1 in Figure 3, which is obtained by running the algorithm for five times on an instance with $n = 10$, $m = 5$, $U = 200$, c randomly drawn from the set $\{1, \ldots, 10\}$, and $\mathbb{E}_{\mathbb{P}}[\zeta_i]$ randomly drawn from the interval $[10, 51]$. In this figure, the solid line represents the difference between each iterate $\phi(x_k, y_k)$ and the final iterate ϕ^*, averaged across the five runs, and the error bar represents the standard deviation of the difference. From this figure, we observe that Algorithm 1 converges at a linear rate in only a few iterations.



(a) $n = 10$ and $m = 5$



(b) $n = 30$ and $m = 20$

Figure 4: Risk envelopes under different risk thresholds

We demonstrate the trade-off between the robustness and the budget in Figure 4, which is obtained by solving instances with $\epsilon \in \{0.1, 0.15, 0.2\}$, $n \in \{10, 30\}$, and $m \in \{5, 20\}$. The vertical axis of this figure represents $\rho(u)$, i.e., the largest Wasserstein radius δ that allows (**P-CC**) to be satisfied. From this figure, we observe that, for fixed ϵ, the largest allowable δ is an S-shaped function of the the budget u. That is, δ remains at zero for small u, and then δ increases with a diminishing momentum as u becomes larger. In addition, for fixed δ, it needs a larger budget u to keep (**P-CC**) satisfied as ϵ decreases.

Appendix A Proof of Lemma 2

Proof. When $\alpha \geq 1$, the result follows from Theorem 1. When $\alpha = 0$, the result was proved in [BBV04] (see Exercise 3.48). When $\alpha = -\infty$, shifting the function along the vertical direction does not affect the convexity of its super level sets. Hence, it suffices to prove the result when $\alpha < 1$ and $\alpha \neq 0$.

We notice that D is convex as it is the super-level set of the quasi-concave function f. Now, for any $x_1, x_2 \in D$ and $\theta \in (0,1)$, the following holds for $x_\theta := \theta x_1 + (1-\theta)x_2$:

$$f(x_\theta) \geq \Big(\theta \cdot (f(x_1))^\alpha + (1-\theta) \cdot (f(x_2))^\alpha\Big)^{1/\alpha}. \tag{13}$$

By Minkowski's Inequality with p set to be α, we have

$$\begin{aligned}\Big(\Big[\theta^{1/\alpha} \cdot f(x_1)\Big]^\alpha + \Big[(1-\theta)^{1/\alpha} \cdot f(x_2)\Big]^\alpha\Big)^{1/\alpha} &\geq \Big(\Big[\theta^{1/\alpha} \cdot (f(x_1) - c)\Big]^\alpha + \Big[(1-\theta)^{1/\alpha} \cdot (f(x_2) - c)\Big]^\alpha\Big)^{1/\alpha} \\ &\quad + \Big(\Big[\theta^{1/\alpha} \cdot c\Big]^\alpha + \Big[(1-\theta)^{1/\alpha} \cdot c\Big]^\alpha\Big)^{1/\alpha},\end{aligned}$$

from which we obtain

$$\Big(\theta \cdot (f(x_1) - c)^\alpha + (1-\theta) \cdot (f(x_2) - c)^\alpha\Big)^{1/\alpha} + c \leq \Big(\theta \cdot (f(x_1))^\alpha + (1-\theta) \cdot (f(x_2))^\alpha\Big)^{1/\alpha}. \tag{14}$$

Combining (13) and (14) concludes the proof:

$$\begin{aligned}f(x_\theta) - c &\geq \Big(\theta \cdot (f(x_1))^\alpha + (1-\theta) \cdot (f(x_2))^\alpha\Big)^{1/\alpha} - c \\ &\geq \Big(\theta \cdot (f(x_1) - c)^\alpha + (1-\theta) \cdot (f(x_2) - c)^\alpha\Big)^{1/\alpha}.\end{aligned}$$

□

Appendix B Proof of Lemma 3

Proof. By definition of CVaR, we have

$$\text{CVaR}_{1-\epsilon}\left(X^-\right) = \text{VaR}_{1-\epsilon}\left(X^-\right) + \frac{1}{\epsilon} \cdot \mathbb{E}\left[X^- - \text{VaR}_{1-\epsilon}(X^-)\right]^+.$$

We discuss two cases:

(i) If $0 < \text{VaR}_{1-\epsilon}(X)$, then $\text{VaR}_{1-\epsilon}(X^-) = 0$, from which

$$\text{CVaR}_{1-\epsilon}(X^-) = 0 + \frac{1}{\epsilon} \cdot \mathbb{E}\left[X^- - 0\right]^+ = 0.$$

(ii) If $0 \geq \text{VaR}_{1-\epsilon}(X)$, then $\text{VaR}_{1-\epsilon}(X^-) = \text{VaR}_{1-\epsilon}(X)$. It follows that

$$\begin{aligned}
\mathbb{E}\left[X^- - \text{VaR}_{1-\epsilon}(X)\right]^+ &= \mathbb{E}\left[\left(X^- - \text{VaR}_{1-\epsilon}(X)\right) \cdot \mathbb{1}\left\{ X^- \geq \text{VaR}_{1-\epsilon}(X) \right\}\right] \\
&= \mathbb{E}\left[\left(X - \text{VaR}_{1-\epsilon}(X) - X^+\right) \cdot \mathbb{1}\left\{ X \geq \text{VaR}_{1-\epsilon}(X) \right\}\right] \\
&= \mathbb{E}\left[\left(X - \text{VaR}_{1-\epsilon}(X)\right) \cdot \mathbb{1}\left\{ X \geq \text{VaR}_{1-\epsilon}(X) \right\}\right] \\
&\quad - \mathbb{E}\left[X^+ \cdot \mathbb{1}\left\{ X \geq \text{VaR}_{1-\epsilon}(X) \right\}\right] \\
&= \mathbb{E}\left[\left(X - \text{VaR}_{1-\epsilon}(X)\right) \cdot \mathbb{1}\left\{ X \geq \text{VaR}_{1-\epsilon}(X) \right\}\right] - \mathbb{E}\left[X^+\right] \\
&= \mathbb{E}\left[X - \text{VaR}_{1-\epsilon}(X)\right]^+ - \mathbb{E}\left[X^+\right],
\end{aligned}$$

where the first equality is by definitions of positive part $[\cdot]^+$ and $\mathbb{1}\{\cdot\}$, the second is due to $\text{VaR}_{1-\epsilon}(X) \leq 0$ and the definitions of positive and negative parts, and the fourth is because $X < \text{VaR}_{1-\epsilon}(X)$ implies $X^+ = 0$. We conclude the proof by noticing that

$$\begin{aligned}
\text{VaR}_{1-\epsilon}\left(X^-\right) + \frac{1}{\epsilon} \cdot \mathbb{E}\left[X^- - \text{VaR}_{1-\epsilon}(X^-)\right]^+ &= \text{VaR}_{1-\epsilon}\left(X\right) + \frac{1}{\epsilon} \cdot \mathbb{E}\left[X - \text{VaR}_{1-\epsilon}(X)\right]^+ - \frac{1}{\epsilon} \cdot \mathbb{E}\left[X^+\right] \\
&= \text{CVaR}_{1-\epsilon}(X) - \frac{1}{\epsilon} \cdot \mathbb{E}\left[X^+\right].
\end{aligned}$$

