Distributionally Robust Joint Chance Constraints with Second-Order Moment Information

Steve Zymler*, Daniel Kuhn, and Berç Rustem

Department of Computing, Imperial College London
180 Queen's Gate, London SW7 2AZ, United Kingdom.

August 13, 2011

Abstract

We develop tractable semidefinite programming (SDP) based approximations for distributionally robust individual and joint chance constraints, assuming that only the first- and second-order moments as well as the support of the uncertain parameters are given. It is known that robust chance constraints can be conservatively approximated by Worst-Case Conditional Value-at-Risk (CVaR) constraints. We first prove that this approximation is exact for robust individual chance constraints with concave or (not necessarily concave) quadratic constraint functions, and we demonstrate that the Worst-Case CVaR can be computed efficiently for these classes of constraint functions. Next, we study the Worst-Case CVaR approximation for joint chance constraints. This approximation affords intuitive dual interpretations and is provably tighter than two popular benchmark approximations. The tightness depends on a set of scaling parameters, which can be tuned via a sequential convex optimization algorithm. We show that the approximation becomes essentially exact when the scaling parameters are chosen optimally and that the Worst-Case CVaR can be evaluated efficiently if the scaling parameters are kept constant. We evaluate our joint chance constraint approximation in the context of a dynamic water reservoir control problem and numerically demonstrate its superiority over the two benchmark approximations.

1 Introduction

A large class of decision problems in engineering and finance can be formulated as *chance constrained programs* of the form

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \mathbb{Q}\left(\boldsymbol{a}_i(\tilde{\boldsymbol{\xi}})^\mathsf{T}\boldsymbol{x} \le b_i(\tilde{\boldsymbol{\xi}}) \quad \forall i=1,\ldots,m\right) \ge 1-\epsilon \\ & \boldsymbol{x}\in\mathcal{X}, \end{array} \tag{1}$$

where $\boldsymbol{x}\in\mathbb{R}^n$ is the decision vector, $\mathcal{X}\subseteq\mathbb{R}^n$ is a convex closed set that can be represented by semidefinite constraints, and $\boldsymbol{c}\in\mathbb{R}^n$ is a cost vector. Without much loss of generality, we assume that $\boldsymbol{c}$ is deterministic. The *chance constraint* in (1) requires a set of m uncertainty-affected inequalities to be jointly satisfied with a probability of at least $1-\epsilon$, where $\epsilon\in(0,1)$ is a desired safety factor specified

*Corresponding author: sz02@doc.ic.ac.uk

by the modeler. The uncertain constraint coefficients $\boldsymbol{a}_i(\tilde{\boldsymbol{\xi}}) \in \mathbb{R}^n$ and $b_i(\tilde{\boldsymbol{\xi}}) \in \mathbb{R}$, $i = 1, \ldots, m$, depend affinely on a random vector $\tilde{\boldsymbol{\xi}} \in \mathbb{R}^k$, whose distribution $\mathbb{Q}$ is assumed to be known. We thus have

$$\boldsymbol{a}_i(\tilde{\boldsymbol{\xi}}) = \boldsymbol{a}_i^0 + \sum_{j=1}^{k} \boldsymbol{a}_i^j \tilde{\xi}_j \quad \text{and} \quad b_i(\tilde{\boldsymbol{\xi}}) = b_i^0 + \sum_{j=1}^{k} b_i^j \tilde{\xi}_j.$$

For ease of notation we introduce auxiliary functions $y_i^j \ : \ \mathbb{R}^n \to \mathbb{R}$, which are defined through

$$y_i^j(\boldsymbol{x}) = (\boldsymbol{a}_i^j)^\mathsf{T} \boldsymbol{x} - b_i^j, \quad i = 1, \ldots, n, \ j = 0, \ldots, k.$$

These functions enable us to rewrite the chance constraint in problem (1) as

$$\mathbb{Q}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1, \ldots, m \right) \geq 1 - \epsilon, \tag{2}$$

where $\boldsymbol{y}_i(\boldsymbol{x}) = [y_i^1(\boldsymbol{x}), \ldots, y_i^k(\boldsymbol{x})]^\mathsf{T}$ is affine in $\boldsymbol{x}$ for $i = 1, \ldots, m$. By convention, (2) is referred to as an *individual* or *joint* chance constraint if $m = 1$ or $m > 1$, respectively. Chance constrained programs were first discussed by Charnes *et al.* [8], Miller and Wagner [18] and Prékopa [23]. Although they have been studied for a long time, they have not found wide application in practice due to the following reasons.

Firstly, computing the optimal solution of a chance constrained program is notoriously difficult. In fact, even checking the feasibility of a fixed decision $\boldsymbol{x}$ requires the computation of a multi-dimensional integral, which becomes increasingly difficult as the dimension k of the random vector $\tilde{\boldsymbol{\xi}}$ increases. Furthermore, even though the inequalities in the chance constraint (2) are biaffine in $\boldsymbol{x}$ and $\tilde{\boldsymbol{\xi}}$, the feasible set of problem (1) is typically nonconvex and sometimes even disconnected.

Secondly, in order to evaluate the chance constraint (2), full and accurate information about the probability distribution $\mathbb{Q}$ of the random vector $\tilde{\boldsymbol{\xi}}$ is required. However, in many practical situations $\mathbb{Q}$ must be estimated from historical data and is therefore itself uncertain. Typically, one has only partial information about $\mathbb{Q}$, e.g. about its moments or its support. Replacing the unknown distribution $\mathbb{Q}$ in (1) by an estimate $\hat{\mathbb{Q}}$ corrupted by measurement errors may lead to over-optimistic solutions which often fail to satisfy the chance constraint under the true distribution $\mathbb{Q}$.

In a few special cases chance constraints can be reformulated as tractable convex constraints. For example, it is known that if the random vector $\tilde{\boldsymbol{\xi}}$ follows a Gaussian distribution and $\epsilon \leq 0.5$, then an individual chance constraint can be equivalently expressed as a single second-order cone constraint. In this case, the chance constrained problem becomes a tractable second-order cone program (SOCP), which can be solved in polynomial time, see Alizadeh and Goldfarb [1]. More generally, Calafiore and El Ghaoui [6] have shown that for $\epsilon \leq 0.5$ individual chance constraints can be converted to second-order cone constraints whenever the random vector $\tilde{\boldsymbol{\xi}}$ is governed by a radial distribution. Tractability results for joint chance constraints are even more scarce. In a seminal paper, Prékopa [23] has shown that joint chance constraints are convex when only the right-hand side coefficients $b_i(\tilde{\boldsymbol{\xi}})$ are uncertain and follow a log-concave distribution. However, under generic distributions, chance constrained programs

are computationally intractable. Indeed, Shapiro and Nemirovski [20] point out that computing the probability of a weighted sum of uniformly distributed variables being nonpositive is already $\mathcal{NP}$-hard.

Recently, Calafiore and Campi [5] as well as Luedtke and Ahmed [17] have proposed to replace the chance constraint (2) by a pointwise constraint that must hold at a finite number of sample points drawn randomly from the distribution $\mathbb{Q}$. A similar approach was suggested by Erdoğan and Iyengar [12]. The advantage of this Monte Carlo approach is that no structural assumptions about $\mathbb{Q}$ are needed and that the resulting approximate problem is convex. Calafiore and Campi [5] showed that one requires $\mathcal{O}(n/\epsilon)$ samples to guarantee that a solution of the approximate problem is feasible in the original chance constrained program. However, this implies that it may be computationally prohibitive to solve large problems or to solve problems for which a small violation probability ϵ is required.

A natural way to *immunize* the chance constraint (2) against uncertainty in the probability distribution is to adopt a distributionally robust approach. To this end, let $\mathcal{P}$ denote the set of all probability distributions on $\mathbb{R}^k$ that are consistent with the known properties of $\mathbb{Q}$, such as its first and second moments and/or its support. Consider now the following *ambiguous* or *distributionally robust chance constraint*.

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1,\ldots,m\right) \geq 1-\epsilon \tag{3}$$

It is easily verified that whenever $\boldsymbol{x}$ satisfies (3) and $\mathbb{Q}\in\mathcal{P}$, then $\boldsymbol{x}$ also satisfies the chance constraint (2) under the true probability distribution $\mathbb{Q}$. Replacing the chance constraint (2) with its distributionally robust counterpart (3) yields the following distributionally robust chance constrained program

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \inf\limits_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1,\ldots,m\right) \geq 1-\epsilon \\ & \boldsymbol{x}\in\mathcal{X}, \end{array} \tag{4}$$

which constitutes a conservative approximation for problem (1) in the sense that it has the same objective function but a smaller feasible set.

A common method to simplify the distributionally robust joint chance constraint (3), which looks even less tractable than (2), is to decompose it into m individual chance constraints by using Bonferroni's inequality. Indeed, by ensuring that the total sum of violation probabilities of the individual chance constraints does not exceed ϵ, the feasibility of the joint chance constraint is guaranteed. Nemirovski and Shapiro [20] propose to divide the overall violation probability ϵ equally among the m individual chance constraints. However, the Bonferroni inequality is not necessarily tight, and the corresponding decomposition could therefore be over-conservative. In fact, for positively correlated constraint functions, the quality of the approximation is known to decrease as m increases [9]. Consequently, the Bonferroni method may result in a poor approximation for problems with joint chance constraints that involve many inequalities.

A recent attempt to improve on the Bonferroni approximation is due to Chen *et al.* [9]. They first

elaborate a convex conservative approximation for a joint chance constraint in terms of a Worst-Case Conditional Value-at-Risk (CVaR) constraint. Then, they rely on a classical inequality in order statistics to determine a tractable conservative approximation for the Worst-Case CVaR and show that the resulting approximation for the joint chance constraint necessarily outperforms the Bonferroni approximation. An attractive feature of this method is that the arising approximate constraints are second-order conic representable. However, the employed probabilistic inequality is not necessarily tight, which may again render the approximation over-conservative.

The principal aim of this paper is to develop new tools and models for approximating robust individual and joint chance constraints under the assumption that only the first- and second-order moments as well as the support of the random vector $\tilde{\boldsymbol{\xi}}$ are known. We embrace the modern approach to approximate robust chance constraints by Worst-Case CVaR constraints, but in contrast to the state-of-the-art methods described above, we find exact semidefinite programming (SDP) reformulations of the Worst-Case CVaR which do not rely on potentially loose probabilistic inequalities. These reformulations are facilitated by the theory of moment problems and by conic duality arguments. We prove that the CVaR approximation is in fact exact for individual chance constraints whose constraint functions are either concave or (possibly nonconcave) quadratic in $\boldsymbol{\xi}$ and for joint chance constraints whose constraint functions depend linearly on $\boldsymbol{\xi}$. We also demonstrate that robust individual chance constraints have manifestly tractable SDP representations in most cases in which the CVaR approximation is exact.

The main contributions of this paper can be summarized as follows:

(1) In Section 2 we review and extend existing approximations for distributionally robust *individual* chance constraints and prove that a robust individual chance constraint is equivalent to a tractable Worst-Case CVaR constraint if the underlying constraint function is either concave or (possibly nonconcave) quadratic in $\boldsymbol{\xi}$. We also demonstrate that this equivalence can fail to hold even if the constraint function is convex and piecewise linear in $\boldsymbol{\xi}$.

(2) In Section 3 we develop a new tractable CVaR approximation for robust *joint* chance constraints and prove that this approximation consistently outperforms the state-of-the-art methods described above. We show that the approximation quality is controlled by a set of scaling parameters and that the CVaR approximation becomes essentially exact if the scaling parameters are chosen optimally. We also present an intuitive dual interpretation for the CVaR approximation in this case.

(3) In Section 4 we analyze the performance of the new joint chance constraint approximation when applied to a dynamic water reservoir control problem.

Notation. We use lower-case bold face letters to denote vectors and upper-case bold face letters to denote matrices. The space of symmetric matrices of dimension n is denoted by $\mathbb{S}^n$. For any two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n$, we let $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{XY})$ be the trace scalar product, while the relation $\mathbf{X} \succcurlyeq \mathbf{Y}$ ($\mathbf{X} \succ \mathbf{Y}$) implies that $\mathbf{X} - \mathbf{Y}$ is positive semidefinite (positive definite). Random variables are always represented by symbols with tildes, while their realizations are denoted by the same symbols without tildes. For

$x \in \mathbb{R}$, we define $x^+ = \max\{x, 0\}$.

