

ISE

Industrial and
Systems Engineering

What is the Largest Sparsity Pattern That Can Be Recovered By 1-Norm Minimization?

MUSTAFA D. KABA, MENGAN ZHAO, RENÉ VIDAL¹, DANIEL P. ROBINSON²,
AND ENRIQUE MALLADA³

¹Fellow, IEEE

²Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA

³Senior Member, IEEE

ISE Technical Report 22T-015



LEHIGH
UNIVERSITY.

What is the Largest Sparsity Pattern that Can Be Recovered by 1-Norm Minimization?

Mustafa D. Kaba, Mengnan Zhao, René Vidal, *Fellow, IEEE*, Daniel P. Robinson, Enrique Mallada, *Senior Member, IEEE*.

Abstract—Much of the existing literature in sparse recovery is concerned with the following question: given a sparsity pattern and a corresponding regularizer, derive conditions on the dictionary under which exact recovery is possible. In this paper, we study the opposite question: given a dictionary and the ℓ_1 -norm regularizer, find the largest sparsity pattern that can be recovered. We show that such a pattern is described by a mathematical object called a “maximum abstract simplicial complex,” and provide two different characterizations of this object: one based on extreme points and the other based on vectors of minimal support. In addition, we show how this new framework is useful in the study of sparse recovery problems when the dictionary takes the form of a graph incidence matrix or a partial discrete Fourier transform. In case of incidence matrices, we show that the largest sparsity pattern that can be recovered is determined by the set of simple cycles of the graph. As a byproduct, we show that standard sparse recovery can be certified in polynomial time, although this is known to be NP-hard for general matrices. In the case of the partial discrete Fourier transform, our characterization of the largest sparsity pattern that can be recovered requires the unknown signal to be real and its dimension to be a prime number.

Index Terms—Compressed sensing, convex optimization, nullspace property, sparse solution of linear equations, sparsity.

I. INTRODUCTION

THE widespread use of sparse recovery methods [2], [3] in data acquisition [4], machine learning [5], [6], [7], [8], medical imaging [9], [10], [11], [12], and networking [13], [14], [15] has made sparse recovery a popular research area and application tool. The goal of sparse recovery is to find an unknown signal $\bar{x} \in \mathbb{R}^n$ from a number of observations $\Phi\bar{x}$, where $\Phi \in \mathbb{R}^{m \times n}$ is a *dictionary* or *measurement matrix* possibly (but not necessarily) with $m \ll n$, under the assumption that the unknown signal \bar{x} is sparse in some sense. For instance, *regular sparsity* assumes that the number

of nonzero entries of \bar{x} is limited by an integer $s \ll n$ [3]. Similarly, *block-sparsity* assumes that there is a collection of non-overlapping groups (i.e. blocks) that cover all entries of \bar{x} , and that when the unknown signal is restricted to these groups, only a small number of such restrictions are nonzero [16]. Further examples include *group-sparsity* [17], a generalization of block-sparsity that removes the assumption of non-overlapping groups, *tree-based sparsity* [18], [19], *model-based compressed sensing* [20], and *sparsity in levels* [21].

When recovering signals with some notion of sparsity, the typical approach is to construct a regularizer that will guarantee the recovery of an unknown sparse signal with high probability. However, in some applications of practical importance where the unknown signal is sparse in a non-regular sense, we still see that ℓ_1 -recovery is the standard recovery method [22] or that ℓ_1 -recovery performs as good as recovery via a tailored regularizer [23]. There are several reasons why recovery via ℓ_1 -minimization might be preferable over recovery via regularizers tailored for the sparsity pattern in hand for some applications. Probably the most important reason is that ℓ_1 -minimization can be cast as a linear program, hence a solution can be found efficiently even for large scale problems [24]. This motivates the exploration of the exact recovery capabilities of ℓ_1 -minimization. In other words, we would like to understand how far the exact recovery capability of ℓ_1 -minimization goes beyond regular sparsity.

In general, a *sparsity pattern* can be defined as a collection of index sets. However, some of the most common sparsity patterns studied in the literature satisfy a noteworthy property, which for regular sparsity can be stated as follows: If a vector is s -sparse with support S , then any vector with support $S' \subseteq S$ is also s -sparse. More generally, we observe that most common sparsity patterns have the following property:

(P). *If S belongs to a sparsity pattern, then $S' \subseteq S$ also belongs to the same sparsity pattern.*

Our claim in this paper is that when the ℓ_1 regularizer is used, property (P) can serve as the basis for a unified treatment of sparsity patterns including those discussed above.

Understanding the exact recovery capabilities of ℓ_1 -minimization has great importance for dictionaries arising in various applications. In this paper we focus on two examples, namely graph incidence matrices and the partial Discrete Fourier Transform (DFT). Our interest in graph incidence matrices stems from their use as a fundamental representation of graphs, and thus as a natural choice when analyzing network flows. In addition, the detection of sparse structural network

A preliminary version of this paper was presented at the 57th IEEE Conference on Decision and Control [1].

This work was supported by NSF grants CCF 1618637, IIS 1704458, AMPS 1736448 and CAREER 1752362.

M.D. Kaba is with eBay Inc. This work was done while he was working at the Department of Applied Mathematics and Statistics of Johns Hopkins University, Baltimore, MD 21218, USA. M. Zhao and E. Mallada are with the Department of Electrical and Computer Engineering, and R. Vidal is with the Mathematical Institute for Data Science and the Department of Biomedical Engineering, at Johns Hopkins University, Baltimore, MD 21218, USA. D.P. Robinson is with the Department of Industrial and Systems Engineering at Lehigh University, Bethlehem, PA 18015, USA. Emails: mkaba@ebay.com, {mzhao21, rvidal, mallada}@jhu.edu and dpr219@lehigh.edu.

Copyright © 2021 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

changes via observations at the nodes can be modeled as an ℓ_1 -recovery problem, where the incidence matrix serves as the dictionary. For instance, this approach is used in [25] to detect physical and cyber attacks on power grids. On the other hand, the DFT is one of the most important discrete transforms that influences applications ranging from image processing to solving partial differential equations [26].

In this paper we answer the following question: Given a dictionary, what is the largest sparsity pattern that can be recovered by ℓ_1 -minimization? Our specific contributions can be summarized as follows:

- 1) We generalize the well-known Nullspace Property to a family of sparsity patterns described by a mathematical object called an *abstract simplicial complex (ASC)* (§ III-A). This leads to a characterization of all sparsity patterns recoverable via ℓ_1 -minimization (§ III-B), including the largest one, which we call *maximum abstract simplicial complex (MASC)*.
- 2) We provide two characterizations of the MASC associated with a dictionary. The first uses the extreme points of the convex set formed by the intersection of the nullspace of the dictionary and the ℓ_1 -ball (§ IV-A). The second one is based on the fact that the extreme points and the vectors of minimal support coincide for sets defined as the intersection of a subspace and the ℓ_1 -ball (§ IV-B).
- 3) For graph incidence matrices, we show that the success of ℓ_1 -minimization is determined by the topology of the graph, specifically by its simple cycles. Moreover, we show that the decision of whether all s -sparse signals can be recovered via ℓ_1 -minimization can be made in polynomial time, although it is NP-hard in general (§ V-B).
- 4) When the dictionary is a partial DFT matrix and the unknown signal is real and its dimension is a prime number, we completely characterize the support sets for which ℓ_1 -recovery is always successful. Under stronger assumptions we show that a computationally more advantageous characterization is possible, and we provide a useful lower bound on the maximum sparsity level for which all signals can be recovered (§ VI-B).
- 5) We illustrate the importance of our results with experiments on incidence (§ V-C) and partial DFT matrices (§ VI-C).

II. PRELIMINARIES

A. Notation

Given a vector $x \in \mathbb{R}^n$, we denote its ℓ_p -norm for $p \geq 1$ by $\|x\|_p := (\sum_{k=0}^{n-1} |x_k|^p)^{1/p}$. We denote the unit ℓ_p -sphere in \mathbb{R}^n by $S_p^{n-1} := \{x \in \mathbb{R}^n : \|x\|_p = 1\}$. Similarly, the unit ℓ_p -ball in \mathbb{R}^n is denoted by $B_p^n := \{x \in \mathbb{R}^n : \|x\|_p \leq 1\}$.¹

We denote the function that counts the number of nonzero entries in a vector $x \in \mathbb{R}^n$ by $\|x\|_0$.² We say that a vector x is s -sparse if $\|x\|_0 \leq s$. To emphasize that the vector has precisely s nonzero entries, we say it is *exactly s -sparse*.

¹In order to avoid confusion, we emphasize once again that throughout this paper, S_p^{n-1} and B_p^n are objects that consist only of real vectors.

²Although $\|\cdot\|_0$ is not a norm, it is common jargon to call it the ℓ_0 -norm.

The *codimension* of a d -dimensional subspace $\mathcal{V} \subseteq \mathbb{R}^n$ is defined to be $n - d$. The *nullspace* of a matrix $\Phi \in \mathbb{R}^{m \times n}$ will be denoted by $\text{Null}(\Phi)$. That is, $\text{Null}(\Phi) = \{x \in \mathbb{R}^n : \Phi x = 0\}$. We define $\mathcal{U}_n := \{0, \dots, n-1\}$. When $S \subseteq \mathcal{U}_n$, we assume that S has the natural ordering, and S_k denotes the k^{th} element of S . For a vector $x \in \mathbb{R}^n$ and index set $S \subseteq \mathcal{U}_n$, we denote the part of x supported on S by x_S , so that $x_S \in \mathbb{R}^{|S|}$, where $|S|$ denotes the cardinality of a set S . When we would like to keep the dimension unchanged, we use the projection map $\text{Pr}_S : \mathbb{R}^n \rightarrow \mathbb{R}^n$, which simply projects vectors onto the coordinates indexed by S . The complement of S in \mathcal{U}_n is denoted by S^c , and the collection of all subsets of S (i.e. the *power set* of S) is denoted by 2^S .

For a nonempty convex set $C \subseteq \mathbb{R}^n$, $\text{Ext}(C)$ denotes the set of extreme points of C , which are precisely the points that cannot be written as a nontrivial convex combination of two distinct points in C . The *affine hull* of C , denoted by $\text{aff}(C)$, is the smallest affine set in \mathbb{R}^n that contains C . Alternatively, it can be characterized as the intersection of all affine sets containing C [27, p.6]. Note that the affine hull of a point in \mathbb{R}^n consists only of the point itself. The *relative interior* of C is denoted by $\text{rinte}(C)$. Formally, it is defined [27, p.44] as

$$\text{rinte}(C) := \{x \in \text{aff}(C) : \exists \varepsilon > 0, (x + \varepsilon B_2^n) \cap \text{aff}(C) \subseteq C\}.$$

Hence, the relative interior of a single point is itself. That is, $\text{rinte}(\{x\}) = \{x\}$ for all $x \in \mathbb{R}^n$. The closure of C is denoted by $\text{clo}(C)$. It is formally defined [27, p.44] as

$$\text{clo}(C) := \bigcap \{C + \varepsilon B_2^n : \varepsilon > 0\}.$$

The *relative boundary* of C is defined as $\partial C = \text{clo}(C) \setminus \text{rinte}(C)$ [27, p.44]. See [27] for additional details.

Throughout paper we introduce several acronyms. In order to help the reader keep track of them, we include a list here.

Acronym	Explanation
DFT	Discrete Fourier Transform
ASC	Abstract Simplicial Complex
MASC	Maximum Abstract Simplicial Complex
MC	Mutual Coherence
NUP	Nullspace Property
RIP	Restricted Isometry Property
GNUP	Generalized Nullspace Property
MRS�	Maximal Recoverable Sparsity Level

B. Sparse Recovery Via ℓ_1 -Minimization

A naive approach to the recovery of s -sparse signals is to pose the ℓ_0 optimization problem

$$\min_{\Phi \bar{x} = \Phi x} \|x\|_0. \quad (1)$$

However, it is well-known that (1) is NP-hard to solve. Hence, the following ℓ_1 convex relaxation is commonly studied

$$\min_{\Phi \bar{x} = \Phi x} \|x\|_1. \quad (2)$$

The minimizer of (2) is unique and coincides with the original signal \bar{x} if the dictionary Φ satisfies certain properties. Popular properties in the literature include the *Restricted Isometry*

Property (RIP) [28], [29], *Mutual Coherence* (MC) [30] and *Nullspace Property* (NUP) [31]. Among these, the RIP and MC are sufficient conditions, whereas the NUP is a necessary and sufficient condition. In this paper we focus on the NUP.

Definition 1 (Nullspace Property (NUP)). *A matrix $\Phi \in \mathbb{R}^{m \times n}$ satisfies the NUP of order s if and only if every $\eta \in \text{Null}(\Phi) \setminus \{0\}$ and index set $S \subseteq \mathcal{U}_n$ with $|S| \leq s$ satisfy*

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

The connection between the NUP and exact sparse recovery is captured by the following result.

Theorem 1. [32, Thm. 4.5] *Let $\Phi \in \mathbb{R}^{m \times n}$. Any s -sparse vector $\bar{x} \in \mathbb{R}^n$ is the unique solution to the optimization problem (2) if and only if the matrix Φ satisfies the NUP of order s .*

Proof. See the proof of [32, Thm. 4.5]. \square

III. SPARSITY PATTERNS AS ABSTRACT SIMPLICIAL COMPLEXES

In this section we introduce the framework in which we study a generalization of the sparse recovery problem. We first introduce a generalization of the NUP, and then show how it naturally leads to the definition of the Maximum Abstract Simplicial Complex (MASC), which encapsulates the sparsity patterns that can be recovered via ℓ_1 -minimization.

A. Generalized Nullspace Property

To generalize Thm. 1, we start by generalizing the NUP. For this purpose, we turn to property (P) and define an Abstract Simplicial Complex (ASC) as any collection of index sets that satisfies a generalization of property (P).

Definition 2 (Abstract Simplicial Complex (ASC)). *Let Ω be a nonempty set. A nonempty set $\mathcal{T} \subseteq 2^\Omega$ is called an ASC if and only if for any $S \in \mathcal{T}$ and $W \subseteq S$, we have $W \in \mathcal{T}$.*

It is easy to see that the collection of supports of all s -sparse, s -block-sparse and s -group-sparse signals each form an ASC. A less trivial sparsity pattern that can be associated with an ASC is found in [22]. There, the authors consider a *convolutional sparse model* $y = Dx$, where the global dictionary $D \in \mathbb{R}^{N \times mN}$ is the concatenation of all shifted versions of a local dictionary $D_L \in \mathbb{R}^{m \times n}$. Due to the special structure of D , a predefined collection of patches (i.e. groups) $\{\Lambda_i\}$ on y defines a collection of groups $\{G_i\}$ (with possible overlap) on x through the relation $y = Dx$. Then, for a given signal $x \in \mathbb{R}^n$, they define the norm

$$\|x\|_0 = \max_i \|x_{G_i}\|_0.$$

The authors are interested in the signals with $\|x\|_0 \leq s$. It can be shown that the collection of the supports of such signals forms an ASC. To see this we note that

$$\|x\|_0 \leq s \iff |\text{Supp}(x) \cap G_i| \leq s \text{ for all } i,$$

and that $\{S \subseteq \mathcal{U}_n : |S \cap G_i| \leq s \text{ for all } i\}$ is an ASC.

