Math. Program., Ser. A (2019) 173:151–192
https://doi.org/10.1007/s10107-017-1212-x

FULL LENGTH PAPER

Ambiguous risk constraints with moment and unimodality information

Bowen Li[1] · **Ruiwei Jiang**[2] ·
Johanna L. Mathieu[1]

Received: 14 September 2016 / Accepted: 12 November 2017 / Published online: 24 November 2017

Abstract Optimization problems face random constraint violations when uncertainty arises in constraint parameters. Effective ways of controlling such violations include risk constraints, e.g., chance constraints and conditional Value-at-Risk constraints. This paper studies these two types of risk constraints when the probability distribution of the uncertain parameters is ambiguous. In particular, we assume that the distributional information consists of the first two moments of the uncertainty and a generalized notion of unimodality. We find that the ambiguous risk constraints in this setting can be recast as a set of second-order cone (SOC) constraints. In order to facilitate the algorithmic implementation, we also derive efficient ways of finding violated SOC constraints. Finally, we demonstrate the theoretical results via computational case studies on power system operations.

Keywords Ambiguity · Chance constraints · Conditional Value-at-Risk · Second-order cone representation · Separation · Golden section search

Mathematics Subject Classification 90C15 Stochastic programming · 90C22 Semidefinite programming · 90C34 Semi-infinite programming

✉ Ruiwei Jiang
ruiwei@umich.edu

Bowen Li
libowen@umich.edu

Johanna L. Mathieu
jlmath@umich.edu

1 Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA

2 Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, USA

1 Introduction

In an uncertain environment, optimization problems often involve making decisions before the uncertainty is realized. In this case, constraints, which may include security criteria and capacity restrictions, may face random violations. For example, we consider a constraint subject to uncertainty taking the form

$$a(x)^\top \xi \leq b(x), \tag{1}$$

where $x \in \{0, 1\}^{n_\text{B}} \times \mathbb{R}^{n-n_\text{B}}$ represents an n-dimensional decision variable, $n_\text{B} \in \{0, 1, \ldots, n\}$ represents the number of binary decisions, $a(x) : \mathbb{R}^n \to \mathbb{R}^T$ and $b(x) : \mathbb{R}^n \to \mathbb{R}$ represent two affine transformations of x, and $\xi \in \mathbb{R}^T$ represents a T-dimensional random vector defined on probability space $(\mathbb{R}^T, \mathcal{B}^T, \mathbb{P}_\xi)$ with Borel σ-algebra $\mathcal{B}^T$. An intuitive way of handling random violations of (1) is to employ chance constraints, which attempt to satisfy (1) with at least a pre-specified probability, i.e.,

$$\mathbb{P}_\xi\{a(x)^\top \xi \leq b(x)\} \geq 1 - \epsilon, \tag{2}$$

where $1 - \epsilon$ represents the confidence level of the chance constraint with ϵ usually taking a small value (e.g., 0.05 or 0.10; see, e.g., [8,23]). Dating back to the 1950s, chance constraints have been applied in a wide range of applications including power system operations (see, e.g., [25,40]), production planning (see, e.g., [6,15]), and chemical processing (see, e.g., [19,20]).

In practice, a decision maker is often interested in not only the violation probability of constraint (1), but also the violation magnitude if any (see, e.g., [28,29]). Indeed, chance constraint (2) offers no guarantees on the magnitude of $a(x)^\top \xi - b(x)$ when it is positive. This motivates an alternative risk measure called the conditional Value-at-Risk (CVaR) that examines the (right) tail of $a(x)^\top \xi - b(x)$. More precisely, the CVaR of a one-dimensional random variable χ with confidence level $1 - \epsilon \in (0, 1)$ is defined as

$$\text{CVaR}_{\mathbb{P}_\chi}^{\epsilon}(\chi) = \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}}[\chi - \beta]_+ \right\}, \tag{3}$$

where $\mathbb{P}_\chi$ represents the probability distribution of χ and $[x]_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$. When the infimum is attained in (3), β represents the Value-at-Risk of χ with confidence level $1 - \epsilon$, that is, $\mathbb{P}_\chi\{\chi \leq \beta\} \geq 1 - \epsilon$ (see [2,29]). As a consequence, $\text{CVaR}_{\mathbb{P}_\chi}^{\epsilon}(\chi)$ measures the conditional expectation of χ on its right ϵ-tail. Hence, chance constraint (2) is implied by the CVaR constraint

$$\text{CVaR}_{\mathbb{P}_\xi}^{\epsilon}(a(x)^\top \xi) \leq b(x). \tag{4}$$

A basic challenge to using risk constraints (2) and (4) is that complete information of probability distribution $\mathbb{P}_\xi$ may not be available. Under many circumstances, we only have structural knowledge of $\mathbb{P}_\xi$ (e.g., symmetry, unimodality, etc.) and possibly a series of historical data that can be considered as samples taken from the true (while ambiguous) distribution. As a result, the solution obtained from a risk-constrained model can be biased, i.e., sensitive to the $\mathbb{P}_\xi$ we employ in constraints (2) and (4), and hence perform poorly in out-of-sample tests. A natural way of addressing this challenge is to employ a set of plausible probability distributions, termed the ambiguity set, rather than a single estimate of $\mathbb{P}_\xi$.

1.1 Ambiguity set with unimodality information

We consider an ambiguity set characterized by the first two moments of ξ and a structural requirement that $\mathbb{P}_\xi$ is unimodal in a generalized sense. By definition, if $T = 1$, then $\mathbb{P}_\xi$ is unimodal about 0 if function $F(z) := \mathbb{P}_\xi(\xi \leq z)$ is convex on $(-\infty, 0)$ and concave on $(0, \infty)$. If ξ admits a density function $f_\xi(z)$, then unimodality is equivalent to $f_\xi(z)$ being nondecreasing on $(-\infty, 0)$ and nonincreasing on $(0, \infty)$. In a multidimensional setting, i.e., if $T > 1$, an intuitive extension of this notion is that $f_\xi(zd)$ is nonincreasing on $(0, \infty)$ for all $d \in \mathbb{R}^T$ and $d \neq 0$. That is, the density function of ξ is nonincreasing along any ray emanating from the mode. The following definitions extend this intuitive notion to also cover the distributions that do not admit density functions.

Definition 1 (*Star-Unimodality; see* [11]) A set $S \subseteq \mathbb{R}^T$ is called *star-shaped* about 0 if, for all $\xi \in S$, the line segment connecting 0 and ξ is completely contained in S. A probability distribution $\mathbb{P}_\xi$ on $\mathbb{R}^T$ is called star-unimodal about 0 if it belongs to the closed convex hull of the set of all uniform distributions on sets in $\mathbb{R}^T$ which are star-shaped about 0.

In this paper, we consider a more general notion than the star-unimodality as follows.

Definition 2 (*α-Unimodality; see* [11]) For any given $\alpha > 0$, a probability distribution $\mathbb{P}_\xi$ is called α-unimodal about 0 if function $G(z) := z^\alpha \mathbb{P}_\xi(S/z)$ is nondecreasing on $(0, \infty)$ for all Borel set $S \in \mathcal{B}^T$.

If ξ admits a density function $f_\xi(z)$, then it can be shown that $\mathbb{P}_\xi$ is α-unimodal about 0 if and only if $z^{T-\alpha} f_\xi(zd)$ is nonincreasing on $(0, \infty)$ for all $d \in \mathbb{R}^T$ and $d \neq 0$ (see [11,36]). As compared to star-unimodal distributions, the density of an α-unimodal distribution can increase along rays emanating from the mode (e.g., when $\alpha > T$), but the increasing rate is controlled by α. Indeed, along any ray d, $f_\xi(zd)$ does not increase faster than $z^{\alpha-T}$ on $(0, \infty)$. Furthermore, when $\alpha = T$, $f_\xi(zd)$ is nonincreasing on $(0, \infty)$ for all d. This implies that α-unimodality reduces to star-unimodality when $\alpha = T$.

Given the first two moments of ξ and α-unimodality, we define the following ambiguity set

$$\mathcal{D}_\xi(\mu, \Sigma, \alpha) := \left\{\mathbb{P}_\xi \in \mathcal{M}_T : \mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu,\ \mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = \Sigma,\ \mathbb{P}_\xi \text{ is } \alpha\text{-unimodal about } 0\right\}, \tag{5}$$

where $\mathcal{M}_T$ represents the set of all probability distributions on $(\mathbb{R}^T, \mathcal{B}^T)$, and μ and Σ represent the first and second moments of ξ, respectively. Without loss of generality, we assume that the mode of ξ is 0 in definition (5) and a general mode m can be modeled by shifting ξ to $\xi - m$ (see, e.g., Example 3.4.4 in [17]). For notational brevity, we often refer to ambiguity set $\mathcal{D}_\xi$ with its dependency on parameters (μ, Σ, α) omitted. Based on $\mathcal{D}_\xi$, we consider an ambiguous chance constraint (ACC)

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \leq b(x)\} \geq 1 - \epsilon, \tag{6}$$

that is, we wish to satisfy chance constraint (2) for all probability distributions $\mathbb{P}_\xi$ in ambiguity set $\mathcal{D}_\xi$. Similarly, we define an ambiguous CVaR constraint (AVC)

$$\sup_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \text{CVaR}^\epsilon_{\mathbb{P}_\xi}(a(x)^\top \xi) \leq b(x), \tag{7}$$

which requires that CVaR constraint (4) is satisfied for all $\mathbb{P}_\xi$ in $\mathcal{D}_\xi$.

1.2 Relations to the prior work

In recent years, distributionally robust optimization (DRO) has become an important tool to handle distributional ambiguity in stochastic programs. Using concepts similar to ACC (6) and AVC (7), DRO aims to optimize or protect a system from the worst-case probability distribution, which belongs to a pre-specified ambiguity set. DRO was first introduced by [30] as a minimax stochastic program for the classical newsvendor problem under an ambiguous demand with only moment information. Following this seminal work, moment information has been widely used for characterizing ambiguity sets in various DRO models (see, e.g., [4,10,44]). A key merit of the DRO approach is that the model can often be recast as tractable convex programs such as semidefinite programs (SDPs) (see, e.g., [10]) or SOC programs (see, e.g., [12]). Recently, [41] successfully identified a class of ambiguity sets that lead to tractable convex program reformulations of general DRO models.

ACCs with moment information (and without structural information) have been well-studied in recent years (see, e.g., [1,7,9,12,18,37,39,43]). In particular, [12], [39], and [7] showed that the ACC can be recast as an SOC constraint if the ambiguity set is characterized by the first two moments of ξ. Later, [43] showed that ACC and AVC are actually equivalent if the same ambiguity set is employed. Recently, [1] and [9] extended the analysis of ACC to the case when variable x involves binary (i.e., 0–1) decisions, and [18] made significant progress on representing the ambiguous joint chance constraints in tractable forms. ACCs with information on the density function have also been studied (see, e.g., [13,14,21]).

In contrast, ACCs and AVCs with both moment and structural information have received less attention. [26] considered general DRO models with ambiguity sets incorporating unimodality, symmetry, and convexity. Recently, by using the Choquet representation of α-unimodal distributions, [35] successfully derived SDPs to quantify the worst-case probability bound in ACC. Furthermore, based on both α-unimodality and γ-monotonicity, [36] extended the analysis to quantifying the worst-case expectation in AVC. The main focus of [35,36] is to evaluate the worst-case expectations in ACC or AVC for a given decision variable x. In contrast, we adjust x to satisfy ACC and AVC. In our prior work [22], we derived approximations of AVC. Here, we obtain an exact representation of AVC and derive tighter approximations than those in [22]. To the best of our knowledge, our results on ACC are most related to [17] (in particular, Example 3.4.4), which employs a different ambiguity set that bounds the second moment of ξ by Σ instead of matching it as in (5). Furthermore, [17] derived a representation of ACC based on SDPs. In contrast, in this paper, we employ a different approach based on projection, which allows us to represent ACC with SOC constraints. SOC constraints are more computationally tractable than SDPs, especially when x involves binary decisions. In addition, many off-the-shelf commercial solvers (e.g., CPLEX and GUROBI) can directly handle mixed-integer SOC programs. Finally, [34] assumed that the mean and the mode of ξ coincide and derived a representation of ACC based on SOC constraints. In contrast, in this paper, we study a more general setting where the mean and the mode may be different.

We summarize our main contributions as follows.

1. We derive equivalent reformulations of ACC (6) and AVC (7) using both moment and unimodality information. Both reformulations are SOC constraints and so can be efficiently handled in commercial solvers. Different from previous results in [43], we find that ACC and AVC are not equivalent after incorporating the unimodality information.
2. Inspired by the separation approach (see, e.g., [24]), we derive efficient ways for finding violated SOC constraints in the reformulations of ACC and AVC. The separation procedures can be used to accelerate the algorithmic implementation of ACC and AVC.
3. We derive conservative and relaxed approximations of ACC and AVC that are asymptotically tight. As demonstrated in the computational case study, these approximations help to provide high-quality bounds for the optimal objective value of the test instances.

The remainder of this paper is organized as follows. Section 2 represents ACC (6) as a set of SOC constraints. Section 3 represents AVC (7) as a set of SOC constraints. In both sections, we derive separation procedures for finding violated SOC constraints based on the golden section search. In Sect. 4, we analyze an extension of ACC and AVC to incorporate the linear unimodality in the ambiguity set. We present computational case studies in Sect. 5 and mention future research directions in Sect. 6.

2 Representation of the ambiguous chance constraint

We show that ACC (6) can be recast as second-order cone (SOC) constraints. To this end, we first simplify the computation of the left-hand side of (6), i.e., $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \le b(x)\}$, by projecting random vector ξ on $\mathbb{R}$ and considering a one-dimensional random variable ζ. We summarize this projection result in the following proposition, whose proof relies on the representation of α-unimodal random vectors in [11].

Proposition 1 *Define scalars* $\mu_1 = a(x)^\top \mu$, $\Sigma_1 = a(x)^\top \Sigma a(x)$, *and ambiguity set* $\mathcal{D}_1 = \{\mathbb{P}_\zeta \in \mathcal{M}_1 : \mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_1,\ \mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_1,\ \mathbb{P}_\zeta$ *is* α*-unimodal about* $0\}$. *Then*

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \le b(x)\} = \inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \le b(x)\}. \tag{8}$$

Proof Theorem 3.5 in [11] states that a random vector $X \in \mathbb{R}^m$ is α-unimodal if and only if there exists a random vector $Z \in \mathbb{R}^m$ such that $X = U^{1/\alpha} Z$, where U is uniform in (0, 1) and independent of Z.

First, pick any ξ such that $\mathbb{P}_\xi \in \mathcal{D}_\xi$. Then, there exists Z_ξ such that $\xi = U^{1/\alpha} Z_\xi$. We define $\zeta = a(x)^\top \xi$. It follows that ζ is α-unimodal because $\zeta = a(x)^\top (U^{1/\alpha} Z_\xi) = U^{1/\alpha}(a(x)^\top Z_\xi)$. Furthermore, $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_1$ and $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_1$. Hence, $\mathbb{P}_\zeta \in \mathcal{D}_1$, and so $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \le b(x)\} \ge \inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \le b(x)\}$.

Second, pick any ζ such that $\mathbb{P}_\zeta \in \mathcal{D}_1$. Then, there exists a Z_ζ such that $\zeta = U^{1/\alpha} Z_\zeta$. It follows that $\mathbb{E}[Z_\zeta] = (\frac{\alpha+1}{\alpha})\mu_1$ and $\mathbb{E}[Z_\zeta^2] = (\frac{\alpha+2}{\alpha})\Sigma_1$. Based on Theorem 1 in [27], there exists a $Z_\xi \in \mathbb{R}^T$ such that $Z_\zeta = a(x)^\top Z_\xi$, $\mathbb{E}[Z_\xi] = (\frac{\alpha+1}{\alpha})\mu$, and $\mathbb{E}[Z_\xi Z_\xi^\top] = (\frac{\alpha+2}{\alpha})\Sigma$. We define $\xi = U^{1/\alpha} Z_\xi$. It follows that ξ is α-unimodal, and furthermore $\mathbb{E}_{\mathbb{P}_\xi}[\xi] = (\frac{\alpha}{\alpha+1})\mathbb{E}[Z_\xi] = \mu$ and $\mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = (\frac{\alpha}{\alpha+2})\mathbb{E}[Z_\xi Z_\xi^\top] = \Sigma$. Therefore, $\mathbb{P}_\xi \in \mathcal{D}_\xi$, and so $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi\{a(x)^\top \xi \le b(x)\} \le \inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \le b(x)\}$. □

Next, we compute the worst-case probability $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \le b(x)\}$, for which we make the following two assumptions in the remainder of this section.

Assumption 1 $\left(\frac{\alpha+2}{\alpha}\right) \Sigma \succ \left(\frac{\alpha+1}{\alpha}\right)^2 \mu\mu^\top$.

Assumption 2 Constraint $a(x)^\top \xi \le b(x)$, and so constraint $\zeta \le b(x)$ as well, is satisfied when ξ takes the value of its mode 0. That is, $b(x) \ge 0$.

Assumption 1 is standard in the literature and ensures that $\mathcal{D}_\xi \ne \emptyset$ (see, e.g., [35]). Assumption 2 is standard in the related literature (see, e.g., [17,35,36]). In fact, as ACC (6) requires that $a(x)^\top \xi \le b(x)$ holds with high probability, it is reasonable to assume that it also holds at the mode of ξ. Additionally, given ACC (6) and Proposition 1, we observe that Assumption 2 holds if $\mathbb{P}_\zeta\{\zeta \le 0\} < 1 - \epsilon$ for each $\mathbb{P}_\zeta \in \mathcal{D}_1$, i.e., if the distributions in $\mathcal{D}_1$ are not extremely negative-skewed. To represent ACC (6), we show an equivalent reformulation of $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \le b(x)\}$ in the following proposition that also sheds light on the worst-case probability distribution.

