Robust Quadratic Programming with Mixed-Integer Uncertainty

Areesh Mittal, Can Gokalp, and Grani A. Hanasusanto

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

December 17, 2018

Abstract

We study robust convex 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. These convex optimization problems are NP-hard but admit a conservative semidefinite programming (SDP) approximation that can be solved efficiently. We prove that the popular approximate $\mathcal{S}$-lemma method—which is valid only in the case of continuous uncertainty—is weaker than our approximation. We also show that all results can be extended to the two-stage robust quadratic 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 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 systems can be formulated as convex quadratic programs of the form

$$
\begin{aligned}
\text{minimize} \quad & \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}) \\
\text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
\tag{1}
$$

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 polytope, $\boldsymbol{\xi} \in \mathbb{R}^K$ is a vector of exogenous problem parameters, $\boldsymbol{A}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^{M \times K}$ and $\boldsymbol{b}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^K$ are matrix- and vector-valued affine functions, respectively, while $c(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}$ is a convex quadratic function. The objective of problem (1) is to determine the best decision $\boldsymbol{x} \in \mathcal{X}$ that minimizes the quadratic function $\|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x})$. The generic formulation (1) includes the class of linear programming problems [42] as a special case (when $\boldsymbol{A} = \boldsymbol{0}$), and has numerous important applications, e.g., in portfolio optimization [37],

least squares regression [26], supervised classification [15], optimal control [41], 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. Thus, they are subject to potentially significant errors that can adversely impact the out-of-sample performance of an 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 $\boldsymbol{\xi}$ lies within a prescribed uncertainty set Ξ and we replace the objective function of (1) with the *worst-case* function given by

$$\sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}). \tag{2}$$

This optimization problem yields a solution $\boldsymbol{x} \in \mathcal{X}$ that minimizes the quadratic objective function under the most adverse uncertain parameter realization $\boldsymbol{\xi} \in \Xi$.

Robust optimization models are appealing as they require minimal assumptions on the description of the uncertain parameters and because they often lead to efficient solution schemes. In 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, 27]. Tractable reformulations for robust quadratic programming problems are derived in [25, 36] 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 become tractable—by virtue of the *exact $\mathcal{S}$-lemma*—if the uncertainty set is defined by an ellipsoid [4, 23]. 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 [36], 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 logistic regression [43], classification problems with

noisy labels [13, 51] and network optimization [1, 48]. If the uncertain parameters contain mixed-integer components then the problem 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 [21]. The corresponding robust version of (1) is tractable only in a few contrived situations, *e.g.*, when the uncertainty set possesses a total unimodularity property or is described by the convex hull of polynomially many integer vectors [4]. Perhaps due to these limitations, there are currently very few results in the literature that provide a systematic and rigorous way to handle generic robust optimization problems with mixed-integer uncertainty. In this paper, 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. An alternate way to handle integer uncertain parameters can be to solve the problem by simply ignoring the integrality assumption. However, doing so adds undesired conservativeness to the uncertainty set. Indeed, in our numerical experiments, we demonstrate that ignoring the integrality assumption on the uncertain parameters leads to overly conservative solutions.

Optimization problems under uncertainty may also involve adaptive recourse decisions which are taken once the uncertain parameters are realized [2, 46]. 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 the uncertainty space with finitely many points [28, 31, 45] or by employing decision rule methods, which restrict the set of all possible recourse decisions to simpler parametric forms in $\boldsymbol{\xi}$ [3, 22, 24]. We refer the reader to [17] for a comprehensive review of recent results in adaptive robust optimization. In this paper, we consider two-stage robust optimization problems with quadratic first- and second-stage objective function and a mixed-integer uncertainty set. We show that if the problem has *complete recourse*, then it can be reformulated as a conic program—which is amenable to tractable approximations.

The conic programming route that we take here to model optimization problems under uncertainty has previously been traversed. In [39], 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 [32, 34, 38]. Recently, equivalent copositive programming reformulations are derived for generic two-stage robust linear programs [29, 50]. 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 [20], the authors

provide completely positive reformulation for a two-stage distributionally robust supply chain risk mitigation problem. They allow some components of $\boldsymbol{\xi}$ to be binary, but assume precise knowledge of the first- and the second-order moments of the distribution of $\boldsymbol{\xi}$. The objective function that they consider is quadratic in the second-stage decision variables but affine in $\boldsymbol{\xi}$. In contrast to [20], we assume no information about the distribution of $\boldsymbol{\xi}$, other than the support. Furthermore, we allow the objective function to be quadratic in the decision variables, as well as in $\boldsymbol{\xi}$, which helps us model a more general class of robust problems, *e.g.*, robust least squares [23].

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 quadratic program can be reformulated as 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 quadratic 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. We prove that even the simplest of these approximations is stronger than the well-known approximate $\mathcal{S}$-lemma method if the problem instance has only continuous uncertain parameters. Furthermore, when some uncertain parameters are restricted to take integer values, the approximate $\mathcal{S}$-lemma method is not applicable, while our method still generates a high-quality conservative solution.

3. We compare our approximation method to other state-of-the-art approximation schemes through extensive numerical experiments. We show that our approximation method generates better estimates of worst-case cost and yields less conservative solutions. We also demonstrate that ignoring the integrality assumption on the uncertainty set may lead to inferior solutions to the robust problem.

4. To the best of our 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 robust quadratic programs.

The remainder of the paper is structured as follows. We formulate and discuss the generic robust quadratic programs in Section 2. We then derive the copositive programming reformulation in Section 3. Section 4 develops a conservative SDP reformulation and provides a theoretical comparison with the popular approximate $\mathcal{S}$-lemma method. In Section 5, we extend the results of Section 3 along several directions including two-stage robust quadratic optimization. We demonstrate the impact of our proposed reformulation via numerical experiments in Section 6, and finally, we conclude in Section 7.

Notation: We use $\mathbb{Z}$ ($\mathbb{Z}_+$) to denote the set of (non-negative) integers. For any positive integer I, we use $[I]$ to denote the index set $\{1, \ldots, I\}$. We use $\|.\|_p$ to denote the l_p-norm. We drop the subscript and write $\|.\|$ when referring to the l_2-norm. 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 $\text{tr}(\boldsymbol{M})$ the trace of a square matrix $\boldsymbol{M}$. For a vector $\boldsymbol{v}$, $\text{diag}(\boldsymbol{v})$ denotes the diagonal matrix with $\boldsymbol{v}$ on its diagonal; whereas for a square matrix $\boldsymbol{M}$, $\text{diag}(\boldsymbol{M})$ 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 the 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 $g(K) = \max\{\binom{K+1}{2} - 4, K\}$ [44]. 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 quadratic programs (RQPs) of the form

$$
\begin{aligned}
\text{minimize} \quad & \sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}) \\
\text{subject to} \quad & \boldsymbol{x} \in \mathcal{X},
\end{aligned}
\tag{3}
$$

where the set $\mathcal{X}$ and the functions $\boldsymbol{A}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^{M \times K}$, $\boldsymbol{b}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^K$, and $c(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}$ 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

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

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 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 mean-variance portfolio optimization problem*

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

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 trade-off 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:

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

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

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

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):

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

This problem constitutes an instance of (3) with the input parameters

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

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 assume that that node with index 1 represents the start of the project and the node with index $|\mathcal{V}|$ represents the end of the project. 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. If the activity duration $d_{ij} - x_{ij}$ represents the length of the arc (i,j), then the project completion time can be determined by computing the length of the longest path from the start node to the end node. We can formulate project crashing as the optimization problem

$$\text{minimize} \quad \sup_{\boldsymbol{z} \in \mathcal{Z}} \sum_{(i,j) \in \mathcal{A}} (d_{ij} - x_{ij}) z_{ij}$$
$$\text{subject to} \quad \boldsymbol{x} \in \mathcal{X},$$

where

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

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

$$\text{minimize} \quad \sup_{\boldsymbol{d} \in \mathcal{D}} \left(\sup_{\boldsymbol{z} \in \mathcal{Z}} \sum_{(i,j) \in \mathcal{A}} (d_{ij} - x_{ij}) z_{ij} \right) \tag{6}$$
$$\text{subject to} \quad \boldsymbol{x} \in \mathcal{X}.$$

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

$$\sup_{\boldsymbol{d} \in \mathcal{D}} \sup_{\boldsymbol{z} \in \mathcal{Z}} \sum_{(i,j) \in \mathcal{A}} (d_{ij} - x_{ij}) z_{ij} = \sup_{(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{q}) \in \Xi} \mathbf{e}^\top \boldsymbol{q} - \boldsymbol{x}^\top \boldsymbol{z}, \tag{7}$$

where

$$\Xi = \left\{ (\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{q}) \in \mathcal{D} \times \mathcal{Z} \times \mathbb{R}_+^{|\mathcal{A}|} : \boldsymbol{q} \leq \boldsymbol{z}, \ \boldsymbol{q} \leq \boldsymbol{d}, \ \boldsymbol{q} \geq \boldsymbol{d} - \mathbf{e} + \boldsymbol{z} \right\}. \tag{8}$$

Using the new objective function (7) and uncertainty set (8), the resulting robust optimization problem constitutes an instance of (3) with the input parameters $\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{0}$, $\boldsymbol{b}(\boldsymbol{x}) = [\boldsymbol{0}^\top \quad -\boldsymbol{x}^\top \quad \mathbf{e}^\top]^\top$, and $c(\boldsymbol{x}) = 0$.