We now use the ASC to state our generalization of the NUP.

Definition 3 (Generalized Nullspace Property (GNUP)). *Let $\mathcal{T} \subseteq 2^{\mathcal{U}_n}$. We say that a matrix $\Phi \in \mathbb{R}^{m \times n}$ satisfies the GNUP with respect to (w.r.t.) \mathcal{T} if and only if every $\eta \in \text{Null}(\Phi) \setminus \{0\}$ and set $S \in \mathcal{T}$ satisfy*

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

The GNUP in Defn. 3 is a *generalization* of the NUP since the latter may be recovered by choosing

$$\mathcal{T} = \mathcal{T}_s := \{S \in 2^{\mathcal{U}_n} : |S| \leq s\} \quad (3)$$

so that \mathcal{T} is an ASC and the GNUP is equivalent to the NUP.

We are now ready to generalize Thm. 1 to the GNUP.

Theorem 2. *Let $\mathcal{T} \subseteq 2^{\mathcal{U}_n}$ be an ASC. Every $\bar{x} \in \mathbb{R}^n$ with $\text{Supp}(\bar{x}) \in \mathcal{T}$ is the unique solution to the optimization problem (2) if and only if Φ satisfies the GNUP w.r.t. \mathcal{T} .*

Proof. This theorem is a generalization of the classical Thm. 1 whose proof follows mutatis mutandis.

\Rightarrow] Let $\eta \in \text{Null}(\Phi)$ be nonzero and $S \in \mathcal{T}$ so that

$$0 = \Phi\eta = \Phi(\text{Pr}_S(\eta) - (-\text{Pr}_{S^c}(\eta))),$$

which shows using $\bar{x} := \text{Pr}_S(\eta)$ that $\Phi\bar{x} = \Phi(-\text{Pr}_{S^c}(\eta))$. Since $\text{Supp}(\bar{x}) \subseteq S \in \mathcal{T}$ and \mathcal{T} is an ASC, it follows that $\text{Supp}(\bar{x}) \in \mathcal{T}$. Combining this with the hypothesis for the direction we are proving shows that \bar{x} as the unique solution to (2). Hence, it follows that $\|\eta_S\|_1 = \|\text{Pr}_S(\eta)\|_1 = \|\bar{x}\|_1 < \|\text{Pr}_{S^c}(\eta)\|_1 = \|\eta_{S^c}\|_1$, which completes the proof.

\Leftarrow] Let $\bar{x} \in \mathbb{R}^n$ satisfy $S = \text{Supp}(\bar{x}) \in \mathcal{T}$. Then, let $x \in \mathbb{R}^n$ be any vector satisfying $\Phi\bar{x} = \Phi x$ and $\bar{x} \neq x$. By setting $\eta = \bar{x} - x$ we see that $\eta \in \text{Null}(\Phi) \setminus \{0\}$ and that

$$\begin{aligned} \|\bar{x}\|_1 &= \|\bar{x} - \text{Pr}_S(x) + \text{Pr}_S(x)\|_1 \\ &\leq \|\bar{x} - \text{Pr}_S(x)\|_1 + \|\text{Pr}_S(x)\|_1 \quad (\text{triangle inequality}) \\ &= \|\text{Pr}_S(\bar{x} - x)\|_1 + \|x_S\|_1 \quad (\text{since } S = \text{Supp}(\bar{x})) \\ &= \|\eta_S\|_1 + \|x_S\|_1 \\ &< \|\eta_{S^c}\|_1 + \|x_S\|_1 \quad (\text{use the GNUP}) \\ &= \|\text{Pr}_{S^c}(\eta)\|_1 + \|x_S\|_1 \quad (\text{since } S = \text{Supp}(\bar{x})) \\ &= \|x\|_1 \end{aligned}$$

which proves that \bar{x} is the unique solution to (2). \square

A useful alternative formulation of the GNUP is now stated.

Lemma 1. *Let $\mathcal{T} \subseteq 2^{\mathcal{U}_n}$. A matrix $\Phi \in \mathbb{R}^{m \times n}$ satisfies the GNUP w.r.t. \mathcal{T} if and only if*

$$\max_{S \in \mathcal{T}} \max_{\eta \in \text{Null}(\Phi)} \max_{B_1^n} \|\eta_S\|_1 < \frac{1}{2}. \quad (4)$$

Proof. By definition of the GNUP, the matrix Φ satisfies the GNUP w.r.t. \mathcal{T} if and only if for all nonzero $\eta \in \text{Null}(\Phi)$ and $S \in \mathcal{T}$ it holds that $\|\eta_S\|_1 < \|\eta_{S^c}\|_1$. Without loss of generality, we can normalize η , and assume $\|\eta_S\|_1 + \|\eta_{S^c}\|_1 = \|\eta\|_1 = 1$. Then, Φ satisfies the GNUP w.r.t. \mathcal{T} if and only if

$$\max_{S \in \mathcal{T}} \max_{\eta \in \text{Null}(\Phi)} \max_{S_1^{n-1}} \|\eta_S\|_1 < \frac{1}{2}.$$

Now, since S_1^{n-1} is the boundary of B_1^n and including the interior of B_1^n in the feasible set does not change the solution, we obtain the desired result. \square

Remark 1. When the ASC is given by $\mathcal{T} = \mathcal{T}_s$ (see (3)), i.e. the ASC is the collection of index sets of cardinality less than or equal to s , the result of the double maximization in (4) is called the nullspace constant, and is NP-hard to compute [33] in general. For this special case, we denote the constant by

$$\text{nsc}(s, \Phi) := \max_{|S| \leq s} \max_{\eta \in \text{Null}(\Phi)} \|\eta_S\|_1. \quad (5)$$

B. Maximum ASC Associated with a Matrix

If \mathcal{T}_1 and \mathcal{T}_2 are ASCs with $\mathcal{T}_1, \mathcal{T}_2 \subseteq 2^{\mathcal{U}_n}$ and $\Phi \in \mathbb{R}^{m \times n}$ satisfies the GNUP w.r.t. \mathcal{T}_1 and \mathcal{T}_2 , then it follows from the definition of the GNUP that Φ also satisfies the GNUP w.r.t. $\mathcal{T}_1 \cup \mathcal{T}_2$. This observation implies that for any matrix Φ , there is a maximum ASC for which Φ satisfies the GNUP. Since the cardinality of \mathcal{U}_n is finite, the maximum ASC can be defined as the union of all ASCs for which Φ satisfies the GNUP.

Definition 4 (Maximum ASC (MASC)). *The union of every ASC in $2^{\mathcal{U}_n}$ for which Φ satisfies the GNUP is called the MASC associated with Φ , and is denoted by $\mathcal{T}_{\max}(\Phi)$.*

The next lemma is a consequence of Defn. 4.

Lemma 2. *Let $\Phi \in \mathbb{R}^{m \times n}$. The following statements hold:*

- (i) Φ satisfies the GNUP w.r.t. $\mathcal{T}_{\max}(\Phi)$.
- (ii) Every $\bar{x} \in \mathbb{R}^n$ with $\text{Supp}(\bar{x}) \in \mathcal{T}_{\max}(\Phi)$ is the unique solution to the optimization problem (2).

Proof. We first prove part (i). Let $\eta \in \text{Null}(\Phi) \setminus \{0\}$ and $S \in \mathcal{T}_{\max}(\Phi)$. Then, it holds from the definition of $\mathcal{T}_{\max}(\Phi)$ that there exists an ASC, say \mathcal{T} , that satisfies $S \in \mathcal{T} \subseteq \mathcal{T}_{\max}(\Phi)$ and that Φ satisfies the GNUP w.r.t. \mathcal{T} , which implies, since $S \in \mathcal{T}$, that $\|\eta_S\|_1 < \|\eta_{S^c}\|_1$. This completes the proof.

Part (ii) follows by setting $\mathcal{T} = \mathcal{T}_{\max}(\Phi)$ in Thm. 2, which is allowed because of Lem. 2(i). \square

Since $\mathcal{T} = \{\emptyset\}$ is always an ASC, it follows that the MASC always contains the empty set. The next example shows that for some matrices, this may be the only set in the MASC.

Example 1. Let $\Phi = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}$ so that the $\text{Null}(\Phi)$ is spanned by $\eta_0 := \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$ and $\eta_1 := \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^T$. Hence, any vector $\eta \in \text{Null}(\Phi)$ is of the form $[\alpha \quad \alpha + \beta \quad \beta]$, where $\{\alpha, \beta\} \subset \mathbb{R}$. Then, it may be shown that for each nonempty set $S \subseteq \mathcal{U}_3$, there exists a nonzero $\eta \in \text{Null}(\Phi)$ such that $\|\eta_S\|_1 \geq \|\eta_{S^c}\|_1$. This shows that if Φ satisfies the GNUP w.r.t. some ASC \mathcal{T} , then $\mathcal{T} = \{\emptyset\}$. It follows that $\mathcal{T}_{\max}(\Phi) = \{\emptyset\}$.

Lem. 2(ii) says that any $\bar{x} \in \mathbb{R}^n$ with $\text{Supp}(\bar{x}) \in \mathcal{T}_{\max}(\Phi)$ is the unique solution to the optimization problem (2). It is natural to ask whether the converse is true. That is, if \bar{x} is the unique solution to the optimization problem (2), then is it true that $\text{Supp}(\bar{x}) \in \mathcal{T}_{\max}(\Phi)$? This converse statement is not necessarily true, and we direct the reader to [32, Cor. 4.29 & Thm. 4.30] for additional details on why it may fail.

Although some well-known sparsity patterns form ASCs, it is not true that every useful sparsity pattern is an ASC (e.g., see [19]). The next result shows that, for any $\mathcal{T} \subseteq \mathcal{T}_{\max}(\Phi)$ (\mathcal{T} need not be an ASC), each vector \bar{x} whose support is in \mathcal{T} is the unique solution to (2). This highlights the fact that the MASC is a special ASC.

Theorem 3. Let $\mathcal{T} \subseteq 2^{\mathcal{U}_n}$ and $\Phi \in \mathbb{R}^{m \times n}$. Every $\bar{x} \in \mathbb{R}^n$ with $\text{Supp}(\bar{x}) \in \mathcal{T}$ is the unique solution to the optimization problem (2) if and only if $\mathcal{T} \subseteq \mathcal{T}_{\max}(\Phi)$.

Proof. We prove both directions of the implication in turn.

[\Leftarrow] Let $\mathcal{T} \subseteq \mathcal{T}_{\max}(\Phi)$ and \bar{x} be so that $\text{Supp}(\bar{x}) \in \mathcal{T} \subseteq \mathcal{T}_{\max}(\Phi)$. The result now follows from Lem. 2(ii).

[\Rightarrow] For a proof by contradiction, let it hold that every $\bar{x} \in \mathbb{R}^n$ with $\text{Supp}(\bar{x}) \in \mathcal{T}$ is the unique solution to the optimization problem (2), but yet $\mathcal{T} \not\subseteq \mathcal{T}_{\max}(\Phi)$. Then, there exists $S \in \mathcal{T}$ such that $S \notin \mathcal{T}_{\max}(\Phi)$. The power set of S , namely 2^S , is the smallest ASC that contains S . Since, $S \notin \mathcal{T}_{\max}(\Phi)$, it holds that $2^S \not\subseteq \mathcal{T}_{\max}(\Phi)$. This means that there exists a nonzero $\eta \in \text{Null}(\Phi)$ and $W \in 2^S$ satisfying $\|\eta_{W^c}\|_1 \leq \|\eta_W\|_1$, which implies that

$$\|\eta_{S^c}\|_1 \leq \|\eta_S\|_1. \quad (6)$$

Using $0 = \Phi\eta = \Phi(\text{Pr}_S(\eta) + \text{Pr}_{S^c}(\eta))$, it follows that $\Phi\text{Pr}_S(\eta) = \Phi(-\text{Pr}_{S^c}(\eta))$. With $\bar{\eta} = \text{Pr}_S(\eta)$ and $\hat{\eta} = -\text{Pr}_{S^c}(\eta)$, it holds using (6) that $\Phi\bar{\eta} = \Phi\hat{\eta}$ and $\|\hat{\eta}\|_1 \leq \|\bar{\eta}\|_1$. This contradicts the above uniqueness of solutions if $\text{Supp}(\bar{\eta}) = S$ since $S \in \mathcal{T}$. Thus, for the remainder of the proof, we only consider the case $\text{Supp}(\bar{\eta}) \subsetneq S$.

Let $\tilde{\eta} \in \mathbb{R}^n$ be any vector satisfying (i) $\text{Supp}(\tilde{\eta}) = S$, (ii) $\tilde{\eta}_k = \bar{\eta}_k$ if $k \in \text{Supp}(\bar{\eta})$, and (iii) $\tilde{\eta}_k$ is any arbitrary nonzero number for all $k \in S \setminus \text{Supp}(\bar{\eta})$. Therefore, it holds that $\|\tilde{\eta} + \bar{\eta}\|_1 = \|\tilde{\eta}\|_1 + \|\bar{\eta}\|_1$ and $\text{Supp}(\tilde{\eta} + \bar{\eta}) = S$. On the other hand, since $\text{Supp}(\tilde{\eta}) \cap \text{Supp}(\hat{\eta}) = \emptyset$, we have $\|\tilde{\eta} + \hat{\eta}\|_1 = \|\tilde{\eta}\|_1 + \|\hat{\eta}\|_1$ as well. Note that $\bar{x} := \tilde{\eta} + \bar{\eta} \neq \tilde{\eta} + \hat{\eta}$ because otherwise $\eta \equiv \bar{\eta} - \hat{\eta} = 0$, which would be a contradiction. Combining these observations shows that

$$\begin{aligned} \|\tilde{\eta} + \hat{\eta}\|_1 &= \|\tilde{\eta}\|_1 + \|\hat{\eta}\|_1 \\ &\leq \|\tilde{\eta}\|_1 + \|\bar{\eta}\|_1 \quad (\text{using (6)}) \\ &= \|\tilde{\eta} + \bar{\eta}\|_1. \end{aligned}$$

It now follows that $\Phi\bar{x} = \Phi(\tilde{\eta} + \bar{\eta}) = \Phi(\tilde{\eta} + \hat{\eta})$ with $\text{Supp}(\bar{x}) = S$, but yet $\|\tilde{\eta} + \hat{\eta}\|_1 \leq \|\bar{x}\|_1$, which contradicts the uniqueness of the solutions since $S \in \mathcal{T}$. \square

As a consequence of Thm. 3, we see that $\mathcal{T}_{\max}(\Phi)$ is the collection of all support sets S for which ℓ_1 -minimization is always successful. In particular, it can be characterized as

$$\mathcal{T}_{\max}(\Phi) = \{S \subseteq \mathcal{U}_n : \|\eta_S\|_1 < \|\eta_{S^c}\|_1, \forall \eta \in \text{Null}(\Phi) \setminus \{0\}\}.$$

Finally, we note that the MASC can be used to obtain a lower bound for the probability of exact recovery.