□

Appendix C Proof of Lemma 4

Proof. We denote the set of points whose distance to $\mathcal{S}^c(x)$ is exactly y by

$$E := \left\{ \zeta \in \Xi \colon \mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right) = y \right\}.$$

We notice that $\mathbf{d}\left(\zeta, \mathcal{S}^c(x)\right) = \mathbf{d}\left(\zeta, \mathbf{cl}\, \mathcal{S}^c(x)\right)$, where $\mathbf{cl}\, \mathcal{S}^c(x)$ denotes the closure of $\mathcal{S}^c(x)$. Then, by the item (1) of [Erd45], we have $\mathbf{Leb}(E) = 0$, which further implies that $\mathbb{P}(E) = 0$ because $\mathbb{P}$ is absolutely continuous with respect to $\mathbf{Leb}(\cdot)$ (see Theorem 2.2 in [Nor93]).

In addition, the Lebesgue measure of the event $\{\zeta \in \Xi : f(x, \zeta) = y\}$ equals zero because $a_i(x) \neq 0$ for all $i \in [m] \setminus I(x)$. It follows that $f(x, \zeta)$ is atomless because $\mathbb{P}$ is absolutely continuous with respect to $\mathbf{Leb}(\cdot)$. □

Appendix D Proof of Remark 3

Proof. First, we define

$$C_p(\epsilon') = \epsilon \text{CVaR}_{1-\epsilon}(Y) - \epsilon' \text{CVaR}_{1-\epsilon'}(Y)$$

and recall from the statement of Theorem 6 that

$$c_p = \inf_{\epsilon' \in (0, \epsilon)} \frac{\delta + C_p(\epsilon')}{\epsilon - \epsilon'}. \tag{15}$$

Since $C_p'(\epsilon') = -\text{VaR}_{1-\epsilon'}(Y)$ and $C_p''(\epsilon') = 1/f_Y(\text{VaR}_{1-\epsilon'}(Y)) > 0$, where f_Y denotes the probability density function of Y, $C_p(\epsilon')$ is convex. Then, the objective function in (15) is quasiconvex because it is a quotient of a convex function over an affine function (see Example 3.38 in [BBV04]). In addition, the first derivative of this objective function reads

$$G(\epsilon') := \frac{C_p'(\epsilon')(\epsilon - \epsilon') + (\delta + C_p(\epsilon'))}{(\epsilon - \epsilon')^2}.$$

It follows that $G(\epsilon')$ tends to $-\infty$ as ϵ' tends to zero and to $+\infty$ as ϵ' tends to ϵ. As a result, to obtain c_p it suffices to solve the equation $G(\epsilon') = 0$, i.e., to search for an ϵ'_* such that

$$\delta + \int_{1-\epsilon}^{1-\epsilon'_*} \mathrm{VaR}_q(Y)\,\mathrm{d}q = \mathrm{VaR}_{1-\epsilon'_*}(Y) \cdot (\epsilon - \epsilon'_*), \tag{16}$$

where $C_p(\epsilon'_*) = \int_{1-\epsilon}^{1-\epsilon'_*} \mathrm{VaR}_q(Y)\,\mathrm{d}q$ was established in the proof of Theorem 6. We make two observations.

(i) The first derivatives of the LHS and the RHS of (16) are $-\mathrm{VaR}_{1-\epsilon'_*}(Y)$ and $-\mathrm{VaR}_{1-\epsilon'_*}(Y) - (\epsilon - \epsilon'_*)/f_Y(-\mathrm{VaR}_{1-\epsilon}(Y))$, respectively. It follows that ϵ'_* increases as δ decreases.

(ii) ϵ'_* converges to ϵ as δ decreases to zero. For contradiction, suppose that ϵ'_* converges to $\overline{\epsilon} < \epsilon$. Then, driving δ down to zero in (16) leads to

$$0 = \int_{1-\epsilon}^{1-\overline{\epsilon}} \mathrm{VaR}_q(Y)\,\mathrm{d}q - \mathrm{VaR}_{1-\overline{\epsilon}}(Y) \cdot (\epsilon - \overline{\epsilon}) > 0$$

as desired, where the inequality is because $\overline{\epsilon} < \epsilon$.

It follows that

$$\lim_{\delta \searrow 0} c_p = \lim_{\epsilon'_* \nearrow \epsilon} \frac{\delta + C_p(\epsilon'_*)}{\epsilon - \epsilon'_*} = \lim_{\epsilon'_* \nearrow \epsilon} \frac{-C'_p(\epsilon'_*)(\epsilon - \epsilon'_*)}{\epsilon - \epsilon'_*} = \mathrm{VaR}_{1-\epsilon}(Y),$$

where the second equality follows from the equation $G(\epsilon'_*) = 0$. This completes the proof. □

Appendix E Proof of Lemma 5

Proof. We show that the hypograph of $\mathrm{VaR}_{1-\epsilon}(f(x,\zeta))$, i.e.,

$$\mathcal{H} := \{\, (x,\theta) \colon \mathrm{VaR}_{1-\epsilon}(f(x,\zeta)) \ge \theta \,\}$$

is convex. To this end, we note that

$$\mathrm{VaR}_{1-\epsilon}(f(x,\zeta)) \ge \theta \iff \mathbb{P}\{\, f(x,\zeta) \le \theta \,\} \le 1-\epsilon \iff \mathbb{P}\{\, f(x,\zeta) - \theta \ge 0 \,\} \ge \epsilon$$

where both equivalences are because $f(x,\zeta)$ is atomless. Since $f(x,\zeta) - \theta$ is jointly concave in (x,ζ,θ) and $\mathbb{P}$ is α-concave, $\mathbb{P}\{\, f(x,\zeta) - \theta \ge 0 \,\}$ is α-concave in (x,θ) on the set

$$\mathcal{H}' := \{\, (x,\theta) \colon \exists\, \zeta \text{ such that } f(x,\zeta) - \theta \ge 0 \,\}$$

by Theorem 7. Now, since $\mathcal{H} \subseteq \mathcal{H}'$, $\mathbb{P}\{\, f(x,\zeta) - \theta \ge 0 \,\}$ is also α-concave on $\mathcal{H}$ and $\mathcal{H}$ is convex because it is a super level set of $\mathbb{P}\{\, f(x,\zeta) - \theta \ge 0 \,\}$. □