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

$$Z(\boldsymbol{x}) = \sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}), \tag{9}$$

which corresponds to the inner subproblem in the objective of (3). We may therefore represent (3) as

$$\text{minimize} \quad Z(\boldsymbol{x})$$
$$\text{subject to} \quad \boldsymbol{x} \in \mathcal{X}.$$

In the next section, we derive exact copositive programming reformulation for evaluating $Z(\boldsymbol{x})$. By substituting $Z(\boldsymbol{x})$ with the emerging copositive program, we obtain an equivalent finite-dimensional convex reformulation for the RQP (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 $\boldsymbol{x} \in \mathcal{X}$, we first derive a copositive upper bound on $Z(\boldsymbol{x})$. We then show that the resulting reformulation is in fact exact under the boundedness assumption on the uncertainty set Ξ.

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

To derive the copositive reformulation, we leverage the following result by Burer [11] which enables us to reduce a generic mixed-binary quadratic program into an equivalent conic program of polynomial size.

Theorem 1 ([11, Theorem 2.6]). *The mixed-binary quadratic program*

$$
\begin{aligned}
\text{maximize} \quad & \boldsymbol{\xi}^\top \boldsymbol{Q} \boldsymbol{\xi} + \boldsymbol{r}^\top \boldsymbol{\xi} \\
\text{subject to} \quad & \boldsymbol{\xi} \in \mathbb{R}_+^P \\
& \boldsymbol{F} \boldsymbol{\xi} = \boldsymbol{g} \\
& \xi_\ell \in \{0, 1\} \qquad \forall \ell \in \mathcal{L}
\end{aligned}
\tag{10}
$$

is equivalent to the completely positive program

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

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 [47].

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 quadratic maximization problem* (9) *coincides with the optimal value of the completely positive program*

$$
\begin{aligned}
Z(\boldsymbol{x}) = \quad \sup \quad & \mathrm{tr}\left(\boldsymbol{\mathcal{A}}(\boldsymbol{x})\boldsymbol{\Omega}\boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top\right) + \boldsymbol{\mathcal{b}}(\boldsymbol{x})^\top \boldsymbol{\xi}' + c(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{\xi}' \in \mathbb{R}_+^{K'},\ \boldsymbol{\Omega} \in \mathbb{S}_+^{K'} \\
& \boldsymbol{\mathcal{S}}\boldsymbol{\xi}' = \boldsymbol{t},\ \mathrm{diag}(\boldsymbol{\mathcal{S}}\boldsymbol{\Omega}\boldsymbol{\mathcal{S}}^\top) = \boldsymbol{t} \circ \boldsymbol{t} \\
& \xi'_\ell = \Omega_{\ell\ell} \quad \forall \ell \in [LQ] \\
& \begin{bmatrix} \boldsymbol{\Omega} & \boldsymbol{\xi}' \\ \boldsymbol{\xi}'^\top & 1 \end{bmatrix} \succeq_{\mathcal{C}^*} \boldsymbol{0},
\end{aligned}
\tag{11}
$$

where

$$
\boldsymbol{\mathcal{S}} = \begin{bmatrix}
\boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{S} \\
-\mathbf{v}_Q^\top & \cdots & \boldsymbol{0}^\top & \boldsymbol{0}^\top & \cdots & \boldsymbol{0}^\top & \mathbf{e}_1^\top \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\boldsymbol{0}^\top & \cdots & -\mathbf{v}_Q^\top & \boldsymbol{0}^\top & \cdots & \boldsymbol{0}^\top & \mathbf{e}_L^\top \\
\mathbb{I} & \cdots & \boldsymbol{0} & \mathbb{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\boldsymbol{0} & \cdots & \mathbb{I} & \boldsymbol{0} & \cdots & \mathbb{I} & \boldsymbol{0}
\end{bmatrix} \in \mathbb{R}^{J' \times K'}, \quad
\boldsymbol{t} = \begin{bmatrix} \boldsymbol{t} \\ 0 \\ \vdots \\ 0 \\ \mathbf{e} \\ \vdots \\ \mathbf{e} \end{bmatrix} \in \mathbb{R}^{J'},
\tag{12}
$$

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

$$
\boldsymbol{\mathcal{b}}(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{0}^\top & \cdots & \boldsymbol{0}^\top & \boldsymbol{0}^\top & \cdots & \boldsymbol{0}^\top & \boldsymbol{b}(\boldsymbol{x})^\top \end{bmatrix}^\top \in \mathbb{R}^{K'},
$$

with

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

Proof. Lemma 1 enables us to reformulate the mixed-integer quadratic program (9) equivalently as the mixed-binary quadratic program

$$
\begin{aligned}
Z(\boldsymbol{x}) = \quad \sup \quad & \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{\xi} \in \mathbb{R}_+^K,\ \boldsymbol{\chi}_\ell \in \{0,1\}^Q \qquad \forall \ell \in [L] \\
& \boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{t} \\
& \xi_\ell = \mathbf{v}_Q^\top \boldsymbol{\chi}_\ell \qquad\qquad \forall \ell \in [L].
\end{aligned}
\tag{13}
$$

We now employ Theorem 1 to derive the equivalent completely positive program for (13). To this end, we first bring the above quadratic program into the standard form (10). 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 $\boldsymbol{\chi}_\ell + \boldsymbol{\eta}_\ell = \mathbf{e}$,

9

$\forall \ell \in [L]$. This yields the equivalent problem

$$
\begin{aligned}
Z(\boldsymbol{x}) = \quad \sup \quad & \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{\xi} \in \mathbb{R}_+^K, \ \boldsymbol{\eta}_\ell \in \mathbb{R}_+^Q, \ \boldsymbol{\chi}_\ell \in \{0,1\}^Q \quad \forall \ell \in [L] \\
& \boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{t} \\
& \xi_\ell = \mathbf{v}_Q^\top \boldsymbol{\chi}_\ell \quad\quad\quad\quad\quad\quad \forall \ell \in [L] \\
& \boldsymbol{\chi}_\ell + \boldsymbol{\eta}_\ell = \mathbf{e} \quad\quad\quad\quad\quad\quad \forall \ell \in [L].
\end{aligned} \tag{14}
$$

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 (14). Together with the augmented parameters (12), we can reformulate (14) concisely as the problem

$$
\begin{aligned}
Z(\boldsymbol{x}) = \quad \sup \quad & \|\boldsymbol{\mathcal{A}}(\boldsymbol{x})\boldsymbol{\xi}'\|^2 + \boldsymbol{\mathscr{b}}(\boldsymbol{x})^\top \boldsymbol{\xi}' + c(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{\xi}' \in \mathbb{R}_+^{K'} \\
& \boldsymbol{\mathcal{S}}\boldsymbol{\xi}' = t \\
& \xi_\ell' \in \{0,1\} \quad\quad\quad\quad\quad\quad \forall \ell \in [LQ].
\end{aligned} \tag{15}
$$

The mixed-binary quadratic program (15) already has the desired standard form (10) with inputs $P = K'$, $\boldsymbol{Q} = \boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}(\boldsymbol{x})$, $\boldsymbol{r} = \boldsymbol{\mathscr{b}}(\boldsymbol{x})$, $\boldsymbol{F} = \boldsymbol{\mathcal{S}}$, $\boldsymbol{g} = \boldsymbol{t}$, and $\mathcal{L} = [LQ]$. We may thus apply Theorem 1 to obtain the equivalent completely positive program (11). This completes the proof. $\qquad\square$

We remark that in view of the concise representation in Lemma 1, the size of the completely positive program (11) remains polynomial in the size of the input data. This completely positive program admits a dual copositive program given by

$$
\begin{aligned}
\overline{Z}(\boldsymbol{x}) = \quad \inf \quad & c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \tau \\
\text{s.t.} \quad & \tau \in \mathbb{R}, \ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{J'}, \ \boldsymbol{\gamma} \in \mathbb{R}^{LQ} \\
& \begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi})\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top \ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathscr{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \mathbf{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathscr{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \mathbf{0}^\top]^\top\right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0}.
\end{aligned} \tag{16}
$$

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

Proposition 2. *For any fixed decision $\boldsymbol{x} \in \mathcal{X}$ we have $\overline{Z}(\boldsymbol{x}) \geq Z(\boldsymbol{x})$.*

3.2   A Copositive Reformulation of RQP

In this section, we demonstrate strong duality for the primal and dual pair (11) and (16), respectively, under the natural boundedness assumption on the uncertainty set Ξ. This exactness result enables us to reformulate the RQP (3) equivalently as a copositive program of polynomial size.

Theorem 2 (Strong Duality). *For any fixed decision $\boldsymbol{x} \in \mathcal{X}$ we have $\overline{Z}(\boldsymbol{x}) = Z(\boldsymbol{x})$.*

We would like to mention that a similar result is proved in a recent paper (Theorem 8 in [9]). However, the two proofs are quite different from one another. While the proof in [9] establishes strong duality by proving the existence of a Slater point for a general copositive program, we show explicitly how to construct a Slater point for the copositive program (16) from input parameters. Because of its constructive nature, we believe our proof is interesting on its own and sheds some light on the geometry of the feasible region of (16).

We note that the primal completely positive program (11) never has an interior [12]. In order to prove Theorem 2, we construct a Slater point for the dual copositive program (16). The construction of the Slater point for problem (16) relies on the following two lemmas. We observe that by construction the boundedness of the uncertainty set Ξ means that the lifted polytope

$$\Xi' = \{\boldsymbol{\xi}' \in \mathbb{R}^{K'} : \boldsymbol{S}\boldsymbol{\xi}' = t, \ \boldsymbol{\xi}' \geq \boldsymbol{0}\} \tag{17}$$

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

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

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

The next lemma, which was proven in [29, 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 $\boldsymbol{M} \succ_{\mathcal{C}} \boldsymbol{0}$ if $\boldsymbol{R} - \boldsymbol{Q}^\top \boldsymbol{P}^{-1} \boldsymbol{Q} \succ_{\mathcal{C}} \boldsymbol{0}$ and $\boldsymbol{P} \succ \boldsymbol{0}$.

