A Distributionally Robust Perspective on Uncertainty Quantification and Chance Constrained Programming

Grani A. Hanasusanto · Vladimir Roitch · Daniel Kuhn · Wolfram Wiesemann

Abstract The objective of uncertainty quantification is to certify that a given physical, engineering or economic system satisfies multiple safety conditions with high probability. A more ambitious goal is to actively influence the system so as to guarantee and maintain its safety, a scenario which can be modeled through a chance constrained program. In this paper we assume that the parameters of the system are governed by an ambiguous distribution that is only known to belong to an ambiguity set characterized through generalized moment bounds and structural properties such as symmetry, unimodality or independence patterns. We delineate the watershed between tractability and intractability in ambiguity-averse uncertainty quantification and chance constrained programming. Using tools from distributionally robust optimization, we derive explicit conic reformulations for tractable problem classes and suggest efficiently computable conservative approximations for intractable ones.

1 Introduction

Consider a physical, engineering or economic system and encode its state through a parameter vector $z \in \mathbb{R}^P$. Suppose that the reliable operation of the

Grani A. Hanasusanto
Department of Computing, Imperial College London, United Kingdom, E-mail: g.hanasusanto11@imperial.ac.uk

Vladimir Roitch
Department of Computing, Imperial College London, United Kingdom, E-mail: vladimir.roitch@imperial.ac.uk

Daniel Kuhn
Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland, E-mail: daniel.kuhn@epfl.ch

Wolfram Wiesemann
Imperial College Business School, Imperial College London, United Kingdom, E-mail: ww@imperial.ac.uk

system requires that some safety constraints $\boldsymbol{Sz} \leq \boldsymbol{t}$ must be satisfied, where $\boldsymbol{S} \in \mathbb{R}^{J \times P}$ is termed the *technology matrix* and $\boldsymbol{t} \in \mathbb{R}^J$ the *right-hand side vector*. Due to measurement errors, limited observability and missing data, the parameter vector is uncertain for the vast majority of systems of practical interest and must therefore be modeled as a random variable $\tilde{\boldsymbol{z}}$ governed by a probability distribution $\mathbb{Q}$. In this situation a problem of great practical importance is to certify that the system is safe with high confidence. Formally, one should ascertain the satisfaction of the inequality $\mathbb{Q}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}] \geq 1-\epsilon$, where $\epsilon \in (0, 1)$ represents a prescribed safety tolerance or violation probability. Examples of societally relevant safety constraints include the prevention of blackouts in electricity grids, the containment of inflation in national economies, the limitation of seismic damages in structural engineering, the assurance of quality of service standards in telecommunication systems, the limitation of the likelihood to develop cancer due to the exposure to a substance etc., see [37].

In most real-life applications, evaluating the exact probability of safe operation is very challenging, if not impossible. On the one hand, the probability distribution $\mathbb{Q}$ is typically unknown as $\mathbb{Q}$ may only be indirectly observable through historical samples, which could be explained by several strikingly different distributions. On the other hand, even if $\mathbb{Q}$ was precisely known, the computation of $\mathbb{Q}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}]$ would require the evaluation of an integral over a possibly high-dimensional polytope, which is computationally cumbersome.

A remedy for the first difficulty is to adopt a distributionally robust approach and to embrace the fact that $\mathbb{Q}$ is merely known to belong to an *ambiguity set $\mathcal{P}$*. This set is typically defined as the family of all distributions that share certain known moments (mean, variance, covariances, higher-order moments, median-absolute deviation etc.) or known structural properties (symmetry, unimodality, multimodality, independence patterns, tail behavior etc.) with the otherwise unknown distribution $\mathbb{Q}$. The ambiguity of $\mathbb{Q}$ prompts us to investigate the *uncertainty quantification problem*

$$\inf_{\mathbb{P} \in \mathcal{P}} \ \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}], \tag{1}$$

which quantifies the worst-case probability of safe operation with respect to all distributions $\mathbb{P} \in \mathcal{P}$. As $\mathbb{Q} \in \mathcal{P}$ by construction, the optimal value of (1) provides a conservative estimate (lower bound) for $\mathbb{Q}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}]$. In order to certify the safety of the system with respect to the true distribution $\mathbb{Q}$, it is thus sufficient to show that the optimal value of (1) exceeds $1 - \epsilon$. Maybe surprisingly, the distributionally robust approach can also mitigate the intractability of evaluating high-dimensional integrals. Using the duality theory for moment problems in conjunction with the rich arsenal of modern robust optimization techniques, one can show that worst-case probability problems of the type (1) are computationally tractable across a wide variety of relevant ambiguity sets.

Rather than passively certifying the safety of a given system, a more ambitious goal would be to actively influence the system so as to maintain its safety. This scenario can conveniently be captured by a *robust chance constrained pro-*

gram, where $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^N$ represents the vector of design decisions.

$$
\begin{aligned}
\text{minimize} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{subject to} \quad & \boldsymbol{x} \in \mathcal{X} \\
& \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\boldsymbol{S}(\boldsymbol{x})\tilde{\boldsymbol{z}} \le \boldsymbol{t}(\boldsymbol{x})] \ge 1 - \epsilon
\end{aligned}
\tag{2}
$$

Here, the technology matrix $\boldsymbol{S}(\boldsymbol{x}) = (\boldsymbol{s}_1(\boldsymbol{x}), \ldots, \boldsymbol{s}_J(\boldsymbol{x}))^\top$ and the right-hand side vector $\boldsymbol{t}(\boldsymbol{x}) = (t_1(\boldsymbol{x}), \ldots, t_J(\boldsymbol{x}))^\top$ may depend on the design decisions in an affine fashion, that is, $\boldsymbol{s}_j(\boldsymbol{x}) = \boldsymbol{S}_j^\top \boldsymbol{x} + \boldsymbol{s}_j$ and $t_j(\boldsymbol{x}) = \boldsymbol{t}_j^\top \boldsymbol{x} + t_j$, where $\boldsymbol{S}_j \in \mathbb{R}^{N \times P}$, $\boldsymbol{s}_j \in \mathbb{R}^P$, $\boldsymbol{t}_j \in \mathbb{R}^N$ and $t_j \in \mathbb{R}$ for all $j \in \mathcal{J} = \{1, \ldots, J\}$. Moreover, $\boldsymbol{c} \in \mathbb{R}^N$ encodes the costs of different design decisions. We assume that problem (2) is tractable if it is stripped of the probabilistic constraint. This is the case, for instance, if $\mathcal{X}$ is a polytope defined by its facets or vertices.

The choice of the ambiguity set $\mathcal{P}$ should be guided by the following principles: *(i)* The set $\mathcal{P}$ must contain $\mathbb{Q}$ with certainty (or at least with high confidence). *(ii)* The structure of $\mathcal{P}$ should facilitate a tractable reformulation (or at least a tractable conservative approximation) of the uncertainty quantification problem (1) and the chance constrained program (2). *(iii)* Among all ambiguity sets satisfying the properties *(i)* and *(ii)*, $\mathcal{P}$ should be chosen as small as possible in the sense of set inclusion. Property *(i)* enables us to certify the safety of the system at hand under the unknown distribution $\mathbb{Q}$ by solving the uncertainty quantification problem (1). Moreover, property *(ii)* ensures that (1) and (2) can be solved efficiently, while property *(iii)* controls the conservatism of the uncertainty quantification problem (1), thereby limiting the risk that a safe system is not recognized as such.

After the fundamental papers [44,53], various ambiguity sets have been studied in the literature on uncertainty quantification [23,37] and distributionally robust optimization [1]. Ambiguity sets of special interest include the *Markov ambiguity set* containing all distributions with known mean and support [48], the *Chebyshev ambiguity set* containing all distributions with known bounds on the first and second-order moments [12,14,22,31,39,46,49,51,52], the *Gauss ambiguity set* containing all unimodal distributions from within the Chebyshev ambiguity set [38,41], various generalized Chebyshev ambiguity sets that specify *asymmetric moments* [12,13,35], *higher-order moments* [7, 30,45] or *marginal moments* [17,18], the *median-absolute deviation ambiguity set* containing all symmetric distributions with known median and mean absolute deviation [24], the *Huber ambiguity set* containing all distributions with known upper bound on the expected Huber loss function [15,48], the *Wasserstein ambiguity set* containing all distributions that are close to the empirical distribution with respect to the Wasserstein metric [19,34,40], the *Kullback-Leibler divergence ambiguity set* and *likelihood ratio ambiguity set* [10,26,27, 31,47] containing all distributions that are sufficiently likely to have generated a given data set, the *Hoeffding ambiguity set* containing all component-wise independent distributions with a box support [3,8,10], the *Bernstein ambiguity set* containing all distributions from within the Hoeffding ambiguity set subject to marginal moment bounds [36], several *φ-divergence-based ambiguity sets* [2,

50] containing all discrete distributions close to a given nominal distribution, *goodness-of-fit ambiguity sets* containing all distributions that pass prescribed statistical tests [6] etc. The proposed terminology associates most ambiguity sets with mathematicians who invented well-known probability inequalities or statistical indicators related to the respective ambiguity sets.

In this paper we endeavor to

- present a unifying framework for formulating and solving uncertainty quantification problems and robust chance constrained programs;
- demonstrate that many of the ambiguity sets listed above represent special cases of a canonical ambiguity set underlying our unifying framework;
- delineate the watershed between tractability and intractability in uncertainty quantification and robust chance constrained programming.

In [48] it has been shown that most moment-based ambiguity sets emerge as special cases of a canonical ambiguity set that contains all distributions under which the probabilities of some conic-representable confidence sets fall between prescribed upper and lower bounds, and the mean values of the uncertain parameters satisfy a linear equality constraint. While [48] describes methods for computing worst-case expectations of biconvex loss functions, the focus of the present paper is to compute worst-case probabilities, that is, worst-case expectations of discontinuous indicator functions. Moreover, while [48] focuses exclusively on moment-based ambiguity sets, the present paper investigates a much richer class of ambiguity sets characterized both in terms of moment constraints and structural information such as symmetry, unimodality, multimodality, independence patterns etc. In particular, we also show that several classical inequalities of probability theory as well as their multidimensional generalizations emerge as special cases of our unified framework.

The overarching objective of this work is to review broad classes of uncertainty quantification and chance constrained programming problems that are computationally tractable. An intimately related secondary objective is to explore the boundaries of tractability. It is thus natural to focus attention on *linear* safety constraints in the probabilistic expressions of (1) and (2). Indeed, the uncertainty quantification problem (1) becomes intractable already in the presence a single convex quadratic safety constraint, even if the underlying ambiguity set contains all distributions supported on a polytope.

Theorem 1 *Evaluating the quadratic uncertainty quantification problem*

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left[\|\boldsymbol{S}\tilde{\boldsymbol{z}}\|_2 \leq t\right] \tag{3}$$

is strongly $\mathcal{NP}$-hard even if the ambiguity set satisfies

$$\mathcal{P} = \left\{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) \,:\, \mathbb{P}\left[\boldsymbol{C}\tilde{\boldsymbol{z}} \leq \boldsymbol{d}\right] = 1\right\},$$

where $\mathcal{P}_0(\mathbb{R}^P)$ denotes the set of all probability distributions on $\mathbb{R}^P$.

The remainder of the paper develops as follows. Section 2 discusses the design of ambiguity sets using moment constraints and structural information. Sections 3 and 4 provide tractable reformulations and complexity results for various uncertainty quantification and chance constrained programming problems, respectively. An efficient approximation algorithm for intractable problems is reported in Section 5, and a summary of the main results is provided in Section 6. All proofs are relegated to an accompanying technical report [25].

Notation. A generalized inequality $\boldsymbol{x} \preccurlyeq_{\mathcal{K}} \boldsymbol{y}$ with respect to a proper (closed, convex, pointed, solid) cone $\mathcal{K}$ implies that $\boldsymbol{y} - \boldsymbol{x} \in \mathcal{K}$. We denote by $\mathbb{S}^P$ ($\mathbb{S}^P_+$) the space (cone) of all symmetric (positive semidefinite) matrices in $\mathbb{R}^{P \times P}$ and use $X \preccurlyeq Y$ as a notational shorthand for the matrix inequality $X \preccurlyeq_{\mathbb{S}^P_+} Y$ where $X, Y \in \mathbb{S}^P$. The cone dual to a proper cone $\mathcal{K}$ is denoted as $\mathcal{K}^\star$. We use $\mathcal{P}_0(B)$ to represent the set of all probability distributions supported on a Borel subset B of $\mathbb{R}^P$. If $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ represents the joint distribution of two random vectors $\tilde{\boldsymbol{z}} \in \mathbb{R}^P$ and $\tilde{\boldsymbol{u}} \in \mathbb{R}^Q$, then $\Pi_{\tilde{\boldsymbol{z}}}\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P)$ denotes the marginal distribution of $\tilde{\boldsymbol{z}}$ under $\mathbb{P}$. We extend this definition to ambiguity sets $\mathcal{P} \subseteq \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ by setting $\Pi_{\tilde{\boldsymbol{z}}}\mathcal{P} = \bigcup_{\mathbb{P} \in \mathcal{P}}\{\Pi_{\tilde{\boldsymbol{z}}}\mathbb{P}\}$. For two sets A and B the relation $A \Subset B$ indicates that A is a subset of the relative interior of B. For a logical expression $\mathcal{E}$, we define $\mathbb{I}_{[\mathcal{E}]} = 1$ if $\mathcal{E}$ is true; $= 0$ otherwise.

2 Ambiguity Sets

We first propose a canonical representation of ambiguity sets as intersections of *moment ambiguity sets* (characterizing features of $\mathbb{Q}$ such as the mean, variance or median-absolute deviation) and *structural ambiguity sets* (describing symmetry, unimodality or independence properties etc.). We define the moment and structural ambiguity sets of interest in Sections 2.1 and 2.2, respectively, and we showcase the expressiveness of our framework in Section 2.3.

2.1 Nested Moment Ambiguity Sets

As in [48], we focus on *nested moment ambiguity sets* of the form

$$\mathcal{P}^n = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) \ : \ \begin{array}{l} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}}\right] = \boldsymbol{b}, \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}_i\right] \in \left[\underline{p}_i, \overline{p}_i\right] \ \ \forall i \in \mathcal{I} \end{array} \right\}, \quad (4)$$

where $\mathbb{P}$ is a joint distribution of the random vector $\tilde{\boldsymbol{z}} \in \mathbb{R}^P$ appearing in the uncertainty quantification and chance constrained programming problems and some auxiliary random vector $\tilde{\boldsymbol{u}} \in \mathbb{R}^Q$. We assume that $\boldsymbol{A} \in \mathbb{R}^{K \times P}$, $\boldsymbol{B} \in \mathbb{R}^{K \times Q}$, $\boldsymbol{b} \in \mathbb{R}^K$ and $\mathcal{I} = \{1, \ldots, I\}$, while the confidence sets $\mathcal{C}_i$ satisfy

$$\mathcal{C}_i = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^P \times \mathbb{R}^Q \ : \ \boldsymbol{C}_i\boldsymbol{z} + \boldsymbol{D}_i\boldsymbol{u} \preccurlyeq_{\mathcal{K}_i} \boldsymbol{d}_i \right\} \quad (5)$$

with $\boldsymbol{C}_i \in \mathbb{R}^{L_i \times P}$, $\boldsymbol{D}_i \in \mathbb{R}^{L_i \times Q}$, $\boldsymbol{d}_i \in \mathbb{R}^{L_i}$ and $\mathcal{K}_i$ being proper cones. Note that the inclusion of the auxiliary vector $\tilde{\boldsymbol{u}}$ in $\mathcal{P}^n$ seems redundant as $\tilde{\boldsymbol{u}}$ could be absorbed in $\tilde{\boldsymbol{z}}$. In the next section, however, we will impose structural requirements on the marginal distribution of $\tilde{\boldsymbol{z}}$ that do not affect $\tilde{\boldsymbol{u}}$. We allow K or Q to be zero, in which case the expectation condition in (4) is void or the random vector $\tilde{\boldsymbol{u}}$ is absent, respectively. We also assume that $\mathcal{C}_i$ is essentially strictly feasible,[1] $\underline{p}_i, \overline{p}_i \in [0, 1]$ and $\underline{p}_i \leq \overline{p}_i$ for all $i \in \mathcal{I}$. Nested moment ambiguity sets of the form (4) have been used in [48] to evaluate worst-case expectations of convex functions. Here, we use them to compute worst-case probabilities of polytopes. As we will see shortly, allowing for multiple confidence sets $\mathcal{C}_i$ enables us to model a rich variety of ambiguity sets. By designing the confidence sets $\mathcal{C}_i$ appropriately, for example, we can prescribe the modality structure of $\tilde{\boldsymbol{z}}$, or we can specify confidence regions for the moments of $\tilde{\boldsymbol{z}}$ if these moments are estimated from historical samples, see [48].

In the remainder we impose the following regularity conditions.

(**B**) The confidence set $\mathcal{C}_I$ has probability one, that is, $\underline{p}_I = \overline{p}_I = 1$, and all other confidence sets $\mathcal{C}_i$, $i = 1, \ldots, I - 1$, are bounded subsets of $\mathcal{C}_I$.

(**N**) For all $i, i' \in \mathcal{I}$, $i \neq i'$, we have either $\mathcal{C}_i \Subset \mathcal{C}_{i'}$, $\mathcal{C}_{i'} \Subset \mathcal{C}_i$ or $\mathcal{C}_i \cap \mathcal{C}_{i'} = \emptyset$.

The boundedness condition (**B**) implies that the confidence set $\mathcal{C}_I$ contains the support of the joint random vector $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})$. This does not restrict generality since we are free to choose $\mathcal{C}_I = \mathbb{R}^P \times \mathbb{R}^Q$, but it simplifies some model formulations in later sections. Condition (**B**) also stipulates that the confidence sets $\mathcal{C}_i$, $i = 1, \ldots, I - 1$, are bounded. This is necessary to obtain tractable formulations for the uncertainty quantification and chance constrained programming problems. The nesting condition (**N**) imposes a strict partial order on the confidence sets $\mathcal{C}_i$ with respect to the $\Subset$-relation, and it also requires that incomparable sets are disjoint. We remark that for two sets $\mathcal{C}_i$ and $\mathcal{C}_{i'}$, the relation $\mathcal{C}_i \Subset \mathcal{C}_{i'}$ can be verified efficiently in many cases, for instance if both sets are polyhedral. All examples studied in this paper satisfy the nesting condition by construction. The importance of the nesting condition is highlighted by the following result, which is proven in [48, Theorem 2].

Theorem 2 *Verifying whether the nested moment ambiguity set $\mathcal{P}^n$ defined in* (4) *is empty is strongly $\mathcal{NP}$-hard even if $\mathcal{P}^n$ does not involve any expectation conditions (i.e., $K = 0$) and there are only two bounded (second-order) conic representable confidence sets $\mathcal{C}_1$, $\mathcal{C}_2$ with $\mathcal{C}_1 \subseteq \mathcal{C}_2$ but $\mathcal{C}_1 \not\Subset \mathcal{C}_2$.*

Theorem 2 implies that if the nesting condition (**N**) is violated, then the uncertainty quantification and chance constrained programming problems are strongly $\mathcal{NP}$-hard. Some results in this paper require that in addition to (**N**), any set $\mathcal{C}_i$ that contains another set $\mathcal{C}_{i'}$, $\mathcal{C}_{i'} \Subset \mathcal{C}_i$, must have an affine dimension of at least 2. This dimensionality condition will be satisfied by all examples in this paper, and it certainly holds for most applications of practical interest.

---

[1] We call $\mathcal{C}_i$ *essentially strictly feasible* if there is $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i$ that satisfies all non-polyhedral constraints in (5) strictly, see [4].

2.2 Structural Ambiguity Sets

As highlighted in [48], nested moment ambiguity sets of the form (4) allow us to model an abundance of (generalized) moment conditions. However, they fail to capture commonly encountered structural properties of the marginal distribution of $\tilde{z}$. In the remainder of the paper, we will thus intersect the nested moment ambiguity set (4) with various structural ambiguity sets that capture features such as symmetry, unimodality or independence. We require all structural ambiguity sets to be convex and weakly closed. For ease of exposition, we temporarily set $Q = 0$, thus assuming that there are no auxiliary random variables $\tilde{u}$, but we will revoke this restriction at the end of this section.

In this work we focus on structural ambiguity sets $\mathcal{P} \subseteq \mathcal{P}_0(\mathbb{R}^P)$ that possess a *Choquet representation*, whereby every distribution $\mathbb{P} \in \mathcal{P}$ can be written as a mixture (i.e., an infinite convex combination) of extremal distributions of $\mathcal{P}$. Thus, for every Borel set $B \in \mathcal{B}(\mathbb{R}^P)$ we require that

$$\mathbb{P}[B] = \int_{\mathcal{V}} \mathbb{V}_{\boldsymbol{v}}(B) \, \mathbb{M}(\mathrm{d}\boldsymbol{v}), \tag{6}$$

where $\mathbb{V}_{\boldsymbol{v}}$, $\boldsymbol{v} \in \mathcal{V} \subseteq \mathbb{R}^V$, represents the family of extremal distributions (extreme points) of $\mathcal{P}$, and $\mathbb{M} \in \mathcal{P}_0(\mathcal{V})$ is the mixture distribution generating $\mathbb{P}$. This implies that the family of extreme points admits a finite-dimensional parameterization. In distributionally robust optimization such structural ambiguity sets were first studied in [41]. We show next that this abstract framework covers several practically relevant classes of structural ambiguity sets.

Symmetry Let $\mathcal{P}^s$ be the set of all point symmetric distributions on $\mathbb{R}^P$ with center $\boldsymbol{m}$. Thus, $\mathbb{P} \in \mathcal{P}^s$ if and only if $\mathbb{P}[B] = \mathbb{P}[2\boldsymbol{m} - B]$ for all Borel sets $B \in \mathcal{B}(\mathbb{R}^P)$. The extremal distributions of $\mathcal{P}^s$ are

$$\mathbb{V}_{\boldsymbol{v}} = \frac{1}{2}\delta_{\boldsymbol{v}} + \frac{1}{2}\delta_{2\boldsymbol{m}-\boldsymbol{v}} \qquad \text{for } \boldsymbol{v} \in \mathbb{R}^P,$$

where $\delta_{\boldsymbol{v}}$ and $\delta_{2\boldsymbol{m}-\boldsymbol{v}}$ denote the Dirac distributions that place all probability mass on the points $\boldsymbol{v}$ and $2\boldsymbol{m} - \boldsymbol{v}$, respectively.

