manuscript

Distributionally Robust Optimization with Infinitely Constrained Ambiguity Sets

Zhi Chen
Imperial College Business School, Imperial College London
zhi.chen@imperial.ac.uk

Melvyn Sim
Department of Analytics & Operations, NUS Business School, National University of Singapore
dscsimm@nus.edu.sg

Huan Xu
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology
huan.xu@isye.gatech.edu

We consider a distributionally robust optimization problem where the ambiguity set of probability distributions is characterized by a tractable conic representable support set and by expectation constraints. We propose a new class of infinitely constrained ambiguity sets for which the number of expectation constraints could be infinite. The description of such ambiguity sets can incorporate the stochastic dominance, dispersion, fourth moment, and our newly proposed "entropic dominance" information about the uncertainty. In particular, we demonstrate that including this entropic dominance can improve the characterization of stochastic independence as compared with a characterization based solely on covariance information. Because the corresponding distributionally robust optimization problem need not lead to tractable reformulations, we adopt a greedy improvement procedure that consists of solving a sequence of tractable distributionally robust optimization subproblems—each of which considers a relaxed and finitely constrained ambiguity set. Our computational study establishes that this approach converges reasonably well.

Key words: distributionally robust optimization; stochastic programming; entropic dominance.

History: May 20, 2018.

---

1. Introduction

In recent years, *distributionally robust optimization* has become a popular approach for addressing optimization problems contaminated by uncertainty. The attractive features of distributionally robust optimization include not only its flexibility in the specification of uncertainty beyond a fixed probability distribution but also its ability to yield computationally tractable models. Its idea can be traced back to minimax stochastic programming, in which optimal solutions are evaluated under the worst-case expectation in response to a family of probability distributions of uncertain parameters (see, for instance, Scarf 1958, Dupačová 1987, Gilboa and Schmeidler 1989, Breton and El Hachem 1995, Shapiro and Kleywegt 2002).

The problem usually studied in distributionally robust optimization is as follows:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right], \tag{1}$$

where $\boldsymbol{x}\in\mathcal{X}\subseteq\mathbb{R}^M$ is the "here-and-now" decision taken before the uncertainty realizes, $\boldsymbol{z}\in\mathbb{R}^N$ is the realization of that uncertainty, and $f(\cdot,\cdot)\colon \mathcal{X}\times\mathbb{R}^N\mapsto\mathbb{R}$ is the objective function. Given the decision $\boldsymbol{x}$, the distributionally robust optimization model (1) evaluates the worst-case expected objective

$$\rho(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \tag{2}$$

with reference to the *ambiguity set*, denoted by $\mathcal{F}$. Distributionally robust optimization is a true generalization of *classical robust optimization* and *stochastic programming*. When the ambiguity set contains only the support information—that is, when $\mathcal{F}=\{\mathbb{P}\in\mathcal{P}_0(\mathbb{R}^N)\mid \tilde{\boldsymbol{z}}\sim\mathbb{P},\ \mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{W}]=1\}$, the evaluation criterion (2) reduces to the worst-case objective

$$\rho(\boldsymbol{x}) = \sup_{\boldsymbol{z}\in\mathcal{W}} f(\boldsymbol{x},\boldsymbol{z})$$

as in classical robust optimization (see, for instance, Soyster 1973, Ben-Tal and Nemirovski 1998, El Ghaoui et al. 1998, Bertsimas and Sim 2004). When the ambiguity set is a singleton (*i.e.*, $\mathcal{F}=\{\mathbb{P}_0\}$), the evaluation criterion (2) recovers the expected objective

$$\rho(\boldsymbol{x}) = \mathbb{E}_{\mathbb{P}_0}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]$$

considered in stochastic programming (see, for instance, Birge and Louveaux 2011).

The introduction of an ambiguity set leads to greater modeling power and allows us to incorporate information about the uncertainty, such as support and descriptive statistics (*e.g.*, mean, moments, and dispersions; see El Ghaoui et al. 2003, Chen et al. 2007, Chen and Sim 2009, Delage and Ye 2010, Wiesemann et al. 2014, Hanasusanto et al. 2017). Recent studies have led to significant progress on computationally tractable reformulations for distributionally robust optimization, which is closely related to the interplay between the ambiguity set and the objective function. In particular, Delage and Ye (2010) propose tractable reformulations for distributionally robust optimization problems, in which the ambiguity set is specified by the first and second moments that themselves could be uncertain. Wiesemann et al. (2014) propose a broad class of ambiguity sets with conic representable expectation constraints and a collection of nested conic representable confidence sets. The authors identify conditions under which the distributionally robust optimization problem has an explicit tractable reformulation. They also comment that from the theoretical standpoint, an optimization problem is considered tractable if it can be solved in polynomial time; yet for practical applications of general interest, the problem is tractable only when formatted as a linear or second-order conic program, or, to a lesser degree, a semidefinite program. Another popular type of ambiguity sets is based on limiting the statistical distance (in terms of certain statistical distance measure) of their members to a reference probability distribution. Through the choice of the statistical distance measure, examples of the statistical-distance-based ambiguity sets include the ϕ-divergence ambiguity set (Ben-Tal et al. 2013, Jiang and Guan 2016, Wang et al. 2016) and the Wasserstein ambiguity set (Pflug and Wozabal 2007, Gao and Kleywegt 2016, Esfahani and Kuhn 2017, Zhao and Guan 2018). More recently, as shown in Chen et al. (2017), these ambiguity sets can be characterized by a finite number of expectation constraints and they belong to the type of scenario-wise ambiguity sets with (conditional) expectation constraints based on generalized moments. While these ambiguity sets lead to tractable reformulations, they may not be able to capture important distributional information such as stochastic independence, which is ubiquitous in characterizing uncertainty.

In this paper, we proposed a new class of ambiguity set that encompasses a potentially infinite number of expectation constraints, which has the benefits of, among other things, tightening the approximation of ambiguity sets with information on stochastic independence. We call this class the *infinitely constrained ambiguity set*, which is complementary to existing ambiguity sets in the literature such as the one in Wiesemann et al. (2014). Although the corresponding distributionally robust optimization problem may not generally lead to tractable reformulations, we propose an algorithm involving a greedy improvement procedure to solve the problem. In each iteration, the algorithm yields a monotonic improvement over the previous solution. Our computational study reveals that this approach converges reasonably well.

The paper's main contributions can be summarized as follows.

1. We propose a new class of infinitely constrained ambiguity sets that would allow ambiguous probability distributions to be characterized by an infinite number of expectation constraints, which can be naturally integrated into existing ambiguity sets. We elucidate the generality of this class for characterizing distributional ambiguity when addressing distributionally robust optimization problems. We also present several examples of infinitely constrained ambiguity sets.

2. To solve the corresponding distributionally robust optimization problem, we propose a greedy improvement procedure to obtain its solution by solving a sequence of subproblems. Each of these subproblems is a tractable distributionally robust optimization problem, in which the ambiguity set is finitely constrained and is a relaxation of the infinitely constrained ambiguity set. At each iteration, we also solve a separation problem that enables us to obtain a tighter relaxation and thus a monotonic improvement over the previous solution.

3. When incorporating covariance and fourth moment information into the ambiguity set, we show that the subproblems are second-order conic programs. This differs from the literature where

the distributionally robust optimization problem with an ambiguity set involving only covariance information is usually reformulated as a semidefinite program. An important advantage of having a second-order conic optimization formulation is the availability of state-of-the-art commercial solvers that also support discrete decision variables.

4. We introduce the *entropic dominance ambiguity set*: an infinitely constrained ambiguity set that incorporates an upper bound on the uncertainty's moment-generating function. This entropic dominance can yield an improved characterization of stochastic independence than can be achieved via the existing approach based solely on covariance information. Although the corresponding separation problem is not convex, we show computationally that the trust region method works well, and that we are able to obtain less conservative solutions if the underlying uncertainties are independently distributed.

The rest of our paper proceeds as follows. We introduce and motivate the infinitely constrained ambiguity set in Section 2. In Section 3, we study the corresponding distributionally robust optimization problem and focus on relaxing the ambiguity set to obtain a tractable reformulation for the worst-case expected objective; toward that end, we propose an algorithm to tighten the relaxation. We address covariance and the fourth moment in Section 4 and Section 5 is devoted to the entropic dominance; in these sections, we also demonstrate potential applications. Proofs not included in the main text are given in the electronic companion.

Notation: Boldface uppercase and lowercase characters represent (respectively) matrices and vectors, and $[\boldsymbol{x}]_i$ or x_i denotes the i-th element of the vector $\boldsymbol{x}$. We denote the n-th standard unit basis vector by $\boldsymbol{e}_n$, the vector of 1s by $\boldsymbol{e}$, ,the vector of 0s by $\boldsymbol{0}$, and the set of positive running indices up to N by $[N] = \{1, 2, \dots, N\}$. For a proper cone $\mathcal{K} \in \mathbb{R}^N$, the set constraint $\boldsymbol{y} - \boldsymbol{x} \in \mathcal{K}$ is equivalent to the general inequality $\boldsymbol{x} \preceq \boldsymbol{y}$. We use $\mathcal{K}^*$ to signify the dual cone of $\mathcal{K}$ such that $\mathcal{K}^* = \{\boldsymbol{y} \mid \boldsymbol{y}'\boldsymbol{x} \geq 0, \forall \boldsymbol{x} \in \mathcal{K}\}$. We denote the second-order cone by $\mathcal{K}_{\text{SOC}}$, i.e., $(a, \boldsymbol{b}) \in \mathcal{K}_{\text{SOC}}$ refers to a second-order conic constraint $||\boldsymbol{b}||_2 \leq a$. For the exponential cone, we write $\mathcal{K}_{\text{EXP}} = \text{cl}\left\{(d_1, d_2, d_3) \mid d_2 > 0, d_2 e^{d_1/d_2} \leq d_3\right\}$ and its dual cone by $\mathcal{K}^*_{\text{EXP}} = \text{cl}\left\{(d_1, d_2, d_3) \mid d_1 < 0, -d_1 e^{d_2/d_1} \leq e d_3\right\}$. We denote the set of all Borel probability distributions on $\mathbb{R}^N$ by $\mathcal{P}_0(\mathbb{R}^N)$. A random variable $\tilde{\boldsymbol{z}}$ is denoted with a tilde sign and we use $\tilde{\boldsymbol{z}} \sim \mathbb{P}$, $\mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right)$ to denote $\tilde{\boldsymbol{z}}$ as an N-dimensional random variable with probability distribution $\mathbb{P}$.

2. Infinitely constrained ambiguity set

Wiesemann et al. (2014) propose a class of conic representable ambiguity sets that results in tractable reformulations of distributionally robust optimization problems with convex piecewise affine objective functions in the form of conic optimization problems. The following tractable conic ambiguity set is an expressive example of the ambiguity set in Wiesemann et al. (2014).

$$\mathcal{F}_{\mathrm{T}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{g}(\tilde{\boldsymbol{z}})\right] \preceq_{\mathcal{K}} \boldsymbol{h} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\} \tag{3}$$

with parameters $\boldsymbol{G} \in \mathbb{R}^{L \times N}$, $\boldsymbol{\mu} \in \mathbb{R}^L$, and $\boldsymbol{h} \in \mathbb{R}^{L_1}$ and with functions $\boldsymbol{g}\colon \mathbb{R}^N \mapsto \mathbb{R}^{L_1}$. The $\mathcal{K}$-epigraph of the function $\boldsymbol{g}$, that is, $\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^N \times \mathbb{R}^{L_1} \mid \boldsymbol{g}(\boldsymbol{z}) \preceq_{\mathcal{K}} \boldsymbol{u}\}$, is tractable conic representable via conic inequalities involving tractable cones. The support set $\mathcal{W}$ is also tractable conic representable. We refer readers to Ben-Tal and Nemirovski (2001) for an excellent introduction to conic representation and we refer to the nonnegative orthant, second-order cone, exponential cone, positive semidefinite cone, and their Cartesian product as the tractable cone.

In the tractable conic ambiguity set $\mathcal{F}_{\mathrm{T}}$, the equality expectation constraint specifies the mean values of $\boldsymbol{G}\tilde{\boldsymbol{z}}$. At the same time, the conic inequality expectation constraint characterizes the distributional ambiguity via a bound on mean, covariance, expected utility, and others. Observe that we can replace the conic inequality expectation constraint by a set of infinitely many expectation constraints as follows:

$$\mathcal{F}_{\mathrm{T}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{q}'\boldsymbol{g}(\tilde{\boldsymbol{z}})\right] \le \boldsymbol{q}'\boldsymbol{h},\ \forall \boldsymbol{q} \in \mathcal{K}^* \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\},$$

where $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$.

In this paper, we propose the infinitely constrained ambiguity set—by extending the tractable conic ambiguity set to incorporate a potentially infinite number of expectation constraints—as follows:

$$\mathcal{F}_{\mathrm{I}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g_i(\boldsymbol{q}_i, \tilde{\boldsymbol{z}})\right] \le h_i(\boldsymbol{q}_i),\ \forall \boldsymbol{q}_i \in \mathcal{Q}_i, \forall i \in [I] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\}; \tag{4}$$

here sets $\mathcal{Q}_i \subseteq \mathbb{R}^{M_i}$, functions $g_i\colon \mathcal{Q}_i \times \mathbb{R}^N \mapsto \mathbb{R}$, and functions $h_i\colon \mathcal{Q}_i \mapsto \mathbb{R}$ for $i \in [I]$. The support set $\mathcal{W}$ is bounded, nonempty, and tractable conic representable. We also assume that the epigraph of the function $g_i(\boldsymbol{q}_i, \boldsymbol{z})$ is tractable conic representable with respect to $\boldsymbol{z}$ for any $\boldsymbol{q}_i \in \mathcal{Q}_i$.

The generality of our infinitely constrained ambiguity set is elucidated in the following result.

Theorem 1. *Let $\mathcal{X}$ be any infinite set in the real space. Suppose that, for any $\boldsymbol{x}\in\mathcal{X}$, the function $f(\boldsymbol{x},\boldsymbol{z})\colon \mathcal{X}\times\mathbb{R}^N\mapsto\mathbb{R}$ is tractable conic representable in the variable $\boldsymbol{z}$. Then for any ambiguity set $\mathcal{F}\subseteq\mathcal{P}_0\left(\mathbb{R}^N\right)$, there exists an infinitely constrained ambiguity set $\mathcal{F}_{\mathrm{I}}\subseteq\mathcal{P}_0\left(\mathbb{R}^N\right)$ such that*

$$\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]=\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{I}}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],\ \forall\boldsymbol{x}\in\mathcal{X}.$$

Proof. We consider the ambiguity set

$$\mathcal{F}_{\mathrm{I}}=\left\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^N\right)\ \middle|\ \begin{array}{l}\tilde{\boldsymbol{z}}\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{q},\tilde{\boldsymbol{z}})\right]\leq h(\boldsymbol{q}),\ \forall\boldsymbol{q}\in\mathcal{Q}\end{array}\right\},$$

where $\mathcal{Q}\triangleq\mathcal{X}$, $h(\boldsymbol{q})\triangleq\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{q},\tilde{\boldsymbol{z}})\right]$, and $g(\boldsymbol{q},\boldsymbol{z})\triangleq f(\boldsymbol{q},\boldsymbol{z})$. Observe that the function $g(\boldsymbol{q},\boldsymbol{z})$ is tractable conic representable in $\boldsymbol{z}$ for any $\boldsymbol{q}\in\mathcal{Q}$ and that the set $\mathcal{Q}$ is infinite; hence the constructive set $\mathcal{F}_{\mathrm{I}}$ is an infinitely constrained ambiguity set.

On the one hand, for any $\mathbb{P}\in\mathcal{F}_{\mathrm{I}}$ we have

$$\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]=\mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]\leq h(\boldsymbol{x})=\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],\ \forall\boldsymbol{x}\in\mathcal{X},$$

which implies that

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{I}}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]\leq\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],\ \forall\boldsymbol{x}\in\mathcal{X}.$$

On the other hand, for any $\mathbb{P}\in\mathcal{F}$ we have

$$\mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]=\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]\leq\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]=h(\boldsymbol{x}),\ \forall\boldsymbol{x}\in\mathcal{X};$$

hence $\mathbb{P}\in\mathcal{F}_{\mathrm{I}}$ and $\mathcal{F}\subseteq\mathcal{F}_{\mathrm{I}}$. As a result,

$$\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]\leq\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{I}}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],\ \forall\boldsymbol{x}\in\mathcal{X}.$$

We can therefore conclude that

$$\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]=\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{I}}}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],\ \forall\boldsymbol{x}\in\mathcal{X}.\quad\square$$

Using Theorem 1, we can represent the distributionally robust optimization problem (1) with any ambiguity set $\mathcal{F}$ as one with an infinitely constrained ambiguity set $\mathcal{F}_{\mathrm{I}}$—provided the objective function $f(\boldsymbol{x},\boldsymbol{z})$ is tractable conic representable in $\boldsymbol{z}$ for any $\boldsymbol{x}\in\mathcal{X}$. This observation reveals the generality of the infinitely constrained ambiguity set for characterizing distributional ambiguity in distributionally robust optimization models. We emphasize that $\mathcal{F}_{\mathrm{I}}$ is not necessarily a substitution of existing ambiguity sets but can also be naturally integrated with them: for any ambiguity set $\mathcal{F}$, we can always consider an intersection $\mathcal{F}\bigcap\mathcal{F}_{\mathrm{I}}$, which shall inherit the modeling power from both ambiguity sets. The infinitely constrained ambiguity set $\mathcal{F}_{\mathrm{I}}$ indeed enables us to characterize some new properties of uncertain probability distributions. We now provide some examples.

Stochastic dominance ambiguity set

Stochastic dominance, particularly in the second order, is a form of stochastic ordering that is prevalent in decision theory and economics.

