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

Samuel Burer* Karthik Natarajan†

April 4, 2025

Abstract

We show that continuous quadratic submodular minimization with bounds is solvable in polynomial time using semidefinite programming, and we apply this result to 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

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

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

*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

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 to maximization, as noted by Yu and Küçükyavuz [2024], 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)*:

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

where Q is a symmetric matrix, c is a column vector, and κ is a constant. The constant is included for convenience later in the paper, and we will often refer to (1) by its triple (Q, c, κ). 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 triple (Q, c, κ) 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.

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, κ) 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], for example. 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, κ) 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] present a probabilistic algorithm for solving a class of CSM over the box that includes QSMB. Their algorithm is linear in n and polynomial in the ratio L/ε, where L is the submodular function's Lipschitz parameter and ε is the additive error in the final optimal value. In particular, the polynomial dependence on L/ε makes their algorithm pseudo-polynomial.

1.1 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 establish that QSMB can in fact be solved in polynomial-time in the following sense: there exists a polynomial-time semidefinite-programming relaxation of (1), whose optimal value equals that of (1). The relaxation is said to be *tight* or to admit *no gap*. Inspired by Kim and Kojima [2003], the relaxation that we employ is a well-known semidefinite relaxation of (1). This result is shown in Section 2 and makes use of a technical lemma presented in Section 6, which proves an important property of the optimal solution set of the semidefinite relaxation. Three examples in multi-product pricing and robust optimization are also discussed in Section 2.

(b) We then apply the semidefinite programming reformulation of QSMB to solve 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 an ambiguity set, which is new in the literature. This ambiguity set models random vectors supported in $[0,1]^n$, where the mean and the second moment for each random variable is fixed and also given are lower bounds on the pairwise cross (mixed) moments of degree two. Our main result for QSMB enables us to solve this class of DRO problems efficiently using SDP. This is closely related to—but different than—the SDP formulation of DRO problems for a popular ambiguity set proposed by Delage and Ye [2010], where the mean is fixed and the cross moments are bounded from above in the positive semidefinite order.

(c) In the second model discussed in Section 4, we compute the tightest upper bound on a nonnegative 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 show that this problem can also be solved efficiently with SDP.

(d) In the numerical results in Section 5, we compare the SDP formulations of Sections 3 and 4 with existing approaches, and in particular, we investigate formulations that compute moment bounds on the energy function defined by Laplacian matrices of graphs. The results illustrate that the bounds from our approaches can indeed be tighter than existing approaches.

We conclude in Section 7 with a discussion on the computational complexity of quadratic minimization both over the box and the Boolean hypercube under assumptions of convexity and submodularity.

Overall, this paper extends polynomial-time results for convex minimization to quadratic submodular minimization over the box (QSMB) via an exact SDP relaxation and examines various applications of this result. As such, this paper lays the groundwork for further study of submodular minimization and indeed can have important ramifications in areas such as robust and distributionally robust optimization.

1.2 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. 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 $\text{diag}(A)$, and the diagonal matrix formed with the entries of the vector a is given by $\text{Diag}(a)$. Given $A, B \in \mathbb{R}^{m \times n}$, we let $A \bullet B = \text{trace}(A^T B)$. A random vector is denoted by $\tilde{\xi}$, and we use $\mathbb{E}_P[\cdot]$ to denote the expectation with respect to $\mathbb{P}$, $\text{Var}[\tilde{\xi}_i]$ to denote the variance of $\tilde{\xi}_i$, and $\text{Cov}[\tilde{\xi}_i, \tilde{\xi}_j]$ to denote the covariance of $\tilde{\xi}_i$ and $\tilde{\xi}_j$.

2 A Tight Semidefinite Relaxation

In this section, we establish our main theoretical result, namely that there exists a tight semidefinite relaxation of (1) that is solvable in polynomial-time. We first review the relevant literature that motivates our approach.

2.1 Existing results

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

Theorem 1 (Theorem 3.1 in Kim and Kojima [2003])**.** *Suppose (Q, c, κ) is submodular with $c \le 0$. Then the optimal value of (1) equals the optimal value of its SDP relaxation*

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

In fact, the authors show that, if (x^*, X^*) is an optimal solution of the SDP, then the vector with entries given by the square root of the diagonal entries of X^* is an optimal solution of (1). Further, 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 2×2 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 $c_j > 0$ for some j such that the SDP relaxation admits a gap.

For the case when (Q, c, κ) is submodular and c is arbitrary, we can utilize a tighter SDP relaxation to attempt to close the gap. A typical approach, which we adopt here, is to include valid inequalities derived from the feasibility condition $x \in [0, 1]^n$. Indeed, one of the most common classes of valid inequalities are the so-called *RLT (reformulation linearization technique) constraints* (or *McCormick inequalities*), which are derived as follows for each pair $1 \le i \le j \le n$:

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

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 the first 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 $\operatorname{diag}(X) \le x$, a strengthening of the constraint $\operatorname{diag}(X) \le e$ in Theorem 1.

2.2 Our result

We show that the following relaxation, which adds only the RLT upper bounds to the relaxation in Theorem 1, is tight for every submodular (Q, c, κ) irrespective of the signs of the entries of c:

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

Our proof is by induction on n with base case $n = 1$.

Lemma 1. *For $n = 1$, let data (Q, c, κ) be given. Then (2) is a tight relaxation of (1).*

Proof. Proof For $n = 1$, problems (1) and (2) respectively read

$$\begin{aligned} v^* &:= \min \left\{ Q_{11} x_1^2 + c_1 x_1 + \kappa : x_1 \in [0, 1] \right\}, \\ r^* &:= \min \left\{ Q_{11} X_{11} + c_1 x_1 + \kappa : x_1^2 \le X_{11} \le x_1 \right\}, \end{aligned}$$

where $r^* \le v^*$ by construction. To show $r^* \ge v^*$, we analyze two subcases. First assume $Q_{11} \ge 0$. The sign of Q_{11}, as well as the fact that X_{11} is only bounded below by x_1^2, ensures that without loss of generality we have $X_{11} = x_1^2$ at optimality, i.e., there exists an optimal (x^*, X^*) of (2) yielding a feasible solution of (1) with the same objective value. Hence, $r^* \ge v^*$. For the second subcase, assume $Q_{11} < 0$. Then, following similar logic, we have $X_{11} = x_1$ at optimality, and so on the optimal solution set, the SDP objective reduces to the linear function $(Q_{11} + c_1)x_1 + \kappa$. Since x_1 ranges over $[0, 1]$, it follows that an optimal solution of (2) occurs at $x_1 \in \{0, 1\}$ such that further $X_{11} = x_1 = x_1^2$ at optimality. As in the previous subcase, we thus have a feasible solution of (1) with the same objective value. Hence $r^* \ge v^*$, as desired. □

Next we prove the induction step. We need a technical result, Lemma 3, whose proof we defer to Section 6.

