# On the Semidefinite Representability of Continuous Quadratic Submodular Minimization With Applications to Pricing and Moment Problems

Samuel Burer* Karthik Natarajan† Rick Willemsen‡

March 2026; Revised September 2026

**Abstract**

We study continuous quadratic submodular minimization with bounds and provide a polynomially sized semidefinite relaxation, which is provably tight for dimension $n \leq 3$. Via an explicit counterexample for $n = 4$, we show that the relaxation is not tight in general, although it remains empirically tight on randomly generated instances. We apply the relaxation to multi-product pricing and two moment problems arising in distributionally robust optimization and the computation of covariance bounds. Accordingly, this research advances the ongoing study of continuous submodular minimization and opens new application areas therein.

## 1 Introduction

Submodular functions have long played an important role in discrete optimization [Lovász, 1983]. In recent years, their relevance for continuous optimization has also grown, for example, in the fields of machine learning and artificial intelligence [Bach, 2019, Bilmes, 2022, Bian et al., 2017, Bunton and Tabuada, 2022, Zhang et al., 2019, Staib and Jegelka, 2017]. A continuous function $f$ defined on $\mathbb{R}^n$ is said to be *submodular* when

$$f(x) + f(y) \geq f(x \vee y) + f(x \wedge y),$$

*Department of Management Sciences, University of Iowa, Iowa City, IA 52242-1994, USA. Email: samuel-burer@uiowa.edu

†Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372. Email: karthik_natarajan@sutd.edu.sg

‡Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372. Email: rick_willemsen@sutd.edu.sg

for all pairs of vectors $x, y \in \mathbb{R}^n$, where the $\vee$ and $\wedge$ operators calculate component-wise maximums and minimums, respectively. If $f$ is twice differentiable, submodularity is equivalent to all mixed second partial derivatives being nonpositive. When the pure second partials are also nonpositive, $f$ is said to be *DR-submodular*, where *DR* stands for *diminishing returns*.

Generally speaking, studies on continuous submodular optimization fall into two types: those addressing the maximization of submodular functions, and those addressing minimization. Relatively more attention has been paid recently to maximization, as noted by Yu and Küçükyavuz [2025], and we refer the reader to this paper for an excellent, recent summary of the literature on optimization with submodular functions. In our paper, we focus on *continuous submodular minimization (CSM)*.

As discussed by Bach [2019] and Axelrod et al. [2020], many of the results for submodular minimization in discrete settings—in particular, minimization over the lattice $\{0,1\}^n$—have natural extensions to CSM over the box $[0,1]^n$. Indeed, the property of submodularity is generally considered to make minimization easier due to a strong connection with convexity. Even still, a number of fundamental questions remain.

In this paper, we specifically study the minimization of a submodular quadratic function over $[0,1]^n$, a problem which we call *quadratic submodular minimization over the box (QSMB)*:

$$\min \left\{ x^T Q x + c^T x : x \in [0,1]^n \right\}, \tag{1}$$

where $Q$ is a symmetric matrix and $c$ is a column vector. We will often refer to (1) by its pair $(Q, c)$. Here, submodularity is equivalent to the off-diagonal entries of $Q$ being nonpositive, and we say that such a $Q$ is a *submodular matrix* and that the pair $(Q, c)$ is *submodular*. Note that, if the constraints were $x \in [l, u]$, then we could convert to the form (1) by a simple affine transformation of $[l, u]$ to $[0,1]^n$ with a corresponding update to the objective function. This transformation preserves certain properties of the Hessian matrix, e.g., the signs of its eigenvalues and its submodularity. Another transformation which preserves submodularity is the flip operation $x \mapsto -x$.

When $Q$ is completely arbitrary, problem (1) is NP-hard [Horst et al., 2000], but it is polynomial-time solvable in certain cases—for example, when $Q$ is positive semidefinite and hence (1) is convex. Submodularity also impacts the computational complexity. In particular, when $(Q, c)$ is submodular and also the diagonal entries of $Q$ are nonpositive, which is DR-submodularity as defined above, then the objective function is concave along each dimension and hence (1) reduces to a special instance of submodular minimization over the lattice $\{0,1\}^n$, which is well-known to be solvable in polynomial time; see Proposition

1.1 in Staib and Jegelka [2017]. Problem (1) with DR-submodularity is in fact solvable using a polynomial sized linear program; see Proposition 10 in Padberg [1989]. Another important case occurs when $(Q, c)$ is submodular and $c \leq 0$. In this case, Kim and Kojima [2003] build on the results of Zhang [2000] to show that (1) can be solved in polynomial time via its basic SDP relaxation; see Section 2 below. In addition, Axelrod et al. [2020] build on results in Bach [2019] to present a probabilistic algorithm for solving CSM over the box that includes QSMB. Their algorithm is linear in $n$ and polynomial in the ratio $L/\varepsilon$, where $L$ is the submodular function's Lipschitz parameter and $\varepsilon$ is the additive error in the final optimal value. In particular, the polynomial dependence on $L/\varepsilon$ makes their algorithm pseudo-polynomial.

Table 1 provides a classification of the complexity of the minimization of quadratic functions versus the type of domain. Note that *QUBO* stands for *quadratic unconstrained binary optimization*, which is standard in the literature. In particular, this paper makes progress on quadratic submodular minimization over the continuous domain using an SDP-based approach.

| | $f$ convex | $f$ submodular |
|---|---|---|
| $\{0,1\}^n$ | NP-hard (QUBO)<br>Garey and Johnson [1979][1] | P (LP)<br>Padberg [1989] |
| $[0,1]^n$ | P (convex QP)<br>Kozlov et al. [1980] | P (LP, when DR-submodular)<br>Padberg [1989] |
| | | P (SDP, when $c \leq 0$)<br>Kim and Kojima [2003] |
| | | P (SDP, when $n \leq 3$)[2]<br>This paper |

Table 1: Complexity of minimizing $f(x) = x^T Q x + c^T x$

## 1.1 Motivating examples

We consider a few examples to motivate the applicability and study of QSMB.

**Example 1** (Multi-product pricing)**.** *Consider $n$ products with the price decision vector $p \in \mathbb{R}^n$, marginal cost vector $m \in \mathbb{R}^n$, and linear demand model given by $d(p) = a - Bp$*

[1]Over $\{0,1\}^n$, $x^T Q x + c^T x = x^T (Q - \lambda_{\min} I) x + (c + \lambda_{\min} e)^T x$ from the identity $x_i^2 = x_i$ where $\lambda_{\min}$ is the minimum eigenvalue of $Q$. Since $Q - \lambda_{\min} I$ is positive semidefinite, the hardness of minimizing a convex quadratic function over $\{0,1\}^n$ follows directly from the hardness of QUBO.

[2]As we will show, tightness of the semidefinite relaxation is specific to $n \leq 3$. In particular, Example 4 exhibits a submodular instance with $n = 4$ for which the relaxation has a strictly positive gap. The complexity of minimizing a submodular quadratic over $[0,1]^n$ for $n \geq 4$ remains open.

*where $a \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n\times n}$. The price vector $p$ is chosen in the box $[l,u]$ where we assume $d(p) \geq 0$ for all $p \in [l,u]$. Then the multi-product pricing problem is formulated as the quadratic maximization problem*

$$\max\left\{(p-m)^T(a-Bp) : p \in [l,u]\right\}.$$

*In this model, depending on the sign of the off-diagonal entries in the matrix $B$, the products might be substitutes or complements. We particularly focus on the pricing of substitute products such as grocery brands, similar electronic products, competing package sizes, or differentiated service plans. Demand models of substitutes require the off-diagonal entries of $B$ to be nonpositive, so that the demand of product $i$ increases as the price of product $j$ increases for $j \neq i$. The pricing problem with linear demand model and substitutable products is then an instance of QSMB, where we symmetrize the quadratic term by replacing $B$ with $(B+B^T)/2$ and minimize the negative of the profit. In the literature, tractability is ensured by assuming additionally that $(B+B^T)/2$ is positive semidefinite, so that the pricing problem becomes a convex quadratic program. See Theorem 8.23 and related demand models in Chapter 8 of Gallego and Topaloglu [2019] for multi-product pricing with substitutes. A second route to tractability is to discretize. Ito and Fujimaki [2016] restrict each price to a finite set of candidate values and encode that choice with binary variables; substitutability then makes the gross profit supermodular, and the pricing problem becomes a submodular binary quadratic minimization. We employ neither convexity nor discretization, in which case solving the pricing problem to optimality requires the solution of QSMB.*

**Example 2** (Ambiguity sets in distributionally robust optimization)**.** *A commonly used moment ambiguity set in distributionally robust optimization is the following Chebyshev ambiguity set (see Bertsimas and Popescu [2005], Kuhn et al. [2024]):*

$$\left\{\mathbb{P} \in \mathcal{P}(\Xi) \;:\; \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu,\; \mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] = \Sigma\right\},$$

*where the support of the random vector $\tilde{\xi}$ is contained in $\Xi$, the mean of the random vector is fixed to $\mu$, and the second moment matrix is fixed to $\Sigma$. The set $\mathcal{P}(\Xi)$ denotes the set of all probability distributions with support contained in $\Xi$. While the Chebyshev ambiguity set is computationally tractable for specific supports—such as when $\Xi$ equals $\mathbb{R}^n$ or is an ellipsoid—checking whether this ambiguity set is nonempty for $\Xi = [0,1]^n$ is known to be NP-hard in general.*[3] *A tractable relaxation of this ambiguity set proposed in Delage and Ye*

[3]By setting the diagonal entries of $\Sigma$ equal to those of $\mu$, the random variables in the ambiguity set are forced to be Bernoulli when $\Xi = [0,1]^n$. Checking if the ambiguity set is nonempty is then equivalent to testing membership in the correlation polytope; see Pitowsky [1991].

*[2010] is*

$$\left\{\mathbb{P} \in \mathcal{P}([0,1]^n) \ : \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu,\ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] \preceq \Sigma\right\},$$

*where the exact second moment matrix condition is relaxed to an upper bound in the positive semidefinite order. In this paper, we will propose an alternative relaxation to the Chebyshev ambiguity set given by*

$$\left\{\mathbb{P} \in \mathcal{P}([0,1]^n) \ : \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu,\ \mathbb{E}_{\mathbb{P}}[\mathrm{diag}(\tilde{\xi}\tilde{\xi}^T)] = \mathrm{diag}(\Sigma),\ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] \geq \Sigma\right\},$$

*where the mean vector and the diagonal entries of the second moment matrix are exactly specified and the off-diagonal entries of the second moment matrix are constrained by entrywise lower bounds. Using duality between the theory of moments and nonnegative polynomials, we will show that QSMB is key to providing a characterization of this relaxed ambiguity set. We discuss the flexibility of modeling uncertainty with this ambiguity set and its application to distributionally robust optimization in Section 3.*

**Example 3** (Covariance bounds)**.** *The Cauchy–Schwarz inequality provides a sharp upper bound on the covariance of a pair of random variables using only the variances of the random variables:*

$$Cov[\tilde{\xi}_1, \tilde{\xi}_2] \leq \sqrt{Var[\tilde{\xi}_1]\, Var[\tilde{\xi}_2]}.$$

*This bound can be sharpened when each random variable is known to lie in the interval* $[0,1]$ *and the means are fixed (see Hössjer and Sjölander [2022]):*

$$Cov[\tilde{\xi}_1, \tilde{\xi}_2] \leq \min\left(\mathbb{E}[\tilde{\xi}_1](1-\mathbb{E}[\tilde{\xi}_2]), \mathbb{E}[\tilde{\xi}_2](1-\mathbb{E}[\tilde{\xi}_1]), \sqrt{Var[\tilde{\xi}_1]\, Var[\tilde{\xi}_2]}\right).$$

*Again, by building on duality theory for moment problems and the connection of the dual formulation to QSMB, we will develop new covariance bounds for an arbitrary number of random variables, thus extending the above inequality.*

## 1.2 Contributions and structure of the paper

The main contributions and the structure of the paper are as follows:

(a) In light of the existing literature, our first goal is to study a polynomial-time semidefinite programming relaxation of (1), which is inspired by Kim and Kojima [2003]. This relaxation is *tight* (i.e., admits no relaxation gap) for $n \leq 3$, a result which we prove in Section 2 making use of technical results established in Section 6. We further show that this result is sharp by exhibiting an explicit submodular instance with $n = 4$ that

has a strictly positive relaxation gap. Nonetheless, the relaxation is empirically tight on a randomly generated panel of instances of varying sizes reported in Section 5.

(b) We then apply the semidefinite programming reformulation of QSMB to approximate—and in low dimensions, solve exactly—two optimization models involving moments of probability measures.

- In the first model discussed in Section 3, we consider a class of distributionally robust optimization (DRO) problems over the new ambiguity set. Our SDP for QSMB enables us to approximate this class of DRO problems efficiently for general $n$ providing an upper bound on the worst-case expected cost with an exact reformulation for $n = 3$.

- In the second model discussed in Section 4, we compute an upper bound on the non-negative weighted sum of covariances of random variables given only the means and the variances of the random variables, again supported in $[0, 1]^n$. This extends the Cauchy–Schwarz inequality to bounded random variables in higher dimensions. We note that the SDP relaxation provides the sharpest upper bound possible in low dimensions ($n \leq 3$). In addition, we provide a new proof for the $n = 2$ case and sharp closed form solutions in special cases for arbitrary number of random variables.

(c) In the numerical results in Section 5, we present our empirical evidence for practical tightness of the relaxation in higher dimensions. We also provide examples in dimension $n = 3$ for multi-product pricing and a distributionally robust optimization problem arising in stratified sampling allocation to illustrate the applicability of QSMB and the value of the SDP formulation. We then compute moment bounds on the energy function defined by Laplacian matrices of graphs. Our results illustrate that the bounds from our approach can be stronger than bounds from existing SDP formulations such as that developed in Delage and Ye [2010] for some of these problems.

We conclude in Section 7. Overall, this paper extends polynomial-time results for convex minimization to quadratic submodular minimization over the box (QSMB) via a provably exact SDP relaxation for $n \leq 3$, shows by an explicit counterexample that exactness fails for $n \geq 4$, and examines various applications of these results. As such, this paper lays the groundwork for further study of continuous submodular minimization and indeed can have important ramifications in areas such as robust and distributionally robust optimization.

### 1.3 Notations

We use $\mathbb{R}^n$ to denote the space of real $n$-dimensional vectors, $\mathbb{S}^n$ to denote the space of symmetric real $n \times n$ matrices, and $\mathbb{R}^{m\times n}$ to denote the space of real $m \times n$ matrices. The vector $x$ is a column vector by default and $x^T$ is the transposed row vector. We use $e$ to denote the vector of ones, whose dimension is determined by the context in which it appears. We write $A \succeq 0$ ($A \succ 0$) to denote that matrix $A$ is positive semidefinite (positive definite) and $A \succeq B$ to denote $A - B \succeq 0$. The vector formed with the diagonal entries of the matrix $A$ is given by $\mathrm{diag}(A)$, and the diagonal matrix formed with the entries of the vector $a$ is given by $\mathrm{Diag}(a)$. Given $A, B \in \mathbb{R}^{m\times n}$, we let $A \bullet B = \mathrm{trace}(A^T B)$. For vectors $x, y \in \mathbb{R}^n$, we use $x \circ y$ to denote their componentwise (Hadamard) product. We use $\sqrt{x}$ to denote a vector with the square root of the entries of the vector $x$. A random vector is denoted by $\tilde{\xi}$, and we use $\mathbb{E}_{\mathbb{P}}[\cdot]$ to denote the expectation with respect to the distribution $\mathbb{P}$, $\mathrm{Var}[\tilde{\xi}_i]$ to denote the variance of $\tilde{\xi}_i$, and $\mathrm{Cov}[\tilde{\xi}_i, \tilde{\xi}_j]$ to denote the covariance of $\tilde{\xi}_i$ and $\tilde{\xi}_j$.

## 2 A Tight Semidefinite Relaxation for $n \leq 3$

In this section, we establish our main theoretical result, namely that there exists a tight semidefinite relaxation of (1) for dimension $n \leq 3$. We first review the relevant literature that motivates our approach.

### 2.1 Existing results

Kim and Kojima [2003] proved the following result for QSMB for general $n$ with additional restrictions on $c$:

**Proposition 1** (Theorem 3.1 in Kim and Kojima [2003])**.** *Let $n$ be arbitrary, and suppose $(Q, c)$ is submodular with $c \leq 0$. Then the optimal value of (1) equals the optimal value of the SDP relaxation*

$$\min \left\{ Q \bullet X + c^T x : \mathrm{diag}(X) \leq e, \ \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0 \right\}.$$

In fact, the authors show that, if $(x^*, X^*)$ is an optimal solution of the SDP, then the vector $\sqrt{\mathrm{diag}(X^*)}$ is an optimal solution of (1). Furthermore, the positive semidefiniteness condition on the $(n+1) \times (n+1)$ matrix can be relaxed to positive semidefiniteness of the $n(n+1)/2$ principal submatrices of size $2 \times 2$ without changing the conclusion of the theorem;

see Theorem 3.5 in Kim and Kojima [2003]. It is also straightforward to construct examples with $n = 2$ and some $c_j > 0$ such that this SDP relaxation admits a gap.