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

1. If $\boldsymbol{\rho} = \boldsymbol{0}$, then $\mathbf{e}^\top \boldsymbol{\xi} = 1$. Therefore $\boldsymbol{\xi} \neq \boldsymbol{0}$, 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, $\boldsymbol{M} \succ_{\mathcal{C}} \mathbf{0}$. $\qquad\square$

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 (16). Specifically, we set $\boldsymbol{\gamma} = \mathbf{0}$, $\boldsymbol{\psi} = \mathbf{0}$, and $\boldsymbol{\phi} = \rho \mathbf{e}$ for some $\rho > 0$. Problem (16) then admits a Slater point if there exist scalars $\rho, \tau > 0$, such that

$$\begin{bmatrix} \rho \boldsymbol{\mathcal{S}}^\top \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}(\boldsymbol{x}) & -\frac{1}{2} \boldsymbol{\mathcal{B}}(\boldsymbol{x}) \\ -\frac{1}{2} \boldsymbol{\mathcal{B}}(\boldsymbol{x})^\top & \tau \end{bmatrix} \succ_{\mathcal{C}} \mathbf{0}. \tag{18}$$

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

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

Using Lemma 3, we may conclude that the strict copositivity constraint in (18) is satisfied by the constructed solution $(\tau, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\gamma})$. Thus, problem (16) admits a Slater point and strong duality indeed holds for the primal and dual pair (11) and (16), respectively. $\qquad\square$

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

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

$$\begin{aligned} \text{minimize} \quad & c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \tau \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}, \ \tau \in \mathbb{R}, \ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{J'}, \ \boldsymbol{\gamma} \in \mathbb{R}^{LQ}, \ \boldsymbol{H} \in \mathbb{S}_+^{K'} \\ & \begin{bmatrix} \mathbb{I} & \boldsymbol{\mathcal{A}}(\boldsymbol{x}) \\ \boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top & \boldsymbol{H} \end{bmatrix} \succeq \mathbf{0} \\ & \begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}) \boldsymbol{\mathcal{S}} - \boldsymbol{H} - \operatorname{diag}\left([\boldsymbol{\gamma}^\top \ \mathbf{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathcal{B}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \mathbf{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathcal{B}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \mathbf{0}^\top]^\top\right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \mathbf{0} \end{aligned} \tag{19}$$

The proof of Theorem 3 relies on the following lemma, which linearizes the quadratic term $\boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}(\boldsymbol{x})$ in the left-hand side matrix of problem (16).

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 $Q \leq R$. Then the copositive inequality*

$$\boldsymbol{M} \succeq_{\mathcal{C}} \begin{bmatrix} \boldsymbol{A}^\top \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{20}$$

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

$$\boldsymbol{M} \succeq_{\mathcal{C}} \begin{bmatrix} \boldsymbol{H} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \quad and \quad \begin{bmatrix} \mathbb{I} & \boldsymbol{A} \\ \boldsymbol{A}^\top & \boldsymbol{H} \end{bmatrix} \succeq \boldsymbol{0}. \tag{21}$$

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 (21) 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 (21) then yields (20). Thus, the claim follows. $\square$

Proof of Theorem 3. Applying Theorem 2, we may replace the objective function of (3) with the corresponding copositive reformulation, we thus find that problem (3) is equivalent to

$$
\begin{aligned}
\text{minimize} \quad & c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \tau \\
\text{subject to} \quad & \boldsymbol{x} \in \mathcal{X},\ \tau \in \mathbb{R},\ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{J'},\ \boldsymbol{\gamma} \in \mathbb{R}^{LQ} \\
& \begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}) \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top\ \boldsymbol{0}^\top]^\top \right) & \frac{1}{2} \left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathscr{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top\ \boldsymbol{0}^\top]^\top \right) \\ \frac{1}{2} \left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathscr{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top\ \boldsymbol{0}^\top]^\top \right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0}.
\end{aligned}
$$

Next, we apply Lemma 4 to linearize the quadratic terms $\boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}(\boldsymbol{x})$, which gives rise to the desired copositive program (19). This completes the proof. $\square$

4 Conservative Semidefinite Programming Approximation

The copositive program (19) is intractable due to its equivalence with generic RQPs over a polyhedral uncertainty set [4]. In the copositive 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 $\mathcal{C}$ [40, 10, 16, 33]. The simplest of these approximations is given by the cone

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

which contains all symmetric matrices that can be decomposed into a sum of positive semidefinite and nonnegative matrices. For dimensions $K \leq 4$ it can be shown that $\mathcal{C}^0 = \mathcal{C}$ [18], while for $K > 4$, C^0 is a strict subset of $\mathcal{C}$.

Replacing the cone $\mathcal{C}$ in (19) with the inner approximation $\mathcal{C}^0$ gives rise to a tractable conservative approximation for the RQP (3). In this case, however, the resulting optimization problem might have no interior or even become infeasible as the Slater point 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 $Z(\boldsymbol{x})$ in (11). 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

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

that contains the lifted set Ξ' in (17). We then consider the following augmented completely positive programming reformulation for the maximization problem (9).

$$
\begin{aligned}
Z(\boldsymbol{x}) = \quad \sup \quad & \mathrm{tr}\left(\boldsymbol{\mathcal{A}}(\boldsymbol{x})\boldsymbol{\Omega}\boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top\right) + \boldsymbol{\mathcal{b}}(\boldsymbol{x})^\top \boldsymbol{\xi}' + c(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{\xi}' \in \mathbb{R}_+^{K'}, \ \boldsymbol{\Omega} \in \mathbb{S}_+^{K'} \\
& \boldsymbol{\mathcal{S}}\boldsymbol{\xi}' = \boldsymbol{t}, \ \mathrm{diag}(\boldsymbol{\mathcal{S}}\boldsymbol{\Omega}\boldsymbol{\mathcal{S}}^\top) = \boldsymbol{t} \circ \boldsymbol{t} \\
& \xi_\ell' = \Omega_{\ell\ell} \quad \forall \ell \in [LQ] \\
& \mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{\Omega}\boldsymbol{Q}^\top\right) - 2\boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{\xi}' + \boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c} \le r^2 \\
& \begin{bmatrix} \boldsymbol{\Omega} & \boldsymbol{\xi}' \\ \boldsymbol{\xi}'^\top & 1 \end{bmatrix} \succeq_{\mathcal{C}^*} \boldsymbol{0}
\end{aligned}
\tag{22}
$$

Here, we have added the redundant constraint $\mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{\Omega}\boldsymbol{Q}^\top\right) - 2\boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{\xi}' + \boldsymbol{c}^\top \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c} \le r^2$ to (11), which arises from linearizing the quadratic constraint

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

where we have set $\boldsymbol{\Omega} = \boldsymbol{\xi}'\boldsymbol{\xi}'^\top$. The dual of the augmented problem (22) is given by the following copositive program.

$$
\begin{aligned}
\overline{Z}(\boldsymbol{x}) = \quad \inf \quad & c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \lambda r^2 - \lambda\|\boldsymbol{Q}\boldsymbol{c}\|^2 + \tau \\
\text{s.t.} \quad & \tau \in \mathbb{R}, \ \lambda \in \mathbb{R}_+, \ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{J'}, \ \boldsymbol{\gamma} \in \mathbb{R}^{LQ}, \ \boldsymbol{h} \in \mathbb{R}^{K'} \\
& \begin{bmatrix} \lambda\boldsymbol{Q}^\top\boldsymbol{Q} + \boldsymbol{\mathcal{S}}^\top \mathrm{diag}(\boldsymbol{\phi})\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}(\boldsymbol{x})^\top\boldsymbol{\mathcal{A}}(\boldsymbol{x}) - \mathrm{diag}\left([\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top\right) & \frac{1}{2}\boldsymbol{h} \\ \frac{1}{2}\boldsymbol{h}^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0} \\
& \boldsymbol{h} = \boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathcal{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top - 2\lambda\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{c}
\end{aligned}
\tag{23}
$$

Note that we have $Z(\boldsymbol{x}) = \overline{Z}(\boldsymbol{x})$ 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 (23) will always yield a problem with a Slater point.

Apart from helping us prove the existence of a Slater point, adding an ellipsoidal constraint to the description of the uncertainty set can also be of help numerically. Although, the constraint is redundant for the exact problem, it might not be redundant for the conservative approximation obtained by replacing $\mathcal{C}$ with $\mathcal{C}^0$. Adding the constraint results in an additional variable λ in the SDP approximation, which can

improve the objective value. Ideally, we would like the volume of the ellipsoid to be as small as possible to get more improvement. However, determining the parameters of the ellipsoid having minimum volume that encloses the set Ξ is NP-hard. A feasible ellipsoid that can be generated tractably is $\{\boldsymbol{\xi} \in \mathbb{R}^K : \|\boldsymbol{\xi}\| \leq \|\boldsymbol{r}\|\}$, where

$$r_k = \sup_{\boldsymbol{\xi} \in \Xi} \xi_k, \quad \forall k \in [K].$$

Note that the parameter $\boldsymbol{r}$ of the ellipsoid can be determined by solving K linear programs. Depending on the specific uncertainty set at hand, it might be possible to find other tighter ellipsoidal approximations.

