BCOL RESEARCH REPORT 18.05

Industrial Engineering & Operations Research
University of California, Berkeley, CA 94720–1777

SPARSE AND SMOOTH SIGNAL ESTIMATION: CONVEXIFICATION OF ℓ_0-FORMULATIONS

ALPER ATAMTÜRK, ANDRÉS GÓMEZ AND SHAONING HAN

Abstract. Signal estimation problems with smoothness and sparsity priors can be naturally modeled as quadratic optimization with ℓ_0-"norm" constraints. Since such problems are non-convex and hard-to-solve, the standard approach is, instead, to tackle their convex surrogates based on ℓ_1-norm relaxations. In this paper, we propose new iterative (convex) conic quadratic relaxations that exploit not only the ℓ_0-"norm" terms, but also the fitness and smoothness functions. The iterative convexification approach substantially closes the gap between the ℓ_0-"norm" and its ℓ_1 surrogate. These stronger relaxations lead to significantly better estimators than ℓ_1-norm approaches and also allow one to utilize affine sparsity priors. In addition, the parameters of the model and the resulting estimators are easily interpretable. Experiments with a tailored Lagrangian decomposition method indicate that the proposed iterative convex relaxations yield solutions within 1% of the exact ℓ_0 approach, and can tackle instances with up to 100,000 variables under one minute.

Keywords Mixed-integer quadratic optimization, conic quadratic optimization, perspective formulation, sparsity.

November 2018; January 2020; August 2020

A. Atamtürk: Department of Industrial Engineering & Operations Research, University of California, Berkeley, CA 94720. atamturk@berkeley.edu

A. Gómez, S. Han: Daniel J. Epstein Department of Industrial & Systems Engineering, University of Southern California, CA 90089. gomezand@usc.edu, shaoning@usc.edu .

1. Introduction

Given nonnegative data $y \in \mathbb{R}^n_+$ corresponding to a noisy realization of an underlying signal, we consider the problem of removing the noise and recovering the original, uncorrupted signal y^*. A successful recovery of the signal requires exploiting *prior* knowledge on the structure and characteristics of the signal effectively.

A common prior knowledge on the underlying signal is *smoothness.* Smoothing considerations can be incorporated in denoising problems through quadratic penalties for deviations in successive estimates [69]. In particular, denoising of a smooth signal can be done by solving an optimization problem of the form

$$\min_{x \in \mathbb{R}^n_+} \|y - x\|_2^2 + \lambda \|Px\|_2^2, \tag{1}$$

where x corresponds to the estimation for y^*, $\lambda > 0$ is a smoothing regularization parameter, $P \in \mathbb{R}^{m \times n}$ is a linear operator, the estimation error term $\|y-x\|_2^2$ measures the *fitness* to data, and the quadratic penalty term $\|Px\|_2^2$ models the smoothness considerations. In its simplest form

$$\|Px\|_2^2 = \sum_{\{i,j\} \in A} (x_i - x_j)^2, \tag{2}$$

where A encodes the notion of adjacency, e.g., consecutive observations in a time series or adjacent pixels in an image. If P is given according to (2), then problem (1) is a convex *Markov Random Fields* problem [46] or *metric labeling problem* [54], commonly used in the image segmentation context [19, 55] and in clustering [48], for which efficient combinatorial algorithms exist. Even in its general form, (1) is a convex quadratic optimization, for which a plethora of efficient algorithms exist.

Another naturally occurring signal characteristic is *sparsity*, i.e., the underlying signal differs from a base value in only a small proportion of the indexes. Sparsity arises in diverse application domains including medical imaging [58], genomic studies [49], face recognition [89], and is at the core of *compressed sensing* methods [30]. In fact, the “bet on sparsity” principle [38] calls for systematically assuming sparsity in high-dimensional statistical inference problems. Sparsity constraints can be modeled using the ℓ_0-“norm”[1], leading to estimation problems of the form

$$\min_{x \in \mathbb{R}^n_+} \|y - x\|_2^2 + \lambda \sum_{\{i,j\} \in A} (x_i - x_j)^2 \text{ subject to } \|x\|_0 \le k, \tag{3}$$

where $k \in \mathbb{Z}_+$ is a target sparsity and $\|x\|_0 = \sum_{i=1}^n \mathbb{1}_{x_i \neq 0}$, where $\mathbb{1}_{(\cdot)}$ is the indicator function equal to 1 if $(\cdot)$ is true and equal to 0 otherwise.

[1] The so-called ℓ_0-“norm” is not a proper norm as it violates homogeneity.

In addition, the indicators can also be used to model *affine sparsity constraints* [27, 28], enforcing more sophisticated priors than simple sparsity; see Section 5.2 for an illustration.

Unlike (1), problem (3) is *non-convex* and hard-to-solve exactly. The regularized version of (3), given by

$$\min_{x\in\mathbb{R}^n_+} \|y-x\|_2^2+\lambda\sum_{\{i,j\}\in A}(x_i-x_j)^2+\mu\|x\|_0 \tag{4}$$

with $\mu \geq 0$, has received (slightly) more attention. Problem (4) corresponds to a Markov Random Fields problem with non-convex deviation functions [see 1, 47], for which a pseudo-polynomial combinatorial algorithm of complexity $O\left(\frac{|A|n}{\epsilon^2}\log\left(\frac{n^2}{\epsilon|A|}\right)\right)$ exists, where ϵ is a precision parameter and $|A|$ is the cardinality of set A; to the best of our knowledge, this algorithm has not been implemented to date. More recently, in the context of signal denoising, Bach [10] proposed another pseudo-polynomial algorithm of complexity $O\left(\left(\frac{n}{\epsilon}\right)^3\log\left(\frac{n}{\epsilon}\right)\right)$, and demonstrated its performance for instances with $n = 50$. The aforementioned algorithms rely on a discretization of the x variables, and their performance depends on how precise the discretization (given by the parameter ϵ) is. Finally, a recent result of Atamtürk and Gómez [6] on quadratic optimization with M-matrices and indicators imply that (4) is equivalent to a submodular minimization problem, which leads to a strongly polynomial-time algorithm of complexity $O(n^7)$. The high complexity by a blackbox submodular minimization algorithm precludes its use except for small instances. No polynomial-time algorithm is known for the constrained problem (3).

In fact, problems (3) and (4) are rarely tackled directly. One of the most popular techniques used to tackle signal estimation problems with sparsity consists of replacing the non-convex term $\|x\|_0$ with the convex ℓ_1-norm, $\|x\|_1 = \sum_{i=1}^n |x_i|$, see Section 2.1 for details. The resulting optimization problems with the ℓ_1-norm can be solved very efficiently, even for large instances; however, the ℓ_1 problems are often weak relaxations of the exact ℓ_0 problem (3), and the estimators obtained may be poor, as a consequence. Alternatively, there is a increasing effort for solving the mixed-integer optimization (MIO) (3) exactly using enumerative techniques, see Section 2.2. While the recovered signals are indeed high quality, exact MIO approaches to-date require at least a few days to solve instances with $n \geq 1,000$, and are inadequate to tackle many realistic instances as a consequence.

Contributions and outline. In this paper, we discuss how to bridge the gap between the easy-to-solve ℓ_1 approximations and the often intractable ℓ_0 problems in a convex optimization framework. Specifically, we construct a set of *iterative convex relaxations* for problems (3) and (4) with increasing strength. These convex relaxations are considerably stronger than the ℓ_1 relaxation, and also significantly improve and generalize other existing convex relations in the literature, including the *perspective relaxation* (see Section 2.3) and recent convex relaxations obtained from simple pairwise quadratic terms (see Section 2.4). The strong convex relaxations can be used to obtain high quality, if not optimal, solutions for (3)–(4), resulting in better performance than the existing methods; in our computations, solutions to instances with $n = 1{,}000$ are obtained with off-the-shelf convex solvers within seconds. For additional scalability, we give an easy-to-parallelize tailored Lagrangian decomposition method that solves instances with $n = 100{,}000$ under one minute. Finally, the proposed formulations are amenable to *conic quadratic* optimization techniques, thus can be tackled using off-the-shelf solvers, resulting in several advantages: *(i)* the methods described here will benefit from the continuous improvements of conic quadratic optimization solvers; *(ii)* the proposed approach is flexible, as it can be used to tackle either (3) or (4), as well as general affine sparsity constraints, by simply changing the objective or adding constraints.

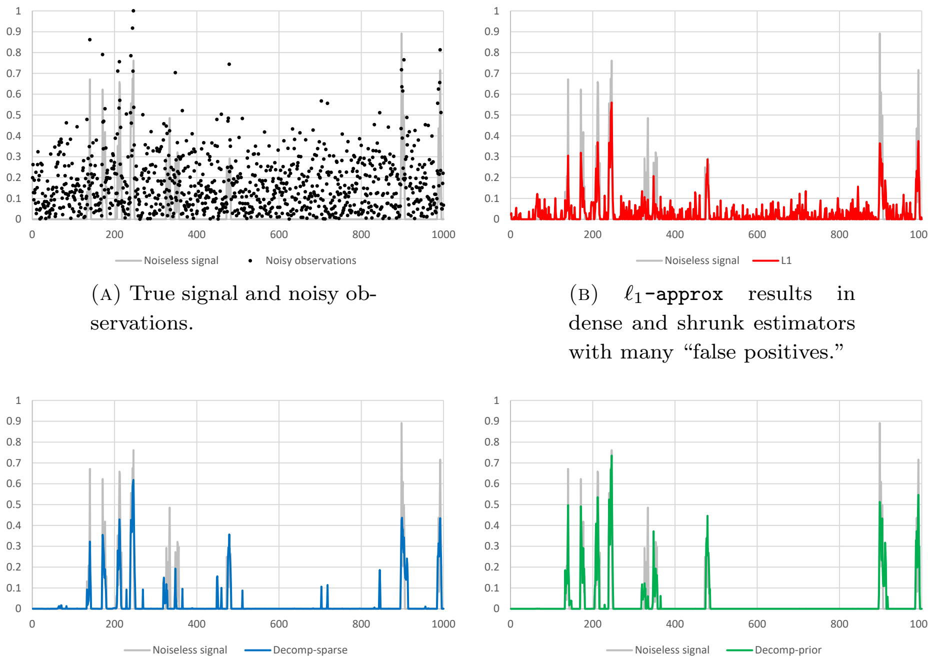
Figure 1 illustrates the performance of the ℓ_1-norm estimator and the proposed strong convex estimators for an instance with $n = 1{,}000$. The new convex estimator, depicted in Figure 1(C), requires only one second to solve; the convex estimator enhanced with additional priors in Figure 1(D) is solved under five seconds.

The rest of the paper is organized as follows. In Section 2 we review the relevant background for the paper. In Section 3 we introduce the strong iterative convex formulations for (3)–(4). In Section 4 we give conic quadratic extended reformulation of the model and describe a scalable Lagrangian decomposition method to solve it. In Section 5 we test the performance of the methods from a computational and statistical perspective, and in Section 6 we conclude the paper with a few final remarks.

Notation. Throughout the paper, we adopt the following convention for division by 0: given $a \geq 0$, $a/0 = \infty$ if $a > 0$ and $a/0 = 0$ if $a = 0$. For a set $X \subseteq \mathbb{R}^n$, let $\text{conv}(X)$ denote the convex hull of X and $\overline{\text{conv}}(X)$ the closure of $\text{conv}(X)$. Given two matrices Q, R of the same dimensions, we denote by $\langle Q, R \rangle$ the inner product of Q and R.

2. Background

In this section, we review formulations relevant to our discussion. First we review the usual ℓ_1-norm approximation (Section 2.1), next we discuss MIO formulations (Section 2.2), then we review the perspective reformulation, a



(A) True signal and noisy observations.

(B) ℓ_1-`approx` results in dense and shrunk estimators with many "false positives."

(C) New strong convex formulation yields better sparse estimators with few "false positives."

(D) Incoporating additional priors further improves the estimators, matching the sparsity pattern of the signal.

FIGURE 1. Estimators from ℓ_1-approximation and the new strong convex formulations (decomp) for signal denoising.

standard technique in the MIO literature, (Section 2.3), and finally pairwise convex relaxations that were recently proposed (Section 2.4).

2.1. **ℓ_1-norm approximations.** A standard technique for signal estimation problems with sparsity is to replace the ℓ_0-norm with the ℓ_1-norm in (3), leading to the convex optimization problem

$$(\ell_1\texttt{-approx}) \qquad \min_{x\in\mathbb{R}^n_+} \|y-x\|_2^2+\lambda\sum_{\{i,j\}\in A}(x_i-x_j)^2 \text{ subject to } \|x\|_1\leq k. \quad (5)$$

The ℓ_1-norm approximation was proposed by Tibshirani [76] in the context of sparse linear regression, and is often referred to as lasso. The main motivation for the ℓ_1-approximation is that the ℓ_1-norm is the convex p-norm closest to the ℓ_0-norm. In fact, for $L=\{x\in[0,1]^n : \|x\|_0\leq 1\}$, it is easy to show that $\text{conv}(L)=\{x\in[0,1]^n : \|x\|_1\leq 1\}$; therefore, the ℓ_1-norm approximation is considered to be the best possible convex relaxation of the ℓ_0-norm.

The ℓ_1-approximation is currently the most commonly used approach for sparsity [40]. It has been applied to a variety of signal estimation problems including signal decomposition and spike detection [e.g., 24, 34, 84, 56], and pervasive in the compressed sensing literature [20, 21, 31]. A common variant is the fused lasso [78], which involves a sparsity-inducing term of the form $\sum_{i=1}^{n-1} |x_{i+1} - x_i|$; the fused lasso was further studied in the context of signal estimation [71], and is often used for digital imaging processing under the name of *total variation denoising* [72, 83, 66]. Several other generalizations of the ℓ_1-approximation exist [77], including the elastic net [94, 65], the adaptive lasso [93], the group lasso [11, 70] and the smooth lasso [43]; related ℓ_1-norm techniques have also been proposed for signal estimation, see [52, 60, 79]. The generalized lasso [80] utilizes the regularization term $\|Ax\|_1$ and is also studied in the context of signal approximation.

Despite its widespread adoption, the ℓ_1-approximation has several drawbacks. First, the ℓ_1-norm term may result in excessive *shrinkage* of the estimated signal, which is undesirable in many contexts [90]. Additionally, the ℓ_1-approximation may struggle to achieve sparse estimators — in fact, solutions to (5) are often dense, and achieving a target sparsity of k requires using a parameter $\hat{k} << k$, inducing additional bias on the estimators. As a consequence, desirable theoretical performance of the ℓ_1-approximation can only be established under stringent conditions [71, 73], which may not be satisfied in practice. Indeed, ℓ_1-approximations have been shown to perform rather poorly in a variety of contexts, e.g., see [51, 63]. To overcome the aforementioned drawbacks, several non-convex approximations have been proposed [33, 41, 61, 91, 92]; more recently, there is also an increasing effort devoted to enforcing sparsity directly with ℓ_0 regularization using enumerative MIO approaches.