Proposition 2 *Define* $\mu_0 = \left(\frac{\alpha+1}{\alpha}\right)\mu_1$ *and* $\Sigma_0 = \left(\frac{\alpha+2}{\alpha}\right)\Sigma_1$. *Then,* $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \le b(x)\}$ *is equivalent to the optimal objective value of optimization problem*

$$\min_{p_1, p_2, z_1, z_2} \; p_1 + \left(\frac{b(x)}{z_2}\right)^\alpha p_2 \tag{9a}$$

$$\text{s.t.} \; p_1 + p_2 = 1, \tag{9b}$$

$$p_1 z_1 + p_2 z_2 = \mu_0, \tag{9c}$$

$$p_1 z_1^2 + p_2 z_2^2 = \Sigma_0, \tag{9d}$$

$$p_1, p_2 \ge 0, \quad z_1 \in \mathbb{R}, \quad z_2 \ge b(x). \tag{9e}$$

Proof First, we rewrite $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \le b(x)\}$ as a functional optimization problem as follows:

$$\min_{\mathbb{P}_\zeta} \mathbb{P}_\zeta\{\zeta \le b(x)\} \tag{10a}$$
$$\text{s.t. } \mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_1, \tag{10b}$$
$$\mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_1, \tag{10c}$$
$$\mathbb{E}_{\mathbb{P}_\zeta}[1] = 1, \tag{10d}$$
$$\mathbb{P}_\zeta \text{ is } \alpha\text{-unimodal}, \tag{10e}$$

where constraints (10b)–(10c) describe the two moments of ζ, and constraint (10d) ensures that $\mathbb{P}_\zeta$ is a probability distribution. Using Theorem 3.5 in [11], since $\mathbb{P}_\zeta$ is α-unimodal, there exists a random variable Z such that $\zeta = U^{1/\alpha}Z$, where U is uniform in (0, 1) and independent of Z. It follows that $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mathbb{E}[U^{1/\alpha}]\mathbb{E}_{\mathbb{P}_Z}[Z] = (\frac{\alpha}{\alpha+1})\mathbb{E}_{\mathbb{P}_Z}[Z]$ and $\mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \mathbb{E}[U^{2/\alpha}]\mathbb{E}_{\mathbb{P}_Z}[Z^2] = (\frac{\alpha}{\alpha+2})\mathbb{E}_{\mathbb{P}_Z}[Z^2]$. Furthermore,

$$\begin{aligned}
\mathbb{P}_\zeta\{\zeta \le b(x)\} &= \mathbb{P}\{U^{1/\alpha}Z \le b(x)\} \\
&= \int_{z=-\infty}^{+\infty} \mathbb{P}\{U^{1/\alpha}z \le b(x)\} d\mathbb{P}_Z(z) \\
&= \int_{z=-\infty}^{b(x)} 1 \, d\mathbb{P}_Z(z) + \int_{z=b(x)}^{+\infty} \mathbb{P}\left\{U^{1/\alpha} \le \frac{b(x)}{z}\right\} d\mathbb{P}_Z(z) \qquad (11a) \\
&= \int_{z=-\infty}^{b(x)} 1 \, d\mathbb{P}_Z(z) + \int_{z=b(x)}^{+\infty} \left(\frac{b(x)}{z}\right)^\alpha d\mathbb{P}_Z(z) \\
&= \int_{z=-\infty}^{+\infty} \left[\frac{b(x)}{\max\{z, b(x)\}}\right]^\alpha d\mathbb{P}_Z(z), \qquad (11b)
\end{aligned}$$

where equality (11a) is because $U^{1/\alpha}z \le b(x)$ when $z \le b(x)$ (note that $b(x) \ge 0$ due to Assumption 2), and in (11b) we designate that $0/0 = 1$ in case $b(x) = 0$. Hence, problem (10a)–(10e) can be recast as

$$\min_{\mathbb{P}_Z} \mathbb{E}_{\mathbb{P}_Z}\left[\frac{b(x)}{\max\{Z, b(x)\}}\right]^\alpha \tag{12a}$$
$$\text{s.t. } \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \tag{12b}$$
$$\mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0, \tag{12c}$$
$$\mathbb{E}_{\mathbb{P}_Z}[1] = 1. \tag{12d}$$

Second, we take the dual of problem (12a)–(12d) to obtain

$$\max_{\pi,\lambda,\gamma} \mu_0\pi + \Sigma_0\lambda + \gamma \tag{13a}$$
$$\text{s.t. } \lambda z^2 + \pi z + \gamma \le \left[\frac{b(x)}{\max\{z, b(x)\}}\right]^\alpha, \quad \forall z \in \mathbb{R}, \tag{13b}$$

where dual variables π, λ, and γ are associated with primal constraints (12b)–(12d), respectively. Meanwhile, dual constraints (13b) are associated with primal variable $\mathbb{P}_Z$.

Strong duality holds between problems (12a)–(12d) and (13a)–(13b) due to Assumption 1 (see Proposition 3.4 in [31]). It follows that there exists an optimal solution $\mathbb{P}_Z^*$ to (12a)–(12d) that is discrete with at most 3 points of support (see Lemma 3.1 in [32]) and a finite optimal solution $(\pi^*, \lambda^*, \gamma^*)$ to (13a)–(13b) (see Proposition 3.4 in [31]).

Third, strong duality yields

$$\mathbb{E}_{\mathbb{P}_Z^*}\left\{\left[\frac{b(x)}{\max\{Z, b(x)\}}\right]^\alpha - (\lambda^* Z^2 + \pi^* Z + \gamma^*)\right\} = \mathbb{E}_{\mathbb{P}_Z^*}\left[\frac{b(x)}{\max\{Z, b(x)\}}\right]^\alpha - (\lambda^*\Sigma_0 + \pi^*\mu_0 + \gamma^*) = 0.$$

It follows that $\mathbb{P}_Z^*$ is supported at those points z such that $[b(x)/\max\{z, b(x)\}]^\alpha = \lambda^* z^2 + \pi^* z + \gamma^*$. From constraints (13b), we note that $\lambda \leq 0$ and so $\lambda z^2 + \pi z + \gamma$ is concave in z. Additionally, $[b(x)/\max\{z, b(x)\}]^\alpha$ is piecewise convex and consists of two pieces, more specifically,

$$\left[\frac{b(x)}{\max\{z, b(x)\}}\right]^\alpha = \begin{cases} 1, & \text{if } z \leq b(x) \\ \left(\frac{b(x)}{z}\right)^\alpha, & \text{if } z > b(x), \end{cases}$$

and both pieces 1 and $(b(x)/z)^\alpha$ are convex in z. Hence, due to constraints (13b), $[b(x)/\max\{z, b(x)\}]^\alpha$ and $\lambda^* z^2 + \pi^* z + \gamma^*$ can meet at at most two points, each located at one piece of $[b(x)/\max\{z, b(x)\}]^\alpha$. It follows that, without loss of optimality, we can shrink the feasible region of formulation (12a)–(12d) to those discrete distributions with at most two points of support, each corresponding to one piece of $[b(x)/\max\{z, b(x)\}]^\alpha$. Therefore, formulations (12a)–(12d) and (9a)–(9e) have the same optimal objective value (note that we relax $z_1 \leq b(x)$ to $z_1 \in \mathbb{R}$ in (9a)–(9e) without loss of optimality, because it is suboptimal that both z_1 and z_2 correspond to the same piece of $[b(x)/\max\{z, b(x)\}]^\alpha$). □

Remark 1 Suppose that $(p_1^*, p_2^*, z_1^*, z_2^*)$ is an optimal solution to problem (9a)–(9e). From the proof of Proposition 2, we observe that problem (9a)–(9e) is solved for the worst-case probability distribution of a random variable Z such that $\zeta = U^{1/\alpha} Z$, where U is uniform on $(0, 1)$ and independent of Z. It follows that the worst-case distribution $\mathbb{P}_\zeta^*$ attaining $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \leq b(x)\}$ is a mixture in the form $\mathbb{P}_\zeta^* = p_1^* \mathbb{P}_\zeta^1 + p_2^* \mathbb{P}_\zeta^2$, where, for $i = 1, 2$, $\mathbb{P}_\zeta^i$ is defined on the interval connecting 0 and z_i^* (i.e., $[0, z_i^*]$ or $[z_i^*, 0]$, depending on the sign of z_i^*) and $\mathbb{P}_\zeta^i\{|\zeta| \leq t|z_i^*|\} = t^\alpha$ for all $t \in [0, 1]$.

Finally, we reformulate ACC (6) by analyzing problem (9a)–(9e). We summarize the main result of this section in the following theorem.

Theorem 1 *ACC* (6) *is equivalent to a set of SOC constraints*

$$\sqrt{\frac{1 - \epsilon - \tau^{-\alpha}}{\epsilon}}\,\|\Lambda a(x)\| \leq \tau b(x) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x), \quad \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}, \tag{14}$$

where $\Lambda := [(\frac{\alpha+2}{\alpha})\Sigma - (\frac{\alpha+1}{\alpha})^2 \mu\mu^\top]^{1/2}$.

Proof We analyze the solutions to problem (9a)–(9e) and identify all possible solutions (p_1, p_2, z_1, z_2) that satisfy constraints (9b)–(9e). To this end, we analyze the following two cases.

Case 1 If $\mu_0 \leq b(x)$, then we parameterize z_2 by defining $z_2 = \tau b(x)$ for $\tau \geq 1$. Accordingly, we parameterize all solutions (p_1, p_2, z_1, z_2) that satisfy constraints (9b)–(9e) by τ as follows:

$$p_1 = \frac{(\tau b(x) - \mu_0)^2}{(\tau b(x) - \mu_0)^2 + \Sigma_0 - \mu_0^2}, \quad p_2 = \frac{\Sigma_0 - \mu_0^2}{(\tau b(x) - \mu_0)^2 + \Sigma_0 - \mu_0^2}, \tag{15a}$$

$$z_1 = \mu_0 - \frac{\Sigma_0 - \mu_0^2}{\tau b(x) - \mu_0}, \text{ and } z_2 = \tau b(x). \tag{15b}$$

Note that, for each $\tau \geq 1$, (p_1, p_2, z_1, z_2) satisfies constraints (9e) because $p_1, p_2 \geq 0$ and $z_2 = \tau b(x) \geq b(x)$. Then, problem (9a)–(9e) can be recast as

$$\min_{\tau \geq 1} \frac{(\tau b(x) - \mu_0)^2 + \tau^{-\alpha}(\Sigma_0 - \mu_0^2)}{(\tau b(x) - \mu_0)^2 + \Sigma_0 - \mu_0^2}.$$

Hence, ACC (6), i.e., $\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta\{\zeta \leq b(x)\} \geq 1 - \epsilon$, can be recast as

$$\frac{(\tau b(x) - \mu_0)^2 + \tau^{-\alpha}(\Sigma_0 - \mu_0^2)}{(\tau b(x) - \mu_0)^2 + (\Sigma_0 - \mu_0^2)} \geq 1 - \epsilon, \quad \forall \tau \geq 1.$$

After simple transformations, this is equivalent to

$$(\tau b(x) - \mu_0)^2 \geq \left(\frac{1 - \epsilon - \tau^{-\alpha}}{\epsilon}\right)(\Sigma_0 - \mu_0^2), \quad \forall \tau \geq 1. \tag{16}$$

As $(\tau b(x) - \mu_0)^2 \geq 0$, we can assume $\tau \geq (1/(1-\epsilon))^{1/\alpha}$ without loss of generality. Furthermore, because $\tau b(x) - \mu_0 \geq 0$ for all $\tau \geq 1$, we can rewrite constraints (16) as (14), using the definitions of μ_0 and Σ_0.

Case 2 If $\mu_0 > b(x)$, then we parameterize z_2 by defining $z_2 = \tau b(x)$ for $\tau \geq 1$. For all $\tau \geq \mu_0/b(x)$, because $z_2 \geq \mu_0$, we parameterize (p_1, p_2, z_1, z_2) by τ as in (15a)–(15b). Similar to *Case* 1, ACC (6) can be recast as

$$\tau b(x) - \mu_0 \geq \sqrt{\left(\frac{1 - \epsilon - \tau^{-\alpha}}{\epsilon}\right)_+} \sqrt{\Sigma_0 - \mu_0^2}, \quad \forall \tau \geq \frac{\mu_0}{b(x)}. \tag{17a}$$

For all $1 \le \tau < \mu_0/b(x)$, because $b(x) \le z_2 < \mu_0$, we parameterize (p_1, p_2, z_1, z_2) by τ as follows:

$$p_1 = \frac{(\mu_0 - \tau b(x))^2}{(\mu_0 - \tau b(x))^2 + \Sigma_0 - \mu_0^2}, \quad p_2 = \frac{\Sigma_0 - \mu_0^2}{(\mu_0 - \tau b(x))^2 + \Sigma_0 - \mu_0^2},$$
$$z_1 = \mu_0 + \frac{\Sigma_0 - \mu_0^2}{\mu_0 - \tau b(x)}, \text{ and } z_2 = \tau b(x).$$

Then, because $\mu_0 > \tau b(x)$, ACC (6) can be recast as

$$\mu_0 - \tau b(x) \ge \sqrt{\left(\frac{1 - \epsilon - \tau^{-\alpha}}{\epsilon}\right)_+} \sqrt{\Sigma_0 - \mu_0^2}, \quad \forall 1 \le \tau < \frac{\mu_0}{b(x)}. \tag{17b}$$

Combining inequalities (17a)–(17b) and the fact that $(1 - \epsilon - \tau^{-\alpha})/\epsilon > 0$ if and only if $\tau > [1/(1-\epsilon)]^{1/\alpha}$, we have $\mu_0/b(x) \le [1/(1-\epsilon)]^{1/\alpha}$ because otherwise, when $\tau = \mu_0/b(x)$, the left-hand side of (17a) equals zero while the right-hand side is strictly positive. It follows that inequalities (17b) are equivalent to $\mu_0 - \tau b(x) \ge 0$ for all $1 \le \tau < \mu_0/b(x)$ and so redundant, and inequalities (17a) are equivalent to (14), using the definitions of μ_0 and Σ_0. $\square$

In computation, directly replacing ACC with constraints (14) involves an infinite number of SOC constraints and so is computationally intractable. An alternative approach is by separation, i.e., (i) obtain a solution $\hat{x}$ from a relaxed formulation, (ii) find a $\hat{\tau} \ge (1/(1-\epsilon))^{1/\alpha}$ such that $\hat{x}$ violates the corresponding SOC constraint (14), and (iii) incorporate the violated SOC constraint to strengthen the formulation. Note that constraints (14) imply that

$$\tau b(x) - \left(\frac{\alpha + 1}{\alpha}\right)\mu^\top a(x) \ge 0, \quad \forall \tau \ge \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}$$

because $\sqrt{(1 - \epsilon - \tau^{-\alpha})/\epsilon}\, \|\Lambda a(x)\| \ge 0$. These inequalities are equivalent to a single linear constraint $(1/(1-\epsilon))^{1/\alpha} b(x) - [(\alpha+1)/\alpha]\mu^\top a(x) \ge 0$, which we assume is always incorporated in the relaxed formulation in Step (i). We discuss how to efficiently conduct Step (ii) of the separation approach, which is equivalent to solving the following problem:

Separation Problem 1: Given $\hat{x}$, does there exist a $\hat{\tau} \ge (1/(1-\epsilon))^{1/\alpha}$ such that $\hat{x}$ violates constraints (14)?

In the following proposition, we show that Separation Problem 1 can be solved by conducting a golden section search on the real line. This search is computationally efficient.

Proposition 3 *Define $\hat{\mu}_0 = (\frac{\alpha+1}{\alpha})\mu^\top a(\hat{x})$ and $\hat{\Sigma}_0 = (\frac{\alpha+2}{\alpha})a(\hat{x})^\top \Sigma a(\hat{x})$. We have the following:*

1. *If $a(\hat{x}) = 0$, then constraints (14) are always satisfied;*

2. *If* $a(\hat{x}) \neq 0$ *and* $b(\hat{x}) = 0$*, then* $\hat{x}$ *violates constraints* (14) *if and only if it violates them at* $\hat{\tau} = +\infty$*, i.e.,* $\sqrt{(1-\epsilon)/\epsilon}\,\|\Lambda a(\hat{x})\| > -[(\alpha+1)/\alpha]\mu^\top a(\hat{x})$*;*
3. *If* $a(\hat{x}) \neq 0$ *and* $b(\hat{x}) > 0$*, then* $\hat{x}$ *violates constraints* (14) *if and only if* $\sqrt{(1-\epsilon-\hat{\tau}^{-\alpha})/\epsilon}\,\|\Lambda a(\hat{x})\| > \hat{\tau} b(\hat{x}) - [(\alpha+1)/\alpha]\mu^\top a(\hat{x})$*, where* $\hat{\tau}$ *represents the minimizer of the one-dimensional problem*

$$\min_{\tau \geq (1/(1-\epsilon))^{1/\alpha}} \left(b(\hat{x})\tau - \hat{\mu}_0\right)^2 - \left(\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}\right)\left(\hat{\Sigma}_0 - \hat{\mu}_0^2\right), \tag{18}$$

whose objective function is strongly convex and can be minimized via a golden section search in the interval $[(1/(1-\epsilon))^{1/\alpha},\ \hat{\mu}_0/b(\hat{x}) + \alpha(1-\epsilon)^{(\alpha+1)/\alpha}(\hat{\Sigma}_0 - \hat{\mu}_0^2)/(2\epsilon b(\hat{x})^2)]$.

Proof First, if $a(\hat{x}) = 0$, then constraints (14) reduce to $\tau b(x) \geq 0$ for all $\tau \geq (1/(1-\epsilon))^{1/\alpha}$, which always holds due to Assumption 2. Second, if $a(\hat{x}) \neq 0$ and $b(\hat{x}) = 0$, then constraints (14) reduce to

$$\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}}\|\Lambda a(\hat{x})\| \leq -\left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(\hat{x}),\quad \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}.$$

As the left-hand side of the above inequality is increasing in τ, constraints (14) are violated if and only if they are violated at $\hat{\tau} = +\infty$. Third, if $a(\hat{x}) \neq 0$ and $b(\hat{x}) > 0$, then constraints (14) are satisfied if and only if

$$\left[\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}}\|\Lambda a(\hat{x})\|\right]^2 \leq \left[\tau b(\hat{x}) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(\hat{x})\right]^2,\quad \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}$$

because both sides of constraints (14) are nonnegative. By the definitions of $\hat{\mu}_0$ and $\hat{\Sigma}_0$, this is equivalent to $(b(\hat{x})\tau - \hat{\mu}_0)^2 - [(1-\epsilon-\tau^{-\alpha})/\epsilon](\hat{\Sigma}_0 - \hat{\mu}_0^2) \geq 0$ for all $\tau \geq (1/(1-\epsilon))^{1/\alpha}$. It follows that the Separation Problem 1 can be answered by checking constraints (14) at the optimal solution $\hat{\tau}$ of problem (18).

Finally, we denote the objective function of problem (18) as $H(\tau)$. It follows that

$$\begin{aligned} H'(\tau) &= 2b(\hat{x})(b(\hat{x})\tau - \hat{\mu}_0) - \left(\frac{\alpha}{\epsilon}\right)\left(\hat{\Sigma}_0 - \hat{\mu}_0^2\right)\tau^{-\alpha-1}, \\ H''(\tau) &= 2\big[b(\hat{x})\big]^2 + \left(\frac{\alpha^2+\alpha}{\epsilon}\right)\left(\hat{\Sigma}_0 - \hat{\mu}_0^2\right)\tau^{-\alpha-2}. \end{aligned}$$

As $H''(\tau) > 0$ for all $\tau \geq (1/(1-\epsilon))^{1/\alpha}$, $H(\tau)$ is strongly convex and can be minimized via a golden section search. More specifically, if $H'((1/(1-\epsilon))^{1/\alpha}) \geq 0$, then $(1/(1-\epsilon))^{1/\alpha}$ is optimal to problem (18). Otherwise, if $H'((1/(1-\epsilon))^{1/\alpha}) < 0$, then problem (18) is optimized at $\hat{\tau}$ such that $H'(\hat{\tau}) = 0$. It follows that $2b(\hat{x})(b(\hat{x})\hat{\tau} -$

$\hat{\mu}_0) = (\alpha/\epsilon)(\hat{\Sigma}_0 - \hat{\mu}_0^2)\hat{\tau}^{-\alpha-1}$. Since $\hat{\tau} \geq (1/(1-\epsilon))^{1/\alpha}$, we have

$$\begin{aligned}
2b(\hat{x})^2\hat{\tau} \ &\leq \ 2b(\hat{x})\hat{\mu}_0 + \left(\frac{\alpha}{\epsilon}\right)\left(\hat{\Sigma}_0 - \hat{\mu}_0^2\right)(1-\epsilon)^{(\alpha+1)/\alpha} \\
\Rightarrow \ \hat{\tau} \ &\leq \ \frac{\hat{\mu}_0}{b(\hat{x})} + \frac{\alpha(1-\epsilon)^{(\alpha+1)/\alpha}}{2\epsilon b(\hat{x})^2}\left(\hat{\Sigma}_0 - \hat{\mu}_0^2\right).
\end{aligned}$$

Hence, the golden section search can be restricted to the interval $[(1/(1-\epsilon))^{1/\alpha}, \hat{\mu}_0/b(\hat{x}) + \alpha(1-\epsilon)^{(\alpha+1)/\alpha}(\hat{\Sigma}_0 - \hat{\mu}_0^2)/(2\epsilon b(\hat{x})^2)]$ without loss of optimality. □

2.1 Approximations of the ambiguous chance constraint

Before closing this section, we derive relaxed and conservative approximations of ACC (6) by using a finite number of SOC constraints. First, based on the exact representation (14) that involves all $\tau \in \big[[1/(1-\epsilon)]^{1/\alpha}, \infty\big)$, we obtain a relaxed approximation by only involving a finite number of τ. We summarize this approximation in the following proposition, whose proof is immediate and so omitted.