2 Distributionally Robust Individual Chance Constraints

It is known that robust individual chance constraints can be conservatively approximated by Worst-Case CVaR constraints. In this section, we first show how the theory of moment problems can be used to reformulate these Worst-Case CVaR constraints in terms of tractable semidefinite constraints. Subsequently, we prove that the Worst-Case CVaR constraints are in fact equivalent to the underlying robust chance constraints for a large class of constraint functions.

Distributional Assumptions. In the remainder of this paper we let $\boldsymbol{\mu} \in \mathbb{R}^k$ be the mean vector and $\boldsymbol{\Sigma} \in \mathbb{S}^k$ be the covariance matrix of the random vector $\tilde{\boldsymbol{\xi}}$ under the true distribution $\mathbb{Q}$. Thus, we implicitly assume that $\mathbb{Q}$ has finite second-order moments. Without loss of generality we also assume that $\boldsymbol{\Sigma} \succ \mathbf{0}$. Furthermore, we let $\mathcal{P}$ denote the set of all probability distributions on $\mathbb{R}^k$ that have the same first- and second-order moments as $\mathbb{Q}$. For notational simplicity, we let

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T} & \boldsymbol{\mu} \\ \boldsymbol{\mu}^\mathsf{T} & 1 \end{bmatrix}$$

be the second-order moment matrix of $\tilde{\boldsymbol{\xi}}$.

2.1 The Worst-Case CVaR Approximation

For $m = 1$, (3) reduces to a distributionally robust *individual* chance constraint

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \leq 0 \right) \geq 1 - \epsilon, \tag{5}$$

whose feasible set is denoted by

$$\mathcal{X}^{\mathrm{ICC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \;:\; \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \leq 0 \right) \geq 1 - \epsilon \right\}.$$

In the remainder of this section we will demonstrate that $\mathcal{X}^{\mathrm{ICC}}$ has a manifestly tractable representation in terms of Linear Matrix Inequalities (LMIs). To this end, we first recall the definition of CVaR due to Rockafellar and Uryasev [24]. For a given measurable loss function $L : \mathbb{R}^k \to \mathbb{R}$, probability distribution $\mathbb{P}$ on $\mathbb{R}^k$, and tolerance $\epsilon \in (0, 1)$, the CVaR at level ϵ with respect to $\mathbb{P}$ is defined as

$$\mathbb{P}\text{-CVaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) = \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_\mathbb{P}\left((L(\tilde{\boldsymbol{\xi}}) - \beta)^+ \right) \right\}, \tag{6}$$

where $\mathbb{E}_\mathbb{P}(\cdot)$ denotes expectation with respect to $\mathbb{P}$. CVaR essentially evaluates the conditional expectation of loss above the $(1 - \epsilon)$-quantile of the loss distribution. It can be shown that CVaR represents a

convex functional of the random variable $L(\tilde{\boldsymbol{\xi}})$.

CVaR can be used to construct convex approximations for chance constraints. Indeed, it is well known that

$$\mathbb{P}\left(L(\tilde{\boldsymbol{\xi}}) \leq \mathbb{P}\text{-CVaR}_\epsilon(L(\tilde{\boldsymbol{\xi}}))\right) \geq 1-\epsilon$$

for any measurable loss function L, see, e.g., Ben-Tal *et al.* [3, §4.3.3]. Thus, $\mathbb{P}\text{-CVaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \leq 0$ is sufficient to imply $\mathbb{P}(L(\tilde{\boldsymbol{\xi}}) \leq 0) \geq 1-\epsilon$. As this implication holds for any probability distribution and loss function, we conclude that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}}\right) \leq 0 \implies \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0\right) \geq 1-\epsilon. \tag{7}$$

Thus, the *Worst-Case CVaR* constraint on the left hand side constitutes a conservative approximation for the distributionally robust chance constraint on the right hand side of (7). The above discussion motivates us to define the feasible set

$$\mathcal{Z}^{\text{ICC}} = \left\{\boldsymbol{x}\in\mathbb{R}^n \;:\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}}\right) \leq 0\right\}, \tag{8}$$

and the implication (7) gives rise to the following elementary result.

Proposition 2.1 *The feasible set $\mathcal{Z}^{\text{ICC}}$ constitutes a conservative approximation for $\mathcal{X}^{\text{ICC}}$, that is, $\mathcal{Z}^{\text{ICC}} \subseteq \mathcal{X}^{\text{ICC}}$.*

We will now show that $\mathcal{Z}^{\text{ICC}}$ has a tractable representation in terms of LMIs.

Theorem 2.1 *The feasible set $\mathcal{Z}^{\text{ICC}}$ can be written as*

$$\mathcal{Z}^{\text{ICC}} = \left\{ \boldsymbol{x}\in\mathbb{R}^n : \begin{array}{l} \exists(\beta,\mathbf{M})\in\mathbb{R}\times\mathbb{S}^{k+1}, \\ \mathbf{M}\succcurlyeq\mathbf{0}, \quad \beta + \frac{1}{\epsilon}\langle\boldsymbol{\Omega},\mathbf{M}\rangle \leq 0, \\ \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x})-\beta \end{bmatrix} \succcurlyeq \mathbf{0} \end{array} \right\}.$$

Proof: By using (6), the Worst-Case CVaR in (8) can be expressed as

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}}\right) &= \sup_{\mathbb{P}\in\mathcal{P}} \inf_{\beta\in\mathbb{R}} \left\{\beta + \frac{1}{\epsilon}\mathbb{E}_\mathbb{P}\left((y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} - \beta)^+\right)\right\} \\ &= \inf_{\beta\in\mathbb{R}} \left\{\beta + \frac{1}{\epsilon}\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}\left((y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} - \beta)^+\right)\right\}, \end{aligned} \tag{9}$$

where the interchange of the maximization and minimization operations is justified by a stochastic saddle point theorem due to Shapiro and Kleywegt [26], see also Delage and Ye [11] or Natarajan *et al.* [19]. We now show that the Worst-Case CVaR (9) of some fixed decision $\boldsymbol{x}\in\mathbb{R}^n$ can be computed by solving a

tractable SDP. To this end, we first derive an SDP reformulation of the worst-case expectation problem

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left((y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} - \beta)^+\right),$$

which can be identified as the subordinate maximization problem in (9). Lemma A.1 in the Appendix enables us to reformulate this worst-case expectation problem as

$$\begin{array}{ll} \inf\limits_{\mathbf{M}\in\mathbb{S}^{k+1}} & \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]\mathbf{M}\left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\xi} - \beta \quad \forall \boldsymbol{\xi}\in\mathbb{R}^k. \end{array} \tag{10}$$

Note that the semi-infinite constraint in (10) can be written as the following LMI.

$$\begin{bmatrix}\boldsymbol{\xi}\\ 1\end{bmatrix}^\mathsf{T}\left(\mathbf{M} - \begin{bmatrix}\mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})\\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x})-\beta\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{\xi}\\ 1\end{bmatrix} \geq 0\ \forall \boldsymbol{\xi}\in\mathbb{R}^k \iff \mathbf{M} - \begin{bmatrix}\mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})\\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x})-\beta\end{bmatrix} \succcurlyeq \mathbf{0}$$

This in turn allows us to reformulate the worst-case expectation problem as

$$\begin{array}{ll} \inf\limits_{\mathbf{M}\in\mathbb{S}^{k+1}} & \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix}\mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})\\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x})-\beta\end{bmatrix} \succcurlyeq \mathbf{0}. \end{array} \tag{11}$$

By replacing the subordinate worst-case expectation problem in (9) by (11), we obtain

$$\begin{array}{rcll} \sup\limits_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}}\right) & = & \inf & \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \\ & & \text{s.t.} & \mathbf{M}\in\mathbb{S}^{k+1}, \quad \beta\in\mathbb{R} \\ & & & \mathbf{M}\succcurlyeq\mathbf{0}, \quad \mathbf{M} - \begin{bmatrix}\mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})\\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x})-\beta\end{bmatrix} \succcurlyeq \mathbf{0}, \end{array} \tag{12}$$

and thus the claim follows. ∎

2.2 Exactness of the Worst-Case CVaR Approximation

So far we have shown that the feasible set $\mathcal{Z}^{\text{ICC}}$ defined in terms of a Worst-Case CVaR constraint constitutes a tractable conservative approximation for $\mathcal{X}^{\text{ICC}}$. We now demonstrate that this approximation is in fact *exact*, that is, we show that the implication (7) is in fact an equivalence. We first recall the nonlinear Farkas Lemma as well as the $\mathcal{S}$-lemma, which are crucial ingredients for the proof of this result. We refer to Pólik and Terlaky [22] for a derivation and an in-depth survey of the $\mathcal{S}$-lemma as well as a review of the Farkas Lemma.

Lemma 2.1 (Farkas Lemma) *Let $f_0,\ldots,f_p : \mathbb{R}^k \to \mathbb{R}$ be convex functions, and assume that there exists a strictly feasible point $\bar{\boldsymbol{\xi}}$ with $f_i(\bar{\boldsymbol{\xi}}) < 0$, $i = 1,\ldots,p$. Then, $f_0(\boldsymbol{\xi}) \geq 0$ for all $\boldsymbol{\xi}$ with $f_i(\boldsymbol{\xi}) \leq 0$,*

$i = 1, \ldots, p$, if and only if there exist constants $\tau_i \geq 0$ such that

$$f_0(\boldsymbol{\xi}) + \sum_{i=1}^{p} \tau_i f_i(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k.$$

Lemma 2.2 ($\mathcal{S}$-lemma) *Let $f_i(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top \mathbf{A}_i \boldsymbol{\xi}$ with $\mathbf{A}_i \in \mathbb{S}^n$ be quadratic functions of $\boldsymbol{\xi} \in \mathbb{R}^n$ for $i = 0, \ldots, p$. Then, $f_0(\boldsymbol{\xi}) \geq 0$ for all $\boldsymbol{\xi}$ with $f_i(\boldsymbol{\xi}) \leq 0$, $i = 1, \ldots, p$, if there exist constants $\tau_i \geq 0$ such that*

$$\mathbf{A}_0 + \sum_{i=1}^{p} \tau_i \mathbf{A}_i \succcurlyeq \mathbf{0}.$$

For $p = 1$, the converse implication holds if there exists a strictly feasible point $\bar{\boldsymbol{\xi}}$ with $f_1(\bar{\boldsymbol{\xi}}) < 0$.

Theorem 2.2 *Let $L : \mathbb{R}^k \to \mathbb{R}$ be a continuous loss function that is either*

(i) concave in $\boldsymbol{\xi}$, or

(ii) (possibly nonconcave) quadratic in $\boldsymbol{\xi}$.

Then, the following equivalence holds.

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon \left(L(\tilde{\boldsymbol{\xi}}) \right) \leq 0 \iff \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq 1 - \epsilon \tag{13}$$

Proof: Consider the *Worst-Case Value-at-Risk* of the loss function L, which is defined as

$$\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \;=\; \inf_{\gamma \in \mathbb{R}} \left\{ \gamma \;:\; \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq \gamma \right) \geq 1 - \epsilon \right\}. \tag{14}$$

By definition, the WC-VaR is indeed equal to the $(1-\epsilon)$-quantile of $L(\tilde{\boldsymbol{\xi}})$ evaluated under some worst-case distribution in $\mathcal{P}$. We first show that the following equivalence holds.

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq 1 - \epsilon \iff \text{WC-VaR}_\epsilon \left(L(\tilde{\boldsymbol{\xi}}) \right) \leq 0 \tag{15}$$

Indeed, if the left hand side of (15) is satisfied, then $\gamma = 0$ is feasible in (14), which implies that $\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \leq 0$. To see that the converse implication holds as well, we note that for any fixed $\mathbb{P} \in \mathcal{P}$, the mapping $\gamma \mapsto \mathbb{P}(L(\tilde{\boldsymbol{\xi}}) \leq \gamma)$ is upper semi-continuous, see [21]. Thus, the related mapping $\gamma \mapsto \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(L(\tilde{\boldsymbol{\xi}}) \leq \gamma)$ is also upper semi-continuous. If $\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \leq 0$, there exists a sequence $\{\gamma_n\}_{n \in \mathbb{N}}$ that converges to zero and is feasible in (14), which implies

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq \limsup_{n \to \infty} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq \gamma_n \right) \geq 1 - \epsilon.$$

Thus, (15) follows.