2.2. **Mixed-integer optimization.** Signal estimation problems with sparsity can be naturally modeled as a mixed-integer quadratic optimization (MIQO) problem. Using indicator variables $z \in \{0,1\}^n$ such that $z_i = \mathbb{1}_{x_i \neq 0}$ for all $i = 1, \ldots, n$, problem (3) can be formulated as

$$\min \sum_{i=1}^{n} (y_i - x_i)^2 + \lambda \sum_{\{i,j\} \in A} (x_i - x_j)^2 \tag{6a}$$

$$\text{s.t. } x_i(1 - z_i) = 0 \tag{6b}$$

$$z \in C \subseteq \{0,1\}^n \tag{6c}$$

$$x \in \mathbb{R}^n_+. \tag{6d}$$

If C is defined by a k-sparsity constraint, i.e., $C = \{z \in \{0,1\}^n : \|z\|_1 \leq k\}$, then problem (6) is the ℓ_0 analog of (5). More generally, C may be defined by other logical (affine sparsity) constraints, which allow the inclusion of additional priors in the inference problem. In this formulation, the nonconvexity of the ℓ_0 regularizer is captured by the complementary constraints (6b) and the binary constraints encoded by set C. Constraints (6b) can be alternatively formulated with the so-called "big-M" constraints with a sufficiently large positive number u,

$$x_i(1 - z_i) = 0 \text{ and } z_i \in \{0, 1\} \Leftrightarrow x_i \leq uz_i \text{ and } z_i \in \{0, 1\}. \tag{7}$$

For the signal estimation problem (6), $u = \|y\|_\infty$ is a valid upper bound for x_i, $i = 1, \dots, n$. Problem (6) is a convex MIQO problem, which can be tackled using off-the-shelf MIO solvers. Estimation problems with a few hundred of variables can be comfortably solved to optimality using such solvers, e.g., see [14, 25, 36, 86]. For high Signal-to-Noise Ratios (SNR), the estimators obtained from solving the exact ℓ_0 problems indeed result in superior statistical performance when compared with the ℓ_1 approximations [15]. For low SNR, however, the lack of *shrinkage* may hamper the estimators obtained from optimal solutions of the ℓ_0 problems [39]; nonetheless, if necessary, shrinkage can be easily added to (6) via conic quadratic regularizations terms [62], resulting again in superior statistical performance over corresponding ℓ_1-approximations. Unfortunately, current MIO solvers are unable to solve larger problems with thousands of variables.

A recent research thrust aims to use MIO formulations and heuristics for sparse learning problems such as (6) [41, 88], which scale to larger instances than MIO solvers but may not provide dual bounds on the quality of the solutions found. Another research direction aims to design tailored exact methods for specialized regression problems [7, 16, 42, 53], which perform substantially better than general-purpose MIO solvers for the problems their are designed to tackle, but may not generalize well to other learning problems; in particular, general constraints such as (6c) are challenging to handle via tailored algorithms.

Finally, we point out the relationship between the ℓ_1-approximation (5) and the MIO formulation (6). It can be verified easily that, if C is defined by a k-sparsity constraint, then there exists an optimal solution z to the simple convex relaxation with big-M constraint, where $z_i = \frac{x_i}{u}$ for all $i = 1, \dots, n$. Therefore, the constraint (6c) reduces to $\|x\|_1 \leq ku$, and we find that (5) is in fact the natural convex relaxation of (6) (for a suitable sparsity parameter). This relaxation is often weak and can be improved substantially.

2.3. **The perspective reformulation.**

A simple strengthening technique to improve the convex relaxation of (6) is the `perspective reformulation` [32], which will be referred to as `persp` in the remainder of the paper for brevity. This reformulation technique can be applied to the estimation error terms in (6a) as follows:

$$\begin{aligned}(y_i - x_i)^2 \leq t \;\Leftrightarrow\; & y_i^2 - 2y_ix_i + x_i^2 \leq t\\ \rightarrow\; & y_i^2 - 2y_ix_i + \frac{x_i^2}{z_i} \leq t. \end{aligned} \tag{8}$$

The term x_i^2/z_i is the closure of the perspective function of the quadratic function x_i^2, and is therefore convex, see p. 160 of [45]. Reformulation (8) is in fact the best possible for *separable* quadratic functions with indicator variables. The perspective terms $\frac{x_i^2}{z_i}$ can be replaced with an auxiliary variable s_i along with rotated cone constraints $x_i^2 \leq s_iz_i$ [2, 37]. Therefore, `persp` relaxations can be easily solved with conic quadratic solvers and is by now a standard technique for mixed-integer quadratic optimization [18, 44, 59, 87]. Additionally, relationships between the `persp` and the sparsity-inducing non-convex penalty functions `minimax concave penalty` [90] and `reverse Huber penalty` [67] have recently been established [29]. In the context of the signal estimation problem (3), the `persp` yields the convex relaxation

$$\begin{aligned} & \sum_{i=1}^n y_i^2 + \min \sum_{i=1}^n (-2y_ix_i + \frac{x_i^2}{z_i}) + \lambda \sum_{\{i,j\}\in A} (x_i - x_j)^2 \\ (\texttt{persp.}) \quad & \text{s.t. } x_i \leq \|y\|_\infty z_i \qquad\qquad i = 1,\dots,n \\ & z \in \bar{C},\; x \in \mathbb{R}^n_+, \end{aligned}$$

where $\bar{C}$ is a valid convex relaxation of C, e.g., $\bar{C} = \text{conv}(C)$. The ℓ_1-approximation model, as discussed in Section 2.1, is the best convex relaxation that considers only the indicators for the ℓ_0 terms. The `persp` approximation is the best convex relaxation that exploits the ℓ_0 indicator variables as well as the separable quadratic estimation error terms; thus, it is stronger than the ℓ_1-approximation. However, `persp` cannot be applied to non-separable quadratic smoothness terms $(x_i - x_j)^2$, as the function $x_i^2/z_i - 2x_ix_j + x_j^2/z_j$ is non-convex due to the bilinear term.

2.4. **Strong formulations for pairwise quadratic terms.** Recently, Jeon et al. [50] gave strong relaxations for the mixed-integer epigraphs of non-separable convex quadratic functions with two variables and indicator variables. Atamtürk and Gómez [6] further strengthened the relaxations for quadratic functions of the form $(x_i - x_j)^2$ corresponding to the smoothness terms in (6). Specifically, let

$$X^2 = \left\{(z,x,s) \in \{0,1\}^2 \times \mathbb{R}^3_+ : (x_1 - x_2)^2 \leq s,\; x_i(1 - z_i) = 0, i = 1,2\right\}$$

and define the function $f : [0,1]^2 \times \mathbb{R}^2_+ \to \mathbb{R}_+$ as

$$f(z,x) = \begin{cases} \frac{(x_1-x_2)^2}{z_1} & \text{if } x_1 \geq x_2 \\ \frac{(x_1-x_2)^2}{z_2} & \text{if } x_1 \leq x_2. \end{cases}$$

Proposition 1 (Atamtürk and Gómez [6]). *The function f is convex and $\overline{conv}(X^2) = \left\{(z,x,s) \in [0,1]^2 \times \mathbb{R}^3_+ : f(z,x) \leq s\right\}$.*

Using `persp` and Proposition 1, one obtains the stronger `pairwise` convex relaxation of (6) as

$$\sum_{i=1}^n y_i^2 + \min \sum_{i=1}^n \left(-2y_i x_i + \frac{x_i^2}{z_i}\right) + \lambda \sum_{\{i,j\}\in A} f(z_i, z_j, x_i, x_j) \tag{9a}$$

$$(\texttt{pairwise}\ ) \qquad \text{s.t. } x_i \leq \|y\|_\infty z_i, \qquad i = 1,\dots,n \tag{9b}$$

$$z \in \bar{C},\ x \in \mathbb{R}^n_+. \tag{9c}$$

Note that f is not differentiable everywhere and it is defined by pieces. Therefore, it cannot be used directly with most convex optimization solvers. Atamtürk and Gómez [6] implement (9) using linear outer approximations of function f: the resulting method performs adequately for instances with $n \leq 400$, but was ineffective in instances with $n \geq 1,000$ as strong linear outer approximations require the addition of a large number of constraints. Moreover, as Example 1 below shows, formulation (9) can be further improved even for $n = 2$.

Example 1. Consider the signal estimation problem (44) with $n = 2$

$$\min\ (0.4 - x_1)^2 + (1 - x_2)^2 + 0.5(x_1 - x_2)^2 + 0.5\,(z_1 + z_2) \tag{10a}$$

$$\text{s.t. } x_i \leq z_i,\ i = 1,2 \tag{10b}$$

$$z \in \{0,1\}^2, x \in \mathbb{R}^2_+. \tag{10c}$$

The optimal solution of (10) is $(z_1^*, z_2^*, x_1^*, x_2^*) = (0.00, 1.00, 0.00, 0.67)$. On the other hand, optimal solutions of the convex relaxations of (10) are:

- **ℓ_1-approx:** Obtained by replacing $z \in \{0,1\}^2$ with $z \in [0,1]^2$. The corresponding optimal solution is $(z_\ell, x_\ell) = (0.30, 0.60, 0.30, 0.60)$, and we find that $\|(z^*, x^*) - (z_\ell, x_\ell)\|_2 = 0.59$.
- **`persp`:** The optimal solution is $(z_p, x_p) = (0.00, 0.82, 0.00, 0.59)$, and $\|(z^*, x^*) - (z_p, x_p)\|_2 = 0.19$.
- **`pairwise`:** The optimal solution is $(z_q, x_q) = (0.11, 1.00, 0.08, 0.69)$, and $\|(z^*, x^*) - (z_q, x_q)\|_2 = 0.14$.

Although `persp` and `pairwise` substantially improve upon the ℓ_1-relaxation, the resulting solutions are still not integral in z. We will give the convex hull of (10) in the next section. □

In this paper, we show how to further improve the `pairwise` formulation to obtain a stronger relaxation of (6). Additionally, we show how to implement the relaxations derived in the paper in a conic quadratic optimization framework. Therefore, the proposed convex relaxations benefit from a growing literature on conic quadratic optimization, e.g., see [3, 5, 8, 57, 64], can be implemented with off-the-shelf solvers, and scale to large instances.

3. Strong convex formulations for signal estimation

In the `pairwise` formulation each single- and two-variable quadratic term is strengthened independently and, consequently, the formulation fails to fully exploit the relationships between different pairs of variables. Observe that problem (6) can be stated as

$$\|y\|_2^2 + \min\ -2y'x + x'Qx \tag{11a}$$

$$\text{s.t. } x_i(1-z_i) = 0, i = 1\ldots,n, \tag{11b}$$

$$z \in C, x \in \mathbb{R}_+^n \tag{11c}$$

where, for $i \neq j$, $Q_{ij} = -\lambda$ if $\{i,j\} \in A$ and $Q_{ij} = 0$ otherwise, and $Q_{ii} = 1 + \lambda|A_i|$ where $A_i = \{j : \{i,j\} \in A\}$. In particular, Q is a symmetric M-matrix, i.e., $Q_{ij} \leq 0$ for $i \neq j$ and $Q \succeq 0$. In this section we derive convex relaxations of (6) that better exploit the M-matrix structure. We briefly review properties of M-matrices and refer the reader to [13, 35, 68, 81] and the references therein for an in-depth discussion on M-matrices.

Proposition 2 (Plemmons [68], characterization 37)**.** *An M-matrix is generalized diagonally dominant, i.e., there exists a positive diagonal matrix D such that DQ is (weakly) diagonally dominant.*

Generalized diagonally dominant matrices are also called scaled diagonally dominant matrices in the literature.

Proposition 3 (Boman et al. [17])**.** *A matrix Q is generalized diagonally dominant iff it has factor width at most two, i.e., there exists a real matrix $V_{n\times m}$ such that $Q = VV^\top$ and each column of V contains at most two non-zeros.*

Proposition 3 implies that if Q is an M-matrix, then the quadratic function $x'Qx$ can be written as a sum of quadratic functions of at most two variables each, i.e., $x'Qx = \sum_{j=1}^m \left(\sum_{i=1}^n V_{ij}x_i\right)^2$ where for any j at most two entries V_{ij} are non-zero. Therefore, to derive stronger formulations for (11), we first study the mixed-integer epigraphs of *parametric* pairwise quadratic functions with indicators.

3.1. **Convexification of the parametric pairwise terms.** Consider the mixed-integer epigraph of a parametric pairwise quadratic term (with parameters d_1, d_2)

$$Z^2 = \Big\{(z,x,s) \in \{0,1\}^2 \times \mathbb{R}^3_+ : d_1x_1^2 - 2x_1x_2 + d_2x_2^2 \leq s, \\ x_i(1-z_i) = 0,\ i = 1,2\Big\},$$

where $d_1d_2 \geq 1$ and $d_1, d_2 > 0$, which is the necessary and sufficient condition for convexity of the function $d_1x_1^2 - 2x_1x_2 + d_2x_2^2$. One may, without loss of generality, assume the cross-product coefficient equals -2, as otherwise the continuous variables and coefficients can be scaled. Clearly, if $d_1 = d_2 = 1$, then Z^2 reduces to X^2.

Consider the two decompositions of the two-variable quadratic function in the definition of Z^2 given by

$$\begin{aligned} d_1x_1^2 - 2x_1x_2 + d_2x_2^2 &= d_1\left(x_1 - \frac{x_2}{d_1}\right)^2 + x_2^2\left(d_2 - \frac{1}{d_1}\right) \\ &= d_2\left(\frac{x_1}{d_2} - x_2\right)^2 + x_1^2\left(d_1 - \frac{1}{d_2}\right). \end{aligned}$$

Intuitively, the decompositions above are obtained by extracting a term $\delta_i x_i^2$ from the quadratic function such that δ_i is as large as possible and the remainder quadratic term is still convex. Then, applying `persp` and Proposition 1 to the separable and pairwise quadratic terms, respectively, one obtains two valid inequalities for Z^2:

$$d_1 f(z_1, z_2, x_1, \frac{x_2}{d_1}) + \frac{x_2^2}{z_2}\left(d_2 - \frac{1}{d_1}\right) \leq s \tag{12}$$

$$d_2 f(z_1, z_2, \frac{x_1}{d_2}, x_2) + \frac{x_1^2}{z_1}\left(d_1 - \frac{1}{d_2}\right) \leq s. \tag{13}$$

Clearly, there are infinitely many such decompositions depending on the values of δ_i, $i = 1, 2$. Surprisingly, Theorem 1 below shows that inequalities (12)–(13) along with the bound constraints are sufficient to describe $\overline{\text{conv}}(Z^2)$.

Theorem 1. $\overline{conv}(Z^2) = \left\{(z,x,s) \in [0,1]^2 \times \mathbb{R}^3_+ : (12) - (13)\right\}.$

Proof. Consider the mixed-integer optimization problem

$$\min_{(z,x,s)\in Z^2} a_1z_1 + a_2z_2 + b_1x_1 + b_2x_2 + \lambda s \tag{14}$$

and the corresponding convex optimization

$$\min a_1z_1+a_2z_2+b_1x_1+b_2x_2+\lambda s \tag{15a}$$

$$\text{s.t. } d_1f(z_1,z_2,x_1,\frac{x_2}{d_1})+\frac{x_2^2}{z_2}\left(d_2-\frac{1}{d_1}\right)\leq s \tag{15b}$$

$$d_2f(z_1,z_2,\frac{x_1}{d_2},x_2)+\frac{x_1^2}{z_1}\left(d_1-\frac{1}{d_2}\right)\leq s \tag{15c}$$

$$z\in[0,1]^2,\ x\in\mathbb{R}_+^2,\ s\in\mathbb{R}_+. \tag{15d}$$

To prove the result it suffices to show that, for any value of (a,b,λ), either (14) and (15) are both unbounded, or that (15) has an optimal solution that is also optimal for (14). We assume, without loss of generality, that $d_1d_2>1$ (if $d_1d_2=1$, the result follows from Proposition 1 by scaling), $\lambda>0$ (if $\lambda<0$, both problems are unbounded by letting $s\to\infty$, and if $\lambda=0$, problem (15) reduces to linear optimization over a integral polytope and optimal solutions are integral in z), and $\lambda=1$ (by scaling). Moreover, since $d_1d_2>1$, there exists an optimal solution for both (14) and (15).

