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 $\alpha$-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 $\xi$ is subject to uncertainty and follows a probability distribution $P_{\text{true}}$, a convenient way of protecting these safety principles is to use chance constraint

$$P_{\text{true}}\left[A(x)\xi \leq b(x)\right] \geq 1 - \epsilon,$$

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 \( \xi_i \). Then, chance constraint

\[
P_{\text{true}} \left[ x^T \xi \geq \eta \right] \geq 1 - \epsilon \quad \text{(PTO)}
\]

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

\[
P_{\text{true}} [Tx \geq \xi] \geq 1 - \epsilon \quad \text{(PP)}
\]

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

In (PP), the random vector \( \xi \) 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, \( \xi \) 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 [Pre90, BR02], and derived valid inequalities to strengthen the ensuing MIP formulations (see, e.g., [Kuc12, 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 \( \xi \) follows a Gaussian distribution, and [LLS01] and [CE06] further extended this result when \( \xi \) follows an elliptical log-concave distribution (see Definition 2). For joint (CC) with RHS uncertainty, [Pre13] (see his Theorem 10.2) proved the convexity of the ensuing feasible region when \( \xi \) 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 \( \alpha \)-concave distributions (see Definition 2).

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

\[
\mathcal{P} := \{ Q \in \mathcal{P}_0 : d_W(Q, P) \leq \delta \},
\]
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

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

(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 $\delta$ is large enough. As a result, the pessimistic counterpart

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

(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):

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

(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 $\xi$. 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 $\xi$. 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 $\xi$, 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 $\xi$ nonlinearly and $\mathcal{P}$ is characterized by a single moment constraint of $\xi$. 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 $\xi$. 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 $\Sigma > 0$ 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 $\alpha$-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 $\alpha$-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 $\mathbb{R}^n$. For a given decision $x$, we denote by $S(x)$ the event $\{\xi: A(x)\xi \leq b(x)\}$ and by $S_c(x)$ 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, $\text{Leb}(\cdot)$ 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 $\alpha$-concave with $\alpha \in \mathbb{R}$ if for all $x, y \in \text{dom} f$ and $\theta \in [0, 1]$

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

where $m_\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is defined as

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

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 \( m_\alpha \) implies that if \( f \) is \( \alpha \)-concave, then it is \( \beta \)-concave for all \( \beta \leq \alpha \). Under certain conditions, summation preserves \( \alpha \)-concavity.

**Theorem 1** (Theorem 4.19 in [SDR09]). If the function \( f : \mathbb{R}^n \to \mathbb{R}_+ \) is \( \alpha \)-concave and the function \( g : \mathbb{R}^n \to \mathbb{R}_+ \) is \( \beta \)-concave, where \( \alpha, \beta \geq 1 \), 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
\[
\theta A + (1 - \theta)B := \{ \theta x + (1 - \theta) y : x \in A, y \in B \}.
\]

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

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

Next, we review the relationship between \( \alpha \)-concave probability measures and their densities.

**Theorem 2** (Theorem 4.15 in [SDR09]). Let \( \Omega \) be a convex subset of \( \mathbb{R}^n \) and \( s \) be the dimension of the smallest affine subspace \( \mathcal{H}(\Omega) \) containing \( \Omega \). The probability measure \( P \) is \( \alpha \)-concave with \( \alpha \leq 1/s \) if and only if its probability density function (PDF) with respect to the Lebesgue measure on \( \mathcal{H} \) is \( \alpha' \)-concave with
\[
\alpha' := \begin{cases} 
\alpha/(1 - s) & \text{if } \alpha \in (-\infty, 1/s), \\
-1/s & \text{if } \alpha = -\infty, \\
+\infty & \text{if } \alpha = 1/s.
\end{cases}
\]

**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 \( \mu \) and \( \Sigma \) 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
\[
f(x) = \frac{1}{\text{vol}(\Omega)} \mathbb{1} \{ x \in \Omega \},
\]
where \( \text{vol}(\Omega) \) represents the volume of \( \Omega \). \( f \) is \( +\infty \)-concave on \( \Omega \). 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
\[
f(x) \geq m_\alpha[f_0(x_0), f_1(x_1); \theta],
\]
whenever \( x = (1 - \theta)x_0 + \theta x_1 \) with \( x_0 \in S_0, x_1 \in S_1; -1/n \leq \alpha \leq +\infty \). Then
\[
\int_{(1-\theta)S_0+\theta S_1} f(x) \, dx \geq m_{\alpha_n^*} \left[ \int_{S_0} f_0(x) \, dx, \int_{S_1} f_1(x) \, dx; \theta \right],
\]
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:
\[
\left( \sum_{i=1}^{n} (a_i + b_i)^p \right)^{1/p} \leq \left( \sum_{i=1}^{n} a_i^p \right)^{1/p} + \left( \sum_{i=1}^{n} b_i^p \right)^{1/p}.
\]
If \( p < 1 \) and \( p \neq 0 \), then the inequality holds with the inequality sign reversed.

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

**Proof.** See Appendix A.

\[ \Box \]

### 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 \( \alpha \)-concave reference distribution \( \mathbb{P} \). The new reformulation leads to convexity proofs for individual (\( \mathbb{P}-\text{CC} \)) with LHS uncertainty and joint (\( \mathbb{P}-\text{CC} \)) 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
\[
\text{VaR}_{1-\epsilon}(X) := \inf \{ x : \mathbb{P}_X[X \leq x] \geq 1 - \epsilon \},
\]
and its \((1 - \epsilon)\)-CVaR is defined through
\[
\text{CVaR}_{1-\epsilon}(X) := \min_{\gamma} \left\{ \gamma + \frac{1}{\epsilon} \mathbb{E} \left[ (X - \gamma)^+ \right] \right\}.
\]

**Theorem 5** (Adapted from Theorem 1 in [Xie19]). For \( \delta > 0 \), it holds that
\[
\mathcal{X}^p = \left\{ x \in \mathbb{R}^n : \frac{\delta}{\epsilon} + \text{CVaR}_{1-\epsilon}\left( -\mathbf{d}(\zeta, \mathcal{S}^c(x)) \right) \leq 0 \right\}.
\]

Here, random variable \( \zeta \) follows the reference distribution \( \mathbb{P} \) and \( \mathbf{d}(\zeta, \mathcal{S}^c(x)) \) represents the distance from \( \zeta \) to the “unsafe” set \( \mathcal{S}^c(x) \) [CKW18],
\[
\mathbf{d}(\zeta, \mathcal{S}^c(x)) := \inf_{\xi \in \Xi} \{ \| \zeta - \xi \| : A(x)\xi \not\in b(x) \},
\]
and \( \Xi \) is the support of \( \xi \).
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 \( a_i(x)^T \) 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 \( f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \)
\[
f(x, \zeta) := \min_{i \in [m] \setminus I(x)} \left\{ \frac{b_i(x) - a_i(x)^T \zeta}{\|a_i(x)\|_\star} \right\},
\]
where \( I(x) := \{ i \in [m] : a_i(x) = 0 \} \). Then, it follows from [Xie19; CKW18] that
\[
d(\zeta, \mathcal{S}^c(x)) = \left( f(x, \zeta) \right)^+.
\]
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 \(-d(\zeta, \mathcal{S}^c(x))\) in (2).

**Lemma 3.** Let \( X \) be a random variable, then
\[
\text{CVaR}_{1-\epsilon}(X^-) = 1 \{ 0 \geq \text{VaR}_{1-\epsilon}(X) \} \cdot \left[ \text{CVaR}_{1-\epsilon}(X) - \frac{1}{\epsilon} \mathbb{E}[X^+] \right].
\]

**Proof.** See Appendix [B] \( \square \)

Combining Theorem [B] and Lemma [B] leads to the following reformulation of \( \mathcal{X}^p \).