Proposition 1. Let $\mathcal{T} \subseteq 2^{\mathcal{U}_n}$ and $\Phi \in \mathbb{R}^{m \times n}$. Assume $\bar{x} \in \mathbb{R}^n$ is randomly chosen from some distribution. Let

$$p := \text{P}(\bar{x} \text{ is the unique solution to (2)} \mid \text{Supp}(\bar{x}) \in \mathcal{T})$$

and

$$q := \text{P}(\text{Supp}(\bar{x}) \in \mathcal{T}_{\max}(\Phi) \mid \text{Supp}(\bar{x}) \in \mathcal{T}),$$

Then, it follows that $p \geq q$.

Proof. Let $\text{Supp}(\bar{x}) \in \mathcal{T}$. If $\text{Supp}(\bar{x}) \in \mathcal{T}_{\max}(\Phi)$, then it follows from Lem. 2(ii) that \bar{x} is the unique solution to (2). It follows from this fact that $p \geq q$, which proves the result. \square

The significance of Prop. 1 is that when \mathcal{T} and $\mathcal{T}_{\max}(\Phi)$ are known, it may be possible to efficiently compute q even when p cannot be efficiently calculated. We provide examples later when we consider special matrices in § V-C and § VI-C.

IV. CHARACTERIZATIONS OF THE GNUP AND THE MASC

The definition of the GNUP involves a condition on *all* nonzero vectors in the nullspace. Since this definition is complicated to use in practice, we seek alternative characterizations that provide new insights and computational advantages, which translate over to the MASC. In this section we provide two such characterizations: the first is based on extreme points and the second is based on vectors of minimal support. An advantage of the characterization based on extreme points is that it is geometrically intuitive, whereas the characterization based on the vectors of minimal support leverages existing results in the literature better. This latter point will become clear when we discuss special classes of matrices in § V-C and § VI-C.

A. Characterization in Terms of Extreme Points

In this section we reformulate the GNUP and MASC in terms of the extreme points of $\text{Null}(\Phi) \cap \mathbb{B}_1^n$.

Lemma 3. *Let $\mathcal{T} \subseteq 2^{\mathcal{U}_n}$. A matrix $\Phi \in \mathbb{R}^{m \times n}$ satisfies the GNUP w.r.t. \mathcal{T} if and only if*

$$\max_{S \in \mathcal{T}} \max_{z \in \text{Ext}(\text{Null}(\Phi) \cap \mathbb{B}_1^n)} \|z_S\|_1 < \frac{1}{2}.$$

Proof. For each $S \in \mathcal{T}$, the problem $\max_{z \in \text{Null}(\Phi) \cap \mathbb{B}_1^n} \|z_S\|_1$ in (4) is the maximization of a convex function over the nonempty, compact, and convex set $\text{Null}(\Phi) \cap \mathbb{B}_1^n$. Hence, the maximum is attained at an extreme point [27]. Thus, it is enough to maximize only over $\text{Ext}(\text{Null}(\Phi) \cap \mathbb{B}_1^n)$, which with Lem. 1 gives the result. \square

Lem. 3 allows for a clearer description of the MASC associated with a matrix Φ in terms of $\text{Ext}(\text{Null}(\Phi) \cap \mathbb{B}_1^n)$.

Proposition 2. *For any matrix $\Phi \in \mathbb{R}^{m \times n}$, we have*

$$\mathcal{T}_{\max}(\Phi) = \{S \subseteq \mathcal{U}_n : \|z_S\|_1 < \frac{1}{2} \forall z \in \text{Ext}(\text{Null}(\Phi) \cap \mathbb{B}_1^n)\}.$$

Proof. We prove that each set is contained in the other.

$[\subseteq]$ It follows from Lem. 2(i) and Lem. 3 that $\|z_S\|_1 < 1/2$ for all $z \in \text{Ext}(\text{Null}(\Phi) \cap \mathbb{B}_1^n)$ and all $S \in \mathcal{T}_{\max}(\Phi)$, which establishes that this inclusion holds.

$[\supseteq]$ It is easy to see that the right-hand side is an ASC, and then Lem. 3 shows that Φ satisfies the GNUP w.r.t. the right-hand side. Therefore, by the definition of $\mathcal{T}_{\max}(\Phi)$, it must hold that the right-hand side is contained in $\mathcal{T}_{\max}(\Phi)$. \square

The extreme points of $\text{Null}(\Phi) \cap \mathbb{B}_1^n$ characterize the GNUP and MASC associated with a dictionary Φ . Although this characterization has a clear geometric interpretation, the computation of extreme points is non-trivial. Therefore, in the next section we provide an alternative mathematical characterization based on vectors of minimal support that can be more easily described, especially for certain classes of matrices.

B. Characterization in Terms of Vectors of Minimal Support

We now delve deeper into the properties of the extreme points of $\text{Null}(\Phi) \cap \mathbb{B}_1^n$. In particular, our aim is to associate them with *vectors of minimal support*,³ which we now define.

Definition 5 (Vectors of minimal support). *Let $\mathcal{V} \subseteq \mathbb{R}^n$ be a subspace and $x \in \mathcal{V}$ be a nonzero vector. We say that x has minimal support in \mathcal{V} if and only if there does not exist a nonzero $\bar{x} \in \mathcal{V}$ whose support is contained in, but not equal to, the support of x , i.e., $\text{Supp}(\bar{x}) \subsetneq \text{Supp}(x)$. We denote the collection of all such vectors in \mathcal{V} by $\text{MinSupp}(\mathcal{V})$.*

To illustrate the concept of vectors of minimal support, let us consider the following example.

Example 2. *Let $\mathcal{V} \subset \mathbb{R}^3$ be the subspace defined by*

$$x_0 + x_1 + x_2 = 0. \quad (7)$$

Let $x \in \text{MinSupp}(\mathcal{V})$. By definition, x is nonzero. Moreover, $\|x\|_0 > 1$ since otherwise (7) cannot be satisfied. On the other hand, $\|x\|_0 \neq 3$ because there exist 2-sparse vectors in \mathcal{V} whose supports are obviously contained in $\{0, 1, 2\}$. Therefore, $\text{MinSupp}(\mathcal{V})$ contains exactly 2-sparse vectors, i.e.,

$$\text{MinSupp}(\mathcal{V}) = \{x \in \mathbb{R}^3 : \|x\|_0 = 2 \text{ and } x_0 + x_1 + x_2 = 0\}.$$

Our next goal is to establish that the sets $\text{Ext}(\mathcal{V} \cap \mathbb{B}_1^n)$ and $\text{MinSupp}(\mathcal{V}) \cap \mathbb{S}_1^{n-1}$ are equal. To prove this result, we must first motivate and prove an auxiliary result. Specifically, let

$$\Delta_e := \{x \in \mathbb{S}_1^{n-1} : \langle e, x \rangle = 1\} \quad (8)$$

be a (closed and convex) simplex defined by a nonzero *sign vector* $e \in \{0, 1, -1\}^n \subset \mathbb{R}^n$. Then, \mathbb{S}_1^{n-1} can be written as a disjoint union of the relative interiors of such simplexes, i.e.,

$$\mathbb{S}_1^{n-1} \equiv \bigcup_e \text{rinte}(\Delta_e). \quad (9)$$

We can use (9) to give a characterization of $\text{Ext}(\mathcal{V} \cap \mathbb{B}_1^n)$.

Lemma 4. *Let $\mathcal{V} \subseteq \mathbb{R}^n$ be a subspace of dimension $d \geq 1$, and define its codimension as $r := n - d$. The following hold.*

- (i) *$z \in \text{Ext}(\mathcal{V} \cap \mathbb{B}_1^n)$ if and only if $\{z\} = \mathcal{V} \cap \text{rinte}(\Delta_e)$ for some nonzero sign vector e .*
- (ii) *If $z \in \text{Ext}(\mathcal{V} \cap \mathbb{B}_1^n)$, then z is $(r + 1)$ -sparse.*

Proof. We first prove part (i). For the only if direction, let $z \in \text{Ext}(\mathcal{V} \cap \mathbb{B}_1^n)$, which means that $z \in \mathcal{V} \cap \mathbb{S}_1^n$ because $d \geq 1$. Since it follows from (9) that

$$\mathcal{V} \cap \mathbb{S}_1^{n-1} = \bigcup_e (\mathcal{V} \cap \text{rinte}(\Delta_e)), \quad (10)$$

we know that there exists a unique nonzero sign vector e such that $z \in \mathcal{V} \cap \text{rinte}(\Delta_e)$. If $\dim(\mathcal{V} \cap \text{rinte}(\Delta_e)) > 0$, then any point in $\mathcal{V} \cap \text{rinte}(\Delta_e)$ can be written as a nontrivial convex combination of two distinct points in $\mathcal{V} \cap \text{rinte}(\Delta_e)$, which is a contradiction. Therefore, it must be that $\dim(\mathcal{V} \cap \text{rinte}(\Delta_e)) = 0$ and that $\{z\} = \mathcal{V} \cap \text{rinte}(\Delta_e)$, as claimed.

For the converse direction, let z be such that $\{z\} = \mathcal{V} \cap \text{rinte}(\Delta_e)$ for some nonzero sign vector e . Since z is the

³In [27], vectors of minimal support are called *elementary vectors*.

unique maximizer of the linear functional $\langle e, \cdot \rangle$ on the compact polyhedral set $\mathcal{V} \cap B_1^n$, we must conclude that $z \in \text{Ext}(\mathcal{V} \cap B_1^n)$.

We now prove part (ii). Let $z \in \text{Ext}(\mathcal{V} \cap B_1^n)$ so that from part (i) we have $\{z\} = \mathcal{V} \cap \text{rinte}(\Delta_e)$ for some nonzero sign vector e . Since $z \in \text{rinte}(\Delta_e)$, we necessarily have

$$\|z\|_0 = \|e\|_0 = \dim(\Delta_e) + 1. \quad (11)$$

Now, to reach a contradiction, suppose that $\dim(\Delta_e) > r$. Combining this with the fact that $\mathcal{V} \cap \text{rinte}(\Delta_e) \neq \emptyset$, shows necessarily that $\dim(\mathcal{V} \cap \text{rinte}(\Delta_e)) > 0$, which is a contradiction. We have shown that $\dim(\Delta_e) \leq r$, which combined with (11) gives the desired result. \square

We now prove the equivalence between $\text{Ext}(\mathcal{V} \cap B_1^n)$ and the intersection of the minimal support vectors of \mathcal{V} and S_1^{n-1} . Fig. 1 illustrates this equivalence for a subspace \mathcal{V} of \mathbb{R}^3 .

Proposition 3. *Let $\mathcal{V} \subseteq \mathbb{R}^n$ be a nontrivial subspace. Then, $\text{MinSupp}(\mathcal{V}) \cap S_1^{n-1} = \text{Ext}(\mathcal{V} \cap B_1^n)$.*

Proof. We prove that each set is contained in the other. $[\subseteq]$ Let $z \in \text{MinSupp}(\mathcal{V}) \cap S_1^{n-1}$. We know from (10) that there exists a unique nonzero sign vector e such that $z \in \mathcal{V} \cap \text{rinte}(\Delta_e)$. If $\dim(\mathcal{V} \cap \text{rinte}(\Delta_e)) = 0$, then it follows from Lem. 4(i) that z is an extreme point of $\mathcal{V} \cap B_1^n$, and we are done. Therefore, in the remainder of the proof, we show that $\dim(\mathcal{V} \cap \text{rinte}(\Delta_e)) > 0$ cannot hold.

For a proof by contradiction, suppose that $\dim(\mathcal{V} \cap \text{rinte}(\Delta_e)) > 0$. Define C as the closure of $\mathcal{V} \cap \text{rinte}(\Delta_e)$ in $\mathcal{V} \cap \Delta_e$ so that the boundary satisfies $\partial C \subseteq \mathcal{V} \cap S_1^{n-1}$. However, for any point $\bar{z} \in \partial C$ we have $\text{Supp}(\bar{z}) \subsetneq \text{Supp}(z)$, which contradicts the fact that z has minimal support.

$[\supseteq]$ Let $z \in \text{Ext}(\mathcal{V} \cap B_1^n)$, which means that $z \in \mathcal{V} \cap S_1^{n-1}$ because \mathcal{V} is a nontrivial subspace by assumption. For a proof by contradiction, suppose that $z \notin \text{MinSupp}(\mathcal{V}) \cap S_1^{n-1}$, which combined with $z \in \mathcal{V} \cap S_1^{n-1}$ implies that $z \notin \text{MinSupp}(\mathcal{V})$, i.e., that there exists a nonzero $\bar{z} \in \mathcal{V} \cap S_1^{n-1}$ such that $\text{Supp}(\bar{z}) \subsetneq \text{Supp}(z)$. Next, we define $s := \|z\|_0$ and $\mathcal{W} \subseteq \mathbb{R}^n$ as the subspace

$$\mathcal{W} := \{x \in \mathbb{R}^n : x_k = 0 \text{ for all } k \notin \text{Supp}(z)\},$$

and note that $\{z, \bar{z}\} \subseteq \mathcal{W} \cap \mathcal{V} \cap S_1^{n-1}$ and that \mathcal{W} can trivially be associated with \mathbb{R}^s . Under this association, $\mathcal{W} \cap B_1^n$ corresponds to $B_1^s \subset \mathbb{R}^s$, and the subspace $\mathcal{W} \cap \mathcal{V}$ corresponds to a subspace $\bar{\mathcal{V}} \subset \mathbb{R}^s$. Under this correspondence, since $z \in \text{Ext}(\mathcal{V} \cap B_1^n)$ it follows that $z \in \text{Ext}(\bar{\mathcal{V}} \cap B_1^s)$. Since z is s -sparse, it follows from Lem. 4 that $s \leq (s - \dim(\bar{\mathcal{V}})) + 1$, which shows that $\dim(\bar{\mathcal{V}}) \leq 1$. Moreover, since $\bar{\mathcal{V}}$ is nonempty, we necessarily have $\dim(\bar{\mathcal{V}}) = 1$ so that $\text{Supp}(z) = \text{Supp}(\bar{z})$, which is a contradiction. \square

Prop. 3 can be used to characterize the GNUP and MASC associated with a matrix in terms of vectors of minimal support, as stated in the next lemma.