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

Next, we show that solving the problem by replacing $\mathcal{C}$ with the simplest inner approximation $\mathcal{C}^0$ is better than the approximate $\mathcal{S}$-lemma method. Since the latter is only valid in the case of continuous uncertain parameters, we restrict the discussion to the case where the bounded uncertainty set contains no integral terms and is given by the polytope $\Xi = \left\{\boldsymbol{\xi} \in \mathbb{R}^K_+ : \boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{t}\right\}$. Here, the extended parameters (12) simplify to

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

while the maximization problem (9) reduces to

$$Z(\boldsymbol{x}) = \sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}). \tag{24}$$

The copositive programming reformulation (23) can then be simplified to

$$
\begin{aligned}
\overline{Z}(\boldsymbol{x}) = \quad & \inf \quad c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \lambda r^2 - \lambda \|\boldsymbol{Q}\boldsymbol{c}\|^2 + \tau \\
& \text{s.t.} \quad \tau \in \mathbb{R}, \ \lambda \in \mathbb{R}_+, \ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^J \\
& \qquad \begin{bmatrix} \lambda \boldsymbol{Q}^\top \boldsymbol{Q} + \boldsymbol{S}^\top \operatorname{diag}(\boldsymbol{\phi})\boldsymbol{S} - \boldsymbol{A}(\boldsymbol{x})^\top \boldsymbol{A}(\boldsymbol{x}) & \frac{1}{2}\left(\boldsymbol{S}^\top \boldsymbol{\psi} - \boldsymbol{b}(\boldsymbol{x}) - 2\lambda \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c}\right) \\ \frac{1}{2}\left(\boldsymbol{S}^\top \boldsymbol{\psi} - \boldsymbol{b}(\boldsymbol{x}) - 2\lambda \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c}\right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0}.
\end{aligned}
\tag{25}
$$

Replacing the cone $\mathcal{C}$ in (25) with its inner approximation $\mathcal{C}^0$, we obtain a tractable SDP reformulation whose optimal value $\overline{Z}^{\mathcal{C}_0}(\boldsymbol{x})$ constitutes an upper bound on $Z(\boldsymbol{x})$. Alternatively, we describe the approximate $\mathcal{S}$-lemma method below, which provides a different conservative SDP approximation for (24).

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 maximization problem (9) is upper bounded by the optimal value of the following semidefinite program:*

$$
\begin{aligned}
\overline{Z}^{\mathcal{S}}(\boldsymbol{x}) = \quad & \inf \quad c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\theta} + \rho r^2 - \rho \|\boldsymbol{Q}\boldsymbol{c}\|^2 + \kappa \\
& \text{s.t.} \quad \kappa \in \mathbb{R}, \ \rho \in \mathbb{R}_+, \ \boldsymbol{\theta} \in \mathbb{R}^J, \ \boldsymbol{\eta} \in \mathbb{R}^J_+ \\
& \qquad \begin{bmatrix} \rho \boldsymbol{Q}^\top \boldsymbol{Q} - \boldsymbol{A}(\boldsymbol{x})^\top \boldsymbol{A}(\boldsymbol{x}) & \frac{1}{2}\left(\boldsymbol{S}^\top \boldsymbol{\theta} - \boldsymbol{b}(\boldsymbol{x}) - \boldsymbol{\eta} - 2\rho \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c}\right) \\ \frac{1}{2}\left(\boldsymbol{S}^\top \boldsymbol{\theta} - \boldsymbol{b}(\boldsymbol{x}) - \boldsymbol{\eta} - 2\rho \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c}\right)^\top & \kappa \end{bmatrix} \succeq \boldsymbol{0}.
\end{aligned}
\tag{26}
$$

15

Proof. The quadratic maximization problem in (24) can be equivalently reformulated as

$$Z(\boldsymbol{x}) = \sup \quad \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x})$$
$$\text{s.t.} \quad \boldsymbol{\xi} \in \mathbb{R}_+^K$$
$$\boldsymbol{S}\boldsymbol{\xi} = \boldsymbol{t}$$
$$\|\boldsymbol{Q}(\boldsymbol{\xi} - \boldsymbol{c})\|^2 \le r^2.$$

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

$$Z(\boldsymbol{x})$$
$$= \sup_{\boldsymbol{\xi}} \inf_{\boldsymbol{\eta} \ge \boldsymbol{0}, \rho \ge 0, \boldsymbol{\theta}} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\theta} - \boldsymbol{\xi}^\top \boldsymbol{S}^\top \boldsymbol{\theta} + \boldsymbol{\xi}^\top \boldsymbol{\eta} + \rho r^2 - \rho \|\boldsymbol{Q}(\boldsymbol{\xi} - \boldsymbol{c})\|^2$$
$$\le \inf_{\boldsymbol{\eta} \ge \boldsymbol{0}, \rho \ge 0, \boldsymbol{\theta}} \sup_{\boldsymbol{\xi}} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\theta} - \boldsymbol{\xi}^\top \boldsymbol{S}^\top \boldsymbol{\theta} + \boldsymbol{\xi}^\top \boldsymbol{\eta} + \rho r^2 - \rho \|\boldsymbol{Q}(\boldsymbol{\xi} - \boldsymbol{c})\|^2$$
$$= \inf_{\boldsymbol{\eta} \ge \boldsymbol{0}, \rho \ge 0, \boldsymbol{\theta}} c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\theta} + \rho r^2 - \rho \|\boldsymbol{Q}\boldsymbol{c}\|^2$$
$$+ \sup_{\boldsymbol{\xi}} \left(\|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} - \boldsymbol{\xi}^\top \boldsymbol{S}^\top \boldsymbol{\theta} + \boldsymbol{\xi}^\top \boldsymbol{\eta} - \rho \|\boldsymbol{Q}\boldsymbol{\xi}\|^2 + 2\rho \boldsymbol{\xi}^\top \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c} \right),$$

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

$$Z(\boldsymbol{x}) \le \inf \quad c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\theta} + \rho r^2 - \rho \|\boldsymbol{Q}\boldsymbol{c}\|^2 + \kappa$$
$$\text{s.t.} \quad \boldsymbol{\theta} \in \mathbb{R}^J, \; \boldsymbol{\eta} \in \mathbb{R}_+^K, \; \rho \in \mathbb{R}_+, \; \kappa \in \mathbb{R}$$
$$\sup_{\boldsymbol{\xi}} \left(\|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} - \boldsymbol{\xi}^\top \boldsymbol{S}^\top \boldsymbol{\theta} + \boldsymbol{\xi}^\top \boldsymbol{\eta} - \rho \|\boldsymbol{Q}\boldsymbol{\xi}\|^2 + 2\rho \boldsymbol{\xi}^\top \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c} \right) \le \kappa.$$

Reformulating the semi-infinite constraint as a semidefinite constraint then yields the desired reformulation (26). This completes the proof. $\square$

The next proposition shows that the approximation resulting from replacing the copositive cone $\mathcal{C}$ in (25) 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.*

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

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 (25).

$$\overline{Z}^{\mathcal{C}_0}(\boldsymbol{x}) = \inf \quad c(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \lambda r^2 - \lambda \|\boldsymbol{Q}\boldsymbol{c}\|^2 + \tau$$
$$\text{s.t.} \quad \tau \in \mathbb{R}, \; , \; \lambda, h \in \mathbb{R}_+, \; \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^J, \; \boldsymbol{F} \in \mathbb{R}_+^{K \times K}, \; \boldsymbol{g} \in \mathbb{R}_+^K$$
$$\begin{bmatrix} \lambda \boldsymbol{Q}^\top \boldsymbol{Q} + \boldsymbol{S}^\top \operatorname{diag}(\boldsymbol{\phi})\boldsymbol{S} - \boldsymbol{A}(\boldsymbol{x})^\top \boldsymbol{A}(\boldsymbol{x}) & \frac{1}{2}\left(\boldsymbol{S}^\top \boldsymbol{\psi} - \boldsymbol{b}(\boldsymbol{x}) - 2\lambda \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c} \right) \\ \frac{1}{2}\left(\boldsymbol{S}^\top \boldsymbol{\psi} - \boldsymbol{b}(\boldsymbol{x}) - 2\lambda \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{c} \right)^\top & \tau \end{bmatrix} \succeq \begin{bmatrix} \boldsymbol{F} & \boldsymbol{g} \\ \boldsymbol{g}^\top & h \end{bmatrix}$$
$$(27)$$

Next, we show that any feasible solution $(\kappa, \rho, \boldsymbol{\theta}, \boldsymbol{\eta})$ to (26) can be used to construct a feasible solution $(\tau, \lambda, h, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{F}, \boldsymbol{g})$ to (27) with the same objective value. Specifically, we set $\tau = \kappa$, $\lambda = \rho$, $h = 0$, $\boldsymbol{\psi} = \boldsymbol{\theta}$, $\boldsymbol{\phi} = \boldsymbol{0}$, $\boldsymbol{F} = \boldsymbol{0}$, and $\boldsymbol{g} = \boldsymbol{\eta}$. The feasibility of the solution $(\kappa, \rho, \boldsymbol{\theta}, \boldsymbol{\eta})$ in (26) then implies that the constructed solution $(\tau, \lambda, h, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{F}, \boldsymbol{g})$ is also feasible in (27). One can verify that these solutions give rise to the same objective function value for the respective problems. Thus, the claim follows. $\qquad\square$

Next, we demonstrate that the inequality in $\overline{Z}^0(\boldsymbol{x}) \leq \overline{Z}^{\mathcal{S}}(\boldsymbol{x})$ in Proposition 4 can often be strict. This affirms that the proposed SDP approximation (27) is indeed stronger than the approximate $\mathcal{S}$-lemma method.