**Corollary 1.** For \( \delta > 0 \), it holds that
\[
\mathcal{X}^p = \left\{ x \in \mathbb{R}^n : \frac{\delta}{\epsilon} + \text{CVaR}_{1-\epsilon}(-f(x, \zeta)) \leq \frac{1}{\epsilon} \mathbb{E} [(-f(x, \zeta))^+] \right\} \quad (3)
\]
\[
0 \geq \text{VaR}_{1-\epsilon}(-f(x, \zeta)) \quad (4)
\]

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

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

**Proof.** See Appendix [C] \( \square \)

**Proposition 1.** Suppose that \( \mathbb{P} \) is \( \alpha \)-concave. Then, for \( \delta > 0 \), it holds that
\[
\mathcal{X}^p = \left\{ x \in \mathbb{R}^n : \delta \leq \int_0^{\text{VaR}_{\epsilon}(f(x, \zeta))} \left( \mathbb{P}[f(x, \zeta) \geq t] - (1 - \epsilon) \right) dt \right\} \quad (5)
\]
\[
\mathbb{P}[A(x)\zeta \leq b(x)] \geq 1 - \epsilon \quad (6)
\]

**Proof.** First, moving the CVaR term to the RHS of (3) yields
\[
\delta \leq \mathbb{E} \left[ f(x, \zeta) \cdot 1 \{ -f(x, \zeta) \geq \text{VaR}_{1-\epsilon}(-f(x, \zeta)) \} \right] - \mathbb{E} \left[ f(x, \zeta) \cdot 1 \{ -f(x, \zeta) \geq 0 \} \right]
\]
\[
= \mathbb{E} \left[ f(x, \zeta) \cdot 1 \{ \text{VaR}_{1-\epsilon}(-f(x, \zeta)) \leq -f(x, \zeta) \leq 0 \} \right]
\]
\[
= \mathbb{E} \left[ f(x, \zeta) \cdot 1 \{ 0 \leq f(x, \zeta) \leq \text{VaR}_{\epsilon}(f(x, \zeta)) \} \right],
\]
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
\[
\mathbb{E} \left[ f(x, \zeta) \cdot 1 \{ 0 \leq f(x, \zeta) \leq \text{VaR}_\epsilon(f(x, \zeta)) \} \right]
\]
\[
= \int_\Xi f(x, \zeta) \cdot 1 \{ 0 \leq f(x, \zeta) \leq \text{VaR}_\epsilon(f(x, \zeta)) \} \, d\mathbb{P}(\zeta)
\]
\[
= \int \int \mathbb{E} \left[ 1 \{ t \leq f(x, \zeta) \} \cdot 1 \{ 0 \leq f(x, \zeta) \leq \text{VaR}_\epsilon(f(x, \zeta)) \} \right] \, dt \, d\mathbb{P}(\zeta)
\]
\[
= \int \int \mathbb{P} \left[ t \leq f(x, \zeta) \leq \text{VaR}_\epsilon(f(x, \zeta)) \right] \, dt \, d\mathbb{P}(\zeta)
\]
\[
= \int_0^{\text{VaR}_\epsilon(f(x, \zeta))} \left( \mathbb{P} \left[ f(x, \zeta) \geq t \right] - (1-\epsilon) \right) \, dt,
\]
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 \( \zeta \in \Xi \),
\[
1 \{ t \leq f(x, \zeta) \} \cdot 1 \{ 0 \leq f(x, \zeta) \leq \text{VaR}_\epsilon(f(x, \zeta)) \} = 1 \{ t \leq f(x, \zeta) \leq \text{VaR}_\epsilon(f(x, \zeta)) \}
\]
holds Lebesgue-almost everywhere for \( t \in \mathbb{R}_+ \). We discuss the following three cases:

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

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

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

The last equality is because
\[
\mathbb{P} \left[ t \leq f(x, \zeta) \leq \text{VaR}_\epsilon(f(x, \zeta)) \right] = \mathbb{P} \left[ t \leq f(x, \zeta) \right] - \mathbb{P} \left[ t \geq \text{VaR}_\epsilon(f(x, \zeta)) \right]
\]
when \( t \in [0, \text{VaR}_\epsilon(f(x, \zeta))] \). This recasts (8) into (5).

Second, constraint (4) is equivalent to \( \mathbb{P}[f(x, \zeta) \geq 0] \geq 1 - \epsilon \) by definition of \( \text{VaR} \), which can be further recast as
\[
\mathbb{P} \left[ a_i(x)^T \zeta \leq b_i(x), \; \forall i \in [m] \setminus I(x) \right] \geq 1 - \epsilon
\]
by definition of \( f(x, \zeta) \). For all \( x \in \mathcal{X}^p \) and \( i \in [m] \), we assume without loss of generality that \( b_i(x) \geq 0 \) whenever \( a_i(x) = 0 \) (because otherwise \( \mathbb{P}[A(x) \zeta \leq b(x)] = 0 \)), and it holds that \( a_i(x)^T \zeta \leq b_i(x) \) for all \( i \in I(x) \). It follows that (4) is equivalent to (6), which completes the proof. \( \square \)

Remark 1. We notice that constraint (6) is simply \( \text{(CC)} \) 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 \( \text{(CC)} \) can be violated. Constraint (5) makes sure that this budget is beyond the radius of \( \mathcal{P} \), i.e., \( \text{(CC)} \) 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

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

We provide sufficient conditions, under which $\mathcal{A}_L^P$ admits a convex and SOC representation.

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

$$
\mathcal{A}_L^P = \left\{ x \in \mathbb{R}^n : a(x)^T \mu + c_p \| \Sigma^{1/2} a(x) \|_2 \leq b(x) \right\},
$$

where

$$
c_p := \inf_{\epsilon' \in (0, \epsilon]} \frac{\delta + \epsilon \text{CVaR}_1(\epsilon) - \epsilon' \text{CVaR}_1(\epsilon' \cdot \epsilon)}{\epsilon - \epsilon'} \geq 0
$$

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

**Proof.** Pick any $x \in \mathcal{A}_L^P$. If $a(x) = 0$, then $\mathbb{Q}[a(x)^T \xi \leq b(x)] = 1$ when $b(x) \geq 0$ and $\mathbb{Q}[a(x)^T \xi \leq b(x)] = 0$ otherwise. It follows that $\mathcal{A}_L^P = \{ x \in \mathbb{R}^n : b(x) \geq 0 \}$, 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

$$
\text{(3)} \iff \frac{\delta}{\epsilon} + \text{CVaR}_{1-\epsilon}(-f(x, \zeta)) \leq \frac{1}{\epsilon} \mathbb{E} \left[ -f(x, \zeta) \cdot 1 \{ -f(x, \zeta) \geq 0 \} \right]
$$

$$
\iff \frac{\delta}{\epsilon} + \text{CVaR}_{1-\epsilon}(-f(x, \zeta)) \leq \frac{1}{\epsilon} \sup_{t \in \mathbb{R}} \mathbb{E} \left[ -f(x, \zeta) \cdot 1 \{ -f(x, \zeta) \geq t \} \right]
$$

$$
\iff \exists t \in \mathbb{R} : \delta + \epsilon \cdot \text{CVaR}_{1-\epsilon}(-f(x, \zeta)) \leq \mathbb{E} \left[ -f(x, \zeta) \cdot 1 \{ -f(x, \zeta) \geq t \} \right]
$$

$$
\iff \exists \epsilon_t > 0 : \delta + \epsilon \cdot \text{CVaR}_{1-\epsilon}(-f(x, \zeta)) \leq \epsilon_t \cdot \text{CVaR}_{1-\epsilon_t}(-f(x, \zeta))
$$

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

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

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

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

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

$$
\mathbb{E} \left[ \frac{a(x)^T (\zeta - \mu)(\zeta - \mu)^T a(x)}{\| a(x) \|_s^2} \right] = \frac{a(x)^T \Sigma a(x)}{a(x)^T \Sigma a(x)} = 1.
$$
It follows from translation invariance and the symmetry of \( Y \) that \( \text{CVaR}_{1-\epsilon}(-f(x, \xi)) = \text{CVaR}_{1-\epsilon}(Y) - m(x) \) and

\[
\exists \epsilon_t \in (0, \epsilon) : \delta + \epsilon \cdot \text{CVaR}_{1-\epsilon}(Y) - \epsilon_t \cdot \text{CVaR}_{1-\epsilon}(Y) \leq m(x)(\epsilon - \epsilon_t)
\]

\[
\iff a(x)^T \mu + c_p \|\Sigma^{1/2} a(x)\|_2 \leq b(x),
\]