When $c$ is arbitrary, we can attempt to close the gap using a tighter SDP relaxation. A typical approach, which we adapt in this paper, is to include valid inequalities derived from the feasibility condition $x \in [0,1]^n$. Indeed, a common class of valid inequalities are the so-called *McCormick inequalities* or *RLT (reformulation linearization technique) constraints*, which are derived for each pair $1 \le i \le j \le n$ as follows [McCormick, 1976]:

$$\begin{aligned} x_i x_j \ge 0 \quad &\Longrightarrow \quad X_{ij} \ge 0 \\ x_i(1-x_j) \ge 0 \quad &\Longrightarrow \quad X_{ij} \le x_i \\ (1-x_i)x_j \ge 0 \quad &\Longrightarrow \quad X_{ij} \le x_j \\ (1-x_i)(1-x_j) \ge 0 \quad &\Longrightarrow \quad X_{ij} \ge x_i + x_j - 1. \end{aligned}$$

In matrix form, these valid constraints are expressed as

$$X \ge 0, \qquad X \ge xe^T + ex^T - ee^T, \qquad X \le xe^T.$$

Note that two of these matrix inequalities provide lower bounds on the components of $X$, whereas the third provides upper bounds. In addition, the right-hand side of $X \le xe^T$ is non-symmetric, and so this single inequality captures both $X_{ij} \le x_i$ and $X_{ij} \le x_j$ for all $i \le j$. Also note that the diagonal inequalities of $X \le xe^T$ ensure $\mathrm{diag}(X) \le x$, a strengthening of the constraint $\mathrm{diag}(X) \le e$ in Proposition 1.

The SDP relaxation, which incorporates the RLT constraints, is known to be tight for $n = 2$ for all data inputs $(Q, c)$:

**Proposition 2** (Theorem 2 in Anstreicher and Burer [2010])**.** *For $n = 2$ and arbitrary $(Q, c)$, the SDP relaxation*

$$\min \left\{ Q \bullet X + c^T x : \begin{array}{l} X \ge 0, \\ X \ge xe^T + ex^T - ee^T, \\ X \le xe^T, \end{array} \quad \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0 \right\}$$

*is a tight relaxation of (1).*

However, the authors also show that this relaxation admits a gap for $n \ge 3$ and general $(Q, c)$.

## 2.2 Additional background

Focusing our attention on the submodular case, we consider the following relaxation, which adds only the RLT upper bounds to the relaxation in Proposition 1:

$$\min \left\{ Q \bullet X + c^T x : X \leq x e^T, \ \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0 \right\}. \tag{2}$$

The intuition for only including the upper bounds comes from the submodularity of $Q$, which implies that the objective function is nonincreasing in the off-diagonal entries of $X$ and hence the RLT lower bounds are less likely to be active at optimality.

Our main result in Theorem 1 states that this relaxation is tight when $n \leq 3$, for every submodular $(Q, c)$ and irrespective of the signs of the entries in $c$. Example 4 below shows that this cannot be extended; tightness fails already at $n = 4$.

Two features of the relaxation (2) are worth emphasizing. First, compared to the full SDP-RLT relaxation of Proposition 2, it retains roughly half the number of linear inequalities. Second—and as a consequence—tightness of (2) as expressed in Theorem 1 is a stronger statement than tightness of the full SDP-RLT relaxation. Because the latter is at least as tight, its exactness follows from the exactness of (2).

We continue by introducing some notation and important concepts. Define

$$Y(x, X) := \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix}$$

to be the symmetric matrix appearing in the SDP relaxation (2). In particular, $Y(x, X) \succeq 0$ is the positive semidefinite condition. Also define

$$\begin{aligned} r(x, X) &:= \operatorname{rank}(Y(x, X)), \\ a(x, X) &:= \left| \{(i, j) : X_{ij} = x_i\} \right|. \end{aligned}$$

In words, $a(x, X)$ is the number of active inequalities in the linear inequality condition $X \leq x e^T$ of (2). In particular, $a(x, X) \leq n^2$ since there are $n^2$ inequalities in $X \leq x e^T$.

The following lemma relates $r(x, X)$ and $a(x, X)$ at extreme points of (2). In particular, the number of active inequalities grows quadratically with the rank.

**Lemma 1.** *Let $(x, X)$ be an extreme point of the feasible set of (2). Define $r := r(x, X)$ and $a := a(x, X)$. It holds that*

$$r(r + 1) \leq 2(a + 1).$$

*Thus, irrespective of $n$, the following low-rank combinations of $r$ and $a$ are possible: $r = 1$ with $a \geq 0$; $r = 2$ with $a \geq 2$; $r = 3$ with $a \geq 5$; and $r = 4$ with $a \geq 9$.*

*Proof.* We recall a celebrated result due to Pataki [1998]:

> Consider an SDP feasible set in block standard form: $F := \{X^j \succeq 0,\ j = 1, \ldots, p : \sum_{j=1}^{p} A_i^j \bullet X^j = b_i,\ i = 1, \ldots, m\}$. Let $(X^1, \ldots, X^p)$ be an extreme point of $F$, and define $r_j := \operatorname{rank}(X^j)$. Then $\sum_{j=1}^{p} r_j(r_j + 1) \leq 2m$.

The feasible set of (2) can be expressed in block standard form with $p = n^2 + 1$ blocks: one PSD block and $n^2$ slack-variables. Moreover, the number of equality constraints is $m = n^2+1$ with $n^2$ defining the slack variables and one defining the top-left entry of $Y(x, X)$ to be 1. The formula now follows by applying Pataki's result and noting that, for a slack-variable, $r_j(r_j + 1) = 0$ if and only if the corresponding slack variable equals 0 and $r_j(r_j + 1) = 2$ if and only if the slack variable is positive. In particular, if $s$ is the number of positive slacks such that $s + a = n^2$, then we have

$$r(r+1) + 2s \leq 2(n^2+1) \quad \Leftrightarrow \quad r(r+1) + 2(n^2 - a) \leq 2(n^2+1) \quad \Leftrightarrow \quad r(r+1) \leq 2(a+1).$$

The final statement of the lemma follows by considering different cases for the rank of $Y(x, X)$. □

Next, we describe *completely positive programming*, which provides a tool for proving tightness of SDP relaxations. Define the cone

$$\mathcal{J} := \left\{ y := \begin{pmatrix} x_0 \\ x \end{pmatrix} \in \mathbb{R}^{n+1} : 0 \leq x \leq x_0 e \right\},$$

which is the homogenization of the unit box $[0, 1]^n$, where $x_0 \in \mathbb{R}_+$ and $x \in \mathbb{R}^n$, and the *completely positive cone with respect to* $\mathcal{J}$:

$$\mathcal{CP}(\mathcal{J}) := \operatorname{conv}\left\{ yy^T : y \in \mathcal{J} \right\}.$$

It is known that the optimal value of (1) equals that of the completely positive program $\min\{Q \bullet X + c^T x : Y(x, X) \in \mathcal{CP}(\mathcal{J})\}$; see Burer [2009, 2012].

The SDP relaxation (2) can be viewed as relaxing complete positivity to $Y(x, X) \in \mathcal{L}$, where

$$\mathcal{L} := \left\{ Y := \begin{pmatrix} X_{00} & x^T \\ x & X \end{pmatrix} \succeq 0 : X \leq xe^T \right\}$$

is the homogenization of the feasible set of (2). The following tightness criterion is used throughout the paper.

**Lemma 2.** *$\mathcal{CP}(\mathcal{J}) \subseteq \mathcal{L}$. Furthermore, if there exists an optimal solution $(x, X)$ of (2) such that $Y(x, X) \in \mathcal{CP}(\mathcal{J})$, then the SDP relaxation is tight.*

Next, we classify the $n^2$ inequalities $X_{ij} \leq x_i$ into three types based on their position in the non-symmetric matrix $X - xe^T$: *s-lower inequalities* ($i > j$), *s-upper inequalities* ($i < j$), and *diagonal inequalities* ($i = j$). The terms *s-lower* and *s-upper* are shorthand for "strictly lower" and "strictly upper."

Finally, given a feasible solution $(x, X)$ of (2), we consider permuting the rows and columns of $Y(x, X)$ so that $x_1 \leq \cdots \leq x_n$, which we call *sorting by $x$*. Sorting by $x$ preserves feasibility in $\mathcal{L}$, the rank $r(x, X)$, the number of active inequalities $a(x, X)$, and complete positivity with respect to $\mathcal{J}$. When $(x, X)$ is sorted by $x$ and $i < j$, feasibility with respect to $\mathcal{L}$ gives $X_{ij} \leq x_i \leq x_j$. Hence an active s-upper inequality ($X_{ij} = x_i$) need not force the s-lower inequality to be active, but an active s-lower inequality ($X_{ij} = x_j$) forces $x_i = x_j$ and hence the s-upper inequality to be active as well. A precise consequence of these observations is recorded in Lemma 7 in Section 6.

## 2.3 Our result

Theorem 1 is stated and proved below by combining several propositions, which are proved in Section 6.

The first proposition addresses optimal solutions of rank 2: either the SDP relaxation is tight, or a proper face of the optimal face exists whose extreme points all have rank 3 or higher.

**Proposition 3.** *If there exists an optimal solution $(x, X)$ of (2) with $r(x, X) = 2$, then: (i) the SDP relaxation is tight; or (ii) there exists a proper face of the optimal face in which all extreme points have rank 3 or higher.*

Our next proposition establishes tightness related to active diagonal and s-upper inequalities.

**Proposition 4.** *Let $(x, X)$ be an optimal solution of (2). After sorting by $x$, suppose one of the following holds:*

- $\operatorname{diag}(X) = x$;
- *all s-upper inequalities are active, and all diagonal inequalities are active except for a single $X_{jj} < x_j$.*

*Then the SDP relaxation is tight.*

The final proposition uses an inductive argument to handle active s-lower inequalities at optimality.

**Proposition 5.** *Suppose the SDP relaxation (2) is tight when applied to all submodular instances of (1) in $n-1$ variables. For dimension $n$, let $(x, X)$ be an optimal solution of (2) for a submodular instance $(Q, c)$. After sorting by $x$, suppose there exists at least one active s-lower inequality. Then the SDP relaxation is tight.*

Using the three propositions above, we can now state and prove the main theorem.

**Theorem 1.** *For $n \leq 3$, the SDP relaxation (2) is tight for all submodular data $(Q, c)$.*

*Proof.* Let $(x, X)$ be an optimal extreme point of (2), and define $Y := Y(x, X)$, $r := r(x, X)$, and $a := a(x, X)$. We prove $n = 1$, $n = 2$, and $n = 3$ in turn.

When $n = 1$, the feasibility condition $X \leq xe^T$ amounts to a single inequality. Hence, $a \leq 1$, and Lemma 1 implies $r = 1$. It follows that $Y \in \mathcal{CP}(\mathcal{J})$. Then the relaxation is tight by Lemma 2.

When $n = 2$, there are $n^2 = 4$ inequalities in $X \leq xe^T$. By Lemma 1, $r(r+1) \leq 2(a+1) \leq 10$, so $r \leq 2$. If $r = 1$, the relaxation is tight as above. If $r = 2$, Proposition 3 implies either (i) the SDP relaxation is tight or (ii) there exists a proper face of the optimal face whose extreme points all have rank at least 3. Since any extreme point of a face is an extreme point of the feasible set, the bound $r \leq 2$ applies throughout, ruling out case (ii). Hence the SDP relaxation is tight.

When $n = 3$, if $r = 1$, then we are done. If $r = 4$, then $a \geq 9$ by Lemma 1, so every inequality is active. In particular, $\operatorname{diag}(X) = x$, and Proposition 4 implies the SDP relaxation is tight. If $r = 2$, then by Proposition 3 either the SDP relaxation is tight, or there exists a proper face of the optimal face whose extreme points all have rank 3 or higher; any such extreme point is handled by the $r = 3$ case next.

This leaves $r = 3$, for which $a \geq 5$ by Lemma 1. Assume without loss of generality that $(x, X)$ is sorted by $x$. Let $l$, $d$, and $u$ be the number of active s-lower, diagonal, and s-upper inequalities, respectively. Then $l + d + u = a \geq 5$ with $l, d, u \leq 3$. If $l \geq 1$, Proposition 5 applies directly. If $l = 0$, then $d + u \geq 5$ with $d, u \leq 3$, so $d \geq 2$. Either $d = 3$ and $\operatorname{diag}(X) = x$, or $d = 2$ and $u = 3$, meaning all s-upper inequalities and all but one diagonal inequality are active. In both subcases, Proposition 4 applies. □

Figure 1 illustrates the geometry of Theorem 1 for $n = 3$. Fixing the values of $x$ and $\operatorname{diag}(X)$ and treating the off-diagonals $(X_{12}, X_{13}, X_{23})$ as free, we project three feasible sets

onto the two-dimensional region defined by the auxiliary variables $(X_{12} + X_{13},\ X_{13} + X_{23})$: (i) the exact hull $\mathcal{CP}(\mathcal{J})$, (ii) the full SDP-RLT relaxation, and (iii) our relaxation (2). A submodular objective maximizes a nonnegative combination of the off-diagonals, so its optimum lies on the "northeast" boundary, precisely where all three bodies coincide, making the relaxation tight even after the lower bounds on $X_{ij}$ are discarded. A non-submodular objective would instead point toward a boundary where the bodies separate and hence a gap appears.

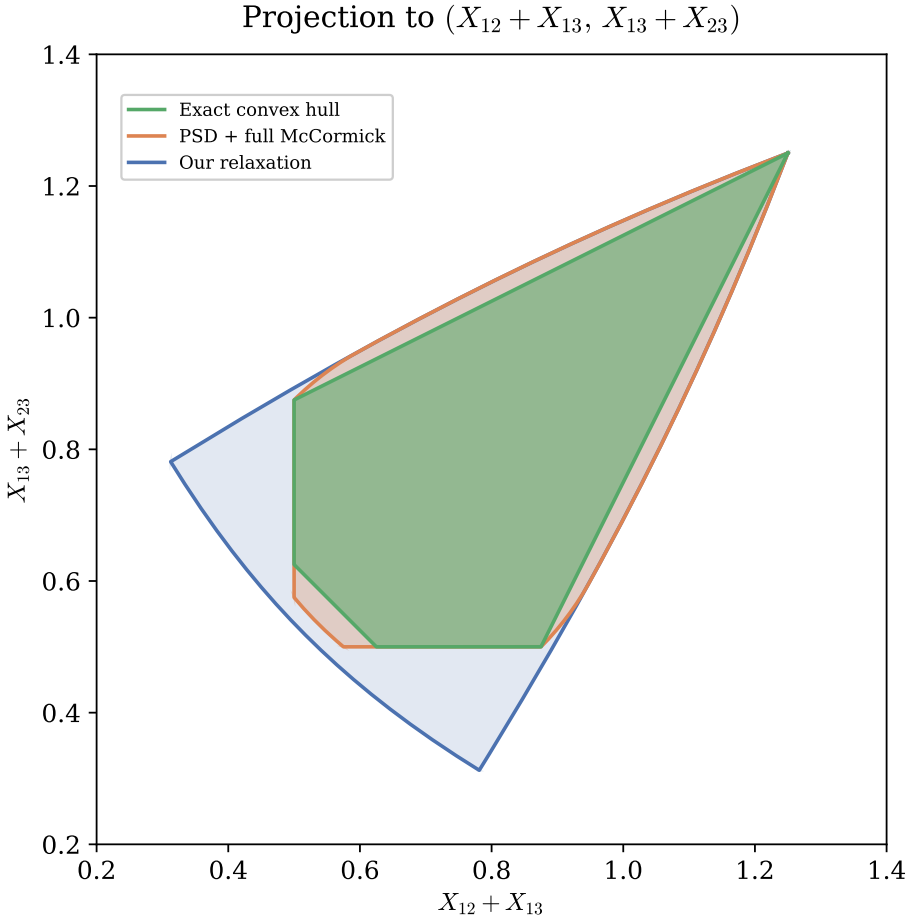


Figure 1: Projection to $(X_{12} + X_{13},\ X_{13} + X_{23})$ of three nested relaxations at fixed values of $x = \mathrm{diag}(X) = (5/8)e$: the exact hull $\mathcal{CP}(\mathcal{J})$, the full SDP-RLT relaxation, and our relaxation (2). The three bodies coincide on the "northeast" boundary optimized by submodular objectives and do not coincide elsewhere, where a non-submodular objective would incur a gap.

