Robust Convex Quadratically Constrained Quadratic Programming with Mixed-Integer Uncertainty

Can Gokalp, Areesh Mittal, and Grani A. Hanasusanto

Graduate Program in Operations Research and Industrial Engineering, The University of Texas at Austin, USA

January 10, 2018

Abstract

We study robust convex quadratically constrained quadratic programs where the uncertain problem parameters can contain both continuous and integer components. Under the natural boundedness assumption on the uncertainty set, we show that the generic problems are amenable to exact copositive programming reformulations of polynomial size. The emerging convex optimization problems are NP-hard but admit a conservative semidefinite programming (SDP) approximation that can be solved efficiently. We prove that this approximation is stronger than the popular approximate $\mathcal{S}$-lemma method for problem instances with only continuous uncertainty. We also show that all results can be extended to the two-stage robust optimization setting if the problem has complete recourse. We assess the effectiveness of our proposed SDP reformulations and demonstrate their superiority over the state-of-the-art solution schemes on stylized instances of least squares, project management, and multi-item newsvendor problems.

1 Introduction

A wide variety of decision making problems in engineering, physical, or economic system can be formulated as convex quadratically constrained quadratic programs of the form

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{A}_0(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_0(\boldsymbol{x})^\top \boldsymbol{\xi} + c_0(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} \\ & \|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) \leq 0 \quad \forall i \in [I]. \end{array} \tag{1}$$

Here, $\mathcal{X} \subseteq \mathbb{R}^D$ is the feasible set of the decision vector $\boldsymbol{x}$ and is assumed to be described by a tractable polytope , $\boldsymbol{\xi} \in \mathbb{R}^K$ is a vector of exogenous problem parameters, $\boldsymbol{A}_i(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^{M \times K}$ and $\boldsymbol{b}_i(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^K$ are matrix- and vector-valued affine functions, respectively, while $c_i(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}$ is a convex quadratic function for every $i \in [I] \cup \{0\}$. The objective of problem (1) is to determine the best decision $\boldsymbol{x} \in \mathcal{X}$ that minimizes the quadratic function $\|\boldsymbol{A}_0(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_0(\boldsymbol{x})^\top \boldsymbol{\xi} + c_0(\boldsymbol{x})$, while ensuring that this solution remains feasible to the constraint system that is described by I quadratic inequalities $\|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) \leq 0$, $i \in [I]$. The generic formulation (1) includes as a special case the class of linear programming problems [38] (when $\boldsymbol{A}_i = \boldsymbol{0}$, $i \in [I] \cup \{0\}$), and has numerous important applications, e.g., in portfolio optimization [33], least squares regression [23], supervised classification [13], optimal control [37], etc. In addition to their exceptional modeling power, quadratic optimization problems of the form (1) are attractive as they can be solved efficiently using standard off-the-shelf solvers.

In many situations of practical interest, the exact values of the parameters $\boldsymbol{\xi}$ are unknown when the decisions are made and can only be estimated through limited historical data or measurements. Thus, they are subject to potential significant errors that can adversely impact the out-of-sample performance of the optimal solution $\boldsymbol{x}$. One popular approach to address decision problems under uncertainty is via *robust optimization* [2]. In this setting, we assume that the vector of uncertain parameters $\boldsymbol{\xi}$ lives within a prescribed uncertainty set Ξ and we replace the functions in the objective and constraints of (1), respectively, with the *worst-case* functions given by

$$\sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) \qquad \forall i \in [I] \cup \{0\}. \tag{2}$$

The emerging optimization problem yields a solution $\boldsymbol{x}$ that minimizes the quadratic objective function under the most adverse uncertain parameter realization $\boldsymbol{\xi} \in \Xi$, while ensuring that the solution remains feasible to the constraint system for all possible values of the uncertain parameter vectors in Ξ.

Robust optimization models are appealing as they require minimal assumptions on the description of uncertainty and as they often lead to efficient solution schemes. In the case of a linear programming setting, the resulting robust optimization problems are tractable for many relevant uncertainty sets and have been broadly applied to problems in engineering, finance, machine learning, and operations management [4, 6, 24]. Tractable reformulations for robust quadratic programming problems are derived in [22, 32] for the particular case when the quadratic functions (in $\boldsymbol{x}$) exhibit a concave dependency in the uncertain parameters $\boldsymbol{\xi}$. When the functions are convex in both $\boldsymbol{x}$ and $\boldsymbol{\xi}$ as we consider in this paper, the corresponding robust problems are generically NP-hard if the uncertainty set is defined by a polytope but becomes tractable—by virtue of the *exact* $\mathcal{S}$-lemma—if the uncertainty set is defined by a simple ellipsoid [20, 4]. Tractable approximation schemes have also been proposed for the standard setting that we consider in this paper. If the uncertainty set is described by a finite intersection of ellipsoids then a conservative semidefinite programming (SDP) reformulation is obtained by leveraging the *approximate* $\mathcal{S}$-lemma [5]. In [7], a special class of functions is introduced to approximate the quadratic terms in (2). The arising robust optimization problems are tractable if the uncertainty sets are defined through affinely transformed norm balls. In [32], conservative and progressive SDP approximations are devised by replacing each quadratic term in (2) with linear upper and lower bounds, respectively.

Most of the existing literature in robust optimization assume that the uncertain problem parameters are continuous and reside in a tractable conic representable set Ξ. However, certain applications require the use of mixed-integer uncertainty. Such decision problems arise prominently in the supply chain context where demands of non-perishable products are more naturally represented as integer quantities and in the discrete choice modeling context where the outcomes are chosen from a discrete set of alternatives. Other pertinent examples include robust optimization applications in machine learning [46, 11, 39] and in network optimization [44, 1]. If the uncertain parameters contain mixed-integer components then the problem immediately becomes computationally formidable even in the simplest setting. Specifically, if all functions are affine in $\boldsymbol{\xi}$ and the uncertain problem parameters are described by binary vectors then computing the worst-case values in (2) is already NP-hard [18]. Only in a few contrived situations is the corresponding robust version of (1) tractable, *e.g.*, when the uncertainty set possesses a totally unimodularity property or is described by the convex hull of polynomially many integer vectors [4]. Perhaps due to these limiting reasons there are currently no results in the literature that provide a systematic and rigorous way to handle generic (linear or quadratic) robust optimization problems with mixed-integer uncertainty. In this paper, we introduce a new method to approximate these intractable problems. We first reformulate the original problem as an equivalent finite-dimensional conic program of polynomial size, which absorbs all the difficulty in its cone, and then replace the cone with tractable inner approximations.

Optimization problems under uncertainty may also involve adaptive recourse decisions which are taken once the realization of the uncertain parameters is observed [2, 42]. This setting gives rise to difficult min-max-min optimization problems which are generically NP-hard even if both the first- and the second-stage cost functions are affine in $\boldsymbol{x}$ and $\boldsymbol{\xi}$ [3]. Thus, they can only be solved approximately, either by employing discretization schemes which approximate the continuum of uncertainty space with finitely many points [25, 27, 41] or by employing decision rules methods which restrict the set of all possible recourse decisions to simpler parametric forms in $\boldsymbol{\xi}$ [3, 19, 21]. We refer the reader to [15] for a comprehensive review of recent results in adaptive robust optimization.

The conic programming route that we take here to model optimization problems under uncertainty has previously been traversed. In [35], completely positive programming reformulations are derived to compute best-case expectations of mixed zero-one linear programs under first- and second-order moment information on the joint distributions of the uncertain parameters. This result has been extended and applied to other pertinent settings such as in stochastic appointment scheduling problems, discrete choice models, random walks and sequencing problems, etc. [28, 34, 30]. Recently, equivalent copositive programming reformulations are derived for generic two-stage robust linear programs [26, 45]. The resulting optimization problems are amenable to conservative semidefinite programming reformulations which are often stronger than the ones obtained from employing quadratic decision rules on the recourse function.

In this paper, we advance the state-of-the-art in robust optimization along several directions. We summarize our main contributions as follows:

1. We prove that any robust convex quadratically constrained quadratic program is equivalent to a copositive program of polynomial size if the uncertainty set is given by a *bounded* mixed-integer polytope. We further show that the exactness result can be extended to the two-stage robust optimization setting if the problem has *complete recourse*.

2. By employing the hierarchies of semidefinite representable cones to approximate the copositive cones, we obtain sequences of tractable conservative approximations for the robust problem. These approximations can be made to have any arbitrary accuracy. Extensive numerical experiments suggest that even the simplest of these approximations distinctly outperforms the state-of-the-art approximation schemes in terms of accuracy.

3. We prove that the simplest approximation is stronger than the well-known approximate $\mathcal{S}$-lemma method if the corresponding problem instance has only continuous uncertain parameters.

4. To our best knowledge, we are the first to provide an exact conic programming reformulation and to propose tractable semidefinite programming approximations for well established classes of one-stage and two-stage quadratically constrained quadratic programs with mixed-integer uncertainty.

The remainder of the paper is structured as follows. We formulate and discuss the generic robust quadratically constrained quadratic programs in Section 2. We then derive the copositive programming reformulations in Section 3. Section 4 develops a conservative SDP reformulation and provides a theoretical comparison with the popular approximate $\mathcal{S}$-lemma method. Finally, we demonstrate the impact of our proposed reformulation via numerical experiments in Section 5.

Notation: For any $I \in \mathbb{N}$, we use $[I]$ to denote the index set $\{1, \ldots, I\}$. The identity matrix and the vector of all ones are denoted by $\mathbb{I}$ and $\mathbf{e}$, respectively. The dimension of such matrices will be clear from the context. We denote by $\mathrm{tr}(\boldsymbol{M})$ the trace of a square matrix $\boldsymbol{M}$. For a vector $\boldsymbol{v}$, $\mathrm{diag}(\boldsymbol{v})$ denotes the diagonal matrix with $\boldsymbol{v}$ on its diagonal; whereas for a square matrix $\boldsymbol{M}$, $\mathrm{diag}(\boldsymbol{M})$ denotes the vector comprising the diagonal elements of $\boldsymbol{M}$. We define $\boldsymbol{P} \circ \boldsymbol{Q}$ as the Hadamard product (element-wise product) of two matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ of same size. For any integer $Q \in \mathbb{Z}_+$, we define $\mathbf{v}_Q = [2^0 \ 2^1 \ \cdots \ 2^{Q-1}]^\top$ as the vector comprising all q-th powers of 2, for $q = 0, 1, \ldots, Q-1$. We define by $\mathbb{S}^K$ ($\mathbb{S}^K_+$) the space of all symmetric (positive semidefinite) matrices in $\mathbb{R}^{K \times K}$. The cone of copositive matrices is denoted by $\mathcal{C} = \{\boldsymbol{M} \in \mathbb{S}^K : \boldsymbol{\xi}^\top \boldsymbol{M} \boldsymbol{\xi} \geq 0 \ \forall \boldsymbol{\xi} \geq \mathbf{0}\}$, while its dual cone, the cone of completely positive matrices, is denoted by $\mathcal{C}^* = \{\boldsymbol{M} \in \mathbb{S}^K : \boldsymbol{M} = \boldsymbol{B}\boldsymbol{B}^\top \text{ for some } \boldsymbol{B} \in \mathbb{R}_+^{K \times g(K)}\}$, where $g(K) = \max\{\binom{K+1}{2} - 4, K\}$ [40]. For any $\boldsymbol{P}, \boldsymbol{Q} \in \mathbb{S}^K$, the relations $\boldsymbol{P} \succeq \boldsymbol{Q}$, $\boldsymbol{P} \succeq_{\mathcal{C}} \boldsymbol{Q}$, and $\boldsymbol{P} \succeq_{\mathcal{C}^*} \boldsymbol{Q}$ indicate that $\boldsymbol{P} - \boldsymbol{Q}$ is an element of $\mathbb{S}_+^K$, $\mathcal{C}$, and $\mathcal{C}^*$, respectively.

2 Problem Formulation

We study robust convex quadratically constrained quadratic programs (RQCQPs) of the form

$$\begin{array}{ll} \text{minimize} & \sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}_0(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_0(\boldsymbol{x})^\top \boldsymbol{\xi} + c_0(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} \\ & \|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) \leq 0 \quad \forall \boldsymbol{\xi} \in \Xi \ \forall i \in [I], \end{array} \tag{3}$$

where the set $\mathcal{X}$ and the functions $\boldsymbol{A}_i(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^{M \times K}$, $\boldsymbol{b}_i(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^K$, and $c_i(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}$ have the same definitions as those in (1). The vector $\boldsymbol{\xi} \in \mathbb{R}^K$ comprises all the uncertain problem parameters and is assumed to belong to the uncertainty set Ξ given by a bounded mixed-integer polyhedral set

$$\Xi = \left\{ \boldsymbol{\xi} \in \mathbb{R}_+^K : \begin{array}{l} \boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{t} \\ \xi_\ell \in \mathbb{Z} \quad \forall \ell \in [L] \end{array} \right\}, \tag{4}$$

where $\boldsymbol{S} \in \mathbb{R}^{J \times K}$ and $\boldsymbol{t} \in \mathbb{R}^J$. We assume without loss of generality that the first L elements of $\boldsymbol{\xi}$ are integer, while the remaining $K - L$ are continuous. Since Ξ is bounded, we may further assume that there exists a scalar integer $Q \in \mathbb{Z}_+$ such that $\xi_l \in \{0, \cdots, 2^Q - 1\}$ for every $\ell \in [L]$. Note that that the quantity Q is bounded by a polynomial function in the bit length of the description of $\boldsymbol{S}$ and $\boldsymbol{t}$.