To prove the postulated equivalence (13), it is now sufficient to show that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(L(\tilde{\boldsymbol{\xi}})\right) = \text{WC-VaR}_\epsilon\left(L(\tilde{\boldsymbol{\xi}})\right).$$

Note that (14) can be rewritten as

$$\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \;=\; \inf_{\gamma\in\mathbb{R}} \left\{ \gamma \;:\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(L(\tilde{\boldsymbol{\xi}}) > \gamma\right) \leq \epsilon \right\}. \tag{16}$$

We proceed by simplifying the subordinate worst-case probability problem $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(L(\tilde{\boldsymbol{\xi}}) > \gamma)$, which, by Lemma A.2 in the Appendix, can be expressed as

$$\inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \left\{ \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \;:\; \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\mathsf{T}\; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\; 1\right]^\mathsf{T} \geq 1 \quad \forall \boldsymbol{\xi} \;:\; \gamma - L(\boldsymbol{\xi}) < 0 \right\}. \tag{17}$$

We will now argue that for all but one value of γ problem (17) is equivalent to

$$\begin{array}{ll} \inf & \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \tau \in \mathbb{R}, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\ & \left[\boldsymbol{\xi}^\mathsf{T}\; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\; 1\right]^\mathsf{T} - 1 + \tau\left(\gamma - L(\boldsymbol{\xi})\right) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{array} \tag{18}$$

For ease of exposition, we define $h = \inf_{\boldsymbol{\xi}\in\mathbb{R}^k} \gamma - L(\boldsymbol{\xi})$. The equivalence of (17) and (18) is proved case by case. Assume first that $h < 0$. Then, the strict inequality in the parameter range of the semi-infinite constraint in (17) can be replaced by a weak inequality without affecting its optimal value. The equivalence then follows from the Farkas Lemma (when $L(\boldsymbol{\xi})$ is concave in $\boldsymbol{\xi}$) or from the $\mathcal{S}$-lemma (when $L(\boldsymbol{\xi})$ is quadratic in $\boldsymbol{\xi}$). Assume next that $h > 0$. Then, the semi-infinite constraint in (17) becomes redundant and, since $\boldsymbol{\Omega} \succ \mathbf{0}$, the optimal solution of (17) is given by $\mathbf{M} = \mathbf{0}$ with a corresponding optimal value of 0. The optimal value of problem (18) is also equal to 0. Indeed, by choosing $\tau = 1/h$, the semi-infinite constraint in (18) is satisfied for any $\mathbf{M} \succcurlyeq \mathbf{0}$. Finally, note that (17) and (18) may be different for $h = 0$.

Since (17) and (18) are equivalent for all but one value of γ and since their optimal values are nonincreasing in γ, we can express $\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}}))$ in (16) as

$$\begin{array}{rcll} \text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) & = & \inf & \gamma \\ & & \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\ & & & \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \leq \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\ & & & \left[\boldsymbol{\xi}^\mathsf{T}\; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\; 1\right]^\mathsf{T} - 1 + \tau\left(\gamma - L(\boldsymbol{\xi})\right) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{array} \tag{19}$$

It can easily be shown that $\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \geq 1$ for any feasible solution of (19) with vanishing τ-component. However, since $\epsilon < 1$, this is in conflict with the constraint $\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \leq \epsilon$. We thus conclude that no feasible point can have a vanishing τ-component. This allows us to divide the semi-infinite constraint in

problem (19) by τ. Subsequently we perform variable substitutions in which we replace τ by $1/\tau$ and $\mathbf{M}$ by $\mathbf{M}/\tau$. This yields the following reformulation of problem (19).

$$\begin{array}{rcll} \text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) & = & \inf & \gamma \\ & & \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\ & & & \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \tau, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\ & & & \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} - \tau + \gamma - L(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \end{array}$$

Note that, since $\boldsymbol{\Omega} \succ \mathbf{0}$ and $\mathbf{M} \succcurlyeq \mathbf{0}$, we have $\frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \geq 0$. This allows us to remove the redundant nonnegativity constraint on τ. We now introduce a new decision variable $\beta = \gamma - \tau$, which allows us to eliminate γ.

$$\begin{array}{rcll} \text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) & = & \inf & \beta + \tau \\ & & \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \tau \in \mathbb{R}, \quad \beta \in \mathbb{R} \\ & & & \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \tau, \quad \mathbf{M} \succcurlyeq \mathbf{0} \\ & & & \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} + \beta - L(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \end{array}$$

Note that at optimality $\tau = \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle$, which finally allows us to express $\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}}))$ as

$$\begin{array}{rcll} \text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) & = & \inf & \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ & & \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R}, \quad \mathbf{M} \succcurlyeq \mathbf{0} \\ & & & \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} + \beta - L(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{array} \tag{20}$$

Recall now that by Lemma A.1 we have

$$\begin{array}{rcll} \sup\limits_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(L(\tilde{\boldsymbol{\xi}})\right) & = & \inf\limits_{\beta\in\mathbb{R}} & \left\{\beta + \dfrac{1}{\epsilon}\sup\limits_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}\left((L(\tilde{\boldsymbol{\xi}}) - \beta)^+\right)\right\} \\ & = & \inf & \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ & & \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R}, \quad \mathbf{M} \succcurlyeq \mathbf{0} \\ & & & \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} + \beta - L(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k, \end{array}$$

which is clearly equivalent to (20). This observation completes the proof. ∎

Corollary 2.1 *The following equivalence holds*

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}}\right) \leq 0 \iff \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0\right) \geq 1 - \epsilon,$$

which implies that $\mathcal{Z}^{\text{ICC}} = \mathcal{X}^{\text{ICC}}$.

Proof: The claim follows immediately from Theorem 2.2 by observing that $L(\boldsymbol{\xi}) = y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\xi}$ is linear (and therefore concave) in $\boldsymbol{\xi}$. ∎

In the following example we demonstrate that the equivalence (13) can fail to hold even if the loss function L is convex and piecewise linear in $\boldsymbol{\xi}$.

Example 2.1 *Let $\tilde{\xi}$ be a scalar random variable with mean $\mu = 0$ and standard deviation $\sigma = 1$. Moreover, let $\mathcal{P}$ be the set of all probability distributions on $\mathbb{R}$ consistent with the given mean and standard deviation. Consider now the loss function $L(\xi) = \max\{\xi - 1, 4\xi - 4\}$, and note that L is strictly increasing and convex in ξ. In particular, L is neither concave nor quadratic and thus falls outside the scope of Theorem 2.2. We now show that for this particular L the Worst-Case CVaR constraint $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_{\frac{1}{2}}(L(\tilde{\xi})) \leq 0$ is violated even though the distributionally robust individual chance constraint $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(L(\tilde{\xi}) \leq 0) \geq 1/2$ is satisfied. To this end, we note that the Chebychev inequality $\mathbb{P}(\tilde{\xi} - \mu \geq \kappa\sigma) \leq 1/(1+\kappa^2)$ for $\kappa = 1$ implies*

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\tilde{\xi} \geq 1\right) \leq \frac{1}{2} &\iff \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(L(\tilde{\xi}) \geq L(1) = 0\right) \leq \frac{1}{2} \\
&\implies \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(L(\tilde{\xi}) > 0\right) \leq \frac{1}{2} \\
&\iff \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(L(\tilde{\xi}) \leq 0\right) \geq \frac{1}{2},
\end{aligned}$$

where the first equivalence follows from the monotonicity of L. Assume now that the true distribution $\mathbb{Q}$ of $\tilde{\xi}$ is discrete and defined through $\mathbb{Q}(\tilde{\xi} = -2) = 1/8$, $\mathbb{Q}(\tilde{\xi} = 0) = 3/4$, and $\mathbb{Q}(\tilde{\xi} = 2) = 1/8$. It is easy to verify that $\mathbb{Q} \in \mathcal{P}$ and that $\mathbb{Q}\text{-CVaR}_{\frac{1}{2}}(L(\tilde{\xi})) = 0.25$. Thus, $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_{\frac{1}{2}}(L(\tilde{\xi})) \geq 0.25 > 0$. We therefore conclude that the Worst-Case CVaR constraint is not equivalent to the robust chance constraint.

2.3 Tractability of the Worst-Case CVaR Approximation

We have already seen that Worst-Case CVaR constraints are equivalent to distributionally robust chance constraints when the loss function is continuous and either concave or quadratic in $\boldsymbol{\xi}$. We now prove that the Worst-Case CVaR can also be computed efficiently for these classes of loss functions.

Theorem 2.3 *Assume that $L : \mathbb{R}^k \to \mathbb{R}$ is either*

(i) concave piecewise affine in $\boldsymbol{\xi}$ with a finite number of pieces or

(ii) (possibly nonconcave) quadratic in $\boldsymbol{\xi}$.

Then, $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon(L(\tilde{\boldsymbol{\xi}}))$ can be computed efficiently as the optimal value of a tractable SDP.

Proof: Assume that (i) holds and that $L(\tilde{\boldsymbol{\xi}}) = \min_{i=1,\ldots,l}\{a_i + \boldsymbol{b}_i^T\tilde{\boldsymbol{\xi}}\}$ for some $a_i \in \mathbb{R}$ and $\boldsymbol{b}_i \in \mathbb{R}^k$, $i = 1, \ldots, l$. Then, the Worst-Case CVaR is representable as

$$\inf_{\beta\in\mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\left[\min_{i=1,\ldots,l} \{a_i + \boldsymbol{b}_i^T\tilde{\boldsymbol{\xi}}\} - \beta \right]^+ \right) \right\}. \tag{21}$$

By Lemma A.1, the subordinate worst-case expectation problem in (21) can be rewritten as

$$\begin{array}{ll} \inf\limits_{\mathbf{M}\in\mathbb{S}^{k+1}} & \langle \mathbf{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \ \geq \ \min\limits_{i=1,\ldots,l} \{a_i + \boldsymbol{b}_i^T \boldsymbol{\xi}\} - \beta \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{array} \tag{22}$$

Noting that

$$\min_{i=1,\ldots,l} \{a_i + \boldsymbol{b}_i^T \boldsymbol{\xi}\} = \min_{\boldsymbol{\lambda}\in\Delta} \sum_{i=1}^{l} \lambda_i (a_i + \boldsymbol{b}_i^T \boldsymbol{\xi}),$$

where $\Delta = \{\boldsymbol{\lambda} \in \mathbb{R}^l : \sum_{i=1}^l \lambda_i = 1,\ \boldsymbol{\lambda} \geq \mathbf{0}\}$ denotes the probability simplex in $\mathbb{R}^l$, we can use techniques developed in [4, Theorem 2.1] to reexpress the semi-infinite constraint in (22) as

$$\begin{array}{lll} & \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} - \min\limits_{\boldsymbol{\lambda}\in\Delta} \sum\limits_{i=1}^{l} \lambda_i (a_i + \boldsymbol{b}_i^T \boldsymbol{\xi}) + \beta \ \geq \ 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \\ \iff & \min\limits_{\boldsymbol{\xi}\in\mathbb{R}^k} \max\limits_{\boldsymbol{\lambda}\in\Delta} \left\{ \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} - \sum\limits_{i=1}^{l} \lambda_i (a_i + \boldsymbol{b}_i^T \boldsymbol{\xi}) + \beta \right\} \ \geq \ 0 \\ \iff & \max\limits_{\boldsymbol{\lambda}\in\Delta} \min\limits_{\boldsymbol{\xi}\in\mathbb{R}^k} \left\{ \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} - \sum\limits_{i=1}^{l} \lambda_i (a_i + \boldsymbol{b}_i^T \boldsymbol{\xi}) + \beta \right\} \ \geq \ 0 \\ \iff & \min\limits_{\boldsymbol{\xi}\in\mathbb{R}^k} \left\{ \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} - \sum\limits_{i=1}^{l} \lambda_i (a_i + \boldsymbol{b}_i^T \boldsymbol{\xi}) + \beta \right\} \ \geq \ 0, \quad \boldsymbol{\lambda} \in \Delta \\ \iff & \mathbf{M} - \begin{bmatrix} \mathbf{0} & \sum_{i=1}^l \frac{\lambda_i}{2} \boldsymbol{b}_i \\ \sum_{i=1}^l \frac{\lambda_i}{2} \boldsymbol{b}_i^\mathsf{T} & \sum_{i=1}^l \lambda_i a_i - \beta \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \boldsymbol{\lambda} \in \Delta. \end{array}$$

The second equivalence in the above expression follows from the classical saddle point theorem. Thus, the Worst-Case CVaR (21) can be rewritten as the optimal value of the following tractable SDP.

$$\begin{array}{ll} \inf & \beta + \frac{1}{\epsilon} \langle \mathbf{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \beta \in \mathbb{R}, \quad \mathbf{M} \in \mathbb{S}^{k+1}, \quad \boldsymbol{\lambda} \in \mathbb{R}^l \\ & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \sum_{i=1}^l \frac{\lambda_i}{2} \boldsymbol{b}_i \\ \sum_{i=1}^l \frac{\lambda_i}{2} \boldsymbol{b}_i^\mathsf{T} & \sum_{i=1}^l \lambda_i a_i - \beta \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \boldsymbol{\lambda} \in \Delta \end{array} \tag{23}$$

Assume now that (ii) holds and that $L(\boldsymbol{\xi}) = \boldsymbol{\xi}^\mathsf{T} \mathbf{Q} \boldsymbol{\xi} + \boldsymbol{q}^\mathsf{T} \boldsymbol{\xi} + q^0$ for some $\mathbf{Q} \in \mathbb{S}^k,\ \boldsymbol{q} \in \mathbb{R}^k$, and $q^0 \in \mathbb{R}$. In this case we have

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \ = \ \inf_{\beta\in\mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P} \left(\left[\tilde{\boldsymbol{\xi}}^\mathsf{T} \mathbf{Q} \tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\mathsf{T} \boldsymbol{q} + q^0 - \beta \right]^+ \right) \right\}. \tag{24}$$

As usual, we first find an SDP reformulation of the subordinate worst-case expectation problem in (24).

