

# ON POLYNOMIAL-TIME SOLVABILITY OF COMBINATORIAL MARKOV RANDOM FIELDS

SHAONING HAN, ANDRÉS GÓMEZ, AND JONG-SHI PANG

September 2022

**ABSTRACT.** The problem of inferring Markov random fields (MRFs) with a sparsity or robustness prior can be naturally modeled as a mixed-integer program. This motivates us to study a general class of convex submodular optimization problems with indicator variables, which we show to be polynomially solvable in this paper. The key insight is that, possibly after a suitable reformulation, indicator constraints preserve submodularity. Fast computations of the associated Lovász extensions are also discussed under certain smoothness conditions, and can be implemented using only linear-algebraic operations in the case of quadratic objectives.

**Keywords.** Submodularity, mixed-integer optimization, Markov random fields, sparsity, robustness

## 1. INTRODUCTION

Markov random fields (MRFs) are popular graphical models pervasively used to represent spatio-temporal processes. They are defined on an undirected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where there is random variable  $X_i$  associated with each vertex  $i \in \mathcal{V}$ . Each edge  $[i, j] \in \mathcal{E}$  represents the a relationship between the variables at their respective nodes  $i$  and  $j$ ; usually, these two variables should take similar values. Moreover, variables not connected by an edge are conditionally independent given realizations of all other variables. In the MRF inference problems we consider, noisy realizations  $\{a_i\}_{i \in \mathcal{V}}$  of the random variables  $\mathbf{X}$  are observed, and the goal is to infer the true values of

---

S. Han, A. Gómez, and J.S. Pang: Daniel J. Epstein Department of Industrial and Systems Engineering, Viterbi School of Engineering, University of Southern California, CA 90089. [shaoning@usc.edu](mailto:shaoning@usc.edu), [gomezand@usc.edu](mailto:gomezand@usc.edu), [jongship@usc.edu](mailto:jongship@usc.edu). This work was based on research partially supported by the National Science Foundation under grant 2006762 and the U.S. Air Force Office of Scientific Research under grants FA9550-22-1-0369 and FA9550-18-1-0382.

**X**. Figure 1 provides a depiction of this problem for three commonly-used structures of MRFs.

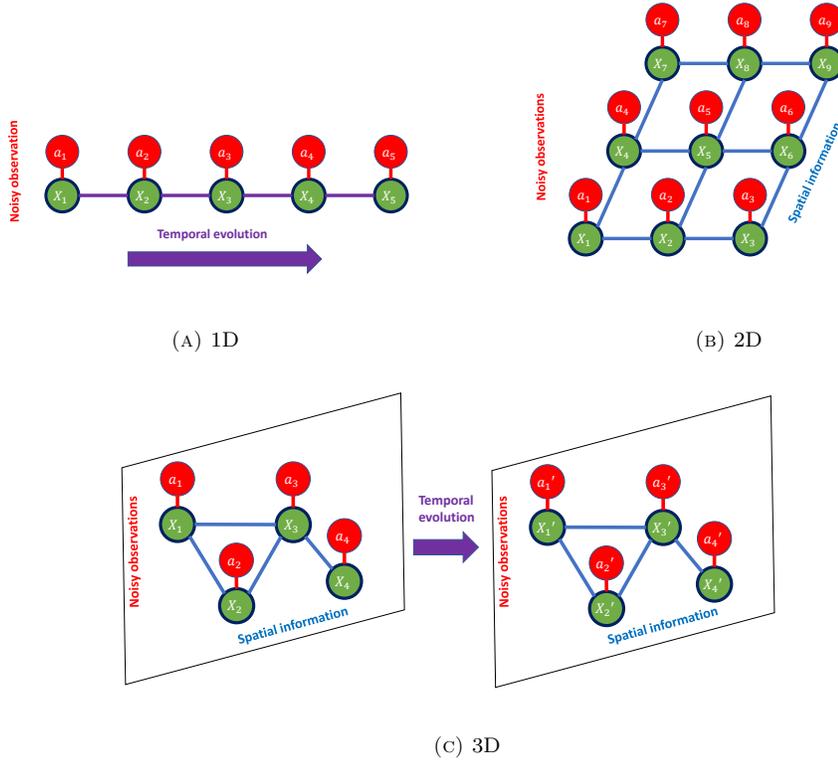


FIGURE 1. Common topologies of MRFs, modeling spatial (blue) and temporal (purple) relationships. The true values of random variables  $X$  (green) are not observed directly, and need to be inferred from the noisy observations  $a$  (red).

One-dimensional MRFs as depicted in Figure 1 (A) are fundamental building blocks in time series analysis and signal processing [3, 34, 44, 45, 52, 53]. They are typically used to model the evolution of a given process or signal over time. Two-dimensional MRFs as depicted in Figure 1 (B) arise pervasively in image denoising [13, 14, 31, 32, 40] and computer vision [25]. Each variable  $X_i$  encodes the “true” value of a pixel in an image, and edges encode the belief that adjacent pixels tend to have similar values. Three-dimensional MRFs as depicted in Figure 1 (C) are used to model spatio-temporal processes [22]. They are used in epidemiology [12, 39, 46] for example to track the spread of a disease over time. In addition, MRFs over general graphs arise in semiconductor manufacturing [21, 33], bioinformatics [20], criminology [41], spam detection [35], among other applications.

*Maximum a posteriori* estimates of the values of  $\mathbf{X}$  can often be obtained as optimal solutions of the (continuous) MRF problem [31]

$$\underset{\ell \leq \mathbf{x} \leq \mathbf{u}}{\text{minimize}} \sum_{i \in \mathcal{V}} h_i(x_i - a_i) + \sum_{[i,j] \in \mathcal{E}} g_{ij}(x_i - x_j), \quad (1)$$

where  $h_i : \mathbb{R} \rightarrow \mathbb{R}_+$  and  $g_{ij} : \mathbb{R} \rightarrow \mathbb{R}_+$  are appropriate convex nonnegative one-dimensional functions such that  $h_i(0) = g_{ij}(0) = 0$ , and  $\ell \leq \mathbf{u}$  are (possibly infinite) lower and upper bounds, respectively, on the values of  $\mathbf{X}$ . Functions  $h_i$  and  $g_{ij}$  are chosen depending on the prior distribution of the random variables and noise. Typically, functions  $h_i$  are quadratic, corresponding to cases with Gaussian noise. The most common choices for functions  $g_{ij}$  are absolute value functions  $g_{ij}(x_i - x_j) = c_{ij}|x_i - x_j|$  with  $c_{ij} \geq 0$ , popular in statistics and signal processing [57, 18] and referred to as total variation denoising problems, and quadratic functions  $g_{ij}(x_i - x_j) = c_{ij}(x_i - x_j)^2$ , in which case the graphical model is a Gaussian MRF (GMRF) and also corresponds to a Besag model [10, 11].

Clearly, problem (1) is convex and can be solved using standard tools in the convex optimization literature. Specialized algorithms have also been proposed [1, 31], whose complexity is strongly polynomial for the special cases of total variation and Besag models (see also [32] and the references therein). In this paper, we study two combinatorial extensions of (1). The first extension corresponds to the situation where  $\mathbf{X}$  is sparse or, more generally, is assumed to take a baseline value (e.g., corresponding to the background of an image or the absence of a disease) in most of its coordinates. In such cases, statistical theory calls for the imposition of an  $\ell_0$  regularization to penalize variables that differ from the baseline value. The second extension corresponds to the situation where the noisy observations are corrupted by a few but potentially gross outliers. In such cases, statistical theory calls for the simultaneous removal of data identified as corrupted and solution of (1). Both extensions involve combinatorial decisions: which random variables differ from the baseline value, and which datapoints should be discarded.

It is well known that linear regression, one of the simplest statistical estimation methods, becomes NP-hard with the inclusion of either sparsity [48] or robustness [7] as described above. Thus, approaches in the literature resort to approximations of the combinatorial problems, heuristics, or expensive mixed-integer optimization approaches to solve the exact problems. *In this paper we show that for the case of (1), the aforementioned combinatorial extensions can in fact be solved in polynomial time by a reduction to submodular minimization.* We point out that an immediate application of submodular minimization techniques [50] results in runtimes of  $\mathcal{O}(n^5 \cdot \text{OE})$ ,

where OE is the complexity of solving problem (1) – resulting for example in strongly polynomial but impractical complexities of  $\mathcal{O}(n^8)$  for the case of total variation and Besag models, but those runtimes can likely be improved (we present such an improvement in this paper). Indeed, the discovery of a (strongly) polynomial time algorithm for a problem has typically been closely followed by highly efficient methods.