Proposition 1. *For $n \ge 2$, suppose (2) is a tight relaxation of (1) for all submodular $(\hat{Q}, \hat{c}, \hat{\kappa}) \in \mathbb{S}^{n-1} \times \mathbb{R}^{n-1} \times \mathbb{R}$. Then (2) is a tight relaxation of (1) for all submodular $(Q, c, \kappa) \in \mathbb{S}^n \times \mathbb{R}^n \times \mathbb{R}$.*

Proof. Proof Let submodular $(Q, c, \kappa) \in \mathbb{S}^n \times \mathbb{R}^n \times \mathbb{R}$ be arbitrary. By Lemma 3, either relaxation (2) is tight—in which case we are done—or there exists optimal (x^*, X^*) of (2) such that $x_j^* \in \{0, 1\}$ for some j. Without loss of generality, suppose $x_n^* \in \{0, 1\}$. To complete the proof, we consider two cases, $x_n^* = 0$ and $x_n^* = 1$.

For the first case, $x_n^* = 0$ and the constraint $x_n^2 \le X_{nn} \le x_n$ imply $X_{nn}^* = 0$. The positive semidefiniteness constraint then ensures that the full $(n+1) \times (n+1)$ matrix

$$\begin{pmatrix} 1 & (x^*)^T \\ x^* & X^* \end{pmatrix}$$

has the block form

$$\begin{pmatrix} 1 & (\hat{x}^*)^T & 0 \\ \hat{x}^* & \hat{X}^* & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\hat{x}^* \in \mathbb{R}^{n-1}$ is the leading subvector of x^* and $\hat{X}^* \in \mathbb{S}^{n-1}$ is the leading principal submatrix of X^*. Optimality of (2) then ensures that $(\hat{x}^*, \hat{X}^*)$ is an optimal solution of the corresponding relaxation (2) of (1) for the data $(\hat{Q}, \hat{c}, \hat{\kappa}) \in \mathbb{S}^{n-1} \times \mathbb{R}^{n-1} \times \mathbb{R}$ in base dimension $n-1$, where $\hat{Q}$ is the leading principal submatrix of Q, $\hat{c}$ is the leading subvector of c, and $\hat{\kappa} := \kappa$. It follows from the induction hypothesis that there exists $\hat{x} \in [0,1]^{n-1}$ such that $x := \binom{\hat{x}}{0} \in [0,1]^n$ satisfies

$$\begin{aligned} x^T Q x + c^T x + \kappa &= \hat{x}^T \hat{Q} \hat{x} + \hat{c}^T \hat{x} + \kappa \\ &= \hat{Q} \bullet \hat{X}^* + \hat{c}^T \hat{x}^* + \kappa \\ &= Q \bullet X^* + c^T x^* + \kappa. \end{aligned}$$

In words, with respect to the base dimension n, the optimal value of (2) equals a feasible value of (1), which guarantees that (2) is tight.

For the second case, $x_n^* = 1$ and the constraint $x_n^2 \le X_{nn} \le x_n$ imply $X_{nn}^* = 1$, which ensures

$$\begin{pmatrix} 1 & (x^*)^T \\ x^* & X^* \end{pmatrix}$$

has the form

$$\begin{pmatrix} 1 & (\hat{x}^*)^T & 1 \\ \hat{x}^* & \hat{X}^* & y^* \\ 1 & (y^*)^T & 1 \end{pmatrix},$$

where we have introduced the variables $y_j := X_{nj}$ for $j = 1, \ldots, n-1$ to simplify notation. Also let q represent the vector formed with the off-diagonal entries of Q that correspond to the vector y. Considering x^* fixed, the linear constraints on y simplify to $y_j \le \min\{x_j^*, 1\} = x_j^*$. Because $q \le 0$ by the assumption of submodularity, we have without loss of generality that $y_j^* = x_j^*$, as long as positive semidefiniteness is preserved. It is indeed preserved because $y_j^* = x_j^*$ for all $j = 1, \ldots, n-1$ then implies that the last row of the matrix above is just a copy of the first. Optimality of (2) then ensures that $(\hat{x}^*, \hat{X}^*)$ is an optimal solution of the corresponding relaxation (2) of (1) for the submodular triple

$$(\hat{Q}, \hat{c} + 2\,q, \kappa + Q_{nn} + c_n) \in \mathbb{S}^{n-1} \times \mathbb{R}^{n-1} \times \mathbb{R},$$

where $\hat{Q}$ and $\hat{c}$ are defined as in the prior paragraph. It follows from the induction hypothesis that there exists $\hat{x} \in [0,1]^{n-1}$ such that $x := \binom{\hat{x}}{1} \in [0,1]^n$ satisfies

$$\begin{aligned} x^T Q x + c^T x + \kappa &= \hat{x}^T \hat{Q} \hat{x} + (\hat{c} + 2q)^T \hat{x} + \kappa + Q_{nn} + c_n \\ &= \hat{Q} \bullet \hat{X}^* + (\hat{c} + 2q)^T \hat{x}^* + \kappa + Q_{nn} + c_n \\ &= Q \bullet X^* + c^T x^* + \kappa. \end{aligned}$$

As in the prior case, this guarantees that (2) is tight. □

We thus have arrived at our main theorem.

Theorem 2. *For all n and all submodular $(Q, c, \kappa) \in \mathbb{S}^n \times \mathbb{R}^n \times \mathbb{R}$, problem (2) is a tight relaxation of (1).*

Given that (2) is tight, it follows that at least one optimal (x, X) solution of (2) is rank-1, i.e., $X = xx^T$, such that one can recover an optimal solution x of (1). However, in general, algorithms for solving the relaxation may return alternative optimal solutions $(\hat{x}, \hat{X})$, where $\text{rank}(\hat{X}) > 1$ and $\hat{x}$ is not optimal for (1). In such cases, an exploration of the SDP optimal face—a so-called rank-reduction procedure—would be necessary to recover an optimal x for (1). We remark that Theorem 2 does not establish that the SDP optimal face is the convex hull of points of the form (x, xx^T), where x is optimal for (1). This means that in general one cannot perform a successful rank-reduction by simply finding an extreme point of the optimal face. In any case, we leave the construction of a rank-reduction procedure for future research; indeed, Sections 3 and 4 only require Theorem 2, i.e., equality between the optimal values of (1) and (2).