Example 1 (Robust Portfolio Optimization)**.** *Consider the classical Markowitz's mean-variance portfolio optimization problem*

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} - \lambda \boldsymbol{\mu}^\top \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x} \in \Delta^K, \end{array} \tag{5}$$

where Δ^K is the unit simplex in $\mathbb{R}^K$, $\lambda \in [0, \infty)$ is the prescribed risk tolerance level of the investor, while $\boldsymbol{\mu} \in \mathbb{R}^K$ and $\boldsymbol{\Sigma} \in \mathbb{S}^K$ are the true mean and covariance matrix of the asset returns, respectively. The objective of this problem is to determine the best vector of weights $\boldsymbol{x} \in \Delta^K$ that maximizes the mean portfolio return $\boldsymbol{\mu}^\top \boldsymbol{x}$ and that also minimizes the portfolio risk that is captured by the variance term $\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}$. Here, the tradeoff between these two terms is controlled by the scalar λ in the objective function.

In practice, the true values of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and can only be estimated by using the available N historical asset returns $\{\hat{\boldsymbol{\xi}}_n\}_{n \in [N]}$, as follows:

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{n \in [N]} \hat{\boldsymbol{\xi}}_n \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \sum_{n \in [N]} \left(\hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\mu}}\right) \left(\hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\mu}}\right)^\top.$$

In the robust optimization setting, we assume that the precise location of each sample point $\hat{\boldsymbol{\xi}}_n$ is uncertain and is only known to belong to a prescribed uncertainty set Ξ_n containing $\hat{\boldsymbol{\xi}}_n$. To bring the resulting problem into the standard form (3)*, we introduce the expanded uncertainty set*

$$\Xi = \left\{ \left((\hat{\boldsymbol{\xi}}_n)_{n\in[N]}, (\hat{\boldsymbol{\chi}}_n)_{n\in[N]} \right) \in \mathbb{R}_+^{NK+NK} : \hat{\boldsymbol{\xi}}_n \in \Xi_n, \;\; \hat{\boldsymbol{\chi}}_n = \hat{\boldsymbol{\xi}}_n - \frac{1}{N} \sum_{n'\in[N]} \hat{\boldsymbol{\xi}}_{n'} \quad \forall n \in [N] \right\} \tag{6}$$

comprising the terms $\hat{\boldsymbol{\xi}}_n$ and $\hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\mu}}$, $n \in [N]$. Using this uncertainty set, we arrive at the following robust version of (5)*.*

$$\begin{array}{ll} \text{minimize} & \displaystyle\sup_{((\hat{\boldsymbol{\xi}}_n)_n, (\hat{\boldsymbol{\chi}}_n)_n) \in \Xi} \left(\frac{1}{N-1} \sum_{n\in[N]} (\hat{\boldsymbol{\chi}}_n^\top \boldsymbol{x})^2 - \frac{\lambda}{N} \sum_{n\in[N]} \hat{\boldsymbol{\xi}}_n^\top \boldsymbol{x} \right) \\ \text{subject to} & \boldsymbol{x} \in \Delta^K \end{array} \tag{7}$$

This problem constitutes an instance of (3) *with the input parameters $I = 0$,*

$$\boldsymbol{A}_0(\boldsymbol{x}) = \frac{1}{\sqrt{N-1}} \begin{bmatrix} \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top \\ \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \boldsymbol{x}^\top & \cdots & \mathbf{0}^\top \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{0}^\top & \cdots & \boldsymbol{x}^\top \end{bmatrix}, \quad \boldsymbol{b}_0(\boldsymbol{x}) = -\frac{\lambda}{N} \begin{bmatrix} \boldsymbol{x} \\ \vdots \\ \boldsymbol{x} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad \textit{and} \quad c_0(\boldsymbol{x}) = 0.$$

Example 2 (Robust Project Crashing)**.** *Consider a project that is described by an activity-on-arc network $\mathcal{N}(\mathcal{V}, \mathcal{A})$, where $\mathcal{V}$ is the set of nodes representing the events, while $\mathcal{A}$ is the set of arcs representing the activities. We define $d_{ij} \in [0, 1]$ to be the nominal duration of the activity $(i, j) \in \mathcal{A}$. Here, we assume that the durations d_{ij}, $(i, j) \in \mathcal{A}$, are already normalized so that they take values in the unit interval.*

The goal of project crashing is to determine the best resource assignments x_{ij}, $(i, j) \in \mathcal{A}$, on the activities that minimize the project completion time or makespan. The project completion time corresponds to the length of the longest path of the network whose arc capacities are given by the activity durations $d_{ij} - x_{ij}$, $(i, j) \in \mathcal{A}$. We can formulate project crashing as the optimization problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sup_{\boldsymbol{z}\in\mathcal{Z}} \sum_{(i,j)\in\mathcal{A}} (d_{ij} - x_{ij}) z_{ij} \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{8}$$

where

$$\mathcal{Z} = \left\{ \boldsymbol{z} \in \{0,1\}^{|\mathcal{A}|} : \sum_{j:(i,j)\in\mathcal{A}} z_{ij} - \sum_{j:(j,i)\in\mathcal{A}} z_{ji} = \left\{ \begin{array}{ll} 1 & \textit{if } i = 1 \\ -1 & \textit{if } i = |\mathcal{V}| \\ 0 & \textit{if otherwise} \end{array} \right. , \quad \forall i \in \mathcal{V} \right\}. \tag{9}$$

If the task durations $\boldsymbol{d}$ are uncertain and are only known to belong to the prescribed uncertainty set $\mathcal{D} \subseteq [0,1]^{|\mathcal{A}|}$, then we arrive at the robust optimization problem

$$\begin{array}{ll} \text{minimize} & \sup_{\boldsymbol{d}\in\mathcal{D}} \left(\sup_{\boldsymbol{z}\in\mathcal{Z}} \sum_{(i,j)\in\mathcal{A}} (d_{ij} - x_{ij}) z_{ij} \right) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}. \end{array}$$

By combining the suprema over $\mathcal{D}$ and $\mathcal{Z}$, and linearizing the bilinear terms $d_{ij}z_{ij}$, $(i,j) \in \mathcal{A}$, we can reformulate the objective of this problem as

$$\sup_{\boldsymbol{d}\in\mathcal{D}} \sup_{\boldsymbol{z}\in\mathcal{Z}} \sum_{(i,j)\in\mathcal{A}} (d_{ij} - x_{ij}) z_{ij} = \sup_{(\boldsymbol{d},\boldsymbol{z},\boldsymbol{q})\in\Xi} \mathbf{e}^\top \boldsymbol{q} - \boldsymbol{x}^\top \boldsymbol{z}, \tag{10}$$

where

$$\Xi = \left\{ (\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{q}) \in \mathcal{D} \times \mathcal{Z} \times \mathbb{R}_+^{|\mathcal{A}|} : \boldsymbol{q} \leq \boldsymbol{z},\ \boldsymbol{q} \leq \boldsymbol{d},\ \boldsymbol{q} \geq \boldsymbol{d} - \mathbf{e} + \boldsymbol{z} \right\}. \tag{11}$$

Using the new objective function (10) *and uncertainty set* (11)*, the resulting robust optimization problem constitutes an instance of* (3) *with the input parameters $I = 0$, $\boldsymbol{A}_0(\boldsymbol{x}) = \mathbf{0}$, $\boldsymbol{b}_0(\boldsymbol{x}) = [\mathbf{0}^\top \;\; -\boldsymbol{x}^\top \;\; \mathbf{e}^\top]^\top$, and $c_0(\boldsymbol{x}) = 0$.*

In the remainder of the paper, for any fixed $\boldsymbol{x} \in \mathcal{X}$, we define the mixed-integer quadratic programs

$$Z_i(\boldsymbol{x}) = \sup_{\boldsymbol{\xi}\in\Xi} \|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) \qquad \forall i \in [I] \cup \{0\}, \tag{12}$$

which correspond to the inner subproblems in the objective and the constraints of (3). As the i-th semi-infinite constraint in (3) can equivalently be reformulated as the constraint $Z_i(\boldsymbol{x}) \leq 0$, we may therefore represent the semi-infinite program (3) as the finite program

$$\begin{array}{ll} \text{minimize} & Z_0(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} \\ & Z_i(\boldsymbol{x}) \leq 0 \quad \forall i \in [I]. \end{array} \tag{13}$$

In this paper, we attempt to derive exact copositive programming reformulations for the terms $Z_i(\boldsymbol{x})$, $i \in [I] \cup \{0\}$, in (13). By substituting the respective terms in (13) with the emerging copositive reformulations, we obtain an equivalent finite-dimensional convex program for the RQCQP (3) that is principally amenable to numerical solution.

3 Copositive Programming Reformulation

In this section, we derive an equivalent copositive programming reformulation for (3) by adopting the following steps. For any fixed $\boldsymbol{x} \in \mathcal{X}$ and $i \in [I] \cup \{0\}$, we first derive a copositive upper bound on $Z_i(\boldsymbol{x})$. We then show that the resulting reformulation is in fact exact in view of the boundedness assumption on the uncertainty set Ξ.

3.1 A Copositive Upper Bound on $Z_i(\boldsymbol{x})$

To derive the copositive reformulation, we leverage the following result by Burer [9] which enables us to reduce a generic mixed-binary quadratic program into an equivalent conic program of polynomial size.

Theorem 1 ([9, Theorem 2.6])**.** *The mixed-binary quadratic program*

$$\begin{array}{ll} \text{maximize} & \boldsymbol{\xi}^\top \boldsymbol{Q} \boldsymbol{\xi} + \boldsymbol{r}^\top \boldsymbol{\xi} \\ \text{subject to} & \boldsymbol{\xi} \in \mathbb{R}^P_+ \\ & \boldsymbol{F}\boldsymbol{\xi} = \boldsymbol{g} \\ & \xi_\ell \in \{0,1\} \qquad \forall \ell \in \mathcal{L} \end{array} \tag{14}$$

is equivalent the completely positive program

$$\begin{array}{ll} \text{maximize} & \operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{Q}) + \boldsymbol{r}^\top \boldsymbol{\xi} \\ \text{subject to} & \boldsymbol{\xi} \in \mathbb{R}^P_+ \\ & \boldsymbol{F}\boldsymbol{\xi} = \boldsymbol{g}, \ \operatorname{diag}(\boldsymbol{F}\boldsymbol{\Omega}\boldsymbol{F}^\top) = \boldsymbol{g} \circ \boldsymbol{g} \\ & \xi_\ell = \Omega_{\ell\ell} \qquad \forall \ell \in \mathcal{L} \\ & \begin{bmatrix} \boldsymbol{\Omega} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\top & 1 \end{bmatrix} \succeq_{\mathcal{C}^*} \boldsymbol{0}, \end{array}$$

where $\mathcal{L} \subseteq [P]$, and it is implicitly assumed that $\xi_\ell \leq 1,\ \ell \in \mathcal{L}$, for any $\boldsymbol{\xi} \in \mathbb{R}^P_+$ satisfying $\boldsymbol{F}\boldsymbol{\xi} = \boldsymbol{g}$.

We also rely on the following standard result which allows us to represent a scalar integer variable using only logarithmically many binary variables [43].

Lemma 1. *If ξ is a scalar integer decision variable taking values in $\{0, \cdots, 2^Q - 1\}$, with $Q \in \mathbb{Z}_+$, then we can reformulate it concisely by employing Q binary decision variables $\chi_1, \cdots, \chi_Q \in \{0,1\}$, as follows:*

$$\xi = \sum_{q \in [Q]} 2^{q-1} \chi_q = \mathbf{v}_Q^\top \boldsymbol{\chi}.$$

Using Theorem 1 and Lemma 1, we are now ready to state our first result.