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

\[
\epsilon \text{CVaR}_{1-\epsilon}(Y) - \epsilon' \text{CVaR}_{1-\epsilon'}(Y) = \mathbb{E}\left[ y \cdot 1\{ y \in [\text{Var}_{1-\epsilon}(Y), \text{Var}_{1-\epsilon'}(Y)] \} \right]
\]

\[
= \int_{\text{Var}_{1-\epsilon}(Y)} y \cdot dF_Y(y)
\]

\[
= \int_{1-\epsilon}^{1-\epsilon'} \text{VaR}_q(Y) \, dq \geq (\epsilon - \epsilon') \text{Var}_{1-\epsilon}(Y),
\]

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

\[
\frac{b(x) - a(x)^T \mu}{\|\Sigma^{1/2} a(x)\|_2} \geq \text{Var}_{1-\epsilon}(Y)
\]

\[
\iff \frac{b(x) - a(x)^T \mu}{\|\Sigma^{1/2} a(x)\|_2} \geq \text{VaR}_{1-\epsilon} \left( \frac{a(x)^T (\xi - \mu)}{\|\Sigma^{1/2} a(x)\|_2} \right) \implies (4),
\]

where the last implication is by translation invariance of \( \text{VaR} \). 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 \( \text{CVaR} \) of the normalized random variable \( a(x)^T (\xi - \mu)/\|a(x)\|_2 \) 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 \( A_L^p \) as the Wasserstein ball shrinks.

Remark 3. Since the Wasserstein ball shrinks to the reference distribution \( \mathbb{P} \) as \( \delta \) tends to zero, one would expect that \( A_L^p \) asymptotically recovers the feasible region of (CC) with respect to \( \mathbb{P} \sim \text{Gaussian}(\mu, \Sigma) \), which reads \( \{ x \in \mathbb{R}^n : a(x)^T \mu + \text{VaR}_{1-\epsilon}(Y)\|\Sigma^{1/2} a(x)\|_2 \leq b(x) \} \). This is indeed the case, since it can be shown that \( c_p \) decreases to \( \text{VaR}_{1-\epsilon}(Y) \) as \( \delta \) 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 \( \xi \) has an \( \alpha \)-concave distribution (particularly, \( \xi \) is log-concave when \( \alpha = 0 \) \[Pre13, SDR09\]).

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

\[
\mathcal{K} := \{ x \in \mathbb{R}^n : \mathbb{P}[h(x, \xi) \geq 0] \geq 1 - \epsilon \}. 
\]
In this subsection, we seek to extend this result to \((\text{P-CC})\).

**Theorem 8.** Suppose that the reference distribution \(P\) of \(P\) is \(\alpha\)-concave with \(\alpha \geq -1\). Then, for \(\delta > 0\) the set

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

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 \(a_i^T\) for all \(i \in [m]\), satisfies

1. \(a_i \neq 0\), because otherwise we can add a deterministic constraint \(b_i(x) \geq 0\) to \(\mathcal{X}_R^P\) and eliminate inequality \(i\) from \((\text{P-CC})\);
2. \(\|a_i\|_* = 1\), 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 \(d(\zeta, \mathcal{S}^c(x))\) to the unsafe set satisfies

\[
d(\zeta, \mathcal{S}^c(x)) = f(x, \zeta) + \psi(x, t) := \mathbb{P}[f(x, \zeta) \geq t] - (1 - \epsilon)
\]

with

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

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

**Lemma 5.** For all \(\epsilon \in (0, 1)\), if \(\zeta\) has an \(\alpha\)-concave distribution with \(\alpha \geq -1\), then \(\text{VaR}_{1-\epsilon}(f(x, \zeta))\) is concave in \(x\) on \(\mathbb{R}^n\).

**Proof.** See Appendix E.

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

\[
\psi(x, t) := \mathbb{P}[f(x, \zeta) \geq t] - (1 - \epsilon) \quad \text{and} \quad \phi(x, y) := \int_0^y \psi(x, t) \, dt
\]

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 \(\mathcal{X}_R^P\) as constraints (5)–(6). For ease of exposition, we denote by \(G(x)\) the RHS of (5).

Second, to show that \(\mathcal{X}_R^P\) 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 \(\text{VaR}_\epsilon(f(x, \zeta))\) is continuous in \(x\) due to its concavity. Then, by Lemma 6 the mapping

\[
x \mapsto \int_0^{\text{VaR}_\epsilon(f(x, \zeta))} \mathbb{P}[f(x, \zeta) \geq t] \, dt
\]

is continuous. It follows that \(G(x)\) is continuous and the feasible region of (5) is closed.
Third, to show that \( X^p \) is convex, it suffices to prove the convexity of the feasible region of (5). Because that of \( \mu \) follows from Theorem 7. To that end, by Theorem 7 and Lemma 2, \( \psi \) is \( \alpha \)-concave in \((x,t)\) on \( \text{dom}\psi:=\{(x,t)\colon \psi(x,t)\ge0\} = \{(x,t)\colon t \le \text{VaR}_\epsilon(f(x,\zeta))\} \), which is convex by Lemma 5. Then, for any \( x_0, x_1 \in X^p \) and any \( t_0 \in S_0 := [0, \text{VaR}_\epsilon(f(x_0,\zeta))] \), \( t_1 \in S_1 := [0, \text{VaR}_\epsilon(f(x_1,\zeta))] \) it holds that

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

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

\[
m_\alpha^* \left[ \int_{S_0} \psi(x_0, t) \, dt, \int_{S_1} \psi(x_1, t) \, dt; \frac{1}{2} \right] \le \int_{\frac{1}{2}S_0 + \frac{1}{2}S_1} \psi(x_{1/2}, t) \, dt \le \int_{S_{1/2}} \psi(x_{1/2}, t) \, dt
\]

where the first inequality is due to Theorem 3 and \( \alpha^*_1 \ge -\infty \) is a function of \( \alpha \) (see Theorem 3), and the second inequality is because \( \frac{1}{2}S_0 + \frac{1}{2}S_1 \subseteq S_{1/2} := [0, \text{VaR}_\epsilon(f(x_{1/2},\zeta))] \). In other words, we obtain that

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

and \( G(x) \) is midpoint \( \alpha^*_1 \)-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.

\textbf{Remark 4.} Although Theorem 7 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 \((x,\zeta)\) and \( f(x,\zeta) \) is atomless for all \( x \).

\section{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: \( \min_{x \in X} \left\{ c^T x : (P-CC) \right\} \), 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

\[
\min_{x \in X} c^T x 
\]

\( \text{s.t. } \delta \le \int_0^{\text{VaR}_\epsilon(f(x,\zeta))} \left( \mathbb{P} \left[ f(x,\zeta) \ge t \right] - (1-\epsilon) \right) \, dt, \tag{10b}
\]

\[
\text{VaR}_\epsilon \left( f(x,\zeta) \right) \ge 0, \tag{10c}
\]

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

\textbf{Proposition 2.} For \( y \ge 0 \), define

\[
\phi(x,y) := \int_0^y \left( \mathbb{P} \left[ f(x,\zeta) \ge t \right] - (1-\epsilon) \right) \, dt.
\]
If $P$ is $\alpha$-concave with $\alpha \geq -1$, then $\phi(x, y)$ is $\alpha^*_1$-concave on

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

where $\alpha^*_1$ is defined in Theorem 3. In addition, $\text{dom } \phi$ is closed and constraints (10a)–(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 $\min_{x \in X} \{ c^T x : (10d) \}$. To address the integral arising from the RHS of constraint (10d), we switch the objective function with the constraint to obtain

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

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 $\delta$ 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 5.2.