Appendix F Proof of Lemma 6

Proof. For any $(\widehat{x}, \widehat{t}) \in \mathbb{R}^n \times \mathbb{R}_+$, consider a sequence $\{\, (x_k, t_k) \,\}_k$ that converges to $(\widehat{x}, \widehat{t})$ as k goes to infinity. Then, for any $\zeta \in \Xi$ such that $f(\widehat{x},\zeta) - \widehat{t} \ne 0$, we have

$$\lim_{k\to\infty} \mathbb{1}\{\, f(x_k,\zeta) \ge t_k \,\} = \mathbb{1}\{\, f(\widehat{x},\zeta) \ge \widehat{t} \,\}$$

because the function $f(x,\zeta)-t$ is continuous in (x,t). Hence, as a function of ζ, $\mathbb{1}\{\, f(x_k,\zeta)\geq t_k \,\}$ converges pointwise to $\mathbb{1}\left\{\, f(\widehat{x},\zeta)\geq \widehat{t}\,\right\}$ on the complement of

$$\mathcal{U}_0 := \{\, \zeta\in\Xi\colon f(\widehat{x},\zeta)=\widehat{t}\,\}\,.$$

It follows that

$$\begin{aligned}
\lim_{k\to\infty}\psi(x_k,t_k)+(1-\epsilon) &= \lim_{k\to\infty}\mathbb{P}\left[f(x_k,\zeta)\geq t_k\right]\\
&= \lim_{k\to\infty}\int_{\Xi\setminus\mathcal{U}_0}\mathbb{1}\{\,\zeta\colon f(x_k,\zeta)\geq t_k\,\}\,\mathrm{d}\mathbb{P}(\zeta)\\
&= \int_{\Xi\setminus\mathcal{U}_0}\lim_{k\to\infty}\mathbb{1}\{\,\zeta\colon f(x_k,\zeta)\geq t_k\,\}\,\mathrm{d}\mathbb{P}(\zeta)\\
&= \int\mathbb{1}\left\{\,\zeta\colon f(\widehat{x},\zeta)\geq \widehat{t}\,\right\}\mathrm{d}\mathbb{P}(\zeta) = \psi(\widehat{x},\widehat{t})+(1-\epsilon),
\end{aligned}$$

where the second and fourth equality are because $\mathbf{Leb}(\mathcal{U}_0(x,t))=0$, and the third equality is by the dominated convergence theorem. The continuity of ϕ can be established in a similar way: let $\{\,(x_k,y_k)\,\}_k$ be a sequence such that converges to $(\widehat{x},\widehat{y})$. Then,

$$\begin{aligned}
\lim_{k\to\infty}\phi(x_k,y_k) &= \int_{\mathbb{R}_+}\lim_{k\to\infty}\psi(x_k,t)\cdot\mathbb{1}\{\,t\leq y_k\,\}\,\mathrm{d}t = \int_{\mathbb{R}_+\setminus\{\,\widehat{y}\,\}}\psi(\widehat{x},t)\cdot\lim_{k\to\infty}\mathbb{1}\{\,t\leq y_k\,\}\,\mathrm{d}t\\
&= \int_{\mathbb{R}_+\setminus\{\,\widehat{y}\,\}}\psi(\widehat{x},t)\cdot\mathbb{1}\{\,t\leq \widehat{y}\,\}\,\mathrm{d}t = \phi(\widehat{x},\widehat{y}),
\end{aligned}$$

where the first equality is by the dominated convergence theorem, and the second equality is because ψ is continuous and $\mathbb{1}\{\,t\leq y_k\,\}$ has a limit as $k\to\infty$ when $t\neq\widehat{y}$. This completes the proof. □

Appendix G Proof of Proposition 2

Proof. We first show the α_1^*-concavity of $\phi(x,y)$ using a similar argument as in the proof of Theorem 8. Recall that $\psi(x,t)=\mathbb{P}[f(x,\zeta)\geq t]-(1-\epsilon)$ and $\phi(x,y)=\int_0^y\psi(x,t)\,\mathrm{d}t$. Pick any $(x_0,y_0),(x_1,y_1)\in\mathbf{dom}\,\phi$, then their midpoint $(x_{1/2},y_{1/2}):=\frac{1}{2}(x_0,y_0)+\frac{1}{2}(x_1,y_1)$ lies in $\mathbf{dom}\,\phi$ because $\mathbf{dom}\,\phi$ is convex by Lemma 5. Define $S_i=[0,y_i]$ and pick any $t_i\in S_i$ for $i=0,1$. Since $\psi(x,t)$ is α-concave by Lemma 2, it holds that

$$\psi(x_{1/2},t_{1/2})\geq m_\alpha\left[\psi(x_0,t_0),\psi(x_0,t_0);\frac{1}{2}\right].$$

It follows from Theorem 3 that

$$\int_{\frac{1}{2}S_0+\frac{1}{2}S_1}\psi(x_{1/2},t)\,\mathrm{d}t\geq m_{\alpha_1^*}\left[\int_{S_0}\psi(x_0,t)\,\mathrm{d}t,\int_{S_1}\psi(x_1,t)\,\mathrm{d}t;\frac{1}{2}\right],$$

or equivalently, $\phi(x_{1/2},y_{1/2})\geq m_{\alpha_1^*}[\phi(x_0,y_0),\phi(x_1,y_1);1/2]$. This shows the midpoint α_1^*-concavity of $\phi(x,y)$, which together with its continuity (see Lemma 6) shows the α_1^*-concavity.

Second, the closedness of $\mathbf{dom}\,\phi$ follows from the continuity of ψ by Lemma 6.