2.3 Three examples

We end this section with three examples where the result of Theorem 2 is immediately applicable.

Example 1 (Model for multi-product pricing with substitutes and linear demand)**.** *Consider n substitute products with the price decision vector $p \in \mathbb{R}^n$ and linear demand model given by $d(p) = a - Bp$ 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 revenue maximization problem is formulated as the quadratic optimization problem

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

Note that, without loss of generality, B is symmetric by replacing B by $(B + B^T)/2$. It is natural to assume the off-diagonal entries of B are nonpositive since $B_{ij} \leq 0$ for $i \neq j$ implies the demand of product i increases as the price of product j increases. While the diagonal entries of B are often assumed to be nonnegative to capture the effect of the demand of a product decreasing as its price increases, it is known that for certain goods, such as luxury goods, this does not hold. Indeed, allowing for general values on the diagonal of B, one can address a larger class of multi-product pricing problems. Using Theorem 2, the multi-product pricing problem is solvable with SDP.

Example 2 (Robust counterpart with quadratic uncertainty)**.** *Consider the following linear constraint in x:*

$$\xi^T A(x)\xi + b^T(x)\xi + c(x) \leq t$$

where $A(x)$, $b(x)$, and $c(x)$ are assumed to depend affinely on the decision x. The robust counterpart of this constraint when ξ is uncertain and varies in $[l, u]$ is given by:

$$\xi^T A(x)\xi + b^T(x)\xi + c \leq t \quad \forall\, \xi \in [l, u].$$

Unfortunately, this robust counterpart over the box is known to be intractable in general; see section 1.4 in Ben-Tal et al. [2009]. However, Theorem 2 implies that when the matrix $-A(x)$ is submodular for all x, then the robust counterpart is a SDP constraint. We extend this to distributionally robust optimization in Section 3.

Example 3 (Quadratic adjustable robust counterpart)**.** *Consider an uncertain linear inequality*

$$a(\xi)^T x + b^T y(\xi) + c(\xi) \leq t \quad \forall\, \xi \in [l, u].$$

where x is a non-adjustable decision vector and y is an adjustable decision vector that depends on the uncertainty ξ. Further assume that $a(\xi)$ and $c(\xi)$ depend affinely on the uncertainty ξ while b is fixed, which corresponds to fixed recourse. Modeling such constraints naturally arise in adjustable robust optimization. If we allow $y(\xi)$ to depend affinely on the uncertainty, this leads to an affinely adjustable robust counterpart of the uncertain linear inequality, which is known to be computationally tractable. If we allow $y(\xi)$ to depend quadratically on ξ, this leads to a quadratically adjustable robust counterpart of the uncertain linear inequality, *which is known to be computationally intractable; see section 14.3.2 in Ben-Tal et al. [2009]. However, if we restrict the quadratic decision rule, we can still obtain tractable SDP reformulations. For simplicity of exposition, we focus on a single adjustable decision variable* $y(\xi) \in \mathbb{R}$ *and* $b \in \mathbb{R}$ *with the result generalizing in a straightforward manner to multiple adjustable decision variables. Let* $y(\xi) = \xi^T Y \xi + y^T \xi + y_0$ *where the adjustable decision variables are transformed to* $Y \in \mathbb{S}^n$, $y \in \mathbb{R}^n$, $y_0 \in \mathbb{R}$. *We can rewrite the quadratic adjustable robust counterpart as:*

$$\max \left\{ a(\xi)^T x + b(\xi^T Y \xi + y^T \xi + y_0) + c(\xi) \ : \ \xi \in [l, u] \right\} \leq t.$$

If we restrict the quadratic decision rule such that the off-diagonal entries of Y *are of the same sign as* b*, then the left-hand side reduces to a submodular minimization, and we can obtain a exact semidefinite representation of the quadratic adjustable robust counterpart using Theorem 2.*

3 Distributionally Robust Optimization

Consider the distributionally robust optimization problem

$$\inf_{x \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[f(x, \tilde{\xi}) := \max_{k=1,\ldots,K} \left(\tilde{\xi}^T A_k(x) \tilde{\xi} + b_k^T(x) \tilde{\xi} + c_k(x) \right) \right], \tag{3}$$

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 ξ and is piecewise quadratic but not necessarily convex in ξ for a fixed x.

Solving (3) 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 is built on Theorem 2 and guarantees computational tractability of (3).

3.1 A new moment ambiguity set

Let $\mathcal{P}([0,1]^n)$ denote the set of all probability distributions with support contained in $[0,1]^n$. Given fixed $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{S}^n$, define

$$\mathcal{P} := \left\{ \mathbb{P} \in \mathcal{P}([0,1]^n) \ : \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu, \ \mathbb{E}_{\mathbb{P}}[\operatorname{diag}(\tilde{\xi}\tilde{\xi}^T)] = \operatorname{diag}(\Sigma), \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] \geq \Sigma \right\}. \tag{4}$$

In this ambiguity set, the support of the random variable $\tilde{\xi}_i$ is contained in the interval $[0,1]$, with mean fixed at $\mu_i = \mathbb{E}[\tilde{\xi}_i]$ and second moment fixed at $\Sigma_{ii} = \mathbb{E}[\tilde{\xi}_i^2]$ for each $i = 1, \ldots, n$. In addition, the pairwise cross moments of degree two are bounded from below term by term with $\mathbb{E}[\tilde{\xi}_i\tilde{\xi}_j] \geq \Sigma_{ij}$ for all $i < j$. Note that Σ is not required to be positive semidefinite; see Section 3.3 below for more discussion.

