

# Sparse Recovery over Graph Incidence Matrices

Mengnan Zhao, M. Devrim Kaba, René Vidal, Daniel P. Robinson, and Enrique Mallada

**Abstract**—Classical results in sparse recovery guarantee the exact reconstruction of  $s$ -sparse signals under assumptions on the dictionary that are either too strong or NP-hard to check. Moreover, such results may be pessimistic in practice since they are based on a worst-case analysis. In this paper, we consider the sparse recovery of signals defined over a graph, for which the dictionary takes the form of an incidence matrix. We derive necessary and sufficient conditions for sparse recovery, which depend on properties of the cycles of the graph that can be checked in polynomial time. We also derive support-dependent conditions for sparse recovery that depend only on the intersection of the cycles of the graph with the support of the signal. Finally, we exploit sparsity properties on the measurements and the structure of incidence matrices to propose a specialized sub-graph-based recovery algorithm that outperforms the standard  $\ell_1$ -minimization approach.

## I. INTRODUCTION

Recently, sparse recovery methods (e.g. see [1], [2]) have become very popular for compressing and processing high-dimensional data. In particular, they have found widespread applications in data acquisition [3], machine learning [4], [5], medical imaging [6]–[9], and networking [10]–[12].

The goal of sparse recovery is to reconstruct a signal  $\bar{x} \in \mathbb{R}^n$  from  $m < n$  linear measurements  $y = \Phi \bar{x} \in \mathbb{R}^m$ , where  $\Phi \in \mathbb{R}^{m \times n}$  is the measurement matrix (also known as dictionary or sparsifying basis). In general, the recovery problem is ill-posed unless  $\bar{x}$  is assumed to satisfy additional assumptions. For example, if we assume that  $\bar{x}$  is  $s$ -sparse<sup>1</sup>, where  $s \ll n$ , then mild conditions on the measurement matrix  $\Phi$  (see Lemma 1) allow for the recovery  $\bar{x}$  from  $y$  using the  $\ell_0$ -minimization problem

$$\min_x \|x\|_0 \text{ s.t. } y = \Phi x, \quad (1)$$

where  $\|x\|_0$  denotes the number of nonzero entries in  $x$ . However, problem (1) is known to be NP-hard [13]. To address this challenge, a common strategy is to solve a convex relaxation of (1) based on  $\ell_1$ -minimization given by

$$\min_x \|x\|_1 \text{ s.t. } y = \Phi x, \quad (2)$$

which can be written equivalently as a linear program. It is known that the sparse signal can be recovered from (2) if the measurement matrix  $\Phi$  satisfies certain conditions. In general, these conditions are either computationally hard to

verify, or too conservative so that false negative certificates are often encountered. For example, the *Nullspace Property* (NUP) [14], which provides necessary and sufficient conditions for sparse recovery, and the *Restricted Isometry Property* (RIP) [15], [16], which is only a sufficient condition for sparse recovery, are both NP-hard [17] to check. On the other hand, the *Mutual Coherence* [18] property is a sufficient condition that can be efficiently verified [17], but is conservative in that it fails to certify that sparse recovery is possible for many measurement matrices seen in practice.

Another major limitation of these recovery guarantees and their associated computational complexity arises because they must allow for the worst case problem in their derivation. For example, the fact that the NUP and RIP are NP-hard to check does not prohibit efficient verification for particular subclasses of matrices. Similarly, even when the NUP is not satisfied for a sparsity level  $s$ , it may still be possible to recover certain subsets of  $s$ -sparse signals by exploiting additional knowledge about their support. These observations suggest studying the sparse recovery problem for subclasses of matrices and signals to obtain conditions that are easier to verify as well as specialized algorithms with improved recovery performance.

The goal of this paper is to study sparse recovery of signals that are defined over graphs when the measurement matrix is the graph's incidence matrix. Our interest in incidence matrices stems from the fact that they are a fundamental representation of graphs, and thus a natural choice for the dictionary when analyzing network flows. In various application areas like communication networks, social networks, and transportation networks, the incidence matrices naturally appear when modeling the flow of information, disease, and goods (e.g., the detection of sparse structural network changes via observations at the nodes can be modeled as (1), where the incidence matrix serves as a measurement matrix).

The main contributions of this paper are as follows:

- 1) We derive a topological characterization of the NUP for incidence matrices. Specifically, we show that the NUP for these matrices is equivalent to a condition on simple cycles of the graph, which is a finite subset of the nullspace of the incidence matrix. As a consequence, we show that for incidence matrices the sparse recovery guarantee depends only on the girth<sup>2</sup> of the underlying graph. This overcomes NP-hardness, as the girth of a graph can be calculated in polynomial time.
- 2) Using the above topological characterization, we further derive necessary and sufficient conditions on the

M. Zhao and E. Mallada are with the Department of Electrical and Computer Engineering, M.D. Kaba and D.P. Robinson are with the Department of Applied Mathematics and Statistics, and R. Vidal is with the Department of Biomedical Engineering, at The Johns Hopkins University, Baltimore, MD 21218, USA. Emails: {mzhao21, mkaba1, rvidal, daniel.p.robinson, mallada}@jhu.edu

This work was supported by NSF grants 1618637 and AMPS 1736448.

<sup>1</sup>The vector  $\bar{x}$  is  $s$ -sparse if and only if at most  $s$  of its entries are nonzero.

<sup>2</sup>Girth: Size of the smallest cycle.

support of the sparse signal that enable its recovery. Specifically, for incidence matrices we show that all signals with a given support can be recovered via (2) if and only if the support consists of strictly less than half of the edges in every simple cycle of the graph. Since our conditions on  $\bar{x}$  depend on its support, we will refer to them as *support-dependent* conditions.

- 3) We propose a specialized algorithm that utilizes the knowledge of the support of the measurements and the structure of incidence matrices to constrain the support of the signal  $\bar{x}$ , and consequently can guarantee sparse recovery of  $\bar{x}$  under even weaker conditions.

The remainder of this paper is organized as follows: In Section II we review basic concepts from convex analysis, graph theory, and sparse recovery. In Section III we formulate sparse recovery problem for incidence matrices, derive conditions for sparse and support-dependent recovery, and propose an efficient algorithm to solve the problem. In Section IV we present numerical experiments that illustrate our theoretical results, and in Section V we provide some conclusions.