Let (z^*,x^*,s^*) be an optimal solution of (15); we show how to construct from (z^*,x^*,s^*) a feasible solution for (14) with same objective value, thus optimal for both problems. Observe that for $\gamma\geq 0$, $f(\gamma z_1,\gamma z_2,\gamma x_1,\gamma x_2)=\gamma f(z_1,z_2,x_1,x_2)$. Thus, if $z_1^*,z_2^*<1$, then $(\gamma z^*,\gamma x^*,\gamma s^*)$ is also feasible for (15) with objective value $\gamma\left(a_1z_1^*+a_2z_2^*+b_1x_1^*+b_2x_2^*+s^*\right)$. In particular, either there exists an (integral) optimal solution with $z^*=x^*=0$ by setting $\gamma=0$, or there exists an optimal solution with one of the z variables equal to one by increasing γ. Thus, assume without loss of generality that $z_1^*=1$. Now consider the optimization problem

$$\min a_2z_2+b_1x_1+b_2x_2+d_1f(1,z_2,x_1,\frac{x_2}{d_1})+\frac{x_2^2}{z_2}\left(d_2-\frac{1}{d_1}\right) \tag{16a}$$

$$z_2\in[0,1],\ x\in\mathbb{R}_+^2, \tag{16b}$$

obtained from (15) by fixing $z_1=1$, dropping constraint (15c), and eliminating variable s since (15b) holds at equality in optimal solutions. An integer optimal solution for (16) is also optimal for (14) and (15). Let $(\hat{z},\hat{x})$ be an optimal solution for (16), and consider the two cases:

Case 1: $\hat{x}_1\leq\hat{x}_2/d_1$: If $0<\hat{z}_2<1$, then the point $(\gamma\hat{z}_2,\gamma\hat{x}_1,\gamma\hat{x}_2)$ with $0\leq\gamma\hat{z}_2\leq 1$ is feasible for (16) with objective value

$$\gamma\left(a_2\hat{z}_2+b_1\hat{x}_1+b_2\hat{x}_2+d_1f(1,\hat{z}_2,\hat{x}_1,\frac{\hat{x}_2}{d_1})+\frac{\hat{x}_2^2}{\hat{z}_2}\left(d_2-\frac{1}{d_1}\right)\right).$$

Therefore, there exists an optimal solution where $\hat{z}_2\in\{0,1\}$. □

Case 2: $\hat{x}_1 > \hat{x}_2/d_1$: In this case, $(\hat{z}_2, \hat{x}_1, \hat{x}_2)$ is an optimal solution of

$$\min\ a_2z_2 + b_1x_1 + b_2x_2 + d_1\left(x_1 - \frac{x_2}{d_1}\right)^2 + \frac{x_2^2}{z_2}\left(d_2 - \frac{1}{d_1}\right) \tag{17a}$$

$$z_2 \in [0,1],\ x \in \mathbb{R}_+^2. \tag{17b}$$

The condition $\hat{x}_1 > \hat{x}_2/d_1$ implies that $\hat{x}_1 > 0$, thus the optimal value of x_1 can be found by taking derivatives and setting to 0. We find

$$\hat{x}_1 = -\frac{b_1}{2d_1} + \frac{x_2}{d_1}\cdot$$

Replacing x_1 with his optimal value in (17) and removing constant terms, we find that (17) is equivalent to

$$\min\ a_2z_2 + \left(\frac{b_1}{d_1} + b_2\right)x_2 + \frac{x_2^2}{z_2}\left(d_2 - \frac{1}{d_1}\right) \tag{18a}$$

$$z_2 \in [0,1],\ x_2 \in \mathbb{R}_+. \tag{18b}$$

If $0 < \hat{z}_2 < 1$, then the point $(\gamma\hat{z}_2, \gamma\hat{x}_2)$ with $0 \leq \gamma\hat{z}_2 \leq 1$ is feasible for (18) with objective value

$$\gamma\left(a_2\hat{z}_2 + \left(\frac{b_1}{d_1} + b_2\right)\hat{x}_2 + \frac{\hat{x}_2^2}{\hat{z}_2}\left(d_2 - \frac{1}{d_1}\right)\right).$$

Therefore, there exists an optimal solution where $\hat{z}_2 \in \{0,1\}$. □

In both cases we find an optimal solution with $z_2 \in \{0,1\}$. Thus, problem (15) has an optimal solution integral in both z_1 and z_2, which is also optimal for (14). □

Example 1 (continued). The relaxation of (10) with only inequality (13):

$$\begin{aligned} 1.16 + \min\ & -0.8x_1 - 2x_2 + 0.5\,(z_1 + z_2) + 0.5s \\ \text{s.t. } & 3f(z_1, z_2, \frac{x_1}{3}, x_2) + \frac{x_1^2}{z_1}\left(3 - \frac{1}{3}\right) \leq s \\ & z \in [0,1]^2, x \in \mathbb{R}_+^2, \end{aligned}$$

is sufficient to obtain the integral optimal solution. Note that the big-M constraints $x_i \leq z_i$ are not needed. □

Given $d_1, d_2 \in \mathbb{R}_+$, define the function $g : [0,1]^2 \times \mathbb{R}_+^2 \to \mathbb{R}_+$ as

$$\begin{aligned} g(z_1, z_2, x_1, x_2; d_1, d_2) = \max\Bigg\{ & d_1 f(z_1, z_2, x_1, \frac{x_2}{d_1}) + \frac{x_2^2}{z_2}\left(d_2 - \frac{1}{d_1}\right), \\ & d_2 f(z_1, z_2, \frac{x_1}{d_2}, x_2) + \frac{x_1^2}{z_1}\left(d_1 - \frac{1}{d_2}\right)\Bigg\}\cdot \end{aligned} \tag{19}$$

For any $d_1, d_2 > 0$ with $d_1 d_2 \geq 1$, function g is the point-wise maximum of two convex functions and is therefore convex. Using the convex function g, Theorem 1 can be restated as

$$\overline{\text{conv}}(Z^2) = \left\{(z,x,s) \in [0,1]^2 \times \mathbb{R}^3_+ : g(z_1,z_2,x_1,x_2;d_1,d_2) \leq s\right\}.$$

Finally, it is easy to verify that if $z_1 \geq z_2$, then the maximum in (19) corresponds to the first term; if $z_1 \leq z_2$, the maximum corresponds to the second term. Thus, an explicit expression of g is

$$g(z,x;d) = \begin{cases} \frac{d_1x_1^2 - 2x_1x_2 + x_2^2/d_1}{z_1} + \frac{x_2^2}{z_2}\left(d_2 - \frac{1}{d_1}\right) & \text{if } z_1 \geq z_2 \text{ and } d_1x_1 \geq x_2 \\ \frac{d_1x_1^2 - 2x_1x_2 + d_2x_2^2}{z_2} & \text{if } z_1 \geq z_2 \text{ and } d_1x_1 \leq x_2 \\ \frac{d_1x_1^2 - 2x_1x_2 + d_2x_2^2}{z_1} & \text{if } z_1 \leq z_2 \text{ and } x_1 \geq d_2x_2 \\ \frac{x_1^2/d_2 - 2x_1x_2 + d_2x_2^2}{z_2} + \frac{x_1^2}{z_1}\left(d_1 - \frac{1}{d_2}\right) & \text{if } z_1 \leq z_2 \text{ and } x_1 \leq d_2x_2. \end{cases}$$

3.2. **Convex relaxations for general M-matrices.** Consider the set

$$Z^n = \left\{(z,x,t) \in \{0,1\}^n \times \mathbb{R}^{n+1}_+ : x'Qx \leq t,\ x_i(1-z_i) = 0, i = 1,\dots,n\right\},$$

where Q is an M-matrix. In this section, we will show how the convex hull descriptions for Z^2 can be used to construct strong convex relaxations for Z^n. We start with the following motivating example.

Example 2. Consider the signal estimation in regularized form with $n = 3$, $(y_1, y_2, y_3) = (0.3, 0.7, 1.0)$, $\lambda = 1$ and $\mu = 0.5$,

$$\zeta = 1.58 + \min\ -0.6x_1 - 1.4x_2 - 2.0x_3 + t + 0.5\,(z_1 + z_2 + z_3) \tag{20a}$$

$$\text{s.t. } x_1^2 + x_2^2 + x_3^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 \leq t \tag{20b}$$

$$x_i \leq z_i, \qquad i = 1,2,3 \tag{20c}$$

$$z \in \{0,1\}^3,\ x \in \mathbb{R}^3_+. \tag{20d}$$

The optimal solution of (20) is $(z^*, x^*) = (0.00, 1.00, 1.00, 0.00, 0.48, 0.74)$ with objective value $\zeta^* = 1.504$. The optimal solutions and the corresponding objective values of the convex relaxations of (20) are as follows:

ℓ_1-**approx**: The opt. solution is $(z_\ell, x_\ell) = (0.24, 0.43, 0.59, 0.24, 0.43, 0.59)$ with value $\zeta_{\ell_1\text{-approx}} = 0.936$, and $\|(z^*, x^*) - (z_\ell, x_\ell)\|_2 = 0.80$.

persp: The opt. solution is $(z_p, x_p) = (0.00, 0.40, 0.82, 0.00, 0.29, 0.58)$ with value $\zeta_{\texttt{persp}} = 1.413$, and $\|(z^*, x^*) - (z_p, x_p)\|_2 = 0.67$.

pairwise: The opt. solution $(z_q, x_q) = (0.18, 0.74, 1.00, 0.13, 0.43, 0.71)$ with value $\zeta_{\texttt{pairwise}} = 1.488$, and $\|(z^*, x^*) - (z_q, x_q)\|_2 = 0.35$.

`decomp.1`**:** The quadratic constraint (20b) can be decomposed and strengthened as follows:

$$\begin{aligned}&\left(2x_1^2-2x_1x_2+x_2^2\right)+\left(2x_2^2-2x_2x_3+2x_3^2\right)\leq t\\ \to\ &g(z_1,z_2,x_1,x_2;2,1)+g(z_2,z_3,x_2,x_3;2,2)\leq t;\end{aligned}$$

leading to solution is $(z_d,x_d)=(0.17,1.00,0.93,0.12,0.53,0.73)$ with value $\zeta_{\texttt{decomp.1}}=1.495$, and $\|(z^*,x^*)-(z_d,x_d)\|_2=0.23$.

`decomp.2`**:** Alternatively, constraint (20b) can also be formulated as $g(z_1,z_2,x_1,x_2;2,2)+g(z_2,z_3,x_2,x_3;1,2)\leq t$, and the resulting convex relaxation has solution $(z^*,x^*)=(0.00,1.00,1.00,0.00,0.48,0.74)$, corresponding to the optimal solution of (20). □

As Example 2 shows, strong convex relaxations of Z^n can be obtained by decomposing $x'Qx$ into sums of two-variable quadratic terms (as Q is an M-matrix) and convexifying each term. However, such a decomposition is not unique and the strength of the relaxation depends on the decomposition chosen. We now discuss how to optimally decompose the matrix Q to derive the strongest lower bound possible for a fixed value of (z,x,t). Then, we show how this decomposition procedure can be embedded in a cutting surface algorithm to obtain a strong convex relaxation of (11).

Consider the *separation problem*: given a point $(z,x,t)\in[0,1]^n\times\mathbb{R}_+^{n+1}$, find a decomposition of Q such that, after strengthening each two-variable term, results in a most violated inequality, which is formulated as follows:

$$\theta(z,x)=\max_d\ \sum_{i=1}^n\sum_{j=i+1}^n|Q_{ij}|g(z_i,z_j,x_i,x_j;d_{ij}^i,d_{ij}^j)\tag{21a}$$

$$\text{s.t.}\ \sum_{j<i}|Q_{ji}|d_{ji}^i+\sum_{j>i}|Q_{ij}|d_{ij}^i=Q_{ii}\qquad\forall i=1,\dots,n\tag{21b}$$

$$d_{ij}^id_{ij}^j\geq1,\ d_{ij}^i\geq0,\ d_{ij}^j\geq0\qquad\forall i<j.\tag{21c}$$

Observe that the variables of the separation problem (21) are the parameters d, and the variables of the estimation problem (z,x) are fixed in the separation problem. In formulation (21) for each (negative) entry Q_{ij}, $i<j$, there is a two-variable quadratic term of the form $|Q_{ij}|\left(d_{ij}^ix_i^2-2x_ix_j+d_{ij}^jx_j^2\right)$; after convexifying each such term, one obtains the objective (21a). Constraints (21b) ensure that the decomposition indeed corresponds to the original matrix Q by ensuring that the diagonal elements coincide, and constraints (21c) ensure that each quadratic term is convex. From Proposition 3, problem (21) is feasible for any M-matrix Q.

For any feasible value of d, the objective (21a) is convex in (z,x); thus the function $\theta:[0,1]^n\times\mathbb{R}_+^n\to\mathbb{R}_+$ defined in (21) is a supremum of convex functions and is convex itself. Moreover, the constraints (21b) and (21c) are linear or rotated cone constraints, thus, are convex in d. As we now show, the objective function (21a) is concave in d, thus (21) is a convex optimization.

Index the variables such that $z_1 \geq z_2 \geq \ldots \geq z_n$. Then, each term in the objective (21a) reduces to

$$g(z_i,z_j,x_i,x_j;d_{ij}^i,d_{ij}^j) = \begin{cases} \frac{d_{ij}^i x_i^2 - 2x_ix_j + x_j^2/d_{ij}^i}{z_i} + \frac{x_j^2}{z_j}\left(d_{ij}^j - \frac{1}{d_{ij}^i}\right) & \text{if } d_{ij}^i x_i \geq x_j \\ \frac{d_{ij}^i x_i^2 - 2x_ix_j + d_{ij}^j x_j^2}{z_j} & \text{if } d_{ij}^i x_i \leq x_j \end{cases}$$

$$= d_{ij}^i\frac{x_i^2}{z_i} + d_{ij}^j\frac{x_j^2}{z_j} + \begin{cases} \frac{-2x_ix_j}{z_i} - \frac{x_j^2}{d_{ij}^i}\left(\frac{1}{z_j} - \frac{1}{z_i}\right) & \text{if } d_{ij}^i x_i \geq x_j \\ \frac{-2x_ix_j}{z_j} + d_{ij}^i x_i^2\left(\frac{1}{z_j} - \frac{1}{z_i}\right) & \text{if } d_{ij}^i x_i \leq x_j. \end{cases}$$

Thus, $g(z,x;d)$ is separable in d_{ij}^i and d_{ij}^j, is linear in d_{ij}^j; and, it is linear in d_{ij}^i for $d_{ij}^i \leq x_j/x_i$, and concave for $d_{ij}^i \geq x_j/x_i$. Moreover, it is easily shown that it is continuous and differentiable (i.e., the derivatives of both pieces of g with respect to d_{ij}^i coincide if $d_{ij}^i x_i = x_j$). Therefore, the separation problem (21) can be solved in polynomial time by first sorting the variables z_i and then by solving a convex optimization problem.

The separation procedure can be embedded in an algorithm that iteratively constructs stronger relaxations of problem (11).

`Simple cutting surface algorithm`:

1. Solve a valid convex relaxation.
2. Solve separation problem (21) using a convex optimization method.
3. Add the inequality obtained from solving the separation problem to the formulation, strengthening the relaxation, and go to step 1.

Below, we illustrate the `simple cutting surface algorithm`.

