On Integer and MPCC Representability of Affine Sparsity

Hongbo Dong

Department of Mathematics and Statistics
Washington State University
Pullman, WA 99164

Abstract

In addition to sparsity, practitioners of statistics and machine learning often wish to promote additional structures in their variable selection process to incorporate prior knowledge. Borrowing the modeling power of linear systems with binary variables, many of such structures can be faithfully modeled by affine sparsity constraints (ASC). In this note we study conditions under which an ASC system can be represented by sets in integer programs and mathematical programs with complementarity conditions (MPCC). Results of this note facilitate developing nonconvex optimization methods for variable selection with structured sparsity.
Keywords. Structured sparsity, Integer program, Mathematical program with complementarity conditions

1. Structured Sparsity and Affine Sparsity Constraints (ASC)

Sparse optimization has received much attention in recent years due to applications in variable selection in statistics, machine learning and signal processing. In many contexts, practitioners may wish to exploit additional *prior knowledge* and promote *structured sparsity*. Some examples include variable selection with hierarchical interaction terms for regression models [3], microarray data analysis exploiting knowledge of biological pathways or networks [11], as well as numerous applications in machine learning [1].

In the case of unstructured sparsity, the underlying optimization is the following cardinality constrained problem (or its corresponding Lagrange form):

$$\min_{(x,y)\in X} \quad f(x,y) \qquad s.t., \quad \sum_{i=1}^{n} |x_i|_0 \leq K, \tag{1}$$

where $x \in \mathbb{R}^n, y \in \mathbb{R}^m$, $X \subseteq \mathbb{R}^{n+m}$, $K > 0$ and $|x_i|_0 = 0$ if $x_i = 0$ and $|x_i|_0 = 1$ if $x_i \neq 0$. This formulation is sometimes referred to as the "ℓ_0-norm formulation". Various convex and nonconvex penalty-based method for sparse optimization can be seen as heuristic methods by approximating the (nonconvex, discontinuous) binary indicator function $|\bullet|_0$ by continuous (but usually nonsmooth) functions. Well-known examples include the lasso and generalizations [8], SCAD [6], MCP [13], capped-ℓ_1 [14], etc.

Current approaches for learning with structured sparsity primarily focus on convex formulations, e.g., overlapping group penalty [9, 1] and latent variable (group) penalty [10] or constrained convex programs by relaxing (nonconvex) complementarity conditions [3]. While effective in some applications, penalty functions are often built upon qualitative, instead of quantitative descriptions of the desired structure in solution, and promote structures and sparsity – two separate concepts – in an (arguably) mixed way. This can cause difficulties in parameter tuning, which are usually done in an *ad hoc* way. It is also known that convex penalties are only loose approximations to the fundamentally nonconvex ℓ_0 problem, and introduce bias in parameter estimation [6]. Without rigorously modeling and studying the "actual" structure in various notions of structured sparsity, it is difficult to develop sharper variable selection methods based on nonconvex optimization.

The recently introduced affine sparsity constraint (ASC) systems [5] is an approach that seeks to first faithfully model the desired structured sparsity and then to approximate systematically. Given a matrix $A \in \mathbb{R}^{s\times n}$ and a continuous mapping $b(\bullet) : \mathbb{R}^m \mapsto \mathbb{R}^s$, the ASC system takes the following form:

$$A\|x\|_0 \leq b(y), \tag{2}$$

where $\| \bullet \|_0 : \mathbb{R}^n \mapsto \{0,1\}^n$ is the binary indicator mapping such that the i-th component of $\|x\|_0$ is $|x_i|_0$. In [5] continuous approximations to (2) are constructed by approximating $| \bullet |_0$ with a class of point-wisely convergent continuous functions. Conditions for set convergence and convergence of stationary points are also provided in [5], laying a mathematical foundation for addressing structured sparsity via ASC systems.

Two other approaches recently explored for directly solving (1) are mathematical programming with complementarity constraints (MPCC) [12, 4, 7] and integer programming [2]. At first sight it may seem straightforward to incorporate constraints (2) into these two solution strategies, as in both MPCC and integer formulations, variables (continuous or binary) are introduced to "represent" $\|x\|_0$. If $b(\bullet)$ is linear (which is the case in many examples in [5] as well as an application to be discussed in Section 3), the system (2) amounts to a polyhedral system, which are typically easy to handle in algorithms. However this is not the whole truth. For one reason, the ASC system may represent sets that are not closed while feasible regions in both MPCC or integer programs are always closed.