## II. PRELIMINARIES

### A. Notation

Given  $x \in \mathbb{R}^n$ , we let  $\|x\|_p$  for  $p \geq 1$  denote the  $\ell_p$ -norm, and denote the function that counts the number of nonzero entries in  $x$  by  $\|x\|_0$ . Although, the latter function is not a norm, we follow the common practice of calling it the  $\ell_0$ -norm. For  $p \geq 1$ , we define the unit  $\ell_p$ -sphere in  $\mathbb{R}^n$  as  $\mathbb{S}_p^{n-1} := \{x \in \mathbb{R}^n : \|x\|_p = 1\}$ . Similarly, the unit  $\ell_p$ -ball in  $\mathbb{R}^n$  is defined as  $\mathbb{B}_p^n := \{x \in \mathbb{R}^n : \|x\|_p \leq 1\}$ .

The *nullspace* of a matrix  $\Phi$  will be denoted by  $\text{Null}(\Phi)$ . As usual,  $|S|$  denotes the cardinality of a set  $S$ . For a vector  $x \in \mathbb{R}^n$  and a nonempty index set  $S \subseteq \{1, \dots, n\}$ , we denote the subvector of  $x$  that corresponds to  $S$  by  $x_S$ , and the associated mapping by  $\Gamma_S : \mathbb{R}^n \rightarrow \mathbb{R}^{|S|}$ , i.e.,  $x_S := \Gamma_S(x)$ . The complement of  $S$  in  $\{1, \dots, n\}$  is denoted by  $S^c$ .

### B. Convex analysis

Critical to our analysis will be the notion of extreme points of a convex set and of quasi-convex functions.

*Definition 1:* An *extreme point* of a convex set  $C \subset \mathbb{R}^n$  is a point in  $C$  that cannot be written as a convex combination of two different points from  $C$ . The set of all extreme points of a convex set  $C \subset \mathbb{R}^n$  is denoted by  $\text{Ext}(C)$ .

*Definition 2:* Let  $C \subseteq \mathbb{R}^n$  be a convex set. A function  $f : C \rightarrow \mathbb{R}$  is called *quasi-convex* if and only if for all  $x, y \in C$  and  $\lambda \in [0, 1]$  it holds that

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}. \quad (3)$$

It is easy to show that a function  $f : C \rightarrow \mathbb{R}$  is quasi-convex if and only if every sublevel set  $S_\alpha := \{x \in C \mid f(x) \leq \alpha\}$  is a convex set. Every convex function is quasi-convex. In particular,  $\ell_p$ -norm functions for  $p \geq 1$  are quasi-convex. The next result on quasi-convex functions [19] is included for completeness.

*Proposition 1:* Let  $C \subset \mathbb{R}^n$  be a compact convex set and  $f : C \rightarrow \mathbb{R}$  be a continuous quasi-convex function. Then  $f$  attains its maximum value at an extreme point of  $C$ .

### C. Graph theory

A directed graph with vertex set  $V = \{v_1, \dots, v_m\}$  and edge set  $E = \{e_1, \dots, e_n\} \subseteq V \times V$  will be denoted by  $\mathcal{G}(V, E)$ . When the edge and vertex sets are irrelevant to the discussion, we will drop them from the notation and denote the graph by  $\mathcal{G}$ . Sometimes the vertex set and the edge set of  $\mathcal{G}$  will be denoted by  $V(\mathcal{G})$  and  $E(\mathcal{G})$  respectively. A graph  $\mathcal{G}$  is called *simple* if there is at most one edge connecting any two vertices and no edge starts and ends at the same vertex, i.e.  $\mathcal{G}$  has no self loops. Henceforth,  $\mathcal{G}$  will always denote a simple directed graph with a finite number of edges and vertices. Although we focus on directed graphs, our analysis only requires the undirected variants of the definitions for paths, cycles, and connectivity.

We say that two vertices  $\{a, b\} \subseteq V$  are *adjacent* if either  $(a, b) \in E$  or  $(b, a) \in E$ . A sequence  $(u_1, \dots, u_r)$  of distinct vertices of a graph  $\mathcal{G}$  such that  $u_i$  and  $u_{i+1}$  are adjacent for all  $i \in \{1, \dots, r-1\}$  is called a *path*. A *connected component*  $\hat{\mathcal{G}}$  of  $\mathcal{G}$  is a subgraph of  $\mathcal{G}$  in which any two distinct vertices are connected to each other by a path and no edge exists between  $V(\hat{\mathcal{G}})$  and  $V(\mathcal{G}) \setminus V(\hat{\mathcal{G}})$ . We will say that the graph  $\mathcal{G}$  is *connected* (for directed graphs this is often referred to as weakly connected) if  $\mathcal{G}$  has a unique connected component. A sequence  $(u_1, \dots, u_r, u_1)$  of adjacent vertices of a graph  $\mathcal{G}$  is called a (*simple*) *cycle* if  $r \geq 3$  and  $u_i \neq u_j$  whenever  $i \neq j$ ; the length of such a cycle is  $r$ . The length of the shortest simple cycle of a graph  $\mathcal{G}$  is called the *girth* of  $\mathcal{G}$ . For an acyclic graph (i.e., a graph with no cycles), the girth is defined to be  $+\infty$ . Since acyclic graphs are not interesting for our purposes, we will assume that the girth is finite. That is, the graph has at least one simple cycle.