Proposition 4 *For given integer $K \geq 1$ and real numbers $[1/(1-\epsilon)]^{1/\alpha} \leq n_1 < n_2 < \cdots < n_K \leq \infty$, ACC (6) implies the SOC constraints*

$$\sqrt{\frac{1-\epsilon-n_k^{-\alpha}}{\epsilon}}\|\Lambda a(x)\| \leq n_k b(x) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x), \quad \forall k = 1, \ldots, K. \tag{19}$$

Second, we obtain a conservative approximation by approximating the left-hand sides of the inequalities (14) by using a piece-wise linear function of τ.

Proposition 5 *Given integer $K \geq 2$ and real numbers $[1/(1-\epsilon)]^{1/\alpha} = n_1 < n_2 < \cdots < n_K = \infty$, we define a piece-wise linear function containing $(K-1)$ pieces:*

$$g(\tau) \ = \ \min_{k=2,\ldots,K}\left\{\sqrt{\frac{1}{\epsilon(1-\epsilon-n_k^{-\alpha})}}\left[\left(\frac{\alpha n_k^{-\alpha-1}}{2}\right)\tau + 1 - \epsilon - \left(1+\frac{\alpha}{2}\right)n_k^{-\alpha}\right]\right\}.$$

Then, $g(\tau) \geq \sqrt{(1-\epsilon-\tau^{-\alpha})/\epsilon}$ for all $\tau \geq [1/(1-\epsilon)]^{1/\alpha}$. Furthermore, denote $m_1 = [1/(1-\epsilon)]^{1/\alpha}$ and let $m_2 < \cdots < m_{K-1}$ represent the $(K-2)$ breakpoints of function $g(\tau)$, i.e.,

$$m_k = \frac{(1-\epsilon)\left(1 - \sqrt{\frac{1-\epsilon-n_k^{-\alpha}}{1-\epsilon-n_{k+1}^{-\alpha}}}\right) + \left(1+\frac{\alpha}{2}\right)\left(n_{k+1}^{-\alpha}\sqrt{\frac{1-\epsilon-n_k^{-\alpha}}{1-\epsilon-n_{k+1}^{-\alpha}}} - n_k^{-\alpha}\right)}{\left(\frac{\alpha}{2}\right)\left(n_{k+1}^{-\alpha-1}\sqrt{\frac{1-\epsilon-n_k^{-\alpha}}{1-\epsilon-n_{k+1}^{-\alpha}}} - n_k^{-\alpha-1}\right)},$$
$$\forall k = 2, \ldots, K-1.$$

Then, ACC (6) is implied by the SOC constraints

$$g(m_k)\|\Lambda a(x)\| \leq m_k b(x) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x), \quad \forall k = 1, \ldots, K-1. \tag{20}$$

Proof Denote $h(\tau) = \sqrt{(1-\epsilon-\tau^{-\alpha})/\epsilon}$. Then, the first derivative $h'(\tau) = \left(\frac{\alpha\tau^{-\alpha-1}}{2}\right)\sqrt{\frac{1}{\epsilon(1-\epsilon-\tau^{-\alpha})}}$ and the tangent of $h(\tau)$ at n_k is

$$\sqrt{\frac{1}{\epsilon(1-\epsilon-n_k^{-\alpha})}}\left[\left(\frac{\alpha n_k^{-\alpha-1}}{2}\right)\tau + 1 - \epsilon - \left(1+\frac{\alpha}{2}\right)n_k^{-\alpha}\right]$$

for all $k = 2, \ldots, K$. It follows that $g(\tau) \geq h(\tau)$ for all $\tau \geq [1/(1-\epsilon)]^{1/\alpha}$ because $h(\tau)$ is concave on the interval $\big[[1/(1-\epsilon)]^{1/\alpha}, \infty\big)$. Hence, ACC (6) is implied by

$$g(\tau)\|\Lambda a(x)\| \leq \tau b(x) - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x), \quad \forall \tau \geq [1/(1-\epsilon)]^{1/\alpha}. \tag{21}$$

Furthermore, given x, as the left-hand side of (21) is piece-wise linear in τ and the right-hand side of (21) is linear in τ, inequalities (21) hold if and only if they hold at the breakpoints of $g(\tau)$. Therefore, ACC (6) is implied by constraints (20). □

Remark 2 In computation, we can use the conservative approximation (20) to find near-optimal solutions. More specifically, suppose that we employ the separation approach to solve problem $\min\{c(x) : x \in X,\ x \text{ satisfies}(6)\}$ and have finished the first K iterations. Then, from these iterations, we obtain a lower bound c_{L}^K of the optimal objective value and $\hat{\tau}_1, \ldots, \hat{\tau}_K$ by iteratively solving Separation Problem 1. By letting $n_1 = [1/(1-\epsilon)]^{1/\alpha}$, $n_{K+2} = \infty$, and $n_k = \hat{\tau}_{k-1}$ for all $k = 2, \ldots, K+1$, we obtain an upper bound c_{U}^K of the optimal objective value by solving problem $\min\{c(x) : x \in X,\ x \text{ satisfies}\,(20) \text{ based on } n_1, \ldots, n_{K+2}\}$, whose optimal solution is denoted x_K^*. If $(c_{\text{U}}^K - c_{\text{L}}^K)/c_{\text{L}}^K$ is small enough, then we can stop the iterations and output x_K^* as a near-optimal solution.

3 Representation of the ambiguous CVaR constraint

To recast AVC (7) as SOC constraints, we adopt a similar method to that described in Sect. 2. Again, we project random vector ξ on $\mathbb{R}$ and consider a one-dimensional random variable ζ. We summarize this result in the following proposition and omit the proof due to its similarity to that of Proposition 1.

Proposition 6 *The following equality holds:*

$$\sup_{\mathbb{P}_\xi \in \mathcal{D}_\xi} CVaR_{\mathbb{P}_\xi}^{\epsilon}(a(x)^\top \xi) = \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} CVaR_{\mathbb{P}_\zeta}^{\epsilon}(\zeta).$$

We compute $\sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathrm{CVaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta)$ by observing that $\mathbb{P}_\zeta$ is α-unimodal and so there exists a random variable Z such that $\zeta = U^{1/\alpha} Z$, where U is uniform in (0, 1) and independent of Z (see Theorem 3.5 in [11]). We summarize this computation in the following proposition, and note that it can also be obtained by following Theorem 2.1 in [36].

Proposition 7 *The following equality holds:*

$$\sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} CVaR^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) \;=\; \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \right\},$$

where $\mathcal{D}(\mu_0, \Sigma_0) := \{\mathbb{P}_Z \in \mathcal{M}_1 : \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0\}$ *and* $f(Z) = \mathbb{1}[\beta < 0] f_-(Z) + \mathbb{1}[\beta \geq 0] f_+(Z)$*, where*

$$f_+(z) = \begin{cases} 0 & \textit{if } z \leq \beta \\ \left(\frac{\alpha}{\alpha+1}\right) z - \beta + \left(\frac{\beta}{\alpha+1}\right)\left(\frac{\beta}{z}\right)^{\alpha} & \textit{if } z > \beta, \end{cases}$$

$$\textit{and } f_-(z) = \begin{cases} -\left(\frac{\beta}{\alpha+1}\right)\left(\frac{\beta}{z}\right)^{\alpha} & \textit{if } z < \beta \\ \left(\frac{\alpha}{\alpha+1}\right) z - \beta & \textit{if } z \geq \beta. \end{cases}$$

Proof First, based on the definition of CVaR, we have

$$\begin{aligned} \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathrm{CVaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) &= \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \inf_{\beta} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ \right\} \\ &= \inf_{\beta} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ \right\}. \end{aligned} \tag{22}$$

To justify the switch of $\inf_\beta$ and $\sup_{\mathbb{P}_\zeta}$ in (22), we observe that $\beta + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+$ is convex in β and concave (actually affine) in $\mathbb{P}_\zeta$. Additionally, we claim that $\beta \in [\mu_1 - \sqrt{(1+\epsilon)(\Sigma_1 - \mu_1^2)/(1-\epsilon)}, \mu_1 + \sqrt{(2-\epsilon)(\Sigma_1 - \mu_1^2)/\epsilon}]$, i.e., β belongs to a compact set, without loss of optimality. Then, the switch follows from the Sion's minimax theorem (see [33]). To prove this claim, we observe that

$$\mathrm{VaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) \;\leq\; \mathrm{argmin}_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ \right\} \;\leq\; \mathrm{VaR}^{\epsilon+}_{\mathbb{P}_\zeta}(\zeta)$$

for all $\mathbb{P}_\zeta \in \mathcal{D}_1$, where $\mathrm{VaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta) := \inf\{\beta : \mathbb{P}_\zeta\{\zeta \leq \beta\} \geq 1 - \epsilon\}$ and $\mathrm{VaR}^{\epsilon+}_{\mathbb{P}_\zeta}(\zeta) := \inf\{\beta : \mathbb{P}_\zeta\{\zeta \leq \beta\} > 1 - \epsilon\}$ (see Theorem 10 in [29]). It follows that we can assume $\beta \in [\mathrm{VaR}^{\epsilon}_{\mathbb{P}_\zeta}(\zeta), \mathrm{VaR}^{\epsilon+}_{\mathbb{P}_\zeta}(\zeta)]$ for all $\mathbb{P}_\zeta \in \mathcal{D}_1$ without loss of optimality. But $\mathcal{D}_1 \subseteq \mathcal{D}_1^{\infty} := \{\mathbb{P}_\zeta \in \mathcal{M}_1 : \mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_1,\ \mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_1\}$, and it follows from Cantelli's inequality that

$$\begin{aligned}
&\inf_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta \left\{ \zeta \leq \mu_1 + \sqrt{\left(\frac{2-\epsilon}{\epsilon}\right)(\Sigma_1 - \mu_1^2)} \right\} \\
&\quad \geq \inf_{\mathbb{P}_\zeta \in \mathcal{D}_1^\infty} \mathbb{P}_\zeta \left\{ \zeta \leq \mu_1 + \sqrt{\left(\frac{2-\epsilon}{\epsilon}\right)(\Sigma_1 - \mu_1^2)} \right\} \geq 1 - \frac{\epsilon}{2}.
\end{aligned}$$

Hence, $\mathrm{VaR}_{\mathbb{P}_\zeta}^{\epsilon+}(\zeta) \leq \mu_1 + \sqrt{(2-\epsilon)(\Sigma_1 - \mu_1^2)/\epsilon}$ for all $\mathbb{P}_\zeta \in \mathcal{D}_1$ because otherwise there exists a $\mathbb{P}_\zeta \in \mathcal{D}_1$ such that $\mathbb{P}_\zeta\{\zeta \leq \mu_1 + \sqrt{(2-\epsilon)(\Sigma_1 - \mu_1^2)/\epsilon}\} \leq 1-\epsilon$, which contradicts $\mathbb{P}_\zeta\{\zeta \leq \mu_1 + \sqrt{(2-\epsilon)(\Sigma_1 - \mu_1^2)/\epsilon}\} \geq 1 - \epsilon/2$. Similarly, application of Cantelli's inequality gives us

$$\begin{aligned}
&\sup_{\mathbb{P}_\zeta \in \mathcal{D}_1} \mathbb{P}_\zeta \left\{ \zeta \leq \mu_1 - \sqrt{\left(\frac{1+\epsilon}{1-\epsilon}\right)(\Sigma_1 - \mu_1^2)} \right\} \\
&\quad \leq \sup_{\mathbb{P}_\zeta \in \mathcal{D}_1^\infty} \mathbb{P}_\zeta \left\{ \zeta \leq \mu_1 - \sqrt{\left(\frac{1+\epsilon}{1-\epsilon}\right)(\Sigma_1 - \mu_1^2)} \right\} \leq 1 - \frac{1+\epsilon}{2}.
\end{aligned}$$

Hence, $\mathrm{VaR}_{\mathbb{P}_\zeta}^{\epsilon}(\zeta) \geq \mu_1 - \sqrt{(1+\epsilon)(\Sigma_1 - \mu_1^2)/(1-\epsilon)}$ for all $\mathbb{P}_\zeta \in \mathcal{D}_1$ because otherwise there exists a $\mathbb{P}_\zeta \in \mathcal{D}_1$ such that $\mathbb{P}_\zeta\{\zeta \leq \mu_1 - \sqrt{(1+\epsilon)(\Sigma_1 - \mu_1^2)/(1-\epsilon)}\} \geq 1-\epsilon$, which contradicts $\mathbb{P}_\zeta\{\zeta \leq \mu_1 - \sqrt{(1+\epsilon)(\Sigma_1 - \mu_1^2)/(1-\epsilon)}\} \leq 1-(1+\epsilon)/2$.

Second, based on the representation $\zeta = U^{1/\alpha} Z$ (see Theorem 3.5 in [11]), we obtain that $\mathbb{E}_{\mathbb{P}_Z}[Z] = (\frac{\alpha+1}{\alpha})\mathbb{E}_{\mathbb{P}_\zeta}[\zeta] = \mu_0$, $\mathbb{E}_{\mathbb{P}_Z}[Z^2] = (\frac{\alpha+2}{\alpha})\mathbb{E}_{\mathbb{P}_\zeta}[\zeta^2] = \Sigma_0$, and

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ &= \mathbb{E}_{\mathbb{P}_Z}[U^{1/\alpha} Z - \beta]_+ \\
&= \int_{z=-\infty}^{+\infty} \int_{u=0}^{1} \left[u^{1/\alpha} z - \beta\right]_+ du \, d\mathbb{P}_Z(z).
\end{aligned}$$

It follows that, when $\beta < 0$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_\zeta}[\zeta - \beta]_+ &= \int_{z=-\infty}^{\beta} \int_{u=0}^{(\beta/z)^\alpha} \left(u^{1/\alpha} z - \beta\right) du \, d\mathbb{P}_Z(z) \\
&\quad + \int_{z=\beta}^{+\infty} \int_{u=0}^{1} \left(u^{1/\alpha} z - \beta\right) du \, d\mathbb{P}_Z(z) \\
&= \int_{z=-\infty}^{\beta} \left(-\frac{1}{\alpha+1}\right)\left(\frac{\beta^{\alpha+1}}{z^\alpha}\right) d\mathbb{P}_Z(z)
\end{aligned}$$

$$+\int_{z=\beta}^{+\infty}\left[\left(\frac{\alpha}{\alpha+1}\right)z-\beta\right]d\mathbb{P}_Z(z)$$
$$=\mathbb{E}_{\mathbb{P}_Z}[f_-(Z)],$$

and, when $\beta \geq 0$,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_\zeta}[\zeta-\beta]_+ &= \int_{z=\beta}^{+\infty}\int_{u=(\beta/z)^\alpha}^{1}\left(u^{1/\alpha}z-\beta\right)du\, d\mathbb{P}_Z(z)\\ &= \int_{z=\beta}^{+\infty}\left[\left(\frac{\alpha}{\alpha+1}\right)z-\beta+\left(\frac{1}{\alpha+1}\right)\left(\frac{\beta^{\alpha+1}}{z^\alpha}\right)\right]d\mathbb{P}_Z(z)\\ &= \mathbb{E}_{\mathbb{P}_Z}[f_+(Z)].\end{aligned}$$

□

Proposition 7 indicates that computing $\sup_{\mathbb{P}_\zeta\in\mathcal{D}_1}\text{CVaR}^\epsilon_{\mathbb{P}_\zeta}(\zeta)$ can be difficult because it needs to evaluate the worst-case expectation of a nonlinear function $f(z)$, i.e., $\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)}\mathbb{E}_{\mathbb{P}_Z}[f(Z)]$. To obtain a computable form, we first present two structural properties of $f(z)$. Lemma 1 proposes two approximations of $f(z)$ from above (termed $f_{\text{U}}(z)$) and below (termed $f_{\text{L}}(z)$), respectively. Both $f_{\text{U}}(z)$ and $f_{\text{L}}(z)$ are convex and consist of two linear pieces. Furthermore, Lemma 2 represents convex functions $f_+(z)$ and $f_-(z)$ by the supporting hyperplanes of their epigraphs.

Lemma 1 *Define* $f_U(z) = \left(\frac{\alpha}{\alpha+1}\right)(z-\beta)_+ + \left(\frac{1}{\alpha+1}\right)(-\beta)_+$ *and* $f_L(z) = \left[\left(\frac{\alpha}{\alpha+1}\right)z-\beta\right]_+$. *Then,* $f_L(z)\leq f(z)\leq f_U(z)$ *for all* $z\in\mathbb{R}$.

Proof First, we prove $f_{\text{L}}(z)\leq f(z)$ by discussing the following four cases:

1. If $z<\beta<0$, then $0\leq(\beta/z)\leq 1$ and $(-\beta)\geq 0$. It follows that $f(z) = -(\beta/(\alpha+1))\,(\beta/z)^\alpha \geq 0$. Additionally, define $H(z) := -(\beta/(\alpha+1))\,(\beta/z)^\alpha$ and then $H(z)$ is a convex function of z on interval $(-\infty,\beta]$. It follows that $H(z)\geq H'(\beta)(z-\beta)+H(\beta)$, i.e.,
$$-\left(\frac{\beta}{\alpha+1}\right)\left(\frac{\beta}{z}\right)^\alpha \geq \left(\frac{\alpha}{\alpha+1}\right)(z-\beta)+\left(-\frac{\beta}{\alpha+1}\right) = \left(\frac{\alpha}{\alpha+1}\right)z-\beta,$$
where the inequality is because $H'(z) = (\alpha/(\alpha+1))(\beta/z)^{\alpha+1}$ and $H(\beta) = (-\beta/(\alpha+1))$. Hence, $-(\beta/(\alpha+1))\,(\beta/z)^\alpha \geq [(\frac{\alpha}{\alpha+1})z-\beta]_+$, i.e., $f(z)\geq f_{\text{L}}(z)$.
2. If $\beta<0$ and $z\geq\beta$, then $(\frac{\alpha}{\alpha+1})z-\beta\geq 0$. It follows that $f_{\text{L}}(z) = (\frac{\alpha}{\alpha+1})z-\beta = f(z)$.
3. If $\beta\geq 0$ and $z\leq\beta$, then $(\frac{\alpha}{\alpha+1})z-\beta<0$. It follows that $f_{\text{L}}(z)=0=f(z)$.