The goal of this note is to address the relations between sets represented by an ASC system and their "straightforward" representations with MPCC and integer programs. Our main results show that while integer formulations based on "big M" always model (a bounded subset of) the closure of ASC-defined sets, a nontrivial condition is required for the MPCC formulation to be "correct". Interestingly, the same condition is used in [5] to prove set convergence of continuous approximations to the ASC system.

2. Integer and MPCC Representations of ASC

Consider the following optimization problem

$$\min_{(x,y)\in X} \quad f(x,y) \qquad s.t.,\ (x,y) \in \mathcal{F}_{ASC} \tag{3}$$

where

$$\mathcal{F}_{ASC} := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \mid A\|x\|_0 \le b(y)\}.$$

A is an $s \times n$ matrix and $b(\bullet) : \mathbb{R}^m \mapsto \mathbb{R}^s$ is a continuous mapping. We use the convention that when $m = 0$, $b(\bullet)$ is a constant vector in $\mathbb{R}^s$. We focus on computable representations of $\mathcal{F}_{ASC}$ while restricting less on $f(\bullet,\bullet)$ except that it is somehow "computationally manageable". Note that $\mathcal{F}_{ASC}$, as well as other sets to be defined shortly, depends on both A and $b(\bullet)$. For convenience we omit this dependence in notation.

In this section, after some discussions on the closedness of $\mathcal{F}_{ASC}$, we show its closure can be represented by simultaneously using binary variables and complementarity conditions. Relaxing either of them leads to more manageable formulations as integer programs or MPCC. The rest of this section then focuses on the conditions under which such relaxations are exact.

It has been observed in [5] that $\mathcal{F}_{ASC}$ is not necessarily closed. The following theorem provides a checkable sufficient and necessary condition for the closedness of $\mathcal{F}_{ASC}$, and is a slight extension of (a part of) [5, Proposition 3], which corresponds to the case of $m = 0$.

Theorem 1. *Given $A \in \mathbb{R}^{s\times n}$ and a continuous mapping $b(\bullet) : \mathbb{R}^m \mapsto \mathbb{R}^s$. Let $\mathcal{Z} := \{(z,y) \in \{0,1\}^n \times \mathbb{R}^m \mid Az \le b(y)\}$. The following two conditions are equivalent:*

(a) $\mathcal{F}_{ASC}$ is a closed set;

(b) For any $(\bar{z},\bar{y}) \in \mathcal{Z}$ and $z \in \{0,1\}^n$, $z \le \bar{z}$, one has $(z,\bar{y}) \in \mathcal{Z}$.

Proof. (a) $\Rightarrow$ (b) follows directly from Proposition 3 in [5] by fixing $y = \bar{y}$. We show that (b) $\Rightarrow$ (a). Let $\{(x^k,y^k)\}_{k=1}^{+\infty}$ be a sequence in $\mathcal{F}_{ASC}$ converging to $(\bar{x},\bar{y})$. Let $\bar{z} \in \{0,1\}^n$ be any cluster point of $\|x^k\|_0$, by the lower semi-continuity of the binary indicator function $| \bullet |_0$, we have

$$|\bar{x}_j|_0 \le \liminf_{k\mapsto+\infty} |x_j^k|_0 \le \bar{z}_j,\ \forall j = 1,...,n.$$

We show that $(\bar{z},\bar{y}) \in \mathcal{Z}$. Let $\{x^{k_j}\}_j$ be a subsequence such that $\lim_j \|x^{k_j}\|_0 = \bar{z}$. Then by continuity of linear transformations and $b(\bullet)$,

$$A\bar{z} = \lim_{j\mapsto+\infty} A\|x^{k_j}\|_0 \le \lim_{j\mapsto+\infty} b\left(y^{k_j}\right) = b(\bar{y}).$$

Therefore $(\bar{z},\bar{y}) \in \mathcal{Z}$, which implies $(\|\bar{x}\|_0,\bar{y}) \in \mathcal{Z}$ by (b). So $(\bar{x},\bar{y}) \in \mathcal{F}_{ASC}$ and $\mathcal{F}_{ASC}$ is closed. □