We would like to identify tractable necessary and sufficient conditions on μ and Σ 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:

$$\mathcal{M} := \left\{ (\lambda_0, \lambda, \Lambda) \ : \ \exists \, m \in \mathcal{M}([0,1]^n) \ \text{ s.t. } \begin{array}{l} \lambda_0 = \int dm(\xi), \\ \lambda = \int \xi dm(\xi), \\ \Lambda \leq \int \xi\xi^T dm(\xi), \\ \operatorname{diag}(\Lambda) = \int \operatorname{diag}(\xi\xi^T) dm(\xi) \end{array} \right\}. \tag{5}$$

We will now argue that $\mathcal{M}$ is semidefinite representable. 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:

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

and

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

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 $\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2$. Hence, $\mathcal{M}$ is clearly a full-dimensional, convex, pointed cone. That $\mathcal{M}$ is closed follows from corollary 1.12 in Dickinson [2013]. The dual cone of $\mathcal{M}$ is given by

$$\begin{aligned}\mathcal{M}^* &= (\mathcal{M}_1 + \mathcal{M}_2)^* = \mathcal{M}_1^* \cap \mathcal{M}_2^* = \mathcal{K}_1 \cap \mathcal{K}_2 \\ &= \left\{(y_0, y, Y) \ : \ y_0 + y^T\xi + \xi^T Y \xi \geq 0 \ \ \forall\, \xi \in [0,1]^n, \ \ Y \text{ submodular}\right\} \\ &=: \mathcal{K}.\end{aligned} \tag{6}$$

It follows that $\mathcal{M} = \mathcal{K}^*$ since $\mathcal{M}$ is a closed convex cone.

This establishes the following proposition.

Proposition 2. *The cones $\mathcal{M}$ and $\mathcal{K}$ 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 representations of $\mathcal{M}$ and $\mathcal{K}$ using submodularity and Theorem 2 of Section 2.

Proposition 3. *$\mathcal{K}$ is semidefinite representable with*

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

and hence

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

Proof. Proof Recall the definition of $\mathcal{K}$ in (6). The semi-infinite constraint therein is equivalent to

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

which in particular implies $y_0 \geq 0$. Since Y is submodular in $\mathcal{K}$, by Theorem 2, we can express

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

Note that strong duality holds, and the optimal solutions are attained for both the primal and dual; see lemma 18 in Qiu and Yıldırım [2024]. Then $\nu \geq 0$ as long as at least one dual solution has nonnegative objective value. Hence,

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

By setting $z_0 = y_0$ without loss of generality, we have the result. The expression for $\mathcal{M}$ follows by standard dual constructions. Note that the auxiliary square matrix $Z \in \mathbb{R}^{n \times n}$ used in the semidefinite representation of $\mathcal{K}$ is not symmetric, while W in the representation of $\mathcal{M}$ is symmetric. □

We summarize the foregoing discussion by concluding that the ambiguity set $\mathcal{P}$ in (4) is nonempty if and only if there exists a matrix W in $\mathbb{S}^n$ such that

$$\Sigma \leq W \leq \mu e^T, \quad \operatorname{diag}(W) = \operatorname{diag}(\Sigma), \quad \begin{pmatrix} 1 & \mu^T \\ \mu & W \end{pmatrix} \succeq 0. \tag{7}$$

3.2 Semidefinite representation

This brings us to the following theorem, which shows that (3) can be solved in polynomial-time as a semidefinite program when the ambiguity set $\mathcal{P}$ is given by (4) .

Proposition 4. *Suppose:*

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

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

(c) $(1, \mu, \Sigma) \in int(\mathcal{M})$ where $int(\mathcal{M})$ is the interior of the moment cone $\mathcal{M}$.

Then the distributionally robust problem (3) *is solvable in polynomial time as the SDP*

$$\begin{array}{lll} \min & y_0 + \mu^T y + \Sigma \bullet Y & \\ \text{s.t.} & x \in \mathcal{X} & \\ & Y \ \textit{submodular} & \\ & Z_k \geq 0 & \forall\ k = 1, \ldots, K \\ & \begin{pmatrix} y_0 & (y^T - e^T Z_k)/2 \\ (y - Z_k^T e)/2 & Y + (Z_k + Z_k^T)/2 \end{pmatrix} \succeq \begin{pmatrix} c_k(x) & b_k^T(x)/2 \\ b_k(x)/2 & A_k(x) \end{pmatrix} & \forall\ k = 1, \ldots, K, \end{array} \tag{8}$$

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* $k = 1, \ldots, K$*.*

Proof. Proof For fixed $x \in \mathcal{X}$, denote the value of the inner supremum in (3) by

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

This problem is strictly feasible under the assumption that $(1, \mu, \Sigma)$ lies in the interior of $\mathcal{M}$, and its dual is

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

Since the dual is strictly feasible, e.g., consider $y = 0$, $Y = -ee^T$ and

$$y_0 > \max \left\{ \xi^T (A_k(x) + ee^T) \xi + b_k^T(x) \xi + c_k(x) : \xi \in [0,1]^n, k = 1, \ldots, K \right\},$$

strong duality holds with $v^*(x) = v_d^*(x)$, and both the primal and dual values are attained for each $x \in \mathcal{X}$. Disaggregating the dual constraints across $k = 1, \ldots, K$ gives

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

Using the semidefinite representation of the cone $\mathcal{K}$ from Proposition 3 and optimizing over $x \in \mathcal{X}$, we get the desired result. □

3.3 Comparison with an existing moment ambiguity set

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

$$\mathcal{Q} = \left\{ \mathbb{P} \in \mathcal{P}(\mathcal{C}) \ : \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu, \ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}\tilde{\xi}^T] \preceq \Sigma \right\}. \tag{9}$$

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

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

It is interesting to compare these conditions with the analogous ones (7) for our ambiguity set $\mathcal{P}$. These two sets express different phenomena, e.g., the matrix Σ in (9) needs to be positive semidefinite for $\mathcal{Q}$ to be nonempty but not for $\mathcal{P}$ to be nonempty. Furthermore, when $\mathcal{C} = [0,1]^n$ and the matrix $A_k(x)$ is negative semidefinite for each k, the distributionally robust optimization (3)—where $\mathcal{P}$ is replaced by $\mathcal{Q}$—can be reformulated as an SDP using duality (see Delage and Ye [2010]):

$$\begin{aligned}
\min \quad & y_0 + \mu^T y + \Sigma \bullet Y \\
\text{s.t.} \quad & x \in \mathcal{X} \\
& Y \succeq 0 \\
& z_k, w_k \geq 0 & \forall\ k = 1, \ldots, K \\
& \begin{pmatrix} y_0 - e^T z_k & (y + z_k - w_k)^T/2 \\ (y + z_k - w_k)/2 & Y \end{pmatrix} \succeq \begin{pmatrix} c_k(x) & b_k^T(x)/2 \\ b_k(x)/2 & A_k(x) \end{pmatrix} & \forall\ k = 1, \ldots, K,
\end{aligned} \tag{10}$$

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 (8) and (10) in Section 5.

