Polyhedral Analysis of Quadratic Optimization Problems with Stieltjes Matrices and Indicators

Peijing Liu * Alper Atamtürk † Andrés Gómez ‡
Simge Küçükyavuz §

April 5, 2024

Abstract

In this paper, we consider convex quadratic optimization problems with indicators on the continuous variables. In particular, we assume that the Hessian of the quadratic term is a Stieltjes matrix, which naturally appears in sparse graphical inference problems and others. We describe an explicit convex formulation for the problem by studying the Stieltjes polyhedron arising as part of an extended formulation and exploiting the supermodularity of a set function defined on its extreme points. Our computational results confirm that the proposed convex relaxation provides an exact optimal solution and may be an effective alternative, especially for instances with large integrality gaps that are challenging with the standard approaches.

1 Introduction

Given vectors $\boldsymbol{a}, \boldsymbol{c} \in \mathbb{R}^n$, a *Stieltjes* matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ (that is, $\boldsymbol{Q} \succ 0$ and $Q_{ij} \leq 0$ for $i \neq j$), and a convex set $C \subseteq \mathbb{R}^{2n}$, consider the mixed-integer quadratic optimization (MIQO) problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}\in\{0,1\}^n} \ \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{z} + \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} \tag{1a}$$

$$\text{s.t. } \boldsymbol{x} \circ (\boldsymbol{e} - \boldsymbol{z}) = \boldsymbol{0} \tag{1b}$$

$$(\boldsymbol{x}, \boldsymbol{z}) \in C, \tag{1c}$$

*Department of Industrial and System Engineering, University of Southern California, peijingl@usc.edu.

†Department of Industrial Engineering and Operations Research, University of California, Berkeley, atamturk@berkeley.edu.

‡Department of Industrial and System Engineering, University of Southern California, gomezand@usc.edu.

§Department of Industrial Engineering and Management Sciences, Northwestern University, simge@northwestern.edu.

where $\boldsymbol{e}$ is an n-dimensional vector of ones. Problem (1) with a Stieltjes matrix $\boldsymbol{Q}$ arises naturally in statistical problems with graphical models, which we discuss in §2, and others.

A fundamental step towards solving (1) effectively is designing strong convex relaxations of the optimization problem. Several approaches have been proposed for MIQO in the literature, by exploiting low-dimensional cases [13, 16] or low-rank cases [4, 5, 15, 24, 25, 26]. In this paper, we turn our attention to a critical substructure given by

$$X_Q \stackrel{\text{def}}{=} \left\{(t, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}^{n+1} \times \{0,1\}^n : t \geq \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x},\ (1\text{b})\right\}.$$

Set X_Q is the mixed-integer epigraph of a quadratic function with a Stieltjes matrix and indicators. Atamtürk and Gómez [3] show that if $\boldsymbol{a}$ has all entries of the same sign, problem (1) with a Stieltjes matrix $\boldsymbol{Q}$ is polynomial-time solvable. However, a tractable convex relaxation of (1) that guarantees optimality under the same conditions has been missing. This paper presents such a convex relaxation.

Outline

The rest of the paper is organized as follows. In §2, we discuss the applications of quadratic optimization with a Stieltjes matrix and indicators. In §3, we provide the necessary background for the paper. In §4, we convexify X_Q by exploiting an underlying polyhedral structure. In §5, we present experimental results illustrating the computational impact of the proposed convexification.

Notation

Given any $n \in \mathbb{Z}_+$, let $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$. We denote vectors and matrices in **bold**. Moreover, $\mathbf{0}$ denotes either a vector or matrix of zeros (whose dimension is clear from the context), $\boldsymbol{e}$ and $\boldsymbol{E}$ denote a vector and matrix of ones, respectively. Given a set $S \subseteq [n]$ we let $\boldsymbol{e_S} \in \mathbb{R}^n$ denote the indicator vector of S, i.e., $(e_S)_i = 1$ if $i \in S$ and $(e_S)_i = 0$ otherwise. Moreover, given $i \in S$, we let $\boldsymbol{e}_i \stackrel{\text{def}}{=} \boldsymbol{e}_{\{i\}}$ denote the i-th standard vector of $\mathbb{R}^n$. Similarly, we let $\boldsymbol{E_{ij}}$ be the square matrix whose dimensions can be inferred from the context, with a 1 in the (i,j)-th position and 0 elsewhere.

We let $\mathbb{S}^n_+$ denote the cone of positive semidefinite $n \times n$ matrices, and $\mathbb{S}^n_{++}$ the set of positive definite $n \times n$ matrices. Given a matrix $\boldsymbol{W}$, we let $\boldsymbol{W}^\dagger$ denote its pseudoinverse. A special case of the pseudoinverse that is used throughout the paper pertains to matrices of the form $\boldsymbol{W} = \begin{pmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ where $\boldsymbol{A}$ is invertible, in which case $\boldsymbol{W}^\dagger = \begin{pmatrix} \boldsymbol{A^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. Given two matrices $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$ of the same dimension, we let $\boldsymbol{W_1} \circ \boldsymbol{W_2}$ denote their Hadamard (entrywise) product.

Given a matrix $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ and set S, denote by $\boldsymbol{W_S} \in \mathbb{R}^{S\times S}$ the submatrix of $\boldsymbol{W}$ induced by S. Moreover, given $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{S} \subseteq [n]$, observe that $\left(\boldsymbol{W} \circ \boldsymbol{e_S e_S^\top}\right) \in \mathbb{R}^{n\times n}$ is the matrix that coincides with $\boldsymbol{W_S}$ in the entries indexed by S and is 0 elsewhere. To simplify the notation, given $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ and $S \subseteq [n]$, we let

$$\boldsymbol{W_S^\star} \stackrel{\text{def}}{=} \left(\boldsymbol{W} \circ \boldsymbol{e_S e_S^\top}\right)^\dagger;$$

if S corresponds to the first indices of $[n]$, then note that $\boldsymbol{W_S^\star} = \begin{pmatrix} \boldsymbol{W_S^{-1}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$.

2 MIQOs with Stieltjes matrices

Problem (1) with a Stieltjes matrix $\boldsymbol{Q}$ arises naturally in statistical problems with graphical models, which we discuss in §2.1. Set X_Q is critical to such problems because a convex relaxation of (1) can be obtained as

$$\begin{aligned} \min_{(t,\boldsymbol{x},\boldsymbol{z})\in\mathbb{R}^{2n+1}} \quad & \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{z} + t \\ \text{s.t. } & (t,\boldsymbol{x},\boldsymbol{z}) \in \operatorname{conv}(X_Q),\ (\boldsymbol{x},\boldsymbol{z}) \in C, \end{aligned}$$

and the relaxation is exact if $C = \mathbb{R}^{2n}$.