Towards representations of $\mathcal{F}_{ASC}$ with MPCC or integer programs, consider the following set involving both binary variables and (half) complementarity conditions. Note that ξ is introduced to "represent" the binary vector $1 - \|x\|_0$.

$$\mathcal{F}_{Bin,CC} := \Big\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \;\Big|\; \exists \xi \in \{0,1\}^n,\; A(1-\xi) \le b(y), \quad x_i\xi_i = 0, \;\; \forall i = 1,...,n\Big\}. \quad (4)$$

This set is in fact exactly the closure of $\mathcal{F}_{ASC}$.

Proposition 1. *Let $\mathcal{F}_{ASC}$ and $\mathcal{F}_{Bin,CC}$ are sets defined in (3) and (4), then*

$$\mathcal{F}_{Bin,CC} = \mathbf{cl}\mathcal{F}_{ASC}.$$

Proof. It is straightforward to see that $\mathcal{F}_{ASC} \subseteq \mathcal{F}_{Bin,CC}$ by choosing $\xi = 1 - \|x\|_0$. Now we show $\mathbf{cl}\mathcal{F}_{ASC} \supseteq \mathcal{F}_{Bin,CC}$. Let $(\bar{x}, \bar{y}) \in \mathcal{F}_{Bin,CC}$ and $\xi \in \{0,1\}^n$ be the corresponding vector in (4). We construct a sequence $\{(x^k, y^k)\}_{k=1}^{\infty}$ in $\mathcal{F}_{ASC}$ (except a finite number of terms) converging to $(\bar{x}, \bar{y})$ as follows,

$$x_i^k = \begin{cases} \bar{x}_i + \frac{1}{k} & \text{if } \xi_i = 0, \\ \bar{x}_i & \text{if } \xi_i = 1, \end{cases}, i = 1, ..., n, \qquad y^k = \bar{y}.$$

Apparently for all i such that $\xi_i = 1$, we must have $\bar{x}_i = 0$ and $|x_i^k|_0 = 0 = 1 - \xi_i$. For i such that $\xi_i = 0$, $|x_i^k|_0 = |\bar{x}_i + \frac{1}{k}|_0 = 1 = 1 - \xi_i$ except at most one k. Therefore $\|x^k\|_0 = 1 - \xi$ except at most n terms. So $A\|x^k\|_0 = A(1-\xi) \le b(\bar{y}) = b(y^k)$ for all k sufficiently large. Hence $(\bar{x}, \bar{y}) \in \mathbf{cl}\mathcal{F}_{ASC}$. Therefore $\mathcal{F}_{ASC} \subseteq \mathcal{F}_{Bin,CC} \subseteq \mathbf{cl}\mathcal{F}_{ASC}$. Since $\mathcal{F}_{Bin,CC}$ is closed, the second inclusion is an equality. ☐

Optimization over a non-closed set is problematic because the optimal solution might not be attained even if the objective function is continuous and the feasible region is bounded. Instead of $\mathcal{F}_{ASC}$, it is reasonable to optimize over $\mathbf{cl}\mathcal{F}_{ASC}$ instead. It is also reasonable to expect the "generic" condition that all optimal (i.e., stationary, local or global) solutions would not appear in $\mathbf{cl}\mathcal{F}_{ASC} \setminus \mathcal{F}_{ASC}$, and this replacement would not cause problem in practice. Rigorously stating such claims require details of the objective function $f(\bullet, \bullet)$ and constraints, and is beyond the scope of this note.

The representation $\mathcal{F}_{Bin,CC}$ involves both binary variables and complementarity conditions – two major sources of computational difficulties. Relaxing or linearizing one of them leads to either integer program or MPCC. A well-known approach in discrete optimization is the "big-M" representation where complementarity conditions are replaced by linear constraints within a bounded region. Given $M > 0$, we define $\mathcal{F}_{Bin,M}$ as follows,

$$\mathcal{F}_{Bin,M} := \Big\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \;\Big|\; \exists z \in \{0,1\}^n \quad Az \le b(y), \; -Mz \le x \le Mz\Big\}. \quad (5)$$

The following straightforward result characterizes the relation between $\mathcal{F}_{Bin,CC}$ and $\mathcal{F}_{Bin,M}$.