Associated with a directed graph  $\mathcal{G} = \mathcal{G}(V, E)$ , we can define the incidence matrix  $A = A(\mathcal{G}) \in \mathbb{R}^{m \times n}$  as

$$a_{ij} = \begin{cases} -1, & \text{if } v_i \text{ is the initial vertex of edge } e_j, \\ 1, & \text{if } v_i \text{ is the terminal vertex of edge } e_j, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

For a nonempty index set  $S \subseteq \{1, \dots, n\}$ , the subgraph of  $\mathcal{G}$  consisting of edges  $\{e_j \mid j \in S\}$  is denoted by  $\mathcal{G}_S$ . The incidence matrix of  $\mathcal{G}_S$  is denoted by  $A_S$ . Let  $\mathcal{C} = (u_1, \dots, u_r, u_{r+1} = u_1)$  be a simple cycle of a simple directed graph  $\mathcal{G}$ . Then,  $\mathcal{C}$  can be associated with a vector  $w(\mathcal{C}) \in \mathbb{R}^n$ , where each coordinate  $w_j$  of  $w(\mathcal{C})$  is defined as

$$w_j = \begin{cases} 1, & \text{if } e_j = (u_i, u_{i+1}) \text{ for some } i \in \{1, \dots, r\} \\ -1, & \text{if } e_j = (u_{i+1}, u_i) \text{ for some } i \in \{1, \dots, r\} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

We now define the *cycle space* of  $\mathcal{G}$  as the subspace spanned by  $\{w(\mathcal{C}) \mid \mathcal{C} \text{ is a simple cycle of } \mathcal{G}\}$ .<sup>3</sup>

*Remark 1:* The cycle space of  $\mathcal{G}$  is exactly the nullspace of the incidence matrix  $A(\mathcal{G})$  [20].

*Remark 2:* The dimension of the nullspace of  $A(\mathcal{G})$  is  $n - m + k$ , where  $k$  is the number of connected components of  $\mathcal{G}$ . Hence, the rank of  $A(\mathcal{G})$  is  $m - k$  [20].

<sup>3</sup>Sometimes this subspace is called the *Flow Space* [20].

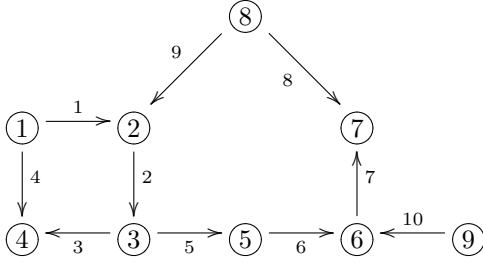


Fig. 1: A connected directed simple graph.

*Example 1:* The incidence matrix of the graph in Fig. 1 is given by

$$A = \begin{bmatrix} -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix},$$

and the vectors from (5) associated with the simple cycles

$$\begin{aligned} \mathcal{C}_1 &= (1, 2, 3, 4, 1), \\ \mathcal{C}_2 &= (2, 3, 5, 6, 7, 8, 2), \\ \mathcal{C}_3 &= (1, 2, 8, 7, 6, 5, 3, 4, 1) \end{aligned}$$

are given by

$$\begin{aligned} w_1 &= [1 \ 1 \ 1 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T \\ w_2 &= [0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \ -1 \ 1 \ 0]^T \\ w_3 &= [1 \ 0 \ 1 \ -1 \ -1 \ -1 \ -1 \ 1 \ -1 \ 0]^T. \end{aligned}$$

Note that  $w_3 = w_1 - w_2$ , so that  $\text{Null}(A(\mathcal{G}))$  must be of dimension 2. The addition/subtraction operations on  $w_j$ 's correspond to "addition/subtraction" operations on edges.

#### D. Sparse Recovery

It is known [21] that every vector  $\bar{x} \in \mathbb{R}^n$  with  $\text{supp}(\bar{x}) \subseteq S$  is the unique solution to (2) if and only if

$$\|\eta_S\|_1 < \|\eta_{S^c}\|_1 \text{ for all } \eta \in \text{Null}(\Phi) \setminus \{0\}. \quad (6)$$

This leads us to the definition of the *Nullspace Property*.

*Definition 3 (Nullspace Property, NUP):* A matrix  $\Phi \in \mathbb{R}^{m \times n}$  is said to satisfy the *nullspace property* (NUP) of order  $s$ , if for any  $\eta \in \text{Null}(\Phi) \setminus \{0\}$ , and any nonempty index set  $S \subseteq \{1, \dots, n\}$  with  $|S| \leq s$ , we have

$$\|\eta_S\|_1 < \|\eta_{S^c}\|_1.$$

Another needed concept is the *spark* of a matrix.

*Definition 4 (Spark):* The spark of a matrix  $\Phi \in \mathbb{R}^{m \times n}$  is the smallest number of linearly dependent columns in  $\Phi$ . Formally, we have  $\text{spark}(\Phi) := \min_{\eta \neq 0} \|\eta\|_0$  s.t.  $\Phi\eta = 0$ .

We note that the rank of a matrix may be used to bound its spark. Specifically, it holds that

$$\text{spark}(\Phi) \leq \text{rank}(\Phi) + 1. \quad (7)$$

The spark may be used to provide a necessary and sufficient condition for uniqueness of sparse solutions [21].

*Lemma 1:* For any matrix  $\Phi \in \mathbb{R}^{m \times n}$ , the  $s$ -sparse signal  $\bar{x} \in \mathbb{R}^n$  is the unique solution to the optimization problem

$$\min_{\Phi\bar{x}=\Phi x} \|x\|_0 \quad (8)$$

if and only if  $\text{spark}(\Phi) > 2s$ .

### III. MAIN RESULTS

It is not difficult to prove that the spark of an incidence matrix  $A$  is equal to the girth of the underlying graph  $\mathcal{G}$ . By combining this fact with Lemma 1, we obtain the following.

*Proposition 2:* Let  $A = A(\mathcal{G}) \in \mathbb{R}^{m \times n}$  be the incidence matrix of a simple, connected graph  $\mathcal{G}$  with girth  $g$ . Then, for every  $s$ -sparse vector  $\bar{x} \in \mathbb{R}^n$ ,  $\bar{x}$  is the unique solution to

$$\min_{A\bar{x}=Ax} \|x\|_0 \quad (9)$$

if and only if  $s < \frac{g}{2}$ .

