

Sparse Recovery of Nonnegative Signals with Minimal Expansion

M. Amin Khajehnejad Alexandros G. Dimakis Weiyu Xu Babak Hassibi

California Institute of Technology

email: {amin,adim,weiyu,hassibi}@caltech.edu *

Abstract

We investigate the problem of reconstructing a high-dimensional nonnegative sparse vector from lower-dimensional linear measurements. While much work has focused on dense measurement matrices, sparse measurement schemes can be more efficient both with respect to signal sensing as well as reconstruction complexity. Known constructions use the adjacency matrices of expander graphs, which often lead to recovery algorithms which are much more efficient than ℓ_1 minimization. However, prior constructions of sparse measurement matrices rely on expander graphs with very high expansion coefficients which make the construction of such graphs difficult and the size of the recoverable sets very small.

In this paper, we introduce sparse measurement matrices for the recovery of nonnegative vectors, using perturbations of the adjacency matrices of expander graphs requiring much smaller expansion coefficients, hereby referred to as *minimal expanders*. We show that when ℓ_1 minimization is used as the reconstruction method, these constructions allow the recovery of signals that are almost three orders of magnitude larger compared to the existing theoretical results for sparse measurement matrices. We provide for the first time tight upper bounds for the so called *weak* and *strong* recovery thresholds when ℓ_1 minimization is used. We further show that the success of ℓ_1 optimization is equivalent to the existence of a “unique” vector in the set of solutions to the linear equations, which enables alternative algorithms for ℓ_1 minimization. We further show that the defined minimal expansion property is *necessary* for all measurement matrices for compressive sensing, (even when the non-negativity assumption is removed) therefore implying that our construction is tight. We finally present a novel recovery algorithm that exploits expansion and is much more computationally efficient compared to ℓ_1 minimization.

*This work was supported in part by the National Science Foundation under grants CCF-0729203, CNS-0932428 and CCF-1018927, by the Office of Naval Research under the MURI grant N00014-08-1-0747, and by Caltech’s Lee Center for Advanced Networking.

1 Introduction

We investigate the problem of reconstructing a k -sparse signal \mathbf{x} from m linear measurements, $\mathbf{y} = \mathbf{A}\mathbf{x}$, where m is smaller than the ambient dimension of the signal n , but larger than k . \mathbf{A} here is the $m \times n$ measurement matrix. In this paper, we focus on the case where the nonzero entries of \mathbf{x} are positive, a special case that is of great practical interest. Signals arising in many problems are naturally nonnegative. Some examples are image processing, DNA micro-arrays, network monitoring, Hidden Markov models and many more applications in which the actual data is of nonnegative nature.

In compressed sensing, \mathbf{A} is often a dense matrix drawn from some ensemble of random matrices (see, e.g., [3]). In this paper, however, we will focus on sparse measurement matrices. This is important for numerous reasons. In several applications, like DNA micro arrays, the cost of each measurement increases with the number of coordinates of \mathbf{x} involved [16, 27]. Besides, sparse measurement matrices often make possible the design of faster decoding algorithms (e.g., [6, 8, 12, 9, 13, 14]) apart from the general linear programming decoder [3, 19].

Unlike Gaussian matrices, where reasonably sharp bounds on the thresholds which guarantee linear programming to recover sparse signals have been obtained [2], such sharp bounds do not exist for expander-graph-based measurements. Finding such sharp bounds is the main focus of the current paper, for the special case where the k -sparse vector is *nonnegative*. Compressed sensing for nonnegative signals has also been studied separately in various papers [4, 11].

In this paper, we carefully examine the connection between linear programming recovery and the fundamental properties of the measurement matrix. In particular, we focus on sparse measurement matrices. For a given measurement matrix, the success of linear programming recovery is often certified by the *Restricted Isometry Property* (RIP) of the matrix. For random dense matrices, these conditions have been studied to a great extent in the past few years. For sparse matrices, however, there are only a handful of promising results, including the recent work of Berinde et al. [5] that showed the adjacency matrices of suitable unbalanced expander graphs satisfy an RIP property for ℓ_1 norm. However, it turns out that RIP conditions are only sufficient conditions for the success of linear programming decoding, and often fail to characterize all the good measurement matrices. A complete characterization of good measurement matrices was recently given in terms of their null spaces [17, 20, 21, 23]. A necessary and sufficient condition for the success of ℓ_1 minimization is therefore called the “*null space property*”¹.

¹Also referred to as the k -neighborly property [1].