Proposition 2. *For any $M > 0$,*

$$\mathcal{F}_{Bin,CC} \cap ([-M, M]^n \times \mathbb{R}^m) = \mathcal{F}_{Bin,M}.$$

Proof. First we show that $\mathcal{F}_{Bin,M} \subseteq \mathcal{F}_{Bin,CC} \cap ([-M, M]^n \times \mathbb{R}^m)$. For any $(\bar{x}, \bar{y}) \in \mathcal{F}_{Bin,M}$, we define $\bar{\xi} = 1 - z$ where z is the associated vector in (5). Then $x_i\bar{\xi}_i = x_i(1 - z_i) = 0$ for all i. So $(\bar{x}, \bar{y}) \in \mathcal{F}_{Bin,CC}$. The fact that $\bar{x} \in [-M, M]^n$ is trivial.

Now suppose that $(\bar{x}, \bar{y}) \in \mathcal{F}_{Bin,CC}$ and $\bar{x} \in [-M, M]^n$. Let $\bar{\xi}$ be the associated vector in (4). We define $\bar{z} = 1 - \bar{\xi}$. Then obviously $\bar{z} \in \{0,1\}^n$ and $A\bar{z} \le b(\bar{y})$. For all i such that $\bar{\xi}_i = 1$, $\bar{z}_i = 0$ and $\bar{x}_i = 0$. For i such that $\bar{\xi}_i = 0$, $\bar{z}_i = 1$. So $-M\bar{z} \le \bar{x} \le M\bar{z}$ and $(\bar{x}, \bar{y}) \in \mathcal{F}_{Bin,M}$. ☐

Combining Proposition 1 and 2, $\mathcal{F}_{Bin,M} = \mathbf{cl}\mathcal{F}_{ASC} \cap ([-M, M]^n \times \mathbb{R}^m)$. Note this equation holds without any condition on A and $b(\bullet)$ except the continuity of $b(\bullet)$. Therefore an integer program formulation that approximates (3) is

$$\min_{x,y} \quad f(x,y) \quad s.t., \quad Az \le b(y), \quad -Mz \le x \le Mz, \; z \in \{0,1\}^n.$$

An alternative way to reduce computational difficulties in (4) is to relax the binary condition $\xi \in \{0,1\}^n$ to $\xi \in [0,1]^n$. This leads to the MPCC approach studied in [12, 4, 7]. Consider the following relaxation of $\mathcal{F}_{Bin,CC}$:

$$\mathcal{F}_{CC} := \Big\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \;\Big|\; \exists \xi \in [0,1]^n, \; A(1-\xi) \le b(y), \; x_i\xi_i = 0, \; \forall i\Big\} \quad (6)$$

$$= \Big\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \;\Big|\; \exists \xi \in [0,1]^n, \; x^+, x^- \in \mathbb{R}_+^n, \quad x = x^+ - x^-, \; A(1-\xi) \le b(y), \; (x^+ + x^-)^T\xi = 0\Big\}. \quad (7)$$

The equivalence of (6) and (7) is well-known and can be realized by the following arguments. Take any (x, y) in the set defined in (6), it is readily in the set in (7) by choosing $x^+ = \max(x, 0)$ and $x^- = -\min(x, 0)$. Now let (x, y) be any vector in the set in (7) and $x^+, x^- \in \mathbb{R}^n_+$ be the associated vectors. For all i such that $\xi_i = 0$, we have $x_i\xi_i = 0$. Otherwise if $\xi_i \neq 0$ we have $x_i^+ + x_i^- = 0$, which implies $x_i = x_i^+ = x_i^- = 0$ and so $x_i\xi_i = 0$. Therefore (x, y) is also in the set in (6).

It is immediate to see that $\mathcal{F}_{Bin,CC} \subseteq \mathcal{F}_{CC}$. However, the reverse inclusion is in general false. In fact the following *facial rounding condition* is needed for $\mathcal{F}_{CC}$ to be a tight relaxation to $\mathcal{F}_{Bin,CC}$. Interestingly, this is exactly one fundamental property needed in [5] for proving set convergence for approximating $\mathcal{F}_{ASC}$ with continuous approximations.

