

# Ambiguous Joint Chance Constraints under Mean and Dispersion Information

Grani A. Hanasusanto<sup>1</sup>, Vladimir Roitch<sup>2</sup>, Daniel Kuhn<sup>3</sup>, and Wolfram Wiesemann<sup>4</sup>

<sup>1</sup>Graduate Program in Operations Research and Industrial Engineering, The University of Texas at Austin, USA

<sup>2</sup>Department of Computing, Imperial College London, United Kingdom

<sup>3</sup>Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland

<sup>4</sup>Imperial College Business School, Imperial College London, United Kingdom

September 29, 2016

## Abstract

We study joint chance constraints where the distribution of the uncertain parameters is only known to belong to an ambiguity set characterized by the mean and support of the uncertainties and by an upper bound on their dispersion. This setting gives rise to pessimistic (optimistic) ambiguous chance constraints, which require the corresponding classical chance constraints to be satisfied for every (for at least one) distribution in the ambiguity set. We demonstrate that the pessimistic joint chance constraints are conic representable if *(i)* the constraint coefficients of the decisions are deterministic, *(ii)* the support set of the uncertain parameters is a cone, and *(iii)* the dispersion function is of first order, that is, it is positively homogeneous. We also show that pessimistic joint chance constrained programs become intractable as soon as either of the conditions *(i)*, *(ii)* or *(iii)* is relaxed in the mildest possible way. We further prove that the optimistic joint chance constraints are conic representable if *(i)* holds, and that they become intractable if *(i)* is violated. We show in numerical experiments that our results allow us to solve large-scale project management and image reconstruction models to global optimality.

## 1 Introduction

The optimal design or control of a physical, engineering or economic system is a ubiquitous problem that arises in numerous practical applications. Many systems of interest are impacted both

by a vector of design decisions  $\mathbf{x}$ , which are to be chosen from within a polytope  $\mathcal{X} \subseteq \mathbb{R}^n$ , as well as an exogenous random vector  $\tilde{\boldsymbol{\xi}}$ , which is governed by a known probability distribution  $\mathbb{Q}$  supported on  $\Xi \subseteq \mathbb{R}^k$ . Suppose that the reliable operation of the system requires the satisfaction of  $m$  uncertainty-affected *safety conditions*  $\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} < \mathbf{u}(\mathbf{x})$ , where  $\mathbf{T}(\mathbf{x}) \in \mathbb{R}^{m \times k}$  and  $\mathbf{u}(\mathbf{x}) \in \mathbb{R}^m$  constitute affine functions of  $\mathbf{x}$ . Then, a popular goal is to design a system at minimum cost  $\mathbf{c}^\top \mathbf{x}$ ,  $\mathbf{c} \in \mathbb{R}^n$ , that satisfies the safety conditions with a probability of at least  $1 - \epsilon$ , where  $\epsilon \in [0, 1)$  reflects the tolerated risk level. This scenario gives rise to the linear *chance constrained program*

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{c}^\top \mathbf{x} \\ & \text{subject to} && \mathbf{x} \in \mathcal{X} \\ & && \mathbb{Q}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} < \mathbf{u}(\mathbf{x})] \geq 1 - \epsilon. \end{aligned} \tag{1}$$

The probabilistic constraint in (1) is termed an *individual chance constraint* in case of a single safety condition ( $m = 1$ ) and a *joint chance constraint* if there are multiple safety conditions ( $m > 1$ ). We refer to the chance constraint in (1) as a *strict chance constraint* as it requires strict satisfaction of the safety conditions. The closely related *weak chance constraints* replace the strict inequality in (1) with a weak one and thus only require weak satisfaction of the safety conditions.

Initiated by the seminal work of Charnes *et al.* [15] and Charnes and Cooper [14], chance constrained programs have been employed in numerous application domains ranging from logistics [20], finance [24], project management [54] and network design [57] to emissions control [1], design optimization [10] and call center staffing [33]. Despite their wide-spread use, chance constrained programs suffer from two shortcomings: they require an exact specification of the distribution  $\mathbb{Q}$ , and they can lead to computationally challenging optimization problems.

Both the conceptual as well as the computational difficulties of model (1) can be alleviated if we replace the classical chance constraint in (1) with an *ambiguous chance constraint* of the form

$$\inf_{\mathbb{P} \in \mathcal{P}} / \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} < \mathbf{u}(\mathbf{x})] \geq 1 - \epsilon.$$

Ambiguous chance constraints acknowledge that the true distribution  $\mathbb{Q}$  is only known to reside within an *ambiguity set*  $\mathcal{P}$ , and they admit a pessimistic or an optimistic formulation. The *pessimistic* version requires the safety conditions to hold with probability  $1 - \epsilon$  in the *worst case*, that is, when the probability of the system being safe is *minimized* over all possible distributions in  $\mathcal{P}$ . In contrast, the *optimistic* version of the ambiguous chance constraint requires the system to be

safe in the *best case*, that is, when the probability of the system being safe is *maximized* over the distributions in  $\mathcal{P}$ . The ambiguity set is chosen so as to contain all distributions that are consistent with the available *a priori* information on  $\mathbb{Q}$ , such as structural characteristics (*e.g.*, support) or statistical properties (*e.g.*, moment bounds). By considering all probability distributions that are deemed possible under the available information, ambiguous chance constraints mitigate the reliance on a precise characterization of the true distribution  $\mathbb{Q}$ .

Note that the *pessimistic* ambiguous chance constraint constitutes a *conservative* approximation (restriction) to the classical chance constraint in (1), while the *optimistic* version represents a *progressive* approximation (relaxation) whenever the true distribution  $\mathbb{Q}$  is contained in the ambiguity set  $\mathcal{P}$ . The existing literature has almost exclusively focused on pessimistic chance constraints, which are also referred to as *robust* chance constraints. They reflect a strict aversion to ambiguity and are appropriate for decision makers who wish to hedge against *any* possible distribution  $\mathbb{P} \in \mathcal{P}$ . In contrast, optimistic chance constraints will appeal to ambiguity-seeking decision makers. They naturally arise, for example, in problems of statistics and machine learning, where the discovery of *some* distribution  $\mathbb{P} \in \mathcal{P}$  that is likely to have generated a given set of observations is an essential part of the decision problem.

Perhaps surprisingly, ambiguous chance constraints are often computationally tractable. In fact, for several classes of ambiguity sets  $\mathcal{P}$ , one can exploit classical probability inequalities to equivalently reformulate or conservatively approximate robust *individual* chance constraints in a tractable way. This has first been observed by Ben-Tal and Nemirovski [4] and Bertsimas and Sim [9], who use Hoeffding’s inequality to derive tractable reformulations of robust individual chance constraints when the components of  $\tilde{\xi}$  are independent, symmetric and bounded random variables. Chen *et al.* [17] also employ the Hoeffding inequality to approximate robust individual chance constraints where the ambiguity set captures asymmetries in the distribution of  $\tilde{\xi}$  via forward and backward deviation bounds. Assuming that the components of  $\tilde{\xi}$  are independent random variables whose joint distribution belongs to a given convex and compact set, Nemirovski and Shapiro [44] and Postek *et al.* [48] use large deviation-type Bernstein bounds to approximate robust individual chance constraints. Calafiore and El Ghaoui [12] employ various statistical bounds to approximate robust individual chance constraints where the ambiguity set specifies structural properties such as radial symmetry, unimodality or independence. Bertsimas *et al.* [6] use statistical hypothesis tests

to derive safe approximations to nonlinear robust individual chance constraints based on hypothesis tests, where the ambiguity set accounts for the ambiguity associated with sampling from historical data. Xu *et al.* [62] employ a generalized Chebyshev inequality to derive tractable reformulations of problems with probabilistic envelope constraints, which enforce robust chance constraints at all tolerance levels  $\epsilon \in [0, 1)$ . Instead of relying on statistical results, one can also employ duality of moment problems [8] to derive tractable reformulations of robust individual chance constraints. This approach was pioneered by El Ghaoui *et al.* [30] for Chebyshev ambiguity sets which contain all distributions sharing a known mean value and covariance matrix.

Ambiguous chance constraints become considerably more challenging if  $m > 1$ , that is, if they involve multiple safety conditions. Nemirovski and Shapiro [44], Bertsimas *et al.* [6] and others employ Bonferroni’s inequality to conservatively approximate a robust joint chance constraint with violation probability  $\epsilon$  by  $m$  robust individual chance constraints whose violation probabilities sum up to  $\epsilon$ . Although this approach has a long tradition in chance constrained programming [49], Chen *et al.* [16] demonstrate that the quality of this approximation can deteriorate substantially with  $m$  if the safety conditions are positively correlated. Instead, they propose to approximate the robust joint chance constraint by a robust conditional value-at-risk (CVaR) constraint. For Chebyshev ambiguity sets, they subsequently approximate the robust CVaR constraint using a classical result from order statistics. Zymler *et al.* [64] prove that the CVaR approximation in [16] becomes exact if certain scaling parameters are chosen optimally. Unfortunately, optimizing simultaneously over the decisions  $\mathbf{x}$  and the scaling parameters seems to be difficult. The authors also propose an exact reformulation of the emerging robust CVaR constraint using moment duality. Van Parys *et al.* [47] use the results of [64] to solve chance constrained finite and infinite horizon control problems. Erdoğan and Iyengar [26] study robust joint chance constraints where the ambiguity set contains all distributions that are within a certain distance of a nominal distribution (in terms of the Prohorov metric). They derive a conservative approximation by sampling from the nominal distribution and enforcing the constraints for all values of  $\xi$  that are ‘close’ to any of the samples. Jiang and Guan [38] and Yamkoğlu and den Hertog [63] study data-driven robust chance constraints, where the ambiguity sets have to be estimated from samples. Conservative approximations to robust chance constraints involving linear matrix inequalities have been studied by Ben-Tal and Nemirovski [5] and Cheung *et al.* [19]. For more detailed surveys of distributionally robust chance

constrained programs, we refer to Ben-Tal *et al.* [3] and Nemirovski [43].

Despite intensive research efforts over the last two decades, only the papers of Hu and Hong [36] and Hu *et al.* [37] appear to derive results that allow for convex reformulations of ambiguous *joint* chance constrained problems. Both papers study pessimistic chance constraints of the form  $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[H(\mathbf{x}, \tilde{\boldsymbol{\xi}}) \leq 0]$  for generic classes of loss function  $H : \mathcal{X} \times \Xi \rightarrow \mathbb{R}$ . The authors show that for ambiguity sets containing all distributions within a certain distance of a nominal distribution, where the distance is measured in terms of the Kullback-Leibler divergence or the likelihood ratio, pessimistic chance constraints are equivalent to non-ambiguous chance constraints with an adjusted confidence level. Applying this result to the loss function  $H(\mathbf{x}, \tilde{\boldsymbol{\xi}}) = \max_i \{[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} - \mathbf{u}(\mathbf{x})]_i\}$ , we see that the (weak version of the) pessimistic joint chance constraint presented earlier reduces to a non-ambiguous joint chance constraint. Moreover, if the nominal probability distribution is log-concave and  $\mathbf{T}(\mathbf{x}) = \mathbf{T}$ , then we can apply Prékopa’s classical result for non-ambiguous joint chance constraints [49] to conclude that pessimistic joint chance constraints over Kullback-Leibler and likelihood ratio ambiguity sets are indeed convex. Note, however, that despite their convexity, checking the feasibility of such pessimistic chance constraints remains  $\sharp\mathbb{P}$ -hard even for individual chance constraints and log-concave nominal distributions; we elaborate on this in Section 3.2. The computational burden is reduced substantially if in addition to the log-concavity of the nominal distribution and the constant technology matrix  $\mathbf{T}$ , we require the components of the vector  $\mathbf{T}\tilde{\boldsymbol{\xi}}$  to be independent. In that case, we can replace the non-ambiguous joint chance constraint with products of non-ambiguous individual chance constraints, which amount to one-dimensional integrations and can thus be evaluated more efficiently. We note, however, that requiring the components of  $\mathbf{T}\tilde{\boldsymbol{\xi}}$  to be independent essentially implies (by a change of variables) that  $\mathbf{T} = \mathbf{I}$ , which is a rather restrictive assumption for most practical applications.

Ambiguous chance constrained programming is closely related to optimal uncertainty quantification, which aims to ascertain whether  $\tilde{\boldsymbol{\xi}}$  satisfies a set of decision-independent safety conditions with high probability for all/some distributions in an ambiguity set  $\mathcal{P}$ . Thus, uncertainty quantification is equivalent to checking whether a fixed decision  $\mathbf{x}$  is feasible in an ambiguous chance constraint. A comprehensive survey of the recent literature on uncertainty quantification has been compiled by Owhadi *et al.* [46]. A powerful method for reducing optimal uncertainty quantification problems to tractable convex programs has been proposed by Han *et al.* [34]. This method relies

on the Richter-Rogosinski theorem [53, Theorem 7.37]. In contrast, we leverage the ‘*primal worst equals dual best*’ duality scheme by Beck and Ben-Tal [2] to convert the uncertainty quantification problems arising in our setting to problems over discrete distributions with  $m$  (worst-case) or 2 (best-case) scenarios. The resulting finite-dimensional reductions are nonconvex and may fail to be solvable. By exploiting ideas by Gorissen *et al.* [32], however, they can be reformulated as conic programs whose optima are always attained. We further elaborate on the relation between this approach and the one by Han *et al.* in Section 3.1.

In this paper, we develop a new approach for solving ambiguous joint chance constrained programs. We highlight the following main contributions of this work.

- We demonstrate that pessimistic joint chance constraints have conic representations if (i) the coefficient matrix  $\mathbf{T}(\mathbf{x})$  is constant in  $\mathbf{x}$ , (ii) the support set  $\Xi$  is a cone, and (iii) the dispersion function  $\mathbf{d}(\tilde{\boldsymbol{\xi}})$  is positively homogeneous. For suitably chosen cones, pessimistic joint chance constraints are thus computationally tractable. This seems to be the first tractability result for pessimistic joint chance constraints.
- We prove that this tractability result is sharp in the sense that pessimistic joint chance constrained programs become strongly NP-hard as soon as either of the conditions (i), (ii) or (iii) is relaxed in the mildest possible way. To our best knowledge this is the first complexity analysis for pessimistic joint chance constraints.
- We show that optimistic joint chance constrained programs are conic representable if (i) holds, and that they become intractable if (i) is violated.
- We showcase that our tractability result enables us to solve ambiguous joint chance constrained programs with more than 320,000 safety conditions using standard optimization solvers. To our best knowledge this problem size is far beyond the capabilities of the existing algorithms in classical chance constrained programming.

For ease of exposition, we focus on strict pessimistic and optimistic chance constraints. Throughout the paper, we outline how our results extend to the closely related weak chance constraints.

In the remainder we first outline our modeling assumptions in Section 2. We then derive conic reformulations and complexity results for *pessimistic* uncertainty quantification and chance constrained programming problems (Section 3). Subsequently we develop the corresponding results

for *optimistic* uncertainty quantification and chance constrained programming problems (Section 4). Finally, we demonstrate the impact of our tractability results in large-scale project management and image reconstruction applications (Section 5). All proofs are relegated to the appendix.

**Notation:** For a proper (*i.e.*, convex, closed, solid and pointed) cone  $\mathcal{D} \subseteq \mathbb{R}^d$  and  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$  the inequality  $\mathbf{v} \preceq_{\mathcal{D}} \mathbf{w}$  ( $\mathbf{v} \prec_{\mathcal{D}} \mathbf{w}$ ) expresses that  $\mathbf{w} - \mathbf{v} \in \mathcal{D}$  ( $\mathbf{w} - \mathbf{v} \in \text{int } \mathcal{D}$ ). A function  $\mathbf{d}(\boldsymbol{\xi})$  mapping  $\mathbb{R}^k$  to  $\mathbb{R}^d$  is called  $\mathcal{D}$ -convex if  $\mathbf{d}(\theta \boldsymbol{\xi}_1 + (1 - \theta) \boldsymbol{\xi}_2) \preceq_{\mathcal{D}} \theta \mathbf{d}(\boldsymbol{\xi}_1) + (1 - \theta) \mathbf{d}(\boldsymbol{\xi}_2)$  for all  $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \mathbb{R}^k$  and  $\theta \in [0, 1]$ . The cone dual to  $\mathcal{D}$  is denoted by  $\mathcal{D}^*$ . An extended real-valued function  $f(\boldsymbol{\xi})$  is proper if  $f(\boldsymbol{\xi}) < +\infty$  for some  $\boldsymbol{\xi}$  and  $f(\boldsymbol{\xi}) > -\infty$  for every  $\boldsymbol{\xi} \in \mathbb{R}^k$ . The conjugate of a proper function  $f(\boldsymbol{\xi})$  is given by  $f^*(\boldsymbol{\nu}) = \sup_{\boldsymbol{\xi} \in \mathbb{R}^k} \boldsymbol{\nu}^\top \boldsymbol{\xi} - f(\boldsymbol{\xi})$ . The indicator function of a set  $\Xi \subseteq \mathbb{R}^k$  is defined as  $\delta_{\Xi}(\boldsymbol{\xi}) = 0$  if  $\boldsymbol{\xi} \in \Xi$ ;  $= \infty$  otherwise, and its conjugate  $\sigma_{\Xi}(\boldsymbol{\nu}) = \sup_{\boldsymbol{\xi} \in \Xi} \boldsymbol{\nu}^\top \boldsymbol{\xi}$  is termed the support function of  $\Xi$ . We define  $\mathbf{e}$  as the vector of all ones, and we let  $\mathbf{e}_i$  be the  $i$ -th standard basis vector of appropriate dimension. All random objects are designated by tilde signs (*e.g.*,  $\tilde{\boldsymbol{\xi}}$ ), while their realizations are denoted by the same symbols without tildes (*e.g.*,  $\boldsymbol{\xi}$ ). The convex cone of nonnegative Borel measures on  $\Xi$  is denoted by  $\mathcal{M}_+(\Xi)$ , and  $\delta_{\boldsymbol{\xi}}$  represents the Dirac measure placing unit mass at  $\boldsymbol{\xi}$ . For a logical expression  $\mathcal{E}$ , we define  $\mathbb{I}_{[\mathcal{E}]} = 1$  if  $\mathcal{E}$  is true;  $= 0$  otherwise.

## 2 Model Formulation

We study pessimistic and optimistic ambiguous joint chance constrained problems of the form

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{c}^\top \mathbf{x} \\ & \text{subject to} && \mathbf{x} \in \mathcal{X} \\ & && \inf_{\mathbb{P} \in \mathcal{P}} / \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x}) \tilde{\boldsymbol{\xi}} < \mathbf{u}(\mathbf{x})] \geq 1 - \epsilon, \end{aligned} \tag{2}$$

where  $\mathbf{c} \in \mathbb{R}^n$ ,  $\mathcal{X} \subseteq \mathbb{R}^n$  is a polytope,  $\mathbf{T}(\mathbf{x}) \in \mathbb{R}^{m \times k}$  and  $\mathbf{u}(\mathbf{x}) \in \mathbb{R}^m$  constitute affine functions of  $\mathbf{x}$ ,  $\epsilon \in [0, 1)$  and the ambiguity set  $\mathcal{P}$  satisfies

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[\mathbf{d}(\tilde{\boldsymbol{\xi}})] \preceq_{\mathcal{D}} \boldsymbol{\sigma} \right\}, \tag{3}$$

where  $\mathcal{P}_0(\Xi)$  is the set of all Borel probability distributions on  $\Xi \subseteq \mathbb{R}^k$ , while  $\boldsymbol{\mu} \in \mathbb{R}^k$  stands for the mean value of  $\tilde{\boldsymbol{\xi}}$ , and  $\boldsymbol{\sigma} \in \mathbb{R}^d$  represents an upper bound on the dispersion measure corresponding to the expectation of the dispersion function  $\mathbf{d}(\tilde{\boldsymbol{\xi}}) \in \mathbb{R}^d$ . Moreover,  $\mathcal{D} \subseteq \mathbb{R}^d$  is a proper cone.

In order to derive conic reformulations for the pessimistic and optimistic uncertainty quantification problems, which evaluate the left-hand side of the probabilistic constraint in (2), we make the following assumptions:

(D) The dispersion function  $\mathbf{d}$  is  $\mathcal{D}$ -convex.

(S) The support set  $\Xi$  is convex, closed and solid.

(A) The ambiguity set  $\mathcal{P}$  satisfies the Slater condition  $\boldsymbol{\mu} \in \text{int } \Xi$  and  $\mathbf{d}(\boldsymbol{\mu}) \prec_{\mathcal{D}} \boldsymbol{\sigma}$ .

Note that (D) and (S) constitute necessary conditions for tractability. If either of them is violated, then it is already strongly NP-hard to check whether  $\mathcal{P}$  is nonempty. Moreover, (A) essentially requires that the ambiguity set  $\mathcal{P}$  contains—but does not solely consist of—the Dirac measure that places unit mass at  $\boldsymbol{\mu}$ . The conditions (D), (S) and (A) are assumed to hold throughout the paper.

Despite its apparent simplicity, the ambiguity set (3) allows us to recover a range of ambiguity sets from the literature. For  $d = 0$ , we obtain ambiguity sets over distributions with known mean and support, which have recently been used in the context of adaptive routing problems [27]. Setting  $\mathbf{d}(\boldsymbol{\xi}) = (\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top$  and identifying  $\mathcal{D}$  with the cone of positive semidefinite matrices, (3) models Chebyshev ambiguity sets closely related to those proposed in [21, 56]. The dispersion measure  $d_i(\boldsymbol{\xi}) = \mathbf{f}_i^\top (\boldsymbol{\xi} - \boldsymbol{\mu})^{m_i/n_i}$ ,  $\mathbf{f}_i \in \mathbb{R}^k$  and  $m_i, n_i \in \mathbb{N}$  with  $m_i > n_i$ ,  $i = 1, \dots, d$ , allows us to impose upper bounds on higher-order moments of  $\tilde{\boldsymbol{\xi}}$  similar to [61]. Information about the distributions' asymmetry can be captured through the choice  $\mathbf{d}(\boldsymbol{\xi}) = \mathbf{f}(\max\{\boldsymbol{\xi} - \boldsymbol{\mu}, \mathbf{0}\}, \max\{\boldsymbol{\mu} - \boldsymbol{\xi}, \mathbf{0}\})$  for  $\mathbf{f} : \mathbb{R}_+^{2k} \rightarrow \mathbb{R}^d$ , where the maximum operators are applied component-wise. The choice  $d_i(\boldsymbol{\xi}) = \xi_i^2/2$  if  $|\xi_i| \leq \delta$ ;  $= \delta(|\xi_i| - \delta/2)$  for  $i = 1, \dots, k$  and  $\delta > 0$  imposes upper bounds on the expected Huber loss function, which is a popular dispersion measure in robust statistics [13]. Finally, separate bounds  $\boldsymbol{\sigma}_i \in \mathbb{R}^{d_i}$  on different dispersion measures  $\mathbf{d}_i : \mathbb{R}^k \rightarrow \mathbb{R}^{d_i}$  and over individual cones  $\mathcal{D}_i$ ,  $i = 1, \dots, s$ , can be combined by setting  $\mathbf{d} = (\mathbf{d}_1^\top, \dots, \mathbf{d}_s^\top)^\top$ ,  $\mathcal{D} = \mathcal{D}_1 \times \dots \times \mathcal{D}_s$  and  $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1^\top, \dots, \boldsymbol{\sigma}_s^\top)^\top$ .

In order to derive a conic reformulation for pessimistic chance constraints, we impose the following additional assumptions, which complement (D), (S) and (A):

(D') The dispersion function  $\mathbf{d}$  is  $\mathcal{D}$ -convex and positively homogeneous of degree 1.

(S') The support set  $\Xi$  is a convex, closed and solid cone.

(T) The technology matrix is constant, that is,  $\mathbf{T}(\mathbf{x}) = \mathbf{T}$  for all  $\mathbf{x} \in \mathbb{R}^n$ .



Although restrictive, these assumptions are satisfied in a number of practically relevant situations.