By Lemma A.1, this problem can be rewritten as

$$\begin{array}{ll} \inf\limits_{\mathbf{M}\in\mathbb{S}^{k+1}} & \langle \mathbf{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \ \geq\ \boldsymbol{\xi}^\mathsf{T}\mathbf{Q}\boldsymbol{\xi} + \boldsymbol{\xi}^\mathsf{T}\boldsymbol{q} + q^0 - \beta \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{array} \tag{25}$$

Note that the semi-infinite constraint in (25) is equivalent to

$$\begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix}^\mathsf{T} \left(\mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 - \beta \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix} \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \quad \Longleftrightarrow \quad \mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0},$$

which enables us to rewrite the Worst-Case CVaR (24) as the optimal value of

$$\begin{array}{ll} \inf & \beta + \frac{1}{\epsilon}\langle \mathbf{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R} \\ & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0}, \end{array}$$

which is indeed a tractable SDP. ∎

Remark If the loss function is concave but not piecewise affine, the Worst-Case CVaR can sometimes still be evaluated efficiently, though not by solving an explicit SDP. Indeed, the Worst-Case CVaR can be computed in polynomial time with an ellipsoid method if $L(\boldsymbol{\xi})$ is concave and if, for any $\boldsymbol{\xi} \in \mathbb{R}^k$, one can evaluate both $L(\boldsymbol{\xi})$ as well as a super-gradient $\nabla_{\boldsymbol{\xi}} L(\boldsymbol{\xi})$ in polynomial time. This is an immediate consequence of a result on the computation of worst-case expectations by Delage and Ye [11, Proposition 2].

3 Distributionally Robust Joint Chance Constraints

We define the feasible set $\mathcal{X}^{\mathrm{JCC}}$ of the distributionally robust *joint* chance constraint (3) as

$$\mathcal{X}^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1, \ldots, m \right) \geq 1 - \epsilon \right\}.$$

The aim of this section is to investigate the structure of $\mathcal{X}^{\mathrm{JCC}}$ and to elaborate tractable conservative approximations. We first review two existing approximations and discuss their benefits and shortcomings.

3.1 The Bonferroni Approximation

A popular approximation for $\mathcal{X}^{\mathrm{JCC}}$ is based on Bonferroni's inequality. Note that the robust joint chance constraint (3) is equivalent to

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\bigcap_{i=1}^m \left\{y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0\right\}\right) \geq 1-\epsilon \iff \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\bigcup_{i=1}^m \left\{y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} > 0\right\}\right) \leq \epsilon.$$

Furthermore, Bonferroni's inequality implies that

$$\mathbb{P}\left(\bigcup_{i=1}^m \left\{y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} > 0\right\}\right) \leq \sum_{i=1}^m \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} > 0\right) \quad \forall \mathbb{P}\in\mathcal{P}.$$

For any vector of safety factors $\boldsymbol{\epsilon}\in\mathcal{E} = \{\boldsymbol{\epsilon}\in\mathbb{R}_+^m \,:\, \sum_{i=1}^m \epsilon_i \leq \epsilon\}$, the system of distributionally robust *individual* chance constraints

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0\right) \geq 1-\epsilon_i \quad \forall i = 1,\ldots,m \tag{26}$$

represents a conservative approximation for the distributionally robust *joint* chance constraint (3). By Theorem 2.1, we can reformulate each of the individual chance constraints in (26) in terms of tractable LMIs. In fact, we can further reduce these LMIs to SOCP constraints, but this further simplification is irrelevant for our purposes. Thus, for any $\boldsymbol{\epsilon}\in\mathcal{E}$, the assertion that $\boldsymbol{x}\in\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon})$, where

$$\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon}) = \left\{ \boldsymbol{x}\in\mathbb{R}^n : \begin{array}{ll} \exists(\beta_i,\mathbf{M}_i)\in\mathbb{R}\times\mathbb{S}^{k+1} & \forall i=1,\ldots,m, \\ \mathbf{M}_i \succcurlyeq \mathbf{0}, \quad \beta_i + \frac{1}{\epsilon_i}\langle\boldsymbol{\Omega},\mathbf{M}_i\rangle \leq 0 & \forall i=1,\ldots,m, \\ \mathbf{M}_i - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & y_i^0(\boldsymbol{x})-\beta_i \end{bmatrix} \succcurlyeq \mathbf{0} & \forall i=1,\ldots,m \end{array} \right\},$$

is a sufficient condition to guarantee that $\boldsymbol{x}$ satisfies the original distributionally robust joint chance constraint (3). The above arguments culminate in the following result.

Theorem 3.1 (Bonferroni Approximation) *For any $\boldsymbol{\epsilon}\in\mathcal{E}$ we have $\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon}) \subseteq \mathcal{X}^{\mathrm{JCC}}$.*

A major shortcoming of the Bonferroni approximation is that the approximation quality depends critically on the choice of $\boldsymbol{\epsilon}\in\mathcal{E}$. Unfortunately, the problem of finding the best $\boldsymbol{\epsilon}\in\mathcal{E}$ for a generic chance constrained problem of type (4) is nonconvex and believed to be intractable [20]. As a result, in most applications of Bonferroni's inequality the "risk budget" ϵ is equally divided among the m individual chance constraints in (26) by setting $\epsilon_i = \epsilon/m$ for $i = 1,\ldots,m$. This approach was first advocated by Nemirovski and Shapiro [20].

The Bonferroni approximation can be overly conservative even if $\boldsymbol{\epsilon}\in\mathcal{E}$ is chosen optimally. The following example, which is adapted from Chen *et al.* [9], highlights this shortcoming.

Example 3.1 *Assume that the inequalities in the chance constraint* (3) *are perfectly positively correlated in the sense that*

$$y_i^0(\boldsymbol{x}) = \delta_i \hat{y}^0(\boldsymbol{x}) \quad \textit{and} \quad y_i(\boldsymbol{x}) = \delta_i \hat{y}(\boldsymbol{x})$$

for some affine functions $\hat{y}^0 : \mathbb{R}^n \to \mathbb{R}$ *and* $\hat{y} : \mathbb{R}^n \to \mathbb{R}^k$ *and for some fixed constants* $\delta_i > 0$ *for* $i = 1, \ldots, m$. *In this case, it can readily be seen that the joint chance constraint* (3) *is equivalent to the robust individual chance constraint*

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0\right) \geq 1 - \epsilon. \tag{27}$$

Thus, the least conservative choice for ϵ_i *which guarantees that* (26) *implies* (3) *is* $\epsilon_i = \epsilon$ *for* $i = 1, \ldots, m$. *However, this means that the* ϵ_i *sum to* $m\epsilon$ *instead of* ϵ *as required by the Bonferroni approximation. In fact, the optimal choice for* $\boldsymbol{\epsilon} \in \mathcal{E}$ *is* $\epsilon_i = \epsilon/m$ *for* $i = 1, \ldots, m$. *This example demonstrates that the quality of the Bonferroni approximation diminishes as* m *increases if the inequalities in the joint chance constraint are positively correlated.*

3.2 Approximation by Chen, Sim, Sun and Teo

In order to mitigate the potential over-conservatism of the Bonferroni approximation, Chen *et al.* [9] proposed an approximation based on a different inequality from probability theory. The starting point is the observation that the joint chance constraint (3) can be reformulated as

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\max_{i=1,\ldots,m}\left\{\alpha_i\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}}\right)\right\} \leq 0\right) \geq 1 - \epsilon \tag{28}$$

for any vector of strictly positive *scaling parameters* $\boldsymbol{\alpha} \in \mathcal{A} = \{\boldsymbol{\alpha} \in \mathbb{R}^m \ : \ \boldsymbol{\alpha} > \mathbf{0}\}$. Note that the choice of $\boldsymbol{\alpha} \in \mathcal{A}$ does not affect the feasible region of the chance constraint (28). Although these scaling parameters are seemingly unnecessary, it turns out that they can be tuned to improve the approximation to be developed below. Chen *et al.* [9] note that (28) represents a distributionally robust *individual* chance constraint, which can be conservatively approximated by a Worst-Case CVaR constraint. Thus, for any $\boldsymbol{\alpha} \in \mathcal{A}$, the requirement

$$\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{\boldsymbol{x} \in \mathbb{R}^n \ : \ \sup_{\mathbb{P}\in\mathcal{P}} \mathrm{CVaR}_\epsilon\left(\max_{i=1,\ldots,m}\left\{\alpha_i\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}}\right)\right\}\right) \leq 0\right\} \tag{29}$$

implies that $\boldsymbol{x} \in \mathcal{X}^{\mathrm{JCC}}$, see Proposition 2.1. It is important to note that, in contrast to the chance constraint (28), the Worst-Case CVaR constraint $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ *does* depend on the choice of $\boldsymbol{\alpha} \in \mathcal{A}$. Thus, the Worst-Case CVaR constraint in (29) is not equivalent to the robust chance constraint (28) since the max function in (28) is convex piecewise linear, see also Theorem 2.2 and Example 2.1.

The following theorem due to Chen *et al.* [9] relies on a classical result in order statistics and provides a tractable SOCP-based conservative approximation for $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$.

Theorem 3.2 (Approximation by Chen *et al.*) *For any $\boldsymbol{\alpha} \in \mathcal{A}$ we have $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}^{\mathrm{JCC}}$ where $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \{\boldsymbol{x} \in \mathbb{R}^n : \hat{\mathcal{J}}(\boldsymbol{x}, \boldsymbol{\alpha}) \leq 0\}$ and*

$$\hat{\mathcal{J}}(\boldsymbol{x}, \boldsymbol{\alpha}) = \min_{w^0 \in \mathbb{R}, \boldsymbol{w} \in \mathbb{R}^k} \left\{ \min_{\beta \in \mathbb{R}} \left[\beta + \frac{1}{\epsilon} \pi \left(w^0 - \beta,\ \boldsymbol{w} \right) \right] + \frac{1}{\epsilon} \left[\sum_{i=1}^{m} \pi \left(\alpha_i y_i^0(\boldsymbol{x}) - w^0,\ \alpha_i y_i(\boldsymbol{x}) - \boldsymbol{w} \right) \right] \right\},$$

where

$$\pi\left(z^0, \boldsymbol{z}\right) = \frac{1}{2}\left(z^0 + \boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{z}\right) + \frac{1}{2} \left\| \left(z^0 + \boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{z},\ \boldsymbol{\Sigma}^{1/2} \boldsymbol{z} \right) \right\|_2$$

Note that, since the feasible set $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a tractable conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$ for any $\boldsymbol{\alpha} \in \mathcal{A}$, the union $\bigcup_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ still constitutes a conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$. Chen *et al.* [9] prove also that their approximation is tighter than the Bonferroni approximation by showing that $\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon}) \subseteq \bigcup_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ for all $\boldsymbol{\epsilon} \in \mathcal{E}$. Unfortunately, similar to the Bonferroni approach, the approximation by Chen *et al.* depends critically on the choice of $\boldsymbol{\alpha}$, while the problem of finding the best $\boldsymbol{\alpha} \in \mathcal{A}$ for a generic chance constrained program of the type (4) is nonconvex and therefore believed to be intractable.

3.3 The Worst-Case CVaR Approximation

Both approximations discussed so far rely on inequalities from probability theory, which are not necessarily tight. In this section we show that the set $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ has in fact an exact tractable representation in terms of LMIs and therefore promises to provide a tight convex approximation for $\mathcal{X}^{\mathrm{JCC}}$.