Lemma 5. *Let $\mathcal{T} \subseteq 2^{U_n}$ and $\Phi \in \mathbb{R}^{m \times n}$ have a nontrivial nullspace. Then, Φ satisfies the GNUP w.r.t. \mathcal{T} if and only if*

$$\max_{S \in \mathcal{T}} \max_{z \in \text{MinSupp}(\text{Null}(\Phi)) \cap S_1^{n-1}} \|z_S\|_1 < \frac{1}{2}.$$

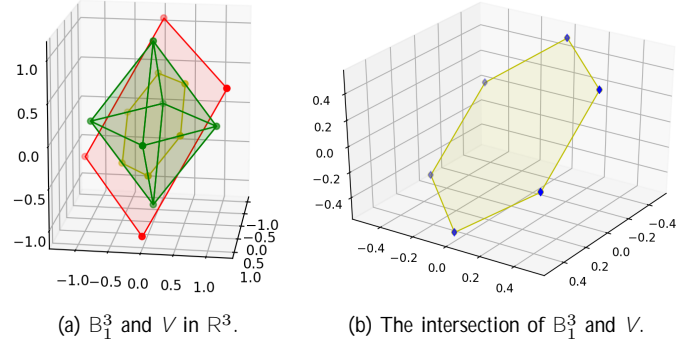


Fig. 1. When a subspace $\mathcal{V} \subseteq \mathbb{R}^n$ intersects the ℓ_1 -ball B_1^n , the sets $\text{Ext}(\mathcal{V} \cap B_1^n)$ and $\text{MinSupp}(\mathcal{V}) \cap S_1^{n-1}$ coincide.

Proof. The result follows from Lem. 3 and Prop. 3 using the identification $\mathcal{V} = \text{Null}(\Phi)$ in Prop. 3. \square

Proposition 4. *If $\Phi \in \mathbb{R}^{m \times n}$ has a nontrivial nullspace, then*

$$\begin{aligned} \mathcal{T}_{\max}(\Phi) \\ = \{S \subseteq U_n : \|z_S\|_1 < \frac{1}{2} \forall z \in \text{MinSupp}(\text{Null}(\Phi)) \cap S_1^{n-1}\}. \end{aligned}$$

Proof. The result follows from Prop. 2 and Prop. 3 using the identification $\mathcal{V} = \text{Null}(\Phi)$ in Prop. 3. \square

We will see in the next two sections that for certain popular dictionaries Φ , the set $\text{MinSupp}(\text{Null}(\Phi))$ has a rather simple characterization. Hence, Prop. 4 will provide an efficient way to characterize $\mathcal{T}_{\max}(\Phi)$ for these dictionaries.

V. APPLICATION TO GRAPH INCIDENCE MATRICES

The first application we consider is the ℓ_1 -minimization problem with a dictionary that takes the form of an incidence matrix A associated with a simple connected graph \mathcal{G} , namely

$$\min_{A\bar{x}=Ax} \|x\|_1. \quad (12)$$

We want to understand the MASC associated with A .

A. Review of Graph Theory

We begin with a review of some concepts from graph theory. (See [34] for additional details.) 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 $\mathcal{G} = \mathcal{G}(V, E)$. Hence, each e_k is represented by an ordered pair of vertices (v_{k_1}, v_{k_2}) . 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. Associated with a directed graph $\mathcal{G} = \mathcal{G}(V, E)$, we define the incidence matrix $A = A(\mathcal{G}) \in \mathbb{R}^{m \times n}$ as

$$a_{ij} = \begin{cases} -1, & \text{if } v_i \text{ is the initial vertex of edge } e_j, \\ 1, & \text{if } v_i \text{ is the terminal vertex of edge } e_j, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

The subspace $\text{Null}(A)$ is called the *flow space* of \mathcal{G} . A *simple cycle* is a closed walk with no repetitions of vertices other than the starting and ending vertex, such as

$\mathcal{C} = (u_1, \dots, u_r, u_{r+1} = u_1)$. The length of the shortest simple cycle in \mathcal{G} is called the *girth* of \mathcal{G} . A simple cycle $\mathcal{C} = (u_1, \dots, u_r, u_{r+1} = u_1)$ can be associated with a vector $w(\mathcal{C}) \in \{-1, 0, 1\}^{|E|}$, where each coordinate is defined as

$$[w(\mathcal{C})]_j = \begin{cases} 1, & \text{if } e_j = (u_i, u_{i+1}) \text{ for some } i \in \{1, \dots, r\}, \\ -1, & \text{if } e_j = (u_{i+1}, u_i) \text{ for some } i \in \{1, \dots, r\}, \\ 0, & \text{otherwise.} \end{cases}$$

The vector $w(\mathcal{C})$ is called the *signed characteristic vector* of the oriented cycle \mathcal{C} [34]. These vectors play a central role.

Remark 2. *The nonzero elements of the flow space with minimal support are nonzero scalar multiples of the signed characteristic vectors of the simple cycles of \mathcal{G} [34, Thm. 14.2.2].*

The signed characteristic vectors of the simple cycles of a simple graph \mathcal{G} span its flow space [34, Cor. 14.2.3]. We denote the set of normalized signed characteristic vectors of the simple cycles of a simple directed graph \mathcal{G} by

$$W_1(\mathcal{G}) := \left\{ \frac{w(\mathcal{C})}{\|w(\mathcal{C})\|_1} : \mathcal{C} \text{ is a simple cycle of } \mathcal{G} \right\}. \quad (14)$$

Rem. 2 states, using our notation, that

$$W_1(\mathcal{G}) = \text{MinSupp}(\text{Null}(A)) \cap S_1^{n-1}, \quad (15)$$

where $A \in \mathbb{R}^{m \times n}$ is the incidence matrix of the graph \mathcal{G} .

B. Characterizing the MASC for Graph Incidence Matrices

We apply the results of Section IV to derive necessary and sufficient conditions for exact recovery. The next corollary provides the link between $\text{Ext}(\text{Null}(A) \cap B_1^n)$ and the simple cycles of the graph \mathcal{G} , thus allowing a clear characterization $\mathcal{T}_{\max}(A)$. This is the main result of this section.

Corollary 1. *Let \mathcal{G} be a simple graph with incidence matrix $A \in \mathbb{R}^{m \times n}$ whose nullspace is nontrivial. The following hold:*

- (i) $\text{Ext}(\text{Null}(A) \cap B_1^n) = W_1(\mathcal{G})$.
- (ii) *The MASC is given by*

$$\mathcal{T}_{\max}(A) = \left\{ S \subseteq \mathcal{U}_n : \frac{|S \cap \text{Supp}(z)|}{\|z\|_0} < \frac{1}{2} \forall z \in W_1(\mathcal{G}) \right\}.$$

Proof. Part (i) follows from Prop. 3 and (15). To prove part (ii), observe for any index set $S \subseteq \mathcal{U}_n$ and $z \in W_1(\mathcal{G})$ that

$$\|z_S\|_1 = \frac{|S \cap \text{Supp}(z)|}{\|z\|_0}. \quad (16)$$

Combining this with Prop. 4 and (15) gives the result. \square

Combining Cor. 1(ii) with Thm. 3 tells us that the following holds when the dictionary is the incidence matrix of a simple graph: every signal whose support is contained in an index set $S \subseteq \mathcal{U}_n$ can be recovered via ℓ_1 -minimization if and only if for each simple cycle, the size of the intersection of S with the indices of the edges appearing in the simple cycle is strictly less than half of the length of the simple cycle. Next, as a concrete example, we calculate MASC for the graph in Fig. 2

Example 3. *Consider the simple directed graph $\widehat{\mathcal{G}}$ and its incidence matrix, $A(\widehat{\mathcal{G}})$, as shown in Fig. 2. We denote the*

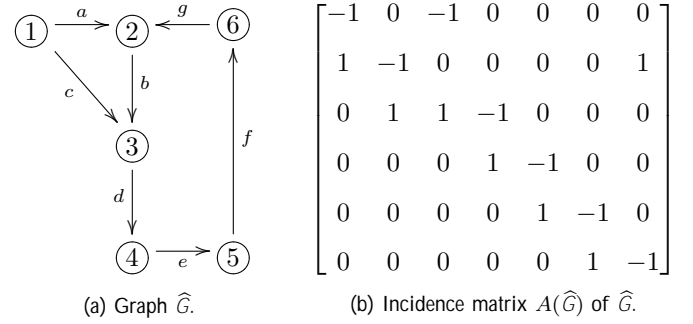


Fig. 2. The graph $\widehat{\mathcal{G}}$ of Exm. 3 and its incidence matrix $A(\widehat{\mathcal{G}})$.

edges $E(\widehat{\mathcal{G}})$ of $\widehat{\mathcal{G}}$ with small letters, whereas the vertices of $\widehat{\mathcal{G}}$ are denoted by numbers. The columns of the incidence matrix $A(\widehat{\mathcal{G}})$ are indexed by the edges of $\widehat{\mathcal{G}}$ in alphabetical order. That is, the first column corresponds to the edge “a”, the second column corresponds to edge “b”, etc. The rows of $A(\widehat{\mathcal{G}})$ are indexed by the vertices of $\widehat{\mathcal{G}}$ in their natural ordering. Note that $\widehat{\mathcal{G}}$ has 3 simple cycles, and—up to a sign change—they correspond to the following vectors:

$$\begin{aligned} v_1 &:= [1 \ 1 \ -1 \ 0 \ 0 \ 0 \ 0] \\ v_2 &:= [0 \ 1 \ 0 \ 1 \ 1 \ 1 \ 1] \\ v_3 &:= [-1 \ 0 \ 1 \ 1 \ 1 \ 1 \ 1]. \end{aligned}$$

As a consequence of Cor. 1(ii), we know that any signal \bar{x} with $\text{Supp}(\bar{x}) = \{a, e\}$ can be recovered via (12), because $\{a, e\}$ contains strictly less than half of the edges in any simple cycle of $\widehat{\mathcal{G}}$. However, if we have $\text{Supp}(\bar{x}) = \{a, b\}$, then (12) may not be able to recover \bar{x} , because the simple cycle that corresponds to v_1 consists of the edges $\{a, b, c\}$ and this $\text{Supp}(\bar{x})$ covers 2/3 of this simple cycle. In fact, Cor. 1(ii) tells us that there are signals \bar{x} with $\text{Supp}(\bar{x}) = \{a, b\}$, which cannot be recovered via (12). Therefore, when the measurement matrix is $A(\widehat{\mathcal{G}})$, the recovery performance of (12) is not uniform on 2-sparse signals. Nonetheless, any 1-sparse signal can always be recovered. A slightly more detailed analysis reveals that there is always some s -sparse signal that cannot be recovered, whenever $s \geq 3$. Hence, MASC of $A(\widehat{\mathcal{G}})$ is

$$\mathcal{T}_{\max}(A(\widehat{\mathcal{G}})) = \{S \in 2^{E(\widehat{\mathcal{G}})} : |S| \leq 1 \text{ or } (|S|=2 \text{ and } |S \cap \{a, b, c\}| \leq 1)\}.$$

Remark 3. *Cor. 1(ii) is especially relevant when studying shortest paths in unweighted graphs. Suppose that we are given a simple, undirected, unweighted graph \mathcal{G} , for which the shortest path between any pair of vertices is unique. Such graphs are called geodetic graphs in the literature [35]. A tree is an example of a geodetic graph. Moreover, a geodetic graph cannot contain a simple cycle of even length. The crucial observation from our view point is that in a geodetic graph, shortest paths can only contain strictly less than half of the edges in any simple cycle of the graph. Hence, if we orient our graph \mathcal{G} by picking directions for the edges, and denote the oriented graph by $\widehat{\mathcal{G}}$, then by Cor. 1(ii), the support of the shortest path between any two vertices of $\widehat{\mathcal{G}}$ is in $\mathcal{T}_{\max}(A(\widehat{\mathcal{G}}))$. Thus, it can be recovered via the ℓ_1 -minimization problem.*

Cor. 1(ii) can be specialized to derive necessary and sufficient conditions for the recovery of s -sparse signals.

Theorem 4. *If $A \in \mathbb{R}^{m \times n}$ is the incidence matrix of a simple graph \mathcal{G} with girth g , then the nullspace constant (5) satisfies*

$$\text{nsc}(s, A) = \min \{1, s/g\}.$$

Hence, every s -sparse vector $\bar{x} \in \mathbb{R}^n$ is the unique solution to the optimization problem (12) if and only if $s < g/2$.

Proof. It follows from the definition of the nullspace constant in (5), Cor. 1, and (16) that

$$\text{nsc}(s, A) = \max_{|S| \leq s} \max_{z \in W_1(\mathcal{G})} \frac{|S \cap \text{Supp}(z)|}{\|z\|_0}. \quad (17)$$

Now, consider any $z \in W_1(\mathcal{G})$. It follows that there exists a simple cycle \mathcal{C} of \mathcal{G} with length $g + \ell(z)$ for some $\ell(z) \geq 0$ such that $z = w(\mathcal{C})/\|w(\mathcal{C})\|_1$, where we have used the fact that the girth of \mathcal{G} is g ; note that $\|z\|_0 = g + \ell(z)$. For this given z , the right-hand side of (17) is maximized by any $S(z)$ satisfying that $|S(z)| \leq s$ and it contains as many indices from the support of z as possible. This choice and (17) show that

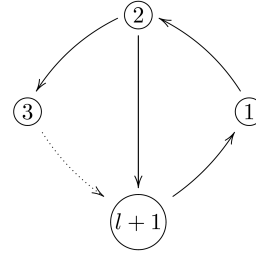
$$\begin{aligned} \text{nsc}(s, A) &= \max_z \max_{W_1(\mathcal{G})} \frac{|S(z) \cap \text{Supp}(z)|}{\|z\|_0} \\ &= \max_z \max_{W_1(\mathcal{G})} \frac{\min\{g + \ell(z), s\}}{g + \ell(z)} \\ &= \max_z \min \left\{ 1, \frac{s}{g + \ell(z)} \right\}, \end{aligned} \quad (18)$$

from which it is now clear that the maximizer is obtained by any z satisfying $\ell(z) = 0$, i.e., by $z = w(\mathcal{C})/\|w(\mathcal{C})\|_1$ for any simple cycle \mathcal{C} whose length is equal to g . Combining this with (18) proves that $\text{nsc}(s, A) = \min\{1, s/g\}$, as claimed.