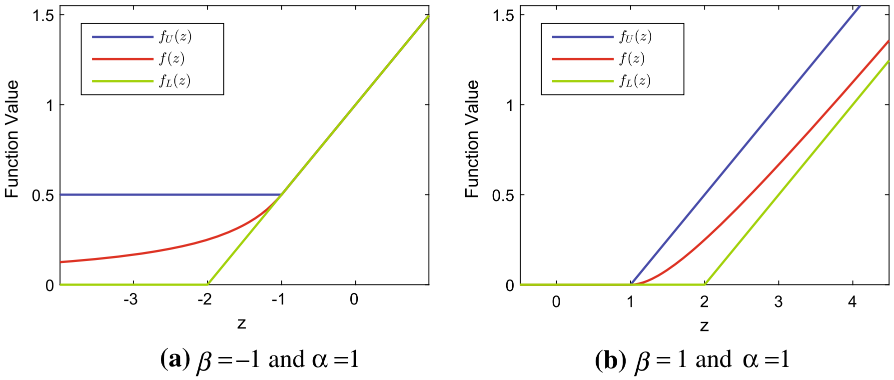


Fig. 1 Examples of function $f(z)$ and its approximations $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$

4. If $z > \beta \geq 0$, then $(\beta/z) \geq 0$. It follows that $f(z) = (\alpha/(\alpha+1))z - \beta + (\beta/(\alpha+1))(\beta/z)^{\alpha} \geq (\frac{\alpha}{\alpha+1})z - \beta$. Additionally, as $-z < -\beta \leq 0$, from *Case* 1 we have

$$-\left(\frac{-\beta}{\alpha+1}\right)\left(\frac{-\beta}{-z}\right)^{\alpha} \geq \left(\frac{\alpha}{\alpha+1}\right)(-z) - (-\beta).$$

In other words, $(\alpha/(\alpha+1))z - \beta + (\beta/(\alpha+1))(\beta/z)^{\alpha} \geq 0$. Hence, $(\alpha/(\alpha+1))z - \beta + (\beta/(\alpha+1))(\beta/z)^{\alpha} \geq [(\frac{\alpha}{\alpha+1})z - \beta]_{+}$, i.e., $f(z) \geq f_{\mathrm{L}}(z)$.

Second, we prove $f(z) \leq f_{\mathrm{U}}(z)$ by discussing the following four cases:

1. If $z < \beta < 0$, then $0 \leq (\beta/z) \leq 1$ and $(-\beta) \geq 0$. It follows that $f(z) = -(\beta/(\alpha+1))(\beta/z)^{\alpha} \leq (\frac{1}{\alpha+1})(-\beta) \leq f_{\mathrm{U}}(z)$.
2. If $\beta < 0$ and $z \geq \beta$, then $(z-\beta)_{+} = z - \beta$ and $(-\beta)_{+} = -\beta$. It follows that $f_{\mathrm{U}}(z) = (\frac{\alpha}{\alpha+1})(z-\beta)_{+} + (\frac{1}{\alpha+1})(-\beta)_{+} = (\frac{\alpha}{\alpha+1})z - \beta = f(z)$.
3. If $\beta \geq 0$ and $z \leq \beta$, then $f(z) = 0 \leq f_{\mathrm{U}}(z)$.
4. If $z > \beta \geq 0$, then $0 \leq (\beta/z) < 1$ and $(z-\beta)_{+} = z - \beta$. It follows that $f(z) = (\alpha/(\alpha+1))z - \beta + (\beta/(\alpha+1))(\beta/z)^{\alpha} \leq (\alpha/(\alpha+1))z - \beta + \beta/(\alpha+1) = f_{\mathrm{U}}(z)$.

□

Figure 1a, b present examples of function $f(z)$ and its approximations $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$.

Lemma 2 *The following two equalities hold:*

$$f_{+}(z) = \sup_{k\geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})z - (1-k^{-\alpha})\beta\right\} \tag{23a}$$

when $\beta \geq 0$, *and*

$$f_{-}(z) = \sup_{k\geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)k^{-\alpha-1}z - k^{-\alpha}\beta\right\}. \tag{23b}$$

when $\beta \le 0$. *Furthermore,* $f_-(z) = f_L(z) \le f_+(z)$ *for all* $z \in \mathbb{R}$ *when* $\beta \ge 0$ *and* $f_+(z) = f_L(z) \le f_-(z)$ *for all* $z \in \mathbb{R}$ *when* $\beta \le 0$.

Proof First, we suppose that $\beta \ge 0$ and pick a $z_0 \ge \beta$. The first derivative of $f_+(z)$ at z_0 is $f'_+(z)\big|_{z=z_0} = (\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]$. It follows that the supporting hyperplane of the epigraph $\{(y,z) \in \mathbb{R}^2 : y \ge f_+(z)\}$ at z_0 is $y \ge (\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]z - \big[1-(\frac{\beta}{z_0})^{\alpha}\big]\beta$. Hence, $f_+(z) = \sup_{z_0 \ge \beta}\big\{(\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]z - \big[1-(\frac{\beta}{z_0})^{\alpha}\big]\beta\big\}$ for all $z \ge \beta$ because $f_+(z)$ is convex. Furthermore, as $f_+(z) = 0$ when $z \le \beta$ and $(\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]z - \big[1-(\frac{\beta}{z_0})^{\alpha}\big]\beta = 0$ when $z_0 = \beta$, we have $f_+(z) = \sup_{z_0 \ge \beta}\big\{(\frac{\alpha}{\alpha+1})\big[1-(\frac{\beta}{z_0})^{\alpha+1}\big]z - \big[1-(\frac{\beta}{z_0})^{\alpha}\big]\beta\big\}$ for all $z \in \mathbb{R}$. Rewriting $z_0 = k\beta$ for $k \ge 1$ leads to representation (23a). The proof of representation (23b) is similar and so omitted.

Second, we suppose that $\beta \ge 0$ and define $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})z - (1-k^{-\alpha})\beta$ for all $k \ge 1$. Then, $f_+(z) = \sup_{k \ge 1}\{f_+^k(z)\}$ and $f_-(z) = \sup_{k\ge1}\{(\frac{\alpha}{\alpha+1})z - \beta - f_+^k(z)\} = (\frac{\alpha}{\alpha+1})z - \beta - \inf_{k\ge1}\{f_+^k(z)\}$. We prove that $\inf_{k\ge1}\{f_+^k(z)\} = -\big[(\frac{\alpha}{\alpha+1})z - \beta\big]_-$ by discussing the following two cases:

1. When $z \le (\frac{\alpha+1}{\alpha})\beta$, we have $z \le (\frac{\alpha+1}{\alpha})k\beta$ as $k \ge 1$ and $\beta \ge 0$. It follows that $(\frac{\alpha}{\alpha+1})(-k^{-\alpha-1})z + k^{-\alpha}\beta \ge 0$ and so $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})z - (1-k^{-\alpha})\beta \ge (\frac{\alpha}{\alpha+1})z - \beta$ for all $k \ge 1$. Hence, $\inf_{k\ge1}\{f_+^k(z)\} \ge (\frac{\alpha}{\alpha+1})z - \beta$. In addition, by letting $k \to +\infty$, we have $f_+^k(z) \to (\frac{\alpha}{\alpha+1})z - \beta$. Therefore, $\inf_{k\ge1}\{f_+^k(z)\} = (\frac{\alpha}{\alpha+1})z - \beta$ when $z \le (\frac{\alpha+1}{\alpha})\beta$.
2. When $z \ge (\frac{\alpha+1}{\alpha})\beta$, we have $(1-k^{-\alpha-1})z \ge (1-k^{-\alpha})(\frac{\alpha+1}{\alpha})\beta$ because $\beta \ge 0$ and $1-k^{-\alpha-1} \ge 1-k^{-\alpha} \ge 0$. It follows that $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})z - (1-k^{-\alpha})\beta \ge 0$ for all $k \ge 1$. Hence, $\inf_{k\ge1}\{f_+^k(z)\} \ge 0$. In addition, by letting $k = 1$, we have $f_+^k(z) = 0$. Therefore, $\inf_{k\ge1}\{f_+^k(z)\} = 0$ when $z \le (\frac{\alpha+1}{\alpha})\beta$.

It follows that $f_-(z) = (\frac{\alpha}{\alpha+1})z - \beta + \big[(\frac{\alpha}{\alpha+1})z - \beta\big]_- = \big[(\frac{\alpha}{\alpha+1})z - \beta\big]_+$. Hence, by Lemma 1, $f_-(z) = f_{\mathrm{L}}(z) \le f_+(z)$ for all $z \in \mathbb{R}$ when $\beta \ge 0$. The proof of $f_+(z) = f_{\mathrm{L}}(z) \le f_-(z)$ when $\beta \le 0$ is similar and so omitted. □

We are now ready to derive a reformulation of the worst-case expectation $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)]$. We summarize this result in the following theorem.

Theorem 2 *For* $\beta \in \mathbb{R}$, $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] = \frac{1}{2}\max\{E_+, E_-\}$, *where*

$$S_{k,\mu_0,\Sigma_0,\beta} = \sqrt{\left[(1-k^{-\alpha})\beta - \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\mu_0\right]^2 + \left(\frac{\alpha}{\alpha+1}\right)^2(1-k^{-\alpha-1})^2(\Sigma_0 - \mu_0^2)},$$

$$E_+ = \sup_{k\ge1}\left\{S_{k,\mu_0,\Sigma_0,\beta} - (1-k^{-\alpha})\beta + \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\mu_0\right\}, \quad \textit{and} \tag{24a}$$

$$E_- = \sup_{k\ge1}\left\{S_{k,\mu_0,\Sigma_0,\beta} - (1+k^{-\alpha})\beta + \left(\frac{\alpha}{\alpha+1}\right)(1+k^{-\alpha-1})\mu_0\right\}. \tag{24b}$$

Proof To avoid clutter, throughout this proof, we assume that $\Sigma_0 > \mu_0^2$ and $\beta \ne 0$. The degenerate cases with $\Sigma_0 = \mu_0^2$ or $\beta = 0$ can be easily verified. First, we suppose that $\beta > 0$ and define $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1 - k^{-\alpha-1})z - (1 - k^{-\alpha})\beta$ for $k \geq 1$. Then, $f(Z) = f_+(Z)$ by Proposition 7 and $f_+(z) = \sup_{k\geq 1}\{f_+^k(z)\}$ by Lemma 2. It follows that $\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] = \sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}]$. We make the following observation on switching the order of two supremum operators.

Observation 1 *For $\beta \in \mathbb{R}$, we have*

$$\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\sup_{k\geq 1}\left\{f_+^k(Z)\right\}\right] = \sup_{k\geq 1}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[f_+^k(Z)\right]_+\right\}.$$

Proof of Observation 1 First, for all $k \geq 1$, it is clear that $\sup_{k\geq 1}\{f_+^k(Z)\} \geq [f_+^k(Z)]_+$ because $\sup_{k\geq 1}\{f_+^k(z)\} = f_+(z) \geq 0$ for all $z \in \mathbb{R}$. It follows that $\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}] \geq \sup_{k\geq 1}\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\}$. We now show the opposite, i.e., $\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}] \leq \sup_{k\geq 1}\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\}$. When $\beta \leq 0$, this holds because

$$\begin{aligned}
\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\sup_{k\geq 1}\left\{f_+^k(Z)\right\}\right] &= \sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[f_{\mathrm{L}}(Z)\right] \\
&= \lim_{k\to\infty}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[f_+^k(Z)\right]_+\right\} \\
&\leq \sup_{k\geq 1}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[f_+^k(Z)\right]_+\right\},
\end{aligned}$$

where the first equality follows from Lemma 2. To prove the second equality, we make the following observation on the monotonicity of function $[f_+^k(z)]_+$ in k and relegate the proof to Appendix A.

Observation 2 $[f_+^{k+1}(z)]_+ \geq [f_+^k(z)]_+$ *for all $z \in \mathbb{R}$ and $k \geq 1$.*

By Observation 2, $f_{\mathrm{L}}(z) = \lim_{k\to\infty}[f_+^k(z)]_+$ for all $z \in \mathbb{R}$. It follows that, for any $\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)$,

$$\mathbb{E}_{\mathbb{P}_Z}\left[f_{\mathrm{L}}(Z)\right] = \mathbb{E}_{\mathbb{P}_Z}\left[\lim_{k\to\infty}[f_+^k(Z)]_+\right] = \lim_{k\to\infty}\mathbb{E}_{\mathbb{P}_Z}\left[[f_+^k(Z)]_+\right],$$

where the second equality follows from the monotone convergence theorem. Hence,

$$\mathbb{E}_{\mathbb{P}_Z}\left[f_{\mathrm{L}}(Z)\right] \leq \lim_{k\to\infty}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\right\}.$$

As this inequality holds for all $\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)$, we have

$$\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] \leq \lim_{k\to\infty}\left\{\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\right\}.$$

On the other hand, as $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] \geq \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+$ for all $k \geq 1$, we have

$$\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] \geq \lim_{k \to \infty} \left\{ \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+ \right\},$$

which proves the second equality. Hence, we focus on the case when $\beta > 0$ in the remainder of this proof.

Second, we present $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k \geq 1}\{f_+^k(Z)\}]$ as the following optimization problem:

$$\begin{aligned} (\mathrm{P}): \quad v_P \;&=\; \max_{\mathbb{P}_Z} \mathbb{E}_{\mathbb{P}_Z}[f_+(Z)] \\ &\quad \text{s.t. } \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \\ &\qquad\;\; \mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0, \\ &\qquad\;\; \mathbb{E}_{\mathbb{P}_Z}[1] = 1, \end{aligned}$$

whose dual is

$$\begin{aligned} (\mathrm{D}): \quad v_D \;&=\; \min_{p,q,r} \mu_0 p + \Sigma_0 q + r \\ &\quad \text{s.t. } qz^2 + pz + r \geq f_+(z), \quad \forall z \in \mathbb{R}. \end{aligned} \tag{25}$$

Strong duality holds between (P) and (D) due to Assumption 1 (see Proposition 3.4 in [31]), i.e., $v_P = v_D$. Furthermore, by Lemma 3.1 in [31], there exists a worst-case probability distribution (i.e., an optimal solution to (P)) with a finite support of at most 3 points. That is, there exists $m \in \{1, 2, 3\}$, $(z_1^*, \ldots, z_m^*) \in \mathbb{R}^m$, and $(\pi_1^*, \ldots, \pi_m^*) \in \mathbb{R}_+^m$ such that $\sum_{i=1}^m \pi_i^* z_i^* = \mu_0$, $\sum_{i=1}^m \pi_i^* (z_i^*)^2 = \Sigma_0$, and $\sum_{i=1}^m \pi_i^* = 1$. Denoting an optimal solution to (D) by (p^*, q^*, r^*), we claim that $q^*(z_i^*)^2 + p^* z_i^* + r^* = f_+(z_i^*)$ for all $i = 1, \ldots, m$, i.e., constraint (25) holds at equality at points $z_1^*, \ldots, z_m^*$. Indeed, if this claim fails to hold, then we have

$$v_P \;=\; \sum_{i=1}^m \pi_i^* f_+(z_i^*) \;<\; \sum_{i=1}^m \pi_i^* [q^*(z_i^*)^2 + p^* z_i^* + r^*] \;=\; q^* \Sigma_0 + p^* \mu_0 + r^* \;=\; v_D, \tag{26}$$

where the inequality follows from constraint (25), and the second equality follows from the definitions of $(z_1^*, \ldots, z_m^*)$ and $(\pi_1^*, \ldots, \pi_m^*)$. As inequality (26) violates the strong duality, the claim holds. In addition, it can be shown that $f_+(z)$ and any quadratic function $qz^2 + pz + r$ satisfying constraint (25) intersect at most once in interval $(-\infty, \beta]$ and at most once in interval $[\beta, \infty)$. It follows that $m \leq 2$, and so $m = 2$ because $\Sigma_0 > \mu_0^2$. Without loss of generality, we assume that $z_1^* \in (-\infty, \beta]$ and $z_2^* \in [\beta, \infty)$.

Third, we define $k^* = z_2^*/\beta$ and consider function $[f_+^{k^*}(z)]_+$ that is tangent to $f_+(z)$ at z_1^* and z_2^* by Lemma 2. Hence, $qz^2 + pz + r \geq [f_+^{k^*}(z)]_+$ for all $z \in \mathbb{R}$ with equality holding only at z_1^* and z_2^*. Consider the primal and dual formulations of $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^{k^*}(Z)]_+$ as follows:

$$
\begin{aligned}
(\mathrm{P}_{k^*}): \quad v_P^{k^*} &= \max_{\mathbb{P}_Z} \ \mathbb{E}_{\mathbb{P}_Z}[f_+^{k^*}(Z)]_+ \\
&\quad \text{s.t.} \quad \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \\
&\qquad\quad\ \ \mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0, \\
&\qquad\quad\ \ \mathbb{E}_{\mathbb{P}_Z}[1] = 1, \\
(\mathrm{D}_{k^*}): \quad v_D^{k^*} &= \min_{p,q,r} \ \mu_0 p + \Sigma_0 q + r \\
&\quad \text{s.t.} \ qz^2 + pz + r \geq [f_+^{k^*}(z)]_+, \quad \forall z \in \mathbb{R}.
\end{aligned}
$$

It is clear that the pair (z_1^*, z_2^*) and (π_1^*, π_2^*) provide a primal feasible solution to (P_{k^*}), and (p^*, q^*, r^*) is a dual feasible solution to (D_{k^*}) because $f_+(z) \geq [f_+^{k^*}(z)]_+$ for all $z \in \mathbb{R}$. Meanwhile, these two solutions share the same objective function value because $\sum_{i=1}^2 \pi_i^*[f_+^{k^*}(z_i^*)]_+ = \sum_{i=1}^2 \pi_i^* f_+(z_i^*) = \mu_0 p^* + \Sigma_0 q^* + r^*$, where the first equality follows from the definition of $[f_+^{k^*}(z)]_+$ and the second equality is due to $v_P = v_D$. It follows that strong duality holds between (P_{k^*}) and (D_{k^*}) and $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}] = \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^{k^*}(Z)]_+$. Therefore, $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[\sup_{k\geq 1}\{f_+^k(Z)\}] \leq \sup_{k\geq 1}\{\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\}$ and so the proof is completed. □

(Proof of Theorem 2 continued) By Observation 1, we have

$$
\begin{aligned}
&\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \\
&= \sup_{k\geq 1}\left\{\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})Z - (1-k^{-\alpha})\beta\right]_+\right\} \\
&= \sup_{k\geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1}) \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[Z - \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)\beta\right]_+\right\} \\
&= \sup_{k\geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\left(\frac{1}{2}\right)\left[\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)\beta - \mu_0\right]^2 + (\Sigma_0 - \mu_0^2)}\right.\right. \\
&\quad \left.\left. - \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)\beta + \mu_0\right]\right\} \qquad (27) \\
&= \frac{1}{2}E_+,
\end{aligned}
$$

where equality (**??**) follows from Observation 3 presented in Appendix B.

Second, we suppose that $\beta < 0$. Then, $f(Z) = f_-(Z)$ by Proposition 7. It follows that

$$
\begin{aligned}
&\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \\
&= \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\sup_{k\geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)k^{-\alpha-1}Z - k^{-\alpha}\beta\right\}\right] \qquad (28a)
\end{aligned}
$$

$$= \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\sup_{k \geq 1}\left\{\max\left\{\left(\frac{\alpha}{\alpha+1}\right)k^{-\alpha-1}Z - k^{-\alpha}\beta,\ \left(\frac{\alpha}{\alpha+1}\right)Z - \beta\right\}\right\}\right] \tag{28b}$$

$$= \sup_{k \geq 1}\left\{\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\max\left\{\left(\frac{\alpha}{\alpha+1}\right)k^{-\alpha-1}Z - k^{-\alpha}\beta,\ \left(\frac{\alpha}{\alpha+1}\right)Z - \beta\right\}\right]\right\} \tag{28c}$$

$$= \sup_{k \geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)k^{-\alpha-1}\mu_0 - k^{-\alpha}\beta + \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+^k(Z)]_+\right\} \tag{28d}$$

$$= \frac{1}{2}E_-, \tag{28e}$$

where equality (28a) follows from Lemma 2, equality (28b) is because $(\frac{\alpha}{\alpha+1})k^{-\alpha-1}z - k^{-\alpha}\beta = (\frac{\alpha}{\alpha+1})z - \beta$ when $k = 1$, equality (28c) is parallel to Observation 1 and can be similarly proved, and equality (28d) follows from the definition of $f_+^k(z)$.