4 Covariance Bounds

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 covariance of the random variables is given by the Cauchy-Schwarz inequality:

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

or equivalently

$$\mathbb{E}[\tilde{\xi}_1\tilde{\xi}_2] \leq \mathbb{E}[\tilde{\xi}_1]\mathbb{E}[\tilde{\xi}_2] + \sqrt{\text{Var}[\tilde{\xi}_1]\text{Var}[\tilde{\xi}]}. \tag{11}$$

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

Suppose each random variable $\tilde{\xi}_i$ has support contained in $[0, 1]$ with known mean μ_i and standard deviation σ_i, or equivalently the variance is σ_i^2. Let (μ, σ) 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

$$v^* = \max \left\{ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}^T A\tilde{\xi}] \;:\; \begin{array}{c} \mathbb{P} \in \mathcal{P}([0,1]^n),\ \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = \mu, \\ \mathbb{E}_{\mathbb{P}}[\text{diag}(\tilde{\xi}\tilde{\xi}^T)] = \text{diag}(\mu\mu^T + \sigma\sigma^T) \end{array} \right\}, \tag{12}$$

i.e., v^* is the tightest 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 feasibility of (12).

Lemma 2. *The moment problem (12) is feasible if and only if* $0 \leq \sigma_i \leq \sqrt{\mu_i(1-\mu_i)}$ *for all* $i = 1, \ldots, n$.

Proof. 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 $\sigma_i^2 \leq \mu_i(1-\mu_i)$ 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 moment problem is efficiently solvable using Theorem 2.

Proposition 5. *Suppose:*

(a) $0 \le \sigma_i^2 \le \mu_i(1-\mu_i)$ *for all* $i = 1, \ldots, n$*;*

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

Then the moment problem (12) *is solvable via the semidefinite program*

$$\begin{aligned} v^* = \max \quad & A \bullet \Sigma \\ \text{s.t.} \quad & \operatorname{diag}(\Sigma) = \operatorname{diag}(\mu\mu^T + \sigma\sigma^T) \\ & \Sigma \le e\mu^T \\ & \begin{pmatrix} 1 & \mu^T \\ \mu & \Sigma \end{pmatrix} \succeq 0. \end{aligned} \tag{13}$$

where the decision variable is $\Sigma \in \mathbb{S}^n$*.*

Proof. Proof Assumption (a) and Lemma 2 imply (12) 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 $z \in \mathbb{R}^n$. Strong duality holds with $v^* = v_d^*$ because the primal is feasible as just established and 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

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

Under assumption (b) and Theorem 2, we have:

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

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 4, we get

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

The primal semidefinite formulation is then given by (13). □

4.1 Closed-form solution in dimension 2

We can also solve problem (13) in closed form when $n = 2$ and

$$A = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{14}$$

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

Proposition 6. *Consider (12) for $n = 2$ and A given by (14). The upper bound is given by*

$$\mathbb{E}_{\mathbb{P}}[\tilde{\xi}_1\tilde{\xi}_2] \leq v^* = \min\left(\mathbb{E}_{\mathbb{P}}[\tilde{\xi}_1], \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_2], \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_1]\mathbb{E}_{\mathbb{P}}[\tilde{\xi}_1] + \sqrt{Var[\tilde{\xi}_1]\, Var[\tilde{\xi}_2]}\right), \tag{15}$$

and this bound is as tight as possible.

Proof. Proof For $n = 2$, the semidefinite program (13) reduces to

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

Using the Schur complement theorem, we can rewrite this as

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

which implies 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 2 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 $\underline{p}_i \leq \overline{p}_i$ for $i = 1, 2$.

Case 1: When the intervals $[\underline{p}_1, \overline{p}_1]$ and $[\underline{p}_2, \overline{p}_2]$ overlap.

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,

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

To verify this bound is tight, assume without loss of generality that $\underline{p}_2 \leq \overline{p}_1$. This is equivalent to $\mu_1\mu_2 + \sigma_1\sigma_2 \leq \mu_1$. It is follows that $\underline{p}_1 \leq \overline{p}_2$ which is equivalent to $\mu_1\mu_2 + \sigma_1\sigma_2 \leq \mu_2$. Hence the bound is tight.

Case 2: When the intervals $[\underline{p}_1, \overline{p}_1]$ and $[\underline{p}_2, \overline{p}_2]$ do not overlap.

Without loss of generality, let $\overline{p}_1 \leq \underline{p}_2$. 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, $\mathbb{E}_{\mathbb{P}^*}[\tilde{\xi}_1\tilde{\xi}_2] = \mu_1$. Here $\mu_1 \leq \mu_2$ and $\mu_1 \leq \mu_1\mu_2 + \sigma_1\sigma_2$. The case $\overline{p}_2 \leq \underline{p}_1$ follows along similar lines. □

The tightness of this upper bound for $n = 2$ has also been recently shown in Hössjer and Sjölander [2022] (see Theorem 2), although their proof technique differs from ours. While the bivariate bound in Proposition 6 can be summed up over all pairs to find an upper bound for the general n case, this need not be tight as we will see in the numerical results in Section 5.

5 Numerical Results

In support of Sections 3 and 4, we next investigate the numerical behavior of our models. In particular, we will consider the test case of random quadratic forms defined by Laplacian matrices of graphs.

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 easily to the case where nonnegative weights are associated with edges. The Laplacian matrix L associated with G is a $|V| \times |V|$ symmetric matrix with: diagonal entries for $i \in V$ given by $L_{ii} = \deg(i)$, where $\deg(i)$ is the degree of vertex i in G; and off-diagonal entries given by $L_{ij} = -1$ when i is adjacent to j and 0 otherwise. The Laplacian matrix is submodular and well-known to be diagonally dominant and hence positive semidefinite.

Associated with the graph G is its associated *quadratic energy*, which is defined using the Laplacian matrix as follows:

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

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

We focus here on moment bounds on the energy function with random $\tilde{\xi}$, and our first example illustrates the results of Section 3.

Example 4. *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])*

$$\sup_x \left(x + \frac{1}{1-\alpha} \cdot \mathbb{E}_{\mathbb{P}} \left[\min \left(0, E(\tilde{\xi}) - x \right) \right] \right). \tag{16}$$

When $\alpha = 0$, the optimal value is the expected value $\mathbb{E}_{\mathbb{P}}[E(\tilde{\xi})]$ and when $\alpha \uparrow 1$, the optimal value converges to the minimum value of $E(\xi)$ over all ξ in the support.