Understanding $\operatorname{conv}(X_Q)$ for Stieltjes matrices can also be helpful in solving MIQO problems with non-Stieltjes quadratic matrices. In such cases, a standard approach in the literature is to decompose $\boldsymbol{Q}$ into simpler matrices of the form $\boldsymbol{Q} = \boldsymbol{Q_0} + \sum_{k\in K} \boldsymbol{Q_k}$ for some index set K, where $\boldsymbol{Q_k}$ are "simple" matrices and $\boldsymbol{Q_0} \succeq 0$ is a remainder "complicated" matrix. Then relaxations of problem (1) can be obtained as

$$\min_{(\boldsymbol{t},\boldsymbol{x},\boldsymbol{z})\in\mathbb{R}^{|K|+2n}} \quad \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{z} + \boldsymbol{x}^\top \boldsymbol{Q_0} \boldsymbol{x} + \sum_{k\in K} t_k \tag{2a}$$

$$\text{s.t. } (t_k,\boldsymbol{x},\boldsymbol{z}) \in \operatorname{conv}(X_{Q_k}) \qquad \forall k \in K \tag{2b}$$

$$(\boldsymbol{x},\boldsymbol{z}) \in C. \tag{2c}$$

The most prevalent relaxation in the literature to tackle (1) is the *perspective relaxation* [1, 11, 12, 14], which is a special case of the approach outlined where $|K| = 1$ and $\boldsymbol{Q_1}$ is a diagonal positive definite matrix. A good understanding of relaxations of X_Q allows one to extend the perspective relaxation to more general (non-diagonal) classes of Stieltjes matrices. In fact, as we discuss in §2.2, matrices $\boldsymbol{Q_k}$ are not required to be Stieltjes to utilize the results of this paper.

2.1 A direct application

Mixed-integer convex quadratic problems with Stieltjes matrices arise, for example, in inference problems with Besag-York-Mollié graphical models [9]. In

this class of graphical models, the vertex set V of graph $\mathcal{G} = (V, E)$ represents latent random variables, Y, and the edge set E represents relationships between random variables. The existence of an edge $[i, j] \in E$ indicates that the random variables i and j should have similar values. Thus, the values of the unobserved random variables can be estimated as the optimal solution of the optimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^V, \boldsymbol{z}\in\{0,1\}^V} \sum_{i\in V} a_i(y_i - x_i)^2 + \sum_{[i,j]\in E} a_{ij}(x_i - x_j)^2 \tag{3a}$$

$$\text{s.t. (1b)} - \text{(1c)}, \tag{3b}$$

where y_i represents some noisy measurement of the value of random variable Y_i, $i \in V$, $\boldsymbol{a} \geq \boldsymbol{0}$ are parameters to be tuned, and the constraints incorporate logical constraints on the estimators. We refer the reader to Han et al. [17] for additional information on this class of combinatorial inference problems. As mentioned in the introduction, if $C = \mathbb{R}^n \times \{0, 1\}^n$ or $C = \mathbb{R}^n_+ \times \{0, 1\}^n$, optimization problems with Stieltjes matrices and indicators can be solved in polynomial time. We refer the reader to Atamtürk and Gómez [3] for the case with $\boldsymbol{a} \leq \boldsymbol{0}$ and $C = \mathbb{R}_+ \times \{0, 1\}^n$.

2.2 Stieljes-equivalent classes

In several cases, quadratic functions with non-Stieltjes matrices can be transformed into equivalent expressions with Stieltjes matrices. Specifically, given a vector $\boldsymbol{x} \in \mathbb{R}^n$, matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ and index set $I \subseteq \{1, \ldots, n\}$, define

$$\bar{x} = \begin{cases} -x_i & \text{if } i \in I \\ x_i & \text{otherwise,} \end{cases} \text{ and } \bar{Q}_{ij} = \begin{cases} -Q_{ij} & \text{if } |\{i, j\} \cap I| = 1 \\ Q_{ij} & \text{otherwise.} \end{cases}$$

Then observe that $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} = \bar{\boldsymbol{x}}^\top \bar{\boldsymbol{Q}}\bar{\boldsymbol{x}}$. Thus, since $(t, \boldsymbol{x}, \boldsymbol{z}) \in X_Q \Leftrightarrow (t, \bar{\boldsymbol{x}}, \boldsymbol{z}) \in X_{\bar{Q}}$, we find that we can convexify sets with non-Stieltjes matrices $\boldsymbol{Q}$ provided that the transformed matrix $\bar{\boldsymbol{Q}}$ is Stieltjes.

Improved formulations for convexifications using Stieltjes-equivalent classes have been presented in the literature. First, there have been notable efforts in studying X_Q where $n = 2$ [8, 13, 18, 27]. Naturally, any positive definite 2×2 matrix is Stieltjes-equivalent: if $Q_{12} \leq 0$, then $\boldsymbol{Q}$ is already Stieljes; otherwise, after the substitution $\bar{x}_1 = -x_1$, we recover an equivalent expression with a Stieltjes matrix. In another line of work, Liu et al. [20] describe the closure of the convex hull of X_Q (in a polynomially-sized extended formulation) when $\boldsymbol{Q}$ is tridiagonal. It can also be verified that tridiagonal matrices are Stieljes equivalent.

3 Preliminaries

In this section, we discuss some preliminary results concerning the convexification of X_Q relevant to the current paper.

Theorem 1 (Wei et al. [27])**.** *Given any matrix $\boldsymbol{Q} \in \mathbb{S}^n_{++}$, the closure of the convex hull of X_Q can be described in an extended formulation as*

$$cl\ conv(X_Q) = \Big\{(t, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}^{2n+1} : \exists \boldsymbol{W} \in \mathbb{R}^{n\times n} \text{ such that } \begin{pmatrix} \boldsymbol{W} & \boldsymbol{x} \\ \boldsymbol{x}^\top & t \end{pmatrix} \in \mathbb{S}^{n+1}_+, \\ (\boldsymbol{z}, \boldsymbol{W}) \in conv(P_Q)\Big\},$$

where $P_Q \stackrel{def}{=} \left\{(\boldsymbol{z}, \boldsymbol{W}) \in \{0,1\}^n \times \mathbb{R}^{n\times n} : \boldsymbol{W} = \left(\boldsymbol{Q} \circ \boldsymbol{z}\boldsymbol{z}^\top\right)^\dagger\right\}.$

Note that P_Q is a finite set and, therefore, conv(P_Q) is a polytope. Consequently, we see from Theorem 1 that convexification of X_Q in a higher dimension reduces to the convexification of a polyhedral set, allowing for the use of theory and techniques from polyhedral theory. However, to utilize Theorem 1 effectively, a major challenge must be overcome: characterizing or approximating set conv(P_Q), which has not been studied in the literature. Our goal in this paper is to close this gap for the special case of Stieltjes polytopes, as defined next.

Definition 1 (Stieltjes polytope)**.** Given a Stieltjes matrix $\boldsymbol{Q}$, the Stieltjes polytope associated with $\boldsymbol{Q}$ is defined as

$$Z_Q = \text{conv}\left(\{(\boldsymbol{e_S}, \boldsymbol{Q^\star_S})\}_{S\subseteq[n]}\right). \tag{4}$$

Example 1. Let $n = 3$, define $\boldsymbol{D} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, $\boldsymbol{q} = \boldsymbol{e}$ and $\boldsymbol{Q} = \boldsymbol{D} - \boldsymbol{e}\boldsymbol{e}^\top = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 2 \end{pmatrix}$. For the Stieltjes matrix $\boldsymbol{Q}$ the Stieltjes polytope Z_Q is the convex hull of the following eight points in $\mathbb{R}^{12}$:

$$\left(\boldsymbol{0}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right), \left(\boldsymbol{e_1}, \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right), \left(\boldsymbol{e_2}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right),$$
$$\left(\boldsymbol{e_3}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}\right), \left(\boldsymbol{e_{\{1,2\}}}, \begin{pmatrix} 3/5 & 1/5 & 0 \\ 1/5 & 2/5 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right), \left(\boldsymbol{e_{\{2,3\}}}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2/5 & 1/5 \\ 0 & 1/5 & 3/5 \end{pmatrix}\right),$$
$$\left(\boldsymbol{e_{\{1,3\}}}, \begin{pmatrix} 2/3 & 0 & 1/3 \\ 0 & 0 & 0 \\ 1/3 & 0 & 2/3 \end{pmatrix}\right), \left(\boldsymbol{e_{\{1,2,3\}}}, \begin{pmatrix} 5/3 & 1 & 4/3 \\ 1 & 1 & 1 \\ 4/3 & 1 & 5/3 \end{pmatrix}\right).$$

■

Since describing Z_Q requires finding a polyhedral description of the inverses of submatrices of $\boldsymbol{Q}$, we review two technical lemmas concerning matrix inversion that will be used throughout the paper.