Third, we show that constraints (10b)–(10c) are equivalent to (10d). To this end, we pick any x that satisfies (10b)–(10c). Then, by letting $y:=\mathrm{VaR}_\epsilon\big(f(x,\zeta)\big)\geq 0$, we obtain $\delta\leq\phi(x,y)$, which implies constraint (10d). On the contrary, pick any x that satisfies (10d). Then, by definition there exists a $y \geq 0$ such that $\delta \leq \phi(x, y)$. Since $\delta > 0$ and $\phi(x, y) = \int_0^y \big(\mathbb{P}\big[f(x,\zeta) \geq t\big] - (1-\epsilon)\big)\, \mathrm{d}t$, there exists a $t \in [0, y]$ such that $\mathbb{P}\big[f(x,\zeta) \geq t\big] \geq (1-\epsilon)$, which implies that $\mathbb{P}\big[f(x,\zeta) \geq 0\big] \geq (1-\epsilon)$, i.e., constraint (10c). Finally, we notice that $\phi(x,y) \leq \phi\big(x, \mathrm{VaR}_\epsilon\big(f(x,\zeta)\big)\big)$ and hence $\delta \leq \phi\big(x, \mathrm{VaR}_\epsilon\big(f(x,\zeta)\big)\big)$, i.e., constraint (10b). This completes the proof. □

Appendix H Proofs of Preparatory Lemmas 8, 9, and 10

Lemma 8. Let $\{\, (x_k, y_k) \,\}_k$ represent a sequence of iterates produced by Algorithm 1. Then, all iterates are feasible, i.e., $(x_k, y_k) \in S$ for all k. In addition, it holds that

$$\lim_{k\to\infty} \phi(x_k, y_k) = \lim_{k\to\infty} \phi(x_{k+1}, y_k).$$

Proof. First, recall that $S \equiv \mathbf{dom}\,\phi \cap \{\, (x,y) \in X \times \mathbb{R}_+ : c^\mathsf{T} x \leq u \,\}$ is compact. Since $\phi(x, y)$ is continuous by Lemma 6, it is bounded on S. In addition, we notice that by construction the ϕ-values of the iterates produced by Algorithm 1 are non-decreasing, i.e.,

$$0 < \phi(x_1, y_1) \leq \phi(x_2, y_1) \leq \phi(x_2, y_2) \leq \cdots \leq \phi(x_k, y_k) \leq \phi(x_{k+1}, y_k) \leq \cdots \tag{17}$$

Hence, this non-decreasing, bounded sequence converges to a finite value. It follows that the two subsequences $\{\, \phi(x_k, y_k) \,\}_k$ and $\{\, \phi(x_{k+1}, y_k) \,\}_k$ converge to the same limit.

Second, we recall that $(x_1, y_1) \in S$ by construction. For all $k \geq 2$, $\phi(x_{k+1}, y_k) > 0$ by (17), which implies that there exists a $t \in [0, y_k]$ such that $\mathbb{P}\big[f(x_{k+1},\zeta) \geq t\big] > 1-\epsilon$. Then, $\mathbb{P}\big[f(x_{k+1},\zeta) \geq 0\big] > 1-\epsilon$, or equivalently, $\mathrm{VaR}_\epsilon\big(f(x_{k+1},\zeta)\big) > 0$. It follows that $y_{k+1} \equiv \mathrm{VaR}_\epsilon\big(f(x_{k+1},\zeta)\big) \geq 0$ and so $(x_{k+1}, y_{k+1}) \in S$. This completes the proof. □

Lemma 9. Let (x^*, y^*) represent a limit point of the sequence $\{\, (x_k, y_k) \,\}_k$. Then, it holds that

$$\phi(x^* + d_x, y^*) \leq \phi(x^*, y^*) \quad \text{and} \quad \phi(x^*, y^* + d_y) \leq \phi(x^*, y^*)$$

for all $d_x \in \mathbb{R}^n, d_y \in \mathbb{R}$ such that $(x^* + d_x, y^*) \in S$ and $(x^*, y^* + d_y) \in S$. In addition, if $(x^* + d_x, y^* + d_y) \in S$, then the directional derivative of $\phi(x, y)$ along (d_x, d_y) satisfies

$$\phi'(x^*, y^*; (d_x, d_y)) := \lim_{s\to 0^+} \frac{1}{s}\Big[\phi(x^* + s d_x, y^* + s d_y) - \phi(x^*, y^*)\Big] \leq 0.$$

Proof. We split the proof into three parts: the perturbation along $(0, d_y)$, the perturbation along $(d_x, 0)$, and the directional derivative $\phi'(x^*, y^*; (d_x, d_y))$. For notation brevity, we assume, by passing to a subsequence if needed, that $\{(x_k, y_k)\}_k$ converges to (x^*, y^*).

(Perturbation along $(0, d_y)$**)** By definition of (x^*, y^*), it holds that

$$\begin{aligned}
\Big|y^* - \mathrm{VaR}_\epsilon\big(f(x^*,\zeta)\big)\Big| &= \Big|\lim_{k\to\infty} y_k - \mathrm{VaR}_\epsilon\big(f(\lim_{k\to\infty} x_k, \zeta)\big)\Big| \\
&= \Big|\lim_{k\to\infty} \big(y_k - \mathrm{VaR}_\epsilon\big(f(x_k,\zeta)\big)\big)\Big| \\
&= \lim_{k\to\infty} |\varepsilon_k| = 0,
\end{aligned}$$

where the second and third equalities are due to the continuity of $\mathrm{VaR}_\epsilon(f(x,\zeta))$ (see Lemma 5) and $|\cdot|$, respectively. Therefore, $\phi(x^*, y^* + d_y) \leq \phi(x^*, y^*)$ because $y^* = \mathrm{VaR}_\epsilon(f(x^*,\zeta))$ is a maximizer of $\phi(x^*, y)$ for fixed x^*.

(Perturbation along $(d_x, 0)$**)** First, suppose that $(x^* + d_x, y^*)$ lies in the interior of S, denoted by $\text{int}(S)$. Then, since $\{(x_k, y_k)\}_k$ converges to (x^*, y^*), there exist neighborhoods $N \subseteq S$ and $N^d \subseteq S$ of (x^*, y^*) and $(x^* + d_x, y^*)$, respectively, such that $(x_k, y_k) \in N$ and $(x_k + d_x, y_k) \in N^d$ for sufficiently large k. Then, by construction it holds that

$$\phi(x_k + d_x, y_k) \leq \max_x \phi(x, y_k) \leq \phi(x_{k+1}, y_k) + \varepsilon_k.$$

Driving k to infinity yields

$$\phi(x^* + d_x, y^*) \leq \phi(x^*, y^*)$$

by continuity of ϕ and Lemma 8.