**Remark 1** (Extending beyond $n = 3$)**.** *To extend Theorem 1 to $n = 4$, in principle the same proof strategy could be applied. For $n = 4$, there are $n^2 = 16$ inequalities in $X \leq xe^T$, decomposed into 6 s-lower, 4 diagonal, and 6 s-upper. Lemma 1 constrains the rank to $r \leq 5$.*

*Among the possible ranks, the cases $r = 1$, $r = 2$, $r = 4$, and $r = 5$ are resolved by the existing propositions together with counting arguments. Specifically, Proposition 3 handles $r = 2$ by reducing to higher rank; for $r = 4$ and $r = 5$, the minimum number of active inequalities ($a \geq 9$ and $a \geq 14$, respectively) is large enough that either an s-lower inequality must be active—invoking Proposition 5—or the diagonal and s-upper conditions of Proposition 4 are met.*

*The bottleneck is $r = 3$, which requires only $a \geq 5$. When no s-lower inequality is active ($l = 0$), one needs $d + u \geq 5$ with $d \leq 4$ and $u \leq 6$. For $n = 3$, the constraints $d, u \leq 3$ left only two subcases—both covered by Proposition 4. For $n = 4$, many additional configurations arise (e.g., $d = 2$, $u = 3$) in which neither all diagonal nor all s-upper inequalities are active, and these are not covered by the current propositions. Resolving these rank-3 configurations with partial activity is the main obstacle to extending Theorem 1 to $n = 4$ and beyond. Example 4 below shows that this obstacle does in fact occur. We find that the relaxation (2) is* not *tight for $n = 4$, and the counterexample occupies precisely the rank-3 configuration just described.*

**Example 4** (The relaxation is not tight for $n \geq 4$)**.** *Let $n = 4$ and consider the data*

$$Q = \begin{pmatrix} 8 & -14 & 0 & 0 \\ -14 & 25 & -25 & 0 \\ 0 & -25 & 25 & -14 \\ 0 & 0 & -14 & 8 \end{pmatrix}, \qquad c = \begin{pmatrix} 12 \\ 29 \\ 0 \\ 0 \end{pmatrix}. \tag{3}$$

*Every off-diagonal entry of $Q$ is nonpositive, so $(Q, c)$ is submodular, and $Q$ is indefinite. The box-constrained problem (1) has optimal value $0$, attained uniquely at $x = 0$, while the relaxation (2) has optimal value approximately $-0.1221$. So (2) has a strictly positive gap, and Theorem 1 does not extend to $n = 4$. Padding $(Q, c)$ with additional variables having zero objective coefficients preserves both optimal values, so the relaxation fails to be tight for every $n \geq 4$.*

*The optimal solution of (2) for this instance has rank $3$, with exactly two diagonal and five s-upper inequalities active and no s-lower inequality active. This is precisely the configuration that Remark 1 identifies as the obstacle to extending the proof beyond $n = 3$. Section 6.4 verifies these claims in exact rational arithmetic and examines whether standard cutting planes close the gap.*

# 3 Distributionally Robust Optimization

Consider the distributionally robust optimization problem

$$\inf_{x\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[f(x,\tilde{\xi}) := \max_{k=1,\ldots,K}\left(\tilde{\xi}^T A_k(x)\tilde{\xi} + b_k^T(x)\tilde{\xi} + c_k(x)\right)\right], \tag{4}$$

where the decision vector $x$ is chosen from a set $\mathcal{X} \subseteq \mathbb{R}^m$ and the random vector $\tilde{\xi}$ is $n$-dimensional. The probability distribution of $\tilde{\xi}$, denoted by $\mathbb{P}$, is ambiguous and lies in a set of probability distributions denoted by $\mathcal{P}$, commonly referred to as the *ambiguity set*. The matrix $A_k(x) \in \mathbb{S}^n$, the vector $b_k(x) \in \mathbb{R}^n$ and the scalar $c_k(x) \in \mathbb{R}$ are assumed to depend affinely on $x$ for each $k = 1, \ldots, K$. We hence assume that the cost function $f(x, \xi)$ is piecewise affine and thus convex in $x$ for a fixed $\xi$ and is piecewise quadratic but not necessarily convex in $\xi$ for a fixed $x$.

Solving (4) corresponds to finding the optimal decision $x$ that minimizes the worst-case expected cost computed over all distributions in the ambiguity set. In this section, our goal is to propose a new moment-based ambiguity set, which incorporates submodular minimization and guarantees a computationally tractable—and empirically tight—bound via the SDP relaxation (2).

## 3.1 A new moment ambiguity set

Given fixed $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{S}^n$, define

$$\mathcal{P} := \left\{\mathbb{P} \in \mathcal{P}([0,1]^n) \;:\; \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu,\ \mathbb{E}_{\mathbb{P}}[\mathrm{diag}(\tilde{\xi}\tilde{\xi}^T)] = \mathrm{diag}(\Sigma),\ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] \geq \Sigma\right\}. \tag{5}$$

As discussed in the Introduction, this ambiguity set can be viewed as a relaxation of the NP-hard Chebyshev ambiguity set, which fixes the means and the diagonal entries of the second moment matrix while relaxing the off-diagonal entries with entrywise lower bounds. Note that the matrix $\Sigma$ is also not necessarily required to be positive semidefinite, which contrasts with the Delage-Ye relaxation also mentioned in the Introduction. This can be a modeling advantage. For example, when the pairwise moments are estimated using different subsets of observations or with missing entries scattered across observations, the estimated entries of $\Sigma$ might not jointly form a positive semidefinite matrix; see Higham [2002].

We would like to identify tractable necessary and sufficient conditions on $\mu$ and $\Sigma$ that guarantee nonemptiness of $\mathcal{P}$. Let $\mathcal{M}([0,1]^n)$ denote the set of all nonnegative finite Borel measures on $[0,1]^n$. As is standard, $\mathcal{P}$ is nonempty if and only if $(1, \mu, \Sigma) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n$ is

a member of the following cone:

$$\mathcal{M}_0 := \left\{ (\lambda_0, \lambda, \Lambda) \ : \ \exists\, m \in \mathcal{M}([0,1]^n) \ \text{ s.t. } \begin{array}{l} \lambda_0 = \int dm(\xi), \\ \lambda = \int \xi dm(\xi), \\ \Lambda \leq \int \xi\xi^T dm(\xi), \\ \mathrm{diag}(\Lambda) = \int \mathrm{diag}(\xi\xi^T) dm(\xi) \end{array} \right\}. \tag{6}$$

We will now argue that $\mathcal{M}_0$ is semidefinite approximable. To this end, we start by defining some relevant convex cones.

Define the *truncated moment cone of degree 2*:

$$\mathcal{M}_1 := \left\{ (\lambda_0, \lambda, \Lambda) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n \ : \ \exists\, m \in \mathcal{M}([0,1]^n) \ \text{ s.t. } \begin{array}{l} \lambda_0 = \int dm(\xi), \\ \lambda = \int \xi dm(\xi), \\ \Lambda = \int \xi\xi^T dm(\xi) \end{array} \right\}.$$

We also define the *cone of nonnegative polynomials of degree 2* on $[0,1]^n$ as

$$\mathcal{K}_1 := \left\{ (y_0, y, Y) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n \ : \ y_0 + y^T\xi + \xi^T Y \xi \geq 0 \ \ \forall\, \xi \in [0,1]^n \right\}.$$

The cones $\mathcal{M}_1$ and $\mathcal{K}_1$ are full dimensional, closed, convex, pointed and dual to each other; see Karlin and Studden [1966], Laurent [2009], Schmudgen [2017]. Also define:

$$\mathcal{M}_2 := \left\{ (0,0,\Lambda) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n \ : \ \Lambda \leq 0, \mathrm{diag}(\Lambda) = 0 \right\}.$$

and

$$\begin{aligned} \mathcal{K}_2 \ &:= \left\{ (y_0, y, Y) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n \ : \ Y_{ij} \leq 0 \ \ \forall\, i < j \right\} \\ &= \left\{ (y_0, y, Y) : Y \text{ submodular} \right\}. \end{aligned}$$

It is easy to check that the cones $\mathcal{M}_2$ and $\mathcal{K}_2$ are closed, convex and dual to each other.

It is now self-evident that $\mathcal{M}_0 = \mathcal{M}_1 + \mathcal{M}_2$. Hence, $\mathcal{M}_0$ is clearly a full-dimensional, convex, pointed cone. Closedness of $\mathcal{M}_0$ follows from Corollary 1.12 in Dickinson [2013]. The dual cone of $\mathcal{M}_0$ is given by

$$\begin{aligned} \mathcal{M}_0{}^* &= (\mathcal{M}_1 + \mathcal{M}_2)^* = \mathcal{M}_1^* \cap \mathcal{M}_2^* = \mathcal{K}_1 \cap \mathcal{K}_2 \\ &= \left\{ (y_0, y, Y) \ : \ y_0 + y^T\xi + \xi^T Y \xi \geq 0 \ \ \forall\, \xi \in [0,1]^n, \ \ Y \text{ submodular} \right\} \\ &=: \mathcal{K}_0. \end{aligned} \tag{7}$$

It follows that $\mathcal{M}_0 = \mathcal{K}_0^*$ since $\mathcal{M}_0$ is a closed convex cone. This establishes the following proposition.

**Proposition 6.** *The cones $\mathcal{M}_0$ and $\mathcal{K}_0$ are dual to each other.*

While the component cones $\mathcal{M}_1$ and $\mathcal{K}_1$ do not have tractable characterizations, we are nevertheless able to provide tractable semidefinite approximations of $\mathcal{M}_0$ and $\mathcal{K}_0$ using submodularity and the empirically tight SDP relaxation (2) of Section 2.

**Proposition 7.** *Define*

$$\mathcal{K}_{\text{SDP}} := \left\{ (y_0, y, Y) : \begin{array}{l} Y \textit{ submodular}, \\ \exists\ Z \in \mathbb{R}^{n\times n} \textit{ s.t. } Z \geq 0,\ \begin{pmatrix} y_0 & (y^T - e^T Z)/2 \\ (y - Z^T e)/2 & Y + (Z + Z^T)/2 \end{pmatrix} \succeq 0 \end{array} \right\}$$

*and*

$$\mathcal{M}_{\text{SDP}} := \left\{ (\lambda_0, \lambda, \Lambda)\ :\ \exists\ W \in \mathbb{S}^n \textit{ s.t. } \begin{array}{l} \Lambda \leq W \leq \lambda e^T, \\ \text{diag}(\Lambda) = \text{diag}(W), \\ \begin{pmatrix} \lambda_0 & \lambda^T \\ \lambda & W \end{pmatrix} \succeq 0 \end{array} \right\}.$$

*Then $\mathcal{K}_{\text{SDP}} \subseteq \mathcal{K}_0$ and $\mathcal{M}_0 \subseteq \mathcal{M}_{\text{SDP}}$ with equality in both when $n \leq 3$.*

*Proof.* Introducing an auxiliary variable $z_0$, we first observe that

$$\mathcal{K}_{\text{SDP}} = \left\{ (y_0, y, Y)\ :\ \begin{array}{l} y_0 - z_0 \geq 0, \\ Z \geq 0,\ \begin{pmatrix} z_0 & (y^T - e^T Z)/2 \\ (y - Z^T e)/2 & Y + (Z + Z^T)/2 \end{pmatrix} \succeq 0 \\ Y \text{ submodular} \end{array} \right\}.$$

Letting $(y_0, y, Y, z_0)$ be an element of the right-hand side, to show $(y_0, y, Y) \in \mathcal{K}_0$, we must demonstrate that $y_0 + y^T\xi + \xi^T Y \xi \geq 0$ for all $\xi \in [0,1]^n$, which is equivalent to showing

$$\nu := \min \left\{ y_0 + y^T\xi + \xi^T Y \xi : \xi \in [0,1]^n \right\} \geq 0,$$

which in particular implies $y_0 \geq 0$. We have

$$\begin{aligned} \nu \geq \nu_{\text{SDP}} &:= \min \left\{ y_0 + y^T\xi + Y \bullet \Xi\ :\ \Xi \leq \xi e^T,\ \begin{pmatrix} 1 & \xi^T \\ \xi & \Xi \end{pmatrix} \succeq 0 \right\} \\ &= \max \left\{ y_0 - z_0\ :\ Z \geq 0,\ \begin{pmatrix} z_0 & (y^T - e^T Z)/2 \\ (y - Z^T e)/2 & Y + (Z + Z^T)/2 \end{pmatrix} \succeq 0 \right\}, \end{aligned}$$

where strong duality holds, for example, by Lemma 18 in Qiu and Yıldırım [2024]. Furthermore, $\nu = \nu_{\text{SDP}}$ holds when $n \leq 3$ by Theorem 1. In particular, $\nu \geq \nu_{\text{SDP}} \geq 0$ holds when

at least one dual solution has nonnegative objective value, which is true by assumption. The expression for $\mathcal{M}_{\text{SDP}}$ follows by standard dual constructions, giving $\mathcal{M}_0 \subseteq \mathcal{M}_{\text{SDP}}$ with equality for $n \leq 3$. $\square$

Note that the auxiliary square matrix $Z \in \mathbb{R}^{n\times n}$ used in $\mathcal{K}_{\text{SDP}}$ is not symmetric, while $W$ in $\mathcal{M}_{\text{SDP}}$ is symmetric.

We summarize the foregoing discussion by concluding that the ambiguity set $\mathcal{P}$ in (5) is nonempty only if there exists a matrix $W$ in $\mathbb{S}^n$ such that

$$\Sigma \leq W \leq \mu e^T, \quad \operatorname{diag}(W) = \operatorname{diag}(\Sigma), \qquad \begin{pmatrix} 1 & \mu^T \\ \mu & W \end{pmatrix} \succeq 0. \tag{8}$$

For $n \leq 3$, the converse also holds, so that (8) is a necessary and sufficient condition for nonemptiness of $\mathcal{P}$.

## 3.2 Semidefinite approximation

The following proposition shows that the optimal value of the distributionally robust problem (4) can be bounded above in polynomial time by a semidefinite program when the ambiguity set $\mathcal{P}$ is given by (5). Moreover, the bound is exact when $n \leq 3$.

**Proposition 8.** *Suppose:*

*(a) $\mathcal{X}$ is semidefinite representable;*

*(b) $-A_k(x)$ is submodular for each $x \in \mathcal{X}$ and each $k = 1, \ldots, K$.*

*Then* (4) *is bounded above in polynomial time by the SDP*

$$\begin{array}{lll} \inf & y_0 + \mu^T y + \Sigma \bullet Y & \\ \text{s.t.} & x \in \mathcal{X} & \\ & Y \text{ submodular} & \\ & Z_k \geq 0 & \forall\ k = 1, \ldots, K \\ & \begin{pmatrix} y_0 & (y^T - e^T Z_k)/2 \\ (y - Z_k^T e)/2 & Y + (Z_k + Z_k^T)/2 \end{pmatrix} \succeq \begin{pmatrix} c_k(x) & b_k^T(x)/2 \\ b_k(x)/2 & A_k(x) \end{pmatrix} & \forall\ k = 1, \ldots, K, \end{array} \tag{9}$$

*where the decision variables are $x \in \mathbb{R}^m$, $(y_0, y, Y) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n$, and $Z_k \in \mathbb{R}^{n\times n}$ for $k = 1, \ldots, K$ with equality holding for $n = 3$.*

*Proof.* For fixed $x \in \mathcal{X}$, denote the value of the inner supremum in (4) by

$$v^*(x) := \sup \left\{ \mathbb{E}_{\mathbb{P}} \left[ \max_{k=1,\ldots,K} \left( \tilde{\xi}^T A_k(x) \tilde{\xi} + b_k^T(x) \tilde{\xi} + c_k(x) \right) \right] : \mathbb{P} \in \mathcal{P} \right\}.$$

The dual is given by

$$\begin{aligned} v_d^*(x) = \inf \quad & y_0 + \mu^T y + Y \bullet \Sigma \\ \text{s.t.} \quad & Y \text{ submodular} \\ & y_0 + y^T\xi + \xi^T Y \xi \geq \max_{k=1,\ldots,K} \left(\xi^T A_k(x)\xi + b_k^T(x)\xi + c_k(x)\right) \quad \forall\, \xi \in [0,1]^n, \end{aligned}$$

where $v^*(x) \leq v_d^*(x)$ holds from weak duality. Disaggregating the dual constraints across $k = 1, \ldots, K$ gives

$$\begin{aligned} v_d^*(x) = \inf \quad & y_0 + \mu^T y + \Sigma \bullet Y \\ \text{s.t.} \quad & Y \text{ submodular} \\ & (y_0 - c_k(x), y - b_k(x), Y - A_k(x)) \in \mathcal{K}_0 \quad \forall\, k = 1, \ldots, K. \end{aligned}$$

Here $Y - A_k(x)$ is submodular for each $k$ and $x$. Replacing $\mathcal{K}_0$ with the inner approximation $\mathcal{K}_{\text{SDP}} \subseteq \mathcal{K}_0$ from Proposition 7 gives

$$\begin{aligned} \overline{v}_d^*(x) = \inf \quad & y_0 + \mu^T y + \Sigma \bullet Y \\ \text{s.t.} \quad & Y \text{ submodular} \\ & (y_0 - c_k(x), y - b_k(x), Y - A_k(x)) \in \mathcal{K}_{\text{SDP}} \quad \forall\, k = 1, \ldots, K, \end{aligned}$$

where $v^*(x) \leq v_d^*(x) \leq \overline{v}_d^*(x)$. Optimizing over $x \in \mathcal{X}$, we obtain the SDP upper bound as claimed. Equality holds for $n \leq 3$ due to Theorem 1 and Proposition 7. □

### 3.3 Comparison with an existing moment ambiguity set

A popular and tractable moment ambiguity set analyzed by Delage and Ye [2010] is

$$\mathcal{Q} = \left\{ \mathbb{P} \in \mathcal{P}(\Xi) \;:\; \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu,\ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] \preceq \Sigma \right\}. \tag{10}$$

Here, the support of the random vector $\tilde{\xi}$ is contained in $\Xi$ which is assumed to be convex and compact, the mean is fixed to $\mu$, and the second moment matrix is bounded from above by $\Sigma$ in the positive semidefinite order. It is well-known that $\mathcal{Q}$ is nonempty if and only if

$$\mu \in \Xi,\ \begin{pmatrix} 1 & \mu^T \\ \mu & \Sigma \end{pmatrix} \succeq 0.$$

It is interesting to compare these conditions with the analogous ones (8) for our ambiguity set $\mathcal{P}$. These two sets express different phenomena, e.g., the matrix $\Sigma$ in (10) needs to be positive semidefinite for $\mathcal{Q}$ to be nonempty but not for $\mathcal{P}$ to be nonempty. Furthermore,

when $\Xi = [0,1]^n$ and the matrix $A_k(x)$ is negative semidefinite for each $k$, the distributionally robust optimization (4)—where $\mathcal{P}$ is replaced by $\mathcal{Q}$—can be reformulated as an SDP using duality (see Delage and Ye [2010]):

$$
\begin{array}{ll}
\min & y_0 + \mu^T y + \Sigma \bullet Y \\
\text{s.t.} & x \in \mathcal{X} \\
& Y \succeq 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (11) \\
& z_k, w_k \geq 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\ k = 1, \ldots, K \\
& \begin{pmatrix} y_0 - e^T z_k & (y + z_k - w_k)^T/2 \\ (y + z_k - w_k)/2 & Y \end{pmatrix} \succeq \begin{pmatrix} c_k(x) & b_k^T(x)/2 \\ b_k(x)/2 & A_k(x) \end{pmatrix} \quad \forall\ k = 1, \ldots, K,
\end{array}
$$

where the decision variables are $x \in \mathbb{R}^m$, $(y_0, y, Y) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n$, and $(z_k, w_k) \in \mathbb{R}^n \times \mathbb{R}^n$ for $k = 1, \ldots, K$. We provide a numerical comparison of the formulations (9) and (11) in Section 5 with two examples that arise in: (i) distributionally robust stratified sampling allocation; and (ii) moment bounds on the quadratic energy of Laplacian matrices where the matrices $-A_k(x)$ are both positive semidefinite and submodular, which allows direct comparison of the SDP relaxations.