Example 2 (Continued)*.* Consider the `persp` relaxation

$$\zeta_1 = 1.58 + \min\ -0.6x_1 - 1.4x_2 - 2.0x_3 + t + 0.5\left(z_1 + z_2 + z_3\right) \tag{22a}$$

$$\text{s.t. } \frac{x_1^2}{z_1} + \frac{x_2^2}{z_2} + \frac{x_3^2}{z_3} + (x_1 - x_2)^2 + (x_2 - x_3)^2 \leq t \tag{22b}$$

$$x_i \leq z_i, \quad i = 1,2,3 \tag{22c}$$

$$z \in [0,1]^3,\ x \in \mathbb{R}_+^3. \tag{22d}$$

with optimal solution $(z,x)_1 = (0.00, 0.40, 0.82, 0.00, 0.29, 0.58)$ with $\zeta_1 = 1.413$ and $\|(z^*,x^*) - (z,x)_1\|_2 = 0.67$. This relaxation can be improved by solving the separation problem (21) at $(z,x)_1$ to obtain the optimal parameters $d_{12}^1 = 2.00$, $d_{12}^2 = 0.51$, $d_{23}^2 = 2.49$ and $d_{23}^3 = 2.00$, leading to the decomposition and the constraint

$$g(z_1,z_2,x_1,x_2;2.00,0.51) + g(z_2,z_3,x_2,x_3;2.49,2.00) \leq t.$$

Adding this constraint to (22) and resolving gives the improved solution $(z,x)_2 = (0.15, 0.70, 1.00, 0.12, 0.43, 0.71)$. This process can be repeated iteratively, resulting in the sequence of solutions

- `iter.2:` $(z,x)_2 = (0.15, 0.70, 1.00, 0.12, 0.43, 0.71)$ with $\zeta_2 = 1.452$ and $\|(z^*,x^*) - (z,x)_2\|_2 = 0.36$. The corresponding separation problem has solution $(d_{12}^1, d_{12}^2, d_{23}^2, d_{23}^3) = (2, 1.06, 1.94, 2)$.
- `iter.3:` $(z,x)_3 = (0.14, 1.00, 1.00, 0.10, 0.52, 0.75)$ with $\zeta_3 = 1.499$ and $\|(z^*,x^*) - (z,x)_3\|_2 = 0.18$. The corresponding separation problem has solution $(d_{12}^1, d_{12}^2, d_{23}^2, d_{23}^3) = (2, 2.5, 0.5, 2)$.
- `iter.4:` $(z,x)_4 = (0.00, 1.00, 1.00, 0.00, 0.48, 0.74)$ with $\zeta_3 = 1.504$. The solution is integral and optimal for (20). □

The iterative separation procedure outlined above ensures that (z,x,t) satisfies the convex relaxation

$$\Theta = \left\{(z,x,t) \in [0,1]^n \times \mathbb{R}_+^{n+1} : \theta(z,x) \le t\right\}$$

of Z^n that dominates the ℓ_1`-approx`, `persp`, and `pairwise` and gives the strong relaxation of problem (11), based on the optimal `decomposition` of matrix Q, given by

$$(\texttt{decomp}) \qquad \|y\|_2^2 + \min_{(z,x)\in[0,1]^n\times\mathbb{R}_+^n} -2y'x + \theta(z,x)\colon z \in \bar{C},\ x_i \le \|y\|_\infty z_i,\ i = 1\ldots,n.$$

In Section 4 we discuss the efficient implementation of `decomp` in a conic quadratic optimization framework.

4. Conic quadratic representation and Lagrangian decomposition

Relaxation `decomp` simultaneously exploits sparsity, fitness and smoothness terms in (3) and, therefore, dominates all of the relaxations discussed in Section 2. However, the convex functions f and g can be pathological, as they are defined by pieces and are not differentiable everywhere. Handling function θ is challenging as it is non-differentiable, but also it is not given in closed form and requires solving optimization problem (21) to evaluate.

In this section, we first show how to tackle `decomp` effectively by formulating it as a conic quadratic optimization problem in an extended space. We then give a tailored Lagrangian decomposition method, which is amenable to parallel computing and highly scalable.

4.1. **Extended formulations.**

The `simple cutting surface algorithm` to solve `decomp`, illustrated in Example 2, is computationally cumbersome since: *(i)* the separation problem (step 2) requires solving a constrained convex optimization problem; *(ii)* each cut added (step 3) is dense (and thus problematic for optimization software); *(iii)* a single cut is generated at each iteration; consequently, the method may require many iterations to converge.

In this section, we show how to address these shortcomings with a conic quadratic extended formulation with auxiliary variables. In particular:

(i) The separation problem can be solved in closed form (Proposition 9)– eliminating the need to solve auxiliary optimization problems.

(ii) Cuts (33) can be added as inequalities with at most four variables. Most conic quadratic optimization solvers are designed to exploit sparsity to improve performance and numerical stability.

(iii) The method may add up to $\mathcal{O}(n^2)$ cuts per round, decreasing the total number of rounds and re-optimizations. More importantly, simple cuts (e.g., sparse or linear cuts) in this extended space may translate into highly nonlinear cuts when projected into the original space of variables, often resulting in additional strength. Indeed, MIO often relies on extended formulations as such formulations lead to substaintial improvements compared to working in the original space of variables [4, 9, 75, 82].

The proposed extended formulation leads to a method at least two orders-of-magnitude faster than the `simple cutting surface algorithm`.

Define additional variables $\Gamma \in \mathbb{R}^{n\times n}$ such that $\Gamma_{ij} = \Gamma_{ji}$; intuitively, variable Γ_{ij} represents the product $x_i x_j$. Given an M-matrix Q, consider the convex optimization problem

$$\begin{aligned}
&\min_{(z,x,\Gamma)} \|y\|_2^2 - 2y'x + \langle \Gamma, Q\rangle && && \text{(23a)}\\
&\quad \text{s.t. } \Gamma_{ii} z_i \geq x_i^2 && \forall i = 1,\dots,n && \text{(23b)}\\
&\qquad 0 \geq \max_{d_{ij}>0} d_{ij} f(z_i, z_j, x_i, \frac{x_j}{d_{ij}}) - \left(d_{ij}\Gamma_{ii} - 2\Gamma_{ij} + \frac{1}{d_{ij}}\Gamma_{jj}\right) && \forall i < j && \text{(23c)}\\
&\qquad 0 \leq x_i \leq \|y\|_\infty z_i && i = 1,\dots,n && \text{(23d)}\\
&\qquad z \in \bar{C},\ x \in \mathbb{R}^n_+,\ \Gamma \in \mathbb{R}^{n\times n}. && && \text{(23e)}
\end{aligned}$$

We will show in this section that problem (23) is equivalent to `decomp` under mild conditions, and can be implemented efficiently via conic quadratic optimization. In order to prove this result, we introduce the auxiliary formulation:

$$\min_{(z,x,\Gamma)} \; \|y\|_2^2 - 2y'x + \langle \Gamma, Q\rangle \tag{24a}$$

$$\text{s.t. } 0 \geq \max_{\substack{d_{ij}^i d_{ij}^j \geq 1\\ d_{ij}^i, d_{ij}^j \geq 0}} g(z_i,z_j,x_i,x_j;d_{ij}^i,d_{ij}^j) - \left(d_{ij}^i\Gamma_{ii} - 2\Gamma_{ij} + d_{ij}^j\Gamma_{jj}\right) \quad \forall i<j \tag{24b}$$

$$0 \leq x_i \leq \|y\|_\infty z_i \qquad i=1,\dots,n \tag{24c}$$

$$z \in \bar{C},\; x \in \mathbb{R}_+^n,\; \Gamma \in \mathbb{R}^{n\times n}. \tag{24d}$$

We first prove that (24) is equivalent to `decomp` (Proposition 5), and then show that (23) and (24) are equivalent (Proposition 6). Before doing so, let us verify that (23)–(24) are indeed relaxations of (11).

Proposition 4. *Problems* (23)–(24) *are valid convex relaxations of* (11).

Proof. We only prove this result for (24); the proof for (23) follows from identical arguments and is omitted for brevity.

First we argue convexity of (24). Clearly, the objective (24a) is linear and constraints (24d) are convex. Moreover, the right hand sides of constraints (24b) are supremum of convex functions, thus convex.

Now we argue that (24) is indeed a relaxation of (11). Suppose that constraints $\Gamma_{ij} = x_ix_j$ and $z \in C$ are added to (24): then $\langle \Gamma, Q\rangle = x'Qx$ and the objective functions of (11) and (24) coincide. Moreover, for any nonnegative d_{ij}^i, d_{ij}^j such that $d_{ij}^i d_{ij}^j \geq 1$, we see that

$$\begin{aligned} g(z_i,z_j,x_i,x_j;d_{ij}^i,d_{ij}^j) &\leq d_{ij}^i x_i^2 - 2x_ix_j + d_{ij}^j x_j^2 && \text{(Theorem 1 -- validity)}\\ &= d_{ij}^i\Gamma_{ii} - 2\Gamma_{ij} + d_{ij}^j\Gamma_{jj}, && (\Gamma_{ij} = x_ix_j)\end{aligned}$$

thus inequalities (24b) are satisfied. So, if constraints $\Gamma_{ij} = x_ix_j$ and $z \in C$ are added, (24) is equivalent to (11). Hence, (24) is a relaxation of (11). □

Proposition 5. *If Q is a positive definite M-matrix, then problems `decomp` and* (24) *are equivalent.*

Proof. Consider the variable Γ_{ij} in (24) for some pair $i<j$: observe that it only appears in the objective with coefficient $Q_{ij} \leq 0$, and a single constraint (24b). It follows that in an optimal solution of (24), variable Γ_{ij} is as large as possible and the corresponding constraint (24b) is binding:

$$2\Gamma_{ij} = \max_{\substack{d_{ij}^i d_{ij}^j \geq 1\\ d_{ij}^i, d_{ij}^j \geq 0}} g(z_i,z_j,x_i,x_j;d_{ij}^i,d_{ij}^j) - \left(d_{ij}^i\Gamma_{ii} + d_{ij}^j\Gamma_{jj}\right).$$

Therefore, we find that problem (24) is equivalent to

$$\min_{z\in\bar{C},x\in\mathbb{R}^n_+,\Gamma\in\mathbb{R}^n}\ \max_d\ \|y\|_2^2-2y'x+\sum_{i=1}^n Q_{ii}\Gamma_{ii}$$

$$+\sum_{i=1}^n\sum_{j=i+1}^n|Q_{ij}|\Big(g(z_i,z_j,x_i,x_j;d_{ij}^i,d_{ij}^j)-d_{ij}^i\Gamma_{ii}-d_{ij}^j\Gamma_{jj}\Big) \tag{25a}$$

$$\text{s.t. } d_{ij}^i d_{ij}^j\geq 1, d_{ij}^i\geq 0,\ d_{ij}^j\geq 0\quad \forall i<j. \tag{25b}$$

Rearranging terms, we see that the objective of the inner maximization problem (25a) is equal to

$$\sum_{i=1}^n\left(Q_{ii}-\sum_{j<i}|Q_{ji}|d_{ji}^i-\sum_{j>i}|Q_{ij}|d_{ij}^i\right)\Gamma_{ii}+\sum_{i=1}^n\sum_{j=i+1}^n|Q_{ij}|g(z_i,z_j,x_i,x_j;d_{ij}^i,d_{ij}^j),$$

where we ignored the constant (in d) term $\|y\|_2^2-2y'x$. In particular, the inner maximization problem is precisely the Lagrangian relaxation of (21), where Γ_{ii} are the dual variables associated with constraints (21b). Therefore, if strong duality holds for problem (21), then problems decomp and (24) are equivalent.

Finally, we verify that Slater's condition and, thus, strong duality for (21) hold for positive definite Q. Since Q is positive definite, we have that $Q=\bar{Q}+\rho I$ for an M-matrix $\bar{Q}$ (with same off-diagonals) and some $\rho>0$ (e.g., let ρ be the minimum eigenvalue of Q). Since $\bar{Q}$ is an M-matrix, there exists a vector δ satisfying

$$\sum_{j<i}|Q_{ji}|\delta_{ji}^i+\sum_{j>i}|Q_{ij}|\delta_{ij}^i=Q_{ii}-\rho<Q_{ii}\qquad\forall i=1,\dots,n$$

$$\delta_{ij}^i\delta_{ij}^j\geq 1,\ \delta_{ij}^i\geq 0,\ \delta_{ij}^j\geq 0\qquad\forall i<j.$$

It follows that letting $d_{ij}^i=\delta_{ij}^i+\epsilon$ and $d_{ij}^j=\delta_{ij}^j+\epsilon$ for all $i<j$ and $\epsilon>0$ small enough, we find a vector d such that $d_{ij}^i d_{ij}^j>1$ and

$$\sum_{j<i}|Q_{ji}|d_{ji}^i+\sum_{j>i}|Q_{ij}|d_{ij}^i\leq Q_{ii}\quad\forall i=1,\dots,n. \tag{26}$$

After increasing additional entries of d until all inequalities (26) are tight, we find an interior point of (21). □

For the signal estimation problem, Q is positive-definite. Nonetheless, if strong duality does not hold, formulation (24) is still a convex relaxation of (11) that is at least as strong as decomp.

Proposition 6. *Problems* (23) *and* (24) *are equivalent.*

Proof. For any $i < j$, we see from Theorem 1 that constraint (24b) is equivalent to the pair of constraints:

$$0 \geq \max_{\substack{d_{ij}^i d_{ij}^j \geq 1 \\ d_{ij}^i, d_{ij}^j \geq 0}} d_{ij}^i f(z_i, z_j, x_i, \frac{x_j}{d_{ij}^i}) + \frac{x_j^2}{z_j}\left(d_{ij}^j - \frac{1}{d_{ij}^i}\right) - d_{ij}^i \Gamma_{ii} + 2\Gamma_{ij} - d_{ij}^j \Gamma_{jj} \tag{27}$$

$$0 \geq \max_{\substack{d_{ij}^i d_{ij}^j \geq 1 \\ d_{ij}^i, d_{ij}^j \geq 0}} d_{ij}^j f(z_i, z_j, \frac{x_i}{d_{ij}^j}, x_j) + \frac{x_i^2}{z_i}\left(d_{ij}^i - \frac{1}{d_{ij}^j}\right) - d_{ij}^i \Gamma_{ii} + 2\Gamma_{ij} - d_{ij}^j \Gamma_{jj}. \tag{28}$$

Observe that $d_{ij}^i f(z_i, z_j, x_i, \frac{x_j}{d_{ij}^i}) \geq 0$ for any $d_{ij}^i > 0$. Therefore, if $\frac{x_j^2}{z_j} > \Gamma_{jj}$, constraint (27) is not satisfied since the right hand side can be made arbitrarily large by letting $d_{ij}^j \to \infty$ and $d_{ij}^i = 1/d_{ij}^j$. Therefore, constraint (27) implies that $\Gamma_{jj} z_j \geq x_j^2$. Similarly, (28) implies that $\Gamma_{ii} z_i \geq x_i^2$.

Now assume that $\Gamma_{jj} z_j \geq x_j^2$ hold for all $j = 1, \ldots, n$. In this case, for any optimal solution of the maximization problem (27) we find that d_{ij}^j is as small as possible; that is, $d_{ij}^j = 1/d_{ij}^i$. Thus, if $\Gamma_{jj} z_j \geq x_j^2$ holds, then constraint (27) reduces to

$$0 \geq \max_{d_{ij}^i > 0} d_{ij}^i f(z_i, z_j, x_i, \frac{x_j}{d_{ij}^i}) - d_{ij}^i \Gamma_{ii} + 2\Gamma_{ij} - \frac{1}{d_{ij}^i} \Gamma_{jj}, \tag{29}$$

which is precisely constraint (23c). Moreover, if $\Gamma_{ii} z_i \geq x_i^2$ holds, then constraint (28) reduces to

$$0 \geq \max_{d_{ij}^j > 0} d_{ij}^j f(z_i, z_j, \frac{x_i}{d_{ij}^j}, x_j) - \frac{1}{d_{ij}^j} \Gamma_{ii} + 2\Gamma_{ij} - d_{ij}^j \Gamma_{jj}. \tag{30}$$

After a change of variable $d_{ij}^i = 1/d_{ij}^j$ and noting that $(1/d_{ij}^i) f(z_i, z_j, d_{ij}^i x_i, x_j) = d_{ij}^i f(z_i, z_j, x_i, x_j/d_{ij}^i)$, we conclude that (30) is equivalent to (29). $\square$

Remark 1. Note that constraints (23c)–(24b) are necessary only if $Q_{ij} \neq 0$. For the signal estimation problem (3), $Q_{ij} = 0$ for $\{i, j\} \notin A$. Thus, the methods developed here are particularly efficient when Q is sparse.

4.2. **Implementation via conic quadratic optimization.** The objectives (23a) and (24a) are linear, and constraints (23b) are rotated cone constraints, and thus can be handled directly by conic quadratic optimization solvers. In Section 4.2.1, we show how constraints (23c) and (24b) can be reformulated as a conic constraints *for a fixed value of* d_{ij}. Then we describe, in Section 4.2.2, a cutting plane method for implementing (23).