Second, suppose that $(x^* + d_x, y^*)$ lies on the boundary of S. Then, for all positive integers M, $(x^*+(1-1/M)d_x, y^*) \in \text{int}(S)$ by convexity of S. It follows that $\phi(x^*+(1-1/M)d_x, y^*) \leq \phi(x^*, y^*)$. Driving M to infinity yields $\phi(x^* + d_x, y^*) \leq \phi(x^*, y^*)$ by continuity of ϕ.

(Directional derivative) Since $\phi(x, y)$ is log-concave and $\phi(x^*, y^*) > 0$, ϕ is directionally differentiable at (x^*, y^*) by Lemma 2.4 in [Nor93]. Hence, $\phi'(x^*, y^*; (d_x, d_y))$ is well-defined. To compute $\phi'(x^*, y^*; (d_x, d_y))$, we define $\varphi(x, t) := \mathbb{P}\left[f(x, \zeta) \geq t\right]$ and recast the finite difference

$$\begin{aligned}
&\phi(x^* + sd_x, y^* + sd_y) - \phi(x^*, y^*) \\
&= \phi(x^* + sd_x, y^* + sd_y) - \phi(x^* + sd_x, y^*) + \phi(x^* + sd_x, y^*) - \phi(x^*, y^*) \\
&= \int_{y^*}^{y^*+sd_y} \Big(\varphi(x^* + sd_x, t) - (1-\epsilon)\Big)\, \mathrm{d}t + \Big(\phi(x^* + sd_x, y^*) - \phi(x^*, y^*)\Big).
\end{aligned} \tag{18}$$

For the second term in (18), we have

$$\lim_{s \to 0^+} \frac{1}{s}\Big[\phi(x^* + sd_x, y^*) - \phi(x^*, y^*)\Big] = \phi'(x^*, y^*; (d_x, 0)) \leq 0$$

because $\phi(x^* + sd_x, y^*) \leq \phi(x^*, y^*)$ for all sufficiently small $s > 0$. In what follows, we address the first term in (18). To that end, we notice that $\varphi(x, t)$ is log-concave on

$$\mathbf{dom}\,\varphi := \{\, (x, t) \in \mathbb{R} \times \mathbb{R}_+ \colon \exists\, \zeta \text{ such that } f(x, \zeta) - t \geq 0 \,\},$$

and $(x^*, y^*) \in \text{int}(\mathbf{dom}\,\varphi)$ because

$$\mathbb{P}\left[f(x^*, \zeta) - y^* > 0\right] = \mathbb{P}\left[f(x^*, \zeta) - y^* \geq 0\right] \geq 1 - \epsilon,$$

which implies that there exists a $\widehat{\zeta} \in \Xi$ such that $f(x^*, \widehat{\zeta}) - y^* > 0$. By continuity of f, we also have $f(x', \widehat{\zeta}) - y' \geq 0$ for all (x', y') sufficiently close to (x^*, y^*). Since $\varphi(x^*, y^*)$ is strictly positive and $\ln \varphi(x, t)$ is concave on $\mathbf{dom}\,\varphi$, $\ln \varphi(x, t)$ is locally Lipschitz at (x^*, y^*), i.e., there exist $M > 0$ and $r > 0$ such that

$$\Big|\ln \varphi(x, t) - \ln \varphi(x^*, y^*)\Big| \leq M \|(x - x^*, t - y^*)\|_2 \qquad \forall (x, t) \in \mathcal{B}\big((x^*, y^*), r\big),$$

where $\mathcal{B}((x^*, y^*), r)$ denotes a Euclidean ball centered around (x^*, y^*) with radius r. For all $s > 0$ sufficiently small such that $s \cdot \|(d_x, d_y)\|_2 \leq r/2$ and all scalar t such that $|t - y^*| < s|d_y|$, we have

$$\begin{aligned}
\big|\ln \varphi(x^* + sd_x, t) - \ln \varphi(x^*, t)\big| &\leq \big|\ln \varphi(x^* + sd_x, t) - \ln \varphi(x^*, y^*)\big| + \big|\ln \varphi(x^*, y^*) - \ln \varphi(x^*, t)\big| \\
&\leq M\|(sd_x, t - y^*)\|_2 + M\|(0, t - y^*)\|_2 \\
&\leq M\|(sd_x, sd_y)\|_2 + M\|(0, sd_y)\|_2 \\
&\leq 2sM\|(d_x, d_y)\|_2,
\end{aligned}$$

where the first inequality is because of the triangle inequality, the second inequality is because $\ln \varphi_0$ is locally Lipschitz around (x^*, y^*), and the third inequality is because

$$\|(sd_x, t - y^*)\|_2^2 = \|sd_x\|_2^2 + |t - y^*|^2 < \|sd_x\|_2^2 + |sd_y|^2 = \|(sd_x, sd_y)\|_2^2.$$

We bound the first term in (18) by discussing the following two cases. First, if $d_y > 0$, then it holds that

$$\begin{aligned}
&\int_{y^*}^{y^*+sd_y} \Big(\varphi(x^* + sd_x, t) - (1-\epsilon)\Big)\,\mathrm{d}t = \int_{y^*}^{y^*+sd_y} \Big(\exp\big[\ln\varphi(x^* + sd_x, t)\big] - (1-\epsilon)\Big)\,\mathrm{d}t \\
\le\ & \int_{y^*}^{y^*+sd_y} \Big(\exp\Big[\ln\varphi(x^*, t) + 2sM\|(d_x, d_y)\|_2\Big] - (1-\epsilon)\Big)\,\mathrm{d}t \\
=\ & \exp\big[2sM\|(d_x, d_y)\|_2\big] \int_{y^*}^{y^*+sd_y} \Big(\varphi(x^*, t) - (1-\epsilon)\exp\left[-2sM\|(d_x, d_y)\|_2\right]\Big)\,\mathrm{d}t \\
=\ & \exp\big[2sM\|(d_x, d_y)\|_2\big] \left(\int_{y^*}^{y^*+sd_y} \big[\varphi(x^*, t) - (1-\epsilon)\big]\,\mathrm{d}t + (1-\epsilon)\big(1 - \exp\left[-2sM\|(d_x, d_y)\|_2\right]\big) sd_y\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\lim_{s\to 0^+} \frac{1}{s}\int_{y^*}^{y^*+sd_y} \big(\varphi(x^* + sd_x, t) - (1-\epsilon)\big)\,\mathrm{d}t \\
\le\ & \left(\lim_{s\to 0^+} \exp\big[2sM\|(d_x, d_y)\|_2\big]\right) \cdot \left(\phi'(x^*, y^*; (0, d_y)) + \lim_{s\to 0^+} \frac{1}{s}(1-\epsilon)\big(1 - \exp\left[-2sM\|(d_x, d_y)\|_2\right]\big) sd_y\right) \\
=\ & \phi'(x^*, y^*; (0, d_y)),
\end{aligned}$$

where the inequality is because

$$\lim_{s\to 0^+} \frac{1}{s}\int_{y^*}^{y^*+sd_y} \big[\varphi(x^*, t) - (1-\epsilon)\big]\,\mathrm{d}t = \lim_{s\to 0^+} \frac{1}{s}\big[\phi(x^*, y^* + sd_y) - \phi(x^*, y^*)\big] = \phi'(x^*, y^*; (0, d_y)).$$