Evaluating $\rho(u)$ is equivalent to maximizing $\phi(x, y)$ over the intersection of $\{ (x, y) \in X \times \mathbb{R}_+ : c^T x \leq u \}$ and $\text{dom } \phi$. Unfortunately, projecting onto $\text{dom } \phi$ 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 $P[A \zeta \leq b(x)] \geq 1 - \epsilon$, i.e., when $x$ satisfies (CC), the maximization over $y$ admits a closed-form solution $y = \text{VaR}_\epsilon(f(x, \zeta))$, that is,

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

because $\phi(x, y)$ is increasing in $y$ on the interval $[0, \text{VaR}_\epsilon(f(x, \zeta))]$ and it becomes decreasing in $y$ when $y > \text{VaR}_\epsilon(f(x, \zeta))$. 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

$$\phi(x, y) = \int_0^y \mathbb{P}[f(x, \zeta) \geq t] dt - y \cdot (1 - \epsilon)$$

$$= y \int_0^1 \mathbb{P}[f(x, \zeta) \geq sy] ds - y \cdot (1 - \epsilon)$$

$$= y \int \int 1 \{ (\zeta, s) : f(x, \zeta) \geq sy \} \cdot 1_{[0, 1]}(s) ds d\mathbb{P}(\zeta) - y \cdot (1 - \epsilon),$$

$$= y \int \int 1 \{ (\zeta, s) : f(x, \zeta) \geq sy \} d\hat{\mathbb{P}}(\zeta, s) - y \cdot (1 - \epsilon),$$

$$= y \cdot \hat{\mathbb{P}}[f(x, \zeta) \geq sy] - y \cdot (1 - \epsilon),$$

13
where the third equality is by Tonelli’s theorem and \( \hat{P} \) represents the product measure of \( P \) and the uniform distribution on \([0,1]\). Since these two distributions are log-concave on \( \Xi \) and \([0,1]\), respectively, \( \hat{P} \) 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 [Prè13; Nor93]. As a result, Algorithm 1 uses the existing solution approach as a building block and assumes that there exists an oracle, denoted by \( O_u(y, \varepsilon) \), which for given \( y \) and \( \varepsilon > 0 \) returns an \( \varepsilon \)-optimal solution \( \hat{x} \in \{ x \in X : c^T x \leq u \} \) such that

\[
\hat{P} \left[ f(\hat{x}, \zeta) \geq sy \right] \geq \max_{x \in X : c^T x \leq u} \left\{ \hat{P} \left[ f(x, \zeta) \geq sy \right] \right\} - \varepsilon.
\]

We are now ready to present Algorithm 1.

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

**Inputs:** budget \( u \), risk level \( \varepsilon \), oracle \( O_u \), a diminishing sequence \( \{ \varepsilon_k \}_k \), and an \( x_1 \) such that \( y_1 := \text{VaR}_\varepsilon(f(x_1, \zeta)) > 0 \).

1. for \( k = 1, 2, \ldots \) do
2. \( x_{k+1} \leftarrow O_u(y_k, \varepsilon_k) \);
3. \( y_{k+1} \leftarrow \text{VaR}_\varepsilon(f(x_{k+1}, \zeta)) \);
4. if stopping criterion is satisfied then
5. return \( \phi(x_{k+1}, y_{k+1}) \).

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

\[
\min_{x \in X} \left\{ 0 : P \left[ f(x, \zeta) \geq \varepsilon_0 \right] \geq 1 - \varepsilon, c^T x \leq u \right\}, \tag{12}
\]

where \( \varepsilon_0 \) is a small positive constant. If formulation (12) is infeasible for all \( \varepsilon > 0 \), then \( \rho(u) = 0 \) because \( \text{VaR}_\varepsilon(f(x, \zeta)) \) always remains non-positive. Numerically, one can solve (12) for a sequence of diminishing \( \varepsilon_0 \)'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 \( P \) is log-concave and, for all \( k \geq 2 \), \( x_k \) and \( y_k \) are \( \varepsilon_k \)-optimal, i.e.,

\[
\max_x \phi(x, y_{k-1}) - \varepsilon_k \leq \phi(x_k, y_{k-1}) \leq \max_x \phi(x, y_{k-1}) \quad \text{and} \quad |y_k - \text{VaR}_\varepsilon(f(x_k, \zeta))| \leq \varepsilon_k
\]

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 II.

First, we define set \( S := \text{dom} \phi \cap \{(x, y) \in X \times \mathbb{R}_+: c^T x \leq u\} \). Then, by compactness of \( X \) and closedness of \( \text{dom} \phi \) (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 $t_\ell \to 0$ 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 $\Delta$ to obtain
\[
\phi'(x^*, y^*; \Delta) = \phi'(x^*, y^*; \lim_{\ell \to \infty} \frac{1}{t_\ell} [(x_\ell, y_\ell) - (x^*, y^*)])
\]
\[
= \lim_{\ell \to \infty} \phi'(x^*, y^*; \frac{1}{t_\ell} [(x_\ell, y_\ell) - (x^*, y^*)])
\]
\[
= \lim_{\ell \to \infty} \frac{1}{t_\ell} \phi'((x^*, y^*; (x_\ell, y_\ell) - (x^*, y^*)) \leq 0,
\]
where the second and third equalities follow from the continuity and positive homogeneity of $\phi'(x^*, y^*; \Delta)$ in $\Delta$, respectively (see Lemma 9), 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. \hfill \Box

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 $\text{dom} \phi$, 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 $\mathcal{X}^\circ$ 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
\[
\mathcal{X}^\circ = \left\{ x \in \mathbb{R}^n : \text{CVaR}_\epsilon \left(- d(\zeta, S(x)) \right) + \frac{\delta}{1 - \epsilon} \geq 0 \right\},
\]
where the CVaR is with respect to the reference distribution $\mathbb{P}$ and $d(\zeta, S(x))$ is the distance from $\zeta \in \mathbb{R}^m$ to the safe set $S(x)$,
\[
d(\zeta, S(x)) := \inf_{\xi \in \Xi} \{ \|\zeta - \xi\| : A(x)\xi \leq b(x) \}.
\]

Proof. See Appendix I. \hfill \Box

5.1 Individual (O-CC) with LHS Uncertainty

When $m = 1$, the feasible region of (O-CC) simplifies to
\[
\mathcal{X}^\circ_L := \left\{ x \in \mathbb{R}^n : \sup_{Q \in \mathbb{P}} \mathbb{Q} \left[ a(x)^T \xi \leq b(x) \right] \geq 1 - \epsilon \right\}.
\]
We show that, under the same sufficient conditions as for individual (P-CC), $\mathcal{X}^\circ_L$ admits a SOC representation.
Theorem 11. Suppose that the reference distribution $\mathbb{P}$ is Gaussian with mean $\mu$ and covariance matrix $\Sigma > 0$. In addition, suppose that the norm $\| \cdot \|$ defining $\mathcal{P}$ satisfies $\| \cdot \| = \| \Sigma^{-1/2} \cdot \|_2$. Then, for $\delta > 0$ it holds that

$$
\mathcal{X}_L^\delta = \left\{ x \in \mathbb{R}^n : a(x)^T \mu + c_o \| \Sigma^{1/2} a(x) \|_2 \leq b(x) \right\},
$$

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

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

Proof. Pick any $x \in \mathcal{X}_L^\delta$. If $a(x) = 0$, then $\mathbb{Q}[a(x)^T \xi \leq b(x)] = 1$ if $b(x) \geq 0$ and $\mathbb{Q}[a(x)^T \xi \leq b(x)] = 0$ otherwise. Hence, in this case $\mathcal{X}_L^\delta = \{ x \in \mathbb{R}^n : b(x) \geq 0 \}$, 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 $d(\zeta, S(x)) = (f(x, \zeta))^+$ \cite{CKW18} with

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

By Theorem \cite{O-CC} (O-CC) is equivalent to

$$
\frac{\delta}{1 - \epsilon} + \text{CVaR}_{\epsilon}\left( - f(x, \zeta) \right) \geq 0,
$$

which by Lemma \cite{L1} is further equivalent to

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

It follows that $\mathcal{X}_L^\delta = \mathcal{X}_{L,1}^\delta \cup \mathcal{X}_{L,2}^\delta$ with

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

$$
\mathcal{X}_{L,2}^\delta := \left\{ x \in \mathbb{R}^n : \frac{\delta}{1 - \epsilon} + \text{CVaR}_{\epsilon}\left( - f(x, \zeta) \right) \geq \frac{1}{1 - \epsilon} \mathbb{E}\left[ \left( - f(x, \zeta) \right)^+ \right] \right\}.
$$