DEFINITION 1. A random outcome $\tilde{z} \sim \mathbb{P}$ dominates another random outcome $\tilde{z}^\dagger \sim \mathbb{P}^\dagger$ in the second order, written $\tilde{z} \succeq_{(2)} \tilde{z}^\dagger$, if $\mathbb{E}_{\mathbb{P}}\left[u(\tilde{z})\right] \geq \mathbb{E}_{\mathbb{P}^\dagger}\left[u(\tilde{z}^\dagger)\right]$ for every concave nondecreasing function $u(\cdot)$, for which these expected values are finite.

The celebrated expected utility hypothesis, introduced by von Neumann and Morgenstern (1947), states that: for every rational decision maker, there exists a utility function $u(\cdot)$ such that she prefers the random outcome $\tilde{z}$ over some random outcome $\tilde{z}^\dagger$ if and only if $\mathbb{E}_{\mathbb{P}}\left[u(\tilde{z})\right] > \mathbb{E}_{\mathbb{P}^\dagger}\left[u(\tilde{z}^\dagger)\right]$. In practice, however, it is almost impossible for a decision maker to elicit her utility function. Instead, Dentcheva and Ruszczyński (2003) propose a safe relaxation of the constraint $\mathbb{E}_{\mathbb{P}}\left[u(\tilde{z})\right] > \mathbb{E}_{\mathbb{P}^\dagger}\left[u(\tilde{z}^\dagger)\right]$ by considering the second-order stochastic dominance constraint $\tilde{z} \succeq_{(2)} \tilde{z}^\dagger$. This constraint ensures that, without articulating her utility function, the decision maker will prefer the payoff $\tilde{z}$ over the benchmark payoff $\tilde{z}^\dagger$ as long as her utility function $u(\cdot)$ is concave and nondecreasing. The proposal is practically useful in many applications. In portfolio optimization problems, for instance, one can construct the reference portfolio from past return rates and then attempt to devise a more nearly optimal portfolio under the constraint that it dominates the reference portfolio (see Dentcheva and Ruszczyński 2006). The second-order stochastic dominance constraint $\tilde{z} \succeq_{(2)} \tilde{z}^\dagger$ has an equivalent representation

$$\mathbb{E}_{\mathbb{P}}[(q - \tilde{z})^+] \leq \mathbb{E}_{\mathbb{P}^\dagger}[(q - \tilde{z}^\dagger)^+], \ \forall q \in \mathbb{R},$$

which can be readily incorporated into an infinitely constrained ambiguity set. In fact, the modeling capability of that set extends to higher-order stochastic dominance relations; for more details, see Dentcheva and Ruszczyński (2003).

Mean-dispersion ambiguity set

In the class of mean-dispersion ambiguity sets, probability distributions of the uncertainty are characterized by the support and mean and by upper bounds on their dispersion. Mean-dispersion ambiguity sets recover a range of ambiguity sets from the literature and have been recently studied in Hanasusanto et al. (2017) for joint chance constraints. A typical mean-dispersion ambiguity set $\mathcal{F}_{\mathrm{D}}$ satisfies

$$\mathcal{F}_{\mathrm{D}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})\right] \leq h(\boldsymbol{q}), \ \forall \boldsymbol{q} \in \mathcal{Q} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}, \tag{5}$$

where the dispersion of the uncertainty $\tilde{\boldsymbol{z}}$ along a direction $\boldsymbol{q} \in \mathcal{Q}$ (in terms of the expectation corresponding to the dispersion function $g(\cdot,\cdot)$) is bounded from above by $h(\boldsymbol{q})$.

By varying the dispersion measure, the mean-dispersion ambiguity set includes many interesting examples. Setting $g(\boldsymbol{q},\boldsymbol{z}) = |\boldsymbol{q}'(\boldsymbol{z}-\boldsymbol{\mu})|$, the mean-dispersion ambiguity set is closely related to the mean absolute deviation ambiguity set (denoted by $\mathcal{F}_{\text{MAD}}$) that has been studied in Postek et al. (2018), where the authors propose to use $h(\boldsymbol{q}) = \sqrt{\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}}$ with $\boldsymbol{\Sigma}$ being the covariance matrix of $\tilde{\boldsymbol{z}}$. Setting $g(\boldsymbol{q},\boldsymbol{z}) = H_\delta(\boldsymbol{q}'(\boldsymbol{z}-\boldsymbol{\mu}))$, where $H_\delta(z)$ is the Huber loss function with the prescribed parameter $\delta > 0$ that takes $z^2/2$ if $|z| \le \delta$ and $\delta(|z|-\delta/2)$ otherwise, the mean-dispersion ambiguity set recovers the Huber ambiguity set (denoted by $\mathcal{F}_{\text{H}}$) that imposes upper bounds on the expected Huber loss function. Through the choice $\mathcal{Q} = \{\boldsymbol{q} \in \mathbb{R}^N \mid \|\boldsymbol{q}\|_2 \le 1\}$, $g(\boldsymbol{q},\boldsymbol{z}) = (\boldsymbol{q}'(\boldsymbol{z}-\boldsymbol{\mu}))^2$, and $h(\boldsymbol{q}) = \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}$, the mean-dispersion ambiguity set becomes

$$\mathcal{F}_{\text{C}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2\right] \le \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q},\ \forall \boldsymbol{q} \in \mathcal{Q} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}. \tag{6}$$

Since a symmetric matrix $\boldsymbol{\Sigma}$ is positive semidefinite if and only if that $\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q} \ge \boldsymbol{0}$ for all $\boldsymbol{q} \in \mathbb{R}^N$, the above ambiguity set is essentially the covariance ambiguity set

$$\mathcal{F}_{\text{C}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})'\right] \preceq \boldsymbol{\Sigma} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}, \tag{7}$$

which is closely related to that proposed in Delage and Ye (2010). This observation implies that the covariance ambiguity set captures the dispersion by an upper bound on the covariance of the uncertainty, or, by an upper bound $\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}$ on the variance of $\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})$ along any direction $\boldsymbol{q} \in \mathbb{R}^N$. The existing literature usually expresses covariance information as in (7) and reformulates the corresponding distributionally robust optimization problem to a semidefinite program. We will show that modeling covariance information from an infinitely constrained perspective as in (6) enables us to obtain a scalable second-order conic relaxation.

For any $\boldsymbol{q} \in \mathbb{R}^N$ and for any $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^N)$, the relation $\mathbb{E}_{\mathbb{P}}\left[|\boldsymbol{q}'(\boldsymbol{z}-\boldsymbol{\mu})|\right] \le \sqrt{\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2\right]}$ follows from Jensen's inequality, which implies that a mean absolute deviation ambiguity set $\mathcal{F}_{\text{MAD}}$ with $h(\boldsymbol{q}) = \sqrt{\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}}$ satisfies $\mathcal{F}_{\text{C}} \subseteq \mathcal{F}_{\text{MAD}}$. Since $H_\delta(\boldsymbol{q}'(\boldsymbol{z}-\boldsymbol{\mu})) \le \frac{1}{2}(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2$ for all $\boldsymbol{q} \in \mathbb{R}^N$, a Huber ambiguity set $\mathcal{F}_{\text{H}}$ with $h(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}$ satisfies $\mathcal{F}_{\text{C}} \subseteq \mathcal{F}_{\text{H}}$. Therefore, both $\mathcal{F}_{\text{MAD}}$ and $\mathcal{F}_{\text{H}}$ are supersets of the covariance ambiguity set. We next present two important examples of mean-dispersion ambiguity sets that are subsets of the covariance ambiguity set and thus can provide a tighter characterization of some properties of the uncertainty than does a covariance ambiguity set.

Fourth moment ambiguity set

Here we consider a class of independently distributed random variables $\{\tilde{z}_n\}_{n\in[N]}$ with identical zero first moments and nonidentical second and fourth moments, which are bounded from above by (respectively) σ_n^2 and κ_n^4. This class of random variables is not uncommon: a typical example is the underlying random factors of the popular factor-based model. Incorporating the fourth moment is not only straightforward but also useful; He et al. (2010) point out that, in many applications, it is relatively easy to bound the fourth moment of a random variable. The authors use the fourth moment, as well as the first and second moments, to derive tighter upper bounds on the probability that a random variable deviates from its mean by a small amount. Moreover, the second and fourth moments of a random variable $\tilde{z}_n$ with zero first moment are closely related to its kurtosis, which is given by κ_n^4/σ_n^4 and is an important measure of tailed-ness of a random variable's probability distribution.

Let $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^N)$ be the joint probability distribution of these random variables. The fourth moment of their linear combination, $\boldsymbol{q}'\tilde{\boldsymbol{z}} = \sum_{n\in[N]} q_n \tilde{z}_n$, is bounded from above by

$$\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^4\right] \leq \sum_{n\in[N]} \left(q_n^4 \kappa_n^4 + 6 \sum_{m\neq n} q_m^2 q_n^2 \sigma_m^2 \sigma_n^2 \right) = \varphi(\boldsymbol{q}).$$

Hence, the fourth moment ambiguity set is given by

$$\mathcal{F}_{\mathrm{F}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \mathbf{0} \\ \mathbb{E}_{\mathbb{P}}[(\hat{\boldsymbol{q}}'\tilde{\boldsymbol{z}})^2] \leq \hat{\boldsymbol{q}}'\boldsymbol{\Sigma}\hat{\boldsymbol{q}},\ \forall \hat{\boldsymbol{q}} \in \mathbb{R}^N : \|\hat{\boldsymbol{q}}\|_2 \leq 1 \\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^4] \leq \varphi(\boldsymbol{q}),\ \forall \boldsymbol{q} \in \mathbb{R}^N : \|\boldsymbol{q}\|_2 \leq 1 \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{W}] = 1 \end{array} \right\}, \tag{8}$$

where $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$ and the support set $\mathcal{W}$ describes (possibly nonidentical) supports of $\{\tilde{z}_n\}_{n\in[N]}$. It is clear that $\mathcal{F}_{\mathrm{F}} \subseteq \mathcal{F}_{\mathrm{C}}$ for some covariance ambiguity set $\mathcal{F}_{\mathrm{C}}$ with $\boldsymbol{\mu} = \mathbf{0}$. Thus, the fourth moment ambiguity set yields a more precise characterization of this class of independently distributed random variables than does a covariance ambiguity set. It is worth mentioning that the fourth moment ambiguity set $\mathcal{F}_{\mathrm{F}}$ relaxes the independence between these random variables to uncorrelation, thus it only approximately captures the statistical independence.

Entropic dominance ambiguity set

We now introduce the entropic dominance ambiguity set, which is defined as follows:

$$\mathcal{F}_{\mathrm{E}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu} \\ \ln \mathbb{E}_{\mathbb{P}}[\exp(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))] \leq \phi(\boldsymbol{q}),\ \forall \boldsymbol{q} \in \mathbb{R}^N \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{W}] = 1 \end{array} \right\}; \tag{9}$$

here $\phi\colon \mathbb{R}^N \mapsto \mathbb{R}$ is some convex and twice continuously differentiable function that satisfies $\phi(\mathbf{0}) = 0$ and $\nabla\phi(\mathbf{0}) = \mathbf{0}$ for $\nabla\phi(\cdot)$ the gradient of $\phi(\cdot)$. The entropic dominance ambiguity set virtually bounds from above the logged moment-generating function of the random variable $\tilde{\boldsymbol{z}}$ as adjusted by its mean $\boldsymbol{\mu}$. Hence, the entropic dominance ambiguity set can encompass common random variables used in statistics—such as *sub-Gaussian* (Wainwright 2015), a class of random variables with exponentially decaying tails. A formal definition follows.

Definition 2. A random variable $\tilde{z}$ with mean μ is *sub-Gaussian* with deviation parameter $\sigma > 0$, if

$$\ln \mathbb{E}_{\mathbb{P}}\left[\exp(q(\tilde{z} - \mu))\right] \le \frac{q^2\sigma^2}{2}, \ \forall q \in \mathbb{R}.$$

In the case of a normal distribution, the deviation parameter σ corresponds to the standard deviation and we have equality in the relation just defined. So if $\{\tilde{z}_n\}_{n\in[N]}$ are independently distributed sub-Gaussian random variables with means $\{\mu_n\}_{n\in[N]}$ and deviation parameters $\{\sigma_n\}_{n\in[N]}$, then we can specify the entropic dominance ambiguity set (9) by letting $\phi(\boldsymbol{q}) = \frac{1}{2}\sum_{n\in[N]} q_n^2\sigma_n^2$.

It is worth noting that the class of infinitely many expectation constraints implicitly specifies the mean, an upper bound on the covariance, and a superset of the support of a probability distribution.

Theorem 2. *For any probability distribution $\mathbb{P}$ that satisfies*

$$\ln \mathbb{E}_{\mathbb{P}}\left[\exp(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))\right] \le \phi(\boldsymbol{q}), \ \forall \boldsymbol{q} \in \mathbb{R}^N,$$

we have: $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}$, $\mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'] \preceq \nabla^2\phi(\mathbf{0})$, *and*

$$\mathcal{W} \subseteq \left\{ \boldsymbol{z} \in \mathbb{R}^N \;\middle|\; \boldsymbol{q}'(\boldsymbol{z} - \boldsymbol{\mu}) \le \lim_{\alpha\to\infty} \frac{1}{\alpha}\phi(\alpha\boldsymbol{q}), \ \forall \boldsymbol{q} \in \mathbb{R}^N \right\}.$$

If $\phi(\boldsymbol{q})$ has an additive form given by $\phi(\boldsymbol{q}) = \sum_{n\in[N]} \phi_n(q_n)$, where $\phi_n(0) = 0$ for all $n \in [N]$, then we have the following simplified results: $\nabla^2\phi(0) = \operatorname{diag}(\phi_1''(0), \ldots, \phi_N''(0))$ *and*

$$\mathcal{W} \subseteq \left\{ \boldsymbol{z} \in \mathbb{R}^N \;\middle|\; \mu_n + \lim_{\alpha\to-\infty} \frac{1}{\alpha}\phi_n(\alpha) \le z_n \le \mu_n + \lim_{\alpha\to\infty} \frac{1}{\alpha}\phi_n(\alpha), \ \forall n \in [N] \right\}.$$

Theorem 2 implies that every entropic dominance ambiguity set is a subset of some covariance ambiguity set; this implication, as we will demonstrate in the sections to follow, is helpful in developing an improved characterization of stochastic independence among uncertain components. We remark that Theorem 2 works only with the class of infinitely many expectation constraints and that one's choice of the function $\phi(\cdot)$ may lead to an "invalid" support (for example, choosing $\phi(q) = \frac{1}{2}\boldsymbol{q}\boldsymbol{q}'$ implies only that the support set $\mathcal{W} \subseteq \mathbb{R}^N$). Therefore, we explicitly stipulate the mean and support in the entropic dominance ambiguity set $\mathcal{F}_{\mathrm{E}}$.

Before proceeding, we showcase yet another application of the entropic dominance ambiguity set in providing a new bound for the expected surplus of an affine function of a set $\{\tilde{z}_n\}_{n\in[N]}$ of independently distributed random variables:

$$\rho_0(x_0,\boldsymbol{x}) = \mathbb{E}_{\mathbb{P}_0}\left[(x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right]. \tag{10}$$

This term frequently appears in the literature on risk management and operations management as in the newsvendor problem and the inventory control problem. Note that computing the exact value of $\rho_0(x_0,\boldsymbol{x})$ involves high-dimensional integration and has been shown to be #P-hard even when all of $\{\tilde{z}_n\}_{n\in[N]}$ are uniformly distributed (Hanasusanto et al. 2016). Even so, its upper bound has been proved useful for providing tractable approximations to stochastic programming and chance-constrained optimization problems (see, for instance, Nemirovski and Shapiro 2006, Chen et al. 2008, Chen et al. 2010, Goh and Sim 2010). A well-known approach that gives an upper bound of Problem (10) is based on the observation that $\omega^+ \leq \eta\exp(\omega/\eta-1)$ for all $\eta>0$. It follows that, under stochastic independence, we can obtain an upper bound on Problem (10) by optimizing the problem

$$\rho_{\mathrm{B}}(x_0,\boldsymbol{x}) = \inf_{\eta>0}\mathbb{E}_{\mathbb{P}_0}\left[\frac{\eta}{e}\exp\left(\frac{x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right] = \inf_{\eta>0}\frac{\eta}{e}\exp\left(\frac{x_0+\boldsymbol{x}'\boldsymbol{\mu}}{\eta}+\sum_{n\in[N]}\phi_n\left(\frac{x_n}{\eta}\right)\right).$$

In this expression, we have $\phi_n(q)=\ln\mathbb{E}_{\mathbb{P}_n}\left[\exp\left(q(\tilde{z}_n-\mu_n)\right)\right]$ for $\mathbb{P}_n$ (resp., μ_n) the marginal distribution (resp., the mean) of $\tilde{z}_n$. The following result presents a new approach—based on the entropic dominance ambiguity set—to bounding $\rho_0(x_0,\boldsymbol{x})$.

Proposition 1. *Consider the entropic dominance ambiguity set*

$$\mathcal{F}_{\mathrm{E}} = \left\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l}\tilde{\boldsymbol{z}}\sim\mathbb{P}\\ \ln\mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})\right)\right]\leq\sum_{n\in[N]}\phi_n(f_n),\ \forall\boldsymbol{q}\in\mathbb{R}^N\end{array}\right\}$$

and let $\rho_{\mathrm{E}}(x_0,\boldsymbol{x})=\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{E}}}\mathbb{E}_{\mathbb{P}}[(x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}})^+]$*. Then, for all* $\boldsymbol{x}\in\mathbb{R}^N$*, we have*

$$\rho_0(x_0,\boldsymbol{x})\leq\rho_{\mathrm{E}}(x_0,\boldsymbol{x})\leq\rho_{\mathrm{B}}(x_0,\boldsymbol{x}).$$

The value of $\rho_{\mathrm{E}}(x_0,\boldsymbol{x})$ does not generally equal the value of $\rho_{\mathrm{B}}(x_0,\boldsymbol{x})$, as we illustrate next.