Second, if $d_y < 0$, then it holds that

$$\begin{aligned}
&\int_{y^*}^{y^*+sd_y} \big(\varphi(x^* + sd_x, t) - (1-\epsilon)\big)\,\mathrm{d}t = \int_{y^*+sd_y}^{y^*} \big(-\exp\left[\ln\varphi(x^* + sd_x, t)\right] + (1-\epsilon)\big)\,\mathrm{d}t \\
\le\ & \int_{y^*+sd_y}^{y^*} \Big(-\exp\Big[\ln\varphi(x^*, t) - 2sM\|(d_x, d_y)\|_2\Big] + (1-\epsilon)\Big)\,\mathrm{d}t \\
=\ & \exp\big[-2sM\|(d_x, d_y)\|_2\big] \int_{y^*+sd_y}^{y^*} \Big(-\varphi(x^*, t) + (1-\epsilon)\exp\left[2sM\|(d_x, d_y)\|_2\right]\Big)\,\mathrm{d}t \\
=\ & \exp\big[-2sM\|(d_x, d_y)\|_2\big] \int_{y^*}^{y^*+sd_y} \Big(\varphi(x^*, t) - (1-\epsilon)\exp\left[2sM\|(d_x, d_y)\|_2\right]\Big)\,\mathrm{d}t \\
=\ & \exp\big[-2sM\|(d_x, d_y)\|_2\big] \left(\int_{y^*}^{y^*+sd_y} \left[\varphi(x^*, t) - (1-\epsilon)\right]\mathrm{d}t + (1-\epsilon)\left(1 - \exp\left[2sM\|(d_x, d_y)\|_2\right]\right) sd_y\right),
\end{aligned}$$

where the inequality is because $\ln\varphi(x^* + sd_x, t) \ge \ln\varphi(x^*, t) - 2sM\|(d_x, d_y)\|_2$ and that the function

$-\exp(\cdot)$ is monotonically decreasing. It follows that

$$
\begin{aligned}
&\lim_{s\to 0^+} \frac{1}{s}\int_{y^*}^{y^*+sd_y} \big(\varphi(x^*+sd_x,t)-(1-\epsilon)\big)\,\mathrm{d}t \\
&\le \Big(\lim_{s\to 0^+} \exp\big[-2sM\|(d_x,d_y)\|_2\big]\Big)\cdot\Big(\phi'(x^*,y^*;(0,d_y)) + \lim_{s\to 0^+}\frac{1}{s}(1-\epsilon)\big(1-\exp\big[2sM\|(d_x,d_y)\|_2\big]\big)sd_y\Big) \\
&= \phi'(x^*,y^*;(0,d_y)).
\end{aligned}
$$

Finally, applying the above analysis on both terms in (18) yields

$$
\begin{aligned}
\phi'(x^*,y^*;(d_x,d_y)) &= \lim_{s\to 0^+}\frac{1}{s}\Big[\phi(x^*+sd_x,y^*+sd_y)-\phi(x^*,y^*)\Big] \\
&\le \phi'\big(x^*,y^*;(0,d_y)\big)+\phi'\big(x^*,y^*;(d_x,0)\big)\le 0,
\end{aligned}
$$

which completes the proof. □

Lemma 10. For all $(x,y)\in \mathbf{dom}\,\phi$ with $\phi(x,y)>0$, the directional derivative $\phi'(x,y;\Delta)$ at (x,y) along direction Δ is continuous and positively homogeneous in Δ.

Proof. For notation brevity, we denote $z=(x,y)$. Then, it holds that

$$
\begin{aligned}
\phi'(z;\Delta) &= \lim_{s\to 0^+}\frac{1}{s}\big[\phi(z+s\Delta)-\phi(z)\big] \\
&= \lim_{s\to 0^+}\left\{\frac{\exp(\ln\phi(z+s\Delta))-\exp(\ln\phi(z))}{\ln\phi(z+s\Delta)-\ln\phi(z)}\cdot\frac{\ln\phi(z+s\Delta)-\ln\phi(z)}{s}\right\} \\
&= \phi(z)\lim_{s\to 0^+}\frac{\ln\phi(z+s\Delta)-\ln\phi(z)}{s} = \phi(z)\cdot(\ln\phi)'(z;\Delta),
\end{aligned}
$$

where the third equality follows from the L'Hôpital's rule. Since $(\ln\phi)'(z;\Delta)$ is convex and positively homogeneous in Δ by Proposition 17.2 in [BC+11], so is $\phi'(z;\Delta)$. The continuity of $\phi'(z;\Delta)$ follows from its convexity, which completes the proof. □

Appendix I Proof of Theorem 10

Proof. By Theorem 1 in [GK16], (**O-CC**) is equivalent to inequality

$$
\sup_{\mathbb{Q}\in\mathcal{P}}\mathbb{Q}[A(x)\xi\le b(x)] \equiv \min_{\lambda\ge 0}\left\{\lambda\delta-\mathbb{E}_{\mathbb{P}}\left[\inf_{\xi\in\Xi}\{\lambda\|\zeta-\xi\|-\mathbb{1}\{\xi\in\mathcal{S}(x)\}\}\right]\right\}\ge 1-\epsilon.
$$