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

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

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

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

16
where the second equality is because (i) the expectation in this equality attains its maximum at \( t = 0 \) and (ii) \( \text{VaR}_t(-f(x, \zeta)) \leq 0 \) by definition of \( \lambda^o_{L,2} \), and the final equality uses a change-of-variable \( \epsilon_t := \mathbb{P}[ -f(x, \zeta) < t ] \) (or equivalently, \( t = \text{VaR}_{\epsilon_t}(-f(x, \zeta)) \)). In addition, we notice that \( (1 - \epsilon_t) \cdot \text{CVaR}_{\epsilon_t}(-f(x, \zeta)) \) decreases as \( \epsilon_t \) converges to 1. As a result, we can drop the candidate solution \( \epsilon_t = 1 \) in the final equality. It follows that

\[
\lambda^o_{L,2} = \left\{ x \in \mathbb{R}^n : m(x) \geq \text{VaR}_t(Y) \right\},
\]

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

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

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

\[
\text{CVaR}_\epsilon(-f(x, \zeta)) = -m(x) + \text{CVaR}_\epsilon(Y),
\]

which implies

\[
\delta + (1 - \epsilon) \cdot \text{CVaR}_\epsilon(-f(x, \zeta)) \geq (1 - \epsilon_t) \cdot \text{CVaR}_{\epsilon_t}(-f(x, \zeta)) \quad \forall \epsilon_t \in (\epsilon, 1)
\]

\[
\iff \delta + (1 - \epsilon) \cdot \left( -m(x) + \text{CVaR}_\epsilon(Y) \right) \geq (1 - \epsilon_t) \cdot \left( -m(x) + \text{CVaR}_{\epsilon_t}(Y) \right) \quad \forall \epsilon_t \in (\epsilon, 1)
\]

\[
\iff -m(x) \geq c_o.
\]

Finally, we notice that \( \lambda^o_{L,2} = \{ x \in \mathbb{R}^n : c_o \leq -m(x) \leq -\text{VaR}_t(Y) \} \) is non-empty because for all \( \epsilon' \in (\epsilon, 1) \)

\[
\frac{(1 - \epsilon')\text{CVaR}_{\epsilon'}(Y) - (1 - \epsilon)\text{CVaR}_\epsilon(Y)}{\epsilon' - \epsilon}
\]

\[
= \frac{\mathbb{E} \left[ Y \cdot 1 \{ Y \geq \text{VaR}_{\epsilon'}(Y) \} \right] - \mathbb{E} \left[ Y \cdot 1 \{ Y \geq \text{VaR}_\epsilon(Y) \} \right]}{\epsilon' - \epsilon}
\]

\[
= - \frac{\mathbb{E} \left[ Y \cdot 1 \{ Y \in \left[ \text{VaR}_\epsilon(Y), \text{VaR}_{\epsilon'}(Y) \right) \} \right]}{\epsilon' - \epsilon}
\]

\[
= - \mathbb{E} \left[ Y \mid Y \in \left[ \text{VaR}_\epsilon(Y), \text{VaR}_{\epsilon'}(Y) \right) \right] \leq -\text{VaR}_t(Y),
\]

which implies that \( c_o \leq -\text{VaR}_t(Y) \). Concatenating \( \lambda^o_{L,1} \) and \( \lambda^o_{L,2} \) yields \( \lambda^o_L = \{ x \in \mathbb{R}^n : -m(x) \geq c_o \} \) and completes the proof.

\[ \square \]

Remark 6. Like in \( \text{(P-CC)} \), 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 \( \epsilon \) and Wasserstein radius \( \delta \), \( c_o \) may take a negative value, rendering \( \lambda^o_L \) non-convex. In Figure 1, we depict the “watershed” of \((\epsilon, \delta)\) combinations that produce a \( c_o = 0 \) (see the solid curve), while any combination lying below the watershed leads to a convex and SOC \( \lambda^o_L \). From this figure, we observe that, for any \( \epsilon \), \( \lambda^o_L \) is convex and SOC as long as \( \delta \) is small enough.
Figure 1: Combinations of the risk threshold $\epsilon$ and the Wasserstein radius $\delta$ that produce a $c_0 = 0$; any $(\epsilon, \delta)$ combination under the curve leads to a convex and SOC $X^o_L$.

Remark 7. Once again, like in (P-CC), it can be shown that $c_0$ increases to $\text{VaR}_{1-\epsilon}(Y)$ as $\delta$ tends to zero. Therefore, $X^o_L$ asymptotically recovers the feasible region of (CC) with respect to $P \sim \text{Gaussian}(\mu, \Sigma)$.

5.2 Joint (O-CC) with RHS uncertainty

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

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

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

Lemma 7. $d(\zeta, S(x)) \equiv \min_{\xi \in \Xi} \left\{ \|\xi - \zeta\| : A\xi \leq b(x) \right\}$ is jointly convex in $(\zeta, x)$ 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 $P$ of $P$ is $\alpha$-concave with $0 \leq \alpha \leq 1/m$. Then, $X^o_R$ is convex and closed for $\delta > 0$.

Proof. By Theorem 11 (O-CC) admits the following reformulations:

$$\text{CVaR}_\epsilon \left\{ -d(\zeta, S(x)) \right\} \geq -\frac{\delta}{1-\epsilon}$$

$$\iff \inf_{\gamma \in \mathbb{R}} \left\{ \gamma + \frac{1}{1-\epsilon} \mathbb{E}_P \left\{ -d(\zeta, S(x)) - \gamma \right\}^{+} \right\} \geq -\frac{\delta}{1-\epsilon}$$

$$\iff \gamma + \frac{1}{1-\epsilon} \mathbb{E}_P \left\{ -d(\zeta, S(x)) - \gamma \right\}^{+} \geq -\frac{\delta}{1-\epsilon} \quad \forall \gamma \in \mathbb{R}.$$

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

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

where $\phi(x, \zeta) := \left[ -d(\zeta, S(x)) - \gamma \right]^+$ and $f_\zeta$ represents the probability density function of $\zeta$. It suffices to show that $\phi(x, \zeta) \cdot f_\zeta(\zeta)$ is jointly log-concave in $(x, \zeta)$ 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 $\zeta$ and $\phi(x, \zeta)$ is jointly log-concave in $(x, \zeta)$.

1. Since $\mathbb{P}$ is $\alpha$-concave, its density function $f_\zeta$ is $\alpha'$-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, $f_\zeta$ 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

$$\phi(x_\theta, \zeta_\theta) = \left( -d(\zeta_\theta, S(x_\theta)) - \gamma \right)^+ \geq \left( m_1(-d(\zeta_1, S(x_1)) - \gamma, -d(\zeta_2, S(x_2)) - \gamma; \theta) \right)^+$$

$$\geq m_0(\phi(x_1, \zeta_1), \phi(x_2, \zeta_2); \theta),$$

where the first inequality is because $d(\zeta, S(x))$ is jointly convex in $(x, \zeta)$. 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 $m_0(\phi(x_1, \zeta_1), \phi(x_2, \zeta_2); \theta)$ equals zero by definition.

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

$$\left( m_1(-d(\zeta_1, S(x_1)) - \gamma, -d(\zeta_2, S(x_2)) - \gamma; \theta) \right)^+ = m_1(\phi(x_1, \zeta_1), \phi(x_2, \zeta_2); \theta)$$

$$\geq m_0(\phi(x_1, \zeta_1), \phi(x_2, \zeta_2); \theta),$$

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 $d(\zeta, S(x))$ remains jointly convex in $(x, \zeta)$.