Unimodality A distribution $\mathbb{P}$ is called unimodal with center $\boldsymbol{m}$ if $\mathbb{P}[t(B - \boldsymbol{m})]/t^P$ is non-increasing in $t > 0$ for all $B \in \mathcal{B}(\mathbb{R}^P)$, see e.g. [16]. The definition implies that if $\mathbb{P}$ has a continuous density function $\rho(\boldsymbol{z})$, then $\mathbb{P}$ is unimodal if and only if $\rho(t(\boldsymbol{z} - \boldsymbol{m}))$ is non-increasing in $t > 0$ for any $\boldsymbol{z} \in \mathbb{R}^P$. The extreme points of the set of unimodal distributions are the radial distributions $\mathbb{V}_{\boldsymbol{v}}$, $\boldsymbol{v} \in \mathbb{R}^P$, that are supported on line segments from $\boldsymbol{m}$ to $\boldsymbol{v}$ and that satisfy

$$\mathbb{V}_{\boldsymbol{v}}\left([\boldsymbol{m}, \boldsymbol{m} + t(\boldsymbol{v} - \boldsymbol{m})]\right) = t^P \qquad \forall t \in [0, 1].$$

α-Unimodality For $\alpha > 0$, let $\mathcal{P}^\alpha$ be the set of α-unimodal distributions with center $\boldsymbol{m}$, that is, $\mathbb{P} \in \mathcal{P}^\alpha$ if and only if $\mathbb{P}[t(B - \boldsymbol{m})]/t^\alpha$ is non-increasing in $t > 0$ for all $B \in \mathcal{B}(\mathbb{R}^P)$, see [16]. Note that an α-unimodal distribution on $\mathbb{R}^P$ is unimodal in the usual sense if $\alpha = P$. Moreover, if an α-unimodal distribution has a continuous density function $\rho(\boldsymbol{z})$, then $t^{\alpha-P}\rho(t(\boldsymbol{z} - \boldsymbol{m}))$ is non-increasing in $t > 0$ for any $\boldsymbol{z} \in \mathbb{R}^P$. The extreme points of $\mathcal{P}^\alpha$ are the radial distributions $\mathbb{V}_{\boldsymbol{v}}$, $\boldsymbol{v} \in \mathbb{R}^P$, that are supported on line segments $[\boldsymbol{m}, \boldsymbol{v}]$ and that satisfy

$$\mathbb{V}_{\boldsymbol{v}}\left([\boldsymbol{m}, \boldsymbol{m} + t(\boldsymbol{v} - \boldsymbol{m})]\right) = t^\alpha \qquad \forall t \in [0, 1].$$

Multimodality Consider the ambiguity set generated by the convex combination $\sum_{r=1}^R \lambda_r \mathcal{P}^\alpha(\boldsymbol{m}_r)$ of α-unimodal ambiguity sets $\mathcal{P}^\alpha(\boldsymbol{m}_r)$ with centers $\boldsymbol{m}_r$, $r = 1, \ldots, R$ and $\sum_{r=1}^R \lambda_r = 1$. This ambiguity set contains all α-*multi*modal distributions with a mode of probability mass $\lambda_r \geq 0$ centered at $\boldsymbol{m}_r$ for each $r = 1, \ldots, R$. The extremal distributions of this ambiguity set are representable as $\mathbb{V}_{\boldsymbol{v}} = \sum_{r=1}^R \lambda_r \mathbb{V}_{r,\boldsymbol{v}_r}$ for $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_R) \in \mathbb{R}^{RP}$, where $\mathbb{V}_{r,\boldsymbol{v}_r}$ for $\boldsymbol{v}_r \in \mathbb{R}^P$ is any extremal distribution of the set $\mathcal{P}^\alpha(\boldsymbol{m}_r)$, $r = 1, \ldots, R$.

Independence Let $\mathcal{P}^i$ be the set of all distributions on $\mathbb{R}^P$ under which the components of the random vector $\tilde{\boldsymbol{z}}$ are independent. One readily verifies that $\mathcal{P}^i$ violates our convexity assumption. Indeed, we have $\delta_{\boldsymbol{0}}, \delta_{\mathbf{e}} \in \mathcal{P}^i$, but the components of $\tilde{\boldsymbol{z}}$ are perfectly correlated under the mixture distribution $\frac{1}{2}\delta_{\boldsymbol{0}} + \frac{1}{2}\delta_{\mathbf{e}}$. We will show in Section 3 that the uncertainty quantification and chance constrained programming problems are typically intractable for ambiguity sets that impose independence among the components of $\tilde{\boldsymbol{z}}$.

Remark 1 In the presence of auxiliary random variables $\tilde{\boldsymbol{u}} \in \mathbb{R}^Q$ with $Q > 0$, the above structural ambiguity sets are redefined as the families of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ whose marginal projections $\Pi_{\tilde{\boldsymbol{z}}}\mathbb{P}$ display the structural properties (e.g., symmetry or α-unimodality etc.) outlined above.

2.3 Examples

The ambiguity sets that can be generated by intersecting a nested moment ambiguity set of the form (4) with a Choquet-representable structural ambiguity set display a remarkable diversity. We now show that many ambiguity sets from the recent literature can be expressed as instances of this class.

Example 1 (Chebyshev Ambiguity Set) Let $\mathcal{P}$ be the ambiguity set of all distributions on $\mathbb{R}^P$ with mean $\boldsymbol{\mu} \in \mathbb{R}^P$ whose covariance matrix is bounded above by $\boldsymbol{\Sigma} \in \mathbb{S}_+^P$, that is,

$$\mathcal{P} = \left\{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu}, \ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})^\top\right] \preccurlyeq \boldsymbol{\Sigma}\right\}. \qquad (7)$$

Consider the following instance of the nested moment ambiguity set (4), which involves the auxiliary random matrix $\tilde{\boldsymbol{U}} \in \mathbb{R}^{P \times P}$.

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^{P \times P}) : \begin{array}{c} \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}, \;\; \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{U}}\right] = \boldsymbol{\Sigma}, \\ \mathbb{P}\left[\begin{pmatrix} 1 & (\tilde{\boldsymbol{z}} - \boldsymbol{\mu})^{\top} \\ (\tilde{\boldsymbol{z}} - \boldsymbol{\mu}) & \tilde{\boldsymbol{U}} \end{pmatrix} \succcurlyeq \boldsymbol{0} \right] = 1 \end{array} \right\} \tag{8}$$

Since $\mathcal{P}'$ only contains one confidence set with probability bounds $\underline{p}_1 = \overline{p}_1 = 1$, the boundedness condition (**B**) and the nesting condition (**N**) are trivially satisfied. Moreover, a Schur complement argument implies that $\Pi_{\tilde{\boldsymbol{z}}} \mathcal{P}' = \mathcal{P}$. The Chebyshev ambiguity set has been studied extensively in distributionally robust optimization [14, 22, 48, 51] and can also be generalized to account for uncertainty in the mean and/or covariance matrix of $\tilde{\boldsymbol{z}}$, see e.g. [14, 48].

If we intersect the nested moment ambiguity set from Example 1 with the structural ambiguity set of all unimodal distributions, then we recover an ambiguity set that has been studied in [38] and is closely related to a tightened Chebyshev-type inequality due to Gauss [21].

Example 2 (Gauss Ambiguity Set) Let $\mathcal{P}$ be the ambiguity set of all unimodal distributions on $\mathbb{R}^P$ with center $\boldsymbol{m} \in \mathbb{R}^P$ and mean $\boldsymbol{\mu} \in \mathbb{R}^P$ whose covariance matrix is bounded above by $\boldsymbol{\Sigma} \in \mathbb{S}_+^P$, that is,

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{array}{c} \mathbb{P} \text{ is unimodal with center } \boldsymbol{m}, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}, \;\; \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})^{\top}\right] \preccurlyeq \boldsymbol{\Sigma} \end{array} \right\}. \tag{9}$$

In analogy to the previous example, the ambiguity set

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^{P \times P}) : \begin{array}{c} \Pi_{\tilde{\boldsymbol{z}}} \mathbb{P} \text{ is unimodal with center } \boldsymbol{m}, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}, \;\; \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{U}}\right] = \boldsymbol{\Sigma}, \\ \mathbb{P}\left[\begin{pmatrix} 1 & (\tilde{\boldsymbol{z}} - \boldsymbol{\mu})^{\top} \\ (\tilde{\boldsymbol{z}} - \boldsymbol{\mu}) & \tilde{\boldsymbol{U}} \end{pmatrix} \succcurlyeq \boldsymbol{0} \right] = 1 \end{array} \right\} \tag{10}$$

satisfies (**B**) and (**N**) as well as $\Pi_{\tilde{\boldsymbol{z}}} \mathcal{P}' = \mathcal{P}$.

Examples 1 and 2 rely on classical statistical indicators—the mean and variance—to characterize the unknown distribution $\mathbb{Q}$. In the next two examples, we describe $\mathbb{Q}$ through location and dispersion measures from robust statistics, namely the median, the median-absolute deviation and the Huber loss function. While reminiscent of the mean and variance, these indicators may be more reliable as they are easier to estimate from data, see [11, 48].

Example 3 (Median-Absolute Deviation Ambiguity Set) Let $\mathcal{P}$ be the ambiguity set of all symmetric distributions on $\mathbb{R}^P$ with center $\boldsymbol{m} \in \mathbb{R}^P$ whose median absolute deviation is bounded above by $\boldsymbol{f} \in \mathbb{R}_+^P$, that is,

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{array}{c} \mathbb{P} \text{ is symmetric with center } \boldsymbol{m}, \\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{z}} - \boldsymbol{m}|] \leq \boldsymbol{f} \end{array} \right\}, \tag{11}$$

where the absolute value is understood to apply component-wise. Note that for symmetric distributions, the median coincides with the center. Consider the following instance of (4), which involves the auxiliary random vector $\tilde{\boldsymbol{u}} \in \mathbb{R}^P$.

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) : \begin{array}{l} \Pi_{\tilde{\boldsymbol{z}}}\mathbb{P} \text{ is symmetric with center } \boldsymbol{m}, \\ \mathbb{P}\left[\tilde{\boldsymbol{u}} \geq \tilde{\boldsymbol{z}} - \boldsymbol{m}, \ \tilde{\boldsymbol{u}} \geq \boldsymbol{m} - \tilde{\boldsymbol{z}}\right] = 1, \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{u}}\right] = \boldsymbol{f} \end{array} \right\} \quad (12)$$

Since $\mathcal{P}'$ involves only a single confidence set with probability bounds $\underline{p}_1 = \overline{p}_1 = 1$, the boundedness condition ($\mathbf{B}$) and the nesting condition ($\mathbf{N}$) are trivially satisfied. Moreover, one readily verifies that $\Pi_{\tilde{\boldsymbol{z}}}\mathcal{P}' = \mathcal{P}$.

Example 4 (Huber Ambiguity Set) Consider the ambiguity set

$$\mathcal{P} = \left\{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}, \ \mathbb{E}_{\mathbb{P}}\left[H_\beta(\boldsymbol{f}^\top[\tilde{\boldsymbol{z}} - \boldsymbol{\mu}])\right] \leq g\right\},$$

where $\boldsymbol{\mu}, \boldsymbol{f} \in \mathbb{R}^P$, $g \in \mathbb{R}_+$ and $H_\beta(z)$ is the Huber loss function with prescribed robustness parameter $\beta > 0$, that is, $H_\beta(z) = \frac{1}{2}z^2$ if $|z| \leq \beta$; $= \beta\left(|z| - \frac{1}{2}\beta\right)$ otherwise, see [28]. The expected Huber loss function represents a robust dispersion measure that generalizes the variance (for $\beta \to \infty$) and the mean absolute deviation (for $\beta \to 0$). Consider now the following instance of (4), which involves the auxiliary random variables $\tilde{u}, \tilde{v}, \tilde{w} \in \mathbb{R}_+$.

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}_+^3) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}, \ \mathbb{E}_{\mathbb{P}}\left[\tilde{w}\right] = g, \\ \mathbb{P}\left[\frac{1}{2}\left(\boldsymbol{f}^\top\tilde{\boldsymbol{z}} + \tilde{u} - \tilde{v}\right)^2 + \beta(\tilde{u} + \tilde{v}) \leq \tilde{w}\right] = 1 \end{array} \right\}$$
$$(13)$$

It has been shown in [48] that $\mathcal{P}'$ satisfies $\Pi_{\tilde{\boldsymbol{z}}}\mathcal{P}' = \mathcal{P}$. Since $\mathcal{P}'$ only involves a single confidence set with probability bounds $\underline{p}_1 = \overline{p}_1 = 1$, the boundedness condition ($\mathbf{B}$) and the nesting condition ($\mathbf{N}$) are trivially satisfied.

We close this section by reviewing a data-driven ambiguity set that can be constructed directly from independent samples of the unknown distribution $\mathbb{Q}$. In contrast to the previous examples, this ambiguity set converges to the singleton $\{\mathbb{Q}\}$ as the number of available samples tends to infinity. For a detailed discussion of this ambiguity set we refer to [34].

Example 5 (Wasserstein Ambiguity Set) Suppose that we have observed independent samples $\hat{\boldsymbol{z}}_i$, $i \in \mathcal{I} = \{1, \ldots, I\}$, of an unknown data-generating distribution $\mathbb{Q}$ with bounded support, and assume for ease of exposition that $\hat{\boldsymbol{z}}_i \neq \hat{\boldsymbol{z}}_j$ for $i, j \in \mathcal{I}$, $i \neq j$. If we denote by $\hat{\mathbb{P}} = \frac{1}{I}\sum_{i=1}^I \delta_{\hat{\boldsymbol{z}}_i}$ the empirical distribution, then the Wasserstein ambiguity set of size $r > 0$ is defined as

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : d_p^{\mathrm{w}}(\mathbb{P}, \hat{\mathbb{P}}) \leq r\} \quad (14)$$

and thus contains all distributions that reside within the ball of radius r around $\hat{\mathbb{P}}$ with respect to the Wasserstein metric of order $p \geq 1$. For any (marginal)

distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ of two P-dimensional random vectors $\tilde{\boldsymbol{z}}_1$ and $\tilde{\boldsymbol{z}}_2$, respectively, the Wasserstein distance $d_p^{\mathrm{w}}(\mathbb{P}_1, \mathbb{P}_2)$ of order p is defined as

$$
\begin{aligned}
d_p^{\mathrm{w}}(\mathbb{P}_1, \mathbb{P}_2) = \inf \quad & \mathbb{E}_{\mathbb{P}}[\|\tilde{\boldsymbol{z}}_1 - \tilde{\boldsymbol{z}}_2\|_p] \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) \\
& \left. \begin{array}{l} \mathbb{P}[\tilde{\boldsymbol{z}}_1 \in B] = \mathbb{P}_1[\tilde{\boldsymbol{z}}_1 \in B] \\ \mathbb{P}[\tilde{\boldsymbol{z}}_2 \in B] = \mathbb{P}_2[\tilde{\boldsymbol{z}}_2 \in B] \end{array} \right\} \quad \forall B \in \mathcal{B}(\mathbb{R}^P),
\end{aligned} \tag{15}
$$

see e.g. [42]. Note that (15) can be viewed as a transportation problem that minimizes the expectation of $\|\tilde{\boldsymbol{z}}_1 - \tilde{\boldsymbol{z}}_2\|_p$ over all possible joint distributions $\mathbb{P}$ of $\tilde{\boldsymbol{z}}_1$ and $\tilde{\boldsymbol{z}}_2$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively. If $\overline{u} \in \mathbb{R}_+$ denotes an upper bound on the Euclidean diameter of the support of $\mathbb{Q}$, then the ambiguity set

$$
\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P \times \mathbb{R}) : \begin{array}{c} \mathbb{E}_{\mathbb{P}}[\tilde{u}_0] = r, \;\; \mathbb{P}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{u}_0) \in \mathbb{R}^{2P+1}] = 1, \\ \mathbb{P}\left[\begin{array}{l} \|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{u}}\|_p \leq \tilde{u}_0, \\ \tilde{u}_0 \leq \overline{u}, \;\; \tilde{\boldsymbol{u}} = \hat{\boldsymbol{z}}_i \end{array} \right] = 1/I \;\; \forall i = 1, \ldots, I \end{array} \right\} \tag{16}
$$

constitutes an instance of (4) that satisfies the boundedness condition (**B**) and the nesting condition (**N**). Moreover, we have $\Pi_{\tilde{\boldsymbol{u}}} \mathcal{P}' = \{\hat{\mathbb{P}}\}$ and $\Pi_{\tilde{\boldsymbol{z}}} \mathcal{P}' = \mathcal{P}$.

3 Uncertainty Quantification

In this section we develop tractable reformulations for instances of the uncertainty quantification problem (1) under the premise that $\mathcal{P}$ constitutes a nested ambiguity set of the form (4) or an intersection of (4) with the set of symmetric or α-unimodal distributions. We also show that the uncertainty quantification problem is generically intractable for ambiguity sets that impose independence among the components of $\tilde{\boldsymbol{z}}$. The results of this section generalize the results of [24] to instances of (4) with $I > 1$ confidence sets.

Our tractability results rely on an interpretation of the uncertainty quantification problem (1) as a generalized moment problem of the type [43, Equation (3.2)] whose semi-infinite dual [43, Equation (3.4)] lends itself for further simplification. Strong duality holds under the Slater condition [43, Equation (3.12)], which we henceforth abbreviate as (**S**). For nonempty ambiguity sets, this condition is non-restrictive and can always be enforced by slightly perturbing the parameters $\boldsymbol{b}$, $\underline{p}_i$ and $\overline{p}_i$, but it is cumbersome to state and verify explicitly for the generic nested ambiguity set (4). Many examples considered in this paper involve only a single confidence set ($I = 1$) and no structural information, in which case the Slater condition (**S**) simplifies to the requirement that $\boldsymbol{b}$ belongs to the interior of the convex set $\{\boldsymbol{Az} + \boldsymbol{Bu} \; : \; (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1\}$.

We are now in the position to state our first tractability result.

Theorem 3 (Moment Ambiguity Sets) *If $\mathcal{P} = \mathcal{P}^n$ is an instance of (4) that satisfies the boundedness condition (**B**), the nesting condition (**N**) and*

the Slater condition (S), then the worst-case probability (1) *coincides with the optimal value of the conic optimization problem*

$$
\begin{aligned}
\sup \quad & \boldsymbol{b}^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} \underline{p}_i \lambda_i - \overline{p}_i \kappa_i \\
\text{s.t.} \quad & \boldsymbol{\gamma} \in \mathbb{R}^K, \ \boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}_+^I, \ \boldsymbol{\phi}_i \in \mathcal{K}_i^\star, \, i \in \mathcal{I} \\
& \tau_{ij} \in \mathbb{R}_+, \, (i,j) \in \mathcal{L}, \ \boldsymbol{\psi}_{ij} \in \mathcal{K}_i^\star, \, (i,j) \in \mathcal{L} \\
& \left. \begin{aligned} & \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) + \boldsymbol{d}_i^\top \boldsymbol{\phi}_i \leq 1 \\ & \boldsymbol{A}^\top \boldsymbol{\gamma} = \boldsymbol{C}_i^\top \boldsymbol{\phi}_i, \ \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_i^\top \boldsymbol{\phi}_i \end{aligned} \right\} \ \forall i \in \mathcal{I} \\
& \left. \begin{aligned} & \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) + \boldsymbol{d}_i^\top \boldsymbol{\psi}_{ij} - \tau_{ij} t_j \leq 0 \\ & \boldsymbol{A}^\top \boldsymbol{\gamma} + \tau_{ij} \boldsymbol{s}_j = \boldsymbol{C}_i^\top \boldsymbol{\psi}_{ij}, \ \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_i^\top \boldsymbol{\psi}_{ij} \end{aligned} \right\} \ \forall (i,j) \in \mathcal{L},
\end{aligned} \tag{17}
$$

where $\mathcal{A}(i) = \{i\} \cup \{i' \in \mathcal{I} : \mathcal{C}_i \Subset \mathcal{C}_{i'}\}$ represents the index set of all supersets (antecedents) of $\mathcal{C}_i$, while $\mathcal{L} = \{(i,j) \in \mathcal{I} \times \mathcal{J} : \exists (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i . \boldsymbol{s}_j^\top \boldsymbol{z} > t_j\}$.

The index set $\mathcal{L}$ contains the pair $(i,j) \in \mathcal{I} \times \mathcal{J}$ if there are realizations $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i$ that violate the j-th constraint in (1). One can verify efficiently whether $(i,j) \in \mathcal{L}$ by checking whether the optimal value of the convex optimization problem $\sup \{\boldsymbol{s}_j^\top \boldsymbol{z} : (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i\}$ is strictly larger than t_j.

Problem (17) is a conic optimization problem whose size scales polynomially in the size of $\boldsymbol{S}$ and $\boldsymbol{t}$ in problem (1) as well as the description of the ambiguity set $\mathcal{P}^n$ in (4). Moreover, if all cones $\mathcal{K}_i$ are polyhedral, then (17) is a linear program. We illustrate Theorem 3 with two examples.

Example 6 (Generalized Chebyshev Bounds) Theorem 3 allows us to compute the worst-case probability of the event $\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}$ if the distribution of $\tilde{\boldsymbol{z}}$ is only known to belong to the Chebyshev ambiguity set (7) from Example 1. Thereby, we recover a generalized multivariate Chebyshev inequality that was discovered in [46]. As $\mathcal{P} = \Pi_{\tilde{\boldsymbol{z}}} \mathcal{P}'$, where $\mathcal{P}'$ is defined in (8), we have $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}] = \inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}]$. For $\boldsymbol{\Sigma} \succ \boldsymbol{0}$, $\mathcal{P}'$ is an instance of (4) that satisfies the conditions (**B**), (**S**) and (**N**). Thus, we can use Theorem 3 to reformulate the uncertainty quantification problem as the semidefinite program

$$
\begin{aligned}
\sup \quad & \beta - \boldsymbol{\mu}^\top \boldsymbol{\gamma} - \langle \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top, \boldsymbol{\Gamma} \rangle \\
\text{s.t.} \quad & \beta \in \mathbb{R}, \ \boldsymbol{\gamma} \in \mathbb{R}^P, \ \boldsymbol{\Gamma} \in \mathbb{S}_+^P, \ \tau_j \in \mathbb{R}_+, \, j \in \mathcal{L} \\
& \begin{pmatrix} 1 - \beta & \frac{1}{2}\boldsymbol{\gamma}^\top \\ \frac{1}{2}\boldsymbol{\gamma} & \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \boldsymbol{0}, \ \begin{pmatrix} \tau_j t_j - \beta & \frac{1}{2}(\boldsymbol{\gamma} + \tau_j \boldsymbol{s}_j)^\top \\ \frac{1}{2}(\boldsymbol{\gamma} + \tau_j \boldsymbol{s}_j) & \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \boldsymbol{0} \quad \forall j \in \mathcal{L},
\end{aligned} \tag{18}
$$

where $\mathcal{L} = \{j \in \mathcal{J} : \boldsymbol{s}_j \neq \boldsymbol{0} \vee t_j < 0\}$.

Example 7 (Data-Driven Uncertainty Quantification) Theorem 3 further allows us to compute the worst-case probability of the event $\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}$ if the distribution of $\tilde{\boldsymbol{z}}$ is an element of the Wasserstein ambiguity set (14) from Example 5. Thereby we recover a data-driven probability inequality first discovered in [34]. As $\mathcal{P} = \Pi_{\tilde{\boldsymbol{z}}} \mathcal{P}'$, where $\mathcal{P}'$ is defined in (16), we have $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}] =$

$\inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}]$. Moreover, $\mathcal{P}'$ is an instance of (4) that satisfies the conditions (**B**) and (**N**). Even though $\mathcal{P}'$ fails to satisfy the Slater condition (**S**), one can show that Theorem 3 remains valid [34]. Thus, the uncertainty quantification problem (1) can be reformulated as the convex optimization problem

$$
\begin{aligned}
\sup \quad & \frac{1}{I} \sum_{i=1}^{I} \beta_i - \gamma r \\
\text{s.t.} \quad & \boldsymbol{\beta} \in \mathbb{R}^I, \ \gamma \in \mathbb{R}_+, \ \boldsymbol{\tau} \in \mathbb{R}_+^{I \times J} \\
& \left. \begin{aligned} \beta_i &\leq 1 \\ \|\tau_{ij}\boldsymbol{s}_j\|_q &\leq \gamma \\ \beta_i + \tau_{ij}\boldsymbol{s}_j^\top\hat{\boldsymbol{z}}_i &\leq \tau_{ij}t_j \end{aligned} \right\} \forall j \in \mathcal{J} \right\} \quad \forall i = 1,\dots,I,
\end{aligned}
\tag{19}
$$

where q is defined through $\frac{1}{p} + \frac{1}{q} = 1$. Note that as $r \to 0$, the variable γ can be driven to infinity at essentially no cost. Thus, the optimal value of (19) converges to the fraction of the samples $\hat{\boldsymbol{z}}_i$, $i = 1,\dots,I$, that satisfy $\boldsymbol{S}\hat{\boldsymbol{z}}_i < \boldsymbol{t}$. Problem (19) reduces to a linear program for $p \in \{1,\infty\}$.