**Outline.** In §2 we discuss the necessary preliminaries for the paper and introduce mixed-integer optimization formulations of the problems we consider. In §3 we show that the aforementioned variants of (1) can be reduced to the minimization of a submodular set function – a class of problems which admits polynomial time algorithms [37, 38]. In §4 we describe parametric procedures that can be used to improve the complexity of submodular minimization algorithms for this class of problems. Finally, in §5 we conclude the paper.

## 2. PRELIMINARIES

In this section we give explicit mixed-integer optimization (MIO) formulations of the problems under consideration, and review the concepts and literature relevant to the paper. First, we introduce the notation used throughout the paper.

**Notation.** We use bold symbols to denote vectors and matrices. Given  $n \in \mathbb{Z}_{++}$ , we let  $[n] \stackrel{\text{def}}{=} \{1, \dots, n\}$ . We denote the vector of all zeros by  $\mathbf{0}$  and the vector of ones by  $\mathbf{1}$  (whose dimensions can be inferred from the context). Given a set  $\mathcal{S} \subseteq \mathbb{R}^n$ , we let  $\mathbf{1}_{\mathcal{S}} \in \{0, 1\}^n$  be the indicator vector of  $\mathcal{S}$ , i.e.,  $(\mathbf{1}_{\mathcal{S}})_i = 1$  if  $i \in \mathcal{S}$  and is zero otherwise. Given  $i \in [n]$ , we also let  $\mathbf{e}^i$  be the  $i$ -th basis vector of  $\mathbb{R}^n$ . Given two vectors  $\mathbf{y}^1$  and  $\mathbf{y}^2 \in \mathbb{R}^n$ , define the *meet*  $\mathbf{y}^1 \wedge \mathbf{y}^2 \in \mathbb{R}^n$  and the *join*  $\mathbf{y}^1 \vee \mathbf{y}^2 \in \mathbb{R}^n$  to be the component-wise minimum and maximum of  $\mathbf{y}^1$  and  $\mathbf{y}^2$ , respectively; we also define  $\mathbf{y}^1 \circ \mathbf{y}^2$  as the Hadamard (entrywise) product. By above notations, a set  $\mathcal{L} \subseteq \mathbb{R}^n$  is called a *lattice* if any  $\mathbf{y}^1, \mathbf{y}^2 \in \mathcal{L}$  implies that  $\mathbf{y}^1 \vee \mathbf{y}^2$  and  $\mathbf{y}^1 \wedge \mathbf{y}^2$  belong to  $\mathcal{L}$ . A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is *submodular* over a lattice  $\mathcal{L}$  if for any  $\mathbf{y}^1$  and  $\mathbf{y}^2 \in \mathcal{L}$ , one has  $f(\mathbf{y}^1) + f(\mathbf{y}^2) \geq f(\mathbf{y}^1 \wedge \mathbf{y}^2) + f(\mathbf{y}^1 \vee \mathbf{y}^2)$ . We denote  $\underline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{-\infty\}$  and  $\overline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{\infty\}$  and we adopt the convention that  $0 \cdot (\pm\infty) = 0$ . For example, given decision variables  $z \in \{0, 1\}$  and  $x \in \mathbb{R}$ , constraint  $|x| \leq uz$  with  $u = \infty$  is equivalent to the complementarity constraint  $x(1 - z) = 0$ .

**2.1. MIO formulations of combinatorial MRF inference problems.** We now formally define the two combinatorial extensions of problem (1) discussed in the introduction: the sparse MRF inference problem and the robust MRF inference problem.

2.1.1. *Sparse MRF inference.* If the underlying statistical process  $\mathbf{X}$  is known to be sparse (e.g., most pixels in an image adopt the background color, or the disease under study is absent from most locations), then a sparsity prior can be included in (1), resulting in problems of the form

$$\underset{\mathbf{x} \in \mathbb{R}^{\mathcal{V}}, \mathbf{z} \in \{0,1\}^{\mathcal{V}}}{\text{minimize}} \quad \sum_{i \in \mathcal{V}} h_i(x_i - a_i) + \sum_{[i,j] \in \mathcal{E}} g_{ij}(x_i - x_j) + \sum_{i \in \mathcal{V}} c_i z_i \quad (2a)$$

$$\text{subject to } \ell \circ \mathbf{z} \leq \mathbf{x} \leq \mathbf{u} \circ \mathbf{z}, \quad (2b)$$

where  $\mathbf{c} \geq 0$  and binary variables  $\mathbf{z}$  are used to indicate the support of  $\mathbf{x}$  – note that while solutions satisfying  $z_i = 1$  and  $x_i = 0$  are feasible, since  $c_i \geq 0$  there always exist an optimal solution where  $z_i = 0$  if  $x_i = 0$ . If all coefficients  $c_i$  are equal, that is,  $\mathbf{c} = \mu \mathbf{1}$  for some  $\mu \geq 0$ , then in optimal solutions of (2) we have that  $\sum_{i \in \mathcal{V}} c_i z_i = \mu \|\mathbf{x}\|_0$ , where  $\|\mathbf{x}\|_0$  stands for the number of nonzero components of  $\mathbf{x}$  and is known as the “ $\ell_0$ -norm” pervasively used in statistics. Alternatively, if priors on the probabilities  $p_i < 0.5$  that variable  $X_i$  is non-zero are available, then one can set  $c_i \propto \ln((1 - p_i)/p_i)$ . Note that if  $\mathbf{X}$  adopts a non-zero baseline value in most of its coordinates, the problem can be transformed into (2) through a change of variables.

Using MIO to model inference problems with sparsity is by now a standard approach in statistics and machine learning [8, 9, 17, 61]. Most existing approaches focus on problems with quadratic functions – probably due to the availability of powerful off-the-shelf MIO solvers capable of handling such functions. State-of-the-art methods revolve around the perspective relaxation [2, 23, 28]: if  $h_i(x_i - a_i) = (x_i - a_i)^2$ , then we can replace such terms with the reformulation  $\hat{h}_i(x_i, z_i) = a_i^2 - 2a_i x_i + x_i^2/z_i$ , where we adopt the following convention of division by 0:  $x^2/z = 0$  if  $x = z = 0$ , and  $x^2/z = \infty$  if  $x \neq 0$  and  $z = 0$ . Indeed, this conic quadratic reformulation is exact if  $z_i \in \{0, 1\}$ , but results in stronger continuous relaxations whenever  $z_i$  is fractional. Tailored branch-and-bound algorithms [29], approximation algorithms [62] and presolving techniques [5] which exploit the perspective reformulation have been proposed in the literature. Finally, Atamtürk et al. [6] derive improved conic relaxations specific to problem (2) for the case of quadratic functions with  $\ell = \mathbf{0}$ .

Two special cases of (2) have been identified to be polynomial-time solvable. First, if graph  $\mathcal{G}$  is a path, then (2) can be solved via dynamic programming [42]. Second, all functions are quadratic,  $u_i = \infty$  for all  $i \in \mathcal{V}$  and  $\mathbf{a} \geq 0$ , then (2) can be reformulated as a binary submodular problem [4] and thus be solved in polynomial time. *In this paper, we show that such a submodular reformulation of (2) is always possible, regardless of the bounds, observations  $\mathbf{a}$  or (convex) functions  $h_i$  and  $g_{ij}$ .*

2.1.2. *Robust MRF inference.* If the noisy observations  $\mathbf{a}$  are corrupted by gross outliers, then the estimates resulting from (1) can be poor. Classical robust estimation methods in statistics [54, 55] call for the removal of outliers such that the objective (1) is minimized, that is, solving the optimization problem

$$\underset{\ell \leq \mathbf{x} \leq \mathbf{u}, \mathbf{z} \in \{0,1\}^{\mathcal{V}}}{\text{minimize}} \sum_{i \in \mathcal{V}} h_i(x_i - a_i)(1 - z_i) + \sum_{[i,j] \in \mathcal{E}} g_{ij}(x_i - x_j) + \sum_{i \in \mathcal{V}} c_i z_i, \quad (3)$$

where  $z_i = 1$  iff observation  $i$  is discarded. Robust estimators such as (3) are, in general, hard to compute [7]. In the context of least squares linear regression, the associated robust estimator is called the Least Trimmed Squares [56], which is even hard to approximate [47]. Exact optimization methods [63, 64] rely on reformulations such as

$$\underset{\mathbf{x}, \mathbf{z}, \mathbf{w}}{\text{minimize}} \sum_{i \in \mathcal{V}} h_i(x_i - w_i - a_i) + \sum_{[i,j] \in \mathcal{E}} g_{ij}(x_i - x_j) + \sum_{i \in \mathcal{V}} c_i z_i \quad (4a)$$

$$\text{subject to } w_i(1 - z_i) = 0 \quad \forall i \in \mathcal{V} \quad (4b)$$

$$\mathbf{x} \in [\ell, \mathbf{u}]^{\mathcal{V}}, \mathbf{z} \in \{0,1\}^{\mathcal{V}}, \mathbf{w} \in \mathbb{R}^{\mathcal{V}}. \quad (4c)$$

Indeed, since  $h$  is nonnegative and  $h(0) = 0$ , we find that if  $z_i = 1$ , then  $w_i = x_i - a_i$  in any optimal solution and the associated term vanishes; on the other hand, if  $z_i = 0$ , then  $w_i = 0$  and  $h_i(x_i - w_i - a_i) = h(x_i - a_i)$  as intended. Constraints (4b) are typically reformulated as big-M constraints; unfortunately, the ensuing continuous relaxation is trivial (e.g.,  $\mathbf{x} = \mathbf{0}$ ,  $\mathbf{z} \rightarrow \mathbf{0}$ ,  $\mathbf{w} = \mathbf{x} - \mathbf{a}$  in optimal solutions of the convex relaxations, and the objective value is almost 0), thus the methods do not scale well. A stronger, big-M free, reformulation was proposed in [26] for the special case where  $\mathcal{G}$  is a path and all functions are convex quadratic.