Proposition 1. *For any fixed decision $\boldsymbol{x} \in \mathcal{X}$ the optimal value of the i-th quadratic maximization problem in* (12) *coincides with the optimal value of the completely positive program*

$$\begin{array}{lll} Z_i(\boldsymbol{x}) = & \sup & \operatorname{tr}\left(\boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) \boldsymbol{\Omega} \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\right) + \boldsymbol{\mathcal{b}}_i(\boldsymbol{x})^\top \boldsymbol{\xi}' + c_i(\boldsymbol{x}) \\ & \text{s.t.} & \boldsymbol{\xi}' \in \mathbb{R}^{K'}_+, \ \boldsymbol{\Omega} \in \mathbb{S}^{K'}_+ \\ & & \boldsymbol{\mathcal{S}}\boldsymbol{\xi}' = \boldsymbol{t}, \ \operatorname{diag}(\boldsymbol{\mathcal{S}}\boldsymbol{\Omega}\boldsymbol{\mathcal{S}}^\top) = \boldsymbol{t} \circ \boldsymbol{t} \\ & & \xi'_\ell = \Omega_{\ell\ell} \quad \forall \ell \in [LQ] \\ & & \begin{bmatrix} \boldsymbol{\Omega} & \boldsymbol{\xi}' \\ \boldsymbol{\xi}'^\top & 1 \end{bmatrix} \succeq_{\mathcal{C}^*} \boldsymbol{0}, \end{array} \tag{15}$$

where

$$\boldsymbol{\mathcal{S}} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{S} \\ -\mathbf{v}_Q^\top & \cdots & \mathbf{0}^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{e}_1^\top \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}^\top & \cdots & -\mathbf{v}_Q^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{e}_L^\top \\ \mathbb{I} & \cdots & \mathbf{0} & \mathbb{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbb{I} & \mathbf{0} & \cdots & \mathbb{I} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{J' \times K'}, \quad \boldsymbol{t} = \begin{bmatrix} \boldsymbol{t} \\ 0 \\ \vdots \\ 0 \\ \mathbf{e} \\ \vdots \\ \mathbf{e} \end{bmatrix} \in \mathbb{R}^{J'}, \tag{16}$$

$$\boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_i(\boldsymbol{x}) \end{bmatrix} \in \mathbb{R}^{M \times K'} \quad \text{and}$$
$$\boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) = \begin{bmatrix} \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \boldsymbol{b}_i(\boldsymbol{x})^\top \end{bmatrix}^\top \in \mathbb{R}^{K'},$$

with

$$J' = LQ + J + L \quad \text{and} \quad K' = 2LQ + K.$$

Proof. Lemma 1 enables us to reformulate the mixed-integer quadratic program (12) equivalently as the mixed-binary quadratic program

$$\begin{array}{rlll} Z_i(\boldsymbol{x}) = & \sup & \|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) & \\ & \text{s.t.} & \boldsymbol{\xi} \in \mathbb{R}_+^K, \ \boldsymbol{\chi}_\ell \in \{0,1\}^Q & \forall \ell \in [L] \\ & & \boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{t} & \\ & & \xi_\ell = \mathbf{v}_Q^\top \boldsymbol{\chi}_\ell & \forall \ell \in [L]. \end{array} \tag{17}$$

We now employ Theorem 1 to derive the equivalent completely positive program for (17). To this end, we first bring the above quadratic program into the standard form (14). We introduce the redundant linear constraints $\boldsymbol{\chi}_\ell \leq \mathbf{e}$, $\ell \in [L]$, which are pertinent for the exactness of the reformulation, and we define new auxiliary slack variables $\boldsymbol{\eta}_\ell$, $\ell \in [L]$, to transform these inequalities into the equality constraints $\boldsymbol{\chi}_\ell + \boldsymbol{\eta}_\ell = \mathbf{e}$, $\forall \ell \in [L]$. This yields the equivalent problem

$$\begin{array}{rlll} Z_i(\boldsymbol{x}) = & \sup & \|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) & \\ & \text{s.t.} & \boldsymbol{\xi} \in \mathbb{R}_+^K, \ \boldsymbol{\eta} \in \mathbb{R}_+^J, \ \boldsymbol{\chi}_\ell \in \{0,1\}^Q & \forall \ell \in [L] \\ & & \boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{t} & \\ & & \xi_\ell = \mathbf{v}_Q^\top \boldsymbol{\chi}_\ell & \forall \ell \in [L] \\ & & \boldsymbol{\chi}_\ell + \boldsymbol{\eta}_\ell = \mathbf{e} & \forall \ell \in [L]. \end{array} \tag{18}$$

We next define the expanded vector

$$\boldsymbol{\xi}' = \begin{bmatrix} \boldsymbol{\chi}_1^\top & \cdots & \boldsymbol{\chi}_L^\top & \boldsymbol{\eta}_1^\top & \cdots & \boldsymbol{\eta}_L^\top & \boldsymbol{\xi}^\top \end{bmatrix}^\top \in \mathbb{R}_+^{K'}$$

that comprises all decision variables in (18). Together with the augmented parameters (16), we can reformulate (18) concisely as the problem

$$
\begin{array}{rll}
Z_i(\boldsymbol{x}) = & \sup & \|\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})\boldsymbol{\xi}'\|^2 + \boldsymbol{\mathscr{b}}_i(\boldsymbol{x})^\top \boldsymbol{\xi}' + c_i(\boldsymbol{x}) \\
& \text{s.t.} & \boldsymbol{\xi}' \in \mathbb{R}^{K'}_+ \\
& & \boldsymbol{\mathcal{S}}\boldsymbol{\xi}' = \boldsymbol{t} \\
& & \xi'_\ell \in \{0,1\} \qquad\qquad\qquad\qquad \forall \ell \in [LQ]\,.
\end{array}
\tag{19}
$$

The mixed-binary quadratic program (19) already has the desired standard form (14) with inputs $P = K'$, $\boldsymbol{Q} = \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})$, $\boldsymbol{r} = \boldsymbol{\mathscr{b}}_i(\boldsymbol{x})$, $\boldsymbol{F} = \boldsymbol{\mathcal{S}}$, $\boldsymbol{g} = \boldsymbol{t}$, and $\mathcal{L} = [LQ]$. We may thus apply Theorem 1 to obtain the equivalent completely positive program (15). This completes the proof. $\square$

We remark that in view of the concise representation in Lemma 1, the size of the completely positive program (15) remains polynomial in the size of the input data. This completely positive program admits a dual copositive program given by

$$
\begin{array}{rll}
\overline{Z}_i(\boldsymbol{x}) = & \inf & c_i(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \tau \\
& \text{s.t.} & \tau \in \mathbb{R},\ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{J'},\ \boldsymbol{\gamma} \in \mathbb{R}^{LQ} \\
& & \begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi})\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top\ \boldsymbol{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathscr{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top\ \boldsymbol{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathscr{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top\ \boldsymbol{0}^\top]^\top\right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0}.
\end{array}
\tag{20}
$$

By weak conic duality, the optimal value of this copositive program constitutes an upper bound on $Z_i(\boldsymbol{x})$.

Proposition 2. *For any $i \in [I] \cup \{0\}$ and fixed decision $\boldsymbol{x} \in \mathcal{X}$ we have $\overline{Z}_i(\boldsymbol{x}) \geq Z_i(\boldsymbol{x})$.*

3.2 A Copositive Reformulation of Problem (3)

In this section, we demonstrate strong duality for the primal and dual pair (15) and (20), respectively, under the natural boundedness assumption on the uncertainty set Ξ. This exactness result enables us to reformulate the RQCQP (3) equivalently as a copositive program of polynomial size.

Theorem 2 (Strong Duality)**.** *For any $i \in [I] \cup \{0\}$ and fixed decision $\boldsymbol{x} \in \mathcal{X}$ we have $\overline{Z}_i(\boldsymbol{x}) = Z_i(\boldsymbol{x})$.*

We note that the primal completely positive program (15) never has an interior [10]. In order to prove Theorem 2, we construct a Slater point for the dual copositive program (20).

Remark 1. *To our best knowledge, Theorem 2 provides the first strong duality result for the completely positive programming reformulation of the generic mixed-integer quadratic programs* (12)*. This encouraging result has important ramifications beyond the robust optimization setting that we consider in this paper. For instance, one could envisage a new interior point algorithm for the mixed-integer quadratic programs* (12) *that solves the equivalent convex reformulation* (20)*.*

The construction of the Slater point for problem (20) relies on the following two lemmas. To this end, we observe that by construction the boundedness of the uncertainty set Ξ means that the lifted polytope

$$\Xi' = \{\boldsymbol{\xi}' \in \mathbb{R}^{K'} : \boldsymbol{\mathcal{S}}\boldsymbol{\xi}' = \boldsymbol{t},\ \boldsymbol{\xi}' \geq \mathbf{0}\} \tag{21}$$

is also bounded. This gives rise to the first lemma on the strict copositivity of the matrix $\boldsymbol{\mathcal{S}}^\top \boldsymbol{\mathcal{S}}$.

Lemma 2. *We have* $\boldsymbol{\mathcal{S}}^\top \boldsymbol{\mathcal{S}} \succ_{\mathcal{C}} \mathbf{0}$.

Proof. The boundedness assumption implies that the recession cone of the set Ξ' coincides with the point $\mathbf{0}$, that is, $\{\boldsymbol{\xi}' \in \mathbb{R}^{K'}_+ : \boldsymbol{\mathcal{S}}\boldsymbol{\xi}' = \mathbf{0}\} = \{\mathbf{0}\}$. Thus, for every $\boldsymbol{\xi}' \geq \mathbf{0}$, $\boldsymbol{\xi}' \neq \mathbf{0}$, we must have $\boldsymbol{\mathcal{S}}\boldsymbol{\xi}' \neq \mathbf{0}$, which further implies that ${\boldsymbol{\xi}'}^\top \boldsymbol{\mathcal{S}}^\top \boldsymbol{\mathcal{S}}\boldsymbol{\xi}' > 0$ for all $\boldsymbol{\xi}' \geq \mathbf{0}$ such that $\boldsymbol{\xi}' \neq \mathbf{0}$. Hence, the matrix $\boldsymbol{\mathcal{S}}^\top \boldsymbol{\mathcal{S}}$ is strictly copositive. $\square$

The next lemma, which was proven in [26, Lemma 4], constitutes an extension of the Schur complements lemma for matrices with a copositive sub-matrix. We include the proof here to keep the paper self-contained.

Lemma 3 (Copositive Schur Complements)**.** *Consider the symmetric matrix*

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{Q} \\ \boldsymbol{Q}^\top & \boldsymbol{R} \end{bmatrix}.$$

We then have $\boldsymbol{M} \succ_{\mathcal{C}} \mathbf{0}$ *if* $\boldsymbol{R} - \boldsymbol{Q}^\top \boldsymbol{P}^{-1}\boldsymbol{Q} \succ_{\mathcal{C}} \mathbf{0}$ *and* $\boldsymbol{P} \succ \mathbf{0}$.

Proof. Consider a non-negative vector $[\boldsymbol{\xi}^\top\ \boldsymbol{\rho}^\top]^\top \in \mathbb{R}^{P+Q}_+$ satisfying $\mathbf{e}^\top\boldsymbol{\xi} + \mathbf{e}^\top\boldsymbol{\rho} = 1$. We have

$$\begin{aligned} [\boldsymbol{\xi}^\top\ \boldsymbol{\rho}^\top]\boldsymbol{M}[\boldsymbol{\xi}^\top\ \boldsymbol{\rho}^\top]^\top &= \boldsymbol{\xi}^\top\boldsymbol{P}\boldsymbol{\xi} + 2\boldsymbol{\xi}^\top\boldsymbol{Q}\boldsymbol{\rho} + \boldsymbol{\rho}^\top\boldsymbol{R}\boldsymbol{\rho} \\ &= (\boldsymbol{\xi} + \boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{\rho})^\top\boldsymbol{P}(\boldsymbol{\xi} + \boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{\rho}) + \boldsymbol{\rho}^\top(\boldsymbol{R} - \boldsymbol{Q}^\top\boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{\rho} \ \geq\ 0. \end{aligned}$$

The final inequality follows from the assumptions $\boldsymbol{P} \succ \mathbf{0}$, $\boldsymbol{R} - \boldsymbol{Q}^\top\boldsymbol{P}^{-1}\boldsymbol{Q} \succ_{\mathcal{C}} \mathbf{0}$ and $\boldsymbol{\rho} \geq \mathbf{0}$. In fact, the inequality will be strict which can be shown by considering the following two cases:

1. If $\boldsymbol{\rho} = \mathbf{0}$, then $\mathbf{e}^\top\boldsymbol{\xi} = 1$. Therefore $\boldsymbol{\xi} \neq \mathbf{0}$ which implies that $(\boldsymbol{\xi} + \boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{\rho})^\top\boldsymbol{P}(\boldsymbol{\xi} + \boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{\rho}) > 0$.

2. If $\boldsymbol{\rho} \neq \mathbf{0}$, then the assumption $\boldsymbol{R} - \boldsymbol{Q}^\top\boldsymbol{P}^{-1}\boldsymbol{Q} \succ_{\mathcal{C}} \mathbf{0}$ implies that $\boldsymbol{\rho}^\top(\boldsymbol{R} - \boldsymbol{Q}^\top\boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{\rho} > 0$.

Therefore, in both cases, by rescaling we have $[\boldsymbol{\xi}^\top\ \boldsymbol{\rho}^\top]\boldsymbol{M}[\boldsymbol{\xi}^\top\ \boldsymbol{\rho}^\top]^\top > 0$ for all $[\boldsymbol{\xi}^\top\ \boldsymbol{\rho}^\top]^\top \in \mathbb{R}^{P+Q}_+$ such that $[\boldsymbol{\xi}^\top\ \boldsymbol{\rho}^\top]^\top \neq \mathbf{0}$. Hence, $\boldsymbol{M} \succ_{\mathcal{C}} \mathbf{0}$. $\square$

Using Lemmas 2 and 3, we are now ready to prove the main strong duality result.