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 the random vector $\tilde{\xi}$ is contained in $[0,1]^n$, the mean fixed is to μ, and the second moment matrix is fixed to Σ. Then the worst-case $(1-\alpha)$-subquantile is given by the optimal value of

$$\sup_x \inf_{\mathbb{P} \in \mathcal{R}} \left(x + \frac{1}{1-\alpha} \cdot \mathbb{E}_{\mathbb{P}} \left[\min \left(0, E(\tilde{\xi}) - x \right) \right] \right). \tag{17}$$

Solving this DRO problem with the ambiguity set $\mathcal{R}$ is computationally intractable. However,

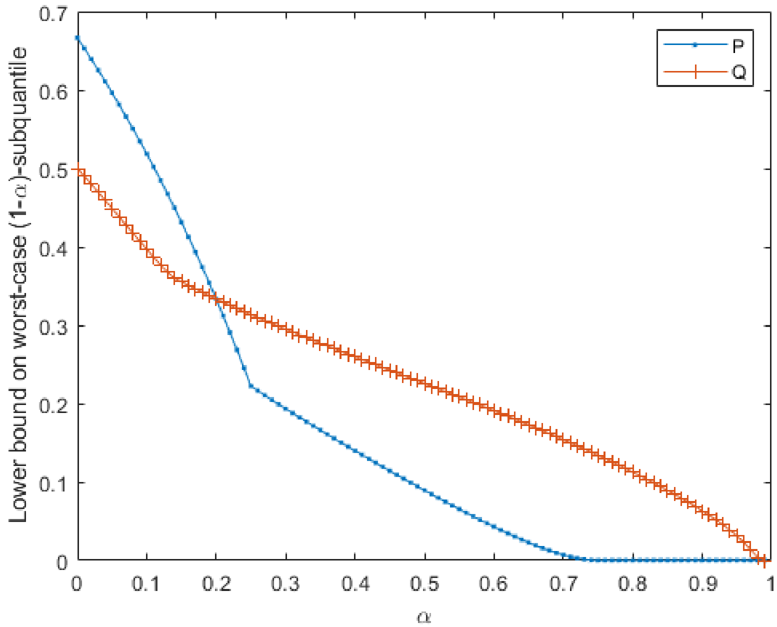

Figure 1: Worst-case $(1-\alpha)$-subquantile for ambiguity sets $\mathcal{Q}$ and $\mathcal{P}$.

we can use $\mathcal{P}$ *defined in (4) as a tractable approximation of* $\mathcal{R}$*. Similarly, we can use* $\mathcal{Q}$ *defined in (9) as a tractable approximation of* $\mathcal{R}$*. These give independently calculated lower bounds on the worst-case* $(1-\alpha)$*-subquantile.*

We compare the two bounds for a path graph $G = (V, E)$ *with* $|V| = 50$*, where* μ *and* Σ *are set to the first two moments of the independent uniform random vector on* $[0, 1]^n$*. It is straightforward to see that formulation* (17) *is equivalent to*

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

and that all the conditions of Section 3 apply, in particular, due to the fact that the Laplacian is both submodular and positive semidefinite. Hence, we can use the SDPs in (8) *and* (10) *to solve this problem for the relaxed ambiguity sets* $\mathcal{P}$ *and* $\mathcal{Q}$*, respectively, to obtain the aforementioned lower bounds on the optimal value of* (17)*, which are depicted for different values of* α *in Figure 1. For* $\alpha \leq 0.2$*, the bound obtained using* $\mathcal{P}$ *is stronger, while for* $\alpha > 0.2$ *it is weaker. This illustrates that, 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 numerical example illustrates the results of Section 4.

Example 5. *Given the mean μ and standard deviation σ of the random vector $\tilde{\xi}$, we compute the minimum expected energy e^* over all distributions supported on the unit hypercube:*

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

The results of Section 4 can be used to calculate e^* via the problem (12) with $A = -L$ where L is the Laplacian of the graph. For comparison, a lower bound $\underline{e}^*$ of e^* can be found by summing up the bivariate bounds for each pair $\tilde{\xi}_i$ and $\tilde{\xi}_j$ using Proposition 6.

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 100 random instances for each type of graph and each value of n, where the mean vector μ and the standard deviation vector σ are randomly generated to satisfy the feasibility conditions in Lemma 2.

In Table 1, 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, as the size of the graph grows, the gap grows as well. Moreover, the average gap is highest for the complete graph followed by the star graph and the path graph.

	Path	Star	Complete
$n = 2$	0 (0)	0 (0)	0 (0)
$n = 10$	0.445 (0.763)	0.870 (1.319)	1.827 (1.079)
$n = 20$	0.489 (0.413)	1.261 (1.578)	2.093 (0.839)
$n = 50$	0.549 (0.267)	1.499 (1.903)	2.294 (0.465)

Table 1: Mean (standard deviation) of the percentage gap $\frac{(e^* - \underline{e}^*)}{e^*} \times 100\%$ computed over 100 instances.

6 Technical Lemma

In this section, we prove the technical lemma used in the proof of Proposition 1. We state a more general setting than strictly required with the hope that this result may find additional use in future research. Immediately after the proof, we verify that the lemma does indeed apply to QSMB.

Lemma 3. *Let $m \leq M$ be positive integers. Given data $(Q_i, c_i, \kappa_i) \in \mathbb{S}^n \times \mathbb{R}^n \times \mathbb{R}$ for all $i = 0, 1, \ldots, m$, consider the nonconvex minimization*

$$v^* := \min\left\{x^T Q_0 x + c_0^T x + \kappa_0 : x \in F\right\}$$

over the feasible set

$$F := \left\{x \in \mathbb{R}^n : x^T Q_i x + c_i^T x + \kappa_i \le 0 \;\; \forall\, i = 1, \ldots, m\right\}.$$

Suppose $Q_i \succeq 0$ for all $i = 1, \ldots, m$ such that F is convex. Suppose also that F is bounded. Given additional data $(Q_i, c_i, \kappa_i) \in \mathbb{S}^n \times \mathbb{R}^n \times \mathbb{R}$ for $i = m+1, \ldots, M$ with Q_i not necessarily positive semidefinite, suppose each additional constraint $x^T Q_i x + c_i^T x + \kappa_i \le 0$ is redundant for F. Suppose moreover that:

(a) there exists $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}^T Q_i x + c_i^T \bar{x} + \kappa_i < 0$ for all $i = 1, \ldots, M$; in particular, int(F) is nonempty;

(b) there exists $\lambda_i > 0$ for $i = 1, \ldots, M$ such that $Q_0 + \sum_{i=1}^M \lambda_i Q_i \succ 0$;

(c) the semidefinite feasible set

$$R := \left\{ (x, X) \in \mathbb{R}^n \times \mathbb{S}^n : \begin{array}{l} Q_i \bullet X + c_i^T x + \kappa_i \le 0 \;\forall\, i = 1, \ldots, M, \\ \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0 \end{array} \right\}$$

is bounded.