Finally, it remains to prove that $E_+ \geq E_-$ when $\beta > 0$ and $E_+ \leq E_-$ when $\beta < 0$. Due to the similarity of proof, we only show the former case, i.e., when $\beta > 0$. To that end, we note that the equalities (28b)–(28e) are independent of the sign of β and so still hold when $\beta > 0$. It follows that $\frac{1}{2}E_- = \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_-(Z)]$ when $\beta > 0$. Similarly, we have $\frac{1}{2}E_+ = \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_+(Z)]$. But $f_-(z) \leq f_+(z)$ for all $z \in \mathbb{R}$ by Lemma 2, and so $\frac{1}{2}E_- \leq \frac{1}{2}E_+$ when $\beta > 0$. □

Theorem 2 leads to an equivalent reformulation of AVC (7). We summarize the main result of this section in the following theorem.

Theorem 3 *AVC (7) is equivalent to a set of SOC constraints*

$$\left\|\begin{bmatrix}(1-k^{-\alpha})\beta - (1-k^{-\alpha-1})\mu^\top a(x) \\ \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\Lambda a(x)\end{bmatrix}\right\| \leq 2\epsilon b(x) - (1-k^{-\alpha-1})\mu^\top a(x) + (1-k^{-\alpha}-2\epsilon)\beta, \tag{29a}$$

$$\left\|\begin{bmatrix}(1-k^{-\alpha})\beta - (1-k^{-\alpha-1})\mu^\top a(x) \\ \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\Lambda a(x)\end{bmatrix}\right\| \leq 2\epsilon b(x) - (1+k^{-\alpha-1})\mu^\top a(x) + (1+k^{-\alpha}-2\epsilon)\beta, \tag{29b}$$

for all $k \geq 1$.

Proof By Propositions 6–7, AVC (7) is equivalent to

$$\inf_{\beta \in \mathbb{R}}\left\{\beta + \frac{1}{\epsilon}\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)]\right\} \leq b(x).$$

Meanwhile, the proof of Proposition 7 shows that there exists a finite β that attains the above infimum. It follows that AVC (7) is satisfied if and only if there exists a $\beta \in \mathbb{R}$

such that $\beta + \frac{1}{\epsilon} \sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)] \leq b(x)$. Then, the conclusion follows from Theorem 2 by the definition of μ_0, Λ, and that

$$S_{k,\mu_0,\Sigma_0,\beta} = \left\| \begin{bmatrix} (1-k^{-\alpha})\beta - (1-k^{-\alpha-1})\mu^\top a(x) \\ \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\Lambda a(x) \end{bmatrix} \right\|, \quad \forall k \geq 1.$$

□

In computation, directly replacing AVC with constraints (29a)–(29b) requires an infinite number of SOC constraints and is so computationally intractable. Like what we described for ACC in Sect. 2, we adopt the separation approach and solve the following problem:

Separation Problem 2: Given $\hat{\beta}$ and $\hat{x}$, does there exist a $\hat{k}$ such that $(\hat{\beta}, \hat{x})$ violate constraints (29a)–(29b)?

In the following proposition, we show that Separation Problem 2 can be solved by conducting a golden section search on the real line. This search is computationally efficient.

Proposition 8 *Define* $\hat{\mu}_0 = (\frac{\alpha+1}{\alpha})\mu^\top a(\hat{x})$, $\hat{\Sigma}_0 = (\frac{\alpha+2}{\alpha})a(\hat{x})^\top \Sigma a(\hat{x})$. *We have the following:*

1. *If* $\hat{\beta} = 0$, *then* $(\hat{\beta}, \hat{x})$ *violate constraints* (29a)–(29b) *if and only if* $(\hat{\beta}, \hat{x})$ *violate them at* $\hat{k} = \infty$*;*
2. *If* $\hat{\beta} \neq 0$ *and* $\hat{\Sigma}_0 = \hat{\mu}_0^2$, *then* $(\hat{\beta}, \hat{x})$ *violate constraints* (29a)–(29b) *if and only if* $(\hat{\beta}, \hat{x})$ *violate them at* $\hat{k} = \max\{\hat{\mu}_0/\hat{\beta}, 1\}$*;*
3. *If* $\hat{\beta} \neq 0$ *and* $\hat{\Sigma}_0 > \hat{\mu}_0^2$, *then* $(\hat{\beta}, \hat{x})$ *violate constraints* (29a)–(29b) *if and only if* $(\hat{\beta}, \hat{x})$ *violate them at the unique root of equation*

$$2\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta\right] = (k-\mu_\beta) - \frac{\Gamma_\beta}{(k-\mu_\beta)} \tag{30}$$

lying within the interval $\left[1+\sqrt{(1-\mu_\beta)^2+\Gamma_\beta},\ 1+1/\alpha+\sqrt{(1-\mu_\beta+1/\alpha)^2+\Gamma_\beta}\right]$, *where* $\mu_\beta = \hat{\mu}_0/\hat{\beta}$ *and* $\Gamma_\beta = (\hat{\Sigma}_0 - \hat{\mu}_0^2)/\hat{\beta}^2$.

Proof For a given $(\hat{\beta}, \hat{x})$, solving Separation Problem 2 is equivalent to finding $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f(Z)]$, i.e., $1/2 \max\{E_+, E_-\}$ defined in Theorem 2. First, if $\hat{\beta} = 0$, then

$$\begin{aligned} S_{k,\hat{\mu}_0,\hat{\Sigma}_0,\hat{\beta}} &= \sqrt{\left[\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\hat{\mu}_0\right]^2 + \left(\frac{\alpha}{\alpha+1}\right)^2(1-k^{-\alpha-1})^2\left(\hat{\Sigma}_0 - \hat{\mu}_0^2\right)} \\ &= \left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\sqrt{\hat{\Sigma}_0}. \end{aligned}$$

It follows that

$$\begin{aligned}\frac{1}{2}E_+ &= \frac{1}{2}\sup_{k\geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\sqrt{\hat{\Sigma}_0}+\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\hat{\mu}_0\right\}\\ &= \frac{1}{2}\sup_{k\geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\left(\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0\right)\right\} \qquad (31a)\\ &= \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)\left(\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0\right), \qquad (31b)\end{aligned}$$

where equality (31b) is because $\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0 \geq 0$ and so $k=\infty$ maximizes (31a). Additionally,

$$\begin{aligned}\frac{1}{2}E_- &= \frac{1}{2}\sup_{k\geq 1}\left\{\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\sqrt{\hat{\Sigma}_0}+\left(\frac{\alpha}{\alpha+1}\right)(1+k^{-\alpha-1})\hat{\mu}_0\right\}\\ &= \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)\sup_{k\geq 1}\left\{\left(\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0\right)+k^{-\alpha-1}\left(\hat{\mu}_0-\sqrt{\hat{\Sigma}_0}\right)\right\} \qquad (31c)\\ &= \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)\left(\sqrt{\hat{\Sigma}_0}+\hat{\mu}_0\right), \qquad (31d)\end{aligned}$$

where equality (31d) is because $\hat{\mu}_0-\sqrt{\hat{\Sigma}_0}\leq 0$ and so $k=\infty$ maximizes (31c). Summing up the above two cases, we have $\hat{k}=\infty$ if $\hat{\beta}=0$.

Second, if $\hat{\beta}\neq 0$ and $\hat{\Sigma}_0=\hat{\mu}_0^2$, then $S_{k,\hat{\mu}_0,\hat{\Sigma}_0,\hat{\beta}}=|(1-k^{-\alpha})\hat{\beta}-(\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})\hat{\mu}_0|$. It follows that

$$\begin{aligned}\frac{1}{2}E_+ &= \frac{1}{2}\sup_{k\geq 1}\left\{\left|(1-k^{-\alpha})\hat{\beta}-\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\hat{\mu}_0\right|\right.\\ &\qquad \left.-(1-k^{-\alpha})\hat{\beta}+\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\hat{\mu}_0\right\}\\ &= \sup_{k\geq 1}\left\{\left[\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\hat{\mu}_0-(1-k^{-\alpha})\hat{\beta}\right]_+\right\} \qquad (32a)\\ &= f_+(\hat{\mu}_0), \qquad (32b)\end{aligned}$$

where equality (32b) results from Lemma 2 and so $k=\max\{\hat{\mu}_0/\hat{\beta},1\}$ maximizes (32a). Meanwhile,

$$\begin{aligned}\frac{1}{2}E_- &= \frac{1}{2}\sup_{k\geq 1}\left\{\left|(1-k^{-\alpha})\hat{\beta}-\left(\frac{\alpha}{\alpha+1}\right)(1-k^{-\alpha-1})\hat{\mu}_0\right|\right.\\ &\qquad \left.-(1+k^{-\alpha})\hat{\beta}+\left(\frac{\alpha}{\alpha+1}\right)(1+k^{-\alpha-1})\hat{\mu}_0\right\}\end{aligned}$$

$$= \sup_{k\geq 1}\left\{\max\left\{\left(\frac{\alpha}{\alpha+1}\right)\hat{\mu}_0 - \hat{\beta},\ \left(\frac{\alpha}{\alpha+1}\right)k^{-\alpha-1}\hat{\mu}_0 - k^{-\alpha}\hat{\beta}\right\}\right\} \tag{32c}$$

$$= f_-(\hat{\mu}_0), \tag{32d}$$

where equality (32d) results from Lemma 2 and so $k = \max\{\hat{\mu}_0/\hat{\beta}, 1\}$ maximizes (32c). Summing up the above two cases, we have $\hat{k} = \max\{\hat{\mu}_0/\hat{\beta}, 1\}$ if $\hat{\beta} \neq 0$ and $\hat{\Sigma}_0 = \hat{\mu}_0^2$.

Third, suppose that $\hat{\beta} \neq 0$ and $\hat{\Sigma}_0 > \hat{\mu}_0^2$. As the case when $\hat{\beta} < 0$ can be similarly derived, we focus on the case when $\hat{\beta} > 0$. In this case, solving Separation Problem 2 is equivalent to finding the maximizer of optimization problem (24a) that defines E_+. To this end, we let $F(k)$ represent the objective function of (24a), i.e., $F(k) := S_{k,\hat{\mu}_0,\hat{\Sigma}_0,\hat{\beta}} - (1-k^{-\alpha})\hat{\beta} + (\frac{\alpha}{\alpha+1})(1-k^{-\alpha-1})\hat{\mu}_0$. It follows that

$$F'(k) = \alpha\hat{\beta}k^{-\alpha-2}\left\{\frac{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta\right](k-\mu_\beta) + \Gamma_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta\right]^2 + \Gamma_\beta}} - (k-\mu_\beta)\right\}.$$

We prove that $F(k)$ is unimodal, and in particular, $F(k)$ is nondecreasing on $[1, \hat{k}]$ and nonincreasing on $[\hat{k}, \infty)$, where $\hat{k}$ represents the root of Eq. (30). The conclusion of this proposition then follows because $\hat{k}$ is the maximizer of $F(k)$ on $[1, \infty)$. To that end, it suffices to show that (i) $\lim_{k\to 1^+} F'(k) > 0$, (ii) there exists a $k \in [1, \infty)$ such that $F'(k) < 0$, and (iii) $\hat{k}$ is the unique root of equation $F'(k) = 0$. We show (i)–(iii) as follows.

(i) As $\lim_{k\to 1^+}\{\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\} = \frac{\alpha}{\alpha+1}$ and $\Gamma_\beta > 0$, we have $\lim_{k\to 1^+} F'(k) = \alpha\hat{\beta}\left[\sqrt{(1-\mu_\beta)^2 + \Gamma_\beta} - (1-\mu_\beta)\right] > 0$.

(ii) We have

$$\frac{F'(k)}{\alpha\hat{\beta}k^{-\alpha-2}} = \left(\frac{\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta\right]^2 + \Gamma_\beta}} - 1\right)(k-\mu_\beta) + \frac{\Gamma_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta\right]^2 + \Gamma_\beta}}.$$

As $\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}} \in \left[\frac{\alpha}{\alpha+1}, 1\right]$ and

$$\frac{\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta\right]^2 + \Gamma_\beta}} - 1 < 0,$$

there exists a sufficiently large k such that $F'(k) < 0$.

(iii) We consider the roots of equation $F'(k) = 0$. As $F'(k) = 0$ is equivalent to

$$\left(1 - \frac{\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta\right]^2 + \Gamma_\beta}}\right)(k - \mu_\beta)$$
$$= \frac{\Gamma_\beta}{\sqrt{\left[\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right) - \mu_\beta\right]^2 + \Gamma_\beta}},$$

any root k satisfies $k - \mu_\beta > 0$ because $\Gamma_\beta > 0$. The above equation can be further simplified to Eq. (30) and so any roots k of Eq. (30) also satisfy $F'(k) = 0$. We now prove the uniqueness of the root. We note that the first derivative of $2\big[(\frac{\alpha+1}{\alpha})(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}) - \mu_\beta\big]$, i.e., the left-hand side of Eq. (30), is always less than 1. To see this, we take the first derivative and denote it

$$Q(k) := 2\Big(\frac{\alpha+1}{\alpha}\Big)\frac{k^{-2\alpha-2} + \alpha k^{-\alpha-1} - (\alpha+1)k^{-\alpha-2}}{(1 - k^{-\alpha-1})^2}.$$

Through basic algebraic manipulations, it follows that $Q(k) \le 1$ if and only if $\overline{Q}(k) := (\alpha+2)k^{-2\alpha-2} + (2\alpha^2 + 4\alpha)k^{-\alpha-1} - (2\alpha^2 + 4\alpha + 2)k^{-\alpha-2} - \alpha \le 0$. As $\overline{Q}(1) = 0$, it suffices to show that $\overline{Q}'(k) \le 0$ for all $k \ge 1$. Noting that $\overline{Q}'(k) = 2(\alpha+1)(\alpha+2)k^{-2\alpha-3}[(\alpha+1)k^\alpha - \alpha k^{\alpha+1} - 1]$, we need to show that $\hat{Q}(k) := (\alpha+1)k^\alpha - \alpha k^{\alpha+1} - 1 \le 0$ for all $k \ge 1$, which holds because $\hat{Q}(1) = 0$ and $\hat{Q}'(k) = \alpha(\alpha+1)k^{\alpha-1}(1-k) \le 0$. Meanwhile, the first derivative of $(k - \mu_\beta) - \frac{\Gamma_\beta}{(k-\mu_\beta)}$, i.e., the right-hand side of Eq. (30), is always greater than 1. Furthermore, $2\big[(\frac{\alpha+1}{\alpha})(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}) - \mu_\beta\big] \in [2(1-\mu_\beta), 2(\frac{\alpha+1}{\alpha} - \mu_\beta)]$, while the range of function $(k - \mu_\beta) - \frac{\Gamma_\beta}{(k-\mu_\beta)}$ is $(-\infty, \infty)$ for $k \in (\mu_\beta, \infty)$. It follows that the two sides of Eq. (30) can meet only once, i.e., this equation has a unique root.

Finally, we provide lower and upper bounds of root $\hat{k}$. As $\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}} \in \big[\frac{\alpha}{\alpha+1}, 1\big]$, we have $2(1-\mu_\beta) \le (\hat{k} - \mu_\beta) - \frac{\Gamma_\beta}{(\hat{k}-\mu_\beta)} \le 2(\frac{\alpha+1}{\alpha} - \mu_\beta)$. It follows that $\hat{k} \in \Big[1 + \sqrt{(1-\mu_\beta)^2 + \Gamma_\beta},\ 1 + 1/\alpha + \sqrt{(1 - \mu_\beta + 1/\alpha)^2 + \Gamma_\beta}\Big]$.

□

3.1 Approximations of the ambiguous CVaR constraint

Before closing this section, we derive approximations of AVC (7). First, in the following proposition, we present a conservative approximation based on $f_U(z)$ and a relaxed one based on $f_L(z)$, both of which are in the form of SOC constraints.

Proposition 9 *AVC (7) is implied by SOC constraints*

$$\left\|\begin{bmatrix}\beta - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x) \\ \Lambda a(x)\end{bmatrix}\right\| \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) - \left[\frac{2\epsilon(\alpha+1)}{\alpha} - 1\right]\beta - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x), \tag{33a}$$

$$\left\|\begin{bmatrix}\beta - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x) \\ \Lambda a(x)\end{bmatrix}\right\| \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) - \left[\frac{(2\epsilon-1)(\alpha+1)-1}{\alpha}\right]\beta - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x). \tag{33b}$$

Furthermore, AVC (7) implies SOC constraint

$$\left\|\begin{bmatrix}\left(\frac{\alpha+1}{\alpha}\right)\beta - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x) \\ \Lambda a(x)\end{bmatrix}\right\| \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) - \left[\frac{(2\epsilon-1)(\alpha+1)}{\alpha}\right]\beta - \left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x). \tag{33c}$$

Proof First, based on Propositions 6–7 and Lemma 1, AVC (7) is implied by constraint $\beta + \frac{1}{\epsilon}\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{U}}(Z)] \leq b(x)$. Furthermore, we have

$$\begin{aligned}
\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{U}}(Z)] &= \sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}\left[\left(\frac{\alpha}{\alpha+1}\right)[Z-\beta]_+ + \left(-\frac{\beta}{\alpha+1}\right)_+\right] \\
&= \left(-\frac{\beta}{\alpha+1}\right)_+ + \left(\frac{\alpha}{\alpha+1}\right)\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+ \\
&= \left(-\frac{\beta}{\alpha+1}\right)_+ + \left(\frac{\alpha}{\alpha+1}\right)\left(\frac{1}{2}\right) \\
&\quad \left[\sqrt{(\beta-\mu_0)^2 + (\Sigma_0 - \mu_0^2)} - \beta + \mu_0\right],
\end{aligned}$$

where the last equality is due to Observation 3 presented in Appendix B. It follows that AVC (7) is implied by

$$\begin{aligned}
&\beta + \left(\frac{1}{\epsilon}\right)\left\{\left(-\frac{\beta}{\alpha+1}\right)_+ + \left(\frac{\alpha}{\alpha+1}\right)\left(\frac{1}{2}\right)\right. \\
&\qquad \left.\left[\sqrt{(\beta-\mu_0)^2 + (\Sigma_0 - \mu_0^2)} - \beta + \mu_0\right]\right\} \leq b(x) \\
\Leftrightarrow\ &\sqrt{(\beta-\mu_0)^2 + (\Sigma_0 - \mu_0^2)} \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha}\right] b(x) \\
&\quad - \left[\frac{2\epsilon(\alpha+1)}{\alpha} - 1\right]\beta - \left(\frac{2}{\alpha}\right)(-\beta)_+ - \mu_0.
\end{aligned}$$

This is equivalent to constraints (33a)–(33b) by the definition of μ_0 and observing that

$$\sqrt{(\beta-\mu_0)^2+(\Sigma_0-\mu_0^2)} \;=\; \left\|\begin{bmatrix}\beta-\left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x)\\ \Lambda a(x)\end{bmatrix}\right\|.$$