Theorem 3.3 *For any fixed $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{\alpha} \in \mathcal{A}$, we have*

$$\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists (\beta, \mathbf{M}) \in \mathbb{R} \times \mathbb{S}^{k+1}, \\ \beta + \frac{1}{\epsilon} \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \\ \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i^{\mathsf{T}} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m \end{array} \right\}. \tag{30}$$

Proof: We note that the constraint $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is equivalent to $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) \leq 0$, where

$$\begin{array}{rcl} \mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) & = & \sup\limits_{\mathbb{P} \in \mathcal{P}} \mathrm{CVaR}_{\epsilon} \left(\max\limits_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^{\mathsf{T}} \tilde{\boldsymbol{\xi}} \right) \right\} \right) \\ & = & \inf\limits_{\beta \in \mathbb{R}} \left\{ \beta + \dfrac{1}{\epsilon} \sup\limits_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\left[\max\limits_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^{\mathsf{T}} \tilde{\boldsymbol{\xi}} \right) \right\} - \beta \right]^+ \right) \right\} \end{array} \tag{31}$$

denotes the Worst-Case CVaR. As in Section 2, the first step towards a tractable reformulation of $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ is to solve the worst-case expectation problem

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\left[\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^{\mathsf{T}} \tilde{\boldsymbol{\xi}} \right) \right\} - \beta \right]^+ \right). \tag{32}$$

For any fixed $\boldsymbol{x} \in \mathcal{X}$, $\beta \in \mathbb{R}$, and $\boldsymbol{\alpha} \in \mathcal{A}$, Lemma A.1 enables us to reformulate (32) as

$$\begin{array}{ll} \inf\limits_{\mathbf{M} \in \mathbb{S}^{k+1}} & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq \max\limits_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \right) \right\} - \beta \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{array} \tag{33}$$

We emphasize that (33) represents a lossless reformulation of the worst-case expectation problem (32). The semi-infinite constraint in (33) can be expanded into m simpler semi-infinite constraints of the form

$$\left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \boldsymbol{\xi} \right) - \beta \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k,\ i = 1, \ldots, m,$$

which can be equivalently expressed as the following system of LMIs.

$$\mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m$$

We can therefore reformulate the worst-case expectation problem (32) as

$$\begin{array}{ll} \inf\limits_{\mathbf{M} \in \mathbb{S}^{k+1}} & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m. \end{array} \tag{34}$$

Substituting (34) into (31) yields

$$\begin{array}{rcll} \mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) & = & \inf & \beta + \dfrac{1}{\epsilon} \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ & & \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R} \\ & & & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m, \end{array} \tag{35}$$

and thus the claim follows. ∎

Theorem 3.3 establishes that $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ has an exact representation in terms of LMIs. We have already seen in Section 3.2 that $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}^{\mathrm{JCC}}$ for all $\boldsymbol{\alpha} \in \mathcal{A}$ and that $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$, see Theorem 3.2. Thus, $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a tractable conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$ which is at least as tight as $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$.

Recall from Section 3.2 that $\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon}) \subseteq \bigcup_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ for all $\boldsymbol{\epsilon} \in \mathcal{E}$. Moreover, we have $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subset \mathcal{X}^{\mathrm{JCC}}$ for all $\boldsymbol{\alpha} \in \mathcal{A}$. This allows us to conclude that our new approximation is at least as tight as the two state-of-the-art approximations discussed above.

Remark 3.1 *In contrast to the classical Bonferroni approximation, the Worst-Case CVaR approximation behaves reasonably in situations in which the m inequalities in the chance constraint* (3) *are positively correlated. Indeed, by choosing $\alpha_i := 1/\delta_i > 0$ for all $i = 1, \ldots, m$ in Example 3.1, the constraint*

$\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ *is equivalent to*

$$\exists \beta \in \mathbb{R},\ \mathbf{M} \in \mathbb{S}^{k+1}\ :\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^{\mathsf{T}} & y^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0},$$

which can easily be identified as the SDP reformulation of the individual chance constraint (27)*. This implies that* $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \mathcal{X}^{\mathrm{ICC}}$ *for all* $\boldsymbol{\alpha} \in \mathcal{A}$ *in Example 3.1, see also Theorem 2.1. Thus, by choosing* $\boldsymbol{\alpha}$ *appropriately, our method can provide tight approximations for distributionally robust joint chance constraints, even in situations when the* m *inequalities are positively correlated.*

3.4 Dual Interpretation of the Worst-Case CVaR Approximation

In this section we explore a different way to find a tractable conservative approximation for the chance constraint (3). Subsequently, we will prove that this approximation is equivalent to the Worst-Case CVaR approximation.

Consider again the robust *individual* chance constraint (28) which is equivalent to the robust *joint* chance constraint (3) for any fixed $\boldsymbol{\alpha} \in \mathcal{A}$. Instead of approximating (28) by a Worst-Case CVaR constraint, we can approximate the max-function in the chance constraint (28) by a quadratic majorant of the form $q(\boldsymbol{\xi}) = \boldsymbol{\xi}^{\mathsf{T}}\mathbf{Q}\boldsymbol{\xi} + \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{q} + q^0$ that satisfies

$$\begin{array}{rl} & q(\boldsymbol{\xi}) \geq \max\limits_{i=1,\ldots,m} \left\{\alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\xi}\right)\right\} \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k, \\ \iff & q(\boldsymbol{\xi}) \geq \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\xi}\right) \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k,\ i = 1, \ldots, m. \end{array} \tag{36}$$

Replacing the max function in (28) by $q(\boldsymbol{\xi})$ yields the distributionally robust (individual) *quadratic* chance constraint

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\tilde{\boldsymbol{\xi}}^{\mathsf{T}}\mathbf{Q}\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^{\mathsf{T}}\boldsymbol{q} + q^0 \leq 0\right) \geq 1 - \epsilon. \tag{37}$$

For further argumentation, we define

$$\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \mathbf{Q} \in \mathbb{S}^k,\ \boldsymbol{q} \in \mathbb{R}^k,\ q^0 \in \mathbb{R} \quad \text{such that} \\ q(\boldsymbol{\xi}) = \boldsymbol{\xi}^{\mathsf{T}}\mathbf{Q}\boldsymbol{\xi} + \boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{q} + q^0 \text{ satisfies (36) and (37)} \end{array} \right\}. \tag{38}$$

Proposition 3.1 *For any fixed* $\boldsymbol{\alpha} \in \mathcal{A}$ *the feasible set* $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ *constitutes a conservative approximation for* $\mathcal{X}^{\mathrm{JCC}}$*, that is,* $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}^{\mathrm{JCC}}$*.*

Proof: Note that any $\boldsymbol{x}$ feasible in (28) is also feasible in (38) since

$$\mathbb{P}\left(\tilde{\boldsymbol{\xi}}^{\mathsf{T}}\mathbf{Q}\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^{\mathsf{T}}\boldsymbol{q} + q^0 \leq 0\right) \ \leq\ \mathbb{P}\left(\max_{i=1,\ldots,m} \left\{\alpha_i(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^{\mathsf{T}}\tilde{\boldsymbol{\xi}})\right\} \leq 0\right) \quad \forall \mathbb{P} \in \mathcal{P}.$$

Since $\boldsymbol{x}$ is feasible in (28) if and only if $\boldsymbol{x} \in \mathcal{X}^{\mathrm{JCC}}$, the claim follows. ∎

Theorem 3.4 *For any fixed $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{\alpha} \in \mathcal{A}$ we have*

$$\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \exists \mathbf{Q} \in \mathbb{S}^k, \quad \boldsymbol{q} \in \mathbb{R}^k, \quad q^0 \in \mathbb{R}, \quad \beta \in \mathbb{R}, \quad \mathbf{M} \in \mathbb{S}^{k+1}, & \\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0}, & \\ \begin{bmatrix} \mathbf{Q} & \frac{1}{2}(\boldsymbol{q} - \alpha_i \boldsymbol{y}_i(\boldsymbol{x})) \\ \frac{1}{2}(\boldsymbol{q} - \alpha_i \boldsymbol{y}_i(\boldsymbol{x}))^\mathsf{T} & q^0 - \alpha_i y_i^0(\boldsymbol{x}) \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m & \end{array} \right\}.$$

Proof: Note that the constraints in (36) are equivalent to

$$\begin{bmatrix} \mathbf{Q} & \frac{1}{2}(\boldsymbol{q} - \alpha_i \boldsymbol{y}_i(\boldsymbol{x})) \\ \frac{1}{2}(\boldsymbol{q} - \alpha_i \boldsymbol{y}_i(\boldsymbol{x}))^\mathsf{T} & q^0 - \alpha_i y_i^0(\boldsymbol{x}) \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m.$$

Moreover, by Theorem 2.2, the robust quadratic chance constraint (37) is equivalent to the Worst-Case CVaR constraint

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}\left(\tilde{\boldsymbol{\xi}}^\mathsf{T} \mathbf{Q} \tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\mathsf{T} \boldsymbol{q} + q^0 \right) \;=\; \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\left[\tilde{\boldsymbol{\xi}}^\mathsf{T} \mathbf{Q} \tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\mathsf{T} \boldsymbol{q} + q^0 - \beta \right]^+ \right) \right\} \;\leq\; 0. \qquad (39)$$

By the proof of part (ii) in Theorem 2.3, we know that (39) can be written as

$$\begin{array}{rcll} 0 & \geq & \inf & \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ & & \text{s.t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R} \\ & & & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0}. \end{array}$$

Thus, the claim follows. ∎

In the following theorem we show that the approximate feasible set $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is equivalent to the set $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ found in Section 3.3. This implies that the approximation of a distributionally robust joint chance constraint by a Worst-Case CVaR constraint is equivalent to the approximation of the max function implied by the joint chance constraint by a quadratic majorant. Note that both approximations depend of the choice of the scaling parameters $\boldsymbol{\alpha}$.

Theorem 3.5 *For any $\boldsymbol{\alpha} \in \mathcal{A}$ we have $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$.*

Proof: By defining the combined variable

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 \end{bmatrix},$$

the set $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ can be rewritten as

$$\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \mathbf{Y} \in \mathbb{S}^k, \quad \beta \in \mathbb{R}, \quad \mathbf{M} \in \mathbb{S}^{k+1}, \\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0} \\ \mathbf{M} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \beta \end{bmatrix} \succcurlyeq \mathbf{Y} \succcurlyeq \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) \end{bmatrix} \quad \forall i = 1, \ldots, m \end{array} \right\},$$

It is easy to see that $\mathbf{Y}$ may be eliminated from the above representation of $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ by rewriting the last constraint group as

$$\mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m.$$

This observation establishes the postulated equivalence. ∎

3.5 Exactness of the Worst-Case CVaR Approximation

So far we have shown that, for any fixed $\boldsymbol{\alpha} \in \mathcal{A}$, the feasible set $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a tractable conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$. This implies that the union $\mathcal{Z}^{\mathrm{JCC}} = \bigcup_{\boldsymbol{\alpha} \in \mathcal{S}} \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ still constitutes a conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$. We now demonstrate that this improved approximation is essentially exact. To this end, we introduce the feasible set

$$\mathcal{X}_\circ^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\bigcap_{i=1}^m \left\{ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} < 0 \right\} \right) \geq 1 - \epsilon \right\}$$

corresponding to a *strict* version of the joint chance constraint.

Theorem 3.6 *The Worst-Case CVaR approximation is essentially exact if $\boldsymbol{\alpha}$ is treated as a decision variable. Formally, we have $\mathcal{X}_\circ^{\mathrm{JCC}} \subseteq \mathcal{Z}^{\mathrm{JCC}} \subseteq \mathcal{X}^{\mathrm{JCC}}$.*

Proof: The theorem can be proved by invoking a Chebyshev-type bound for the worst-case probability of a random vector to lie in the intersection of a set of quadratic (or, a fortiori, linear) inequalities, see Vandenberghe *et al.* [28]. To keep this paper self-contained, we provide here an elementary proof which is reminiscent of the exactness proof in Section 3.5.