Donoho et al. [1] were the first to prove the validity of this condition with high probability for random i.i.d. Gaussian matrices, and were able to compute fairly tight thresholds on when linear-programming-based compressed sensing works [2]. The first analysis of the null space for sparse matrices has been done by Berinde et al. [10], where in particular they consider measurement matrices that are adjacency matrices of *expander graphs*. It is shown that every $(2k, \epsilon)$ expander graph² with $\epsilon \leq \frac{1}{6}$ satisfies the null space property, and therefore every k -sparse vector can be recovered from the corresponding linear measurements. The recovery thresholds given by this result, namely the relationship between $\frac{k}{n}$ and $\frac{m}{n}$ for which reconstruction is successful, are governed by the extents at which the suitable expander graphs exist. Expander graphs have either random or explicit construction (See for example [18] for explicit constructions of expander graphs). In either case, the resulting thresholds of [10] are very small (e.g. in the order of 10^{-5} for $\frac{m}{n} = 0.5$), due to the high expansion assumption, i.e. $\epsilon \leq 1/6$. This paper designs sparse measurement matrices that obtain a threshold of 0.01, which however holds only for non-negative signals. The null space characterization and its use in compressed sensing has also been studied in [11], from a quite different perspective. In that paper, the authors show that a so called “coherence” measure on the measurement matrix is related to the null space property. Unfortunately, when applied to sparse matrices, this result does not yield very sharp bounds for the recovery threshold either.

Finally it is worth mentioning that there are other related works that address the problem of sparse signal recovery in the case of sparse matrices, but for different recovery methods. One example can be found in [28], which provides a theoretical analysis (based on Density Evolution technique) of the message passing algorithm for recovering sparse signals measured by sparse measurement matrices. Another example is [29], which considers the same problem, but for nonnegative signals. A shortcoming of density evolution technique is that it can only determine asymptotic (infinite blocklength) results and relies on an asymptotic limit exchange. Furthermore, it should be clear that unlike these papers, we focus only on ℓ_1 minimization recovery. Finally we provide a strong bound, namely the bound for *all* nonnegative signals, which was not provided in [29].

Contributions of the current paper. We introduce sparse measurement matrices that result from adding perturbations to the adjacency matrices of expander graphs with a small critical expansion coefficient, hereby referred to as *minimal expanders*. We show that when ℓ_1 minimization is used to reconstruct nonnegative vectors, these constructions allow the recovery of sparse signals with way more nonzero

²We shall formally define (k, ϵ) expander graphs shortly.

entries- almost three orders of magnitude - than the existing theoretical results for sparse measurement matrices, namely the results of [10] and [11]. Please refer to Figure 3 for details. We provide theoretical upper bounds for the so called *weak* and *strong* recovery thresholds when ℓ_1 minimization is used. These bounds are very close (order-wise) to the bounds of Gaussian matrices for the nonnegative case (Section 5 and Figures 1 and 2). Furthermore, by carefully examining ℓ_1 minimization for sparse matrices, we deduce certain uniqueness results for the nonnegative solution of the linear equation when constant column sum matrices are used (see Section 3.1). We exploit this fact later to find faster alternatives to ℓ_1 minimization. In particular we present a novel recovery algorithm that directly leverages the minimal expansion property, and we prove that it is both optimal and robust to noise (Section 6).

One critical innovation of our work is that for expander graphs in the context of compressed sensing, we require a much smaller expansion coefficient in order to be effective. Throughout the literature, several sparse matrix constructions rely on adjacency matrices of expander graphs [5, 10, 12, 14, 15]. In these works, the technical arguments require very large expansion coefficients, in particular, $1 - \epsilon \geq 3/4$, in order to guarantee a large number of unique neighbors [22] to the expanding sets. Our analysis is the first to obtain error-correction results through small expansion ($\epsilon > 1/2$). In fact we show that the *minimal expansion* we use in our constructions is actually necessary for any matrix that works for compressive sensing (even when the non-negativity assumption is removed). See Section 4.2 for more details on this.