# 4 Covariance Bounds

This section studies sharp upper bounds on covariance terms when only first and second moments are known and the support is $[0,1]^n$. We first formulate a moment problem and characterize feasibility, then derive an SDP upper bound that is exact for $n \leq 3$ under submodularity. We then provide closed-form bounds and explicit extremal constructions for $n = 2$ and then for general $n$ under an additional overlap condition on marginal moments.

## 4.1 A moment problem and SDP relaxation

Given only the mean and variance of two random variables $\tilde{\xi}_1$ and $\tilde{\xi}_2$, with no restrictions on their support, a well-known upper bound on the pairwise moment is given by:

$$\mathbb{E}[\tilde{\xi}_1 \tilde{\xi}_2] \leq \mathbb{E}[\tilde{\xi}_1]\mathbb{E}[\tilde{\xi}_2] + \sqrt{\mathrm{Var}[\tilde{\xi}_1]\mathrm{Var}[\tilde{\xi}_2]}. \tag{12}$$

This bound is known to be best possible. We consider a multivariate version of this problem and develop sharp covariance bounds for bounded random variables where the means and variances are known.

For $i = 1, \ldots, n$, suppose random variable $\tilde{\xi}_i$ has support contained in $[0, 1]$ with known mean $\mu_i$ and standard deviation $\sigma_i$, or equivalently the variance is $\sigma_i^2$. Let $(\mu, \sigma)$ represent the vector of means and standard deviations of the random vector $\tilde{\xi}$. Given a matrix $A \in \mathbb{S}^n$, consider the moment problem

$$v^* = \max \left\{ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}^T A \tilde{\xi}] \ : \ \begin{array}{c} \mathbb{P} \in \mathcal{P}([0,1]^n), \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu, \\ \mathbb{E}_{\mathbb{P}}[\mathrm{diag}(\tilde{\xi}\tilde{\xi}^T)] = \mathrm{diag}(\mu\mu^T + \sigma\sigma^T) \end{array} \right\}, \tag{13}$$

i.e., $v^*$ is the best possible upper bound on the expectation of $\tilde{\xi}^T A \tilde{\xi}$ given only the means and variances in $\tilde{\xi}$ with support contained in the unit hypercube.

The next lemma provides necessary and sufficient conditions for the feasibility of (13).

**Lemma 3.** *The moment problem* (13) *is feasible if and only if* $0 \leq \sigma_i \leq \sqrt{\mu_i(1-\mu_i)}$*for all* $i = 1, \ldots, n$.

*Proof.* Necessity of $\sigma_i \geq 0$ arises from $\mathbb{E}_{\mathbb{P}}[(\tilde{\xi}_i - \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i])^2] \geq 0$. Necessity of $\sigma_i \leq \sqrt{\mu_i(1-\mu_i)}$ arises from $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i(1 - \tilde{\xi}_i)] \geq 0$ since $\xi_i \in [0, 1]$. Sufficiency follows from considering the two point random variable $\tilde{\xi}_i$:

$$\tilde{\xi}_i = \begin{cases} \mu_i - \sigma_i \sqrt{\frac{p_i}{1-p_i}} & \text{w.p. } 1 - p_i, \\ \mu_i + \sigma_i \sqrt{\frac{1-p_i}{p_i}} & \text{w.p. } p_i \end{cases}$$

where $p_i \in [\sigma_i^2/((1-\mu_i)^2 + \sigma_i^2), \mu_i^2/(\mu_i^2 + \sigma_i^2)]$ to ensure that the support of the random variable is in $[0, 1]$ with $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i] = \mu_i$ and $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i^2] = \mu_i^2 + \sigma_i^2$. The random vector $\tilde{\xi}$ can be constructed from the marginal distributions using independence. □

We next show that under the assumption that $-A$ is submodular, the covariance-bounding moment problem admits a tractable SDP upper bound, and this bound is exact in dimensions $n \leq 3$.

**Proposition 9.** *Suppose:*

*(a)* $0 \leq \sigma_i \leq \sqrt{\mu_i(1-\mu_i)}$ *for all* $i = 1, \ldots, n$*;*

*(b) the matrix* $-A$ *is submodular.*

*Then the semidefinite program*

$$\begin{aligned} v^*_{\mathrm{SDP}} = \max \quad & A \bullet \Sigma \\ \text{s.t.} \quad & \operatorname{diag}(\Sigma) = \operatorname{diag}(\mu\mu^T + \sigma\sigma^T) \\ & \Sigma \le e\mu^T \\ & \begin{pmatrix} 1 & \mu^T \\ \mu & \Sigma \end{pmatrix} \succeq 0. \end{aligned} \tag{14}$$

*where the decision variable is $\Sigma \in \mathbb{S}^n$, provides an upper bound $v^* \le v^*_{\mathrm{SDP}}$ on the moment problem* (13)*, with equality for $n \le 3$.*

In (14), the diagonal constraint fixes the second moments, the componentwise inequality enforces $\Sigma_{ij} \le \mu_j$, and the block positive-semidefinite constraint is the usual moment-matrix consistency condition.

*Proof.* We proceed in four steps: dual reformulation of (13), replacement of the QSMB constraint by its SDP relaxation, dualization of the inner SDP, and recovery of the primal semidefinite program (14).

**Step 1 (dual reformulation).** Assumption (a) and Lemma 3 imply (13) is feasible. Its dual is

$$\begin{aligned} v^*_d = \inf \quad & y_0 + \mu^T y + \operatorname{diag}(\mu\mu^T + \sigma\sigma^T)^T z \\ \text{s.t.} \quad & y_0 + \xi^T y + \xi^T \operatorname{Diag}(z)\xi \ge \xi^T A \xi \quad \forall\, \xi \in [0,1]^n, \end{aligned}$$

where the decision variables are $y_0 \in \mathbb{R}, y \in \mathbb{R}^n$, and $z \in \mathbb{R}^n$. Then strong duality holds with $v^* = v^*_d$ because the dual is strictly feasible. For example, we can set $y = z = 0$ and then take $y_0$ larger than the maximum value of $\xi^T A \xi$ over $\xi \in [0,1]^n$. We next reformulate the dual as

$$\begin{aligned} v^* = \inf \quad & y_0 + \mu^T y + \operatorname{diag}(\mu\mu^T + \sigma\sigma^T)^T z \\ \text{s.t.} \quad & v(z) := \min \left\{ y_0 + y^T\xi + \xi^T(\operatorname{Diag}(z) - A)\xi \ : \ \xi \in [0,1]^n \right\} \ge 0. \end{aligned}$$

**Step 2 (SDP relaxation of QSMB).** Under assumption (b), $\operatorname{Diag}(z) - A$ is submodular. Replacing the QSMB constraint with the SDP relaxation from Theorem 1 gives the following optimization:

$$\begin{aligned} \rho^* = \inf \quad & y_0 + \mu^T y + \operatorname{diag}(\mu\mu^T + \sigma\sigma^T)^T z \\ \text{s.t.} \quad & \rho(z) := \min \left\{ y_0 + y^T\xi + (\operatorname{Diag}(z) - A) \bullet \Xi \ : \ \Xi \le \xi e^T, \begin{pmatrix} 1 & \xi^T \\ \xi & \Xi \end{pmatrix} \succeq 0 \right\} \ge 0. \end{aligned}$$

Note that $v(z) \ge \rho(z)$ with equality for $n \le 3$ by Theorem 1, which means the SDP

constraint is more restrictive than the original QSMB constraint. Hence, $v^* \leq \rho^*$ with equality for $n \leq 3$.

**Step 3 (dualization of the inner SDP).** Taking the dual of the minimization in the constraint, where both the primal and dual semidefinite programs are strictly feasible, and following a similar argument as in the proof of Proposition 8, we get

$$\begin{aligned} \rho^* = \inf \quad & \lambda + \mu^T y + \operatorname{diag}(\mu\mu^T + \sigma\sigma^T)^T z \\ \text{s.t.} \quad & Z \geq 0 \\ & \begin{pmatrix} \lambda & (y - Z^T e)^T/2 \\ (y - Z^T e)/2 & \operatorname{Diag}(z) + (Z + Z^T)/2 \end{pmatrix} \succeq \begin{pmatrix} 0 & 0^T \\ 0 & A \end{pmatrix}. \end{aligned}$$

Here, $\lambda$ is an auxiliary scalar variable that plays the role of $y_0$. It arises as the top-left entry in the dual of the inner SDP, exactly as $z_0$ does in the proof of Proposition 8, and the constraint $\rho(z) \geq 0$ forces $\lambda = y_0$ at optimality. In particular, the block positive-semidefinite constraint implies $\lambda \geq 0$, consistent with $y_0 \geq 0$.

**Step 4 (recover the primal SDP).** The primal semidefinite formulation is then given by (14). In particular, $v^*_{\text{SDP}} = \rho^* \geq v^*$ with equality for $n \leq 3$. □

## 4.2 Closed-form solution in dimension 2

We solve the moment problem (13) in closed form when $n = 2$ and

$$A = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{15}$$

by explicitly constructing the extremal distribution that attains the upper bound.

**Proposition 10.** *Consider (13) for $n = 2$ and $A$ given by (15). The upper bound is given by*

$$\mathbb{E}_{\mathbb{P}}[\tilde{\xi}_1 \tilde{\xi}_2] \leq v^* = v^*_{\text{SDP}} = \min\left(\mu_1, \mu_2, \mu_1\mu_2 + \sigma_1\sigma_2\right), \tag{16}$$

*and this bound is best possible for all feasible parameter values $\mu_i := \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i]$ and $\sigma_i := \sqrt{Var[\tilde{\xi}_i]}$ for $i = 1, 2$.*

*Proof.* For $n = 2$, the semidefinite program (14) reduces to

$$v^* = v^*_{\text{SDP}} = \max \left\{ \Sigma_{12} \ : \ \Sigma_{12} \leq \mu_1, \ \Sigma_{12} \leq \mu_2, \ \begin{pmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \mu_1^2 + \sigma_1^2 & \Sigma_{12} \\ \mu_2 & \Sigma_{12} & \mu_2^2 + \sigma_2^2 \end{pmatrix} \succeq 0 \right\}.$$

Using the Schur complement theorem, we rewrite this as

$$\begin{aligned} v^* &= \max\left\{\Sigma_{12} \;:\; \Sigma_{12} \le \mu_1,\ \Sigma_{12} \le \mu_2,\ \begin{pmatrix} \sigma_1^2 & \Sigma_{12} - \mu_1\mu_2 \\ \Sigma_{12} - \mu_1\mu_2 & \sigma_2^2 \end{pmatrix} \succeq 0\right\} \\ &= \max\left\{\Sigma_{12} \;:\; \Sigma_{12} \le \mu_1,\ \Sigma_{12} \le \mu_2,\ \mu_1\mu_2 - \sigma_1\sigma_2 \le \Sigma_{12} \le \mu_1\mu_2 + \sigma_1\sigma_2\right\}. \end{aligned}$$

It follows that $v^* = \min\{\mu_1, \mu_2, \mu_1\mu_2 + \sigma_1\sigma_2\}$.

We now construct the extremal joint distribution using the two-point marginal distributions from Lemma 3 to show attainment of the bound. Define $\underline{p}_i = \sigma_i^2/((1-\mu_i)^2 + \sigma_i^2)$ and $\overline{p}_i = \mu_i^2/(\mu_i^2 + \sigma_i^2)$ for $i = 1, 2$, and note that $\underline{p}_i \le \overline{p}_i$ for $i = 1, 2$.

**Case 1 (overlap).** Suppose the intervals $[\underline{p}_1, \overline{p}_1]$ and $[\underline{p}_2, \overline{p}_2]$ overlap, and let $p$ be any value in the overlapping region. Consider the two-point joint distribution with perfect positive dependence:

$$(\tilde{\xi}_1, \tilde{\xi}_2) = \begin{cases} \left(\mu_1 - \sigma_1\sqrt{\frac{p}{1-p}}, \mu_2 - \sigma_2\sqrt{\frac{p}{1-p}}\right) & \text{w.p.} \quad 1-p, \\ \left(\mu_1 + \sigma_1\sqrt{\frac{1-p}{p}}, \mu_2 + \sigma_2\sqrt{\frac{1-p}{p}}\right) & \text{w.p.} \quad p. \end{cases}$$

In this case,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^*}[\tilde{\xi}_1\tilde{\xi}_2] &= (1-p)\left(\mu_1 - \sigma_1\sqrt{\tfrac{p}{1-p}}\right)\left(\mu_2 - \sigma_2\sqrt{\tfrac{p}{1-p}}\right) + p\left(\mu_1 + \sigma_1\sqrt{\tfrac{1-p}{p}}\right)\left(\mu_2 + \sigma_2\sqrt{\tfrac{1-p}{p}}\right) \\ &= \mu_1\mu_2 + \sigma_1\sigma_2. \end{aligned}$$

To verify that this bound is attained, note that the overlap condition implies both $\underline{p}_2 \le \overline{p}_1$ and $\underline{p}_1 \le \overline{p}_2$, which are equivalent to $\mu_1\mu_2 + \sigma_1\sigma_2 \le \mu_1$ and $\mu_1\mu_2 + \sigma_1\sigma_2 \le \mu_2$, respectively. Hence $v^* = \min\{\mu_1, \mu_2, \mu_1\mu_2 + \sigma_1\sigma_2\} = \mu_1\mu_2 + \sigma_1\sigma_2$, and the bound is attained.

**Case 2 (non-overlap).** Suppose the intervals $[\underline{p}_1, \overline{p}_1]$ and $[\underline{p}_2, \overline{p}_2]$ do not overlap and, without loss of generality, suppose $\overline{p}_1 \le \underline{p}_2$. Consider the three-point joint distribution with perfect positive dependence:

$$(\tilde{\xi}_1, \tilde{\xi}_2) = \begin{cases} (0, (\mu_2 - \mu_2^2 - \sigma_2^2)/(1-\mu_2)) & \text{w.p.} \quad 1 - \underline{p}_2, \\ (0, 1) & \text{w.p.} \quad \underline{p}_2 - \overline{p}_1, \\ (\mu_1 + \sigma_1^2/\mu_1, 1) & \text{w.p.} \quad \overline{p}_1. \end{cases}$$

In this case, $\mathbb{E}_{\mathbb{P}^*}[\tilde{\xi}_1\tilde{\xi}_2] = \mu_1$. Since $\overline{p}_1 \le \underline{p}_2$, we have $\mu_1 \le \mu_1\mu_2 + \sigma_1\sigma_2$. On the other hand, the chain $\underline{p}_1 \le \overline{p}_1 \le \underline{p}_2 \le \overline{p}_2$ shows that $\underline{p}_1 \le \overline{p}_2$ holds, which is equivalent to $\mu_1\mu_2 + \sigma_1\sigma_2 \le \mu_2$. Combining these two inequalities yields $\mu_1 \le \mu_1\mu_2 + \sigma_1\sigma_2 \le \mu_2$. Hence $v^* = \min\{\mu_1, \mu_2, \mu_1\mu_2 + \sigma_1\sigma_2\} = \mu_1$, and the bound is attained. □

The attainability of this upper bound for $n = 2$ has also been recently shown in Hössjer and Sjölander [2022] (see Theorem 2 therein), although their proof technique differs from ours.

The $n = 2$ result above shows when the bivariate upper bound is attainable exactly. We now extend this perspective to general $n$ and identify a condition under which all pairwise upper bounds are simultaneously attainable.

## 4.3 Closed-form solution in dimension $n$

We can also solve problem (13) in closed form for general $n$ under additional moment conditions when

$$A = \frac{1}{2}\begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{pmatrix} = \frac{1}{2}(ee^T - I) \tag{17}$$

by explicitly constructing the extremal distribution that attains the upper bound.