The positive homogeneity condition  $(\mathbf{D}')$  supersedes the convexity condition  $(\mathbf{D})$ , and it restricts us to first-order dispersion measures. First-order dispersion measures are commonly used in robust statistics [13] as they are less affected by outliers and deviations from the classical statistical modeling assumptions (e.g., normality). In optimization, first-order dispersion measures have recently been employed in inventory management problems [41], where they have been shown to possess favorable properties over other statistical indicators especially when few historical samples are available. Moreover, first-order and robust dispersion measures are used in portfolio optimization [23, 39], where they enjoy computational advantages and help to immunize the portfolio weights against outliers in the historical return samples. Condition  $(\mathbf{D}')$  is satisfied by ambiguity sets that specify the mean and an upper bound on  $\mathbb{E}_{\mathbb{P}}[\|\tilde{\xi} - \mu\|]$ , where  $\|\cdot\|$  is any norm on  $\mathbb{R}^k$ . Indeed, the substitution  $\tilde{\xi} \leftarrow \tilde{\xi} - \mu$  allows us to choose  $\mu = \mathbf{0}$  and  $d(\xi) = \|\xi\|$ , and the condition  $(\mathbf{D}')$  is implied by the absolute homogeneity of norms. Likewise, condition  $(\mathbf{D}')$  is satisfied by ambiguity sets that specify the mean and an upper bound on the component-wise mean absolute deviation  $|\tilde{\xi} - \mu|$  of the random vector  $\tilde{\xi}$ . Ambiguity sets with mean and mean absolute deviation information have recently been studied in [48]. Condition  $(\mathbf{D}')$  is also satisfied by ambiguity sets that specify the mean value  $\mu$  and separate upper bounds on the lower and upper mean semi-deviations,  $d(\xi) = (\max\{\xi - \mu, \mathbf{0}\}^\top, \max\{\mu - \xi, \mathbf{0}\}^\top)^\top$ , which have been proposed in [61].

The conic support condition  $(\mathbf{S}')$  supersedes the convexity condition  $(\mathbf{S})$ . The two most natural choices of support sets  $\Xi$  that satisfy condition  $(\mathbf{S}')$  are  $\Xi = \mathbb{R}^k$  and  $\Xi = \mathbb{R}_+^k$ . If  $\tilde{\xi}$  is known to be supported on a non-conic subset  $\Xi$  of  $\mathbb{R}^k$ , then we can replace  $\Xi$  with its conic hull in order to satisfy condition  $(\mathbf{S}')$ . The resulting outer approximation of the ambiguity set allows us to derive a conic reformulation that constitutes a conservative approximation to the pessimistic chance constraint.

Condition  $(\mathbf{T})$  requires that  $\tilde{\xi}$  and  $\mathbf{x}$  appear on different sides of the safety conditions. This condition is also instrumental for Prékopa's classical convexity result for non-ambiguous joint chance constraints subject to log-concave probability distributions [49]. The condition is satisfied, among others, in resource allocation problems on temporal networks (see [59] and Section 5.1), production planning, scheduling and inventory management problems [29], as well as uncertain binary optimization problems, where products of decision variables and parameters can be linearized exactly.

Our conic reformulation for optimistic chance constraints only requires the assumptions  $(\mathbf{D})$ ,

(**S**), (**A**) and (**T**) to hold, that is, the stricter assumptions (**D'**) and (**S'**) on the dispersion measure and the support of the random vector  $\tilde{\boldsymbol{\xi}}$  are not needed.

By itself, the existence of a conic reformulation for an optimization problem does not guarantee computational tractability in the sense of polynomial time solvability. To this end, we require

- (**X**) The support set  $\Xi$  and the epigraph of the dispersion function  $\mathbf{d}$  can be represented through polynomially many linear, conic quadratic and/or semidefinite constraints. Moreover,  $\mathcal{D}$  is either the nonnegative orthant, the second-order cone or the semidefinite cone.

Together with the other assumptions, (**X**) will turn out to be *sufficient* but not *necessary* for computational tractability of the uncertainty quantification and chance constrained problems. While it is possible to replace (**X**) with a necessary *and* sufficient condition, we prefer to use (**X**) as it covers most of the practically relevant settings and avoids technicalities.

### 3 Pessimistic Chance Constraints

This section focuses on pessimistic chance constraints of the form

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} < \mathbf{u}(\mathbf{x})] \geq 1 - \epsilon, \quad (\text{PCC})$$

where the ambiguity set  $\mathcal{P}$  is defined as in (3). The conditions (**D**), (**S**) and (**A**) are tacitly assumed to hold throughout this section. In order to keep the notation clean, we use the shorthand  $\mathcal{I} = \{1, \dots, m\}$  to denote the index set of all safety conditions. Moreover, for any fixed  $\mathbf{x}$ , we let

$$\mathcal{I}(\mathbf{x}) = \left\{ i \in \mathcal{I} : \mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\xi} \geq u_i(\mathbf{x}) \text{ for some } \boldsymbol{\xi} \in \Xi \right\}$$

contain the indices of those safety conditions that can be violated. We also set  $\mathcal{I}_0 = \mathcal{I} \cup \{0\}$  and  $\mathcal{I}_0(\mathbf{x}) = \mathcal{I}(\mathbf{x}) \cup \{0\}$ . Note that the worst-case probability problem on the left-hand side of (PCC) constitutes a pessimistic uncertainty quantification problem. Below we discuss the solution of this uncertainty quantification problem (Section 3.1), derive a conic reformulation for pessimistic chance constraints (Section 3.2) and explore the limits of tractability (Section 3.3).

#### 3.1 The Uncertainty Quantification Problem

We first prove that the pessimistic uncertainty quantification problem in (PCC) admits a finite-dimensional reduction.

**Theorem 1** (Finite-Dimensional Reduction). *The worst-case probability on the left hand side of (PCC) coincides with the optimal value of the finite-dimensional optimization problem*

$$\begin{aligned}
& \underset{\lambda_i, \xi_i}{\text{minimize}} && \lambda_0 \\
& \text{subject to} && \lambda_i \in \mathbb{R}_+, \quad \xi_i \in \Xi, \quad i \in \mathcal{I}_0(\mathbf{x}) \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i = 1 \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i \xi_i = \boldsymbol{\mu} \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i \mathbf{d}(\xi_i) \preceq_{\mathcal{D}} \boldsymbol{\sigma} \\
& && \mathbf{t}_i(\mathbf{x})^\top \xi_i \geq u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}(\mathbf{x}).
\end{aligned} \tag{4}$$

**Remark 1.** *Observe that (4) can be viewed as a variant of the uncertainty quantification problem on the left-hand side of (PCC) that minimizes only over discrete distributions from within the ambiguity set  $\mathcal{P}$  with atoms or scenarios  $\xi_i$  and associated probabilities  $\lambda_i$ ,  $i \in \mathcal{I}_0(\mathbf{x})$ . Clearly, any discrete distribution corresponding to some feasible solution  $(\{\lambda_i\}_i, \{\xi_i\}_i)$  satisfies the moment conditions of the ambiguity set  $\mathcal{P}$ . The last constraint set in (4) implies that scenario  $\xi_i$  violates the  $i$ -th safety condition for  $i \in \mathcal{I}(\mathbf{x})$ . Moreover, in Proposition 3 we will show that scenario  $\xi_0$  satisfies all safety conditions at optimality if the minimum of (4) is attained and strictly positive.*

Using a continuity argument, one can show that problem (4) also quantifies the worst-case probability of the weak variant of (PCC) if we replace the index set  $\mathcal{I}(\mathbf{x})$  with  $\bar{\mathcal{I}}(\mathbf{x}) = \{i \in \mathcal{I} : \mathbf{t}_i(\mathbf{x})^\top \xi > u_i(\mathbf{x}) \text{ for some } \xi \in \Xi\}$  and  $\mathcal{I}_0(\mathbf{x})$  with  $\bar{\mathcal{I}}_0(\mathbf{x}) = \bar{\mathcal{I}}(\mathbf{x}) \cup \{0\}$ .

Problem (4) is easily interpretable because all of its feasible solutions correspond to discrete distributions from within the ambiguity set  $\mathcal{P}$ . However, it is not suitable for numerical solution. Indeed, as exemplified below, the infimum of (4) may not even be attained.

**Example 1** (Non-Existence of Optimal Solutions). *Let  $\mathcal{P}$  be the ambiguity set of all univariate distributions  $\mathbb{P}$  with mean  $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = 0$  and unrestricted support. Theorem 1 then implies that the worst-case probability  $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\tilde{\xi} < 1]$  coincides with the infimum of the following finite-dimensional*

optimization problem over two-point distributions.

$$\begin{aligned}
& \inf_{\lambda_i, \xi_i} && \lambda_0 \\
& \text{s.t.} && \lambda_i \in \mathbb{R}_+, \xi_i \in \mathbb{R}, i \in \{0, 1\} \\
& && \lambda_0 + \lambda_1 = 1 \\
& && \lambda_0 \xi_0 + \lambda_1 \xi_1 = 0 \\
& && \xi_1 \geq 1
\end{aligned} \tag{5}$$

The infimum of (5) is zero, which is attained asymptotically by the sequence  $\lambda_0(t) = \frac{1}{1+t}$ ,  $\lambda_1(t) = \frac{t}{1+t}$ ,  $\xi_0(t) = -t$  and  $\xi_1(t) = 1$  as  $t$  grows. However, the infimum is not attained by any single feasible solution as no distribution with mean zero can assign zero probability to  $\{\xi \in \mathbb{R} : \xi < 1\}$ .

It follows from [52, Corollary 3.1] that the infimum of (4) is attained if the support set  $\Xi$  is compact. Even then, however, computing a minimizer may be difficult or impossible as (4) constitutes a nonconvex optimization problem. Indeed, the moment constraints involve bilinearities which render the feasible set of (4) nonconvex. In Proposition 1 below we demonstrate that the nonconvex program (4) can be reformulated as a convex program that is often amenable to efficient numerical solution. Maybe surprisingly, the infimum of this convex program is always attained, and any minimizer can be used systematically to construct asymptotically optimal distributions for the original uncertainty quantification problem.

**Definition 1** (Perspective Functions). *The perspective function  $\mathbf{g}(\boldsymbol{\chi}, \lambda) = \lambda \mathbf{d}(\boldsymbol{\chi}/\lambda)$  of  $\mathbf{d}(\boldsymbol{\xi})$  is defined for  $\lambda > 0$ . If  $\mathbf{d}$  is  $\mathcal{D}$ -convex, proper and lower semicontinuous, then we can extend  $\mathbf{g}$  to  $\lambda = 0$  [50, Corollary 8.5.2]. In this case,  $\mathbf{g}(\boldsymbol{\chi}, 0)$  is interpreted as the recession function  $\lim_{\lambda \downarrow 0} \lambda \mathbf{d}(\boldsymbol{\chi}/\lambda)$ . Likewise, we interpret  $\boldsymbol{\chi}/0 \in \Xi$  as the requirement that  $\boldsymbol{\chi}$  belongs to the recession cone  $\text{recc}(\Xi) = \{\boldsymbol{\chi} \in \mathbb{R}^k : \boldsymbol{\xi} + \lambda \boldsymbol{\chi} \in \Xi \ \forall \boldsymbol{\xi} \in \Xi, \lambda \geq 0\}$ . Thus,  $\mathbf{0}/0 \in \Xi$  holds true whenever  $\Xi \neq \emptyset$ .*

**Proposition 1** (Convex Reformulation). *The reduced worst-case probability problem (4) has the*

same optimal value as the following convex optimization problem.

$$\begin{aligned}
& \underset{\lambda_i, \boldsymbol{\chi}_i}{\text{minimize}} && \lambda_0 \\
& \text{subject to} && \lambda_i \in \mathbb{R}_+, \boldsymbol{\chi}_i \in \mathbb{R}^k, i \in \mathcal{I}_0 \\
& && \sum_{i \in \mathcal{I}_0} \lambda_i = 1 \\
& && \frac{\boldsymbol{\chi}_i}{\lambda_i} \in \Xi \quad \forall i \in \mathcal{I}_0 \\
& && \sum_{i \in \mathcal{I}_0} \boldsymbol{\chi}_i = \boldsymbol{\mu} \\
& && \sum_{i \in \mathcal{I}_0} \lambda_i \mathbf{d} \left( \frac{\boldsymbol{\chi}_i}{\lambda_i} \right) \preceq_{\mathcal{D}} \boldsymbol{\sigma} \\
& && \mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\chi}_i \geq \lambda_i u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}
\end{aligned} \tag{6}$$

**Remark 2** (Convexity). *Note that the perspective function  $g(\boldsymbol{\chi}, \lambda) = \lambda \mathbf{d}(\boldsymbol{\chi}/\lambda)$  is  $\mathcal{D}$ -convex. Indeed, for all  $\lambda_1, \lambda_2 > 0$ ,  $\boldsymbol{\chi}_1, \boldsymbol{\chi}_2 \in \mathbb{R}^k$  and  $\theta \in [0, 1]$ , we have*

$$\begin{aligned}
& g(\theta \boldsymbol{\chi}_1 + (1 - \theta) \boldsymbol{\chi}_2, \theta \lambda_1 + (1 - \theta) \lambda_2) \\
&= (\theta \lambda_1 + (1 - \theta) \lambda_2) \mathbf{d} \left( \frac{\theta \boldsymbol{\chi}_1 + (1 - \theta) \boldsymbol{\chi}_2}{\theta \lambda_1 + (1 - \theta) \lambda_2} \right) \\
&= (\theta \lambda_1 + (1 - \theta) \lambda_2) \mathbf{d} \left( \frac{\theta \lambda_1}{\theta \lambda_1 + (1 - \theta) \lambda_2} \cdot \frac{\boldsymbol{\chi}_1}{\lambda_1} + \frac{(1 - \theta) \lambda_2}{\theta \lambda_1 + (1 - \theta) \lambda_2} \cdot \frac{\boldsymbol{\chi}_2}{\lambda_2} \right) \\
&\preceq_{\mathcal{D}} \theta \lambda_1 \mathbf{d} \left( \frac{\boldsymbol{\chi}_1}{\lambda_1} \right) + (1 - \theta) \lambda_2 \mathbf{d} \left( \frac{\boldsymbol{\chi}_2}{\lambda_2} \right) = \theta g(\boldsymbol{\chi}_1, \lambda_1) + (1 - \theta) g(\boldsymbol{\chi}_2, \lambda_2),
\end{aligned}$$

where the inequality in the last line follows from the  $\mathcal{D}$ -convexity of  $\mathbf{d}(\boldsymbol{\xi})$ . For  $\lambda_1, \lambda_2 \geq 0$  the above inequality still holds by virtue of a limiting argument, which applies as the cone  $\mathcal{D}$  is closed and as the perspective function  $\lambda \mathbf{d}(\boldsymbol{\chi}/\lambda)$  is continuous, owing to the  $\mathcal{D}$ -convexity of  $\mathbf{d}(\boldsymbol{\xi})$  and due to our definition of the perspective function for  $\lambda = 0$ . Similarly, one can show that the constraint  $\boldsymbol{\chi}/\lambda \in \Xi$  has a convex feasible set whenever  $\Xi$  is convex. This implies that (6) is indeed a convex program.

Even though the infimum of problem (4) may not be attained (see Example 1), its convex reformulation (6) is always solvable.

**Proposition 2.** *The minimum of the convex program (6) is always attained.*

Propositions 1 and 2 carry over to the weak variant of (PCC) if we replace  $\mathcal{I}$  with  $\bar{\mathcal{I}}(\mathbf{x})$  and  $\mathcal{I}_0$  with  $\bar{\mathcal{I}}_0(\mathbf{x})$ . Problem (6) thus remedies the two major shortcomings of problem (4): its minimum is

always attained, and a minimizer can be computed by leveraging convex optimization algorithms. In particular, problem (6) can be solved in polynomial time whenever condition  $(\mathbf{X})$  is satisfied.

We now show that the minimizers of (6) can be used to construct (near-)optimal distributions for the original uncertainty quantification problem (4). Indeed, Lemma 8 in the appendix shows that (6) admits a Slater-type point  $(\{\lambda'_i\}_i, \{\chi'_i\}_i)$ , that is, a feasible solution for (6) with  $\lambda'_i > 0$  for all  $i \in \mathcal{I}_0(\mathbf{x})$ . Any such Slater-type point corresponds to a discrete distribution in  $\mathcal{P}$  with a scenario of weight  $\lambda'_i > 0$  at  $\chi'_i/\lambda'_i$  for  $i \in \mathcal{I}_0(\mathbf{x})$ . Moreover, scenario  $\chi'_i/\lambda'_i$  violates the  $i$ -th safety condition for  $i \in \mathcal{I}(\mathbf{x})$ , and therefore the safety conditions are satisfied with a probability of at most  $\lambda'_0$  under this discrete distribution. As problem (6) has a convex feasible set, we can use convex combinations of  $(\{\lambda'_i\}_i, \{\chi'_i\}_i)$  with a minimizer  $(\{\lambda^*_i\}_i, \{\chi^*_i\}_i)$  of (6) to construct a sequence of points converging to  $(\{\lambda^*_i\}_i, \{\chi^*_i\}_i)$ . These points correspond to a sequence of discrete distributions in  $\mathcal{P}$  under which the probability that all safety conditions are satisfied converges to  $\lambda^*_0$  (recall that the safety conditions are satisfied with probability at least  $\lambda^*_0$  under any distribution in  $\mathcal{P}$  because  $\lambda^*_0$  equals the infimum of the original uncertainty quantification problem). Thus, the resulting sequence of discrete distributions is asymptotically optimal in the uncertainty quantification problem.

**Example 1** (cont'd). *Proposition 1 provides the following convex reformulation for the worst-case uncertainty quantification problem studied in Example 1:*

$$\begin{aligned}
& \inf_{\lambda_i, \chi_i} && \lambda_0 \\
& \text{s.t.} && \lambda_i \in \mathbb{R}_+, \chi_i \in \mathbb{R}, i \in \{0, 1\} \\
& && \lambda_0 + \lambda_1 = 1, \chi_0 + \chi_1 = 0 \\
& && \chi_1 \geq \lambda_1
\end{aligned} \tag{7}$$

One readily verifies that  $(\boldsymbol{\lambda}^*, \boldsymbol{\chi}^*) = ([0, 1], [-1, 1])$  minimizes problem (7), and that  $(\boldsymbol{\lambda}', \boldsymbol{\chi}') = ([\frac{1}{2}, \frac{1}{2}], [-\frac{1}{2}, \frac{1}{2}])$  constitutes a Slater-type point. The asymptotically optimal sequence  $(\boldsymbol{\lambda}(t), \boldsymbol{\xi}(t)) = ([\frac{1}{1+t}, \frac{t}{1+t}], [-t, 1])$  to the nonconvex worst-case uncertainty quantification problem (5) then corresponds to the convex combinations  $\frac{t-1}{1+t}(\boldsymbol{\lambda}^*, \boldsymbol{\chi}^*) + \frac{2}{1+t}(\boldsymbol{\lambda}', \boldsymbol{\chi}')$  in the convex problem (7).

**Proposition 3.** *If the minimum of the convex program (6) is strictly positive, then every optimal solution  $(\{\lambda_i\}_i, \{\chi_i\}_i)$  satisfies the extra constraints*

$$\mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\chi}_0 < \lambda_0 u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}. \tag{8}$$

The findings of this section are closely related to the results by Han *et al.* [34] for uncertainty quantification problems. Indeed, Han *et al.* find a similar finite reduction for ambiguity sets involving linear (instead of conic) moment conditions. However, their result relies on the critical assumption that  $\mathcal{I}(\mathbf{x}) = \mathcal{I}$  for all  $\mathbf{x} \in \mathcal{X}$ , that is, that none of the safety conditions are redundant. While this assumption is natural in uncertainty quantification, it is not tenable in chance constrained programming, where safety conditions may be redundant for some—but not all—choices of  $\mathbf{x} \in \mathcal{X}$ . Moreover, our proof of the finite reduction theorem leverages the ‘*primal worst equals dual best*’ duality scheme by Beck and Ben-Tal [2] and the convexification technique by Gorissen *et al.* [32], and it is thus fundamentally different from the derivations in [34], which rely on the Richter-Rogosinski theorem [53, Theorem 7.37]. We believe that our proof reveals a possibly fruitful connection between distributionally robust optimization and the duality scheme by Beck and Ben-Tal, which may have further ramifications beyond chance constrained programming.

### 3.2 The Chance Constrained Program

In this section, we assume that the assumptions  $(\mathbf{D}')$ ,  $(\mathbf{S}')$ ,  $(\mathbf{A})$  and  $(\mathbf{T})$  are satisfied. Our conic reformulation relies on the following preparatory lemma.

**Lemma 1.** *For any fixed  $\boldsymbol{\gamma} \in \mathcal{D}^*$ , the conjugate of the convex function  $\boldsymbol{\gamma}^\top \mathbf{d}(\boldsymbol{\xi})$  is given by*

$$(\boldsymbol{\gamma}^\top \mathbf{d})^*(\boldsymbol{\nu}) = \sigma_{\text{epi}(\mathbf{d})}(\boldsymbol{\nu}, -\boldsymbol{\gamma}),$$

where  $\text{epi}(\mathbf{d}) = \{(\boldsymbol{\xi}, \boldsymbol{\eta}) \in \mathbb{R}^k \times \mathbb{R}^d : \mathbf{d}(\boldsymbol{\xi}) \preceq_{\mathcal{D}} \boldsymbol{\eta}\}$  denotes the  $\mathcal{D}$ -epigraph of  $\mathbf{d}(\boldsymbol{\xi})$ .

**Theorem 2** (Pessimistic Chance Constraints). *The pessimistic chance constraint (PCC) is satisfied if and only if there exist  $\boldsymbol{\beta} \in \mathbb{R}^k$ ,  $\boldsymbol{\gamma} \in \mathcal{D}^*$ ,  $\tau_i \in \mathbb{R}_+$ ,  $i \in \mathcal{I}$ , and  $\boldsymbol{\nu}_i \in \mathbb{R}^k$ ,  $i \in \mathcal{I}_0$ , such that*

$$\begin{aligned} 1 + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} &\geq 1 - \epsilon \\ \boldsymbol{\nu}_0 - \boldsymbol{\beta} &\in \Xi^* \\ \boldsymbol{\nu}_i - \boldsymbol{\beta} - \tau_i \mathbf{t}_i &\in \Xi^* && \forall i \in \mathcal{I} \\ (-\boldsymbol{\nu}_i, \boldsymbol{\gamma}) &\in \text{epi}(\mathbf{d})^* && \forall i \in \mathcal{I}_0 \\ \left\| \begin{pmatrix} 2 \\ \tau_i - u_i(\mathbf{x}) \end{pmatrix} \right\|_2 &\leq \tau_i + u_i(\mathbf{x}) && \forall i \in \mathcal{I}. \end{aligned} \tag{9}$$

Since  $\mathbf{u}(\mathbf{x})$  is affine, (9) is a system of linear, conic quadratic and/or semidefinite constraints whenever  $(\mathbf{X})$  is satisfied. In this case, linear programs involving pessimistic chance constraints are computationally tractable and can be solved efficiently with modern interior point algorithms.

Under the assumption that the strict pessimistic chance constrained program (2) is feasible, Theorem 2 applies in the same way to weak chance constraints. In particular, the strict and weak pessimistic chance constrained programs thus share the same optimal value.

Due to its reliance on the condition  $(\mathbf{T})$ , which requires the decisions and uncertain parameters to appear on different sides of the safety conditions, Theorem 2 is reminiscent of Prékopa's celebrated convexity result for classical joint chance constraints subject to log-concave probability distributions [49]. We emphasize, however, that checking the feasibility of a classical chance constraint is hard even if there is only a single safety condition satisfying  $(\mathbf{T})$  and even in the simplest probabilistic setting. For example, assume that  $\tilde{\xi}$  follows the uniform distribution on the standard hypercube in  $\mathbb{R}^k$ , which is evidently log-concave. Then, the probability that the safety condition is satisfied coincides with the volume of the knapsack polytope that emerges from intersecting the hypercube containing all possible scenarios with the halfspace containing all safe scenarios. However, computing the volume of a knapsack polytope is  $\#P$ -hard [25]. Thus, checking the feasibility of a classical individual (not even joint) chance constraint is  $\#P$ -hard even when Prékopa's conditions are all satisfied and the chance constraint has a convex feasible set.