Proposition 2. *Let* $N=1$ *and* $\mathbb{P}_0$ *be the standard normal distribution. Then*

$$\rho_0(0,1)=\frac{1}{\sqrt{2\pi}},\qquad\rho_{\mathrm{E}}(0,1)=\frac{1}{2},\qquad\rho_{\mathrm{B}}(0,1)=\frac{1}{\sqrt{e}}.$$

We can extend this approach to the case of distributional ambiguity if the underlying random variables are independently distributed. Suppose $\tilde{z}_n$ has the ambiguous distribution $\mathbb{P}_n\in\mathcal{F}_n$. In that case, we can define the function ϕ_n as the tightest upper bound on its logged moment-generating function, as adjusted by its mean, in the way: $\phi_n(q)=\sup_{\mathbb{P}\in\mathcal{F}_n}\ln\mathbb{E}_{\mathbb{P}}\left[\exp(q(\tilde{z}_n-\mu_n))\right]$. For example, if $\tilde{z}_n$ were an ambiguous sub-Gaussian random variable with mean μ_n and deviation parameter σ_n, then we would define $\phi_n(q)=\frac{1}{2}q^2\sigma_n^2$.

3. Distributionally robust optimization model and solution procedure

Our aim in this paper is to obtain the optimal solution (or a reasonable approximation thereof) to the distributionally robust optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_{\rm I}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] = \min_{\boldsymbol{x}\in\mathcal{X}} \rho_{\rm I}(\boldsymbol{x}), \tag{11}$$

where $\mathcal{X}\subseteq\mathbb{R}^M$ is the feasible set and $\rho_{\rm I}(\boldsymbol{x})$ is the worst-case expected objective over the convex and compact infinitely constrained ambiguity set $\mathcal{F}_{\rm I}$. Observe that Problem (11) may not be tractable, because standard reformulation approaches based on duality results (Isii 1962, Shapiro 2001, Bertsimas and Popescu 2005) could lead to infinitely many dual variables associated with the expectation constraints. So instead, we consider the following relaxed ambiguity set:

$$\mathcal{F}_{\rm R} = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}}\sim\mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g_i(\boldsymbol{q}_i,\tilde{\boldsymbol{z}})\right] \le h_i(\boldsymbol{q}_i),\ \forall \boldsymbol{q}_i\in\bar{\mathcal{Q}}_i, \forall i\in[I] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}}\in\mathcal{W}\right] = 1 \end{array} \right\}; \tag{12}$$

here, for $i\in[I]$, the set $\bar{\mathcal{Q}}_i = \{\boldsymbol{q}_{ij} \mid j\in[J_i]\}\subseteq\mathcal{Q}_i$ is a finite approximation of $\mathcal{Q}_i$ that leads to $\mathcal{F}_{\rm I}\subseteq\mathcal{F}_{\rm R}$. To obtain an explicit formulation in the tractable conic optimization format, we define the epigraphs of $g_i, i\in[I]$ and the support set $\mathcal{W}$ as follows:

$$\bar{\mathcal{W}} = \left\{ (\boldsymbol{z},\boldsymbol{u})\in\mathbb{R}^N\times\mathbb{R}^J \mid \boldsymbol{z}\in\mathcal{W},\ g_i(\boldsymbol{q}_{ij},\boldsymbol{z})\le u_{ij},\ \forall i\in[I], \forall j\in[J_i] \right\},$$

where $J=\sum_{i\in[I]} J_i$. Here we have used the concept of conic representation and also make the following assumption.

Assumption 1. *The conic representation of the system*

$$\begin{array}{l} \boldsymbol{G}\boldsymbol{z} = \boldsymbol{\mu} \\ g_i(\boldsymbol{q}_{ij},\boldsymbol{z}) \le h_i(\boldsymbol{q}_{ij}),\ \forall i\in[I], \forall j\in[J_i] \\ \boldsymbol{z}\in\mathcal{W} \end{array}$$

satisfies Slater's condition (see, Ben-Tal and Nemirovski 2001, Theorem 1.4.2).

With respect to the relaxed ambiguity set $\mathcal{F}_{\rm R}$, we now consider the relaxed distributionally robust optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_{\rm R}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] = \min_{\boldsymbol{x}\in\mathcal{X}} \rho_{\rm R}(\boldsymbol{x}). \tag{13}$$

Note that $\rho_{\rm I}(\boldsymbol{x})\le\rho_{\rm R}(\boldsymbol{x})$ and recall that our goal is to achieve sequentially better solutions by tightening the relaxation of $\mathcal{F}_{\rm I}$ (i.e., $\mathcal{F}_{\rm R}$). We shall focus on the convex and piecewise affine objective function

$$f(\boldsymbol{x},\boldsymbol{z}) = \max_{k\in[K]} f_k(\boldsymbol{x},\boldsymbol{z}) = \max_{k\in[K]} \left\{ \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} + b_k(\boldsymbol{x}) \right\},$$

where for all $k \in [K]$, $\boldsymbol{a}_k : \mathcal{X} \mapsto \mathbb{R}^N$ and $b_k : \mathcal{X} \mapsto \mathbb{R}$ are some affine functions. The requirement on the objective function can be extended to a richer class of functions such that $f_k(\boldsymbol{x}, \boldsymbol{z})$ is convex in $\boldsymbol{x}$ (given $\boldsymbol{z}$) and is concave in $\boldsymbol{z}$ (given $\boldsymbol{x}$). In such cases, the resulting reformulation of Problem (13) would vary accordingly; for detailed discussions, we refer interested readers to Wiesemann et al. (2014).

In the following result, we recast the inner supremum problem $\rho_{\mathrm{R}}(\boldsymbol{x})$ as a minimization problem, which can be performed jointly with the outer minimization over $\boldsymbol{x}$.

THEOREM 3. *Given the relaxed ambiguity set $\mathcal{F}_{\mathrm{R}}$, the relaxed worst-case expected objective $\rho_{\mathrm{R}}(\boldsymbol{x})$ is the same as the optimal value of the following tractable conic optimization problem:*

$$\begin{array}{lll} \inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{i \in [I]} \sum_{j \in [J_i]} \gamma_{ij} h_i(\boldsymbol{q}_{ij}) & \\ \text{s.t.} & \alpha - b_k(\boldsymbol{x}) - t_k \geq 0, & \forall k \in [K] \\ & \boldsymbol{G}'\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k = \boldsymbol{0}, & \forall k \in [K] \\ & \boldsymbol{\gamma} - \boldsymbol{s}_k = \boldsymbol{0}, & \forall k \in [K] \\ & (\boldsymbol{r}_k, \boldsymbol{s}_k, t_k) \in \mathcal{K}^*, & \forall k \in [K] \\ & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^L, \boldsymbol{\gamma} \in \mathbb{R}^J_+, & \end{array} \tag{14}$$

where $\mathcal{K}^*$ is the dual cone of $\mathcal{K} = \mathrm{cl}\{(\boldsymbol{z}, \boldsymbol{u}, t) \in \mathbb{R}^N \times \mathbb{R}^J \times \mathbb{R} \mid (\boldsymbol{z}/t, \boldsymbol{u}/t) \in \bar{\mathcal{W}},\, t > 0\}$. Furthermore, Problem (14) is solvable—that is, its optimal value is attainable.

Proof. Introducing dual variables α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ that correspond to the respective probability and expectation constraints in $\mathcal{F}_{\mathrm{R}}$, we obtain the dual of $\rho_{\mathrm{R}}(\boldsymbol{x})$ as follows:

$$\begin{array}{rll} \rho_0(\boldsymbol{x}) = & \inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{i \in [I]} \sum_{j \in [J_i]} \gamma_{ij} h_i(\boldsymbol{q}_{ij}) \\ & \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{G}\boldsymbol{z} + \sum_{i \in [I]} \sum_{j \in [J_i]} \gamma_{ij} g_i(\boldsymbol{q}_{ij}, \boldsymbol{z}) \geq f(\boldsymbol{x}, \boldsymbol{z}),\ \forall \boldsymbol{z} \in \mathcal{W} \\ & & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^L, \boldsymbol{\gamma} \in \mathbb{R}^J_+, \end{array} \tag{15}$$

which provides an upper bound on $\rho_{\mathrm{R}}(\boldsymbol{x})$. Indeed, consider any $\mathbb{P} \in \mathcal{F}_{\mathrm{R}}$ and any feasible solution $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ in Problem (15); the robust counterpart in the dual implies that

$$\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq \mathbb{E}_{\mathbb{P}}\left[\alpha + \boldsymbol{\beta}'\boldsymbol{G}\tilde{\boldsymbol{z}} + \sum_{i \in [I]} \sum_{j \in [J_i]} \gamma_{ij} g_i(\boldsymbol{q}_{ij}, \tilde{\boldsymbol{z}})\right] \leq \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{i \in [I]} \sum_{j \in [J_i]} \gamma_{ij} h_i(\boldsymbol{q}_{ij}).$$

Thus weak duality follows; that is, $\rho_{\mathrm{R}}(\boldsymbol{x}) \leq \rho_0(\boldsymbol{x})$. We next establish the equality $\rho_{\mathrm{R}}(\boldsymbol{x}) = \rho_0(\boldsymbol{x})$.

The robust counterpart can be written more compactly as a set of robust counterparts

$$\alpha + \boldsymbol{\beta}'\boldsymbol{G}\boldsymbol{z} + \sum_{i \in [I]} \sum_{j \in [J_i]} \gamma_{ij} g_i(\boldsymbol{q}_{ij}, \boldsymbol{z}) - \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} - b_k(\boldsymbol{x}) \geq 0,\ \forall \boldsymbol{z} \in \mathcal{W}, \forall k \in [K].$$

After introducing the auxiliary variable $\boldsymbol{u}$, we can state that each robust counterpart is not violated if and only if

$$\begin{aligned} \delta = \inf\ & (\boldsymbol{G}'\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\boldsymbol{z} + \sum_{i\in[I]}\sum_{j\in[J_i]} \gamma_{ij}u_{ij} \\ \text{s.t.}\ & (\boldsymbol{z},\boldsymbol{u},1)\in\mathcal{K}, \end{aligned}$$

is not less than $b_k(\boldsymbol{x}) - \alpha$. The dual of the preceding problem is given by

$$\begin{aligned} \delta' = \sup\ & -t_k \\ \text{s.t.}\ & \boldsymbol{G}'\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k = \boldsymbol{0} \\ & \boldsymbol{\gamma} - \boldsymbol{s}_k = \boldsymbol{0} \\ & (\boldsymbol{r}_k, \boldsymbol{s}_k, t_k) \in \mathcal{K}^*, \end{aligned}$$

where $\delta' \le \delta$ follows from weak duality. Re-injecting the dual formulation into Problem (15), we obtain a problem with a smaller feasible set of $(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma})$ and hence a larger objective value:

$$\begin{aligned} \rho_0(\boldsymbol{x}) \le \rho_1(\boldsymbol{x}) = \inf\ & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{i\in[I]}\sum_{j\in[J_i]}\gamma_{ij}h_i(\boldsymbol{q}_{ij}) \\ \text{s.t.}\ & \alpha - b_k(\boldsymbol{x}) - t_k \ge 0, && \forall k\in[K] \\ & \boldsymbol{G}'\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k = \boldsymbol{0}, && \forall k\in[K] \\ & \boldsymbol{\gamma} - \boldsymbol{s}_k = \boldsymbol{0}, && \forall k\in[K] \\ & (\boldsymbol{r}_k,\boldsymbol{s}_k,t_k)\in\mathcal{K}^*, && \forall k\in[K] \\ & \alpha\in\mathbb{R}, \boldsymbol{\beta}\in\mathbb{R}^L, \boldsymbol{\gamma}\in\mathbb{R}^J_+. \end{aligned} \tag{16}$$

By conic duality, the dual of Problem (16) is given by

$$\begin{aligned} \rho_2(\boldsymbol{x}) = \sup\ & \sum_{k\in[K]} (\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k) \\ \text{s.t.}\ & \sum_{k\in[K]} \eta_k = 1 \\ & \sum_{k\in[K]} \boldsymbol{G}\boldsymbol{\xi}_k = \boldsymbol{\mu} \\ & \sum_{k\in[K]} [\boldsymbol{\zeta}_k]_{ij} \le h_i(\boldsymbol{q}_{ij}), && \forall i\in[I], \forall j\in[J_i] \\ & (\boldsymbol{\xi}_k,\boldsymbol{\zeta}_k,\eta_k)\in\mathcal{K}, && \forall k\in[K] \\ & \boldsymbol{\xi}_k\in\mathbb{R}^N, \boldsymbol{\zeta}_k\in\mathbb{R}^J, \eta_k\in\mathbb{R}_+,\ \forall k\in[K], \end{aligned} \tag{17}$$

where $\boldsymbol{\xi}_k, \boldsymbol{\zeta}_k, \eta_k$, $k\in[K]$ are dual variables associated with the respective constraints and where the conic constraint follows from $\mathcal{K}^{**} = \mathcal{K}$.

Under Assumption 1, Slater's condition holds for Problem (17); hence strong duality holds (i.e., $\rho_1(\boldsymbol{x}) = \rho_2(\boldsymbol{x})$). In fact, there exists a sequence $\{(\bar{\boldsymbol{\xi}}_k^l, \bar{\eta}_k^l)_{k\in[K]}\}_{l\ge 0}$ of strictly feasible solutions to Problem (17) such that

$$\lim_{l\to\infty} \sum_{k\in[K]} \left(\boldsymbol{a}_k(\boldsymbol{x})'\bar{\boldsymbol{\xi}}_k + b_k(\boldsymbol{x})\bar{\eta}_k\right) = \rho_2(\boldsymbol{x}).$$

We can therefore construct a sequence of discrete probability distributions $\{\mathbb{P}_l \in \mathcal{P}_0(\mathbb{R}^N)\}_{l\geq 0}$ on the random variable $\tilde{\boldsymbol{z}} \in \mathbb{R}^N$ as follows:

$$\mathbb{P}_l\left[\tilde{\boldsymbol{z}} = \frac{\bar{\boldsymbol{\xi}}_k^l}{\bar{\eta}_k^l}\right] = \bar{\eta}_k^l, \ \forall k \in [K].$$

For all $l \geq 0$, we have: $\mathbb{E}_{\mathbb{P}_l}[\boldsymbol{G}\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}$; $\mathbb{E}_{\mathbb{P}_l}[g_i(\boldsymbol{q}_{ij}, \tilde{\boldsymbol{z}})] \leq h_i(\boldsymbol{q}_{ij})$ for all $i \in [I]$ and for all $j \in [J_i]$; and $\boldsymbol{\xi}_k^l/\eta_k^l \in \mathcal{W}$ for all $k \in [K]$. It follows that $\mathbb{P}_l \in \mathcal{F}_{\mathrm{R}}$. In addition, we have

$$\begin{aligned}
\rho_2(\boldsymbol{x}) &= \lim_{l\to\infty} \sum_{k\in[K]} \bar{\eta}_k^l \Big(\boldsymbol{a}_k'(\boldsymbol{x})\frac{\bar{\boldsymbol{\xi}}_k^l}{\bar{\eta}_k^l} + b_k(\boldsymbol{x})\Big) \\
&\leq \lim_{l\to\infty} \sum_{k\in[K]} \bar{\eta}_k^l \Big(\max_{n\in[K]}\Big\{\boldsymbol{a}_n(\boldsymbol{x})'\frac{\bar{\boldsymbol{\xi}}_k^l}{\bar{\eta}_k^l} + b_n(\boldsymbol{x})\Big\}\Big) \\
&= \lim_{l\to\infty} \mathbb{E}_{\mathbb{P}_l}\Big[\max_{n\in[K]}\{\boldsymbol{a}_n(\boldsymbol{x})'\tilde{\boldsymbol{z}} + b_n(\boldsymbol{x})\}\Big] \\
&\leq \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] \\
&= \rho_{\mathrm{R}}(\boldsymbol{x}).
\end{aligned}$$

Because $\rho_{\mathrm{R}}(\boldsymbol{x}) \leq \rho_0(\boldsymbol{x}) \leq \rho_1(\boldsymbol{x})$, we obtain $\rho_{\mathrm{R}}(\boldsymbol{x}) \leq \rho_0(\boldsymbol{x}) \leq \rho_1(\boldsymbol{x}) = \rho_2(\boldsymbol{x}) \leq \rho_{\mathrm{R}}(\boldsymbol{x})$. It now follows from Ben-Tal and Nemirovski (2001, Theoremm 1.4.2) that Problem (14), which is the dual of Problem (17), is solvable. □

Theorem 3 provides two equivalent reformulations of the worst-case expected objective $\rho_{\mathrm{R}}(\boldsymbol{x})$. The minimization reformulation $\rho_1(\boldsymbol{x})$ enables us to solve the relaxed distributionally robust optimization (13) as a single conic optimization problem with an order of $K(N+J)$ decision variables and an order of KJ conic constraints, both linearly growing in the number J of expectation constraints. The maximization reformulation $\rho_2(\boldsymbol{x})$ allows us to represent Problem (13), which is of the minimax type (Sion 1958), as yet another minimax problem: $\min_{\boldsymbol{x}\in\mathcal{X}} \rho_2(\boldsymbol{x})$. By exploring that representation, we can determine a worst-case distribution in the relaxed ambiguity set that corresponds to the optimal solution $\boldsymbol{x}^*$ for Problem (13), which plays a crucial role in our proposed algorithm for tightening the relaxation.