The second inclusion follows immediately from the known conservativeness of the CVaR approximation. Therefore, it is sufficient to prove the first inclusion. By using similar arguments as in Section 3.1, we can rewrite $\mathcal{X}_\circ^{\mathrm{JCC}}$ as

$$\mathcal{X}_\circ^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\bigcup_{i=1}^m \left\{ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \geq 0 \right\} \right) \leq \epsilon \right\}.$$

By Lemma A.2 in the Appendix we may thus conclude that

$$\mathcal{X}_\circ^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \mathbf{M} \in \mathbb{S}^{k+1}, \quad \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \\ \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq 1 \quad \forall \boldsymbol{\xi} \in \bigcup_{i=1}^m \left\{ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \boldsymbol{\xi} \geq 0 \right\} \end{array} \right\}.$$

The semi-infinite constraint in the above representation of $\mathcal{X}_\circ^{\mathrm{JCC}}$ can be reexpressed as

$$\left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq 1 \quad \forall \boldsymbol{\xi} \ : \ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \boldsymbol{\xi} \geq 0, \quad \forall i = 1, \ldots, m,$$

which, by the $\mathcal{S}$-lemma, is equivalent to

$$\exists \boldsymbol{\alpha} \geq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) + 1 \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m.$$

Thus, the feasible set $\mathcal{X}_\circ^{\mathrm{JCC}}$ can be written as

$$\mathcal{X}_\circ^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \mathbf{M} \in \mathbb{S}^{k+1}, \quad \boldsymbol{\alpha} \in \mathbb{R}^m, \\ \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \boldsymbol{\alpha} > \mathbf{0}, \\ \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) + 1 \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m \end{array} \right\}. \tag{40}$$

Note that we require here without loss of generality that $\boldsymbol{\alpha}$ is strictly positive. Indeed, it can be shown that no feasible $\boldsymbol{\alpha}$ has any vanishing components. By Theorem 3.3, we have

$$\mathcal{Z}^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \beta \in \mathbb{R}, \quad \mathbf{M} \in \mathbb{S}^{k+1}, \quad \boldsymbol{\alpha} \in \mathcal{A} \\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \\ \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m \end{array} \right\}. \tag{41}$$

It is now clear that $\mathcal{X}_\circ^{\mathrm{JCC}} \subseteq \mathcal{Z}^{\mathrm{JCC}}$ since we are free to set $\beta = -1$ in (41) and since $-1 + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0$ is equivalent to $\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon$. This observation completes the proof. ∎

Remark 3.2 *Note that $\mathcal{Z}^{\mathrm{JCC}} = \mathcal{X}^{\mathrm{JCC}}$ for $m = 1$; see Corollary 2.1. In general, however, both inclusions in Theorem 3.6 can be strict. If there is no degenerate constraint function with $(y_i^0(\boldsymbol{x}), \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T})^\mathsf{T} = \mathbf{0}$ $\forall \boldsymbol{x} \in \mathbb{R}^n$, then $\mathcal{N} = \bigcup_{i=1}^m \left\{ \boldsymbol{x} \in \mathbb{R}^n : (y_i^0(\boldsymbol{x}), \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T})^\mathsf{T} = \mathbf{0} \right\}$ constitutes a Lebesgue null set as it is a finite union of strict affine subspaces of $\mathbb{R}^n$. By using similar arguments as in the proof of Theorem 3.6 one can show that $\mathcal{X}^{\mathrm{JCC}} \backslash \mathcal{X}_\circ^{\mathrm{JCC}} \subseteq \mathcal{N}$, which implies that $\mathcal{X}^{\mathrm{JCC}}$ and $\mathcal{X}_\circ^{\mathrm{JCC}}$ differ at most by a Lebesgue null set for well-specified chance constraints.*

Theorem 3.6 implies that the original joint chance constrained program

$$\underset{\boldsymbol{x} \in \mathcal{X} \cap \mathcal{X}^{\mathrm{JCC}}}{\text{minimize}} \quad \boldsymbol{c}^\mathsf{T} \boldsymbol{x}$$

$$\begin{array}{ll} \underset{\substack{\boldsymbol{x}\in\mathcal{X}\cap\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})\\ \boldsymbol{\alpha}\in\mathcal{A}}}{\text{minimize}} & \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} \end{array} \tag{42}$$

attain the same optimal value except in degenerate cases. Unfortunately, optimizing jointly over $\boldsymbol{x} \in \mathcal{X} \cap \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ and $\boldsymbol{\alpha} \in \mathcal{A}$ in (42) involves Bilinear Matrix Inequalities (BMIs). It is known that generic BMI constrained problems are $\mathcal{NP}$-hard, see [27]. Similar nonconvexities arise also in the approximations discussed in Sections 3.1 and 3.2, which underlines the general perception that problems with distributionally robust joint chance constraints are hard to solve.

Recall, however, that for any fixed $\boldsymbol{\alpha} \in \mathcal{A}$, the set $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is representable in terms of tractable LMI constraints involving the auxiliary variables β and $\mathbf{M}$. In particular, the constraints in (41) are convex in $\beta, \mathbf{M}$, and $\boldsymbol{x}$ for any fixed $\boldsymbol{\alpha}$, and convex in $\boldsymbol{\alpha}$ for any fixed $\beta, \mathbf{M}$, and $\boldsymbol{x}$. In Section 3.7 we will use this property to propose an algorithm for solving (42) approximately.

3.6 Injecting Support Information

In many practical applications the support of the (true) distribution $\mathbb{Q}$ of $\tilde{\boldsymbol{\xi}}$ is known to be a strict subset of $\mathbb{R}^k$. Disregarding this information in the definition of $\mathcal{P}$ can result in unnecessarily conservative robust chance constraints. In this section we briefly outline how support information can be used to tighten robust joint chance constraints and their approximations developed in Section 3. To this end, we first revise our distributional assumptions.

Distributional Assumptions. The random vector $\tilde{\boldsymbol{\xi}}$ has a distribution $\mathbb{Q}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} \succ \mathbf{0}$. We assume that $\mathbb{Q}$ is supported on $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}^k \ : \ [\boldsymbol{\xi}^{\mathsf{T}}\ 1]\mathbf{W}_i[\boldsymbol{\xi}^{\mathsf{T}}\ 1]^{\mathsf{T}} \leq 0\ \forall i = 1, \ldots, l\}$, where $\mathbf{W}_i \in \mathbb{S}^{k+1}$ for all $i = 1, \ldots, l$.[1] Thus, we have $\mathbb{Q}(\tilde{\boldsymbol{\xi}} \in \Xi) = 1$. We define $\mathcal{P}_\Xi$ as the set of all probability distributions supported on Ξ that have the same first- and second-order moments as $\mathbb{Q}$.

In this section we are interested in tractable conservative approximations for the feasible set

$$\mathcal{X}_\Xi^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \inf_{\mathbb{P}\in\mathcal{P}_\Xi} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^{\mathsf{T}}\tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1, \ldots, m \right) \geq 1 - \epsilon \right\}.$$

As before, we study approximate feasible sets of the form

$$\mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \sup_{\mathbb{P}\in\mathcal{P}_\Xi} \mathrm{CVaR}_\epsilon\left(\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^{\mathsf{T}}\tilde{\boldsymbol{\xi}} \right) \right\} \right) \leq 0 \right\}$$

for $\boldsymbol{\alpha} \in \mathcal{A}$. By using similar arguments as in Section 2.1, one can show that $\mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}_\Xi^{\mathrm{JCC}}$ for all $\boldsymbol{\alpha} \in \mathcal{A}$. However, the sets $\mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha})$ have no longer an exact representation in terms of LMIs. Instead, they need to be conservatively approximated.

[1] Note that every finite intersection of half-spaces and ellipsoids in $\mathbb{R}^k$ is representable as a set of the form Ξ.

Theorem 3.7 *For any fixed $\boldsymbol{\alpha} \in \mathcal{A}$, we have $\mathcal{Y}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}_\Xi^{\mathrm{JCC}}$, where $\mathcal{Y}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha})$ has the following tractable reformulation in terms of LMIs.*

$$\mathcal{Y}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \exists \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R}, \quad \boldsymbol{\tau}_i \in \mathbb{R}^l, & \\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \boldsymbol{\tau}_i \geq \mathbf{0} & \forall i = 0, \ldots, m \\ \mathbf{M} + \sum_{j=1}^l \tau_{0,j} \mathbf{W}_j \succcurlyeq \mathbf{0} & \\ \mathbf{M} + \sum_{j=1}^l \tau_{i,j} \mathbf{W}_j - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} & \forall i = 1, \ldots, m \end{array} \right\}. \quad (43)$$

Furthermore, for $l = 1$, we have $\mathcal{Y}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha})$.

Proof: The proof widely parallels the proof of Theorem 3.3. The only difference is that $\mathbb{R}^k$ is replaced by Ξ and that we use the $\mathcal{S}$-lemma to approximate (for $l > 1$) or reformulate (for $l = 1$) the semi-infinite constraints over Ξ by LMI constraints.

Remark 3.3 *While $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is exactly representable in terms of LMIs in the absence of support information, Theorem 3.7 only provides a conservative LMI approximation for $\mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha})$. Nevertheless, it is easily verified that $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Y}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha})$ and therefore $\mathcal{Y}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a better approximation for $\mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha})$ than $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$. In fact, by setting $\boldsymbol{\tau}_i = \mathbf{0}$ for all $i = 0, \ldots, m$, (43) reduces to (35).*

Remark 3.4 *Support information can also be used in a straightforward way to tighten the approximations discussed in Sections 3.1 and 3.2.*

3.7 Optimizing over the Scaling Parameters

By Theorem 3.6, the original distributionally robust chance constrained program (4) can be written as

$$\begin{array}{ll} \underset{\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{\alpha} \in \mathcal{A}}{\text{minimize}} & \boldsymbol{c}^\mathsf{T} \boldsymbol{x} \\ \text{subject to} & \mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) \leq 0 \\ & \boldsymbol{x} \in \mathcal{X}, \end{array} \quad (44)$$

where the Worst-Case CVaR functional $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ is defined as in (31). Unfortunately, as discussed in Section 3.3, $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ is merely biconvex, but not jointly convex in $\boldsymbol{x}$ and $\boldsymbol{\alpha}$. Thus, optimization problem (44) is nonconvex. By Theorem 3.3, however, the problem becomes convex and tractable when the values of the scaling parameters $\boldsymbol{\alpha}$ are frozen.

For the further argumentation we define the set $\bar{\mathcal{A}} = \{\boldsymbol{\alpha} \ : \ \boldsymbol{\alpha} \geq \delta \boldsymbol{e}\}$, where $\boldsymbol{e}$ denotes the vector of ones and $\delta > 0$ represents a small tolerance, which we set to 10^{-7}. Note that, unlike $\mathcal{A}$, the set $\bar{\mathcal{A}}$ is

closed. Consider now the following optimization model where $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$ is fixed.

$$\begin{array}{ll} \min\limits_{\boldsymbol{x} \in \mathbb{R}^n} & \boldsymbol{c}^\mathsf{T} \boldsymbol{x} \\ \text{s.t.} & \mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) \leq 0 \\ & \boldsymbol{x} \in \mathcal{X} \end{array} \tag{45}$$

We emphasize again that by Theorem 3.3 (45) is equivalent to a tractable SDP and that any $\boldsymbol{x}$ feasible in (45) is also feasible in the original chance constrained problem (4). In the remainder of this section we develop an algorithm that repeatedly solves (45) while systematically improving the scaling parameters $\boldsymbol{\alpha}$.

The main idea of this approach, which is inspired by [9], is to minimize $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ over $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$ with the aim of enlarging the feasible region of problem (45) and thereby improving the objective value. To this end, we introduce the following optimization model which depends parametrically on $\boldsymbol{x} \in \mathcal{X}$.

$$\begin{array}{ll} \min\limits_{\boldsymbol{\alpha} \in \mathbb{R}^m} & \mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) \\ \text{s.t.} & \boldsymbol{\alpha} \in \bar{\mathcal{A}} \end{array} \tag{46}$$

Theorem 3.3 implies that (46) can also be expressed as a tractable SDP.

Assume that $\boldsymbol{x}^*$ is an optimal solution of problem (45) for a given $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$. By the feasibility of $\boldsymbol{x}^*$ in (45) we know that $\mathcal{J}(\boldsymbol{x}^*, \boldsymbol{\alpha}) \leq 0$. Keeping $\boldsymbol{x}^*$ fixed, we then solve problem (46) to obtain the optimal scaling parameters $\boldsymbol{\alpha}^*$ corresponding to $\boldsymbol{x}^*$. By construction, we find

$$\mathcal{J}(\boldsymbol{x}^*, \boldsymbol{\alpha}^*) \;\leq\; \mathcal{J}(\boldsymbol{x}^*, \boldsymbol{\alpha}) \;\leq 0. \tag{47}$$

The above inequalities imply that the optimal objective value of problem (45) with input $\boldsymbol{\alpha}^*$ must not exceed $\boldsymbol{c}^\mathsf{T} \boldsymbol{x}^*$. Therefore, by solving the problems (45) and (46) in alternation, we obtain a sequence of monotonically decreasing objective values. This motivates the following algorithm, which relies on the availability of an initial feasible solution $\boldsymbol{x}_{\text{init}}$ for problem (45).