2 Problem Formulation

The goal in compressed sensing is to recover a sparse vector from a set of under-determined linear equations. In many real world applications the original data vector is nonnegative, which is the case that we will focus on in this paper. The original problem of compressed sensing for the nonnegative input vectors is the following:

$$\min_{\mathbf{Ax}=\mathbf{y}, \mathbf{x} \geq 0} \|\mathbf{x}\|_0, \tag{1}$$

where $\mathbf{A}^{m \times n}$ is the measurement matrix, $\mathbf{y}^{m \times 1}$ is called the observation vector, $\mathbf{x}^{n \times 1}$ is the unknown vector which is known to be k -sparse, i.e. to have only k nonzero entries, and where $\|\cdot\|_0$ is ℓ_0 norm, i.e. the number of nonzero entries of a given vector. (1) solves for the sparsest nonnegative solution in the constraint set $\{\mathbf{x} \mid \mathbf{Ax} = \mathbf{y}\}$. The typical situation is that $n > m > k$. Donoho and Tanner have shown in [4] that, for a class of matrices \mathbf{A} maintaining a so-called outwardly k -neighborly property and \mathbf{x} being

at most k -sparse, the solution to (1) is unique and can be recovered via the following linear programming problem:

$$\min_{\mathbf{Ax}=\mathbf{y}, \mathbf{x} \geq 0} \|\mathbf{x}\|_1. \quad (2)$$

They also show that i.i.d. Gaussian random $m \times n$ matrices with $m = n/2$ are outwardly $m/8$ -neighborly with high probability, and thus allow the recovery of $n/16$ -sparse vectors \mathbf{x} via linear programming. They further define a *weak* neighborly notion, based upon which they show that the same Gaussian random matrices will allow the recovery of *almost* all $0.279n$ -sparse nonnegative vectors \mathbf{x} via ℓ_1 optimization for sufficiently large n .

In this paper, we primarily seek the answer to a similar question when the measurement matrix \mathbf{A} is sparse and, in particular when \mathbf{A} is the adjacency matrix of an unbalanced bipartite graph with constant left degree d . The aim is to analyze the outwardly neighborly conditions for this class of matrices and come up with sparse structures that allow the recovery of vectors with sparsity proportional to the number of equations.

3 Analysis of ℓ_1 Minimization

We begin by stating an equivalent version of the outwardly neighborly condition which is in fact similar to the null space property that was mentioned in the introduction, but for the nonnegative case. Later we show that this has a much more mundane interpretation for the special case of regular bipartite graphs, namely a combinatorial null space condition. We leverage this condition to derive bounds for the successful recovery of sparse signals when particular sparse matrices are used.

3.1 Null Space and Uniqueness Conditions

As mentioned in the introduction, the success of ℓ_1 minimization in recovering sparse signals can be characterized by the null space condition. This condition has been previously stated for the general nondefinite sign signals in a couple of papers [17, 20, 21, 23]. We present a similar condition for the success of ℓ_1 minimization in recovering nonnegative signals. We also show that under the assumption that the measurement matrix has constant column sum, this condition is equivalent to the uniqueness of *any* nonnegative solution to the under-determined system of linear equations.

We present the first theorem in the same style as in [4].

Theorem 3.1. *Let \mathbf{A} be an $m \times n$ matrix and k be a positive integer. The following two statements are equivalent:*

1. *For every nonnegative vector \mathbf{x}_0 with at most k nonzero entries, \mathbf{x}_0 is the unique solution to (2) with $\mathbf{y} = \mathbf{A}\mathbf{x}_0$.*
2. *For every vector $\mathbf{w} \neq 0$ in the null space of \mathbf{A} , and every index set $S \subset \{1, 2, \dots, n\}$ with $|S| = k$ such that $\mathbf{w}_{S^c} \geq 0$, it holds that $\sum_{i=1}^n w_i > 0$.*

Here S^c is the complement set of S in $\{1, 2, \dots, n\}$ and \mathbf{w}_S denotes the sub-vector of \mathbf{w} constructed by those elements indexed in S . $|S|$ means the cardinality of the set S .

Theorem 3.1 is in fact the counterpart of Theorem 1 of [17] for nonnegative vectors. It gives a necessary and sufficient condition on the matrix \mathbf{A} , such that all the k -sparse vectors \mathbf{x}_0 can be recovered using (2). The condition is essentially that if a nonzero vector in the null space of \mathbf{A} happens to have $n - k$ nonnegative entries, then the sum of all its entries must be positive. We call this property the nonnegative null space property.

Proof. Suppose \mathbf{A} has the nonnegative null space property. We assume \mathbf{x}_0 is k -sparse and show that under the mentioned null space condition, the solution to (2) produces \mathbf{x}_0 . We denote by \mathbf{x}_1 the solution to (2). Let S be the support set of \mathbf{x}_0 . We can write:

$$\begin{aligned} \|\mathbf{x}_1\|_1 &= \|\mathbf{x}_0 + (\mathbf{x}_1 - \mathbf{x}_0)\|_1 \\ &= \sum_{i=1}^n (\mathbf{x}_0)_i + (\mathbf{x}_1 - \mathbf{x}_0)_i \end{aligned} \tag{3}$$