Proof of Theorem 2. We construct a Slater point $(\tau, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma})$ for problem (20). Specifically, we set $\boldsymbol{\gamma} = \mathbf{0}$, $\boldsymbol{\psi} = \mathbf{0}$, and $\boldsymbol{\phi} = \rho\mathbf{e}$ for some $\rho > 0$. Problem (20) then admits a Slater point if there exist scalars $\rho, \tau > 0$, such that

$$\begin{bmatrix} \rho\boldsymbol{\mathcal{S}}^\top\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) & -\frac{1}{2}\boldsymbol{\mathscr{b}}_i(\boldsymbol{x}) \\ -\frac{1}{2}\boldsymbol{\mathscr{b}}_i(\boldsymbol{x})^\top & \tau \end{bmatrix} \succ_{\mathcal{C}} \mathbf{0}. \tag{22}$$

Lemma 2 implies that for a sufficiently large ρ the matrix $\rho\boldsymbol{\mathcal{S}}^\top\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})$ is strictly copositive. Thus, we can choose a positive τ to ensure that

$$\rho\boldsymbol{\mathcal{S}}^\top\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) - \frac{1}{4\tau}\boldsymbol{\mathcal{b}}_i(\boldsymbol{x})\boldsymbol{\mathcal{b}}_i(\boldsymbol{x})^\top \succ_{\mathcal{C}} \mathbf{0}.$$

Using Lemma 3, we may conclude that the strict copositivity constraint in (22) is satisfied by the constructed solution $(\tau, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma})$. Thus, problem (20) admits a Slater point and strong duality indeed holds for the primal and dual pair (15) and (20), respectively. $\square$

The exactness result portrayed in Theorem 2 enables us to derive the equivalent copositive programming reformulation for (3).

Theorem 3. *The RQCQP* (3) *is equivalent to the following copositive program.*

$$\begin{array}{lll}
\text{minimize} & c_0(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\psi}_0 + (\boldsymbol{t}\circ\boldsymbol{t})^\top\boldsymbol{\phi}_0 + \tau_0 & \\
\text{subject to} & \boldsymbol{x}\in\mathcal{X},\ \tau_i\in\mathbb{R},\ \boldsymbol{\psi}_i,\boldsymbol{\phi}_i\in\mathbb{R}^{J'},\ \boldsymbol{\gamma}_i\in\mathbb{R}^{LQ},\ \boldsymbol{H}_i\in\mathbb{S}_+^{K'} & \forall i\in[I]\cup\{0\} \\
& c_i(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\psi}_i + (\boldsymbol{t}\circ\boldsymbol{t})^\top\boldsymbol{\phi}_i + \tau_i \le 0 & \forall i\in[I] \\
& \begin{bmatrix} \mathbb{I} & \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) \\ \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top & \boldsymbol{H}_i \end{bmatrix} \succeq \mathbf{0} & \forall i\in[I]\cup\{0\} \\
& \begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}_i)\boldsymbol{\mathcal{S}} - \boldsymbol{H}_i - \operatorname{diag}\left([\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right)^\top & \tau_i \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0} & \forall i\in[I]\cup\{0\}
\end{array} \tag{23}$$

The proof of Theorem 3 relies on the following lemma which linearizes the quadratic term $\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})$ in the left-hand side matrix of problem (20).

Lemma 4. *Let $\boldsymbol{M}\in\mathbb{S}^R$ be a symmetric matrix and $\boldsymbol{A}\in\mathbb{R}^{P\times Q}$ be an arbitrary matrix with $Q\le R$. Then the copositive inequality*

$$\boldsymbol{M} \succeq_{\mathcal{C}} \begin{bmatrix} \boldsymbol{A}^\top\boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{24}$$

is satisfied if and only if there exists a positive semidefinite matrix $\boldsymbol{H}\in\mathbb{S}_+^Q$ such that

$$\boldsymbol{M} \succeq_{\mathcal{C}} \begin{bmatrix} \boldsymbol{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbb{I} & \boldsymbol{A} \\ \boldsymbol{A}^\top & \boldsymbol{H} \end{bmatrix} \succeq \mathbf{0}. \tag{25}$$

Proof. The only if statement is satisfied immediately by setting $\boldsymbol{H} = \boldsymbol{A}^\top\boldsymbol{A}$. To prove the converse statement, assume that there exists such a positive semidefinite matrix $\boldsymbol{H}\in\mathbb{S}_+^Q$. Then by the Schur complement the semidefinite inequality in (25) implies that $\boldsymbol{H}\succeq\boldsymbol{A}^\top\boldsymbol{A}$ and, *a fortiori*, $\boldsymbol{H}\succeq_{\mathcal{C}}\boldsymbol{A}^\top\boldsymbol{A}$. Combining this with the copositive inequality in (25) then yields (24). Thus, the claim follows. $\square$

Proof of 3. Applying Theorem 2, we may replace the quadratic maximization problem in the i-th constraint of (3) with a copositive minimization problem. This yields the equivalent constraint

$$
\begin{array}{rll}
0 \geq & \inf & c_i(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi}_i + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi}_i + \tau_i \\
& \text{s.t.} & \tau_i \in \mathbb{R},\ \boldsymbol{\psi}_i, \boldsymbol{\phi}_i \in \mathbb{R}^{J'},\ \boldsymbol{\gamma}_i \in \mathbb{R}^{LQ} \\
& & \begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}_i)\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right)^\top & \tau_i \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0},
\end{array}
$$

which is satisfied if and only if there exist decision variables $\tau_i \in \mathbb{R}$, $\boldsymbol{\psi}_i$, $\boldsymbol{\phi}_i \in \mathbb{R}^{J'}$, and $\boldsymbol{\gamma}_i \in \mathbb{R}^{LQ}$ such that the constraint system

$$
\begin{array}{l}
c_i(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi}_i + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi}_i + \tau_i \leq 0 \qquad \text{and} \\
\begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}_i)\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top\ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right)^\top & \tau_i \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0}
\end{array}
$$

is satisfied. By replacing the objective function of (3) with the corresponding copositive reformulation, we thus find that problem (3) is equivalent to

$$
\begin{array}{lll}
\text{minimize} & c_0(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi}_0 + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi}_0 + \tau_0 & \\
\text{subject to} & \boldsymbol{x} \in \mathcal{X},\ \tau_i \in \mathbb{R},\ \boldsymbol{\psi}_i, \boldsymbol{\phi}_i \in \mathbb{R}^{J'},\ \boldsymbol{\gamma}_i \in \mathbb{R}^{LQ} & \forall i \in [I] \cup \{0\} \\
& c_i(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi}_i + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi}_i + \tau_i \leq 0 & \forall i \in [I] \\
& \begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}_i)\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right)^\top & \tau_i \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0} & \\
& & \forall i \in [I] \cup \{0\}.
\end{array}
$$

Next, we apply Lemma 4 to linearize the quadratic terms $\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})$, $i \in [I] \cup \{0\}$, which gives rise to the desired copositive program (23). This completes the proof. $\square$

Remark 2. *All exactness results in this paper extend immediately to the following setting where the objective and constraint functions in* (3) *involve non-convex quadratic terms in the uncertainty* $\boldsymbol{\xi}$.

$$
Z_i(\boldsymbol{x}) = \sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{\xi}^\top \boldsymbol{D}_i(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) \qquad \forall i \in [I] \cup \{0\}
$$

Here, $\boldsymbol{D}_i(\boldsymbol{x}) : \mathcal{X} \to \mathbb{S}^K$, $i \in [I] \cup \{0\}$, *are matrix-valued affine functions in* $\boldsymbol{x}$. *In this case, the copositive programming reformulation is obtained by replacing the last constraint system in* (23) *with the copositive constraints*

$$
\begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}_i)\boldsymbol{\mathcal{S}} - \boldsymbol{H}_i - \boldsymbol{\mathcal{D}}_i(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top\ \mathbf{0}^\top]^\top\right)^\top & \tau_i \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0} \quad \forall i \in [I] \cup \{0\},
$$

where

$$\boldsymbol{\mathcal{D}}_i(\boldsymbol{x}) = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{D}_i(\boldsymbol{x}) \end{bmatrix} \in \mathbb{S}^{K'} \qquad \forall i \in [I] \cup \{0\}.$$

3.3 Extension to the Two-Stage Robust Optimization Setting

In this section, we briefly study the generalized two-stage robust optimization problem of the form

$$\begin{array}{ll} \text{minimize} & \sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}. \end{array} \tag{26}$$

Here, for any fixed decision $\boldsymbol{x} \in \mathcal{X}$ and uncertain parameter realization $\boldsymbol{\xi} \in \Xi$, the second-stage cost $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi})$ coincides with the optimal value of the convex quadratic program given by

$$\begin{array}{rll} \mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi}) = & \inf & \|\boldsymbol{P}\boldsymbol{y}\|^2 + (\boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{r})^\top \boldsymbol{y} \\ & \text{s.t.} & \boldsymbol{y} \in \mathbb{R}^{D_2} \\ & & \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{W}\boldsymbol{y}, \end{array} \tag{27}$$

where $\boldsymbol{T}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^{T \times K}$ and $\boldsymbol{h}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^T$ are matrix- and vector-valued affine functions, respectively.

Example 3 (Support Vector Machines with Noisy Labels)**.** *Consider the following soft-margin support vector machines (SVM) model for data classification.*

$$\begin{array}{ll} \text{minimize} & \lambda\|\boldsymbol{w}\|^2 + \sum_{n \in [N]} \max\left\{0, 1 - \hat{\xi}_n(\boldsymbol{w}^\top \hat{\boldsymbol{\chi}}_n - w_0)\right\} \\ \text{subject to} & \boldsymbol{w} \in \mathbb{R}^K,\ w_0 \in \mathbb{R} \end{array} \tag{28}$$

Here, for every index $n \in [N]$, the vector $\hat{\boldsymbol{\chi}}_n \in \mathbb{R}^K$ is a data point that has been labeled as $\hat{\xi}_n \in \{-1, 1\}$. The objective of problem (28) is to find a hyperplane $\{\boldsymbol{\chi} \in \mathbb{R}^K : \boldsymbol{w}^\top \boldsymbol{\chi} = w_0\}$ that separates all points labeled $+1$ with the ones labeled -1. If the hyperplane satisfies $\hat{\xi}_n(\boldsymbol{w}^\top \hat{\boldsymbol{\chi}}_n - w_0) > 1$, $n \in [N]$, then the data points are linearly separable. In practice, however, these data points may not be linearly separable. We thus seek for the best linear separator that minimizes the number of incorrect classifications. This non-convex objective is captured by employing the hinge loss term $\sum_{n \in [N]} \max\left\{0, 1 - \hat{\xi}_n(\boldsymbol{w}^\top \hat{\boldsymbol{\chi}}_n - w_0)\right\}$ in (28) as a convex surrogate. Here, the term $\lambda\|\boldsymbol{w}\|^2$ in the objective function constitutes a regularizer for the coefficent $\boldsymbol{w}$.

If the labels $\{\hat{\xi}_n\}_{n \in [N]}$ are erroneous, then one could envisage a robust optimization model that seeks for the best linear separator in view the most adverse realization of the labels. To this end, we assume that the vector of labels $\boldsymbol{\xi}$ is only known to reside in a prescribed binary uncertainty set $\Xi \subseteq \{-1, 1\}^N$. Then, an

SVM model that is robust against uncertainty in the labels can be formulated as

$$\begin{array}{ll}\text{minimize} & \lambda\|\boldsymbol{w}\|^2 + \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{R}(\boldsymbol{w}, w_0, \boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{w} \in \mathbb{R}^K,\ w_0 \in \mathbb{R},\end{array}$$

where

$$\mathcal{R}(\boldsymbol{w}, w_0, \boldsymbol{\xi}) = \begin{array}[t]{ll}\inf & \mathbf{e}^\top \boldsymbol{y} \\ \text{s.t.} & \boldsymbol{y} \in \mathbb{R}^N_+ \\ & y_n \geq 1 - \xi_n(\boldsymbol{w}^\top \hat{\boldsymbol{\chi}}_n - w_0) \quad \forall n \in [N].\end{array} \tag{29}$$

This problem constitutes an instance of (26) *with the decision vector* $\boldsymbol{x} = (\boldsymbol{w}, w_0)$*, and the input parameters*

$$\begin{array}{l} I{=}0, \quad \boldsymbol{A}(\boldsymbol{x}) = \mathbf{0}, \quad \boldsymbol{b}(\boldsymbol{x}) = \mathbf{0}, \quad c(\boldsymbol{x}) = \lambda\|\boldsymbol{w}\|, \quad \boldsymbol{P} = \mathbf{0}, \quad \boldsymbol{R} = \mathbf{0}, \quad \boldsymbol{r} = \mathbf{e}, \\ \boldsymbol{T}(\boldsymbol{x}) = -\operatorname{diag}\left(\begin{bmatrix} \boldsymbol{w}^\top \hat{\boldsymbol{\chi}}_1 \\ \vdots \\ \boldsymbol{w}^\top \hat{\boldsymbol{\chi}}_N \end{bmatrix}\right) - w_0\mathbb{I}, \quad \boldsymbol{h}(\boldsymbol{x}) = \mathbf{e}, \quad \textit{and} \quad \boldsymbol{W} = \mathbb{I}. \end{array}$$

The exactness result portrayed in Theorems 2 and 3 can be extended to the two-stage robust optimization problem (26). Specifically, if the problem has a *complete recourse*[1] then, by employing Theorem 2 and extending the techniques developed in [26, Theorem 4], the two-stage problem (26) can be reformulated as a copositive program of polynomial size.