4.2.1. *Conic quadratic reformulation of functions f and g.* We now show how to formulate convex models involving functions f and g as conic quadratic optimization problems. Specifically, we show how to model the epigraph of functions f and g in Propositions 7 and 8, respectively.

Proposition 7 (Extended formulation of $\overline{\text{conv}}(X^2)$)**.** *A point* $(z,x,s) \in \overline{conv}(X^2)$ *if and only if* $(z,x,s) \in [0,1]^2 \times \mathbb{R}^3_+$ *and there exists* $v, w \in \mathbb{R}$ *such that the set of inequalities*

$$v \geq x_1 - x_2,\ v^2 \leq sz_1,\ w \geq x_2 - x_1,\ w^2 \leq sz_2 \tag{31}$$

are satisfied.

Proof. Suppose, without loss of generality, that $x_1 \geq x_2$ and that (z,x) satisfies the bound constraints. If $(z,x,s) \in \overline{\text{conv}}(X^2)$ then $\frac{(x_1-x_2)^2}{z_1} \leq s$; setting $v = x_1 - x_2$ and $w = 0$, we find a feasible solution for (31). Conversely, if (31) is feasible, then $\frac{(x_1-x_2)^2}{z_1} \leq \frac{v^2}{z_1} \leq s$ and $(z,x,s) \in \overline{\text{conv}}(X^2)$. □

Proposition 8 (Extended formulation of $\overline{\text{conv}}(Z^2)$)**.** *A point* $(z,x,s) \in \overline{conv}(Z^2)$ *if and only if* $(z,x,s) \in [0,1]^2 \times \mathbb{R}^3_+$ *and there exists* $s_1, s_2, q_1, q_2 \in \mathbb{R}_+$ *and* $v_1, v_2, w_1, w_2 \in \mathbb{R}_+$ *such that the set of inequalities*

$$\begin{aligned}
&x_1^2 \leq s_1 z_1,\ x_2^2 \leq s_2 z_2 && (\texttt{persp})\\
&d_1 v_1 \geq d_1 x_1 - x_2,\ v_1^2 \leq q_1 z_1 && (z_1 \geq z_2 \text{ and } d_1 x_1 \geq x_2)\\
&d_1 v_2 \geq -d_1 x_1 + x_2,\ v_2^2 \leq q_1 z_2 && (z_1 \geq z_2 \text{ and } d_1 x_1 \leq x_2)\\
&d_1 q_1 + s_2\left(d_2 - \frac{1}{d_1}\right) \leq s && (z_1 \geq z_2)\\
&d_2 w_1 \geq x_1 - d_2 x_2,\ w_1^2 \leq q_2 z_1 && (z_1 \leq z_2 \text{ and } x_1 \geq d_2 x_2)\\
&d_2 w_2 \geq -x_1 + d_2 x_2,\ w_2^2 \leq q_2 z_2 && (z_1 \leq z_2 \text{ and } x_1 \leq d_2 x_2)\\
&d_2 q_2 + s_1\left(d_1 - \frac{1}{d_2}\right) \leq s && (z_1 \leq z_2)
\end{aligned}$$

are satisfied.

Proof. Follows from using the system (31) with inequalities (12)–(13). □

4.2.2. *Improved cutting surface method.* Our implementation of the strong relaxation `decomp` is based on formulation (23), implemented in a cutting surface method. Consider the relaxation of (23) given by

$$\begin{aligned}
\min_{(z,x,\Gamma)}\ & \|y\|_2^2 - 2y'x + \langle \Gamma, Q\rangle && && (32a)\\
\text{s.t. } & \Gamma_{ii} z_i \geq x_i^2 && \forall i = 1,\dots,n && (32b)\\
& 0 \geq d \cdot f(z_i, z_j, x_i, \frac{x_j}{d}) - \left(d\Gamma_{ii} - 2\Gamma_{ij} + \frac{1}{d}\Gamma_{jj}\right) && \forall i<j,\ \forall d \in \Delta_{ij} && (32c)\\
& 0 \leq x_i \leq \|y\|_\infty z_i && i = 1,\dots,n && (32d)\\
& z \in \bar{C},\ x \in \mathbb{R}^n_+,\ \Gamma \in \mathbb{R}^{n\times n}, && && (32e)
\end{aligned}$$

where each Δ_{ij} is a finite subset of $\mathbb{R}$. From Proposition 7, each constraint (32c) can be formulated by introducing new variables $s, v, w \geq 0$ as the system

$$0 \geq ds - \left(d\Gamma_{ii} - 2\Gamma_{ij} + \frac{1}{d}\Gamma_{jj}\right), \tag{33a}$$

$$v \geq x_1 - \frac{x_2}{d},\ v^2 \leq sz_1,\ w \geq \frac{x_2}{d} - x_1,\ w^2 \leq sz_2. \tag{33b}$$

Therefore, relaxation (32) can be solved using a conic quadratic solver.

In the proposed cutting surface method, formulation (32) is iteratively refined by adding additional elements to sets Δ_{ij}, as outlined in Algorithm 1. First, all sets Δ_{ij} are initialized to the singleton $\{1\}$ (line 1). At each iteration of the algorithm, a relaxation of the form (32) is solved to optimality (line 3). Then, for each pair of indexes $i < j$ where the relaxation induced by (32c) is weak, the set Δ_{ij} is enlarged to improve the relaxation (line 7); Remark 2 and Proposition 9 below show to efficiently check whether the relaxation needs to be refined and how to do so, respectively.

Algorithm 1 Algorithm to solve formulation `decomp`

```
Output: $(\hat{x}, \hat{z}, \hat{\Gamma})$ optimal for decomp
 1: $\Delta_{ij} \leftarrow \{1\}$ for all $i < j$
 2: while Stopping criterion not met do
 3:     $(\hat{x}, \hat{z}, \hat{\Gamma}) \leftarrow$ Solve (32)
 4:     for all $i < j$ do
 5:         if Constraint (23c) is not satisfied then
 6:             Compute optimal $d^*_{ij}$ for maximization (23c)    ▷ See Proposition 9
 7:             $\Delta_{ij} \leftarrow \Delta_{ij} \cup \{d^*_{ij}\}$
 8:         end if
 9:     end for
10: end while
11: return $(\hat{x}, \hat{z}, \hat{\Gamma})$
```

Proposition 9. *For any $i < j$, the optimal solution of the inner maximization problem* (23c) *is obtained as follows:*

(1) *If $x_i^2/\Gamma_{ii} \geq x_j^2/\Gamma_{jj}$ then:*
   (a) *If $\Gamma_{ii} - x_i^2/z_i = 0$, then $d_{ij} \to \infty$ is optimal.*
   (b) *Otherwise, $d_{ij} = \sqrt{\frac{\Gamma_{jj} - \frac{x_j^2}{z_i}}{\Gamma_{ii} - \frac{x_i^2}{z_i}}}$ is optimal.*

(2) *If $x_i^2/\Gamma_{ii} \leq x_j^2/\Gamma_{jj}$ then:*
   (a) *If $\Gamma_{ii} - x_i^2/z_j = 0$, then $d_{ij} \to \infty$ is optimal.*
   (b) *Otherwise, $d_{ij} = \sqrt{\frac{\Gamma_{jj} - \frac{x_j^2}{z_j}}{\Gamma_{ii} - \frac{x_i^2}{z_j}}}$ is optimal.*

Proof. Suppose $x_i^2/\Gamma_{ii} \geq x_j^2/\Gamma_{jj}$. Note that $\Gamma_{ii} \geq x_i^2/z_i$ holds from constraints (23b). Moreover, we find that

$$\Gamma_{jj} - \frac{x_j^2}{z_i} \geq \Gamma_{ii}\frac{x_j^2}{x_i^2} - \frac{x_j^2}{z_i} = x_j^2\left(\frac{\Gamma_{ii}}{x_i^2} - \frac{1}{z_i}\right) \geq 0.$$

We now show that there exists a stationary point of (23c) satisfying $d_{ij}x_i \geq x_j$. In this case, optimization problem (23c) reduces to

$$\begin{aligned} &0 \geq \max_{d_{ij}>0} \frac{d_{ij}x_i^2 - 2x_ix_j + x_j^2/d_{ij}}{z_i} - \left(d_{ij}\Gamma_{ii} - 2\Gamma_{ij} + \frac{1}{d_{ij}}\Gamma_{jj}\right) \\ \Leftrightarrow\ &0 \geq 2\left(\Gamma_{ij} - \frac{x_ix_j}{z_i}\right) + \max_{d_{ij}>0}\left\{-d_{ij}\left(\Gamma_{ii} - \frac{x_i^2}{z_i}\right) - \frac{1}{d_{ij}}\left(\Gamma_{jj} - \frac{x_j^2}{z_i}\right)\right\}. \end{aligned} \tag{34}$$

If $\Gamma_{ii} - x_i^2/z_i = 0$, then $d_{ij}^* \to \infty$ is an optimal solution to (34). If $\Gamma_{jj} - x_j^2/z_i = 0$, then $d_{ij}^* = 0$ is optimal. Moreover, if both $\Gamma_{ii} - x_i^2/z_i > 0$ and $\Gamma_{jj} - x_j^2/z_i > 0$, then taking derivatives with respect to d_{ij} we find that

$$d_{ij}^* = \sqrt{\frac{\Gamma_{jj} - \frac{x_j^2}{z_i}}{\Gamma_{ii} - \frac{x_i^2}{z_i}}}. \tag{35}$$

Finally, we verify that the condition $d_{ij}^*x_i \geq x_j$ holds. Indeed, this condition reduces to

$$\left(\frac{\Gamma_{jj} - \frac{x_j^2}{z_i}}{\Gamma_{ii} - \frac{x_i^2}{z_i}}\right)x_i^2 \geq x_j^2 \Leftrightarrow \left(\Gamma_{jj} - \frac{x_j^2}{z_i}\right)x_i^2 \geq \left(\Gamma_{ii} - \frac{x_i^2}{z_i}\right)x_j^2 \Leftrightarrow \frac{x_i^2}{\Gamma_{ii}} \geq \frac{x_j^2}{\Gamma_{jj}},$$

which is satisfied. The proof for the case $x_i^2/\Gamma_{ii} \leq x_j^2/\Gamma_{jj}$ is analogous. □

Remark 2. By replacing d_{ij} with its optimal value in (23c), we find that this constraint can be written explicitly as the piecewise constraint

$$0 \geq \begin{cases} \Gamma_{ij} - \frac{x_ix_j}{z_i} - \sqrt{\left(\Gamma_{ii} - \frac{x_i^2}{z_i}\right)\left(\Gamma_{jj} - \frac{x_j^2}{z_i}\right)} & \text{if } \frac{x_i^2}{\Gamma_{ii}} \geq \frac{x_j^2}{\Gamma_{jj}} \\ \Gamma_{ij} - \frac{x_ix_j}{z_j} - \sqrt{\left(\Gamma_{ii} - \frac{x_i^2}{z_j}\right)\left(\Gamma_{jj} - \frac{x_j^2}{z_j}\right)} & \text{if } \frac{x_i^2}{\Gamma_{ii}} \leq \frac{x_j^2}{\Gamma_{jj}}. \end{cases} \tag{36}$$

However, constraint (36) is not conic quadratic.

Remark 3. In our computations, we use the following stopping criterion in line 3. Let ζ_{old} and ζ_{new} be the optimal objective value of the relaxation (line 3). The algorithm is terminated when the relative improvement of the relaxation $(\zeta_{\text{new}} - \zeta_{\text{old}})/\zeta_{\text{new}} \leq 5 \times 10^{-5}$.

Remark 4. Using Proposition 8, one can extend the ideas discussed in this section to tackle (24) in a conic quadratic optimization framework as well.

However, we prefer formulation (23) since the conic quadratic representation of function f is simpler and more compact.

4.3. **Lagrangian methods for estimation with regularized objective.**

The cutting surface method introduced in Section 4.2 requires solving a sequence of progressively larger conic quadratic optimization problems. Based on our computations, this method can handle a variety of constraints (encoded by set $\bar{C}$), and solve the instances with $n \leq 10,000$ within seconds. For better scalability, in this section, we develop a Lagrangian relaxation-based method for the estimation problem with regularization objective:

$$\begin{aligned}
\min_{x,z}\ & \|y-x\|_2^2 + \lambda\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2 + \mu\sum_{i=1}^{n} z_i && && \text{(37a)}\\
\text{s.t. } & 0 \leq x_i \leq \|y\|_\infty z_i && i=1,\dots,n && \text{(37b)}\\
& x\in\mathbb{R}^n_+,\ z\in\{0,1\}^n && && \text{(37c)}
\end{aligned}$$

where $\mu \geq 0$ is a regularization parameter controlling the sparsity of the target signal. Let $L = \{\ell_1,\dots,\ell_m,\ell_{m+1}\} \subseteq \{1,\dots,n\}$ be any subset of the indexes such that $1 = \ell_1 < \ldots < \ell_m < \ell_{m+1} = n+1$. With the introduction of additional variables $w_j = x_{\ell_j} - x_{\ell_j - 1}$, problem (37) can be equivalently written as

$$\begin{aligned}
\min_{x,z}\ & \sum_{j=1}^{m}\left(\sum_{i=\ell_j}^{\ell_{j+1}-1}(y_i-x_i)^2 + \lambda\sum_{i=\ell_j}^{\ell_{j+1}-2}(x_{i+1}-x_i)^2 + \mu\sum_{i=\ell_j}^{\ell_{j+1}-1} z_i\right) + \sum_{j=2}^{m} w_j^2 && && \text{(38a)}\\
\text{s.t. } & w_j = x_{\ell_j} - x_{\ell_j-1} && j=2,\dots,m && \text{(38b)}\\
& 0 \leq x_i \leq \|y\|_\infty z_i && i=1,\dots,n && \text{(38c)}\\
& x\in\mathbb{R}^n_+,\ z\in\{0,1\}^n,\ w\in\mathbb{R}^{m-1}. && && \text{(38d)}
\end{aligned}$$

Without the coupling constraints (38b), problem (38) decomposes into m independent problems, each with variables indexed in $[\ell_j, \ell_{j+1}-1]$ for $j = 1,\dots,m$. Letting γ_j be the Lagrange multiplier for constraint $w_j = x_{\ell_j} - x_{\ell_j-1}$, we obtain the Lagrangian dual problem

$$\begin{aligned}
\max_{\gamma\in\mathbb{R}^{m-1}}\min_{x,z,w}\ & \sum_{j=1}^{m}\left(\sum_{i=\ell_j}^{\ell_{j+1}-1}(y_i-x_i)^2 + \lambda\sum_{i=\ell_j}^{\ell_{j+1}-2}(x_{i+1}-x_i)^2 + \mu\sum_{i=\ell_j}^{\ell_{j+1}-1} z_i\right) + \sum_{j=2}^{m} w_j^2\\
& + \gamma_j(w_j - x_{\ell_j} + x_{\ell_j-1}) && && \text{(39a)}\\
\text{s.t. } & 0 \leq x_i \leq \|y\|_\infty z_i && i=1,\dots,n && \text{(39b)}\\
& x\in\mathbb{R}^n_+,\ z\in\{0,1\}^n,\ w\in\mathbb{R}^{m-1}. && && \text{(39c)}
\end{aligned}$$

Observe that $w_j = -\gamma_j/2$ holds for an optimal solution of the inner minimization problem. Moreover, to obtain a strong convex relaxation, we can reformulate each independent inner minimization problem using the formulations discussed in Section 3.2, yielding the convex relaxation

$$\begin{aligned}\max_{\gamma\in\mathbb{R}^{m-1}}\ \min_{x,z}\ &\sum_{j=1}^{m}\left(\sum_{i=\ell_j}^{\ell_{j+1}-1}(y_i^2-2y_ix_i)+\theta_j(z,x)+\mu\sum_{i=\ell_j}^{\ell_{j+1}-1}z_i\right)-\sum_{j=2}^{m}\frac{\gamma_j^2}{4}\\ &+\gamma_j(x_{\ell_j-1}-x_{\ell_j}) && \text{(40a)}\\ \text{s.t. }\ &0\le x_i\le\|y\|_\infty z_i \qquad i=1,\dots,n && \text{(40b)}\\ &x\in\mathbb{R}^n_+,\ z\in[0,1]^n, && \text{(40c)}\end{aligned}$$

where $\theta_j(z,x)$ is the convexification of the epigraph of the term

$$\sum_{i=\ell_j}^{\ell_{j+1}-1}x_i^2+\lambda\sum_{i=\ell_j}^{\ell_{j+1}-2}(x_{i+1}-x_i)^2.$$

Implementation. Problem (40) can be solved via a primal-dual method: for any fixed γ, the inner minimization problem can be solved by solving m independent sub-problems (in parallel), and each sub-problem is solved using Algorithm 1. We now describe our implementation of the "main" outer maximization problem.