Noting that for any fixed $x\in\mathbb{R}^n$ and $\zeta\in\Xi$

$$
\begin{aligned}
\inf_{\xi\in\Xi}\{\lambda\|\zeta-\xi\|-\mathbb{1}\{\xi\in\mathcal{S}(x)\}\} &= \begin{cases} -1 & \text{if } \zeta\in\mathcal{S}(x) \\ \min\{\lambda\cdot\mathbf{d}(\zeta,\mathcal{S}(x))-1,0\} & \text{if } \zeta\notin\mathcal{S}(x)\end{cases} \\
&= \min\{\lambda\cdot\mathbf{d}(\zeta,\mathcal{S}(x))-1,0\},
\end{aligned}
$$

we recast $\mathcal{X}^o$ as

$$
\lambda\delta+\mathbb{E}_{\mathbb{P}}\Big[\max\{1-\lambda\cdot\mathbf{d}(\zeta,\mathcal{S}(x)),0\}\Big]\ge 1-\epsilon \qquad \forall\lambda\ge 0.
$$

We notice that the above inequality automatically holds when $\lambda = 0$ because, in this case, the LHS equals one. Hence, we can drop this case and assume that $\lambda > 0$. Then, we divide both sides by λ and denote $\gamma = 1/\lambda$ to obtain

$$\delta + \mathbb{E}_{\mathbb{P}}\left[(\gamma - \mathbf{d}\left(\zeta, \mathcal{S}(x)\right), 0)^{+}\right] \geq (1-\epsilon)\gamma \qquad \forall \gamma \geq 0.$$

We notice that the above inequality holds for all $\gamma < 0$ because, in that case, the LHS is positive and the RHS is negative. Hence, we expand the domain of γ to be the whole real line and finish the proof as follows:

$$\begin{aligned}
& (-\gamma) + \frac{1}{1-\epsilon}\mathbb{E}_{\mathbb{P}}\left[(-\mathbf{d}\left(\zeta, \mathcal{S}(x)\right) - (-\gamma), 0)^{+}\right] \geq -\frac{\delta}{1-\epsilon} \qquad \forall \gamma \in \mathbb{R} \\
\iff\ & \inf_{-\gamma \in \mathbb{R}} \left\{ (-\gamma) + \frac{1}{1-\epsilon}\mathbb{E}_{\mathbb{P}}\left[(-\mathbf{d}\left(\zeta, \mathcal{S}(x)\right) - (-\gamma), 0)^{+}\right] \right\} \geq \frac{-\delta}{1-\epsilon} \\
\iff\ & \mathrm{CVaR}_{\epsilon}\Big(- \mathbf{d}\left(\zeta, \mathcal{S}(x)\right) \Big) + \frac{\delta}{1-\epsilon} \geq 0.
\end{aligned}$$

□

Appendix J Proof of Lemma 7

Proof. Since $\mathbf{d}\left(\zeta, \mathcal{S}(x)\right)$ is defined through a convex program, in which the Slater's condition holds, we take the dual to obtain

$$\mathbf{d}\left(\zeta, \mathcal{S}(x)\right) = \max_{\lambda \leq 0} \left\{ \lambda^{\mathsf{T}}[b(x) - A\zeta] \colon \|A^{\mathsf{T}}\lambda\|_{*} \leq 1 \right\}.$$

This completes the proof. □

References

[Aus76] A. Auslender. "Optimisation". In: *Méthodes numériques* (1976) (page 13).

[BBV04] S. Boyd, S. P. Boyd, and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004 (pages 23, 24).

[BC+11] H. H. Bauschke, P. L. Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*. Vol. 408. Springer, 2011 (page 30).

[Bec15] A. Beck. "On the Convergence of Alternating Minimization for Convex Programming With Applications To Iteratively Reweighted Least Squares and Decomposition Schemes". In: *SIAM Journal on Optimization* 25.1 (Jan. 2015), pp. 185–209. ISSN: 1095-7189 (pages 13, 15).

[Ber97] D. P. Bertsekas. "Nonlinear Programming". In: *Journal of the Operational Research Society* 48.3 (1997), pp. 334–334 (page 13).

[BR02] P. Beraldi and A. Ruszczyński. "The Probabilistic Set-Covering Problem". In: *Operations Research* 50.6 (2002), pp. 956–967 (page 2).

[BT13] A. Beck and L. Tetruashvili. "On the Convergence of Block Coordinate Descent Type Methods". In: *SIAM Journal on Optimization* 23.4 (Jan. 2013), pp. 2037–2060. ISSN: 1095-7189 (page 13).

[Bul13] P. S. Bullen. *Handbook of means and their inequalities*. Vol. 560. Springer Science & Business Media, 2013 (page 6).

[CC59] A. Charnes and W. W. Cooper. "Chance-Constrained Programming". In: *Management Science* 6.1 (1959), pp. 73–79 (page 1).

[CCS58] A. Charnes, W. W. Cooper, and G. H. Symonds. "Cost Horizons and Certainty Equivalents: an Approach To Stochastic Programming of Heating Oil". In: *Management Science* 4.3 (1958), pp. 235–263. ISSN: 00251909, 15265501 (page 1).

[CE06] G. C. Calafiore and L. El Ghaoui. "On Distributionally Robust Chance-Constrained Linear Programs". In: *Journal of Optimization Theory and Applications* 130.1 (2006), pp. 1–22 (pages 2, 3, 10).

[CKW18] Z. Chen, D. Kuhn, and W. Wiesemann. "Data-Driven Chance Constrained Programs Over Wasserstein Balls". In: *arXiv preprint arXiv:1809.00210* (2018) (pages 1, 3, 6, 7, 16).

[DS16] Y. Deng and S. Shen. "Decomposition Algorithms for Optimizing Multi-Server Appointment Scheduling With Chance Constraints". In: *Mathematical Programming* 157.1 (2016), pp. 245–276 (page 1).

[EOO03] L. El Ghaoui, M. Oks, and F. Oustry. "Worst-Case Value-At-Risk and Robust Portfolio Optimization: a Conic Programming Approach". In: *Operations Research* 51.4 (2003), pp. 543–556 (page 3).

[Erd45] P. Erdös. "Some Remarks on the Measurability of Certain Sets". In: *Bulletin of the American Mathematical Society* 51.10 (1945), pp. 728–731 (page 24).

[GK16] R. Gao and A. J. Kleywegt. "Distributionally Robust Stochastic Optimization With Wasserstein Distance". In: (2016). arXiv: `1604.02199 [math.OC]` (page 30).

[GS99] L. Grippof and M. Sciandrone. "Globally Convergent Block-Coordinate Techniques for Unconstrained Optimization". In: *Optimization Methods and Software* 10.4 (1999), pp. 587–637 (page 13).

[Gup80] S. D. Gupta. "Brunn-Minkowski Inequality and Its Aftermath". In: *Journal of Multivariate Analysis* 10.3 (1980), pp. 296–318. ISSN: 0047-259X (page 5).