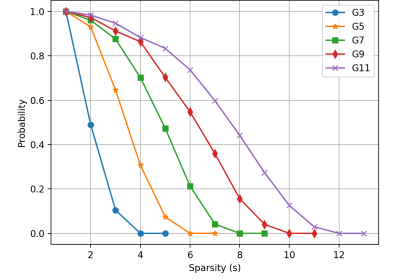
To prove the second part, we first observe using the first part that $\text{nsc}(s, A) = \min\{1, s/g\} < 1/2$ if and only if $s < g/2$. The result follows by combining this fact with Lem. 1, and recalling that $\mathcal{T} = \mathcal{T}_s$ in the current setting (see (3)). \square

Thm. 4 reveals that when the dictionary is a graph incidence matrix, the maximum sparsity level that can always be recovered via ℓ_1 -minimization is determined by the girth of the graph. Therefore, for graphs with small girth, we expect recovery experiments to fail for certain signals of relatively small sparsity. On the other hand, although one can not recover all s -sparse signals for a given s , due to Cor. 1, the successful recovery rate can still be relatively high. Because, even though the girth is small, if the graph consists mainly of large simple cycles, $\mathcal{T}_{\max}(A)$ can still contain a large number of index sets of cardinality s , for which ℓ_1 -recovery will always be successful. We will illustrate this intuition with experiments in § V-C.

Remark 4. *Thm. 4 has an important consequence regarding the computational complexity of the calculation of nullspace constant for incidence matrices. We already mentioned in Rem. 1 that it is NP-hard to calculate the nullspace constant in general. However, there are algorithms that can calculate the girth of a graph exactly in $\mathcal{O}(mn)$. Therefore, Thm. 4 reveals that the nullspace constant can be calculated—hence the NUP*



(a) Graph \mathcal{G}_l for $l = 3$.



(b) Sparse recovery results for \mathcal{G}_l .

Fig. 3. The graphs $\{\mathcal{G}_l\}$ used in §V-C1 and the sparse recovery results for \mathcal{G}_l with $\ell \in \{3, 5, 7, 9, 11\}$ using (19).

can be verified—for graph incidence matrices in polynomial time.

C. Experiments on Graph Incidence Matrices

All experiments are implemented in Python 3.6. In these experiments, we often run an ℓ_1 -recovery experiment that requires sampling a random vector \bar{x} of a given sparsity level s . For this purpose we use the random function from Python's `sci.py.sparse` class. The support of the sparse vector \bar{x} is chosen uniformly at random, and its nonzero values are chosen from the standard normal distribution. The resulting vector is subsequently normalized to have unit ℓ_2 norm, thus forming the final \bar{x} . For this \bar{x} , a vector \hat{x} is then obtained as a solution to the optimization problem (2). The attempted ℓ_1 -recovery is deemed successful if $\|\hat{x} - \bar{x}\|_2 \leq 10^{-6}$. It is well-known that problem (2) can be cast as a linear program, which we solve using the `linprog` function in the `sci.py.optimize` class.

In this section, we consider numerical experiments that are based on two different families of graph incidence matrices.

1) *Graphs from Fig. 3a:* Consider the family of graphs $\{\mathcal{G}_l\}$ depicted in Fig. 3a. Graph \mathcal{G}_l consists of a simple cycle of length three and a simple cycle of length $l + 1$ for $l \in \{3, 5, 7, 9, 11\}$. For each of these graphs, we tested the ℓ_1 -sparsity recovery capabilities by solving the problem

$$\min_x \|x\|_1 \text{ subject to } A(\mathcal{G}_l)\bar{x} = A(\mathcal{G}_l)x, \quad (19)$$

where $x \in \mathbb{R}^{l+2}$ and $A(\mathcal{G}_l)$ denotes the incidence matrix of the graph \mathcal{G}_l .

For each pair of sparsity level s and incidence matrix $A(\mathcal{G}_l)$, this experiment was repeated 2000 times for different choices of \bar{x} , and the ratio of the successful trials is reported in Fig. 3b.

Since the girth is 3 for each \mathcal{G}_l considered in this experiment, it follows from Thm. 4 that only 1-sparse vectors can be recovered exactly, which is observed in Fig. 3b. Moreover, due to Cor. 1(ii), we expect to have better exact recovery performance for graphs with larger simple cycles, which in our case means larger values of l . This increase in recovery probability can be observed in Fig. 3b for sparsity levels $s > 1$.

2) *Erdős-Rényi Graphs:* The previous experiment showed that the success of ℓ_1 -recovery for incidence matrices of deterministic graphs relies on the topology of the graph. The simple cycles of the graph completely determine the recovery capability for signals with a given support. Now, we investigate

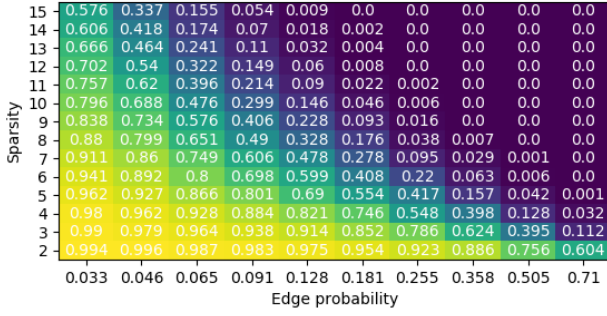


Fig. 4. Exact recovery probability of ℓ_1 -minimization via (19) when the dictionary is the incidence matrix of an Erdős-Rényi graph with 100 vertices. The horizontal axis shows the edge probability p and the vertical axis is the sparsity of the vector \bar{x} that is attempted to be recovered. The critical edge probability value as described in Section V-C2 is $p_{\text{crit}} = 0.046$.

the exact recovery probability for incidence matrices arising from randomly generated Erdős-Rényi graphs.

Consider Erdős-Rényi graphs on 100 vertices. Any two vertices have an edge linking them with probability $p \in (0, 1)$. For a fixed p , we sample 20 such graphs $\{\mathcal{G}_l\}_{l=1}^{19}$, calculate their incidence matrices $\{A(\mathcal{G}_l)\}_{l=1}^{19}$ and consider the ℓ_1 -recovery problem (19).

For a given sparsity level s and incidence matrix $A(\mathcal{G}_l)$, we sample 100 vectors \bar{x} and determine whether ℓ_1 sparse recovery using (19) is successful for each \bar{x} , as described in the first paragraph of this section. We record how many times successful recovery occurs.

We repeat this experiment for different values of p . A critical value for p is $p_{\text{crit}} := \frac{\ln(100)}{100} \simeq 0.046$. When $p < p_{\text{crit}}$ the random graph is almost surely disconnected, and when $p > p_{\text{crit}}$ the random graph is almost surely connected. To highlight the influence of p_{crit} , we select $p \in \{p_{\text{crit}}^{k/9} : 1 \leq k \leq 10\}$.

The result of this experiment is shown in Fig. 4. For a given p , the expected number of edges in an Erdős-Rényi graph with 100 vertices is $\binom{100}{2}p = 4950p$, and the expected number of cycles of length three is $\binom{100}{3}p^3 = 161700p^3$. For instance, for $p = p_{\text{crit}}$, the expected number of edges is approximately 228 and the expected number of triangles is approximately 15, which explains the high successful recovery rate observed in Fig. 4 for $p = p_{\text{crit}}$. On the other hand, when $p = p_{\text{crit}}^{1/9} \simeq 0.71$, the expected number of triangles is approximately 57960, which is much larger than the expected number of edges (approximately 3516). This explains the extremely poor recovery rates observed in Fig. 4.

VI. APPLICATION TO PARTIAL DFT MATRICES

The second application we present comes from the Discrete Fourier Transform (DFT) of dimension n , which we denote by \mathcal{F}_n . We consider the partial DFT matrices, each denoted by \mathcal{F}_Ω for some $\Omega \subseteq \mathcal{U}_n$, which are the submatrices of the DFT matrix consisting of the rows indexed by Ω with $m := |\Omega| \geq 1$. Our goal is to understand the exact recovery capabilities of

$$\min \|x\|_1 \text{ subject to } \mathcal{F}_\Omega \bar{x} = \mathcal{F}_\Omega x \text{ and } x, \bar{x} \in \mathbb{R}^n. \quad (20)$$

A. Notation and Preliminaries for DFT matrices

Before we present the second application in detail, we would like to make a few remarks about the notation. Up until this section we only dealt with real matrices and real vector spaces. Here we generalize some of the previous concepts to the complex domain. For a complex vector $c \in \mathbb{C}^n$, $\|c\|_0$ denotes the number of nonzero entries of the vector c . If $\Phi \in \mathbb{C}^{m \times n}$, then $\text{Null}(\Phi)$ denotes the complex nullspace. That is, $\text{Null}(\Phi) = \{z \in \mathbb{C}^n : \Phi z = 0\}$. In this subsection, when we talk about the dimension of a subspace $\mathcal{V} \subseteq \mathbb{C}^n$, we mean its dimension as a complex subspace. For any $\Omega, \Gamma \subseteq \mathcal{U}_n$ and any matrix $\Phi \in \mathbb{C}^{m \times n}$, we use $\Phi_{\Omega, \Gamma}$ to denote the submatrix of Φ consisting of the rows and columns from Φ whose indices are in Ω and Γ , respectively. For a complex number $z \in \mathbb{C}$, we let $|z|$ denote the norm of z .

For a positive integer n and $\xi := e^{-\frac{2\pi}{n}i}$, the DFT matrix $\mathcal{F}_n \in \mathbb{C}^{n \times n}$ is defined entrywise as

$$[\mathcal{F}_n]_{k,l} := \frac{1}{n}(\xi^{kl}) \text{ for } 0 \leq k, l \leq n-1. \quad (21)$$

The following remark gives useful properties of the DFT matrix in the case that n is a prime number.

Remark 5. Tao [36] shows that when n is a prime number, any $k \times k$ minor of \mathcal{F}_n is nonzero [36, Lem. 1.3] for all $1 \leq k \leq n$. Also, as a consequence of [36, Thm. 1.1], any sparse polynomial $\sum_{j=0}^k c_j z^{n_j}$ with $k+1$ nonzero coefficients and $0 \leq n_0 < \dots < n_k < n$, can have at most k of the n^{th} roots of unity as zeros, i.e., such a polynomial cannot vanish at more than k of the n^{th} roots of unity. For an analogous result where n need not be a prime number, we refer to [37].

This remark has the following immediate consequences.

Lemma 6. Let n be a prime number and $\Omega, \Gamma \subseteq \mathcal{U}_n$ be nonempty index sets satisfying $|\Gamma| = |\Omega| + 1$.

- (i) All nonzero $c \in \text{Null}(\mathcal{F}_{\Omega, \Gamma})$ satisfy $\|c\|_0 = |\Gamma|$.
- (ii) It holds that $\dim(\text{Null}(\mathcal{F}_{\Omega, \Gamma})) = 1$. Moreover, the nonzero vector $\nu \in \mathbb{C}^{|\Gamma|}$ defined entrywise as

$$\nu_t := (-1)^t \det(\mathcal{F}_{\Omega, \Gamma \setminus \{\Gamma_t\}}) \text{ for all } t \in \mathcal{U}_{|\Gamma|}$$

spans $\text{Null}(\mathcal{F}_{\Omega, \Gamma})$.

Proof. To prove part (i) let $c \in \text{Null}(\mathcal{F}_{\Omega, \Gamma}) \setminus \{0\}$. If $\|c\|_0 < |\Gamma|$, then $\sum_{k \in \Gamma} c_k z^k$ is a polynomial with $\|c\|_0 \leq |\Omega|$ nonzero coefficients that vanishes at $|\Omega|$ roots of unity. This contradicts Tao's result in Rem. 5, so that $\|c\|_0 = |\Gamma|$ holds.

To prove the first part of (ii), we use an argument by contradiction. Suppose that $\dim(\text{Null}(\mathcal{F}_{\Omega, \Gamma})) > 1$, in which case there exists two linearly independent vectors $\{\tilde{c}, \hat{c}\} \subset \text{Null}(\mathcal{F}_{\Omega, \Gamma})$, which from part (i) must satisfy $\|\tilde{c}\|_0 = \|\hat{c}\|_0 = |\Gamma|$. Next, define $\tilde{c} := (1/\tilde{c}_1)\tilde{c} - (1/\hat{c}_1)\hat{c}$ so that $\tilde{c} \in \text{Null}(\mathcal{F}_{\Omega, \Gamma})$, $\tilde{c}_1 = 0$, and $\tilde{c} \neq 0$, which contradicts part (i).

To prove the second part of (ii), we note that for any $l \in \Omega$, the determinant of the $|\Gamma| \times |\Gamma|$ matrix $[\mathcal{F}_{\Omega, \Gamma}^T \mathcal{F}_{\Omega, \Gamma}^T]^T$ is zero because it has a repeated row. However, if we expand the determinant w.r.t. the first row we see that

$$0 = \sum_{t=0}^{|\Gamma|-1} (-1)^t \mathcal{F}_{\{l\}, \{\Gamma_t\}} \det(\mathcal{F}_{\Omega, \Gamma \setminus \{\Gamma_t\}}).$$

Since this holds for all $l \in \Omega$, we get $\nu \in \text{Null}(\mathcal{F}_{\Omega, \Gamma})$. Moreover, it follows from the definition of ν that every entry in ν is nonzero because all square minors of \mathcal{F}_n are nonzero (see Rem. 5). This completes the proof. \square

B. Characterization of the MASC for Partial DFT Matrices

We will characterize the exact recovery capabilities of (20) when n is a prime number. This is achieved by characterizing $\mathcal{T}_{\max}(\mathcal{F}_{\Omega})$. By definition of the MASC, $\mathcal{T}_{\max}(\mathcal{F}_{\Omega})$ is still a collection of supports of certain real signals, not complex ones. In fact, an index set $S \in \mathcal{T}_{\max}(\mathcal{F}_{\Omega})$ if and only if $\|z_S\|_1 < \frac{1}{2}$ for all $z \in \text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n) \cap S_1^{n-1}$. Therefore, our intermediate goal is to provide a description of the vectors of minimal support in $\text{Null}(\mathcal{F}_{\Omega})$, i.e. $\text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n)$. First note that if we define $\bar{\Omega} = \Omega \cup \{n - k : k \in \Omega \setminus \{0\}\}$, then $\text{Null}(\mathcal{F}_{\Omega}) \cap B_1^n = \text{Null}(\mathcal{F}_{\bar{\Omega}}) \cap B_1^n$, which holds because if a polynomial with real coefficients vanishes at an n^{th} root of unity ξ^k , then it also vanishes at its complex conjugate ξ^{n-k} . Thus, for the purposes of the ℓ_1 -minimization problem (20), without loss of generality, we can assume that

$$k \in \Omega \implies n - k \in \Omega \text{ for } k = 1, \dots, n - 1 \quad (22)$$

since the unknown signal is assumed to be real.