**Proposition 11.** *Consider (13) for general $n$ and $A$ given by (17), where the marginal moments satisfy the condition*

$$\max_{i=1,\dots,n} \frac{\sigma_i^2}{(1-\mu_i)^2 + \sigma_i^2} \leq \min_{i=1,\dots,n} \frac{\mu_i^2}{\mu_i^2 + \sigma_i^2}. \tag{18}$$

*In words, condition* (18) *guarantees that all intervals $[\underline{p}_i, \overline{p}_i]$ have a common overlap, which allows one comonotone two-point construction to attain all pairwise upper bounds at once. Then the best possible upper bound is given by*

$$v^* = \sum_{i<j} \left(\mu_i \mu_j + \sigma_i \sigma_j\right). \tag{19}$$

*Proof.* For general $n$, an upper bound on $v^*$ (not necessarily attained) is obtained by adding up the respective bivariate bounds from Proposition 10:

$$v^* \leq \sum_{(i,j):i<j} \min\left(\mu_i, \mu_j, \mu_i \mu_j + \sigma_i \sigma_j\right). \tag{20}$$

We now construct the extremal joint distribution that attains this bound. Define $\underline{p}_i = \sigma_i^2/((1-\mu_i)^2 + \sigma_i^2)$ and $\overline{p}_i = \mu_i^2/(\mu_i^2 + \sigma_i^2)$, and note that $\underline{p}_i \leq \overline{p}_i$. Under condition (18), the intervals $[\underline{p}_i, \overline{p}_i]$ overlap for $i = 1, \dots, n$. Let $p$ be any value in the overlapping region. Consider the following two-point marginal distribution for each random variable with perfect

positive dependence:

$$\tilde{\xi}_i = \begin{cases} \mu_i - \sigma_i\sqrt{\frac{p}{1-p}} & \text{w.p. } 1-p, \\ \mu_i + \sigma_i\sqrt{\frac{1-p}{p}} & \text{w.p. } p. \end{cases}$$

In this case for all $(i,j)$ with $i<j$, we have:

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^*}[\tilde{\xi}_i\tilde{\xi}_j] &= (1-p)\left(\mu_i - \sigma_i\sqrt{\tfrac{p}{1-p}}\right)\left(\mu_j - \sigma_j\sqrt{\tfrac{p}{1-p}}\right) + p\left(\mu_i + \sigma_i\sqrt{\tfrac{1-p}{p}}\right)\left(\mu_j + \sigma_j\sqrt{\tfrac{1-p}{p}}\right) \\ &= \mu_i\mu_j + \sigma_i\sigma_j. \end{aligned}$$

Since condition (18) ensures that the intervals $[\underline{p}_i, \overline{p}_i]$ overlap for all pairs $(i,j)$, the Case 1 analysis from Proposition 10 gives $\min\{\mu_i, \mu_j, \mu_i\mu_j + \sigma_i\sigma_j\} = \mu_i\mu_j + \sigma_i\sigma_j$ for each pair. Hence

$$v^* \leq \sum_{(i,j):i<j} \left(\mu_i\mu_j + \sigma_i\sigma_j\right), \tag{21}$$

and the constructed distribution attains this bound, establishing equality. □

We have the following immediate corollary of Proposition 11 because the overlap condition (18) is satisfied when all $\mu_i$ are equal and all $\sigma_i$ are equal.

**Corollary 1.** *The bound in Proposition 11 is attained for all $n$ when $\mu_i = \mu_0$ and $\sigma_i = \sigma_0$ for some $\mu_0$ and $\sigma_0$ and all $i = 1, \ldots, n$.*

# 5 Numerical Results

In support of Sections 2, 3, and 4, we next investigate the numerical behavior of our SDP relaxation for QSMB. Specifically, we first examine its behavior in higher dimensions and then present numerical studies based on three examples in multi-product pricing, distributionally robust stratified sampling allocation, and moment bounds on the Laplacian quadratic energy.

## 5.1 Empirical behavior of the relaxation when $n \geq 4$

Theorem 1 establishes that the SDP relaxation (2) is tight for $n \leq 3$ but generally inexact for $n \geq 4$. To investigate how empirical tightness extends to higher dimensions, we conducted a numerical study using 1,000 randomly generated submodular instances with dimensions $n$ sampled uniformly from $\{4, 5, \ldots, 20\}$.

Each instance is constructed as follows. A symmetric matrix $Q \in \mathbb{R}^{n\times n}$ is generated by drawing entries from the standard normal distribution and symmetrizing, after which all off-diagonal entries are replaced by $-|Q_{ij}|$ to enforce the submodular structure $Q_{ij} \leq 0$ for

$i \neq j$. A cost vector $c \in \mathbb{R}^n$ is drawn independently from the standard normal distribution. Note that no restrictions are placed on the signs of $c$ or the diagonal of $Q$, so that these instances are not covered by the existing polynomial-time results of Kim and Kojima [2003] (which require $c \leq 0$) or those for DR-submodular functions (which require $\mathrm{diag}(Q) \leq 0$).

For each instance, we solve the QP (1) to global optimality using Gurobi [Gurobi Optimization, LLC, 2024] and then solve the SDP relaxation (2) using Mosek [MOSEK ApS, 2024], both with stringent solver tolerances ($10^{-8}$ for the QP and $10^{-10}$ for the SDP). We measure the quality of the relaxation via the relative gap

$$\text{relative gap} := \frac{p^* - d^*}{\max\Big\{1, |p^* + d^*|/2\Big\}},$$

where $p^*$ is the optimal value of the QP and $d^*$ is the dual objective value of the SDP. We use the dual objective because it provides a rigorous lower bound via weak duality. By weak duality, $p^* \geq d^*$, so the relative gap is nonnegative, and a gap of zero indicates tightness.

Figure 2 displays the distribution of relative gaps across all 1,000 instances using a kernel density estimate with gap values plotted on a logarithmic (base-10) scale. The gaps range from approximately $10^{-14}$ to $10^{-9}$, with the density peaking near $10^{-10}$. These values are well within solver precision, indicating that the SDP relaxation (2) is tight to machine accuracy for every instance tested. In particular, every relative gap is several orders of magnitude below $10^{-4}$, a conservative threshold for declaring numerical tightness.

While these results suggest tightness occurs often for every $n$, Example 4 of course shows that tightness is not a generic property. The experiment is therefore best understood as evidence that violations of tightness are rare.

A further search at each fixed dimension $n \in \{4, 5, 6, 7\}$, comprising 34,000 instances generated by the same random submodular construction, likewise produced no relative gap exceeding the $10^{-4}$ threshold. These runs also reveal a structural regularity. In all but one of the 34,000 instances the optimal SDP solution satisfied $\mathrm{rank}(Y(x, X)) = 1$, and no s-lower inequality was active at any of them. Any rank-one solution lies in $\mathcal{CP}(\mathcal{J})$, so tightness follows immediately from Lemma 2, without recourse to the more delicate rank arguments of Propositions 3–5. In particular, none of the higher-rank configurations identified in Remark 1 arose in any of these randomly generated instances even though, by Example 4, such configurations do occur and can produce a gap. That configuration is therefore highly non-generic, which explains how the relaxation can fail in principle while being tight on every instance sampled here.

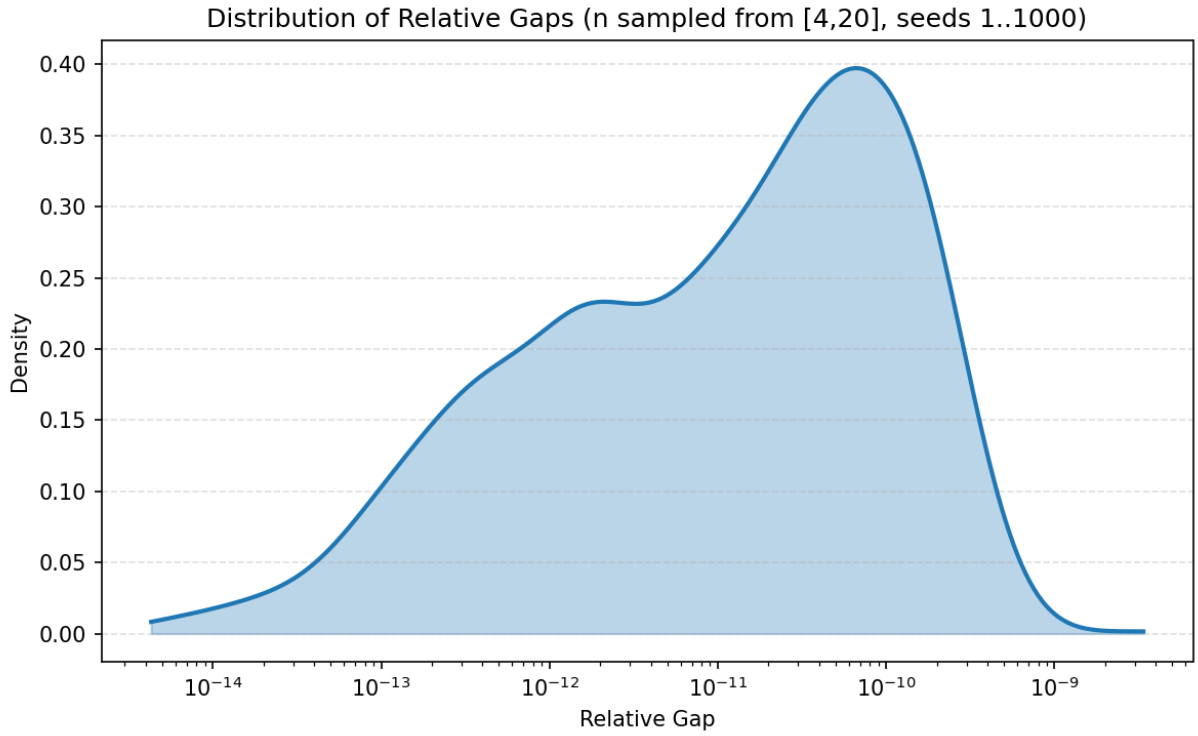


Figure 2: Distribution of relative gaps between the QP optimal value and the SDP dual bound over 1,000 random submodular instances with $n$ sampled uniformly from $\{4, \ldots, 20\}$. All gaps are within solver precision. Example 4 shows that tightness nevertheless fails for some submodular instances.

| Bundle | Marginal cost | Baseline price | Baseline demand | Price range |
|---|---|---|---|---|
| 10 GB | \$3 | \$8 | 400 | [\$6, \$10] |
| 50 GB | \$8 | \$18 | 180 | [\$14, \$22] |
| 100 GB | \$14 | \$32 | 98.5 | [\$24, \$36] |

Table 2: Data for the product pricing problem where demand is measured in thousands.

| Bundle | Optimal price | Optimal demand | Optimal margin | Profit (in millions) |
|---|---|---|---|---|
| 10 GB | 10 | 428 | \$7 | \$2.996 |
| 50 GB | 22 | 169.5 | \$14 | \$2.373 |
| 100 GB | 30 | 117.5 | \$16 | \$1.880 |

Table 3: Optimal prices, demands, margins and product-level profits.

## 5.2 Pricing with $n = 3$ products

We consider a pricing problem faced by a mobile data provider, who sells three prepaid data bundles of 10 GB, 50 GB and 100 GB. The linear demand model is given by $d(p) = a - Bp$, where

$$a = \begin{pmatrix} 208 \\ 220 \\ 342 \end{pmatrix}, \qquad B = \begin{pmatrix} 25 & -20 & -1 \\ -3 & 4 & -\frac{1}{4} \\ -1 & -\frac{1}{4} & 8 \end{pmatrix}.$$

The price variable $p$ is measured in dollars, and demand $d(p)$ is measured in thousands of bundles. The marginal costs in dollars, baseline prices in dollars, baseline demands in thousands per month, and admissible price ranges are provided in Table 2. Over the admissible price ranges, the demands are nonnegative. The matrix $(B + B^T)/2$ is indefinite, and so the pricing problem is not directly solvable as a convex quadratic program. The baseline profit, i.e., before optimization, equals \$5.573 million per month. Since $n = 3$, Theorem 1 guarantees that the SDP relaxation (2) is tight, and solving it returns the exact optimum of the pricing problem:

$$p^* = \begin{pmatrix} 10 \\ 22 \\ 30 \end{pmatrix}, \qquad d(p^*) = \begin{pmatrix} 428 \\ 169.5 \\ 117.5 \end{pmatrix}.$$

The corresponding optimal profit equals \$7.249 million per month, an increase of approximately 30.07% over the baseline profit. The product-level profit calculation is shown in Table 3, where the optimal margin is the optimal price less the marginal cost, and the profit is the margin times the optimal demand, expressed in millions of dollars per month. In the optimal solution, the prices of the small and medium bundles are raised to their upper bounds, while the price of the largest bundle is lowered \$2 from its baseline.

## 5.3 Distributionally robust stratified sampling allocation

Allocation problems in stratified sampling are solvable using optimization methods; see Neyman [1934], Srikantan [1963]. In this subsection, we focus on a distributionally robust variant. Consider a population of individuals where each individual is associated with a random Bernoulli outcome variable and the probability of the Bernoulli outcome taking a value of 1 is itself random. Such random variables are typically modeled using the Beta-Bernoulli distribution, but we consider a distributionally robust approach instead of requiring the probability to follow a Beta distribution.

Assume the population is divided into $n$ strata indexed by $i$. Let $\tilde{\xi}_i \in [0,1]$ denote the probability of the Bernoulli outcome taking a value of 1 in stratum $i$, which is itself random. Conditional on $\tilde{\xi}_i$, the Bernoulli outcome for individual $j$ from stratum $i$ is

$$\tilde{b}_{ij}|\tilde{\xi}_i \sim \text{Bernoulli}(\tilde{\xi}_i).$$

When the outcomes are conditionally independent within each stratum and across strata, the sample-proportion estimator is given by

$$\tilde{b}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} \tilde{b}_{ij},$$

where $m_i$ is the number of samples allocated to stratum $i$. The conditional mean and variance of the estimator for stratum $i$ are given by

$$\mathbb{E}[\tilde{b}_i|\tilde{\xi}_i] = \tilde{\xi}_i, \qquad \text{Var}[\tilde{b}_i|\tilde{\xi}_i] = \frac{\tilde{\xi}_i(1-\tilde{\xi}_i)}{m_i}.$$

Suppose the decision maker has a sampling budget $B$, and suppose $c_i > 0$ is the unit sampling cost for stratum $i$. When the joint distribution of the probability vector $\tilde{\xi}$ is known and given by $\mathbb{P}$, one can find an optimal allocation which minimizes the ex-ante maximum variance by solving

$$\min\left\{\mathbb{E}\left[\max_{i\in[n]} \frac{\tilde{\xi}_i(1-\tilde{\xi}_i)}{m_i}\right] :\ c^T m \le B, m \ge 0\right\}. \tag{22}$$

See Krause et al. [2008] for similar models that minimize the maximum posterior variance in Gaussian process regression. Further, we consider a distributionally robust version of this problem where the joint distribution of the probabilities is unknown and the moment

ambiguity set $\mathcal{P}$ is defined as in (5):

$$\min \left\{ \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[ \max_{i\in[n]} \frac{\tilde{\xi}_i(1-\tilde{\xi}_i)}{m_i} \right] : \; c^T m \leq B, m \geq 0 \right\}. \tag{23}$$

For comparison, we also solve the problem with the ambiguity set $\mathcal{Q}$ defined in (10).

Since $\Xi = [0,1]^n$, the ambiguity sets $\mathcal{P}$ and $\mathcal{Q}$ are relevant in this context. The objective function inside the expectation is given by the maximum of $n$ random quadratic terms where the corresponding $A_i$ matrices are defined by diagonal matrices with negative entries along the diagonals. This satisfies the sufficient conditions discussed in Section 3 for both ambiguity sets. It is also straightforward to see that the formulation is convex in $m$ by a simple transformation where we replace $1/m_i$ by $m_i$.

For a fixed allocation $m$, the inner worst-case expectation is precisely a moment problem of the form studied in Section 3, and we evaluate it by solving the associated semidefinite program. Since $n = 3$ and the objective is submodular, Theorem 1 guarantees that this semidefinite program returns the exact worst-case value rather than a bound. The outer minimization over $m$ is then handled by enumeration. Although (22) is convex in $m$ after the substitution just described, sample sizes must in practice be integers, and for the instance below there are only 406 integer allocations to consider, so we simply solve the inner semidefinite program at each of them and select the best.

For a numerical example, consider $n = 3$ strata with equal sampling costs $c = (1,1,1)^T$ and budget $B = 30$. The $\mu$ vector and the $\Sigma$ matrix are given by

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 0.49 & 0.25 & 0.25 \\ 0.25 & 0.29 & 0.25 \\ 0.25 & 0.25 & 0.26 \end{pmatrix}.$$

We enumerate over all possible positive sampling integer allocations with $m_1+m_2+m_3 = 30$, which yields

$$m^*_{\mathcal{P}} = \begin{pmatrix} 4 \\ 13 \\ 13 \end{pmatrix}, \qquad m^*_{\mathcal{Q}} = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}.$$

Table 4 reports the worst-case objective value for each ambiguity set at each of these two allocations: a row indexes the ambiguity set over which the worst case is taken, and a column indexes the allocation at which it is evaluated. The diagonal entries, shown in bold, are the optimal values. Each off-diagonal entry is what one model delivers when the other model's allocation is used.

| Ambiguity set | $m^*_{\mathcal{P}}$ | $m^*_{\mathcal{Q}}$ |
|---|---|---|
| $\mathcal{P}$ | **0.0210** | 0.0250 |
| $\mathcal{Q}$ | 0.0625 | **0.0250** |

Table 4: Objective functions for a given ambiguity set and given sampling allocation (optimal solutions are highlighted in bold).