**Remark 3** (Violations of Model Assumptions). *In the following, we show that any violation of the conditions  $(\mathbf{D}')$ ,  $(\mathbf{S}')$  and  $(\mathbf{T})$  renders the pessimistic chance constrained program NP-hard. Nevertheless, one can readily adapt the results of this section to derive nonconvex reformulations for general pessimistic chance constrained programs that only satisfy the weaker conditions  $(\mathbf{D})$ ,  $(\mathbf{S})$  and  $(\mathbf{A})$ . The resulting reformulations can be solved approximately with a sequential convex optimization scheme, see [35, 40]. We will make use of such a scheme in Section 5.1.*

### 3.3 Complexity Analysis

We now show that the tractability result from the previous section is tight, that is, any violation of the conditions  $(\mathbf{D}')$ ,  $(\mathbf{S}')$  and  $(\mathbf{T})$  leads to an NP-hard optimization problem, even if all other conditions are satisfied. In the following three sections, we investigate each condition in turn.



### 3.3.1 Intractability of Nonhomogeneous Dispersion Measures

In this section, we consider instances of the problem (2) that satisfy the conditions **(A)**, **(S')**, **(T)** and **(X)** but violate the condition **(D')**. We show that such instances give rise to strongly NP-hard problems even if they satisfy the weaker condition **(D)** from Section 1. To this end, we recall the strongly NP-hard Integer Programming (IP) Feasibility problem [28]:

<p style="text-align: center; margin: 0;">INTEGER PROGRAMMING FEASIBILITY.</p> <p><b>Instance.</b> Given are <math>\mathbf{A} \in \mathbb{Z}^{m \times n}</math> and <math>\mathbf{b} \in \mathbb{Z}^m</math>.</p> <p><b>Question.</b> Is there a vector <math>\mathbf{x} \in \{0, 1\}^n</math> such that <math>\mathbf{Ax} \leq \mathbf{b}</math>?</p>
---

In fact, the IP Feasibility problem is solvable whenever there is a fractional solution whose components are sufficiently close to 0 or 1. Similar results have been reported in [60, 61]; we include the proof to keep the paper self-contained.

**Lemma 2.** *Fix any  $\kappa < \min_i \{(\sum_j |A_{ij}|)^{-1}\}$  that satisfies  $\kappa < \frac{1}{2}$ . The IP Feasibility problem has an affirmative answer if and only if there is a vector  $\mathbf{y} \in ([0, \kappa] \cup [1 - \kappa, 1])^n$  such that  $\mathbf{Ay} \leq \mathbf{b}$ .*

For a fixed instance  $(\mathbf{A}, \mathbf{b})$  of the IP Feasibility problem, consider the following instance of (2)

$$\begin{aligned}
 & \underset{\mathbf{x}}{\text{minimize}} && 0 \\
 & \text{subject to} && \mathbf{x} \in [-1, 1]^n \\
 & && \mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b} \\
 & && \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( -3\mathbf{e} < \mathbf{x} + \tilde{\boldsymbol{\xi}} < 3\mathbf{e} \right) \geq 1 - \epsilon,
 \end{aligned} \tag{10}$$

where  $\epsilon \in (0, 1)$  and the ambiguity set is defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^k) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \mathbf{0}, \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{\xi}}|] \leq \frac{4(\epsilon - \delta)}{n} \mathbf{e}, \mathbb{E}_{\mathbb{P}}[\mathbf{d}(\tilde{\boldsymbol{\xi}})] \leq \frac{\delta^2}{n} \mathbf{e} \right\} \tag{11}$$

with  $\mathbf{d}(\boldsymbol{\xi}) = \max\{\boldsymbol{\xi} - (4 - \delta)\mathbf{e}, \mathbf{0}, -\boldsymbol{\xi} - (4 - \delta)\mathbf{e}\}$  and  $\delta \in (0, \frac{\epsilon}{8}]$ . Here, the absolute value and maximum operators apply component-wise. Problem (10) is a feasibility problem that evaluates to zero if it is feasible and to  $+\infty$  otherwise. One readily verifies that the problem satisfies the conditions **(A)**, **(D)**, **(S')**, **(T)** and **(X)**, but it violates the condition **(D')**.

We now show that there is a one-to-one correspondence between solutions  $\mathbf{z} \in \{0, 1\}^n$  to the IP Feasibility problem and solutions  $\mathbf{x} = 2\mathbf{z} - \mathbf{e}$  to problem (10). To this end, note that the last condition in (11) stipulates that most of the probability mass of distributions  $\mathbb{P} \in \mathcal{P}$  is placed on

realizations  $\boldsymbol{\xi} \in [-(4 - \delta), 4 - \delta]^n$ . We can exploit this observation to show that a decision  $\boldsymbol{x}$  is ‘nearly binary’ if and only if it satisfies the chance constraint in (10).

**Lemma 3.** *For  $\boldsymbol{x} \in \{-1, 1\}^n$ , the worst-case probability on the left-hand side of the chance constraint in (10) amounts to at least  $1 - \epsilon$ .*

**Lemma 4.** *For  $\boldsymbol{x} \in [-1, 1]^n \setminus ([-1, -1 + \delta] \cup [1 - \delta, 1])^n$ , the worst-case probability on the left-hand side of the chance constraint in (10) is strictly less than  $1 - \epsilon$ .*

We are now in the position to prove the NP-hardness of the chance constrained program (10).

**Theorem 3.** *The pessimistic joint chance constrained program (2) is strongly NP-hard whenever the condition  $(\mathbf{D}')$  is replaced with  $(\mathbf{D})$ , even if the conditions  $(\mathbf{A})$ ,  $(\mathbf{S}')$ ,  $(\mathbf{T})$  and  $(\mathbf{X})$  are satisfied.*

### 3.3.2 Intractability of Nonconic Supports

We now consider instances of the problem (2) that satisfy the conditions  $(\mathbf{A})$ ,  $(\mathbf{D}')$ ,  $(\mathbf{T})$  and  $(\mathbf{X})$  but violate the condition  $(\mathbf{S}')$ . We show that such instances are strongly NP-hard even in the absence of any information about the distributions’ dispersion and even if the support  $\Xi$  is a hyperrectangle. To this end, we reduce the IP Feasibility problem from the previous section to the problem

$$\begin{aligned} & \underset{\boldsymbol{x}, y}{\text{minimize}} && y \\ & \text{subject to} && \boldsymbol{x} \in [-1, 1]^n, \quad y \in \mathbb{R}_+ \\ & && \mathbf{A}(\boldsymbol{x} + \mathbf{e})/2 \leq \mathbf{b} \\ & && \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( -\tilde{\xi}_{n+i} - y < x_i < \tilde{\xi}_i + y \quad \forall i = 1, \dots, n \right) \geq 1 - \epsilon, \end{aligned} \tag{12}$$

where  $\epsilon \in (0, 1)$ ,  $\mathbf{A} \in \mathbb{Z}^{m \times n}$  and  $\mathbf{b} \in \mathbb{Z}^m$  correspond to the respective input parameters of the IP Feasibility problem, and the ambiguity set  $\mathcal{P}$  satisfies

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0([-1, 3]^{2n}) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \left( 3 - \frac{2\epsilon}{n} \right) \mathbf{e} \right\}. \tag{13}$$

One readily verifies that (12) satisfies the conditions  $(\mathbf{A})$ ,  $(\mathbf{D}')$ ,  $(\mathbf{T})$  and  $(\mathbf{X})$ , but it violates  $(\mathbf{S}')$ .

We now establish a one-to-one correspondence between solutions  $\boldsymbol{z} \in \{0, 1\}^n$  to the IP Feasibility problem and solutions  $\boldsymbol{x} = 2\boldsymbol{z} - \mathbf{e}$  to problem (12) that achieve an optimal value of zero. To this end, we first prove that a weak variant of the chance constraint in (12) enforces binarity of  $\boldsymbol{x}$ .

**Lemma 5.** For  $\mathbf{x} \in [-1, 1]^n$ , the following weak variant of the chance constraint in (12) satisfies:

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( -\tilde{\xi}_{n+i} \leq x_i \leq \tilde{\xi}_i \quad \forall i = 1, \dots, n \right) = \max \left\{ 0, 1 - \epsilon - \frac{2\epsilon}{n} \left( \sum_{i: x_i \in (-1, 1)} \frac{3 + x_i^2}{18 - 2x_i^2} \right) \right\} \quad (14)$$

Lemma 5 essentially shows that  $\mathbf{x} \in [-1, 1]^n$  achieves an objective value of zero in problem (12) if and only if  $\mathbf{x} \in \{-1, 1\}^n$ . Equipped with the insight from Lemma 5, we can now prove the postulated intractability result.

**Theorem 4.** The pessimistic joint chance constrained program (2) is strongly NP-hard whenever the condition (S') is replaced with (S), even if the conditions (A), (D'), (T) and (X) are satisfied.

### 3.3.3 Intractability of Left-Hand Side Uncertainty

We now study instances of the problem (2) that satisfy the conditions (A), (D'), (S') and (X) but violate the condition (T). We show that such instances are computationally intractable even in the absence of support information and even if the mean absolute deviation—arguably one of the most basic dispersion measures—is used to quantify the distributions' spread. To this end, we reduce the IP Feasibility problem from Section 3.3.1 to the following problem

$$\begin{aligned} & \underset{\mathbf{x}, y}{\text{minimize}} && y \\ & \text{subject to} && \mathbf{x} \in [0, 1]^n, \quad y \geq 0 \\ & && \mathbf{Ax} \leq \mathbf{b} \\ & && \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( (\tilde{\xi}_i + 1)x_i + y > 0, \quad (\tilde{\xi}_{n+i} + 1)(1 - x_i) + y > 0 \quad \forall i = 1, \dots, n \right) \geq 1 - \epsilon, \end{aligned} \quad (15)$$

where  $\epsilon \in (0, 1/2)$ ,  $\mathbf{A} \in \mathbb{Z}^{m \times n}$  and  $\mathbf{b} \in \mathbb{Z}^m$  correspond to the respective input parameters of the IP Feasibility problem, and the ambiguity set  $\mathcal{P}$  is defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{2n}) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \mathbf{0}, \quad \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{\xi}}|] \leq \frac{2\epsilon}{n} \mathbf{e} \right\}. \quad (16)$$

One verifies that the problem satisfies the conditions (A), (D'), (S') and (X), but it violates (T).

We again derive an analytical expression for a weak variant of the chance constraint in (15).

**Lemma 6.** For  $\mathbf{x} \in [0, 1]^n$ , the following weak variant of the chance constraint in (15) satisfies:

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( (\tilde{\xi}_i + 1)x_i \geq 0, \quad (\tilde{\xi}_{n+i} + 1)(1 - x_i) \geq 0 \quad \forall i = 1, \dots, n \right) = 1 - \epsilon - \frac{\epsilon}{n} |\{i \in \{1, \dots, n\} : x_i \in (0, 1)\}| \quad (17)$$

Lemma 6 essentially implies that  $\mathbf{x} \in [0, 1]^n$  achieves an objective value of zero in problem (15) if and only if  $\mathbf{x} \in \{0, 1\}^n$ . This implies the postulated intractability result.

**Theorem 5.** *The pessimistic joint chance constrained program (2) is strongly NP-hard whenever the condition  $(\mathbf{T})$  is violated, even if the conditions  $(\mathbf{A})$ ,  $(\mathbf{D}')$ ,  $(\mathbf{S}')$  and  $(\mathbf{X})$  are satisfied.*

## 4 Optimistic Chance Constraints

Next, we investigate optimistic chance constraints of the form

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} < \mathbf{u}(\mathbf{x})] \geq 1 - \epsilon. \quad (\text{OCC})$$

Note that the best-case probability problem on the left-hand side of (OCC) constitutes an optimistic uncertainty quantification problem. In analogy to Section 3, we discuss the solution of this uncertainty quantification problem (Section 4.1), derive a conic reformulation for the optimistic chance constraint (Section 4.2) and explore the limits of tractability (Section 4.3). The conditions  $(\mathbf{D})$ ,  $(\mathbf{S})$  and  $(\mathbf{A})$  are tacitly assumed to hold throughout this section.

### 4.1 The Uncertainty Quantification Problem

We start by proving that the optimistic uncertainty quantification problem in (OCC) admits a finite-dimensional reduction.

**Theorem 6** (Finite-Dimensional Reduction). *Assume that there exists  $\boldsymbol{\xi} \in \Xi$  with  $\mathbf{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})$  as otherwise (OCC) is not satisfiable. Then, the best-case probability on the left-hand side of (OCC) is given by the optimal value of the finite-dimensional optimization problem*

$$\begin{aligned} & \underset{\lambda_i, \boldsymbol{\xi}_i}{\text{maximize}} && \lambda_0 \\ & \text{subject to} && \lambda_i \in \mathbb{R}_+, \quad \boldsymbol{\xi}_i \in \Xi, \quad i \in \{0, 1\} \\ & && \sum_{i \in \{0, 1\}} \lambda_i = 1 \\ & && \sum_{i \in \{0, 1\}} \lambda_i \boldsymbol{\xi}_i = \boldsymbol{\mu} \\ & && \sum_{i \in \{0, 1\}} \lambda_i \mathbf{d}(\boldsymbol{\xi}_i) \preceq_{\mathcal{D}} \boldsymbol{\sigma} \\ & && \mathbf{T}(\mathbf{x})\boldsymbol{\xi}_0 \leq \mathbf{u}(\mathbf{x}). \end{aligned} \quad (18)$$

**Remark 4.** *Observe that (18) can be viewed as a variant of the uncertainty quantification problem on the left-hand side of (OCC) that minimizes only over two-point distributions from within the ambiguity set  $\mathcal{P}$  with scenarios  $\xi_i$  and associated probabilities  $\lambda_i$ ,  $i \in \{0, 1\}$ . Clearly, any two-point distribution corresponding to some feasible solution  $(\{\lambda\}_i, \{\xi\}_i)$  satisfies the moment conditions of the ambiguity set  $\mathcal{P}$ . The last constraint in (18) implies that scenario  $\xi_0$  satisfies all safety conditions. Moreover, in Proposition 6 we will show that scenario  $\xi_1$  violates at least one safety condition at optimality if the maximum of (18) is attained and strictly smaller than 1.*

Theorem 6 immediately extends to weak optimistic chance constraints if we assume that there is a realization  $\xi \in \Xi$  that satisfies the safety condition weakly, that is, if  $\xi$  satisfies  $\mathbf{T}(\mathbf{x})\xi \leq \mathbf{u}(\mathbf{x})$ .

Problem (18) is easily interpretable because all of its feasible solutions correspond to two-point distributions from within  $\mathcal{P}$ . However, as in the case of the pessimistic uncertainty quantification problem studied in Section 3.1, problem (18) is nonconvex and may not even be solvable.

**Example 2** (Non-Existence of Optimal Solutions). *Let  $\mathcal{P}$  be the ambiguity set of all univariate distributions  $\mathbb{P}$  with mean  $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = 0$ , mean-absolute deviation  $\mathbb{E}_{\mathbb{P}}[|\tilde{\xi}|] \leq 2$  and unrestricted support. Theorem 6 then implies that the best-case probability  $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\tilde{\xi} < -1]$  coincides with the supremum of the following finite-dimensional optimization problem over two-point distributions.*

$$\begin{aligned}
& \sup_{\lambda_i, \xi_i} && \lambda_0 \\
& \text{s.t.} && \lambda_i \in \mathbb{R}_+, \xi_i \in \mathbb{R}, i \in \{0, 1\} \\
& && \lambda_0 + \lambda_1 = 1 \\
& && \lambda_0 \xi_0 + \lambda_1 \xi_1 = 0 \\
& && \lambda_0 |\xi_0| + \lambda_1 |\xi_1| \leq 2 \\
& && \xi_0 \leq -1
\end{aligned} \tag{19}$$

*The supremum of (19) is 1, which is attained asymptotically by the sequence  $\lambda_0(t) = \frac{t}{1+t}$ ,  $\lambda_1(t) = \frac{1}{1+t}$ ,  $\xi_0(t) = -1$  and  $\xi_1(t) = t$  as  $t$  grows. However, the supremum is not attained by any single feasible solution as no distribution with mean zero can assign probability 1 to  $\{\xi \in \mathbb{R} : \xi < -1\}$ .*

As in the pessimistic case, it follows from [52, Corollary 3.1] that the supremum of (18) is attained if the support set  $\Xi$  is compact. Even then, however, computing a maximizer may be difficult as (18) constitutes a nonconvex optimization problem. In Proposition 4 below we demonstrate that

the nonconvex program (18) can be reformulated as a convex program whose maximum is attained and that is amenable to numerical solution.

**Proposition 4** (Convex Reformulation). *If there exists  $\boldsymbol{\xi} \in \Xi$  with  $\mathbf{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})$ , then problem (18) has the same optimal value as the following convex optimization problem.*

$$\begin{aligned}
& \underset{\lambda_i, \boldsymbol{\chi}_i}{\text{maximize}} && \lambda_0 \\
& \text{subject to} && \lambda_i \in \mathbb{R}_+, \quad \boldsymbol{\chi}_i \in \mathbb{R}^k, \quad i \in \{0, 1\} \\
& && \sum_{i \in \{0, 1\}} \lambda_i = 1 \\
& && \frac{\boldsymbol{\chi}_i}{\lambda_i} \in \Xi && \forall i \in \{0, 1\} \\
& && \sum_{i \in \{0, 1\}} \boldsymbol{\chi}_i = \boldsymbol{\mu} \\
& && \sum_{i \in \{0, 1\}} \lambda_i \mathbf{d} \left( \frac{\boldsymbol{\chi}_i}{\lambda_i} \right) \preceq_{\mathcal{D}} \boldsymbol{\sigma} \\
& && \mathbf{T}(\mathbf{x})\boldsymbol{\chi}_0 \leq \lambda_0 \mathbf{u}(\mathbf{x})
\end{aligned} \tag{20}$$

The reasoning in Remark 2 can be used to show that (20) is indeed a convex program and is therefore a good candidate for numerical solution. Indeed, problem (20) can be solved in polynomial time whenever condition **(X)** is satisfied. Moreover, even though the supremum of problem (18) may not be attained (see Example 2), its convex reformulation (20) is always solvable.

**Proposition 5.** *The maximum of the convex program (20) is always attained.*

Propositions 4 and 5 carry over to weak optimistic chance constraints. In that case, we only need to assume in Proposition 4 that there is a  $\boldsymbol{\xi} \in \Xi$  that satisfies the safety condition weakly.

Next, we argue that the maximizers of (20) can be used to construct (near-)optimal distributions for the original uncertainty quantification problem. Indeed, whenever there is  $\boldsymbol{\xi} \in \Xi$  with  $\mathbf{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})$ , Lemma 10 in the appendix shows that (20) admits a Slater-type point  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$ , that is, a feasible solution to (20) with  $\lambda'_i > 0$  for  $i \in \{0, 1\}$ . Any such Slater-type point corresponds to a two-point distribution in  $\mathcal{P}$  with a scenario of strictly positive weight  $\lambda'_i$  at  $\boldsymbol{\chi}'_i/\lambda'_i$  for  $i \in \{0, 1\}$ . We can then use convex combinations of  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$  with a minimizer  $(\{\lambda^*_i\}_i, \{\boldsymbol{\chi}^*_i\}_i)$  of (20) to construct a sequence of points converging to  $(\{\lambda^*_i\}_i, \{\boldsymbol{\chi}^*_i\}_i)$ . These points correspond to two-point distributions in  $\mathcal{P}$  under which the probability that all safety conditions are satisfied converges

to  $\lambda_0^*$ . Thus, the resulting sequence of two-point distributions is asymptotically optimal in the uncertainty quantification problem.

**Example 2** (cont'd). *Proposition 4 provides the following convex reformulation for the best-case uncertainty quantification problem studied in Example 2:*

$$\begin{aligned}
& \sup_{\lambda_i, \chi_i} \lambda_0 \\
& \text{s.t.} \quad \lambda_i \in \mathbb{R}_+, \chi_i \in \mathbb{R}, i \in \{0, 1\} \\
& \quad \lambda_0 + \lambda_1 = 1, \chi_0 + \chi_1 = 0, |\chi_0| + |\chi_1| \leq 2 \\
& \quad \chi_0 \leq -\lambda_0
\end{aligned} \tag{21}$$

One readily verifies that  $(\boldsymbol{\lambda}^*, \boldsymbol{\chi}^*) = ([1, 0], [-1, 1])$  minimizes problem (21), and that  $(\boldsymbol{\lambda}', \boldsymbol{\chi}') = ([\frac{1}{2}, \frac{1}{2}], [-\frac{1}{2}, \frac{1}{2}])$  constitutes a Slater-type point. The asymptotically optimal sequence  $(\boldsymbol{\lambda}(t), \boldsymbol{\xi}(t)) = ([\frac{t}{1+t}, \frac{1}{1+t}], [-1, t])$  to the nonconvex best-case uncertainty quantification problem (19) then corresponds to the convex combinations  $\frac{t-1}{1+t}(\boldsymbol{\lambda}^*, \boldsymbol{\chi}^*) + \frac{2}{1+t}(\boldsymbol{\lambda}', \boldsymbol{\chi}')$  in the convex problem (21).

**Proposition 6.** *If the supremum of the convex program (20) is strictly smaller than 1, then every optimal solution satisfies the (nonconvex) extra constraint*

$$\mathbf{T}(\mathbf{x})\boldsymbol{\chi}_1 \not\leq \lambda_1 \mathbf{u}(\mathbf{x}). \tag{22}$$

## 4.2 The Chance Constrained Program

In addition to the assumptions **(D)**, **(S)** and **(A)**, our conic reformulation for optimistic chance constraints only requires satisfaction of the separability condition **(T)**, which we assume to hold throughout this section. The assumptions **(D')** and **(S')**, which were crucial to derive the conic reformulation of pessimistic chance constraints, are no longer necessary.

**Theorem 7** (Optimistic Chance Constraints). *The best-case chance constraint (OCC) for  $\epsilon \in [0, 1)$  is satisfied if and only if there exist  $\lambda \in \mathbb{R}_+$  and  $\boldsymbol{\xi}_i \in \mathbb{R}^k$ ,  $i \in \{0, 1\}$ , with*

$$\begin{aligned}
\lambda &\leq \frac{\epsilon}{1-\epsilon}, & \mathbf{T}\boldsymbol{\xi}_0 &\leq \mathbf{u}(\mathbf{x}) \\
\boldsymbol{\xi}_0 &\in \Xi, & \frac{\boldsymbol{\xi}_1}{\lambda} &\in \Xi \\
\sum_{i \in \{0,1\}} \boldsymbol{\xi}_i &= (1+\lambda)\boldsymbol{\mu}, & \mathbf{d}(\boldsymbol{\xi}_0) + \lambda \mathbf{d}\left(\frac{\boldsymbol{\xi}_1}{\lambda}\right) &\preceq_{\mathcal{D}} (1+\lambda)\boldsymbol{\sigma}.
\end{aligned} \tag{23}$$

**Remark 5.** *If the assumptions  $(\mathbf{D}')$ ,  $(\mathbf{S}')$ , and  $(\mathbf{T})$  hold, which were needed in Theorem 2 to ensure the tractability of the worst-case chance constraint, then we may set  $\lambda = \epsilon/(1 - \epsilon)$ , whereby the convex reformulation (23) of the best-case chance constraint simplifies to*

$$\begin{aligned} \mathbf{T}\boldsymbol{\xi}_0 &\leq (1 - \epsilon)\mathbf{u}(\mathbf{x}), & \boldsymbol{\xi}_i &\in \Xi, i \in \{0, 1\} \\ \sum_{i \in \{0,1\}} \boldsymbol{\xi}_i &= \boldsymbol{\mu}, & \sum_{i \in \{0,1\}} \mathbf{d}(\boldsymbol{\xi}_i) &\preceq_{\mathcal{D}} \boldsymbol{\sigma}. \end{aligned} \tag{24}$$

Theorem 7 carries over to weak optimistic chance constraints without any modifications. Note that (23) and (24) can be solved in polynomial time whenever condition  $(\mathbf{X})$  is satisfied.