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**: $F = \text{Fixed Return}$, $S1 = \text{Least Profitable/Risky}$, $S10 = \text{Most Prof- itable/Risky}$

(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 \( S_1\text{--}S_{10} \), with random rate of returns \( R_i \) and 
\[
R_i := R_{0,i} + r \quad \forall i \in [10],
\]
where \( R_{0,i} \sim \mathcal{N}(1 + 0.01i, (0.03i)^2) \) represents the randomness of stock \( i \) and \( r \sim \mathcal{N}(0, (0.01)^2) \) denotes the market effect on all stocks. Hence, \( S_1 \) is the least profitable/risky and \( S_{10} \) is the most profitable/risky. We consider the following models based on (P-CC) and (O-CC), respectively:
\[
\begin{align*}
\max_{x \in \mathbb{R}^n_+} & \quad \mathbb{E}\left[ \sum_{i=1}^{n} R_i x_i \right] \\
\text{s.t.} & \quad \inf_{Q \in \mathcal{P}} \mathbb{Q}\left( \sum_{i=1}^{n} R_i x_i \geq \eta \right) \geq 1 - \epsilon,
\end{align*}
\]
where \( \eta = 1.0 \) represents the target return and \( \epsilon \) 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):
\[
\begin{align*}
\min_{x} & \quad c^T x, \\
\text{s.t.} & \quad \inf_{Q \in \mathcal{P}} \mathbb{Q}\left[ T x \geq \xi \right] \geq 1 - \epsilon, \\
& \quad 0 \leq x_i \leq U, \forall i \in [n],
\end{align*}
\]
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
\[
\rho(u) = \max_{x \in \mathbb{R}^n_+, y \in \mathbb{R}_+} \phi(x, y) \equiv \int_0^y \left( \mathbb{P}\left[ \min_{t \in [m]} (T_i x - \zeta_i) \geq t \right] - (1 - \epsilon) \right) dt
\]
\[
\text{s.t.} \quad c^T x \leq u, \\
& \quad 0 \leq x_i \leq U, \forall i \in [n],
\]

21
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 $O_u(y_k, \varepsilon_k)$ 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 $\mathbb{E}_P[\zeta_i]$ 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 $\phi^*$, 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.

Figure 4: Risk envelopes under different risk thresholds

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

(b) $n = 30$ and $m = 20$
We demonstrate the trade-off between the robustness and the budget in Figure 1, 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 $\delta$ that allows (P-CC) to be satisfied. From this figure, we observe that, for fixed $\epsilon$, the largest allowable $\delta$ is an S-shaped function of the budget $u$. That is, $\delta$ remains at zero for small $u$, and then $\delta$ increases with a diminishing momentum as $u$ becomes larger. In addition, for fixed $\delta$, it needs a larger budget $u$ to keep (P-CC) satisfied as $\epsilon$ 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 $x_\theta := \theta x_1 + (1 - \theta) x_2$:

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

By Minkowski’s Inequality with $p$ set to be $\alpha$, we have

$$\left( \left[ \theta^{1/\alpha} \cdot f(x_1) \right]^\alpha + \left[ (1 - \theta)^{1/\alpha} \cdot f(x_2) \right]^\alpha \right)^{1/\alpha} \geq \left( \left[ \theta^{1/\alpha} \cdot (f(x_1) - c) \right]^\alpha + \left[ (1 - \theta)^{1/\alpha} \cdot (f(x_2) - c) \right]^\alpha \right)^{1/\alpha}$$

$$+ \left( \left[ \theta^{1/\alpha} \cdot c \right]^\alpha + \left[ (1 - \theta)^{1/\alpha} \cdot c \right]^\alpha \right)^{1/\alpha},$$

from which we obtain

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

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

$$f(x_\theta) - c \geq \left( \theta \cdot (f(x_1))^\alpha + (1 - \theta) \cdot (f(x_2))^\alpha \right)^{1/\alpha} - c$$

$$\geq \left( \theta \cdot (f(x_1) - c)^\alpha + (1 - \theta) \cdot (f(x_2) - c)^\alpha \right)^{1/\alpha}.$$

\hfill $\Box$

**Appendix B  Proof of Lemma 3**

*Proof.* By definition of CVaR, we have

$$\text{CVaR}_{1-\epsilon} (X^-) = \text{VaR}_{1-\epsilon} (X^-) + \frac{1}{\epsilon} \cdot \mathbb{E} [X^- - \text{VaR}_{1-\epsilon} (X^-)]^+.$$  \hfill (23)

We discuss two cases:

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

$$\text{CVaR}_{1-\epsilon} (X^-) = 0 + \frac{1}{\epsilon} \cdot \mathbb{E} [X^- - 0]^+ = 0.$$
(ii) If $0 \geq \text{VaR}_{1-\epsilon}(X)$, then $\text{VaR}_{1-\epsilon}(X^-) = \text{VaR}_{1-\epsilon}(X)$. It follows that
\[
\mathbb{E} [X^- - \text{VaR}_{1-\epsilon}(X)]^+ = \mathbb{E} \left( (X^- - \text{VaR}_{1-\epsilon}(X)) \cdot 1 \{ X^- \geq \text{VaR}_{1-\epsilon}(X) \} \right)
= \mathbb{E} \left( (X - \text{VaR}_{1-\epsilon}(X) - X^+) \cdot 1 \{ X \geq \text{VaR}_{1-\epsilon}(X) \} \right)
= \mathbb{E} \left( (X - \text{VaR}_{1-\epsilon}(X)) \cdot 1 \{ X \geq \text{VaR}_{1-\epsilon}(X) \} \right) - \mathbb{E} [X^+]
\]
where the first equality is by definitions of positive part $[.]^+$ and $1 \{ \cdot \}$, 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
\[
\text{VaR}_{1-\epsilon}(X^-) + \frac{1}{\epsilon} \cdot \mathbb{E} [X^- - \text{VaR}_{1-\epsilon}(X^-)]^+ = \text{VaR}_{1-\epsilon}(X) + \frac{1}{\epsilon} \cdot \mathbb{E} [X - \text{VaR}_{1-\epsilon}(X)]^+ - \frac{1}{\epsilon} \cdot \mathbb{E} [X^+]
\]
\[
= \text{CVaR}_{1-\epsilon}(X) - \frac{1}{\epsilon} \cdot \mathbb{E} [X^+].
\]

Appendix C  Proof of Lemma \[\text{Lemma 4}\]

Proof. We denote the set of points whose distance to $S^c(x)$ is exactly $y$ by
\[
E := \{ \zeta \in \Xi : d(\zeta, S^c(x)) = y \}.
\]
We notice that $d(\zeta, S^c(x)) = d(\zeta, \text{cl} S^c(x))$, where $\text{cl} S^c(x)$ denotes the closure of $S^c(x)$. Then, by
the item (1) of \[\text{Erd45}\], we have $\text{Leb}(E) = 0$, which further implies that $\mathbb{P}(E) = 0$ because $\mathbb{P}$ is
absolutely continuous with respect to $\text{Leb}(\cdot)$ (see Theorem 2.2 in \[\text{Nor93}\]).

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 $\text{Leb}(\cdot)$.

Appendix D  Proof of Remark \[\text{Remark 3}\]

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 \[\text{3}\] that
\[
c_p = \inf_{\epsilon' \in (0, \epsilon)} \frac{\delta + C_p(\epsilon')}{\epsilon - \epsilon'}.
\]
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 \[\text{15}\] is quasiconvex because
it is a quotient of a convex function over an affine function (see Example 3.38 in \[\text{BBV04}\]). 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 $\epsilon'$ tends to zero and to $+\infty$ as $\epsilon'$ tends to $\epsilon$. As a result, to obtain $c_p$ it suffices to solve the equation $G(\epsilon') = 0$, i.e., to search for an $\epsilon'_*$ such that

$$\delta + \int_{1-\epsilon}^{1-\epsilon'} \text{VaR}_q(Y) \, dq = \text{VaR}_{1-\epsilon'_*}(Y) \cdot (\epsilon - \epsilon'_*),$$

(16)

where $C_p(\epsilon'_*) = \int_{1-\epsilon}^{1-\epsilon'} \text{VaR}_q(Y) \, dq$ 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 $-\text{VaR}_{1-\epsilon'_*}(Y)$ and $-\text{VaR}_{1-\epsilon'_*}(Y) - (\epsilon - \epsilon'_*)/f_Y(-\text{VaR}_{1-\epsilon'_*}(Y))$, respectively. It follows that $\epsilon'_*$ increases as $\delta$ decreases.