Even though Proposition 2 is useful as a uniqueness result, it does not come in handy when one would like to recover the original signal  $\bar{x}$  from the measurements  $A\bar{x}$ . For this purpose, an  $\ell_1$ -relaxation of the optimization problem is preferable. Hence, one needs a theorem akin to Proposition 2 that addresses the solutions of the optimization problem

$$\min_{A\bar{x}=Ax} \|x\|_1. \quad (10)$$

This leads us to study the NUP for incidence matrices. Specifically, in this section we answer the following questions about the incidence matrix  $A \in \mathbb{R}^{m \times n}$  of a simple connected graph  $\mathcal{G}$  with  $n$  edges and  $m$  vertices, and an  $s$ -sparse vector  $\bar{x} \in \mathbb{R}^n$ :

- 1) What are necessary and sufficient conditions for  $A$  to satisfy the NUP of order  $s$ ? Such conditions would guarantee sparse recovery, i.e., that any  $s$ -sparse signal  $\bar{x}$  can be recovered as the unique solution to (10).
- 2) If traditional sparse recovery is not possible, can we characterize subclasses of sparse signals that are recoverable via (10) in terms of the support of the signal and the topology of the graph  $\mathcal{G}(V, E)$ ?
- 3) Can we use the support of the measurement  $A\bar{x}$  and the structure of  $A$  to derive constraints on the support of  $\bar{x}$  that allow us to modify (10) and successfully recover the sparse signal  $\bar{x}$ ?

#### A. Topological Characterization of the Nullspace Property for the Class of Incidence Matrices

Before addressing the questions above in detail, we would like to build a simple framework to study them. We will start with a reformulation of the NUP.

*Lemma 2:* A matrix  $\Phi \in \mathbb{R}^{m \times n}$  satisfies the NUP of order  $s$  if and only if

$$\max_{|S| \leq s} \max_{\eta \in \text{Null}(\Phi) \cap \mathbb{B}_1^S} \|\eta_S\|_1 < \frac{1}{2}. \quad (11)$$

*Proof:* For  $\eta \in \text{Null}(\Phi) \setminus \{0\}$ , using  $\|\eta\|_1 = \|\eta_S\|_1 + \|\eta_{S^c}\|_1$ , we get

$$\|\eta_S\|_1 < \|\eta_{S^c}\|_1 \text{ if and only if } \frac{\|\eta_S\|_1}{\|\eta\|_1} < \frac{1}{2}.$$

Using this inequality and Definition 3, it follows that  $\Phi$  satisfies NUP of order  $s$  if and only if

$$\max_{|S| \leq s} \sup_{\eta \in \text{Null}(\Phi) \setminus \{0\}} \frac{\|\eta_S\|_1}{\|\eta\|_1} < \frac{1}{2}. \quad (12)$$

Since the objective function in (12) is independent of the scale of  $\eta$ , the condition in (12) is equivalent to

$$\max_{|S| \leq s} \max_{\eta \in \text{Null}(\Phi) \cap \mathbb{S}_1^{n-1}} \|\eta_S\|_1 < \frac{1}{2}. \quad (13)$$

Since we can replace  $\mathbb{S}_1^{n-1}$  with  $\mathbb{B}_1^n$  (this does not change the result of the maximization problem), we find that (13) is equivalent to (11), as claimed. ■

*Remark 3:* The value of the left hand side of (11) is called the *nullspace constant* in the literature [17]. The calculation of the nullspace constant for an arbitrary matrix  $\Phi$  and sparsity  $s$  is known to be NP-hard [17].

The reformulation of the NUP in Lemma 2 has certain benefits. For a fixed index set  $S \subseteq \{1, \dots, n\}$ , it draws our attention to the optimization problem

$$\max_{\eta \in \text{Null}(\Phi) \cap \mathbb{B}_1^n} \|\eta_S\|_1 \quad (14)$$

which is the maximization of a continuous convex function  $\|\Gamma_S(\cdot)\|_1$  over a compact convex set  $\text{Null}(\Phi) \cap \mathbb{B}_1^n$ . Thus, it follows from Proposition 1 that the maximum is attained at an extreme point of  $\text{Null}(\Phi) \cap \mathbb{B}_1^n$ . This leads us to want to understand the extreme points of this set, which can be a computationally involved task for arbitrary matrices  $\Phi$ . Nonetheless, one can still set a bound on the sparsity of the extreme points of  $\text{Null}(\Phi) \cap \mathbb{B}_1^n$ .

*Lemma 3:* If  $\Phi \in \mathbb{R}^{m \times n}$  is an  $m \times n$  matrix of rank  $r$ , then extreme points of  $\text{Null}(\Phi) \cap \mathbb{B}_1^n$  are at most  $(r+1)$ -sparse.

*Proof:* If  $\dim(\text{Null}(\Phi)) \equiv n - r \leq 1$ , then the result of the lemma holds trivially since  $r+1 \geq n$ .

The unit  $\ell_1$ -sphere in  $\mathbb{R}^n$ , namely  $\mathbb{S}_1^{n-1}$ , may be written as the union of  $(n-1)$ -dimensional simplices. Hence, if  $\dim(\text{Null}(\Phi)) \equiv n - r \geq 1$ , each extreme point is contained in an  $(n-1)$ -dimensional simplex.

In particular, if  $\dim(\text{Null}(\Phi)) \equiv n - r > 1$ , we argue that, no extreme point of  $\text{Null}(\Phi) \cap \mathbb{B}_1^n$  lies in the interior of these  $(n-1)$ -dimensional simplices. This is because in this case the intersection of  $\text{Null}(\Phi)$  and the interior of the  $(n-1)$ -dimensional simplex is either empty, or a non-singleton open convex subset of  $\mathbb{R}^n$ , where every point is a convex combination of two distinct points. Hence, no extreme point can live there. So, the extreme points must lie in the boundary of  $(n-1)$ -dimensional simplices, which are  $(n-2)$ -dimensional simplices. Moreover, these  $(n-2)$ -dimensional simplices are where one coordinate becomes zero. Hence, at least one coordinate of an extreme point living in an  $(n-2)$ -dimensional simplex is zero.

As long as the sum of the dimension of the simplex, which contains the extreme point, and the dimension of  $\text{Null}(\Phi)$  is strictly larger than  $n$ , we could repeat the argument in the previous paragraph. This argument stops when the dimension of the simplex containing the extreme point is  $r$ . Thus, at least  $n - r - 1$  coordinates of the extreme point have to be zero, so that each extreme point is at most  $(r+1)$ -sparse. ■

Although Lemma 3 is analogous to (7), it is stronger in the sense that the statement bounds the sparsity of all the extreme points of the convex set  $\text{Null}(\Phi) \cap \mathbb{B}_1^n$ , not just the sparsest vectors in the  $\text{Null}(\Phi)$ , as is implied by (7). Armed with Lemma 3, it turns out that for incidence matrices, these extreme points have a nice characterization in terms of the properties of the underlying graph.