**Remark 6** (Violations of Model Assumptions). *In analogy to Remark 3, one can readily adapt the results of this section to derive a nonconvex reformulation for general optimistic chance constrained programs that violate the condition  $(\mathbf{T})$ . The resulting reformulation can again be solved approximately with a sequential convex optimization scheme [35, 40].*

### 4.3 Complexity Analysis

We now show that instances of the problem (OCC) that violate the assumption  $(\mathbf{T})$  give rise to intractable optimization problems even in the absence of any information about the distributions' spread and even if the support  $\Xi$  is a hyperrectangle. To this end, we reduce the IP Feasibility problem from Section 3.3.1 to the feasibility problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} & \quad 0 \\ \text{subject to} & \quad \mathbf{x} \in [-1, 1]^n \\ & \quad \mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b} \\ & \quad \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( \tilde{\boldsymbol{\xi}}^\top \mathbf{x} > n - 2\kappa \right) \geq 1/2, \end{aligned} \tag{25}$$

where  $\mathbf{A} \in \mathbb{Z}^{m \times n}$  and  $\mathbf{b} \in \mathbb{Z}^m$  correspond to the respective input parameters of the IP Feasibility problem,  $\kappa$  is chosen as prescribed by Lemma 2, and the ambiguity set  $\mathcal{P}$  is defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0([-1, 1]^n) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \mathbf{0} \right\}.$$

One easily verifies that problem (25) satisfies the conditions  $(\mathbf{D})$ ,  $(\mathbf{S})$  and  $(\mathbf{A})$ , but it violates  $(\mathbf{T})$ .

**Theorem 8.** *The optimistic joint chance constrained program (2) becomes strongly NP-hard whenever the condition  $(\mathbf{T})$  is violated, even if the conditions  $(\mathbf{D})$ ,  $(\mathbf{S})$  and  $(\mathbf{A})$  are satisfied.*



## 5 Numerical Experiments

We now compare our exact reformulations of the pessimistic and optimistic chance constraints with popular approximations from the literature in the context of project management (Section 5.1) and image reconstruction (Section 5.2). The experiments also provide insights into the efficacy of a well-known sequential convex optimization scheme in problems where the ambiguity set violates our regularity conditions. All optimization problems are solved using CPLEX 12.5 on an 8-core 3.4GHz computer with 16 GB RAM.<sup>1</sup>

### 5.1 Project Management

We identify a project with a directed, acyclic graph  $G = (V, E)$  whose nodes  $V = \{1, \dots, n\}$  represent the tasks and whose arcs  $E \subseteq V \times V$  denote temporal precedences between the tasks: if  $(i, j) \in E$ , then task  $j$  can only start after task  $i$  has been completed. The duration of task  $i \in V$  is a random quantity given by  $d_i(\mathbf{x}; \tilde{\boldsymbol{\xi}}) = (1 + \tilde{\xi}_i)d_i^0 - x_i$ , where  $d_i^0$  denotes the nominal task duration,  $\tilde{\xi}_i$  represents exogenous fluctuations (*e.g.*, due to weather conditions, machine downtimes or staff shortage) and  $x_i$  is the amount of a nonrenewable resource (*e.g.*, capital or manpower) that is used to expedite the task. We assume that the probability distribution governing the uncertain fluctuations  $\tilde{\boldsymbol{\xi}}$  of all tasks belongs to the ambiguity set

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) : \mathbb{E}_{\mathbb{P}} [\tilde{\boldsymbol{\xi}}] = \mathbf{0}, \mathbb{E}_{\mathbb{P}} \left[ \max \left\{ -s^- \tilde{\boldsymbol{\xi}}, s^+ \tilde{\boldsymbol{\xi}} \right\} \right] \leq \boldsymbol{\sigma}, \mathbb{P} \left( \tilde{\boldsymbol{\xi}} \in [\underline{\boldsymbol{\xi}}, \bar{\boldsymbol{\xi}}] \right) = 1 \right\}, \quad (26)$$

where  $s^-, s^+ \in \mathbb{R}_+$ ,  $\boldsymbol{\sigma} \in \mathbb{R}_+^n$  and  $\underline{\boldsymbol{\xi}}, \bar{\boldsymbol{\xi}} \in \mathbb{R}^n$ . Our model of the task durations is reminiscent of the classical PERT model, which fits optimistic, pessimistic and most likely estimates of the non-expedited task durations  $(1 + \tilde{\xi}_i)d_i^0$  to Beta distributions [22]. In our setting, we identify the expected duration with  $d_i^0$ , while the support  $[\underline{\xi}_i, \bar{\xi}_i]$  and the semi-deviation parameters  $s^-, s^+$  and  $\sigma_i$  express our beliefs about the optimistic and pessimistic task durations. This resembles the use of forward and backward deviations, which have been employed in [18] to describe durations of project tasks that are governed by asymmetric distributions.

We seek for a resource allocation  $\mathbf{x}$  that minimizes the worst-case value-at-risk of the project's makespan (*i.e.*, the time required to complete all tasks). Following [7, 58], this problem can be

---

<sup>1</sup>IBM ILOG CPLEX Optimizer: <http://www.ibm.com/software/commerce/optimization/cplex-optimizer/>.

formulated as the following instance of problem (2)

$$\begin{aligned}
& \underset{\tau, \mathbf{x}}{\text{minimize}} && \tau \\
& \text{subject to} && \tau \in \mathbb{R}_+, \quad \mathbf{x} \in \mathcal{X} \\
& && \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( \tau > \sum_{i \in T_l} d_i(\mathbf{x}; \tilde{\xi}) \quad \forall T_l \in \mathcal{T} \right) \geq 1 - \epsilon,
\end{aligned} \tag{27}$$

where  $\epsilon$  is the decision maker's risk tolerance level,  $\mathcal{X}$  denotes the set of feasible resource allocations and  $\mathcal{T}$  is the set of all inclusion-maximal paths in the project network. A path  $T = \{i_1, \dots, i_t\} \subseteq V$ ,  $(i_1, i_2), \dots, (i_{t-1}, i_t) \in E$ , is called inclusion-maximal if no other path  $T' = \{i'_1, \dots, i'_t\}$  satisfies  $T' \supset T$  and  $T' \neq T$ . The chance constraint thus requires  $\tau$  to exceed the duration of the longest task path in the project network, which coincides with the project's makespan, with high probability. Note that the size of the set  $\mathcal{T}$  and hence the number of safety conditions inside the chance constraint in (27) typically grows exponentially in the description  $G$  of the project [59]. This is not surprising as it has been shown in [59, Theorem 2.1] that problem (27) is strongly NP-hard even if  $\epsilon = 0$ .

We compare our reformulation of problem (27) with two popular bounding schemes from the literature. The first one (hereafter called 'Bonferroni approximation') uses Bonferroni's inequality to conservatively approximate the joint chance constraint in (27) by individual chance constraints

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( \tau > \sum_{i \in T_l} d_i(\mathbf{x}; \tilde{\xi}) \right) \geq 1 - \epsilon_l \quad \forall T_l \in \mathcal{T},$$

where the individual risk tolerances  $\epsilon_l$  become additional decision variables that have to satisfy  $\sum_{T_l \in \mathcal{T}} \epsilon_l = \epsilon$ , see [58]. This problem fails to be convex, and we solve it approximately using a sequential convex optimization scheme, see [6, 16]. The second bounding scheme ('LDR approximation') first conservatively approximates the task start times by linear decision rules and then applies the Bonferroni approximation to the resulting constraints, see [18]. In this scheme, too, we employ a sequential convex optimization scheme to optimize over the individual risk levels  $\epsilon_l$ .

In our first experiment, we omit the support constraint in the ambiguity set (26) and choose the risk tolerance  $\epsilon = 0.1$ . The nominal task durations are fixed to  $\mathbf{d}^0 = \mathbf{e}$ , the dispersion function coefficients to  $(s^-, s^+) = (2, 1)$ , and we select  $\boldsymbol{\sigma}$  uniformly at random from  $[\mathbf{0}, \frac{5}{100}\mathbf{e}]^N$ . We assume that feasible resource allocations need to meet task-wise and cumulative resource budgets and set  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in [\mathbf{0}, \frac{1}{2}\mathbf{e}], \mathbf{e}^\top \mathbf{x} \leq \frac{3}{8}n\}$ . Thus, at most 75% of the tasks can be assigned the maximum resource allocation  $x_i = \frac{1}{2}$ . The resulting instances of problem (27) satisfy all of our

$n$	OS	# Paths	Solution times (secs)			Bound gap		Suboptimality	
			Exact	LDR	Bonf.	LDR	Bonf.	LDR	Bonf.
30	0.25	65.7	0.55	11.02	0.16	344%	677%	8%	11%
30	0.5	155.1	6.14	4.60	0.40	637%	1,899%	11%	14%
30	0.75	411.4	34.11	2.51	1.70	531%	5,499%	13%	13%
40	0.25	125.9	2.30	48.06	0.41	551%	1,184%	7%	12%
40	0.5	390.2	71.90	21.74	2.03	972%	4,427%	9%	14%
40	0.75	2,089.5	552.37	6.55	23.47	852%	27,282%	15%	15%
50	0.25	209.5	49.46	200.61	1.08	1,141%	1,862%	6%	11%
50	0.5	904.3	355.67	159.12	6.60	1,206%	9,725%	9%	15%
50	0.75	7,520.3	5,306.18	43.86	204.40	1,122%	104,097%	14%	15%

**Table 1.** Numerical results for our exact reformulation (‘Exact’), the LDR approximation (‘LDR’) and the Bonferroni approximation (‘Bonf.’) in the project management problem (27) without support. The ‘bound gap’ quantifies the increase in the estimated worst-case makespan relative to the makespan of the exact problem, and the ‘suboptimality’ quantifies the increase in the makespan of the exact problem if we replace the optimal solution with the determined resource allocation.

regularity conditions and can hence be solved exactly. Table 1 presents average results for 100 random project networks of size  $n \in \{30, 40, 50\}$  and order strength 0.25, 0.5 and 0.75. The order strength (OS) denotes the fraction of all  $n(n-1)/2$  possible precedences between the project tasks that are enforced (directly or via transitivity) through the arcs of the project graph.

The table shows that networks of 50 tasks give rise to problems with up to 7,500 safety conditions inside the chance constraint. Nevertheless, using the convex reformulation (9), the chance constrained problem (27) can be solved exactly within 1.5 hours. While the LDR and Bonferroni bounds can be computed much faster, they lead to overly pessimistic estimates of the resulting worst-case project makespan. Indeed, the naive Bonferroni bound overestimates the value-at-risk of the project’s makespan by up to 100,000%, and the state-of-the-art LDR bound reports makespans that are up to 10 times too high. Moreover, these inflated worst-case makespan estimates result in resource allocations that also underperform when evaluated in the exact problem (27).

We now repeat the experiment with support constraints. To this end, we set  $(\underline{\xi}, \bar{\xi}) = (-\frac{1}{2}\mathbf{e}, +\frac{1}{2}\mathbf{e})$  in (26) and choose the other parameters as in the previous experiment. Note that the support constraints violate the assumption  $(\mathbf{S}')$ . Following the discussion in Remark 3, we use our results from Section 3 to formulate a nonconvex optimization problem that we solve approximately using a sequential convex optimization scheme ('SCO'). We compare this approach with our exact reformulation applied to an outer approximation of the ambiguity set that disregards the support constraint ('No support'), as well as the LDR and Bonferroni bounds. As expected, all approaches require more computation time due to the presence of the support constraints, with the SCO scheme taking up to 2.5 hours for the largest instances. We observe that the SCO scheme provides the tightest bounds and that the 'No support' bound typically outperforms both the LDR and the Bonferroni approximations, despite the fact that the latter bounds account for the support of  $\bar{\xi}$ . We remark that the degree of conservatism is much smaller than in the previous experiment, which is due to the fact that the support constraints remove many pathological distributions that place small probabilities on very large task durations.

## 5.2 Image Reconstruction

A fundamental problem in image processing is the restoration of noisy images, where the noise is caused by the image recording or transmission process. In this section, we consider a discrete version of this problem where a noisy  $m \times n$ -grayscale image is represented by a vector  $\mathbf{f} \in [0, 1]^{mn}$  of pixel light intensities. The goal is to decompose this image into a restored image  $\mathbf{x} \in [0, 1]^{mn}$  and an additive noise realization  $\boldsymbol{\xi} \in [-1, 1]^{mn}$  that explain the observed image.

A unique decomposition of a noisy image  $\mathbf{f}$  into a restored image  $\mathbf{x}$  and a noise realization  $\boldsymbol{\xi}$  requires further information. It has been observed that pristine images often contain large regions of smooth color gradients which are separated by sharp edges. Thus, those images possess a low total variation, which we define as a functional of the image's intensity gradient:

$$TV(\mathbf{x}) = \sum_{\substack{1 \leq i < m \\ 1 \leq j < n}} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2},$$

where  $x_{i,j}$  corresponds to the pixel in row  $i$  and column  $j$  of the image  $\mathbf{x}$ . Accounting for this empirical observation, Goldfarb and Yin [31] propose to minimize the total variation of the recovered

$N$	OS	Solution times (secs)			Bound gap			Suboptimality		
		SCO	LDR	Bonf.	No support	LDR	Bonf.	No support	LDR	Bonf.
30	0.25	3.60	3.00	0.36	28%	40%	40%	1%	4%	8%
30	0.5	9.72	4.42	1.14	15%	39%	42%	1%	4%	9%
30	0.75	39.41	3.27	6.66	5%	45%	48%	1%	2%	6%
40	0.25	13.63	8.36	1.20	40%	41%	40%	2%	4%	11%
40	0.5	60.73	8.35	6.18	23%	39%	41%	2%	3%	11%
40	0.75	474.21	8.82	67.23	9%	45%	48%	1%	2%	8%
50	0.25	46.50	23.45	3.17	53%	37%	37%	7%	8%	12%
50	0.5	291.64	25.50	24.59	30%	40%	40%	2%	3%	13%
50	0.75	8,795.87	16.83	539.93	13%	44%	46%	2%	3%	7%

**Table 2.** Numerical results for the SCO scheme (‘SCO’), our exact reformulation applied to an outer approximation of the ambiguity set that disregards the support constraints (‘No support’), the LDR approximation (‘LDR’) and the Bonferroni approximation (‘Bonf.’) in the project management problem (27) with support. The ‘bound gap’ quantifies the increase in the estimated worst-case makespan relative to the makespan of the SCO problem, and the ‘suboptimality’ quantifies the increase in the makespan of the SCO problem if we replace the optimal solution with the determined resource allocation.

original image  $\mathbf{x}$ , subject to an upper bound on the noise realization  $\boldsymbol{\xi}$ :

$$\begin{aligned}
& \underset{\mathbf{x}, \boldsymbol{\xi}}{\text{minimize}} && TV(\mathbf{x}) \\
& \text{subject to} && \mathbf{x} \in [0, 1]^{mn}, \quad \boldsymbol{\xi} \in \mathbb{R}^{mn} \\
& && \mathbf{x} + \boldsymbol{\xi} = \mathbf{f} \\
& && \|\boldsymbol{\xi}\|_2 \leq \sigma
\end{aligned} \tag{28}$$

Much of the literature on image reconstruction assumes, either explicitly or implicitly, that the components of  $\boldsymbol{\xi}$  are realizations of independent normal random variables with zero mean and known fixed variance. This is reflected by the last constraint in (28), which can be interpreted as a likelihood bound under normality assumption. As the pixel intensities are confined to  $[0, 1]$ , however, this is a strict approximation, the quality of which deteriorates with increasing noise levels. Indeed, the noise distribution cannot be normal and must have a nonzero mean as well as a reduced variance in the very dark and light areas of the image. Unless the pristine image is known, the distribution of the random noise vector  $\tilde{\boldsymbol{\xi}}$  is therefore ambiguous even if we knew that it is a normal distribution truncated to ensure that the observed noisy image remains within  $[0, 1]^{mn}$ .

With the methods proposed in this paper, we can faithfully model the additive noise  $\tilde{\boldsymbol{\xi}}$  as a random vector that follows an ambiguous distribution. It is then natural to impose that the difference between the observed and the reconstructed image should be explained by a ‘sufficiently likely’ noise realization  $\tilde{\boldsymbol{\xi}}$  under any of the potential noise distributions contained in the ambiguity set. This gives rise to the following optimistic chance constrained program:

$$\begin{aligned}
& \underset{\mathbf{x}}{\text{minimize}} && TV(\mathbf{x}) \\
& \text{subject to} && \mathbf{x} \in [0, 1]^{mn} \\
& && \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\mathbf{x} + \tilde{\boldsymbol{\xi}} = \mathbf{f}) \geq 1 - \epsilon,
\end{aligned} \tag{29}$$

where the ambiguity set  $\mathcal{P}$  is defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{mn}) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}[\|\tilde{\boldsymbol{\xi}}\|_2] \leq \sigma \right\}.$$

As we pointed out in Section 4, Theorem 7 carries over to weak chance constraints without any modifications. We can therefore employ Theorem 7 to reformulate problem (29) as a second-order cone program. One can show that this optimization problem constitutes a generalization of (28), where the last constraint is replaced with  $\|\boldsymbol{\xi} - \boldsymbol{\mu}\|_2 \leq \frac{\sigma}{2-2\epsilon}$ . While Goldfarb and Yin

implicitly set  $\boldsymbol{\mu} = \mathbf{0}$ , we know from the previous discussion that  $\boldsymbol{\mu}$  must have nonzero components at least in the very dark and light areas of the image. To account for this phenomenon, we set  $\mu_{i,j} = \mathbb{E}[[s_{i,j} + \tilde{\xi}_{i,j}^0]_{[0,1]} - s_{i,j}]$ , where  $\mathbf{s} \in \mathbb{R}^{mn}$  is a smoothed version of the noisy image obtained from applying a circular moving average filter to  $\mathbf{f}$ ,  $\tilde{\xi}^0$  is a zero-mean, normally distributed random variable designed to match the noise prior to truncation, and  $[\cdot]_{[0,1]}$  denotes the projection onto  $[0, 1]$ . Contrary to the truncated noise  $\tilde{\xi}$ , the distribution of  $\tilde{\xi}^0$  can often be estimated reliably [45, 42].

Table 3 and Figure 1 compare our formulation with Goldfarb and Yin’s model (28) on a variety of standard benchmark images. Note that both formulations rely on a single design parameter ( $\sigma$  in Goldfarb and Yin’s model;  $\sigma/(2 - 2\epsilon)$  in our approach). To facilitate a fair comparison, we choose the best value of this parameter in both models for each image. Table 3 presents the normalized distances  $\|\mathbf{x} - \mathbf{x}^0\|_2 / \|\mathbf{f} - \mathbf{x}^0\|_2$  of the reconstructed images  $\mathbf{x}$  from the original images  $\mathbf{x}^0$  for our chance-constrained program (29) with true  $\boldsymbol{\mu}$  (‘CC’) and with estimated  $\boldsymbol{\mu}$  as outlined above (‘CC-E’), as well as Goldfarb and Yin’s model (‘GY’). For comparison purposes, we also include the normalized distances of the smoothed images resulting from the moving average filter that we employ to estimate  $\boldsymbol{\mu}$  in CC-E (‘MAF’). The table also presents the relative improvements of CC, CC-E and MAF over GY in terms of these normalized distances. Figure 1 presents the average normalized distances of the four approaches over all benchmark images as a function of the standard deviation of  $\tilde{\xi}^0$ , and a specific solution is shown in Figure 2. The results indicate that a faithful modeling of the ambiguity about the truncated noise  $\tilde{\xi}$  can lead to consistent improvements over Goldfarb and Yin’s model, and that these improvements increase with the noise level.

**Acknowledgements** This research was supported by the Swiss National Science Foundation grant BSCGI0-157733 and the EPSRC grants EP/M028240/1 and EP/M027856/1.

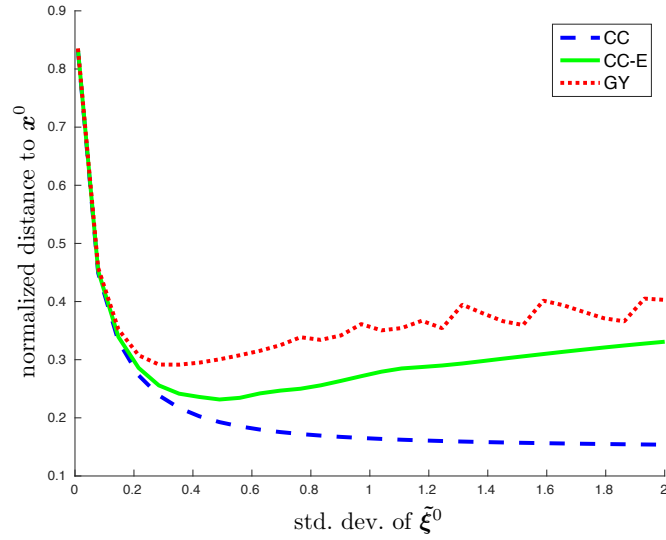
## References

- [1] H. An and J. W. Eheart. A screening technique for joint chance-constrained programming for air-quality management. *Operations Research*, 55(4):792–798, 2007.
- [2] A. Beck and A. Ben-Tal. Duality in robust optimization: Primal worst equals dual best. *Operations Research Letters*, 37(1):1–6, 2009.
- [3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.
- [4] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming A*, 88(3):411–424, 2000.
- [5] A. Ben-Tal and A. Nemirovski. On safe tractable approximations of chance-constrained linear matrix inequalities. *Mathematics of Operations Research*, 34(1):1–25, 2009.

Image	Norm. Dist.				Rel. Impr.				Image	Norm. Dist.				Rel. Impr.			
	CC	CC-E	MAF	GY	CC	CC-E	MAF	GY		Image	CC	CC-E	MAF	GY	CC	CC-E	MAF
K-1	42.5%	42.9%	52.1%	43.1%	1.292%	0.461%	-20.839%		K-16	26.9%	27.1%	47.2%	27.8%	3.307%	2.263%	-69.805%	
K-2	23.2%	23.7%	46.6%	23.8%	2.459%	0.055%	-96.067%		K-17	31.0%	31.8%	48.8%	33.9%	8.630%	6.118%	-43.958%	
K-3	23.0%	23.4%	46.5%	24.0%	4.165%	2.519%	-93.828%		K-18	38.4%	39.3%	51.2%	39.8%	3.666%	1.359%	-28.554%	
K-4	26.7%	27.0%	46.8%	27.5%	2.890%	1.852%	-70.232%		K-19	35.7%	36.1%	50.1%	36.3%	1.708%	0.433%	-37.928%	
K-5	46.6%	47.8%	55.9%	48.6%	4.078%	1.577%	-15.071%		K-20	29.5%	32.5%	57.0%	43.8%	32.653%	25.962%	-30.016%	
K-6	36.9%	37.5%	51.1%	39.0%	5.262%	3.726%	-31.202%		K-21	35.0%	35.5%	50.0%	35.9%	2.518%	1.182%	-39.176%	
K-7	33.0%	33.4%	47.5%	33.7%	2.170%	1.052%	-40.719%		K-22	30.1%	30.4%	48.1%	30.8%	2.361%	1.501%	-55.875%	
K-8	49.6%	51.3%	57.9%	52.1%	4.652%	1.456%	-11.236%		K-23	24.5%	24.8%	46.8%	25.5%	4.259%	2.808%	-83.374%	
K-9	27.9%	28.3%	47.3%	28.9%	3.243%	1.949%	-63.612%		K-24	38.7%	39.7%	53.2%	40.8%	5.055%	2.677%	-30.454%	
K-10	28.1%	28.6%	47.4%	29.3%	3.987%	2.576%	-61.743%		Dantzig	27.2%	28.4%	47.7%	30.5%	10.781%	6.984%	-56.545%	
K-11	32.8%	33.3%	49.4%	34.4%	4.764%	3.164%	-43.710%		4.2.03	46.2%	46.4%	53.9%	46.4%	0.463%	-0.062%	-16.068%	
K-12	24.0%	24.7%	47.3%	25.9%	7.318%	4.566%	-82.644%		4.2.04	27.7%	27.6%	46.6%	27.8%	0.672%	0.807%	-67.201%	
K-13	47.4%	48.4%	56.9%	48.8%	2.759%	0.902%	-16.628%		4.2.06	35.3%	35.8%	48.2%	36.3%	2.879%	1.282%	-32.841%	
K-14	37.3%	37.9%	50.1%	38.4%	2.962%	1.380%	-30.407%		5.2.08	32.0%	32.9%	48.8%	33.4%	4.201%	1.496%	-46.220%	
K-15	27.3%	28.7%	50.0%	33.0%	17.369%	13.145%	-51.623%		5.2.10	41.8%	42.4%	51.1%	43.1%	2.984%	1.482%	-18.750%	

**Table 3.** Results for benchmark instances images scaled to 160,000 pixels. Instances starting with ‘K’ are from the Kodak Photo CD Sampler, 1995, PCD0992; ‘Dantzig’ is taken from <http://news.stanford.edu/news/2006/june7/memldant-060706.html>, and the remaining images are taken from the USC-SIPI Image Database (<http://sipi.usc.edu/database/database.php>).