Then the SDP relaxation

$$r^* := \min\left\{Q_0 \bullet X + c_0^T x + \kappa_0 : (x, X) \in R\right\} \tag{18}$$

has the following property: $r^* = v^*$, or there exists an optimal solution (x^*, X^*) such that $x^* \in \mathrm{bd}(F)$.

Proof. Proof To prove the result, we express the relaxation (18) in an alternative form. First note that

$$\begin{aligned} (x, X) \in R \quad &\Longrightarrow \quad Q_i \bullet X + c_i^T x + \kappa_i \le 0 \;\; \forall\, i = 1, \ldots, m \\ &\Longleftrightarrow \quad Q_i \bullet (X - xx^T) + x^T Q_i x + c_i^T x + \kappa_i \le 0 \;\; \forall\, i = 1, \ldots, m \\ &\Longrightarrow \quad x^T Q_i x + c_i^T x + \kappa_i \le 0 \;\; \forall\, i = 1, \ldots, m \\ &\Longleftrightarrow \quad x \in F, \end{aligned}$$

where the third implication follows because $Q_i \bullet (X - xx^T) \ge 0$ by the positive semidefiniteness of Q_i and $X - xx^T$. Then, by introducing a new auxiliary variable $Y := X - xx^T$, we see that (18) is equivalent to

$$\min_{x\in F}\left(x^TQ_0x + c_0^Tx + \kappa_0 + \varphi(x)\right),$$

where

$$\varphi(x) := \min_{Y}\left\{Q_0 \bullet Y : \begin{array}{l} Q_i \bullet Y \le -(x^TQ_ix + c_i^Tx + \kappa_i) \ \ \forall\ i = 1,\dots,M, \\ Y \succeq 0 \end{array}\right\}. \tag{19}$$

Note in particular that the value $\varphi(x)$ is attained at some $Y = X - xx^T$ because X is bounded when x is fixed by assumption (c). By strong duality, we have

$$\varphi(x) = \sup\left\{\sum_{i=1}^{M}\lambda_i\left(x^TQ_ix + c_i^Tx + \kappa_i\right) : \begin{array}{l} \lambda_i \ge 0 \ \ \forall\ i = 1,\dots,M, \\ Q_0 + \sum_{i=1}^{M}\lambda_iQ_i \succeq 0 \end{array}\right\},$$

where we have used assumption (b), i.e., that the dual problem here has an interior point. Hence, (18) can be rewritten

$$\min_{x\in F}\sup_{\lambda}\left\{\left(x^TQ_0x + c_0^Tx + \kappa_0\right) + \sum_{i=1}^{M}\lambda_i\left(x^TQ_ix + c_i^Tx + \kappa_i\right) : \begin{array}{l} \lambda_i \ge 0\ \forall\ i = 1,\dots,M, \\ Q_0 + \sum_{i=1}^{M}\lambda_iQ_i \succeq 0 \end{array}\right\}.$$

Introducing auxiliary variables (S, s, σ) as well as the dual feasible set

$$D := \left\{(\lambda, S, s, \sigma) : \begin{array}{l} \lambda_i \ge 0 \ \ \forall\ i = 1,\dots,M, \\ S = Q_0 + \sum_{i=1}^{M}\lambda_iQ_i \succeq 0, \\ s = c_0 + \sum_{i=1}^{M}\lambda_ic_i, \\ \sigma = \kappa_0 + \sum_{i=1}^{M}\lambda_i\kappa_i \end{array}\right\},$$

we can further express (18) as

$$\min_{x\in F}\sup_{(\lambda,S,s,\sigma)\in D}\left(x^TSx + s^Tx + \sigma\right). \tag{20}$$

Now consider the following minimax theorem (see proposition 5.5.7 in Bertsekas [2009]):

> Let F and D by nonempty convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^k$, respectively, and let $\phi(x,z) : F\times D \to \mathbb{R}$ be a function such that $\phi(\cdot,z) : F \to \mathbb{R}$ is convex and closed for each $z \in D$, and $-\phi(x,\cdot) : D \to \mathbb{R}$ is convex and closed for each $x \in F$. Suppose (i) F is compact, and (ii) for some $\bar{x} \in F$ and $\bar{\nu} \in \mathbb{R}$, the level set $\{z \in D : \phi(\bar{x}, z) \ge \bar{\nu}\}$ is nonempty and compact. Then $\inf_{x\in F}\sup_{z\in D}\phi(x,z) =$

$\sup_{z\in D}\inf_{x\in F}\phi(x,z)$, and the set of saddle points is nonempty and compact. In particular, *inf* and *sup* may be replaced by *min* and *max* in the preceding equation.

In our case at hand, $\phi(x,z)$ corresponds to the function $x^T Sx + s^T x + \sigma$, which clearly satisfies the assumptions of the theorem. In addition, F is compact by assumption, so that (i) is satisfied. To see that (ii) is satisfied, let $\bar{x} \in F$ be any interior point, which exists by assumption (a), and take $\bar{\nu} := \bar{x}^T \bar{S}\bar{x} + \bar{s}^T\bar{x} + \bar{\sigma}$ for some arbitrary $(\bar{\lambda}, \bar{S}, \bar{s}, \bar{\sigma}) \in D$. Then consider the (nonempty) level set

$$\left\{ (\lambda, S, s, \sigma) : \begin{array}{l} \lambda_i \geq 0 \ \ \forall\, i = 1, \ldots, M, \\ S = Q_0 + \sum_{i=1}^M \lambda_i Q_i \succeq 0, \\ s = c_0 + \sum_{i=1}^M \lambda_i c_i, \\ \sigma = \kappa_0 + \sum_{i=1}^M \lambda_i \kappa_i, \\ \bar{x}^T S\bar{x} + s^T\bar{x} + \sigma \geq \bar{\nu} \end{array} \right\},$$

or—after eliminating the auxiliary variables (S, s, σ)—the equivalent level set

$$\left\{ \lambda : \begin{array}{l} \lambda_i \geq 0 \ \ \forall\, i = 1, \ldots, M, \\ Q_0 + \sum_{i=1}^M \lambda_i Q_i \succeq 0, \\ (\bar{x}^T Q_0 \bar{x} + c_0^T \bar{x} + \kappa_0) + \sum_{i=1}^M \lambda_i \left(\bar{x}^T Q_i \bar{x} + c_i^T \bar{x} + \kappa_i \right) \geq \bar{\nu} \end{array} \right\}.$$