Example 3. *Consider the following quadratic maximization problem:*

$$
\begin{aligned}
Z(\boldsymbol{x}) = \quad &\sup \quad \xi_1^2 \\
&\text{s.t.} \quad \boldsymbol{\xi} \in \mathbb{R}_+^2 \\
&\qquad\quad 2\xi_1 + \xi_2 = 2.
\end{aligned}
\tag{28}
$$

A simple analysis shows that $Z(\boldsymbol{x}) = 1$, which is attained at the solution $(\xi_1, \xi_2) = (1, 0)$. The problem (28) constitutes an instance of problem (24) with the parameterizations

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

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 $\boldsymbol{S} = \begin{bmatrix} 2 & 1 \end{bmatrix}$ and $\boldsymbol{t} = 2$. Replacing the cone $\mathcal{C}$ with its inner approximation $\mathcal{C}^0$ in the copositive programming reformulation of (28), we find that the resulting semidefinite program yields the same optimal objective value of $\overline{Z}^{\mathcal{C}_0}(\boldsymbol{x}) = 1$. Meanwhile, the corresponding approximate $\mathcal{S}$-lemma method yields an optimal objective value $Z^{\mathcal{S}}(\boldsymbol{x}) = 4$. Thus, while the SDP approximation of the copositive program (25) is tight, the approximate $\mathcal{S}$-lemma generates an inferior objective value for the simple instance (28).

5 Extensions

In this section, we discuss several extensions to the RQP (3) which are also amenable to exact copositive programming reformulation. In Section 5.1, we study two-stage robust optimization with mixed-integer uncertainty set where the objective is quadratic in the first- and the second-stage decision variables. In Section 5.2, we develop an extension to the case when the model has robust quadratic constraints. Finally, in Section 5.3, we discuss the case where the objective function contains quadratic terms which are not convex in the uncertain parameter vector $\boldsymbol{\xi}$.

5.1   Two-Stage Robust Quadratic Optimization

In this section, we study the two-stage robust quadratic optimization problems of the form

$$
\begin{aligned}
\text{minimize} \quad & \sup_{\boldsymbol{\xi}\in\Xi} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}(\boldsymbol{x})^\top\boldsymbol{\xi} + c(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x},\boldsymbol{\xi}) \\
\text{subject to} \quad & \boldsymbol{x}\in\mathcal{X}.
\end{aligned}
\tag{29}
$$

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

$$
\begin{aligned}
\mathcal{R}(\boldsymbol{x},\boldsymbol{\xi}) = \quad \inf \quad & \|\boldsymbol{P}\boldsymbol{y}\|^2 + (\boldsymbol{R}\boldsymbol{\xi}+\boldsymbol{r})^\top\boldsymbol{y} \\
\text{s.t.} \quad & \boldsymbol{y}\in\mathbb{R}^{D_2} \\
& \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}+\boldsymbol{h}(\boldsymbol{x}) \le \boldsymbol{W}\boldsymbol{y},
\end{aligned}
\tag{30}
$$

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

$$
\begin{aligned}
\text{minimize} \quad & \lambda\|\boldsymbol{w}\|^2 + \sum_{n\in[N]} \max\left\{0, 1-\hat{\xi}_n(\boldsymbol{w}^\top\hat{\boldsymbol{\chi}}_n - w_0)\right\} \\
\text{subject to} \quad & \boldsymbol{w}\in\mathbb{R}^K,\ w_0\in\mathbb{R}
\end{aligned}
\tag{31}
$$

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 (31) 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 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 (31) as a convex surrogate. Here, the term $\lambda\|\boldsymbol{w}\|^2$ in the objective function constitutes a regularizer for the coefficient $\boldsymbol{w}$.

If the labels $\{\hat{\xi}_n\}_{n\in[N]}$ are erroneous, then one could envisage a robust optimization model that seeks the best linear separator in view of 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

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

where

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

18

This problem constitutes an instance of (29) *with the decision vector* $\boldsymbol{x} = (\boldsymbol{w}, w_0)$, *and the input parameters*

$$\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{0}, \quad \boldsymbol{b}(\boldsymbol{x}) = \boldsymbol{0}, \quad c(\boldsymbol{x}) = \lambda\|\boldsymbol{w}\|^2, \quad \boldsymbol{P} = \boldsymbol{0}, \quad \boldsymbol{R} = \boldsymbol{0}, \quad \boldsymbol{r} = \mathbf{e},$$

$$\boldsymbol{T}(\boldsymbol{x}) = -\operatorname{diag}\left(\begin{bmatrix} \boldsymbol{w}^\top \hat{\chi}_1 \\ \vdots \\ \boldsymbol{w}^\top \hat{\chi}_N \end{bmatrix}\right) - w_0 \mathbb{I}, \quad \boldsymbol{h}(\boldsymbol{x}) = \mathbf{e}, \quad and \quad \boldsymbol{W} = \mathbb{I}.$$

The exactness result portrayed in Theorems 2 and 3 can be extended to the two-stage robust optimization problem (29). Specifically, if the problem has a *complete recourse*[1] then, by employing Theorem 2 and extending the techniques developed in [29, Theorem 4], the two-stage problem (29) 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* (29) *is equivalent to the copositive program*

$$\text{minimize} \quad c(\boldsymbol{x}) - \frac{1}{4}\boldsymbol{r}^\top (\boldsymbol{P}^\top \boldsymbol{P})^{-1}\boldsymbol{r} + \boldsymbol{t}^\top \boldsymbol{\psi} + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi} + \tau$$

$$\text{subject to} \quad \boldsymbol{x} \in \mathcal{X}, \ \tau \in \mathbb{R}, \ \boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{J'}, \ \boldsymbol{\gamma} \in \mathbb{R}^{LQ}, \ \boldsymbol{H} \in \mathbb{S}_+^{K'}$$

$$\begin{bmatrix} \mathbb{I} & \mathcal{A}(\boldsymbol{x}) \\ \mathcal{A}(\boldsymbol{x})^\top & \boldsymbol{H} \end{bmatrix} \succeq \boldsymbol{0} \tag{32}$$

$$\begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi})\boldsymbol{\mathcal{S}} - \boldsymbol{H} - \boldsymbol{\mathcal{P}}(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{b}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{b}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top\right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0},$$

where

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

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

---

[1]The two-stage problem (29) has complete recourse if there exists $\boldsymbol{y}^+ \in \mathbb{R}^{D_2}$ with $\boldsymbol{W}\boldsymbol{y}^+ > \boldsymbol{0}$, which implies that the second-stage subproblem is feasible for every $\boldsymbol{x} \in \mathbb{R}^{D_1}$ and $\boldsymbol{\xi} \in \mathbb{R}^K$.

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

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

with

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

Proof. Since P has full column rank, the matrix $P^\top P$ is positive definite. Thus, for any fixed $x \in \mathcal{X}$ and $\xi \in \Xi$, the recourse problem (30) admits a dual quadratic program given by

$$
\begin{aligned}
\mathcal{R}(x, \xi) = \quad \sup \quad & -\frac{1}{4}\left((W^\top\theta - R\xi - r)^\top(P^\top P)^{-1}(W^\top\theta - R\xi - r)\right) + h(x)^\top\theta + \xi^\top T(x)^\top\theta \\
\text{s.t.} \quad & \theta \in \mathbb{R}_+^T.
\end{aligned}
\tag{33}
$$

Strong duality holds as the two-stage problem (29) has complete recourse. Substituting the dual formulation (33) into the objective of (29) yields

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

Thus, for any fixed $x \in \mathcal{X}$, the objective value of the two-stage problem (29) 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 (32). This completes the proof. $\square$

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

$$\begin{bmatrix} \Lambda_+ & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where Λ_+ is a diagonal matrix whose main diagonal comprises the non-zero eigenvalues of $P^\top P$. Next, by using the constructed eigendecomposition and performing the change of variable $z \leftarrow U^{-1}y$, we can reformulate the recourse problem (30) equivalently as

$$
\begin{aligned}
\mathcal{R}(x, \xi) = \quad \inf \quad & z_+^\top \Lambda_+ z_+ + (R\xi + r)^\top U_+ z_+ + (R\xi + r)^\top U_0 z_0 \\
\text{s.t.} \quad & (z_+, z_0) \in \mathbb{R}^{D_2} \\
& T(x)\xi + h(x) \leq WU_+ z_+ + WU_0 z_0,
\end{aligned}
$$

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:*

$$\mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi}) = \quad \sup \quad -\frac{1}{4}\left((\boldsymbol{W}^\top \boldsymbol{\theta} - \boldsymbol{R}\boldsymbol{\xi} - \boldsymbol{r})^\top \boldsymbol{U}_+^\top \boldsymbol{\Lambda}_+^{-1} \boldsymbol{U}_+ (\boldsymbol{W}^\top \boldsymbol{\theta} - \boldsymbol{R}\boldsymbol{\xi} - \boldsymbol{r})\right) + \boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\theta} + \boldsymbol{\xi}^\top \boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\theta}$$

$$\text{s.t.} \quad \boldsymbol{\theta} \in \mathbb{R}_+^T$$

$$\boldsymbol{U}_0^\top (\boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{r}) = \boldsymbol{U}_0^\top \boldsymbol{W}^\top \boldsymbol{\theta}.$$

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.