**Figure 1.** Normalized distances of the reconstructed images to their true counterparts when the reconstructed images are obtained from Goldfarb and Yin’s model (28) (dotted red line) and the chance constrained program (29) with true and estimated mean values (dashed blue and solid green line, respectively).



**Figure 2.** Original, noisy and reconstructed versions of the image K-15 from Table 3. The noisy image is obtained by adding Gaussian noise with mean zero and standard deviation 0.15 to the original image and projecting the sum to  $[0, 1]^{mn}$ . The reconstructed image is the solution of the ambiguous chance constrained program CC-E.

- [6] D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. Available on Optimization Online, 2013.
- [7] D. Bertsimas, K. Natarajan, and C. P. Teo. Applications of semidefinite optimization in stochastic project scheduling. *High Performance Computation for Engineered Systems (HPCES)*, 2002.
- [8] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal of Optimization*, 15(3):780–804, 2004.
- [9] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.
- [10] R. F. Bordley and S. M. Pollock. A decision-analytic approach to reliability-based design optimization. *Operations Research*, 57(5):1262–1270, 2009.
- [11] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [12] 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.
- [13] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Thomson Learning, 2nd edition, 2002.
- [14] A. Charnes and W. W. Cooper. Chance-constrained programming. *Management Science*, 6(1):73–79, 1959.
- [15] A. Charnes, W. W. Cooper, and G. H. Symonds. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science*, 4(3):235–263, 1958.
- [16] 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.
- [17] X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.
- [18] X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.
- [19] S.-S. Cheung, A. M.-C. So, and K. Wang. Linear matrix inequalities with stochastically dependent perturbations and applications to chance-constrained semidefinite optimization. *SIAM Journal on Optimization*, 22(4):1394–1430, 2012.
- [20] J.-F. Cordeau, G. Laporte, M. W. P. Savelsbergh, and D. Vigo. Vehicle routing. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, chapter 6, pages 367–428. North-Holland, Amsterdam, The Netherlands, 2007.
- [21] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.
- [22] E. L. Demeulemeester and W. S. Herroelen. *Project Scheduling – A Research Handbook*. Kluwer Academic Publishers, 2002.
- [23] V. DeMiguel and F. J. Nogales. Portfolio selection with robust estimation. *Operations Research*, 57(3):560–577, 2009.
- [24] C. Dert and B. Oldenkamp. Optimal guaranteed return portfolios and the casino effect. *Operations Research*, 48(5):768–775, 2000.
- [25] M. E. Dyer and A. M. Frieze. On the complexity of computing the volume of a polyhedron. *SIAM Journal on Computing*, 17(5):967–974, 1988.
- [26] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming B*, 107(1–2):37–61, 2006.
- [27] A. Flajolet, S. Blandin, and P. Jaillet. Robust adaptive routing under uncertainty. Available on arXiv, 2014.
- [28] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
- [29] A. Georghiou, W. Wiesemann, and D. Kuhn. The decision rule approach to optimisation under uncertainty: Methodology and applications in operations management. Available on Optimization Online, 2011.
- [30] 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.

- [31] D. Goldfarb and W. Yin. Second-order cone programming methods for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 27(2):622–645, 2005.
- [32] B. L. Gorissen, H. Blanc, D. den Hertog, and A. Ben-Tal. Technical note—Deriving robust and globalized robust solutions of uncertain linear programs with general convex uncertainty sets. *Operations Research*, 62(3):672–679, 2014.
- [33] I. Gurvich, J. Luedtke, and T. Tezcan. Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. *Management Science*, 56(7):1093–1115, 2010.
- [34] S. Han, M. Tao, U. Topcu, H. Owhadi, and R. M. Murray. Convex optimal uncertainty quantification. *SIAM Journal on Optimization*, 25(3):1368–1387, 2015.
- [35] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62, 2015.
- [36] Z. Hu and L. J. Hong. Kullback-Leibler divergence constrained distributionally robust optimization. Available on Optimization Online, 2012.
- [37] Z. Hu, L. J. Hong, and A. M.-C. So. Ambiguous probabilistic programs. Available on Optimization Online, 2013.
- [38] R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. Available on Optimization Online, 2012.
- [39] H. Konno and H. Yamazaki. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management Science*, 37(5):519–531, 1991.
- [40] J. Korsi, 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.
- [41] R. Levi, G. Perakis, and J. Uichanco. Regret optimization for stochastic inventory models with spread information. Technical Report, Massachusetts Institute of Technology, 2011.
- [42] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang. Noise estimation from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 901–908. IEEE Computer Society, 2006.
- [43] A. Nemirovski. On safe tractable approximations of chance constraints. *European Journal of Operational Research*, 219(3):707–718, 2012.
- [44] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.
- [45] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.
- [46] H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. Optimal uncertainty quantification. *SIAM Review*, 55(2):271–345, 2013.
- [47] 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.
- [48] K. Postek, A. Ben-Tal, D. den Hertog, and B. Melenberg. Exact robust counterparts of ambiguous stochastic constraints under mean and dispersion information. Available on Optimization Online, 2015.
- [49] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.
- [50] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1997.
- [51] R. T. Rockafellar and R. J-B Wets. *Variational Analysis*. Springer Science & Business Media, 2009.
- [52] A. Shapiro. On duality theory of conic linear problems. In *Semi-Infinite Programming*, chapter 7, pages 135–165. Kluwer Academic Publishers, 2001.
- [53] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modelling and Theory*. SIAM, 2009.
- [54] S. Shen, J. C. Smith, and S. Ahmed. Expectation and chance-constrained models and algorithms for insuring critical paths. *Management Science*, 56(10):1794–1814, 2010.
- [55] J. Stoer and C. Witzgall. *Convexity and Optimization in Finite Dimensions I*, volume 163. Springer Science & Business Media, 2012.

- [56] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64, 2007.
- [57] J. Wang. The  $\beta$ -reliable median on a network with discrete probabilistic demand weights. *Operations Research*, 55(5):966–975, 2007.
- [58] W. Wiesemann, D. Kuhn, and B. Rustem. Multi-resource allocation in stochastic project scheduling. *Annals of Operations Research*, 193(1):193–220, 2012.
- [59] W. Wiesemann, D. Kuhn, and B. Rustem. Robust resource allocations in temporal networks. *Mathematical Programming A*, 135(1):437–471, 2012.
- [60] W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2013.
- [61] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.
- [62] H. Xu, C. Caramanis, and S. Mannor. Optimization under probabilistic envelope constraints. *Operations Research*, 60(3):682–699, 2012.
- [63] İ. Yanıkoğlu and D. den Hertog. Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing*, 25(4):666–681, 2013.
- [64] 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.

## Appendix: Proofs

In the remainder we let  $\mathbb{B}_r(\mathbf{c})$  be the closed Euclidean ball of radius  $r \geq 0$  centered at  $\mathbf{c}$ .

**Lemma 7** (Strong Duality). *Under the assumptions (A) and (D), strong duality holds between the uncertainty quantification problem*

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} < \mathbf{u}(\mathbf{x})]$$

and its dual semi-infinite program

$$\begin{aligned} & \underset{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}}{\text{maximize}} && \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \\ & \text{subject to} && \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}^k, \quad \boldsymbol{\gamma} \in \mathcal{D}^* \\ & && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \mathbf{d}(\boldsymbol{\xi})^\top \boldsymbol{\gamma} \leq \mathbb{I}_{[\mathbf{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})]} \quad \forall \boldsymbol{\xi} \in \Xi. \end{aligned}$$

*Proof.* We need to show that the point  $(1, \boldsymbol{\mu}, \boldsymbol{\sigma})$  resides in the interior of the convex cone

$$\mathcal{V} = \left\{ (a, \mathbf{b}, \mathbf{c}) \in \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^d : \exists \mu \in \mathcal{M}_+(\Xi) \text{ such that } \begin{aligned} & \int \mu(d\boldsymbol{\xi}) = a, \\ & \int \boldsymbol{\xi} \mu(d\boldsymbol{\xi}) = \mathbf{b}, \\ & \int \mathbf{d}(\boldsymbol{\xi}) \mu(d\boldsymbol{\xi}) \preceq_{\mathcal{D}} \mathbf{c} \end{aligned} \right\},$$

see [52, Proposition 3.4]. To this end, choose any point  $(s, \mathbf{m}, \mathbf{s}) \in \mathbb{B}_\kappa(1) \times \mathbb{B}_\kappa(\boldsymbol{\mu}) \times \mathbb{B}_\kappa(\boldsymbol{\sigma})$ , where  $\kappa > 0$  is chosen sufficiently small, and consider the scaled Dirac measure  $s \cdot \delta_{\mathbf{m}/s}$  that places mass  $s$  at  $\mathbf{m}/s$ . By construction, this measure satisfies  $\int s \cdot \delta_{\mathbf{m}/s}(d\boldsymbol{\xi}) = s$  and  $\int \boldsymbol{\xi} s \cdot \delta_{\mathbf{m}/s}(d\boldsymbol{\xi}) = \mathbf{m}$ . Moreover, for sufficiently small  $\kappa$  the measure is supported on  $\Xi$  (since  $\boldsymbol{\mu} \in \text{int } \Xi$ ) and satisfies  $\int \mathbf{d}(\boldsymbol{\xi}) s \cdot \delta_{\mathbf{m}/s}(d\boldsymbol{\xi}) \preceq_{\mathcal{D}} \mathbf{s}$  (since  $\mathbf{d}(\boldsymbol{\mu}) \prec_{\mathcal{D}} \boldsymbol{\sigma}$  and the dispersion function  $\mathbf{d}$  is continuous).  $\square$

**Proof of Theorem 1.** The worst-case probability in (PCC) is given by the optimal value of the moment problem

$$\begin{aligned} & \underset{\mu}{\text{minimize}} && \int_{\Xi} \mathbb{I}_{[\mathbf{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})]} \mu(d\boldsymbol{\xi}) \\ & \text{subject to} && \mu \in \mathcal{M}_+(\Xi) \\ & && \int_{\Xi} \mu(d\boldsymbol{\xi}) = 1 \\ & && \int_{\Xi} \boldsymbol{\xi} \mu(d\boldsymbol{\xi}) = \boldsymbol{\mu} \\ & && \int_{\Xi} \mathbf{d}(\boldsymbol{\xi}) \mu(d\boldsymbol{\xi}) \preceq_{\mathcal{D}} \boldsymbol{\sigma}, \end{aligned}$$

where  $\mathcal{M}_+(\Xi)$  denotes the convex cone of nonnegative Borel measures on  $\Xi$ . Note that the first moment constraint forces  $\mu$  to be a probability measure. The above moment problem constitutes in fact a conic program involving infinitely many decisions and finitely many constraints. Its dual is therefore a semi-infinite conic program with finitely many decision variables (the Lagrange multipliers  $\alpha$ ,  $\beta$  and  $\gamma$  corresponding to the normalization, location and dispersion conditions, respectively) and infinitely many constraints parameterized by the uncertainty realizations  $\xi \in \Xi$ :

$$\begin{aligned}
& \underset{\alpha, \beta, \gamma}{\text{maximize}} && \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \\
& \text{subject to} && \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}^k, \quad \boldsymbol{\gamma} \in \mathcal{D}^* \\
& && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \mathbf{d}(\boldsymbol{\xi})^\top \boldsymbol{\gamma} \leq \mathbb{I}_{[\mathcal{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})]} \quad \forall \boldsymbol{\xi} \in \Xi
\end{aligned} \tag{30}$$

Strong duality holds due to the assumptions **(A)** and **(D)**; see Lemma 7 in the appendix for more details. Thus, the optimal value of (30) coincides with that of the primal moment problem.

By expanding the indicator function, the discontinuous semi-infinite constraint in (30) can be decomposed into several continuous semi-infinite constraints:

$$\begin{aligned}
& \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \mathbf{d}(\boldsymbol{\xi})^\top \boldsymbol{\gamma} \leq 1 && \forall \boldsymbol{\xi} \in \Xi \\
& \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} - \mathbf{d}(\boldsymbol{\xi})^\top \boldsymbol{\gamma} \leq 0 && \forall i \in \mathcal{I}, \forall \boldsymbol{\xi} \in \Xi : \mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\xi} \geq u_i(\mathbf{x})
\end{aligned} \tag{31}$$

This implies that the worst-case probability in (PCC) reduces to

$$\begin{aligned}
& \underset{\alpha, \beta, \gamma}{\text{maximize}} && \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\gamma}^\top \boldsymbol{\sigma} \\
& \text{subject to} && \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}^k, \quad \boldsymbol{\gamma} \in \mathcal{D}^* \\
& && \alpha + \boldsymbol{\xi}_0^\top \boldsymbol{\beta} - \mathbf{d}(\boldsymbol{\xi}_0)^\top \boldsymbol{\gamma} \leq 1 && \forall \boldsymbol{\xi}_0 \in \Xi_0(\mathbf{x}) \\
& && \alpha + \boldsymbol{\xi}_i^\top \boldsymbol{\beta} - \mathbf{d}(\boldsymbol{\xi}_i)^\top \boldsymbol{\gamma} \leq 0 && \forall \boldsymbol{\xi}_i \in \Xi_i(\mathbf{x}), \quad i \in \mathcal{I}(\mathbf{x}),
\end{aligned} \tag{32}$$

where  $\Xi_0(\mathbf{x}) = \Xi$  and  $\Xi_i(\mathbf{x}) = \{\boldsymbol{\xi} \in \Xi : \mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\xi} \geq u_i(\mathbf{x})\}$  for all  $i \in \mathcal{I}(\mathbf{x})$ . Note that (32) can be interpreted as the robust counterpart of an uncertain convex program with constraint-wise uncertainty. Thus, problem (32) is solved by a decision maker choosing  $\alpha$ ,  $\beta$  and  $\gamma$  under the worst possible data  $\boldsymbol{\xi}_i$ ,  $i \in \mathcal{I}(\mathbf{x})$ . It has been shown in [2, Theorem 4.1] that this problem is equivalent to the dual of the uncertain convex program where the decision maker operates under the best

possible data. Specifically, (32) is equivalent to<sup>2</sup>

$$\begin{aligned}
& \underset{\lambda_i, \boldsymbol{\xi}_i}{\text{minimize}} && \lambda_0 \\
& \text{subject to} && \lambda_i \in \mathbb{R}_+, \quad \boldsymbol{\xi}_i \in \Xi_i(\mathbf{x}), \quad i \in \mathcal{I}_0(\mathbf{x}) \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i = 1 \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i \boldsymbol{\xi}_i = \boldsymbol{\mu} \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i \mathbf{d}(\boldsymbol{\xi}_i) \preceq_{\mathcal{D}} \boldsymbol{\sigma},
\end{aligned} \tag{33}$$

which can easily be reformulated as (4). □

**Lemma 8.** *Problem (6) admits a Slater-type point  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$ , that is, a feasible point that satisfies the nonnegativity constraints strictly for all  $i \in \mathcal{I}_0(\mathbf{x})$ .*

**Proof of Lemma 8.** For  $i \in \mathcal{I}(\mathbf{x})$  and  $\kappa \in (0, |\mathcal{I}|^{-1})$ , we set  $\lambda'_i = \kappa$  and  $\boldsymbol{\chi}'_i = \lambda'_i \boldsymbol{\xi}_i$  for any  $\boldsymbol{\xi}_i \in \Xi$  that satisfies  $\mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\xi}_i \geq u_i(\mathbf{x})$ . We also set  $\lambda'_i = 0$  and  $\boldsymbol{\chi}'_i = \mathbf{0}$  for  $i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})$ , as well as  $\lambda'_0 = 1 - |\mathcal{I}(\mathbf{x})|\kappa$  and  $\boldsymbol{\chi}'_0 = \boldsymbol{\mu} - \sum_{i \in \mathcal{I}(\mathbf{x})} \boldsymbol{\chi}'_i$ . For  $\kappa$  sufficiently small, the assumptions **(A)** and **(D)** imply that  $\boldsymbol{\chi}'_0/\lambda'_0 \in \Xi$  (since  $\boldsymbol{\mu} \in \text{int } \Xi$ ) and  $\sum_{i \in \mathcal{I}_0} \lambda'_i \mathbf{d}(\boldsymbol{\chi}'_i/\lambda'_i) \preceq_{\mathcal{D}} \boldsymbol{\sigma}$  (since  $\mathbf{d}(\boldsymbol{\mu}) \prec_{\mathcal{D}} \boldsymbol{\sigma}$  and  $\mathbf{d}$  is continuous). One readily verifies that  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$  also satisfies the other constraints. □

**Proof of Proposition 1.** For any feasible solution  $(\{\lambda'_i\}_i, \{\boldsymbol{\xi}'_i\}_i)$  to problem (4), the solution  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  with  $(\lambda_i, \boldsymbol{\chi}_i) = (\lambda'_i, \lambda'_i \boldsymbol{\xi}'_i)$ ,  $i \in \mathcal{I}_0(\mathbf{x})$ , and  $(\lambda_i, \boldsymbol{\chi}_i) = (0, \mathbf{0})$ ,  $i \in \mathcal{I}_0 \setminus \mathcal{I}_0(\mathbf{x})$ , is feasible in (6) and attains the same objective value; see also Definition 1. We thus conclude that the optimal value of (6) provides a lower bound on the optimal value of (4).

Next, we show that the optimal value of (6) also provides an upper bound on the optimal value of (4). To this end, we note that (6) is solvable by Proposition 2 below. We now proceed in two steps. We first show that any feasible solution to problem (6) can be transformed into a feasible

---

<sup>2</sup>To be precise, in [2] it is shown that the equivalence of (32) and (33) holds if all uncertainty sets  $\Xi_i(\mathbf{x})$ ,  $i \in \mathcal{I}_0(\mathbf{x})$ , are compact. As (33) constitutes a reformulation of the worst-case probability problem in (PCC), whose optimal value must lie in the interval  $[0, 1]$ , one can show that the equivalence extends to unbounded uncertainty sets.

solution to the related problem

$$\begin{aligned}
& \underset{\lambda_i, \boldsymbol{\chi}_i}{\text{minimize}} && \lambda_0 \\
& \text{subject to} && \lambda_i \in \mathbb{R}_+, \boldsymbol{\chi}_i \in \mathbb{R}^k, i \in \mathcal{I}_0(\mathbf{x}) \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i = 1 \\
& && \frac{\boldsymbol{\chi}_i}{\lambda_i} \in \Xi \quad \forall i \in \mathcal{I}_0(\mathbf{x}) \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \boldsymbol{\chi}_i = \boldsymbol{\mu} \\
& && \sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i \mathbf{d}\left(\frac{\boldsymbol{\chi}_i}{\lambda_i}\right) \preceq_{\mathcal{D}} \boldsymbol{\sigma} \\
& && \mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\chi}_i \geq \lambda_i u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}(\mathbf{x}),
\end{aligned} \tag{34}$$

which employs the same decision-dependent index sets  $\mathcal{I}(\mathbf{x})$  and  $\mathcal{I}_0(\mathbf{x})$  as problem (4). Afterwards, we prove that any solution to (34) can in turn be transformed into a feasible solution to problem (4).

In view of the first step, assume that  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$  is feasible in (6). We show that  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  defined through

$$\lambda_0 = \lambda'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \lambda'_i, \quad \lambda_i = \lambda'_i \quad \forall i \in \mathcal{I}(\mathbf{x}), \quad \boldsymbol{\chi}_0 = \boldsymbol{\chi}'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \boldsymbol{\chi}'_i, \quad \boldsymbol{\chi}_i = \boldsymbol{\chi}'_i \quad \forall i \in \mathcal{I}(\mathbf{x})$$

is feasible in (34) and attains the same objective value. Clearly,  $\lambda_i \geq 0$ ,  $i \in \mathcal{I}_0(\mathbf{x})$ , and we have

$$\sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i = \sum_{i \in \mathcal{I}_0} \lambda'_i = 1 \quad \text{and} \quad \sum_{i \in \mathcal{I}_0(\mathbf{x})} \boldsymbol{\chi}_i = \sum_{i \in \mathcal{I}_0} \boldsymbol{\chi}'_i = \boldsymbol{\mu}.$$

Similarly, we find

$$\begin{aligned}
\sum_{i \in \mathcal{I}_0(\mathbf{x})} \lambda_i \mathbf{d}\left(\frac{\boldsymbol{\chi}_i}{\lambda_i}\right) &= \left( \lambda'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \lambda'_i \right) \mathbf{d}\left(\frac{\boldsymbol{\chi}'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \boldsymbol{\chi}'_i}{\lambda'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \lambda'_i}\right) + \sum_{i \in \mathcal{I}(\mathbf{x})} \lambda'_i \mathbf{d}\left(\frac{\boldsymbol{\chi}'_i}{\lambda'_i}\right) \\
&\preceq_{\mathcal{D}} \sum_{i \in \mathcal{I}_0} \lambda'_i \mathbf{d}\left(\frac{\boldsymbol{\chi}'_i}{\lambda'_i}\right) \preceq_{\mathcal{D}} \boldsymbol{\sigma},
\end{aligned}$$

where the first inequality follows from the  $\mathcal{D}$ -convexity of the dispersion function  $\mathbf{d}(\boldsymbol{\xi})$ . Next, we have  $\boldsymbol{\chi}_i/\lambda_i = \boldsymbol{\chi}'_i/\lambda'_i \in \Xi$  for all  $i \in \mathcal{I}(\mathbf{x})$ , and the convexity of  $\Xi$  further implies that

$$\frac{\boldsymbol{\chi}_0}{\lambda_0} = \frac{\boldsymbol{\chi}'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \boldsymbol{\chi}'_i}{\lambda'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \lambda'_i} = \frac{\lambda'_0}{\lambda'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \lambda'_i} \cdot \frac{\boldsymbol{\chi}'_0}{\lambda'_0} + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \frac{\lambda'_i}{\lambda'_0 + \sum_{i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})} \lambda'_i} \cdot \frac{\boldsymbol{\chi}'_i}{\lambda'_i} \in \Xi.$$



It is also easy to verify that

$$\mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\chi}_i = \mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\chi}'_i \geq \lambda'_i u_i(\mathbf{x}) = \lambda_i u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}(\mathbf{x}).$$

Thus,  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  is feasible in (34). Moreover, the objective value of  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  in (34) equals the objective value of  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$  in (6). Indeed, we have  $\lambda'_i = 0$  for all  $i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x})$  for otherwise  $\boldsymbol{\xi}'_i = \boldsymbol{\chi}'_i / \lambda'_i \in \Xi$  satisfies  $\mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\xi}'_i \geq u_i(\mathbf{x})$ , which is in conflict with the definition of  $\mathcal{I}(\mathbf{x})$ .