Theorem 4. *Given the relaxed ambiguity set $\mathcal{F}_{\mathrm{R}}$ and the optimal solution $\boldsymbol{x}^*$ for Problem* (13), *let $(\boldsymbol{\xi}_k^*, \eta_k^*)_{k\in[K]}$ be an optimal solution to the following tractable conic optimization problem*:

$$\begin{aligned}
\sup \ & \sum_{k\in[K]} (\boldsymbol{a}_k(\boldsymbol{x}^*)'\boldsymbol{\xi}_k + b_k(\boldsymbol{x}^*)\eta_k) \\
\text{s.t.} \ & \sum_{k\in[K]} \eta_k = 1 \\
& \sum_{k\in[K]} \boldsymbol{G}\boldsymbol{\xi}_k = \boldsymbol{\mu} \\
& \sum_{k\in[K]} [\boldsymbol{\zeta}_k]_{ij} \leq h_i(\boldsymbol{q}_{ij}), && \forall i \in [I], \forall j \in [J_i] \\
& (\boldsymbol{\xi}_k, \boldsymbol{\zeta}_k, \eta_k) \in \mathcal{K}, && \forall k \in [K] \\
& \boldsymbol{\xi}_k \in \mathbb{R}^N, \boldsymbol{\zeta}_k \in \mathbb{R}^J, \eta_k \in \mathbb{R}_+, && \forall k \in [K];
\end{aligned} \tag{18}$$

here if the feasible set $\mathcal{X}$ is convex and compact, we add further a robust constraint

$$\sum_{k\in[K]} (\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k) \geq \sum_{k\in[K]} (\boldsymbol{a}_k(\boldsymbol{x}^*)'\boldsymbol{\xi}_k + b_k(\boldsymbol{x}^*)\eta_k), \ \forall \boldsymbol{x}\in\mathcal{X}. \tag{19}$$

Then the worst-case distribution in $\mathcal{F}_{\mathrm{R}}$ for the worst-case expected objective $\rho_{\mathrm{R}}(\boldsymbol{x}^*)$ is given by

$$\mathbb{P}_v\left[\tilde{\boldsymbol{z}} = \frac{\boldsymbol{\xi}_k^*}{\eta_k^*}\right] = \eta_k^*, \ \forall k\in[K]\colon \eta_k^* > 0.$$

Proof. For the constructive discrete distribution we note that if $\eta_{k'}^* = 0$ for some $k'\in[K]$, then $\boldsymbol{\xi}_{k'}^* = \mathbf{0}$; otherwise, $\boldsymbol{\xi}_{k'}^*/\eta_{k'}^* \in \mathcal{W}$ would contradict the compactness of the support set $\mathcal{W}$. We can verify that $\mathbb{P}_v[\tilde{\boldsymbol{z}}\in\mathcal{W}] = 1$, $\mathbb{E}_{\mathbb{P}_v}[\boldsymbol{G}\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}$, and $\mathbb{E}_{\mathbb{P}_v}[g_i(\boldsymbol{q}_{ij},\tilde{\boldsymbol{z}})] \leq h_i(\boldsymbol{q}_{ij})$ for all $i\in[I]$ and for all $j\in[J_i]$; hence $\mathbb{P}_v\in\mathcal{F}_{\mathrm{R}}$. In addition, we observe that

$$\mathbb{E}_{\mathbb{P}_v}[f(\boldsymbol{x}^*,\tilde{\boldsymbol{z}})] = \sum_{k\in[K]} \eta_k^* \max_{n\in[K]}\left\{\boldsymbol{a}_n(\boldsymbol{x}^*)'\frac{\boldsymbol{\xi}_k^*}{\eta_k^*} + b_n(\boldsymbol{x})\right\} \geq \sum_{k\in[K]} (\boldsymbol{a}_k(\boldsymbol{x}^*)'\boldsymbol{\xi}_k^* + b_k(\boldsymbol{x}^*)'\eta_k^*) = \rho_{\mathrm{R}}(\boldsymbol{x}^*),$$

where the last equality follows from $\rho_2(\boldsymbol{x}^*) = \rho_{\mathrm{R}}(\boldsymbol{x}^*)$. Therefore, we can conclude that $\mathbb{P}_v$ is the worst-case distribution.

We next discuss the case when the feasible set $\mathcal{X}$ is convex and compact. For ease of notation, we write $(\boldsymbol{\xi},\boldsymbol{\eta}) = (\boldsymbol{\xi}_k,\eta_k)_{k\in[K]}$ and $\bar{f}(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\eta}) = \sum_{k\in[K]}(\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k)$. Let $\mathcal{C}$ be the convex and compact set of the feasible solution $(\boldsymbol{\xi},\boldsymbol{\eta})$ in Problem (18). By Theorem 3, the relaxed distributionally robust optimization problem (13) is equivalent to the minimax problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{C}} \bar{f}(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\eta}), \tag{20}$$

which, by Assumption 1, satisfies

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{C}} \bar{f}(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\eta}) = \sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{C}} \min_{\boldsymbol{x}\in\mathcal{X}} \bar{f}(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{\eta}).$$

Because $\mathcal{X}$ is convex and compact, by the minimax theorem, $\boldsymbol{x}^*$ is the optimal solution to Problem (20) if and only if there exists an $(\boldsymbol{\xi}^*,\boldsymbol{\eta}^*)\in\mathcal{C}$ such that $(\boldsymbol{x}^*,\boldsymbol{\xi}^*,\boldsymbol{\eta}^*)$ is a saddle point—that is, if and only if

$$\sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{C}} \bar{f}(\boldsymbol{x}^*,\boldsymbol{\xi},\boldsymbol{\eta}) \leq \bar{f}(\boldsymbol{x}^*,\boldsymbol{\xi}^*,\boldsymbol{\eta}^*) \leq \min_{\boldsymbol{x}\in\mathcal{X}} \bar{f}(\boldsymbol{x},\boldsymbol{\xi}^*,\boldsymbol{\eta}^*).$$

This saddle-point property ensures that given $\boldsymbol{x}^*$, we can add the constraint (19) into Problem (18) without influencing its optimal objective value. □

Theorem 4 asserts that $\mathbb{P}_v$ is the worst-case distribution in $\mathcal{F}_{\mathrm{R}}$. However, $\mathbb{P}_v$ might not belong to the infinitely constrained ambiguity set $\mathcal{F}_{\mathrm{I}}$ because it may violate those classes of expectation constraints for some $\boldsymbol{q}_i^*\in\mathcal{Q}_i$, $i\in[I]$. In that event, the worst-case distribution $\mathbb{P}_v$ is essentially a violating distribution in $\mathcal{F}_{\mathrm{I}}$. This observation motivates us to solve the separation problem

$$Z_i(\mathbb{P}_v) = \min_{\boldsymbol{q}\in\mathcal{Q}_i}\left\{\psi_i(h_i(\boldsymbol{q})) - \psi_i(\mathbb{E}_{\mathbb{P}_v}[g_i(\boldsymbol{q},\tilde{\boldsymbol{z}})])\right\} \tag{21}$$

to check whether each of those classes of expectation constraints holds. Note that $\psi_i : \mathbb{R} \mapsto \mathbb{R}$ is some increasing function, which could (in principle) be linear—although a nonlinear ψ_i may increase precision when we evaluate the separation problem's objective. When we consider an entropic dominance ambiguity set, for instance, $h_i(\boldsymbol{q})$ would be an exponential function; in that case, we would not (owing to the 64 bit floating-point precision) be able to evaluate the value of $\exp(x)$ for $x \geq 710$. For such circumstances, we can choose ψ_i to be a logarithmic function to improve precision when solving the separation problem.

We now offer a formal statement of our algorithm, which utilizes the separation problem (21) to tighten the relaxation.

Algorithm Greedy Improvement Procedure (GIP).

Input: Initial finite subsets $\bar{\mathcal{Q}}_i \subseteq \mathcal{Q}_i$ for all $i \in [I]$.

1. Solve Problem (13) and obtain an optimal solution $\boldsymbol{x}$.
2. Solve Problem (18) and obtain the worst-case distribution $\mathbb{P}_v$.
3. For all $i \in [I]$, solve Problem (21): if $Z_i(\mathbb{P}_v) < 0$, obtain the optimal solution $\boldsymbol{q}_i$ and update $\bar{\mathcal{Q}}_i = \bar{\mathcal{Q}}_i \cup \{\boldsymbol{q}_i\}$.
4. If $Z_i(\mathbb{P}_v) \geq 0$ for all $i \in [I]$, then STOP. Otherwise return to Step 1.

Output: Solution $\boldsymbol{x}$.

Theorem 5. *In Algorithm GIP, the sequence of objectives of Problem* (13) *is nonincreasing. If in addition, the feasible set $\mathcal{X}$ is convex and compact, then the sequence of solutions in Algorithm GIP converges to the optimal solution of the distributionally robust optimization problem* (11).

Proof. The proof of the theorem's first part is straightforward and so is omitted. We next proof the theorem's second part.

Let $\mathcal{F}_{\mathrm{R}}^t$ be the ambiguity set at the t-th iteration of Algorithm GIP and let $\mathbb{P}_v^t$ and $\boldsymbol{x}^t$ denote the optimal solutions to (respectively) Problem (18) and Problem (13). Observe that $\{\mathcal{F}_{\mathrm{R}}^t\}_{t\geq 1}$ is a nonincreasing sequence of sets, all of which contain $\mathcal{F}_{\mathrm{I}}$; hence there exists a $\mathcal{F}_{\mathrm{R}}^*$ such that $\lim_{t\to\infty} \mathcal{F}_{\mathrm{R}}^t = \mathcal{F}_{\mathrm{R}}^*$ and $\mathcal{F}_{\mathrm{I}} \subseteq \mathcal{F}_{\mathrm{R}}^*$. The steps in Algorithm GIP indicate that $(\boldsymbol{x}^t, \mathbb{P}_v^t)$ is a saddle point of the minimax type problem $\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}^t} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$, from which it follows that

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}^t} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}^t, \tilde{\boldsymbol{z}})\right] \leq \mathbb{E}_{\mathbb{P}_v^t}\left[f(\boldsymbol{x}^t, \tilde{\boldsymbol{z}})\right] \leq \min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}_v^t}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right], \tag{22}$$

Because $\mathcal{X}$ and $\mathcal{F}$ are compact, the sequence $\{(\boldsymbol{x}^t, \mathbb{P}_v^t)\}_{t\geq 1}$ has a subsequence that converges to a limit point $(\boldsymbol{x}^*, \mathbb{P}_v^*)$. Hence, inequality (22) becomes

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}^*} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}^*, \tilde{\boldsymbol{z}})\right] \leq \mathbb{E}_{\mathbb{P}_v^*}\left[f(\boldsymbol{x}^*, \tilde{\boldsymbol{z}})\right] \leq \min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}_v^*}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right],$$

which implies that $(\boldsymbol{x}^*, \mathbb{P}_v^*)$ is a saddle point for the problem $\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}^*} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$. Since $\mathcal{F}_{\mathrm{I}} \subseteq \mathcal{F}_{\mathrm{R}}^*$, we obtain

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{I}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}^*, \tilde{\boldsymbol{z}})\right] \leq \mathbb{E}_{\mathbb{P}_v^*}\left[f(\boldsymbol{x}^*, \tilde{\boldsymbol{z}})\right] \leq \min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}_v^*}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right].$$

Moreover, it follows that $\mathbb{P}_v^* \in \mathcal{F}_{\mathrm{I}}$; otherwise there must be some i such that the corresponding separation problem yields $Z_i(\mathbb{P}_v^*) < 0$ for some optimal solution $\boldsymbol{q}^*$.

Let $\boldsymbol{q}^t$ be the optimal solution for the separation problem at Algorithm GIP's t-th iteration:

$$\boldsymbol{q}^t \in \arg\min_{\boldsymbol{q}\in\mathcal{Q}_i} \left\{ \psi_i\left(h_i(\boldsymbol{q})\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^t}\left[g_i(\boldsymbol{q}, \tilde{\boldsymbol{z}})\right]\right) \right\}.$$

Then the algorithm's procedure ensures that

$$\psi_i\left(h_i(\boldsymbol{q}^t)\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^*}\left[g_i(\boldsymbol{q}^t, \tilde{\boldsymbol{z}})\right]\right) \geq 0, \ \forall t \geq 1. \tag{23}$$

By the definition of $\boldsymbol{q}^t$, for every t we have:

$$\psi_i\left(h_i(\boldsymbol{q}^t)\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^t}\left[g_i(\boldsymbol{q}^t, \tilde{\boldsymbol{z}})\right]\right) \leq \psi_i\left(h_i(\boldsymbol{q}^*)\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^t}\left[g_i(\boldsymbol{q}^*, \tilde{\boldsymbol{z}})\right]\right).$$

Let $t \to \infty$, then $\psi_i\left(h_i(\bar{\boldsymbol{q}})\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^*}\left[g_i(\bar{\boldsymbol{q}}, \tilde{\boldsymbol{z}})\right]\right) \leq \psi_i\left(h_i(\boldsymbol{q}^*)\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^*}\left[g_i(\boldsymbol{q}^*, \tilde{\boldsymbol{z}})\right]\right) = Z_i(\mathbb{P}_v^*)$ for some accumulation point $\bar{\boldsymbol{q}}$ of the sequence $\{\boldsymbol{q}^t\}_{t\geq 1}$. However, from inequality (23), it follows that $\psi_i\left(h_i(\bar{\boldsymbol{q}})\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^*}\left[g_i(\bar{\boldsymbol{q}}, \tilde{\boldsymbol{z}})\right]\right) \geq 0$, which contradicts $Z_i(\mathbb{P}_v^*) < 0$. Therefore, $(\boldsymbol{x}^*, \mathbb{P}_v^*)$ is a saddle point for the problem $\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{I}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$ and we can conclude that $\boldsymbol{x}^*$ is an optimal solution to Problem (11). □

The key to establishing the convergence result is that we are able to derive a saddle point for the relaxed problem in each iteration whenever the feasible set $\mathcal{X}$ is convex and compact. If the solver adopts the primal-dual approach to solve Problem (13) and provides the corresponding optimal dual solutions, then we can construct the worst-case distribution using these associated dual variables without solving Problem (18). Since the separation problem (21) is generally not convex, it could be computationally challenging to obtain its optimal solution. For the case of the covariance ambiguity set we can solve the separation problem to optimality in polynomial time because it is a minimal eigenvalue problem. For instances of the fourth moment ambiguity set and the entropic dominance ambiguity set, we can efficiently obtain solutions that attain negative objective values using the trust region method (Conn et al. 2000) and then proceed to the next iteration by using these solutions to tighten the current relaxation. In this way, we can still obtain iteratively a monotonic improvement.

4. Covariance and fourth moment

In this section, we first consider the distributionally robust optimization model

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \tag{24}$$

with $\mathcal{F}_{\mathrm{C}}$ being the covariance ambiguity set (7), which—as the next proposition shows—can be reformulated into a semidefinite program.

PROPOSITION 3. *Problem* (24) *is equivalent to*

$$\begin{array}{lll}\inf & \alpha+\boldsymbol{\beta}'\boldsymbol{\mu}+\langle\boldsymbol{\Gamma},\boldsymbol{\Sigma}\rangle & \\ \text{s.t.} & \alpha-\boldsymbol{b}_k(\boldsymbol{x})+2\boldsymbol{\chi}_k'\boldsymbol{\mu}-\delta_k-t_k\geq 0, & \forall k\in[K]\\ & \boldsymbol{\beta}-\boldsymbol{a}_k(\boldsymbol{x})-\boldsymbol{r}_k-2\boldsymbol{\chi}_k=\boldsymbol{0}, & \forall k\in[K]\\ & \boldsymbol{\Gamma}=\boldsymbol{\Gamma}_k, & \forall k\in[K]\\ & (\boldsymbol{r}_k,t_k)\in\mathcal{K}^*(\mathcal{W}), & \forall k\in[K]\\ & \begin{pmatrix}\delta_k & \boldsymbol{\chi}_k'\\ \boldsymbol{\chi}_k & \boldsymbol{\Gamma}_k\end{pmatrix}\succeq\boldsymbol{0}, & \forall k\in[K]\\ & \alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^N,\boldsymbol{x}\in\mathcal{X}. & \end{array} \tag{25}$$

In this problem, $\mathcal{K}^*(\mathcal{W})$ *is the dual cone of* $\mathcal{K}(\mathcal{W})=\mathrm{cl}\{(\boldsymbol{z},t)\in\mathbb{R}^N\times\mathbb{R}\mid \boldsymbol{z}/t\in\mathcal{W},t>0\}$ *and* $\langle\cdot,\cdot\rangle$ *denotes the trace inner product of two matrices.*

As mentioned previously, a covariance ambiguity set can be represented as in (6), in the form of an infinitely constrained ambiguity set. With that representation, we now consider the corresponding relaxed ambiguity set

$$\mathcal{G}_{\mathrm{C}}=\left\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^N\right)\;\middle|\;\begin{array}{l}\tilde{\boldsymbol{z}}\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2\right]\leq\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q},\;\forall\boldsymbol{q}\in\bar{\mathcal{Q}}\\ \mathbb{P}\left[\tilde{\boldsymbol{z}}\in\mathcal{W}\right]=1\end{array}\right\}, \tag{26}$$

where $\bar{\mathcal{Q}}=\{\boldsymbol{q}_j\mid j\in[J]\}$ for some $\boldsymbol{q}_j\in\mathbb{R}^N$ with $\|\boldsymbol{q}_j\|_2\leq 1$. The relaxed distributionally robust optimization is then given by

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{G}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right], \tag{27}$$

which—in contrast to Problem (24)—can be reformulated as a second-order conic program. In addition, the optimization problem for determining the worst-case distribution required in solving the separation problem is also a second-order conic program.