Theorem 4. *Assume that* $\boldsymbol{P}$ *has full column rank. Then the two-stage robust optimization problem* (26) *is equivalent to the copositive program*

$$\begin{array}{ll}\text{minimize} & c(\boldsymbol{x}) - \dfrac{1}{4}\boldsymbol{r}^\top(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{r} + \boldsymbol{t}^\top\boldsymbol{\psi} + (\boldsymbol{t}\circ\boldsymbol{t})^\top\boldsymbol{\phi} + \tau \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X},\ \tau \in \mathbb{R},\ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{J'},\ \boldsymbol{\gamma} \in \mathbb{R}^{LQ},\ \boldsymbol{H} \in \mathbb{S}^{K'}_+ \\ & \begin{bmatrix} \mathbb{I} & \boldsymbol{\mathcal{A}}(\boldsymbol{x}) \\ \boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top & \boldsymbol{H} \end{bmatrix} \succeq \mathbf{0} \\ & \begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi})\boldsymbol{\mathcal{S}} - \boldsymbol{H} - \boldsymbol{\mathcal{P}}(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top\ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\psi} - \boldsymbol{\mathcal{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top\ \mathbf{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\psi} - \boldsymbol{\mathcal{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top\ \mathbf{0}^\top]^\top\right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0},\end{array} \tag{30}$$

[1]The two-stage problem (26) has complete recourse if there exists $\boldsymbol{y}^+ \in \mathbb{R}^{D_2}$ with $\boldsymbol{W}\boldsymbol{y}^+ > \mathbf{0}$, which implies that the second-stage subproblem is feasible for every $\boldsymbol{x} \in \mathbb{R}^{D_1}$ and $\boldsymbol{\xi} \in \mathbb{R}^K$.

where

$$\boldsymbol{\mathcal{S}} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{S} & \mathbf{0} \\ -\mathbf{v}_Q^\top & \cdots & \mathbf{0}^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{e}_1^\top & \mathbf{0}^\top \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}^\top & \cdots & -\mathbf{v}_Q^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{e}_L^\top & \mathbf{0}^\top \\ \mathbb{I} & \cdots & \mathbf{0} & \mathbb{I} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbb{I} & \mathbf{0} & \cdots & \mathbb{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{J' \times K'}, \quad \boldsymbol{t} = \begin{bmatrix} \boldsymbol{t} \\ 0 \\ \vdots \\ 0 \\ \mathbf{e} \\ \vdots \\ \mathbf{e} \end{bmatrix} \in \mathbb{R}^{J'},$$

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & -\frac{1}{4}\boldsymbol{R}^\top(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{R} & \frac{1}{2}\left(\boldsymbol{T}(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{W}(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{R}\right)^\top \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \frac{1}{2}\left(\boldsymbol{T}(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{W}(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{R}\right) & -\frac{1}{4}\boldsymbol{W}(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{W}^\top \end{bmatrix} \in \mathbb{S}^{K'},$$

$$\boldsymbol{\mathcal{A}}(\boldsymbol{x}) = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}(\boldsymbol{x}) & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{M \times K'}, \quad \textit{and}$$

$$\boldsymbol{\mathcal{b}}(\boldsymbol{x}) = \begin{bmatrix} \mathbf{0}^\top & \cdots & \mathbf{0}^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top & (\boldsymbol{b}(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{R}^\top(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{r})^\top & (\boldsymbol{h}(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{W}(\boldsymbol{P}^\top\boldsymbol{P})^{-1}\boldsymbol{r})^\top \end{bmatrix}^\top \in \mathbb{R}^{K'},$$

with

$$J' = LQ + J + L \quad \textit{and} \quad K' = 2LQ + K + T.$$

Proof. Since $\boldsymbol{P}$ has full column rank, the matrix $\boldsymbol{P}^\top\boldsymbol{P}$ is positive definite. Thus, for any fixed $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$, the recourse problem (27) admits a dual quadratic program given by

$$\begin{array}{rll} \mathcal{R}(\boldsymbol{x},\boldsymbol{\xi}) = & \sup & -\dfrac{1}{4}\left((\boldsymbol{W}^\top\boldsymbol{\theta} - \boldsymbol{R}\boldsymbol{\xi} - \boldsymbol{r})^\top(\boldsymbol{P}^\top\boldsymbol{P})^{-1}(\boldsymbol{W}^\top\boldsymbol{\theta} - \boldsymbol{R}\boldsymbol{\xi} - \boldsymbol{r})\right) + \boldsymbol{h}(\boldsymbol{x})^\top\boldsymbol{\theta} + \boldsymbol{\xi}^\top\boldsymbol{T}(\boldsymbol{x})^\top\boldsymbol{\theta} \\ & \text{s.t.} & \boldsymbol{\theta} \in \mathbb{R}^T_+. \end{array} \tag{31}$$

Strong duality holds as the two-stage problem (26) has complete recourse. Substituting the dual formulation (31) into the objective of (26) yields

$$\begin{aligned} & \sup_{\boldsymbol{\xi}\in\Xi} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top\boldsymbol{\xi} + c(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x},\boldsymbol{\xi}) \\ = \quad & \sup_{\boldsymbol{\xi}\in\Xi,\boldsymbol{\theta}\in\mathbb{R}^T_+} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top\boldsymbol{\xi} + c(\boldsymbol{x}) - \frac{1}{4}\left((\boldsymbol{W}^\top\boldsymbol{\theta} - \boldsymbol{R}\boldsymbol{\xi} - \boldsymbol{r})^\top(\boldsymbol{P}^\top\boldsymbol{P})^{-1}(\boldsymbol{W}^\top\boldsymbol{\theta} - \boldsymbol{R}\boldsymbol{\xi} - \boldsymbol{r})\right) \\ & \qquad\qquad + \boldsymbol{h}(\boldsymbol{x})^\top\boldsymbol{\theta} + \boldsymbol{\xi}^\top\boldsymbol{T}(\boldsymbol{x})\boldsymbol{\theta}. \end{aligned}$$

Thus, for any fixed $\boldsymbol{x} \in \mathcal{X}$, the objective value of the two-stage problem (26) coincides with the optimal value of a quadratic maximization problem, which is amenable to an exact completely positive programming reformulation similar to the one derived in Proposition 1. We can then follow the same steps taken in the proofs of Theorems 2 and 3 to obtain the equivalent copositive program (23). This completes the proof. $\square$

Remark 3. *The assumption that $\boldsymbol{P}$ has full column rank in Theorem 4 can be relaxed. If $\boldsymbol{P}$ does not have full column rank then the symmetric matrix $\boldsymbol{P}^\top \boldsymbol{P}$ is not positive definite but admits the eigendecomposition $\boldsymbol{P}^\top \boldsymbol{P} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}$, where $\boldsymbol{U}$ is an orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{P}^\top \boldsymbol{P}$, while $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues of $\boldsymbol{P}^\top \boldsymbol{P}$ on its main diagonal. We assume without loss of generality that the matrix $\boldsymbol{\Lambda}$ has the block diagonal form*

$$\begin{bmatrix} \boldsymbol{\Lambda}_+ & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where $\boldsymbol{\Lambda}_+$ is a diagonal matrix whose main diagonal comprises the non-zero eigenvalues of $\boldsymbol{P}^\top \boldsymbol{P}$. Next, by using the constructed eigendecomposition and performing the change of variable $\boldsymbol{z} \leftarrow \boldsymbol{U}^{-1}\boldsymbol{y}$, we can reformulate the recourse problem (27) *equivalently as*

$$\begin{aligned} \mathcal{R}(\boldsymbol{x},\boldsymbol{\xi}) = \quad & \inf \quad \boldsymbol{z}_+^\top \boldsymbol{\Lambda}_+ \boldsymbol{z}_+ + (\boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{r})^\top \boldsymbol{U}_+ \boldsymbol{z}_+ + (\boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{r})^\top \boldsymbol{U}_0 \boldsymbol{z}_0 \\ & \text{s.t.} \quad (\boldsymbol{z}_+, \boldsymbol{z}_0) \in \mathbb{R}^{D_2} \\ & \qquad \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{W}\boldsymbol{U}_+ \boldsymbol{z}_+ + \boldsymbol{W}\boldsymbol{U}_0 \boldsymbol{z}_0, \end{aligned}$$

where $\boldsymbol{U} = [\boldsymbol{U}_+ \quad \boldsymbol{U}_0]$ and $\boldsymbol{z} = [\boldsymbol{z}_+^\top \quad \boldsymbol{z}_0^\top]^\top$. The dual of this problem is given by the following quadratic program with a linear constraint system.

$$\begin{aligned} \mathcal{R}(\boldsymbol{x},\boldsymbol{\xi}) = \quad & \sup \quad -\frac{1}{4}\left((\boldsymbol{W}^\top \boldsymbol{\theta} - \boldsymbol{R}\boldsymbol{\xi} - \boldsymbol{r})^\top \boldsymbol{U}_+^\top \boldsymbol{\Lambda}_+^{-1} \boldsymbol{U}_+ (\boldsymbol{W}^\top \boldsymbol{\theta} - \boldsymbol{R}\boldsymbol{\xi} - \boldsymbol{r})\right) + \boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\theta} + \boldsymbol{\xi}^\top \boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\theta} \\ & \text{s.t.} \quad \boldsymbol{\theta} \in \mathbb{R}_+^T \\ & \qquad \boldsymbol{U}_0^\top (\boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{r}) = \boldsymbol{U}_0^\top \boldsymbol{W}^\top \boldsymbol{\theta} \end{aligned}$$

We can then repeat the same steps in the proof of Theorem 4 to obtain an equivalent copositive programming reformulation. We omit this result for the sake of brevity.

4 Conservative Semidefinite Programming Approximation

Due to the equivalence with generic RQCQPs over a polyhedral uncertainty set, the copositive program (23) is intractable to solve [4]. In the new reformulation, however, all the difficulty of the original problem (3) is shifted into the copositive cone $\mathcal{C}$ which has been well studied in the literature. Specifically, there exists a hierarchy of increasingly tight semidefinite representable inner approximations that converge in finitely many iterations to $\mathcal{C}$ [36, 8, 14, 29]. The simplest of these approximations is given by the cone

$$\mathcal{C}^0 = \left\{ \boldsymbol{M} \in \mathbb{S}^K : \boldsymbol{M} = \boldsymbol{P} + \boldsymbol{N},\ \boldsymbol{P} \succeq \mathbf{0},\ \boldsymbol{N} \geq \mathbf{0} \right\},$$

which contains all symmetric matrices that can be decomposed into a sum of positive semidefinite and non-negative matrices. For dimensions $K \leq 4$ it can be shown that $\mathcal{C}^0 = \mathcal{C}$ [16], while for $K > 4$, C^0 is a strict subset of $\mathcal{C}$.

Replacing the cone $\mathcal{C}$ in (23) with the inner approximation $\mathcal{C}^0$ gives rise to a tractable conservative approximation for the RQCQP (3). In this case, however, the resulting optimization problem might have no interior or even become infeasible as the Slater point that we have constructed in Theorem 2 can fail to be a Slater point to the restricted problem. Indeed, the strict copositivity of the matrix $\boldsymbol{\mathcal{S}}^\top \boldsymbol{\mathcal{S}}$ is in general insufficient to ensure that the matrix is also strictly positive definite. To remedy this shortcoming, we suggest the following simple modification to the primal completely positive formulation of $Z_i(\boldsymbol{x})$ in (15). Specifically, we assume that there exists a non-degenerate ellipsoid centered at $\boldsymbol{c} \in \mathbb{R}_+^{K'}$ with radius $r \in \mathbb{R}_{++}$ and shape parameter $\boldsymbol{Q} \in \mathbb{S}_{++}^{K'}$ given by

$$\mathcal{B}(r, \boldsymbol{Q}, \boldsymbol{c}) = \left\{ \boldsymbol{\xi}' \in \mathbb{R}^{K'} : \|\boldsymbol{Q}(\boldsymbol{\xi}' - \boldsymbol{c})\| \leq r \right\}$$

that contains the lifted set Ξ' in (21). We then consider the following augmented completely positive programming reformulation for the i-th maximization problem in (12).

$$\begin{array}{rll} Z_i(\boldsymbol{x}) = & \sup & \operatorname{tr}\left(\boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) \boldsymbol{\Omega} \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\right) + \boldsymbol{\mathcal{b}}_i(\boldsymbol{x})^\top \boldsymbol{\xi}' + c_i(\boldsymbol{x}) \\ & \text{s.t.} & \boldsymbol{\xi}' \in \mathbb{R}_+^{K'},\ \boldsymbol{\Omega} \in \mathbb{S}_+^{K'} \\ & & \boldsymbol{\mathcal{S}} \boldsymbol{\xi}' = \boldsymbol{t},\ \operatorname{diag}(\boldsymbol{\mathcal{S}} \boldsymbol{\Omega} \boldsymbol{\mathcal{S}}^\top) = \boldsymbol{t} \circ \boldsymbol{t} \\ & & \xi'_\ell = \Omega_{\ell\ell} \quad \forall \ell \in [LQ] \\ & & \operatorname{tr}\left(\boldsymbol{Q} \boldsymbol{\Omega} \boldsymbol{Q}^\top\right) - 2\boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{\xi}' + \boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{c} \leq r^2 \\ & & \begin{bmatrix} \boldsymbol{\Omega} & \boldsymbol{\xi}' \\ \boldsymbol{\xi}'^\top & 1 \end{bmatrix} \succeq_{\mathcal{C}^*} \mathbf{0} \end{array} \tag{32}$$