We now consider instances of the uncertainty quantification problem (1) where $\mathcal{P}$ emerges from the intersection of a nested moment ambiguity set $\mathcal{P}^n$ of the form (4) with the set of all symmetric distributions $\mathcal{P}^s$ centered around $\boldsymbol{m}$. In order to derive a tractable reformulation for this problem class, we require that $\mathcal{P}^n$ satisfies the following technical dimensionality condition:

(**D**) The ambiguity set satisfies $I = 1$ (support only), or it satisfies $Q > 0$ (the vector $\tilde{\boldsymbol{u}}$ is not absent) and all confidence sets $\mathcal{C}_i$, $i \in \mathcal{I}$, are bounded.

Moreover, for the multi-indices $\mathbf{i} = (i^+, i^-) \in \mathcal{I}^2$ ranging over pairs of confidence sets and $\mathbf{j} = (j^+, j^-) \in (\mathcal{J} \cup \{0\})^2$ indexing pairs of inequalities in the uncertainty quantification problem, we define the set

$$
\mathcal{D}_{\mathbf{ij}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q \ : \ \begin{aligned} &(\boldsymbol{z}, \boldsymbol{u}^+) \in \mathcal{C}_{i^+}, \ (-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_{i^-} \\ &j^+ > 0 \implies \boldsymbol{s}_{j^+}^\top \boldsymbol{z} > t_{j^+} \\ &j^- > 0 \implies -\boldsymbol{s}_{j^-}^\top \boldsymbol{z} > t_{j^-} \end{aligned} \right\},
\tag{20}
$$

and we impose the following feasibility condition.

(**F**) For any $\mathbf{i} = (i^+, i^-) \in \mathcal{I}^2$ and $\mathbf{j} = (j^+, j^-) \in (\mathcal{J} \cup \{0\})^2$, if $\mathcal{D}_{\mathbf{ij}} \neq \emptyset$, then $\mathcal{D}_{\mathbf{ij}}$ is essentially strictly feasible.

Condition (**F**) is a mild technical condition that is satisfied, for example, if the cones $\mathcal{K}_i$, $i \in \mathcal{I}$, are polyhedral.

Theorem 4 (Symmetry) *Assume that $\mathcal{P} = \mathcal{P}^n \cap \mathcal{P}^s$, where $\mathcal{P}^n$ is an instance of the nested moment ambiguity set (4) that satisfies the boundedness condition (**B**), the nesting condition (**N**), the dimensionality condition (**D**) and the feasibility condition (**F**), whereas $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ under which $\Pi_{\tilde{\boldsymbol{z}}}\mathbb{P}$ is point symmetric around $\boldsymbol{m} = \boldsymbol{0}$. If the Slater condition*

(S) holds, then the worst-case probability (1) *coincides with the optimal value of the conic program*

$$
\begin{aligned}
\sup \quad & \boldsymbol{b}^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} \underline{p}_i \lambda_i - \bar{p}_i \kappa_i \\
\text{s.t.} \quad & \boldsymbol{\gamma} \in \mathbb{R}^K, \ \boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}_+^I \\
& \boldsymbol{\psi}_{\mathbf{ij}}^+ \in \mathcal{K}_{i^+}^\star, \ \boldsymbol{\psi}_{\mathbf{ij}}^- \in \mathcal{K}_{i^-}^\star, \ \chi_{\mathbf{ij}}^+, \chi_{\mathbf{ij}}^- \in \mathbb{R}_+, \ (\mathbf{i}, \mathbf{j}) \in \mathcal{L} \\
& \left.\begin{aligned}
& \sum_{i \in \mathcal{A}(i^+)} (\lambda_i - \kappa_i) + \sum_{i \in \mathcal{A}(i^-)} (\lambda_i - \kappa_i) \\
& \quad + \boldsymbol{d}_{i^+}^\top \boldsymbol{\psi}_{\mathbf{ij}}^+ + \boldsymbol{d}_{i^-}^\top \boldsymbol{\psi}_{\mathbf{ij}}^- - \chi_{\mathbf{ij}}^+ t_{j^+} - \chi_{\mathbf{ij}}^- t_{j^-} \le w_{\mathbf{j}} \\
& \boldsymbol{C}_{i^+}^\top \boldsymbol{\psi}_{\mathbf{ij}}^+ - \boldsymbol{C}_{i^-}^\top \boldsymbol{\psi}_{\mathbf{ij}}^- = \chi_{\mathbf{ij}}^+ \boldsymbol{s}_{j^+} - \chi_{\mathbf{ij}}^- \boldsymbol{s}_{j^-} \\
& \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_{i^+}^\top \boldsymbol{\psi}_{\mathbf{ij}}^+ = \boldsymbol{D}_{i^-}^\top \boldsymbol{\psi}_{\mathbf{ij}}^-
\end{aligned}\right\} \quad \forall (\mathbf{i}, \mathbf{j}) \in \mathcal{L},
\end{aligned}
\tag{21}
$$

where $\mathbf{i} = (i^+, i^-) \in \mathcal{I}^2$ *and* $\mathbf{j} = (j^+, j^-) \in (\mathcal{J} \cup \{0\})^2$ *are multi-indices,*

$$
\mathcal{L} = \left\{ (\mathbf{i}, \mathbf{j}) \in \mathcal{I}^2 \times (\mathcal{J} \cup \{0\})^2 : \mathcal{D}_{\mathbf{ij}} \ne \emptyset \right\}
$$

with $\mathcal{D}_{\mathbf{ij}}$ *defined in* (20), $w_{\mathbf{j}} = \mathbb{I}_{[j^+ = 0]} + \mathbb{I}_{[j^- = 0]}$, $\boldsymbol{s}_0 = \mathbf{0}$ *and* $t_0 = 0$.

In analogy to Theorem 3, Theorem 4 provides a reformulation that scales polynomially in the input data. Note also that (21) reduces to a linear program whenever the cones $\mathcal{K}_i$ for $i \in \mathcal{I}$ are polyhedral.

We emphasize that in Theorem 4, the center of symmetry is set to $\mathbf{0}$ merely to simplify the exposition. This is without loss of generality and can always be accomplished by a coordinate shift. Moreover, we note that one can efficiently verify whether $(\mathbf{i}, \mathbf{j}) \in \mathcal{L}$. If $\mathbf{j} > \mathbf{0}$, for example, we have $(\mathbf{i}, \mathbf{j}) \in \mathcal{L}$ if and only if the optimal value of the convex optimization problem

$$
\begin{aligned}
\sup \quad & \min \{ \boldsymbol{s}_{j^+}^\top \boldsymbol{z} - t_{j^+}, -\boldsymbol{s}_{j^-}^\top \boldsymbol{z} - t_{j^-} \} \\
\text{s.t.} \quad & (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q \\
& (\boldsymbol{z}, \boldsymbol{u}^+) \in \mathcal{C}_{i^+}, \ (-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_{i^-}
\end{aligned}
$$

is strictly positive. We illustrate Theorem 4 with two examples.

Example 8 (Uncertainty Quantification with Robust Dispersion Measures) Let $\mathcal{P}$ be the median-absolute deviation ambiguity set (11) from Example 3. Note that $\mathcal{P} = \Pi_{\tilde{\boldsymbol{z}}} \mathcal{P}'$, where $\mathcal{P}'$ is defined in (12). Moreover, $\mathcal{P}'$ can be viewed as the intersection of a moment ambiguity set $\mathcal{P}^n$ of the form (4) with $I = 1$ and a structural ambiguity set $\mathcal{P}^s$ containing all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P)$ under which $\Pi_{\tilde{\boldsymbol{z}}} \mathbb{P}$ is symmetric around $\boldsymbol{m}$. For $\boldsymbol{f} > \mathbf{0}$, $\mathcal{P}'$ satisfies the conditions (**B**), (**N**), (**D**), (**F**) and (**S**). Thus, we can employ a coordinate shift and use Theorem 4 to reformulate the uncertainty quantification problem (1) as the tractable linear program

$$
\begin{aligned}
\sup \quad & \alpha - \boldsymbol{f}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \alpha \in \mathbb{R}, \ \boldsymbol{\gamma} \in \mathbb{R}^K, \ \chi_{\mathbf{j}}^+, \chi_{\mathbf{j}}^- \in \mathbb{R}_+, \ \mathbf{j} \in \mathcal{L} \\
& \left.\begin{aligned}
& 2\alpha + \chi_{\mathbf{j}}^+ (\boldsymbol{s}_{j^+}^\top \boldsymbol{m} - t_{j^+}) + \chi_{\mathbf{j}}^- (\boldsymbol{s}_{j^-}^\top \boldsymbol{m} - t_{j^-}) \le w_{\mathbf{j}} \\
& \chi_{\mathbf{j}}^+ \boldsymbol{s}_{j^+} - \chi_{\mathbf{j}}^- \boldsymbol{s}_{j^-} \le 2\boldsymbol{\gamma}, \ \chi_{\mathbf{j}}^- \boldsymbol{s}_{j^-} - \chi_{\mathbf{j}}^+ \boldsymbol{s}_{j^+} \le 2\boldsymbol{\gamma}
\end{aligned}\right\} \quad \forall \mathbf{j} \in \mathcal{L},
\end{aligned}
$$

where $\mathbf{j} = (j^+, j^-) \in (\mathcal{J} \cup \{0\})^2$ is a multi-index ranging over

$$\mathcal{L} = \left\{ \mathbf{j} \in (\mathcal{J} \cup \{0\})^2 : \left[\exists \boldsymbol{z} \in \mathbb{R}^P : \begin{array}{l} j^+ > 0 \implies \boldsymbol{s}_{j^+}^\top (\boldsymbol{m} + \boldsymbol{z}) > t_{j^+} \\ j^- > 0 \implies \boldsymbol{s}_{j^-}^\top (\boldsymbol{m} - \boldsymbol{z}) > t_{j^-} \end{array} \right] \right\},$$

$w_{\mathbf{j}} = \mathbb{I}_{[j^+=0]} + \mathbb{I}_{[j^-=0]}$, $\boldsymbol{s}_0 = \mathbf{0}$ and $t_0 = 0$.

Example 9 (Data-Driven Uncertainty Quantification, cont'd) Let $\mathcal{P} = \mathcal{P}' \cap \mathcal{P}^s$, where $\mathcal{P}'$ is the lifted Wasserstein ambiguity set (16) from Example 5 and $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P \times \mathbb{R})$ with the property that $\Pi_{\tilde{\boldsymbol{z}}}\mathbb{P}$ is point symmetric around $\boldsymbol{m}$. If $\overline{u}$ is a strict upper bound on the Euclidean diameter of the support of $\mathbb{Q}$, then the ambiguity set $\mathcal{P}'$ satisfies the conditions (**B**), (**N**), (**D**) and (**F**). Even though $\mathcal{P}'$ fails to satisfy (**S**), it can be shown that Theorem 4 remains valid if r is sufficiently large. Thus, we can reformulate the uncertainty quantification problem as the convex optimization problem

$$\begin{aligned}
\sup \quad & \frac{1}{I} \sum_{i \in \mathcal{I}} \beta_i - \gamma r \\
\text{s.t.} \quad & \boldsymbol{\beta} \in \mathbb{R}^I, \ \gamma \in \mathbb{R}_+ \\
& \chi_{\mathbf{ij}}^+, \chi_{\mathbf{ij}}^- \in \mathbb{R}_+, \ \boldsymbol{\nu}_{\mathbf{ij}}^+, \boldsymbol{\nu}_{\mathbf{ij}}^- \in \mathbb{R}^P, \ (\mathbf{i}, \mathbf{j}) \in \mathcal{L} \\
& \left. \begin{array}{l} \|\boldsymbol{\nu}_{\mathbf{ij}}^+\|_q \leq \gamma, \quad \|\boldsymbol{\nu}_{\mathbf{ij}}^-\|_q \leq \gamma, \quad \boldsymbol{\nu}_{\mathbf{ij}}^+ + \boldsymbol{\nu}_{\mathbf{ij}}^- = \chi_{\mathbf{ij}}^+ \boldsymbol{s}_{j^+} - \chi_{\mathbf{ij}}^- \boldsymbol{s}_{j^-} \\ \beta_{i^+} + \beta_{i^-} + \boldsymbol{\nu}_{\mathbf{ij}}^{+\top}(\hat{\boldsymbol{z}}_{i^+} - \boldsymbol{m}) + \boldsymbol{\nu}_{\mathbf{ij}}^{-\top}(\boldsymbol{m} - \hat{\boldsymbol{z}}_{i^-}) \\ \quad + \chi_{\mathbf{ij}}^+(\boldsymbol{s}_{j^+}^\top \boldsymbol{m} - t_{j^+}) + \chi_{\mathbf{ij}}^-(\boldsymbol{s}_{j^-}^\top \boldsymbol{m} - t_{j^-}) \leq w_{\mathbf{j}} \end{array} \right\} \forall (\mathbf{i}, \mathbf{j}) \in \mathcal{L},
\end{aligned}$$

where $\mathbf{i} = (i^+, i^-)$ and $\mathbf{j} = (j^+, j^-)$ are multi-indices ranging over

$$\mathcal{L} = \left\{ (\mathbf{i}, \mathbf{j}) \in \mathcal{I}^2 \times (\mathcal{J} \cup \{0\})^2 : \left[\exists \boldsymbol{z} \in \mathbb{R}^P \ : \ \begin{array}{l} j^+ > 0 \implies \boldsymbol{s}_{j^+}^\top (\boldsymbol{m} + \boldsymbol{z}) > t_{j^+} \\ j^- > 0 \implies \boldsymbol{s}_{j^-}^\top (\boldsymbol{m} - \boldsymbol{z}) > t_{j^-} \end{array} \right] \right\},$$

$w_{\mathbf{j}} = \mathbb{I}_{[j^+=0]} + \mathbb{I}_{[j^-=0]}$, $\boldsymbol{s}_0 = \mathbf{0}$, $t_0 = 0$, and $q \geq 1$ is defined through $\frac{1}{p} + \frac{1}{q} = 1$. Problem (19) is a linear program if we choose $p \in \{1, \infty\}$.

Next, we consider instances of the uncertainty quantification problem (1) where $\mathcal{P}$ emerges from an intersection of the nested moment ambiguity set (4) with the set of all α-unimodal distributions $\mathcal{P}^\alpha$. In order to facilitate tractable reformulations, we restrict our attention to the subclass of *epigraphic moment ambiguity sets*. An epigraphic moment ambiguity set is an instance of (4) with $I = 1$ (i.e., it contains no confidence set other than the support), where the interaction between $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{u}}$ is captured through the epigraph constraint in

$$\mathcal{C}_1 = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \boldsymbol{C}_1 \boldsymbol{z} \preccurlyeq_{\mathcal{K}_1} \boldsymbol{d}_1, \ \boldsymbol{g}(\boldsymbol{z}) \preccurlyeq_{\mathcal{K}_2} \boldsymbol{u} \right\}, \tag{22}$$

where $\boldsymbol{C}_1 \in \mathbb{R}^{L_1 \times P}$, $\boldsymbol{d}_1 \in \mathcal{K}_1$ and $\mathcal{K}_1 \subseteq \mathbb{R}^{L_1}$, $\mathcal{K}_2 \subseteq \mathbb{R}^Q$ are proper cones. We require that the function $\boldsymbol{g} : \mathbb{R}^P \to \mathbb{R}^Q$ is $\mathcal{K}_2$-convex [9, § 3.6.2] and that the set $\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \boldsymbol{g}(\boldsymbol{z}) \preccurlyeq_{\mathcal{K}_2} \boldsymbol{u}\}$ is representable through conic inequalities and additional auxiliary variables [48]. We remark that any confidence set of the form (5) can be expressed as an instance of (22) and vice versa. Epigraphic ambiguity sets become a special case of (4) with $I = 1$, however, if structural properties are imposed on the marginal projection $\Pi_{\tilde{\boldsymbol{z}}}\mathbb{P}$.

Theorem 5 (Unimodality)   *Assume that $\mathcal{P} = \mathcal{P}^e \cap \mathcal{P}^\alpha$ where $\mathcal{P}^e$ is an epigraphic moment ambiguity set and $\mathcal{P}^\alpha$ is the structural ambiguity set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\tilde{z}}\mathbb{P}$ is α-unimodal around $\boldsymbol{z} = \boldsymbol{0}$. If the Slater condition (S) holds and $\boldsymbol{t} \geq \boldsymbol{0}$, then the uncertainty quantification problem (1) is equivalent to the semi-infinite program*

$$
\begin{aligned}
\sup \quad & \beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \beta \in \mathbb{R}, \ \boldsymbol{\gamma} \in \mathbb{R}^K, \ \tau_j \in \mathbb{R}_+, \ \boldsymbol{\psi}_0, \boldsymbol{\psi}_j \in \mathcal{K}_1^\star, \ j \in \mathcal{L} \\
& -\boldsymbol{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^\star \\
& \left. \begin{aligned}
& \beta + \left(\frac{\alpha}{\alpha+1} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{g}^\alpha(\boldsymbol{z}) \right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi}_0 \leq 1 \\
& \beta + \left(\frac{\alpha}{\alpha+1} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{g}^\alpha(\boldsymbol{z}) \right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi}_j \\
& \quad + \tau_j \boldsymbol{s}_j^\top \boldsymbol{z} - \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau_j t_j)^{\frac{\alpha}{\alpha+1}} \leq 0 \quad \forall j \in \mathcal{L}
\end{aligned} \right\} \quad \forall \boldsymbol{z} \in \mathbb{R}^P,
\end{aligned}
\tag{23}
$$

where $\mathcal{L} = \left\{ j \in \mathcal{J} \ : \ \exists (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha . \, \boldsymbol{s}_j^\top \boldsymbol{z} > 0 \right\}$ and $\boldsymbol{g}^\alpha(\boldsymbol{z}) = \int_0^1 \boldsymbol{g}(t\boldsymbol{z}) \, \alpha t^{\alpha-1} \mathrm{d}t$.

Unlike the previous results in this section, Theorem 5 does not provide a tractable reformulation *per se*. Instead, the tractability of problem (23) is determined by the properties of the function $\boldsymbol{g}^\alpha(\cdot)$ that appears in the semi-infinite constraints of (23). In the following, we present two examples for which problem (23) has a tractable reformulation.

Example 10 (Generalized Gauss Bounds) Theorem 5 allows us to compute the worst-case probability of the event $\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}$ if the distribution of $\tilde{\boldsymbol{z}}$ belongs to the Gauss ambiguity set $\mathcal{P}$ defined in equation (9) of Example 2. Since $\mathcal{P} = \Pi_{\tilde{\boldsymbol{z}}}\mathcal{P}'$ for the ambiguity set $\mathcal{P}'$ defined in (10), we have $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}] = \inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}]$. Moreover, the lifted ambiguity set $\mathcal{P}'$ satisfies the conditions of Theorem 5 if $\boldsymbol{\Sigma} \succ \boldsymbol{0}$, $\boldsymbol{S}\boldsymbol{m} \leq \boldsymbol{t}$ and $\mathcal{C}_1$ is defined as

$$
\mathcal{C}_1 = \left\{ (\boldsymbol{z}, \boldsymbol{U}) \in \mathbb{R}^P \times \mathbb{R}^{P \times P} : \boldsymbol{g}(\boldsymbol{z}) \preccurlyeq \boldsymbol{U} \right\},
$$

where $\boldsymbol{g}(\boldsymbol{z}) = \boldsymbol{z}\boldsymbol{z}^\top$ is $\mathbb{S}_+^P$-convex. Applying a coordinate shift and employing Theorem 5 allows us to reformulate the uncertainty quantification problem (1) over the Gauss ambiguity set as a semi-infinite program. Thereby, we recover a generalized multivariate Gauss inequality that was discovered in [38]. Since $\boldsymbol{g}^\alpha(\boldsymbol{z}) = \frac{\alpha}{\alpha+2} \boldsymbol{z}\boldsymbol{z}^\top$ can be computed explicitly, we may use standard robust optimization techniques to further simplify this semi-infinite program to the finite convex program

$$
\begin{aligned}
\sup \quad & \beta - (\boldsymbol{\mu} - \boldsymbol{m})^\top \boldsymbol{\gamma} - \langle \boldsymbol{\Sigma} + (\boldsymbol{\mu} - \boldsymbol{m})(\boldsymbol{\mu} - \boldsymbol{m})^\top, \boldsymbol{\Gamma} \rangle \\
\text{s.t.} \quad & \beta \in \mathbb{R}, \ \boldsymbol{\gamma} \in \mathbb{R}^P, \ \boldsymbol{\Gamma} \in \mathbb{S}_+^P, \ \tau_j \in \mathbb{R}_+, \ j \in \mathcal{L} \\
& \begin{pmatrix} 1 - \beta & \frac{1}{2}\frac{\alpha}{\alpha+1}\boldsymbol{\gamma}^\top \\ \frac{1}{2}\frac{\alpha}{\alpha+1}\boldsymbol{\gamma} & \frac{\alpha}{\alpha+2}\boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \boldsymbol{0} \\
& \begin{pmatrix} \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha}(\tau_j(t_j - \boldsymbol{s}_j^\top \boldsymbol{m}))^{\frac{\alpha}{\alpha+1}} - \beta & \frac{1}{2}(\frac{\alpha}{\alpha+1}\boldsymbol{\gamma} - \tau_j \boldsymbol{s}_j)^\top \\ \frac{1}{2}(\frac{\alpha}{\alpha+1}\boldsymbol{\gamma} - \tau_j \boldsymbol{s}_j) & \frac{\alpha}{\alpha+2}\boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \boldsymbol{0} \ \ \forall j \in \mathcal{L},
\end{aligned}
\tag{24}
$$

where $\mathcal{L} = \{j \in \mathcal{J} : \boldsymbol{s}_j \neq \boldsymbol{0}\}$. Note that (24) is equivalent to a tractable semidefinite program if α is rational, in which case the nonlinear term in the second matrix inequality can be linearized by using a well-known conic expansion of power functions, see e.g. [5, § 2.3.1].