Algorithm 3.1 *Sequential Convex Optimization Procedure*

1. ***Initialization***. *Let $\boldsymbol{x}_{\text{init}}$ be some feasible solution of problem (45). Set the current solution to $\boldsymbol{x}^0 \leftarrow \boldsymbol{x}_{\text{init}}$, the current objective value to $f^0 \leftarrow \boldsymbol{c}^\mathsf{T} \boldsymbol{x}^0$, and the iteration counter to $t \leftarrow 1$.*

2. ***Scaling Parameter Optimization***. *Solve problem (46) with input $\boldsymbol{x}^{t-1}$ and let $\boldsymbol{\alpha}^*$ denote an optimal set of scaling parameters. Set $\boldsymbol{\alpha}^t \leftarrow \boldsymbol{\alpha}^*$.*

3. ***Decision Optimization***. *Solve problem (45) with input $\boldsymbol{\alpha}^t$ and let $\boldsymbol{x}^*$ denote an optimal solution. Set $\boldsymbol{x}^t \leftarrow \boldsymbol{x}^*$ and $f^t \leftarrow \boldsymbol{c}^\mathsf{T} \boldsymbol{x}^t$.*

4. ***Termination***. *If $(f^t - f^{t-1})/|f^{t-1}| \leq \gamma$ (where γ is a given small tolerance), output $\boldsymbol{x}^t$ and stop. Otherwise, set $t \leftarrow t + 1$ and go back to Step 2.*

Theorem 3.8 *Assume that $\boldsymbol{x}_{\text{init}}$ is feasible in problem (45) for some $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$. Then, the sequence of objective values $\{f^t\}$ generated by Algorithm 3.1 is monotonically decreasing. If the set $\mathcal{X}$ is bounded, then the sequence $\{\boldsymbol{x}^t\}$ is also bounded, while the sequence $\{f^t\}$ converges to a finite limit.*

Proof: By the inequality (47), an update of the scaling parameters from $\boldsymbol{\alpha}^{t-1}$ to $\boldsymbol{\alpha}^t$ in Step 2 of the algorithm preserves the feasibility of $\boldsymbol{x}^{t-1}$ in problem (45). This guarantees that the sequence of objective values $\{f^t\}$ is monotonically decreasing. Furthermore, it is readily seen that the solution sequence $\{\boldsymbol{x}^t\}$ is bounded if the feasible set $\mathcal{X}$ is bounded. Since (45) has a continuous objective function, the monotonicity of the objective value sequence implies that $\{f^t\}$ has a finite limit. ∎

Remark 3.5 *Algorithm 3.1 can also be used in the presence of support information as discussed in Section 3.6. In this case, the Worst-Case CVaR functional $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ has to be redefined in the obvious way. Algorithm 3.1 can further be used in the context of the approximation by Chen* et al.*, see Section 3.2. In this case, $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ is replaced by its conservative approximation $\hat{\mathcal{J}}(\boldsymbol{x}, \boldsymbol{\alpha})$ defined in Theorem 3.2. Details are omitted for brevity of exposition.*

We emphasize that Algorithm 3.1 does not necessarily find the global optimum of problem (44). Nevertheless, as confirmed by the numerical results in the next section, the method can perform well in practice.

4 Numerical Results

We consider a dynamic water reservoir control problem for hydro power generation, which is inspired by a model due to Andrieu *et al.* [2]. Let $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1,\ \tilde{\xi}_2, \ldots,\ \tilde{\xi}_T)$ denote the sequence of stochastic inflows (precipitation) into the reservoir at time instances $t = 1, \ldots, T$. The history of inflows up to time t is denoted by $\tilde{\boldsymbol{\xi}}^t = (\tilde{\xi}_1, \ldots, \tilde{\xi}_t)$, where $\tilde{\boldsymbol{\xi}}^T = \tilde{\boldsymbol{\xi}}$. We let $\boldsymbol{\mu} \in \mathbb{R}^T$ and $\boldsymbol{\Sigma} \in \mathbb{S}^T$ denote the mean vector and covariance matrix of $\tilde{\boldsymbol{\xi}}$, respectively. Furthermore, $\tilde{\boldsymbol{\xi}}$ is supported on a rectangle of the form $\Xi = [\boldsymbol{l},\ \boldsymbol{u}]$. However, we assume that no further information about the true distribution of $\tilde{\boldsymbol{\xi}}$ is available. As usual, we let $\mathcal{P}_\Xi$ denote the set of all distributions supported on Ξ with matching first- and second-order moments. We denote by $x_t(\tilde{\boldsymbol{\xi}}^t)$ the amount of water released from the reservoir in period t. Note that the decision $x_t(\tilde{\boldsymbol{\xi}}^t)$ is selected at time t after $\tilde{\boldsymbol{\xi}}^t$ has been observed and is therefore a function of the observation history. We require $x_t(\tilde{\boldsymbol{\xi}}^t) \geq 0$ almost surely for all $\mathbb{P} \in \mathcal{P}_\Xi$ and $t = 1, \ldots, T$. The water level at time t is computed as the sum of the initial level l_0 and the cumulative inflows minus the cumulative releases up to time t, that is,

$$l_0 + \sum_{i=1}^{t} \tilde{\xi}_i - \sum_{i=1}^{t} x_t(\tilde{\boldsymbol{\xi}}^t).$$

We require that the water level remains between some upper threshold l_{high} (flood reserve) and some lower threshold l_{low} (dead storage) over all time periods $t = 1, \ldots, T$ with probability $1 - \epsilon$, where $\epsilon \in (0, 1)$. The water released in any period t is used to produce electric energy which is sold at a

periodic price

$$c_t = 10 + 5\sin\left[\frac{\pi(1-t)}{3}\right] \quad \forall t = 1, \ldots, T.$$

The worst-case expected profit over all time periods is computed as

$$\inf_{\mathbb{P}\in\mathcal{P}_\Xi} \mathbb{E}_{\mathbb{P}}\left(\sum_{t=1}^{T} c_t x_t(\tilde{\boldsymbol{\xi}}^t)\right).$$

In order to determine an admissible control strategy that maximizes the worst-case profit, we must solve the following distributionally robust joint chance constrained problem.

$$\begin{array}{ll} \underset{x_1(\cdot),\ldots,x_T(\cdot)}{\text{maximize}} & \inf\limits_{\mathbb{P}\in\mathcal{P}_\Xi} \mathbb{E}_{\mathbb{P}}\left(\sum\limits_{t=1}^{T} c_t x_t(\tilde{\boldsymbol{\xi}}^t)\right) \\ \text{subject to} & \inf\limits_{\mathbb{P}\in\mathcal{P}_\Xi} \mathbb{P}\left(l_{\text{low}} \le l_0 + \sum\limits_{i=1}^{t}\tilde{\xi}_i - \sum\limits_{i=1}^{t} x_t(\tilde{\boldsymbol{\xi}}^t) \le l_{\text{high}} \quad \forall t = 1,\ldots,T\right) \ge 1-\epsilon \\ & x_t(\tilde{\boldsymbol{\xi}}^t) \ge 0 \quad \mathbb{P}\text{-a.s.} \quad \forall \mathbb{P}\in\mathcal{P}_\Xi,\ t = 1,\ldots,T \end{array} \tag{48}$$

Note that (48) is an infinite dimensional problem since the control decisions $x_t(\cdot)$ are generic measurable functionals of the uncertain inflows. To reduce the problem complexity, we focus on policies that are affine functions of $\tilde{\boldsymbol{\xi}}$. Thus, we optimize over affine disturbance feedback policies of the form

$$x_t(\tilde{\boldsymbol{\xi}}^t) = x_t^0 + \boldsymbol{x}_t^{\mathsf{T}}\mathbf{P}_t\tilde{\boldsymbol{\xi}} \quad \forall t = 1,\ldots,T, \tag{49}$$

where $x_t^0 \in \mathbb{R}$, $\boldsymbol{x}_t \in \mathbb{R}^t$ and $\mathbf{P}_t : \mathbb{R}^T \to \mathbb{R}^t$ is a truncation operator that maps $\tilde{\boldsymbol{\xi}}$ to $\tilde{\boldsymbol{\xi}}^t$. By focusing on affine control policies we conservatively approximate the infinite dimensional dynamic problem (48) by a problem with a polynomial number of variables, namely, the coefficients $\{x_t^0, \boldsymbol{x}_t\}_{t=1}^T$. For more details on the use of affine control policies in robust control and stochastic programming, see, e.g., Ben-Tal *et al.* [3], Chen *et al.* [10], and Kuhn *et al.* [15].

By applying now standard robust optimization techniques [3], the requirement that $x_t(\tilde{\boldsymbol{\xi}}^t) \ge 0$ holds almost surely can be expressed as

$$\begin{array}{ll} & x_t^0 + \boldsymbol{x}_t^{\mathsf{T}}\mathbf{P}_t\boldsymbol{\xi} \ge 0 \quad \forall \boldsymbol{\xi}\in\Xi \\ \iff & 0 \le \min\limits_{\boldsymbol{\xi}\in\mathbb{R}^T}\left\{x_t^0 + \boldsymbol{x}_t^{\mathsf{T}}\mathbf{P}_t\boldsymbol{\xi} \ :\ \boldsymbol{l} \le \boldsymbol{\xi} \le \boldsymbol{u}\right\} \\ \iff & 0 \le \max\limits_{\boldsymbol{\lambda}_t\in\mathbb{R}^T}\left\{x_t^0 + \boldsymbol{x}_t^{\mathsf{T}}\mathbf{P}_t\boldsymbol{u} + \boldsymbol{\lambda}_t^{\mathsf{T}}(\boldsymbol{l}-\boldsymbol{u}) \ :\ \boldsymbol{\lambda}_t \ge \mathbf{P}_t^{\mathsf{T}}\boldsymbol{x}_t,\ \boldsymbol{\lambda}_t \ge \mathbf{0}\right\} \\ \iff & \exists \boldsymbol{\lambda}_t \in \mathbb{R}^T \ :\ x_t^0 + \boldsymbol{x}_t^{\mathsf{T}}\mathbf{P}_t\boldsymbol{u} + \boldsymbol{\lambda}_t^{\mathsf{T}}(\boldsymbol{l}-\boldsymbol{u}) \ge 0,\ \boldsymbol{\lambda}_t \ge \mathbf{P}_t^{\mathsf{T}}\boldsymbol{x}_t,\ \boldsymbol{\lambda}_t \ge \mathbf{0}. \end{array}$$

By substituting (49) into (48) we thus obtain the following conservative approximation for (48).

$$
\begin{array}{ll}
\text{maximize} & \displaystyle\sum_{t=1}^{T} c_t \left(x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \boldsymbol{\mu}\right) \\
\text{subject to} & \boldsymbol{\lambda}_t \in \mathbb{R}^T, \quad \boldsymbol{x}_t \in \mathbb{R}^t \quad \forall t = 1, \ldots, T \\
& \displaystyle\inf_{\mathbb{P} \in \mathcal{P}_\Xi} \mathbb{P} \left(\begin{array}{l} l_0 - l_{\text{high}} + \displaystyle\sum_{i=1}^{t} \tilde{\xi}_i - \left(\sum_{i=1}^{t} x_i^0 + \boldsymbol{x}_i^\mathsf{T} \mathbf{P}_i \tilde{\boldsymbol{\xi}} \right) \leq 0 \quad \forall t = 1, \ldots, T \\ l_{\text{low}} - l_0 - \displaystyle\sum_{i=1}^{t} \tilde{\xi}_i + \left(\sum_{i=1}^{t} x_i^0 + \boldsymbol{x}_i^\mathsf{T} \mathbf{P}_i \tilde{\boldsymbol{\xi}} \right) \leq 0 \quad \forall t = 1, \ldots, T \end{array} \right) \geq 1 - \epsilon \\
& \left. \begin{array}{l} x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \boldsymbol{u} + \boldsymbol{\lambda}_t^\mathsf{T} (\boldsymbol{l} - \boldsymbol{u}) \geq 0 \\ \boldsymbol{\lambda}_t \geq \mathbf{P}_t^\mathsf{T} \boldsymbol{x}_t, \ \boldsymbol{\lambda}_t \geq \mathbf{0} \end{array} \right\} \quad \forall t = 1, \ldots, T
\end{array}
\tag{50}
$$

Note that the joint chance constraint in (50) involves $2T$ inequalities that are bilinear in the decisions $\{\boldsymbol{x}_t\}_{t=1}^T$ and the random vector $\tilde{\boldsymbol{\xi}}$. Problem (50) can therefore be identified as a special instance of problem (4) and is amenable to the approximation methods described in Section 3. In the remainder of this section, we compare the performance of these approximation methods.