Note that NP-hardness of robust estimators in general, and Trimmed Least Squares in particular, does not imply that (4) is NP-hard. *In fact, we show in this paper that it is polynomial-time solvable for arbitrary convex functions  $h_i$  and  $g_{ij}$  and arbitrary graphs  $\mathcal{G}$ .*

2.2. **Submodular minimization with indicators.** Problems (2) and (4) are both special cases of *continuous submodular minimization problems with indicators*, which are the focus of this paper. Specifically, the problems we consider are of the form

$$\underset{\mathbf{x} \in \mathbb{R}^n, \mathbf{z} \in \{0,1\}^n}{\text{minimize}} \left\{ f(\mathbf{x}) + \mathbf{c}^\top \mathbf{z} : \ell \circ \mathbf{z} \leq \mathbf{x} \leq \mathbf{u} \circ \mathbf{z} \right\} \quad (5)$$

where:

- (1) Function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex and (continuous) submodular. Moreover, we assume that the minimum value of (5) is finite.

(2) Bounds  $\ell \in \underline{\mathbb{R}}^n$  and  $\mathbf{u} \in \overline{\mathbb{R}}^n$  are possibly infinite, and satisfy  $\ell \leq \mathbf{u}$ .

Observe that problem (5) assumes there is an indicator variable associated with each continuous variable. This assumption is without loss of generality, as it is always possible to introduce artificial binary variables with cost  $c_i = 0$  to transform general problems into the form (5). Also note that we do not assume that  $\ell_i \leq 0 \leq u_i$  for any  $i \in [n]$ , and thus (5) is general enough to include the constraint that either a continuous variable is zero, or it is bounded away from zero.

Submodular functions of binary variables arise pervasively in combinatorial optimization [49, 19], including for example cut capacity functions of networks and rank functions of matroids, and are often associated with tractable discrete optimization problems [24, 36]. However, less effort has been devoted to studying structured submodular problems involving both continuous and discrete variables such as (5). Proposition 1 below provides several equivalent definitions of submodular functions; see [59] for their reference.

**Proposition 1.** *The following statements hold true.*

- (Zeroth-order definition) *A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is submodular if and only if for all  $c_i, c_j > 0, i \neq j$  and  $\mathbf{y} \in \mathbb{R}^n$ , it holds that*

$$f(\mathbf{y} + c_i \mathbf{e}^i) + f(\mathbf{y} + c_j \mathbf{e}^j) \geq f(\mathbf{y}) + f(\mathbf{y} + c_i \mathbf{e}^i + c_j \mathbf{e}^j).$$

- (First-order definition) *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable, then  $f$  is submodular if and only if  $\frac{\partial f}{\partial y_i}(\mathbf{y} + c_1 \mathbf{e}^i) \leq \frac{\partial f}{\partial y_i}(\mathbf{y} + c_2 \mathbf{e}^i)$  for all  $\mathbf{y} \in \mathbb{R}^n$ ,  $i \neq j$  and  $c_1 \geq c_2$ .*
- (Second-order definition) *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is twice differentiable, then  $f$  is submodular if and only if  $\frac{\partial^2 f(\mathbf{y})}{\partial y_i \partial y_j} \leq 0$  for all  $\mathbf{y} \in \mathbb{R}^n$  and  $i \neq j$ .*

From the second-order definition we find that any function of the form  $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{Q} \mathbf{x}$  is (strongly) convex submodular if  $\mathbf{Q}$  a Stieltjes matrix, that is,  $\mathbf{Q} \succ 0$  and  $Q_{ij} \leq 0$  for all  $i \neq j$ . Moreover, functions  $g_{ij}(x_i - x_j)$  and  $h_i(x_i - w_i - a_i)$  appearing in (2) and (4) are compositions of a convex function and a difference function, which can be easily verified to be submodular; see [60]; since a sum of submodular functions is submodular, it follows that the objective functions of both (2) and (4) are submodular.

### 3. EQUIVALENCE WITH BINARY SUBMODULAR MINIMIZATION

In this section, we show that (5) can be reduced to a binary submodular minimization problem (under additional mild conditions). Our derivations are based on the fact that complementarity constraints preserve (to some degree) the lattice structure, and rely on the following lemma.

**Lemma 1** (Topkis [59], Theorem 4.2). *Given lattices  $\mathcal{U}$  and  $\mathcal{W}$ , assume function  $\phi : \mathcal{U} \times \mathcal{W} \rightarrow \mathbb{R}$  is submodular on a sublattice  $\mathcal{L} \subseteq \mathcal{U} \times \mathcal{W}$ . If*

$$\psi(\mathbf{w}) \stackrel{\text{def}}{=} \underset{\mathbf{u}}{\text{minimum}} \{ \phi(\mathbf{u}, \mathbf{w}) : (\mathbf{u}, \mathbf{w}) \in \mathcal{L} \} > -\infty \quad \forall \mathbf{w} \in \mathcal{W},$$

*then the marginal function  $\psi$  is submodular on the lattice  $\text{proj}_{\mathcal{W}}(\mathcal{L}) \stackrel{\text{def}}{=} \{ \mathbf{w} \in \mathcal{W} : \exists \mathbf{u} \in \mathcal{U} \text{ s.t. } (\mathbf{u}, \mathbf{w}) \in \mathcal{L} \}$ .*

**3.1. Nonnegative case.** In this subsection, we assume  $\ell \geq 0$ , that is,  $\mathbf{x}$  is nonnegative. Given  $u \in \overline{\mathbb{R}}$ , define

$$\mathcal{L}_+ = \{ (x, z) \in \mathbb{R} \times \{0, 1\} : \ell z \leq x \leq uz \} \quad (6)$$

**Lemma 2.** *If  $0 \leq \ell \leq u$ , then set  $\mathcal{L}_+$  is a lattice.*

*Proof.* Consider any  $(x_1, z_1), (x_2, z_2) \in \mathcal{L}_+$ . It suffices to prove the case of  $\min\{z_1, z_2\} = 0$  since the other case where  $\min\{z_1, z_2\} = 1$  is trivial. If  $\min\{z_1, z_2\} = 0$ , then  $z_1 = 0$  or  $z_2 = 0$ , which implies  $x_1 = 0$  or  $x_2 = 0$ . Since  $0 \leq x_1, x_2 \leq u$ , one can deduce that  $\min\{x_1, x_2\} = 0$  and  $\max\{x_1, x_2\} \leq u$ ; thus,  $(x_1, z_1) \wedge (x_2, z_2) \in \mathcal{L}_+$  and  $(x_1, z_1) \vee (x_2, z_2) \in \mathcal{L}_+$ . Therefore,  $\mathcal{L}_+$  is a lattice.  $\square$

**Theorem 1.** *If  $\ell \in \mathbb{R}_+^n$ , then the function*

$$v_+(z) \stackrel{\text{def}}{=} \underset{\mathbf{x} \in \mathbb{R}^n, \ell \circ z \leq \mathbf{x} \leq u \circ z}{\text{minimum}} f(\mathbf{x})$$

*is submodular on  $\mathcal{Z}$ .*

*Proof.* Note that the feasible region is a Cartesian product of  $n$  lattices and thus is a lattice itself. The conclusion follows from Lemma 1.  $\square$

*Remark 1.* Atamtürk and Gómez [4] show that problem

$$\underset{\mathbf{x} \in \mathbb{R}^n, \mathbf{z} \in \{0, 1\}^n}{\text{minimize}} \left\{ \mathbf{a}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{c}^\top \mathbf{z} : \mathbf{x} \geq 0, x_i(1 - z_i) = 0, \forall i \in [n] \right\}$$

reduces to a submodular optimization problem provided that  $\mathbf{Q}$  is a Stieltjes matrix and  $\mathbf{a} \leq 0$ . Theorem 1 is a direct generalization, as it does not impose conditions on  $\mathbf{a}$ , allows for arbitrary (nonnegative) variable lower bounds and arbitrary (finite or infinite) upper bounds on the continuous variables, and it holds for arbitrary (possibly non-quadratic) submodular functions.  $\square$

**3.2. General Case.** If  $\ell \not\geq 0$ , then the statement of Lemma 2 does not hold, and function  $v_+$  is not necessarily submodular. In this subsection, we allow the continuous variables to be positive or negative.