Example 11 (Mean-Absolute Deviation & Unimodality) Let $\mathcal{P}$ be the ambiguity set of all distributions on $\mathbb{R}^P$ that are α-unimodal with mean and center $\boldsymbol{m}$ and whose mean-absolute deviation is bounded above by $\boldsymbol{f}$, that is,

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{array}{l} \mathbb{P} \text{ is } \alpha\text{-unimodal with center } \boldsymbol{m} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{m}, \ \ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{z}} - \boldsymbol{m}|] \leq \boldsymbol{f} \end{array} \right\}.$$

Since $\Pi_{\tilde{\boldsymbol{z}}} \mathcal{P}' = \mathcal{P}$ for the ambiguity set $\mathcal{P}'$ defined as

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) : \begin{array}{l} \Pi_{\tilde{\boldsymbol{z}}} \mathbb{P} \text{ is } \alpha\text{-unimodal with center } \boldsymbol{m}, \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{m}, \ \ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{u}}] = \boldsymbol{f}, \ \ \mathbb{P}[|\tilde{\boldsymbol{z}} - \boldsymbol{m}| \leq \tilde{\boldsymbol{u}}] = 1 \end{array} \right\},$$

we have $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}] = \inf_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}]$. Moreover, $\mathcal{P}'$ satisfies the conditions of Theorem 5 if we define $\boldsymbol{g}(\boldsymbol{z}) = |\boldsymbol{z} - \boldsymbol{m}|$ and require that $\boldsymbol{S}\boldsymbol{m} \leq \boldsymbol{t}$ and $\boldsymbol{f} > \boldsymbol{0}$. Applying a coordinate shift and using Theorem 5 in conjunction with standard robust optimization techniques, we can thus reformulate the uncertainty quantification problem (1) as the finite convex program

$$
\begin{array}{ll}
\sup & \beta - \boldsymbol{f}^\top \boldsymbol{\eta} \\
\text{s.t.} & \beta \in \mathbb{R}, \ \boldsymbol{\theta} \in \mathbb{R}^P, \ \boldsymbol{\eta} \in \mathbb{R}^P_+, \ \tau_j \in \mathbb{R}_+, \ j \in \mathcal{L} \\
& \boldsymbol{\eta} \geq \boldsymbol{\theta} \geq -\boldsymbol{\eta}, \ \beta \leq 1 \\
& \dfrac{\alpha}{\alpha + 1} \boldsymbol{\eta} \geq \dfrac{\alpha}{\alpha + 1} \boldsymbol{\theta} + \tau_j \boldsymbol{s}_j \geq -\dfrac{\alpha}{\alpha + 1} \boldsymbol{\eta} \\
& \beta \leq \dfrac{(\alpha + 1)^{\alpha + 1}}{\alpha^\alpha} (\tau_j(t_j - \boldsymbol{s}_j^\top \boldsymbol{m}))^{\frac{\alpha}{\alpha + 1}}
\end{array} \ \right\} \ \forall j \in \mathcal{L},
\tag{25}
$$

where $\mathcal{L} = \{j \in \mathcal{J} : \boldsymbol{s}_j \neq \boldsymbol{0}\}$. Note that (25) can be reformulated as a tractable second-order cone program if α is rational, see [5, § 2.3.1].

In addition to evaluating the worst-case probability (1), practical applications often require simulation runs under the distribution that attains the worst-case probability. We can construct such worst-case distributions from the dual problems of (17), (21) and (23). In the interest of space, we do not embark on this pathway and instead refer the interested reader to [24, 46].

To conclude this section, we investigate the tractability of the uncertainty quantification problem (1) when the components of $\tilde{\boldsymbol{z}}$ are mutually independent under every distribution within the ambiguity set $\mathcal{P}$.

Theorem 6 (Independence) *The uncertainty quantification problem* (1) *over the ambiguity set $\mathcal{P}$ defined as*

$$\mathcal{P} = \mathcal{P}^n \cap \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \text{ the components of } \tilde{\boldsymbol{z}} \text{ are independent under } \mathbb{P}\},$$

where $\mathcal{P}^n$ is the moment ambiguity set (4)*, is strongly $\mathcal{NP}$-hard even if $(\boldsymbol{B})$ and $(\boldsymbol{N})$ are satisfied and $K = 0$ (no expectation constraints).*

One can show that if in addition to the assumptions of Theorem 6, the ambiguity set $\mathcal{P}^n$ satisfies $I = 1$ (that is, there are no confidence sets other than the support), then the uncertainty quantification problem (1) reduces to a robust feasibility problem that can be solved efficiently. Also, if the moment ambiguity set in Theorem 6 satisfies $K > 0$ but $I = 1$ and $\mathcal{C}_1 = \mathbb{R}^P \times \mathbb{R}^Q$, then problem (1) evaluates to 1 if both $\boldsymbol{S} = \boldsymbol{0}$ and $\boldsymbol{t} \geq \boldsymbol{0}$ and to 0 otherwise.

4 Chance Constrained Programming

We now study chance constrained programs of the form (2), where the safety of the underlying system can be actively enhanced by adjusting the design decisions $\boldsymbol{x} \in \mathbb{R}^N$. It turns out that the tractability of (2) is intimately related to the number of rows J of the technology matrix $\boldsymbol{S}(\boldsymbol{x})$ and the right-hand side vector $\boldsymbol{t}(\boldsymbol{x})$. Hence, Section 4.1 is devoted to *individual chance constraints* where $J = 1$, and Section 4.2 studies *joint chance constraints* where $J > 1$.

4.1 Individual Chance Constraints

As $J = 1$ throughout this section, we can simplify the notation if we denote the technology matrix by $\boldsymbol{s}(\boldsymbol{x})^\top$ and the right-hand side vector by $t(\boldsymbol{x})$.

We first study instances of the chance constrained program (2) where the ambiguity set is of the form (4). To derive a tractable reformulation for such problems, we restrict our attention to the subclass of *Markov ambiguity sets*, which are defined as instances of (4) with $I = 1$ and $\underline{p}_1 = \overline{p}_1 = 1$. Note that these ambiguity sets involve no confidence sets other than the support. We emphasize that apart from the Wasserstein ambiguity set all other examples of Section 2.3 constitute either Markov ambiguity sets or result from intersections of Markov ambiguity sets with structural ambiguity sets. Note that Markov ambiguity sets satisfy the conditions (**B**) and (**N**).

Theorem 7 (Moment Ambiguity Sets) *If $\mathcal{P}$ is a Markov ambiguity set satisfying (S) and $J = 1$, then the chance constraint in (2) is satisfied if and only if there are $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^\star$, $\tau \in \mathbb{R}_+$ such that*

$$
\begin{array}{lll}
\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq (1 - \epsilon)\tau, & \beta + \boldsymbol{d}_1^\top \boldsymbol{\phi} \leq \tau, & \beta + \boldsymbol{d}_1^\top \boldsymbol{\psi} \leq t(\boldsymbol{x}) \\
\boldsymbol{A}^\top \boldsymbol{\gamma} = \boldsymbol{C}_1^\top \boldsymbol{\phi}, & \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\phi} & \\
\boldsymbol{A}^\top \boldsymbol{\gamma} + \boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{C}_1^\top \boldsymbol{\psi}, & \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\psi}.
\end{array}
\tag{26}
$$

Note that (26) is a system of linear constraints that scales polynomially in the description of problem (2). The next result shows that the restriction to Markov ambiguity sets in Theorem 7 is necessary.

Theorem 8 *If $\mathcal{P}$ is an instance of (4) with $I > 1$, then the chance constrained program (2) is strongly $\mathcal{NP}$-hard even if $J = 1$, $K = 0$ (no expectation constraints) and the boundedness condition (B) as well as the nesting condition (N) are satisfied, while $\mathcal{X}$ is a polyhedron.*

We illustrate the tractable reformulation of Theorem 7 with two examples.

Example 12 (Chebyshev Ambiguity Set) We can use Theorem 7 to derive a tractable reformulation for individual chance constraints over the Chebyshev ambiguity set $\mathcal{P}$ defined in Example 1 with $\boldsymbol{\Sigma} \succ \mathbf{0}$. We have $\mathcal{P} = \Pi_{\tilde{\boldsymbol{z}}}\mathcal{P}'$ for the Markov ambiguity set $\mathcal{P}'$ defined in (8). Theorem 7 thus implies that the individual chance constraint $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left[\boldsymbol{s}(\boldsymbol{x})^\top \tilde{\boldsymbol{z}} \leq t(\boldsymbol{x})\right]$ is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^P$, $\boldsymbol{\Gamma} \in \mathbb{S}_+^P$ and $\tau \in \mathbb{R}_+$ with

$$\beta - \boldsymbol{\mu}^\top\boldsymbol{\gamma} - \langle \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top, \boldsymbol{\Gamma}\rangle \geq (1-\epsilon)\tau$$
$$\begin{pmatrix} \tau - \beta & \frac{1}{2}\boldsymbol{\gamma}^\top \\ \frac{1}{2}\boldsymbol{\gamma} & \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0}$$
$$\begin{pmatrix} t_1(\boldsymbol{x}) - \beta & \frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{s}(\boldsymbol{x}))^\top \\ \frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{s}(\boldsymbol{x})) & \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0}.$$

Moreover, one can show that this constraint system is satisfied if and only if

$$\sqrt{\frac{1-\epsilon}{\epsilon}}\|\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{s}(\boldsymbol{x})\|_2 + \boldsymbol{\mu}^\top\boldsymbol{s}(\boldsymbol{x}) \leq t(\boldsymbol{x}),$$

which is a second-order cone constraint, see [22].

Example 13 (Huber Ambiguity Set) Theorem 7 allows us to derive a tractable reformulation for individual chance constraints over the Huber ambiguity set $\mathcal{P}$ defined in Example 4, assuming that $\boldsymbol{f}^\top\boldsymbol{\mu} < g$. We have $\mathcal{P} = \Pi_{\tilde{\boldsymbol{z}}}\mathcal{P}'$ for the Markov ambiguity set $\mathcal{P}'$ defined in (13). Theorem 7 thus implies that the individual chance constraint $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left[\boldsymbol{s}(\boldsymbol{x})^\top \tilde{\boldsymbol{z}} \leq t(\boldsymbol{x})\right]$ is satisfied if and only if there are $\alpha, \nu_0, \nu_1 \in \mathbb{R}$, $\boldsymbol{\psi} \in \mathbb{R}^P$ and $\tau, \phi, \lambda_0, \lambda_1 \in \mathbb{R}_+$ that satisfy the following semi-definite constraints.

$$\alpha - \phi g \geq (1-\epsilon)\tau$$
$$\begin{pmatrix} \lambda_0 & \nu_0 \\ \nu_0 & \frac{1}{2}\phi \end{pmatrix} \succcurlyeq \mathbf{0}, \quad \begin{pmatrix} \lambda_1 & \nu_1 \\ \nu_1 & \frac{1}{2}\phi \end{pmatrix} \succcurlyeq \mathbf{0}$$
$$\alpha + \lambda_0 \leq \tau, \quad \boldsymbol{\psi} = 2\nu_0\boldsymbol{f}, \quad -\beta\phi \leq 2\nu_0 \leq \beta\phi$$
$$\alpha + \lambda_1 - t(\boldsymbol{x}) + \boldsymbol{s}(\boldsymbol{x})^\top\boldsymbol{\mu} \leq 0, \quad \boldsymbol{\psi} + \boldsymbol{s}(\boldsymbol{x}) = 2\nu_1\boldsymbol{f}, \quad -\beta\phi \leq 2\nu_1 \leq \beta\phi$$

We now consider instances of the chance constrained program (2) where $\mathcal{P}$ is generated by intersecting a Markov ambiguity set with the set of all symmetric distributions $\mathcal{P}^s$.

Theorem 9 (Symmetry) *Assume that $\mathcal{P} = \mathcal{P}^m \cap \mathcal{P}^s$, where $\mathcal{P}^m$ is a Markov ambiguity set that satisfies the dimensionality condition (**D**) as well as the feasibility condition (**F**), and $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\tilde{\boldsymbol{z}}}\mathbb{P}$ is point symmetric around $\boldsymbol{m} = \mathbf{0}$. If the Slater condition (**S**) holds, then the chance constraint in (2) with $J = 1$ and $\epsilon \in (0, 1/2)$ is satisfied if and only if there are $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi}, \boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^\star$ and $\tau \in \mathbb{R}_+$ such that*

$$\begin{array}{ll} \beta + \boldsymbol{b}^\top\boldsymbol{\gamma} \geq (1-\epsilon)\tau, & \beta + \boldsymbol{d}_1^\top\boldsymbol{\phi} \leq \tau \\ 2\beta + \boldsymbol{d}_1^\top(\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) - t(\boldsymbol{x}) \leq \tau, & \boldsymbol{C}_1^\top(\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) = \boldsymbol{s}(\boldsymbol{x}) \\ \boldsymbol{B}^\top\boldsymbol{\gamma} = \boldsymbol{D}_1^\top\boldsymbol{\psi}^+ = \boldsymbol{D}_1^\top\boldsymbol{\psi}^- = \boldsymbol{D}_1^\top\boldsymbol{\phi}. \end{array} \tag{27}$$

In analogy to Theorem 7, Theorem 9 reexpresses the individual chance constraint as a finite set of tractable constraints. One can show that a Markov ambiguity set $\mathcal{P}^m$ satisfies the feasibility condition (**F**) whenever the projected support $\{z \in \mathbb{R}^P : \exists u \in \mathbb{R}^Q . (z, u) \in \mathcal{C}_1\}$ is point symmetric around $m = 0$. The restriction to Markov ambiguity sets in Theorem 9 is again necessary.

Theorem 10 *If $\mathcal{P} = \mathcal{P}^n \cap \mathcal{P}^s$ where $\mathcal{P}^n$ is an instance of (4) with $I > 1$ and $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\tilde{z}}\mathbb{P}$ is point symmetric around $m = 0$, then the chance constrained program (2) is strongly $\mathcal{NP}$-hard even if $J = 1$, $K = 0$ (no expectation constraints) and the conditions (**B**), (**N**), (**D**) and (**F**) are satisfied, while $\mathcal{X}$ is a polyhedron.*

We illustrate the tractable reformulation of Theorem 9 with an example.

Example 14 (Median-Absolute Deviation & Symmetry) We can use Theorem 9 to derive a tractable reformulation for individual chance constraints over the median-absolute deviation ambiguity set $\mathcal{P}$ defined in Example 3 with $f > 0$. Note that we have $\mathcal{P} = \Pi_{\tilde{z}}\mathcal{P}'$ for $\mathcal{P}'$ defined in (12). If we assume that $\epsilon \in (0, 1/2)$ and apply a coordinate shift, then all conditions of Theorem 9 are met and the chance constraint in (2) is satisfied if and only if

$$\frac{1}{2\epsilon} f^\top |s(x)| + m^\top s(x) \le t(x),$$

which can be expressed by a system of $2P + 1$ linear inequalities.

Next, we consider instances of the chance constrained program (2) where the ambiguity set $\mathcal{P}$ emerges from the intersection of an epigraphic moment ambiguity set with the set of all unimodal distributions $\mathcal{P}^\alpha$.

Theorem 11 (Unimodality) *Assume that $\mathcal{P} = \mathcal{P}^e \cap \mathcal{P}^\alpha$ where $\mathcal{P}^e$ is an epigraphic moment ambiguity set, whereas $\mathcal{P}^\alpha$, $\alpha > 1$, is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\tilde{z}}\mathbb{P}$ is α-unimodal around $z = 0$. If the Slater condition (**S**) holds, $J = 1$ and $t(x) \ge 0$ for all $x \in \mathcal{X}$, then the chance constraint in (2) is satisfied if and only if there are $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$, $\phi, \psi \in \mathcal{K}_1^\star$ such that*

$$\left.\begin{aligned}
&\beta + b^\top \gamma \ge (1 - \epsilon)\tau, \quad -B^\top \gamma \in \mathcal{K}_2^\star \\
&\beta + \left(\frac{\alpha}{\alpha + 1} Az + Bg^\alpha(z)\right)^\top \gamma + (d_1 - C_1 z)^\top \phi \le \tau \\
&\beta + \left(\frac{\alpha}{\alpha + 1} Az + Bg^\alpha(z)\right)^\top \gamma + (d_1 - C_1 z)^\top \psi \\
&\qquad + s(x)^\top z - \frac{(\alpha + 1)^{\alpha + 1}}{\alpha^\alpha} \tau^{\frac{1}{\alpha + 1}} t(x)^{\frac{\alpha}{\alpha + 1}} \le 0
\end{aligned}\right\} \quad \forall z \in \mathbb{R}^P. \quad (28)$$

We remark that the term $-\tau^{\frac{1}{\alpha + 1}} t(x)^{\frac{\alpha}{\alpha + 1}}$ is convex on $\{(\tau, x) \in \mathbb{R}_+ \times \mathcal{X}\}$ since $\alpha > 0$ and $t(x) \ge 0$ for all $x \in \mathcal{X}$. If α is a rational number exceeding 1, that is, if $\alpha = p/q$ for some $p, q \in \mathbb{N}$ with $p \ge q$, then the epigraph of this term can be expressed through $\mathcal{O}(p)$ second-order conic constraints, see e.g. [38,

Lemma 4.2]. In analogy to the formulation (23) of Theorem 5, the tractability of (28) depends on the functional form of $\boldsymbol{g}^\alpha(\cdot)$. We close this section with an example for which the constraint system (28) has a tractable reformulation.

Example 15 (Gauss Ambiguity Set) Theorem 11 allows us to derive a tractable reformulation for individual chance constraints over the Gauss ambiguity set $\mathcal{P}$ defined in Example 2 whenever $\boldsymbol{\Sigma} \succ \mathbf{0}$. Since $\mathcal{P} = \Pi_{\tilde{\boldsymbol{z}}}\mathcal{P}'$ for $\mathcal{P}'$ defined in (10), we can replace $\mathcal{P}$ in the chance constrained program (2) with $\mathcal{P}'$. The lifted ambiguity set $\mathcal{P}'$ satisfies the conditions of Theorem 11 if we set

$$\mathcal{C}_1 = \left\{ (\boldsymbol{z}, \boldsymbol{U}) \in \mathbb{R}^P \times \mathbb{R}^{P \times P} : \boldsymbol{g}(\boldsymbol{z}) \preccurlyeq \boldsymbol{U} \right\},$$

where $\boldsymbol{g}(\boldsymbol{z}) = \boldsymbol{z}\boldsymbol{z}^\top$ is $\mathbb{S}_+^P$-convex, and require that the mode $\boldsymbol{m}$ satisfies $\boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{m} \leq t(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}$. Applying a coordinate shift, using Theorem 11 and employing standard robust optimization techniques allows us to conclude that the chance constraint in (2) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^P$, $\boldsymbol{\Gamma} \in \mathbb{S}_+^P$ and $\tau \in \mathbb{R}_+$ such that

$$\beta - (\boldsymbol{\mu} - \boldsymbol{m})^\top \boldsymbol{\gamma} - \langle \boldsymbol{\Sigma} + (\boldsymbol{\mu} - \boldsymbol{m})(\boldsymbol{\mu} - \boldsymbol{m})^\top, \boldsymbol{\Gamma} \rangle \geq 1 - \epsilon$$

$$\begin{pmatrix} \tau - \beta & \frac{1}{2}\frac{\alpha}{\alpha+1}\boldsymbol{\gamma}^\top \\ \frac{1}{2}\frac{\alpha}{\alpha+1}\boldsymbol{\gamma} & \frac{\alpha}{\alpha+2}\boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0}$$

$$\begin{pmatrix} \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha}\tau^{\frac{1}{\alpha+1}}(t(\boldsymbol{x}) - \boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{m})^{\frac{\alpha}{\alpha+1}} - \beta & \frac{1}{2}(\frac{\alpha}{\alpha+1}\boldsymbol{\gamma} - \boldsymbol{s}(\boldsymbol{x}))^\top \\ \frac{1}{2}(\frac{\alpha}{\alpha+1}\boldsymbol{\gamma} - \boldsymbol{s}(\boldsymbol{x})) & \frac{\alpha}{\alpha+2}\boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0}.$$

4.2 Joint Chance Constraints

We now study joint chance constrained programs of the form (2) where $J > 1$. Until recently, such problems were suspected to be generically intractable, and the majority of the literature focused on conservative approximations via Bonferroni's inequality [36], distributionally robust conditional value-at-risk constraints [12,51] and component-wise quasi-concave functions [33]. In the following, we present an *exact* tractable reformulation of problem (2) for a specific class of ambiguity sets, and we argue that this tractability result is unlikely to extend to more general settings. The results of this section originate from [24], and thus we do not repeat the proofs here.

We first consider instances of (2) where the ambiguity set is of the form (4). To obtain tractable reformulations for such problems, we restrict our attention to *conic moment ambiguity sets*, which we define as Markov ambiguity sets with $\boldsymbol{d}_1 = \mathbf{0}$. Conic moment ambiguity sets form a strict subclass of the epigraphic moment ambiguity sets introduced in Section 3. Nevertheless, they are expressive enough to capture moment constraints involving many interesting dispersion measures such as the mean-absolute deviation $\mathbb{E}\,|\cdot|$, the mean-semideviation $\mathbb{E}\,[\cdot]_+$ and the mean-maximum deviation $\mathbb{E}\,\|\cdot\|_\infty$.

Theorem 12 *Assume that the technology matrix of the chance constraint in (2) is fixed, that is, $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{S}$. Let $\mathcal{P}$ be a conic moment ambiguity set*

satisfying (S), and assume that $(\boldsymbol{s}_j, \boldsymbol{0}) \notin \mathrm{pol}(\mathcal{C}_1)$ for all $j \in \mathcal{J}$, where $\mathrm{pol}(\mathcal{C}_1)$ denotes the polar cone of $\mathcal{C}_1$. Then the chance constraint in (2) is satisfied if and only if there exist $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi} \in \mathcal{K}_1^\star$, $\tau_j \in \mathbb{R}_+$ and $\boldsymbol{\psi}_j \in \mathcal{K}_1^\star$, $j \in \mathcal{J}$, with

$$\left.\begin{array}{l} 1 + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, \;\; \boldsymbol{A}^\top \boldsymbol{\gamma} = \boldsymbol{C}_1^\top \boldsymbol{\phi}, \;\; \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\phi} \\[4pt] \boldsymbol{A}^\top \boldsymbol{\gamma} + \tau_j \boldsymbol{s}_j = \boldsymbol{C}_1^\top \boldsymbol{\psi}_j, \;\; \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\psi}_j \\[4pt] \left\| \begin{bmatrix} 2 \\ \tau_j - t_j(\boldsymbol{x}) \end{bmatrix} \right\|_2 \leq \tau_j + t_j(\boldsymbol{x}) \end{array}\right\} \;\; \forall j \in \mathcal{J}, \tag{29}$$

where $\boldsymbol{s}_j$, $j = 1, \ldots, J$, denotes the j-th row of $\boldsymbol{S}$ as a column vector.

Note that (29) is a system of conic-quadratic constraints that scales polynomially with the description of the chance constraint (2). In contrast to all other results in this paper, joint chance constraints thus require a conic-quadratic reformulation even for ambiguity sets with a polyhedral support.

The following two results show that both the restriction to conic moment ambiguity sets and to fixed technology matrices is critical in Theorem 12.