Lemma 1 (Blockwise inversion, Lu and Shiou [22])**.** *The inverse of a non-singular square matrix* $\boldsymbol{R} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}$ *is given by*

$$\boldsymbol{R^{-1}} = \begin{pmatrix} \boldsymbol{A^{-1}} + \boldsymbol{A^{-1}B}(\boldsymbol{D} - \boldsymbol{CA^{-1}B})^{-1}\boldsymbol{CA^{-1}} & -\boldsymbol{A^{-1}B}(\boldsymbol{D} - \boldsymbol{CA^{-1}B})^{-1} \\ (\boldsymbol{D} - \boldsymbol{CA^{-1}B})^{-1}\boldsymbol{CA^{-1}} & (\boldsymbol{D} - \boldsymbol{CA^{-1}B})^{-1} \end{pmatrix}.$$

We apply Lemma 1 to a special class of matrices. We summarize the relevant results in the following corollary.

Corollary 1. *The inverse of a positive definite matrix* $\boldsymbol{R} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{v} \\ \boldsymbol{v}^\top & d \end{pmatrix}$*, where* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$*,* $v \in \mathbb{R}^n$ *and* $d \in \mathbb{R}_+$ *is given by*

$$\boldsymbol{R^{-1}} = \begin{pmatrix} \boldsymbol{A^{-1}} + \boldsymbol{A^{-1}v}(d - \boldsymbol{v}^\top\boldsymbol{A^{-1}v})^{-1}\boldsymbol{v}^\top\boldsymbol{A^{-1}} & -\boldsymbol{A^{-1}v}(d - \boldsymbol{v}^\top\boldsymbol{A^{-1}v})^{-1} \\ (d - \boldsymbol{v}^\top\boldsymbol{A^{-1}v})^{-1}\boldsymbol{v}^\top\boldsymbol{A^{-1}} & (d - \boldsymbol{v}^\top\boldsymbol{A^{-1}v})^{-1} \end{pmatrix}.$$

Moreover, letting $\boldsymbol{u} \stackrel{def}{=} \begin{pmatrix} -\boldsymbol{A^{-1}v} \\ 1 \end{pmatrix}$*, the identity*

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{v} \\ \boldsymbol{v}^\top & d \end{pmatrix}^\dagger - \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0}^\top & 0 \end{pmatrix}^\dagger = \frac{1}{d - \boldsymbol{v}^\top\boldsymbol{A^{-1}v}}\boldsymbol{uu}^\top \tag{5}$$

holds.

Note that since the difference in (5) is a rank-one matrix and, therefore, it is positive semi-definite. A recursive application of this property also wields that if $T \supseteq S$, then $\boldsymbol{W^\star_T} \succeq \boldsymbol{W^\star_S}$. The second lemma we introduce concerns the inverses of Stieltjes matrices in particular.

Lemma 2 (Supermodular inverses, Atamtürk and Gómez [3])**.** *Given a matrix* $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ *and a pair of indices* $i, j \in [n]$*, define the set function* $\theta_{ij}(S) \stackrel{def}{=} (\boldsymbol{Q^\star_S}^\dagger)_{ij}$ *as the* (i,j)*-th entry of matrix* $\boldsymbol{Q^\star_S}$*. If* $\boldsymbol{Q}$ *is a Stieltjes matrix, then* θ_{ij} *is a non-decreasing supermodular function for all pairs of indices* $i, j \in [n]$*.*

Observe that since $\theta_{ij}(\emptyset) = 0$ and θ_{ij} is non-decreasing, Lemma 2 implies that inverses of Stieltjes matrices are non-negative, a fact that is observed in [23] (along with several other properties of Stieltjes matrices).

4 Facial structure of Stieltjes polytopes

In this section, we study convexifications of Stieltjes polytopes as defined in Definition 1. Throughout, matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is assumed to be Stieltjes.

Proposition 1. *Any point $(\boldsymbol{z}, \boldsymbol{W}) \in Z_Q$ satisfies the following properties:*

1. $W_{ij} = W_{ji}$ *for all* $1 \leq i < j \leq n$,
2. $\sum_{j=1}^{n} Q_{ij} W_{ij} = z_i$ *for all* $i = 1, \ldots, n$,
3. $W_{ij} = 0$ *for all* i, j *such that* $Q^{-1}_{ij} = 0$,
4. $W_{ij} \geq 0$ *for all* i, j.

Proof. We first check the validity of the equalities. The first set of equalities follows since all extreme points of Z_Q have symmetric matrices. The second set follows from [27, Proposition 6]. For the third set, from Lemma 2 (non-negative and non-decreasing θ) it follows that extreme points satisfy $0 \leq W_{ij} \leq Q^{-1}_{ij}$: clearly, if $Q^{-1}_{ij} = 0$, then $W_{ij} = 0$ holds. Finally, the fourth set of inequalities follows directly from the non-negativity of inverses of Stieltjes matrices. □

It is evident from Proposition 1 that Z_Q is not full-dimensional. In this paper, we study the relaxation induced by the upper-bound constraints

$$P^{\leq}_{Q} \stackrel{\text{def}}{=} \left\{ (\boldsymbol{z}, \boldsymbol{W}) \in \{0,1\}^n \times \mathbb{R}^{n \times n} : \boldsymbol{W} \leq \left(\boldsymbol{Q} \circ \boldsymbol{z}\boldsymbol{z}^\top \right)^\dagger \right\}.$$

Defining $Z^{\leq}_{Q} \stackrel{\text{def}}{=} \text{conv}(P^{\leq}_{Q})$, it is easy to show that $Z^{\leq}_{Q}$ is full-dimensional.

Given $i, j \in [n]$ and $S \subseteq [n]$, let $\theta_{ij}(S) = (\boldsymbol{Q}^{\star}_{\boldsymbol{S}})_{ij}$ be the function introduced in Lemma 2, and given $k \in [n] \setminus S$, let

$$\rho_{ij}(k; S) \stackrel{\text{def}}{=} \theta_{ij}(S \cup \{k\}) - \theta_{ij}(S).$$

Finally, let $\boldsymbol{R}(k; S)$ be the matrix that collects functions ρ in its entries; that is, $R(k; S)_{ij} = \rho_{ij}(k; S)$. Recall, from Corollary 1 and the discussion immediately thereafter, that $\boldsymbol{R}$ is a rank-one matrix for all values of k and S. Supermodularity of θ_{ij} immediately leads to a class of valid inequalities.