Theorem 5. *Let n be a prime number and $\Omega \subseteq \mathcal{U}_n$ be nonempty index set satisfying (22). Then the following hold:*

- (i) *If $\Gamma \subseteq \mathcal{U}_n$ satisfies $|\Gamma| = |\Omega| + 1$, then $\text{Null}(\mathcal{F}_{\Omega, \Gamma})$ is spanned by a real vector. Moreover, it must be the case that $\text{Null}(\mathcal{F}_{\Omega}) \cap \text{Null}(\text{Pr}_{\Gamma^c}) \cap S_1^{n-1} \neq \emptyset$.*
- (ii) *The ℓ_1 unit norm minimum support vectors satisfy*

$$\begin{aligned} & \text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n) \cap S_1^{n-1} \\ &= \bigcup_{\substack{\Gamma \subseteq \mathcal{U}_n \\ |\Gamma| = |\Omega| + 1}} (\text{Null}(\mathcal{F}_{\Omega}) \cap \text{Null}(\text{Pr}_{\Gamma^c}) \cap S_1^{n-1}). \end{aligned}$$

Proof. To prove part (i), we start by letting $\Gamma \subseteq \mathcal{U}_n$ satisfy $|\Gamma| = |\Omega| + 1$. It follows from Lem. 6(ii) that $\dim(\text{Null}(\mathcal{F}_{\Omega, \Gamma})) = 1$, so we can find a nonzero $\nu \in \mathbb{C}^{\Gamma}$ that spans $\text{Null}(\mathcal{F}_{\Omega, \Gamma})$. Since Ω satisfies (22), the complex conjugate of ν is also in the nullspace, i.e., $\bar{\nu} \in \text{Null}(\mathcal{F}_{\Omega, \Gamma})$. If all entries of ν are purely imaginary, then we can divide all entries of ν by i to obtain a nonzero real vector ν_{real} that spans $\text{Null}(\mathcal{F}_{\Omega, \Gamma})$. Otherwise, the vector $\nu_{\text{real}} := \nu + \bar{\nu}$ is nonzero, real, and spans $\text{Null}(\mathcal{F}_{\Omega, \Gamma})$. Using $\nu_{\text{real}} \in \mathbb{R}^{\Gamma}$, we then define $\bar{\nu}_{\text{real}} \in \mathbb{R}^n$ via $[\bar{\nu}_{\text{real}}]_{\Gamma} = \nu_{\text{real}} / \|\nu_{\text{real}}\|_1$ and $[\bar{\nu}_{\text{real}}]_{\Gamma^c} = 0$, so that $\bar{\nu}_{\text{real}} \in \text{Null}(\mathcal{F}_{\Omega}) \cap \text{Null}(\text{Pr}_{\Gamma^c}) \cap S_1^{n-1}$, completing the proof. To prove part (ii), we show both inclusions in turn.

\supseteq Let $z \in \text{Null}(\mathcal{F}_{\Omega}) \cap \text{Null}(\text{Pr}_{\Gamma^c}) \cap S_1^{n-1}$ for some $\Gamma \subseteq \mathcal{U}_n$ satisfying $|\Gamma| = |\Omega| + 1$. In particular, this means that $z \neq 0$, $z \in \text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n$, and $\text{Supp}(z) \subseteq \Gamma$. We also claim that $z \in \text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n)$, which can be seen as follows. Suppose, in order to reach a contradiction, that z is not in $\text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n)$, i.e., that there exists a nonzero vector \bar{z} satisfying $\text{Supp}(\bar{z}) \subseteq \text{Supp}(z) \subseteq \Gamma$ and $\bar{z} \in \text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n$. It follows that $c := [\bar{z}]_{\Gamma}$ satisfies $c \in \text{Null}(\mathcal{F}_{\Omega, \Gamma}) \setminus \{0\}$ and $\|c\|_0 < \Gamma$, which contradicts Lem. 6(i). Thus, $z \in \text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n) \cap S_1^{n-1}$, as claimed.

\subseteq Let $z \in \text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n) \cap S_1^{n-1}$ and define its support as $\Gamma = \text{Supp}(z)$. We claim that $|\Gamma| \leq |\Omega| + 1$. For a proof by contradiction, suppose that $|\Gamma| > |\Omega| + 1$. Then choose any $\bar{\Gamma} \subset \Gamma$ satisfying $|\bar{\Gamma}| = |\Omega| + 1$, and then note that $\text{Null}(\mathcal{F}_{\Omega, \bar{\Gamma}})$ is nontrivial so that there exists a vector $\bar{\eta} \in \mathbb{R}^{\bar{\Gamma}}$ satisfying $\bar{\eta} \in \text{Null}(\mathcal{F}_{\Omega, \bar{\Gamma}}) \cap S_1^{|\bar{\Gamma}|}$. Defining $\bar{z} \in \mathbb{R}^n$ as $[\bar{z}]_{\bar{\Gamma}} = \bar{\eta}$ and $[\bar{z}]_{\bar{\Gamma}^c} = 0$, it follows that $\bar{z} \in S^{n-1} \cap \text{Null}(\mathcal{F}_{\Omega})$ and $\text{Supp}(\bar{z}) \subseteq \bar{\Gamma} \subset \Gamma = \text{Supp}(z)$, which contradicts $z \in \text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n)$. Thus, we know that $|\Gamma| \leq |\Omega| + 1$ holds. Finally, choose any $\hat{\Gamma} \subseteq \mathcal{U}_n$ satisfying $\Gamma \subseteq \hat{\Gamma}$ and $|\hat{\Gamma}| = |\Omega| + 1$. It follows that $z \in \text{Null}(\mathcal{F}_{\Omega}) \cap \text{Null}(\text{Pr}_{\hat{\Gamma}^c}) \cap S_1^{n-1}$ with $\hat{\Gamma} \subseteq \mathcal{U}_n$ and $|\hat{\Gamma}| = |\Omega| + 1$, thus completing the proof. \square

Thm. 5 provides an interesting correspondence between $\text{MinSupp}(\text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n) \cap S_1^{n-1}$ and the index sets $\Gamma \subseteq \mathcal{U}_n$ with $|\Gamma| = |\Omega| + 1$. A consequence of Thm. 5 and Prop. 3 is

$$\text{Ext}(\text{Null}(\mathcal{F}_{\Omega}) \cap B_1^n) = \bigcup_{\substack{\Gamma \subseteq \mathcal{U}_n \\ |\Gamma| = |\Omega| + 1}} (\text{Null}(\mathcal{F}_{\Omega}) \cap \text{Null}(\text{Pr}_{\Gamma^c}) \cap S_1^{n-1})$$

anytime $\text{Null}(\mathcal{F}_{\Omega})$ is nontrivial. We are now ready to state the main result of this section.

Theorem 6. *Let n be a prime number, $\Omega \subseteq \mathcal{U}_n$ satisfy (22), and $S \subseteq \mathcal{U}_n$. Then, the following statements are equivalent*

- 1) $S \in \mathcal{T}_{\max}(\mathcal{F}_{\Omega})$.
- 2) For all $\Gamma \subseteq \mathcal{U}_n$ with $|\Gamma| = |\Omega| + 1$, we have

$$\sum_{k \in S} |\det(\mathcal{F}_{\Omega, \Gamma \setminus \{k\}})| < \sum_{k \in S^c} |\det(\mathcal{F}_{\Omega, \Gamma \setminus \{k\}})|.$$

- 3) Every $\bar{x} \in \mathbb{R}^n$ with $\text{Supp}(\bar{x}) \subseteq S$ is the unique solution to (20).

Proof. The equivalence of (1) and (2) follows from Prop. 4, Thm. 5(ii), and Lem. 6(ii). The equivalence of (1) and (3) follows from Thm. 3. \square

We discuss the implications of Thm. 6 with an example.

Example 4. *Let $n = 11$ and consider a 5×11 partial DFT matrix, \mathcal{F}_{Ω} , where $\Omega = \{0, 2, 4, 7, 9\}$. From Thm. 5, a vector $x \in \text{Null}(\mathcal{F}_{\Omega}) \cap \mathbb{R}^n$ has minimal support if and only if $\|x\|_0 = 5 + 1 = 6$. Furthermore, by Thm. 6, we have $S \in \mathcal{T}_{\max}(\mathcal{F}_{\Omega})$ if and only if for all $\Gamma \subseteq \mathcal{U}_{11}$ with $|\Gamma| = 6$, we have*

$$\sum_{k \in S} |\det(\mathcal{F}_{\Omega, \Gamma \setminus \{k\}})| < \sum_{k \in S^c} |\det(\mathcal{F}_{\Omega, \Gamma \setminus \{k\}})|.$$

Then, a simple computer program can be coded to check this condition, which reveals that MASC is given by

$$\begin{aligned} \mathcal{T}_{\max}(\mathcal{F}_{\Omega}) &= \{S \in 2^{\mathcal{U}_{11}} : |S| \leq 1\} \\ &\cup \left\{ \{\alpha, \beta\} \in 2^{\mathcal{U}_{11}} : \left| |\alpha - \beta| - \frac{11}{2} \right| \geq \frac{3}{2} \right\}, \end{aligned}$$

which has 56 elements (including the empty set). The recovery over 1-sparse signals is uniform. However, 2-sparse signals are not recovered uniformly. For instance, $\{0, 5\} \notin \mathcal{T}_{\max}(\mathcal{F}_{\Omega})$. Hence, there are signals $\bar{x} \in \mathbb{R}^{11}$ with $\text{Supp}(\bar{x}) = \{0, 5\}$ which cannot be recovered via (20). Furthermore, $\mathcal{T}_{\max}(\mathcal{F}_{\Omega})$ does not contain any index set of cardinality 3 or higher.

The determinants appearing in Thm. 6 are a consequence of the computation of a vector spanning the one dimensional subspace $\text{Null}(\mathcal{F}_{\Omega,\Gamma})$ when $|\Gamma| = |\Omega| + 1$, as indicated by Lem. 6(ii). Unfortunately, this characterization of the $\text{Null}(\mathcal{F}_{\Omega,\Gamma})$ is not useful for computational purposes since it requires the computation of a possibly large number of determinants of large matrices. One approach for significantly reducing the computation is to make the following assumption.

Assumption 1. *Let n be a prime number that defines the size of the DFT matrix \mathcal{F}_n . The index set Ω takes the form*

$$\Omega = \{0, 1, \dots, \bar{m}, n - \bar{m}, \dots, n - 1\} \quad (23)$$

for some $\bar{m} \in \{1, 2, \dots, (n-3)/2\}$ so that $|\Omega| = 2\bar{m} + 1 \leq n-2$ since p is a prime number. This implies that any $\Gamma \subseteq \mathcal{U}_n$ satisfying $|\Gamma| = |\Omega| + 1$ must also satisfy $|\Gamma| \leq n-1$.

Given Ω satisfying Assumption 1, we define, for any $\Gamma \subseteq \mathcal{U}_n$ satisfying $|\Gamma| = |\Omega| + 1$, the polynomial

$$f_\Gamma(z) := \prod_{k \in \Gamma} (z - \xi^k) \equiv \sum_{i=0}^{n-1} c_i(\Gamma) z^i \quad (24)$$

for some vector $c(\Gamma) \in \mathbb{R}^n$ of the form

$$c(\Gamma) = [c_0(\Gamma) \ c_1(\Gamma) \ \dots \ c_{|\Gamma|}(\Gamma) \ 0 \ 0 \ \dots \ 0]^T. \quad (25)$$

The action of the DFT on $c(\Gamma)$ is given componentwise by

$$[\mathcal{F}_n c(\Gamma)]_k = \sum_{i=0}^{n-1} c_i(\Gamma) \xi^{ik} = \sum_{i=0}^{n-1} c_i(\Gamma) (\xi^k)^i = f_\Gamma(\xi^k), \quad (26)$$

which is a useful property. It follows from (24) and (26) that

$$[\mathcal{F}_n c(\Gamma)]_\Gamma = 0. \quad (27)$$

The definition in (24) also allows us to rewrite the determinants needed in Thm. 6 in a convenient form, as we now show.

Proposition 5. *Let n be a prime number and $\Omega \subseteq \mathcal{U}_n$ satisfy (23). If $\Gamma \subseteq \mathcal{U}_n$ satisfies $|\Gamma| = |\Omega| + 1$, then*

$$\det(\mathcal{F}_{\Omega, \Gamma \setminus \{k\}}) = \pm \beta(\Gamma) \frac{1}{\xi^{k(n-\bar{m})} f_\Gamma(\xi^k)} \quad \text{for all } k \in \Gamma,$$

where $f_\Gamma(z)$ is the derivative of $f_\Gamma(z)$ and

$$\beta(\Gamma) := \xi^{(n-\bar{m}) \sum_{l \in \Gamma} l} \prod_{l \in \Gamma, r \in \Gamma, l < r} (\xi^l - \xi^r) \quad (28)$$

is a complex constant independent of k .

Proof. Let $k \in \Gamma$ and define $\bar{\Gamma} := \Gamma \setminus \{k\}$. If, for each $l \in \bar{\Gamma}$, we divide the corresponding column of $\mathcal{F}_{\Omega, \bar{\Gamma}}$ by the constant $\xi^{l(n-\bar{m})}$, then we end up with a Vandermonde matrix, $V(\bar{\Gamma}) \in \mathbb{C}^{(2\bar{m}+1) \times (2\bar{m}+1)}$, whose entries are of the form $[V(\bar{\Gamma})]_{i,j} = (\xi^{\bar{\Gamma}_j})^i$. Using properties of determinants and the formula for the determinant of the Vandermonde matrix, we get

$$\det(\mathcal{F}_{\Omega, \bar{\Gamma}}) = \xi^{(n-\bar{m}) \sum_{l \in \bar{\Gamma}} l} \prod_{l \in \bar{\Gamma}, r \in \bar{\Gamma}, l < r} (\xi^l - \xi^r).$$

Note that, up to a sign change, we have

$$\prod_{l \in \Gamma, r \in \Gamma, l < r} (\xi^l - \xi^r) = \pm \left(\prod_{l \in \bar{\Gamma}, r \in \bar{\Gamma}, l < r} (\xi^l - \xi^r) \right) \left(\prod_{l \in \Gamma, l=k} (\xi^k - \xi^l) \right).$$

But since

$$\prod_{l \in \Gamma, l=k} (\xi^k - \xi^l) = f_\Gamma(\xi^k),$$

up to a sign change, we have

$$\det(\mathcal{F}_{\Omega, \Gamma}) = \pm \frac{\xi^{(n-\bar{m}) \sum_{l \in \Gamma} l} \prod_{l \in \Gamma, r \in \Gamma, l < r} (\xi^l - \xi^r)}{\xi^{k(n-\bar{m})} f_\Gamma(\xi^k)}.$$

The result now follows from the definition of $\beta(\Gamma)$ in (28). \square

We can further develop Prop. 5 by observing that

$$\begin{aligned} nz^{n-1} &= \frac{d}{dz}(z^n - 1) = \frac{d}{dz}(f_\Gamma(z) f_{\Gamma^c}(z)) \\ &= f_\Gamma(z) f_{\Gamma^c}'(z) + f_\Gamma'(z) f_{\Gamma^c}(z), \end{aligned}$$

which combined with $f_\Gamma(\xi^k) = 0$ for all $k \in \Gamma$, gives

$$f_\Gamma(\xi^k) = \frac{n \xi^{k(n-1)}}{f_{\Gamma^c}'(\xi^k)} \quad \text{for all } k \in \Gamma.$$

Combining this identity with Prop. 5 gives the following result.