Theorem 13 *The joint chance constrained program* (2) *is strongly $\mathcal{NP}$-hard even if $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{S}$ and $\mathcal{P}$ is an epigraphic moment ambiguity set whose support $\mathcal{C}_1$ represents a hypercube.*

Theorem 14 *If the technology matrix $\boldsymbol{S}(\boldsymbol{x})$ depends on $\boldsymbol{x}$, then the joint chance constrained program* (2) *is strongly $\mathcal{NP}$-hard even if $\mathcal{P}$ constitutes a conic moment ambiguity set.*

Chance constraints with fixed technology matrices arise naturally, for example, in inventory control, cash matching, unit commitment or airline revenue management problems etc. We illustrate the tractable reformulation offered by Theorem 12 with two examples.

Example 16 (Mean-Absolute Deviation) Theorem 12 allows us to derive a tractable reformulation of joint chance constraints over mean-absolute deviation ambiguity sets $\mathcal{P}$ of the form

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) \, : \, \mathbb{E}_\mathbb{P}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}, \;\; \mathbb{E}_\mathbb{P}[|\tilde{\boldsymbol{z}} - \boldsymbol{\mu}|] \leq \boldsymbol{f} \right\},$$

where the absolute value is understood to apply component-wise and $\boldsymbol{f} > \boldsymbol{0}$. Consider now an instance of (4) involving the auxiliary random vector $\tilde{\boldsymbol{u}} \in \mathbb{R}^P$.

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) \, : \, \begin{array}{l} \mathbb{E}_\mathbb{P}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}, \;\; \mathbb{E}_\mathbb{P}[\tilde{\boldsymbol{u}}] = \boldsymbol{f}, \\ \mathbb{P}[\tilde{\boldsymbol{u}} \geq \tilde{\boldsymbol{z}} - \boldsymbol{\mu}, \; \tilde{\boldsymbol{u}} \geq \boldsymbol{\mu} - \tilde{\boldsymbol{z}}] = 1 \end{array} \right\}$$

One readily verifies that $\mathcal{P} = \Pi_{\tilde{\boldsymbol{z}}} \mathcal{P}'$ and the lifted ambiguity set $\mathcal{P}'$ satisfies the conditions of Theorem 12. We thus conclude that the joint chance constraint in (2) is satisfied if and only if there exist $\boldsymbol{\gamma}, \boldsymbol{\beta} \in \mathbb{R}^P$ and $\tau_j \in \mathbb{R}_+$, $j \in \mathcal{L}$, with

$$\left.\begin{array}{l} 1 + \boldsymbol{f}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, \;\; -\boldsymbol{\gamma} \geq \boldsymbol{\beta} \geq \boldsymbol{\gamma} \\[4pt] -\boldsymbol{\gamma} \geq \boldsymbol{\beta} + \tau_j \boldsymbol{s}_j \geq \boldsymbol{\gamma} \\[4pt] \left\| \begin{bmatrix} 2 \\ \tau_j - t_j(\boldsymbol{x}) + \boldsymbol{s}_j^\top \boldsymbol{\mu} \end{bmatrix} \right\| \leq \tau_j + t_j(\boldsymbol{x}) - \boldsymbol{s}_j^\top \boldsymbol{\mu} \end{array}\right\} \;\; \forall j \in \mathcal{J}.$$

Example 17 (Mean Semi-Deviation) Theorem 12 also allows us to derive a tractable reformulation of joint chance constraints over more general mean semi-deviation ambiguity sets $\mathcal{P}$ of the form

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) \, : \, \mathbb{E}_{\mathbb{P}}\left[\tilde{z}\right] = \mu, \ \mathbb{E}_{\mathbb{P}}\left[\tilde{z} - \mu\right]_+ \leq f^+, \ \mathbb{E}_{\mathbb{P}}\left[\mu - \tilde{z}\right]_+ \leq f^- \right\},$$

where the operator $[\cdot]_+ = \max\{\cdot, 0\}$ applies component-wise and $\mathbf{f}^+, \mathbf{f}^- > \mathbf{0}$. Consider the following instance of (4), which involves the auxiliary random vectors $\tilde{u}^+, \tilde{u}^- \in \mathbb{R}^P$.

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P \times \mathbb{R}^P) \, : \, \begin{array}{c} \mathbb{E}_{\mathbb{P}}\left[\tilde{z}\right] = \mu, \ \mathbb{P}\left[\tilde{u}^+, \tilde{u}^- \geq 0\right] = 1, \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{u}^+\right] = f^+, \ \mathbb{E}_{\mathbb{P}}\left[\tilde{u}^-\right] = f^-, \\ \mathbb{P}\left[\tilde{u}^+ \geq \tilde{z} - \mu, \ \tilde{u}^- \geq \mu - \tilde{z}\right] = 1 \end{array} \right\}$$

One readily verifies that $\mathcal{P} = \Pi_{\tilde{z}}\mathcal{P}'$ and the lifted ambiguity set $\mathcal{P}'$ satisfies the conditions of Theorem 12. We thus conclude that the joint chance constraint in (2) is satisfied if and only if there exist $\boldsymbol{\beta}, \boldsymbol{\gamma}^+, \boldsymbol{\gamma}^- \in \mathbb{R}^P$, $\boldsymbol{\theta}_j, \boldsymbol{\eta}_j \in \mathbb{R}_+^P$, $j \in \mathcal{J} \cup \{0\}$, and $\tau_j \in \mathbb{R}_+$, $j \in \mathcal{J}$, with

$$\left. \begin{array}{l} 1 + (\boldsymbol{f}^+)^\top \boldsymbol{\gamma}^+ + (\boldsymbol{f}^-)^\top \boldsymbol{\gamma}^- \geq 1 - \epsilon \\ \boldsymbol{\beta} = \boldsymbol{\theta}_0 - \boldsymbol{\eta}_0, \quad \boldsymbol{\gamma}^+ \leq -\boldsymbol{\theta}_0, \ \boldsymbol{\gamma}^- \leq -\boldsymbol{\eta}_0 \\ \boldsymbol{\beta} + \tau_j \boldsymbol{s}_j = \boldsymbol{\theta}_j - \boldsymbol{\eta}_j, \quad \boldsymbol{\gamma}^+ \leq -\boldsymbol{\theta}_j, \ \boldsymbol{\gamma}^- \leq -\boldsymbol{\eta}_j \\ \left\| \begin{bmatrix} 2 \\ \tau_j - t_j(\boldsymbol{x}) + \boldsymbol{s}_j^\top \boldsymbol{\mu} \end{bmatrix} \right\| \leq \tau_j + t_j(\boldsymbol{x}) - \boldsymbol{s}_j^\top \boldsymbol{\mu} \end{array} \right\} \ \forall j \in \mathcal{J}.$$

If the ambiguity set of Theorem 12 is restricted to distributions $\mathbb{P} \in \mathcal{P}$ with point symmetric marginals $\Pi_{\tilde{z}}\mathbb{P}$, then the joint chance constrained program (2) is strongly $\mathcal{NP}$-hard even if the support set $\mathcal{C}_1$ is a cone.

Theorem 15 *If the ambiguity set $\mathcal{P}$ satisfies $\mathcal{P} = \mathcal{P}^c \cap \mathcal{P}^s$ where $\mathcal{P}^c$ is a conic moment ambiguity set and $\mathcal{P}^s$ is the set of all distributions $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ with the property that $\Pi_{\tilde{z}}\mathbb{P}$ is point symmetric around $\boldsymbol{m} = \boldsymbol{0}$, then the joint chance constrained program (2) is strongly $\mathcal{NP}$-hard even if $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{S}$.*

Finally, if the ambiguity set $\mathcal{P}$ in Theorem 12 is restricted to contain only distributions $\mathbb{P} \in \mathcal{P}$ under which $\Pi_{\tilde{z}}\mathbb{P}$ is α-unimodal, then the intractability results of Theorems 13 and 14 remain valid since any distribution can be approximated arbitrarily well—in the weak sense—by an α-unimodal distribution with sufficiently large $\alpha > 0$, see e.g. [38]. To our best knowledge, the complexity of the joint chance constrained program (2) has not been settled for $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{S}$ and for $\mathcal{P}$ generated by the intersection of a conic moment ambiguity set and the structural ambiguity set $\mathcal{P}^\alpha$ of all unimodal distributions.

5 Approximation Algorithm

While the uncertainty quantification problem (1) can be solved efficiently for a broad range of ambiguity sets (see Section 3), the associated chance constrained programs (2) frequently become intractable (see Section 4). In this

Ambiguity set	Coupled decisions $\boldsymbol{\chi}$	Uncoupled decisions $\boldsymbol{\rho}$
Moment information (Theorem 3)	$(\{\tau_{ij}\}_{ij})$	$(\boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{\kappa}, \{\boldsymbol{\phi}_i\}_i, \{\boldsymbol{\psi}_{ij}\}_{ij})$
Moments + symmetry (Theorem 4)	$(\{\chi_{\mathbf{ij}}^+\}_{\mathbf{ij}}, \{\chi_{\mathbf{ij}}^-\}_{\mathbf{ij}})$	$(\boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{\kappa}, \{\boldsymbol{\psi}_{\mathbf{ij}}^+\}_{\mathbf{ij}}, \{\boldsymbol{\psi}_{\mathbf{ij}}^-\}_{\mathbf{ij}})$
Moments + unimodality (Example 10)	$(\{\tau_j\}_j)$	$(\beta, \boldsymbol{\gamma}, \boldsymbol{\Gamma})$
Moments + unimodality (Example 11)	$(\{\tau_j\}_j)$	$(\beta, \boldsymbol{\theta}, \boldsymbol{\eta})$

Table 1 Definitions of $\boldsymbol{\chi}$ and $\boldsymbol{\rho}$ for some of the ambiguity sets studied in this paper.

section, we therefore report a heuristic for chance constrained programs that determines 'good' but in general suboptimal decisions even if the associated instance of (2) is intractable. The key idea is to decompose problem (2) into an *uncertainty quantification problem* that evaluates the worst-case probability of the chance constraint in (2) for a fixed decision $\boldsymbol{x}$ and a *policy improvement problem* that aims to improve the current decision $\boldsymbol{x}$.

To this end, we introduce a unified notation for the uncertainty quantification problems (17), (21) and (23). We denote the objective function of the unified uncertainty quantification problem by $Q(\boldsymbol{\psi})$, where we combine all decision variables to a single vector $\boldsymbol{\psi}$. Likewise, we represent the constraints of the unified uncertainty quantification problem as $\boldsymbol{\psi} \in \mathcal{Q}(\boldsymbol{x})$, where we replace the coefficient matrix $\boldsymbol{S} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_J)^\top$ with $\boldsymbol{S}(\boldsymbol{x}) = (\boldsymbol{s}_1(\boldsymbol{x}), \ldots, \boldsymbol{s}_J(\boldsymbol{x}))^\top$ and the right-hand side vector $\boldsymbol{t}$ with $\boldsymbol{t}(\boldsymbol{x}) = (t_1(\boldsymbol{x}), \ldots, t_J(\boldsymbol{x}))^\top$. Thus, a decision $\boldsymbol{x} \in \mathcal{X}$ is feasible in the chance constrained program (2) if and only if there is $\boldsymbol{\psi} \in \mathcal{Q}(\boldsymbol{x})$ such that $Q(\boldsymbol{\psi}) \geq 1 - \epsilon$. This chance constrained program is nonconvex in general as the constraint system $\boldsymbol{\psi} \in \mathcal{Q}(\boldsymbol{x})$ may involve bilinear couplings between $\boldsymbol{x}$ and some components of $\boldsymbol{\psi}$. We thus decompose $\boldsymbol{\psi}$ into a subvector $\boldsymbol{\chi}$ that contains all variables which are coupled with $\boldsymbol{x}$ and the subvector $\boldsymbol{\rho}$ that contains the remaining variables. The chance constrained program (2) can then be formulated as follows.

$$
\begin{aligned}
\text{minimize} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}, \ (\boldsymbol{\chi}, \boldsymbol{\rho}) \in \mathcal{Q}(\boldsymbol{x}) \\
& Q(\boldsymbol{\chi}, \boldsymbol{\rho}) \geq 1 - \epsilon
\end{aligned}
\tag{30}
$$

Table 1 exemplifies the definitions of $\boldsymbol{\chi}$ and $\boldsymbol{\rho}$ for some popular ambiguity sets, and Fig. 1 presents a block coordinate descent algorithm for solving (30). The algorithm requires a feasible point $\boldsymbol{x}^0$ as input. Note that the optimization problems solved in Steps 2 and 3 of the algorithm are convex and can thus be solved efficiently. For any threshold $\delta > 0$, the algorithm terminates after finitely many iterations to a partial optimum of the chance constrained program (30), that is, a feasible point $(\boldsymbol{x}^\star, \boldsymbol{\chi}^\star, \boldsymbol{\rho}^\star)$ where $(\boldsymbol{x}^\star, \boldsymbol{\rho}^\star)$ maximizes (30) for fixed $\boldsymbol{\chi}^\star$ and $Q(\boldsymbol{\chi}^\star, \boldsymbol{\rho}^\star)$ represents the worst-case probability that the system is safe under the fixed decision $\boldsymbol{x}^\star$. For a convergence proof we refer to [32].

BLOCK COORDINATE DESCENT ALGORITHM.

1. **Initialization.** For a given initial feasible solution $\boldsymbol{x}^0$, set the objective value to $f^0 \leftarrow \boldsymbol{c}^\top \boldsymbol{x}^0$ and the iteration counter to $t \leftarrow 1$.

2. **Uncertainty Quantification.** Let $(\boldsymbol{\chi}^\star, \boldsymbol{\rho}^\star)$ be an optimal solution to

$$\sup_{\boldsymbol{\chi}, \boldsymbol{\rho}} \left\{ Q(\boldsymbol{\chi}, \boldsymbol{\rho}) \,:\, (\boldsymbol{\chi}, \boldsymbol{\rho}) \in \mathcal{Q}(\boldsymbol{x}^{t-1}) \right\}$$

and set $\boldsymbol{\chi}^t \leftarrow \boldsymbol{\chi}^\star$.

3. **Policy Improvement.** Let $(\boldsymbol{x}^\star, \boldsymbol{\rho}^\star)$ be an optimal solution to

$$\inf_{\boldsymbol{x}, \boldsymbol{\rho}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \,:\, \boldsymbol{x} \in \mathcal{X}, \; Q(\boldsymbol{\chi}^t, \boldsymbol{\rho}) \geq 1 - \epsilon, \; (\boldsymbol{\chi}^t, \boldsymbol{\rho}) \in \mathcal{Q}(\boldsymbol{x}) \right\}$$

and set $\boldsymbol{x}^t \leftarrow \boldsymbol{x}^\star$, $\boldsymbol{\rho}^t \leftarrow \boldsymbol{\rho}^\star$ and $f^t \leftarrow \boldsymbol{c}^\top \boldsymbol{x}^\star$.

4. **Termination Criterion.** If $|f^t - f^{t-1}| \leq \delta$, where $\delta > 0$ is a small convergence threshold, then terminate with the solution $\boldsymbol{x}^t$. Otherwise, set $t \leftarrow t + 1$ and go to Step 2.
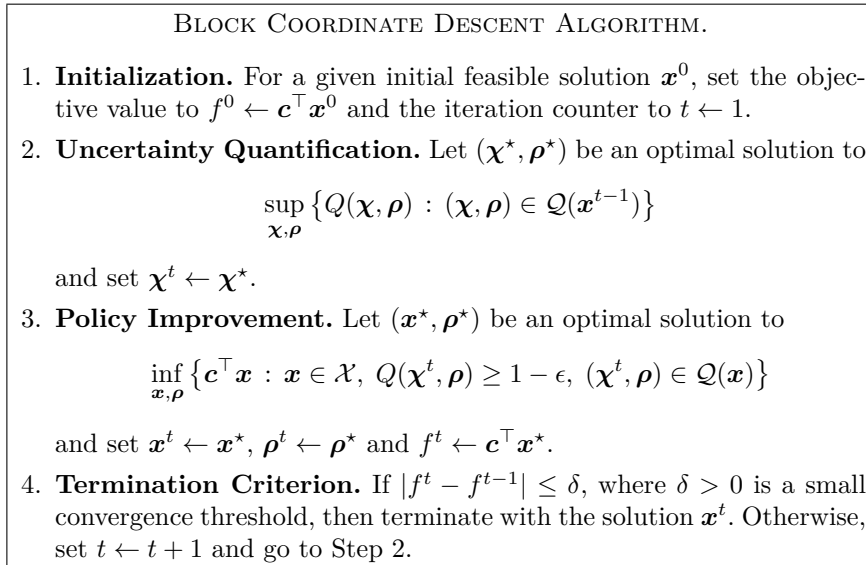
Fig. 1 Block coordinate descent algorithm for problem (30).

6 Summary

Table 2 summarizes the results of this paper. We reiterate that conic moment ambiguity sets form a strict subclass of epigraphic moment ambiguity sets and the class of epigraphic moment ambiguity sets coincides with the class of Markov ambiguity sets. Neither statement is true, however, if the corresponding moment ambiguity set is intersected with a structural ambiguity set.

References

1. A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.
2. A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.
3. A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming A*, 88(3):411–424, 2000.
4. A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. SIAM, 2001.
5. A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, 2001.
6. D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. *Available on arXiv.org*, 2013.
7. D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal of Optimization*, 15(3):780–804, 2004.
8. D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.
9. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Table 2 Summary of the results in this paper. Checks (✔) and crosses (✘) highlight tractable reformulations and intractability results, respectively.

Ambiguity set	Constraint	Moment information	Moments + symmetry	Moments + unimodality
Conic ambiguity set $\mathcal{P}^c$ $$\left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \begin{array}{l} \mathbb{E}_\mathbb{P}[A\tilde{z} + B\tilde{u}] = b, \\ \mathbb{P}[(\tilde{z}, \tilde{u}) \in \mathcal{C}_1] = 1 \end{array} \right\}$$ with $$\mathcal{C}_1 = \left\{ (z, u) \in \mathbb{R}^P \times \mathbb{R}^Q : C_1 z + D_1 u \preceq_{\mathcal{K}_1} 0 \right\}$$	**UQ:**	✔ Theorem 3	✔ Theorem 4	
	ICC:	✔ Theorem 7	✔ Theorem 9	
	JCC/FT:	✔ Theorem 12	✘ Theorem 15	
	JCC/RT:	✘ Theorem 14	✘ Theorem 15	
Epigraphic ambiguity set $\mathcal{P}^e$ $$\left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \begin{array}{l} \mathbb{E}_\mathbb{P}[A\tilde{z} + B\tilde{u}] = b, \\ \mathbb{P}[(\tilde{z}, \tilde{u}) \in \mathcal{C}_1] = 1 \end{array} \right\}$$ with $$\mathcal{C}_1 = \left\{ (z, u) \in \mathbb{R}^P \times \mathbb{R}^Q : C_1 z \preceq_{\mathcal{K}_1} d_1,\ g(z) \preceq_{\mathcal{K}_2} u \right\}$$	**UQ:**	✔ Theorem 3	✔ Theorem 4	✔ **UQ:** Theorem 5
	ICC:	✔ Theorem 7	✔ Theorem 9	✔ **ICC:** Theorem 11
	JCC/FT:	✘ Theorem 13	✘ Theorem 15	✘ **JCC/FT:** Theorem 13
	JCC/RT:	✘ Theorem 14	✘ Theorem 15	✘ **JCC/RT:** Theorem 14
Markov ambiguity set $\mathcal{P}^m$ $$\left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \begin{array}{l} \mathbb{E}_\mathbb{P}[A\tilde{z} + B\tilde{u}] = b, \\ \mathbb{P}[(\tilde{z}, \tilde{u}) \in \mathcal{C}_1] = 1 \end{array} \right\}$$ with $$\mathcal{C}_1 = \left\{ (z, u) \in \mathbb{R}^P \times \mathbb{R}^Q : C_1 z + D_1 u \preceq_{\mathcal{K}_1} d_1 \right\}$$	**UQ:**	✔ Theorem 3	✔ Theorem 4	
	ICC:	✔ Theorem 7	✔ Theorem 9	
	JCC/FT:	✘ Theorem 13	✘ Theorem 15	
	JCC/RT: (sym: **JCC/RLHS:**)	✘ Theorem 14	✘ Theorem 15	
Nested ambiguity set $\mathcal{P}^n$ $$\left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \begin{array}{l} \mathbb{E}_\mathbb{P}[A\tilde{z} + B\tilde{u}] = b, \\ \mathbb{P}[(\tilde{z}, \tilde{u}) \in \mathcal{C}_i] \in \left[\underline{p}_i, \overline{p}_i\right] \ \forall i \in \mathcal{I} \end{array} \right\}$$ with $$\mathcal{C}_i = \left\{ (z, u) \in \mathbb{R}^P \times \mathbb{R}^Q : C_i z + D_i u \preceq_{\mathcal{K}_i} d_i \right\}$$	**UQ:**	✔ Theorem 3	✔ Theorem 4	
	ICC:	✘ Theorem 8	✘ Theorem 10	
	JCC/FT:	✘ Theorem 13	✘ Theorem 15	
	JCC/RT:	✘ Theorem 14	✘ Theorem 15	

Abbreviations: UQ = Uncertainty quantification,   ICC = Individual chance constraint,
JCC/FT = Joint chance constraint with fixed technology matrix,
JCC/RT = Joint chance constraint with random technology matrix.

10. G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.
11. G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Thomson Learning, 2nd edition, 2002.
12. W. Chen, M. Sim, J. Sun, and C.-P. Teo. From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, 58(2):470–485, 2010.
13. X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.
14. E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):596–612, 2010.
15. V. DeMiguel and F. J. Nogales. Portfolio selection with robust estimation. *Operations Research*, 57(3):560–577, 2009.
16. S. W. Dharmadhikari and K. Joag-Dev. *Unimodality, Convexity, and Applications*, volume 27 of *Probability and Mathematical Statistics*. Academic Press, 1988.
17. X. V. Doan, X. Li, and K. Natarajan. Robustness to dependency in portfolio optimization using overlapping marginals. Available on Optimization Online, 2013.
18. X. V. Doan and K. Natarajan. On the complexity of nonoverlapping multivariate marginal bounds for probabilistic combinatorial optimization problems. *Operations Research*, 60(1):138–149, 2012.
19. E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming B*, 107(1–2):37–61, 2006.
20. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
21. C. F. Gauss. Theoria combinationis observationum erroribus minimis obnoxiae, pars prior. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores*, 33:321–327, 1821.
22. L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.
23. S. Han, M. Tao, U. Topcu, H. Owhadi, and R. M. Murray. Convex optimal uncertainty quantification. Available on arXiv.org, 2013.
24. G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. Distributionally robust joint chance constraints with conic dispersion measures. *Working Paper, Imperial College London and École Polytechnique Fédérale de Lausanne*, 2015.
25. G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Technical Report, Imperial College London and École Polytechnique Fédérale de Lausanne*, 2015.
26. Z. Hu and L. J. Hong. Kullback-Leibler divergence constrained distributionally robust optimization. Available on Optimization Online, 2012.
27. Z. Hu, L. J. Hong, and A. M.-C. So. Ambiguous probabilistic programs. Available on Optimization Online, 2013.
28. P. J. Huber. Robust estimation of a location parameter. *Annals of Statistics*, 53(1):73–101, 1964.
29. K. Isii. On sharpness of Tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14(1):185–197, 1962.
30. A. Jasour, N. S. Aybat, and C. Lagoa. Semidefinite programming for chance optimization over semialgebraic sets. Available on arXiv.org, 2014.
31. R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. Available on Optimization Online, 2012.
32. J. Korski, F. Pfeuffer, and K. Klamroth. Biconvex sets and optimization with biconvex functions: A survey and extensions. *Mathematical Methods of Operations Research*, 66(3):373–407, 2007.
33. S.-W. Lam, T. S. Ng, M. Sim, and J.-H. Song. Multiple objectives satisficing under uncertainty. *Operations Research*, 61(1):214–227, 2013.
34. P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Working Paper, École Polytechnique Fédérale de Lausanne*, 2015.