*Proposition 3:* Let  $A \in \mathbb{R}^{m \times n}$  be the incidence matrix of a simple connected graph  $\mathcal{G}$  that has at least one simple cycle, and let  $W_1$  denote the set of normalized simple cycles of  $\mathcal{G}$ , i.e.

$$W_1 := \left\{ \frac{w(\mathcal{C})}{\|w(\mathcal{C})\|_1} \mid \mathcal{C} \text{ is a simple cycle of } \mathcal{G} \right\}. \quad (15)$$

Then, we have  $\text{Ext}(\text{Null}(A) \cap \mathbb{B}_1^n) \subseteq W_1$ .

*Proof:* Let  $z$  be an extreme point of  $\text{Null}(A) \cap \mathbb{B}_1^n$  with support  $S$ . We necessarily have  $\|z\|_1 = 1$ . Suppose that  $\mathcal{G}_S$  has  $\bar{m}$  vertices and  $\bar{k}$  connected components. Note that  $z_S \in \text{Null}(A_S) \cap \mathbb{B}_1^{|S|}$  has only nonzero entries. We claim that  $z_S$  is an extreme point of  $\text{Null}(A_S) \cap \mathbb{B}_1^{|S|}$ . For a proof by contradiction, suppose that  $z_S$  could be written as a convex combination of two distinct vectors in  $\text{Null}(A_S) \cap \mathbb{B}_1^{|S|}$ . In this case, one could pad those vectors with zeros at the coordinates in  $S^c$  to get two distinct vectors in  $\text{Null}(A) \cap \mathbb{B}_1^n$ , whose convex combination would give  $z$ . Since this would contradict the fact that  $z$  is an extreme point, we must conclude that  $z_S$  is an extreme point of  $\text{Null}(A_S) \cap \mathbb{B}_1^{|S|}$ .

Using the fact that  $z_S$  is an extreme point of  $\text{Null}(A_S) \cap \mathbb{B}_1^{|S|}$ , Lemma 3, and Remark 2, it follows that

$$|S| \leq \text{rank}(A_S) + 1 = \bar{m} - \bar{k} + 1. \quad (16)$$

Combining (16) with the rank-nullity theorem yields

$$\dim(\text{Null}(A_S)) = |S| - \bar{m} + \bar{k} \leq 1.$$

This may be combined with  $0 \neq z_S \in \text{Null}(A_S)$  to conclude that  $\dim(\text{Null}(A_S)) = 1$ ; thus  $\text{Null}(A_S)$  is spanned by  $z_S$ . It now follows from Remark 1 that  $z_S$  corresponds to the unique simple cycle in  $\mathcal{G}_S$ . In addition, since  $z_S$  has no zero entries, the corresponding simple cycle includes every edge of  $\mathcal{G}_S$ . It follows that  $z$  corresponds to a simple cycle in  $\mathcal{G}$ , which combined with  $\|z\|_1 = 1$  proves  $z \in W_1$ . ■

Next we show that Proposition 3 provides a way to study the maximization problem (14).

*Theorem 1:* Let  $A \in \mathbb{R}^{m \times n}$  be the incidence matrix of a simple, connected graph  $\mathcal{G}$ , and let  $W_1$  denote the set of normalized simple cycles of  $\mathcal{G}$  as in (15). Let  $S \subseteq \{1, \dots, n\}$  be a nonempty index set. Then

$$\max_{\eta \in \text{Null}(A) \cap \mathbb{B}_1^n} \|\eta_S\|_1 = \max_{z \in W_1} \|z_S\|_1. \quad (17)$$

*Proof:* Since  $W_1 \subset \text{Null}(A) \cap \mathbb{B}_1^n$ , we obviously have

$$\max_{\eta \in \text{Null}(A) \cap \mathbb{B}_1^n} \|\eta_S\|_1 \geq \max_{z \in W_1} \|z_S\|_1.$$

For the converse inequality we argue as follows: Since  $\|\Gamma_S(\cdot)\|_1$  is a continuous convex function (thus also quasi-convex), and  $\text{Null}(A) \cap \mathbb{B}_1^n$  is a compact convex set, the maximum in the left hand side of (17) will be attained at an extreme point of  $\text{Null}(A) \cap \mathbb{B}_1^n$  (see Proposition 1). That is,

$$\max_{\eta \in \text{Null}(A) \cap \mathbb{B}_1^n} \|\eta_S\|_1 = \max_{\nu \in \text{Ext}(\text{Null}(A) \cap \mathbb{B}_1^n)} \|\nu_S\|_1. \quad (18)$$

But  $\text{Ext}(\text{Null}(A) \cap \mathbb{B}_1^n) \subseteq W_1$  by Proposition 3. Hence,

$$\max_{\nu \in \text{Ext}(\text{Null}(A) \cap \mathbb{B}_1^n)} \|\nu_S\|_1 \leq \max_{z \in W_1} \|z_S\|_1.$$

Combining this with (18) we get the converse inequality, and hence the result.  $\blacksquare$

Theorem 1 builds a connection between the algebraic condition NUP and a topological property of the graph, namely its simple cycles. This connection is what we will primarily exploit in the rest of the paper.

Let us make a simple but important observation.

*Lemma 4:* Let  $\mathcal{G}$  and  $W_1$  be as in Theorem 1. Then, for any index set  $S$  and  $z \in W_1$ , it holds that

$$\|z_S\|_1 = \frac{|S \cap \text{supp}(z)|}{\|z\|_0}.$$

*Proof:* For any simple cycle  $\mathcal{C}$ , the entries of  $w(\mathcal{C})$  are in the set  $\{0, -1, 1\}$ . Hence, entries of any  $z \in W_1$  are from the set  $\{0, \frac{1}{\|z\|_0}, \frac{-1}{\|z\|_0}\}$ , from which the result follows.  $\blacksquare$

### B. Polynomial Time Guarantees for Sparse Recovery

An answer to the first question posed at the beginning of Section III is given by Theorem 2.