5.2 Robust Quadratically Constrained Quadratic Programming (RQCQP)

The setting that we consider can be extended to the case where, in addition to the robust quadratic objective function, there are several robust quadratic constraints of the form

$$\sup_{\boldsymbol{\xi} \in \Xi} \left\{ \|\boldsymbol{A}_i(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{b}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + c_i(\boldsymbol{x}) \right\} \leq 0 \qquad \forall i \in [I]. \tag{34}$$

In this case, the goal is to find a decision $\boldsymbol{x} \in \mathcal{X}$ which minimizes the worst-case objective function, while ensuring that the quadratic constraints are satisfied for all possible uncertain parameter vectors in Ξ.

For every $i \in [I]$, we define $\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})$ and $\boldsymbol{\mathcal{b}}_i(\boldsymbol{x})$ similarly to the definitions of the extended parameters $\boldsymbol{\mathcal{A}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{b}}(\boldsymbol{x})$ in (12). By applying Theorem 2, the quadratic maximization problem in the i-th constraint of (34) can be replaced with a copositive minimization problem, which yields the constraint

$$0 \geq \quad \inf \quad c_i(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi}_i + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi}_i + \tau_i$$

$$\text{s.t.} \quad \tau_i \in \mathbb{R}, \ \boldsymbol{\psi}_i, \boldsymbol{\phi}_i \in \mathbb{R}^{J'}, \ \boldsymbol{\gamma}_i \in \mathbb{R}^{LQ}$$

$$\begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}_i)\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}_i^\top \ \boldsymbol{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top \ \boldsymbol{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top \ \boldsymbol{0}^\top]^\top\right)^\top & \tau_i \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0}.$$

The constraint 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

$$c_i(\boldsymbol{x}) + \boldsymbol{t}^\top \boldsymbol{\psi}_i + (\boldsymbol{t} \circ \boldsymbol{t})^\top \boldsymbol{\phi}_i + \tau_i \leq 0,$$

$$\begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi}_i)\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top \ \boldsymbol{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi}_i - \boldsymbol{\mathcal{b}}_i(\boldsymbol{x}) + [\boldsymbol{\gamma}_i^\top \ \boldsymbol{0}^\top]^\top\right)^\top & \tau_i \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0} \tag{35}$$

is satisfied. Therefore the i-th constraint of (34) can be replaced by the constraint system (35). The procedure for linearization of the quadratic terms $\boldsymbol{\mathcal{A}}_i(\boldsymbol{x})^\top \boldsymbol{\mathcal{A}}_i(\boldsymbol{x})$ is analogous to the method presented in Theorem 3.

5.3 Non-Convex Terms in the Objective Function

All exactness results in this paper extend immediately to the setting where the objective function in (3) involves non-convex quadratic terms in the uncertainty $\boldsymbol{\xi}$. Specifically, we consider the objective function

$$Z(\boldsymbol{x}) = \sup_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\xi}\|^2 + \boldsymbol{\xi}^\top \boldsymbol{D}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\xi} + c(\boldsymbol{x}),$$

where $\boldsymbol{D}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{S}^K$ is a matrix-valued affine function of $\boldsymbol{x}$. We can still use Theorem 1 to reformulate $Z(\boldsymbol{x})$ as the optimal value of a copositive program. By following the steps of Proposition 1 and Theorem 3, the copositive programming reformulation is obtained by replacing the last constraint in (19) with the copositive constraint

$$\begin{bmatrix} \boldsymbol{\mathcal{S}}^\top \operatorname{diag}(\boldsymbol{\phi})\boldsymbol{\mathcal{S}} - \boldsymbol{H} - \boldsymbol{\mathcal{D}}(\boldsymbol{x}) - \operatorname{diag}\left([\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top\right) & \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathscr{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top\right) \\ \frac{1}{2}\left(\boldsymbol{\mathcal{S}}^\top \boldsymbol{\psi} - \boldsymbol{\mathscr{b}}(\boldsymbol{x}) + [\boldsymbol{\gamma}^\top \ \boldsymbol{0}^\top]^\top\right)^\top & \tau \end{bmatrix} \succeq_{\mathcal{C}} \boldsymbol{0},$$

where

$$\boldsymbol{\mathcal{D}}(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{D}(\boldsymbol{x}) \end{bmatrix} \in \mathbb{S}^{K'}.$$

We omit the details for the sake of brevity.

6 Numerical Experiments

In this section, we assess the performance of the SDP approximations presented in Section 4. All optimization problems are solved using the YALMIP interface [35] on a 16-core 3.4 GHz computer with 32 GB RAM. We use MOSEK 8.1 to solve SDP formulations, and CPLEX 12.8 to solve integer programs and non-convex quadratic programs.

6.1 Least Squares

The classical least squares problem seeks an approximate solution $\boldsymbol{x}$ to an overdetermined linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ which minimizes the *residual* $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$. This yields the following quadratic program:

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \\ \text{subject to} \quad & \boldsymbol{x} \in \mathbb{R}^N. \end{aligned}$$

The solution to this problem can be very sensitive to perturbations in the input data $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ and $\boldsymbol{b} \in \mathbb{R}^M$ [19, 26]. To address the issue of parameter uncertainty, El Ghaoui and Lebret [23] recommend

solving the following robust optimization problem:

$$\begin{array}{ll}\text{minimize} & \sup_{(\boldsymbol{U},\boldsymbol{v})\in\mathcal{U}} \|(\boldsymbol{A}+\boldsymbol{U})\boldsymbol{x} - (\boldsymbol{b}+\boldsymbol{v})\|^2 \\ \text{subject to} & \boldsymbol{x}\in\mathbb{R}^N.\end{array} \tag{36}$$

Here, the goal is to find a solution $\boldsymbol{x}$ that minimizes the worst-case residual when the matrix $\boldsymbol{U}$ and the vector $\boldsymbol{v}$ can vary within the prescribed uncertainty set $\mathcal{U}$. A tractable SDP reformulation of this problem is derived in [23] for problem instances where the uncertainty set is given by the Frobenius norm ball

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

We consider the case when the uncertainty set is a polytope, and compare our SDP scheme with the state-of-the-art approximate $\mathcal{S}$-lemma method described in Section 4.1. We also compare our method with the approximation scheme proposed by Bertsimas and Sim [7], where the worst-case quadratic term in (2) is replaced with an upper bounding function. Minimizing this upper-bounding function over $\boldsymbol{x}$ yields an approximate solution to the RQP. We note that the robust least squares problem can be solved to optimality using Benders' constraint generation method [8]. However, doing so entails solving a non-convex quadratic optimization problem at each step to generate a valid cut, which becomes intractable when M and N become large.

In our experiment, we consider the case where the uncertainty affects only the right-hand side vector $\boldsymbol{b}$ (*i.e.*, $\boldsymbol{U} = \boldsymbol{0}$). We assume that the uncertain parameter $\boldsymbol{v}$ depends affinely on N_f *factors* represented by $\boldsymbol{\xi} \in \mathbb{R}^{N_f}$, where $N_f < M$. Specifically, we consider the uncertainty set

$$\mathcal{U} = \left\{ \boldsymbol{v}\in\mathbb{R}^M : \boldsymbol{v} = \boldsymbol{F}\boldsymbol{\xi},\ \boldsymbol{\xi}\in\mathbb{R}^{N_f},\ \|\boldsymbol{\xi}\|_\infty \leq 1,\ \|\boldsymbol{\xi}\|_1 \leq \rho N_f \right\},$$

where $\boldsymbol{F} \in \mathbb{R}^{M\times N_f}$ is the factor matrix and ρ lies in the interval $[0,1]$. By substituting $\boldsymbol{U} = \boldsymbol{0}$ and $\boldsymbol{v} = \boldsymbol{F}\boldsymbol{\xi}$ into (36), the resulting robust problem constitutes an instance of RQP (3) with the following input parameters:

$$\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{F}, \quad \boldsymbol{b}(\boldsymbol{x}) = -2\boldsymbol{F}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}), \quad c(\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}),$$

$$\Xi = \left\{ \boldsymbol{\xi}\in\mathbb{R}^{N_f} : \|\boldsymbol{\xi}\|_\infty \leq 1,\ \|\boldsymbol{\xi}\|_1 \leq \rho N_f \right\}.$$

In order to solve the problem using our method, we modify the formulation discussed in Section 4 slightly, which leads to a tremendous reduction in the solution time. We discuss this modification in Appendix A.

We perform an experiment on problem instances of dimensions $M = 200$, $N = 20$ and $N_f = 30$. The experimental results are averaged over 100 random trials generated in the following manner. In each trial, we sample the matrix $\boldsymbol{A}$ and the vector $\boldsymbol{b}$ from the uniform distribution on $[-0.5, 0.5]^{M\times N}$ and $[-0.5, 0.5]^N$, respectively. Each row of the matrix $\boldsymbol{F}$ is sampled randomly from a standard simplex, and ρ is generated uniformly at random from the interval $[0.1, 0.25]$. For problems of this size, we are unable to solve the problem to optimality using Benders' method as the solver runs out of memory. Therefore, we put a time

limit of 120 seconds for each iteration of the Benders' method. By doing so, Benders' method yields a lower bound to the optimal worst-case residual, which we use as a baseline to compute the objective gaps for the approximation methods.

Table 1 summarizes the optimality gaps of the approximation methods. The results show that our method significantly outperforms the other two approximations in terms of the estimates of the worst-case residuals. While the other two approximations generate overly pessimistic estimates of the resulting worst-case residuals (with a relative difference of about 100%), the worst-case residuals estimated using our method have negligible objective gaps.