In the subsequent tests, we set $T = 5$, $l_0 = 1$, $l_{\text{low}} = 1$, and $l_{\text{high}} = 5$. The mean value of $\tilde{\xi}_t$ is assumed to be 1, while its standard deviation is set to 10%, over all time periods. Furthermore, we set the correlation of different stochastic inflows to 25% for adjacent time periods and 0% otherwise. Finally, we assume that $\Xi = [0, 2]^T$. All tests are run for a range of reliability levels ϵ between 1% and 10% in steps of 1%.

We first solve problem (50) using the Bonferroni approximation by decomposing the joint chance constraint into $2T$ individual chance constraints with reliability factors $\epsilon_i = \epsilon/(2T)$ for $i = 1, \ldots, 2T$. The resulting optimal objective value is denoted by V^B, and the associated optimal solution is used to initialize Algorithm 3.1. We run the algorithm using the Worst-Case CVaR approximation as well as the approximation by Chen *et al.* described in Section 3.2. We denote the resulting optimal objective values by V^M and V^U, respectively. In both cases the algorithm's convergence threshold is set to $\gamma = 10^{-6}$. All SDPs arising from the Worst-Case CVaR approximation are solved with SDPT3 using the YALMIP interface [16], while all SOCPs arising from the Bonferroni approximation and the approximation by Chen *et al.* are solved with MOSEK using the algebraic modeling toolbox ROME [13].

Table 1 reports the optimal objective values and the improvement of V^M relative to V^U and V^B. As expected, all three methods yield optimal objective values that increase with ϵ because the joint chance constraint becomes less restrictive as ϵ grows. At $\epsilon = 1\%$ the objective values of the different approximations coincide. However, V^M exceeds V^U and V^B for all the other values of ϵ. In this particular example, our method outperforms the Bonferroni approximation by up to 25% and the approximation by Chen *et al.* by up to 12%. Table 1 also reports the runtimes of the different algorithms. All instances based on the Worst-Case CVaR approximation are solved in less then 20 seconds, while the instances based on the approximation by Chen *et al.* and the Bonferroni approximation are solved in less then 5

ϵ	V^M	V^U	V^B	$(V^M - V^U)/V^U$	$(V^M - V^B)/V^B$	R^M	R^U	R^B
1%	44.3	44.3	44.3	0.0%	0.0%	2.18	2.50	0.82
2%	44.9	44.3	44.3	1.4%	1.3%	17.47	2.51	0.82
3%	49.4	44.4	44.3	11.3%	11.4%	14.99	4.19	0.81
4%	52.4	46.7	44.5	12.2%	17.6%	14.14	4.17	0.82
5%	54.5	49.0	45.2	11.2%	20.5%	15.79	4.18	0.81
6%	56.3	50.9	46.0	10.6%	22.5%	17.30	4.24	0.82
7%	57.8	53.0	46.7	9.1%	23.6%	15.98	4.54	0.86
8%	58.9	54.7	47.3	7.7%	24.5%	13.82	4.62	0.82
9%	59.9	56.0	47.8	7.0%	25.2%	17.70	4.16	0.82
10%	60.7	57.1	48.8	6.3%	24.5%	14.29	4.24	0.81

Table 1: Optimal objective values of the water reservoir control problem for the Worst-Case CVaR approximation (V^M), the approximation by Chen *et al.* (V^U), and the Bonferroni approximation (V^B). The table also reports the percentage gaps $(V^M - V^U)/V^U$ and $(V^M - V^B)/V^B$ as well as the runtimes for the three algorithms (R^M, R^U, R^B) in seconds.

and 1 seconds, respectively. Thus, as expected, the improved solution quality offered by the (SDP-based) Worst-Case CVaR approximation over the two (SOCP-based) benchmark approximations comes at an increased computational overhead.

Acknowledgements. The authors are indebted to Prof. A. Ben-Tal for valuable discussions on the topic of this paper and would also like to thank EPSRC for financial support under grant EP/H0204554/1.

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[2] L. Andrieu, R. Henrion, and W. Römisch. A model for dynamic chance constraints in hydro power reservoir management. 207(2):579–589, 2010.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

[4] D. Bertsimas, X. Doan, K. Natarajan, and C.P. Teo. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602, 2010.

[5] G. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753.

[6] G. Calafiore and L. El Ghaoui. Distributionally robust chance-constrained linear programs with applications. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[7] G. Calafiore, U. Topcu, and L. El Ghaoui. Parameter estimation with expected and residual-at-risk criteria. *Systems & Control Letters*, 58(1):39–46, 2009.

[8] 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.

[9] W. Chen, M. Sim, J. Sun, and Teo C.P. From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, 58(2):470–485, 2010.

[10] X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071.

[11] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 2010.

[12] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming, Series B*, 107:37–61, 2006.

[13] J. Goh and M. Sim. Robust optimization made easy with ROME. *Operations Research*, 2010.

[14] K. Isii. The extrema of probability determined by generalized moments (i) bounded random variables. *Annals of the Institute of Statistical Mathematics*, 12(2):119–134, 1960.

[15] D. Kuhn, W. Wiesemann, and A. Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 2009.

[16] J. Löfberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[17] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.

[18] L.B. Miller and H. Wagner. Chance-constrained programming with joint constraints. *Operations Research*, 13(6):930–945, 1965.

[19] K. Natarajan, D. Pachamanova, and M. Sim. Constructing risk measures from uncertainty sets. *Operations Research*, 57(5):1129–1141, 2009.

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

[21] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications*, 142(2), 2009.

[22] I. Polik and T. Terlaky. A survey of the S-lemma. *SIAM Review*, 49(3):371–481, 2007.

[23] A. Prékopa. On probabilistic constrained programming. In *Proceedings of the Princeton Symposium on Mathematical Programming*, pages 113–138, Princeton University Press, Princeton, 1970.

[24] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2002.

[25] A. Shapiro. *On duality theory of conic linear problems.* Semi-infinite programming: recent advances (M.A.Goberna and M.A. Lopez, eds.), Kluwer Academic Publishers, 2001.

[26] A. Shapiro and A.J. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, (17):523–542, 2002.

[27] O. Toker and H. Ozbay. On the NP-hardness of solving bilinear matrix inequalities and simultaneous stabilization with static output feedback. In *Proceedings of the American Control Conference*, pages 2525–2526, Seatle, Washington, 1995.

[28] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49:52–64, 2007.

[29] S. Zymler, D. Kuhn, and B. Rustem. Worst-case value-at-risk of non-linear portfolios. *Available from Optimization Online*, 2009.

A Worst-Case Expectation and Probability Problems

Lemma A.1 *Let $f : \mathbb{R}^k \to \mathbb{R}$ be a measurable function, and define the worst-case expectation $\theta_{\rm wc}$ as*

$$\theta_{\rm wc} = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \left((f(\tilde{\boldsymbol{\xi}}))^+ \right),$$

where $\mathcal{P}$ represents the usual set of all probability distributions on $\mathbb{R}^k$ with given mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then,

$$\theta_{\rm wc} = \inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \left\{ \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \;:\; \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\mathsf{T} \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T} \; 1\right]^\mathsf{T} \geq f(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \right\},$$

where $\boldsymbol{\Omega}$ is the second-order moment matrix of $\tilde{\boldsymbol{\xi}}$.

Proof: The worst-case expectation $\theta_{\rm wc}$ can equivalently be expressed as

$$\begin{array}{rcll} \theta_{\rm wc} & = & \sup\limits_{\mu\in\mathcal{M}_+} & \displaystyle\int_{\mathbb{R}^k} \max\{0, f(\boldsymbol{\xi})\}\mu(\mathrm{d}\boldsymbol{\xi}) \\ & & \text{s.t.} & \displaystyle\int_{\mathbb{R}^k} \mu(\mathrm{d}\boldsymbol{\xi}) = 1 \\ & & & \displaystyle\int_{\mathbb{R}^k} \boldsymbol{\xi}\mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\mu} \\ & & & \displaystyle\int_{\mathbb{R}^k} \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}\mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}, \end{array} \tag{51}$$

where $\mathcal{M}_+$ represents the cone of nonnegative Borel measures on $\mathbb{R}^k$. The optimization variable of the semi-infinite linear program (51) is the nonnegative measure μ. Note that the first constraint forces μ to be a probability measure. The other two constraints enforce consistency with the given first- and second-order moments, respectively. We now assign dual variables $y_0 \in \mathbb{R}$, $\boldsymbol{y} \in \mathbb{R}^k$, and $\mathbf{Y} \in \mathbb{S}^k$ to the equality constraints in (51), respectively, and introduce the following dual problem (see, e.g., [25]).

$$\begin{array}{ll} \inf & y_0 + \boldsymbol{y}^\mathsf{T}\boldsymbol{\mu} + \langle \mathbf{Y}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T} \rangle \\ \text{s.t.} & y_0 \in \mathbb{R}, \quad \boldsymbol{y} \in \mathbb{R}^k, \quad \mathbf{Y} \in \mathbb{S}^k \\ & y_0 + \boldsymbol{y}^\mathsf{T}\boldsymbol{\xi} + \langle \mathbf{Y}, \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} \rangle \geq \max\{0, f(\boldsymbol{\xi})\} \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \end{array} \tag{52}$$

Because $\boldsymbol{\Sigma} \succ \mathbf{0}$, it can be shown that strong duality holds [14]. Therefore, the worst-case probability $\theta_{\rm wc}$ coincides with the optimal value of the dual problem (52). By defining the combined variable

$$\mathbf{M} = \begin{bmatrix} \mathbf{Y} & \frac{1}{2}\boldsymbol{y} \\ \frac{1}{2}\boldsymbol{y}^\mathsf{T} & y_0 \end{bmatrix},$$

problem (52) reduces to

$$\begin{array}{ll} \inf\limits_{\mathbf{M}\in\mathbb{S}^{k+1}} & \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]\mathbf{M}\left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]^{\mathsf{T}} \geq \max\{0, f(\boldsymbol{\xi})\} \quad \forall \boldsymbol{\xi}\in\mathbb{R}^k. \end{array} \tag{53}$$

Note that the semi-infinite constraint in (53) can be expanded in terms of two equivalent semi-infinite constraints.

$$\left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]\mathbf{M}\left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]^{\mathsf{T}} \geq 0 \qquad \forall \boldsymbol{\xi}\in\mathbb{R}^k \tag{54a}$$

$$\left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]\mathbf{M}\left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]^{\mathsf{T}} \geq f(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi}\in\mathbb{R}^k \tag{54b}$$

Since (54a) is equivalent to $\mathbf{M} \succcurlyeq \mathbf{0}$, the claim follows. ∎

Lemma A.2 *Let $\mathcal{S} \subseteq \mathbb{R}^k$ be any Borel measurable set (which is not necessarily convex), and define the worst-case probability π_{wc} as*

$$\pi_{\text{wc}} = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{S}\}, \tag{55}$$

Then,

$$\pi_{\text{wc}} = \inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \left\{ \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \ :\ \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]\mathbf{M}\left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]^{\mathsf{T}} \geq 1 \quad \forall \boldsymbol{\xi}\in\mathcal{S} \right\}.$$

.

Proof: The proof is due to Calafiore *et al.* [7], see also Zymler *et al.* [29]. A sketch of the proof is provided here to keep this paper self-contained. Define the indicator function of the set $\mathcal{S}$ as

$$\mathbb{I}_{\mathcal{S}}(\boldsymbol{\xi}) \quad = \quad \begin{cases} 1 & \text{if } \boldsymbol{\xi}\in\mathcal{S}, \\ 0 & \text{otherwise.} \end{cases}$$

The worst-case probability problem (55) can equivalently be expressed as

$$\begin{array}{rcll} \pi_{\text{wc}} & = & \sup\limits_{\mu\in\mathcal{M}_+} & \int_{\mathbb{R}^k} \mathbb{I}_{\mathcal{S}}(\boldsymbol{\xi})\mu(\mathrm{d}\boldsymbol{\xi}) \\ & & \text{s.t.} & \int_{\mathbb{R}^k} \mu(\mathrm{d}\boldsymbol{\xi}) = 1 \\ & & & \int_{\mathbb{R}^k} \boldsymbol{\xi}\mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\mu} \\ & & & \int_{\mathbb{R}^k} \boldsymbol{\xi}\boldsymbol{\xi}^T\mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T. \end{array}$$

By dualizing this problem and applying similar manipulations as in the proof of Lemma A.1 we obtain the postulated result. ∎