Proposition 4. *Problem (27) is equivalent to*

$$\begin{aligned}
\inf\ & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j \boldsymbol{q}_j'\boldsymbol{\Sigma}\boldsymbol{q}_j \\
\text{s.t.}\ & \alpha - b_k(\boldsymbol{x}) + \sum_{j\in[J]} (m_{jk} - l_{jk} + 2n_{jk}\boldsymbol{q}_j'\boldsymbol{\mu}) - t_k \geq 0, && \forall k\in[K] \\
& \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - \sum_{j\in[J]} 2n_{jk}\boldsymbol{q}_j = \boldsymbol{0}, && \forall k\in[K] \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*(\mathcal{W}), && \forall k\in[K] \\
& \gamma_j = l_{jk} + m_{jk}, && \forall j\in[J], \forall k\in[K] \\
& (l_{jk}, m_{jk}, n_{jk}) \in \mathcal{K}_{\text{SOC}}, && \forall j\in[J], \forall k\in[K] \\
& \boldsymbol{l}_k, \boldsymbol{m}_k, \boldsymbol{n}_k \in \mathbb{R}^J, && \forall k\in[K] \\
& \alpha\in\mathbb{R}, \boldsymbol{\beta}\in\mathbb{R}^N, \boldsymbol{\gamma}\in\mathbb{R}^J_+,\ \boldsymbol{x}\in\mathcal{X}.
\end{aligned} \tag{28}$$

Given the relaxed ambiguity set $\mathcal{F}_{\text{R}}$ *and the optimal solution* $\boldsymbol{x}^*$ *for Problem* (27), *let* $(\boldsymbol{\xi}_k^*, \eta_k^*)_{k\in[K]}$ *be an optimal solution to the following tractable conic optimization problem*:

$$\begin{aligned}
\sup\ & \sum_{k\in[K]} (\boldsymbol{a}_k(\boldsymbol{x}^*)'\boldsymbol{\xi}_k + b_k(\boldsymbol{x}^*)\eta_k) \\
\text{s.t.}\ & \sum_{k\in[K]} \eta_k = 1 \\
& \sum_{k\in[K]} \boldsymbol{\xi}_k = \boldsymbol{\mu} \\
& \sum_{k\in[K]} \zeta_{kj} \leq \boldsymbol{q}_j'\boldsymbol{\Sigma}\boldsymbol{q}_j, && \forall j\in[J] \\
& (\boldsymbol{\xi}_k, \eta_k) \in \mathcal{K}(\mathcal{W}), && \forall k\in[K] \\
& \big(\zeta_{kj} + \eta_k, \zeta_{kj} - \eta_k, 2\boldsymbol{q}_j'(\boldsymbol{\xi}_k - \eta_k\boldsymbol{\mu})\big) \in \mathcal{K}_{\text{SOC}}, && \forall j\in[J], \forall k\in[K] \\
& \boldsymbol{\xi}_k\in\mathbb{R}^N, \boldsymbol{\zeta}_k\in\mathbb{R}^J, \eta_k\in\mathbb{R}_+, && \forall k\in[K];
\end{aligned}$$

here if the feasible set $\mathcal{X}$ *is convex and compact, we add further a robust constraint*

$$\sum_{k\in[K]} (\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k) \geq \sum_{k\in[K]} (\boldsymbol{a}_k(\boldsymbol{x}^*)'\boldsymbol{\xi}_k + b_k(\boldsymbol{x}^*)\eta_k),\ \forall \boldsymbol{x}\in\mathcal{X}.$$

Then the worst-case distribution is given by

$$\mathbb{P}_v\left[\tilde{\boldsymbol{z}} = \frac{\boldsymbol{\xi}_k^*}{\eta_k^*}\right] = \eta_k^*,\ \forall k\in[K]\colon \eta_k^* > 0.$$

Given the worst-case distribution, $\mathbb{P}_v$, the corresponding separation problem would be

$$\min_{\boldsymbol{q}\colon \|\boldsymbol{q}\|_2\leq 1} \boldsymbol{q}'\big(\boldsymbol{\Sigma} - \mathbb{E}_{\mathbb{P}_v}[(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})']\big)\boldsymbol{q};$$

this is a classical minimal eigenvalue problem that can be solved efficiently using numerical techniques such as the power iteration. Whenever the minimal eigenvalue is negative, we can add the corresponding eigenvector to the set $\bar{\mathcal{Q}}$, which would end up tightening the relaxed covariance ambiguity set $\mathcal{G}_{\text{C}}$. It is of considerable interest that there exists an ambiguity set $\mathcal{G}_{\text{C}}$ with $J = N$ such that Problems (24) and (27) have the same objective value, as shown in the following result.

Theorem 6. *Consider any function $f : \mathbb{R}^N \mapsto \mathbb{R}$, for which the problem*

$$\sup_{\mathbb{P}\in\mathcal{F}_{\text{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right]$$

is finitely optimal. There exists a relaxed ambiguity set $\mathcal{G}^*_{\text{C}}$ such that $J = N$ and

$$\sup_{\mathbb{P}\in\mathcal{F}_{\text{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right] = \sup_{\mathbb{P}\in\mathcal{G}^*_{\text{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right].$$

Proof. Consider the problem $Z_{\text{A}} = \sup_{\mathbb{P}\in\mathcal{F}_{\text{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right]$. By duality, we have

$$\begin{aligned} Z_{\text{A}} = \inf\ & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma}\rangle \\ \text{s.t.}\ & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \left\langle \boldsymbol{\Gamma}, (\boldsymbol{z}-\boldsymbol{\mu})(\boldsymbol{z}-\boldsymbol{\mu})'\right\rangle \geq f(\boldsymbol{z}),\ \forall \boldsymbol{z}\in\mathcal{W} \\ & \alpha\in\mathbb{R}, \boldsymbol{\beta}\in\mathbb{R}^N, \boldsymbol{\Gamma}\succeq \boldsymbol{0}, \end{aligned} \tag{29}$$

whose optimal solution we denote by $(\alpha^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^*)$. Using the eigendecomposition, we write $\boldsymbol{\Gamma}^* = \sum_{n\in[N]} \lambda^*_n \boldsymbol{q}^*_n (\boldsymbol{q}^*_n)'$, where the $\{\lambda^*_n\}_{n\in[N]}$ are eigenvalues of $\boldsymbol{\Gamma}^*$ and where $\{\boldsymbol{q}^*_n\}_{n\in[N]}$ are the corresponding eigenvectors. Note that $\boldsymbol{\Gamma}^* \succeq \boldsymbol{0}$ implies that $\lambda^*_n \geq 0$ for all $n\in[N]$.

Let $\mathcal{G}^*_{\text{C}}$ be the particular relaxed ambiguity set with $\bar{\mathcal{Q}} = \{\boldsymbol{q}^*_1, \ldots, \boldsymbol{q}^*_N\}$ and consider the problem $Z_{\text{B}} = \sup_{\mathbb{P}\in\mathcal{G}^*_{\text{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right]$, whose dual is given as follows:

$$\begin{aligned} Z_{\text{B}} = \inf\ & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{n\in[N]} \gamma_n (\boldsymbol{q}^*_n)'\boldsymbol{\Sigma}\boldsymbol{q}^*_n \\ \text{s.t.}\ & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{n\in[N]} \gamma_n((\boldsymbol{q}^*_n)'(\boldsymbol{z}-\boldsymbol{\mu}))^2 \geq f(\boldsymbol{z}),\ \forall \boldsymbol{z}\in\mathcal{W} \\ & \alpha\in\mathbb{R}, \boldsymbol{\beta}\in\mathbb{R}^N, \boldsymbol{\gamma}\in\mathbb{R}^N_+. \end{aligned} \tag{30}$$

We can exploit the optimal solution to Problem (29) to construct a feasible solution $(\alpha^{\text{B}}, \boldsymbol{\beta}^{\text{B}}, \boldsymbol{\gamma}^{\text{B}})$ to Problem (30) by letting $\alpha_{\text{B}} = \alpha^*$, $\boldsymbol{\beta}_{\text{B}} = \boldsymbol{\beta}^*$, and $\boldsymbol{\gamma}_{\text{B}} = \boldsymbol{\lambda}^*$. Since $\langle \boldsymbol{\Gamma}^*, \boldsymbol{\Sigma}\rangle = \sum_{n\in[N]} \langle \lambda^*_n \boldsymbol{q}^*_n (\boldsymbol{q}^*_n)', \boldsymbol{\Sigma}\rangle = \sum_{n\in[N]} \gamma_n (\boldsymbol{q}^*_n)'\boldsymbol{\Sigma}\boldsymbol{q}^*_n$, it follows that $Z_{\text{B}} \leq Z_{\text{A}}$. Yet we already know that $Z_{\text{A}} \leq Z_{\text{B}}$ for any relaxed covariance ambiguity set, so our claim holds. □

Having a relaxation of Problem (24) in the form of a second-order conic program is especially useful when the feasible set $\mathcal{X}$ has integral constraints because such a format is already supported by commercial solvers such as CPLEX and Gurobi. In such cases, the reformulation of Problem (24) would become a semidefinite program involving discrete decision variables, which are harder to solve. Instead, we can first consider the linear relaxation of $\mathcal{X}$ (denoted by $\bar{\mathcal{X}}$), solve the problem

$$\min_{\boldsymbol{x}\in\bar{\mathcal{X}}} \sup_{\mathbb{P}\in\mathcal{F}_{\text{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right],$$

and obtain $\mathcal{G}_{\mathrm{C}}^*$ through Theorem 6. Afterward, we implement Algorithm GIP with $\mathcal{G}_{\mathrm{C}}^*$ being the initial relaxed ambiguity set and solve a sequence of mixed-integer second-order conic programs to obtain approximate solutions to the problem of interest.

Fourth moment ambiguity set

The fourth moment ambiguity set (8) has the following relaxation

$$\mathcal{G}_{\mathrm{F}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{0} \\ \mathbb{E}_{\mathbb{P}}\left[(\hat{\boldsymbol{q}}'\tilde{\boldsymbol{z}})^2\right] \le \hat{\boldsymbol{q}}'\boldsymbol{\Sigma}\hat{\boldsymbol{q}},\ \forall \hat{\boldsymbol{q}} \in \hat{\mathcal{Q}} \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^4\right] \le \varphi(\boldsymbol{q}),\ \ \forall \boldsymbol{q} \in \bar{\mathcal{Q}} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}, \tag{31}$$

where $\hat{\mathcal{Q}} = \{\hat{\boldsymbol{q}}_j \mid j \in [\hat{J}]\}$ and $\bar{\mathcal{Q}} = \{\boldsymbol{q}_j \mid j \in [J]\}$ for some $\hat{\boldsymbol{q}}_j, \boldsymbol{q}_j \in \mathbb{R}^N$ and $\|\hat{\boldsymbol{q}}_j\|_2, \|\boldsymbol{q}_j\|_2 \le 1$. Note that the function $(\boldsymbol{q}'\boldsymbol{z})^4$ is also second-order conic representable in $\boldsymbol{z}$. Following from a similar derivation as in Proposition 4, the relaxed distributionally robust optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{G}_{\mathrm{F}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$$

can be reformulated as a second-order conic program and so does the optimization problem for determining the worst-case distribution. Given that the worst-case distribution $\mathbb{P}_v$ takes value $\boldsymbol{z}_k$ with probability p_k, the separation problem that corresponds to the second set of infinitely many expectation constraints in $\mathcal{F}_{\mathrm{F}}$ is unconstrained and takes the following form:

$$\min_{\boldsymbol{q}} \Psi(\boldsymbol{q}) = \min_{\boldsymbol{q}} \left\{ \sum_{n\in[N]} \left(q_n^4 \kappa_n^4 + 6 \sum_{m\neq n} q_m^2 q_n^2 \sigma_m^2 \sigma_n^2 \right) - \sum_{k\in[K]} p_k (\boldsymbol{q}'\boldsymbol{z}_k)^4 \right\},$$

where the objective function is a multivariate polynomial of $\{q\}_{n\in[N]}$. In general, this separation problem is NP-hard (Nesterov 2000) but admits a semidefinite relaxation if we use the sum of squares tool (cf. Choi et al. 1995, Lasserre 2001, Parrilo 2003). This kind of semidefinite relaxations may be successively improved and has been applied in many areas including option pricing (Bertsimas and Popescu 2002, Lasserre et al. 2006). An alternative approach is to adopt the trust region method (TRM), which is useful for finding local minimums of this separation problem. The TRM is one of the most important numerical optimization methods in solving nonlinear programming problems. It works by solving a sequence of subproblems; in each subproblem, the TRM approximates the objective function $\Psi(\cdot)$ by a quadratic function (usually obtained by taking Taylor series up to second order) and searches for an improving direction in a region around the current solution. Note that solving an unconstrained nonlinear optimization problem via the TRM is well supported in MATLAB (Mathworks 2017); we need only provide the gradient and the Hessian matrix of $\Psi(\cdot)$.

An experimental study

We now conduct a numerical experiment, using randomly generated instances, for three purposes:

(i) To investigate the convergence speed of Algorithm GIP;

(ii) To study the effect of that algorithm's initialization;

(iii) To elaborate the advantage of the fourth moment ambiguity set.

Thus, we study the distributionally robust optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right].$$

The setting of this experiment is described as follows. The feasible set $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^M_+ \mid \boldsymbol{e}'\boldsymbol{x} = 1\}$, and the cost function is given by $f(\boldsymbol{x},\boldsymbol{z}) = \max_{k\in[K]}\left\{(\boldsymbol{S}_k'\boldsymbol{x} + \boldsymbol{t}_k)'\boldsymbol{z} + \boldsymbol{s}_k'\boldsymbol{x} + t_k\right\}$. We consider the class of independently distributed random variables $\{\tilde{z}_n\}_{n\in[N]}$ with identical zero mean and support $[-\underline{z},\bar{z}]$. For every $n\in[N]$, the second and fourth moments of $\tilde{z}_n$ are bounded by (respectively) σ_n^2 and κ_n^4. For different sizes of the problem (as determined by K, M, and N), we fix $\bar{z} = \underline{z} = 50$ and investigate 100 random instances by generating components in $(\boldsymbol{S}_k, \boldsymbol{s}_k, \boldsymbol{t}_k, t_k)_{k\in[K]}$ from the standard normal distribution, thereby generating σ_n from the uniform distribution over $[0,50]$ and setting $\kappa_n = \varepsilon_n\sigma_n$; ε_n is randomly drawn from the uniform distribution over $[1,2]$. For the family of joint probability distributions of these random variables, we consider the covariance ambiguity set (7) (i.e., $\mathcal{F} = \mathcal{F}_{\rm C}$) and the fourth moment ambiguity set (8) (i.e., $\mathcal{F} = \mathcal{F}_{\rm F}$).

In view of the purposes (i) and (ii), we consider $\mathcal{F} = \mathcal{F}_{\rm C}$ and solve the distributionally robust optimization problem exactly by using Proposition 3. We also implement Algorithm GIP by using, in (26), four different relaxed covariance ambiguity set as follows.

- Marginal moment ambiguity set $\mathcal{G}_{\rm M}$: $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n \mid n \in [N]\}$.
- Relaxed ambiguity set $\mathcal{G}_{\rm A}$: $|\bar{\mathcal{Q}}| = N$ and the components of each element in $\bar{\mathcal{Q}}$ are independently and uniformly generated from $[-1,1]$.
- Partial cross-moment ambiguity set $\mathcal{G}_{\rm P}$: $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n \mid n\in[N]\} \cup \{\boldsymbol{e} - \boldsymbol{e}_n \mid n \in [N]\}$.
- Relaxed ambiguity set $\mathcal{G}_{\rm B}$: $|\bar{\mathcal{Q}}| = 2N$ and the components of each element in $\bar{\mathcal{Q}}$ are independently and uniformly generated from $[-1,1]$.

We terminate Algorithm GIP once the approximate second-order conic solution is within 0.5% of the optimal semidefinite solution (or after 100 iterations). Results for different initializations among randomly generated instances are summarized in Table 1. For $(K,M,N) = (6,8,10)$ or $(K,M,N) = (12,10,8)$, Algorithm GIP terminates when it finds (as it does in most instances) a second-order conic solution with high approximation accuracy. For $(K,M,N) = (10,12,15)$ or $(K,M,N) = (15,10,20)$, the algorithm usually terminates owing to the cap on iterations; however, it is still able to provide a good approximate solution for both the general instance and the worst instance. These results suggest that Algorithm GIP converges reasonably well. Table 1 reveals no clear effect on the obtained results under different initializations for Algorithm GIP, and we observe a similar outcome in our other numerical studies. That being said, we believe that—at least for deriving the relaxed covariance ambiguity set—the marginal moment ambiguity set and the partial cross-moment ambiguity set are good candidates for the initial relaxed ambiguity set because they have meaningful physical interpretation; see an application in Bertsimas et al. (2018).

(K,M,N)	$(6,8,10)$	$(10,12,15)$	$(12,10,8)$	$(15,10,20)$
$\mathcal{G}_{\mathrm{M}}$	0.5% (0.4%, 0.5%, 0.9%)	0.9% (0.5%, 0.9%, 1.7%)	0.5% (0.3%, 0.5%, 0.5%)	1.3% (0.8%, 1.4%, 2.2%)
$\mathcal{G}_{\mathrm{A}}$	0.5% (0.4%, 0.5%, 0.9%)	0.9% (0.5%, 0.9%, 1.9%)	0.5% (0.3%, 0.5%, 0.5%)	1.4% (0.9%, 1.5%, 2.4%)
$\mathcal{G}_{\mathrm{P}}$	0.5% (0.4%, 0.5%, 0.8%)	0.9% (0.5%, 0.9%, 1.7%)	0.5% (0.4%, 0.5%, 0.5%)	1.3% (0.7%, 1.3%, 2.2%)
$\mathcal{G}_{\mathrm{B}}$	0.5% (0.4%, 0.5%, 0.8%)	0.8% (0.5%, 0.9%, 1.7%)	0.5% (0.3%, 0.5%, 0.5%)	1.3% (0.8%, 1.4%, 2.3%)

Table 1 Median (minimal, average, maximal) relative gap against the exact semidefinite solution for different initializations.