Set L: Given a target number of subproblems $m\in\mathbb{Z}_+$, we let $\ell_j=1+(j-1)\lfloor n/m\rfloor$ for $j=1,\dots,m$.

Subgradient method: Given $\gamma\in\mathbb{R}^{m-1}$, a subgradient of the objective (40a) at γ is given by $\xi(\gamma)_j=-\frac{\gamma_j}{2}+(x^*_{\ell_j-1}-x^*_{\ell_j})$, where x^* is an optimal solution of the inner minimization problem at γ. Thus, letting γ^h be the value of γ at iteration $h\in Z_+$, we use the update rule $\gamma^{h+1}=\gamma_h+(1/h)\xi(\gamma^h)$.

Initial point: We start the algorithm with the initial point $\gamma^0=0$. Note that if $x_{\ell_j-1}=x_{\ell_j}=0$ (which is the case for large μ), then $\gamma_j=0$ is optimal.

Stopping criterion: We terminate the algorithm when $\|\xi(\gamma^h)\|_\infty<\epsilon$ ($\epsilon=10^{-3}$ in our computations) or when the number of iterations reaches h_{max} ($h_{max}=100$ in our computations).

Additional considerations: In the first iteration, we need to solve m subproblems. However, the subsequent iterations often require solving fewer subproblems: if $\xi(\gamma^h)_j=\xi(\gamma^{h+1})_j$ and $\xi(\gamma^h)_{j+1}=\xi(\gamma^{h+1})_{j+1}$, then at iteration $h+1$ solution of subproblem j does not change from the previous iteration. For problem instances with large μ, the number of subproblems solved in subsequent iterations reduces considerably.

On Lagrangian methods for general adjacency graphs. We point out that the tailored Lagrangian relaxation-based decomposition can be applied to higher-dimensional adjacency graphs. For example, when the adjacency graph is a two-dimensional mesh, one can divide the mesh into rectangular regions and apply a similar decomposition approach. Theoretically, it is possible to use a similar method for arbitrary adjacency graphs. However, key to efficiency of the Lagrangian method, described in "additional considerations" above, is the observation that if the dual variables corresponding to the border of a subregion do not change, then subproblems need not be resolved. With higher-dimensional graphs, where borders are no longer defined by two points but by rectangles or other objects, more subproblems may need to be resolved *unless regions are defined carefully*, depending on the data y. Developing such methods to partition the adjacency graph is beyond the scope of this paper.

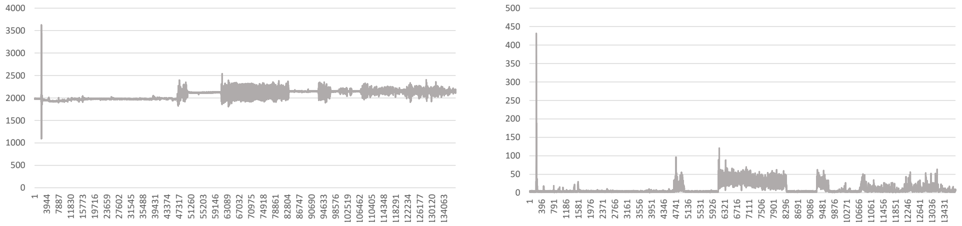
5. Computations

In this section we present experiments with utilizing the strong convex relaxations based on the pairwise convexification methods proposed in the paper. In Section 5.1, we perform experiments to evaluate whether the convex model `decomp` provides a good approximation to the non-convex problem (11). In Section 5.2 we test the merits of formulation `decomp` (with a variety of constraints $\bar{C}$) compared to the usual ℓ_1-approximation from an inference perspective. Finally, in Section 5.3 we test the Lagrangian relaxation-based method proposed in Section 4.3. We use Mosek 8.1.0 (with default settings) to solve the conic quadratic optimization problems. All computations are performed on a laptop with eight Intel(R) Core(TM) i7-8550 CPUs and 16GB RAM. All data and code used in the computations are available at `https://sites.google.com/usc.edu/gomez/data`.

5.1. **Relaxation quality.**

This section is devoted to testing how well the proposed convex relaxations are able to approximate the ℓ_0 optimization problems using real data.

5.1.1. *Data.* Consider the accelerometer data depicted in Figure 2 (A), used in [22, 23] and downloaded from the UCI Machine Learning Repository [26]. The time series corresponds to the "x acceleration" of participant 2 of the "Activity Recognition from Single Chest-Mounted Accelerometer Dataset". This participant was "working at computer" until time stamp 44,149; "standing up, walking and going upstairs" until time stamp 47,349; "standing" from time stamp 47,350 to 58,544, from 80,720 to 90,439, and from time 90,441 to 97,199; "walking" from 58,545 to 80,719; "going up or down stairs" from 90,440 to 94,349; "walking and talking with someone" from 97,200 to 104,300; and "talking while standing" from 104,569 to 138,000 (status between 104,301 and 104,568 is unknown).



(A) Original data (B) Transformed data

Figure 2. Underlying signals and noisy observations.

Several machine learning methods have been proposed to use accelerometer data to discriminate between activities, e.g., see [12] and the references therein. Variations of the acceleration can help to discriminate between activities [22]. Moreover, as pointed out in [85], behaviors can be identified (at a simplistic level) from frequencies and amplitudes of wave patterns in a single axis of the accelerometer. Therefore, we consider a rudimentary approach to identify activities from the accelerometer data: we partition the dataset into windows of 10 samples each, and for each window we compute the mean absolute value of the successive differences, obtaining the dataset plotted in Figure 2 (B)[2]. Finally, we scale the data so that $\|y\|_\infty = 1$.

Given an optimal solution x^* of the estimation problem (6) or a suitable relaxation of it, periods with little or no physical activity can be naturally associated with time stamps i where $x_i^* = 0$, and values $x_i^* > 0$ can be used as a proxy for the energy expenditure due to physical activity [74].

5.1.2. *Methods.* We compare the following two relaxations of the ℓ_0-problem

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n_+}\ & \sum_{i=1}^n (y_i - x_i)^2 + \lambda \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 && \text{(41a)}\\
\text{s.t.}\ & \sum_{i=1}^n z_i \leq k && \text{(41b)}\\
& x \leq z, && \text{(41c)}\\
& z \in \{0,1\}^n. && \text{(41d)}
\end{aligned}
$$

L1: The natural convex relaxation of (41), obtained by relaxing the integrality constraints to $z \in [0,1]^n$.

[2] One of the key features identified in [22] for activity recognition are the minmax sums of 52-sample windows, computed as the sums of successive differences of consecutive "peaks". The time series we obtain follows a similar intuition, but is larger and noisier due to smaller windows.

- **Decomp:** The convex model `decomp` –equivalently, (23)– implemented using Algorithm 1.

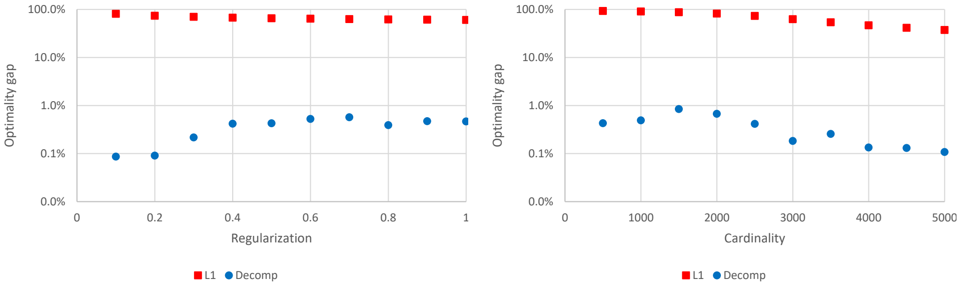
The convex formulations used are relaxations of the ℓ_0 problem; so, their optimal objective values ζ_{LB} provide lower bounds on the optimal objective value ζ of (41). We use a simple *thresholding* heuristic to construct a feasible solution for (41): for a given solution $\hat{x}$ to a convex relaxation, let $\hat{x}_{(k)}$ denote the k-th largest value, and $\bar{x}$ be the solution given by

$$\bar{x}_i = \begin{cases} \hat{x}_i & \text{if } \hat{x}_i \geq \hat{x}_{(k)} \\ 0 & \text{otherwise.} \end{cases}$$

By construction $\bar{x}$ is feasible for (41), and its objective value ζ_{UB} provides an upper bound on ζ. Thus, the optimality gap of the heuristic is

$$\texttt{gap} = 100 \times \frac{\zeta_{\text{UB}} - \zeta_{\text{LB}}}{\zeta_{\text{UB}}} \cdot . \tag{42}$$

5.1.3. *Results.* We test the convex formulations with the accelerometer data using $\lambda = 0.1t$ and $k = 500t$ for $t = 1, \ldots, 10$ for all 100 combinations. Figure 3(A) presents the optimality gaps obtained by each method for each value of λ (averaging over all values of k), and Figure 3(B) presents the optimality gaps for each value of k (averaging over all values of λ). We see that `decomp` substantially improves upon the natural ℓ_1 relaxation. Indeed, the gaps from the ℓ_1 relaxation are 66.7% on average, and can be very close to 100%; in contrast, the strong relaxations derived in this paper yields optimality gaps of 0.4% on average.



(A) Gap as a function of λ. (B) Gap as a function of k.

FIGURE 3. Optimality gaps of the ℓ_1 relaxation (red) and the proposed convexification `decomp` (blue).

Figure 4 presents the distribution of the time required for each method to solve the respective convex model. We see that the improvement of relaxation quality of the new relaxations comes at the cost of computational efficiency: while the ℓ_1 relaxation is solved in approximately one second, the proposed convexification requires on average 54 seconds. Although the vast majority of the instances are solved under 100 seconds using Algorithm 1, a couple of instances require close to 10 minutes. Nonetheless, an average time of under minute to solve the instances to near-optimality (less than 1% optimality gap) is adequate for most practical settings. Moreover, as shown in Section 5.3, the computation times can be improved substantially using the decomposition method proposed in Section 4.3.

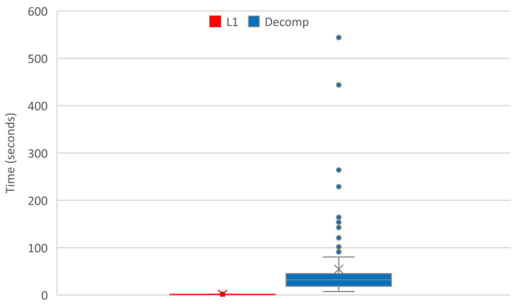


Figure 4. Distribution of CPU times for each method.

Finally, we present a brief comparison with mixed-integer optimization methods. Specifically, we use the perspective reformulation of (41), i.e.,

$$\begin{aligned}\min_{x\in\mathbb{R}^n_+}\; & \sum_{i=1}^{n}\left(y_i^2-2y_ix_i+\frac{x_i^2}{z_i}\right)+\lambda\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2\\ \text{s.t.}\; & \sum_{i=1}^{n} z_i\le k\\ & x\le z,\\ & z\in\{0,1\}^n\end{aligned}$$

and solve the problems using Gurobi 8.0 with a one hour time limit. Table 1 presents the results: it shows, for $k \in \{2000, 4000\}$ and $\lambda \in \{0.1, 0.2\}$, the gaps and solution times for ℓ_1-approx, decomp , and the branch-and-bound solver (b&b)as well as the number of branch-and-bound nodes explored. The mixed-integer optimizer is unable to solve to optimality instances of this size within the time limit of one hour. While Gurobi is able to produce good optimality gaps, improving substantially upon the simple ℓ_1-relaxation (in part due to the use of the perspective reformulation), decomp produces better optimality gaps by an order-of-magnitude and requires only a small fraction of the computational effort.

Table 1. Comparison with exact branch-and-bound method.

k	λ	ℓ_1-`approx`		`decomp`		`b&b`		
		gap	time(s)	gap	time(s)	gap	time(s)	nodes
2000	0.1	91.2	1	0.3	60	2.7	3,600	4,047
	0.2	87.0	1	0.6	42	7.4	3,600	1,445
4000	0.1	68.0	1	0.0	7	3.4	3,600	1,350
	0.2	56.7	1	0.1	19	2.8	3,600	4,268

5.2. **Statistical performance – modeling with priors.** In Section 5.1 we established that the convex model derived in this paper indeed provides a much closer approximation for ℓ_0 signal estimation problem than the usual ℓ_1 relaxation. In this section we demonstrate that using the proposed convexification leads to better statistical performance than relying on the ℓ_1-approximation alone. We also show how additional priors other than sparsity can be seamlessly integrated into the new convex models, and the benefits of doing so.

5.2.1. *Data.* We now describe how we generate test instances. First, the "true" sparse signal $\hat{y}$ is generated as follows. Let n be the number of time epochs, let s be a parameter controlling the number of "spikes" of the signal and let h be a parameter controlling the length of each spike. Initially, the true signal is fully sparse, $\hat{y} = 0$. Then we iteratively repeat the following process to generate s spikes of non-zero values:

(1) We select an index ℓ uniformly between 1 and $n+1-h$, corresponding to the start of a given spike.
(2) We sample an h-dimensional vector v for a multivariate Gaussian distribution with mean 0 and covariance matrix B, where $B_{ij} = \frac{i(h+1-j)}{h+1}$ for $i \leq j$. Thus v is a realization of a *Brownian bridge* process.
(3) We update $\hat{y}_{\ell+i} \leftarrow \hat{y}_{\ell+i} + |v_i|$.

Note that two different spikes may overlap, in which case the true signal $\hat{y}$ would have a single spike with larger intensity. Also note that the true signal $\hat{y}$ generated in this way has at most hs non-zeros and at most s spikes, but may have fewer if overlaps occur. Then, given a noise parameter σ, we generate the noisy observations $y_i = \hat{y}_i + \varepsilon_i$, where ε_i follows a truncated normal distribution with mean 0, variance σ_i^2 and lower bound $-\hat{y}_i$. Finally, we scale the data so that $\|y\|_\infty = 1$.

5.2.2. *Methods.* We compare the following methods:

L1: Corresponds to solving the ℓ_1-`approx` problem

$$\min_{x\in\mathbb{R}^n_+} \|y-x\|_2^2+\lambda\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2+\mu\|x\|_1.$$

Decomp-sparse: Enforces the prior that the signal has a most hs non-zeros, by solving the convex optimization problem

$$\begin{aligned}\min_{x\in\mathbb{R}^n_+,z\in[0,1]^n}\;& \|y\|_2^2-2\sum_{i=1}^n y_ix_i+\theta(z,x)+\mu\|x\|_1\\ \text{s.t.}\;& \sum_{i=1}^n z_i\le hs\\ & 0\le x\le \|y\|_\infty z\end{aligned}$$

using Algorithm 1.