Here, we have added the redundant constraint $\operatorname{tr}\left(\boldsymbol{Q} \boldsymbol{\Omega} \boldsymbol{Q}^\top\right) - 2\boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{\xi}' + \boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{c} \leq r^2$ to (15), which arises from linearizing the quadratic constraint

$$\|\boldsymbol{Q}(\boldsymbol{\xi}' - \boldsymbol{c})\|^2 = \operatorname{tr}\left(\boldsymbol{Q} \boldsymbol{\xi}' \boldsymbol{\xi}'^\top \boldsymbol{Q}^\top\right) - 2\boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{\xi}' + \boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{c} \leq r^2,$$

where we have set $\boldsymbol{\Omega} = \boldsymbol{\xi}' \boldsymbol{\xi}'^\top$. The dual of the augmented problem (32) is given by the following copositive program.

$$\begin{array}{rll} \overline{Z}_i(\boldsymbol{x}) = & \inf & c_i(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \lambda r^2 - \lambda \|\boldsymbol{Q} \boldsymbol{c}\|^2 + \tau \\ & \text{s.t.} & \tau \in \mathbb{R},\ \lambda \in \mathbb{R}_+,\ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{J'},\ \boldsymbol{\gamma} \in \mathbb{R}^{LQ},\ \boldsymbol{h} \in \mathbb{R}^{K'} \\ & & \begin{bmatrix} \lambda \boldsymbol{Q}^\top \boldsymbol{Q} + \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}) \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top\ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\boldsymbol{h} \\ \frac{1}{2}\boldsymbol{h}^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0} \\ & & \boldsymbol{h} = \boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top\ \mathbf{0}^\top]^\top - 2\lambda \boldsymbol{Q}^\top \boldsymbol{Q} c \end{array} \tag{33}$$

Note that we have $Z_i(\boldsymbol{x}) = \overline{Z}_i(\boldsymbol{x})$ since all the new additional terms are redundant for the original reformulations. Nevertheless, since the ellipsoid $\mathcal{B}(r, \boldsymbol{Q}, \boldsymbol{c})$ is non-degenerate, we find that the matrix $\boldsymbol{Q}^\top \boldsymbol{Q}$ is positive definite. We can thus set all eigenvalues of the scaled matrix $\lambda \boldsymbol{Q}^\top \boldsymbol{Q}$ to any arbitrarily large positive values by controlling the scalar $\lambda \in \mathbb{R}_+$. This suggests that replacing the cone $\mathcal{C}$ with its inner approximation $\mathcal{C}^0$ in (33) will always yield a problem with a Slater point.

4.1 Comparison with the Approximate $\mathcal{S}$-Lemma Method

In this section, we consider the case where the bounded uncertainty set contains no integral terms and is given by the simple polytope $\Xi = \left\{\boldsymbol{\xi} \in \mathbb{R}^K_+ : \boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{t}\right\}$. Here, the extended parameters (16) simplify to

$$\boldsymbol{\mathcal{S}} = \boldsymbol{S}, \quad \boldsymbol{t} = \boldsymbol{t}, \quad \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) = \boldsymbol{A}_i(\boldsymbol{x}), \quad \text{and} \quad \boldsymbol{\mathscr{b}}_i(\boldsymbol{x}) = \boldsymbol{b}_i(\boldsymbol{x}),$$

while the i-th maximization problem in (12) reduces to

$$Z_i(\boldsymbol{x}) = \sup_{\boldsymbol{\xi}\in\Xi} \|\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top\boldsymbol{\xi} + c_i(\boldsymbol{x}). \tag{34}$$

The copositive programming reformulation (33) can then be simplified to

$$\begin{array}{rll} \overline{Z}_i(\boldsymbol{x}) = & \inf & c_i(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\psi} + (\boldsymbol{t}\circ\boldsymbol{t})^\top\boldsymbol{\phi} + \lambda r^2 - \lambda\|\boldsymbol{Q}\boldsymbol{c}\|^2 + \tau \\ & \text{s.t.} & \tau \in \mathbb{R}, \ , \ \lambda \in \mathbb{R}_+, \ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^J \\ & & \begin{bmatrix} \lambda\boldsymbol{Q}^\top\boldsymbol{Q} + \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi})\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\psi} - \boldsymbol{\mathscr{b}}_i(\boldsymbol{x}) - 2\lambda\boldsymbol{Q}^\top\boldsymbol{Q}c\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\psi} - \boldsymbol{\mathscr{b}}_i(\boldsymbol{x}) - 2\lambda\boldsymbol{Q}^\top\boldsymbol{Q}c\right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0}. \end{array} \tag{35}$$

Replacing the cone $\mathcal{C}$ in (35) with its inner approximation $\mathcal{C}^0$, we obtain a tractable SDP reformulation whose optimal value $\overline{Z}_i^{\mathcal{C}_0}(\boldsymbol{x})$ constitutes an upper bound on $Z_i(\boldsymbol{x})$. Alternatively, there is also a well-known conservative SDP approximation for (34) called the approximate $\mathcal{S}$-lemma method.

Proposition 3 (Approximate $\mathcal{S}$-lemma Method [4])**.** *Assume that the uncertainty set is a bounded polytope and there is an ellipsoid centered at $\boldsymbol{c} \in \mathbb{R}^K_+$ of radius r given by $\mathcal{B}(r, \boldsymbol{Q}, \boldsymbol{c}) = \{\boldsymbol{\xi} \in \mathbb{R}^K : \|\boldsymbol{Q}(\boldsymbol{\xi} - \boldsymbol{c})\| \leq r\}$ that contains the set Ξ. Then, for any fixed $\boldsymbol{x} \in \mathcal{X}$, the i-th maximization problem in* (12) *is upper bounded by the optimal value of the following semidefinite program.*

$$\begin{array}{rll} \overline{Z}_i^{\mathcal{S}}(\boldsymbol{x}) = & \inf & c_i(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\theta} + \rho r^2 - \rho\|\boldsymbol{Q}\boldsymbol{c}\|^2 + \kappa \\ & \text{s.t.} & \kappa \in \mathbb{R}, \ \rho \in \mathbb{R}_+, \ \boldsymbol{\theta} \in \mathbb{R}^J, \ \boldsymbol{\eta} \in \mathbb{R}^J_+ \\ & & \begin{bmatrix} \rho\boldsymbol{Q}^\top\boldsymbol{Q} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\theta} - \boldsymbol{\mathscr{b}}_i(\boldsymbol{x}) - \boldsymbol{\eta} - 2\rho\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{c}\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\theta} - \boldsymbol{\mathscr{b}}_i(\boldsymbol{x}) - \boldsymbol{\eta} - 2\rho\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{c}\right)^\top & \kappa \end{bmatrix} \succeq \boldsymbol{0} \end{array} \tag{36}$$

Proof. The quadratic maximization problem in (34) can be equivalently reformulated as

$$\begin{array}{rll} Z_i(\boldsymbol{x}) = & \sup & \|\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{\mathscr{b}}_i(\boldsymbol{x})^\top\boldsymbol{\xi} + c_i(\boldsymbol{x}) \\ & \text{s.t.} & \boldsymbol{\xi} \in \mathbb{R}^K_+ \\ & & \boldsymbol{\mathcal{S}}\boldsymbol{\xi} = t \\ & & \|\boldsymbol{Q}(\boldsymbol{\xi} - \boldsymbol{c})\|^2 \leq r^2. \end{array}$$

Here, the last constraint is added without loss generality since $\Xi \subseteq \mathcal{B}(r, \boldsymbol{Q}, \boldsymbol{c})$. Reformulating the problem into its Lagrangian form then yields

$$
\begin{aligned}
& Z_i(\boldsymbol{x}) \\
= \quad & \sup_{\boldsymbol{\xi}} \quad \inf_{\boldsymbol{\eta}\geq \mathbf{0},\rho\geq 0,\boldsymbol{\theta}} \ \|\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{\mathcal{b}}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\theta} - \boldsymbol{\xi}^\top\boldsymbol{\mathcal{S}}^\top\boldsymbol{\theta} + \boldsymbol{\xi}^\top\boldsymbol{\eta} + \rho r^2 - \rho\|\boldsymbol{Q}(\boldsymbol{\xi}-\boldsymbol{c})\|^2 \\
\leq \quad & \inf_{\boldsymbol{\eta}\geq \mathbf{0},\rho\geq 0,\boldsymbol{\theta}} \quad \sup_{\boldsymbol{\xi}} \ \|\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{\mathcal{b}}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\theta} - \boldsymbol{\xi}^\top\boldsymbol{\mathcal{S}}^\top\boldsymbol{\theta} + \boldsymbol{\xi}^\top\boldsymbol{\eta} + \rho r^2 - \rho\|\boldsymbol{Q}(\boldsymbol{\xi}-\boldsymbol{c})\|^2 \\
= \quad & \inf_{\boldsymbol{\eta}\geq \mathbf{0},\rho\geq 0,\boldsymbol{\theta}} \quad c_i(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\theta} + \rho r^2 - \rho\|\boldsymbol{Q}\boldsymbol{c}\|^2 \\
& \qquad\qquad + \sup_{\boldsymbol{\xi}} \left(\|\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{\mathcal{b}}_i(\boldsymbol{x})^\top \boldsymbol{\xi} - \boldsymbol{\xi}^\top\boldsymbol{\mathcal{S}}^\top\boldsymbol{\theta} + \boldsymbol{\xi}^\top\boldsymbol{\eta} - \rho\|\boldsymbol{Q}\boldsymbol{\xi}\|^2 + 2\rho\boldsymbol{\xi}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{c}\right),
\end{aligned}
$$

where the inequality follows from the weak Lagrangian duality. We next introduce an epigraphical variable κ that shifts the supremum in the objective function into the constraint. We have

$$
\begin{aligned}
Z_i(\boldsymbol{x}) \leq \ \ & \inf \quad c_i(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\theta} + \rho r^2 - \rho\|\boldsymbol{Q}\boldsymbol{c}\|^2 + \kappa \\
& \text{s.t.} \quad \boldsymbol{\theta} \in \mathbb{R}^J,\ \boldsymbol{\eta} \in \mathbb{R}^K_+,\ \rho \in \mathbb{R}_+,\ \kappa \in \mathbb{R} \\
& \qquad \sup_{\boldsymbol{\xi}} \left(\|\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{\mathcal{b}}_i(\boldsymbol{x})^\top \boldsymbol{\xi} - \boldsymbol{\xi}^\top\boldsymbol{\mathcal{S}}^\top\boldsymbol{\theta} + \boldsymbol{\xi}^\top\boldsymbol{\eta} - \rho\|\boldsymbol{Q}\boldsymbol{\xi}\|^2 + 2\rho\boldsymbol{\xi}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{c}\right) \leq \kappa.
\end{aligned}
$$

Reformulating the semi-infinite constraint as a semidefinite constraint then yields the desired reformulation (36). This completes the proof. $\square$

The next proposition shows that the approximation resulting from replacing the copositive cone $\mathcal{C}$ in (35) with its coarsest inner approximation $\mathcal{C}^0$ is stronger than the state-of-art approximate $\mathcal{S}$-lemma method.

Proposition 4. *The following relation holds.*

$$
Z_i(\boldsymbol{x}) = \overline{Z}_i(\boldsymbol{x}) \leq \overline{Z}_i^{\mathcal{C}_0}(\boldsymbol{x}) \leq \overline{Z}_i^{\mathcal{S}}(\boldsymbol{x})
$$

Proof. The equality and the first inequality hold by construction. To prove the second inequality, we consider the following semidefinite program that arises from replacing the cone $\mathcal{C}$ with the inner approximation $\mathcal{C}^0$ in (35).

$$
\begin{aligned}
\overline{Z}_i^{\mathcal{C}_0}(\boldsymbol{x}) = \ \ & \inf \quad c_i(\boldsymbol{x}) + \boldsymbol{t}^\top\boldsymbol{\psi} + (\boldsymbol{t}\circ\boldsymbol{t})^\top\boldsymbol{\phi} + \lambda r^2 - \lambda\|\boldsymbol{Q}\boldsymbol{c}\|^2 + \tau \\
& \text{s.t.} \quad \tau \in \mathbb{R},\ ,\ \lambda, h \in \mathbb{R}_+,\ \boldsymbol{\psi},\boldsymbol{\phi} \in \mathbb{R}^J,\ \boldsymbol{F} \in \mathbb{R}^{K\times K}_+,\ \boldsymbol{g} \in \mathbb{R}^K_+ \\
& \qquad \begin{bmatrix} \lambda\boldsymbol{Q}^\top\boldsymbol{Q} + \boldsymbol{\mathcal{S}}^\top\operatorname{diag}(\boldsymbol{\phi})\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top\boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\psi} - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) - 2\lambda\boldsymbol{Q}^\top\boldsymbol{Q}c\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top\boldsymbol{\psi} - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) - 2\lambda\boldsymbol{Q}^\top\boldsymbol{Q}c\right)^\top & \tau \end{bmatrix} \succeq \begin{bmatrix} \boldsymbol{F} & \boldsymbol{g} \\ \boldsymbol{g}^\top & h \end{bmatrix}
\end{aligned} \tag{37}
$$