*Theorem 2:* Let  $A \in \mathbb{R}^{m \times n}$  be the incidence matrix of a simple, connected graph  $\mathcal{G}$  with girth  $g$ . Then, every  $s$ -sparse vector  $\bar{x} \in \mathbb{R}^n$  is the unique solution to

$$\min_{A\bar{x}=Ax} \|\bar{x}\|_1$$

if and only if  $s < \frac{g}{2}$ .

*Proof:* From Lemma 2 and Theorem 1 it follows that the NUP of order  $s$  is satisfied if and only if

$$\max_{|S| \leq s} \max_{z \in W_1} \|z_S\|_1 < \frac{1}{2}. \quad (19)$$

From Lemma 4, the maximum on the left of (19) is equal to

$$\max_{|S| \leq s} \max_{z \in W_1} \frac{|S \cap \text{supp}(z)|}{\|z\|_0}. \quad (20)$$

If  $s \geq g$ , then (20) is equal to 1. If  $s < g$ , the maximum is attained at the smallest simple cycle when  $S$  picks edges only from this cycle, and the maximum is equal to  $\frac{s}{g}$ . Hence,

$$\max_{|S| \leq s} \max_{z \in W_1} \|z_S\|_1 = \min\{s/g, 1\}. \quad (21)$$

The desired result now follows from (19) and (21).  $\blacksquare$

*Remark 4:* As was mentioned in Remark 3, calculating the nullspace constant is NP-hard in general. However, there are algorithms that can calculate the girth of a graph exactly

in  $\mathcal{O}(mn)$  or sufficiently accurately for our purposes in  $\mathcal{O}(m^2)$  (see [22]). Therefore, Theorem 2 reveals that the nullspace constant can be calculated—hence the NUP can be verified—for graph incidence matrices in polynomial time.

### C. Guarantees for Support-dependent Sparse Recovery

The sparse recovery result given by Theorem 2 depends on the cardinality  $s$  of the support of the unknown signal. For a graph with small girth (e.g.,  $g = 3$ ), this result establishes sparse recovery guarantees only for signals with low sparsity levels (specifically,  $s < g/2$ ). It is natural, therefore, to ask the following question: Given a graph with relatively small girth, can we identify a class of signals with sparsity  $s \geq g/2$  that can be recovered? This leads us to the study of support-dependent sparse recovery, where we not only focus on the cardinality of the support of the unknown signal  $\bar{x}$ , but also take the location of the support into account. More precisely, we have the following result.

*Theorem 3:* Let  $A \in \mathbb{R}^{m \times n}$  be the incidence matrix of a simple, connected graph  $\mathcal{G}$ , let  $W_1$  denote the set of normalized simple cycles of  $\mathcal{G}$  as in (15), and let  $\emptyset \neq S \subseteq \{1, \dots, n\}$ . Then, every vector  $\bar{x} \in \mathbb{R}^n$  with  $\text{supp}(\bar{x}) \subseteq S$  is the unique solution to the optimization problem

$$\min_{A\bar{x}=Ax} \|\bar{x}\|_1 \quad (22)$$

if and only if

$$\max_{z \in W_1} \frac{|S \cap \text{supp}(z)|}{\|z\|_0} < \frac{1}{2}. \quad (23)$$

*Proof:* From arguments in the proof of Lemma 2, we know that (6) is equivalent to

$$\max_{\eta \in \text{Null}(A) \cap \mathbb{B}_1^n} \|\eta_S\|_1 < \frac{1}{2},$$

which with Theorem 1 and Lemma 4 gives the result.  $\blacksquare$

*Remark 5:* In comparison to Theorem 2, Theorem 3 provides a deeper insight into the vectors  $\bar{x}$  that can be recovered from the observations  $A\bar{x}$  via  $\ell_1$ -minimization. Specifically, any vector whose support within each simple cycle has size strictly less than half the size of that cycle, can be successfully recovered.

As an example consider the graph in Fig. 1, which has girth  $g = 4$ . Thus, it follows from Theorem 2 that any 1-sparse signal can be recovered. However, in reality, some signals that are supported on more than one edge can also be recovered. For instance, Theorem 3 guarantees that a signal  $\bar{x}$  supported on edges  $\{2, 7\}$  or  $\{4, 6, 8\}$  can be recovered as the unique solution to the  $\ell_1$  problem (22) since for each of the three simple cycles, its intersection with the support of the unknown vector  $\bar{x}$  is less than half of the length of the simple cycle. This reveals that different sparsity patterns of the same sparsity level can have different recovery performances.

*Remark 6:* To use Theorem 3 as a way of providing a support-dependent recovery guarantee, one needs to compute the intersection of the support of  $\bar{x}$  with all simple cycles in the graph. A number of algorithms with a polynomial time complexity bound for cycle enumeration are known [23].

#### D. Using Measurement Sparsity to Aid Recovery

When there is additional information about the support of the unknown signal, Theorem 3 gives a necessary and sufficient condition for exact recovery. In practice, this information is usually missing. However, the special structure of the incidence matrix and its connection to the graph can help us circumvent this difficulty. Notice that the columns of any incidence matrix are always 2-sparse, which means that the measurement  $y := A\bar{x}$  will be  $2s$ -sparse for any  $s$ -sparse signal  $\bar{x}$ . Therefore, one can seek to obtain a superset of  $\text{supp}(\bar{x})$  by observing  $\text{supp}(y)$ , i.e. the vertices with nonzero measurements. Typically, this observation reduces the size of the problem, and gives rise to Algorithm 1.