The two allocations differ because the two ambiguity sets encode different information. Under $\mathcal{P}$, the diagonal second moments are specified exactly, so the expected conditional variance in each stratum is known exactly as well:

$$\mathbb{E}[\tilde{\xi}_i(1-\tilde{\xi}_i)] = \mu_i - \Sigma_{ii} = 0.01,\ 0.21,\ 0.24 \quad \text{for } i = 1, 2, 3.$$

Stratum 1 is an extreme case. Its probability $\tilde{\xi}_1$ varies a great deal across the population, with $\mathrm{Var}[\tilde{\xi}_1] = 0.24$, and yet conditional on $\tilde{\xi}_1$ the Bernoulli outcome is nearly deterministic, so very few samples are needed there. The allocation $m^*_{\mathcal{P}} = (4, 13, 13)$ reflects precisely this, moving samples away from stratum 1 and toward strata 2 and 3.

Under $\mathcal{Q}$, by contrast, only the positive semidefinite bound $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] \preceq \Sigma$ is imposed, and the deterministic distribution placing all of its mass at $\tilde{\xi} = (0.5, 0.5, 0.5)$ is feasible, because $\Sigma - 0.25\, ee^T \succeq 0$. Under that distribution every stratum attains the largest conditional variance possible, $\tilde{\xi}_i(1-\tilde{\xi}_i) = 1/4$, so all strata look equally noisy and the equal allocation $m^*_{\mathcal{Q}} = (10, 10, 10)$ is optimal. The same distribution accounts for the $\mathcal{Q}$ row of Table 4. At any allocation it gives $\max_i 1/(4m_i)$, which is 0.0625 at $m^*_{\mathcal{P}}$ and 0.0250 at $m^*_{\mathcal{Q}}$.

Two features of Table 4 deserve comment. First, at $m^*_{\mathcal{P}}$ the two models differ by a factor of three, so the additional information encoded in $\mathcal{P}$ translates into a materially better allocation. Second, the two models agree exactly at $m^*_{\mathcal{Q}}$. This is not a coincidence: $\max_i \tilde{\xi}_i(1-\tilde{\xi}_i) \leq 1/4$ holds for every distribution supported on $[0,1]^n$, so $1/(4 \cdot 10) = 0.0250$ is an upper bound at the equal allocation for *any* ambiguity set, and both $\mathcal{P}$ and $\mathcal{Q}$ contain a distribution attaining it. In other words, the equal allocation is the one place where the extra information in $\mathcal{P}$ cannot help. Taken together, the example shows that the new ambiguity set can deliver strictly tighter bounds, and correspondingly better decisions, than the existing one.

## 5.4 Quadratic energy of Laplacian matrices

Let $G = (V, E)$ be a simple undirected graph, where $V$ is the set of vertices and $E$ is the set of edges. We focus on unweighted graphs, but the results extend directly to nonnegative edge weights. The Laplacian matrix $L$ associated with $G$ is a symmetric $|V| \times |V|$ matrix

with diagonal entries $L_{ii} = \deg(i)$ for $i \in V$ and off-diagonal entries $L_{ij} = -1$ when $i$ is adjacent to $j$ and 0 otherwise. The Laplacian is submodular. It is also diagonally dominant, and therefore positive semidefinite.

Associated with graph $G$ is the *quadratic energy* defined by

$$E(\xi) := \xi^T L \xi = \sum_{(i,j)\in E} (\xi_i - \xi_j)^2,$$

where $\xi \in \mathbb{R}^{|V|}$. Minimization of $E(\xi)$ under appropriate constraints on $\xi$ has found applications in graph machine learning and spring and resistor networks, and there are fast algorithms available to solve these problems [Spielman, 2010].

In spring and resistor networks, it is natural to *ground* some of the vertices, that is, to hold them at a fixed reference value. In a resistor network, a grounded vertex is held at zero potential, while in a spring network, it is a fixed anchor point rather than a free mass. Accordingly, given a subset $S \subseteq V$ of grounded vertices, let $G_S$ denote the graph $G$ augmented with a single ground vertex $g$ joined to every $i \in S$, and pin $\xi_g = 0$. Eliminating the pinned coordinate, the energy of $G_S$ becomes the *grounded quadratic energy*

$$E_S(\xi) := \xi^T L_S \xi = \sum_{(i,j)\in E} (\xi_i - \xi_j)^2 + \sum_{i\in S} \xi_i^2, \qquad L_S := L + \mathrm{Diag}(\chi_S),$$

where $\chi_S \in \{0,1\}^{|V|}$ is the indicator vector of $S$. Grounding alters the diagonal only, so $L_S$ still has nonpositive off-diagonal entries and is still diagonally dominant, and hence is again both submodular and positive semidefinite. The ungrounded case is recovered by taking $S = \emptyset$, for which $L_\emptyset = L$ and $E_\emptyset = E$. Grounding does, however, change the null space: if $G$ is connected and $S \neq \emptyset$, then $L_S$ is positive definite, whereas $L$ is always singular along the constant vector $\xi = te$. This distinction drives Example 5 below.

Because $L_S$ is both submodular and positive semidefinite for every $S \subseteq V$, it provides a natural setting in which to apply the results of Sections 3 and 4. We present two examples: the first concerns distributionally robust subquantile bounds on the energy of a grounded graph (illustrating Section 3), and the second concerns minimum expected energy of an ungrounded graph under marginal moment information (illustrating Section 4). Moment bounds on the Laplacian energy of graphs have also been studied in the signal processing community; see Wang et al. [2022].

**Example 5.** *Given a distribution $\mathbb{P}$ for $\tilde{\xi}$ and a parameter $\alpha \in [0,1)$, the expectation in the lower $(1-\alpha)$-tail distribution (or $(1-\alpha)$-subquantile) of the quadratic energy is given by the*

*optimal value (see Rockafellar and Uryasev [2000])*

$$\sup_x \left( x + \frac{1}{1-\alpha} \cdot \mathbb{E}_{\mathbb{P}} \left[ \min \left( 0, E_S(\tilde{\xi}) - x \right) \right] \right). \tag{24}$$

*Intuitively, the* $(1-\alpha)$*-subquantile captures the expected energy in the lower tail of the distribution. When* $\alpha = 0$*, the optimal value is the expected value* $\mathbb{E}_{\mathbb{P}}[E_S(\tilde{\xi})]$ *and when* $\alpha \uparrow 1$*, the optimal value converges to the minimum value of* $E_S(\xi)$ *over all* $\xi$ *in the support.*

*Now assume the distribution* $\mathbb{P}$ *lies in the ambiguity set*

$$\mathcal{R} = \left\{ \mathbb{P} \in \mathcal{P}([0,1]^n) \ : \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu, \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] = \Sigma \right\},$$

*where the support of* $\tilde{\xi}$ *is contained in* $[0,1]^n$*, the mean is fixed at* $\mu$*, and the second moment matrix is fixed at* $\Sigma$*. Then the worst-case* $(1-\alpha)$*-subquantile is given by the optimal value of*

$$\sup_x \inf_{\mathbb{P} \in \mathcal{R}} \left( x + \frac{1}{1-\alpha} \cdot \mathbb{E}_{\mathbb{P}} \left[ \min \left( 0, E_S(\tilde{\xi}) - x \right) \right] \right). \tag{25}$$

*Solving this DRO problem with ambiguity set* $\mathcal{R}$ *is computationally intractable. However, we can use* $\mathcal{P}$ *defined in (5) and* $\mathcal{Q}$ *defined in (10) as tractable approximations of* $\mathcal{R}$*. These yield independently computed lower bounds on the worst-case* $(1-\alpha)$*-subquantile.*

*We compare the two bounds on the path graph* $1 - 2 - \cdots - 50$ *with* $n = 50$ *vertices, grounded at its two endpoints, that is, with* $S = \{1, 50\}$*. In the spring and resistor network interpretation, this anchors the two ends of the path. We set* $\mu$ *and* $\Sigma$ *to the first two moments of the independent uniform random vector on* $[0,1]^n$*. To apply the framework of Section 3, we rewrite* (25) *equivalently as*

$$-\inf_x \sup_{\mathbb{P} \in \mathcal{R}} \left( -x + \frac{1}{1-\alpha} \cdot \mathbb{E}_{\mathbb{P}} \left[ \max \left( 0, x - E_S(\tilde{\xi}) \right) \right] \right)$$

*and all conditions of Section 3 apply, because the grounded Laplacian* $L_S$ *is both submodular and positive semidefinite. Hence, we use the SDPs in* (9) *and* (11) *for the relaxed ambiguity sets* $\mathcal{P}$ *and* $\mathcal{Q}$*, respectively, to obtain the corresponding lower bounds on* (25)*. The results for different values of* $\alpha$ *are shown in Figure 3. For* $\alpha \leq 0.2$*, the bound from* $\mathcal{P}$ *is stronger; for* $\alpha > 0.2$*, it is weaker. Thus, while both ambiguity sets provide tractable relaxations, neither subsumes the other.*

*One could also combine* $\mathcal{P}$ *and* $\mathcal{Q}$ *to develop tighter bounds for* $\mathcal{R}$*. Such an approach using infimal convolution has been adopted in the DRO literature [Wiesemann et al., 2014], but we leave a detailed analysis for future research.*

Our next example illustrates the results of Section 4 by comparing the full SDP bound to

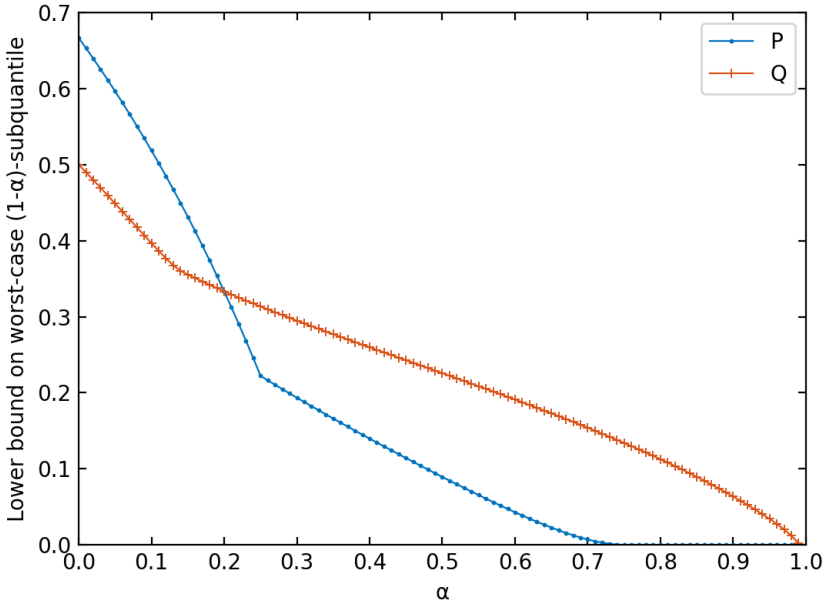


Figure 3: Lower bounds on the worst-case $(1-\alpha)$-subquantile for ambiguity sets $\mathcal{Q}$ and $\mathcal{P}$.

the simpler sum of bivariate bounds, showing how the gap between them varies with graph type and size.

**Example 6.** *Given the mean $\mu$ and standard deviation $\sigma$ of the random vector $\tilde{\xi}$, we compute the minimum expected energy $e^*$ over all distributions supported on the unit hypercube:*

$$e^* := \min \left\{ \mathbb{E}_{\mathbb{P}}[E(\tilde{\xi})] \ : \ \mathbb{P} \in \mathcal{P}([0,1]^n), \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu, \mathbb{E}_{\mathbb{P}}[\mathrm{diag}(\tilde{\xi}\tilde{\xi}^T)] = \mathrm{diag}(\mu\mu^T + \sigma\sigma^T) \right\}.$$

*Here we take $S = \emptyset$, so that the energy is the ungrounded $E(\xi) = \xi^T L \xi$ and decomposes as a pure sum over edges. The results of Section 4 can be used to calculate $e^*$ via the problem (13) with $A = -L$ where $L$ is the Laplacian of the graph. For comparison, since the Laplacian energy decomposes as a sum over edges, a lower bound $\underline{e}^*$ of $e^*$ can be obtained by independently bounding each edge term $(\tilde{\xi}_i - \tilde{\xi}_j)^2$ using the bivariate result of Proposition 10 and summing.*

*In the following numerical experiments, we consider three types of graphs—the path graph, the star graph, and the complete graph, each with $n = 2, 10, 20,$ and $50$ vertices. We generated 1000 random instances for each type of graph and each value of $n$, where the mean vector $\mu$ and the standard deviation vector $\sigma$ are randomly generated to satisfy the feasibility conditions in Lemma 3.*

*In Table 5, we report the percentage gap between the $e^*$ and its lower bound $\underline{e}^*$, i.e.,*

$(e^* - \underline{e}^*)/e^* \times 100\%$. *The results show that, for the star and complete graphs, the gap grows as the size of the graph grows, while for the path graph it remains roughly constant near* 0.6%. *Moreover, at every size the average gap is highest for the complete graph, followed by the star graph and then the path graph.*

| | Path | Star | Complete |
|---|---|---|---|
| $n = 2$ | 0 (0) | 0 (0) | 0 (0) |
| $n = 10$ | 0.577 (0.872) | 1.253 (2.007) | 1.765 (1.052) |
| $n = 20$ | 0.628 (0.603) | 1.501 (1.892) | 2.094 (0.746) |
| $n = 50$ | 0.609 (0.345) | 1.776 (2.046) | 2.329 (0.455) |

Table 5: Mean (standard deviation) of the percentage gap $\frac{(e^* - \underline{e}^*)}{e^*} \times 100\%$ computed over 1000 instances.

# 6 Additional Proofs

In this section, we present proofs of Propositions 3, 4, and 5, which are used to establish Theorem 1.

## 6.1 Proof of Proposition 3

To prove Proposition 3, we need two lemmas. We start by showing that a rank-2 feasible solution $Y(x, X)$, which satisfies an additional (nonlinear) inequality condition, is completely positive with respect to $\mathcal{J}$.

**Lemma 4.** *Let $(x, X)$ be feasible for (2) with $r(x, X) = 2$ and $X \geq xx^T$. Then $Y(x, X) \in \mathcal{CP}(\mathcal{J})$.*

*Proof.* Denote $Y := Y(x, X)$. It is well-known that $\operatorname{rank}(Y) = \operatorname{rank}(X - xx^T) + 1$, which implies $\operatorname{rank}(X - xx^T) = 1$. Combining $X - xx^T \succeq 0$ and $X - xx^T \geq 0$, we see $X - xx^T = ww^T$ where $w = \sqrt{\operatorname{diag}(X) - x \circ x} \geq 0$. (Indeed, a rank-1 PSD matrix can be written as $ss^T$; entrywise nonnegativity then implies all entries of $s$ have the same sign, and we take $w = |s|$.) Then

$$Y = \begin{pmatrix} 1 & x^T \\ x & xx^T + ww^T \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ x & w \end{pmatrix} \begin{pmatrix} 1 & 0 \\ x & w \end{pmatrix}^T.$$

We wish to apply an orthogonal rotation of the form

$$\begin{pmatrix} \alpha & \beta \\ u & v \end{pmatrix} := \begin{pmatrix} 1 & 0 \\ x & w \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \qquad \theta \in (0, \pi/2), \qquad \begin{pmatrix} \alpha \\ u \end{pmatrix}, \begin{pmatrix} \beta \\ v \end{pmatrix} \in \mathcal{J}. \tag{26}$$

Note that $\alpha = \cos\theta > 0$ and $\beta = \sin\theta > 0$ for $\theta \in (0, \pi/2)$. This rotation would establish that $Y \in \mathcal{CP}(\mathcal{J})$ because this construction ensures

$$Y = \begin{pmatrix} \alpha & \beta \\ u & v \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ u & v \end{pmatrix}^T.$$

It thus remains to show that system (26) is feasible by demonstrating $0 \leq u \leq \alpha\, e$ and $0 \leq v \leq \beta\, e$, which are equivalent to

$$\begin{aligned} 0 \leq \alpha x - \beta w \leq \alpha e \quad &\Leftrightarrow \quad 0 \leq x - (\beta/\alpha)w \leq e \\ 0 \leq \beta x + \alpha w \leq \beta e \quad &\Leftrightarrow \quad 0 \leq x + (\alpha/\beta)w \leq e. \end{aligned}$$

Because $0 \leq x \leq e$ and $\beta/\alpha > 0$, two of these four vector inequalities necessarily hold; the remaining ones to verify are

$$0 \leq x - (\beta/\alpha)w \qquad \text{and} \qquad x + (\alpha/\beta)w \leq e.$$

Whenever $w_i = 0$, both vector inequalities hold. Furthermore, when $x_i \in \{0, 1\}$, we must have $x_i^2 = X_{ii} = x_i$ and hence $w_i = \sqrt{X_{ii} - x_i^2} = 0$. So we may assume $w > 0$ and $0 < x < e$ without loss of generality in which case the two inequalities read

$$(e - x)^{-1} \circ w \leq \frac{\beta}{\alpha} \leq x \circ w^{-1},$$

where inverses are componentwise and are well-defined because $0 < x < e$ and $w > 0$, and where $\beta/\alpha = \tan\theta$. Define

$$L := \max_i\, (1 - x_i)^{-1} w_i, \qquad U := \min_i\, x_i w_i^{-1}.$$

Then it suffices to show $L \leq U$, i.e., $(1 - x_i)^{-1} w_i \leq x_j w_j^{-1}$ for all $(i, j)$, which is true because

$$w_i w_j = X_{ij} - x_i x_j \leq x_j - x_i x_j = (1 - x_i) x_j,$$

where $X_{ij} \leq x_j$ follows from symmetry and feasibility: $X_{ij} = X_{ji} \leq x_j$ since $X \leq x e^T$. Hence $L \leq U$, and we can choose $\tan\theta \in [L, U]$, which yields a feasible $\theta \in (0, \pi/2)$. □

Our next lemma considers what happens when the nonlinear condition $X \geq xx^T$ of the previous lemma is not satisfied.