As we show in Theorem 2, (5) can still be reformulated as a submodular minimization problem with the introduction of additional binary variables. Towards this goal, given  $\ell \in \mathbb{R}$  and  $u \in \overline{\mathbb{R}}$ , define additional sets

$$\mathcal{L}_- = \{(x, z) \in \mathbb{R} \times \{0, 1\} : \ell(1 - z) \leq x \leq u(1 - z)\} \quad (7)$$

$$\mathcal{L}_\pm = \{(x, z^+, z^-) \in \mathbb{R} \times \{0, 1\} \times \{0, 1\} : \ell(1 - z^-) \leq x \leq uz^+\}. \quad (8)$$

**Lemma 3.** *If  $\ell \leq u \leq 0$ , then  $\mathcal{L}_-$  is a lattice. If  $\ell < 0 < u$ , then  $\mathcal{L}_\pm$  is a lattice.*

*Proof.* We prove just the result for  $\mathcal{L}_\pm$ , as the proof of  $\mathcal{L}_-$  is analogous to the one of Lemma 2. If  $\ell \leq 0$  and  $u \geq 0$  are finite, then

$$\mathcal{L}_\pm = \{(x, z^+, z^-) : -\ell z^- + \ell \leq x \leq uz^+\} \cap (\mathbb{R} \times \{0, 1\}^2)$$

as the intersection of two closed lattices is a closed lattice itself. In the general case where  $\ell$  and  $u$  are allowed to take infinite values, consider any  $(x_1, z_1^+, z_1^-), (x_2, z_2^+, z_2^-) \in \mathcal{L}_\pm$ . Let  $\hat{u} = x_1 \vee x_2 \leq u, \hat{\ell} = x_1 \wedge x_2 \geq \ell$ . Then  $(x_i, z_i^+, z_i^-) \in \hat{\mathcal{L}}_\pm := \{(x, z^+, z^-) \in \mathbb{R} \times \{0, 1\}^2 : \hat{\ell}(1 - z^-) \leq x \leq \hat{u}z^+\}$  for  $i = 1, 2$ . The conclusion follows from the lattice property of  $\hat{\mathcal{L}}_\pm$  and the inclusion  $\hat{\mathcal{L}}_\pm \subseteq \mathcal{L}_\pm$ .  $\square$

To reformulate (5), define  $\mathcal{N}_+ \stackrel{\text{def}}{=} \{i \in [n] : 0 \leq \ell_i \leq u_i\}$ ,  $\mathcal{N}_- \stackrel{\text{def}}{=} \{i \in [n] : \ell_i \leq u_i \leq 0\}$  and  $\mathcal{N}_\pm \stackrel{\text{def}}{=} \{i \in [n] : \ell_i < 0 < u_i\}$ . For each  $i \in \mathcal{N}_\pm$  introduce binary variables  $z_i^+ = 1$  if  $x_i > 0$  and  $z_i^- = 0$  if  $x_i < 0$ , so that we can substitute  $z_i = z_i^+ + (1 - z_i^-)$ —note that we need to add constraint  $z_i^- \geq z_i^+$  to rule out the impossible case where both  $x_i > 0$  and  $x_i < 0$ ; for convenience, for  $i \in \mathcal{N}_+$  we rename  $z_i = z_i^+$  and for  $i \in \mathcal{N}_-$  we rename  $z_i = 1 - z_i^-$ . After performing the substitutions above, we find that (5) can be formulated as

$$\underset{\mathbf{x}, \mathbf{z}^+, \mathbf{z}^-}{\text{minimize}} f(\mathbf{x}) + \sum_{i \in \mathcal{N}_+} c_i z_i^+ + \sum_{i \in \mathcal{N}_-} c_i (1 - z_i^-) + \sum_{i \in \mathcal{N}_\pm} c_i (z_i^+ + 1 - z_i^-) \quad (9a)$$

$$\text{subject to } \ell_i z_i^+ \leq x_i \leq u_i z_i^+ \quad \forall i \in \mathcal{N}_+ \quad (9b)$$

$$\ell_i (1 - z_i^-) \leq x_i \leq u_i (1 - z_i^-) \quad \forall i \in \mathcal{N}_- \quad (9c)$$

$$\ell_i (1 - z_i^-) \leq x_i \leq u_i z_i^+, \quad z_i^- \geq z_i^+ \quad \forall i \in \mathcal{N}_\pm \quad (9d)$$

$$\mathbf{x} \in \mathbb{R}^n, \mathbf{z}^+ \in \{0, 1\}^{\mathcal{N}_+ \cup \mathcal{N}_\pm}, \mathbf{z}^- \in \{0, 1\}^{\mathcal{N}_- \cup \mathcal{N}_\pm}. \quad (9e)$$

**Proposition 2.** *The set defined by constraints (9b)-(9e) is a sublattice of  $\mathbb{R}^n \times \{0, 1\}^{\mathcal{N}_+ \cup \mathcal{N}_\pm} \times \{0, 1\}^{\mathcal{N}_- \cup \mathcal{N}_\pm}$ .*

*Proof.* Each constraint involving  $(x_i, z_i)$  jointly defines a lattice by Lemma 2 and Lemma 3, and so does the constraint  $z_i^- \geq z_i^+$ .  $\square$

**Theorem 2.** *Function*

$$v_{\pm}(\mathbf{z}^+, \mathbf{z}^-) \stackrel{\text{def}}{=} \underset{\mathbf{x}}{\text{minimum}} \{f(\mathbf{x}) : (9\text{b}) - (9\text{e})\}$$

is submodular on  $\{0, 1\}^{\mathcal{N}_+ \cup \mathcal{N}_{\pm}} \times \{0, 1\}^{\mathcal{N}_- \cup \mathcal{N}_{\pm}}$ .

*Proof.* Follows directly from Lemma 1 and Proposition 2.  $\square$

*Remark 2.* In fact, Theorem 1 and Theorem 2 holds true for an arbitrary submodular (not necessarily convex) function  $f$ . However, if  $f$  is not convex, the evaluation of the value function  $v_{\pm}$  is in general not an easy task.

*Remark 3.* Observe that since  $\mathbf{c} \geq 0$ , constraints  $z_i^- \geq z_i^+$ ,  $\forall i \in \mathcal{N}_{\pm}$  can be dropped from the formulation. Indeed, if the constraints are removed and  $z_i^+ = 1$ ,  $z_i^- = 0$  in an optimal solution of the resulting problem, then setting  $z_i^+ = 0$  if  $x_i \leq 0$  or  $z_i^- = 1$  if  $x_i \geq 0$  results in a feasible with equal or better objective value.  $\square$

#### 4. FAST COMPUTATION OF THE LOVÁSZ EXTENSION

As seen in Theorems 1 and 2, solving optimization problem (5) reduces to an binary submodular minimization problem of the form

$$\underset{\mathbf{z} \in \{0,1\}^m}{\text{minimize}} v(\mathbf{z}). \quad (10)$$

Polynomial time algorithms for submodular minimization problems are often expressed in terms of the number of calls to an *evaluation oracle* for  $v$ . In the settings considered, evaluating  $v$  is an expensive process as it requires solving a (convex submodular) minimization problem. For reference, in the case of the problems discussed in §2.1, computing  $v$  requires solving a problem of the form (1), whose complexity in the context of total variation and Besag models is  $\mathcal{O}(|\mathcal{V}|^3)$  if  $\mathcal{G}$  is a complete graph, although faster algorithms with complexity  $\mathcal{O}\left(|\mathcal{V}||\mathcal{E}|\log\frac{|\mathcal{V}|^2}{|\mathcal{E}|}\right)$  exist if  $|\mathcal{E}|$  is small [32].

As pointed out in [50], rather than actual evaluations of function  $v$ , algorithms need to evaluate its Lovász extension, as defined next.