Definition 1 (Facial Rounding Condition)**.** For fixed $A \in \mathbb{R}^{s\times n}$ and continuous mapping $b(\bullet) : \mathbb{R}^m \mapsto \mathbb{R}^s$, we say the *facial rounding condition* (FRC) holds at $\bar{y} \in \mathbb{R}^m$ if for any subset $S \subseteq \{1, ..., n\}$, $z \in [0, 1]^n$ such that $z_i = 1$, $\forall i \in S$ and $Az \leq b(\bar{y})$, one can find a "rounded" vector $\hat{z} \in \{0, 1\}^n$ such that $\hat{z}_i = 1$, $\forall i \in S$ and $A\hat{z} \leq b(\bar{y})$. By convention we also say FRC holds at $\bar{y}$ if $\{z \in [0, 1]^n \mid Az \leq b(\bar{y})\}$ is empty.

Remark 1 (Sufficient conditions for FRC). It was mentioned in [5] that if matrix A has the *columnwise uni-sign* property, i.e., for each j, $\{A_{ij}\}_i$ are either all non-negative or all non-positive, then FRC holds at all y. Also if at some $\bar{y}$ such that

$$\begin{aligned} &\{z \in [0, 1]^n \mid Az \leq b(\bar{y})\} \\ =\;& \mathbf{ConvexHull}\{z \in \{0, 1\}^n \mid Az \leq b(\bar{y})\}. \end{aligned}$$

Then FRC holds at $\bar{y}$.

Theorem 2. *Let $\mathcal{F}_{Bin,CC}$ and $\mathcal{F}_{CC}$ be sets as defined in (4) and (6), we have*

$$(\mathbf{cl}\mathcal{F}_{ASC} =)\mathcal{F}_{Bin,CC} \subseteq \mathcal{F}_{CC}.$$

Furthermore, $\mathcal{F}_{Bin,CC} = \mathcal{F}_{CC}$ if and only if the FRC holds at all $y \in \mathbb{R}^m$.

Proof. By definition it is straightforward that $\mathcal{F}_{Bin,CC} \subseteq \mathcal{F}_{CC}$. Take $(\bar{x}, \bar{y}) \in \mathcal{F}_{CC}$ and $\bar{\xi} \in [0, 1]^n$ the associated vector in (6). Suppose that the FRC holds at $\bar{y}$, then there exists $\hat{\xi} \in \{0, 1\}^n$ such that $A(1 - \hat{\xi}) \leq b(\bar{y})$ and $\hat{\xi}_i = 0$ whenever $\bar{\xi}_i = 0$. Therefore $\bar{x}_i\hat{\xi}_i = 0$ for all $i = 1, ..., n$ and $(\bar{x}, \bar{y}) \in \mathcal{F}_{Bin,CC}$.

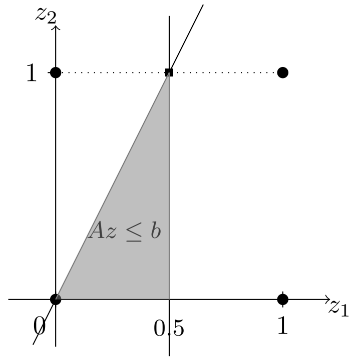

Figure 1: FRC fails at $[0.5, 1]^T$ (cannot round to left or right while remain in the gray region). $\mathcal{F}_{ASC} = \{(0, 0)\}$ while $\mathcal{F}_{CC} = \{(0, t) \mid \forall t \in \mathbb{R}\}$.

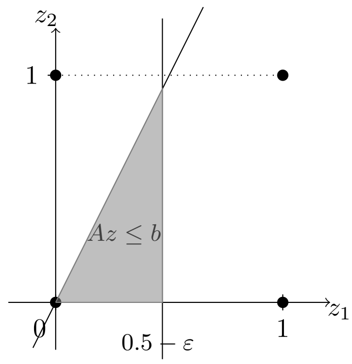

Figure 2: FRC holds everywhere. $\mathcal{F}_{ASC} = \mathcal{F}_{CC} = \{(0, 0)\}$.