Decomp-prior: In addition to the sparsity prior as before, it incorporates the information that the underlying signal has a most s spikes and that each spike has at least h non-zeros. These two priors can be enforced by solving the optimization problem

$$\min_{x\in\mathbb{R}^n_+,z\in[0,1]^n}\; \|y\|_2^2-2\sum_{i=1}^n y_ix_i+\theta(z,x)+\mu\|x\|_1 \tag{43a}$$

$$\text{s.t.}\; \sum_{i=1}^n z_i\le hs \tag{43b}$$

$$\sum_{i=1}^{n-1}|z_{i+1}-z_i|\le 2s \tag{43c}$$

$$\sum_{i=\max\{1,\ell-h\}}^{\min\{n,\ell+h\}} z_i\ge hz_\ell \qquad \ell=1,\dots,n \tag{43d}$$

$$0\le x\le \|y\|_\infty z. \tag{43e}$$

Constraint (43c) states that the process can transition from a zero value to a non-zero value at most $2s$ times, thus can have at most s spikes. Each constraint (43d) states that, if $z_\ell=1$, then there must be at least $h-1$ neighboring non-zero points, thus non-zero indexes occur in patches of at least h elements.

Observe that, following the results in [62], we keep an ℓ_1-regularization for shrinkage to improve performance in low signal-noise-ratio regimes.

5.2.3. *Computational setting.* For the computations in this section, we generate instances with $n=1,000$, $s=10$, and $h=10$; so, each signal is zero in approximately 90% of the time. Moreover, we test noise levels $\sigma = 0.1t$, $t = 1, \dots, n$, and for each σ we generate 10 different instances as follows:

(1) For each parameter combination, two signals are randomly generated: one signal for training, the other for testing.

(2) For all methods, we solve the corresponding optimization problem for the training signal with 10 values of the smoothness parameter λ and 10 values of the shrinkage parameter μ, a total of 100 combinations. We consider two criteria for choosing a pair (λ, μ):

Error: The pair that best fits the true signal with the respect to the estimation error, i.e., combination minimizing $\|\hat{y} - x^*\|_2^2$, where x^* is the solution for corresponding optimization.

Sparsity: The pair that best matches the sparsity pattern of the true signal[3], i.e., combination minimizing $\sum_{i=1}^{n} \big||\hat{y}_i|_0 - |x_i^*|_0\big|$. This setting is of practical interest in cases where the training data is partially labeled: the location of the spikes is known but the actual value of the signal is not.

(3) We solve the optimization problem for the testing signal with parameters (λ, μ) chosen in (2), and report the results (averaged over the 10 instances).

For the instances considered, Table 2 shows the average Signal-to-Noise Ratio (SNR) as a function of σ, computed as SNR $= \frac{\|\hat{y}\|_2^2}{\|\hat{y}-y\|_2^2}$.

Table 2. Signal-to-Noise Ratio for different values of the noise.

σ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
SNR	2,200	138	27	8.6	3.5	1.7	0.9	0.5	0.3	0.2

5.2.4. *Results with respect to the error criterion.* We now present the results when the true values of $\hat{y}$ are known in training. Figure 5 depicts the out-of-sample error of each method and SNR, computed as error $= \frac{\|\hat{y}_{test}-x^*\|_2^2}{\|\hat{y}_{test}\|_2^2}$ where $\hat{y}_{test}$ is the true testing signal, and x^* is the estimator. Figure 6 depicts how accurately the estimator obtained in testing matches the sparsity pattern of the true signal. We observe that the standard ℓ_1-norm approach results in dense signals with a substantial number of false positives, and is outperformed by the approaches that enforce priors in terms of error as well. The inclusion of the sparsity prior results in a notable improvement in terms of the error across all SNRs, and reducing it by half or more for SNR$\geq$ 3. This prior also yields an order-of-magnitude improvement in terms

[3] A point x_i is considered non-zero if $|x_i| > 10^{-3}$.

of matching the sparsity pattern for SNR≥ 3, although for low SNRs the improvement in matching the sparsity pattern is less pronounced (and is worse for SNR=0.5). The inclusion of additional priors for the number and length of each spike yields further improvements (especially for low SNRs), and yields a good match for the sparsity pattern in all cases.

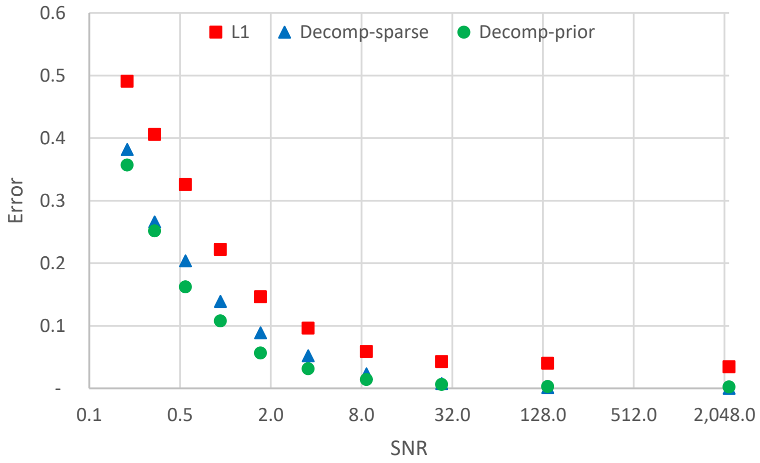


Figure 5. Average out-of-sample error as a function of SNR (in log-scale).

Figure 7 provides detailed information about the distribution of the out-of-sample errors for three different SNRs. We see that in high SNR regimes, the inclusion of the sparsity prior consistently outperforms the ℓ_1-norm method, and the inclusion of additional priors consistently outperforms using only the sparsity prior. In contrast, in low SNR regimes, while the inclusion of additional priors yields better results on average, the improvement is not as consistent.

Finally, Figure 8 depicts the average time required to solve the optimization problems as a function of the SNR. As expected, the ℓ_1-norm approximation is the fastest method. Optimization problems with the sparsity prior are solved under two seconds, and optimization problems with all priors are solved under 10 seconds. We see that time required to solve the problems based on the stronger relaxations increases as the SNR decreases.

5.2.5. *Results with respect to the sparsity pattern criterion.* We now present the results when, for the training data, the true values of $\hat{y}$ are unknown, but its sparsity pattern is known. Figure 9 depicts the out-of-sample error of each method for each SNR and Figure 10 depicts how accurately the estimator obtained in validation matches the sparsity pattern of the true

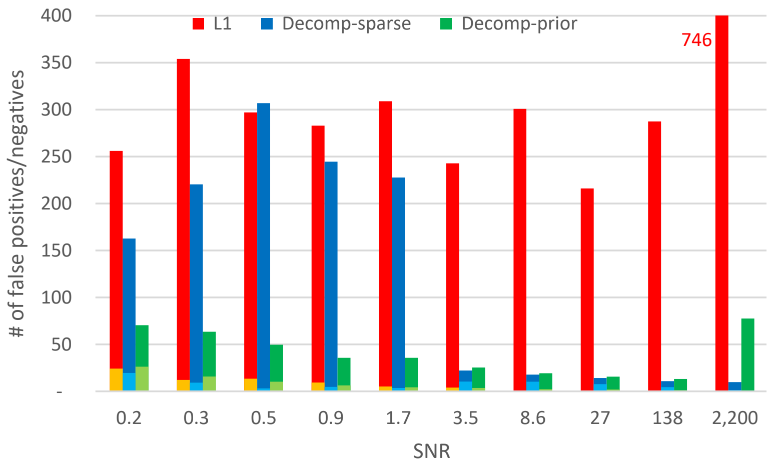
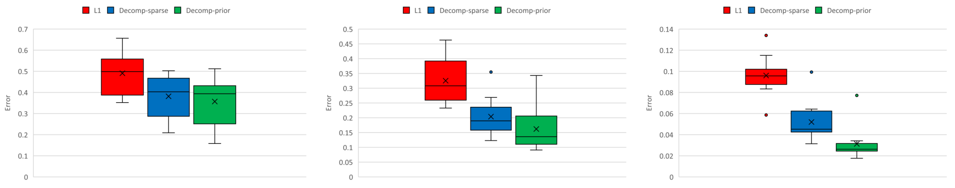


Figure 6. Average out-of-sample number of false positives (red/dark blue/dark green) and false negatives (orange/light blue/light green) as a function of SNR. The number of false positives for L1 with SNR=2,200 is 746.



(a) SNR=0.2.

(b) SNR=0.5.

(c) SNR=3.5.

Figure 7. Distribution of the out-of-sample errors for different SNRs when the true values of the signal used in training are available.

signal. Naturally, as the true values of the training signal are unknown, all methods perform worse in terms of the out-of-sample error. The ℓ_1-norm method in particular performs very poorly in low SNR regimes: the estimator is $x \approx 0$, resulting a large error close to one and several false negatives (with no false positives, since few or no indexes are non-zero). In contrast the methods that enforce priors result in significantly reduced error across all SNRs while simultaneously improving the detection of the sparsity in low SNRs regimes, correctly detecting several spikes. In this setting, we did not observe a substantial difference between methods Decomp-sparse and Decomp-prior. From Figure 11,which depicts the distributions of the errors,

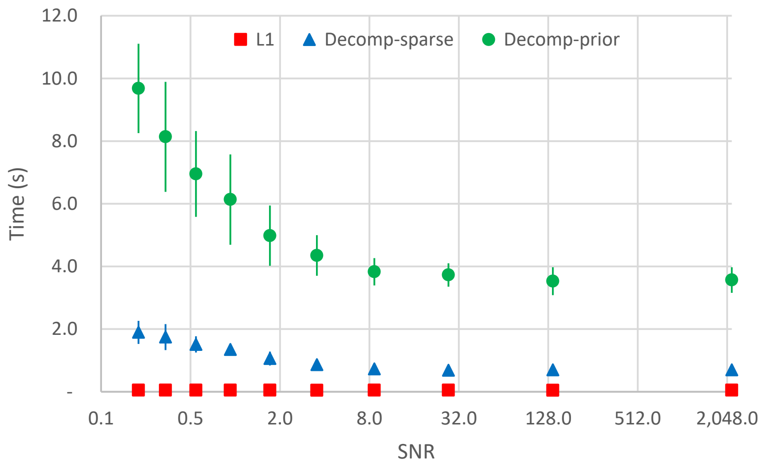


FIGURE 8. CPU time in seconds as a function of SNR (in log-scale). The error bars correspond to ±1 stdev.

we see that the new convexification-based methods consistently outperform the ℓ_1-method.

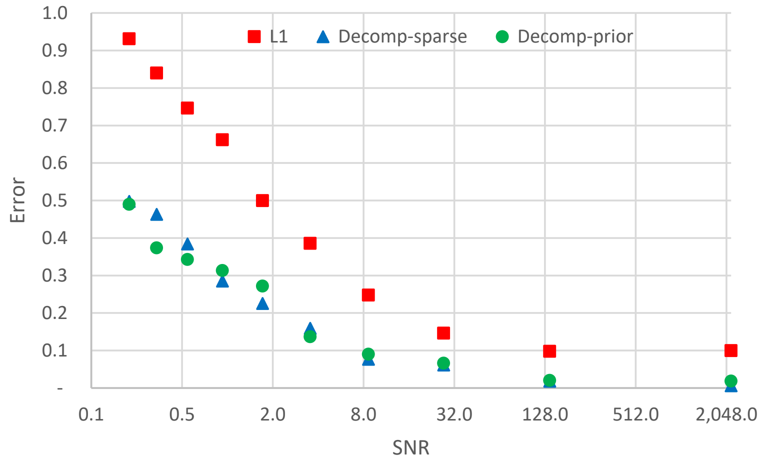


FIGURE 9. Average out-of-sample error as a function of SNR (in log-scale) when only the sparsity pattern of the training signal is known.

5.3. **Computational experiments - Lagrangian methods.** We now report on the performance of the Lagrangian method given in Section 4.3 for

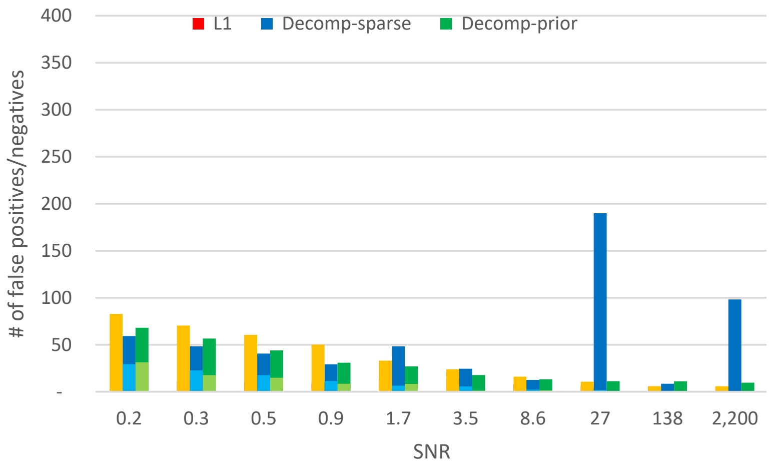
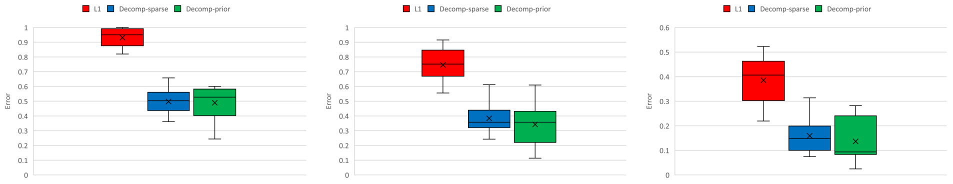


FIGURE 10. Average out-of-sample number of false positives (red/dark blue/dark green) and false negatives (orange/light blue/light green) as a function of SNR when only the sparsity pattern of the training signal is known.



(A) SNR=0.2.

(B) SNR=0.5.

(C) SNR=3.5.

FIGURE 11. Distribution of the out-of-sample errors for different SNRs when only the sparsity pattern of the training signal is known.

larger signals with $n = 100,000$, $\sigma = 0.5$, $s = 10$ and $h = 100$ (so approximately 1% of the signal values are non-zero). We denoise the signal by solving the optimization problem

$$\min_{x\in\mathbb{R}^n_+, z\in[0,1]^n} \; \|y\|_2^2 - 2\sum_{i=1}^n y_i x_i + \theta(z,x) + \mu\|x\|_1 + \kappa\|z\|_1 \tag{44a}$$

$$\text{s.t. } 0 \leq x \leq \|y\|_\infty z. \tag{44b}$$

In these experiments we use synthetic instances generated as in Section 5.2 with $\lambda = 0.3$ and $\mu = 0$[4] and varying $\kappa \in \{0.0005, 0.001, 0.002, 0.005, 0.01, 0.02\}$. We solve (44) using the Lagrangian method with $m \in \{1, 10, 100, 1000\}$ subproblems ($m = 1$ corresponds to no decomposition). The independent subproblems are solved in parallel in the same laptop computer.

Table 3 presents the results, both in terms of statistical and computational performance. For each value of κ and m, it shows the error between the true signal and the estimated signal[5], the number of non-zero values $\|x^*\|_0$ of the resulting estimator, the time required to solve the problem; the number of subgradient iterations used, and the actual number of subproblems solved.

We observe that for the smallest value of $\kappa = 0.0005$ (corresponding to a sparsity of $\|x^*\|_0 \approx 30,000$), the method without decomposition ($m = 1$) is the fastest and is able to solve the problems in approximately three minutes. However, as the value of the ℓ_0 regularization parameter κ increases, the Lagrangian methods solve the problems increasingly faster. In particular, for values of $\kappa \geq 0.005$ (sparsity of $\|x^*\|_0 \leq 900$), the Lagrangian method with $m = 1,000$ solves the problems in under one minute whereas a direct implementation via Algorithm 1 may require an hour or more. Indeed, we see that as the ℓ_0 regularization parameter increases, the number of iterations and number of subproblems solved decreases considerably. In fact, if $\kappa \geq 0.01$, the Lagrangian method with $m = 10$ is solved to optimality without performing any subgradient iterations. Finally, we point out that in terms of the estimation error, all methods return comparable errors (except for $\kappa = 0.0005$, where the maximum number of 100 iterations is reached and the Lagrangian methods do not solve the problems to optimality).