**Definition 1** (Lovász extension [43]). For any  $\mathbf{z} \in [0, 1]^n$  with  $z_1 \geq z_2 \geq \dots \geq z_n$ , the Lovász extension of  $v$  evaluated at  $\mathbf{z}$  is given by

$$\hat{v}(\mathbf{z}) \stackrel{\text{def}}{=} v(\mathbf{0}) + \sum_{i=1}^m (v(\mathbf{1}_{[i]}) - v(\mathbf{1}_{[i-1]})) z_i, \quad (11)$$

where (we recall)  $\mathbf{1}_{[i]}$  is the indicator vector of  $[i] = \{1, \dots, i\}$ .

Clearly, the Lovász extension can be computed with  $n$  evaluations of  $v$ : thus, the algorithm that requires  $\mathcal{O}(n^4)$  evaluation of the Lovász extension [50] immediately translates to an algorithm requiring  $\mathcal{O}(n^5)$  evaluations of

$v$ . In this section, we show that in some cases it is possible to compute (11) in the same complexity as a single evaluation of  $v$  using a parametric algorithm, ultimately reducing the complexity of minimization algorithms by a factor of  $n$ .

In particular, in this section we focus on the special case of (5), which we repeat for convenience,

$$\underset{\mathbf{x} \in \mathbb{R}^n, \mathbf{z} \in \{0,1\}^n}{\mathbf{minimize}} \left\{ f(\mathbf{x}) + \mathbf{c}^\top \mathbf{z} : \boldsymbol{\ell} \circ \mathbf{z} \leq \mathbf{x} \leq \mathbf{u} \circ \mathbf{z} \right\}, \quad (12)$$

where additionally we assume that  $f$  is a strictly convex differentiable submodular function on  $[\mathbf{0}, \mathbf{u}]$ ,  $\mathbf{0} \leq \boldsymbol{\ell} \leq \mathbf{u}$  and  $\|\mathbf{u}\|_\infty < \infty$ . Most of the additional assumptions are made for simplicity but are not strictly required: the proposed method traces a solution path of minima of

$$v(\mathbf{z}) = \underset{\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\ell} \circ \mathbf{z} \leq \mathbf{x} \leq \mathbf{u} \circ \mathbf{z}}{\mathbf{minimum}} f(\mathbf{x})$$

as  $\mathbf{z}$  varies: if function  $f$  is not strictly convex, then multiple minima may exist for a given variable of  $\mathbf{z}$ , and the solution path needs to choose between these optimal solutions; if additionally  $u_i = \infty$  for some indexes, then additional steps are necessary to verify whether  $v(\mathbf{z}) > -\infty$ —which can be accomplished by verifying  $v(\mathbf{1}) > -\infty$ ; finally, we discuss in §4.2 how to adapt the algorithm to cases with  $\boldsymbol{\ell} \not\leq \mathbf{0}$ .

Our goal is to evaluate all functions  $v(\mathbf{1}_{[i]})$  inductively for all  $i = 0, \dots, n$ . Suppose that we have already computed  $v(\mathbf{1}_{[k]})$  for some  $k = 0, \dots, n-1$ , and let  $\mathbf{x}^*(k)$  denote an associated optimal solution. Our goal is to compute  $v(\mathbf{1}_{[k+1]})$  and its optimal solution  $\mathbf{x}^*(k+1)$ , using  $\mathbf{x}^*(k)$  as a warm-start. Given  $k \in \{0, \dots, n-1\}$ , define function  $f_k : \mathbb{R}^k \rightarrow \mathbb{R}$  as  $f_k(\mathbf{y}) \stackrel{\text{def}}{=} f(\mathbf{y}, \mathbf{0})$  where in this case  $\mathbf{0}$  is an  $(n-k)$ -dimensional vector of zeros—if  $k=0$ , then  $f_0 \equiv f(\mathbf{0})$ . Moreover, define  $\xi_{k+1} : \mathbb{R} \rightarrow \mathbb{R}$  as

$$\xi_{k+1}(x) = \underset{\mathbf{y} \in \mathbb{R}^k}{\mathbf{minimum}} \left\{ f_{k+1}(\mathbf{y}, x) : \ell_i \leq y_i \leq u_i, i = 1, \dots, k \right\}. \quad (13)$$

Observe that  $v(\mathbf{1}_{[k]}) = \xi_{k+1}(0)$ , and

$$v(\mathbf{1}_{[k+1]}) = \underset{\ell_{k+1} \leq x \leq u_{k+1}}{\mathbf{minimum}} \xi_{k+1}(x). \quad (14)$$

Finally, given  $x \in [0, u]$ , let  $\bar{\mathbf{y}}_x^k$  denote the optimal solution of (13) corresponding to value  $x$ —this solution exists and is unique due to the assumptions we imposed in (12)—, and let  $\bar{x}^{k+1}$  be the optimal solution of (14).

The key property in the proposed algorithm follows directly from [59, Theorem 6.1].

**Lemma 4.** *If  $x_1 \leq x_2$ , then  $\bar{\mathbf{y}}_{x_1}^k \leq \bar{\mathbf{y}}_{x_2}^k$ .*

Thus to solve (14), we trace the path of all optimal solutions  $\bar{\mathbf{y}}_x^k$  as  $x$  varies first from 0 to  $\ell_{k+1}$ , and then as  $x$  varies from  $\ell_{k+1}$  to  $\bar{x}^{k+1}$ . Due to Proposition 4, as  $x$  increases, the  $i$ -th coordinate of  $\bar{\mathbf{y}}_x^k$  is non-decreasing and moves from  $(\bar{y}_0^k)_i = x^*(k)_i$  to  $x^*(k+1)_i$ . Given any fixed coordinate  $i$ , two potential *breakpoints* in the solution path may occur: if  $x^*(k)_i = \ell_i$  and  $x^*(k+1)_i > \ell_i$ , then a breakpoint occurs at  $\max\{x : (\bar{y}_x^k)_i = \ell_i\}$ ; if  $x^*(k+1)_i = u_i$  and  $x^*(k)_i < u_i$ , then a breakpoint occurs at  $\min\{x : (\bar{y}_x^k)_i = u_i\}$ .

Define  $\text{trace\_path}_{a \rightarrow b}^{k+1}(\mathbf{x})$  as a routine that receives as input  $\mathbf{x} \in \mathbb{R}^n$  whose first  $k$  coordinates correspond to  $\bar{\mathbf{y}}_a^k$ , the  $(k+1)$ -coordinate is  $x_{k+1} = a$  and remaining coordinates are 0, i.e.,  $\mathbf{x} = (\bar{\mathbf{y}}_a^k, a, \mathbf{0})$ , and outputs a vector  $(\bar{\mathbf{y}}_b^k, b, \mathbf{0}) \in \mathbb{R}^n$  by tracing the solution path  $\bar{\mathbf{y}}_x^k$  for all  $a \leq x \leq b$ . Now consider Algorithm 1, which sequentially calls the  $\text{trace\_path}$  routine to solve (14) for all values of  $k$ .

---

**Algorithm 1** Procedure to compute the optimal solutions corresponding to  $v(\mathbf{1}_{[k]})$  for all  $k = 1, \dots, n$

---

```

1:  $\mathbf{x}_B^0 \leftarrow \mathbf{0} \in \mathbb{R}^n$  ▷ All variables are initialized to 0
2: for  $k = 0, \dots, n-1$  do
3:    $\mathbf{x}_A^{k+1} \leftarrow \text{trace\_path}_{0 \rightarrow \ell_{k+1}}^{k+1}(\mathbf{x}_B^k)$  ▷ Finds feas. sol. for (14)
4:    $\mathbf{x}_B^{k+1} \leftarrow \text{trace\_path}_{\ell_{k+1} \rightarrow \bar{x}^{k+1}}^{k+1}(\mathbf{x}_A^{k+1})$  ▷ Solves (14)
5: end for
6: return  $\{f(\mathbf{x}_B^k)\}_{k \in [n]}$ 

```

---

**Proposition 3.** *Algorithm 1 computes all optimal objective values of problems  $v(\mathbf{1}_{[k]})\}_{k \in [n]}$ , and encounters  $\mathcal{O}(n)$  breakpoints during its execution.*

*Proof.* The first statement follows from the definition of the  $\text{trace\_path}$  routine: letting  $x^* = \bar{x}^{k+1}$ , solutions  $\mathbf{x}_B^{k+1}$  obtained at line 4 are of the form  $\mathbf{x}_B^{k+1} = (\bar{\mathbf{y}}_{x^*}^k, x^*, \mathbf{0})$ , which by definition are the optimal solutions of  $v(\mathbf{1}_{[k+1]})$  (padding the missing entries with 0s).