We will now show that any feasible solution $(\kappa, \rho, \boldsymbol{\theta}, \boldsymbol{\eta})$ to (36) can be used to construct a feasible solution $(\tau, \lambda, h, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{F},\ \boldsymbol{g})$ to (37) with the same objective value. Specifically, we set $\tau = \kappa$, $\lambda = \rho$, $h = 0$, $\boldsymbol{\psi} = \boldsymbol{\theta}$, $\boldsymbol{\phi} = \mathbf{0}$, $\boldsymbol{F} = \mathbf{0}$, and $\boldsymbol{g} = \boldsymbol{\eta}$. The feasibility of the solution $(\kappa, \rho, \boldsymbol{\theta}, \boldsymbol{\eta})$ in (36) then implies that the constructed solution $(\tau, \lambda, h, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{F},\ \boldsymbol{g})$ is also feasible in (37). Furthermore, one could verify that these solutions give rise to the same objective value for the respective problems. Thus, the claim follows. $\square$

In practice, the tightest approximation is obtained by employing the smallest volume ellipsoid $\mathcal{B}(r, \boldsymbol{Q}, \boldsymbol{c})$ that encloses the set Ξ. Unfortunately, determining the parameters $\boldsymbol{Q}$, $\boldsymbol{c}$, and r of such a set is generically NP-hard. In the following, we will instead consider the simple norm ball $\mathcal{B}(r) = \{\boldsymbol{\xi} \in \mathbb{R}^K : \|\boldsymbol{\xi}\| \leq r\}$ of radius r centered at the origin as a surrogate for the ellipsoid $\mathcal{B}(r, \boldsymbol{Q}, \boldsymbol{c})$. This simpler scheme has recently been employed in [45] to ensure a Slater point for the conservative SDP reformulation of two-stage robust optimization problems with right-hand side uncertainty. While determining the smallest radius r remains intractable, a reasonable approximation can typically be obtained by solving K tractable linear programs. Specifically, we will set

$$r = \sqrt{K} \max_{k \in [K]} \left(\sup_{\boldsymbol{\xi} \in \Xi} \xi_k \right).$$

We will now demonstrate that the inequality in $\overline{Z}_i^0(\boldsymbol{x}) \leq \overline{Z}_i^{\mathcal{S}}(\boldsymbol{x})$ in Proposition 4 can often be strict. This affirms that the proposed SDP reformulation (37) is indeed a stronger approximation than the approximate $\mathcal{S}$-lemma method.

Example 4. *Consider the following quadratic maximization problem.*

$$\begin{aligned} Z_0(\boldsymbol{x}) = \quad \sup \quad & \xi_1^2 \\ \text{s.t.} \quad & \boldsymbol{\xi} \in \mathbb{R}_+^2 \\ & 2\xi_1 + \xi_2 = 2 \end{aligned} \tag{38}$$

A simple analysis shows that $Z_0(\boldsymbol{x}) = 1$, which is attained at the solution $(\xi_1, \xi_2) = (1, 0)$. The problem (38) *constitutes an instance of problem* (34) *with the parameterizations*

$$\boldsymbol{\mathcal{A}}_0(\boldsymbol{x}) = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \boldsymbol{\mathscr{b}}_0(\boldsymbol{x}) = \boldsymbol{0}, \quad \text{and} \quad c_0(\boldsymbol{x}) = 0.$$

Here, the uncertainty set is given by the polytope $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}_+^2 : 2\xi_1 + \xi_2 = 2\}$ which corresponds to the inputs $\boldsymbol{\mathcal{S}} = [2\ 1]$ and $\boldsymbol{t} = 2$. Replacing the cone $\mathcal{C}$ with its inner approximation $\mathcal{C}^0$ in the copositive programming reformulation of (38)*, we find that the resulting semidefinite program yields the same optimal objective value of $\overline{Z}_0^{\mathcal{C}_0}(\boldsymbol{x}) = 1$. Meanwhile, the corresponding approximate $\mathcal{S}$-lemma method yields an optimal objective value $Z_0^{\mathcal{S}}(\boldsymbol{x}) = 4$. Thus, while SDP approximation of the copositive program* (35) *is tight, the approximate $\mathcal{S}$-lemma generates an inferior objective value for the simple instance* (38)*.*

5 Numerical Experiments

In this section, we assess the performance of the SDP approximations studied in Section 4. All optimization problems are solved with MOSEK v8 using the YALMIP interface [31] on a 16-core 3.4 GHz computer with 32 GB RAM.

5.1 Least Squares

The classical least squares problem seeks for an approximate solution $\boldsymbol{x}$ to an overdetermined linear system $\boldsymbol{F}\boldsymbol{x} = \boldsymbol{g}$ by solving the following convex quadratic program whose objective is given by the residual $\|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{g}\|^2$.

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{g}\|^2 \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^D. \end{array}$$

Solutions to this problem can however be very sensitive to perturbations in the input data $\boldsymbol{F} \in \mathbb{R}^{M \times D}$ and $\boldsymbol{g} \in \mathbb{R}^M$ [17, 23]. In order to mitigate the instability of these solutions, El Ghaoui and Lebret [20] propose to solve the robust optimization problem given by

$$\begin{array}{ll} \text{minimize} & \sup_{(\boldsymbol{U}, \boldsymbol{v}) \in \mathcal{U}} \|(\boldsymbol{F} + \boldsymbol{U})\boldsymbol{x} - (\boldsymbol{g} + \boldsymbol{v})\|^2 \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^D. \end{array} \tag{39}$$

Here, the residual in the objective is evaluated in view of the most adverse perturbation matrix $\boldsymbol{U}$ and vector $\boldsymbol{v}$ from within the prescribed uncertainty set $\mathcal{U}$. A tractable SDP reformulation of this problem is derived in [20] for problem instances where the uncertainty set is given by the Frobenius norm ball

$$\mathcal{B}(r) = \left\{ (\boldsymbol{U}, \boldsymbol{v}) \in \mathbb{R}^{M \times D} \times \mathbb{R}^D : \left\| \left[\boldsymbol{U}^\top \ \boldsymbol{v} \right] \right\|_F \leq r \right\}. \tag{40}$$

In our experiments, we consider the case where the uncertainty affects only the left-hand side matrix $\boldsymbol{F}$ (*i.e.*, $\boldsymbol{v} = \boldsymbol{0}$). The resulting robust problem constitutes an instance of (3) with the input parameters

$$I = 0, \quad \boldsymbol{A}_0(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{x}^\top & \boldsymbol{0}^\top & \cdots & \boldsymbol{0}^\top & \boldsymbol{f}_1^\top \boldsymbol{x} - g_1 \\ \boldsymbol{0}^\top & \boldsymbol{x}^\top & \cdots & \boldsymbol{0}^\top & \boldsymbol{f}_2^\top \boldsymbol{x} - g_2 \\ \boldsymbol{0}^\top & \boldsymbol{0}^\top & \ddots & \boldsymbol{0}^\top & \vdots \\ \boldsymbol{0}^\top & \boldsymbol{0}^\top & \cdots & \boldsymbol{x}^\top & \boldsymbol{f}_M^\top \boldsymbol{x} - g_M \end{bmatrix}, \quad \boldsymbol{b}_0(\boldsymbol{x}) = \boldsymbol{0}, \quad c_0(\boldsymbol{x}) = 0,$$

and the uncertainty set

$$\Xi = \left\{ (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_M, 1) \in \mathbb{R}^{MD+1} : \boldsymbol{U} \in \mathcal{U} \right\},$$

where for any matrix $\boldsymbol{H} \in \mathbb{R}^{M \times D}$, we denote its m-th row as $\boldsymbol{h}_m \in \mathbb{R}^D$. We compare the performance our proposed SDP approximation described in Section 4 with the SDP scheme of [20] and the approximate $\mathcal{S}$-lemma method described in Section 4.1. All experimental results are averaged over 100 random trials generated in the following manner. In each trial, we sample the matrix $\boldsymbol{F}$ and the vector $\boldsymbol{g}$ from the uniform distribution on $[0, 1]^{M \times D}$ and $[0, 1]^D$, respectively, and we fix the uncertainty set to the polytope

$$\mathcal{U} = \left\{ \boldsymbol{U} \in \mathbb{R}^{M \times D} : -\hat{\boldsymbol{U}} \circ \boldsymbol{F} \leq \boldsymbol{U} \leq \hat{\boldsymbol{U}} \circ \boldsymbol{F} \right\}, \tag{41}$$

where the matrix $\hat{\boldsymbol{U}}$ is chosen uniformly at random from $[0, 1]^{M \times D}$. As the SDP scheme of [20] is only able to handle simple ellipsoidal uncertainty sets, in the experiments we set the radius r in (40) to $\|\hat{\boldsymbol{U}} \circ \boldsymbol{F}\|_F$ so that the resulting Frobenius norm ball is the smallest one enclosing the polytope (41). This gives rise to a conservative SDP approximation for the original problem. We can also solve the robust least squares problem exactly by enumerating all extreme points of the polytope (41) and formulating a convex quadratically constrained quadratic program that incorporates all these enumerated points in its constraint system. The resulting problem however is generically intractable as its size scales exponentially in the dimensions M and D.

Table 1 reports average results over 100 randomly generated instances for problem dimensions $M = 4$ and $D = 3$. The table shows that our proposed SDP approximation is in fact exact as it reports worst-case residual estimates that coincide with the optimal ones. On the other hand, both the approximate $\mathcal{S}$-lemma method and the SDP scheme of [20] generate overly pessimistic estimates of the resulting worst-case residual. Indeed, the SDP scheme of [20] overestimates the residual by up to 82%. The approximate $\mathcal{S}$-lemma method performs slightly better because it incoporates the polyhedral description of the uncertainty set in its reformulation, with estimations that are still up to 37% higher than the optimal ones. These inflated worst-case residual estimates further result in solutions that also underperform when evaluated in the exact problem ('Suboptimality' column).

	Bound gap			Suboptimality		
Statistic	SDP	$\mathcal{S}$-Lemma	EGL	SDP	$\mathcal{S}$-Lemma	EGL
Mean	0.0%	36.7%	82.1%	0.0%	26.5%	40.8%
10th Percentile	0.0%	1.8%	22.8%	0.0%	1.0%	6.6%
90th Percentile	0.0%	100.8%	181.4%	0.0%	71.0%	98.7%

Table 1. Numerical results for the proposed SDP approximation ('SDP'), the approximate $\mathcal{S}$-lemma method ('$\mathcal{S}$-Lemma'), and the SDP scheme in [20] ('EGL'). The 'bound gap' quantifies the increase in the estimated worst-case residual relative to the residual of the exact problem, and the 'suboptimality' quantifies the increase in the residual of the exact problem if we replace the optimal solution with the determined solution.

Table 2 reports experimental results on larger problem instances. Here, we are unable to solve the robust least squares problem *exactly* as the solver runs out of memory, and only report the relative improvements of the estimates generated from the proposed SDP approximation over the ones from the approximate $\mathcal{S}$-lemma method and the SDP scheme in [20], respectively. The table shows that the results presented in Table 1 generalize to these larger instances where our proposed SDP approximation distinctly outperforms the other state-of-the-art schemes.

	$M=6,\ D=4$		$M=10,\ D=4$		$M=10,\ D=6$		$M=12,\ D=5$	
Statistic	$\mathcal{S}$-Lemma	EGL	$\mathcal{S}$-Lemma	EGL	$\mathcal{S}$-Lemma	EGL	$\mathcal{S}$-Lemma	EGL
Mean	26.9%	68.9%	15.9%	51.0%	21.6%	60.2%	16.1%	52.0%
10th Percentile	4.8%	26.2%	3.8%	26.5%	7.5%	33.0%	4.9%	30.4%
90th Percentile	57.0%	118.2%	34.2%	81.0%	39.2%	94.5%	34.0%	82.9%

Table 2. Percentage improvements of the estimated worst-case residual generated from the proposed SDP approximation relative to the ones generated from the approximate $\mathcal{S}$-lemma method ('$\mathcal{S}$-Lemma') and the SDP scheme in [20] ('EGL'), respectively.