4.1 Valid inequalities for $P^{\leq}_{Q}$

We now study the facial structure of $Z^{\leq}_{Q}$, which bounds matrix $\boldsymbol{W}$ from above. The facets of $Z^{\leq}_{Q}$ are given by the *polymatroid inequalities* [6], and are a direct consequence of supermodularity of θ. For any permutation $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$ of $[n]$, define $S^{\pi}_{0} \stackrel{\text{def}}{=} \emptyset$ and for $k \in [n]$, define set

$$S^{\pi}_{k} \stackrel{\text{def}}{=} \{\pi_1, \pi_2, \ldots, \pi_k\}.$$

Proposition 2. *For any $i, j \in [n]$ and any permutation $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$ of $[n]$, the inequality*

$$W_{ij} \leq \sum_{k=1}^{n} \rho_{ij}(\pi_k; S^{\pi}_{k-1}) z_{\pi_k} \tag{6}$$

is valid for $Z^{\leq}_{Q}$.

Proof. Inequality (6) is a polymatroid inequality, necessary to describe the Lovász extension, which describes the concave envelope of a supermodular function [7, 21]. □

Observe that, given permutation $\boldsymbol{\pi}$, inequalities (6) can be written compactly (for all $i, j \in [n]$) as the matrix inequalities

$$\boldsymbol{W} \leq \sum_{k=1}^{n} \boldsymbol{R}(\pi_k; S^{\pi}_{k-1}) z_{\pi_k}. \tag{7}$$

Example 1 (Continued). For matrix $\boldsymbol{Q} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 2 \end{pmatrix}$ and permutation $\boldsymbol{\pi} = (1, 2, 3)$, inequalities (7) reduce to

$$\begin{pmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{pmatrix} \leq \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} z_1 + \begin{pmatrix} 1/10 & 1/5 & 0 \\ 1/5 & 2/5 & 0 \\ 0 & 0 & 0 \end{pmatrix} z_2 + \begin{pmatrix} 16/15 & 4/5 & 4/3 \\ 4/5 & 3/5 & 1 \\ 4/3 & 1 & 5/3 \end{pmatrix} z_3. \blacksquare$$

4.2 Strength of the inequalities

We now show that inequalities (7) are indeed strong.

Proposition 3. *Inequality* (6) *is facet-defining for conv*$(P^{\leq}_{\overline{Q}})$.

Proof. In Table 1, we provide $(n + n^2)$ affinely independent points in $P^{\leq}_{\overline{Q}}$ such that (6) holds at equality.

Table 1: Affinely independent points in $P^{\leq}_{\overline{Q}}$ satisfying (6) at equality.

#	Indices	$\boldsymbol{z}$	$\boldsymbol{W}$
1	-	$\boldsymbol{e}$	$\boldsymbol{Q^{-1}}$
2	$\forall k, \ell \in [n]$ with $(k, \ell) \neq (i, j)$ and $Q^{-1}_{ij} \neq 0$	$\boldsymbol{e}$	$\boldsymbol{Q^{-1} - E_{k\ell}}$
3	$\forall k \in [n]$	$\boldsymbol{e_{S^{\pi}_k}}$	$\boldsymbol{Q^{\star}_{S^{\pi}_k}}$

• Point #1 belongs to P_Q and, therefore, it also belongs to $P^{\leq}_{\overline{Q}}$; if $\boldsymbol{z} = \boldsymbol{e}$, then inequality (6) reduces to $W_{ij} \leq \theta_{ij}(\boldsymbol{e}) = Q^{-1}_{ij}$, and thus the inequality is satisfied at equality at this point.

• Points #2 correspond to $n^2 - 1$ points which are obtained by subtracting a non-negative quantity from the feasible point #1 and, therefore, also belong to $P^{\leq}_{\overline{Q}}$; the inequality is tight for the same reason as point #1. Points #2 are affinely independent from previously introduced points because they are the first points with $W_{k\ell} \neq Q^{-1}_{k\ell}$.

• Points #3 correspond to n points belonging to P_Q; therefore, they also belong to $P^{\leq}_{\overline{Q}}$; if $\boldsymbol{z} = \boldsymbol{e_{S^{\pi}_k}}$, then inequality (6) reduces to $W_{ij} \leq \theta_{ij}(\boldsymbol{e_{S^{\pi}_k}}) = \left(\boldsymbol{Q^{\star}_{S^{\pi}_k}}\right)_{ij}$, and thus the inequality is satisfied at equality at this point. Finally,

points #3 are affinely independent from others because they are the first points with $z_k \neq 1$. □

From Proposition 3, we see that all inequalities (6) for all combinations of $i, j \in [n]$ and all permutations $\boldsymbol{\pi}$ are necessary to describe $Z_Q^{\leq}$. We now show that they are, along with the bound inequalities, sufficient.

Theorem 2. *Inequalities* (7) *(for all permutations* $\boldsymbol{\pi}$*) and bound constraints* $\mathbf{0} \leq \boldsymbol{z} \leq \boldsymbol{e}$ *describe conv*$(P_Q^{\leq})$.

Proof. We show that, for any $\boldsymbol{c} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$, the optimization problems

$$\min_{\boldsymbol{z}, \boldsymbol{W}} \boldsymbol{c}^\top \boldsymbol{z} + \langle \boldsymbol{\Sigma}, \boldsymbol{W} \rangle \text{ s.t. } (\boldsymbol{z}, \boldsymbol{W}) \in P_Q^{\leq} \tag{8}$$

$$\min_{\boldsymbol{z}, \boldsymbol{W}} \boldsymbol{c}^\top \boldsymbol{z} + \langle \boldsymbol{\Sigma}, \boldsymbol{W} \rangle \text{ s.t. } (7),\ \mathbf{0} \leq \boldsymbol{z} \leq \boldsymbol{e} \tag{9}$$

are equivalent; that is, either both are unbounded, or there exists an optimal solution of (9) that is feasible for (8).

Note that if $\Sigma_{ij} > 0$ for any $i, j \in [n]$, then both problems are unbounded by letting $W_{ij} \to -\infty$. Therefore, we assume $\boldsymbol{\Sigma} \leq \mathbf{0}$. In this case, in optimal solutions of (8), $\boldsymbol{W}$ will be set to its upper bound, i.e., $\boldsymbol{W} = \left(\boldsymbol{Q} \circ \boldsymbol{z}\boldsymbol{z}^\top\right)^\dagger$ holds. By abuse of notation, let $\theta_{ij}(\boldsymbol{z}) \stackrel{\text{def}}{=} \theta_{ij}(S_z)$, where $S_z = \{i \in [n] : z_i = 1\}$; since $\left(\left(\boldsymbol{Q} \circ \boldsymbol{z}\boldsymbol{z}^\top\right)^\dagger\right)_{ij} = \theta_{ij}(\boldsymbol{z})$, we find that (8) is equivalent to

$$\min_{\boldsymbol{z} \in \{0,1\}^n} \boldsymbol{c}^\top \boldsymbol{z} + \sum_{i=1}^n \sum_{j=1}^n \Sigma_{ij} \theta_{ij}(\boldsymbol{z}). \tag{10}$$

Recalling that functions θ_{ij} are supermodular (Lemma 2) and $\Sigma_{ij} \leq 0$ for all i, j, we find that $\Theta(\boldsymbol{z}) \stackrel{\text{def}}{=} \sum_{i=1}^n \sum_{j=1}^n \Sigma_{ij} \theta_{ij}(\boldsymbol{z})$ is a submodular function. Therefore, it follows that (10) is equivalent to minimization over its Lovász extension [21], which is precisely (9). □