The second statement follows since the solutions produced satisfy  $\mathbf{x}_B^0 \leq \mathbf{x}_A^1 \leq \mathbf{x}_B^1 \leq \mathbf{x}_A^2 \leq \dots \leq \mathbf{x}_B^{n-1} \leq \mathbf{x}_A^n \leq \mathbf{x}_B^n$  (due to a sequential application of Lemma 4). Thus, any one coordinate  $x_i$  “leaves” its lower bound  $\ell_i$  at most once, and “reaches” its upper bound  $u_i$  at most once.  $\square$

Implementing the  $\text{trace\_path}$  routine requires either identifying the next breakpoint that will be encountered in the solution path, or identifying that an optimal solution  $\bar{x}^{k+1}$  occurs before the next breakpoint. In general, numerically identifying the next breakpoint or optimal solution is in spirit similar to solving a system of nonlinear equations defined by a so-called  $Z$ -function, which dates back to Tamir [58] and Chandrasekaran [15]. In some

cases, and notably for the case where  $f$  is quadratic, tracing the solution path can be done analytically, as we discuss next.

Given a set  $\mathcal{S}$  of variables currently fixed to their lower bounds and  $\mathcal{T}$  of variables currently fixed to their upper bounds, the optimality conditions for  $\bar{\mathbf{y}}_x^k$  associated with problem (13) can be stated as

$$\frac{\partial f_{k+1}}{\partial y_i}(\bar{\mathbf{y}}_x^k, x) \geq 0 \quad \forall i \in \mathcal{S} \quad (15a)$$

$$\frac{\partial f_{k+1}}{\partial y_i}(\bar{\mathbf{y}}_x^k, x) \leq 0 \quad \forall i \in \mathcal{T} \quad (15b)$$

$$\frac{\partial f_{k+1}}{\partial y_i}(\bar{\mathbf{y}}_x^k, x) = 0 \quad \forall i \in \mathcal{R} \stackrel{\text{def}}{=} [n] \setminus (\mathcal{S} \cup \mathcal{T}). \quad (15c)$$

$$\ell_{[k]} \leq \bar{\mathbf{y}}_x^k \leq \mathbf{u}_{[k]} \quad (15d)$$

Thus, finding the next breakpoint (assuming we are tracing the solution path by increasing  $x$ ) amounts to finding the minimum value of  $x$  such that either a constraint (15a) or an upper-bound constraint in (15d) can no longer be satisfied. Observe that from the first-order definition of submodularity in Proposition 1 and Lemma 4, we find that if (15b) is satisfied, then it will still be satisfied as we keep increasing  $(\mathbf{y}, x)$  while tracing the solution path. Thus, any variable fixed to its upper bound can be ignored (i.e., treated as a constant) when computing the solution path. Similarly, checking whether a given  $\ell_{k+1} < x < u_{k+1}$  is optimal can be obtained by verifying whether

$$\frac{\partial f_{k+1}}{\partial x}(\bar{\mathbf{y}}_x^k, x) = 0 \quad (16)$$

holds.

**4.1. Tracing solutions paths in the quadratic case.** Assume  $f(\mathbf{x}) = -\mathbf{a}^\top \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x}$ , where  $\mathbf{Q}$  is a Stieltjes matrix. Then (13) can be stated as

$$\begin{aligned} \xi_{k+1}(x) = & -a_{k+1}x + \frac{1}{2}Q_{k+1,k+1}x^2 \\ & + \mathbf{minimum}_{\mathbf{y} \in \mathbb{R}^k} \left\{ \frac{1}{2} \mathbf{y}^\top \mathbf{Q}^k \mathbf{y} + \sum_{i=1}^k (Q_{i,k+1}x - a_i) y_i : \ell_i \leq y_i \leq u_i, \forall i \in [k] \right\} \end{aligned} \quad (17)$$

where  $\mathbf{Q}^k \in \mathbb{R}^k \times \mathbb{R}^k$  is the  $k$ -th leading principal submatrix of  $\mathbf{Q}$ . Moreover, for any  $\mathcal{S}, \mathcal{T} \subseteq [n]$ , we denote the submatrix of  $\mathbf{Q}$  indexed by  $\mathcal{S}$  and  $\mathcal{T}$  by  $\mathbf{Q}_{\mathcal{S}, \mathcal{T}}$ , and rewrite  $\mathbf{Q}_{\mathcal{S}, \mathcal{S}}^{-1} = (\mathbf{Q}_{\mathcal{S}, \mathcal{S}})^{-1}$  for short. Similarly, given a vector  $\mathbf{y} \in \mathbb{R}^n$  and set  $\mathcal{S} \subseteq [n]$ , we let  $\mathbf{y}_{\mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$  be the subvector of  $\mathbf{y}$  induced by  $\mathcal{S}$ .

Assuming that variables  $y_i = \ell_i$  for all  $i \in \mathcal{S}$ ,  $y_i = u_i$  for all  $i \in \mathcal{T}$ , and bounds on variables  $y_i$  for  $i \in \mathcal{R}$  are not relevant, we find that the solution

of system (15c) is given by

$$\mathbf{y}_{\mathcal{R}}(x) = \mathbf{Q}_{\mathcal{R},\mathcal{R}}^{-1} (\mathbf{a}_{\mathcal{R}} - \mathbf{Q}_{\mathcal{R},\mathcal{S}} \ell_{\mathcal{S}} - \mathbf{Q}_{\mathcal{R},\mathcal{T}} u_{\mathcal{T}}) - \left( \mathbf{Q}_{\mathcal{R},\mathcal{R}}^{-1} \mathbf{Q}_{\mathcal{R},\{k+1\}} \right) x.$$

In particular, entries of  $\mathbf{y}_{\mathcal{R}}(x)$  are affine functions of  $x$ . Thus, given  $i \in \mathcal{R}$ , solving the linear equalities (with one unknown)  $\mathbf{y}_{\mathcal{R}}(x)_i = u_i$  can also be done analytically. Moreover, given  $i \in \mathcal{S}$ , solving the equality  $-a_i + \frac{\partial f_{k+1}}{\partial y_i}(\mathbf{y}_{\mathcal{R}}(x), x) = 0$  can be done analytically. The minimum of the solutions of these linear equalities corresponds to the next breakpoint. Since the minimum optimality condition  $-a_{k+1} + \frac{\partial f_{k+1}}{\partial x}(\mathbf{y}_{\mathcal{R}}(x), x) = 0$  admits a closed form as well, one can also identify whether an optimal solution is reached before the next breakpoint. We formalize the above discussion as Algorithm 2 in Appendix A.

An efficient implementation to update the quantities in the algorithm is closely tied to pivoting methods for solving linear complementarity problems; see Chapter 4 of [16]. The complexity is  $\mathcal{O}(n^2)$  time per step, see [27, 30, 51] and the references therein for details. Combining the quadratic complexity per step with the linear number of steps (Proposition 3), we obtain the overall complexity of the method.

**Proposition 4.** *If  $f(\mathbf{x}) = -\mathbf{a}^{\top} \mathbf{x} + \frac{1}{2} \mathbf{x}^{\top} \mathbf{Q} \mathbf{x}$  with  $\mathbf{Q}$  being a Stieltjes matrix, Algorithm 2 can terminate in  $\mathcal{O}(n^3)$  time.*

Notably, in this case, the cubic complexity matches the best known complexity of computing  $v(\mathbf{1})$ , thus the Lovász extension (11) can be computed in the same complexity as an evaluation of the submodular function.

**4.2. General lower bounds.** Now consider problem (12), but we now assume  $\ell < \mathbf{0} < \mathbf{u}$  for notational simplicity. Following §3.2, we rewrite (12) as

$$\underset{\mathbf{x} \in \mathbb{R}^n, (\mathbf{z}^+, \mathbf{z}^-) \in \{0,1\}^{2n}}{\text{minimize}} \left\{ f(\mathbf{x}) + \mathbf{c}^{\top} (\mathbf{z}^+ - \mathbf{z}^-) : \ell \circ (\mathbf{1} - \mathbf{z}^-) \leq \mathbf{x} \leq \mathbf{u} \circ \mathbf{z}^+ \right\}, \quad (18)$$

ensuring the submodularity of function  $v(\mathbf{z}^+, \mathbf{z}^-)$ . As before, we want to compute  $\{v(\mathbf{1}_{[k]})\}_{k \in [2n]}$ , but in this case  $\mathbf{1}_{[k]} \in \{0,1\}^{2n}$  is a reordering of  $(\mathbf{z}^+, \mathbf{z}^-)$ . We now briefly discuss how to adapt the parametric algorithm in this case.

First, observe that

$$v(\mathbf{0}) = \underset{\mathbf{x} \in \mathbb{R}^n: \ell \leq \mathbf{x} \leq \mathbf{0}}{\text{minimum}} f(\mathbf{x}); \quad (19)$$

thus, instead of initializing  $\mathbf{x}_B^0$  to  $\mathbf{0}$  in line 1 of Algorithm 1, we set it to the optimal solution of (19), which –for the Stieltjes case– can be computed in  $\mathcal{O}(n^3)$ .