As for the second step, assume that  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  is feasible in (34). A straightforward adaption of Lemma 8 shows that (34) admits a Slater-type point  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$  with  $\lambda'_i > 0$  for all  $i \in \mathcal{I}_0(\mathbf{x})$ . Since (34) is convex, we can construct convex combinations of  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$  and  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  to generate a sequence of feasible solutions  $(\{\lambda_i^k\}_i, \{\boldsymbol{\chi}_i^k\}_i)_k$  that converge to  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  and that satisfy  $\lambda_i^k > 0$ ,  $i \in \mathcal{I}_0(\mathbf{x})$ . The corresponding solutions  $(\{\lambda_i^k\}_i, \{\boldsymbol{\xi}_i^k\}_i)$  with  $\boldsymbol{\xi}_i^k = \boldsymbol{\chi}_i^k / \lambda_i^k$  are feasible in (4) and attain the same objective values, which concludes the proof.  $\square$

**Proof of Proposition 2.** Problem (6) can be reformulated as

$$\begin{aligned} & \underset{\lambda_i, \boldsymbol{\chi}_i, \boldsymbol{\nu}_i}{\text{minimize}} && \lambda_0 \\ & \text{subject to} && \lambda_i \in \mathbb{R}_+, \quad \boldsymbol{\chi}_i \in \mathbb{R}^k, \quad \boldsymbol{\nu}_i \in \mathbb{R}^d, \quad i \in \mathcal{I}_0 \\ & && (\lambda_i, \boldsymbol{\chi}_i, \boldsymbol{\nu}_i) \in \mathcal{K} \quad \forall i \in \mathcal{I}_0 \\ & && \sum_{i \in \mathcal{I}_0} (\lambda_i, \boldsymbol{\chi}_i, \boldsymbol{\nu}_i) = (1, \boldsymbol{\mu}, \boldsymbol{\sigma}) \\ & && \mathbf{t}_i(\mathbf{x})^\top \boldsymbol{\chi}_i \geq \lambda_i u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}, \end{aligned}$$

where  $\boldsymbol{\nu}_i$ ,  $i \in \mathcal{I}_0$ , are epigraphical auxiliary variables and where the cone

$$\mathcal{K} = \left\{ (\lambda, \boldsymbol{\chi}, \boldsymbol{\nu}) \in \mathbb{R}_+ \times \mathbb{R}^k \times \mathbb{R}^d : \frac{\boldsymbol{\chi}}{\lambda} \in \Xi, \lambda \mathbf{d} \left( \frac{\boldsymbol{\chi}}{\lambda} \right) \preceq_{\mathcal{D}} \boldsymbol{\nu} \right\}$$

is convex due to assumption **(D)** and due to the convexity of  $\Xi$ . As  $\mathcal{K}$  is also nonempty due to assumption **(A)**, it admits a decomposition of the form  $\mathcal{K} = \mathcal{L} + \mathcal{K}^\perp$ , where  $\mathcal{L} = \mathcal{K} \cap -\mathcal{K}$  is the lineality space of  $\mathcal{K}$  and  $\mathcal{K}^\perp = \mathcal{K} \cap \mathcal{L}^\perp$  is a pointed convex cone in the orthogonal complement of  $\mathcal{L}$  [55, Theorem 2.10.5]. Note that any  $(\lambda, \boldsymbol{\chi}, \boldsymbol{\nu}) \in \mathcal{K}$  can thus be written as  $(\lambda, \boldsymbol{\chi}, \boldsymbol{\nu}) = (\lambda^\parallel, \boldsymbol{\chi}^\parallel, \boldsymbol{\nu}^\parallel) + (\lambda^\perp, \boldsymbol{\chi}^\perp, \boldsymbol{\nu}^\perp)$ , where  $(\lambda^\parallel, \boldsymbol{\chi}^\parallel, \boldsymbol{\nu}^\parallel) \in \mathcal{L}$  and  $(\lambda^\perp, \boldsymbol{\chi}^\perp, \boldsymbol{\nu}^\perp) \in \mathcal{K}^\perp$ . Moreover, as  $\lambda$  must be nonnegative

by the definition of  $\mathcal{K}$ , it is clear that  $\lambda^\parallel = 0$  and  $\lambda^\perp = \lambda$ . We can thus rewrite (6) equivalently as

$$\begin{aligned} & \text{minimize} && \lambda_0^\perp && (35a) \\ & \lambda_i^\perp, \boldsymbol{\chi}_i^\parallel, \boldsymbol{\chi}_i^\perp, \boldsymbol{\nu}_i^\parallel, \boldsymbol{\nu}_i^\perp \end{aligned}$$

$$\text{subject to} \quad \lambda_i^\perp \in \mathbb{R}_+, \boldsymbol{\chi}_i^\parallel, \boldsymbol{\chi}_i^\perp \in \mathbb{R}^k, \boldsymbol{\nu}_i^\parallel, \boldsymbol{\nu}_i^\perp \in \mathbb{R}^d, i \in \mathcal{I}_0 \quad (35b)$$

$$(0, \boldsymbol{\chi}_i^\parallel, \boldsymbol{\nu}_i^\parallel) \in \mathcal{L} \quad \forall i \in \mathcal{I}_0 \quad (35c)$$

$$(\lambda_i^\perp, \boldsymbol{\chi}_i^\perp, \boldsymbol{\nu}_i^\perp) \in \mathcal{K}^\perp \quad \forall i \in \mathcal{I}_0 \quad (35d)$$

$$\sum_{i \in \mathcal{I}_0} (\lambda_i^\parallel, \boldsymbol{\chi}_i^\parallel, \boldsymbol{\nu}_i^\parallel) = (0, \boldsymbol{\mu}^\parallel, \boldsymbol{\sigma}^\parallel) \quad (35e)$$

$$\sum_{i \in \mathcal{I}_0} (\lambda_i^\perp, \boldsymbol{\chi}_i^\perp, \boldsymbol{\nu}_i^\perp) = (1, \boldsymbol{\mu}^\perp, \boldsymbol{\sigma}^\perp) \quad (35f)$$

$$\boldsymbol{t}_i(\boldsymbol{x})^\top (\boldsymbol{\chi}_i^\parallel + \boldsymbol{\chi}_i^\perp) \geq \lambda_i^\perp u_i(\boldsymbol{x}) \quad \forall i \in \mathcal{I}, \quad (35g)$$

where  $(1, \boldsymbol{\mu}, \boldsymbol{\sigma})$  is decomposed into  $(0, \boldsymbol{\mu}^\parallel, \boldsymbol{\sigma}^\parallel) \in \mathcal{L}$  and  $(1, \boldsymbol{\mu}^\perp, \boldsymbol{\sigma}^\perp) \in \mathcal{K}^\perp$ . For any  $i \in \mathcal{I}_0$ , the triplet  $(\lambda_i^\perp, \boldsymbol{\chi}_i^\perp, \boldsymbol{\nu}_i^\perp)$  belongs to  $\mathcal{C} = \mathcal{K}^\perp \cap (\{(1, \boldsymbol{\mu}^\perp, \boldsymbol{\sigma}^\perp)\} - \mathcal{K}^\perp)$  as

$$(\lambda_i^\perp, \boldsymbol{\chi}_i^\perp, \boldsymbol{\nu}_i^\perp) = (1, \boldsymbol{\mu}^\perp, \boldsymbol{\sigma}^\perp) - \sum_{j \in \mathcal{I}_0, j \neq i} (\lambda_j^\perp, \boldsymbol{\chi}_j^\perp, \boldsymbol{\nu}_j^\perp).$$

Note that  $\mathcal{C}$  is both convex and compact. Indeed,  $\mathcal{C}$  inherits convexity and closedness from  $\mathcal{K}^\perp$ . Moreover,  $\mathcal{C}$  is bounded for otherwise it would have a nonzero recession direction, which would also be a recession direction for both  $\mathcal{K}^\perp$  and  $-\mathcal{K}^\perp$  [50, Corollary 8.3.2]. This, however, would contradict the pointedness of  $\mathcal{K}^\perp$ .

Consider now a sequence of feasible decisions

$$(\{\lambda_i^\perp(t)\}_i, \{\boldsymbol{\chi}_i^\parallel(t)\}_i, \{\boldsymbol{\chi}_i^\perp(t)\}_i, \{\boldsymbol{\nu}_i^\parallel(t)\}_i, \{\boldsymbol{\nu}_i^\perp(t)\}_i), \quad t \in \mathbb{N},$$

that attain the infimum in (35) as  $t \in \mathbb{N}$  tends to infinity. By passing to a subsequence if necessary, we may assume without loss of generality that  $(\lambda_i^\perp(t), \boldsymbol{\chi}_i^\perp(t), \boldsymbol{\nu}_i^\perp(t))$  converges to  $(\hat{\lambda}_i^\perp, \hat{\boldsymbol{\chi}}_i^\perp, \hat{\boldsymbol{\nu}}_i^\perp) \in \mathcal{C}$  for every  $i \in \mathcal{I}_0$  because  $\mathcal{C}$  is compact. By construction,  $(\{\hat{\lambda}_i^\perp\}_i, \{\hat{\boldsymbol{\chi}}_i^\perp\}_i, \{\hat{\boldsymbol{\nu}}_i^\perp\}_i)$  satisfies the constraints (35d) and (35f), while  $\hat{\lambda}_0^\perp$  is equal to the infimum of (35). To prove the solvability of (35), thus, it remains to be shown that there exist  $(\{\hat{\boldsymbol{\chi}}_i^\parallel\}_i, \{\hat{\boldsymbol{\nu}}_i^\parallel\}_i)$  satisfying (35c), (35e) and (35g) for  $(\{\lambda_i^\perp\}_i, \{\boldsymbol{\chi}_i^\perp\}_i, \{\boldsymbol{\nu}_i^\perp\}_i) = (\{\hat{\lambda}_i^\perp\}_i, \{\hat{\boldsymbol{\chi}}_i^\perp\}_i, \{\hat{\boldsymbol{\nu}}_i^\perp\}_i)$ . In other words, we need to show that the projection of the polytope defined by (35c), (35e) and (35g) on the variables  $(\{\lambda_i^\perp\}_i, \{\boldsymbol{\chi}_i^\perp\}_i, \{\boldsymbol{\nu}_i^\perp\}_i)$  contains the point  $(\{\hat{\lambda}_i^\perp\}_i, \{\hat{\boldsymbol{\chi}}_i^\perp\}_i, \{\hat{\boldsymbol{\nu}}_i^\perp\}_i)$ . This, however, follows immediately from the fact that

the class of closed convex polytopes is closed under projections [50, Theorem 19.3] and because  $(\{\hat{\lambda}_i^\perp\}_i, \{\hat{\chi}_i^\perp\}_i, \{\hat{\nu}_i^\perp\}_i)$  is the limit of  $(\{\lambda_i^\perp(t)\}_i, \{\chi_i^\perp(t)\}_i, \{\nu_i^\perp(t)\}_i)$ , which can be extended to a feasible solution of (35) for each  $t \in \mathbb{N}$ . Thus, (35) and, *a fortiori*, (6) are solvable.  $\square$

**Proof of Proposition 3.** Let  $(\{\lambda_i\}_i, \{\chi_i\}_i)$  be any optimal solution of (6), whose existence is guaranteed by Proposition 2, and assume that  $\lambda_0 > 0$ . For the sake of argument, assume further that (8) is violated, that is,  $\mathbf{t}_j(\mathbf{x})^\top \chi_0 \geq \lambda_0 u_j(\mathbf{x})$  for some  $j \in \mathcal{I}$ . Next, define  $(\{\lambda_i^*\}_i, \{\chi_i^*\}_i)$  as

$$\begin{aligned} \lambda_0^* &= 0, & \lambda_j^* &= \lambda_0 + \lambda_j, & \lambda_i^* &= \lambda_i, \quad i \in \mathcal{I}, \quad i \neq j, \\ \chi_0^* &= \mathbf{0}, & \chi_j^* &= \chi_0 + \chi_j, & \chi_i^* &= \chi_i, \quad i \in \mathcal{I}, \quad i \neq j. \end{aligned}$$

We show that  $(\{\lambda_i^*\}_i, \{\chi_i^*\}_i)$  is feasible in (6). Since  $(\{\lambda_i^*\}_i, \{\chi_i^*\}_i)$  adopts a smaller objective value (*i.e.*,  $\lambda_0^* = 0$ ) than  $(\{\lambda_i\}_i, \{\chi_i\}_i)$ , this will contradict the optimality of  $(\{\lambda_i\}_i, \{\chi_i\}_i)$ .

Clearly,  $\lambda_i^* \geq 0$  for all  $i \in \mathcal{I}_0$ , and we have

$$\sum_{i \in \mathcal{I}_0} \lambda_i^* = \sum_{i \in \mathcal{I}_0} \lambda_i = 1 \quad \text{and} \quad \sum_{i \in \mathcal{I}_0} \chi_i^* = \sum_{i \in \mathcal{I}_0} \chi_i = \boldsymbol{\mu}.$$

Similarly, we find

$$\sum_{i \in \mathcal{I}_0} \lambda_i^* \mathbf{d} \left( \frac{\chi_i^*}{\lambda_i^*} \right) = (\lambda_0 + \lambda_j) \mathbf{d} \left( \frac{\chi_0 + \chi_j}{\lambda_0 + \lambda_j} \right) + \sum_{i \in \mathcal{I}, i \neq j} \lambda_i \mathbf{d} \left( \frac{\chi_i}{\lambda_i} \right) \preceq_{\mathcal{D}} \sum_{i \in \mathcal{I}_0} \lambda_i \mathbf{d} \left( \frac{\chi_i}{\lambda_i} \right) \preceq_{\mathcal{D}} \boldsymbol{\sigma},$$

where the first inequality follows from the  $\mathcal{D}$ -convexity of the dispersion function  $\mathbf{d}(\boldsymbol{\xi})$ . Next, we have  $\chi_0^*/\lambda_0^* = \mathbf{0}/0 \in \Xi$ , which holds because  $\Xi$  is nonempty, and we have  $\chi_i^*/\lambda_i^* = \chi_i/\lambda_i \in \Xi$  for all  $i \in \mathcal{I} : i \neq j$ . The convexity of  $\Xi$  further implies that

$$\frac{\chi_j^*}{\lambda_j^*} = \frac{\chi_0 + \chi_j}{\lambda_0 + \lambda_j} = \frac{\lambda_0}{\lambda_0 + \lambda_j} \cdot \frac{\chi_0}{\lambda_0} + \frac{\lambda_j}{\lambda_0 + \lambda_j} \cdot \frac{\chi_j}{\lambda_j} \in \Xi.$$

Finally, it is easy to verify that

$$\mathbf{t}_i(\mathbf{x})^\top \chi_i^* = \mathbf{t}_i(\mathbf{x})^\top \chi_i \geq \lambda_i u_i(\mathbf{x}) = \lambda_i^* u_i(\mathbf{x}) \quad \forall i \in \mathcal{I} : i \neq j$$

and

$$\mathbf{t}_j(\mathbf{x})^\top \chi_j^* = \mathbf{t}_j(\mathbf{x})^\top (\chi_0 + \chi_j) \geq (\lambda_0 + \lambda_j) u_j(\mathbf{x}) = \lambda_j^* u_j(\mathbf{x}),$$

where the inequality in the last expression follows from our assumption that  $\mathbf{t}_j(\mathbf{x})^\top \chi_0 \geq \lambda_0 u_j(\mathbf{x})$ . Thus,  $(\{\lambda_i^*\}_i, \{\chi_i^*\}_i)$  is feasible in (6), which contradicts the optimality of  $(\{\lambda_i\}_i, \{\chi_i\}_i)$ . Therefore, every optimal solution of (6) must satisfy (8).  $\square$

**Proof of Lemma 1.** By the definition of conjugacy, we have

$$(\gamma^\top \mathbf{d})^*(\boldsymbol{\nu}) = \sup_{\boldsymbol{\xi} \in \mathbb{R}^k} \left\{ \boldsymbol{\nu}^\top \boldsymbol{\xi} - \gamma^\top \mathbf{d}(\boldsymbol{\xi}) \right\} = \sup_{\boldsymbol{\xi} \in \mathbb{R}^k, \boldsymbol{\eta} \in \mathbb{R}^d} \left\{ \boldsymbol{\nu}^\top \boldsymbol{\xi} - \gamma^\top \boldsymbol{\eta} : \mathbf{d}(\boldsymbol{\xi}) \preceq_{\mathcal{D}} \boldsymbol{\eta} \right\} = \sigma_{\text{epi}(\mathbf{d})}(\boldsymbol{\nu}, -\boldsymbol{\gamma}),$$

where the epigraph reformulation in the second equality holds because  $\boldsymbol{\gamma} \in \mathcal{D}^*$ .  $\square$

**Proof of Theorem 2.** By Proposition 1 the worst-case probability problem on the left-hand side of (PCC) is equivalent to (6). The conditions **(D')**, **(S')** and **(T)** then imply that (6) simplifies to

$$\begin{aligned} & \underset{\lambda_i, \boldsymbol{\chi}_i}{\text{minimize}} && \lambda_0 \\ & \text{subject to} && \lambda_i \in \mathbb{R}_+, \boldsymbol{\chi}_i \in \Xi, i \in \mathcal{I}_0 \\ & && \sum_{i \in \mathcal{I}_0} \lambda_i = 1 \\ & && \sum_{i \in \mathcal{I}_0} \boldsymbol{\chi}_i = \boldsymbol{\mu} \\ & && \sum_{i \in \mathcal{I}_0} \mathbf{d}(\boldsymbol{\chi}_i) \preceq_{\mathcal{D}} \boldsymbol{\sigma} \\ & && \mathbf{t}_i^\top \boldsymbol{\chi}_i \geq \lambda_i u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}. \end{aligned} \tag{36}$$

Similar techniques as in Proposition 1 show that the strong Lagrangian dual of (36) is

$$\begin{aligned} & \underset{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau_i}{\text{maximize}} && g(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \{\tau_i\}_{i \in \mathcal{I}}) \\ & \text{subject to} && \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\gamma} \in \mathcal{D}^*, \tau_i \in \mathbb{R}_+, i \in \mathcal{I}, \end{aligned} \tag{37}$$

where the dual objective function is representable as

$$\begin{aligned} g(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \{\tau_i\}_{i \in \mathcal{I}}) &= \inf_{\lambda_i \in \mathbb{R}_+, \boldsymbol{\chi}_i \in \Xi, i \in \mathcal{I}_0} \lambda_0 + \alpha \left[ 1 - \sum_{i \in \mathcal{I}_0} \lambda_i \right] + \boldsymbol{\beta}^\top \left[ \boldsymbol{\mu} - \sum_{i \in \mathcal{I}_0} \boldsymbol{\chi}_i \right] \\ &\quad - \boldsymbol{\gamma}^\top \left[ \boldsymbol{\sigma} - \sum_{i \in \mathcal{I}_0} \mathbf{d}(\boldsymbol{\chi}_i) \right] + \sum_{i \in \mathcal{I}} \tau_i \left[ \lambda_i u_i(\mathbf{x}) - \mathbf{t}_i^\top \boldsymbol{\chi}_i \right] \\ &= \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} + \inf_{\lambda_0 \in \mathbb{R}_+} \left\{ \lambda_0 (1 - \alpha) \right\} + \sum_{i \in \mathcal{I}} \inf_{\lambda_i \in \mathbb{R}_+} \left\{ \lambda_i (\tau_i u_i(\mathbf{x}) - \alpha) \right\} \\ &\quad + \inf_{\boldsymbol{\chi}_0 \in \Xi} \left\{ \boldsymbol{\gamma}^\top \mathbf{d}(\boldsymbol{\chi}_0) - \boldsymbol{\beta}^\top \boldsymbol{\chi}_0 \right\} + \sum_{i \in \mathcal{I}} \inf_{\boldsymbol{\chi}_i \in \Xi} \left\{ \boldsymbol{\gamma}^\top \mathbf{d}(\boldsymbol{\chi}_i) - \boldsymbol{\beta}^\top \boldsymbol{\chi}_i - \tau_i \mathbf{t}_i^\top \boldsymbol{\chi}_i \right\}. \end{aligned}$$

The last expression can be further simplified by noting that

$$\begin{aligned}
\inf_{\boldsymbol{\chi}_0 \in \Xi} \left\{ \boldsymbol{\gamma}^\top \mathbf{d}(\boldsymbol{\chi}_0) - \boldsymbol{\beta}^\top \boldsymbol{\chi}_0 \right\} &= - \sup_{\boldsymbol{\chi}_0 \in \mathbb{R}^k} \left\{ \boldsymbol{\beta}^\top \boldsymbol{\chi}_0 - \boldsymbol{\gamma}^\top \mathbf{d}(\boldsymbol{\chi}_0) - \delta_\Xi(\boldsymbol{\chi}_0) \right\} \\
&= - \inf_{\boldsymbol{\nu}_0 \in \mathbb{R}^k} \left\{ (\boldsymbol{\gamma}^\top \mathbf{d})^*(\boldsymbol{\nu}_0) + \sigma_\Xi(\boldsymbol{\beta} - \boldsymbol{\nu}_0) \right\} \\
&= \sup_{\boldsymbol{\nu}_0 \in \mathbb{R}^k} \left\{ -\sigma_{\text{epi}(\mathbf{d})}(\boldsymbol{\nu}_0, -\boldsymbol{\gamma}) - \sigma_\Xi(\boldsymbol{\beta} - \boldsymbol{\nu}_0) \right\},
\end{aligned}$$

where the second equality holds due to [51, Theorem 11.23(a)], whereby the conjugate function of  $\boldsymbol{\gamma}^\top \mathbf{d}(\boldsymbol{\chi}_0) + \delta_\Xi(\boldsymbol{\chi}_0)$  is given by the inf-convolution of the conjugates of  $\boldsymbol{\gamma}^\top \mathbf{d}(\boldsymbol{\chi}_0)$  and  $\delta_\Xi(\boldsymbol{\chi}_0)$ . The third equality follows from Lemma 1. Similarly, one can show that

$$\inf_{\boldsymbol{\chi}_i \in \Xi} \left\{ \boldsymbol{\gamma}^\top \mathbf{d}(\boldsymbol{\chi}_i) - \boldsymbol{\beta}^\top \boldsymbol{\chi}_i - \tau_i \mathbf{t}_i^\top \boldsymbol{\chi}_i \right\} = \sup_{\boldsymbol{\nu}_i \in \mathbb{R}^k} \left\{ -\sigma_{\text{epi}(\mathbf{d})}(\boldsymbol{\nu}_i, -\boldsymbol{\gamma}) - \sigma_\Xi(\boldsymbol{\beta} + \tau_i \mathbf{t}_i - \boldsymbol{\nu}_i) \right\}$$

for every  $i \in \mathcal{I}$ . In summary, the dual problem (37) can thus be reformulated as

$$\begin{aligned}
&\underset{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau_i, \boldsymbol{\nu}_i}{\text{maximize}} && \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} - \sum_{i \in \mathcal{I}_0} \sigma_{\text{epi}(\mathbf{d})}(\boldsymbol{\nu}_i, -\boldsymbol{\gamma}) - \sigma_\Xi(\boldsymbol{\beta} - \boldsymbol{\nu}_0) - \sum_{i \in \mathcal{I}} \sigma_\Xi(\boldsymbol{\beta} + \tau_i \mathbf{t}_i - \boldsymbol{\nu}_i) \\
&\text{subject to} && \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\gamma} \in \mathcal{D}^*, \tau_i \in \mathbb{R}_+, i \in \mathcal{I}, \boldsymbol{\nu}_i \in \mathbb{R}^k, i \in \mathcal{I}_0 \\
&&& \alpha \leq 1 \\
&&& \alpha \leq \tau_i u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}.
\end{aligned}$$

By [51, Example 11.4(b)], the support function of any closed convex cone  $\mathcal{K}$  coincides with the indicator function of the negative dual (*i.e.*, polar) cone  $-\mathcal{K}^*$ . As  $\Xi$  and  $\text{epi}(\mathbf{d})$  are closed convex cones by assumptions **(S')** and **(D')**, respectively, we thus conclude that  $\sigma_\Xi(\boldsymbol{\nu}) = \delta_{\Xi^*}(-\boldsymbol{\nu})$  and  $\sigma_{\text{epi}(\mathbf{d})}(\boldsymbol{\nu}) = \delta_{\text{epi}(\mathbf{d})^*}(-\boldsymbol{\nu})$ . Therefore, problem (37) simplifies to

$$\begin{aligned}
&\underset{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau_i, \boldsymbol{\nu}_i}{\text{maximize}} && \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} - \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \\
&\text{subject to} && \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\gamma} \in \mathcal{D}^*, \tau_i \in \mathbb{R}_+, i \in \mathcal{I}, \boldsymbol{\nu}_i \in \mathbb{R}^k, i \in \mathcal{I}_0 \\
&&& \boldsymbol{\nu}_0 - \boldsymbol{\beta} \in \Xi^* \\
&&& \boldsymbol{\nu}_i - \boldsymbol{\beta} - \tau_i \mathbf{t}_i \in \Xi^* \quad \forall i \in \mathcal{I} \\
&&& (-\boldsymbol{\nu}_i, \boldsymbol{\gamma}) \in \text{epi}(\mathbf{d})^* \quad \forall i \in \mathcal{I}_0 \\
&&& \alpha \leq 1 \\
&&& \alpha \leq \tau_i u_i(\mathbf{x}) \quad \forall i \in \mathcal{I}.
\end{aligned} \tag{38}$$

Note that the variable  $\lambda_0$  in the primal problem (36) is the Lagrange multiplier of the constraint  $\alpha \leq 1$  in the dual problem (38). By complementary slackness, which applies because both (36) and

(38) are solvable,<sup>3</sup> any pair of optimal primal and dual solutions must satisfy  $\lambda_0(1 - \alpha) = 0$ . As  $\lambda_0$  also constitutes the objective function of (36), we have that  $\alpha = 1$  whenever the two optimization problems have a strictly positive optimal value.