**Lemma 5.** *Let $(x, X)$ be feasible for (2) with $r(x, X) = 2$ and $X \not\geq xx^T$. Then there exists*

*another feasible* $(\hat{x}, \hat{X})$ *with no worse objective value such that one of the following holds:*

*(i)* $r(\hat{x}, \hat{X}) = 3$ *and* $a(\hat{x}, \hat{X}) > a(x, X)$*;*

*(ii)* $r(\hat{x}, \hat{X}) = 2$ *and* $\hat{X} \geq \hat{x}\hat{x}^T$*.*

*Proof.* Our proof strategy is to construct a line segment $\{Y(\lambda) : \lambda \in [0, 1]\}$, which starts at $Y(0) := Y := Y(x, X)$ and ends at another rank-2, positive semidefinite matrix $Y(1)$. In addition, $Y(\lambda)$ is feasible for (2) for sufficiently small $\lambda > 0$ and has non-increasing objective value along the segment. Furthermore, $\text{rank}(Y(\lambda)) = 3$ for all $\lambda \in (0, 1)$. Then we define $\lambda^*$ to be the supremum of $\lambda \in [0, 1]$ such that $Y(\lambda)$ is feasible and show that $Y(\lambda^*)$ satisfies either (i) or (ii).

To construct the line segment, we first identify the symmetric set of indices where $X_{ij} < x_i x_j$, which is nonempty by assumption:

$$\mathcal{S} := \{(i, j) : X_{ij} < x_i x_j\} \neq \emptyset.$$

Also,

$$X_{ij} < x_i x_j \leq \min\{x_i, x_j\} \quad \forall (i, j) \in \mathcal{S},$$

so the corresponding linear inequalities $X_{ij} \leq \min\{x_i, x_j\}$ are inactive on $\mathcal{S}$.

As in the proof of Lemma 4, since $\text{rank}(Y) = 2$, we have $\text{rank}(X - xx^T) = 1$ and $X - xx^T \succeq 0$, so $X - xx^T = ww^T$ for some vector $w$ with $|w| = \sqrt{\text{diag}(X) - x \circ x}$. Then

$$Y = \begin{pmatrix} 1 & x^T \\ x & xx^T + ww^T \end{pmatrix},$$

and since $X \not\geq xx^T$, the vector $w$ must have mixed signs.

For $\lambda \in [0, 1]$, define the desired line segment via the convex combination

$$Y(\lambda) := (1 - \lambda)Y + \lambda Y(1), \qquad \text{where} \qquad Y(1) := \begin{pmatrix} 1 & x^T \\ x & xx^T + |w||w|^T \end{pmatrix}.$$

These relationships can be equivalently expressed via the definitions

$$\Delta := |w||w|^T - ww^T, \qquad x(\lambda) := x, \qquad X(\lambda) := X + \lambda\Delta, \qquad Y(\lambda) := Y(x(\lambda), X(\lambda)).$$

Note that $Y(1)$ is positive semidefinite and rank-2 and satisfies $X(1) \geq x(1)x(1)^T$. Note also that the $x(\lambda)$ is fixed along the segment, while $X(\lambda)$ changes. In particular, let $P := \{i : w_i > 0\}$ and $N := \{i : w_i < 0\}$. Since $X_{ij} = x_i x_j + w_i w_j$, we have $X_{ij} < x_i x_j$ if and only if

$w_i w_j < 0$, i.e., $(i,j) \in P \times N \cup N \times P$. Hence $\mathcal{S} = P \times N \cup N \times P$, and in particular no diagonal pair belongs to $\mathcal{S}$. Moreover,

$$\Delta_{ij} > 0 \text{ for } (i,j) \in \mathcal{S}, \qquad \Delta_{ij} = 0 \text{ for } (i,j) \notin \mathcal{S}.$$

This ensures that $X(\lambda)_{ij}$ is strictly increasing in $\lambda$ for $(i,j) \in \mathcal{S}$, while all other entries of $X(\lambda)$ remain fixed.

We investigate the feasibility of $Y(\lambda)$. Because $Y(\lambda)$ is a convex combination of $Y$ and $Y(1)$, which are both positive semidefinite, $Y(\lambda)$ is positive semidefinite for all $\lambda \in [0,1]$. Furthermore, because $w$ has mixed signs, $w$ and $|w|$ are linearly independent, so $(1-\lambda)ww^T + \lambda|w||w|^T$ has rank 2 for every $\lambda \in (0,1)$. By the standard Schur complement rank identity, it follows that

$$\operatorname{rank}(Y(\lambda)) = 1 + \operatorname{rank}\left((1-\lambda)ww^T + \lambda|w||w|^T\right) = 3 \qquad \forall \lambda \in (0,1).$$

With regards to the linear inequalities defining the feasible region, the sign pattern of $\Delta$ discussed in the prior paragraph as well as the fact that $X_{ij} < \min\{x_i, x_j\}$ for all $(i,j) \in \mathcal{S}$ ensures $X(\lambda) \leq x(\lambda)e^T = xe^T$ for sufficiently small $\lambda > 0$. We conclude that, for sufficiently small $\lambda > 0$, $Y(\lambda)$ is feasible.

Next we consider the change in objective value along the line segment. Since $x(\lambda) = x$ is constant, the change in objective value is determined by

$$Q \bullet X(\lambda) = Q \bullet X + \lambda\, Q \bullet \Delta,$$

and therefore the sign pattern of $\Delta \geq 0$ discussed above, which is supported only on off-diagonal indices in $\mathcal{S}$, implies $Q \bullet \Delta \leq 0$ by submodularity. So the objective value is non-increasing along the segment.

Now define

$$\lambda^* := \sup\{\lambda \in [0,1] : X(\lambda) \leq xe^T\},$$

and set $(\hat{x}, \hat{X}) := (x, X(\lambda^*))$. If $\lambda^* < 1$, then at least one inequality indexed by $\mathcal{S}$ is active at $(\hat{x}, \hat{X})$, so $a(\hat{x}, \hat{X}) > a(x, X)$, and since $\lambda^* \in (0,1)$ we have $r(\hat{x}, \hat{X}) = 3$; this is case (i). If $\lambda^* = 1$, then $X(1) \geq xx^T$ and $\operatorname{rank}(Y(1)) = 2$ as noted above; this is case (ii). □

By combining the above two lemmas, we can now prove Proposition 3.

*Proof of Proposition 3.* If $X \geq xx^T$, then Lemma 4 implies that $Y(x, X)$ is completely positive with respect to $\mathcal{J}$, and hence by Lemma 2, the relaxation is tight. On the other hand, if $X \not\geq xx^T$, Lemma 5 implies that there exists an alternative optimal solution $(\hat{x}, \hat{X})$ with

either $r(\hat{x}, \hat{X}) = 2$ and $\hat{X} \geq \hat{x}\hat{x}^T$ or $r(\hat{x}, \hat{X}) = 3$ and more active inequalities. In the former case, Lemma 4 again implies that the relaxation is tight.

In the latter case, consider the minimal face defined by $(\hat{x}, \hat{X})$, which necessarily excludes $(x, X)$ because there are more active inequalities. Within that face, consider an extreme point with rank at most 2. If such a point exists, repeat the same argument at that new rank-2 point: either the SDP relaxation is tight, or we obtain another rank-3 optimal point with more active inequalities. Since the number of inequalities is finite, this process terminates, yielding a proper subface of the optimal face in which all extreme points have rank at least 3. □

## 6.2 Proof of Proposition 4

We start with a key lemma showing that $Y(x, X)$ is completely positive with respect to $\mathcal{J}$ whenever all s-upper inequalities are active and at most one diagonal inequality is inactive. We remark that the first statement in the lemma does not assume $Y(x, X) \succeq 0$.

**Lemma 6.** *Let $(x, X)$ be sorted by $x$ with all s-upper inequalities active. If all diagonal inequalities are active, then $Y(x, X) \in \mathcal{CP}(\mathcal{J})$. The same conclusion holds if $(x, X)$ is feasible for (2) and all diagonal inequalities are active except for a single $X_{jj} < x_j$.*

*Proof.* Let $Y := Y(x, X)$. We have $X_{ij} = \min\{x_i, x_j\}$ for all index pairs except possibly for a single $X_{jj} < x_j$. We prove the result by induction on $n$.

For the base case $n = 1$, if $X_{11} = x_1$, we verify that

$$Y = \begin{pmatrix} 1 & x_1 \\ x_1 & x_1 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}^T + (1 - x_1) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^T \in \mathcal{CP}(\mathcal{J}).$$

If $X_{11} < x_1$ and $x_1^2 = X_{11}$, then $Y$ is rank-1 and hence in $\mathcal{CP}(\mathcal{J})$. If $X_{11} < x_1$ and $x_1^2 < X_{11}$, then $Y \succ 0$, and choosing $\epsilon > 0$ maximally such that $Y_\epsilon := Y - \epsilon\, ee^T \succeq 0$ with $e := (1, 1)^T$ yields $Y_\epsilon = uu^T$ for some vector $u$. The linear inequalities in $\mathcal{L}$ are preserved as $Y$ shifts to $Y_\epsilon$ because $ee^T$ satisfies all inequalities in $\mathcal{L}$ with equality. Hence, $Y_\epsilon \in \mathcal{L}$, and thus $u \in \mathcal{J}$, which in turn implies $Y_\epsilon = uu^T + \epsilon\, ee^T \in \mathcal{CP}(\mathcal{J})$.

Now consider $n > 1$. If $x_1 = 1$, then $x = e$ and $Y = \binom{1}{e}\binom{1}{e}^T \in \mathcal{CP}(\mathcal{J})$. Suppose $x_1 < 1$ and either all diagonal inequalities are active or $j \neq 1$, where $j$ is the index of the inactive diagonal. In both situations, $X_{11} = x_1$, and so the second column of $Y$ equals $x_1 e$. Define $\tilde{x}_i := (x_i - x_1)/(1 - x_1)$ for $i = 2, \ldots, n$. Then

$$Y = x_1\, ee^T + (1 - x_1)\, \tilde{Y},$$

where $\tilde{Y}$ has zero second row and column. After deleting them, $\tilde{Y}$ satisfies the conditions of the lemma with respect to $\tilde{x}$ (note $0 \le \tilde{x}_2 \le \cdots \le \tilde{x}_n \le 1$): when all diagonals are active, $\tilde{Y}$ again has all diagonals active; when at most one diagonal is inactive, the inactive diagonal persists in $\tilde{Y}$; and $\tilde{Y} \succeq 0$ because $Y - x_1\, ee^T$ is the Schur complement of $X_{11} = x_1$ in $Y \succeq 0$. By the induction hypothesis, $\tilde{Y} \in \mathcal{CP}(\mathcal{J})$, and since $ee^T \in \mathcal{CP}(\mathcal{J})$, the result follows.

Finally, suppose $x_1 < 1$ and there exists an inactive diagonal inequality at $j = 1$. Note $x_1 > 0$, since feasibility gives $X_{11} \ge x_1^2$ and $X_{11} < x_1$ forces $x_1 > 0$. If $x_n = 1$, the first and last rows of $Y$ coincide because $X_{nn} = x_n = 1$ and $X_{in} = x_i$ for all $i$. Then the result follows by induction on the reduced matrix. If $x_n < 1$, the difference of the first and last columns of $Y$ is $(1 - x_n)\, v$ with $v := (1, 0, \ldots, 0)^T$, so $v$ is in the range space of $Y$. Choosing $\epsilon > 0$ maximally so that $Y - \epsilon\, vv^T \succeq 0$ reduces the rank while changing only the top-left entry, preserving all linear inequalities in $\mathcal{L}$. After rescaling by the top-left entry, the resulting matrix retains the same structure, so the process can be repeated until the rank reaches 1—giving $Y \in \mathcal{CP}(\mathcal{J})$ by construction—or $x_n = 1$ after rescaling, which is handled above. □

We are now ready to prove Proposition 4.

*Proof of Proposition 4.* Given optimal $(x, X)$ with $\operatorname{diag}(X) = x$, assume $(x, X)$ is sorted by $x$ without loss of generality. Define $(\hat{x}, \hat{X})$ by $\hat{x} := x$ and $\hat{X}_{ij} := \min\{x_i, x_j\}$ for all $i, j$, so that all s-upper and diagonal inequalities are active for $(\hat{x}, \hat{X})$. Then $Y(\hat{x}, \hat{X})$ is completely positive with respect to $\mathcal{J}$ by Lemma 6 and hence feasible for (2). Moreover, since $(\hat{x}, \hat{X})$ differs from $(x, X)$ only in the off-diagonal entries of $\hat{X}$ such that $\operatorname{offdiag}(X) \le \operatorname{offdiag}(\hat{X})$, the submodularity of $Q$ implies $(\hat{x}, \hat{X})$ is optimal for (2) as well. This proves the proposition by Lemma 2. The second case follows directly from Lemma 6.

□

## 6.3 Proof of Proposition 5

The proof of Proposition 5 relies on the following lemma, which shows that if an s-lower inequality is active, then the corresponding columns of $Y(x, X)$ are identical.

**Lemma 7.** *Let $(x, X)$ be feasible for (2). After sorting by $x$, suppose that, for $i < j$, the s-lower inequality $X_{ij} \le x_j$ is active. Then the columns of $Y(x, X)$ indexed by $i$ and $j$ are identical.*

*Proof.* Since $X_{ij} = x_j$ and $X_{ij} \le x_i \le x_j$, we have $x_i = x_j = X_{ij}$. The $3 \times 3$ principal

submatrix of $Y := Y(x, X)$ indexed by $0, i, j$ becomes

$$\begin{pmatrix} 1 & x_i & x_i \\ x_i & X_{ii} & x_i \\ x_i & x_i & X_{jj} \end{pmatrix} \succeq 0.$$

Positive semidefiniteness gives $X_{ii}X_{jj} \geq x_i^2$, while the diagonal inequalities give $X_{ii} \leq x_i$ and $X_{jj} \leq x_i$. Together, $X_{ii} = X_{jj} = x_i$. Now let $v := e_i - e_j \in \mathbb{R}^{n+1}$, where $e_i$ and $e_j$ are standard basis vectors in the index space of $Y$. Then

$$v^T Y v = X_{ii} - 2X_{ij} + X_{jj} = x_i - 2x_i + x_i = 0.$$

Since $Y \succeq 0$, this implies $Yv = 0$, which gives the result. □

*Proof of Proposition 5.* By assumption, after sorting by $x$, there exists $i < j$ such that the s-lower inequality $X_{ij} \leq x_j$ is active. By Lemma 7, the columns of $Y(x, X)$ indexed by $i$ and $j$ are identical.

Define the $(n-1)$-dimensional data $(\hat{Q}, \hat{c})$ by merging indices $i$ and $j$. In particular, set $\hat{Q}_{ii} := Q_{ii} + Q_{jj} + 2Q_{ij}$, $\hat{Q}_{ik} := Q_{ik} + Q_{jk}$ for $k \neq i, j$, and other entries of $\hat{Q}$ unchanged. Also set $\hat{c}_i := c_i + c_j$, while keeping other entries of $\hat{c}$ unchanged. Then delete index $j$ throughout. Since the off-diagonal entries of $Q$ are nonpositive, $(\hat{Q}, \hat{c})$ is submodular. Let $(\hat{x}, \hat{X})$ denote the reduced pair obtained from $(x, X)$ by deleting index $j$. Then $(\hat{x}, \hat{X})$ is feasible for the $(n-1)$-dimensional SDP relaxation with data $(\hat{Q}, \hat{c})$. In particular, $Y(\hat{x}, \hat{X})$ is a principal submatrix of $Y(x, X) \succeq 0$, and $\hat{X} \leq \hat{x}e^T$ is inherited from $X \leq xe^T$, where abusing notation $e$ is the vector of all ones of the appropriate size in each case.