Now suppose  $v(\mathbf{1}_{[k]})$  has been computed, and we seek to compute  $v(\mathbf{1}_{[k+1]})$ . On the one hand, if the  $(k+1)$ -th variable corresponds to  $\mathbf{z}^-$ , then variable  $x_i$  is now forced to be non-negative: a feasible solution can be recovered from the optimal solution associated with  $v(\mathbf{1}_{[k]})$  by tracing the solution path resulting from increasing this variables to 0, analogously to the `trace_path` routine in line 3. On the other hand, if the  $(k+1)$ -th variable corresponds to  $\mathbf{z}^+$ , then variable  $x_i$  is now allowed to be positive: an optimal solution can be recovered from the optimal solution associated with  $v(\mathbf{1}_{[k]})$  by tracing the solution path resulting from increasing this variables from 0, analogously to the `trace_path` routine in line 4. In both cases, for the Stieltjes case, the per iteration complexity is  $\mathcal{O}(n^2)$ .

Finally, the number of breakpoints encountered by the parametric algorithm may be doubled, as breakpoints now correspond to: a variable  $x_i \neq \ell_i$  for the first time; a variable  $x_i = 0$  for the first time; a variable  $x_i > 0$  for the first time; and a variable  $x_i = u_i$  for the first time. Nonetheless, there are still  $\mathcal{O}(n)$  breakpoints, resulting in the same overall complexity of  $\mathcal{O}(n^3)$ .

## 5. CONCLUSION

In this paper, we study a class of convex submodular minimization problems with indicator and lattice constraints, of which the inference of Markov random fields with sparsity and robustness priors is a special case. Such a problem can be solved as a binary submodular minimization problem and thus in (strongly) polynomial time provided that for each fixed binary variable, the resulting convex optimization subproblem is (strongly) polynomially solvable. When applied to quadratic cases, it extends known results in the literature. More efficient implementations are also proposed by exploiting the isotonicity of the solution mapping in parametric settings.

## 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] Angelov, S., Harb, B., Kannan, S., and Wang, L.-S. (2006). Weighted isotonic regression under the L1 norm. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm*, pages 783–791.

- [4] Atamtürk, A. and Gómez, A. (2018). Strong formulations for quadratic optimization with m-matrices and indicator variables. *Mathematical Programming*, 170(1):141–176.
- [5] Atamturk, A. and Gómez, A. (2020). Safe screening rules for L0-regression from perspective relaxations. In *International Conference on Machine Learning*, pages 421–430. PMLR.
- [6] Atamtürk, A., Gómez, A., and Han, S. (2021). Sparse and smooth signal estimation: Convexification of L0-formulations. *Journal of Machine Learning Research*, 22(52):1–43.
- [7] Bernholt, T. (2006). Robust estimators are hard to compute. Technical report, Technical report.
- [8] Bertsimas, D. and King, A. (2015). OR forum – an algorithmic approach to linear regression. *Operations Research*, 64:2–16.
- [9] Bertsimas, D., King, A., Mazumder, R., et al. (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44:813–852.
- [10] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.
- [11] Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.
- [12] Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.
- [13] Boykov, Y. and Funka-Lea, G. (2006). Graph cuts and efficient ND image segmentation. *International Journal of Computer Vision*, 70(2):109–131.
- [14] 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.
- [15] Chandrasekaran, R. (1970). A special case of the complementary pivot problem. *Opsearch*, 7:263–268.
- [16] Cottle, R. W., Pang, J.-S., and Stone, R. E. (2009). *The linear complementarity problem*. SIAM.
- [17] Cozad, A., Sahinidis, N. V., and Miller, D. C. (2014). Learning surrogate models for simulation-based optimization. *AIChE Journal*, 60:2211–2227.
- [18] Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution. *The Annals of Statistics*, 29(1):1–65.
- [19] Edmonds, J. (1970). Submodular functions, matroids, and certain polyhedra, combinatorial structures and their applications, r. guy, h. hanani, n. sauer, and j. schonheim, eds. *New York*, pages 69–87.

- [20] Eilers, P. H. and De Menezes, R. X. (2005). Quantile smoothing of array CGH data. *Bioinformatics*, 21(7):1146–1153.
- [21] Ezzat, A. A., Liu, S., Hochbaum, D. S., and Ding, Y. (2021). A graph-theoretic approach for spatial filtering and its impact on mixed-type spatial pattern recognition in wafer bin maps. *IEEE Transactions on Semiconductor Manufacturing*, 34(2):194–206.
- [22] Fattahi, S. and Gomez, A. (2021). Scalable inference of sparsely-changing gaussian markov random fields. *Advances in Neural Information Processing Systems*, 34:6529–6541.
- [23] Frangioni, A. and Gentile, C. (2006). Perspective cuts for a class of convex 0–1 mixed integer programs. *Mathematical Programming*, 106(2):225–236.
- [24] Fujishige, S. (2005). *Submodular functions and optimization*. Elsevier.
- [25] Geman, S. and Graffigne, C. (1986). Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicians*, volume 1, page 2. Berkeley, CA.
- [26] Gómez, A. (2021). Outlier detection in time series via mixed-integer conic quadratic optimization. *SIAM Journal on Optimization*, 31(3):1897–1925.
- [27] Gómez, A., He, Z., and Pang, J.-S. (2022). Linear-step solvability of some folded concave and singly-parametric sparse optimization problems. *Mathematical Programming*, pages 1–42.
- [28] Günlük, O. and Linderoth, J. (2010). Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Mathematical Programming*, 124:183–205.
- [29] Hazimeh, H., Mazumder, R., and Saab, A. (2021). Sparse regression at scale: Branch-and-bound rooted in first-order optimization. *Mathematical Programming*, pages 1–42.
- [30] He, Z., Han, S., Gómez, A., Cui, Y., and Pang, J.-S. (2021). Comparing solution paths of sparse quadratic minimization with a stieltjes matrix. [http://www.optimization-online.org/DB\\_HTML/2021/09/8608.html](http://www.optimization-online.org/DB_HTML/2021/09/8608.html).
- [31] Hochbaum, D. S. (2001). An efficient algorithm for image segmentation, Markov random fields and related problems. *Journal of the ACM (JACM)*, 48:686–701.
- [32] 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(1):169–198.
- [33] 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.

- [34] Hochbaum, D. S. and Lu, C. (2017). A faster algorithm solving a generalization of isotonic median regression and a class of fused lasso problems. *SIAM Journal on Optimization*, 27(4):2563–2596.
- [35] Hochbaum, D. S., Spaen, Q., and Velednitsky, M. (2019). Detecting aberrant linking behavior in directed networks. In *KDIR*, pages 72–82.
- [36] Iwata, S. (2008). Submodular function minimization. *Mathematical Programming*, 112(1):45–64.
- [37] Iwata, S., Fleischer, L., and Fujishige, S. (2001). A combinatorial strongly polynomial algorithm for minimizing submodular functions. *Journal of the ACM (JACM)*, 48(4):761–777.
- [38] Iwata, S. and Orlin, J. B. (2009). A simple combinatorial algorithm for submodular function minimization. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1230–1237. SIAM.
- [39] Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine*, 17(18):2045–2060.
- [40] Kolmogorov, V. and Zabini, R. (2004). What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:147–159.
- [41] Law, J., Quick, M., and Chan, P. (2014). Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level. *Journal of Quantitative Criminology*, 30(1):57–78.
- [42] Liu, P., Fattahi, S., Gómez, A., and Küçükyavuz, S. (2022). A graph-based decomposition method for convex quadratic optimization with indicators. *Mathematical Programming*, pages 1–33.
- [43] Lovász, L. (1983). Submodular functions and convexity. In *Mathematical Programming the State of the Art*, pages 235–257. Springer.
- [44] Lu, C. and Hochbaum, D. S. (2022). A unified approach for a 1D generalized total variation problem. *Mathematical Programming*, 194(1):415–442.
- [45] Mammen, E., van de Geer, S., et al. (1997). Locally adaptive regression splines. *The Annals of Statistics*, 25:387–413.
- [46] Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., and DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in STAN. *Spatial and Spatio-Temporal Epidemiology*, 31:100301.
- [47] Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y. (2014). On the least trimmed squares estimator. *Algorithmica*, 69(1):148–183.
- [48] Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234.