Second, based on Propositions 6–7 and Lemma 1, AVC (7) implies constraint $\beta+\frac{1}{\epsilon}\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)}\mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)]\le b(x)$. Furthermore, we have

$$\begin{aligned}
\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)}\mathbb{E}_{\mathbb{P}_Z}[f_{\mathrm{L}}(Z)] &= \sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)}\mathbb{E}_{\mathbb{P}_Z}\left[\left(\frac{\alpha}{\alpha+1}\right)Z-\beta\right]_+\\
&= \left(\frac{\alpha}{\alpha+1}\right)\sup_{\mathbb{P}_Z\in\mathcal{D}(\mu_0,\Sigma_0)}\mathbb{E}_{\mathbb{P}_Z}\left[Z-\left(\frac{\alpha+1}{\alpha}\right)\beta\right]_+\\
&= \left(\frac{\alpha}{\alpha+1}\right)\left(\frac{1}{2}\right)\\
&\quad\left[\sqrt{\left(\left(\frac{\alpha+1}{\alpha}\right)\beta-\mu_0\right)^2+(\Sigma_0-\mu_0^2)}-\left(\frac{\alpha+1}{\alpha}\right)\beta+\mu_0\right],
\end{aligned}$$

where the last equality is due to Observation 3. It follows that AVC (7) implies

$$\begin{aligned}
&\beta+\left(\frac{1}{\epsilon}\right)\left(\frac{\alpha}{\alpha+1}\right)\\
&\quad\left(\frac{1}{2}\right)\left[\sqrt{\left(\left(\frac{\alpha+1}{\alpha}\right)\beta-\mu_0\right)^2+(\Sigma_0-\mu_0^2)}-\left(\frac{\alpha+1}{\alpha}\right)\beta+\mu_0\right]\le b(x)\\
&\Leftrightarrow\ \sqrt{\left(\left(\frac{\alpha+1}{\alpha}\right)\beta-\mu_0\right)^2+(\Sigma_0-\mu_0^2)}\le\left[\frac{2\epsilon(\alpha+1)}{\alpha}\right]b(x)\\
&\quad-\left[\frac{(2\epsilon-1)(\alpha+1)}{\alpha}\right]\beta-\mu_0.
\end{aligned}$$

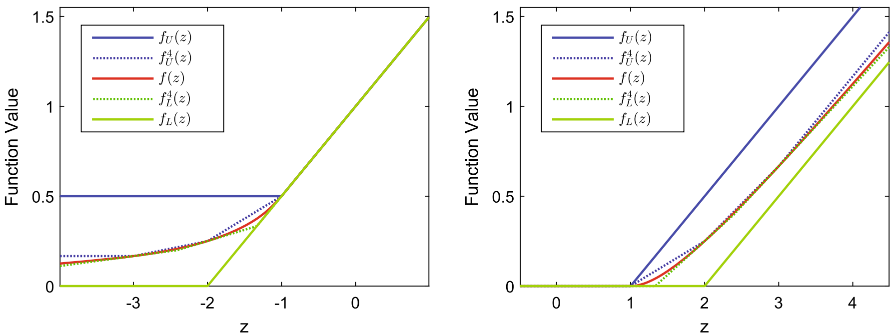
This is equivalent to constraints (33c) by the definition of μ_0 and observing that

$$\sqrt{\left(\left(\frac{\alpha+1}{\alpha}\right)\beta-\mu_0\right)^2+(\Sigma_0-\mu_0^2)} \;=\; \left\|\begin{bmatrix}\left(\frac{\alpha+1}{\alpha}\right)\beta-\left(\frac{\alpha+1}{\alpha}\right)\mu^\top a(x)\\ \Lambda a(x)\end{bmatrix}\right\|.$$

□

Second, we derive tighter approximations of AVC (7) based on tighter approximations of function $f(z)$. Note that both $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$ approximate $f(z)$ based on two linear pieces (see Fig. 2a, b). We generalize $f_{\mathrm{U}}(z)$ and $f_{\mathrm{L}}(z)$ by defining K-piece approximations as follows.

Definition 3 Given integer $K\ge 3$ and real numbers $1=n_1<n_2<\cdots<n_K=\infty$, we define $f_{\mathrm{U}}^K(z)=\mathbb{1}[\beta<0]f_{\mathrm{U}-}^K(z)+\mathbb{1}[\beta\ge 0]f_{\mathrm{U}+}^K(z)$ and $f_{\mathrm{L}}^K(z)=\mathbb{1}[\beta<$



(a) $\beta=-1$ and $\alpha=1$ **(b)** $\beta=1$ and $\alpha=1$

Fig. 2 K-piece approximations of $f(z)$ with $K=4$, $n_1=1$, $n_2=2$, $n_3=3$, and $n_4=\infty$

$0]f^K_{\mathrm{L}-}(z)+\mathbb{1}[\beta\geq 0]f^K_{\mathrm{L}+}(z)$, where

$$
\begin{aligned}
f^K_{\mathrm{U}+}(z) &= \max\Bigg\{0,\ \max_{k=1,\dots,K-1}\Bigg\{\Bigg[\left(\frac{\alpha}{\alpha+1}\right)+\frac{n_{k+1}^{-\alpha}-n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Bigg]z \\
&\quad +\Bigg[\frac{n_{k+1}n_k^{-\alpha}-n_kn_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}-1\Bigg]\beta\Bigg\}\Bigg\}, \\
f^K_{\mathrm{U}-}(z) &= \max\Bigg\{\left(\frac{\alpha}{\alpha+1}\right)z-\beta,\quad \max_{k=1,\dots,K-1}\Bigg\{\Bigg[\frac{n_k^{-\alpha}-n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Bigg]z \\
&\quad -\Bigg[\frac{n_{k+1}n_k^{-\alpha}-n_kn_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\Bigg]\beta\Bigg\}\Bigg\}, \\
f^K_{\mathrm{L}+}(z) &= \max_{k=1,\dots,K}\left\{\left(\frac{\alpha}{\alpha+1}\right)(1-n_k^{-\alpha-1})z-(1-n_k^{-\alpha})\beta\right\}, \quad \text{and} \\
f^K_{\mathrm{L}-}(z) &= \max_{k=1,\dots,K}\left\{\left(\frac{\alpha}{\alpha+1}\right)n_k^{-\alpha-1}z-n_k^{-\alpha}\beta\right\}.
\end{aligned}
$$

We note that $f^K_{\mathrm{U}}(z)$ is the linear interpolation of points $\{(n_k, f(n_k\beta))\}_{k=1,\dots,K}$ and $f^K_{\mathrm{L}}(z)$ is the pointwise maximum of the tangents of $f(z)$ at these points (see Fig. 2a, b). Due to the convexity of $f(z)$, it follows that $f^K_{\mathrm{U}}(z)$ and $f^K_{\mathrm{L}}(z)$ are convex, and $f^K_{\mathrm{L}}(z)\leq f(z)\leq f^K_{\mathrm{U}}(z)$. Furthermore, we observe that $f^K_{\mathrm{L}+}(z)\leq f_+(z)$ by definition. Based on Lemma 2, $f^K_{\mathrm{L}+}(z)\leq f_{\mathrm{L}}(z)\leq f^K_{\mathrm{L}-}(z)$ when $\beta<0$. Similarly, we have $f^K_{\mathrm{L}-}(z)\leq f^K_{\mathrm{L}+}(z)$ when $\beta\geq 0$. It follows that $f^K_{\mathrm{L}}(z)=\max\{f^K_{\mathrm{L}+}(z), f^K_{\mathrm{L}-}(z)\}$. We formalize and extend this observation to $f^K_{\mathrm{U}}(z)$ in the following lemma.

Lemma 3 *We have* $f^K_L(z)=\max\{f^K_{L+}(z), f^K_{L-}(z)\}$ *for all* $z\in\mathbb{R}$. *Furthermore,* $f^K_{U+}(z)\leq f(z)$ *when* $\beta<0$ *and* $f^K_{U-}(z)\leq f(z)$ *when* $\beta\geq 0$. *It follows that* $f^K_U(z)=\max\{f^K_{U+}(z), f^K_{U-}(z)\}$.

Proof We first show that $f^K_{\mathrm{U}+}(z) \leq f_{\mathrm{L}}(z) = \left[\left(\frac{\alpha}{\alpha+1}\right)z - \beta\right]_+$ when $\beta < 0$. Assuming this is true, we have $f^K_{\mathrm{U}+}(z) \leq f(z)$ based on Lemma 2. To this end, we define $f^K_{\mathrm{U}+}(z) = \max\left\{0, \max_{k=1,\ldots,K-1}\ g^k_{\mathrm{U}+}(z)\right\}$, where

$$g^k_{\mathrm{U}+}(z) := \left[\left(\frac{\alpha}{\alpha+1}\right) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)}\right] z + \left[\frac{n_{k+1} n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)} - 1\right] \beta.$$

For each $k = 1, \ldots, K-1$, we prove $g^k_{\mathrm{U}+}(\frac{\alpha+1}{\alpha}\beta) \leq 0$ by the following chain of equivalences:

$$\begin{aligned}
&\left[\left(\frac{\alpha}{\alpha+1}\right) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)}\right]\left(\frac{\alpha+1}{\alpha}\right)\beta + \left[\frac{n_{k+1} n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)} - 1\right]\beta \leq 0 \\
&\Leftrightarrow\ \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{\alpha(n_{k+1} - n_k)} + \frac{n_{k+1} n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)} \geq 0 \\
&\Leftrightarrow\ n_{k+1}^{\alpha}(\alpha n_{k+1} - \alpha - 1) \geq n_k^{\alpha}(\alpha n_k - \alpha - 1),
\end{aligned}$$

where the last line holds because function $g(y) := y^{\alpha}(\alpha y - \alpha - 1)$ is nondecreasing when $y \geq 1$. Indeed, $g'(y) = (\alpha^2 + \alpha)y^{\alpha-1}(y-1) \geq 0$ when $y \geq 1$. It follows that $f^K_{\mathrm{U}+}(\frac{\alpha+1}{\alpha}\beta) = 0$. In addition, we note that $0 \leq \left(\frac{\alpha}{\alpha+1}\right) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \leq \frac{\alpha}{\alpha+1}$. On the one hand, $\left(\frac{\alpha}{\alpha+1}\right) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \leq \frac{\alpha}{\alpha+1}$ because $n_{k+1} > n_k$ and $n_{k+1}^{-\alpha} - n_k^{-\alpha} < 0$. On the other hand, $\left(\frac{\alpha}{\alpha+1}\right) + \frac{n_{k+1}^{-\alpha} - n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \geq 0$ follows from the following equivalence:

$$\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)} \leq \frac{\alpha}{\alpha+1} \quad \Leftrightarrow \quad n_{k+1}^{-\alpha} + \alpha n_{k+1} \geq n_k^{-\alpha} + \alpha n_k,$$

where the right-hand side holds because function $h(y) := y^{-\alpha} + \alpha y$ is nondecreasing when $y \geq 1$. Indeed, $h'(y) = \alpha(1 - y^{-\alpha-1}) \geq 0$ when $y \geq 1$. Hence, the slope of $g^k_{\mathrm{U}+}(z)$ is within interval $[0, \frac{\alpha+1}{\alpha}]$ for all k. It follows that $f^K_{\mathrm{U}+}(z) \leq f_{\mathrm{L}}(z)$ for all $z \in \mathbb{R}$ because (i) $f^K_{\mathrm{U}+}(\frac{\alpha+1}{\alpha}\beta) = f_{\mathrm{L}}(\frac{\alpha+1}{\alpha}\beta) = 0$, (ii) $f^K_{\mathrm{U}+}(z) \leq 0 = f_{\mathrm{L}}(z)$ when $z < \frac{\alpha+1}{\alpha}\beta$ because the slopes of all affine functions making up $f^K_{\mathrm{U}+}(z)$ are nonnegative, and (iii) $f^K_{\mathrm{U}+}(z) \leq f_{\mathrm{L}}(z)$ when $z > \frac{\alpha+1}{\alpha}\beta$ because the slopes of all affine functions making up $f^K_{\mathrm{U}+}(z)$ are smaller than or equal to that of $f_{\mathrm{L}}(z)$, i.e., $\frac{\alpha}{\alpha+1}$.

Second, we show that $f^K_{\mathrm{U}-}(z) \leq f_{\mathrm{L}}(z) = \left[\left(\frac{\alpha}{\alpha+1}\right)z - \beta\right]_+$ when $\beta \geq 0$. Assuming this is true, we have $f^K_{\mathrm{U}-}(z) \leq f(z)$ based on Lemma 2. To this end, we define $f^K_{\mathrm{U}-}(z) = \max\left\{\frac{\alpha}{\alpha+1}z - \beta, \max_{k=1,\ldots,K-1}\ g^k_{\mathrm{U}-}(z)\right\}$, where

$$g^k_{\mathrm{U}-}(z) := \left[\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)}\right] z - \left[\frac{n_{k+1} n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)}\right] \beta.$$

For each $k = 1, \ldots, K-1$, we prove $g^k_{\mathrm{U}-}(\frac{\alpha+1}{\alpha}\beta) \le 0$ by the following chain of equivalences:

$$\left[\frac{n_k^{-\alpha} - n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)}\right]\left(\frac{\alpha+1}{\alpha}\right)\beta - \left[\frac{n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1} - n_k)}\right]\beta \le 0$$
$$\Leftrightarrow \ (\alpha+1)(n_k^{-\alpha} - n_{k+1}^{-\alpha}) \le \alpha(n_{k+1}n_k^{-\alpha} - n_k n_{k+1}^{-\alpha})$$
$$\Leftrightarrow \ n_{k+1}^{\alpha}(\alpha n_{k+1} - \alpha - 1) \ge n_k^{\alpha}(\alpha n_k - \alpha - 1),$$

where the last line has been shown above. It follows that $f^K_{\mathrm{U}-}(\frac{\alpha+1}{\alpha}\beta) = 0$. In addition, we note that $0 \le \frac{n_k^{-\alpha}-n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \le \frac{\alpha}{\alpha+1}$ since $0 \le \left(\frac{\alpha}{\alpha+1}\right) + \frac{n_{k+1}^{-\alpha}-n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} \le \frac{\alpha}{\alpha+1}$. It follows that $f^K_{\mathrm{U}-}(z) \le f_{\mathrm{L}}(z)$ for all $z \in \mathbb{R}$ because (i) $f^K_{\mathrm{U}-}(\frac{\alpha+1}{\alpha}\beta) = f_{\mathrm{L}}(\frac{\alpha+1}{\alpha}\beta) = 0$, (ii) $f^K_{\mathrm{U}-}(z) \le 0 = f_{\mathrm{L}}(z)$ when $z < \frac{\alpha+1}{\alpha}\beta$ because the slopes of all affine functions making up $f^K_{\mathrm{U}-}(z)$ are nonnegative, and (iii) $f^K_{\mathrm{U}-}(z) \le f_{\mathrm{L}}(z)$ when $z > \frac{\alpha+1}{\alpha}\beta$ because the slopes of all affine functions making up $f^K_{\mathrm{U}-}(z)$ are smaller than or equal to that of $f_{\mathrm{L}}(z)$, i.e., $\frac{\alpha}{\alpha+1}$. □

In the following proposition, we present conservative approximations based on $f^K_{\mathrm{U}}(z)$ and relaxed ones based on $f^K_{\mathrm{L}}(z)$, both of which are in the form of linear matrix inequalities. We note that these approximations are asymptotically tight as K grows to infinity. We omit the proof here because it follows from the standard duality approach. Interested readers are referred to [12] and [43].

Proposition 10 *Define $(T+1)\times(T+1)$ matrix $\Omega := \begin{bmatrix} \left(\frac{\alpha+2}{\alpha}\right)\Sigma & \left(\frac{\alpha+1}{\alpha}\right)\mu \\ \left(\frac{\alpha+1}{\alpha}\right)\mu^\top & 1 \end{bmatrix}$. Then, for given integer $K \ge 3$ and real numbers $1 = n_1 < n_2 < \cdots < n_K = \infty$, AVC (7) is satisfied if there exists a symmetric matrix $M_U \in \mathbb{R}^{(T+1)\times(T+1)}$ such that*

$$\beta + \frac{1}{\epsilon}M_U \cdot \Omega \le b(x), \quad M_U \succeq 0, \quad M_U \succeq \begin{bmatrix} 0 & \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)a(x) \\ \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)a(x)^\top & -\beta \end{bmatrix}, \tag{34a}$$

$$M_U \succeq \begin{bmatrix} 0 & \frac{1}{2}\left[\left(\frac{\alpha}{\alpha+1}\right) + \frac{n_{k+1}^{-\alpha}-n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\right]a(x) \\ \frac{1}{2}\left[\left(\frac{\alpha}{\alpha+1}\right) + \frac{n_{k+1}^{-\alpha}-n_k^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\right]a(x)^\top & \left[\frac{n_{k+1}n_k^{-\alpha}-n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)} - 1\right]\beta \end{bmatrix},$$
$$\forall k = 1, \ldots, K-1, \tag{34b}$$

$$M_U \succeq \begin{bmatrix} 0 & \frac{1}{2}\left[\frac{n_k^{-\alpha}-n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\right]a(x) \\ \frac{1}{2}\left[\frac{n_k^{-\alpha}-n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\right]a(x)^\top & -\left[\frac{n_{k+1}n_k^{-\alpha}-n_k n_{k+1}^{-\alpha}}{(\alpha+1)(n_{k+1}-n_k)}\right]\beta \end{bmatrix}, \quad \forall k = 1, \ldots, K-1, \tag{34c}$$

where $\cdot$ represents the Frobenius product of matrices. Furthermore, AVC (7) implies that there exists a symmetric matrix $M_L \in \mathbb{R}^{(T+1)\times(T+1)}$ such that

$$\beta + \frac{1}{\epsilon} M_L \cdot \Omega \le b(x), \quad M_L \succeq 0, \quad M_L \succeq \begin{bmatrix} 0 & \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)a(x) \\ \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)a(x)^\top & -\beta \end{bmatrix},$$

$$M_L \succeq \begin{bmatrix} 0 & \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)(1-n_k^{-\alpha-1})a(x) \\ \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)(1-n_k^{-\alpha-1})a(x)^\top & -(1-n_k^{-\alpha})\beta \end{bmatrix}, \quad \forall k = 1, \ldots, K-1,$$

$$M_L \succeq \begin{bmatrix} 0 & \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)n_k^{-\alpha-1}a(x) \\ \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)n_k^{-\alpha-1}a(x)^\top & -n_k^{-\alpha}\beta \end{bmatrix}, \quad \forall k = 1, \ldots, K-1.$$

Note that, as $n_K = \infty$, constraints (34b)–(34c) reduce to

$$M_{\mathrm{U}} \succeq \begin{bmatrix} 0 & \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)a(x) \\ \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)a(x)^\top & \left(\frac{n_{K-1}^{-\alpha}}{\alpha+1} - 1\right)\beta \end{bmatrix},$$

$$M_{\mathrm{U}} \succeq \begin{bmatrix} 0 & 0 \\ 0 & -\left(\frac{n_{K-1}^{-\alpha}}{\alpha+1}\right)\beta \end{bmatrix}$$

when $k = K-1$.