Proposition 6. *Let n be a prime number and $\Omega \subseteq \mathcal{U}_n$ satisfy (23). If $\Gamma \subseteq \mathcal{U}_n$ satisfies $|\Gamma| = |\Omega| + 1$, then*

$$\det(\mathcal{F}_{\Omega, \Gamma \setminus \{k\}}) = \pm \beta(\Gamma) \frac{f_{\Gamma^c}'(\xi^k)}{n \xi^{(n-\bar{m}-1)k}} \quad \text{for all } k \in \Gamma,$$

where $\beta(\Gamma) \in \mathbb{C}$ is the constant defined in (28).

With the help of these propositions, we can provide another characterization of $\mathcal{T}_{\max}(\mathcal{F}_\Omega)$ when Ω satisfies (23) and the dimension of the unknown real signal is a prime number.

Corollary 2. *Let n be a prime number and the set $\Omega \subseteq \mathcal{U}_n$ satisfy (23). It follows that $S \subseteq \mathcal{U}_n$ is in $\mathcal{T}_{\max}(\mathcal{F}_\Omega)$ if and only if for all $\Gamma \subseteq \mathcal{U}_n$ satisfying $|\Gamma| = |\Omega| + 1$, one of the following equivalent conditions holds:*

$$\begin{aligned} (C1) \quad & \sum_{k \in S} \sum_{\Gamma} |f_\Gamma(\xi^k)|^{-1} < \sum_{k \in S^c} \sum_{\Gamma} |f_\Gamma(\xi^k)|^{-1}. \\ (C2) \quad & \sum_{k \in S} \sum_{\Gamma} |f_{\Gamma^c}'(\xi^k)| < \sum_{k \in S^c} \sum_{\Gamma} |f_{\Gamma^c}'(\xi^k)|. \end{aligned}$$

Proof. These results follow by combining Thm. 6 with Prop. 5 (for (C1)) and Prop. 6 (for (C2)). \square

The identity in (26) shows that (C2) can be interpreted as a condition on the DFT acting on the vectors $c(\Gamma^c)$ defined in (25). Hence, existing results can be used to bound the level of sparsity for which exact recovery is possible. To take advantage of this, we require the following definition.

Definition 6 (coherence). *Let $\Phi \in \mathbb{C}^{p \times q}$ have ℓ_2 -normalized columns. The coherence of Φ is defined as*

$$\mu(\Phi) := \max_{k=l} |\Phi_k \Phi_l|,$$

where Φ_k denotes the k th column of Φ .

The following known result uses the coherence to show that a nonzero vector cannot be arbitrarily well concentrated with respect to two different orthonormal bases.

Lemma 7 ([38]). *Let $A, B \in \mathbb{C}^{n \times n}$ be unitary and $S, T \subseteq \mathcal{U}_n$ be nonempty. If there exists a nonzero $\bar{z} \in \mathbb{C}^n$ satisfying*

$\|\bar{z} - \text{Pr}_S(\bar{z})\|_1 \leq \varepsilon \|\bar{z}\|_1$ for some $\varepsilon \in [0, 1]$, and a nonzero $\bar{y} \in \mathbb{C}^n$ with $\text{Supp}(\bar{y}) = T$ such that $A\bar{z} = B\bar{y}$, then

$$|S||T| \geq \frac{1 - \varepsilon}{\mu^2(\begin{bmatrix} A \\ B \end{bmatrix})}, \quad (29)$$

where the coherence $\mu(\cdot)$ is defined in Defn. 6.

Using this lemma we can get a bound on the sparsity of signals that can be recovered exactly via (20).

Corollary 3. *Let n be a prime number and $\Omega \subseteq \mathcal{U}_n$ satisfy (23). Every s -sparse signal $\bar{x} \in \mathbb{R}^n$ is the unique solution to the optimization problem in (20) if*

$$s < \frac{n}{2(n-|\Omega|)}. \quad (30)$$

Proof. Define $A := I_n \in \mathbb{C}^{n \times n}$ (the $n \times n$ identity matrix) and $B := \mathcal{F}_n \in \mathbb{C}^{n \times n}$, and then observe using Defn. 6 that

$$\mu(\begin{bmatrix} I_n & \mathcal{F}_n \end{bmatrix}) = 1/\sqrt{n}. \quad (31)$$

Also note that since Ω is assumed to satisfy (23), we know that $|\Omega| = 2\bar{m} + 1$ for some $\bar{m} \in \{1, 2, \dots, (n-3)/2\}$.

Next, let $T := \mathcal{U}_{n-2\bar{m}-1}$ and let $S \subseteq \mathcal{U}_n$ be any set satisfying $|S| \leq s$. Combining this with (30) we find that

$$|S| \leq s < \frac{n}{2(n-|\Omega|)} = \frac{n}{2(n-2\bar{m}-1)} = \frac{n}{2|T|},$$

which combined with (31) and choosing $\varepsilon = 1/2$ yields

$$|S||T| < \frac{n}{2} = \frac{1}{2\mu^2(\begin{bmatrix} I_n & \mathcal{F}_n \end{bmatrix})} = \frac{1-\varepsilon}{\mu^2(\begin{bmatrix} I_n & \mathcal{F}_n \end{bmatrix})}.$$

Since this shows that (29) does not hold, we must conclude from Lem. 7 that there does not exist a pair of nonzero vectors $\bar{y}, \bar{z} \in \mathbb{C}^n$ satisfying $\|\bar{z} - \text{Pr}_S(\bar{z})\|_1 \leq \frac{1}{2}\|\bar{z}\|_1$, $\text{Supp}(\bar{y}) = T$, and $\bar{z} = \mathcal{F}_n \bar{y}$; this is equivalent to saying that all vectors $\bar{y} \in \mathbb{C}^n$ with $\text{Supp}(\bar{y}) = T$ must satisfy $\|\mathcal{F}_n \bar{y} - \text{Pr}_S(\mathcal{F}_n \bar{y})\|_1 > \frac{1}{2}\|\mathcal{F}_n \bar{y}\|_1$, which itself may equivalently be stated as

$$\|(\mathcal{F}_n \bar{y})_{S^c}\|_1 > \|(\mathcal{F}_n \bar{y})_S\|_1 \quad \forall \bar{y} \text{ with } \text{Supp}(\bar{y}) = T. \quad (32)$$

Next, consider any $\Gamma \subseteq \mathcal{U}_n$ satisfying $|\Gamma| = |\Omega| + 1 = 2(\bar{m} + 1)$, which implies that $|\Gamma^c| = n - 2(\bar{m} + 1)$. It follows from this equality, (25), and the definition of T that

$$\text{Supp}(c(\Gamma^c)) = \mathcal{U}_{\Gamma^c/+1} = \mathcal{U}_{n-2\bar{m}-1} = T.$$

Thus, using (26), (27), and $\bar{y} = c(\Gamma^c)$ in (32), we have

$$\begin{aligned} \sum_{k \in S^c} |f_{\Gamma^c}(\xi^k)| &= \sum_{k \in S^c} |[\mathcal{F}_n(c(\Gamma^c))]_k| \\ &= \sum_{k \in S^c} |[\mathcal{F}_n(c(\Gamma^c))]_k| \\ &= \|(\mathcal{F}_n c(\Gamma^c))_{S^c}\|_1 > \|(\mathcal{F}_n c(\Gamma^c))_S\|_1 \\ &= \sum_{k \in S} |[\mathcal{F}_n(c(\Gamma^c))]_k| \\ &= \sum_{k \in S} |[\mathcal{F}_n(c(\Gamma^c))]_k| = \sum_{k \in S} |f_{\Gamma^c}(\xi^k)|, \end{aligned}$$

so that (C2) holds. It now follows from Cor. 2 that $S \in \mathcal{T}_{\max}(\mathcal{F}_\Omega)$, i.e., $\{S\} \subset \mathcal{T}_{\max}(\mathcal{F}_\Omega)$, which in turn implies via Thm.3 that any \bar{x} satisfying $\text{Supp}(\bar{x}) = S$ is the unique solution to the optimization problem (20). Since S satisfied $|S| \leq s$ but was otherwise arbitrary, the proof is complete. \square

From Cor. 3, the partial DFT matrices we consider in Assumption 1, coherence based bounds on maximum sparsity that can be recovered are extremely simple to calculate. We note this as the biggest merit of this corollary. More generally, the results we provide in this subsection show how the machinery we built in the previous section can be used to describe the collection of all support sets for which ℓ_1 -recovery is always successful, when the dictionary is the partial Fourier transform - probably the most commonly studied dictionary in the compressed sensing/sparse recovery literature.

C. Experiments on Partial DFT Matrices

In this section we perform two numerical experiments aimed at better understanding the theoretical results obtained for sparse recovery when the dictionary is a partial DFT matrix. We refer to § V-C for implementation details, in particular how the sampling of each random vector \bar{x} is performed.

1) *Tightness of the Bound in Cor. 3:* The bound in Cor. 3 on the sparsity for which recovery is guaranteed is not expected to be tight. We illustrate this with a simple experiment using a partial DFT matrix \mathcal{F}_Ω with $n = 19$ and Ω chosen to satisfy (23) with $\bar{m} = 7$ so that $|\Omega| = 15$. By Cor. 3, we know that if $s < 19/8 = 2.375$, then every s -sparse signal $\bar{x} \in \mathbb{R}^{19}$ can be recovered as a solution of the ℓ_1 -minimization problem

$$\min_x \|x\|_1 \text{ subject to } \mathcal{F}_\Omega \bar{x} = \mathcal{F}_\Omega x. \quad (33)$$

Thus, the corollary guarantees recovery only for $s \in \{1, 2\}$.

To test this conclusion, for each sparsity level s , we sampled 1000 signals $\bar{x} \in \mathbb{R}^{19}$ and tested whether ℓ_1 -recovery via solving (33) was successful. The probability of recovering the true signal for each sparsity level is illustrated in Fig. 5. In that same figure, we also plot the probability that a support set of cardinality s is in $\mathcal{T}_{\max}(\mathcal{F}_\Omega)$, which is calculated using Cor. 2. (We terminate the plot once the probabilities hit zero.) From this second plot, we see that the highest sparsity level that can be recovered is three, thus showing that the bound provided by Cor. 3 is not tight. Also, although the probability of an index set being in $\mathcal{T}_{\max}(\mathcal{F}_\Omega)$ drops rapidly, the drop rate for the exact recovery probability using ℓ_1 -minimization is much slower. This indicates that for a given support set $S \notin \mathcal{T}_{\max}(\mathcal{F}_\Omega)$, the signals with support S that are recoverable via ℓ_1 -minimization can still be abundant. Moreover, this plot gives a summary of the structure of $\mathcal{T}_{\max}(\mathcal{F}_\Omega)$ in terms of support sets of cardinality s .

2) *The Maximal Recoverable Sparsity Level (MRSL):* Let n be a prime number. For any index set $\Omega \subseteq \mathcal{U}_n$ that satisfies (23), what is the maximal sparsity level for which every vector of that sparsity level can be recovered via ℓ_1 -minimization? We call this number the MRSL and denote it by $s_{\max}(\Omega)$. Cor. 3 gives a lower bound for $s_{\max}(\Omega)$. In principle, $s_{\max}(\Omega)$ can be obtained by computing every extreme point of $\text{Null}(\mathcal{F}_\Omega) \cap \mathbb{B}_1^n$ and checking whether (C2) in Cor. 2 is satisfied. More precisely, let us first define

$$E(\Omega) := \{\Gamma \subseteq \mathcal{U}_n : |\Gamma| = |\Omega| + 1\}.$$

For each $\Gamma \in E(\Omega)$, we sort $\{|f_{\Gamma^c}(\xi^k)|\}_k$ into decreasing order and define $s_{\max}(\Omega, \Gamma)$ as the maximum integer

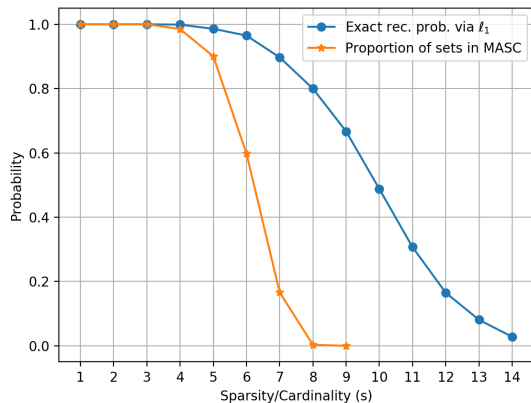


Fig. 5. The exact recovery probability (blue line) via ℓ_1 -minimization using (33) when the signal \bar{x} is real-valued of dimension $n = 19$ and the measurement set Ω for the partial DFT matrix F_Ω satisfies (23) with $\bar{m} = 7$. We also plot (orange line) the probability that a randomly selected support set of cardinality s from U_n is in $T_{\max}(F_\Omega)$.

so that the sum of the first $s_{\max}(\Omega, \Gamma)$ largest elements of $\{|f_{\Gamma^c}(\xi^k)|\}_k$ does not exceed half of the total sum $\sum_{k=0}^{n-1} |f_{\Gamma^c}(\xi^k)|$, so that (C2) in Cor. 2 holds. It follows that

$$s_{\max}(\Omega) \equiv \min_{\Gamma \in \mathcal{E}(\Omega)} s_{\max}(\Omega, \Gamma). \quad (34)$$

For large n , the cardinality of $\mathcal{E}(\Omega)$ (i.e., the number of extreme points) is huge, making the calculation in (34) impractical. This motivates the need for an upper bound on $s_{\max}(\Omega)$ that is tractable. We describe such a strategy next.

Let $\hat{E}(\Omega)$ denote any randomly selected subset of $\mathcal{E}(\Omega)$. For such a choice, it follows from (34) that

$$\widehat{s_{\max}}(\Omega) := \min_{\Gamma \in \hat{\mathcal{E}}(\Omega)} s_{\max}(\Omega, \Gamma) \geq s_{\max}(\Omega), \quad (35)$$

which means that $\widehat{s_{\max}}(\Omega)$ is an upper bound for $s_{\max}(\Omega)$. Moreover, the computation of $\widehat{s_{\max}}(\Omega)$ will be efficient provided we select $\hat{E}(\Omega)$ to be a relatively small subset of $\mathcal{E}(\Omega)$. To evaluate this strategy, we use the following set of matrices.

Test 1. Let $n = 1009$ (a prime number) and choose the collection of sets $\{\Omega_j\}_{j=0}^{38} \subset \mathcal{U}_{1009}$ to satisfy (23) with sizes $|\Omega_j| = 247 + 20j$ for each $0 \leq j \leq 38$.⁴

For each Ω_j described in Test 1, we randomly select a subset $\hat{E}(\Omega_j) \subset \mathcal{E}(\Omega_j)$ of size 1000 and proceed to compute $\widehat{s_{\max}}(\Omega_j)$ using (35) as an upper bound on the MRSL when using the ℓ_1 -recovery problem (33) with $\Omega = \Omega_j$. The results are presented in Fig. 6. For each Ω_j , we plot the sampling-based upper bound $\widehat{s_{\max}}(\Omega_j)$ and the coherence-based lower bound $\frac{n}{2(n-|\Omega_j|)}$ derived in Cor. 3. Note that for $|\Omega_j| \leq 600$ it holds that $s_{\max}(\Omega_j) \leq 2$ so that the maximal sparsity level that can be recovered is small.