5.2 Project Management

We consider a project crashing problem described in Example 2, where the duration of activity $(i,j) \in \mathcal{A}$ is given by the uncertain quantity $d_{ij} = (1 + r_{ij})d_{ij}^0$. Here, d_{ij}^0 is the nominal activity duration and r_{ij} represents exogenous fluctuations. Let x_{ij} be the amount of resources that is used to expedite the activity (i,j). We consider randomly generated project networks of size $|\mathcal{V}| = 25$ and order strength 0.75,[2] which gives rise to projects with an average of 50 activities. We fix the feasible set of the resource allocation vector to $\mathcal{X} = \{\boldsymbol{x} \in [0,1]^{|\mathcal{A}|} : \mathbf{e}^\top \boldsymbol{x} \le \frac{3}{4}|\mathcal{A}|\}$, so that at most 75% of the activities can receive the maximum resource allocation. We further set the nominal task durations to $\boldsymbol{d}^0 = \mathbf{e}$. The uncertainty set of $\boldsymbol{d}$ is defined through a factor model as follows:

$$\mathcal{D} = \left\{ \boldsymbol{d} \in \mathbb{R}^{|\mathcal{A}|} : d_{ij} = (1 + \boldsymbol{f}_{ij}^\top \boldsymbol{\chi})d_{ij}^0 \text{ for some } \boldsymbol{\chi} \in [0,1]^F, \quad \forall (i,j) \in \mathcal{A} \right\},$$

where the factor size is fixed to $F = |\mathcal{V}|$. For each activity $(i,j) \in \mathcal{A}$ we sample the factor loading vector $\boldsymbol{f}_{ij}$ from the uniform distribution on $[-\frac{1}{2F}, \frac{1}{2F}]^F$, which ensures that the duration of each activity can deviate by up to 50% of its nominal value. We can form the final mixed-integer uncertainty set Ξ from $\mathcal{D}$ using (11). In the experiments, we compare the performance the proposed SDP approximation for (9) with a popular linear decision rules bounding scheme (LDR) in the literature [12, 44]. The project crashing problem can also be solved exactly by enumerating all the paths in the network and formulating an equivalent robust optimization problem which incorporates the description of all these paths in its constraint system [44]. While this equivalent robust problem can be reduced to a linear program, the resulting problem is intractable as its size grows exponentially in the description of the project crashing instance.

Table 3 presents average results over 100 randomly generated project networks. While our proposed SDP approximation consistently provides near optimal estimates of the worst-case project makespan ($\sim$2% gaps), the LDR bound generates overly pessimistic estimates ($\sim$24% gaps). The 10th and 90th percentiles

[2] The order strength denotes the fraction of all $|\mathcal{V}|(|\mathcal{V}|-1)/2$ possible precedences between the nodes that are enforced in the graph.

of the bound gaps further indicate that the estimates generated from our SDP approximation stochastically dominate those generated from the LDR approximation. In addition, these excessive worst-case makespan estimates give rise to poor resource allocations that also underperform when evaluated in the exact problem.

	Bound gap		Suboptimality	
Statistic	SDP	LDR	SDP	LDR
Mean	2.2%	24.3%	1.4%	9.5%
10th Percentile	1.7%	20.3%	1.0%	6.3%
90th Percentile	2.8%	29.0%	1.9%	13.3%

Table 3. Numerical results for the proposed SDP approximation ('SDP') and the linear decision rules approximation ('LDR'). The 'bound gap' quantifies the increase in the estimated worst-case makespan relative to the makespan of the exact problem, and the 'suboptimality' quantifies the increase in the makespan of the exact problem if we replace the optimal solution with the determined resource allocation.

5.3 Multi-Item Newsvendor

We now assess the performance of the SDP approximation described in Section 4 on a multi-item newsvendor problem, where an inventory planner has to determine the vector $\boldsymbol{x} \in \mathbb{R}^K_+$ of order quantities for K different products at the beginning of a planning period. Demands $\boldsymbol{\xi} \in \mathbb{Z}^K_+$ of these products are uncertain and are assumed to belong to a prescribed discrete uncertainty set Ξ. We assume that there are no ordering costs on the products but the total order quantity $\mathbf{e}^\top \boldsymbol{x}$ must not exceed a given budget B. Excess inventory of the k-th product incurs a per-unit holding cost of g_k, while unmet demand incurs a per-unit stock-out cost of s_k.

For any realization of the demand vector $\boldsymbol{\xi}$, the total cost of a fixed order $\boldsymbol{x}$ is given by

$$\begin{aligned}
\mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi}) &= \mathbf{e}^\top \max\left\{\mathrm{diag}(\boldsymbol{g})(\boldsymbol{x} - \boldsymbol{\xi}), \mathrm{diag}(\boldsymbol{s})(\boldsymbol{\xi} - \boldsymbol{x})\right\} \\
&= \inf_{\boldsymbol{y} \in \mathbb{R}^K_+} \left\{\mathbf{e}^\top \boldsymbol{y} \;:\; \boldsymbol{y} \geq \mathrm{diag}(\boldsymbol{g})(\boldsymbol{x} - \boldsymbol{\xi}),\ \boldsymbol{y} \geq \mathrm{diag}(\boldsymbol{s})(\boldsymbol{\xi} - \boldsymbol{x})\right\}.
\end{aligned}$$

The objective of a risk-averse inventory planner is then to determine the best order $\boldsymbol{x}$ that minimizes the worst-case total cost $\sup_{\boldsymbol{\xi} \in \Xi} \mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi})$. This gives rise to the optimization problem

$$\begin{array}{ll}
\text{minimize} & \sup\limits_{\boldsymbol{\xi} \in \Xi} \mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi}) \\
\text{subject to} & \boldsymbol{x} \in \mathbb{R}^K_+ \\
& \mathbf{e}^\top \boldsymbol{x} \leq B.
\end{array} \tag{42}$$

This problem constitutes an instance of the two-stage robust optimization problem (26) corresponding to the input parameters

$$\boldsymbol{A}_0(\boldsymbol{x}) = \mathbf{0},\ \boldsymbol{b}_0(\boldsymbol{x}) = \mathbf{0},\ c_0(\boldsymbol{x}) = 0,\ \boldsymbol{P} = \mathbf{0},\ \boldsymbol{R} = \mathbf{0},\ \boldsymbol{r} = \mathbf{e},\ \boldsymbol{T}(\boldsymbol{x}) = \begin{bmatrix} -\operatorname{diag}(\boldsymbol{g}) \\ \operatorname{diag}(\boldsymbol{s}) \end{bmatrix},\ \text{and } \boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \operatorname{diag}(\boldsymbol{g})\boldsymbol{x} \\ -\operatorname{diag}(\boldsymbol{s})\boldsymbol{x} \end{bmatrix}.$$

In our experiments, we compare the performance of the SDP approximation of (42) that explicitly models $\boldsymbol{\xi}$ as a discrete vector with the one that naïvely relaxes $\boldsymbol{\xi}$ as a continuous vector. To this end, we consider a $K = 5$ item newsvendor problem with the following setting. We fix the vectors of holding and stock-out costs to $\boldsymbol{g} = \mathbf{e}$ and $\boldsymbol{s} = 5\mathbf{e}$, respectively, and we set the ordering budget to $B = 20$. All experimental results are averaged over 100 random trials generated in the following manner. In each trial, we sample the vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ independently from the uniform distribution on the unit hypercube $[0,1]^K$. We then set the uncertainty sets for the exact and the naïve models to

$$\Xi = \left\{\boldsymbol{\xi} \in \mathbb{Z}_+^K : \boldsymbol{\xi} \le 15\mathbf{e},\ \boldsymbol{w}_1^\top \boldsymbol{\xi} \le 1,\ \boldsymbol{w}_2^\top \boldsymbol{\xi} \le 1\right\} \quad \text{and} \quad \Xi_{\text{Cont}} = \left\{\boldsymbol{\xi} \in \mathbb{R}_+^K : \boldsymbol{\xi} \le 15\mathbf{e},\ \boldsymbol{w}_1^\top \boldsymbol{\xi} \le 1,\ \boldsymbol{w}_2^\top \boldsymbol{\xi} \le 1\right\},$$

respectively, and we solve the SDP approximations of (42) with inputs Ξ and Ξ_{Cont}. Alternatively, since the uncertainty set Ξ is discrete, we can enumerate all possible demand vectors $\boldsymbol{\xi}$ in Ξ and solve the exact linear programming reformulation of (42) that incorporates all these enumerated demand vectors in its constraint system. We remark that while the aforementioned SDP approximations are polynomial time solvable, the exact linear programming reformulation is generically intractable as its size scales exponentially in the dimension K. In the case of $K = 5$, for instance, we find that the number of constraints in resulting linear program is already in the order of $15^5 \approx 7.5 \times 10^5$.

Table 4 reports the statistics of the optimality gaps. While the naïve model can on average incur a large optimality gap of more than 25%, the exact model consistently provides near optimal solutions with an average gap of less than 5%. Thus, modeling a discrete uncertainty set properly using the proposed scheme will often generate high quality solutions from the resulting SDP approximation.

	Bound gap		Suboptimality	
Statistic	SDP Exact	SDP Naïve	SDP Exact	SDP Naïve
Mean	4.8%	25.4%	4.8%	25.4%
10th Percentile	0.1%	11.7%	0.0%	11.7%
90th Percentile	12.1%	41.1%	12.0%	41.0%

Table 4. Numerical results for the SDP approximations to the model that imposes integrality on the uncertainty set ('SDP Exact') and the model that ignores the integrality ('SDP Naïve'). The 'bound gap' quantifies the increase in the estimated worst-case cost relative to the cost of the exact problem, and the 'suboptimality' quantifies the increase in the cost of the exact problem if we replace the optimal solution with the determined order quantities.

References

[1] S. D. Ahipasaoglu, K. Natarajan, and D. Shi. Distributionally robust project crashing with partial or no correlation information. 2016.

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

[3] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming A*, 99(2):351–376, 2004.

[4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[5] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2002.

[6] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.

[7] D. Bertsimas and M. Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming B*, 107(1-2):5–36, 2006.

[8] I. M. Bomze and E. de Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24(2):163–185, 2002.

[9] S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming A*, 120(2):479–495, 2009.

[10] S. Burer. Copositive programming. In *Handbook on Semidefinite, Conic and Polynomial Optimization*, pages 201–218. Springer, 2012.

[11] C. Caramanis, S. Mannor, and H. Xu. Robust optimization in machine learning. *Optimization for machine learning*, page 369, 2011.

[12] X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.

[13] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[14] E. de Klerk and D. V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization*, 12(4):875–892, 2002.

[15] E. Delage and D. A. Iancu. Robust multistage decision making. In *The Operations Research Revolution*, pages 20–46. INFORMS, 2015.

[16] P. H. Diananda. On non-negative forms in real variables some or all of which are non-negative. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 58, pages 17–25, 1962.

[17] L. Eldén. Perturbation theory for the least squares problem with linear equality constraints. *SIAM Journal on Numerical Analysis*, 17(3):338–350, 1980.

[18] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[19] A. Georghiou, W. Wiesemann, and D. Kuhn. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming*, 152(1-2):301–338, 2015.

[20] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

[21] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010.

[22] D. Goldfarb and G. Iyengar. Robust convex quadratically constrained programs. *Mathematical Programming*, 97(3):495–515, 2003.

[23] G. H. Golub and C. Van Loan. *Matrix Computations*. John Hopkins Press, 2012.

[24] B. L. Gorissen, I. Yanıkoğlu, and D. den Hertog. A practical guide to robust optimization. *Omega*, 53:124–137, 2015.

[25] M. J. Hadjiyiannis, P. J. Goulart, and D. Kuhn. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In *IEEE Conference on Decision and Control and European Control Conference*, pages 7386–7391, 2011.

[26] G. A. Hanasusanto and D. Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *arXiv preprint arXiv:1609.07505*, 2016.

[27] A. J. Kleywegt, A. Shapiro, and T. Homem de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[28] Q. Kong, C.-Y. Lee, C.-P. Teo, and Z. Zheng. Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research*, 61(3):711–726, 2013.

[29] J. B. Lasserre. Convexity in semialgebraic geometry and polynomial optimization. *SIAM Journal on Optimization*, 19(4):1995–2014, 2009.

[30] X. Li, K. Natarajan, C.-P. Teo, and Z. Zheng. Distributionally robust mixed integer linear programs: Persistency models with applications. *European Journal of Operational Research*, 233(3):459–473, 2014.

[31] J. Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *IEEE International Symposium on Computer Aided Control Systems Design*, pages 284–289, 2004.

[32] A. Marandi, A. Ben-Tal, D. den Hertog, and B. Melenberg. Extending the scope of robust quadratic optimization. *Available on Optimization Online*, 2017.

[33] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[34] K. Natarajan and C.-P. Teo. On reduced semidefinite programs for second order moment bounds with applications. *Forthcoming in Mathematical Programming A*, 2016.

[35] K. Natarajan, C.-P. Teo, and Z. Zheng. Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Operations Research*, 59(3):713–728, 2011.

[36] P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[37] R. T. Rockafellar and R. J. B. Wets. Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time. *SIAM Journal on Control and Optimization*, 28(4):810–822, 1990.

[38] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.

[39] S. Shafieezadeh-Abadeh, P. Mohajerin Esfahani, and D. Kuhn. Distributionally robust logistic regression. In *Advances in Neural Information Processing Systems*, pages 1576–1584, 2015.

[40] N. Shaked-Monderer, A. Berman, I. M. Bomze, F. Jarre, and W. Schachinger. New results on the cp-rank and related properties of co(mpletely)positive matrices. *Linear and Multilinear Algebra*, 63(2):384–396, 2015.

[41] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Handbook in Operations Research and Management Science*, volume 10, pages 353–425. Elsevier, 2003.

[42] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2014.

[43] L. J. Watters. Reduction of integer polynomial programming problems to zero-one linear programming problems. *Operations Research*, 15(6):1171–1174, 1967.

[44] W. Wiesemann, D. Kuhn, and B. Rustem. Robust resource allocations in temporal networks. *Mathematical programming*, pages 1–35.

[45] G. Xu and S. Burer. A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides. *arXiv preprint arXiv:1609.07402*, 2016.

[46] H. Xu, C. Caramanis, and S. Mannor. Robustness and regularization of support vector machines. *Journal of Machine Learning Research*, 10(Jul):1485–1510, 2009.