Table 2 reports the solution times of finding the exact solution (using Benders' method) and the upper bounds provided by various approximation methods. It can be observed that the improvement in solution quality given by our method comes at the cost of longer solution times compared to other approximation methods. However, our method is still significantly faster than the exact Benders' method. We also note that while the approximation scheme described in [7] can be solved quickly, it is only valid when the uncertainty set is defined as a *norm-bounded set* ($l_1 \cap l_\infty$ norm in our experiment). Our method, on the other hand, is applicable for general polyhedral uncertainty sets.

	Objective gap		
Statistic	SDP	$\mathcal{S}$-lemma	B&S
Mean	0.0%	108.4%	99.7%
10th Percentile	0.0%	93.9%	80.5%
90th Percentile	0.0%	119.6%	115.3%

Table 1. Numerical results comparing the proposed SDP approximation ('SDP'), the approximate $\mathcal{S}$-lemma method ('$\mathcal{S}$-lemma') and the approximation scheme proposed by Bertsimas and Sim [7] ('B&S') for the least squares problem. The 'objective gap' quantifies the increase in the worst-case residuals estimated using the approximation methods relative to the Benders' lower bound.

	Benders	SDP	$\mathcal{S}$-lemma	B&S
Mean solution time (in secs)	626.9	10.2	0.45	0.004

Table 2. Solution times for the Benders' constraint generation method ('Benders'), the proposed SDP approximation ('SDP'), the approximate $\mathcal{S}$-lemma method ('$\mathcal{S}$-lemma') and the approximation scheme proposed by Bertsimas and Sim [7] ('B&S') for the least squares problem.

6.2 Project Management

In this experiment, we consider the 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. We consider randomly generated project networks of size $|\mathcal{V}| = 30$ and *order strength* 0.75,[2] which gives rise to projects with an average of 67 activities. Let x_{ij} be the amount of resources that are used to expedite the activity (i,j). We fix the feasible set of the resource allocation vector to $\mathcal{X} = \{\boldsymbol{x} \in [0,1]^{|\mathcal{A}|} : \mathbf{e}^\top \boldsymbol{x} \leq \frac{3}{4}|\mathcal{A}|\}$, so that at most 75% of the activities can receive the maximum resource allocation. The uncertainty set of $\boldsymbol{d}$ is defined through a factor model as follows:

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

where the factor size is fixed to $N_f = |\mathcal{V}|$. We set the nominal task durations to $\boldsymbol{d}^0 = \mathbf{e}$. In each trial, we sample the factor loading vector $\boldsymbol{f}_{ij}$ from the uniform distribution on $[-\frac{1}{2N_f}, \frac{1}{2N_f}]^{N_f}$, 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 the procedure described in Example 2 (Equation (8)).

In our experiment, we compare the performance of our proposed SDP approximation with *linear decision rules* (LDR) approximation scheme discussed in [14, 48] which we describe below. In Example 2, for our reformulation, we model the second stage problem as the maximization problem over the binary variables $\boldsymbol{z}$ (See Equation (6)). Alternatively, the second-stage problem can be written as the following minimization problem:

$$\begin{aligned} \text{minimize} \quad & \rho_{|\mathcal{V}|} - \rho_1 \\ \text{subject to} \quad & \boldsymbol{\rho} \in \mathbb{R}^{|\mathcal{V}|}, \\ & \rho_j - \rho_i \geq d_{ij} - x_{ij}, \forall (i,j) \in \mathcal{A}. \end{aligned}$$

Here, $\boldsymbol{\rho}$ is second-stage variable which depends on the realization of the uncertain $\boldsymbol{d}$. In the LDR approximation scheme, $\boldsymbol{\rho}$ is restricted to be an affine function of $\boldsymbol{d}$, which yields a tractable conservative approximation. To assess the suboptimality of our SDP and the LDR approximation scheme, we solve the problem to optimality using Benders' constraint generation method.

Table 3 presents the optimality gaps of the two approximation methods for 100 randomly generated project networks. The solution times of all the methods are reported in Table 4. It can be observed that our proposed SDP approximation consistently provides near-optimal estimates of the worst-case project makespan ($\sim 2.7\%$ gaps). On the other hand, while the LDR bound can be computed quickly, the bounds are too pessimistic ($\sim 27\%$ gaps). The 10th and 90th percentiles of the objective gaps further indicate that the estimated makespan generated from our SDP approximation *stochastically dominates* the makespan generated from the LDR approximation. In addition to a higher estimate of the worst-case makespan, the

---

[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 (either directly or through transitivity).

actual makespan of the resource allocation $\boldsymbol{x}$ generated by the LDR approximation is also higher than the ones generated by our method, as shown in the "Suboptimality" column in Table 3. The experimental results demonstrate that our method generates near-optimal solutions to the project crashing problem faster than solving the problem to optimality using Benders' method.

Statistic	Objective gap		Suboptimality	
	SDP	LDR	SDP	LDR
Mean	2.7%	26.9%	1.7%	10.0%
10th Percentile	2.0%	23.8%	1.3%	7.1%
90th Percentile	3.2%	30.2%	2.2%	12.8%

Table 3. Numerical results for the proposed SDP approximation ('SDP') and the linear decision rules approximation ('LDR') for the project crashing problem. The 'objective gap' quantifies the increase in the worst-case makespan estimated using the approximation methods relative to the optimal worst-case makespan. The 'suboptimality' quantifies the increase in the actual worst-case makespan of the resource allocations found using the approximation methods relative to the optimal worst-case makespan.

	Benders	SDP	LDR
Mean solution time (in secs)	518.0	85.0	0.16

Table 4. Solution times for the Benders' constraint generation method ('Benders'), the proposed SDP approximation ('SDP') and the linear decision rules approximation ('LDR') for the project crashing problem.

6.3 Multi-Item Newsvendor

We now demonstrate the advantage of using a mixed-integer uncertainty set over using a continuous uncertainty set in a variant of the multi-item newsvendor problem, where an inventory planner must determine the vector $\boldsymbol{x} \in \mathbb{R}_+^N$ of order quantities for N different raw-materials at the beginning of a planning period. The raw materials are used to make K different types of products which are then sold to customers. The matrix $\boldsymbol{F} \in \mathbb{R}^{N \times K}$ is such that F_{nk} represents the amount of raw material n required to make 1 unit of product k. The demands $\boldsymbol{\xi} \in \mathbb{Z}_+^K$ for 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 raw materials but the total order quantity must not exceed a given budget B. Excess inventory of the n-th raw material incurs a per-unit holding cost of g_n, while the unmet demand incurs a quadratic penalty with coefficient λ. The quadratic

penalty on the unmet demand is added to discourage stock-outs [30, 49].

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

$$
\begin{aligned}
\mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi}) &= \sum_{n=1}^{N} g_n \left(\boldsymbol{x}_n - \sum_{k=1}^{K} F_{nk} \xi_k \right)^+ + \lambda \sum_{n=1}^{N} \left(\left(\sum_{k=1}^{K} F_{nk} \xi_k - \boldsymbol{x}_n \right)^+ \right)^2 \\
&= \inf_{\boldsymbol{y}_1 \in \mathbb{R}^N, \boldsymbol{y}_2 \in \mathbb{R}^N} \left\{ \boldsymbol{g}^\top \boldsymbol{y}_1 + \lambda \boldsymbol{y}_2^\top \boldsymbol{y}_2 \ : \ \boldsymbol{y}_1 \geq \boldsymbol{x} - \boldsymbol{F}\boldsymbol{\xi}, \ \boldsymbol{y}_1 \geq \boldsymbol{0}, \ \boldsymbol{y}_2 \geq \boldsymbol{F}\boldsymbol{\xi} - \boldsymbol{x}, \ \boldsymbol{y}_2 \geq \boldsymbol{0} \right\}.
\end{aligned}
$$

Here, we use the notation z^+ to denote $\max\{z, 0\}$. The objective of a risk-averse inventory planner is then to determine a vector of order quantities $\boldsymbol{x}$ 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

$$
\begin{aligned}
\text{minimize} \quad & \sup_{\boldsymbol{\xi} \in \Xi} \mathcal{R}(\boldsymbol{x}, \boldsymbol{\xi}) \\
\text{subject to} \quad & \boldsymbol{x} \in \mathbb{R}_+^N \\
& \mathbf{e}^\top \boldsymbol{x} \leq B.
\end{aligned}
\tag{37}
$$

This problem constitutes an instance of the two-stage robust quadratic optimization problem (29) with parameters

$$
\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{0}, \ \boldsymbol{b}(\boldsymbol{x}) = \boldsymbol{0}, \ c(\boldsymbol{x}) = 0, \ \boldsymbol{P} = \sqrt{\lambda} \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \mathbb{I} \end{bmatrix}, \ \boldsymbol{R} = \boldsymbol{0}, \ \boldsymbol{r} = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix},
$$

$$
\boldsymbol{T}(\boldsymbol{x}) = \begin{bmatrix} -\boldsymbol{F} \\ \boldsymbol{0} \\ \boldsymbol{F} \\ \boldsymbol{0} \end{bmatrix}, \ \boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{0} \\ -\boldsymbol{x} \\ \boldsymbol{0} \end{bmatrix}, \ \text{and} \ \boldsymbol{W} = \begin{bmatrix} \mathbb{I} & \boldsymbol{0} \\ \mathbb{I} & \boldsymbol{0} \\ \boldsymbol{0} & \mathbb{I} \\ \boldsymbol{0} & \mathbb{I} \end{bmatrix}.
$$