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

$$0 = \int_{1-\epsilon}^{1-\epsilon'} \text{VaR}_q(Y) \, dq - \text{VaR}_{1-\epsilon'}(Y) \cdot (\epsilon - \overline{\epsilon}) > 0$$

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

It follows that

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

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 $\text{VaR}_{1-\epsilon}(f(x, \zeta))$, i.e.,

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

is convex. To this end, we note that

$$\text{VaR}_{1-\epsilon}(f(x, \zeta)) \geq \theta \iff \mathbb{P}\{f(x, \zeta) \leq \theta\} \leq 1 - \epsilon \iff \mathbb{P}\{f(x, \zeta) - \theta \geq 0\} \geq \epsilon$$

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

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

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

**Appendix F  Proof of Lemma 6**

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

$$\lim_{k \to \infty} \mathbbm{1}\{f(x_k, \zeta) \geq t_k\} = \mathbbm{1}\{f(\hat{x}, \zeta) \geq \hat{t}\}$$
because the function $f(x, \zeta) - t$ is continuous in $(x, t)$. Hence, as a function of $\zeta$, $\mathbb{1}\{f(x_k, \zeta) \geq t_k\}$ converges pointwise to $\mathbb{1}\{f(\hat{x}, \zeta) \geq \hat{t}\}$ on the complement of
\[
\mathcal{U}_0 := \{ \zeta \in \Xi : f(\hat{x}, \zeta) = \hat{t} \}.
\]
It follows that
\[
\lim_{k \to \infty} \psi(x_k, t_k) + (1 - \epsilon) = \lim_{k \to \infty} \mathbb{P}[f(x_k, \zeta) \geq t_k]
= \lim_{k \to \infty} \int_{\Xi \setminus \mathcal{U}_0} \mathbb{1}\{\zeta : f(x_k, \zeta) \geq t_k\} \, d\mathbb{P}(\zeta)
= \int_{\Xi \setminus \mathcal{U}_0} \lim_{k \to \infty} \mathbb{1}\{\zeta : f(x_k, \zeta) \geq t_k\} \, d\mathbb{P}(\zeta)
= \int \mathbb{1}\{\zeta : f(\hat{x}, \zeta) \geq \hat{t}\} \, d\mathbb{P}(\zeta) = \psi(\hat{x}, \hat{t}) + (1 - \epsilon),
\]
where the second and fourth equality are because $\text{Leb}(\mathcal{U}_0(x, t)) = 0$, and the third equality is by the dominated convergence theorem. The continuity of $\phi$ can be established in a similar way: let $\{(x_k, y_k)\}_k$ be a sequence such that converges to $(\hat{x}, \hat{y})$. Then,
\[
\lim_{k \to \infty} \phi(x_k, y_k) = \int_{\mathbb{R}^+} \lim_{k \to \infty} \psi(x_k, t) \cdot \mathbb{1}\{t \leq y_k\} \, dt = \int_{\mathbb{R}^+ \setminus \{\hat{y}\}} \psi(\hat{x}, t) \cdot \lim_{k \to \infty} \mathbb{1}\{t \leq y_k\} \, dt
= \int_{\mathbb{R}^+ \setminus \{\hat{y}\}} \psi(\hat{x}, t) \cdot \mathbb{1}\{t \leq \hat{y}\} \, dt = \phi(\hat{x}, \hat{y}),
\]
where the first equality is by the dominated convergence theorem, and the second equality is because $\psi$ is continuous and $\mathbb{1}\{t \leq y_k\}$ has a limit as $k \to \infty$ when $t \neq \hat{y}$. This completes the proof.

**Appendix G  Proof of Proposition 2**

**Proof.** We first show the $\alpha^*_x$-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 $\phi(x, y) = \int_{\mathbb{R}} \psi(x, t) \, dt$. Pick any $(x_0, y_0), (x_1, y_1) \in \text{dom} \phi$, 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 $\text{dom} \phi$ because $\text{dom} \phi$ 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 $\alpha$-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 8 that
\[
\int_{S_0 + \frac{1}{2} S_1} \psi(x_{1/2}, t) \, dt \geq m_\alpha^* \left[\int_{S_0} \psi(x_0, t) \, dt, \int_{S_1} \psi(x_1, t) \, dt; \frac{1}{2}\right],
\]
or equivalently, $\phi(x_{1/2}, y_{1/2}) \geq m_\alpha^* [\phi(x_0, y_0), \phi(x_1, y_1); 1/2]$. This shows the midpoint $\alpha^*_x$-concavity of $\phi(x, y)$, which together with its continuity (see Lemma 6) shows the $\alpha^*_x$-concavity.

Second, the closedness of $\text{dom} \phi$ follows from the continuity of $\psi$ 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 $y := \text{VaR}_e(f(x, \zeta)) \geq 0$, 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 $\phi(x,y) = \int_0^y (P[f(x,\zeta) \geq t] - (1-\epsilon)) dt$, there exists a $t \in [0,y]$ such that $P[f(x,\zeta) \geq t] \geq (1-\epsilon)$, which implies that $P[f(x,\zeta) \geq 0] \geq (1-\epsilon)$, i.e., constraint (10c). Finally, we notice that $\phi(x,y) \leq \phi(x,\text{VaR}_\epsilon(f(x,\zeta)))$ and hence $\delta \leq \phi(x,\text{VaR}_\epsilon(f(x,\zeta)))$, i.e., constraint (10d). 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 $S \equiv \text{dom} \phi \cap \{(x,y) \in X \times \mathbb{R}_+ : c^\top x \leq u\}$ is compact. Since $\phi(x,y)$ is continuous by Lemma 6, it is bounded on $S$. In addition, we notice that by construction the $\phi$-values of the iterates produced by Algorithm 1 are non-decreasing, i.e.,

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

Hence, this non-decreasing, bounded sequence converges to a finite value. It follows that the two subsequences $\{(x_k, y_k)\}_k$ 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 $P[f(x_{k+1}, \zeta) \geq t] > 1-\epsilon$. Then, $P[f(x_{k+1}, \zeta) \geq 0] > 1-\epsilon$, or equivalently, $\text{VaR}_\epsilon(f(x_{k+1}, \zeta)) > 0$. It follows that $y_{k+1} \equiv \text{VaR}_\epsilon(f(x_{k+1}, \zeta)) \geq 0$ 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

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

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

$$\left| y^* - \text{VaR}_\epsilon(f(x^*, \zeta)) \right| = \lim_{k \to \infty} |y_k - \text{VaR}_\epsilon(f(x_k, \zeta))|$$

$$= \lim_{k \to \infty} |(y_k - \text{VaR}_\epsilon(f(x_k, \zeta)))| = \lim_{k \to \infty} |\epsilon_k| = 0,$$

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

27
(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

\[
\phi(x_k + d_x, y_k) \leq \max_x \phi(x, y_k) \leq \phi(x_{k+1}, y_k) + \epsilon_k.
\]

Driving \(k\) to infinity yields

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

by continuity of \(\phi\) 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 \(\phi\).

(Directional derivative) Since \(\phi(x, y)\) is log-concave and \(\phi(x^*, y^*) > 0\), \(\phi\) 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 \(\varphi(x, t) := \mathbb{P} \{ f(x, \zeta) \geq t \}\) and recast the finite difference

\[
\phi(x^* + s d_x, y^* + s d_y) - \phi(x^*, y^*) = \phi(x^* + s d_x, y^*) + \phi(x^* + s d_x, y^*) - \phi(x^*, y^*)
\]

\[
= \int_{y^*}^{y^* + s d_y} \left( \varphi(x^* + s d_x, t) - (1 - \epsilon) \right) dt + \left( \phi(x^* + s d_x, y^*) - \phi(x^*, y^*) \right).
\]

(18)

For the second term in \(\text{LHS}\), we have

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

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

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

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

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

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

\[
\left| \ln \varphi(x, t) - \ln \varphi(x^*, y^*) \right| \leq M \|(x - x^*, t - y^*)\|_2 \quad \forall (x, t) \in \mathcal{B}((x^*, y^*), r),
\]

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

\[
\left| \ln \varphi(x^* + s d_x, t) - \ln \varphi(x^*, t) \right| \leq \left| \ln \varphi(x^* + s d_x, t) - \ln \varphi(x^*, y^*) \right| + \left| \ln \varphi(x^*, y^*) - \ln \varphi(x^*, t) \right|
\]

\[
\leq M \|(s d_x, t - y^*)\|_2 + M \|(0, t - y^*)\|_2
\]

\[
\leq M \|(s d_x, s d_y)\|_2 + M \|(0, s d_y)\|_2
\]

\[
\leq 2sM \|(d_x, d_y)\|_2,
\]

28
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

\[
\| (sd_x, t - y^*) \|_2^2 = \| sd_x \|_2^2 + |t - y^* |^2 \leq \| sd_y \|_2^2 = \| (sd_x, sd_y) \|_2^2.
\]