Although Table 1 showcases the convergence of our proposed algorithm, we do not mean to suggest that solving a sequence of second-order conic programs is necessarily preferable to solving a semidefinite program, which can be efficiently solved (if the problem is not too large and does not involve discrete decision variables) using Mosek, SDPT3, or SeDuMi. In Table 2, we compare the solution time for solving the exact semidefinite program (SDP_time) and the accumulative solution time of all iterations when Algorithm GIP is implemented (SOCP_time). Algorithm GIP takes much longer to obtain approximate solutions when the exact semidefinite program can be efficiently solved (in fewer than 12 second for all our instances). On the other hand, solving a sequence of second-order conic programs can be helpful when the exact semidefinite program is hard to solve: for instances, (a) when the problem size is large and (b) when the feasible set has integral constraints. We conduct two numerical experiments for these two circumstances in Appendix A and Appendix B and show the benefits from implementing Algorithm GIP.

With regard to the purpose (iii), we consider $\mathcal{F} = \mathcal{F}_{\mathrm{F}}$ and implement Algorithm GIP using an initialization (31) such that $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n \mid n \in [N]\}$ and $\hat{\mathcal{Q}} = \{\boldsymbol{\lambda}_n^* \mid n \in [N]\}$; here we utilize Theorem 6 and choose $\{\boldsymbol{\lambda}_n^*\}_{n\in[N]}$ to be the eigenvectors of the optimal $\boldsymbol{\Gamma}^*$ when $\mathcal{F} = \mathcal{F}_{\mathrm{C}}$. In each iteration, we leverage the trust region method for local minimums of the separation problem. We use a randomly generated start and stop at a local minimum, and if that minimum attains a negative

(K,M,N)	$(6,8,10)$	$(10,12,15)$	$(12,10,8)$	$(15,10,20)$
$\mathcal{G}_{\mathrm{M}}$	4.5	4.5	0.7	4.5
$\mathcal{G}_{\mathrm{A}}$	5.0	4.6	1.1	4.9
$\mathcal{G}_{\mathrm{P}}$	5.5	6.1	0.9	7.0
$\mathcal{G}_{\mathrm{B}}$	5.4	5.8	1.2	6.7

Table 2 Average SOCP_time/SDP_time **for different initializations.**

objective value for the separation problem then we find a violating expectation constraint and proceed to the next iteration; otherwise, we continue with another random start. We terminate Algorithm GIP if no violating expectation constraint is found after 100 trials for a given iteration (or after 50 iterations). The relative improvement from considering the fourth moment ambiguity set is reported in Table 3, which shows the clear improvement for all four problem sizes. We remark that this improvement accounts only for the best approximation we have sought—because finding the separation problem's optimal solution is hard and we have to terminate Algorithm GIP prematurely.

(K,M,N)	$(6,8,10)$	$(10,12,15)$	$(12,10,8)$	$(15,10,20)$
Relative	1.8%	3.9%	4.5%	5.1%
improvement	(0.3%, 1.9%, 5.1%)	(1.4%, 4.2%, 7.5%)	(1.6%, 4.6%, 9.9%)	(2.0%, 5.2%, 7.2%)

Table 3 Median (minimal, average, maximal) relative improvement over the semidefinite solution.

5. Entropic dominance

In this section, we focus on the entropic dominance ambiguity set $\mathcal{F}_{\mathrm{E}}$ in (9) and show that it provides a better characterization of stochastic independence than does the covariance ambiguity set. In particular, we consider the relaxed distributionally robust optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{G}_{\mathrm{E}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \tag{32}$$

with the relaxed entropic dominance ambiguity set

$$\mathcal{G}_{\mathrm{E}} = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}}\sim\mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right]=\boldsymbol{\mu} \\ \ln\mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})\right)\right]\le\phi(\boldsymbol{q}),\ \forall\boldsymbol{q}\in\bar{\mathcal{Q}} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}}\in\mathcal{W}\right]=1 \end{array} \right\},$$

where $\bar{\mathcal{Q}}=\{\boldsymbol{q}_j \mid j\in[J]\}$ for some $\boldsymbol{q}_j\in\mathbb{R}^N$. As mentioned in Section 2, we explicitly specify the mean and support because they are not implied when there are only a finite number of expectation constraints in $\mathcal{G}_{\mathrm{E}}$. Problem (32) can be reformulated as a conic optimization problem involving the exponential cone, as can the problem for obtaining the worst-case distribution.

PROPOSITION 5. *Problem* (32) *is equivalent to*

$$
\begin{array}{lll}
\inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[j]} \gamma_j & \\
\text{s.t.} & \alpha - b_k(\boldsymbol{x}) + \sum_{j\in[J]} l_{kj}\left(\boldsymbol{q}_j'\boldsymbol{\mu} + \phi(\boldsymbol{q}_j)\right) - \sum_{j\in[J]} m_{kj} - t_k \geq 0, & \forall k \in [K] \\
& \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - \sum_{j\in[J]} l_{kj}\boldsymbol{q}_j = \boldsymbol{0}, & \forall k \in [K] \\
& \boldsymbol{\gamma} - \boldsymbol{n}_k = \boldsymbol{0}, & \forall k \in [K] \\
& (l_{kj}, m_{kj}, n_{kj}) \in \mathcal{K}^*_{\mathrm{EXP}}, & \forall j \in [J], \forall k \in [K] \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*(\mathcal{W}), & \forall k \in [K] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}^J_+, \ \boldsymbol{x} \in \mathcal{X}. &
\end{array}
$$

Given the relaxed ambiguity set $\mathcal{F}_{\mathrm{R}}$ *and the optimal solution* $\boldsymbol{x}^*$ *for Problem* (32), *let* $(\boldsymbol{\xi}_k^*, \eta_k^*)_{k\in[K]}$ *be an optimal solution to the following tractable conic optimization problem*:

$$
\begin{array}{lll}
\sup & \sum_{k\in[K]} \left(\boldsymbol{a}_k(\boldsymbol{x}^*)'\boldsymbol{\xi}_k + b_k(\boldsymbol{x}^*)\eta_k\right) & \\
\text{s.t.} & \sum_{k\in[K]} \eta_k = 1 & \\
& \sum_{k\in[K]} \boldsymbol{\xi}_k = \boldsymbol{\mu} & \\
& \sum_{k\in[K]} \boldsymbol{\zeta}_k \leq \boldsymbol{e} & \\
& (\boldsymbol{\xi}_k, \eta_k) \in \mathcal{K}(\mathcal{W}), & \forall k \in [K] \\
& \left(\boldsymbol{q}_j'(\boldsymbol{\xi}_k - \eta_k\boldsymbol{\mu}) - \eta_k\phi(\boldsymbol{q}_j), \eta_k, \zeta_{kj}\right) \in \mathcal{K}_{\mathrm{EXP}}, & \forall j \in [J], k \in [K] \\
& \boldsymbol{\xi}_k \in \mathbb{R}^N, \boldsymbol{\zeta}_k \in \mathbb{R}^J, \eta_k \in \mathbb{R}_+, & \forall k \in [K];
\end{array}
$$

here if the feasible set $\mathcal{X}$ *is convex and compact, we add further a robust constraint*

$$
\sum_{k\in[K]} \left(\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k\right) \geq \sum_{k\in[K]} \left(\boldsymbol{a}_k(\boldsymbol{x}^*)'\boldsymbol{\xi}_k + b_k(\boldsymbol{x}^*)\eta_k\right), \ \forall \boldsymbol{x} \in \mathcal{X}.
$$

Then the worst-case distribution is given by

$$
\mathbb{P}_v\left[\tilde{\boldsymbol{z}} = \frac{\boldsymbol{\xi}_k^*}{\eta_k^*}\right] = \eta_k^*, \ \forall k \in [K] : \eta_k^* > 0.
$$

Note that an exponential conic constraint can be approximated fairly accurately via a small number of second-order conic constraints (see Chen and Sim 2009, Appendix B). This kind of successive approximation methods is supported in MATLAB toolboxes, such as ROME (Goh and Sim 2011) and CVX (Grant and Boyd 2008, Grant et al. 2008). An exponential conic program can also be efficiently and exactly solved by interior-point methods (cf. Chares 2009, Skajaa and Ye 2015). For the entropic dominance ambiguity set, the corresponding separation problem is unconstrained and takes the following form:

$$\min_{\boldsymbol{q}} \Phi(\boldsymbol{q}) = \min_{\boldsymbol{q}} \{\phi(\boldsymbol{q}) - \ln \mathbb{E}_{\mathbb{P}_v}[\exp(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))]\} = \min_{\boldsymbol{q}} \left\{ \phi(\boldsymbol{q}) - \ln \sum_{k\in[K]} p_k \exp\left(\boldsymbol{q}'(\boldsymbol{z}^k - \boldsymbol{\mu})\right) \right\};$$

here the worst-case distribution $\mathbb{P}_v$ takes value $\boldsymbol{z}_k$ with probability p_k. We can also use the TRM for efficiently detection of the local minimums.

An application in portfolio selection

In this numerical example, we illustrate how our new bound for the expected surplus $\mathbb{E}_{\mathbb{P}}[(\cdot)^+]$ can be leveraged in practice. For that purpose, we study the distributionally robust portfolio optimization problem under the worst-case conditional value at risk (CVaR) of Rockafellar and Uryasev (2002). We consider the case of N assets, each with independently distributed random return premium $\mu_n + \sigma_n \tilde{z}_n$ that are affected by the uncertainty $\tilde{z}_n$ with mean 0 and support $[-1, 1]$. The parameters used in our study are $N = 50$, $\mu_n = n/250$, and $\sigma_n = (N\sqrt{2n})/1000$. Thus, the asset with a higher return premium is more risky. The feasible investment lies in the simplex $\boldsymbol{x} \in \mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^N_+ \mid \boldsymbol{e}'\boldsymbol{x} = 1\}$. The total return premium obtained from an investment $\boldsymbol{x}$ under realization $\boldsymbol{z}$ is $L(\boldsymbol{x}, \boldsymbol{z}) = \sum_{n\in[N]} x_n(\mu_n + \sigma_n z_n)$. Inspired by Zymler et al. (2013), we consider the distributionally robust optimization model, which looks for an investment plan with minimal worst-case CVaR:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\text{-CVaR}_{\varepsilon}\left(-L(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right), \tag{33}$$

where the CVaR at level ε with respect to a probability distribution $\mathbb{P}$ is

$$\mathbb{P}\text{-CVaR}_{\varepsilon}\left(-L(\boldsymbol{x}, \boldsymbol{x})\right) = \min_{\theta\in\mathbb{R}} \left\{ \theta + \frac{1}{\varepsilon} \mathbb{E}_{\mathbb{P}}\left[(-L(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \theta)^+ \right] \right\}.$$

By the stochastic minimax theorem in Shapiro and Kleywegt (2002), we can rewrite the worst-case CVaR in the objective function of Problem (33) as

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\text{-CVaR}_{\varepsilon}\left(-L(\boldsymbol{x}, \boldsymbol{x})\right) = \min_{\theta\in\mathbb{R}} \left\{ \theta + \frac{1}{\varepsilon} \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[(-L(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \theta)^+ \right] \right\}.$$

As a result, the worst-case CVaR portfolio selection problem (33) becomes

$$\min_{\boldsymbol{x}\in\mathcal{X}, \theta\in\mathbb{R}} \left\{ \theta + \frac{1}{\varepsilon} \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[(-L(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \theta)^+ \right] \right\}.$$

We consider the following covariance ambiguity set, which encompasses the family of distributions of these independently distributed uncertainties $\{\tilde{z}_n\}_{n\in[N]}$:

$$\mathcal{F}_{\mathrm{C}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{0} \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^2\right] \le \boldsymbol{q}'\boldsymbol{q},\ \forall \boldsymbol{q} \in \mathbb{R}^N : \|\boldsymbol{q}\|_2 \le 1 \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{W}] = 1 \end{array} \right\},$$

where the support set is modeled as $\mathcal{W} = \{\boldsymbol{z} \in \mathbb{R}^N \mid \|\boldsymbol{z}\|_\infty \leq 1\}$. Because $\tilde{z}_n$ is sub-Gaussian with zero mean and unit deviation parameter for $n \in [N]$, we also consider the following entropic ambiguity set (which captures the sub-Gaussianity):

$$\mathcal{F}_{\mathrm{G}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{0} \\ \ln \mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'\tilde{\boldsymbol{z}}\right)\right] \leq \dfrac{1}{2}\boldsymbol{q}'\boldsymbol{q},\ \forall \boldsymbol{q} \in \mathbb{R}^N \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\};$$

here we explicitly specify the support set $\mathcal{W}$ because, by Theorem 2, the chosen function $\phi(\cdot)$ implies only that $\mathcal{W}$ lies in $\mathbb{R}^N$. As shown in the following result, the entropic dominance ambiguity set improves upon the covariance ambiguity set in terms of capturing independently distributed random variables with known mean and support.

Proposition 6. *Let $\tilde{z}_1, \ldots, \tilde{z}_N$ be independently distributed random variables of zero mean and with support $\mathcal{W} = \{\boldsymbol{z} \in \mathbb{R}^N \mid \|\boldsymbol{z}\|_\infty \leq 1\}$. The minimal covariance ambiguity set that encompasses this family of distributions is $\mathcal{F}_{\mathrm{C}}$. The entropic dominance ambiguity set $\mathcal{F}_{\mathrm{G}}$ yields a better characterization of this family of distributions: that is, $\mathcal{F}_{\mathrm{G}} \subseteq \mathcal{F}_{\mathrm{C}}$.*

We now investigate the numerical performance of the relaxed entropic dominance ambiguity set

$$\mathcal{G}_{\mathrm{G}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{0} \\ \ln \mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'\tilde{\boldsymbol{z}}\right)\right] \leq \dfrac{1}{2}\boldsymbol{q}'\boldsymbol{q},\ \forall \boldsymbol{q} \in \bar{\mathcal{Q}} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\}$$

against the covariance ambiguity set $\mathcal{F}_{\mathrm{C}}$. We initialize Algorithm GIP with $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n \mid n \in [N]\}$. In each iteration, we solve the separation problem by the trust region method with a randomly generated initial solution. If the resulting local minimum attains a negative objective value, then we find a violating expectation constraint and proceed to the next iteration; otherwise, we continue with the TRM while using another randomly generated initial solution. The algorithm terminates if no violating expectation constraint is found after 100 trials.

We report the objective values for various confidence levels, $\varepsilon \in \{0.01, 0.02, 0.05, 0.08, 0.1\}$, in Table 4. Note that Z_{C} denotes the objective value obtained by the covariance ambiguity set and that Z_{G}^i denotes the objective value achieved at the i-th iteration when using the entropic dominance ambiguity set. The relaxed entropic dominance solutions yield significantly lower objective values than those obtained from the covariance ambiguity set. Moreover, the entropic dominance approach converges reasonably well: it terminates in at most 10 iterations. As a robustness check, we used different initializations for the relaxed entropic dominance ambiguity set; we arrived at the same solutions and observed similar convergence results.

ε	0.01	0.02	0.05	0.08	0.1
Z_{C}	0.0674	0.0674	0.0674	0.0579	0.0430
Z_{G}^1	0.0674	0.0674	0.0674	0.0674	0.0674
Z_{G}^2	0.0674	0.0674	0.0674	0.0661	0.0661
Z_{G}^3	0.0474	0.0674	0.0484	0.0251	0.0569
Z_{G}^4	0.0443	0.0408	0.0247	0.0126	0.0517
Z_{G}^5	0.0442	0.0350	0.0201	0.0126	0.0486
Z_{G}^6	—	0.0347	0.0199	0.0106	0.0163
Z_{G}^7	—	—	0.0199	0.0105	0.0069
Z_{G}^8	—	—	—	0.0105	0.0055
Z_{G}^9	—	—	—	—	0.0054
Z_{G}^{10}	—	—	—	—	—

Table 4 **Objective values of the covariance approach and the entropic dominance approach.**

6. Future work

In this paper, we work with static distributionally robust optimization problems. In our numerical studies, we choose the marginal moment ambiguity set and the partial cross-moment ambiguity set as the relaxed covariance ambiguity set and identify "violating" expectation constraints to improve the solution. In a recent work, Bertsimas et al. (2018) study adaptive distributionally robust optimization problems and show the improvement of the partial cross-moment ambiguity set over the marginal moment ambiguity set. We believe that the extension of our approach to adaptive problems with infinitely constrained ambiguity sets and to systematically improve the partial cross-moment ambiguity set (or any relaxed ambiguity set) will be useful.

Acknowledgments

The authors would like to thank Daniel Kuhn and Wolfram Wiesemann for the helpful discussions and comments. The authors also thank Zhenyu Hu and Karthik Natarajan for their feedbacks. The research of Melvyn Sim is supported by the Singapore National Research Foundation National Cybersecurity R&D Grant under Grant R-263-000-C74-281 and Grant NRF2015NCR-NCR003-006. The research of Huan Xu is partly supported by the startup fund from the H. Milton Stewart School of Industrial and Systems Engineering of Georgia Institute of Technology.

References

Ben-Tal, Aharon, Dick Den Hertog, Anja De Waegenaere, Bertrand Melenberg, Gijs Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* **59**(2) 341–357.

Ben-Tal, Aharon, Arkadi Nemirovski. 1998. Robust convex optimization. *Mathematics of operations research* **23**(4) 769–805.

Ben-Tal, Aharon, Arkadi Nemirovski. 2001. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM.

Bertsimas, Dimitris, Ioana Popescu. 2002. On the relation between option and stock prices: a convex optimization approach. *Operations Research* **50**(2) 358–374.

Bertsimas, Dimitris, Ioana Popescu. 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* **15**(3) 780–804.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations research* **52**(1) 35–53.