Let $\rho_n$ and $\rho_{n-1}$ denote the optimal SDP values in dimensions $n$ and $n-1$ for data $(Q, c)$ and $(\hat{Q}, \hat{c})$, respectively. Also let $v_n$ and $v_{n-1}$ denote the corresponding optimal values of the underlying nonconvex QPs. The duplicate structure ensures $\hat{Q} \bullet \hat{X} + \hat{c}^T\hat{x} = Q \bullet X + c^T x = \rho_n$, so $\rho_{n-1} \leq \rho_n$. Since the reduced QP is equivalent to minimizing $x^T Q x + c^T x$ over $[0, 1]^n$ with $x_i = x_j$, we have $v_{n-1} \geq v_n$. By the induction hypothesis, $\rho_{n-1} = v_{n-1}$. Therefore,

$$\rho_n \leq v_n \leq v_{n-1} = \rho_{n-1} \leq \rho_n,$$

and all inequalities are equalities. In particular, $\rho_n = v_n$, so the SDP relaxation is tight. □

## 6.4 Verification of the counterexample

This section verifies Example 4 for the instance (3). Throughout, all arithmetic is exact and carried out over the rational numbers, except where a floating-point solver is used.

**Proposition 12.** *For the data $(Q, c)$ of (3), problem (1) has optimal value 0, attained uniquely at $x = 0$, while the relaxation (2) has optimal value at most $-109/1024$. In particular, the relaxation gap is positive.*

*Proof.* Every global minimizer of (1) is a KKT point, and every KKT point has an active-set pattern assigning each variable to $\{0\}$, $\{1\}$, or the open interval $(0, 1)$. Enumerating all $3^4 = 81$ patterns and solving each stationarity system over the rationals is therefore exhaustive. Four of the 81 reduced Hessians are singular, but in each of those cases the stationarity system is inconsistent, so no face carries a degenerate set of stationary points.[4] The minimum over all KKT points equals 0 and is attained only at $x = 0$.

For the relaxation it suffices to exhibit a single feasible point of negative objective value. Take

$$\bar{x} = \begin{pmatrix} 3/32 \\ 3/16 \\ 9/16 \\ 3/4 \end{pmatrix}, \qquad \bar{X} = \begin{pmatrix} 3/32 & 3/32 & 3/32 & 61/1024 \\ 3/32 & 125/1024 & 3/16 & 3/16 \\ 3/32 & 3/16 & 231/512 & 9/16 \\ 61/1024 & 3/16 & 9/16 & 3/4 \end{pmatrix}.$$

All sixteen inequalities $\bar{X}_{ij} \le \bar{x}_i$ hold, and the leading principal minors of $Y(\bar{x}, \bar{X})$ are all strictly positive. Hence $Y(\bar{x}, \bar{X}) \succ 0$ by Sylvester's criterion, and $(\bar{x}, \bar{X})$ is feasible for (2). Its objective value is exactly $c^T\bar{x} + Q \bullet \bar{X} = -109/1024$. □

Solving (2) numerically with Mosek [MOSEK ApS, 2024] gives an optimal value of approximately $-0.1220566$, and the computed optimal solution satisfies $\operatorname{rank}(Y(x, X)) = 3$. Its vector $x$ is already increasing, so no sorting is required, and exactly seven of the sixteen inequalities $X_{ij} \le x_i$ are active: the first and last diagonal inequalities, and five of the six s-upper inequalities, the pair $(1, 4)$ being the only slack one. In the notation of Section 2,

$$l = 0, \qquad d = 2, \qquad u = 5, \qquad a = l + d + u = 7,$$

and the counting bound of Lemma 1 is comfortably satisfied, since $r(r+1) = 12 \le 2(a+1) = 16$. This configuration lies outside the hypotheses of every proposition used in the proof of Theorem 1, and it is precisely the rank-3 configuration with partial activity anticipated in Remark 1.

[4] Had such a set existed, it would have been harmless: on an affine set of stationary points $f$ is constant, since a null direction $v$ of the reduced Hessian satisfies both $\nabla f \cdot v = 0$ and $v^T Q v = 0$.

It is natural to ask whether the gap can be closed by strengthening (2). For $1 \le i < j < k \le n$ the triangle inequalities

$$\begin{aligned} X_{ij} + X_{ik} &\le x_i + X_{jk}, & X_{ij} + X_{jk} &\le x_j + X_{ik}, \\ X_{ik} + X_{jk} &\le x_k + X_{ij}, & x_i + x_j + x_k &\le X_{ij} + X_{ik} + X_{jk} + 1 \end{aligned} \tag{27}$$

are valid for the Boolean quadric polytope [Padberg, 1989], and hence valid on $\mathcal{CP}(\mathcal{J})$ as well; see Burer and Letchford [2009], extending Yajima and Fujie [1998]. Table 6 reports the optimal value of (2) strengthened successively by the full RLT inequalities and by all sixteen triangle inequalities available at $n = 4$.

| Relaxation | Optimal value |
|---|---|
| PSD + RLT upper bounds, i.e. (2) | $-0.1220566$ |
| PSD + full RLT | $-0.1220566$ |
| PSD + full RLT + triangle inequalities | $-0.0001432$ |

Table 6: Optimal values of successively stronger relaxations for the instance (3). The true optimal value of (1) is 0. The triangle inequalities remove more than 99.9% of the gap but do not close it.

We see that the gap is not an artifact of having discarded the RLT lower bounds in (2); the point $(\bar{x}, \bar{X})$ of Proposition 12 satisfies the lower bounds $X_{ij} \ge 0$ and $X_{ij} \ge x_i + x_j - 1$ as well, so the full SDP-RLT relaxation of Proposition 2 attains the same value. In addition, the triangle inequalities remove most, but not all, of the gap. This surviving gap deserves an exact certificate. Mixing the computed optimum with the strictly feasible point $x = e/2$, $X = \frac{1}{4}ee^T + \frac{1}{8}I$ and rounding to denominator $10^8$ yields a rational point satisfying, exactly, the semidefiniteness condition, all sixteen RLT inequalities in both directions, and all sixteen triangle inequalities, with objective value

$$-\frac{14113}{100{,}000{,}000} \;=\; -1.4113 \times 10^{-4} \;<\; 0.$$

Since the optimal value of (1) is 0 by Proposition 12, the relaxation strengthened by the full RLT and triangle inequalities retains a positive gap on this instance.

Figure 4 depicts this instance in a two-dimensional projection that makes both the size of the gap and the effect of the triangle inequalities visible at once. Rather than fixing marginals as in Figure 1, it projects the full bodies onto the two affine functionals

$$t(x, X) \;=\; c^T x + Q \bullet X \qquad \text{and} \qquad g(x, X) \;=\; x_3 + X_{14} - X_{13} - X_{34},$$

the second of which is the slack of one of the two violated triangle inequalities. Both are affine, so the projection is linear and the inclusion $\mathcal{CP}(\mathcal{J}) \subseteq \mathcal{L}$ is preserved in the image. The horizontal dimension is then self-interpreting: the leftmost point of each shadow is the optimal value of the corresponding problem, so the horizontal offset between the two tips is exactly the relaxation gap. Vertically, the exact hull lies entirely above the line $g = 0$, since the triangle inequality is valid on $\mathcal{CP}(\mathcal{J})$, whereas the relaxation protrudes below it.

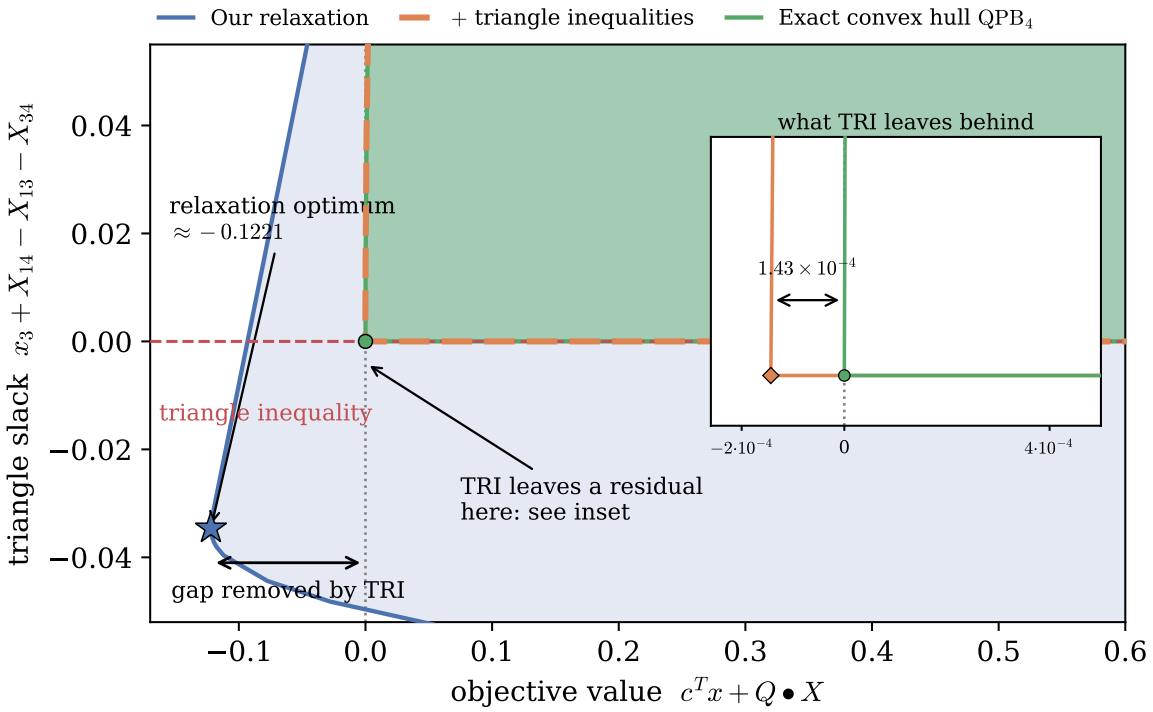


Figure 4: The instance (3) projected onto (objective value, triangle-inequality slack). Adding the triangle inequalities cuts the body at $g = 0$ and removes more than 99.9% of the gap, but not all of it. Inset (magnified roughly 1000×): the strengthened relaxation still reaches $-1.43 \times 10^{-4}$, whereas the exact hull stops at 0.

# 7 Conclusions

We now return to the complexity landscape summarized in Table 1 of the Introduction, characterizing the complexity of minimizing a general quadratic function $f(x) := x^T Q x + c^T x$ over the sets $\{0,1\}^n$ and $[0,1]^n$. In particular, the polynomial-time solvable submodular case over $\{0,1\}^n$ is well-known as discussed in the Introduction, while the complexity over $[0,1]^n$ has not been fully established. Two special cases where polynomial time solvability is known over $[0,1]^n$ is when additional restrictions are imposed on the signs of the diagonal entries of $Q$

or the entries of $c$. This paper adds to this landscape by establishing new tractability for the submodular case over $[0,1]^n$ when $n \leq 3$ by removing these additional restrictions. Example 4 delimits that contribution precisely. The relaxation (2) is not tight for $n \geq 4$, and the gap it exhibits is closed neither by the RLT lower bounds nor by the triangle inequalities. What fails at $n \geq 4$ is therefore this particular semidefinite certificate, not necessarily tractability itself. The complexity of minimizing a submodular quadratic over $[0,1]^n$ for $n \geq 4$ remains open, and settling it will require either a different certificate or a hardness reduction. The experiments of Section 5.1 indicate that instances exhibiting a gap are highly non-generic, so the relaxation continues to perform well in practice even though it is not tight in general.

Finally, it is natural to ask whether sparsity in $Q$ can be exploited to simplify the relaxation (2). Since the entry $X_{ij}$ enters the objective only through $Q_{ij}$, the RLT upper bound $X_{ij} \leq \min\{x_i, x_j\}$ is intuitively unnecessary whenever $Q_{ij} = 0$. When $Q$ is block separable, this holds exactly: the problem decomposes across blocks, and only the within-block bounds are needed. For general sparse $Q$, our experiments indicate that dropping the bounds associated with zero entries of $Q$ preserves the optimal value in the large majority of instances, though not always. A precise characterization of when such bounds are redundant—and hence of the smallest sufficient set of RLT inequalities—is a promising direction for future work.


## Acknowledgements

We thank Shaoning Han for pointing out an error in an earlier version of this paper.


# References


K. M. Anstreicher and S. Burer. Computable representations for convex hulls of low-dimensional quadratic forms. *Mathematical Programming (series B)*, 124:33–43, 2010.

B. Axelrod, Y. P. Liu, and A. Sidford. Near-optimal approximate discrete and continuous submodular function minimization. In *Proceedings of the Fourteenth Annual SIAM Symposium on Discrete Algorithms (SODA)*, pages 837–853, 2020.

F. Bach. Submodular functions: from discrete to continuous domains. *Mathematical Programming*, 175(1):419–459, 2019.

D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005.

A. Bian, K. Levy, A. Krause, and J. M. Buhmann. Continuous dr-submodular maximization: Structure and algorithms. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, volume 30, pages 486–496, 2017.

J. Bilmes. Submodularity in machine learning and artificial intelligence, 2022. URL `https://arxiv.org/abs/2202.00132`.

J. Bunton and P. Tabuada. Joint continuous and discrete model selection via submodularity. *Journal of Machine Learning Research*, 23(329):1–42, 2022.

S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120:479–495, 2009.

S. Burer. *Copositive Programming*, pages 201–218. Springer US, New York, NY, 2012. ISBN 978-1-4614-0769-0. doi: 10.1007/978-1-4614-0769-0_8. URL `https://doi.org/10.1007/978-1-4614-0769-0_8`.

S. Burer and A. N. Letchford. On nonconvex quadratic programming with box constraints. *SIAM Journal on Optimization*, 20(2):1073–1089, 2009. doi: 10.1137/080729529.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

P. Dickinson. *The Copositive Cone, the Completely Positive Cone and their Generalisations*. PhD thesis, University of Groningen, 2013.

G. Gallego and H. Topaloglu. *Revenue management and pricing analytics*, volume 279 of *International Series in Operations Research and Management Science*. Springer, 2019.

M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.

Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*, 2024. URL `https://www.gurobi.com`.

N. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002.

R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to global optimization*, volume 48 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, second edition, 2000.

O. Hössjer and A. Sjölander. Sharp lower and upper bounds for the covariance of bounded random variables. *Statistics and Probability Letters*, 182, 2022.

S. Ito and R. Fujimaki. Large-scale price optimization via network flow. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL `https://proceedings.neurips.cc/paper_files/paper/2016/file/cb79f8fa58b91d3af6c9c991f63962d3-Paper.pdf`.

S. Karlin and K. Studden. *Tchebycheff systems: With Applications in Analysis and Statistics*. Interscience Publisher, 1966.

S. Kim and M. Kojima. Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. *Computational Optimization and Applications*, 26(2): 143–154, 2003.

M. K. Kozlov, S. P. Tarasov, and L. G. Khachiyan. The polynomial solvability of convex quadratic programming. *USSR Computational Mathematics and Mathematical Physics*, 20(5):223–228, 1980.

A. Krause, H. B. McMahan, C. Guestrin, and A. Gupta. Robust submodular observation selection. *Journal of Machine Learning Research*, 9(93):2761–2801, 2008.

D. Kuhn, S. Shafiee, and W. Wiesemann. Distributionally robust optimization. *Acta Numerica*, 2024.

M. Laurent. Sums of squares, moment matrices and optimization over polynomials. *Emerging applications of algebraic geometry*, pages 157–270, 2009.

L. Lovász. Submodular functions and convexity. *Mathematical Programming The State of the Art: Bonn 1982*, pages 235–257, 1983.

G. P. McCormick. Computability of global solutions to factorable nonconvex programs. I. Convex underestimating problems. *Math. Programming*, 10(2):147–175, 1976.

MOSEK ApS. *MOSEK Optimization Toolbox*, 2024. URL `https://www.mosek.com`.

J. Neyman. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4):585–625, 1934.

M. Padberg. The boolean quadric polytope: Some characteristics, facets and relatives. *Mathematical Programming*, 45:139–172, 1989.

G. Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Math. Oper. Res.*, 23:339–358, 1998.

I. Pitowsky. Correlation polytopes: Their geometry and complexity. *Mathematical Programming*, 50:395–414, 1991.

Y. Qiu and E. A. Yıldırım. On exact and inexact RLT and SDP-RLT relaxations of quadratic programs with box constraints. *Journal of Global Optimization*, 90(2):293–322, 2024.

R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

K. Schmudgen. *The Moment Problem*. Graduate Texts in Mathematics (GTM, volume 277). Springer, 2017.

D. Spielman. Algorithms, graph theory, and linear equations in laplacian matrices. In *Proceedings of the International Congress of Mathematicians*, 2010.

K. S. Srikantan. A problem in optimum allocation. *Operations Research*, 11(2):265–273, 1963. doi: 10.1287/opre.11.2.265.

M. Staib and S. Jegelka. Robust budget allocation via continuous submodular functions. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 3230–3240, 2017.

X. Wang, Y.-M. Pun, and A. M.-C. So. Distributionally robust graph learning from smooth signals under moment uncertainty. *IEEE Transactions on Signal Processing*, 70:6216–6231, 2022.

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

Y. Yajima and T. Fujie. A polyhedral approach for nonconvex quadratic programming problems with box constraints. *Journal of Global Optimization*, 13(2):151–170, 1998. doi: 10.1023/A:1008230200679.

Q. Yu and S. Küçükyavuz. On constrained mixed-integer dr-submodular minimization. *Mathematics of Operations Research*, 50(2):871–909, 2025.

M. Zhang, L. Chen, H. Hassani, and A. Karbasi. Online continuous submodular maximization: From full-information to bandit feedback. In *33rd Conference on Neural Information Processing Systems*, volume 32, 2019.

S. Zhang. Quadratic maximization and semidefinite relaxation. *Mathematical Programming*, 87(3):453–465, 2000. doi: 10.1007/s101070050006.