Now suppose that FRC fails at some $\bar{y} \in \mathbb{R}^m$ with subset $S \subseteq \{1, ..., n\}$, we construct a point in $\mathcal{F}_{CC} \setminus \mathcal{F}_{Bin,CC}$ as follows. Suppose $\bar{z} \in [0, 1]^n$ is a vector such that $\bar{z}_i = 1$ for all $i \in S$ and $A\bar{z} \leq b(\bar{y})$. Define $\bar{\xi} = 1 - \bar{z}$. Then $\bar{\xi} \in [0, 1]^n$ and $A(1 - \bar{\xi}) \leq b(\bar{y})$. Choose $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_i = 0$ for all $\bar{\xi}_i \neq 0$ and $\bar{x}_i \neq 0$ for all $\bar{\xi}_i = 0$, then $(\bar{x}, \bar{y}) \in \mathcal{F}_{CC}$. However, $(\bar{x}, \bar{y})$ cannot be in $\mathcal{F}_{Bin,CC}$ because that would require $\hat{\xi} \in \{0, 1\}^n$ such that $A(1 - \hat{\xi}) \leq b(\bar{y})$ and $\bar{x}_i\hat{\xi}_i = 0$ for all i. The existence of such $\hat{\xi}$ would imply that $\hat{\xi}_i = 0$ whenever $\bar{\xi}_i = 0$, i.e., $1 - \hat{\xi}_i = 1$ whenever $\bar{z}_i = 1$ (e.g., for all $i \in S$). This contradicts with our assumption that FRC fails at $\bar{y} \in \mathbb{R}^m$ with subset $S \subseteq \{1, ..., n\}$. ☐

We provide a numerical example illustrating that when FRC does not hold, $\mathcal{F}_{CC}$ can be significantly larger than $\mathbf{cl}\mathcal{F}_{ASC}$.

Example 1. *The authors of [5] provided a two-variable example where FRC fails, when $n = 2$ and $m = 0$ (i.e., $b(\bullet)$ is a constant vector b). With* *a slight modification, consider* $A = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$ *and* $b = [0.5, 0]^T$*, then* $\mathcal{F}_{ASC} = \{0\}$*. It is easy to see that* $\bar{z} = [0.5, 1]^T$ *satisfies* $A\bar{z} \le b$*, however, neither* $[0, 1]^T$ *nor* $[1, 1]^T$ *satisfies the same inequality. Take* $x = [0, t]^T$ *(with arbitrary* $t \neq 0$*) and* $\xi = [0.5, 0]^T$*, it can be verified that* $x \in \mathcal{F}_{CC}$*. So* $\mathcal{F}_{CC}$ *is much larger than* $\mathbf{cl}\mathcal{F}_{ASC}$*. See Figure 1 for illustration.*

Now if we perturb b slightly to be $\left[0.5 - \varepsilon, 0\right]^T$ *with some* $0 < \varepsilon < 0.5$*. We verify that* $\mathcal{F}_{CC} = \mathcal{F}_{ASC} = \{0\}$*. To see this, the solution of*

$$\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 - \xi_1 \\ 1 - \xi_2 \end{bmatrix} \le \begin{bmatrix} 0.5 - \varepsilon \\ 0 \end{bmatrix} \Longrightarrow \begin{cases} \xi_1 \in [0.5 + \varepsilon, 1], \\ \xi_2 \in [2\xi_1 - 1, 1]. \end{cases}$$

Obviously $\xi_i > 0$ *for* $i = 1, 2$*. So the only possible* (x^+, x^-) *in* $\mathbb{R}^2_+$ *satisfying the complementarity condition* $(x^+ + x^-)^T \xi = 0$ *is the zero vector. Therefore* $\mathcal{F}_{CC} = \{0\}$ *in this case. See Figure 2 for illustration.*

3. Application to One-dimensional Contiguous Patterns

We consider a specific case of structured sparsity with the prior assumption that n predictor variables are organized in a sequence, and selected variables should form a continuous block of the sequence. This structure is studied in [9], where convex approximations are proposed. Contiguous structures on more complex networks are also of interests in applications such as predicting clinical outcomes with gene expression profiles and a gene network (e.g., [11]). For simplicity we focus on the one-dimensional sequence structure in this section.