[Han+15] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. "A Distributionally Robust Perspective on Uncertainty Quantification and Chance Constrained Programming". In: *Mathematical Programming* 151.1 (2015), pp. 35–62 (page 3).

[Han+17] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. "Ambiguous Joint Chance Constraints Under Mean and Dispersion Information". In: *Operations Research* 65.3 (2017), pp. 751–767 (page 3).

[JL20] R. Ji and M. A. Lejeune. "Data-Driven Distributionally Robust Chance-Constrained Optimization With Wasserstein Metric". In: *Journal of Global Optimization* 79.4 (Nov. 2020), pp. 779–811. ISSN: 1573-2916 (page 3).

[KJ21] S. Küçükyavuz and R. Jiang. "Chance-Constrained Optimization: A Review of Mixed-Integer Conic Formulations and Applications". In: *arXiv preprint arXiv:2101.08746* (2021) (page 2).

[Küç12] S. Küçükyavuz. "On Mixing Sets Arising in Chance-Constrained Programming". In: *Mathematical Programming* 132.1-2 (2012), pp. 31–56 (page 2).

[LA08] J. Luedtke and S. Ahmed. "A Sample Approximation Approach for Optimization With Probabilistic Constraints". In: *SIAM Journal on Optimization* 19.2 (2008), pp. 674–699 (page 2).

[LAN08] J. Luedtke, S. Ahmed, and G. L. Nemhauser. "An Integer Programming Approach for Linear Programs With Probabilistic Constraints". In: *Mathematical Programming* 122.2 (Oct. 2008), pp. 247–272. ISSN: 1436-4646 (page 2).

[Li95] S. Li. "An Insurance and Investment Portfolio Model Using Chance Constrained Programming". In: *Omega* 23.5 (1995), pp. 577–585 (page 1).

[LJM19] B. Li, R. Jiang, and J. L. Mathieu. "Ambiguous Risk Constraints With Moment and Unimodality Information". In: *Mathematical Programming* 173.1-2 (2019), pp. 151–192 (page 3).

[LLS01] C. Lagoa, X. Li, and M. Sznaier. "Application of probabilistically constrained linear programs to risk-adjusted controller design". In: *Proceedings of the 2001 American Control Conference. (Cat. No.01CH37148)*. Vol. 2. 2001, 738–743 vol.2 (page 2).

[LT93] Z.-Q. Luo and P. Tseng. "Error Bounds and Convergence Analysis of Feasible Descent Methods: a General Approach". In: *Annals of Operations Research* 46.1 (1993), pp. 157–178 (page 13).

[Lue14] J. Luedtke. "A Branch-And-Cut Decomposition Algorithm for Solving Chance-Constrained Mathematical Programs With Finite Support". In: *Mathematical Programming* 146.1-2 (2014), pp. 219–244 (page 2).

[MG06] P. A. Miranda and R. A. Garrido. "A Simultaneous Inventory Control and Facility Location Model With Stochastic Capacity Constraints". In: *Networks and Spatial Economics* 6.1 (2006), pp. 39–53 (page 1).

[MK18] P. Mohajerin Esfahani and D. Kuhn. "Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations". In: *Mathematical Programming* 171.1 (Sept. 2018), pp. 115–166. ISSN: 1436-4646 (page 3).

[MW65] B. L. Miller and H. M. Wagner. "Chance Constrained Programming With Joint Constraints". In: *Operations Research* 13.6 (1965), pp. 930–945 (page 1).

[Nor93] V. I. Norkin. "The Analysis and Optimization of Probability Functions". In: (1993) (pages 14, 18, 22, 24, 28).

[PP63] C. van de Panne and W. Popp. "Minimum-Cost Cattle Feed Under Probabilistic Protein Constraints". In: *Management Science* 9.3 (Apr. 1963), pp. 405–430. ISSN: 1526-5501 (page 2).

[Pré13] A. Prékopa. *Stochastic Programming*. Vol. 324. Springer Science & Business Media, 2013 (pages 2, 10, 14).

[Pré70] A. Prékopa. "On probabilistic constrained programming". In: *Proceedings of the Princeton Symposium on Mathematical Programming*. Vol. 113. Princeton, NJ. 1970, p. 138 (page 1).

[Pré90] A. Prékopa. "Dual Method for the Solution of a One-Stage Stochastic Programming Problem With Random RHS Obeying a Discrete Probability Distribution". In: *ZOR Zeitschrift für Operations Research Methods and Models of Operations Research* 34.6 (Nov. 1990), pp. 441–461. ISSN: 1432-5217 (page 2).

[RU99] R. T. Rockafellar and S. Uryasev. "Optimization of Conditional Value-At-Risk". In: *Journal of Risk* (1999) (page 6).

[SDR09] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009 (pages 2, 4, 5, 10).

[SG83] W. R. Stewart Jr and B. L. Golden. "Stochastic Vehicle Routing: a Comprehensive Approach". In: *European Journal of Operational Research* 14.4 (1983), pp. 371–385 (page 1).

[SW14] A. Saumard and J. A. Wellner. "Log-Concavity and Strong Log-Concavity: a Review". In: *Statistics Surveys* 8 (2014), p. 45 (page 19).

[WGW11] Q. Wang, Y. Guan, and J. Wang. "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output". In: *IEEE Transactions on Power Systems* 27.1 (2011), pp. 206–215 (page 1).

[XA16] W. Xie and S. Ahmed. "On Deterministic Reformulations of Distributionally Robust Joint Chance Constrained Optimization Problems". In: *SIAM Journal on Optimization* 28 (Sept. 2016) (page 3).

[XA20] W. Xie and S. Ahmed. "Bicriteria Approximation of Chance-Constrained Covering Problems". In: *Operations Research* (Feb. 2020). ISSN: 1526-5463 (page 3).

[XCM12] H. Xu, C. Caramanis, and S. Mannor. "Optimization Under Probabilistic Envelope Constraints". In: *Operations Research* 60.3 (2012), pp. 682–699 (page 21).

[Xie19] W. Xie. "On Distributionally Robust Chance Constrained Programs With Wasserstein Distance". In: *Mathematical Programming* (2019) (pages 1, 3, 6, 7, 15).

[ZKR11] S. Zymler, D. Kuhn, and B. Rustem. "Distributionally Robust Joint Chance Constraints With Second-Order Moment Information". In: *Mathematical Programming* 137.1-2 (2011), pp. 167–198 (page 3).