Consider now a restriction of problem (38) with the additional constraint  $\alpha = 1$ . By the above discussion, the restricted problem shares the common optimal value of (36) and (38) whenever this value is strictly positive. However, as  $\epsilon < 1$  by assumption, the chance constraint (PCC) is infeasible for any  $\mathbf{x}$  associated with a zero worst-case probability. We may thus set  $\alpha = 1$  at no loss. This gives rise to the hyperbolic constraints  $1 \leq \tau_i u_i(\mathbf{x})$ ,  $i \in \mathcal{I}$ , which can be reformulated as explicit second-order cone constraints.<sup>4</sup> The claim then follows from the observation that (PCC) is satisfied whenever (38) has a feasible solution whose objective value exceeds  $1 - \epsilon$ .  $\square$

**Proof of Lemma 2.** The ‘only if’ direction is immediate. As for the ‘if’ direction, fix a fractional vector  $\mathbf{y}$  as described in the statement and let  $\mathbf{y}'$  be the closest binary vector, that is,  $y'_j := 1$  if  $y_j \geq 1 - \kappa$ ;  $:= 0$  if  $y_j \leq \kappa$ . We then observe that  $\sum_j A_{ij} y'_j \leq \sum_j A_{ij} y_j + \sum_j |A_{ij}| \kappa < \sum_j A_{ij} y_j + 1 \leq b_i + 1$  for all  $i = 1, \dots, m$ . Due to the integrality of  $\mathbf{A}$ ,  $\mathbf{y}'$  and  $\mathbf{b}$ , we thus conclude that  $\mathbf{A}\mathbf{y}' \leq \mathbf{b}$ .  $\square$

**Proof of Lemma 3.** We first argue that the worst-case probability on the left-hand side of the chance constraint in (10) is identical for all  $\mathbf{x} \in \{-1, 1\}^n$ . To see this, choose any  $\mathbf{x}, \mathbf{x}' \in \{-1, 1\}^n$ . Then for any  $\mathbb{P} \in \mathcal{P}$ , we have

$$\mathbb{P}\left(-3\mathbf{e} < \mathbf{x} + \tilde{\boldsymbol{\xi}} < 3\mathbf{e}\right) = \mathbb{Q}\left(-3\mathbf{e} < \mathbf{x}' + \tilde{\boldsymbol{\xi}} < 3\mathbf{e}\right)$$

for the distribution  $\mathbb{Q} \in \mathcal{P}$  that satisfies  $\mathbb{Q}(\tilde{\boldsymbol{\xi}} \in A) = \mathbb{P}(f(\tilde{\boldsymbol{\xi}}) \in A)$  for all Borel-measurable sets  $A \subseteq \mathbb{R}^n$ , where  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined through  $f_i(\boldsymbol{\xi}) = \xi_i$  if  $x_i = x'_i$ ;  $= -\xi_i$  otherwise. Thus, in the remainder of the proof we assume without loss of generality that  $\mathbf{x} = -\mathbf{e}$ , in which case the worst-case probability on the left-hand side of the chance constraint in (10) simplifies to

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(-2\mathbf{e} < \tilde{\boldsymbol{\xi}} < 4\mathbf{e}\right).$$

From the proof of Theorem 1 we know that this expression equals the optimal value of the problem

<sup>3</sup>We stress that (36) is solvable by Proposition 1. One can show that its dual (38) is also solvable.

<sup>4</sup>Note that by the nonnegativity of  $\tau_i$ , the constraint  $1 \leq \tau_i u_i(\mathbf{x})$  is only satisfiable if  $\tau_i > 0$  as well as  $u_i(\mathbf{x}) > 0$ .

$$\begin{aligned}
& \underset{\alpha, \beta, \gamma, \zeta}{\text{maximize}} && \alpha - \frac{4(\epsilon - \delta)}{n} \gamma^\top \mathbf{e} - \frac{\delta^2}{n} \zeta^\top \mathbf{e} \\
& \text{subject to} && \alpha \in \mathbb{R}, \beta \in \mathbb{R}^n, \gamma, \zeta \in \mathbb{R}_+^n \\
& && \alpha + \beta^\top \boldsymbol{\xi} - \gamma^\top |\boldsymbol{\xi}| - \zeta^\top \mathbf{d}(\boldsymbol{\xi}) \leq 1 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n \\
& && \alpha + \beta^\top \boldsymbol{\xi} - \gamma^\top |\boldsymbol{\xi}| - \zeta^\top \mathbf{d}(\boldsymbol{\xi}) \leq 0 \quad \forall i \in \mathcal{I}, \forall \boldsymbol{\xi} \in \mathbb{R}^n : \xi_i \geq 4 \\
& && \alpha + \beta^\top \boldsymbol{\xi} - \gamma^\top |\boldsymbol{\xi}| - \zeta^\top \mathbf{d}(\boldsymbol{\xi}) \leq 0 \quad \forall i \in \mathcal{I}, \forall \boldsymbol{\xi} \in \mathbb{R}^n : \xi_i \leq -2
\end{aligned} \tag{39}$$

We now show that  $\alpha = 1$ ,  $\beta = \gamma = \mathbf{e}/4$  and  $\zeta = \mathbf{e}/\delta$  is feasible in problem (39). The statement of the lemma then follows since this solution attains the objective value  $1 - \epsilon$  in (39).

For the postulated solution  $(\alpha, \beta, \gamma, \zeta)$ , the function  $\alpha + \beta^\top \boldsymbol{\xi} - \gamma^\top |\boldsymbol{\xi}| - \zeta^\top \mathbf{d}(\boldsymbol{\xi})$  appearing on the left-hand sides of the constraints in (39) satisfies

$$\alpha + \beta^\top \boldsymbol{\xi} - \gamma^\top |\boldsymbol{\xi}| - \zeta^\top \mathbf{d}(\boldsymbol{\xi}) = 1 + \frac{1}{4} \mathbf{e}^\top (\boldsymbol{\xi} - |\boldsymbol{\xi}|) - \frac{1}{\delta} \mathbf{e}^\top \max \{ \boldsymbol{\xi} - (4 - \delta) \mathbf{e}, \mathbf{0}, -\boldsymbol{\xi} - (4 - \delta) \mathbf{e} \}.$$

Over  $\mathbb{R}^n$ , this function attains its maximum value of 1 at any point  $\boldsymbol{\xi}^* \in [0, 4 - \delta]^n$ . Hence,  $(\alpha, \beta, \gamma, \zeta)$  satisfies the first semi-infinite constraint in (39).

In view of the second semi-infinite constraint in (39), we observe that for each  $i \in \mathcal{I}$ , we have

$$\begin{aligned}
& \max_{\boldsymbol{\xi}} \left\{ 1 + \frac{1}{4} \mathbf{e}^\top (\boldsymbol{\xi} - |\boldsymbol{\xi}|) - \frac{1}{\delta} \mathbf{e}^\top \max \{ \boldsymbol{\xi} - (4 - \delta) \mathbf{e}, \mathbf{0}, -\boldsymbol{\xi} - (4 - \delta) \mathbf{e} \} : \boldsymbol{\xi} \in \mathbb{R}^n, \xi_i \geq 4 \right\} \\
& \leq \min_{\tau, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi}} \left\{ \begin{array}{l} \tau \in \mathbb{R}_+, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi} \in \mathbb{R}_+^n \\ 1 + (4 - \delta) \mathbf{e}^\top (\boldsymbol{\phi} + \boldsymbol{\psi}) - 4\tau : \mathbf{e}/4 - \boldsymbol{\phi} + \boldsymbol{\psi} + \tau \mathbf{e}_i = \boldsymbol{\theta} - \boldsymbol{\eta} \\ \boldsymbol{\theta} + \boldsymbol{\eta} = \mathbf{e}/4, \boldsymbol{\phi} + \boldsymbol{\psi} \leq \mathbf{e}/\delta \end{array} \right\}
\end{aligned}$$

due to weak LP duality. The minimum on the right-hand side of the second inequality is nonpositive since  $\tau = 1/\delta$ ,  $\boldsymbol{\theta} = \mathbf{e}/4$ ,  $\boldsymbol{\eta} = \mathbf{0}$ ,  $\boldsymbol{\phi} = \mathbf{e}_i/\delta$  and  $\boldsymbol{\psi} = \mathbf{0}$  is feasible in the minimization problem and attains an objective value of 0.

Applying the same argument to the third semi-infinite constraint in (39), we observe that the  $i$ -th constraint,  $i \in \mathcal{I}$ , is satisfied whenever

$$\min_{\tau, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi}} \left\{ \begin{array}{l} \tau \in \mathbb{R}_+, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi} \in \mathbb{R}_+^n \\ 1 + (4 - \delta) \mathbf{e}^\top (\boldsymbol{\phi} + \boldsymbol{\psi}) - 2\tau : \mathbf{e}/4 - \boldsymbol{\phi} + \boldsymbol{\psi} - \tau \mathbf{e}_i = \boldsymbol{\theta} - \boldsymbol{\eta} \\ \boldsymbol{\theta} + \boldsymbol{\eta} = \mathbf{e}/4, \boldsymbol{\phi} + \boldsymbol{\psi} \leq \mathbf{e}/\delta \end{array} \right\}$$

is nonpositive, which holds since  $\tau = 1/2$ ,  $\boldsymbol{\theta} = (\mathbf{e} - \mathbf{e}_i)/4$ ,  $\boldsymbol{\eta} = \mathbf{e}_i/4$ ,  $\boldsymbol{\phi} = \mathbf{0}$  and  $\boldsymbol{\psi} = \mathbf{0}$  is feasible in the minimization problem and attains an objective value of 0.  $\square$

**Proof of Lemma 4.** We denote by  $\mathcal{J}(\mathbf{x}) = \{i \in \{1, \dots, n\} : x_i \in [-1, -1 + \delta) \cup (1 - \delta, 1]\}$  the components of  $\mathbf{x}$  that are close to  $-1$  or  $1$ , and we consider the distribution  $\mathbb{P}^*$  defined by

$$\mathbb{P}^*(\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}) = \begin{cases} \phi & \text{for } \boldsymbol{\xi} = \boldsymbol{\varphi} \\ \frac{2(\epsilon - \delta)}{n(3 + x_i)} & \text{for } \boldsymbol{\xi} = (-3 - x_i)\mathbf{e}_i \quad \forall i \in \mathcal{I} \\ \frac{(x_i + 1)\epsilon + 2\delta}{n(3 + x_i)} & \text{for } \boldsymbol{\xi} = (3 - x_i)\mathbf{e}_i \quad \forall i \in \mathcal{I} \\ \frac{2(\epsilon - \delta)(3 - x_i)}{9n} & \text{for } \boldsymbol{\xi} = (-3 - x_i)\mathbf{e}_i \quad \forall i \in \mathcal{I} \setminus \mathcal{J}(\mathbf{x}) \\ \frac{2(\epsilon - \delta)(3 + x_i)}{9n} & \text{for } \boldsymbol{\xi} = (3 - x_i)\mathbf{e}_i \quad \forall i \in \mathcal{I} \setminus \mathcal{J}(\mathbf{x}), \end{cases} \quad (40)$$

where  $\phi = 1 - |\mathcal{J}(\mathbf{x})|\frac{\epsilon}{n} - (n - |\mathcal{J}(\mathbf{x})|)\frac{4(\epsilon - \delta)}{3n}$  and the components of  $\boldsymbol{\varphi}$  satisfy

$$\varphi_i = \frac{12(\epsilon - \delta) + \epsilon(x_i^2 - 9)}{\phi n(3 + x_i)}$$

if  $i \in \mathcal{J}(\mathbf{x})$ ;  $= 0$  otherwise. By construction, we have  $\mathbb{P}^* \in \mathcal{P}$  and

$$\mathbb{P}^* \left( -3\mathbf{e} < \mathbf{x} + \tilde{\boldsymbol{\xi}} < 3\mathbf{e} \right) = \phi$$

since the realizations  $\boldsymbol{\xi} \in \{(-3 - x_1)\mathbf{e}_1, (3 - x_1)\mathbf{e}_1, \dots, (-3 - x_n)\mathbf{e}_n, (3 - x_n)\mathbf{e}_n\}$  all violate the constraints inside the probability expression. The assumption that  $\delta \leq \frac{\epsilon}{8}$ , however, implies that  $\frac{\epsilon}{n} < \frac{4(\epsilon - \delta)}{3n}$  and therefore  $\phi < 1 - \epsilon$ , which concludes the proof.  $\square$

**Proof of Theorem 3.** We set  $\delta = \min\{\frac{\epsilon}{8}, 2\kappa\}$  in the definition of the ambiguity set (11), where  $\kappa$  is chosen as prescribed by Lemma 2. We show that problem (10) is feasible if and only if the IP Feasibility problem has an affirmative answer.

Assume first that there is  $\mathbf{z} \in \{0, 1\}^n$  with  $\mathbf{A}\mathbf{z} \leq \mathbf{b}$ . Lemma 3 then implies that the chance constraint in (10) is satisfied by  $\mathbf{x} = 2\mathbf{z} - \mathbf{e}$ , that is, problem (10) is feasible.

Assume now that there is no binary solution to the IP feasibility problem. Lemma 2 then implies that there is no  $\mathbf{z} \in ([0, \kappa] \cup [1 - \kappa, 1])^n$  that satisfies  $\mathbf{A}\mathbf{z} \leq \mathbf{b}$ . For any other solution  $\mathbf{z} \in [0, 1]^n \setminus ([0, \kappa] \cup [1 - \kappa, 1])^n$ , however, Lemma 4 implies that  $\mathbf{x} = 2\mathbf{z} - \mathbf{e}$  violates the chance constraint in (10). We thus conclude that problem (10) is infeasible.  $\square$



**Proof of Lemma 5.** The feasible region of the safety condition in (14) is monotonically increasing in the realizations of  $\tilde{\boldsymbol{\xi}}$  with respect to set inclusion. Thus, the worst-case probability in (14) does not change if we increase the ambiguity set  $\mathcal{P}$  in (13) to

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0([-1, 3]^{2n}) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] \geq \left(3 - \frac{2\epsilon}{n}\right) \mathbf{e} \right\},$$

where we have relaxed the expectation constraint to an inequality. For the new ambiguity set  $\mathcal{P}'$ , similar arguments as in the proof of Theorem 1 show that the worst-case probability in (14) affords the semi-infinite dual formulation

$$\begin{aligned} & \underset{\alpha, \boldsymbol{\beta}}{\text{maximize}} && \alpha + \left(3 - \frac{2\epsilon}{n}\right) \mathbf{e}^\top \boldsymbol{\beta} \\ & \text{subject to} && \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}_+^{2n} \\ & && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 1 && \forall \boldsymbol{\xi} \in \Xi \\ & && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 0 && \forall i \in \mathcal{I}, \forall \boldsymbol{\xi} \in \Xi : \xi_i < x_i \\ & && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 0 && \forall i \in \mathcal{I}, \forall \boldsymbol{\xi} \in \Xi : \xi_{n+i} < -x_i, \end{aligned}$$

where  $\Xi = [-1, 3]^{2n}$ . Note that the second constraint vanishes whenever  $x_i = -1$ , and a continuity argument allows us to replace the strict inequality with a weak one whenever  $x_i > -1$ . Likewise, the third constraint vanishes whenever  $x_i = 1$ , and we can replace the strict inequality with a weak one whenever  $x_i < 1$ . The worst-case probability in (14) is thus equivalent to

$$\begin{aligned} & \underset{\alpha, \boldsymbol{\beta}}{\text{maximize}} && \alpha + \left(3 - \frac{2\epsilon}{n}\right) \mathbf{e}^\top \boldsymbol{\beta} \\ & \text{subject to} && \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}_+^{2n} \\ & && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 1 && \forall \boldsymbol{\xi} \in \Xi \\ & && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 0 && \forall i \in \underline{\mathcal{I}}(\mathbf{x}), \forall \boldsymbol{\xi} \in \Xi : \xi_i \leq x_i \\ & && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} \leq 0 && \forall i \in \overline{\mathcal{I}}(\mathbf{x}), \forall \boldsymbol{\xi} \in \Xi : \xi_{n+i} \leq -x_i, \end{aligned}$$

where we have introduced the shorthand notations  $\underline{\mathcal{I}}(\mathbf{x}) = \{i \in \{1, \dots, n\} : x_i > -1\}$  and  $\overline{\mathcal{I}}(\mathbf{x}) = \{i \in \{1, \dots, n\} : x_i < 1\}$  for  $\mathbf{x} \in [-1, 1]^n$ . Since  $\boldsymbol{\beta} \geq \mathbf{0}$ , we can replace each semi-infinite constraint

with a single constraint where  $\boldsymbol{\xi}$  attains its component-wise largest value:

$$\begin{aligned}
& \underset{\alpha, \boldsymbol{\beta}}{\text{maximize}} && \alpha + \left(3 - \frac{2\epsilon}{n}\right) \mathbf{e}^\top \boldsymbol{\beta} \\
& \text{subject to} && \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}_+^{2n} \\
& && \alpha + 3\mathbf{e}^\top \boldsymbol{\beta} \leq 1 \\
& && \alpha + 3\mathbf{e}^\top \boldsymbol{\beta} - (3 - x_i)\beta_i \leq 0 \quad \forall i \in \underline{\mathcal{I}}(\mathbf{x}) \\
& && \alpha + 3\mathbf{e}^\top \boldsymbol{\beta} - (3 + x_i)\beta_{n+i} \leq 0 \quad \forall i \in \bar{\mathcal{I}}(\mathbf{x})
\end{aligned} \tag{41}$$

As in the proof of Theorem 2 we can employ a complementary slackness argument to conclude that either the optimal value of (41) is 0 or the first constraint in (41) is binding. In the second case, we can replace  $\alpha$  with  $1 - 3\mathbf{e}^\top \boldsymbol{\beta}$  to obtain the equivalent reformulation

$$\begin{aligned}
& \underset{\boldsymbol{\beta}}{\text{maximize}} && 1 - \frac{2\epsilon}{n} \cdot \mathbf{e}^\top \boldsymbol{\beta} \\
& \text{subject to} && \boldsymbol{\beta} \in \mathbb{R}_+^{2n} \\
& && 1/(3 - x_i) \leq \beta_i \quad \forall i \in \underline{\mathcal{I}}(\mathbf{x}) \\
& && 1/(3 + x_i) \leq \beta_{n+i} \quad \forall i \in \bar{\mathcal{I}}(\mathbf{x}).
\end{aligned}$$

Since the objective function is strictly monotonically decreasing in  $\boldsymbol{\beta}$ , this problem has the optimal solution  $\beta_i^* = 1/(3 - x_i)$  if  $i \in \underline{\mathcal{I}}(\mathbf{x})$ ;  $= 1/(3 + x_i)$  if  $i \in \bar{\mathcal{I}}(\mathbf{x})$ ;  $= 0$  otherwise. In summary, the optimal value of (41) is given by the maximum of 0 and

$$1 - \frac{2\epsilon}{n} \left( \sum_{i \in \underline{\mathcal{I}}(\mathbf{x})} \frac{1}{3 - x_i} + \sum_{i \in \bar{\mathcal{I}}(\mathbf{x})} \frac{1}{3 + x_i} \right),$$

which concludes the proof.  $\square$

The proof of Theorem 4 relies on the following auxiliary result, which we prove first.

**Lemma 9.** *For any instance of the pessimistic chance constrained program (2),  $\mathbf{x} \in \mathcal{X}$  satisfies*

$$\begin{aligned}
& \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} < \mathbf{u}(\mathbf{x}) + y\mathbf{e}] \geq 1 - \epsilon \quad \forall y > 0 \\
& \iff \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} \leq \mathbf{u}(\mathbf{x}) + y\mathbf{e}] \geq 1 - \epsilon \quad \forall y > 0 \\
& \iff \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} \leq \mathbf{u}(\mathbf{x})] \geq 1 - \epsilon.
\end{aligned} \tag{42}$$

**Proof.** One readily verifies the first equivalence in (42). To prove the second equivalence, we show that the mapping  $y \mapsto \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} - \mathbf{u}(\mathbf{x}) \leq y\mathbf{e}]$  is right-continuous. Note that the

related mapping  $y \mapsto \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} - \mathbf{u}(\mathbf{x}) \leq y\mathbf{e}]$ , which can be interpreted as the distribution function of the random variable  $\max_i \{\mathbf{t}_i(\mathbf{x})^\top \tilde{\boldsymbol{\xi}} - u_i(\mathbf{x})\}$ , is both right-continuous and non-decreasing, and is thus upper semicontinuous. Since the infimum over  $\mathbb{P} \in \mathcal{P}$  preserves monotonicity and upper semicontinuity, the mapping  $y \mapsto \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\mathbf{T}(\mathbf{x})\tilde{\boldsymbol{\xi}} - \mathbf{u}(\mathbf{x}) \leq y\mathbf{e}]$  is indeed right-continuous.  $\square$

**Proof of Theorem 4.** We show that the optimal value of problem (12) is zero if and only if the IP Feasibility problem has an affirmative answer. To this end, suppose that there is  $\mathbf{z} \in \{0, 1\}^n$  with  $\mathbf{A}\mathbf{z} \leq \mathbf{b}$ . We have that  $\mathbf{x} = 2\mathbf{z} - \mathbf{e} \in \{-1, 1\}^n$ ,  $\mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b}$ , and Lemma 5 implies that

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( -\tilde{\xi}_{n+i} \leq x_i \leq \tilde{\xi}_i \quad \forall i = 1, \dots, n \right) \geq 1 - \epsilon.$$

Lemma 9 then allows us to conclude that  $(\mathbf{x}, y)$  is feasible in (12) for any  $y > 0$ , that is, the optimal value of (12) is zero. Assume now that problem (12) has a feasible solution  $\mathbf{x}$  that achieves an objective value of zero, that is, for all  $y > 0$ , we have

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( -\tilde{\xi}_{n+i} - y < x_i < \tilde{\xi}_i + y \quad \forall i = 1, \dots, n \right) \geq 1 - \epsilon.$$

Lemma 9 then implies that

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( -\tilde{\xi}_{n+i} \leq x_i \leq \tilde{\xi}_i \quad \forall i = 1, \dots, n \right) \geq 1 - \epsilon,$$

and Lemma 5 allows us to conclude that  $\mathbf{x} \in \{-1, 1\}^n$ . Since  $\mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b}$  by construction,  $\mathbf{z} = (\mathbf{x} + \mathbf{e})/2$  is binary and satisfies  $\mathbf{A}\mathbf{z} \leq \mathbf{b}$ , that is, it solves the IP Feasibility problem.  $\square$