$$= \|\mathbf{x}_0\|_1 + \sum_{i=1}^n (\mathbf{x}_1 - \mathbf{x}_0)_i, \tag{4}$$

where $(\mathbf{x}_0)_i$ and $(\mathbf{x}_1 - \mathbf{x}_0)_i$ are the i^{th} entry of \mathbf{x}_0 and $\mathbf{x}_1 - \mathbf{x}_0$ respectively. The reason (3) and (4) are true is that \mathbf{x}_1 and \mathbf{x}_0 are both nonnegative vectors and their ℓ_1 -norm is simply the sum of their entries. Now, if \mathbf{x}_1 and \mathbf{x}_0 are not equal, since $\mathbf{x}_1 - \mathbf{x}_0$ is in the null space of \mathbf{A} and is nonnegative on S^c (because S is the support set of \mathbf{x}_0) we can write:

$$\sum_{i=1}^n (\mathbf{x}_1 - \mathbf{x}_0)_i > 0, \tag{5}$$

which implies

$$\|\mathbf{x}_1\|_1 > \|\mathbf{x}_0\|_1.$$

But we know that $\|\mathbf{x}_1\|_1 \leq \|\mathbf{x}_0\|_1$ from the construction. This means we should have $\mathbf{x}_1 = \mathbf{x}_0$.

Conversely, suppose there is a nonzero vector \mathbf{w} in the null space of \mathbf{A} and a subset $S \subset \{1, 2, \dots, n\}$ of size k with $\mathbf{w}_{S^c} \geq 0$ and $\sum_{i=1}^n w_i \leq 0$. We construct a nonnegative vector \mathbf{x}_0 supported on S , and show that there exists another nonnegative vector $\mathbf{x}_1 \neq \mathbf{x}_0$ such that $\mathbf{A}\mathbf{x}_0 = \mathbf{A}\mathbf{x}_1$ and $\|\mathbf{x}_1\|_1 \leq \|\mathbf{x}_0\|_1$. This means that \mathbf{x}_0 is not the unique solution to (2) with $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ and will complete the proof. For simplicity we may assume $S = \{1, 2, \dots, k\}$. We construct a nonnegative vector \mathbf{x}_0 supported on S that cannot be recovered via ℓ_1 minimization of (2). Without loss of generality we write

$$\mathbf{w} = (-\mathbf{w}_{S^-} \ \mathbf{w}_{S^+} \ \mathbf{w}_{S^c})^T, \tag{6}$$

where \mathbf{w}_{S^-} and \mathbf{w}_{S^+} are both nonnegative vectors. Now set

$$\mathbf{x}_0 = (\mathbf{w}_{S^-} \ \mathbf{w}_{S^+} \ 0)^T, \ \mathbf{x}_1 = (0 \ 2\mathbf{w}_{S^+} \ \mathbf{w}_{S^c})^T. \tag{7}$$

■

3.2 Null Space of Adjacency Matrices

In this paper we will be considering measurement matrices \mathbf{A} with two main properties: sparse and constant column sum. This class of matrices includes measurement matrices obtained from the adjacency matrices of regular left degree bipartite graphs (i.e. 0-1 matrices with a constant number of ones in each column), as well as the perturbed expanders introduced in section 4.1. For this class of matrices we actually show that the condition for the success of ℓ_1 recovery is simply the condition for there being a “unique” vector in the constraint set $\{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}_0, \mathbf{x} \geq 0\}$. To this end, we prove the following lemma and theorem.

Lemma 3.1. *Let $\mathbf{A}^{m \times n}$ be a matrix with constant column sum d . For any vector \mathbf{w} in the null space of \mathbf{A} , the following is true*

$$\sum_{i=1}^n w_i = 0. \tag{8}$$

Proof. Let $\mathbf{1} = (1, 1, \dots, 1)^T$ be the $1 \times m$ vector of all 1's. We have:

$$\mathbf{A}\mathbf{w} = 0 \Rightarrow \mathbf{1}^T \mathbf{A}\mathbf{w} = 0 \Rightarrow d \sum_{i=1}^n w_i = 0, \quad (9)$$

where d is the column sum of \mathbf{A} . ■

Theorem 3.2. *Let $\mathbf{A} \in \mathcal{R}^{m \times n}$ be a matrix with constant column sum. Then the following three statements are equivalent.*