---

**Algorithm 1** Recovering the unknown signal by using sparse vertex measurements.

---

- 1: **INPUT:**  $y \in \mathbb{R}^m$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $\mathcal{G}(V, E)$ .
- 2: Set  $\hat{x} = 0$ .
- 3: Define  $T := \text{supp}(y)$ .
- 4: Define  $S := \{j \mid e_j = (v_i, v_k) \in E \text{ and } v_i, v_k \in T\}$ , the indices of the edges of  $\mathcal{G}$  connecting vertices with nonzero measurements.
- 5: Construct  $\mathcal{G}_S$ , the subgraph of  $\mathcal{G}$  consisting of the edges indexed by  $S$ .
- 6: Construct the incidence matrix  $A_S$  of the subgraph  $\mathcal{G}_S$ .
- 7: Solve the  $\ell_1$ -minimization problem

$$\tilde{x} = \arg \min_{y_T = A_S x} \|x\|_1$$

- 8: Set  $\hat{x}_S = \tilde{x}$ .
  - 9: **return**  $\hat{x} \in \mathbb{R}^n$
- 

*Remark 7:* There are cases where Algorithm 1 may fail. For instance, when the unknown nonzero signal  $\bar{x}$  is in the nullspace of one of the rows of  $A$ , say  $a_k$ , and  $\text{supp}(a_k) \cap \text{supp}(\bar{x}) \neq \emptyset$  (i.e. when the non-zero values of the signal  $\bar{x}$  cancel each other at a vertex  $k$ ), Algorithm 1 would simply ignore the corresponding vertex, and hence, the edges connected to it. This leads to a wrong subgraph  $\mathcal{G}_S$  and possibly an incorrect outcome  $\hat{x}$ . Fortunately, this case happens with zero probability when the signal has random support and random values.

*Corollary 1:* Let  $A \in \mathbb{R}^{m \times n}$  be the incidence matrix of a simple, connected graph  $\mathcal{G}$ . Given a random signal  $\bar{x}$ , let  $y := A\bar{x}$  and  $\mathcal{G}_S$  be the subgraph associated with  $y$  in Algorithm 1. Then, Algorithm 1 will almost surely recover the signal  $\bar{x}$  if and only if  $\text{supp}(\bar{x})$  picks strictly less than half of the edges from each simple cycle of  $\mathcal{G}_S$ .

The proof of Corollary 1 follows from Theorem 3 and the fact that  $\bar{x}$  is assumed to be randomly generated.

In comparison to  $\ell_1$ -minimization on the original graph, Algorithm 1 may have improved support-dependent recovery performance because the formation of the subgraph may eliminate cycles; this is especially true for large-scale networks. For concreteness, let us give a simple example.

*Example 2:* Consider the graph depicted in Fig. 1. Let  $\bar{x}$  be a random signal that is supported on the edges

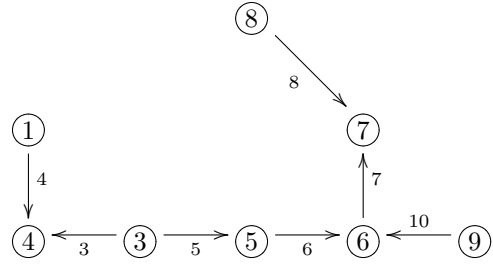


Fig. 2: Acyclic subgraph  $\mathcal{G}_S$  induced by an unknown signal with support  $S = \{3, 4, 5, 6, 7, 8, 10\}$ .

$\{3, 4, 5, 6, 7, 8, 10\}$ , so that (with probability one) there are nonzero measurements on nodes  $\{1, 3, 4, 5, 6, 7, 8, 9\}$ . Then,  $x$  cannot be recovered using the incidence matrix of the original graph. However, the subgraph  $\mathcal{G}_S$  is acyclic, as shown in Fig. 2, which allows exact recovery via Algorithm 1.

#### IV. SIMULATIONS

In this section we provide numerical simulation results on the recovery performance associated with incidence matrices. In all experiments, the signal  $\bar{x}$  has random support with each nonzero entry drawn i.i.d. from the standard normal distribution. We use the CVX software package [24] to solve the optimization problems. A vector is declared to be recovered if the 2-norm error is less than or equal to  $10^{-5}$ .

*Experiment 1:* Fig. 4 shows the probability of exact recovery of signals via  $\ell_1$ -relaxation over a sequence of graphs containing two cycles (see Fig. 3). Each graph has a cycle of length 3 and a larger one of varying lengths 3, 5, 7, 9, and 11, respectively. For each sparsity level, 1000 trials were performed. According to Theorem 2, for all graphs we can recover 1-sparse signals since the girth is 3 for each graph in the sequence. When the sparsity level is increased, we expect that the probability of exact recovery will increase for graphs with larger cycles because it becomes less likely that the support of the random signal will consist of more than half of the edges for one of the simple cycles in the graph. Note that this agrees with our observation in Fig. 4.

*Experiment 2:* We now evaluate the performance of Algorithm 1 against the  $\ell_1$ -minimization method in (10) on

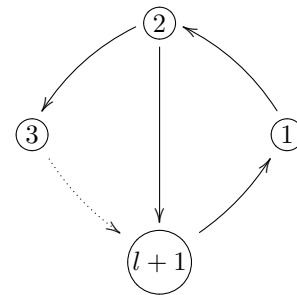


Fig. 3: Graph used in Experiment 1, which consists of a cycle of length 3 and a cycle of varying length  $l \in \{3, 5, 7, 9, 11\}$ .

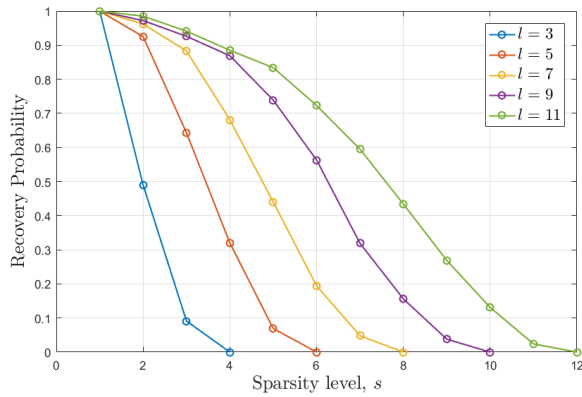


Fig. 4: Probability of exact recovery as a function of the sparsity level for the graphs of various loop sizes in Fig. 3.