In this experiment, we compare the performance of the SDP approximation of the optimization problem (37) when $\boldsymbol{\xi}$ is explicitly modeled as a discrete vector versus the model where the integer restriction on $\boldsymbol{\xi}$ is ignored. We consider problems with $N = 8$ raw materials and $K = 5$ products. We fix the vector of holding costs to $\boldsymbol{g} = \mathbf{e}$, the ordering budget to $B = 20$, and the penalty constant to $\lambda = 10$. All experimental results are averaged over 100 random trials generated in the following manner. We assume that every product uses one unit each of two randomly chosen raw materials. In each trial, we generate every element of $\boldsymbol{G} \in \mathbb{R}^{2 \times K}$ uniformly at random from the interval $[0, 1]$. We define the actual discrete uncertainty set (Ξ_{True}) and the set formed by ignoring the integrality assumption (Ξ_{Cont}) as:

$$
\Xi_{\text{True}} = \left\{ \boldsymbol{\xi} \in \mathbb{Z}_+^K : \boldsymbol{\xi} \leq 15\mathbf{e}, \ \boldsymbol{G}\boldsymbol{\xi} \leq 0.75\mathbf{e} \right\} \quad \text{and} \quad \Xi_{\text{Cont}} = \left\{ \boldsymbol{\xi} \in \mathbb{R}_+^K : \boldsymbol{\xi} \leq 15\mathbf{e}, \ \boldsymbol{G}\boldsymbol{\xi} \leq 0.75\mathbf{e} \right\},
$$

and solve the SDP approximations of (37) with inputs Ξ_{True} and Ξ_{Cont}. We use the Benders' constraint generation method to solve the problem to optimality.

The statistics of the optimality gaps generated by the models using Ξ_{True} and Ξ_{Cont} are reported in Table 5. The solution times of all the methods are presented in Table 6. We observe that the model using

Ξ_{True} as the uncertainty set provides much better estimates of the worst-case cost ($\sim 13\%$ average gap) than the model using Ξ_{Cont} ($\sim 85\%$ average gap). Furthermore, our proposed SDP approximation can be solved much faster than solving the problem exactly using Benders' method. For problems with integer uncertainty, these experimental results suggest that the SDP approximation which utilizes the integer restriction gives high-quality solutions in comparison to the approximation which neglects these restrictions.

	Objective gap		Suboptimality	
Statistic	SDP True	SDP Cont	SDP True	SDP Cont
Mean	13.1%	85.2%	13.0%	84.9%
10th Percentile	0.0%	25.8%	0.0%	25.7%
90th Percentile	28.4%	173.7%	27.6%	173.5%

Table 5. Numerical results for the SDP approximations for the newsvendor model with integer uncertainty set ('SDP True') and the model that ignores the integrality restriction ('SDP Cont'). The 'objective gap' quantifies the increase in the worst-case cost estimated using the approximation methods relative to the optimal worst-case cost. The 'suboptimality' quantifies the increase in the actual worst-case cost of the order quantities found using the approximation methods relative to the optimal worst-case cost.

	Benders	SDP True	SDP Cont
Mean solution time (in secs)	52.9	11.3	0.63

Table 6. Solution times for the Benders' constraint generation method for the newsvendor problem ('Benders'), the SDP approximations for the model with integer uncertainty set ('SDP True') and the model that ignores the integrality restriction ('SDP Cont').

7 Conclusion

The paper aims at developing a near-optimal approximation method for one- and two-stage robust quadratic programs with mixed-integer uncertain parameters. The approximation method developed in the paper is not only more general than the current state-of-the-art approximate $\mathcal{S}$-lemma method—since the latter only handles continuous uncertain parameters—but is guaranteed to yield a better estimate of the optimal value. Furthermore, our numerical experiments show that the difference in the performance of the two approximation method can be quite significant. Our experimental results also demonstrate the disadvantage of ignoring the integer restrictions on the uncertain parameters. In the future, it would be interesting to

extend the model to the distributionally robust setting, where additional information about the distribution of the uncertain parameters is explicitly incorporated.

A Implementation of Least Squares in Section 6.1

In this section, we limit the discussion to the case when there are no discrete uncertain parameters. In the paper, we consider the uncertainty set to be of the standard form $\Xi_S := \{\boldsymbol{\xi} \geq \mathbf{0} : \boldsymbol{S\xi} = \boldsymbol{t}\}$. However, in some cases, the uncertainty sets are more naturally represented in the inequality form $\Xi_I := \{\boldsymbol{\xi} : \boldsymbol{S\xi} \leq \boldsymbol{t}\}$. Transforming the uncertainty set in standard form involves introducing additional variables and constraints which increases the problem size. As an example, in the least squares experiment in Section 6.1, we consider the uncertainty set $\Xi = \left\{\boldsymbol{\xi} \in \mathbb{R}^{N_f} : \|\boldsymbol{\xi}\|_\infty \leq 1, \|\boldsymbol{\xi}\|_1 \leq \rho N_f \right\}$. By lifting, the uncertainty set can be equivalently written as $\Xi_{LS} = \left\{(\boldsymbol{\xi}, \boldsymbol{\gamma}) : \boldsymbol{\xi} \in \mathbb{R}^{N_f}, \boldsymbol{\gamma} \in \mathbb{R}^{N_f}, -\boldsymbol{\xi} \leq \boldsymbol{\gamma}, \boldsymbol{\xi} \leq \boldsymbol{\gamma}, \boldsymbol{\gamma} \leq \mathbf{e}, \mathbf{e}^\top \boldsymbol{\gamma} \leq \rho N_f \right\}$, which is of the form Ξ_I. The paper [12] presents a *generalized copositive programming* (GCP) reformulation of non-convex quadratic programs over conic representable sets. In [50], the authors consider a conservative approximation when the cone is polyhedral, which is relevant for the polyhedral uncertainty sets that we consider. Utilizing this GCP-based approximation, the robust least squares problem

$$
\begin{aligned}
\text{minimize} \quad & \sup_{\boldsymbol{\xi} \in \Xi_I} \|\boldsymbol{Ax} - (\boldsymbol{b} + \boldsymbol{F\xi})\|^2 \\
\text{subject to} \quad & \boldsymbol{x} \in \mathbb{R}^N
\end{aligned}
$$

that we consider in Section 6.1 yields the following conservative SDP approximation:

$$
\begin{aligned}
\text{minimize} \quad & \tau + (\boldsymbol{Ax} - \boldsymbol{b})^\top (\boldsymbol{Ax} - \boldsymbol{b}) \\
\text{subject to} \quad & \boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{\mu} \in \mathbb{R}^J, \boldsymbol{N} \in \mathbb{R}^{J \times J}, \tau \in \mathbb{R} \\
& \boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{N} \geq \mathbf{0}, \\
& \begin{bmatrix} -\boldsymbol{F}^\top \boldsymbol{F} & \boldsymbol{F}^\top(\boldsymbol{Ax} - \boldsymbol{b}) \\ (\boldsymbol{Ax} - \boldsymbol{b})^\top \boldsymbol{F} & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{S}^\top \boldsymbol{\mu} \\ \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{S} & \tau - \boldsymbol{\mu}^\top \boldsymbol{t} \end{bmatrix} \succeq \begin{bmatrix} -\boldsymbol{S}^\top \\ \boldsymbol{t}^\top \end{bmatrix} \boldsymbol{N} \begin{bmatrix} -\boldsymbol{S} & \boldsymbol{t} \end{bmatrix}.
\end{aligned}
\tag{38}
$$

We use this formulation with the uncertainty set Ξ_{LS} for our experiment in Section 6.1. Skipping the conversion to the standard form generates same the objective value, but reduces the average solution time from 85 seconds to about 10 seconds. We emphasize that this alternate formulation might not be valid when some of the components of $\boldsymbol{\xi}$ are restricted to be integers. Therefore, it is not straightforward to apply it to the project management and the newsvendor experiments, both of which contain discrete uncertain parameters.

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] J.W. Blankenship and J. E. Falk. Infinitely constrained optimization problems. *Journal of Optimization Theory and Applications*, 19(2):261–281, 1976.

[9] I. M. Bomze, J. Cheng, P. J. C. Dickinson, and L. Abdel. A fresh CP look at mixed-binary QPs: new formulations and relaxations. *Mathematical Programming*, 166(1-2):159–184, 2017.

[10] 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.

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

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

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

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

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

[16] 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.

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

[18] 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.

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

[20] S.Y. Gao, D. Simchi-Levi, C.-P. Teo, and Z. Yan. Disruption Risk Mitigation in Supply Chains-The Risk Exposure Index Revisited. *To appear in Operations Research*, 2018.

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

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

[23] 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.

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

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

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

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

[28] 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.

[29] G. A. Hanasusanto and D. Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research*, 2018.

[30] G. A. Hay. Production, price, and inventory theory. *The American Economic Review*, 60(4):531–545, 1970.

[31] 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.

[32] 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.

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

[34] 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.

[35] 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.

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

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

[38] K. Natarajan and C.-P. Teo. On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming*, 161(1-2):487–518, 2017.

[39] 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.

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

[41] 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.

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

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

[44] 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.

[45] 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.

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

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

[48] W. Wiesemann, D. Kuhn, and B. Rustem. Robust resource allocations in temporal networks. *Mathematical programming*, 135(1-2):437–471, 2012.

[49] Y. Xia, M.-H. Yang, B. Golany, S.M. Gilbert, and G. Yu. Real-time disruption management in a two-stage production and inventory system. *IIE transactions*, 36(2):111–125, 2004.

[50] G. Xu and S. Burer. A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides. *Computational Optimization and Applications*, 70(1):33–59, 2018.

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