Let $\{x_1, ..., x_n\}$ be the parameters under estimation for ordered predictor variables. Suppose we would like to enforce the structure that the nonzero parameters form a contiguous block of length at most K (a positive integer no larger than n). This structure can be characterized by

$$|x_1|_0 + |x_n|_0 + \sum_{i=1}^{n-1} \Big| |x_i|_0 - |x_{i+1}|_0 \Big| \le 2, \quad \sum_{i=1}^{n} |x_i|_0 \le K.$$

The first inequality implies that there are at most two "jumps" in the ordered sequence $\{0, |x_1|_0, |x_2|_0, ..., |x_n|_0, 0\}$ (with dummy zeros added at the front and the end). Hence enforcing the contiguity of nonzero parameters.

An ASC system can be defined with additional continuous variables y,

$$\begin{aligned} \mathcal{F}_{ASC} = \Big\{ (x, y) \in \mathbb{R}^{2n-1} \ \Big| \ & \sum_{i=1}^{n} |x_i|_0 \le K \\ & |x_{i+1}|_0 - |x_i|_0 \le y_i, \\ & |x_i|_0 - |x_{i+1}|_0 \le y_i, \\ & |x_1|_0 + |x_n|_0 \le 2 - \sum_{i=1}^{n-1} y_i \Big\}. \end{aligned} \tag{8}$$

It is easy to see by Theorem 1 that $\mathcal{F}_{ASC}$ does not define a closed set. For example when $K = 3$ and $n = 4$, then a sequence $\{x^k\}_{k=1}^{+\infty}$ with $(|x_1^k|_0, ..., |x_4^k|_0) = (1, 1, 1, 0)$ may converge to a point with infeasible nonzero pattern $(1, 0, 1, 0)$. By Proposition 1 and 2, the following set equals $\mathbf{cl}\mathcal{F}_{ASC}$ with bounds on x,

$$\begin{aligned} \mathcal{F}_{Bin,M} = \Big\{ (x, y) \in \mathbb{R}^{2n-1} \ \Big| \ & \exists z \in \{0, 1\}^n, \\ & |z_{i+1} - z_i| \le y_i, \forall 1 \le i \le n-1, \ \sum_{i=1}^{n} z_i \le K \\ & -Mz \le x \le Mz, \ z_1 + z_n + \sum_{i=1}^{n-1} y_i \le 2 \Big\}. \end{aligned}$$

However we show that this system may fail the FRC. So the corresponding MPCC formulation, i.e., set $\mathcal{F}_{CC}$ as in (7) does not equal to $\mathbf{cl}\mathcal{F}_{ASC}$ in general. To see this, we define set $\mathcal{M}$ as follows,

$$\begin{aligned} \mathcal{M} := \Big\{ (z, y) \in [0, 1]^n \times \mathbb{R}^{n-1} \ \Big| \ & \sum_{i=1}^{n} z_i \le K, \\ & |z_{i+1} - z_i| \le y_i, \forall 1 \le i \le n-1, \\ & z_1 + z_n + \sum_{i=1}^{n-1} y_i \le 2 \ \Big\}. \end{aligned}$$

Take $n = 3$, $K = 2$, $\bar{z} = [0.5, 1, 0.5]^T$ and $\bar{y} = [0.5, 0.5]^T$. Apparently $(\bar{z}, \bar{y}) \in \mathcal{M}$. Take $S = \{2\}$. One cannot find $\hat{z} \in \{0, 1\}^3$ such that $\hat{z}_2 = 1$ and $(\hat{z}, \bar{y}) \in \mathcal{M}$. The problem is that the two inequalities $|\hat{z}_1 - \hat{z}_2| \le 0.5$, $|\hat{z}_2 - \hat{z}_3| \le 0.5$, together with the binary conditions, imply $\hat{z} = [1, 1, 1]^T$, which violates the cardinality condition.