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

\[
\int_{y^*}^{y^* + sd_y} (\varphi(x^* + sd_x, t) - (1 - \epsilon)) \, dt = \int_{y^*}^{y^* + sd_y} \left( \exp \left[ \ln \varphi(x^* + sd_x, t) \right] - (1 - \epsilon) \right) \, dt \\
\leq \int_{y^*}^{y^* + sd_y} \left( \exp \left[ \ln \varphi(x^*, t) + 2sM\|(d_x, d_y)\|_2 \right] - (1 - \epsilon) \right) \, dt \\
= \exp \left[ 2sM\|(d_x, d_y)\|_2 \right] \left( \int_{y^*}^{y^* + sd_y} \left[ \varphi(x^*, t) - (1 - \epsilon) \right] \, dt + (1 - \epsilon)(1 - \exp [-2sM\|(d_x, d_y)\|_2]) \, sd_y \right).
\]

It follows that

\[
\lim_{s \to 0^+} \frac{1}{s} \int_{y^*}^{y^* + sd_y} (\varphi(x^* + sd_x, t) - (1 - \epsilon)) \, dt \\
\leq \left( \lim_{s \to 0^+} \exp \left[ 2sM\|(d_x, d_y)\|_2 \right] \right) \cdot \left( \phi'(x^*, y^*; (0, d_y)) + \lim_{s \to 0^+} \frac{1}{s}(1 - \epsilon)(1 - \exp [-2sM\|(d_x, d_y)\|_2]) \, sd_y \right) \\
= \phi'(x^*, y^*; (0, d_y)),
\]

where the inequality is because

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

Second, if \( d_y < 0 \), then it holds that

\[
\int_{y^*}^{y^* + sd_y} (\varphi(x^* + sd_x, t) - (1 - \epsilon)) \, dt = \int_{y^*}^{y^*} (\exp \left[ \ln \varphi(x^* + sd_x, t) \right] - (1 - \epsilon)) \, dt \\
\leq \int_{y^*}^{y^*} \left( - \exp \left[ \ln \varphi(x^*, t) + 2sM\|(d_x, d_y)\|_2 \right] + (1 - \epsilon) \right) \, dt \\
= \exp \left[ - 2sM\|(d_x, d_y)\|_2 \right] \left( \int_{y^*}^{y^* + sd_y} \left[ - \varphi(x^*, t) + (1 - \epsilon) \exp [2sM\|(d_x, d_y)\|_2] \right] \, dt \right) \\
= \exp \left[ - 2sM\|(d_x, d_y)\|_2 \right] \left( \int_{y^*}^{y^* + sd_y} \left[ \varphi(x^*, t) - (1 - \epsilon) \exp [2sM\|(d_x, d_y)\|_2] \right] \, dt \right) \\
= \exp \left[ - 2sM\|(d_x, d_y)\|_2 \right] \left( \int_{y^*}^{y^* + sd_y} \left[ \varphi(x^*, t) - (1 - \epsilon) \right] \, dt + (1 - \epsilon)(1 - \exp [2sM\|(d_x, d_y)\|_2]) \, sd_y \right),
\]

where the inequality is because \( \ln \varphi(x^* + sd_x, t) \geq \ln \varphi(x^*, t) - 2sM\|(d_x, d_y)\|_2 \) and that the function
− \exp(\cdot) is monotonically decreasing. It follows that
\[
\lim_{s \to 0^+} \frac{1}{s} \int_{y^*}^{y^* + s d y} (\varphi(x^* + s d x, t) - (1 - \epsilon)) \, dt
\]
\[
\leq \left( \lim_{s \to 0^+} \exp \left[ -2 s M \|d_x, d_y\|_2 \right] \right) \cdot \left( \phi'(x^*, y^*; (0, d_y)) + \lim_{s \to 0^+} \frac{1}{s} (1 - \epsilon) \right) \exp \left[ 2 s M \|d_x, d_y\|_2 \right] s d y
\]
= \phi'(x^*, y^*; (0, d_y)).

Finally, applying the above analysis on both terms in (18) yields
\[
\phi'(x^*, y^*; (d_x, d_y)) = \lim_{s \to 0^+} \frac{1}{s} \left[ \phi(x^* + s d x, y^* + s d y) - \phi(x^*, y^*) \right]
\]
\[
\leq \phi'(x^*, y^*; (0, d_y)) + \phi'(x^*, y^*; (d_x, 0)) \leq 0,
\]
which completes the proof.

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

Proof. For notation brevity, we denote \(z = (x, y)\). Then, it holds that
\[
\phi'(z; \Delta) = \lim_{s \to 0^+} \frac{1}{s} \left[ \phi(z + s \Delta) - \phi(z) \right]
\]
= \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\}
\[
= \phi(z) \lim_{s \to 0^+} \frac{\ln \phi(z + s \Delta) - \ln \phi(z)}{s} = \phi(z) \cdot (\ln \phi)'(z; \Delta),
\]
where the third equality follows from the L’Hôpital’s rule. Since \((\ln \phi)'(z; \Delta)\) is convex and positively homogeneous in \(\Delta\) 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], \((O-\text{CC})\) is equivalent to inequality
\[
\sup_{Q \in \mathcal{P}} \mathbb{Q}[A(x) \xi \leq b(x)] \equiv \min_{\lambda \geq 0} \left\{ \lambda \delta - \mathbb{E}_\mathcal{P} \left[ \inf_{\xi \in \Xi} \left\{ \lambda \|\zeta - \xi\| - 1 \{ \xi \in \mathcal{S}(x) \} \right\} \right] \right\} \geq 1 - \epsilon.
\]
Noting that for any fixed \(x \in \mathbb{R}^n\) and \(\zeta \in \Xi\)
\[
\inf_{\xi \in \Xi} \left\{ \lambda \|\zeta - \xi\| - 1 \{ \xi \in \mathcal{S}(x) \} \right\} = \begin{cases} -1 & \text{if } \zeta \in \mathcal{S}(x) \\ \min \{ \lambda \cdot d(\zeta, \mathcal{S}(x)) - 1, 0 \} & \text{if } \zeta \notin \mathcal{S}(x) \end{cases}
\]
= \min \{ \lambda \cdot d(\zeta, \mathcal{S}(x)) - 1, 0 \},
we recast \(X^*\) as
\[
\lambda \delta + \mathbb{E}_\mathcal{P} \left[ \max \{ 1 - \lambda \cdot d(\zeta, \mathcal{S}(x)), 0 \} \right] \geq 1 - \epsilon \quad \forall \lambda \geq 0.
\]
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 \( \lambda \) and denote \( \gamma = 1/\lambda \) to obtain
\[
\delta + \mathbb{E}_P \left[ (\gamma - d(\zeta, S(x)), 0)^+ \right] \geq (1 - \epsilon)\gamma \quad \forall \gamma \geq 0.
\]

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 \( \gamma \) to be the whole real line and finish the proof as follows:
\[
(-\gamma) + \frac{1}{1 - \epsilon} \mathbb{E}_P \left[ (-d(\zeta, S(x)) - (-\gamma), 0)^+ \right] \geq -\frac{\delta}{1 - \epsilon} \quad \forall \gamma \in \mathbb{R}
\]
\[
\iff \inf_{-\gamma \in \mathbb{R}} \left\{ (-\gamma) + \frac{1}{1 - \epsilon} \mathbb{E}_P \left[ (-d(\zeta, S(x)) - (-\gamma), 0)^+ \right] \right\} \geq -\frac{\delta}{1 - \epsilon}
\]
\[
\iff \text{CVaR}_\epsilon \left( -d(\zeta, S(x)) \right) + \frac{\delta}{1 - \epsilon} \geq 0.
\]

\[\square\]

**Appendix J  Proof of Lemma 7**

**Proof.** Since \( d(\zeta, S(x)) \) is defined through a convex program, in which the Slater’s condition holds, we take the dual to obtain
\[
d(\zeta, S(x)) = \max_{\lambda \leq 0} \left\{ \Lambda^T [b(x) - A\zeta] : \|A^T \lambda\|_* \leq 1 \right\}.
\]

This completes the proof. \[\square\]
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 3).

[CC59] A. Charnes and W. W. Cooper. “Chance-Constrained Programming”. In: *Management Science* 6.1 (1959), pp. 73–79 (page 11).

[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. Erdős. “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. Grippo 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 2).

[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 5).