**Proof of Lemma 6.** The proof of Theorem 1 implies that the worst-case probability on the left-hand side of (17) equals the optimal value of the nonconvex optimization problem

$$\begin{aligned} & \underset{\lambda_0, \lambda_i^+, \lambda_i^0, \boldsymbol{\xi}_0, \boldsymbol{\xi}_i^+, \boldsymbol{\xi}_i^-}{\text{minimize}} && \lambda_0 \\ & \text{subject to} && \lambda_0 \in \mathbb{R}_+, \quad \lambda_i^+ \in \mathbb{R}_+, i \in \mathcal{I}^+(\mathbf{x}), \quad \lambda_i^- \in \mathbb{R}_+, i \in \mathcal{I}^-(\mathbf{x}) \\ & && \boldsymbol{\xi}_0 \in \mathbb{R}^{2n}, \quad \boldsymbol{\xi}_i^+ \in \mathbb{R}^{2n}, i \in \mathcal{I}^+(\mathbf{x}), \quad \boldsymbol{\xi}_i^- \in \mathbb{R}^{2n}, i \in \mathcal{I}^-(\mathbf{x}) \\ & && \lambda_0 + \sum_{i \in \mathcal{I}^+(\mathbf{x})} \lambda_i^+ + \sum_{i \in \mathcal{I}^-(\mathbf{x})} \lambda_i^- = 1 \\ & && \lambda_0 \boldsymbol{\xi}_0 + \sum_{i \in \mathcal{I}^+(\mathbf{x})} \lambda_i^+ \boldsymbol{\xi}_i^+ + \sum_{i \in \mathcal{I}^-(\mathbf{x})} \lambda_i^- \boldsymbol{\xi}_i^- = \mathbf{0} \\ & && \lambda_0 |\boldsymbol{\xi}_0| + \sum_{i \in \mathcal{I}^+(\mathbf{x})} \lambda_i^+ |\boldsymbol{\xi}_i^+| + \sum_{i \in \mathcal{I}^-(\mathbf{x})} \lambda_i^- |\boldsymbol{\xi}_i^-| \leq \frac{2\epsilon}{n} \mathbf{e} \\ & && \boldsymbol{\xi}_{i,i}^+ + 1 \leq 0, i \in \mathcal{I}^+(\mathbf{x}), \quad \boldsymbol{\xi}_{i,n+i}^- + 1 \leq 0, i \in \mathcal{I}^-(\mathbf{x}), \end{aligned} \tag{43}$$

where  $\mathcal{I}^+(\mathbf{x}) = \{i \in \{1, \dots, n\} : x_i > 0\}$  and  $\mathcal{I}^-(\mathbf{x}) = \{i \in \{1, \dots, n\} : x_i < 1\}$ . Performing the substitutions  $\xi_0 \leftarrow \xi_0 - \mathbf{e}$ ,  $\xi_i^+ \leftarrow \xi_i^+ - \mathbf{e}$ , and  $\xi_i^- \leftarrow \xi_i^- - \mathbf{e}$ , we obtain the equivalent problem

$$\begin{aligned}
& \underset{\lambda_0, \lambda_i^+, \lambda_i^0, \xi_0, \xi_i^+, \xi_i^-}{\text{minimize}} && \lambda_0 \\
& \text{subject to} && \lambda_0 \in \mathbb{R}_+, \lambda_i^+ \in \mathbb{R}_+, i \in \mathcal{I}^+(\mathbf{x}), \lambda_i^- \in \mathbb{R}_+, i \in \mathcal{I}^-(\mathbf{x}) \\
& && \xi_0 \in \mathbb{R}^{2n}, \xi_i^+ \in \mathbb{R}^{2n}, i \in \mathcal{I}^+(\mathbf{x}), \xi_i^- \in \mathbb{R}^{2n}, i \in \mathcal{I}^-(\mathbf{x}) \\
& && \lambda_0 + \sum_{i \in \mathcal{I}^+(\mathbf{x})} \lambda_i^+ + \sum_{i \in \mathcal{I}^-(\mathbf{x})} \lambda_i^- = 1 \\
& && \lambda_0 \xi_0 + \sum_{i \in \mathcal{I}^+(\mathbf{x})} \lambda_i^+ \xi_i^+ + \sum_{i \in \mathcal{I}^-(\mathbf{x})} \lambda_i^- \xi_i^- = \mathbf{e} \\
& && \lambda_0 |\xi_0 - \mathbf{e}| + \sum_{i \in \mathcal{I}^+(\mathbf{x})} \lambda_i^+ |\xi_i^+ - \mathbf{e}| + \sum_{i \in \mathcal{I}^-(\mathbf{x})} \lambda_i^- |\xi_i^- - \mathbf{e}| \leq \frac{2\epsilon}{n} \mathbf{e} \\
& && \xi_{i,i}^+ \leq 0, i \in \mathcal{I}^+(\mathbf{x}), \xi_{i,n+i}^- \leq 0, i \in \mathcal{I}^-(\mathbf{x}),
\end{aligned} \tag{44}$$

The third constraint line in (44) implies that the objective value of (44) decreases when any of the components of  $\lambda^+$  or  $\lambda^-$  increases. Moreover, we argue that every feasible solution to (44) satisfies  $\lambda_i^+ \leq \epsilon/n$ ,  $i \in \mathcal{I}^+(\mathbf{x})$ , and  $\lambda_i^- \leq \epsilon/n$ ,  $i \in \mathcal{I}^-(\mathbf{x})$ . Indeed, assume to the contrary that  $\lambda_i^+ > \epsilon/n$  for some  $i \in \mathcal{I}^+(\mathbf{x})$ . As  $\xi_{i,i}^+ \leq 0$ , the fifth constraint line in (44) then implies that

$$\begin{aligned}
& \lambda_0 |\xi_{0,i} - 1| + \sum_{\substack{j \in \mathcal{I}^+(\mathbf{x}), \\ j \neq i}} \lambda_j^+ |\xi_{j,i}^+ - 1| + \sum_{j \in \mathcal{I}^-(\mathbf{x})} \lambda_j^- |\xi_{j,i}^- - 1| \leq \frac{2\epsilon}{n} - \lambda_i^+ \\
\implies & \lambda_0 (\xi_{0,i} - 1) + \sum_{\substack{j \in \mathcal{I}^+(\mathbf{x}), \\ j \neq i}} \lambda_j^+ (\xi_{j,i}^+ - 1) + \sum_{j \in \mathcal{I}^-(\mathbf{x})} \lambda_j^- (\xi_{j,i}^- - 1) \leq \frac{2\epsilon}{n} - \lambda_i^+ \\
\iff & \lambda_0 \xi_{0,i} + \sum_{\substack{j \in \mathcal{I}^+(\mathbf{x}), \\ j \neq i}} \lambda_j^+ \xi_{j,i}^+ + \sum_{j \in \mathcal{I}^-(\mathbf{x})} \lambda_j^- \xi_{j,i}^- \leq 1 + \frac{2\epsilon}{n} - 2\lambda_i^+
\end{aligned}$$

If  $\lambda_i^+ > \epsilon/n$ , however, this implies that the fourth constraint line in (44) is violated since  $\lambda_i^+ \xi_{i,i}^+ \leq 0$ . Since a similar argument can be constructed for the case where  $\lambda_i^- > \epsilon/n$ , we conclude that every feasible solution to problem (44) indeed satisfies  $\lambda_i^+ \leq \epsilon/n$  and  $\lambda_i^- \leq \epsilon/n$ .

One readily verifies that the solution

$$\left\{ \begin{array}{ll}
(\lambda_i^+, \xi_i^+) = \left(\frac{\epsilon}{n}, \mathbf{e} - \mathbf{e}_i\right) & \text{if } x_i = 1, \\
(\lambda_i^-, \xi_i^-) = \left(\frac{\epsilon}{n}, \mathbf{e} - \mathbf{e}_{n+i}\right) & \text{if } x_i = 0, \\
(\lambda_i^+, \xi_i^+) = \left(\frac{\epsilon}{n}, \mathbf{e} - \mathbf{e}_i + \mathbf{e}_{n+i}\right), \\
(\lambda_i^-, \xi_i^-) = \left(\frac{\epsilon}{n}, \mathbf{e} - \mathbf{e}_{n+i} + \mathbf{e}_i\right) & \text{if } x_i \in (0, 1),
\end{array} \right\}$$

as well as  $\lambda_0 = 1 - \frac{\epsilon}{n}(|\mathcal{I}^+(\mathbf{x})| + |\mathcal{I}^-(\mathbf{x})|)$  and

$$\boldsymbol{\xi}_0 = \mathbf{e} + \frac{\epsilon}{n\lambda_0} \left( \sum_{i \in \mathcal{I}^+(\mathbf{x}) \setminus \mathcal{I}^-(\mathbf{x})} \mathbf{e}_i + \sum_{i \in \mathcal{I}^-(\mathbf{x}) \setminus \mathcal{I}^+(\mathbf{x})} \mathbf{e}_{n+i} \right)$$

is feasible in (44). Our previous discussion implies that this solution is indeed optimal, and the value of  $\lambda_0$  coincides with the expression on the right-hand side of (17).  $\square$

**Proof of Theorem 5.** As in the proof of Theorem 4, we show that the optimal value of problem (15) is zero if and only if the IP Feasibility problem has an affirmative answer. To this end, suppose that there is  $\mathbf{z} \in \{0, 1\}^n$  with  $\mathbf{A}\mathbf{z} \leq \mathbf{b}$ . In that case, Lemma 6 implies that

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( (\tilde{\xi}_i + 1)x_i \geq 0, (\tilde{\xi}_{n+i} + 1)(1 - x_i) \geq 0 \quad \forall i = 1, \dots, n \right) \geq 1 - \epsilon.$$

Lemma 9 then allows us to conclude that  $(\mathbf{x}, y)$  is feasible in (15) for any  $y > 0$ , that is, the optimal value of (15) is zero. Assume now that problem (15) has a feasible solution  $\mathbf{x}$  that achieves an objective value of zero, that is, we have

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( (\tilde{\xi}_i + 1)x_i + y > 0, (\tilde{\xi}_{n+i} + 1)(1 - x_i) + y > 0 \quad \forall i = 1, \dots, n \right) \geq 1 - \epsilon$$

for all  $y > 0$ . Lemma 9 then implies that

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left( (\tilde{\xi}_i + 1)x_i \geq 0, (\tilde{\xi}_{n+i} + 1)(1 - x_i) \geq 0 \quad \forall i = 1, \dots, n \right) \geq 1 - \epsilon,$$

and Lemma 6 allows us to conclude that  $\mathbf{x} \in \{0, 1\}^n$ . Since  $\mathbf{A}\mathbf{x} \leq \mathbf{b}$  by construction,  $\mathbf{x}$  solves the IP Feasibility problem.  $\square$

**Proof of Theorem 6.** As in the proof of Theorem 1, the best-case probability in (OCC) can be expressed as the optimal value of a moment problem, whose dual semi-infinite linear program is representable as

$$\begin{aligned} & \underset{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}}{\text{minimize}} && \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \\ & \text{subject to} && \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\gamma} \in \mathcal{D}^* \\ & && \alpha + \boldsymbol{\xi}^\top \boldsymbol{\beta} + \mathbf{d}(\boldsymbol{\xi})^\top \boldsymbol{\gamma} \geq \mathbb{I}_{[\mathbf{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})]} \quad \forall \boldsymbol{\xi} \in \Xi. \end{aligned} \tag{45}$$

Strong duality holds due to the assumptions **(A)** and **(D)**, and thus the optimal value of (45) coincides with the best-case probability in (OCC). By decomposing the indicator function in the

semi-infinite constraint, (45) can be reformulated as

$$\begin{aligned}
& \underset{\alpha, \beta, \gamma}{\text{minimize}} && \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \\
& \text{subject to} && \alpha \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{R}^k, \quad \boldsymbol{\gamma} \in \mathcal{D}^\star \\
& && \alpha + \boldsymbol{\xi}_0^\top \boldsymbol{\beta} + \mathbf{d}(\boldsymbol{\xi}_0)^\top \boldsymbol{\gamma} \geq 1 \quad \forall \boldsymbol{\xi}_0 \in \Xi_0(\mathbf{x}) \\
& && \alpha + \boldsymbol{\xi}_1^\top \boldsymbol{\beta} + \mathbf{d}(\boldsymbol{\xi}_1)^\top \boldsymbol{\gamma} \geq 0 \quad \forall \boldsymbol{\xi}_1 \in \Xi_1(\mathbf{x}),
\end{aligned} \tag{46}$$

where  $\Xi_0(\mathbf{x}) = \{\boldsymbol{\xi} \in \Xi : \mathbf{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})\}$  and  $\Xi_1(\mathbf{x}) = \Xi$ . Note that  $\Xi_0(\mathbf{x})$  is nonempty by assumption. As  $\boldsymbol{\gamma} \in \mathcal{D}^\star$  and  $\mathbf{d}(\boldsymbol{\xi})$  is  $\mathcal{D}$ -convex by assumption **(D)**,  $\alpha + \boldsymbol{\xi}_0^\top \boldsymbol{\beta} - \mathbf{d}(\boldsymbol{\xi}_0)^\top \boldsymbol{\gamma}$  is concave in  $\boldsymbol{\xi}$  [11, p. 110] and, *a fortiori*, continuous [50, Theorem 10.1]. We can therefore replace  $\Xi_0(\mathbf{x})$  in problem (46) with its closure  $\text{cl } \Xi_0(\mathbf{x}) = \{\boldsymbol{\xi} \in \Xi : \mathbf{T}(\mathbf{x})\boldsymbol{\xi} \leq \mathbf{u}(\mathbf{x})\}$ . As in the proof of Theorem 1, we may then interpret (46) as the robust counterpart of an uncertain convex program with constraint-wise uncertainty, which is solved by a decision maker choosing  $\alpha$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  under the worst possible data  $\boldsymbol{\xi}_i$ ,  $i \in \{0, 1\}$ . By [2, Theorem 4.1], this problem is equivalent to the dual of the uncertain convex program where the decision maker operates under the best possible data.<sup>5</sup> Thus, (46) reduces to

$$\begin{aligned}
& \underset{\lambda_i, \boldsymbol{\xi}_i}{\text{maximize}} && \lambda_0 \\
& \text{subject to} && \lambda_i \in \mathbb{R}_+, \quad i \in \{0, 1\}, \quad \boldsymbol{\xi}_0 \in \text{cl } \Xi_0(\mathbf{x}), \quad \boldsymbol{\xi}_1 \in \Xi_1(\mathbf{x}) \\
& && \sum_{i \in \{0, 1\}} \lambda_i = 1 \\
& && \sum_{i \in \{0, 1\}} \lambda_i \boldsymbol{\xi}_i = \boldsymbol{\mu} \\
& && \sum_{i \in \{0, 1\}} \lambda_i \mathbf{d}(\boldsymbol{\xi}_i) \preceq_{\mathcal{D}} \boldsymbol{\sigma},
\end{aligned} \tag{47}$$

which is evidently equivalent to (18). □

**Lemma 10.** *If there is  $\boldsymbol{\xi} \in \Xi$  with  $\mathbf{T}(\mathbf{x})\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})$ , then (20) admits a Slater-type point  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$ , that is, a feasible point that satisfies all nonnegativity constraints strictly.*

**Proof of Lemma 10.** We set  $\lambda'_0 = \kappa$  and  $\lambda'_1 = 1 - \kappa$  for some  $\kappa \in (0, 1)$ . Next, we set  $\boldsymbol{\chi}'_0 = \lambda'_0 \boldsymbol{\xi}_0$  for any  $\boldsymbol{\xi}_0 \in \Xi$  that satisfies  $\mathbf{T}(\mathbf{x})\boldsymbol{\xi}_0 \leq \mathbf{u}(\mathbf{x})$ , as well as  $\boldsymbol{\chi}'_1 = \boldsymbol{\mu} - \boldsymbol{\chi}'_0$ . For  $\kappa$  sufficiently small, **(A)**

<sup>5</sup>As the best-case probability in (OCC) must lie in the interval  $[0, 1]$ , one can show that the equivalence of (46) and (47) holds even for unbounded uncertainty sets.

and **(D)** imply that  $\boldsymbol{\chi}'_1/\lambda'_1 \in \Xi$  (since  $\boldsymbol{\mu} \in \text{int } \Xi$ ) and  $\sum_{i \in \{0,1\}} \lambda'_i \mathbf{d}(\boldsymbol{\chi}'_i/\lambda'_i) \preceq_{\mathcal{D}} \boldsymbol{\sigma}$  (since  $\mathbf{d}(\boldsymbol{\mu}) \prec_{\mathcal{D}} \boldsymbol{\sigma}$  and  $\mathbf{d}$  is continuous). Moreover,  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$  satisfies the other constraints by construction.  $\square$

**Proof of Proposition 4.** For any feasible solution  $(\{\lambda_i\}_i, \{\boldsymbol{\xi}_i\}_i)$  to problem (18), the solution  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  with  $\boldsymbol{\chi}_i = \lambda_i \boldsymbol{\xi}_i$ ,  $i \in \{0,1\}$ , is feasible in (20) and attains the same objective value. We thus conclude that the optimal value of (20) gives an upper bound on the optimal value of (18).

We now show that the optimal value of (20) also provides a lower bound on the optimal value of (18). Lemma 10 implies that (20) admits a Slater-type point  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$ . Since (20) is convex, we can construct convex combinations of  $(\{\lambda'_i\}_i, \{\boldsymbol{\chi}'_i\}_i)$  and an optimal solution to (20) to generate a sequence of feasible solutions  $(\{\lambda_i^k\}_i, \{\boldsymbol{\chi}_i^k\}_i)_k$  that converge to the optimal solution and that satisfy  $\lambda_i^k > 0$ ,  $i \in \{0,1\}$ . The corresponding solutions  $(\{\lambda_i^k\}_i, \{\boldsymbol{\xi}_i^k\}_i)$  with  $\boldsymbol{\xi}_i^k = \boldsymbol{\chi}_i^k/\lambda_i^k$  are feasible in (18) and attain the same objective values, which concludes the proof.  $\square$

**Proof of Proposition 5.** The proof widely parallels that of Proposition 2 and is therefore omitted.  $\square$

**Proof of Proposition 6.** Let  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  be any optimal solution of (20), whose existence is guaranteed by Proposition 5. As the supremum of (20) is strictly smaller than 1, we have  $\lambda_0 < 1$ . For the sake of argument, assume that (22) is violated, that is,  $\mathbf{T}(\mathbf{x})\boldsymbol{\chi}_1 < \lambda_1 \mathbf{u}(\mathbf{x})$ . Next, define  $\lambda_0^* = 1$ ,  $\lambda_1^* = 0$ ,  $\boldsymbol{\chi}_0^* = \boldsymbol{\mu}$  and  $\boldsymbol{\chi}_1^* = \mathbf{0}$ . We show that  $(\{\lambda_i^*\}_i, \{\boldsymbol{\chi}_i^*\}_i)$  is feasible in (20). Since  $\lambda_0^* = 1$ , this solution would achieve a strictly larger objective value than  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$ , which contradicts the optimality of  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$ . Clearly,  $\lambda_0^*$  and  $\lambda_1^*$  are nonnegative and sum to 1. Moreover, we have  $\boldsymbol{\chi}_0^* + \boldsymbol{\chi}_1^* = \boldsymbol{\mu}$ , and assumption **(A)** implies that  $\sum_{i \in \{0,1\}} \lambda_i^* \mathbf{d}\left(\frac{\boldsymbol{\chi}_i^*}{\lambda_i^*}\right) = \mathbf{d}(\boldsymbol{\mu}) \preceq_{\mathcal{D}} \boldsymbol{\sigma}$ . Next, we have  $\boldsymbol{\chi}_0^*/\lambda_0^* = \boldsymbol{\mu} \in \Xi$ , which holds again by assumption **(A)**, and we have  $\boldsymbol{\chi}_1^*/\lambda_1^* = \mathbf{0}/0 \in \Xi$ , which holds because  $\Xi$  is nonempty. Finally, we have

$$\mathbf{T}(\mathbf{x})\boldsymbol{\chi}_0^* = \mathbf{T}(\mathbf{x})(\boldsymbol{\chi}_0 + \boldsymbol{\chi}_1) < (\lambda_0 + \lambda_1)\mathbf{u}(\mathbf{x}) = \lambda_0^*\mathbf{u}(\mathbf{x})$$

where the first equality holds as  $\boldsymbol{\chi}_0 + \boldsymbol{\chi}_1 = \boldsymbol{\mu}$ , while the inequality follows from the feasibility of  $(\{\lambda_i\}_i, \{\boldsymbol{\chi}_i\}_i)$  in (20) and our assumption that  $\mathbf{T}(\mathbf{x})\boldsymbol{\chi}_1 < \lambda_1 \mathbf{u}(\mathbf{x})$ .  $\square$

**Proof of Theorem 7.** We may assume that there exists  $\boldsymbol{\xi} \in \Xi$  with  $\mathbf{T}\boldsymbol{\xi} < \mathbf{u}(\mathbf{x})$ . Otherwise, both (23) and the best-case chance constraint (OCC) are infeasible and therefore trivially equivalent.

By Proposition 4, the best-case probability on the left-hand side of (OCC) is given by the optimal value of the maximization problem (20). The chance constraint (OCC) thus holds if and only if there exist  $\lambda_i \in \mathbb{R}_+$  and  $\boldsymbol{\chi}_i \in \mathbb{R}^k$ ,  $i \in \{0, 1\}$ , with

$$\begin{aligned} \lambda_0 &\geq 1 - \epsilon, & \sum_{i \in \{0,1\}} \lambda_i &= 1 \\ \mathbf{T}\boldsymbol{\chi}_0 &\leq \lambda_0 \mathbf{u}(\mathbf{x}), & \frac{\boldsymbol{\chi}_i}{\lambda_i} &\in \Xi, i \in \{0, 1\} \\ \sum_{i \in \{0,1\}} \boldsymbol{\chi}_i &= \boldsymbol{\mu}, & \lambda_0 \mathbf{d}\left(\frac{\boldsymbol{\chi}_0}{\lambda_0}\right) + \lambda_1 \mathbf{d}\left(\frac{\boldsymbol{\chi}_1}{\lambda_1}\right) &\preceq_{\mathcal{D}} \boldsymbol{\sigma}. \end{aligned}$$

As any feasible  $\lambda_0$  is strictly positive, the above constraint system is equivalent to

$$\begin{aligned} 1 &\leq \frac{1}{\lambda_0} \leq \frac{1}{1 - \epsilon}, & \mathbf{T}\frac{\boldsymbol{\chi}_0}{\lambda_0} &\leq \mathbf{u}(\mathbf{x}) \\ \frac{\boldsymbol{\chi}_0}{\lambda_0} &\in \Xi, & \frac{\boldsymbol{\chi}_1}{1 - \lambda_0} &\in \Xi \\ \sum_{i \in \{0,1\}} \frac{\boldsymbol{\chi}_i}{\lambda_0} &= \frac{\boldsymbol{\mu}}{\lambda_0}, & \mathbf{d}\left(\frac{\boldsymbol{\chi}_0}{\lambda_0}\right) + \frac{1 - \lambda_0}{\lambda_0} \mathbf{d}\left(\frac{\boldsymbol{\chi}_1}{1 - \lambda_0}\right) &\preceq_{\mathcal{D}} \frac{\boldsymbol{\sigma}}{\lambda_0}. \end{aligned}$$

The claim then follows from the variable substitution  $\lambda \leftarrow \frac{1 - \lambda_0}{\lambda_0}$ ,  $\boldsymbol{\xi}_0 \leftarrow \frac{\boldsymbol{\chi}_0}{\lambda_0}$  and  $\boldsymbol{\xi}_1 \leftarrow \frac{\boldsymbol{\chi}_1}{\lambda_0}$ .  $\square$

**Proof of Theorem 8.** We show that (25) is feasible if and only if the IP Feasibility problem has an affirmative answer. To this end, assume first that there is  $\mathbf{y} \in \{0, 1\}^n$  such that  $\mathbf{A}\mathbf{y} \leq \mathbf{b}$ . In this case,  $\mathbf{x} = 2\mathbf{y} - \mathbf{e}$  and  $\mathbb{P} \in \mathcal{P}$  defined through  $\mathbb{P}(\tilde{\boldsymbol{\xi}} = \mathbf{x}) = \mathbb{P}(\tilde{\boldsymbol{\xi}} = -\mathbf{x}) = 1/2$  are feasible in (25).

Assume now that there is no binary solution to the IP Feasibility problem. This implies that any  $\mathbf{x} \in [-1, 1]^n$  satisfying  $\mathbf{A}(\mathbf{x} + \mathbf{e})/2 \leq \mathbf{b}$  must satisfy  $x_i \in (-1 + 2\kappa, 1 - 2\kappa)$  for at least one  $i \in \{1, \dots, n\}$ . Any such choice of  $\mathbf{x}$  violates the chance constraint in (25), however, since every realization  $\boldsymbol{\xi} \in [-1, 1]^n$  of the random vector  $\tilde{\boldsymbol{\xi}}$  satisfies  $\boldsymbol{\xi}^\top \mathbf{x} < (n - 1) + (1 - 2\kappa) = n - 2\kappa$ .  $\square$