Remark 3 In computation, we can use the conservative approximation (34a)–(34c) to find near-optimal solutions. More specifically, suppose that we employ the separation approach to solve problem $\min\{c(x) : x \in X, \ x \text{ satisfies } (7)\}$ and have finished the first K iterations. Then, from these iterations, we obtain a lower bound c_{L}^K of the optimal objective value and K outputs, denoted $\varphi_1, \ldots, \varphi_K$, by iteratively solving Separation Problem 2. By letting $n_1 = 1$, $n_{K+2} = \infty$, and $n_k = \varphi_{k-1}$ for all $k = 2, \ldots, K+1$, we obtain an upper bound c_{U}^K of the optimal objective value by solving problem $\min\{c(x) : \ x \in X, \ x \text{ satisfies } (34a)\text{–}(34c) \text{ based on } n_1, \ldots, n_{K+2}\}$, whose optimal solution is denoted x_K^*. If $(c_{\mathrm{U}}^K - c_{\mathrm{L}}^K)/c_{\mathrm{L}}^K$ is small enough, then we can stop the iterations and output x_K^* as a near-optimal solution.

4 Extension to linear unimodality

In this section, we consider an extension of ACC (6) and AVC (7) based on a related structural property called *linear unimodality*.

Definition 4 (*Linear Unimodality; see* [11]) A probability distribution $\mathbb{P}_\xi$ is called linear unimodal about 0 if for all $a \in \mathbb{R}^T$, the linear combination $a^\top \xi$ is univariate unimodal about 0.

Analogous to (5), we define the alternative ambiguity set based on linear unimodality as

$$\mathcal{D}_\xi^{\mathrm{LU}}(\mu, \Sigma) := \left\{\mathbb{P}_\xi \in \mathcal{M}_T : \mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu,\ \mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = \Sigma,\ \mathbb{P}_\xi \text{ is linear unimodal about } 0\right\}. \tag{35}$$

We now show an equivalence between ambiguity sets $\mathcal{D}_\xi^{\mathrm{LU}}(\mu, \Sigma)$ and $\mathcal{D}_\xi(\mu, \Sigma, \alpha)$ with $\alpha = 1$. It follows that all results derived in Sects. 2 and 3, with α set to be 1, remain valid under $\mathcal{D}_\xi^{\mathrm{LU}}(\mu, \Sigma)$.

Proposition 11 *For any Borel measurable function $h : \mathbb{R} \to \mathbb{R}$, we have*

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi(\mu,\Sigma,1)} \mathbb{E}_{\mathbb{P}_\xi}\big[h(a(x)^\top \xi)\big] = \inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi^{LU}(\mu,\Sigma)} \mathbb{E}_{\mathbb{P}_\xi}\big[h(a(x)^\top \xi)\big].$$

Proof By Theorem 3.5 in [11], a random variable X is 1-unimodal if and only if there exists a random variable Z such that $X = UZ$, where U is uniform in (0, 1) and independent of Z.

First, pick any ξ such that $\mathbb{P}_\xi \in \mathcal{D}_\xi(\mu, \Sigma, 1)$. As $a^\top \xi$ is univariate 1-unimodal for all $a \in \mathbb{R}^T$ because $\mathbb{P}_\xi$ is 1-unimodal, $\mathbb{P}_\xi \in \mathcal{D}_\xi^{\mathrm{LU}}(\mu, \Sigma)$. It follows that $\mathcal{D}_\xi(\mu, \Sigma, 1) \subseteq \mathcal{D}_\xi^{\mathrm{LU}}(\mu, \Sigma)$ and so $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi(\mu,\Sigma,1)} \mathbb{E}_{\mathbb{P}_\xi}[h(a(x)^\top \xi)] \geq \inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi^{\mathrm{LU}}(\mu,\Sigma)} \mathbb{E}_{\mathbb{P}_\xi}[h(a(x)^\top \xi)]$.

Second, pick any ξ such that $\mathbb{P}_\xi \in \mathcal{D}_\xi^{\mathrm{LU}}(\mu, \Sigma)$. Then, $\zeta := a(x)^\top \xi$ is 1-unimodal because $\mathbb{P}_\xi$ is linear unimodal. Hence, there exists a Z_ζ such that $\zeta = UZ_\zeta$. It follows that $\mathbb{E}[Z_\zeta] = 2\mu_1$ and $\mathbb{E}[Z_\zeta^2] = 3\Sigma_1$. Based on Theorem 1 in [27], there exists a $Z_\xi \in \mathbb{R}^T$ such that $Z_\zeta = a(x)^\top Z_\xi$, $\mathbb{E}[Z_\xi] = 2\mu$, and $\mathbb{E}[Z_\xi Z_\xi^\top] = 3\Sigma$. It follows that UZ_ξ is 1-unimodal, and meanwhile $\mathbb{E}_{\mathbb{P}_\xi}[UZ_\xi] = \frac{1}{2}\mathbb{E}[Z_\xi] = \mu$ and $\mathbb{E}_{\mathbb{P}_\xi}[(UZ_\xi)(UZ_\xi)^\top] = \frac{1}{3}\mathbb{E}[Z_\xi Z_\xi^\top] = \Sigma$. Furthermore, $a(x)^\top \xi = a(x)^\top (UZ_\xi)$. Therefore, the probability distribution of UZ_ξ belongs to $\mathcal{D}_\xi(\mu, \Sigma, 1)$, and so $\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi(\mu,\Sigma,1)} \mathbb{E}_{\mathbb{P}_\xi}[h(a(x)^\top \xi)] \leq \inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi^{\mathrm{LU}}(\mu,\Sigma)} \mathbb{E}_{\mathbb{P}_\xi}[h(a(x)^\top \xi)]$. $\square$

5 Computational case studies

In this section, we conduct two case studies. In Sect. 5.1, we evaluate the theoretical results derived in Sects. 2 and 3 based on a risk-constrained economic dispatch (RCED) problem in power system operation. In Sect. 5.2, we compare the computational performance of our ACC representation using SOC constraints with that of the SDP reformulation derived in [17] based on a simplified RCED problem with a varying uncertainty dimension.

5.1 The RCED case study

We present a nominal RCED model as follows:

$$\min_{g,d,r^{\text{U}},r^{\text{D}}} \sum_{i\in\mathcal{I}\setminus\mathcal{I}_{\text{R}}} \left[c_{i2}g_i^2 + c_{i1}g_i + c_i^{\text{R}}(r_i^{\text{U}} + r_i^{\text{D}})\right] \tag{36a}$$

$$\text{s.t.} \sum_{i\in\mathcal{I}\setminus\mathcal{I}_{\text{R}}} g_i + \sum_{i\in\mathcal{I}_{\text{R}}} f_i = \sum_{b=1}^{B} L_b, \tag{36b}$$

$$r_i = -\Big(\sum_{i\in\mathcal{I}_{\text{R}}} w_i\Big)d_i, \quad \forall i \in \mathcal{I}\setminus\mathcal{I}_{\text{R}}, \tag{36c}$$

$$\sum_{i\in\mathcal{I}\setminus\mathcal{I}_{\text{R}}} d_i = 1, \tag{36d}$$

$$-r_i^{\text{D}} \le r_i \le r_i^{\text{U}}, \quad \forall i \in \mathcal{I}\setminus\mathcal{I}_{\text{R}}, \tag{36e}$$

$$g_i^{\text{MIN}} \le g_i + r_i \le g_i^{\text{MAX}}, \quad \forall i \in \mathcal{I}\setminus\mathcal{I}_{\text{R}}, \tag{36f}$$

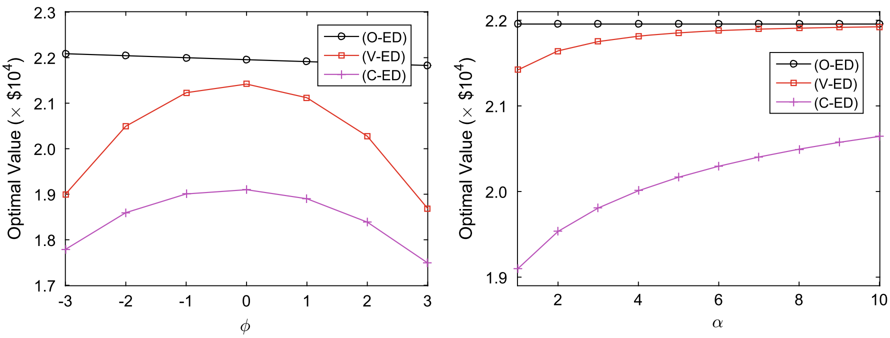
$$-C_\ell \le \sum_{b=1}^{B} D_\ell^b\Big[\sum_{i\in\mathcal{G}_b}(g_i + r_i) + \sum_{i\in\mathcal{H}_b}(f_i + w_i) - L_b\Big] \le C_\ell, \quad \forall \ell \in \mathcal{L}, \tag{36g}$$

where B represents the number of buses in the power system, $\mathcal{I}$ represents the set of generating units (conventional and renewable), $\mathcal{I}_{\text{R}}$ represents the set of renewable units, $\mathcal{L}$ represents the set of transmission lines, $\mathcal{G}_b$ represents the set of conventional units at bus b, $\mathcal{H}_b$ represents the set of renewable units at bus b, c_{i2} and c_{i1} represent cost parameters of conventional unit i, c_i^{R} represents the unit cost for up/down reserve capacity of conventional unit i, L_b represents the load at bus b, and C_ℓ represents the capacity of transmission line ℓ. For each renewable unit $i \in \mathcal{I}_{\text{R}}$, f_i and w_i represent the forecasted power output and the forecast error, respectively. For each conventional unit $i \in \mathcal{I}\setminus\mathcal{I}_{\text{R}}$, g_i and r_i represent the planned generation amount and the adjustment amount, respectively, and d_i represents the portion of total generation-load mismatch to be offset by this unit (see, e.g., [5,38]). Constraint (36b) describes the power balance requirement for generation and loads (we assume that the loads are deterministic), constraints (36c) describe the proportional distribution of mismatches, constraint (36d) requires that all proportions sum to 1, constraints (36e) limit the adjustment amount by the reserve capacities r^{U} and r^{D}, constraints (36f) bound the generation amount by the generation capacity, and constraints (36g) describe the transmission capacity limits based on the dc power flow approximation where D_ℓ^b maps power injections to power flows (see, e.g., [3] and [16]).

Our case study uses the IEEE 30-bus system [42]. We increase all electric loads by 50% and add two wind farms at buses 5 and 22. The forecasted power output from each wind farm is 30 MW. The transmission line between buses 1 and 2 has a capacity of 30 MW, while all other line flows are unconstrained. Other cost and capacity coefficients are reported in Table 1. We assume random forecast errors and describe the uncertainty by an uncorrelated random vector $w := [w_1, w_2]^\top$ with mean μ_{w} and covariance matrix $\Gamma_{\text{w}} = \text{diag}(9, 9)$. Additionally, we assume that w is α-unimodal about $[0, 0]^\top$. To handle random violations of constraints (36e)–(36g), we replace them by ACC (6) and AVC (7), and term the resultant RCED model (C-ED) and (V-ED), respectively. For example, in (C-

Table 1 Coefficients of the case study

Conventional unit	Bus index	c_{i1} (\$/MW)	c_{i2} (\$/MW^2)	c_i^{R} (\$/MW)	g_i^{MIN} (MW)	g_i^{MAX} (MW)
1	1	20	0.04	200	0	360
2	2	40	0.25	400	0	140
3	5	40	0.01	400	0	100
4	8	40	0.01	400	0	100
5	11	40	0.01	400	0	100
6	13	40	0.01	400	0	100



(a) Optimal Value versus ϕ **(b)** Optimal Value versus α

Fig. 3 Optimal values of (O-ED), (C-ED), and (V-ED) with various ϕ and α

ED), we replace constraints (36e) by $\inf_{\mathbb{P}_{\mathrm{w}} \in \mathcal{D}_{\mathrm{w}}}\{d_i \sum_{i \in \mathcal{I}_{\mathrm{R}}} w_i \leq r_i^{\mathrm{D}}\} \geq 1-\epsilon$ and $\inf_{\mathbb{P}_{\mathrm{w}} \in \mathcal{D}_{\mathrm{w}}}\{-d_i \sum_{i \in \mathcal{I}_{\mathrm{R}}} w_i \leq r_i^{\mathrm{U}}\} \geq 1-\epsilon$, where $\mathcal{D}_{\mathrm{w}} = \{\mathbb{P}_{\mathrm{w}} \in \mathcal{M}_2 : \mathbb{E}_{\mathbb{P}_{\mathrm{w}}}[w] = \mu_{\mathrm{w}}, \mathbb{E}_{\mathbb{P}_{\mathrm{w}}}[ww^\top] = \mu_{\mathrm{w}}\mu_{\mathrm{w}}^\top + \mathrm{diag}(9,9), \mathbb{P}_{\mathrm{w}} \text{ is } \alpha\text{-unimodal about } 0\}$. In contrast, in (V-ED), we replace constraints (36e) by $\sup_{\mathbb{P}_{\mathrm{w}} \in \mathcal{D}_{\mathrm{w}}} \mathrm{CVaR}_{\mathbb{P}_{\mathrm{w}}}^{\epsilon}(d_i \sum_{i \in \mathcal{I}_{\mathrm{R}}} w_i) \leq r_i^{\mathrm{D}}$ and $\sup_{\mathbb{P}_{\mathrm{w}} \in \mathcal{D}_{\mathrm{w}}} \mathrm{CVaR}_{\mathbb{P}_{\mathrm{w}}}^{\epsilon}(-d_i \sum_{i \in \mathcal{I}_{\mathrm{R}}} w_i) \leq r_i^{\mathrm{U}}$. Throughout this case study, we set $1-\epsilon = 95\%$. Lastly, when the requirement of α-unimodality is relaxed from $\mathcal{D}_{\mathrm{w}}$, (C-ED) and (V-ED) become equivalent and we term this model (O-ED).

By using (O-ED) as a benchmark, we test (C-ED) and (V-ED) under various selections of μ_{w} and α values. First, we fix $\alpha = 1$ and let $\mu_{\mathrm{w}} = \phi[1,1]^\top$ with $\phi \in \{-3, -2, \ldots, 3\}$. We report the optimal objective values of the three models in Fig. 3a. From this figure, we observe that the optimal value of (O-ED) is consistently larger than that of (V-ED), which is consistently larger than that of (C-ED). This demonstrates that incorporating α-unimodality makes the RCED model less conservative and hence decreases the cost of economic dispatch. Meanwhile, unlike in (O-ED), ACC (6) and AVC (7) are not equivalent when α-unimodality is incorporated in the ambiguity set. Furthermore, we observe that the discrepancy between (O-ED) and (C-ED)/(V-ED) amplifies as ϕ deviates from 0. This indicates that α-unimodality plays a more important role in $\mathcal{D}_{\mathrm{w}}$ as the difference between μ_{w} and the mode increases.

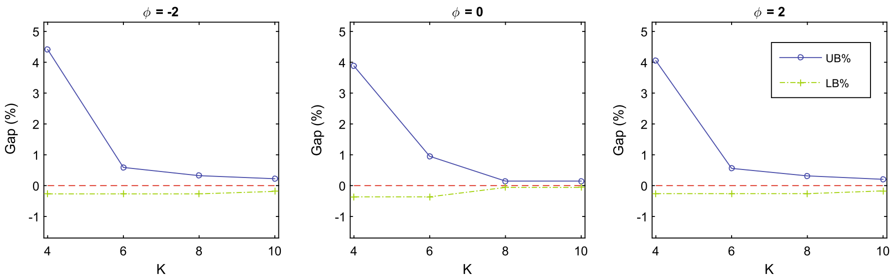


Fig. 4 Gaps between the optimal objective value and the relaxed and conservative approximations of (C-ED)

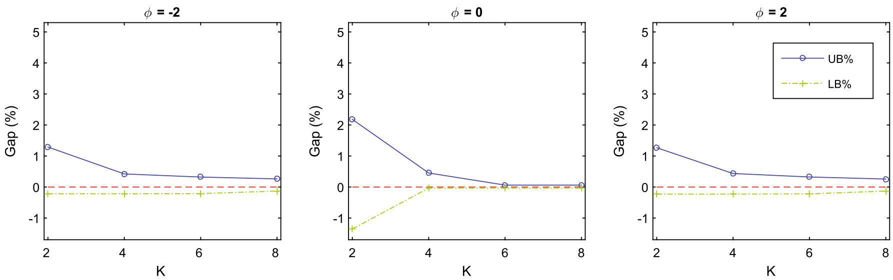
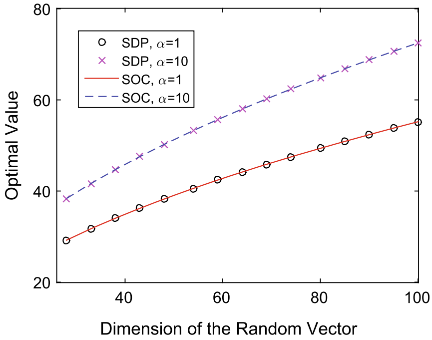


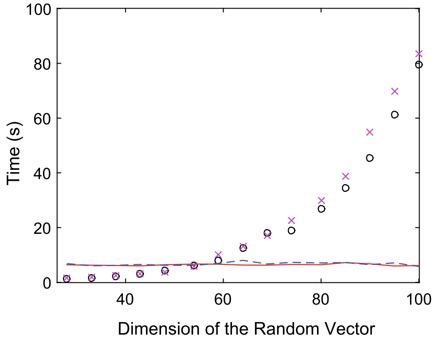
Fig. 5 Gaps between the optimal objective value and the relaxed and conservative approximations of (V-ED)

Second, we fix $\mu_{\mathrm{W}} = [0, 0]^\top$ and let α increase from 1 to 10. We report the optimal objective values of the three models in Fig. 3b. From this figure, we observe that the discrepancy between (O-ED) and (C-ED)/(V-ED) shrinks as α grows. This is as expected because the requirement of α-unimodality weakens as α grows. Although not shown in this figure, the convergence of (V-ED) to (O-ED) takes place when $\alpha \geq 40$, while the convergence of (C-ED) takes place when $\alpha \geq 10^4$. The slow convergence indicates that unimodality information can significantly influence the structure of $\mathcal{D}_{\mathrm{W}}$ and the worst-case probability distribution.

Third, we let $\alpha = 1$, $\mu_{\mathrm{W}} = \phi[1, 1]^\top$ with $\phi \in \{-2, 0, 2\}$, and evaluate the tightness of the approximations of ACC and AVC derived in Propositions 4, 5 and 10, respectively. In this test, we follow Remarks 2, 3 to choose the interpolation points $n_1, \ldots, n_K$ in these approximations. In Fig. 4, we report the gap between the optimal objective value $v^*_{\text{(C-ED)}}$ of (C-ED) and the upper bound v_{UB} obtained from the conservative approximation, and the gap between $v^*_{\text{(C-ED)}}$ and the lower bound v_{LB} obtained from the relaxed approximation, for $K \in \{4, 6, 8, 10\}$. The gaps are obtained by computing UB% $= (v_{\text{UB}} - v^*_{\text{(C-ED)}})/v^*_{\text{(C-ED)}} \times 100\%$ and LB% $= (v^*_{\text{(C-ED)}} - v_{\text{LB}})/v^*_{\text{(C-ED)}} \times 100\%$. Similarly, in Fig. 5, we report the gap between the optimal objective value $v^*_{\text{(V-ED)}}$ of (V-ED) and those of its K-piece approximations with $K \in \{2, 4, 6, 8\}$. From Figs. 4, 5, we observe that the gaps quickly shrink as K increases and the approximations become near-optimal (e.g., UB% + LB% $< 1\%$) when $K \geq 8$.