Therefore, we conclude that the proposed Lagrangian method is able to efficiently tackle large-scale problems when the target sparsity is small compared to the dimension of the problem, and can solve the problems by two-orders of magnitude faster compared to default method. The drawback is that the decomposition method is unable to incorporate additional priors using constraints.

6. Conclusions

In this paper we derived strong iterative convex relaxations for quadratic optimization problems with M-matrices and indicators, of which signal estimation with smoothness and sparsity is a special case. The relaxations are based on convexification of quadratic functions on two variables, and

[4] In the experiments reported in Section 5.2.4 with $\sigma = 0.5$ and method decomp-sparse, the combination $(\lambda, \mu) = (0.32, 0)$ was chosen in 4/10 instances and was the combination more often selected in training.

[5] Since we do perform cross-validation, we report the in-sample error.

Table 3. Performance of the Lagrangian method for signals with $n = 100,000$. **Bold entries** correspond to the best estimation error and the fastest solution time.

κ	m	signal quality		computational performance		
		error	$\|x^*\|_0$	time	# iter	# sub
0.0005	1	**0.172**	29,690	**172**	1	1
	10	0.188	30,650	1,825	89	10+362
	100	0.187	30,648	656	100	100+3,252
	1,000	0.187	30,670	631	100	1,000+31,932
0.001	1	0.091	10,790	506	1	1
	10	0.091	10,789	3,391	62	10+257
	100	0.091	10,781	613	100	100+2,455
	1,000	**0.090**	10,760	**475**	100	1,000+23,174
0.002	1	**0.027**	2,523	1,390	1	1
	10	**0.027**	2,511	1,703	22	10+94
	100	**0.027**	2,510	280	100	100+848
	1,000	**0.027**	2,502	**173**	100	1,000+7,861
0.005	1	**0.008**	878	5,579	1	1
	10	**0.008**	877	309	12	10+20
	100	**0.008**	877	51	17	100+54
	1,000	**0.008**	878	**31**	61	1,000+480
0.01	1	**0.013**	758	2,141	1	1
	10	**0.013**	758	174	1	10+0
	100	**0.013**	759	81	18	100+38
	1,000	**0.013**	761	**49**	71	1,000+347
0.02	1	**0.028**	648	2,184	1	1
	10	0.030	637	185	1	10+0
	100	**0.028**	646	89	14	100+27
	1,000	**0.028**	649	**44**	62	1,000+275

optimal decompositions of an M-matrix into pairwise terms. We also gave extended conic quadratic formulations of the convex relaxations, allowing the use of off-the-shelf conic solvers. The approach is general enough to permit the addition of multiple priors in the form of additional constraints.

The proposed iterative convexification approach substantially closes the gap between the ℓ_0-"norm" and its ℓ_1 surrogate and results in significantly better estimators than the standard approaches using ℓ_1 approximations. In fact, near-optimal solution of the ℓ_0-problems are obtained in seconds for instances with over 10,000 variables, and the method scales to instances with 100,000 variables using tailored algorithms.

In addition to better inference properties, the proposed models and resulting estimators are easily *interpretable*. On the one hand, unlike ℓ_1-approximations and related estimators, the sparsity of the proposed estimators is close to the target sparsity parameter k. Thus, a prior on the sparsity of the signal can be naturally fed to the inference problems. On the other hand, the proposed strong convex relaxations compare favorably to ℓ_1-approximations in classification or spike inference purposes: the 0-1 variables can be easily used to assign a category to each observation via simple rounding heuristics, and resulting in high-quality solutions.

Acknowledgments

A. Atamtürk is supported, in part, by grant FA9550-10-1-0168 from the Office of the Assistant Secretary of Defense for Research and Engineering and grant 1807260 from the National Science Foundation. A. Gómez is supported, in part, by the National Science Foundation under Grant No. 1818700.

References

[1] Ahuja, R. K., Hochbaum, D. S., and Orlin, J. B. (2004). A cut-based algorithm for the nonlinear dual of the minimum cost network flow problem. *Algorithmica*, 39:189–208.

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

[3] Alizadeh, F. and Goldfarb, D. (2003). Second-order cone programming. *Mathematical Programming*, 95:3–51.

[4] Atamtürk, A. (2007). Strong formulations of robust mixed 0-1 programming. *Mathematical Programming*, 108:235–250.

[5] Atamtürk, A. and Gómez, A. (2016). Submodularity in conic quadratic mixed 0-1 optimization. *arXiv preprint arXiv:1705.05918*. BCOL Research Report 16.02, UC Berkeley. Forthcoming in *Operations Research*.

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

[7] Atamtürk, A. and Gómez, A. (2020). Safe screening rules for ℓ_0-regression. *Forthcoming in 2020 International Conference of Machine Learning.*

[8] Atamtürk, A. and Narayanan, V. (2007). Cuts for conic mixed-integer programming. In Fischetti, M. and Williamson, D. P., editors, *Integer Programming and Combinatorial Optimization*, pages 16–29, Berlin, Heidelberg. Springer.

[9] Atamtürk, A. and Narayanan, V. (2010). Conic mixed-integer rounding cuts. *Mathematical Programming*, 122:1–20.

[10] Bach, F. (2016). Submodular functions: from discrete to continuous domains. *Mathematical Programming*, 175:1–41.

[11] Bach, F. R. (2008). Consistency of the group lasso and multiple kernel learning. *Journal of Machine Learning Research*, 9:1179–1225.

[12] Bao, L. and Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, pages 1–17. Springer.

[13] Berman, A. and Plemmons, R. J. (1994). *Nonnegative matrices in the mathematical sciences*, volume 9. SIAM.

[14] Bertsimas, D. and King, A. (2015). OR forum – an algorithmic approach to linear regression. *Operations Research*, 64:2–16.

[15] Bertsimas, D., King, A., Mazumder, R., et al. (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44:813–852.

[16] Bertsimas, D., Van Parys, B., et al. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *The Annals of Statistics*, 48(1):300–323.

[17] Boman, E. G., Chen, D., Parekh, O., and Toledo, S. (2005). On factor width and symmetric H-matrices. *Linear Algebra and Its Applications*, 405:239–248.

[18] Bonami, P., Lodi, A., Tramontani, A., and Wiese, S. (2015). On mathematical programming with indicator constraints. *Mathematical Programming*, 151:191–223.

[19] Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23:1222–1239.

[20] Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25:21–30.

[21] Candes, E. J., Wakin, M. B., and Boyd, S. P. (2008). Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905.

[22] Casale, P., Pujol, O., and Radeva, P. (2011). Human activity recognition from accelerometer data using a wearable device. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 289–296. Springer.

[23] Casale, P., Pujol, O., and Radeva, P. (2012). Personalization and user verification in wearable systems using biometric walking patterns. *Personal and Ubiquitous Computing*, 16:563–580.

[24] Chen, S. S., Donoho, D. L., and Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM review*, 43:129–159.

[25] Cozad, A., Sahinidis, N. V., and Miller, D. C. (2014). Learning surrogate models for simulation-based optimization. *AIChE Journal*, 60:2211–2227.

[26] Dheeru, D. and Karra Taniskidou, E. (2017). UCI machine learning repository.

[27] Dong, H. (2019). On integer and MPCC representability of affine sparsity. *Operations Research Letters*, 47(3):208–212.

[28] Dong, H., Ahn, M., and Pang, J.-S. (2019). Structural properties of affine sparsity constraints. *Mathematical Programming*, 176(1-2):95–135.

[29] Dong, H., Chen, K., and Linderoth, J. (2015). Regularization vs. relaxation: A conic optimization perspective of statistical variable selection. *arXiv preprint arXiv:1510.06083*.

[30] Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306.

[31] Donoho, D. L., Elad, M., and Temlyakov, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52:6–18.

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

[33] Frank, L. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35:109–135.

[34] Friedrich, J., Zhou, P., and Paninski, L. (2017). Fast online deconvolution of calcium imaging data. *PLoS computational biology*, 13.

[35] Gao, Y.-m. and Wang, X.-h. (1992). Criteria for generalized diagonally dominant matrices and M-matrices. *Linear Algebra and its Applications*, 169:257–268.

[36] Gómez, A. and Prokopyev, O. (2018). A mixed-integer fractional optimization approach to best subset selection. *http://www.optimization-online.org/DB_HTML/2018/08/6791.html*.

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

[38] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 1. Springer series in statistics New York, NY, USA:.

[39] Hastie, T., Tibshirani, R., and Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv preprint arXiv:1707.08692*.

[40] Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC press.

[41] Hazimeh, H. and Mazumder, R. (2018). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. *arXiv preprint arXiv:1803.01454*.

[42] Hazimeh, H., Mazumder, R., and Saab, A. (2020). Sparse regression at scale: Branch-and-bound rooted in first-order optimization. *arXiv preprint arXiv:2004.06152*.

[43] Hebiri, M., Van De Geer, S., et al. (2011). The smooth-lasso and other $\ell_1 + \ell_2$-penalized methods. *Electronic Journal of Statistics*, 5:1184–1226.

[44] Hijazi, H., Bonami, P., Cornuéjols, G., and Ouorou, A. (2012). Mixed-integer nonlinear programs featuring "on/off" constraints. *Computational Optimization and Applications*, 52:537–558.

[45] Hiriart-Urruty, J.-B. and Lemaréchal, C. (2013). *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media.

[46] Hochbaum, D. S. (2001). An efficient algorithm for image segmentation, Markov random fields and related problems. *Journal of the ACM (JACM)*, 48:686–701.

[47] Hochbaum, D. S. (2013). Multi-label markov random fields as an efficient and effective tool for image segmentation, total variations and regularization. *Numerical Mathematics: Theory, Methods and Applications*, 6:169–198.

[48] Hochbaum, D. S. and Liu, S. (2018). Adjacency-clustering and its application for yield prediction in integrated circuit manufacturing. *Operations Research*, 66(6):1571–1585.

[49] Huang, J., Ma, S., and Zhang, C.-H. (2008). Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica*, pages 1603–1618.

[50] Jeon, H., Linderoth, J., and Miller, A. (2017). Quadratic cone cutting surfaces for quadratic programs with on–off constraints. *Discrete Optimization*, 24:32–50.

[51] Jewell, S. and Witten, D. (2017). Exact spike train inference via ℓ_0 optimization. *arXiv preprint arXiv:1703.08644*.

[52] Kim, S.-J., Koh, K., Boyd, S., and Gorinevsky, D. (2009). ℓ_1 trend filtering. *SIAM review*, 51:339–360.

[53] Kimura, K. and Waki, H. (2018). Minimization of Akaike's information criterion in linear regression analysis via mixed integer nonlinear program. *Optimization Methods and Software*, 33(3):633–649.

[54] Kleinberg, J. and Tardos, E. (2002). Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields. *Journal of the ACM (JACM)*, 49:616–639.

[55] Kolmogorov, V. and Zabin, R. (2004). What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:147–159.

[56] Lin, X., Pham, M., and Ruszczyński, A. (2014). Alternating linearization for structured regularization problems. *Journal of Machine Learning Research*, 15:3447–3481.

[57] Lobo, M. S., Vandenberghe, L., Boyd, S., and Lebret, H. (1998). Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228.

[58] Lustig, M., Donoho, D., and Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58:1182–1195.

[59] Mahajan, A., Leyffer, S., Linderoth, J., Luedtke, J., and Munson, T. (2017). Minotaur: A mixed-integer nonlinear optimization toolkit. Technical report, ANL/MCS-P8010-0817, Argonne National Lab.

[60] Mammen, E., van de Geer, S., et al. (1997). Locally adaptive regression splines. *The Annals of Statistics*, 25:387–413.

[61] Mazumder, R., Friedman, J. H., and Hastie, T. (2011). Sparsenet: Coordinate descent with nonconvex penalties. *Journal of the American Statistical Association*, 106:1125–1138.

[62] Mazumder, R., Radchenko, P., and Dedieu, A. (2017). Subset selection with shrinkage: Sparse linear modeling when the SNR is low. *arXiv preprint arXiv:1708.03288.*

[63] Miller, A. (2002). *Subset Selection in Regression.* CRC Press.

[64] Nemirovski, A. S. and Todd, M. J. (2008). Interior-point methods for optimization. *Acta Numerica*, 17(1):191–234.

[65] Nevo, D. and Ritov, Y. (2017). Identifying a minimal class of models for high-dimensional data. *Journal of Machine Learning Research*, 18:797–825.

[66] Padilla, O. H. M., Sharpnack, J., Scott, J. G., and Tibshirani, R. J. (2018). The DFS fused lasso: Linear-time denoising over general graphs. *Journal of Machine Learning Research*, 18(176):1–36.

[67] Pilanci, P., Wainwright, M. J., and El Ghaoui, L. (2015). Sparse learning via boolean relaxations. *Mathematical Programming*, 151:63–87.

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

[69] Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317:314.

[70] Qin, Z. and Goldfarb, D. (2012). Structured sparsity via alternating direction methods. *Journal of Machine Learning Research*, 13:1435–1468.

[71] Rinaldo, A. et al. (2009). Properties and refinements of the fused lasso. *The Annals of Statistics*, 37:2922–2952.

[72] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60:259–268.

[73] Shen, X., Pan, W., Zhu, Y., and Zhou, H. (2013). On constrained and regularized high-dimensional regression. *Annals of the Institute of Statistical Mathematics*, 65:807–832.

[74] Shepard, E. L., Wilson, R. P., Quintana, F., Laich, A. G., Liebsch, N., Albareda, D. A., Halsey, L. G., Gleiss, A., Morgan, D. T., Myers, A. E., et al. (2008). Identification of animal movement patterns using tri-axial accelerometry. *Endangered Species Research*, 10:47–60.

[75] Tawarmalani, M. and Sahinidis, N. V. (2005). A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming*, 103(2):225–249.

[76] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

[77] Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73:273–282.

[78] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67:91–108.

[79] Tibshirani, R. J. et al. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42:285–323.

[80] Tibshirani, R. J. and Taylor, J. (2011). The solution path of the generalized lasso. *Annals of Statistics*, 39:1335–1371.

[81] Varga, R. S. (1976). On recurring theorems on diagonal dominance. *Linear Algebra and its Applications*, 13(1-2):1–9.

[82] Vielma, J. P., Dunning, I., Huchette, J., and Lubin, M. (2017). Extended formulations in mixed integer conic quadratic programming. *Mathematical Programming Computation*, 9(3):369–418.

[83] Vogel, C. R. and Oman, M. E. (1996). Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17:227–238.

[84] Vogelstein, J. T., Packer, A. M., Machado, T. A., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104:3691–3704.

[85] Wilson, R. P., Shepard, E., and Liebsch, N. (2008). Prying into the intimate details of animal lives: Use of a daily diary on animals. *Endangered Species Research*, 4:123–137.

[86] Wilson, Z. T. and Sahinidis, N. V. (2017). The ALAMO approach to machine learning. *Computers & Chemical Engineering*, 106:785–795.

[87] Wu, B., Sun, X., Li, D., and Zheng, X. (2017). Quadratic convex reformulations for semicontinuous quadratic programming. *SIAM Journal on Optimization*, 27:1531–1553.

[88] Xie, W. and Deng, X. (2018). Scalable algorithms for the sparse ridge regression. *arXiv preprint arXiv:1806.03756.*

[89] Yang, A. Y., Sastry, S. S., Ganesh, A., and Ma, Y. (2010). Fast ℓ_1-minimization algorithms and an application in robust face recognition: A review. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1849–1852. IEEE.

[90] Zhang, C.-H. et al. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38:894–942.

[91] Zhang, Y., Wainwright, M. J., and Jordan, M. I. (2014). Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In *Conference on Learning Theory*, pages 921–948.

[92] Zheng, Z., Fan, Y., and Lv, J. (2014). High dimensional thresholded regression and shrinkage effect. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76:627–649.

[93] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429.

[94] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67:301–320.