Bertsimas, Dimitris, Melvyn Sim, Meilin Zhang. 2018. Adaptive distributionally robust optimization. *Management Science* .

Birge, John R, Francois Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.

Breton, Michèle, Saeb El Hachem. 1995. Algorithms for the solution of stochastic dynamic minimax problems. *Computational Optimization and Applications* **4**(4) 317–345.

Chares, Robert. 2009. Cones and interior-point algorithms for structured convex optimization involving powers and exponentials. *PhD thesis* .

Chen, Wenqing, Melvyn Sim. 2009. Goal-driven optimization. *Operations Research* **57**(2) 342–357.

Chen, Wenqing, Melvyn Sim, Jie Sun, Chung-Piaw Teo. 2010. From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations research* **58**(2) 470–485.

Chen, Xin, Melvyn Sim, Peng Sun. 2007. A robust optimization perspective on stochastic programming. *Operations Research* **55**(6) 1058–1071.

Chen, Xin, Melvyn Sim, Peng Sun, Jiawei Zhang. 2008. A linear decision-based approximation approach to stochastic programming. *Operations Research* **56**(2) 344–357.

Chen, Zhi, Melvyn Sim, Peng Xiong. 2017. Tractable distributionally robust optimization with data. URL `http://www.optimization-online.org/DB_FILE/2017/06/6055.pdf`.

Choi, Man-Duen, Tsit Yuen Lam, Bruce Reznick. 1995. Sums of squares of real polynomials. *Proceedings of Symposia in Pure mathematics*, vol. 58. American Mathematical Society, 103–126.

Conn, Andrew R, Nicholas IM Gould, Philippe L Toint. 2000. *Trust region methods*. SIAM.

Delage, Erick, Yinyu Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research* **58**(3) 595–612.

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

Dentcheva, Darinka, Andrzej Ruszczyński. 2006. Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance* **30**(2) 433–451.

Dupačová, Jitka. 1987. The minimax approach to stochastic programming and an illustrative application. *Stochastics: An International Journal of Probability and Stochastic Processes* **20**(1) 73–88.

El Ghaoui, Laurent, Maksim Oks, Francois Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations research* **51**(4) 543–556.

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

Esfahani, Peyman Mohajerin, Daniel Kuhn. 2017. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 1–52.

Gao, Rui, Anton J Kleywegt. 2016. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199* .

Gilboa, Itzhak, David Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of mathematical economics* **18**(2) 141–153.

Goh, Joel, Melvyn Sim. 2010. Distributionally robust optimization and its tractable approximations. *Operations research* **58**(4-part-1) 902–917.

Goh, Joel, Melvyn Sim. 2011. Robust optimization made easy with rome. *Operations Research* **59**(4) 973–985.

Grant, Michael, Stephen Boyd, Yinyu Ye. 2008. Cvx: Matlab software for disciplined convex programming.

Grant, Michael C, Stephen P Boyd. 2008. Graph implementations for nonsmooth convex programs. *Recent advances in learning and control*. Springer, 95–110.

Hanasusanto, Grani A, Daniel Kuhn, Wolfram Wiesemann. 2016. A comment on computational complexity of stochastic programming problems. *Mathematical Programming* **159**(1-2) 557–569.

Hanasusanto, Grani A, Vladimir Roitch, Daniel Kuhn, Wolfram Wiesemann. 2017. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research* **65**(3) 751–767.

He, Simai, Jiawei Zhang, Shuzhong Zhang. 2010. Bounding probability of small deviation: A fourth moment approach. *Mathematics of Operations Research* **35**(1) 208–232.

Isii, Keiiti. 1962. On sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics* **14**(1) 185–197.

Jiang, Ruiwei, Yongpei Guan. 2016. Data-driven chance constrained stochastic program. *Mathematical Programming* **158**(1-2) 291–327.

Kaas, Rob, Marc Goovaerts, Jan Dhaene, Michel Denuit. 2008. *Modern actuarial risk theory: using R*, vol. 128. Springer Science & Business Media.

Lasserre, Jean B. 2001. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization* **11**(3) 796–817.

Lasserre, Jean-Bernard, Tomas Prieto-Rumeau, Mihail Zervos. 2006. Pricing a class of exotic options via moments and sdp relaxations. *Mathematical Finance* **16**(3) 469–494.

Mathworks. 2017. Unconstrained nonlinear optimization algorithms. URL `www.mathworks.com/help/optim/ug/unconstrained-nonlinear-optimization-algorithms.html`.

Nemirovski, Arkadi, Alexander Shapiro. 2006. Convex approximations of chance constrained programs. *SIAM Journal on Optimization* **17**(4) 969–996.

Nesterov, Yurii. 2000. Squared functional systems and optimization problems. *High performance optimization*. Springer, 405–440.

Parrilo, Pablo A. 2003. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming* **96**(2) 293–320.

Pflug, Georg, David Wozabal. 2007. Ambiguity in portfolio selection. *Quantitative Finance* **7**(4) 435–442.

Postek, Krzysztof, Aharon Ben-Tal, Dick den Hertog, Bertrand Melenberg. 2018. Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research* .

Rockafellar, R Tyrrell, Stanislav Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of banking & finance* **26**(7) 1443–1471.

Scarf, Herbert. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* .

Shapiro, Alexander. 2001. On duality theory of conic linear problems. *Semi-infinite programming*. Springer, 135–165.

Shapiro, Alexander, Anton Kleywegt. 2002. Minimax analysis of stochastic problems. *Optimization Methods and Software* **17**(3) 523–542.

Sion, Maurice. 1958. On general minimax theorems. *Pacific Journal of mathematics* **8**(1) 171–176.

Skajaa, Anders, Yinyu Ye. 2015. A homogeneous interior-point algorithm for nonsymmetric convex conic optimization. *Mathematical Programming* **150**(2) 391–422.

Soyster, Allen L. 1973. Technical noteconvex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research* **21**(5) 1154–1157.

von Neumann, John, Oskar Morgenstern. 1947. *Theory of games and economic behavior*. Princeton university press.

Wainwright, JM. 2015. *High-dimensional statistics: A non-asymptotic viewpoint*. Preparation. University of California, Berkeley.

Wang, Zizhuo, Peter W Glynn, Yinyu Ye. 2016. Likelihood robust optimization for data-driven problems. *Computational Management Science* **13**(2) 241–261.

Wiesemann, Wolfram, Daniel Kuhn, Melvyn Sim. 2014. Distributionally robust convex optimization. *Operations Research* **62**(6) 1358–1376.

Zhao, Chaoyue, Yongpei Guan. 2018. Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters* **46**(2) 262–267.

Zymler, Steve, Daniel Kuhn, Berç Rustem. 2013. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* 1–32.

Eletronic Companion

Appendix A

We conduct an experiment using randomly generated instances of the following distributionally robust optimization problem

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{F}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right],$$

where $\mathcal{F}_{\mathrm{C}}$ is a covariance ambiguity set. The setting of this experiment is identical to that in the paper. We increase the size of the problem (as determined by K, M, and N) so that the computer can not solve the exact semidefinite reformulation within an hour. We implement Algorithm GIP by using the partial cross-moment ambiguity set with $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n \mid n \in [N]\} \cup \{\boldsymbol{e}\}$ as the initial relaxed ambiguity set. We terminate Algorithm GIP after 50 iterations (the accumulative solution time of all iterations is less than an hour).

Let Z_{C}^i denote the objective value achieved at the i-th iteration when using the relaxed ambiguity set. We report in Table EC.1 the relative gap (over 10 random instances for each size of the problem) between the initial solution Z_{C}^1 and the terminal solution Z_{C}^{50}, which is given by

$$\left| \frac{Z_{\mathrm{C}}^1 - Z_{\mathrm{C}}^{50}}{Z_{\mathrm{C}}^{50}} \right| \times 100\%.$$

The result reveals that by solving a sequence of second-order conic subproblems, Algorithm GIP yields iteratively an improved solution, even when the exact semidefinite program can not be solved within the time limit. This experiment supports that having a second-order conic relaxation can be helpful when the exact semidefinite program is of a large size.

(K, M, N)	$(110, 20, 50)$	$(100, 20, 60)$	$(80, 20, 70)$
Relative improvement	12.0% (10.5%, 12.1%, 13.9%)	12.8% (10.9%, 12.9%, 14.4%)	14.3% (13.3%, 14.4%, 16.2%)

Table EC.1 Median (minimal, average, maximal) relative gap.

Appendix B

We conduct an experiment using randomly generated instances of the following distributionally robust optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],$$

where $\mathcal{F}_{\mathrm{C}}$ is a covariance ambiguity set. The setting of this experiment is identical to that in the paper except that now the feasible set $\mathcal{X}=\{\boldsymbol{x}\in\mathbb{Z}_+^M \mid \boldsymbol{e}'\boldsymbol{x}\leq M\}$. In such cases, the reformulation would become a semidefinite optimization problem involving discrete decision variables and is hard to solve. Instead, we can first consider the linear relaxation of $\mathcal{X}$ (denoted by $\bar{\mathcal{X}}$), solve the problem

$$\min_{\boldsymbol{x}\in\bar{\mathcal{X}}} \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],$$

and obtain the $\mathcal{G}_{\mathrm{C}}^*$ through Theorem 6. Afterward, we implement Algorithm GIP by using $\mathcal{G}_{\mathrm{C}}^*$ as the initial relaxed ambiguity set and solve a sequence of mixed-integer second-order conic programs to obtain approximate solutions to the problem of interest. The problem size is given by $(K,M,N)=(10,12,15)$ and we terminate Algorithm GIP after 50 iterations.

Let Z_{C}^i denote the objective value achieved at the i-th iteration when using the relaxed ambiguity set. We report in Table EC.2 the relative gap (over 100 random instances) between the intermediate solution Z_{C}^i and the terminal solution Z_{C}^{50}, which is given by

$$\left|\frac{Z_{\mathrm{C}}^i - Z_{\mathrm{C}}^{50}}{Z_{\mathrm{C}}^{50}}\right| \times 100\%.$$

The result reveals that by solving a sequence of mixed-integer second-order conic subproblems, Algorithm GIP yields iteratively an improved solution. We also observe that the improvements in the first few iterations are larger than those in the later iterations. This experiment supports that having a second-order conic relaxation can be helpful when the exact semidefinite program involves discrete decision variables.

i	1	10	20	30	40
Relative improvement	10.4% (1.2%, 10.6%, 21.0%)	3.6% (0.4%, 3.7%, 8.6%)	1.7% (0.2%, 1.7%, 4.4%)	0.9% (0.1%, 0.9%, 2.2%)	0.4% (0.0%, 0.4%, 0.9%)

Table EC.2 Median (minimal, average, maximal) relative gap.

Appendix C

Proof of Theorem 2. Consider any probability distribution $\mathbb{P} \in \mathcal{F}_{\mathrm{E}}$ and observe that

$$\mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n - \mu_n\right] = \lim_{\alpha \to 0} \frac{\mathbb{E}_{\mathbb{P}}\left[\exp\left(\alpha(\tilde{z}_n - \mu_n)\right)\right] - 1}{\alpha}.$$

Let $\alpha \to 0^+ (0^-)$; then

$$\lim_{\alpha \to 0^+(0^-)} \frac{\mathbb{E}_{\mathbb{P}}\left[\exp\left(\alpha(\tilde{z}_n - \mu_n)\right)\right] - 1}{\alpha} \leq (\geq) \lim_{\alpha \to 0^+} \frac{\exp\left(\phi(\alpha \boldsymbol{e}_n)\right) - \exp\left(\phi(\mathbf{0})\right)}{\alpha} = \left.\frac{\partial \phi\left(\boldsymbol{q}\right)}{\partial q_n}\right|_{\boldsymbol{q}=\mathbf{0}}.$$

Thus, for every $n \in [N]$ we have

$$\mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n - \mu_n\right] = \left.\frac{\partial \phi\left(\boldsymbol{q}\right)}{\partial q_n}\right|_{\boldsymbol{q}=\mathbf{0}} = 0,$$

which implies that $\mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu}$.

For any $\alpha \in \mathbb{R}$ and $\boldsymbol{q} \in \mathbb{R}^N$, we have $\mathbb{E}_{\mathbb{P}}\left[\exp\left(\alpha \boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)\right] \leq \exp\left(\phi\left(\alpha \boldsymbol{q}\right)\right)$; this inequality implies

$$\lim_{\alpha \to 0} \frac{\mathbb{E}_{\mathbb{P}}\left[\exp\left(\alpha \boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)\right] - \exp\left(\phi\left(\alpha \boldsymbol{q}\right)\right)}{\alpha^2} \leq 0, \ \forall \boldsymbol{q} \in \mathbb{R}^N. \tag{EC.1}$$

Taking Taylor expansion up to second order at 0, the first term in the numerator of (EC.1) becomes

$$1 + \alpha \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right] + \frac{1}{2}\alpha^2 \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2\right] + o\left(\alpha^2\right) = 1 + \frac{1}{2}\alpha^2 \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2\right] + o\left(\alpha^2\right)$$

and the second term becomes

$$\begin{aligned}
\exp(\phi(\alpha \boldsymbol{q})) &= \exp\left(\phi(\mathbf{0}) + \alpha \nabla \phi(\mathbf{0})' \boldsymbol{q} + \frac{1}{2}\alpha^2 \boldsymbol{q}' \nabla^2 \phi(\mathbf{0}) \boldsymbol{q} + o(\alpha^2)\right) \\
&= \exp\left(\frac{1}{2}\alpha^2 \boldsymbol{q}' \nabla^2 \phi(\mathbf{0}) \boldsymbol{q} + o(\alpha^2)\right) \\
&= 1 + \frac{1}{2}\alpha^2 \boldsymbol{q}' \nabla^2 \phi(\mathbf{0}) \boldsymbol{q} + o(\alpha^2).
\end{aligned}$$

Hence, inequality (EC.1) becomes

$$\begin{aligned}
&\lim_{\alpha \to 0} \frac{1 + \frac{1}{2}\alpha^2 \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2\right] - 1 - \frac{1}{2}\alpha^2 \boldsymbol{q}' \nabla^2 \phi(\mathbf{0}) \boldsymbol{q} + o\left(\alpha^2\right)}{\alpha^2} \\
&= \lim_{\alpha \to 0} \frac{\alpha^2 \left(\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2\right] - \boldsymbol{q}' \nabla^2 \phi(\mathbf{0}) \boldsymbol{q}\right) + o\left(\alpha^2\right)}{2\alpha^2} \\
&= \frac{\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2\right] - \boldsymbol{q}' \nabla^2 \phi(\mathbf{0}) \boldsymbol{q}}{2} \leq 0, \ \forall \boldsymbol{q} \in \mathbb{R}^N,
\end{aligned}$$

which implies that $\mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'\right] \preceq \nabla^2 \phi(\mathbf{0})$.

It is well known (for instance, Kaas et al. 2008) that the function $\ln \mathbb{E}_{\mathbb{P}}\left[\exp(\alpha \tilde{z})\right] / \alpha$ is increasing in α and satisfies

$$\lim_{\alpha \to \infty} \ln \mathbb{E}_{\mathbb{P}}\left[\exp(\alpha \tilde{z})\right] / \alpha = \operatorname{ess\,sup}\left(\tilde{z}\right).$$

Hence, taking the supremum over the distributional ambiguity set $\mathcal{F}_{\rm E}$ preserves the monotonicity and we have

$$\sup_{\mathbb{P}\in\mathcal{F}_{\rm E}} \operatorname{ess\,sup}(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})) = \inf\{\delta \mid \mathbb{P}\left[\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}) \le \delta\right] = 1, \forall \mathbb{P}\in\mathcal{F}_{\rm E}\}.$$

The superset of the support set then follows from

$$\operatorname{ess\,sup}\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})\right) = \lim_{\alpha\to\infty}\frac{1}{\alpha}\ln\mathbb{E}_{\mathbb{P}}\left[\exp(\alpha\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))\right] \le \lim_{\alpha\to\infty}\frac{1}{\alpha}\phi(\alpha\boldsymbol{q}), \forall\tilde{\boldsymbol{z}}\sim\mathbb{P}, \mathbb{P}\in\mathcal{F}_{\rm E}.$$

If $\phi(\boldsymbol{q})$ has the additive form then we have $\dfrac{\partial\phi(\boldsymbol{q})}{\partial q_n} = \phi_n'(q_n)$ for all $n\in[N]$. Therefore,

$$\frac{\partial^2\phi(\boldsymbol{q})}{\partial q_l \partial q_n} = \begin{cases} \phi_n''(q_n) & \text{for all } l,n\in[N], l=n \\ 0 & \text{otherwise,} \end{cases}$$

from which the equality $\nabla^2\phi(0) = \operatorname{diag}\left(\phi_1''(0),\ldots,\phi_N''(0)\right)$ naturally follows. Furthermore, the class of constraints in the support set is satisfied if and only if

$$\max_{\boldsymbol{q}\in\mathbb{R}^N}\left\{\boldsymbol{q}'(\boldsymbol{z}-\boldsymbol{\mu}) - \lim_{\alpha\to\infty}\frac{1}{\alpha}\sum_{n\in[N]}\phi_n(\alpha q_n)\right\} \le 0.$$

Because the maximization operates additively in $\boldsymbol{q}$, the preceding inequality is equivalent to

$$\sum_{n\in[N]}\left(\max_{q_n\in\mathbb{R}}\left\{q_n(z_n-\mu_n) - \lim_{\alpha\to\infty}\frac{\phi_n(\alpha q_n)}{\alpha}\right\}\right) \le 0. \tag{EC.2}$$