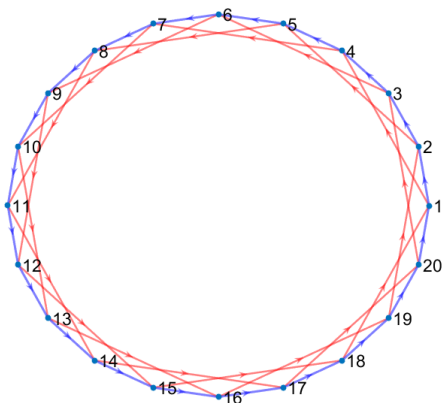


Fig. 5: Graphs with 20 nodes for Experiment 2.  $\mathcal{G}_B$  consists of blue edges only and  $\mathcal{G}_{BR}$  consists of blue and red edges.

two graphs with 20 nodes (see Fig. 5). The first graph,  $\mathcal{G}_B$ , is a simple cycle connecting node 1 to node 20 in order (blue edges in Fig. 5). The second graph,  $\mathcal{G}_{BR}$ , consists of both blue and red edges. The red edges connect each node to its third neighbor, i.e.,  $(1, 4), (2, 5) \cdots (17, 20) \cdots (20, 3)$ . Fig. 6 and Fig. 7 show the performance of both algorithms on  $\mathcal{G}_B$  and  $\mathcal{G}_{BR}$  respectively. For each sparsity level, the experiments are repeated 1000 times to compute recovery probabilities. For  $\mathcal{G}_B$ , Algorithm 1 outperforms  $\ell_1$ -minimization since the reduced graph in Algorithm 1 will be acyclic in some cases. For  $\mathcal{G}_{BR}$ , Algorithm 1 will eliminate small cycles when forming the subgraph and lead to higher recovery probability for fixed  $s$  due to a larger girth.

## V. CONCLUSION

In this paper we studied sparse recovery for the class of graph incidence matrices. For such matrices we provided a characterization of the NUP in terms of the topological properties of the underlying graph. This characterization allows one to verify sparse recovery guarantees in polynomial time for the class of incidence matrices. Moreover, we showed

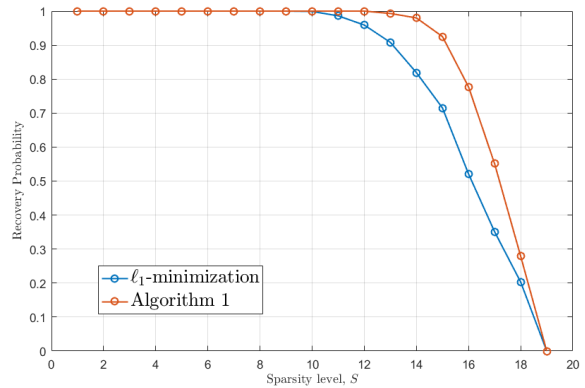


Fig. 6: Probability of exact recovery as a function of the sparsity level for Algorithm 1 and  $\ell_1$ -minimization on  $\mathcal{G}_B$ .

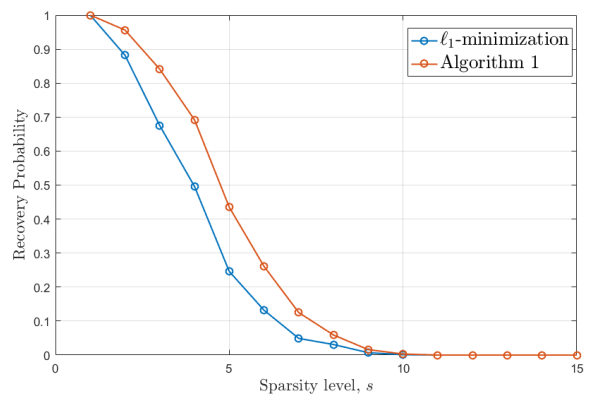


Fig. 7: Probability of exact recovery as a function of the sparsity level for Algorithm 1 and  $\ell_1$ -minimization on  $\mathcal{G}_{BR}$ .

that support-dependent recovery performance can also be analyzed in terms of the simple cycles of the graph. Finally, by exploiting the structure of incidence matrices and sparsity of the measurements, we proposed an efficient algorithm for recovery and evaluated its performance numerically.

## REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [3] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [4] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2790–2797, IEEE, 2009.
- [5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [6] M. Lustig, D. Donoho, and J. Pauly, "Rapid MR imaging with compressed sensing and randomly under-sampled 3DFT trajectories," in *Proc. 14th Ann. Meeting ISMRM*, Citeseer, 2006.
- [7] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

- [8] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [9] R. Liégeois, B. Mishra, M. Zorzi, and R. Sepulchre, "Sparse plus low-rank autoregressive identification in neuroimaging time series," in *IEEE Conference on Decision and Control*, pp. 3965–3970, 2015.
- [10] M. Coates, Y. Pointurier, and M. Rabbat, "Compressed network monitoring," in *IEEE/SP Workshop on Statistical Signal Processing*, pp. 418–422, IEEE, 2007.
- [11] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 92–101, 2008.
- [12] W. Xu, E. Mallada, and A. Tang, "Compressive sensing over graphs," in *IEEE INFOCOM*, pp. 2087–2095, IEEE, 2011.
- [13] R. G. Michael and S. J. David, "Computers and intractability: a guide to the theory of NP-completeness," *WH Free. Co., San Fr*, pp. 90–91, 1979.
- [14] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best  $k$ -term approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, 2009.
- [15] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [16] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [17] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1248–1259, 2014.
- [18] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.
- [19] F. Flores-Bazán, F. Flores-Bazán, and C. Vera, "Maximizing and minimizing quasiconvex functions: related properties, existence and optimality conditions via radial epiderivatives," *Journal of Global Optimization*, vol. 63, pp. 99–123, Sep 2015.
- [20] C. Godsil and G. F. Royle, *Algebraic graph theory*. Graduate text in mathematics, Springer, New York, 2001.
- [21] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, vol. 1. Birkhäuser Basel, 2013.
- [22] A. Itai and M. Rodeh, "Finding a minimum circuit in a graph," *SIAM Journal on Computing*, vol. 7, no. 4, pp. 413–423, 1978.
- [23] P. Mateti and N. Deo, "On algorithms for enumerating all circuits of a graph," *SIAM Journal on Computing*, vol. 5, no. 1, pp. 90–99, 1976.
- [24] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2008.