35. K. Natarajan, D. Pachamanova, and M. Sim. Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, 54(3):573–585, 2008.

36. A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

37. H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. Optimal uncertainty quantification. *SIAM Review*, 55(2):271–345, 2013.

38. B. P. G. Van Parys, P. J. Goulart, and D. Kuhn. Generalized Gauss inequalities via semidefinite programming. *Available on Optimization Online*, 2014.

39. B. P. G. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari. Distributionally robust control of constrained stochastic systems. Available on Optimization Online, 2013.

40. G. Pflug, A. Pichler, and D. Wozabal. The $1/n$ investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417, 2012.

41. I. Popescu. An SDP approach to optimal moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 50(3):632–657, 2005.

42. S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models*. Wiley, 1991.

43. A. Shapiro. On duality theory of conic linear problems. In *Semi-Infinite Programming*, chapter 7, pages 135–165. Kluwer Academic Publishers, 2001.

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

45. H. Sun and H. Xu. Asymptotic convergence analysis for distributional robust optimization and equilibrium problems. Available on Optimization Online, 2013.

46. L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64, 2007.

47. W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2013.

48. W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

49. H. Xu, C. Caramanis, and S. Mannor. Optimization under probabilistic envelope constraints. *Operations Research*, 60(3):682–699, 2012.

50. İ. Yanıkoğlu and D. Den Hertog. Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing*, 25(4):666–681, 2013.

51. S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming A*, 137(1–2):167–198, 2013.

52. S. Zymler, D. Kuhn, and B. Rustem. Worst-case value-at-risk of non-linear portfolios. *Management Science*, 59(1):172–188, 2013.

53. J. Žáčková. On minimax solutions of stochastic linear programming problems. *Časopis pro pěstování matematiky*, 91(4):423–430, 1966.

Appendix: Proofs

For the proof of Theorem 1, we recall that the strongly $\mathcal{NP}$-hard 0/1 Integer Programming (IP) problem [20] is defined as follows.

0/1 INTEGER PROGRAMMING.

Instance. Given are $\boldsymbol{E} \in \mathbb{Z}^{M \times P}$ and $\boldsymbol{f} \in \mathbb{Z}^{M}$.

Question. Is there a vector $\boldsymbol{y} \in \{0,1\}^{P}$ such that $\boldsymbol{E}\boldsymbol{y} \leq \boldsymbol{f}$?

Assume that $\boldsymbol{y}' \in [0,1]^{P}$ constitutes a fractional vector that satisfies $\boldsymbol{E}\boldsymbol{y}' \leq \boldsymbol{f}$. We now show that we can obtain an integral vector $\boldsymbol{y} \in \{0,1\}^{P}$ that satisfies $\boldsymbol{E}\boldsymbol{y} \leq \boldsymbol{f}$ by rounding $\boldsymbol{y}'$ if its components are 'close enough' to 0 or 1.

Lemma 1 *Assume that $\boldsymbol{y}' \in ([0,\epsilon] \cup (1-\epsilon,1])^{P}$ satisfies $\boldsymbol{E}\boldsymbol{y}' \leq \boldsymbol{f}$ for $0 < \epsilon < \min_m \left\{ \left(\sum_p |E_{mp}| \right)^{-1} \right\}$. Then $\boldsymbol{E}\boldsymbol{y} \leq \boldsymbol{f}$ for $\boldsymbol{y} \in \{0,1\}^{P}$, where $y_p = 1$ if $y'_p > 1 - \epsilon$ and $y_p = 0$ otherwise.*

Remark 2 A proof of Lemma 1 can be found in [47,48]. To keep the paper self-contained, we repeat the proof here.

Proof of Lemma 1. We have that $\boldsymbol{e}_m^{\top}\boldsymbol{y} \leq \boldsymbol{e}_m^{\top}\boldsymbol{y}' + \sum_p |E_{mp}| \epsilon < \boldsymbol{e}_m^{\top}\boldsymbol{y}' + 1 \leq f_m + 1$ for all $m \in \{1, \ldots, M\}$, where $\boldsymbol{e}_m^{\top}$ denotes the m-th row of $\boldsymbol{E}$. Due to the integrality of $\boldsymbol{E}$, $\boldsymbol{f}$ and $\boldsymbol{y}$, we thus conclude that $\boldsymbol{E}\boldsymbol{y} \leq \boldsymbol{f}$. $\qquad\square$

Proof of Theorem 1. Fix an instance $(\boldsymbol{E}, \boldsymbol{f})$ of the IP problem and assume without loss of generality that the LP relaxation of $(\boldsymbol{E}, \boldsymbol{f})$ has a solution, that is, there is $\boldsymbol{y} \in [0,1]^{P}$ such that $\boldsymbol{E}\boldsymbol{y} \leq \boldsymbol{f}$. Consider the following instance of the quadratic uncertainty quantification problem (3), where $\epsilon \in (0, 1/2)$ is chosen as prescribed in Lemma 1:

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left[\left\| \tilde{\boldsymbol{z}} - \frac{1}{2}\mathbf{e} \right\|_2 \leq \sqrt{\frac{P}{4} - \epsilon(1 - \epsilon)} \right], \tag{31}$$

where

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) \,:\, \mathbb{P}\left[\tilde{\boldsymbol{z}} \in [0,1]^{P}, \;\; \boldsymbol{E}\tilde{\boldsymbol{z}} \leq \boldsymbol{f} \right] = 1 \right\}.$$

We want to show that (31) evaluates to 0 if and only if the IP feasibility problem has an affirmative answer. To this end, assume first that (31) evaluates to 0. One readily verifies that the infimum in (31) must be attained by a probability distribution $\mathbb{P} \in \mathcal{P}$. By construction, $\mathbb{P}$ places all probability mass on realizations $\boldsymbol{z} \in \mathbb{R}^P$ that satisfy $\boldsymbol{z} \in [0,1]^{P}$, $\boldsymbol{E}\boldsymbol{z} \leq \boldsymbol{f}$ and $\left\| \boldsymbol{z} - \frac{1}{2}\mathbf{e} \right\|_2 > \sqrt{P/4 - \epsilon(1 - \epsilon)}$, that is, $z_p \in [0,\epsilon) \cup (1-\epsilon,1]$ for all $p = 1, \ldots, P$. We can then use Lemma 1 to round any such $\boldsymbol{z}$ to a solution of the IP problem.

Assume now that the instance of the IP problem is feasible, that is, there is $\boldsymbol{z} \in \{0,1\}^{P}$ such that $\boldsymbol{E}\boldsymbol{z} \leq \boldsymbol{f}$. By construction, $\boldsymbol{z}$ satisfies $\left\| \boldsymbol{z} - \frac{1}{2}\mathbf{e} \right\|_2 =$

$\sqrt{P/4} > \sqrt{P/4 - \epsilon(1-\epsilon)}$. Thus, we have $\delta_{\boldsymbol{z}} \in \mathcal{P}$, where $\delta_{\boldsymbol{z}}$ represents the Dirac distribution that concentrates unit mass at $\boldsymbol{z}$, and we observe that

$$
\begin{aligned}
0 \;\le\; & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left[\left\|\tilde{\boldsymbol{z}} - \frac{1}{2}\mathbf{e}\right\|_2 \le \sqrt{\frac{P}{4} - \epsilon(1-\epsilon)}\right] \\
\;\le\; & \mathbb{P}_{\boldsymbol{\delta_z}}\left[\left\|\tilde{\boldsymbol{z}} - \frac{1}{2}\mathbf{e}\right\|_2 \le \sqrt{\frac{P}{4} - \epsilon(1-\epsilon)}\right] \;=\; 0,
\end{aligned}
$$

that is, problem (31) indeed evaluates to 0.   □

In the remainder, we denote by $\mathcal{D}(i) = \{i' \in \mathcal{I} : \mathcal{C}_{i'} \Subset \mathcal{C}_i\}$ the set of all strict subsets (descendants) of the confidence set $\mathcal{C}_i$, and we define $\overline{\mathcal{C}}_i = \mathcal{C}_i \setminus \bigcup_{i' \in \mathcal{D}(i)} \mathcal{C}_{i'}$, $i \in \mathcal{I}$ as the confidence set $\mathcal{C}_i$ stripped of all its descendants.

Proof of Theorem 3. The uncertainty quantification problem (1) over the nested moment ambiguity set (4) coincides with the optimal value of the following problem, where $\mathcal{M}_+(\mathbb{R}^P \times \mathbb{R}^Q)$ denotes the cone of all nonnegative measures on $\mathbb{R}^P \times \mathbb{R}^Q$.

$$
\begin{aligned}
\inf \quad & \int_{\mathcal{C}_I} \mathbb{I}_{[\boldsymbol{S}\boldsymbol{z} \le \boldsymbol{t}]} \, \mathrm{d}\mu(\boldsymbol{z}, \boldsymbol{u}) \\
\text{s.t.} \quad & \mu \in \mathcal{M}_+(\mathbb{R}^P \times \mathbb{R}^Q) \\
& \int_{\mathcal{C}_I} [\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u}] \, \mathrm{d}\mu(\boldsymbol{z}, \boldsymbol{u}) = \boldsymbol{b} \\
& \left.\begin{aligned}
\int_{\mathcal{C}_I} \mathbb{I}_{[(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i]} \, \mathrm{d}\mu(\boldsymbol{z}, \boldsymbol{u}) \ge \underline{p}_i \\
\int_{\mathcal{C}_I} \mathbb{I}_{[(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i]} \, \mathrm{d}\mu(\boldsymbol{z}, \boldsymbol{u}) \le \overline{p}_i
\end{aligned}\right\} \forall i \in \mathcal{I}
\end{aligned}
\tag{32}
$$

Note that $\underline{p}_I = \overline{p}_I = 1$ due to condition (**B**). Hence, every measure μ feasible in (32) constitutes a probability measure $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ supported on $\mathcal{C}_I$.

The moment problem (32) admits the following semi-infinite dual problem.

$$
\begin{aligned}
\sup \quad & \boldsymbol{b}^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} \underline{p}_i \lambda_i - \overline{p}_i \kappa_i \\
\text{s.t.} \quad & \boldsymbol{\gamma} \in \mathbb{R}^K, \; \boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}^I_+ \\
& (\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i)\mathbb{I}_{[(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i]} \le \mathbb{I}_{[\boldsymbol{S}\boldsymbol{z} \le \boldsymbol{t}]} \quad \forall(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_I
\end{aligned}
\tag{33}
$$

Strong duality holds due to the Slater condition (**S**), see [29, 43]. By expanding the right-hand side indicator function as the pointwise minimum of indicator functions of halfspaces, the semi-infinite constraint in (33) can be split up into the following system of semi-infinite constraints.

$$
\left.\begin{aligned}
(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i)\mathbb{I}_{[(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i]} \le 1 \\
(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i)\mathbb{I}_{[(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i]} \le 0 \;\; \forall j \in \mathcal{J} : \boldsymbol{s}_j^\top \boldsymbol{z} > t_j
\end{aligned}\right\} \forall(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_I.
$$

The nesting condition (**N**) implies that we can partition the support set $\mathcal{C}_I$ into I nonempty disjoint sets, $i \in \mathcal{I}$. We may thus reformulate the above semi-infinite constraints as

$$
\left.
\begin{aligned}
(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} + \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) \leq 1 \\
(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} + \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) \leq 0 \ \ \forall j \in \mathcal{J} \ : \ \boldsymbol{s}_j^\top \boldsymbol{z} > t_j
\end{aligned}
\right\}
\begin{aligned}
& \forall i \in \mathcal{I}, \\
& \forall (\boldsymbol{z}, \boldsymbol{u}) \in \overline{\mathcal{C}}_i.
\end{aligned}
\tag{34}
$$

As the constraint left-hand sides are linear in $(\boldsymbol{z}, \boldsymbol{u})$, we may replace the sets $\overline{\mathcal{C}}_i$ and $\{(\boldsymbol{z}, \boldsymbol{u}) \in \overline{\mathcal{C}}_i : \boldsymbol{s}_j^\top \boldsymbol{z} > t_j\}$ in (34) with their convex hull $\mathrm{co}\,(\overline{\mathcal{C}}_i)$ and $\mathrm{co}\,(\{(\boldsymbol{z}, \boldsymbol{u}) \in \overline{\mathcal{C}}_i : \boldsymbol{s}_j^\top \boldsymbol{z} > t_j\})$. Due to the nesting condition (**N**), these convex hulls coincide with the sets $\mathcal{C}_i$ and $\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i : \boldsymbol{s}_j^\top \boldsymbol{z} > t_j\}$, respectively. By strong conic duality, which holds since $\mathcal{C}_i$ is essentially strictly feasible, the first semi-infinite constraint in (34) is then satisfied for $i \in \mathcal{I}$ if and only if there exists $\boldsymbol{\phi}_i \in \mathcal{K}_i^\star$ with

$$
\sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) + \boldsymbol{d}_i^\top \boldsymbol{\phi}_i \leq 1, \ \ \boldsymbol{A}^\top \boldsymbol{\gamma} = \boldsymbol{C}_i^\top \boldsymbol{\phi}_i, \ \ \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_i^\top \boldsymbol{\phi}_i. \tag{35}
$$

Note that for $(i, j) \in \mathcal{I} \times \mathcal{J}$, the second semi-infinite constraint in (34) is void and can thus be omitted whenever $(i, j) \notin \mathcal{L}$. For $(i, j) \in \mathcal{L}$, however, we can replace the strict inequality in (34) with a weak one since the expression

$$
(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} + \sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'})
$$

is continuous in $(\boldsymbol{z}, \boldsymbol{u})$ and the set $\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i : \boldsymbol{s}_j^\top \boldsymbol{z} > t_j\}$ is nonempty. By strong conic duality, which holds since $\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i : \boldsymbol{s}_j^\top \boldsymbol{z} > t_j\}$ is essentially strictly feasible, we then conclude that the constraint is satisfied if and only if there exist $\boldsymbol{\psi}_{ij} \in \mathcal{K}_i^\star$ and $\tau_{ij} \in \mathbb{R}_+$ with

$$
\sum_{i' \in \mathcal{A}(i)} (\lambda_{i'} - \kappa_{i'}) + \boldsymbol{d}_i^\top \boldsymbol{\psi}_{ij} - \tau_{ij} t_j \leq 0, \ \ \boldsymbol{A}^\top \boldsymbol{\gamma} + \tau_{ij} \boldsymbol{s}_j = \boldsymbol{C}_i^\top \boldsymbol{\psi}_{ij}, \ \ \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_i^\top \boldsymbol{\psi}_{ij}. \tag{36}
$$

The claim now follows from substituting (35) and (36) into (33). $\qquad \square$

The proof of Theorem 4 relies on the following two auxiliary results.

Lemma 2 *If (**B**), (**D**) and (**N**) hold, then for any $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i$, $i \in \mathcal{I}$, there exist $(\boldsymbol{z}, \boldsymbol{u}_1), (\boldsymbol{z}, \boldsymbol{u}_2) \in \overline{\mathcal{C}}_i$ and $\lambda \in [0, 1]$ with $(\boldsymbol{z}, \boldsymbol{u}) = \lambda(\boldsymbol{z}, \boldsymbol{u}_1) + (1 - \lambda)(\boldsymbol{z}, \boldsymbol{u}_2)$.*

Proof. Under condition (**D**), the statement trivially holds if $I = 1$. Assume now that $Q > 0$ and $\mathcal{C}_I$ is bounded, and fix a confidence set $\mathcal{C}_i$, $i \in \mathcal{I}$, and

parameter realization $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_i$. Under conditions (**B**) and (**D**), the auxiliary variable u_1 exists and is uniformly bounded across all $(\boldsymbol{z}', \boldsymbol{u}') \in \mathcal{C}_i \subseteq \mathcal{C}_I$. Thus,

$$\underline{u}_1 = \max_{\theta \in \mathbb{R}_+} \{\theta : (\boldsymbol{z}, \boldsymbol{u} - \theta \mathbf{e}_1) \in \mathcal{C}_i\} \quad \text{and} \quad \overline{u}_1 = \max_{\theta \in \mathbb{R}_+} \{\theta : (\boldsymbol{z}, \boldsymbol{u} + \theta \mathbf{e}_1) \in \mathcal{C}_i\}$$

both exist and are finite. The nesting condition (**N**) implies that $\boldsymbol{u}_1 = (\boldsymbol{z}, \boldsymbol{u} - \underline{u}_1 \mathbf{e}_1)$ and $\boldsymbol{u}_2 = (\boldsymbol{z}, \boldsymbol{u} + \overline{u}_1 \mathbf{e}_1)$ belong to the boundary of $\mathcal{C}_i$ and are thus elements of $\overline{\mathcal{C}}_i$. The claim now follows by setting $\lambda = \overline{u}_1 / (\underline{u}_1 + \overline{u}_1)$. $\qquad \square$

Lemma 3 *Let Z be a convex subset of $\mathbb{R}^P$ and define the sets*

$$\mathcal{D}_\mathbf{i} = \left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (\boldsymbol{z}, \boldsymbol{u}^+) \in \mathcal{C}_{i^+}, \ (-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_{i^-} \right\},$$
$$\overline{\mathcal{D}}_\mathbf{i} = \left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (\boldsymbol{z}, \boldsymbol{u}^+) \in \overline{\mathcal{C}}_{i^+}, \ (-\boldsymbol{z}, \boldsymbol{u}^-) \in \overline{\mathcal{C}}_{i^-} \right\}$$

*for $\mathbf{i} = (i^+, i^-) \in \mathcal{I}^2$. If (**B**), (**D**) and (**N**) hold, then $\mathrm{co}(\overline{\mathcal{D}}_\mathbf{i}) = \mathcal{D}_\mathbf{i}, \ \mathbf{i} \in \mathcal{I}^2$.*

Proof. We note that for a fixed $\mathbf{i} \in \mathcal{I}^2$, we have that

$$
\begin{aligned}
\mathrm{co}(\overline{\mathcal{D}}_\mathbf{i}) \ = \ & \mathrm{co}\Big(\left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (\boldsymbol{z}, \boldsymbol{u}^+) \in \overline{\mathcal{C}}_{i^+} \right\} \cap \\
& \qquad \left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (-\boldsymbol{z}, \boldsymbol{u}^-) \in \overline{\mathcal{C}}_{i^-} \right\} \Big) \\
\subseteq \ & \mathrm{co}\Big(\left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (\boldsymbol{z}, \boldsymbol{u}^+) \in \overline{\mathcal{C}}_{i^+} \right\} \Big) \cap \\
& \mathrm{co}\Big(\left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in Z \times \mathbb{R}^Q \times \mathbb{R}^Q : (-\boldsymbol{z}, \boldsymbol{u}^-) \in \overline{\mathcal{C}}_{i^-} \right\} \Big) \ = \ \mathcal{D}_\mathbf{i},
\end{aligned}
$$

where the last inclusion follows from the relation $\mathrm{co}(\overline{\mathcal{C}}_i) = \mathcal{C}_i$ for all $i \in \mathcal{I}$, which is implied by the nesting condition (**N**).

To prove that $\mathcal{D}_\mathbf{i} \subseteq \mathrm{co}(\overline{\mathcal{D}}_\mathbf{i})$, fix some $(\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathcal{D}_\mathbf{i}$. Thus, $(\boldsymbol{z}, \boldsymbol{u}^+) \in \mathcal{C}_{i^+}$ and $(-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_{i^-}$. By Lemma 2 there exist $\boldsymbol{u}_1^+, \boldsymbol{u}_2^+, \boldsymbol{u}_1^-, \boldsymbol{u}_2^- \in \mathbb{R}^Q$ and $\lambda^+, \lambda^- \in [0, 1]$ with $(\boldsymbol{z}, \boldsymbol{u}_1^+), (\boldsymbol{z}, \boldsymbol{u}_2^+) \in \overline{\mathcal{C}}_{i^+}$ and $(-\boldsymbol{z}, \boldsymbol{u}_1^-), (-\boldsymbol{z}, \boldsymbol{u}_2^-) \in \overline{\mathcal{C}}_{i^-}$ with

$$\boldsymbol{u}^+ = \lambda^+ \boldsymbol{u}_1^+ + (1 - \lambda^+) \boldsymbol{u}_2^+ \quad \text{and} \quad \boldsymbol{u}^- = \lambda^- \boldsymbol{u}_1^- + (1 - \lambda^-) \boldsymbol{u}_2^-.$$

By construction, the vectors $(\boldsymbol{z}, \boldsymbol{u}_1^+, \boldsymbol{u}_1^-), (\boldsymbol{z}, \boldsymbol{u}_1^+, \boldsymbol{u}_2^-), (\boldsymbol{z}, \boldsymbol{u}_2^+, \boldsymbol{u}_1^-)$ and $(\boldsymbol{z}, \boldsymbol{u}_2^+, \boldsymbol{u}_2^-)$ belong to $\overline{\mathcal{D}}_\mathbf{i}$, and we have

$$
\begin{aligned}
(\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) = \ & \lambda^+ \lambda^- (\boldsymbol{z}, \boldsymbol{u}_1^+, \boldsymbol{u}_1^-) + \lambda^+ (1 - \lambda^-)(\boldsymbol{z}, \boldsymbol{u}_1^+, \boldsymbol{u}_2^-) \\
& + (1 - \lambda^+) \lambda^- (\boldsymbol{z}, \boldsymbol{u}_2^+, \boldsymbol{u}_1^-) + (1 - \lambda^+)(1 - \lambda^-)(\boldsymbol{z}, \boldsymbol{u}_2^+, \boldsymbol{u}_2^-).
\end{aligned}
$$

This implies that $(\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathrm{co}(\overline{\mathcal{D}}_\mathbf{i})$. As $(\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathcal{D}_\mathbf{i}$ was chosen arbitrarily, we have $\mathcal{D}_\mathbf{i} \subseteq \mathrm{co}(\overline{\mathcal{D}}_\mathbf{i})$. We may thus conclude that $\mathrm{co}(\overline{\mathcal{D}}_\mathbf{i}) = \mathcal{D}_\mathbf{i}$. $\qquad \square$