1. *For all nonnegative k -sparse \mathbf{x}_0 with an arbitrary (a particular) support set S , it holds that $\{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}_0, \mathbf{x} \geq 0\} = \{\mathbf{x}_0\}$.*
2. *For every vector $\mathbf{w} \neq 0$ in the null space of \mathbf{A} , and every (a particular) index set $S \subset \{1, 2, \dots, n\}$ with $|S| = k$ such that $\mathbf{w}_{S^c} \geq 0$, it holds that $\sum_{i=1}^n w_i > 0$.*
3. *For every (a particular) subset $S \subset \{1, 2, \dots, n\}$ with $|S| = k$, there exists no vector $\mathbf{w} \neq 0$ in the null space of \mathbf{A} such that $\mathbf{w}_{S^c} \geq 0$.*

Theorems 3.1 and 3.2 show that for the class of matrices with constant column sum, the condition for the success of ℓ_1 recovery is simply the condition for there being a “unique” vector in the constraint set $\{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}_0, \mathbf{x} \geq 0\}$. In this case, *any* optimization problem, e.g. $\min_{\mathbf{x} \geq 0, \mathbf{A}\mathbf{x} = \mathbf{y}} \|\mathbf{x}\|_2$, would also recover the desired \mathbf{x}_0 .

Proof. First, we show that for any matrix \mathbf{A} , the statements 1 and 3 of Theorem 3.2 are equivalent. Suppose that condition 3 holds for a specific subset $S \subset \{1, 2, \dots, n\}$. Consider a nonnegative $n \times 1$ vector \mathbf{x}_0 supported on S . If there exists another nonnegative vector \mathbf{x}_1 with the property that $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_0$, then $\mathbf{x}_1 - \mathbf{x}_0$ would be a vector in the null space of \mathbf{A} which is also nonnegative on S^c , due to the nonnegativity of \mathbf{x}_1 and the fact that S is the support set of \mathbf{x}_0 . This contradicts the earlier assumption of condition 3.

The proof of the converse is also straightforward. Suppose that condition 1 holds for a specific subset S and all nonnegative vectors \mathbf{x}_0 supported on S . Let's say one can find a nonzero vector \mathbf{w} in the null space of \mathbf{A} with $\mathbf{w}_{S^c} \geq 0$. As in the proof of Theorem 3.1, we may write \mathbf{w} as

$$\mathbf{w} = (-\mathbf{w}_{S^-} \ \mathbf{w}_{S^+} \ \mathbf{w}_{S^c})^T, \quad (10)$$

where \mathbf{w}_{S^-} and \mathbf{w}_{S^+} are both nonnegative vectors. Now if

$$\mathbf{x}_0 = (\mathbf{w}_{S^-} \ \mathbf{w}_{S^+} \ 0)^T, \quad \mathbf{x}_1 = (0 \ 2\mathbf{w}_{S^+} \ \mathbf{w}_{S^c})^T, \quad (11)$$

then \mathbf{x}_0 and \mathbf{x}_1 are distinct nonzero vectors and belong to the set $\{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}_0, \mathbf{x} \geq 0\}$. This is a contradiction to the assumption we earlier made.

So far we have shown that for any matrix \mathbf{A} the two statements 1 and 3 are equivalent. Now we show that for matrices with constant column sum the two statements 2 and 3 are equivalent. We make use of Lemma 3.1, that for this special class of matrices with constant column sum, every vector in the null space has a zero sum of entries. Therefore, statement 2 can be true only if there is no \mathbf{w} in the null space of \mathbf{A} with $\mathbf{w}_{S^c} \geq 0$. Conversely if the condition in statement 3 holds, then there is no $\mathbf{w} \in \mathcal{N}(\mathbf{A}) \setminus \{0\}$ such that \mathbf{w}_{S^c} is nonnegative and therefore statement 2 is also true. ■

Corollary 1 (Corollary of Theorem 3.2). *The three conditions of Theorem 3.2 are equivalent to the following statement.*

- Every nonzero vector \mathbf{w} in the null space of \mathbf{A} has at least $k + 1$ negative entries.

Proof. Follows directly from the third statement of Theorem 3.2. ■

The results of this section show how the structure of the null space of the measurement matrix is related to the recoverability of sparse vectors. Thus, to achieve our primary goal of constructing optimal sparse measurement matrices, we need to find appropriate bipartite graphs, the adjacency matrices of which satisfy the nonnegative null space properties up to a maximal sparsity. We define expander graphs in the next section and introduce perturbed expanders. Using a linear algebraic view of expanders, we will be able to make a probabilistic analysis of the null space property for this class of sparse matrices.

4 Expander Graphs and Their Linear Algebraic View

Before proceeding, let us consider the following definitions, whose relation to the main topic will be shortly made apparent.