To illustrate that our strategy is better than a naive approach that repeatedly solves (33) to estimate $s_{\max}(\Omega)$ (see Algorithm 1 below), consider the following setup.

⁴To illustrate how large the number of extreme points can be, we remark that when $|\Omega| = 507$, the number of extreme points is $\binom{1009}{508} \approx 1.34 \times 10^{302}$.

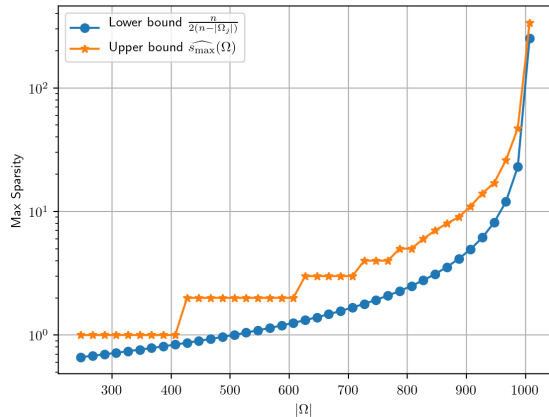


Fig. 6. Bounds for the MRSL via (33) for the subsets $\{\Omega_j\}_{j=0}^{38}$ in Test 1. The lower bound $\frac{n}{2(n-|\Omega_j|)}$ (blue line) is given by Cor. 3 and the upper bound $s_{\max}(\Omega)$ by (35) (orange line).

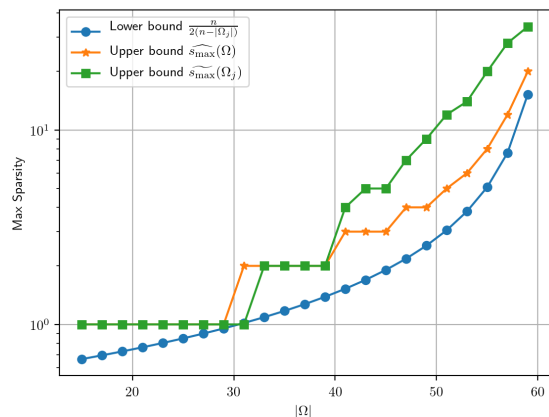


Fig. 7. Bounds for the MRSL via (33) for the matrices $\{\Omega_j\}_{j=0}^{22}$ in Test 2. The lower bound $\frac{n}{2(n-|\Omega_j|)}$ (blue line) is given by Cor. 3. The upper bound $s_{\max}(\Omega)$ (orange line) is given by (35) and the upper bound $\widehat{s_{\max}}(\Omega_j)$ (green line) is returned by Algorithm 1 with inputs $\Omega = \Omega_j$ and $K = 1000$.

Test 2. Let $n = 61$ (a prime number) and define $\{\Omega_j\}_{j=0}^{22} \subset \mathcal{U}_n$ by choosing $\bar{m} \in \{7, 8, \dots, 29\}$ in (23), which means that $|\Omega_j| = 15 + 2j$ for $0 \leq j \leq 22$.

For each Ω_j described in Test 2, we use Algorithm 1 with inputs $\Omega = \Omega_j$ and $K = 1000$ to obtain $\widehat{s_{\max}}(\Omega_j)$ as an upper bound to $s_{\max}(\Omega_j)$. The results are shown in Fig. 7, where we also include the upper bound $\widehat{s_{\max}}(\Omega)$ in (35) obtained via sampling of the extreme points. For large values of $|\Omega_j|$, the upper bound $\widehat{s_{\max}}(\Omega_j)$ is significantly better. Moreover, in terms of the computation time the estimation obtained via sampling of the extreme points is more than an order of magnitude faster in comparison to the naive approach we described in Algorithm 1.

VII. CONCLUSION

In this paper we introduced a new framework for the recovery of arbitrary sparsity patterns via ℓ_1 -minimization. We showed that there is a maximal recoverable sparsity pattern for each dictionary $\Phi \in \mathbb{R}^{m \times n}$, which we called the *maximum abstract simplicial complex (MASC) associated with Φ* . We

Algorithm 1 A simple upper bound for the MRSL.

```

1: inputs:  $\Omega \subset \mathcal{U}_n$  and sampling size  $K \geq 1$ .
2: set  $\tilde{s} \leftarrow n$ 
3: loop
4:   Randomly select  $K$  vectors in  $\mathbb{R}^n$  that are  $\tilde{s}$ -sparse.
5:   if all  $K$  vectors are correctly recovered via (33) then
6:     return  $\widetilde{s_{\max}}(\Omega) \leftarrow \tilde{s}$ 
7:   end if
8:   set  $\tilde{s} \leftarrow \tilde{s} - 1$ 
9: end loop

```

provided a characterization of the MASC using the extreme points of $\text{Null}(\Phi) \cap \mathbb{B}_1^n$, and a second characterization using the vectors of minimal support of $\text{Null}(\Phi)$. Furthermore, we showed how our approach benefits the analysis of sparse recovery when the dictionary is an incidence matrix associated with a simple graph or a $m \times p$ partial DFT matrix, where p is a prime number and the unknown signal is real. In particular, we gave a complete characterization of the MASC associated with these matrix classes, which allowed us to characterize the collection of all support sets for which exact recovery via ℓ_1 -minimization is always possible. Interestingly, we showed that when the dictionary is an incidence matrix, the *Nullspace Property* can be verified in polynomial time, although this condition is NP-hard to check for matrices in general. We also showed that a computationally more advantageous characterization can be achieved if stronger assumptions are imposed on the measurement indices of the partial DFT matrix.

Our framework opens the door for new directions of research. In particular, the connection between extreme points, vectors of minimal support, and the GNUF can be exploited to study sparse recovery problems for new classes of dictionaries.

REFERENCES

- [1] M. Zhao, M. D. Kaba, R. Vidal, D. P. Robinson, and E. Mallada, "Sparse recovery over graph incidence matrices," in *57th IEEE Conference on Decision and Control (CDC)*, 2018, pp. 364–371.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [4] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [5] E. Elhamifar and R. Vidal, "Sparse manifold clustering and embedding," in *Neural Information Processing and Systems*, 2011.
- [6] E. Elhamifar, G. Sapiro, and R. Vidal, "Finding exemplars from pairwise dissimilarities via simultaneous sparse recovery," in *Neural Information Processing and Systems*, 2012.
- [7] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, 2013.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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*, 2015, pp. 3965–3970.
- [12] E. Schwab, B. D. Haeffele, R. Vidal, and N. Charon, "Global optimality in separable dictionary learning with applications to the analysis of diffusion mri," *SIAM Journal of Imaging Sciences*, 2019.
- [13] M. Coates, Y. Pointurier, and M. Rabbat, "Compressed network monitoring," in *IEEE/SP Workshop on Statistical Signal Processing*. IEEE, 2007, pp. 418–422.
- [14] 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.
- [15] W. Xu, E. Mallada, and A. Tang, "Compressive sensing over graphs," in *IEEE INFOCOM*. IEEE, 2011, pp. 2087–2095.
- [16] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [17] G. Obozinski, L. Jacob, and J.-P. Vert, "Group lasso with overlaps: the latent group lasso approach," National Institute for Research in Digital Science and Technology (INRIA), Research Report, Oct. 2011.
- [18] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3488–3497, 2009.
- [19] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *Journal of Machine Learning Research*, vol. 12, pp. 3371–3412, 2011.
- [20] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on information theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [21] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," in *Forum of Mathematics, Sigma*, vol. 5. Cambridge University Press, 2017.
- [22] V. Pappas, J. Sulam, and M. Elad, "Working locally thinking globally: Theoretical guarantees for convolutional sparse coding," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5687–5701, Nov 2017.
- [23] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4094–4107, 2012.
- [24] C. You, D. Robinson, and R. Vidal, "Scalable sparse subspace clustering by orthogonal matching pursuit," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [25] S. Soltan, M. Yannakakis, and G. Zussman, "Joint cyber and physical attacks on power grids: Graph theoretical approaches for information recovery," in *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS '15. New York, NY, USA: ACM, 2015, pp. 361–374.
- [26] S. W. Smith, *The scientist and engineer's guide to digital signal processing*. California Technical Pub. San Diego, 1997.
- [27] R. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
- [28] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Birkhäuser Basel, 2013, vol. 1, no. 3.
- [33] 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.
- [34] C. Godsil and G. F. Royle, *Algebraic graph theory*, ser. Graduate text in mathematics. Springer, New York, 2001.
- [35] O. Ore, *Theory of Graphs*, ser. AMS colloquium publications. American Mathematical Society, 1962, no. pt. 1.
- [36] T. Tao, "An uncertainty principle for cyclic groups of prime order," *Mathematical Research Letters*, vol. 12, no. 1, pp. 121–127, 2005.
- [37] R. Meshulam, "An uncertainty inequality for finite abelian groups," *European Journal of Combinatorics*, vol. 27, no. 1, pp. 63–67, 2006.
- [38] E. Riegler and H. Bölcskei, "Uncertainty relations and sparse signal recovery," *arXiv preprint arXiv:1811.03996*, 2018.

Mustafa D. Kaba graduated from Middle East Technical University (METU) (Ankara, Turkey) in 2004, with a B.Sc. degree in Electrical and Electronics Engineering (Double major in Mathematics). He received Ph.D. degrees in Mathematics, first from METU, in 2011, with his thesis in arithmetic algebraic geometry, and then from University of Groningen (UoG) (Groningen, The Netherlands) in 2014, with his thesis in systems and control theory. His Ph.D. work at METU was supported by the Turkish Scientific and Technical Research Council (TÜBİTAK) and during his studies at UoG, he was a recipient of the Ubbo Emmius Fellowship. After receiving his Ph.D. from UoG, he worked as a postdoctoral researcher, lead research scientist, and assistant research scientist at Northeastern University (Boston, MA, USA), General Electric Global Research Center (Niskayuna, NY, USA), and Johns Hopkins University (Baltimore, MD, USA) respectively. Currently, he is working as an applied researcher at eBay Inc.

Mengnan Zhao received the M.S. degrees in Electrical and Computer Engineering and Applied Mathematics from Johns Hopkins University in 2019, M.S. degree in Applied Physics from Cornell University in 2015, and B.S. degree in Optical Information Science and Technology from Sun Yat-sen University, Guangzhou, China in 2013. His research interests lie in the areas of signal processing, machine learning, and mathematical data science.

René Vidal received his B.S. degree in Electrical Engineering (valedictorian) from the Pontificia Universidad Católica de Chile in 1997 and his M.S. and Ph.D. degrees in Electrical Engineering and Computer Science from the University of California at Berkeley in 2000 and 2003, respectively. He is currently the Director of the Mathematical Institute for Data Science (MINDS) and the Hershel L. Seder Professor of Department of Biomedical Engineering at The Johns Hopkins University, where he has been since 2004. He is co-author of the book "Generalized Principal Component Analysis" (Springer 2016), co-editor of the book "Dynamical Vision" (Springer 2006) and co-author of over 250 articles in machine learning, computer vision, signal and image processing, biomedical image analysis, hybrid systems, robotics and control. He is or has been Associate Editor in Chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence and Computer Vision and Image Understanding, Associate Editor or Guest Editor of Medical Image Analysis, the IEEE Transactions on Pattern Analysis and Machine Intelligence, the SIAM Journal on Imaging Sciences, Computer Vision and Image Understanding, the Journal of Mathematical Imaging and Vision, the International Journal on Computer Vision and Signal Processing Magazine. He has received numerous awards for his work, including the 2016 D'Alembert Faculty Fellowship, the 2012 IAPR J.K. Aggarwal Prize, the 2009 ONR Young Investigator Award, the 2009 Sloan Research Fellowship and the 2005 NSF CAREER Award. He is a Fellow of the IEEE, Fellow of IAPR, Fellow of AIMBE, and a member of the ACM and SIAM.

Daniel P. Robinson received his Ph.D. from the University of California at San Diego in 2007. He spent the next three years working with Nicholas I. M. Gould and Jorge Nocedal as a Postdoctoral Researcher in the Mathematical Institute at the University of Oxford and the Department of Industrial Engineering and Management Sciences at Northwestern University. In 2011 he joined the Department of Applied Mathematics and Statistics in the Whiting School of Engineering at Johns Hopkins University, and then in 2019 he joined the Department of Industrial and Systems Engineering in the PC Rossin College of Engineering & Applied Science at Lehigh University. His primary research area is optimization with specific interest in the design, analysis, and implementation of efficient algorithms for large-scale convex and nonconvex problems, with particular interest in applications related to computer vision and medicine/healthcare. He is a member of the Society of Industrial and Applied Mathematics (SIAM), the SIAM Activity Group (SIAG) in Optimization, the Mathematical Optimization Society (MOS), and the Institute for Operations Research and the Management Sciences (INFORMS), in addition to being the INFORMS Vice-Chair for Nonlinear Optimization (2014-2016). Daniel has also served as the cluster chair for the 2016 International Conference on Continuous Optimization (ICCOPT) in Tokyo, Japan, a Program Committee member for the 2017 AAAI Conference on Artificial Intelligence in San Francisco, California, and is currently an Associated Editor for Computational Optimization and Applications, Optimization Methods and Software, and Journal of Scientific Computing. Finally, Daniel has been awarded the Professor Joel Dean Award for Excellence in Teaching at Johns Hopkins University in 2012 and 2018, and awarded a Best Paper Prize from Optimization Letters for his co-authored 2018 paper titled "Concise Complexity Analyses for Trust-Region Methods".

Enrique Mallada (S'09-M'13-SM') is an Assistant Professor of Electrical and Computer Engineering at Johns Hopkins University. Prior to joining Hopkins in 2016, he was a Post-Doctoral Fellow in the Center for the Mathematics of Information at Caltech from 2014 to 2016. He received his Ingeniero en Telecomunicaciones degree from Universidad ORT, Uruguay, in 2005 and his Ph.D. degree in Electrical and Computer Engineering with a minor in Applied Mathematics from Cornell University in 2014. Dr. Mallada was awarded the NSF CAREER award in 2018, the ECE Director's PhD Thesis Research Award for his dissertation in 2014, the Center for the Mathematics of Information (CMI) Fellowship from Caltech in 2014, and the Cornell University Jacobs Fellowship in 2011. His research interests lie in the areas of control, dynamical systems and optimization, with applications to engineering networks such as power systems and the Internet.