Proof of Theorem 4.   The general duality theory for moment problems allows us to express the worst-case probability (1) as the optimal value of

$$
\sup \quad \boldsymbol{b}^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} \underline{p}_i \lambda_i - \overline{p}_i \kappa_i
$$

$$
\text{s.t.} \quad \boldsymbol{\gamma} \in \mathbb{R}^K, \; \boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}^I_+
$$

$$
\mathbb{E}_\mathbb{P} \left[(\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}})^\top \boldsymbol{\gamma} + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i) \mathbb{I}_{[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}_i]} \right] \leq \mathbb{E}_\mathbb{P} \left[\mathbb{I}_{[\boldsymbol{S}\tilde{\boldsymbol{z}} \leq \boldsymbol{t}]} \right] \quad \forall \mathbb{P} \in \mathcal{P}^s(\mathcal{C}_I),
$$

$$
\tag{37}
$$

where $\mathcal{P}^s(\mathcal{C}_I)$ contains all distributions $\mathbb{P} \in \mathcal{P}_0(\mathcal{C}_I)$ for which $\Pi_{\tilde{\boldsymbol{z}}}\mathbb{P}$ is symmetric around $\boldsymbol{m} = \boldsymbol{0}$, see [43, § 3] or [41, Theorem 3.1]. Strong duality holds due to the Slater condition (**S**). The extreme points of the structural ambiguity set $\mathcal{P}^s(\mathcal{C}_I)$ are the discrete distributions $\frac{1}{2}\delta_{(\boldsymbol{z},\boldsymbol{u}^+)} + \frac{1}{2}\delta_{(-\boldsymbol{z},\boldsymbol{u}^-)}$ where $(\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q$ with $(\boldsymbol{z}, \boldsymbol{u}^+) \in \mathcal{C}_I$ and $(-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_I$. The semi-infinite constraint in (37) is thus equivalent to

$$
\boldsymbol{\gamma}^\top \boldsymbol{B}(\boldsymbol{u}^+ + \boldsymbol{u}^-) + \sum_{i \in \mathcal{I}} (\lambda_i - \kappa_i)(\mathbb{I}_{[(\boldsymbol{z},\boldsymbol{u}^+) \in \mathcal{C}_i]} + \mathbb{I}_{[(-\boldsymbol{z},\boldsymbol{u}^-) \in \mathcal{C}_i]}) \leq \mathbb{I}_{[\boldsymbol{S}\boldsymbol{z} \leq \boldsymbol{t}]} + \mathbb{I}_{[-\boldsymbol{S}\boldsymbol{z} \leq \boldsymbol{t}]}
$$

$$
\forall (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : (\boldsymbol{z}, \boldsymbol{u}^+) \in \mathcal{C}_I, \; (-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_I.
$$

As the support $\mathcal{C}_I$ can be partitioned into the subsets $\overline{\mathcal{C}}_i$, $i \in \mathcal{I}$, we may reformulate the above semi-infinite constraint as

$$
\boldsymbol{\gamma}^\top \boldsymbol{B}(\boldsymbol{u}^+ + \boldsymbol{u}^-) + \sum_{i \in \mathcal{A}(i^+)} (\lambda_i - \kappa_i) + \sum_{i \in \mathcal{A}(i^-)} (\lambda_i - \kappa_i) \leq \mathbb{I}_{[\boldsymbol{S}\boldsymbol{z} \leq \boldsymbol{t}]} + \mathbb{I}_{[-\boldsymbol{S}\boldsymbol{z} \leq \boldsymbol{t}]}
$$

$$
\forall \mathbf{i} \in \mathcal{I}^2, \; \forall (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : (\boldsymbol{z}, \boldsymbol{u}^+) \in \overline{\mathcal{C}}_{i^+}, \; (-\boldsymbol{z}, \boldsymbol{u}^-) \in \overline{\mathcal{C}}_{i^-}.
$$

By expanding the right-hand side indicator functions as the pointwise minima of indicator functions on halfspaces, the semi-infinite constraint reduces to

$$
\boldsymbol{\gamma}^\top \boldsymbol{B}(\boldsymbol{u}^+ + \boldsymbol{u}^-) + \sum_{i \in \mathcal{A}(i^+)} (\lambda_i - \kappa_i) + \sum_{i \in \mathcal{A}(i^-)} (\lambda_i - \kappa_i) \leq w_{\mathbf{j}}
$$

$$
\forall (\mathbf{i}, \mathbf{j}) \in \mathcal{I}^2 \times (\mathcal{J} \cup \{0\})^2, \; \forall (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \overline{\mathcal{D}}_{\mathbf{ij}},
$$

$$
\tag{38}
$$

where

$$
\overline{\mathcal{D}}_{\mathbf{ij}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q \; : \; \begin{array}{r} (\boldsymbol{z}, \boldsymbol{u}^+) \in \overline{\mathcal{C}}_{i^+}, \; (-\boldsymbol{z}, \boldsymbol{u}^-) \in \overline{\mathcal{C}}_{i^-} \\ j^+ > 0 \implies \boldsymbol{s}_{j^+}^\top \boldsymbol{z} > t_{j^+} \\ j^- > 0 \implies -\boldsymbol{s}_{j^-}^\top \boldsymbol{z} > t_{j^-} \end{array} \right\}
$$

for all $\mathbf{i} \in \mathcal{I}^2$. As the semi-infinite constraint (38) is affine in the uncertain parameters $(\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-)$, we can use Lemma 3 to replace the parameter range $\overline{\mathcal{D}}_{\mathbf{ij}}$ with its convex hull $\mathcal{D}_{\mathbf{ij}}$, which is defined in (20). If $\mathcal{D}_{\mathbf{ij}} = \emptyset$, then the subconstraint indexed by $(\mathbf{i}, \mathbf{j})$ is vacuously satisfied. Conversely, if $\mathcal{D}_{\mathbf{ij}} \neq \emptyset$, then we can use a standard continuity argument to replace $\mathcal{D}_{\mathbf{ij}}$ with its closure

$$
\mathrm{cl}(\mathcal{D}_{\mathbf{ij}}) = \left\{ (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q \; : \; \begin{array}{r} (\boldsymbol{z}, \boldsymbol{u}^+) \in \mathcal{C}_{i^+}, \; (-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_{i^-} \\ j^+ > 0 \implies \boldsymbol{s}_{j^+}^\top \boldsymbol{z} \geq t_{j^+} \\ j^- > 0 \implies -\boldsymbol{s}_{j^-}^\top \boldsymbol{z} \geq t_{j^-} \end{array} \right\}.
$$

The feasibility condition (**F**) implies that $\mathrm{cl}(\mathcal{D}_{\mathbf{ij}})$ is essentially strictly feasible, and the conic Farkas lemma [9, Example 2.26] allows us to reformulate (38) as

$$
\begin{aligned}
&\forall (\mathbf{i}, \mathbf{j}) \in \mathcal{L} \; \exists \boldsymbol{\psi}_{\mathbf{ij}}^+ \in \mathcal{K}_{i+}^\star, \; \boldsymbol{\psi}_{\mathbf{ij}-}^- \in \mathcal{K}_{i-}^\star, \; \chi_{\mathbf{ij}}^+, \chi_{\mathbf{ij}}^- \in \mathbb{R}_+ : \\
&\boldsymbol{\gamma}^\top \boldsymbol{B} (\boldsymbol{u}^+ + \boldsymbol{u}^-) + \sum_{i \in \mathcal{A}(i^+)} (\lambda_i - \kappa_i) + \sum_{i \in \mathcal{A}(i^-)} (\lambda_i - \kappa_i) \\
&\quad - \boldsymbol{\psi}_{\mathbf{ij}}^{+\top} (\boldsymbol{C}_{i+} \boldsymbol{z} + \boldsymbol{D}_{i+} \boldsymbol{u}^+ - \boldsymbol{d}_{i+}) - \boldsymbol{\psi}_{\mathbf{ij}}^{-\top} (-\boldsymbol{C}_{i-} \boldsymbol{z} + \boldsymbol{D}_{i-} \boldsymbol{u}^- - \boldsymbol{d}_{i-}) \\
&\quad + \chi_{\mathbf{ij}}^+ (\boldsymbol{s}_{j+}^\top \boldsymbol{z} - t_{j+}) + \chi_{\mathbf{ij}}^- (-\boldsymbol{s}_{j-}^\top \boldsymbol{z} - t_{j-}) \le w_{\mathbf{j}} \quad \forall (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q.
\end{aligned}
$$

The claim now follows from the observation that an affine function is bounded from above uniformly on the entire space if and only if it is bounded above at the origin and if its gradient vanishes.   □

Proof of Theorem 5. The duality theory for moment problems allows us to express the worst-case probability (1) as the optimal value of the problem

$$
\begin{aligned}
\sup \quad & \beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \beta \in \mathbb{R}, \; \boldsymbol{\gamma} \in \mathbb{R}^K \\
& \mathbb{E}_{\mathbb{P}} \left[\beta + (\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}})^\top \boldsymbol{\gamma} \right] \le \mathbb{E}_{\mathbb{P}} \left[\mathbb{I}_{[\boldsymbol{S}\tilde{\boldsymbol{z}} \le \boldsymbol{t}]} \right] \quad \forall \mathbb{P} \in \mathcal{P}^\alpha(\mathcal{C}_1),
\end{aligned} \tag{39}
$$

where $\mathcal{P}^\alpha(\mathcal{C}_1)$ contains all distributions $\mathbb{P} \in \mathcal{P}_0(\mathcal{C}_1)$ for which $\Pi_{\tilde{\boldsymbol{z}}} \mathbb{P}$ is α-unimodal around $\boldsymbol{m} = \boldsymbol{0}$. Strong duality holds due to the Slater condition (**S**), which simplifies in the current setting to the requirement that

$$
\mathcal{C}_1^\alpha = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \boldsymbol{C}_1 \boldsymbol{z} \preccurlyeq_{\mathcal{K}_1} \boldsymbol{d}_1, \; \boldsymbol{g}^\alpha(\boldsymbol{z}) \preccurlyeq_{\mathcal{K}_2} \boldsymbol{u} \right\}
$$

contains the vector $\boldsymbol{b}$ in its interior, where $\boldsymbol{g}^\alpha(\boldsymbol{z}) = \int_0^1 \boldsymbol{g}(t\boldsymbol{z}) \alpha t^{\alpha-1} \mathrm{d}t$. Any distribution in $\mathcal{P}^\alpha(\mathcal{C}_1)$ can be represented as a mixture of distributions $\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}$ parameterized by $(\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^P \times \mathbb{R}^Q$ with the following properties:

(P1) $\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}$ is supported on $\mathcal{C}_1$;
(P2) $\Pi_{\tilde{\boldsymbol{z}}} \mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}$ is the radial α-unimodal distribution on the interval $[\boldsymbol{0}, \boldsymbol{z}]$;
(P3) $\mathbb{E}_{\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}} [\tilde{\boldsymbol{u}}] = \boldsymbol{u}$.

Note that if a distribution $\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}$ satisfies (P1) and (P2) for some $(\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^P \times \mathbb{R}^Q$, then $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha$ as $\mathbb{E}_{\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}} [\boldsymbol{g}(\tilde{\boldsymbol{z}})] = \boldsymbol{g}^\alpha(\boldsymbol{z})$ due to (P2). Conversely, for every $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha$ there exists a distribution $\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}$ that satisfies (P1) and (P2). The constraint in (39) is therefore equivalent to

$$
\mathbb{E}_{\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}} \left[\beta + (\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}})^\top \boldsymbol{\gamma} \right] \le \mathbb{E}_{\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}} \left[\mathbb{I}_{[\boldsymbol{S}\tilde{\boldsymbol{z}} \le \boldsymbol{t}]} \right] \qquad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha. \tag{40}
$$

We emphasize that $\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}$ is not uniquely determined by $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha$, but all distributions $\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}$ satisfying (P1)–(P3) for a given $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha$ have the same effect on constraint (40). Note that $\mathbb{E}_{\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}} [\tilde{\boldsymbol{z}}] = \frac{\alpha}{\alpha+1} \boldsymbol{z}$ and $\mathbb{E}_{\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}} [\tilde{\boldsymbol{u}}] = \boldsymbol{u}$ by (P2) and (P3), respectively. As $\boldsymbol{t} \ge \boldsymbol{0}$, (P2) further implies

$$
\mathbb{E}_{\mathbb{V}_{(\boldsymbol{z}, \boldsymbol{u})}} \left[\mathbb{I}_{[\boldsymbol{S}\tilde{\boldsymbol{z}} \le \boldsymbol{t}]} \right] = \int_0^1 \mathbb{I}_{[\boldsymbol{S}\ell\boldsymbol{z} \le \boldsymbol{t}]} \alpha \ell^{\alpha-1} \mathrm{d}\ell = \int_0^{F(\boldsymbol{z})} \alpha \ell^{\alpha-1} \mathrm{d}\ell = F(\boldsymbol{z})^\alpha,
$$

where $F(\boldsymbol{z}) = \min\{1, \inf\{t_j/\boldsymbol{s}_j^\top \boldsymbol{z} \,:\, j \in \mathcal{J} \,:\, \boldsymbol{s}_j^\top \boldsymbol{z} > 0\}\}$. Therefore, constraint (40) can be decomposed into the following semi-infinite constraints.

$$
\left.
\begin{aligned}
&\beta + (\tfrac{\alpha}{\alpha+1}\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} \leq 1 \\
&\beta + (\tfrac{\alpha}{\alpha+1}\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} \leq (t_j/\boldsymbol{s}_j^\top \boldsymbol{z})^\alpha \;\; \forall j \in \mathcal{L} \,:\, \boldsymbol{s}_j^\top \boldsymbol{z} > 0
\end{aligned}
\right\} \;\; \forall (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha
$$

(41)

By the conic Farkas lemma, which applies since $\mathcal{C}_1^\alpha$ inherits essential strict feasibility from $\mathcal{C}_1$, the first constraint in (41) holds if and only if there is $\boldsymbol{\phi} \in \mathcal{K}_2^\star$, $\boldsymbol{\psi}_0 \in \mathcal{K}_1^\star$ with

$$
\beta + \left(\tfrac{\alpha}{\alpha+1}\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u}\right)^\top \boldsymbol{\gamma} + (\boldsymbol{u} - \boldsymbol{g}^\alpha(\boldsymbol{z}))^\top \boldsymbol{\phi} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi}_0 \leq 1
$$
$$
\forall (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^P \times \mathbb{R}^Q,
$$

which is equivalent to the existence of $\boldsymbol{\psi}_0 \in \mathcal{K}_1^\star$ with $-\boldsymbol{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^\star$ and

$$
\beta + \left(\tfrac{\alpha}{\alpha+1}\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{g}^\alpha(\boldsymbol{z})\right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi}_0 \leq 1 \qquad \forall \boldsymbol{z} \in \mathbb{R}^P. \qquad (42)
$$

Note that $\boldsymbol{g}^\alpha(\boldsymbol{z})$ inherits $\mathcal{K}_2$-convexity from $\boldsymbol{g}(\boldsymbol{z})$ since $\mathcal{K}_2$ is convex. As $\boldsymbol{B}^\top \boldsymbol{\gamma} \in -\mathcal{K}_2^\star$, the left-hand side of the above constraint is thus concave in $\boldsymbol{z}$.

For $j \in \mathcal{L}$ one can use a continuity argument to show that the strict inequality $\boldsymbol{s}_j^\top \boldsymbol{z} > 0$ in (41) can be replaced by a weak inequality. By the conic Farkas lemma, which applies as $j \in \mathcal{L}$, the j-th constraint in the second line of (41) thus holds if and only if there exists $\boldsymbol{\psi}_j \in \mathcal{K}_1^\star$ with $-\boldsymbol{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^\star$ and

$$
\beta + \left(\tfrac{\alpha}{\alpha+1}\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{g}^\alpha(\boldsymbol{z})\right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi}_j \leq (t_j/\boldsymbol{s}_j^\top \boldsymbol{z})^\alpha \quad \forall \boldsymbol{z} \in \mathbb{R}^P : \boldsymbol{s}_j^\top \boldsymbol{z} \geq 0.
$$

Introducing the concave auxiliary function

$$
f_j(\boldsymbol{z}) = \beta + \left(\tfrac{\alpha}{\alpha+1}\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{g}^\alpha(\boldsymbol{z})\right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi}_j
$$

to simplify the notation, this semi-infinite constraint can be reformulated as

$$
\begin{aligned}
0 &\geq \sup_{\boldsymbol{z} \in \mathbb{R}^P, \kappa_j \in \mathbb{R}_+} \; \inf_{\tau_j \in \mathbb{R}_+} \; f_j(\boldsymbol{z}) - (t_j/\kappa_j)^\alpha + \tau_j(\boldsymbol{s}_j^\top \boldsymbol{z} - \kappa_j) \\
&= \inf_{\tau_j \in \mathbb{R}_+} \; \sup_{\boldsymbol{z} \in \mathbb{R}^P, \kappa_j \in \mathbb{R}_+} \; f_j(\boldsymbol{z}) - (t_j/\kappa_j)^\alpha + \tau_j(\boldsymbol{s}_j^\top \boldsymbol{z} - \kappa_j) \\
&= \inf_{\tau_j \in \mathbb{R}_+} \; \sup_{\boldsymbol{z} \in \mathbb{R}^P} \; f_j(\boldsymbol{z}) + \tau_j \boldsymbol{s}_j^\top \boldsymbol{z} - \tfrac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha}(\tau_j t_j)^{\frac{\alpha}{\alpha+1}},
\end{aligned}
$$

where the inner minimization over τ_j in the first line enforces the implicit constraint $\boldsymbol{s}_j^\top \boldsymbol{z} \geq \kappa_j$ for the outer maximization, while the interchange of minimization and maximization in the second line is justified by the classical saddle point theorem, which applies because $j \in \mathcal{L}$.

In summary, the j-th constraint in the second line of (41) holds if and only if there exist $\boldsymbol{\psi}_j \in \mathcal{K}_1^\star$ and $\tau_j \in \mathbb{R}_+$ with $-\boldsymbol{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^\star$ and

$$
\begin{aligned}
\beta + \left(\tfrac{\alpha}{\alpha+1} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{g}^\alpha(\boldsymbol{z}) \right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi}_j \\
+ \tau_j \boldsymbol{s}_j^\top \boldsymbol{z} - \tfrac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau_j t_j)^{\frac{\alpha}{\alpha+1}} \le 0 \quad \forall \boldsymbol{z} \in \mathbb{R}^P .
\end{aligned}
\tag{43}
$$

The claim now follows by substituting (42) and (43) into (39).   □

In order to prove Theorem 6, we first derive the following auxiliary result.

Lemma 4 *The ambiguity set*

$$
\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) \; : \; \begin{array}{l} \mathbb{P}[\tilde{\boldsymbol{z}} = \boldsymbol{0}] = 2^{-P}, \; \mathbb{P}[\tilde{\boldsymbol{z}} = \mathbf{e}] = 2^{-P} \\ \textit{the components of } \tilde{\boldsymbol{z}} \textit{ are independent} \end{array} \right\}
\tag{44}
$$

satisfies $\mathcal{P} = \{\mathbb{P}^0\}$ *for the distribution* $\mathbb{P}^0$ *which places all probability mass uniformly on the vertices of the unit hypercube in* $\mathbb{R}^P$:

$$
\mathbb{P}^0 = \frac{1}{2^P} \sum_{\boldsymbol{x} \in \{0,1\}^P} \delta_{\boldsymbol{x}} .
$$

Proof of Lemma 4. By construction, we have $\mathbb{P}^0 \in \mathcal{P}$. To show that $\mathbb{P}^0$ is the only distribution in $\mathcal{P}$, fix any distribution $\mathbb{P} \in \mathcal{P}$ and let $\rho_p^0 = \mathbb{P}[\tilde{z}_p = 0]$ and $\rho_p^1 = \mathbb{P}[\tilde{z}_p = 1]$, $p = 1, \ldots, P$. By construction, we have $\boldsymbol{0} \le \boldsymbol{\rho}^0 + \boldsymbol{\rho}^1 \le \mathbf{e}$. Since $\mathbb{P} \in \mathcal{P}$, the components of $\tilde{\boldsymbol{z}}$ are independent, and we obtain that

$$
\mathbb{P}[\tilde{\boldsymbol{z}} = \boldsymbol{0}] \; = \; \prod_{p=1}^{P} \mathbb{P}[\tilde{z}_p = 0] \; = \; \prod_{p=1}^{P} \rho_p^0 \; = \; 2^{-P}
$$

and

$$
\mathbb{P}[\tilde{\boldsymbol{z}} = \mathbf{e}] \; = \; \prod_{p=1}^{P} \mathbb{P}[\tilde{z}_p = 1] \; = \; \prod_{p=1}^{P} \rho_p^1 \; = \; 2^{-P} .
$$

We show via induction on P that $\boldsymbol{\rho}^0 = \boldsymbol{\rho}^1 = \mathbf{e}/2$ is the unique maximizer to

$$
\min \left\{ \prod_{p=1}^{P} \rho_p^0, \; \prod_{p=1}^{P} \rho_p^1 \right\}
\tag{45}
$$

over $\{(\boldsymbol{\rho}^0, \boldsymbol{\rho}^1) \in \mathbb{R}^{2P} : \boldsymbol{0} \le \boldsymbol{\rho}^0 + \boldsymbol{\rho}^1 \le \mathbf{e}\}$. The statement of the lemma then follows from the fact that $\prod_p \rho_p^0 = \prod_p \rho_p^1 = 2^{-P}$ for $\boldsymbol{\rho}^0 = \boldsymbol{\rho}^1 = \mathbf{e}/2$.