Definition 1. For a matrix $\mathbf{A}^{m \times n}$ we define the Complete Rank of \mathbf{A} (denoted by $\mathcal{H}(\mathbf{A})$) to be the maximum integer r_0 with the property that every r_0 columns of \mathbf{A} are linearly independent. In other words, $\mathcal{H}(\mathbf{A}) = \min_{\mathbf{w} \in \mathcal{N}(\mathbf{A}), \mathbf{w} \neq 0} (|\text{Supp}(\mathbf{w})| - 1)$, where $\text{Supp}(\mathbf{w})$ is the support set of \mathbf{w} .

This notion is also known in linear algebra as “Kruskal rank” (see [26]). It has also been given other names in the literature. The complete rank of a matrix \mathbf{A} is equivalent to the minimum Hamming distance (weight) of the null space of \mathbf{A} minus one [30]. It is also referred to as the “spark” of the matrix minus one [31].

Definition 2. *A left regular bipartite graph (X, Y, d) with X and Y as the set of left and right vertices ($|X| = n, |Y| = m$) and d as the regular left degree is called a (k, ϵ) -unbalanced expander if for every $S \subset X$ with $|S| \leq k$, the following holds: $|\Gamma(S)| \geq |S|d(1 - \epsilon)$, where $\Gamma(S)$ is the set of neighbors of S in Y .*

We also define generalized bipartite graphs that allow us to consider adjacency matrices with nonunitary entries.

Definition 3. *Let \mathbf{A} be a nonnegative $m \times n$ matrix. Consider the weighted bipartite graph $G = (X, Y, E, W)$ where $X = \{v_1, v_2, \dots, v_n\}$ and $Y = \{u_1, u_2, \dots, u_m\}$ are the sets of nodes, E is the set of edges, and W is the set of weights assigned to the edges, where G has the following property: for every nonzero entry \mathbf{A}_{ij} of the matrix \mathbf{A} , v_j and u_i are connected together with an edge e_{ij} of weight $w_{ij} = \mathbf{A}_{ij}$. For the zero entries of \mathbf{A} , there is no edge connecting the corresponding nodes in G . We call G the generalized bipartite graph of \mathbf{A} , and refer to \mathbf{A} as the generalized adjacency matrix of G .*

Note that the notions of “neighbor” of a vertex and “expansion” in a weighted bipartite graph are the same as in a unitary bipartite graph. So for instance, the neighbors of a node v in this graph are the set of nodes to which v is connected with an edge of nonzero weight. The following lemma connects the two notions of expansion of a (generalized) bipartite graph and the complete rank of its (generalized) adjacency matrix:

Lemma 4.1. *Let \mathbf{A} be a nonnegative matrix with exactly d nonzero entries in each column. The generalized bipartite graph of \mathbf{A} is a $(\mathcal{H}(\mathbf{A}), \frac{d-1}{d})$ expander.*

Proof. Let $G = (X, Y, E, W)$ be the generalized bipartite graph of \mathbf{A} . If $S \subset X$ with $|S| \leq \mathcal{H}(\mathbf{A})$ then the columns of \mathbf{A} corresponding to the elements of S are linearly independent. So the sub-matrix of \mathbf{A} produced by only those columns which correspond to S must be of full rank. Therefore, it must have at least $|S|$ nonzero rows, which is to say $|\Gamma(S)| \geq |S| = |S|d(1 - \frac{d-1}{d})$. ■

A direct consequence of the proof of this theorem is that:

$$\forall S \subseteq X, |\Gamma(S)| \geq \min(|S|, \mathcal{H}(\mathbf{A})) \quad (12)$$

The notion of complete rank is tightly related to the expansion property. It is also related to the null space characterization we are shooting for. The following theorem sheds some light on this fact.

Theorem 4.1. *Let \mathbf{A} be a nonnegative matrix with exactly d nonzero entries in each column. For every nonzero vector \mathbf{w} in the null space of \mathbf{A} , the number of negative elements of \mathbf{w} is at least $\frac{\mathcal{H}(\mathbf{A})}{d}$.*