Now consider the relaxation of (1) induced by Theorem 1, but using only inequalities (7) and bound constraints instead of the full description of P_Q:

$$\min_{\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{W}, t} \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{z} + t \tag{11a}$$

$$\text{s.t. } \begin{pmatrix} \boldsymbol{W} & \boldsymbol{x} \\ \boldsymbol{x}^\top & t \end{pmatrix} \in \mathbb{S}_+^{n+1} \tag{11b}$$

$$\boldsymbol{W} \leq \sum_{k=1}^n \boldsymbol{R}(\pi_k; S_{k-1}^\pi) z_{\pi_k} \quad \text{for all permutations } \boldsymbol{\pi} \text{ of } [n] \tag{11c}$$

$$(\boldsymbol{x}, \boldsymbol{z}) \in C \tag{11d}$$

$$\boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{z} \in [0,1]^n,\ t \in \mathbb{R}_+,\ \boldsymbol{W} \in \mathbb{R}^{n \times n}. \tag{11e}$$

Because constraints (11c) are a relaxation of constraints $(\boldsymbol{z}, \boldsymbol{W}) \in P_Q$, we find from Theorem 1 that (11) is indeed a valid relaxation. In general, (11) can be weak: in fact, it can be unbounded unless constraints $\boldsymbol{W} \geq \boldsymbol{0}$ are also added. Nonetheless, as we now show, under the specific conditions stated in [3] for polynomial-time solvability of (1), the relaxation (11) is exact. Given a Stieltjes matrix $\boldsymbol{Q}$, define the optimization

$$\min_{\substack{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}\in\{0,1\}^n \\ \boldsymbol{x}\circ(\boldsymbol{e}-\boldsymbol{z})=0}} \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{z} + \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}, \tag{12}$$

which is the special case of (1) with $C = \mathbb{R}^{2n}$.

Proposition 4. *If $\boldsymbol{a} \leq \boldsymbol{0}$ or $\boldsymbol{a} \geq \boldsymbol{0}$, and $C = \mathbb{R}^{2n}$, then there exists an optimal solution of* (11) *that is also optimal for* (12)*, with the same objective value.*

Proof. The start of the proof follows the steps in [27, Theorem 1], which we repeat for completeness. Constraint (11b) is equivalent to the system [2]

$$\boldsymbol{W} \succeq 0,\ t \geq \boldsymbol{x}^\top \boldsymbol{W}^\dagger \boldsymbol{x},\ \text{and}\ \boldsymbol{W}\boldsymbol{W}^\dagger \boldsymbol{x} = \boldsymbol{x}.$$

Therefore, variable t can be easily projected out since any optimal solution satisfies $t = \boldsymbol{x}^\top \boldsymbol{W}^\dagger \boldsymbol{x}$. We can restate problem (11) as

$$\min_{\boldsymbol{x},\boldsymbol{z},\boldsymbol{W}} \boldsymbol{a}^\top \boldsymbol{W}\boldsymbol{W}^\dagger \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{z} + \boldsymbol{x}^\top \boldsymbol{W}^\dagger \boldsymbol{x} \tag{13a}$$

$$\text{s.t. } \boldsymbol{W}\boldsymbol{W}^\dagger \boldsymbol{x} = \boldsymbol{x},\ \boldsymbol{W} \in \mathbb{S}^n_+ \tag{13b}$$

$$(11\text{c}) - (11\text{e}). \tag{13c}$$

Note that we use the equality in (13b) to rewrite a linear term in the objective. We now project out variables $\boldsymbol{x}$: The KKT conditions associated with the continuous variables are

$$\begin{aligned} &\boldsymbol{W}\boldsymbol{W}^\dagger \boldsymbol{x} = \boldsymbol{x} \\ &\boldsymbol{a}^\top \boldsymbol{W}\boldsymbol{W}^\dagger + 2\boldsymbol{W}^\dagger x + \boldsymbol{\lambda}^\top \left(\boldsymbol{W}\boldsymbol{W}^\dagger - \boldsymbol{I}\right) = \boldsymbol{0}, \end{aligned}$$

which are satisfied by setting $\boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{W}\boldsymbol{a}$ and $\boldsymbol{\lambda}^* = 0$. Substituting $\boldsymbol{x}$ with its optimal value, we find that problem (13) further simplifies to

$$\min_{\boldsymbol{z},\boldsymbol{W}} \langle -\frac{1}{4}\boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{W}\rangle + \boldsymbol{c}^\top \boldsymbol{z} \tag{14a}$$

$$\text{s.t. } (11\text{c}), \boldsymbol{W} \in \mathbb{S}^n_+, \boldsymbol{0} \leq \boldsymbol{z} \leq \boldsymbol{e}. \tag{14b}$$

In the last step of the proof (which does not follow from [27]), we study the relaxation of (14) obtained by removing constraint $\boldsymbol{W} \in \mathbb{S}^n_+$. In particular, we show that there exist optimal solutions $(\bar{\boldsymbol{z}}, \bar{\boldsymbol{W}}, \bar{t})$ of the relaxation such that: *(i)* $\bar{\boldsymbol{z}}$ is integral; *(ii)* $\bar{\boldsymbol{W}}$ is positive semidefinite; *(iii)* $\bar{t} = -\frac{1}{4}\boldsymbol{a}_{\boldsymbol{S}}^\top \boldsymbol{Q}_{\boldsymbol{S}}^{-\mathbf{1}} \boldsymbol{a}_{\boldsymbol{S}}$, where $S = \{i \in [n] : \bar{z}_i = 1\}$. As a consequence, $(\bar{\boldsymbol{z}}, \bar{\boldsymbol{W}}, \bar{t})$ is also optimal for (14) and (12), concluding the proof.

Observe that $-\frac{1}{4}\boldsymbol{a}\boldsymbol{a}^\top \leq \mathbf{0}$ due to the assumption that $\boldsymbol{a}$ is of the same sign. Therefore, if $\boldsymbol{W} \in \mathbb{S}_+^n$ is removed, then in optimal solutions of the relaxation, $\bar{\boldsymbol{W}}$ is equal to its upper bound, i.e.,

$$\bar{\boldsymbol{W}} = \min_{\pi \in \Pi} \sum_{k=1}^{n} \boldsymbol{R}(\pi_k; S_{k-1}^{\pi}) z_{\pi_k}, \tag{15}$$

where Π is the set of all permutations of $[n]$. In other words, the relaxation is equivalent to

$$\min_{\mathbf{0} \leq \boldsymbol{z} \leq \mathbf{1}} \min_{\pi \in \Pi} \left\{ \sum_{k=1}^{n} \left(\langle -\frac{1}{4}\boldsymbol{a}\boldsymbol{a}^\top, \boldsymbol{R}(\pi_k; S_{k-1}^{\pi}) \rangle \right) z_{\pi_k} \right\} + \boldsymbol{c}^\top \boldsymbol{z}. \tag{16}$$

Recognizing (16) as a linear optimization problem over the Lovász extension of a submodular function, we conclude that $\bar{\boldsymbol{z}} \in \{0,1\}^n$ in optimal solutions, proving *(i)*. Since optimal permutations for (15) correspond to nondecreasing orders of $\bar{\boldsymbol{z}}$ [10], letting $\tau = \|\bar{\boldsymbol{z}}\|_0$ we conclude that

$$\bar{\boldsymbol{W}} = \sum_{k=1}^{\tau} \boldsymbol{R}(\pi_k; S_{k-1}) = \sum_{k=1}^{\tau} \left(\boldsymbol{W}_{\boldsymbol{S_k}}^{\star} - \boldsymbol{W}_{\boldsymbol{S_{k-1}}}^{\star} \right) = \boldsymbol{W}_{\boldsymbol{S_\tau}}^{\star},$$

where we defined $\boldsymbol{W}_{\boldsymbol{S_0}}^{\star} = \mathbf{0}$. In particular, $\bar{\boldsymbol{W}} \in \mathbb{S}_+^n$, proving *(ii)*, and substituting $\boldsymbol{W}$ with its optimal value in (16) we can prove *(iii)*. □