We now consider an alternative formulation modeling the same structure. We define the following two binary sets

$$\mathcal{P} = \Big\{ z \in \{0, 1\}^n \ \Big| \ \text{all 1's in } z \text{ form one single contiguous block of length at most } K \Big\},$$

$$\mathcal{Q} = \Big\{ z \in \{0,1\}^n \;\Big|\; \text{all 1's in } z \text{ appear in one single contiguous block of length at most } K \Big\}.$$

For example, when $n = 4$ and $K = 3$, $[0, 1, 1, 1]^T$ is in both $\mathcal{P}$ and $\mathcal{Q}$, while $[0, 1, 0, 1]^T$ is in $\mathcal{Q}$ but not in $\mathcal{P}$. Let us define $\mathcal{F}_\mathcal{P} = \{x \in \mathbb{R}^n \mid \|x\|_0 \in \mathcal{P}\}$ and $\mathcal{F}_\mathcal{Q} = \{x \in \mathbb{R}^n \mid \|x\|_0 \in \mathcal{Q}\}$. Note that both $\mathcal{P}$ and $\mathcal{Q}$ can be characterized by finitely many linear inequalities with binary variables, so both $\mathcal{F}_\mathcal{P}$ and $\mathcal{F}_\mathcal{Q}$ can be represented as ASC systems with fixed right hand side. By [5, Proposition 4] it is clear that $\mathcal{F}_\mathcal{Q} = \mathbf{cl}\mathcal{F}_\mathcal{P}$. In fact, $\mathcal{F}_\mathcal{Q}$ can be represented as

$$\mathcal{F}_\mathcal{Q} = \Big\{ x \in \mathbb{R}^n \;\Big|\; |x_i|_0 + |x_j|_0 \le 1, \quad \forall 1 \le i < j \le n,\ j \ge i + K \Big\}.$$

This is an ASC system with nonnegative coefficient matrix, so FRC holds (see Remark 1). By Theorem 2,

$$\mathcal{F}_\mathcal{Q} = \Big\{ x \in \mathbb{R}^n \;\Big|\; \exists \xi \in [0,1]^n, x^+, x^- \in \mathbb{R}^n_+ \\ \xi_i + \xi_j \ge 1, \forall 1 \le i < j \le n,\ j \ge i + K \\ x = x^+ - x^-,\ (x^+ + x^-)^T \xi = 0 \Big\}.$$

This formulation can be used to derive an MPCC formulation for variable selection with the structured sparsity of one-dimensional contiguous patterns.

Acknowledgement. This paper is motivated by a question from Uday V. Shanbhag on the 2018 INFORMS optimization society conference at Denver, Colorado.

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Structured sparsity through convex optimization. *Statist. Sci.*, 27:450–468, 2012.

[2] Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44(2):813–852, 2016.

[3] Jacob Bien, Jonathan Taylor, and Robert Tibshirani. A lasso for hierarchical interactions. *Ann Statistics*, 41(3):1111–1141, 2013.

[4] Oleg P. Burdakov, Christian Kanzow, and Alexandra Schwartz. Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. *SIAM J. Optim.*, 26(1):397–425, 2016.

[5] Hongbo Dong, Miju Ahn, and Jong-Shi Pang. Structural Properties of Affine Sparsity Constraints. Submitted. Available at `http://www.optimization-online.org/DB_HTML/2017/04/5965.html`, April 2017.

[6] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc*, 96(456):1348–1360, 2001.

[7] M. Feng, J. E. Mitchell, J.S. Pang, X. Shen, and A. Wächter. Complementarity formulations of l0-norm optimization problems. Technical report, Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA., `http://www.optimization-online.org/DB_FILE/2013/09/4053.pdf`, 2013.

[8] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC, 2015.

[9] R. Jenatton, J.-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. *Journal of Machine Learning Research*, 12:2777–2824, 2011.

[10] Guillaume Obozinski, Laurent Jacob, and Jean-Philippe Vert. Group lasso with overlaps: the latent group lasso approach. [Research Report] 2011, pp.60. inria-00628498.

[11] Wei Pan, Benhuai Xie, and Xiaotong Shen. Incorporating Predictor Network in Penalized Regression with Application to Microarray Data. *Biometrics*, 66(2):474–484, 2010.

[12] Y. Xie and U. V. Shanbhag. On the resolution of ℓ_0 minimization problems via alternating Lagrangian schemes. `https://arxiv.org/pdf/1710.04613.pdf`, Dec. 2017.

[13] Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, 38(2):894–942, 2010.

[14] Tong Zhang. Analysis of Multi-stage Convex Relaxation for Sparse Regularization. *Journal of Machine Learning Research*, 11:1081–1107, March 2010.