The statement trivially holds for $P = 1$. Assume now that $\boldsymbol{\rho}^0, \boldsymbol{\rho}^1 \in \mathbb{R}^{P+1}$ is a maximizer to (45) over $\{(\boldsymbol{\rho}^0, \boldsymbol{\rho}^1) \in \mathbb{R}^{2P+2} : \boldsymbol{0} \le \boldsymbol{\rho}^0 + \boldsymbol{\rho}^1 \le \mathbf{e}\}$ that does not satisfy $\boldsymbol{\rho}^0 = \boldsymbol{\rho}^1 = \mathbf{e}/2$. Without loss of generality, assume that $\rho_{P+1}^0 = \delta < 1/2$ and $\rho_{P+1}^1 = 1 - \delta$. We then have that

$$
\prod_{p=1}^{P+1} \rho_p^0 \; = \; \delta \cdot \prod_{p=1}^{P} \rho_p^0 \; \ge \; \frac{1}{2^{P+1}} \quad \Longleftrightarrow \quad \prod_{p=1}^{P} \rho_p^0 \; \ge \; \frac{1}{2^{P+1}} \cdot \frac{1}{\delta}
$$

and

$$\prod_{p=1}^{P+1} \rho_p^1 = (1-\delta) \cdot \prod_{p=1}^{P} \rho_p^1 \geq \frac{1}{2^{P+1}} \iff \prod_{p=1}^{P} \rho_p^1 \geq \frac{1}{2^{P+1}} \cdot \frac{1}{1-\delta}.$$

Consider $\hat{\boldsymbol{\rho}}^0$ and $\hat{\boldsymbol{\rho}}^1$ defined through $\hat{\rho}_p^0 = \hat{\rho}_p^1 = \sqrt{\rho_p^0 \rho_p^1}$, $p = 1, \ldots, P$. Since $\sqrt{\rho_p^0 \rho_p^1} \leq \sqrt{\rho_p^0 (1-\rho_p^0)} \leq 1/2$, we have that $\boldsymbol{0} \leq \hat{\boldsymbol{\rho}}^0 + \hat{\boldsymbol{\rho}}^1 \leq \mathbf{e}$. We now see that

$$\prod_{p=1}^{P} \hat{\rho}_p^0 = \prod_{p=1}^{P} \sqrt{\rho_p^0 \rho_p^1} = \sqrt{\prod_{p=1}^{P} \rho_p^0} \sqrt{\prod_{p=1}^{P} \rho_p^1} \geq \sqrt{\frac{1}{2^{P+1}} \cdot \frac{1}{\delta}} \sqrt{\frac{1}{2^{P+1}} \cdot \frac{1}{1-\delta}}$$

$$= \frac{1}{2^P} \sqrt{\frac{1}{4\delta(1-\delta)}} > \frac{1}{2^P}$$

for $\delta < 1/2$ and similarly $\prod_{p=1}^{P} \hat{\rho}_p^1 > 2^{-P}$, which contradicts the fact that $\boldsymbol{\rho}^0 = \boldsymbol{\rho}^1 = \mathbf{e}/2 \in \mathbb{R}^P$ is the unique maximizer to (45) over $\{(\boldsymbol{\rho}^0, \boldsymbol{\rho}^1) \in \mathbb{R}^{2P} : \boldsymbol{0} \leq \boldsymbol{\rho}^0 + \boldsymbol{\rho}^1 \leq \mathbf{e}\}$. A similar argument can be constructed if $\rho_{P+1}^0 = \delta > 1/2$, which concludes the proof. $\qquad\square$

Proof of Theorem 6. Fix an instance $(\boldsymbol{E}, \boldsymbol{f})$ of the IP problem and consider the following instance of the uncertainty quantification problem (1):

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\boldsymbol{E}\tilde{\boldsymbol{z}} \leq \boldsymbol{f}], \tag{46}$$

where the ambiguity set $\mathcal{P}$ is defined as in (44). Note that this ambiguity set satisfies the boundedness condition (**B**) and the nesting condition (**N**) since the two singleton sets $\{\boldsymbol{0}\}$ and $\{\mathbf{e}\}$ are disjoint and both contained in the interior of $\mathbb{R}^P$. Moreover, the bit-length of the probabilities $\underline{p}_i = \overline{p}_i = 2^{-P}$ in the ambiguity set $\mathcal{P}$ is bounded from above by the bit-length of $\boldsymbol{E} \in \mathbb{R}^{M \times P}$. From Lemma 4 we conclude that problem (46) evaluates to

$$\frac{1}{2^P} \sum_{\boldsymbol{x} \in \{0,1\}^P} \delta_{\boldsymbol{x}}(\boldsymbol{E}\tilde{\boldsymbol{z}} \leq \boldsymbol{f}) = \frac{1}{2^P} \left| \{\boldsymbol{x} \in \{0,1\}^P : \boldsymbol{E}\boldsymbol{x} \leq \boldsymbol{f}\} \right|,$$

that is, (46) evaluates to 0 if and only if the IP problem has a negative answer. The statement now follows since the IP problem is strongly $\mathcal{NP}$-hard. $\qquad\square$

Proof of Theorem 7. We can employ Theorem 3 to conclude that the worst-case probability of the chance constraint in (2) is given by

$$\begin{aligned}
\sup \quad & \beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \beta \in \mathbb{R}, \ \boldsymbol{\gamma} \in \mathbb{R}^K, \ \boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^\star, \ \tau \in \mathbb{R}_+ \\
& \beta + \boldsymbol{d}_1^\top \boldsymbol{\phi} \leq 1, \ \boldsymbol{A}^\top \boldsymbol{\gamma} = \boldsymbol{C}_1^\top \boldsymbol{\phi}, \ \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\phi} \\
& \left.\begin{aligned} & \beta + \boldsymbol{d}_1^\top \boldsymbol{\psi} - \tau t(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{A}^\top \boldsymbol{\gamma} + \tau \boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{C}_1^\top \boldsymbol{\psi}, \ \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\psi} \end{aligned}\right\} \text{ if } \boldsymbol{x} \in \mathcal{X}^+,
\end{aligned} \tag{47}$$

where $\mathcal{X}^+ = \left\{ \boldsymbol{x} \in \mathcal{X} : \exists (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1 \,.\, \boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} > t(\boldsymbol{x}) \right\}$ and where we have substituted $\beta = \lambda_1 - \kappa_1$.

In the remainder we will separately investigate the cases $\boldsymbol{x} \in \mathcal{X}^+$ and $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^+$. If $\boldsymbol{x} \in \mathcal{X}^+$, then the chance constraint in (2) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^\star$ and $\tau \in \mathbb{R}_+$ such that

$$
\begin{array}{lll}
\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, & \beta + \boldsymbol{d}_1^\top \boldsymbol{\phi} \leq 1, & \beta + \boldsymbol{d}_1^\top \boldsymbol{\psi} - \tau t(\boldsymbol{x}) \leq 0 \\
\boldsymbol{A}^\top \boldsymbol{\gamma} = \boldsymbol{C}_1^\top \boldsymbol{\phi}, & \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\phi} & \\
\boldsymbol{A}^\top \boldsymbol{\gamma} + \tau \boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{C}_1^\top \boldsymbol{\psi}, & \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\psi}. &
\end{array}
\tag{48}
$$

Any τ feasible in (48) must be strictly positive. Indeed, assume to the contrary that $\tau = 0$. Then all constraints involving $\boldsymbol{\phi}$ are redundant as we may set $\boldsymbol{\phi} = \boldsymbol{\psi}$ without loss of generality. Thus, (48) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$ and $\boldsymbol{\psi} \in \mathcal{K}_1^\star$ such that

$$
\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, \quad \beta + \boldsymbol{d}_1^\top \boldsymbol{\psi} \leq 0, \quad \boldsymbol{A}^\top \boldsymbol{\gamma} = \boldsymbol{C}_1^\top \boldsymbol{\psi}, \quad \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\psi},
$$

and the conic Farkas lemma implies that this set of equations is satisfiable if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$ and $\boldsymbol{\psi} \in \mathcal{K}_1^\star$ such that

$$
\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, \quad \beta + (\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^\top \boldsymbol{\gamma} \leq 0 \quad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1.
$$

Taking expectation with respect to any probability distribution $\mathbb{P} \in \mathcal{P}$, we thus obtain $\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} = \mathbb{E}_\mathbb{P}[\beta + (\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}})^\top \boldsymbol{\gamma}] \leq 0$, which contradicts the first constraint in the above expression. We conclude that any feasible τ must be strictly positive. Dividing all constraints in (48) by $\tau > 0$ and performing the change of variables $\beta \leftarrow \beta/\tau$, $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma}/\tau$, $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/\tau$, $\boldsymbol{\psi} \leftarrow \boldsymbol{\psi}/\tau$ and $\tau \leftarrow 1/\tau$ shows that the chance constraint in (2) is satisfied if and only if (26) holds.

Assume now that $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^+$, that is, $\boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} \leq t(\boldsymbol{x})$ for all $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1$. In this case, the worst-case probability in (2) evaluates to 1 and the chance constraint is vacuously satisfied. We show that (26) is vacuously satisfied as well. Indeed, if we set $(\beta, \boldsymbol{\gamma}, \boldsymbol{\phi}, \tau) = (0, \boldsymbol{0}, \boldsymbol{0}, 0)$, then (26) reduces to

$$
\exists \boldsymbol{\psi} \in \mathcal{K}_1^\star : \boldsymbol{d}_1^\top \boldsymbol{\psi} \leq t(\boldsymbol{x}), \quad \boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{C}_1^\top \boldsymbol{\psi}, \quad \boldsymbol{D}_1^\top \boldsymbol{\psi} = \boldsymbol{0}.
$$

The conic Farkas lemma implies that this statement is equivalent to the assertion that $\boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} \leq t(\boldsymbol{x})$ for all $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1$, which is true since $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^+$.

In summary, we conclude that the chance constraint in (2) is equivalent to (26) for all $\boldsymbol{x} \in \mathcal{X}$. Thus, the claim follows. $\square$

Proof of Theorem 8. Fix an instance $(\boldsymbol{E}, \boldsymbol{f})$ of the IP problem and consider the following instance of (2), which is parameterized in $n = 1, \ldots, P$:

$$
\begin{array}{ll}
\sup & 0 \\
\text{s.t.} & \boldsymbol{x} \in [0, 1]^P, \;\; \boldsymbol{E}\boldsymbol{x} \leq \boldsymbol{f}, \;\; \mathbf{e}^\top \boldsymbol{x} = n \\
& \displaystyle\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left[\tilde{\boldsymbol{z}}^\top \boldsymbol{x} \geq 1\right] \geq \frac{n}{P},
\end{array}
\tag{49}
$$

where the ambiguity set $\mathcal{P}$ satisfies

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) \ : \ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in [-2, 2]^P \right] = 1, \quad \mathbb{P}\left[\tilde{\boldsymbol{z}} = \mathbf{e}_p \right] = \frac{1}{P} \quad \forall p = 1, \dots P \right\}.$$

Here, $\mathbf{e}$ and $\mathbf{e}_p$ denote the vector of all ones in $\mathbb{R}^P$ and the p-th unit basis vector in $\mathbb{R}^P$, respectively. The ambiguity set $\mathcal{P}$ is an instance of (4) that satisfies the conditions (**B**) and (**N**) since the confidence sets $\{\mathbf{e}_p\}$, $p = 1, \dots, P$, are pairwise disjoint subsets of the interior of the support $[-2, 2]^P$.

We want to show that problem (49) is feasible if and only if the IP problem has a solution $\boldsymbol{x}$ with exactly n components being one. Indeed, we observe that $\mathcal{P} = \{\mathbb{P}_U\}$, where $\mathbb{P}_U$ denotes the uniform distribution over the vertices of the probability simplex in $\mathbb{R}^P$. Thus, one readily verifies that if there is a solution $\boldsymbol{x}$ to the IP problem with $\mathbf{e}^\top \boldsymbol{x} = n$, then $\boldsymbol{x}$ is a feasible solution to problem (49).

Conversely, assume that $\boldsymbol{x}$ solves (49). The box constraint $\boldsymbol{x} \in [0, 1]^P$ and the chance constraint in (49) then imply that at least n components of $\boldsymbol{x}$ are one. From the constraint $\mathbf{e}^\top \boldsymbol{x} = n$ we then conclude that exactly n components of $\boldsymbol{x}$ are one and all other components are zero, that is, $\boldsymbol{x}$ is a solution of the IP problem with exactly n components being one.

In conclusion, problem (49) is feasible if and only if the IP problem has a solution $\boldsymbol{x}$ with exactly n components being one. We can therefore decide the IP problem by checking whether $\boldsymbol{E0} \leq \boldsymbol{f}$ and otherwise sequentially verifying the feasibility of problem (49) for $n = 1, \dots, P$. $\qquad\square$

Proof of Theorem 9. We can employ Theorem 4 and conduct some simple algebraic manipulations to conclude that the worst-case probability of the chance constraint in (2) is given by

$$
\begin{aligned}
\sup \quad & \beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \beta \in \mathbb{R}, \ \ \boldsymbol{\gamma} \in \mathbb{R}^K, \ \ \boldsymbol{\phi}, \boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^\star, \ \ \tau \in \mathbb{R}_+ \\
& \beta + \boldsymbol{d}_1^\top \boldsymbol{\phi} \leq 1, \ \ \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\phi} \\
& \left. \begin{array}{l} 2\beta + \boldsymbol{d}_1^\top(\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) - \tau t(\boldsymbol{x}) \leq 1 \\ \boldsymbol{C}_1^\top(\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) = \tau \boldsymbol{s}(\boldsymbol{x}) \\ \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\psi}^+ = \boldsymbol{D}_1^\top \boldsymbol{\psi}^- \end{array} \right\} \ \ \text{if } \boldsymbol{x} \in \mathcal{X}^\pm,
\end{aligned}
$$

where $\mathcal{X}^\pm = \{\boldsymbol{x} \in \mathcal{X} \ : \exists (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1 \ . \ \boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} < -t(\boldsymbol{x}) \vee \boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} > t(\boldsymbol{x})\}$.

In the remainder we will separately investigate the cases $\boldsymbol{x} \in \mathcal{X}^\pm$ and $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^\pm$. If $\boldsymbol{x} \in \mathcal{X}^\pm$, then the chance constraint in (2) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\boldsymbol{\phi}, \boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^\star$ and $\tau \in \mathbb{R}_+$ such that

$$
\begin{array}{ll}
\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, & \beta + \boldsymbol{d}_1^\top \boldsymbol{\phi} \leq 1 \\
2\beta + \boldsymbol{d}_1^\top(\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) - \tau t(\boldsymbol{x}) \leq 1, & \boldsymbol{C}_1^\top(\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) = \tau \boldsymbol{s}(\boldsymbol{x}) \\
\boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\psi}^+ = \boldsymbol{D}_1^\top \boldsymbol{\psi}^- = \boldsymbol{D}_1^\top \boldsymbol{\phi}
\end{array} \tag{50}
$$

Any τ feasible in (50) must be strictly positive. Indeed, assume to the contrary that $\tau = 0$. We can then conclude that all constraints involving $\boldsymbol{\phi}$ are redundant as we may set $\boldsymbol{\phi} = (\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-)/2$ without loss of generality. Thus, the

constraint system (50) is satisfiable if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$ and $\boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^\star$ such that

$$\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, \qquad 2\beta + \boldsymbol{d}_1^\top (\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) \leq 1$$
$$\boldsymbol{C}_1^\top (\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) = \boldsymbol{0}, \quad \boldsymbol{B}^\top \boldsymbol{\gamma} = \boldsymbol{D}_1^\top \boldsymbol{\psi}^+ = \boldsymbol{D}_1^\top \boldsymbol{\psi}^-,$$

and the conic Farkas lemma implies that this set of equations is satisfiable if and only if there exist $\beta \in \mathbb{R}$ and $\boldsymbol{\gamma} \in \mathbb{R}^K$ such that

$$\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon$$
$$2\beta + (\boldsymbol{B}\boldsymbol{u}^+ + \boldsymbol{B}\boldsymbol{u}^-)^\top \boldsymbol{\gamma} \leq 1 \quad \forall (\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q : \qquad (51)$$
$$(\boldsymbol{z}, \boldsymbol{u}^+), (-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_1.$$

Note that every distribution $\mathbb{P} \in \mathcal{P}$ is a mixture of extremal distributions of the form $\frac{1}{2}\delta_{(\boldsymbol{z},\boldsymbol{u}^+)} + \frac{1}{2}\delta_{(-\boldsymbol{z},\boldsymbol{u}^-)}$, where $(\boldsymbol{z}, \boldsymbol{u}^+), (-\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_1$. The semi-infinite constraint in (51) implies that $\mathbb{E}_{\mathbb{P}}[2\beta + 2(\boldsymbol{B}\tilde{\boldsymbol{u}})^\top \boldsymbol{\gamma}] \leq 1$ for all extremal distributions $\mathbb{P} \in \mathcal{P}$, and we thus obtain that $\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} = \mathbb{E}_{\mathbb{P}}[\beta + (\boldsymbol{B}\tilde{\boldsymbol{u}})^\top \boldsymbol{\gamma}] \leq 1/2$, which contradicts the first constraint in (51) since $\epsilon < 1/2$. We thus conclude that any feasible τ must be strictly positive. Dividing all constraints in (50) by $\tau > 0$ and performing the change of variables $\beta \leftarrow \beta/\tau$, $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma}/\tau$, $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/\tau$, $\boldsymbol{\psi}^+ \leftarrow \boldsymbol{\psi}^+/\tau$, $\boldsymbol{\psi}^- \leftarrow \boldsymbol{\psi}^-/\tau$ and $\tau \leftarrow 1/\tau$ shows that the chance constraint in (2) is satisfied if and only if (27) holds.

Assume now that $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^\pm$, that is, $-t(\boldsymbol{x}) \leq \boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} \leq t(\boldsymbol{x})$ for all $(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1$. In this case, the worst-case probability in (2) evaluates to 1 and the chance constraint is vacuously satisfied. We show that (27) is vacuously satisfied as well. Indeed, if we set $(\beta, \boldsymbol{\gamma}, \boldsymbol{\phi}, \tau) = (0, \boldsymbol{0}, \boldsymbol{0}, 0)$, then (27) becomes

$$\exists \boldsymbol{\psi}^+, \boldsymbol{\psi}^- \in \mathcal{K}_1^\star : \begin{cases} \boldsymbol{d}_1^\top (\boldsymbol{\psi}^+ + \boldsymbol{\psi}^-) \leq t(\boldsymbol{x}), \\ \boldsymbol{C}_1^\top (\boldsymbol{\psi}^+ - \boldsymbol{\psi}^-) = \boldsymbol{s}(\boldsymbol{x}), \\ \boldsymbol{D}_1^\top \boldsymbol{\psi}^+ = \boldsymbol{D}_1^\top \boldsymbol{\psi}^- = \boldsymbol{0}. \end{cases}$$

The conic Farkas lemma implies that this statement is equivalent to the assertion that $\boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} \leq t(\boldsymbol{x})$ for all $(\boldsymbol{z}, \boldsymbol{u}^+, \boldsymbol{u}^-) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{R}^Q$ such that $(\boldsymbol{z}, \boldsymbol{u}^+), (\boldsymbol{z}, \boldsymbol{u}^-) \in \mathcal{C}_1$, which is true since $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^\pm$.

In summary, we conclude that the chance constraint in (2) is equivalent to (27) for all $\boldsymbol{x} \in \mathcal{X}$. Thus, the claim follows.   □

Proof of Theorem 10. The proof follows the same argument as the proof of Theorem 8. We only need to replace $\mathcal{P}$ with the symmetric ambiguity set

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}) : \begin{array}{l} \mathbb{P}\left[\begin{pmatrix} \tilde{\boldsymbol{z}} \\ \tilde{u} \end{pmatrix} \in [-2,2]^P \times \mathbb{R}\right] = 1, \\ \mathbb{P}\left[\begin{pmatrix} \tilde{\boldsymbol{z}} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_p \\ 0 \end{pmatrix}\right] = \dfrac{1}{2P} \\ \mathbb{P}\left[\begin{pmatrix} \tilde{\boldsymbol{z}} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} -\mathbf{e}_p \\ 0 \end{pmatrix}\right] = \dfrac{1}{2P} \end{array} \right\} \forall p = 1, \ldots P \right\}$$

and set the right-hand side of the chance constraint in (49) to $n/(2P)$.   □

Proof of Theorem 11.   By Theorem 5, the chance constraint in (2) is satisfied if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$ and $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^\star$ with

$$
\left.
\begin{aligned}
&\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, \quad -\boldsymbol{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^\star \\
&\beta + \left(\frac{\alpha}{\alpha+1} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}g^\alpha(\boldsymbol{z}) \right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\phi} \leq 1 \\
&\beta + \left(\frac{\alpha}{\alpha+1} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}g^\alpha(\boldsymbol{z}) \right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi} \\
&\qquad + \tau \boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} - \frac{(\alpha+1)^{\alpha+1}}{\alpha^\alpha} (\tau t(\boldsymbol{x}))^{\frac{\alpha}{\alpha+1}} \leq 0 \ \text{ if } \boldsymbol{x} \in \mathcal{X}^+
\end{aligned}
\right\} \forall \boldsymbol{z} \in \mathbb{R}^P,
$$

$$(52)$$

where $\mathcal{X}^+ = \left\{ \boldsymbol{x} \in \mathcal{X} : \exists (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha . \, \boldsymbol{s}_1(\boldsymbol{x})^\top \boldsymbol{z} > 0 \right\}$.

In the remainder, we will separately investigate the cases $\boldsymbol{x} \in \mathcal{X}^+$ and $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^+$. If $\boldsymbol{x} \in \mathcal{X}^+$, then any τ feasible in (52) must be strictly positive. Indeed, assume to the contrary that $\tau = 0$. Then (52) is satisfiable by any $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$ and $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^\star$ if and only if it is satisfiable by any $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$ and $\boldsymbol{\phi}, \boldsymbol{\psi} \in \mathcal{K}_1^\star$ with $\boldsymbol{\phi} = \boldsymbol{\psi}$. We may thus set $\boldsymbol{\phi} = \boldsymbol{\psi}$, in which case the first semi-infinite constraint becomes redundant. Thus, (52) is satisfiable if and only if there exist $\beta \in \mathbb{R}$, $\boldsymbol{\gamma} \in \mathbb{R}^K$, $\tau \in \mathbb{R}_+$, $\boldsymbol{\psi} \in \mathcal{K}_1^\star$ with

$$
\begin{aligned}
&\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon, \quad -\boldsymbol{B}^\top \boldsymbol{\gamma} \in \mathcal{K}_2^\star \\
&\beta + \left(\tfrac{\alpha}{\alpha+1} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}g^\alpha(\boldsymbol{z}) \right)^\top \boldsymbol{\gamma} + (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi} \leq 0 \quad \forall \boldsymbol{z} \in \mathbb{R}^P.
\end{aligned}
$$

Using the conic Farkas lemma, one can show that these constraints hold if and only if there exists $\beta \in \mathbb{R}$ and $\boldsymbol{\gamma} \in \mathbb{R}^K$ with

$$
\begin{aligned}
&\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} \geq 1 - \epsilon \\
&\beta + \left(\tfrac{\alpha}{\alpha+1} \boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u} \right)^\top \boldsymbol{\gamma} \leq 0 \quad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{C}_1^\alpha.
\end{aligned}
$$

Taking the expectation of the second constraint with respect to any $\mathbb{P} \in \mathcal{P}$ implies that $\beta + \boldsymbol{b}^\top \boldsymbol{\gamma} = \mathbb{E}_\mathbb{P}[\beta + (\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}})^\top \boldsymbol{\gamma}] \leq 0$, which contradicts the first constraint since $\epsilon < 1$. We thus conclude that any τ feasible in (52) must be strictly positive. Dividing all constraints in (52) by $\tau > 0$ and performing the change of variables $\beta \leftarrow \beta/\tau$, $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma}/\tau$, $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}/\tau$, $\boldsymbol{\psi} \leftarrow \boldsymbol{\psi}/\tau$, and $\tau \leftarrow 1/\tau$ shows that the chance constraint in (2) is satisfied if and only if (28) holds.

Assume next that $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^+$, which implies that $\boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} \leq 0 \leq t(\boldsymbol{x})$ for all $\boldsymbol{z} \in \{ \boldsymbol{z} \in \mathbb{R}^P : \boldsymbol{C}_1 \boldsymbol{z} \preccurlyeq_{\mathcal{K}_1} \boldsymbol{d}_1 \}$. In this case, the chance constraint is vacuously satisfied as $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left[\boldsymbol{s}(\boldsymbol{x})^\top \tilde{\boldsymbol{z}} \leq t(\boldsymbol{x}) \right] = 1$. We show that (28) is vacuously satisfied as well. Indeed, if we set $(\beta, \boldsymbol{\gamma}, \boldsymbol{\phi}, \tau) = (0, \boldsymbol{0}, \boldsymbol{0}, 0)$, then (28) reduces to

$$
\exists \boldsymbol{\psi} \in \mathcal{K}_1^\star : (\boldsymbol{d}_1 - \boldsymbol{C}_1 \boldsymbol{z})^\top \boldsymbol{\psi} + \boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} \leq 0 \quad \forall \boldsymbol{z} \in \mathbb{R}^P.
$$

Strong conic duality implies that this statement is equivalent to the assertion that $\boldsymbol{s}(\boldsymbol{x})^\top \boldsymbol{z} \leq 0$ for all $\boldsymbol{z} \in \{ \boldsymbol{z} \in \mathbb{R}^P : \boldsymbol{C}_1 \boldsymbol{z} \preccurlyeq_{\mathcal{K}_1} \boldsymbol{d}_1 \}$, which is true since $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^+$. In summary, we conclude that the chance constraint in (2) is equivalent to (28) for all $\boldsymbol{x} \in \mathcal{X}$. Thus, the claim follows.   □