4.3 Separation algorithm

We now discuss the separation problem: given any fixed point $(\bar{\boldsymbol{z}}, \bar{\boldsymbol{W}}) \in \mathbb{R}^n \times \mathbb{R}^{n \times n}$, how to find a violated inequality (6) if there exists one. Since inequality (6) is a polymatroid inequality, it follows that the most violated inequalities correspond to permutations $(\pi_1, \pi_2, \ldots, \pi_n)$ such that $\bar{z}_{\pi_1} \geq \bar{z}_{\pi_2} \geq \cdots \geq \bar{z}_{\pi_n}$ [10]. In particular, since the permutation depends only on the values of $\bar{\boldsymbol{z}}$, we find that most violated permutations coincide for all values of indices i, j in (6). Therefore, using the matrix notation (7), we see that a straightforward way to compute a violated inequality is simply to compute matrices $\boldsymbol{Q}_{\boldsymbol{S_k^\pi}}^{\star}$ for all $k \in [n]$, which requires inverting $\mathcal{O}(n)$ matrices. A faster approach to compute the coefficients of inequalities (7) consists of using a Cholesky decomposition.

Indeed, consider the properties of coefficient matrices $\{\boldsymbol{R}(\pi_k; S_{k-1})\}_{k=1}^n$ in inequalities (7). First, they are rank-one matrices; that is, there exists $\boldsymbol{v_k} \in \mathbb{R}^n$ such that $\boldsymbol{R}(\pi_k; S_{k-1}) = \boldsymbol{v_k}\boldsymbol{v_k}^\top$. Second, they add up to $\boldsymbol{Q}^{-1}$; that is, $\sum_{k=1}^n \boldsymbol{R}(\pi_k; S_{k-1}) = \sum_{k=1}^n \boldsymbol{v_k}\boldsymbol{v_k}^\top = \boldsymbol{Q^{-1}}$. Third, the entries corresponding to indices that have not appeared yet in the permutation vanish; that is, $R(\pi_k; S_{k-1})_{ij} = 0$ if $\max\{i,j\} > k$, or equivalently $(v_k)_i = 0$ if $i > k$. If we define a matrix $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ such that its i-th column is precisely $\boldsymbol{v_i}$, these properties are equivalent to $\boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{Q^{-1}}$ and $\boldsymbol{V}$ is upper triangular (where elements are ordered according to the permutation π). Note that these properties

are similar to the ones corresponding to a Cholesky decomposition of $\boldsymbol{Q^{-1}}$, except that the matrix in the Cholesky decomposition is lower triangular instead of upper triangular. To account for this difference, it suffices to compute the Cholesky decomposition in the *reverse order* of the permutation, and because $\boldsymbol{Q^{-1}} \in \mathbb{S}^n_{++}$, the Cholesky decomposition is the unique matrix satisfying the required properties, and thus indeed coincides with the coefficients in inequalities (7). We summarize the separation algorithm in Proposition 5 below.

Proposition 5. *Algorithm 1 produces up to $\mathcal{O}(n^2)$ violated inequalities, one for each combination of elements $i, j \in [n]$, if there exists any.*

Algorithm 1 Separation procedure

Input: Point $\bar{\boldsymbol{z}} \in \mathbb{R}^n$.
Output: Most violated inequalities.

1: Find a permutation $\boldsymbol{\pi}$ satisfying $\bar{z}_{\pi_1} \leq \bar{z}_{\pi_2} \leq \cdots \leq \bar{z}_{\pi_n}$ ▷ Sorting
2: Compute Cholesky decomposition $\boldsymbol{Q^{-1}} = \boldsymbol{V}\boldsymbol{V}^\top$ according to order $\boldsymbol{\pi}$
3: **return** inequalities

$$\boldsymbol{W} \leq \sum_{k=1}^{n} \left(\boldsymbol{v_{\pi_k}} \boldsymbol{v_{\pi_k}}^\top\right) z_{\pi_k}$$

where $\boldsymbol{v_{\pi_i}}$ denotes the i-th column of $\boldsymbol{V}$.

We emphasize that the order in line 1 of Algorithm 1 is non-decreasing, instead of the natural non-increasing order that arises often with polymatroid inequalities, since we use the *reverse order* of the permutation in computing the Cholesky decomposition as we discussed in the preceding paragraph.

We now discuss the runtime of Algorithm 1. Since sorting the variables (line 1) can be done in $\mathcal{O}(n \log n)$, we find that the complexity of Algorithm 1 is dominated by the cost of computing a Cholesky decomposition (line 2), which is usually $\mathcal{O}(n^3)$. Note that line 2 also requires inverting matrix $\boldsymbol{Q}$, but this operation (with cubic runtime as well) needs to be performed only once as preprocessing. In some cases, the runtime for a general Stieltjes matrix can be improved. For example, Lemma 1 can be called recursively to compute the coefficients, requiring $\mathcal{O}(n)$ calls to a routine for matrix-vector product and matrix-matrix subtraction: if $\boldsymbol{Q}$ is sparse, then the vectors appearing in the computations are sparse as well, potentially improving runtimes for the multiplications (with appropriate data structures). Finally, we point out that the output of Algorithm 1 consists of $\mathcal{O}(n^2)$ inequalities with up to n nonzeros per inequalities, thus formulating the inequalities in a solver already requires processing a cubic number of nonzeros, thus it is not possible to improve this runtime (at least with an off-the-shelf solver).

5 Computational results

In this section, we describe computational experiments performed to test the effectiveness of the proposed convexification. For the computational study, we consider problems of the form (3), arising from the inference of sparse Besag-York-Mollié graphical models as discussed in §2.1. Specifically, given a graph $\mathcal{G} = (V, E)$, we consider problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^V, \boldsymbol{z}\in\{0,1\}^V} \frac{1}{\sigma^2}\sum_{i\in V}(y_i - x_i)^2 + \sum_{[i,j]\in E}(x_i - x_j)^2 + \mu\sum_{i\in V} z_i \tag{17a}$$

$$\text{s.t.} \sum_{i\in V} z_i \leq k,\ \boldsymbol{x}\circ(\boldsymbol{e}-\boldsymbol{z}) = \boldsymbol{0}, \tag{17b}$$

where $\sigma, \mu, k \in \mathbb{R}_+$ are given parameters. In our experiments, we solve *cardinality-penalized problems* with $\mu > 0$ and $k = |V|$ (modeling situations where the density of the model is penalized) as well as *cardinality-constrained problems* with $\mu = 0$ and $k < |V|$ (modeling situations where the sparsity is directly specified). In either case, we solve (17) for different values of σ, a parameter that is connected to the noise of the underlying model.

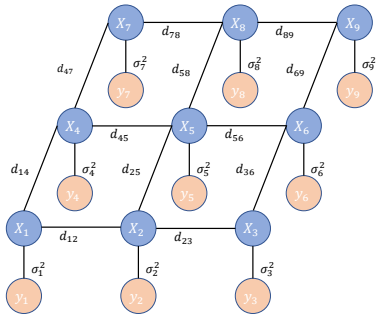

Figure 1: Grid graph for modeling the Besag-York-Mollié process; from [19].