Observe that

$$q_n(z_n-\mu_n) - \lim_{\alpha\to\infty}\frac{\phi_n(\alpha q_n)}{\alpha} = \begin{cases} q_n\Big(z_n-\mu_n-\lim\limits_{\alpha\to\infty}\dfrac{\phi_n(\alpha)}{\alpha}\Big) & q_n\ge 0 \\ q_n\Big(z_n-\mu_n-\lim\limits_{\alpha\to-\infty}\dfrac{\phi_n(\alpha)}{\alpha}\Big) & q_n\le 0, \end{cases}$$

thus the inequality (EC.2) holds if and only if

$$\mu_n + \lim_{\alpha\to-\infty}\frac{\phi_n(\alpha)}{\alpha} \le z_n \le \mu_n + \lim_{\alpha\to\infty}\frac{\phi_n(\alpha)}{\alpha},\ \forall n\in[N]. \quad \square$$

Proof of Proposition 1. Because $\{\tilde{z}_n\}_{n\in[N]}$ are independent, we have $\mathbb{P}_0\in\mathcal{F}_{\rm E}$; this implies that

$$\rho_0(x_0,\boldsymbol{x}) = \mathbb{E}_{\mathbb{P}_0}\left[(x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right] \le \sup_{\mathbb{P}\in\mathcal{F}_{\rm E}}\mathbb{E}_{\mathbb{P}}\left[(x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right] = \rho_{\rm E}(x_0,\boldsymbol{x}).$$

Since $\omega^+ \le \eta\exp\left(\omega/\eta - 1\right)$ for all $\eta > 0$, it follows that

$$\mathbb{E}_{\mathbb{P}}\left[(x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right] \le \mathbb{E}_{\mathbb{P}}\left[\frac{\eta}{e}\exp\left(\frac{x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right],\ \forall\mathbb{P}\in\mathcal{F}_{\rm E}, \forall\eta>0.$$

By the definition of the entropic dominance ambiguity set, we then have

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{\boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right] \le \exp\left(\frac{\boldsymbol{x}'\boldsymbol{\mu}}{\eta} + \sum_{n\in[N]}\phi_n\left(\frac{x_n}{\eta}\right)\right),\ \forall\mathbb{P}\in\mathcal{F}_{\rm E}.$$

This inequality implies that, for any $\mathbb{P} \in \mathcal{F}_{\mathrm{E}}$,

$$\mathbb{E}_{\mathbb{P}}\left[\frac{\eta}{e}\exp\left(\frac{x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right] \leq \frac{\eta}{e}\exp\left(\frac{x_0 + \boldsymbol{x}'\boldsymbol{\mu}}{\eta} + \sum_{n\in[N]}\phi_n\left(\frac{x_n}{\eta}\right)\right),\ \forall \eta > 0.$$

Combining these inequalities, we obtain

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{E}}}\mathbb{E}_{\mathbb{P}_0}\left[(x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right] \leq \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{E}}}\mathbb{E}_{\mathbb{P}}\left[\frac{\eta}{e}\exp\left(\frac{x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right] \leq \inf_{\eta>0}\frac{\eta}{e}\exp\left(\frac{x_0 + \boldsymbol{x}'\boldsymbol{\mu}}{\eta} + \sum_{n\in[N]}\phi_n\left(\frac{x_n}{\eta}\right)\right);$$

that is, $\rho_{\mathrm{E}}(x_0, \boldsymbol{x}) \leq \rho_{\mathrm{B}}(x_0, \boldsymbol{x})$. □

Proof of Proposition 2. We can derive $\rho_0(0,1)$ and $\rho_{\mathrm{B}}(0,1)$ as follows:

$$\rho_0(0,1) = \int_{-\infty}^{\infty} z^+ \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{1}{2\sqrt{2\pi}}\int_0^{\infty} e^{-y/2} dy = \frac{1}{\sqrt{2\pi}},$$

and

$$\rho_{\mathrm{B}}(0,1) = \inf_{\eta>0}\mathbb{E}_{\mathbb{P}_0}\left(\frac{\eta}{e}\exp\left(\frac{\tilde{z}}{\eta}\right)\right) = \inf_{\eta>0}\frac{\eta}{e}\exp\left(\frac{1}{2\eta^2}\right) = \frac{1}{\sqrt{e}}.$$

Since $\mathbb{P}_0$ is the standard normal distribution, the entropic dominance ambiguity set becomes

$$\mathcal{F}_{\mathrm{E}} = \left\{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}) \ \middle|\ \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \ln \mathbb{E}_{\mathbb{P}}\left[\exp(q\tilde{z})\right] \leq \dfrac{q^2}{2},\ \forall q \in \mathbb{R} \end{array}\right\},$$

which implies (by Theorem 2) that $\mathbb{E}_{\mathbb{P}}[\tilde{z}] = 0$ and $\mathbb{E}_{\mathbb{P}}[\tilde{z}^2] \leq 1$. Thus, for $\mathbb{P} \in \mathcal{F}_{\mathrm{E}}$ we have

$$\mathbb{E}_{\mathbb{P}}\left[\tilde{z}^+\right] = \frac{\mathbb{E}_{\mathbb{P}}\left[|\tilde{z}| + \tilde{z}\right]}{2} = \frac{\mathbb{E}_{\mathbb{P}}\left[|\tilde{z}|\right]}{2} \leq \frac{\sqrt{\mathbb{E}_{\mathbb{P}}\left[\tilde{z}^2\right]}}{2} \leq \frac{1}{2},$$

which implies $\rho_{\mathrm{E}}(0,1) \leq 1/2$. Observe that this inequality is binding with a two-point distribution that has weight 0.5 at -1 and weight 0.5 at 1. To conclude that $\rho_{\mathrm{E}}(0,1) = 1/2$, it is sufficient to show that this two-point distribution lies in $\mathcal{F}_{\mathrm{E}}$. Toward that end, we shall establish that

$$\ln\frac{e^q + e^{-q}}{2} \leq \frac{q^2}{2},\ \forall q \in \mathbb{R}.$$

In fact, by considering Taylor series we obtain

$$\frac{e^q + e^{-q}}{2} = \sum_{n=0}^{\infty}\frac{q^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty}\frac{q^{2n}}{2^n n!} = \sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{q^2}{2}\right)^n = \exp\left(\frac{q^2}{2}\right),\ \forall q \in \mathbb{R},$$

where the inequality follows because $(2n)! = 2n(2n-1)\cdots(n+1)n! \geq 2^n n!$ for all $n \in \mathbb{Z}_+$. □

Proof of Proposition 3. The inner supremum in Problem (24) is given by the following classical robust optimization problem:

$$\begin{aligned} \inf\ & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma}\rangle \\ \text{s.t.}\ & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \langle \boldsymbol{\Gamma}, \boldsymbol{U}\rangle \geq \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} + b_k(\boldsymbol{x}),\ \forall(\boldsymbol{z}, 1) \in \mathcal{K}(\mathcal{W}), \boldsymbol{U} \succeq (\boldsymbol{z} - \boldsymbol{\mu})(\boldsymbol{z} - \boldsymbol{\mu})', \forall k \in [K] \\ & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\Gamma} \succeq \boldsymbol{0}. \end{aligned}$$

Here we have introduced the auxiliary variable $\boldsymbol{U} \in \mathbb{R}^{N\times N}$ and represent the constraint $\boldsymbol{z} \in \mathcal{W}$ by $(\boldsymbol{z},1) \in \mathcal{K}(\mathcal{W})$. Using the Schur complement, we know that each k-th robust counterpart is not violated if and only if

$$
\begin{aligned}
\inf\ & (\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\boldsymbol{z} + \langle \boldsymbol{\Gamma}, \boldsymbol{U}\rangle \\
\text{s.t. } & \begin{pmatrix} 1 & (\boldsymbol{z}-\boldsymbol{\mu})' \\ (\boldsymbol{z}-\boldsymbol{\mu}) & \boldsymbol{U} \end{pmatrix} \succeq \boldsymbol{0} \\
& (\boldsymbol{z},1) \in \mathcal{K}_{\mathcal{W}}
\end{aligned}
$$

is not less than $b_k(\boldsymbol{x}) - \alpha$. The dual of the above problem is

$$
\begin{aligned}
\sup\ & 2\boldsymbol{\chi}_k'\boldsymbol{\mu} - \delta_k - t_k \\
\text{s.t. } & \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - 2\boldsymbol{\chi}_k = \boldsymbol{0} \\
& \boldsymbol{\Gamma} = \boldsymbol{\Gamma}_k \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*_{\mathcal{W}} \\
& \begin{pmatrix} \delta_k & \boldsymbol{\chi}_k' \\ \boldsymbol{\chi}_k & \boldsymbol{\Gamma}_k \end{pmatrix} \succeq \boldsymbol{0}.
\end{aligned}
$$

Substituting this dual formulation and then performing the outer and inner minimizations jointly, we obtain the formulation (25). □

Proof of Proposition 4. The inner supremum in Problem (27) can be reformulated as

$$
\begin{aligned}
\inf\ & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j \boldsymbol{q}_j'\boldsymbol{\Sigma}\boldsymbol{q}_j \\
\text{s.t. } & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{j\in[J]} \gamma_j (\boldsymbol{q}_j'(\boldsymbol{z}-\boldsymbol{\mu}))^2 \geq \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} + b_k(\boldsymbol{x}),\ \forall (\boldsymbol{z},1) \in \mathcal{K}(\mathcal{W}), \forall k \in [K] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}^J_+.
\end{aligned}
$$

In this formulation, each k-th constraint is not violated if and only if the optimal value of the problem

$$
\begin{aligned}
\inf\ & (\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\,\boldsymbol{z} + \boldsymbol{\gamma}'\boldsymbol{u} \\
\text{s.t. } & u_j \geq \left(\boldsymbol{q}_j'(\boldsymbol{z}-\boldsymbol{\mu})\right)^2,\ \forall j \in [J] \\
& (\boldsymbol{z},1) \in \mathcal{W}
\end{aligned}
$$

is not less than $b_k(\boldsymbol{x}) - \alpha$. Here, we have introduced an auxiliary vector $\boldsymbol{u}$ such that the u_j is associated with the epigraph of quadratic function $\left(\boldsymbol{q}_j'(\boldsymbol{z}-\boldsymbol{\mu})\right)^2$. The first set of constraints has an equivalent second-order conic representation as $\big(u_j+1, u_j-1, 2\boldsymbol{y}_j'(\boldsymbol{z}-\boldsymbol{\mu})\big) \in \mathcal{K}_{\text{SOC}}$, which enables us to derive the dual problem as follows:

$$
\begin{aligned}
\sup\ & \sum_{j\in[J]} (m_{jk} - l_{jk} + 2n_{jk}\boldsymbol{q}_j'\boldsymbol{\mu}) - t_k \\
\text{s.t.}\ & \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - \sum_{j\in[J]} 2n_{jk}\boldsymbol{q}_j = \boldsymbol{0} \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*(\mathcal{W}) \\
& \gamma_j = l_{jk} + m_{jk}, && \forall j \in [J] \\
& (l_{jk}, m_{jk}, n_{jk}) \in \mathcal{K}_{\text{SOC}}, && \forall j \in [J] \\
& \boldsymbol{l}_k, \boldsymbol{m}_k, \boldsymbol{n}_k \in \mathbb{R}^J.
\end{aligned}
$$

Note that set of conic constraints follows from the self-dual nature of the second-order cone; that is, $\mathcal{K}_{\text{SOC}} = \mathcal{K}_{\text{SOC}}^*$. Re-injecting this dual into the preceding reformulation of the inner supremum and then combining it with the outer minimization over $\boldsymbol{x} \in \mathcal{X}$, we obtain (28). The proof of the proposition's second part is a direct implementation of Theorem 4 and so is omitted. □

Proof of Proposition 5. Observe that the set of expectation constraints can be equivalently represented as

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}) - \phi(\boldsymbol{q})\right)\right] \le 1.$$

The inner supremum in Problem (32) can be reformulated as the following optimization problem:

$$
\begin{aligned}
\inf\ & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j \\
\text{s.t.}\ & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{j\in[J]} \gamma_j \exp\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}) - \phi(\boldsymbol{q})\right) \ge f(\boldsymbol{x}, \boldsymbol{z}),\ \forall \boldsymbol{z} \in \mathcal{W} \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}_+^J.
\end{aligned} \tag{EC.3}
$$

The first constraint is equivalent to a set of robust counterparts such that, for all $k \in [K]$,

$$\alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{j\in[J]} \gamma_j \exp\left(\boldsymbol{q}_j'(\boldsymbol{z} - \boldsymbol{\mu}) - \phi(\boldsymbol{q}_j)\right) \ge \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} + b_k(\boldsymbol{x}),\ \forall \boldsymbol{z} \in \mathcal{W};$$

each of these counterparts is satisfied if and only if the optimal value of the problem

$$\inf_{\boldsymbol{z}\in\mathcal{W}} \left\{ (\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\boldsymbol{z} + \sum_{j\in[J]} \gamma_j \exp(\boldsymbol{q}_j'(\boldsymbol{z} - \boldsymbol{\mu}) - \phi(\boldsymbol{q}_j)) \right\} \tag{EC.4}$$

is not less than $b_k(\boldsymbol{x}) - \alpha$. Introducing an auxiliary vector $\boldsymbol{v} \in \mathbb{R}^J$ and then re-expressing $\boldsymbol{z} \in \mathcal{W}$ as $(\boldsymbol{z}, 1) \in \mathcal{K}(\mathcal{W})$, we obtain a representation that is equivalent to Problem (EC.4) as follows:

$$
\begin{aligned}
\inf\ & (\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\boldsymbol{z} + \boldsymbol{\gamma}'\boldsymbol{v} \\
\text{s.t.}\ & v_j \ge \exp(\boldsymbol{q}_j'(\boldsymbol{z} - \boldsymbol{\mu}) - \phi(\boldsymbol{q}_j)),\ \forall j \in [J] \\
& (\boldsymbol{z}, 1) \in \mathcal{K}(\mathcal{W}).
\end{aligned} \tag{EC.5}
$$

We can represent the exponential constraints as set constraints $\left(\boldsymbol{q}_j'(\boldsymbol{z}-\boldsymbol{\mu})-\phi(\boldsymbol{q}_j),1,v_j\right)\in\mathcal{K}_{\text{EXP}}$ and obtain the dual of Problem (EC.5) as

$$\begin{aligned}
\sup\ & \sum_{j\in[J]} l_{kj}(\boldsymbol{q}_j'\boldsymbol{\mu}+\phi(\boldsymbol{q}_j))-\sum_{j\in[J]} m_{kj}-t_k\\
\text{s.t.}\ & \boldsymbol{\beta}-\boldsymbol{a}_k(\boldsymbol{x})-\boldsymbol{r}_k-\sum_{j\in[J]} l_{kj}\boldsymbol{q}_j=\boldsymbol{0}\\
& \boldsymbol{\gamma}-\boldsymbol{n}_k=\boldsymbol{0}\\
& (l_{kj},m_{kj},n_{kj})\in\mathcal{K}_{\text{EXP}}^*,\ \forall j\in[J]\\
& (\boldsymbol{r}_k,t_k)\in\mathcal{K}^*(\mathcal{W}).
\end{aligned}$$

If we substitute this dual reformulation into (EC.3) and then perform the outer and inner minimizations jointly, the result is the conic reformulation. The proof for obtaining the worst-case distribution follows from a straightforward application of Theorem 4 and is therefore omitted. □

Proof of Proposition 6. Observe that for any $\boldsymbol{q}\in\mathbb{R}^N$,

$$\mathbb{E}_{\mathbb{P}}[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^2]=\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in[N]}\sum_{j\in[N]} q_iq_j\tilde{z}_i\tilde{z}_j\right]=\sum_{n\in[N]} q_n^2\mathbb{E}_{\mathbb{P}}[\tilde{z}_n^2];$$

here the second equality follows from the mutual independence and zero means of random variables. To show that $\mathcal{F}_{\text{C}}$ is the required minimal covariance ambiguity set, we need only to establish that

$$\boldsymbol{q}'\boldsymbol{q}=\sum_{n\in[N]} q_n^2\sup_{\mathbb{P}\in\mathcal{F}_n}\mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n^2\right],\ \forall\boldsymbol{q}\in\mathbb{R}^N, \tag{EC.6}$$

where for every $n\in[N]$ we define $\mathcal{F}_n=\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^N\right)\mid\tilde{z}\sim\mathbb{P},\ \mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n\right]=0,\ \mathbb{P}\left[|\tilde{z}_n|\le 1\right]=1\}$. By weak duality, we have

$$\sup_{\mathbb{P}\in\mathcal{F}_n}\mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n^2\right]\le\inf\{\alpha\mid\alpha,\beta\in\mathbb{R},\ \alpha+\beta z_n\ge z_n^2,\ \forall|z_n|\le 1\}.$$

Since $z_n^2-\beta z_n$ is convex in z_n, it follows that

$$\alpha\ge\max_{|\tilde{z}_n|\le 1,\beta}\left(z_n^2-\beta z_n\right)=\max_{\beta}\left\{1-\beta,1+\beta\right\};$$

this expression implies that $\sup_{\mathbb{P}\in\mathcal{F}_n}\mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n^2\right]\le\inf\alpha=1$. Strong duality holds because the displayed equality can be achieved by a two-point distribution in $\mathcal{F}_n$ that has weight 0.5 at -1 and weight 0.5 at 1. Hence we conclude that the equality (EC.6) holds.

We now consider any probability distribution $\mathbb{P}\in\mathcal{F}_{\text{G}}$. According to Theorem 2, the means of random variables are given by $\mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right]=\nabla\phi(\boldsymbol{0})=\boldsymbol{0}$ and their covariance is bounded by the identity matrix: $\mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}'\right]\preceq\nabla^2\phi(\boldsymbol{0})=\boldsymbol{I}$; that is, $\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^2\right]\le\boldsymbol{q}'\boldsymbol{q},\forall\boldsymbol{q}\in\mathbb{R}^N$. It now follows that $\mathcal{F}_{\text{G}}\subseteq\mathcal{F}_{\text{C}}$. □