This level set is compact if and only if its recession cone

$$\left\{ \Delta\lambda : \begin{array}{l} \Delta\lambda_i \geq 0 \ \ \forall\, i = 1, \ldots, M, \\ \sum_{i=1}^M \Delta\lambda_i Q_i \succeq 0, \\ \sum_{i=1}^M \Delta\lambda_i \left(\bar{x}^T Q_i \bar{x} + c_i^T \bar{x} + \kappa_i \right) \geq 0 \end{array} \right\}$$

is trivial. Since every value $\bar{x}^T Q_i \bar{x} + c_i^T \bar{x} + \kappa_i$ is negative by assumption (a), it follows that $\Delta\lambda = 0$, i.e., the recession cone is indeed trivial. So the minimax theorem implies (20) and hence also (18) are equivalent to

$$\max_{(\lambda,S,s,\sigma)\in D} \min_{x\in F} \left(x^T Sx + s^T x + \sigma \right).$$

Moreover, the set of saddle points is nonempty and compact.

We are now ready to prove the result. Let $(\lambda^*, S^*, s^*, \sigma^*, x^*)$ be a saddle point. In particular, x^* minimizes the convex quadratic $x^T S^* x + (s^*)^T x + \sigma^*$ over F. If $x^* \in \text{bd}(F)$, then we are done. Otherwise, if $x^* \in \text{int}(F)$, then in fact x^* minimizes $x^T S^* x + (s^*)^T x + \sigma^*$

over the entire space $\mathbb{R}^n$ such that the gradient vanishes at x^*, i.e.,

$$2S^*x^* + s^* = 0.$$

We now consider two subcases. First, if $\text{Null}(S^*)$ is non-trivial, then there exists a non-zero direction $d \in \text{Null}(S)$ that yields a line of alternative optimal solutions to the inner minimization problem via the equation

$$\begin{aligned}
&(x^* + td)^T S^*(x^* + td) + (s^*)^T(x^* + td) + \sigma^* \\
&\qquad = (x^*)^T S^* x^* + (s^*)^T x^* + \sigma^* + (2S^*x^* + s^*)^T d \\
&\qquad = (x^*)^T S^* x^* + (s^*)^T x^* + \sigma^*,
\end{aligned}$$

which holds for all $t \in \mathbb{R}$. In particular, t can be taken such that $x^* + td \in \text{bd}(F)$, as desired. Second, if $S^* \succ 0$, then by complementary slackness, the corresponding optimal $Y^* = X^* - x^*(x^*)^T$ in (19) must be 0, i.e., the optimal X^* is rank-1. This proves (18) is tight, i.e., that $r^* = v^*$. □

To see that QSMB satisfies the conditions of Lemma 3, we express (1) and (2) in the form of the lemma. In this case, $F = [0, 1]^n$, and the full set of quadratic constraints considered in the lemma is

$$-x_j \le 0 \;\; \forall\, j, \qquad x_j - 1 \le 0 \;\; \forall\, j,$$

which define the compact convex F, and

$$x_i x_j - x_i \le 0 \;\; \forall\, 1 \le i, j \le n,$$

which are redundant for F and correspond to the RLT upper bound constraints. Clearly, assumption (a) of the lemma holds, i.e., there exists an interior point of F that satisfies all of the quadratic constraints—even the redundant ones—strictly. Also, assumption (b) is satisfied because the n redundant constraints $x_i^2 - x_i \le 0$ have Hessians that sum to the identity matrix. Finally, assumption (c) is satisfied because the feasible set of the relaxation implies that $x \in [0, 1]^n$, i.e., that x is bounded, and the constraint $\text{diag}(X) \le x$ along with the positive semidefiniteness constraint imply X is bounded as well.

7 Conclusions

We conclude the paper by summarizing the complexity of minimizing a general quadratic function $f(x) := x^TQx + c^Tx$ over the sets $\{0, 1\}^n$ and $[0, 1]^n$; see Table 2. In particular, the polynomial time solvable cases are well-known as discussed in the Introduction, and this paper establishes tractability for the submodular case.

Let us discuss the NP-hardness results provided in the table. First, the equivalence of minimizing a convex quadratic f over $\{0,1\}^n$ with QUBO (quadratic unconstrained binary optimization) follows from the equality

$$x^T Qx + c^T x = x^T (Q - \lambda_{min} I)x + (c + \lambda_{min} e)^T x$$

for $x \in \{0,1\}^n$, where $x_i^2 = x_i$ and λ_{min} is the minimum eigenvalue of Q. Note that the matrix $Q - \lambda_{min} I$ positive semidefinite. Since QUBO is NP-hard, so is the former. Second, consider $Q = Q_1 + Q_2$ with Q_1 positive semidefinite and Q_2 submodular. The equivalence of minimizing f over $[0,1]^n$ in this case with QP over the box follows from a splitting argument as follows. Given a general symmetric matrix Q, split it as $Q = Q_+ + Q_-$ where Q_+ has all the nonnegative entries and Q_- has all the negative entries. Then

$$x^T Qx + c^T x = x^T (Q_+ - \gamma_{min} I)x + x^T (Q_- + \gamma_{min} I)x + c^T x$$

where γ_{min} is the minimum eigenvalue of Q_+. This makes the matrix $Q_+ - \gamma_{min} I$ convex and $Q_- + \gamma_{min} I$ submodular. Since QP over the box is NP-hard, so is the former.

	f convex	f submodular	f sum of convex and submodular
$\{0,1\}^n$	NP-hard (QUBO) Garey and Johnson [1979]	P (LP) Padberg [1989]	NP-hard (by setting $Q_2 = 0$)
$[0,1]^n$	P (convex QP) Kozlov et al. [1980]	P (SDP) (This paper)	NP-hard (QP over box) Horst et al. [2000]

Table 2: Complexity of minimizing general $f(x) = x^T Qx + c^T x$ over $\{0,1\}^n$ and $[0,1]^n$

References

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.

A. Ben-Tal, A. Nemirovski, and L. E. Ghaoui. *Robust optimization*. Princeton University Press, 2009.

D. P. Bertsekas. *Convex Optimization Theory*. Athena Scientific, Belmont, MA, 2009.

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.

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.

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

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. 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.

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.

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

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.

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.

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

Q. Yu and S. Küçükyavuz. On constrained mixed-integer dr-submodular minimization. *Mathematics of Operations Research*, Ahead of Print, 2024.

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*, (3):453–465, 2000.