The data are generated following [19] and are available online at `https://sites.google.com/usc.edu/gomez/data`. Figure 1 depicts the graph $\mathcal{G}$ used to model spatial processes. Graph $\mathcal{G}$ is given by a two-dimensional lattice; that is, vertices $V = [n]$ are arranged in a grid, with edges between horizontally and vertically adjacent vertices. We consider instances with grid sizes 10×10, thus resulting in instances with $|V| = 100$. To generate the data, we create a sparse "true" signal $\{X_i\}_{i\in V}$ with three non-negative spikes, each spike affecting a 3×3 block of elements in V. Thus, the generated signal can have at most 27 non-zeros (the construction of the underlying true signal is explained in detail in Appendix A). We then generate noisy observations as $y_i = |X_i + \epsilon_i|$, where ϵ_i are i.i.d. samples from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Note that since $\boldsymbol{y} \geq \boldsymbol{0}$,

linear coefficients of variables $\boldsymbol{x}$ obtained by expanding the quadratic terms in (17a) are non-positive.

In § 5.1, we describe the models tested, and in § 5.2, we report the computational experiments.

5.1 Models

We compare three reformulations or relaxations of the problem (1). The former two are commonly used perspective formulations [11, 14], while the latter is based on the valid inequalities proposed in §4. We describe these formulations next.

`pers-c`: Perspective relaxation with continuous variable $\boldsymbol{z}$ and big-M constraints:

$$\min_{\boldsymbol{x},\boldsymbol{z}} \frac{1}{\sigma^2}\|\boldsymbol{y}\|_2^2 + \frac{1}{\sigma^2}\sum_{i\in V}\left(\frac{x_i^2}{z_i} - 2y_i x_i\right) + \sum_{[i,j]\in E}(x_i - x_j)^2 + \mu\sum_{i\in V} z_i \tag{18a}$$

$$\text{s.t.} \sum_{i\in V} z_i \le k,\ -M\boldsymbol{z} \le \boldsymbol{x} \le M\boldsymbol{z} \tag{18b}$$

$$\boldsymbol{x} \in \mathbb{R}^V, \boldsymbol{z} \in [0,1]^V, \tag{18c}$$

where the value M is large enough so that the big-M constraints are redundant when $z_i = 1$. In our computations, the bounds are set to $M = 10$.

`pers-b`: Perspective reformulation `pers-c` with binary variables $\boldsymbol{z} \in \{0,1\}^V$. Using this formulation, problems are solved to optimality with branch-and-bound, closing the gap from the perspective relaxation.

`poly`: Polymatroid relaxation with the equalities and lower bounds given in Proposition 1 and the polymatroid inequalities as described in (11):

$$\min_{\boldsymbol{x},\boldsymbol{z},\boldsymbol{W},t} \frac{1}{\sigma^2}\|\boldsymbol{y}\|_2^2 - 2\frac{1}{\sigma^2}\sum_{i\in V} y_i x_i + \mu\sum_{i\in V} z_i + t \tag{19a}$$

$$\text{s.t.} \sum_{i\in V} z_i \le k \tag{19b}$$

$$\begin{pmatrix} \boldsymbol{W} & \boldsymbol{x} \\ \boldsymbol{x}^\top & t \end{pmatrix} \succeq 0 \tag{19c}$$

$$W_{ij} = W_{ji} \quad \text{for all } i,j \in V \tag{19d}$$

$$\sum_{j\in V} Q_{ij} W_{ij} = z_i \quad \text{for all } i \in V \tag{19e}$$

$$\boldsymbol{W} \le \sum_{k\in V} \boldsymbol{R}(\pi_k; S^\pi_{k-1}) z_{\pi_k} \quad \text{for all permutations } \boldsymbol{\pi} \text{ of } V \tag{19f}$$

$$\boldsymbol{x} \in \mathbb{R}^V,\ \boldsymbol{z} \in [0,1]^V,\ t \in \mathbb{R}_+,\ \boldsymbol{W} \in \mathbb{R}_+^{V\times V}, \tag{19g}$$

where $\boldsymbol{Q}$ is the Stieltjes matrix obtained by collecting all nonlinear terms in the objective and inequalities (19f) are those required to describe $\text{conv}(P_Q^{\leq})$. Note that (19) is a reformulation of (1) under the conditions of Proposition 4, that is, in the cardinality-penalized instances, and is a relaxation otherwise.

Remark 1 (Implementation of `poly`). As the total number of polymatroid inequalities is $|V|!$, listing all of them in (19) is impractical. Instead, we add (19f) as cutting planes; that is, we solve the `poly` problem (19) by adding polymatroid inequalities (19f) as cutting planes iteratively using the separation procedure 1 and stopping when the difference of the objective values between two subsequent iterations is small enough ($< 10^{-3}$).

5.2 Results

The perspective formulations (`pers-c`, `pers-b`) are solved with Gurobi version 9.0.2 using a single thread. The time limit for the computations is set to one hour; all other configurations are set to default values. The polymatroid relaxation (`poly`) is solved with Mosek version 9.3 with the default settings. All experiments are run on a Lenovo laptop with a 1.9 GHz Intel®Core™ i7-8650U CPU and 16 GB main memory. For cardinality-constrained instances, we set $k = 20$, and for cardinality-penalized instances, we choose μ so that the number of non-zeros of the estimator approximately matches the number of non-zeros of the underlying signal.

Tables 2 and 3 present results for values of $\sigma^2 \in \{0.5, 1.0, 2.0, 5.0, 10.0\}$ and cardinality-penalized and cardinality-constrained instances, respectively. Each row represents the average over five instances generated with identical parameters. The gap corresponding to each model compares the lower bound produced by the model with the best upper bound found by Gurobi with model `pers-b`. For model `pers-b`, we also report, under the #Opt column, the number of instances that can be solved to optimality before the time limit, and, under the Gap column, we report the average optimality gap at termination. The number of iterations for model `poly` is the number of cutting plane rounds required before termination, with each iteration resulting in the addition of up to $\binom{|V|}{2}$ cuts. All times reported are in seconds.

Table 2: Experiments with cardinality-penalized instances.

		`pers-c`		`pers-b`			`poly`		
σ^2	μ	Time	Gap	#Opt	Time	Gap	#Iter	Time	Gap
0.5	0.25	0.1	2.4%	5	2.1	0.0%	4	102.4	$3 \cdot 10^{-8}$
1.0	0.12	0.1	5.7%	1	2,998.3	1.7%	6	235.4	$7 \cdot 10^{-4}$
2.0	0.12	0.1	5.1%	2	2,937.1	0.5%	6	218.3	$2 \cdot 10^{-7}$
5.0	0.12	0.1	5.1%	1	3,021.9	0.9%	6	235.6	$3 \cdot 10^{-6}$
10.0	0.12	0.1	5.1%	2	2,990.6	0.8%	6	219.0	$9 \cdot 10^{-8}$

Table 3: Experiments with cardinality-constrained instances.

	poly		pers-b			poly		
σ^2	Time	Gap	#Opt	Time	Gap	#Iter	Time	Gap
0.5	0.1	3.3%	5	13.1	0.0%	4	181.3	$4 \cdot 10^{-7}$
1.0	0.1	7.4%	0	3,600.0	1.5%	7	285.3	$9 \cdot 10^{-4}$
2.0	0.1	7.1%	1	3,017.3	1.3%	7	277.5	$3 \cdot 10^{-3}$
5.0	0.1	7.1%	1	2,990.3	1.4%	7	281.7	$2 \cdot 10^{-3}$
10.0	0.1	7.1%	1	2,988.9	1.3%	7	260.0	$3 \cdot 10^{-3}$

In both cases, we observe that the relaxations from the perspective relaxation can be solved in a fraction of a second and yield gaps between 2% and 7%: the gaps are larger for cardinality-constrained instances and for problems with larger values of σ. Indeed, for a larger value of σ, terms x_i^2/z_i (crucial to the strength of the perspective relaxation) in the objective of (18) represent a relatively smaller portion of the objective, leading to larger gaps. When used in a branch-and-bound algorithm, it is possible to efficiently prove optimality if $\sigma^2 = 0.5$, but most instances with larger values of σ cannot be solved within the time limit of one hour.