(a) Optimal Value versus n



(b) Solution Time versus n

Fig. 6 Comparisons between our approach and the SDP approach in [17]

5.2 Comparison with the SDP reformulation

To demonstrate the scalability of our ACC representation with regard to an increasing uncertainty dimension, we consider a simplified RCED model where we combine all conventional generating units into a single unit without generation capacity limits and relax all transmission capacity limits. As a result, the mismatch offset proportion of the combined unit is $d = 1$. Additionally, the planned generation amount of the combined unit is deterministic, i.e., $g = \sum_{b=1}^{B} L_b - \sum_{i \in \mathcal{I}_{\mathrm{R}}} f_i$, and so is the corresponding generation cost. Accordingly, we can remove variable g and the corresponding generating cost from the formulation and present the simplified RCED model as follows:

$$\min_{r^{\mathrm{U}}, r^{\mathrm{D}}} r^{\mathrm{U}} + r^{\mathrm{D}} \tag{37a}$$

$$\text{s.t.} \quad \inf_{\mathbb{P}_{\mathrm{w}} \in \mathcal{D}_{\mathrm{w}}} \mathbb{P}_{\mathrm{w}} \left\{ \sum_{i=1}^{n} w_i \le r^{\mathrm{D}} \right\} \ge 1 - \epsilon, \tag{37b}$$

$$\inf_{\mathbb{P}_{\mathrm{w}} \in \mathcal{D}_{\mathrm{w}}} \mathbb{P}_{\mathrm{w}} \left\{ - \sum_{i=1}^{n} w_i \le r^{\mathrm{U}} \right\} \ge 1 - \epsilon, \tag{37c}$$

where $n = |\mathcal{I}_{\mathrm{R}}|$, $\mathcal{D}_{\mathrm{w}} = \{\mathbb{P}_{\mathrm{w}} \in \mathcal{M}_n : \mathbb{E}_{\mathbb{P}_{\mathrm{w}}}[w] = 0, \mathbb{E}_{\mathbb{P}_{\mathrm{w}}}[ww^\top] = \mathrm{diag}(1, \ldots, 1), \mathbb{P}_{\mathrm{w}}$ is α-unimodal about $0\}$, and $1 - \epsilon = 95\%$. Our case study tests $\alpha = 1$, 10 and 15 values of n ranging from 25 to 100. For each instance of the simplified RCED model, we evaluate the optimal objective value and solution time of the following two approaches:

1. Our ACC representation using SOC constraints based on Theorem 1;
2. The ACC representation using SDP based on Theorem 3.4.5 and Example 3.4.4 in [17].[1]

[1] The ACC representation of Example 3.4.4 has typos and, for completeness, we present the corrected representation in Appendix C. We test the corrected ACC representation in this case study.

We report the evaluation results in Fig. 6. On the one hand, from Fig. 6a, we observe that these two approaches yield the same optimal values in all test instances. This indicates that the ACC representations derived in these two approaches are equivalent. On the other hand, from Fig. 6b, we observe that the solution time of our approach remains approximately constant as the dimension n of the random vector w increases, while that of the SDP approach increases superlinearly with n. This indicates that our approach using SOC constraints has a computational advantage, especially when involving a large number of random variables.

6 Future research

An interesting future research direction is to derive a computationally tractable reformulation of ACC (6) without making Assumption 2. Another future research direction is to bound the errors of relaxed approximation (19) and conservative approximation (20) for ACC (6).

Acknowledgements This research has been supported in part by the National Science Foundation (NSF) under Grants CMMI-1555983 and CCF-1442495.

Appendix A: proof of Observation 2

Proof As $f_+^k(z) = \left(\frac{\alpha}{\alpha+1}\right)(1 - k^{-\alpha-1})z - (1 - k^{-\alpha})\beta$, we have

$$\left[f_+^k(z)\right]_+ = \begin{cases} 0, & \text{if } z < z_0(k) := \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)\beta \\ \left(\frac{\alpha}{\alpha+1}\right)(1 - k^{-\alpha-1})z - (1 - k^{-\alpha})\beta, & \text{if } z \geq z_0(k) \end{cases}$$

for all $k \geq 1$. As $[f_+^{k+1}(z)]_+ \geq 0$ for all $z \in \mathbb{R}$, to show that $[f_+^{k+1}(z)]_+ \geq [f_+^k(z)]_+$, it suffices to prove that $[f_+^{k+1}(z)]_+ \geq f_+^k(z)$ for all $z \in \mathbb{R}$. First, as $\beta \leq 0$ and $\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}$ increases in k, we have $\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-k^{-\alpha}}{1-k^{-\alpha-1}}\right)\beta \geq \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{1-(k+1)^{-\alpha}}{1-(k+1)^{-\alpha-1}}\right)\beta$, i.e., $z_0(k) \geq z_0(k+1)$. It follows that, when $z < z_0(k)$, $f_+^k(z) \leq 0$ and hence $[f_+^{k+1}(z)]_+ \geq f_+^k(z)$. Second, when $z \geq z_0(k)$, $f_+^{k+1}(z) \geq 0$ because $z \geq z_0(k) \geq z_0(k+1)$ and $f_+^{k+1}(z)$ increases in z. As both $f_+^k(z)$ and $f_+^{k+1}(z)$ are affine functions of z, we have $f_+^{k+1}(z) = f_+^{k+1}(z_0(k)) + (\frac{\alpha}{\alpha+1})(1 - (k+1)^{-\alpha-1})(z - z_0(k))$ and $f_+^k(z) = (\frac{\alpha}{\alpha+1})(1 - k^{-\alpha-1})(z - z_0(k))$ for $z \geq z_0(k)$. It follows that $f_+^{k+1}(z) - f_+^k(z) = f_+^{k+1}(z_0(k)) + (\frac{\alpha}{\alpha+1})[k^{-\alpha-1} - (k+1)^{-\alpha-1}](z - z_0(k)) \geq 0$. Hence, $[f_+^{k+1}(z)]_+ \geq f_+^k(z)$ when $z \geq z_0(k)$ and the proof is complete. □

Appendix B

For random variable Z and constant $\beta \in \mathbb{R}$, we make the following observation on the worst-case expectation $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[Z - \beta]_+$. Note that this observation can be made following the derivations in [30], and we present a proof below for completeness.

Observation 3 *Given $\beta \in \mathbb{R}$, we have*

$$\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+ = \frac{1}{2}\left[\sqrt{(\beta-\mu_0)^2 + (\Sigma_0 - \mu_0^2)} - \beta + \mu_0\right].$$

Proof We represent $\sup_{\mathbb{P}_Z \in \mathcal{D}(\mu_0, \Sigma_0)} \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+$ as the following optimization problem

$$\begin{aligned}
v_P = \ & \max_{\mathbb{P}_Z} \ \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+ \\
\text{(P)} \qquad & \text{s.t. } \mathbb{E}_{\mathbb{P}_Z}[Z] = \mu_0, \\
& \qquad \mathbb{E}_{\mathbb{P}_Z}[Z^2] = \Sigma_0, \\
& \qquad \mathbb{E}_{\mathbb{P}_Z}[1] = 1,
\end{aligned}$$

whose dual is

$$\begin{aligned}
v_D = \ & \min_{q,p,r} \ \mu_0 p + \Sigma_0 q + r \\
\text{(D)} \qquad & \text{s.t. } qz^2 + pz + r \ge [z-\beta]_+, \quad \forall z \in \mathbb{R}.
\end{aligned}$$

The weak duality between (P) and (D), i.e., $v_D \le v_P$, holds because $\mu_0 p + \Sigma_0 q + r = \mathbb{E}_{\mathbb{P}_Z}[qZ^2 + pZ + r] \le \mathbb{E}_{\mathbb{P}_Z}[Z-\beta]_+$ for any feasible solution (q, p, r) to (D) and feasible solution $\mathbb{P}_Z$ to (P). Now we prove the strong duality by constructing two feasible solutions to (P) and (D), respectively, that have the same objective value. On the one hand, the primal solution $\hat{\mathbb{P}}_Z$ is supported on two points z_1 and z_2 with probability masses p_1 and p_2, respectively, where $\Delta = \sqrt{(\beta-\mu_0)^2 + (\Sigma_0 - \mu_0^2)}$ and

$$p_1 = \frac{\beta - \mu_0 + \Delta}{2\Delta}, \quad p_2 = \frac{\mu_0 - \beta + \Delta}{2\Delta}, \quad z_1 = \beta - \Delta, \quad \text{and} \quad z_2 = \beta + \Delta.$$

We have $p_1, p_2 \ge 0$ because $\Delta \ge |\beta - \mu_0|$. Meanwhile, we have

$$p_1 z_1 + p_2 z_2 = \frac{(\beta - \mu_0 + \Delta)(\beta - \Delta)}{2\Delta} + \frac{(\mu_0 - \beta + \Delta)(\beta + \Delta)}{2\Delta} = \mu_0,$$

and

$$\begin{aligned}
p_1 z_1^2 + p_2 z_2^2 &= \frac{(\beta - \mu_0 + \Delta)(\beta - \Delta)^2}{2\Delta} + \frac{(\mu_0 - \beta + \Delta)(\beta + \Delta)^2}{2\Delta} \\
&= \frac{(\beta - \mu_0)\left[(\beta-\Delta)^2 - (\beta+\Delta)^2\right] + \Delta\left[(\beta-\Delta)^2 + (\beta+\Delta)^2\right]}{2\Delta} \\
&= -\beta^2 + 2\mu_0\beta + \Delta^2 = -\beta^2 + 2\mu_0\beta + (\beta-\mu_0)^2 + (\Sigma_0 - \mu_0^2) = \Sigma_0.
\end{aligned}$$

Hence, $\hat{\mathbb{P}}_Z$ is feasible to (P). On the other hand, the dual solution $(\hat{q}, \hat{p}, \hat{r})$ is such that

$$\hat{q} = \frac{1}{4\Delta}, \quad \hat{p} = \frac{\Delta - \beta}{2\Delta}, \quad \text{and} \quad \hat{r} = \frac{(\Delta - \beta)^2}{4\Delta}.$$

Hence, $\hat{q}z^2 + \hat{p}z + \hat{r} = \frac{1}{4\Delta}(z + \Delta - \beta)^2$. It follows that $\hat{q}z^2 + \hat{p}z + \hat{r} \geq 0$ for all $z \in \mathbb{R}$. Meanwhile, $(\hat{q}z^2 + \hat{p}z + \hat{r}) - (z - \beta) = \frac{1}{4\Delta}(z - \beta - \Delta)^2 \geq 0$, i.e., $\hat{q}z^2 + \hat{p}z + \hat{r} \geq z - \beta$. Thus, $\hat{q}z^2 + \hat{p}z + \hat{r} \geq [z - \beta]_+$ and so $(\hat{q}, \hat{p}, \hat{r})$ is feasible to (D).

Finally, the primal objective value associated with $\hat{\mathbb{P}}_Z$ is $p_2(z_2 - \beta) = \frac{(\mu_0 - \beta + \Delta)\Delta}{2\Delta} = \frac{1}{2}(\Delta - \beta + \mu_0)$. Meanwhile, the dual objective value associated with $(\hat{q}, \hat{p}, \hat{r})$ is

$$\begin{aligned} &\mu_0\left(\frac{\Delta - \beta}{2\Delta}\right) + \Sigma_0\left(\frac{1}{4\Delta}\right) + \frac{(\Delta - \beta)^2}{4\Delta} \\ &= \frac{\Delta^2 + (\beta^2 - 2\mu_0\beta + \mu_0^2) + (\Sigma_0 - \mu_0^2) + 2\mu_0\Delta - 2\Delta\beta}{4\Delta} \\ &= \frac{2\Delta^2 + 2\mu_0\Delta - 2\Delta\beta}{4\Delta} = \frac{1}{2}(\Delta - \beta + \mu_0), \end{aligned}$$

which coincides with the primal objective value associated with $\hat{\mathbb{P}}_Z$. □

Appendix C: corrected ACC representation of Example 3.4.4 in [17]

The ACC representation (3.50) in Example 3.4.4 in [17] has typos and is corrected as follows:

$$\begin{aligned} &\beta - (\mu - m)^\top \gamma - \langle \Sigma + (\mu - m)(\mu - m)^\top, \ \Gamma \rangle \geq (1 - \epsilon)\tau, \\ &\begin{bmatrix} \tau - \beta & \frac{1}{2}\frac{\alpha}{\alpha+1}\gamma^\top \\ \frac{1}{2}\frac{\alpha}{\alpha+1}\gamma & \frac{\alpha}{\alpha+2}\Gamma \end{bmatrix} \succeq 0, \\ &\begin{bmatrix} \left[\alpha^{\frac{1}{\alpha+1}} + \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha+1}}\right]\tau^{\frac{1}{\alpha+1}}\left(b(x) - m^\top a(x)\right)^{\frac{\alpha}{\alpha+1}} - \beta & \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\gamma - a(x)\right)^\top \\ \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\gamma - a(x)\right) & \frac{\alpha}{\alpha+2}\Gamma \end{bmatrix} \succeq 0. \end{aligned}$$

References

1. Ahmed, S., Papageorgiou, D.: Probabilistic set covering with correlations. Oper. Res. **61**(2), 438–452 (2013)
2. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999)
3. Bergen, A.R., Vittal, V.: Power Systems Analysis, 2nd edn. Prentice Hall, Upper Saddle River (1999)
4. Bertsimas, D., Doan, X.V., Natarajan, K., Teo, C.-P.: Models for minimax stochastic linear optimization problems with risk aversion. Math. Oper. Res. **35**(3), 580–602 (2010)
5. Bienstock, D., Chertkov, M., Harnett, S.: Chance-constrained optimal power flow: risk-aware network control under uncertainty. SIAM Rev. **56**(3), 461–495 (2014)
6. Bookbinder, J.H., Tan, J.-Y.: Strategies for the probabilistic lot-sizing problem with service-level constraints. Manag. Sci. **34**(9), 1096–1108 (1988)
7. Calafiore, G., El Ghaoui, L.: On distributionally robust chance-constrained linear programs. J. Optim. Theory Appl. **130**(1), 1–22 (2006)

8. Charnes, A., Cooper, W., Symonds, G.: Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. Manag. Sci. **4**(3), 235–263 (1958)
9. Cheng, J., Delage, E., Lisser, A.: Distributionally robust stochastic knapsack problem. SIAM J. Optim. **24**(3), 1485–1506 (2014)
10. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**(3), 595–612 (2010)
11. Dharmadhikari, S., Joag-Dev, K.: Unimodality, Convexity, and Applications. Elsevier, Amsterdam (1988)
12. El Ghaoui, L., Oks, M., Oustry, F.: Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. Oper. Res. **51**(4), 543–556 (2003)
13. Erdoğan, E., Iyengar, G.: Ambiguous chance constrained problems and robust optimization. Math. Program. **107**(1), 37–61 (2006)
14. Esfahani, P. M., Kuhn, D.: Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. Available on optimization-online. http://www.optimization-online.org/DB_FILE/2015/05/4899.pdf (2015)
15. Gade, D., Küçükyavuz, S.: Formulations for dynamic lot sizing with service levels. Nav. Res. Logist. **60**(2), 87–101 (2013)
16. Gómez-Expósito, A., Conejo, A.J., Cañizares, C.: Electric Energy Systems: Analysis and Operation. CRC Press, Boca Raton (2008)
17. Hanasusanto, G. A.: Decision making under uncertainty: robust and data-driven approaches. Ph.D. thesis, Imperial College London (2015)
18. Hanasusanto, G. A., Roitch, V., Kuhn, D., Wiesemann, W.: Ambiguous joint chance constraints under mean and dispersion information. Available on optimization-online. http://www.optimization-online.org/DB_FILE/2015/11/5199.pdf (2015)
19. Henrion, R., Li, P., Möller, A., Steinbach, M.C., Wendt, M., Wozny, G.: Stochastic optimization for operating chemical processes under uncertainty. In: Grötschel, M., Krunke, S.O., Rambau, J. (eds.) Online Optimization of Large Scale Systems, pp. 457–478. Springer, Berlin (2001)
20. Henrion, R., Möller, A.: Optimization of a continuous distillation process under random inflow rate. Comput. Math. Appl. **45**(1), 247–262 (2003)
21. Jiang, R., Guan, Y.: Data-driven chance constrained stochastic program. Math. Program. **158**(1–2), 291–327 (2016)
22. Li, B., Jiang, R., Mathieu, J. L.: Distributionally robust risk-constrained optimal power flow using moment and unimodality information. In: 55th IEEE Conference on Decision and Control (CDC). IEEE (2016)
23. Miller, B., Wagner, H.: Chance constrained programming with joint constraints. Oper. Res. **13**(6), 930–945 (1965)
24. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. Wiley, Hoboken (1999)
25. Ozturk, U., Mazumdar, M., Norman, B.: A solution to the stochastic unit commitment problem using chance constrained programming. IEEE Trans. Power Syst. **19**(3), 1589–1598 (2004)
26. Popescu, I.: A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. Math. Oper. Res. **30**(3), 632–657 (2005)
27. Popescu, I.: Robust mean–covariance solutions for stochastic optimization. Oper. Res. **55**(1), 98–112 (2007)
28. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–42 (2000)
29. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Finance **26**(7), 1443–1471 (2002)
30. Scarf, H.: A min-max solution of an inventory problem. In: Arrow, K., Karlin, S., Scarf, H. (eds.) Studies in the Mathematical Theory of Inventory and Production, pp. 201–209. Stanford University Press, Palo Alto (1958)
31. Shapiro, A.: On duality theory of conic linear problems. In: Goberna, M.Á., López M.A. (eds.) Semi-Infinite Programming, pp. 135–165. Kluwer Academic Publishers, Dordrecht (2001)
32. Shapiro, A., Kleywegt, A.: Minimax analysis of stochastic problems. Optim. Methods Softw. **17**(3), 523–542 (2002)
33. Sion, M.: On general minimax theorems. Pac. J. Math. **8**(1), 171–176 (1958)
34. Stellato, B.: Data-driven chance constrained optimization. Master's thesis, ETH Zürich (2014)
35. Van Parys, B. P. G., Goulart, P. J., Kuhn, D.: Generalized Gauss inequalities via semidefinite programming. Math. Program. **156**(1–2), 271–302 (2015a)

36. Van Parys, B. P. G., Goulart, P. J., Morari, M.: Distributionally robust expectation inequalities for structured distributions. Available on optimization-online. http://www.optimization-online.org/DB_FILE/2015/05/4896.pdf (2015b)
37. Vandenberghe, L., Boyd, S., Comanor, K.: Generalized Chebyshev bounds via semidefinite programming. SIAM Rev. **49**(1), 52–64 (2007)
38. Vrakopoulou, M., Margellos, K., Lygeros, J., Andersson, G.: A probabilistic framework for reserve scheduling and $N - 1$ security assessment of systems with high wind power penetration. IEEE Trans. Power Syst. **28**(4), 3885–3896 (2013)
39. Wagner, M.: Stochastic 0–1 linear programming under limited distributional information. Oper. Res. Lett. **36**(2), 150–156 (2008)
40. Wang, Q., Guan, Y., Wang, J.: A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. IEEE Trans. Power Syst. **27**(1), 206–215 (2012)
41. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Oper. Res. **62**(6), 1358–1376 (2014)
42. Zimmerman, R.D., Murillo-Sánchez, C.E., Thomas, R.J.: MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans. Power Syst. **26**(1), 12–19 (2011)
43. Zymler, S., Kuhn, D., Rustem, B.: Distributionally robust joint chance constraints with second-order moment information. Math. Program. **137**(1–2), 167–198 (2013a)
44. Zymler, S., Kuhn, D., Rustem, B.: Worst-case value at risk of nonlinear portfolios. Manag. Sci. **59**(1), 172–188 (2013b)