- [49] Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294.
- [50] Orlin, J. B. (2009). A faster strongly polynomial time algorithm for submodular function minimization. *Mathematical Programming*, 118(2):237–251.
- [51] Pang, J.-S. and Han, S. (2021). Some strongly polynomially solvable convex quadratic programs with bounded variables. <https://arxiv.org/abs/2112.03886>.
- [52] Restrepo, A. and Bovik, A. C. (1993). Locally monotonic regression. *IEEE Transactions on Signal Processing*, 41(9):2796–2810.
- [53] Rinaldo, A. et al. (2009). Properties and refinements of the fused lasso. *The Annals of Statistics*, 37:2922–2952.
- [54] Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388):871–880.
- [55] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust regression and outlier detection*. John Wiley & Sons.
- [56] Rousseeuw, P. J. and Van Driessen, K. (2006). Computing LTS regression for large data sets. *Data Mining and Knowledge Discovery*, 12(1):29–45.
- [57] Sharpnack, J., Singh, A., and Rinaldo, A. (2012). Sparsistency of the edge lasso over graphs. In *Artificial Intelligence and Statistics*, pages 1028–1036. PMLR.
- [58] Tamir, A. (1974). Minimality and complementarity properties associated with Z-functions and M-functions. *Mathematical Programming*, 7(1):17–31.
- [59] Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations Research*, 26(2):305–321.
- [60] Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton University Press.
- [61] Wilson, Z. T. and Sahinidis, N. V. (2017). The ALAMO approach to machine learning. *Computers & Chemical Engineering*, 106:785–795.
- [62] Xie, W. and Deng, X. (2020). Scalable algorithms for the sparse ridge regression. *SIAM Journal on Optimization*, 30(4):3359–3386.
- [63] Zioutas, G. and Avramidis, A. (2005). Deleting outliers in robust regression with mixed integer programming. *Acta Mathematicae Applicatae Sinica*, 21(2):323–334.
- [64] Zioutas, G., Pitsoulis, L., and Avramidis, A. (2009). Quadratic mixed integer programming and support vectors for deleting outliers in robust regression. *Annals of Operations Research*, 166(1):339–353.

## APPENDIX A. SOLUTION TRACING PROCEDURE IN QUADRATIC CASES

In this section, we assume  $\ell \geq 0$  and  $f(\mathbf{x}) = -\mathbf{a}^\top \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x}$ , where  $\mathbf{Q}$  is a Stieltjes matrix. In this case, the analytical form of the solution to (15) is available, based on which, we specialize Algorithm 1 as Algorithm 2. Note that Line 3 and Line 4 of Algorithm 1 are unified into Line 15 of Algorithm 2.

---

**Algorithm 2** Procedure to compute the optimal solutions corresponding to  $v(\mathbf{1}_{[k]})$  for all  $k = 1, \dots, n$  in the quadratic case

---

```

1: for  $k \leftarrow 1, \dots, n$  do
2:    $x \leftarrow 0, \bar{\mathbf{y}}_x^{k-1} \leftarrow (\mathbf{x}^*(k-1))_{[k-1]}, \mathbf{r} \leftarrow \infty \in \mathbb{R}^{k-1} \triangleright r_i$  is a potential pivoting value
3:    $\mathcal{R} \leftarrow \{i : \ell_i < (\bar{\mathbf{y}}_x^{k-1})_i < u_i\}, \mathcal{S} \leftarrow \{i : (\bar{\mathbf{y}}_x^{k-1})_i = \ell_i\}, \mathcal{T} \leftarrow \{i : (\bar{\mathbf{y}}_x^{k-1})_i = u_i\}$ 
       $\triangleright$  Initialize variables and  $[k-1] = \mathcal{R} \cup \mathcal{S} \cup \mathcal{T}$  for  $k$ -th problem
4:   repeat
5:     for  $i \leftarrow 1, \dots, k-1$  do
6:       if  $i \in \mathcal{R}$  and  $(\mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1})_{\{i\}, \mathcal{R}} \mathbf{Q}_{\mathcal{R}, \{k\}} \neq 0$  then
7:          $r_i \leftarrow \frac{u_i - (\mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1})_{\{i\}, \mathcal{R}} (\mathbf{a}_{\mathcal{R}} - \mathbf{Q}_{\mathcal{R}, \mathcal{S}} \ell_{\mathcal{S}} - \mathbf{Q}_{\mathcal{R}, \mathcal{T}} \mathbf{u}_{\mathcal{T}})}{- (\mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1})_{\{i\}, \mathcal{R}} \mathbf{Q}_{\mathcal{R}, \{k\}}}$   $\triangleright$  by  $(\bar{\mathbf{y}}_x^{k-1})_i \leq u_i$ 
8:       else if  $i \in \mathcal{S}$  and  $\mathbf{Q}_{ik} - \mathbf{Q}_{\{i\}, \mathcal{R}} \mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1} \mathbf{Q}_{\mathcal{R}, \{k\}} < 0$  then
9:          $r_i \leftarrow \frac{a_i + \mathbf{Q}_{\{i\}, \mathcal{R}} \mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1} (\mathbf{Q}_{\mathcal{R}, \mathcal{S}} \ell_{\mathcal{S}} + \mathbf{Q}_{\mathcal{R}, \mathcal{T}} \mathbf{u}_{\mathcal{T}} - \mathbf{a}_{\mathcal{R}}) - \mathbf{Q}_{\{i\}, \mathcal{S}} \ell_{\mathcal{S}} - \mathbf{Q}_{\{i\}, \mathcal{T}} \mathbf{u}_{\mathcal{T}}}{\mathbf{Q}_{ik} - \mathbf{Q}_{\{i\}, \mathcal{R}} \mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1} \mathbf{Q}_{\mathcal{R}, \{k\}}}$ 
       $\triangleright$  due to (15a)  $\frac{\partial f_k}{\partial y_i} \geq 0$ 
10:      else
11:         $r_i \leftarrow \infty$ 
12:      end if
13:    end for
14:     $\bar{x} \leftarrow \frac{a_k - (\mathbf{Q}_{\mathcal{R}, \mathcal{S}} \ell_{\mathcal{S}} + \mathbf{Q}_{\mathcal{R}, \mathcal{T}} \mathbf{u}_{\mathcal{T}} - \mathbf{a}_{\mathcal{R}})^\top \mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1} \mathbf{Q}_{\mathcal{R}, \{k\}} - (\ell_{\mathcal{S}})^\top \mathbf{Q}_{\mathcal{S}, k} - (\mathbf{u}_{\mathcal{T}})^\top \mathbf{Q}_{\mathcal{T}, k}}{\mathbf{Q}_{kk} - \mathbf{Q}_{\{k\}, \mathcal{R}} \mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1} \mathbf{Q}_{\mathcal{R}, \{k\}}}$   $\triangleright$  Sol. of (16)
15:     $x \leftarrow \min \left\{ \max\{\ell_k, \bar{x}\}, u_k, \min_{i \in \mathcal{R} \cup \mathcal{S}} r_i \right\}$   $\triangleright$  Ratio test
16:    if  $x = \min_{i \in \mathcal{R}} r_i$  then
17:       $i^* \leftarrow \arg \min_{i \in \mathcal{R}} r_i, \mathcal{R} \leftarrow \mathcal{R} \setminus \{i^*\}, \mathcal{T} \leftarrow \mathcal{T} \cup \{i^*\}$ 
18:    else if  $x = \min_{i \in \mathcal{S}} r_i$  then
19:       $i^* \leftarrow \arg \min_{i \in \mathcal{S}} r_i, \mathcal{S} \leftarrow \mathcal{S} \setminus \{i^*\}, \mathcal{R} \leftarrow \mathcal{R} \cup \{i^*\}$ 
20:    end if
21:    until  $x = \ell_k, \bar{x}$  or  $u_k$ 
22:     $(\bar{\mathbf{y}}_x^{k-1})_{\mathcal{R}} \leftarrow \mathbf{Q}_{\mathcal{R}, \mathcal{R}}^{-1} (\mathbf{a}_{\mathcal{R}} - \mathbf{Q}_{\mathcal{R}, \mathcal{S}} \ell_{\mathcal{S}} - \mathbf{Q}_{\mathcal{R}, \mathcal{T}} \mathbf{u}_{\mathcal{T}} - x \mathbf{Q}_{\mathcal{R}, \{k\}})$   $\triangleright$  Sol. of (15c)
23:     $(\bar{\mathbf{y}}_x^{k-1})_{\mathcal{S}} \leftarrow \ell_{\mathcal{S}}, (\bar{\mathbf{y}}_x^{k-1})_{\mathcal{T}} \leftarrow \mathbf{u}_{\mathcal{T}}, \mathbf{x}^*(k) \leftarrow (\bar{\mathbf{y}}_x^{k-1}, x, \mathbf{0})$   $\triangleright$  Sol. of  $k$ -th subproblem
24:  end for
25: return  $\{f(\mathbf{x}^*(k))\}_{k \in [n]}$ 

```

---