For cardinality-penalized instances, model `poly` delivers optimal solutions on average in about 200 seconds. The runtimes are larger than those required to solve the perspective relaxation `pers-c` due to expenses associated with solving a semidefinite program with a large number of linear inequalities via cutting planes. Nonetheless, for the more challenging instances with $\sigma^2 \geq 1$, the runtimes of `poly` are at least an order of magnitude less than those arising from the branch-and-bound method to solve `pers-b`; for these instances, both `poly` and `pers-b` are exact models. The runtimes of `poly` for cardinality-constrained instances are similar. Although in this case, due to the additional cardinality constraint, `poly` is not guaranteed to deliver exact solutions to problem (17), the gaps proved are almost 0% and much smaller than those obtained after one hour of branching with `pers-b`.

Acknowledgments

Alper Atamtürk is supported, in part, by NSF AI Institute for Advances in Optimization Award 2112533 and the Office of Naval Research grant N00014-24-1-2149. Andrés Gómez is supported, in part, by grant FA9550-22-1-0369 from the Air Force Office of Scientific Research. Simge Küçükyavuz is supported, in part, by grant #2007814 from the National Science Foundation.

References

[1] M. S. Aktürk, A. Atamtürk, and S. Gürel. A strong conic quadratic reformulation for machine-job assignment with controllable processing times. *Operations Research Letters*, 37:187–191, 2009.

[2] A. Albert. Conditions for positive and nonnegative definiteness in terms of pseudoinverses. *SIAM Journal on Applied Mathematics*, 17(2):434–440, 1969.

[3] A. Atamtürk and A. Gómez. Strong formulations for quadratic optimization with M-matrices and indicator variables. *Mathematical Programming*, 170:141–176, 2018.

[4] A. Atamtürk and A. Gómez. Rank-one convexification for sparse regression. *arXiv preprint arXiv:1901.10334*, 2019.

[5] A. Atamtürk and A. Gómez. Supermodularity and valid inequalities for quadratic optimization with indicators. *Mathematical Programming*, 201: 295–338, 2023.

[6] A. Atamtürk and V. Narayanan. Polymatroids and mean-risk minimization in discrete optimization. *Operations Research Letters*, 36:618–622, 2008.

[7] A. Atamtürk and V. Narayanan. Submodular function minimization and polarity. *Mathematical Programming*, 196:57–67, 2022.

[8] A. Atamtürk, A. Gómez, and S. Han. Sparse and smooth signal estimation: Convexification of ℓ_0 formulations. *Journal of Machine Learning Research*, 22:1–43, 2021.

[9] J. Besag, J. York, and A. Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, 1991.

[10] J. Edmonds. Submodular functions, matroids, and certain polyhedra. In R. Guy, H. Hanani, N. Sauer, and J. Schönenheim, editors, *Combinatorial Structures and Their Applications*, pages 69–87. Gordon and Breach, 1970.

[11] A. Frangioni and C. Gentile. Perspective cuts for a class of convex 0–1 mixed integer programs. *Mathematical Programming*, 106:225–236, 2006.

[12] A. Frangioni and C. Gentile. SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. *Operations Research Letters*, 35:181–185, 2007.

[13] A. Frangioni, C. Gentile, and J. Hungerford. Decompositions of semidefinite matrices and the perspective reformulation of nonseparable quadratic programs. *Mathematics of Operations Research*, 45(1):15–33, 2020.

[14] O. Günlük and J. Linderoth. Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Mathematical Programming*, 124:183–205, 2010.

[15] S. Han and A. Gómez. Compact extended formulations for low-rank functions with indicator variables. *arXiv preprint arXiv:2110.14884*, 2021.

[16] S. Han, A. Gómez, and A. Atamtürk. 2x2 convexifications for convex quadratic optimization with indicator variables. *arXiv preprint arXiv:2004.07448*, 2020.

[17] S. Han, A. Gómez, and J.-S. Pang. On polynomial-time solvability of combinatorial Markov random fields. *arXiv preprint arXiv:2209.13161*, 2022.

[18] S. Han, A. Gómez, and A. Atamtürk. 2× 2-convexifications for convex quadratic optimization with indicator variables. *Mathematical Programming*, 202:95–134, 2023.

[19] Z. He, S. Han, A. Gómez, Y. Cui, and J.-S. Pang. Comparing solution paths of sparse quadratic minimization with a Stieltjes matrix. *Mathematical Programming*, 204:517–566, 2024.

[20] P. Liu, S. Fattahi, A. Gómez, and S. Küçükyavuz. A graph-based decomposition method for convex quadratic optimization with indicators. *Mathematical Programming*, 200(2):669–701, 2023.

[21] L. Lovász. Submodular functions and convexity. In *Mathematical programming the state of the art*, pages 235–257. Springer, 1983.

[22] T.-T. Lu and S.-H. Shiou. Inverses of 2× 2 block matrices. *Computers & Mathematics with Applications*, 43(1-2):119–129, 2002.

[23] R. J. Plemmons. M-matrix characterizations. I – nonsingular M-matrices. *Linear Algebra and its Applications*, 18:175–188, 1977.

[24] S. Shafieezadeh-Abadeh and F. Kılınç-Karzan. Constrained optimization of rank-one functions with indicator variables. *Mathematical Programming*, 2024. doi: https://doi.org/10.1007/s10107-023-02047-y. Article in Advance.

[25] L. Wei, A. Gómez, and S. Küçükyavuz. On the convexification of constrained quadratic optimization problems with indicator variables. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 433–447. Springer, 2020.

[26] L. Wei, A. Gómez, and S. Küçükyavuz. Ideal formulations for constrained convex optimization problems with indicator variables. *Mathematical Programming*, 192(1-2):57–88, 2022.

[27] L. Wei, A. Atamtürk, A. Gómez, and S. Küçükyavuz. On the convex hull of convex quadratic optimization problems with indicators. *Mathematical Programming*, 204(1-2):703–737, 2024.

A Construction of true signals in computational experiments

Here we discuss how to construct the true signal $\{X_i\}_{i\in V}$. Note that since each node in $\mathcal{G}$ is uniquely determined by its coordinates $1 \leq k,\ \ell \leq m$ in the $m \times m$ grid, we let $X_{(k,\ell)}$ denote the value of the signal at those coordinates.

The true signal is initially set to 0 at all its coordinates. We then repeat the following process three times:

1. Uniformly pick coordinates $2 \leq k,\ \ell \leq 9$.
2. Generate a 9-dimensional Gaussian spike $\boldsymbol{s} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta^{-1}})$, where
$$\Theta = \begin{pmatrix} 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 4 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 \end{pmatrix}.$$
3. For all $0 \leq j_1, j_2 \leq 2$, let $h = 1 + 3j_1 + j_2$ and update
$$X_{(k-1+j_1,\ell-q+j_2)} \leftarrow X_{(k-1+j_1,\ell-1+j_2)} + |s_h|\,.$$