Proof. Let X and Y be the sets of left and right vertices of the generalized bipartite graph of \mathbf{A} . Let $S_{\mathbf{w}}^+$ be the set of vertices in X corresponding to the positive elements of \mathbf{w} , and $S_{\mathbf{w}}^-$ be the set of vertices corresponding to the negative elements.³ Let $S_{\mathbf{w}} = S_{\mathbf{w}}^+ \cup S_{\mathbf{w}}^-$. Since $\mathbf{A}\mathbf{w} = 0$, we must have $\Gamma(S_{\mathbf{w}}^+) = \Gamma(S_{\mathbf{w}}^-) = \Gamma(S_{\mathbf{w}})$, since otherwise, there exists a vertex in Y connected to exactly one of the sets $S_{\mathbf{w}}^+$ or $S_{\mathbf{w}}^-$. Therefore, the coordinate of the vector $\mathbf{A}\mathbf{w}$ corresponding to this node will not sum up to zero. On the other hand, from the definition of $\mathcal{H}(\mathbf{A})$, we must have $|S_{\mathbf{w}}| > \mathcal{H}(\mathbf{A})$. The number of edges emanating from $S_{\mathbf{w}}^-$ is $d|S_{\mathbf{w}}^-|$, which is at least as large as the number of its neighbors $|\Gamma(S_{\mathbf{w}}^-)|$. Hence:

$$d|S_{\mathbf{w}}^-| \geq |\Gamma(S_{\mathbf{w}}^-)| = |\Gamma(S_{\mathbf{w}})| \geq \mathcal{H}(\mathbf{A}),$$

where the last inequality is a consequence of (12). ■

We now turn to the task of constructing adjacency matrices with complete rank proportional to the dimension n . Throughout this paper, all the thresholds that we achieve are asymptotic, i.e., they hold for the regime of very large n and m .

4.1 Perturbed Expanders

When n and $m = \beta n$ are large, we are interested in constructing 0-1 matrices $\mathbf{A}^{m \times n}$ with d (constant) 1's in each column such that $\mathcal{H}(\mathbf{A})$ is proportional to n . Furthermore, the maximum achievable value of $\frac{\mathcal{H}(\mathbf{A})}{nd}$ is critical. This is a very difficult question to address. However, it turns out to be much easier if we allow for a *small* perturbation of the nonzero entries of \mathbf{A} .

³We interchangeably use S and its variations to denote a set of vertices or a support set of a vector.

Lemma 4.2. *For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ which is the adjacency matrix of a bipartite left d -regular graph, if in the submatrix formed by any r_0 columns of \mathbf{A} , every $r \leq r_0$ columns have at least r nonzero rows (namely by Hall's theorem, there exists a perfect matching between some r_0 rows and the r_0 columns of the submatrix), then it is possible to perturb the nonzero entries of \mathbf{A} and obtain another nonnegative matrix $\tilde{\mathbf{A}}$ through this procedure, with $\mathcal{H}(\tilde{\mathbf{A}}) \geq r_0$. Furthermore, the perturbations can be done in a way that the sum of each column remains a constant d , and all perturbations are rational numbers representable by a finite number of bits.*

Proof. We add a random set of perturbations $\Delta = \{\delta_{i,j} \mid 1 \leq i \leq m, 1 \leq j \leq n, \mathbf{A}_{i,j} \neq 0\}$ to the non-zero elements of \mathbf{A} , while leaving the zero elements of \mathbf{A} intact. We denote the perturbed matrix by $\tilde{\mathbf{A}}$. The way the random perturbations are generated is as follows. Suppose without loss of generality that we scale the matrix up by a prime integer q before adding perturbations, so that the nonzero entries of \mathbf{A} are equal to q , and suppose that $\delta_{i,j}$'s are integers in $\{0, 1, \dots, q-1\}$. For each nonzero entry $\mathbf{A}_{i,j}$ of \mathbf{A} , we independently choose $\delta_{i,j}$ uniformly at random from the set $\{0, 1, \dots, q\}$. We look at the submatrix \mathbf{A}_{sub} formed by arbitrary r_0 distinct columns of \mathbf{A} . Without loss of generality, we index these columns by the set $N_{sub} = \{1, \dots, r_0\}$. Then there exists a perfect matching between these r_0 columns and some r_0 rows, which we denote by the set $M_{sub} = \{i_1, i_2, \dots, i_{r_0}\}$ accordingly. First we bound the probability that Δ makes the submatrix $\tilde{\mathbf{A}}_{sub}$ have rank smaller than r_0 , i.e. $\mathbb{P}\{rank(\tilde{\mathbf{A}}_{sub}) < r_0\}$. In order for the submatrix $\tilde{\mathbf{A}}_{sub}$ to have rank smaller than r_0 , the determinant of the square matrix $\tilde{\mathbf{A}}_{M_{sub}, N_{sub}}$ must be zero, namely $\det(\tilde{\mathbf{A}}_{M_{sub}, N_{sub}}) = 0$. By definition, $\det(\tilde{\mathbf{A}}_{M_{sub}, N_{sub}})$ is a polynomial over the variables of Δ , say $p(\Delta)$, which contains a product term $\delta_{1,i_1} \delta_{2,i_2} \dots \delta_{r_0,i_{r_0}}$ which certainly has a nonzero coefficient (actually its coefficient is either 1 or -1). Therefore, invoking Schwartz-Zippel lemma we can assert that $\mathbb{P}\{p(\Delta) = 0\} \leq \frac{r_0}{q}$. Furthermore, the number of ways the submatrix \mathbf{A}_{sub} can be chosen is $\binom{n}{r_0}$. Applying a union bound over all possible choices of \mathbf{A}_{sub} , we can write:

$$\mathbb{P}\{\mathcal{H}(\tilde{\mathbf{A}}) \leq r_0\} \leq \binom{n}{r_0} \frac{r_0}{q} \quad (13)$$

For given n and r_0 , we can choose a finite q large enough such that the right hand side of (13) is sufficiently small. Therefore, there exists a choice of perturbations $\delta_{i,j}$ so that the resulting $\tilde{\mathbf{A}}$ satisfies $\mathcal{H}(\tilde{\mathbf{A}}) \geq r_0$. Furthermore, after scaling down the perturbed matrix by q , each perturbation $\delta_{i,j}$ is a rational number of the form $\frac{t}{q}$, $0 \leq t < q$, and is therefore representable by a finite number of bits.

■

It is worth noticing that, after modifying \mathbf{A} based on perturbations of Lemma 4.2, Theorems 3.1, 3.2 and 4.1, and Lemmas 3.1 and 4.1 all continue to hold for this class of matrices $\tilde{\mathbf{A}}$. The reason is as follows. First of all, note that Lemma 3.1 and Theorem 3.2 require only that \mathbf{A} be constant column sum, which is true for $\tilde{\mathbf{A}}$. Theorem 3.1 assumes no restriction on the matrix. Finally, Lemma 4.1 and Theorem 4.1 are valid, since they hold for nonnegative matrices with a constant number of nonzero entries in each column, and $\tilde{\mathbf{A}}$ is such a matrix.

The conclusion of this section so far is that if one starts off with the adjacency matrix \mathbf{A} of a regular bipartite graph and perturb its nonzero entries to obtain a nonzero constant column sum matrix $\tilde{\mathbf{A}}$ with $\mathcal{H}(\tilde{\mathbf{A}}) \geq r_0$, then the following guarantee exists: ℓ_1 minimization perfectly recovers a $\lceil \frac{r_0}{d} - 1 \rceil$ -sparse nonnegative vector \mathbf{x}_0 from the measurements $\tilde{\mathbf{A}}\mathbf{x}_0$. Our goal now becomes constructing $(r_0, \frac{d-1}{d})$ expanders with the ratio $\frac{r_0}{nd}$ as large as possible. In Section 5, we use a probabilistic method to show that the desired $(r_0 = \mu n, \frac{d-1}{d})$ expanders exist and provide thresholds for $\frac{\mu}{d}$. Before continuing, note that we are using a $1 - \epsilon \geq \frac{1}{d}$ expansion coefficient for perfect recovery, which is very small compared to other schemes that use expanders (see, e.g., [12, 5, 6, 10, 13, 14]) and require expansion coefficients at least larger than $1 - \epsilon \geq \frac{3}{4}$. $1 - \epsilon \geq \frac{1}{d}$ is indeed the *critical* expansion coefficient. We shortly digress in a subsection to discuss this a little further.

4.2 Necessity of Expansion for Compressive Sensing

Let's first consider the following definition:

Definition 4. Let $G = (X, Y, E, W)$ be a weighted bipartite graph and $S \subset X$ be a set of nodes in X . Let M be a subset of edges of G . M is called a *partial perfect matching saturating S* , if:

- $|M| = |S|$.
- each node of S is incident to exactly one edge in M .
- no node in Y is incident to more than one edge in M .

We prove that any *good* sparse measurement matrix corresponds to an expander graph with a critical expansion coefficient.

Theorem 4.2. Let \mathbf{A} be a $m \times n$ nonnegative matrix, with d nonzero entries per column, and assume that \mathbf{A} allows the recovery of all r_0 -sparse vectors. Then the generalized bipartite graph of \mathbf{A} is an $(r_0, 1 - \frac{1}{d})$ expander.

Proof. The statement holds for any recovery algorithm. In fact we show that if the generalized bipartite graph of \mathbf{A} is not an $(r_0, 1 - \frac{1}{d})$ expander, then even a stronger recovery algorithm that magically knows the support of the vector, fails to recover some r_0 -sparse vectors. Assume that the bipartite graph is not a $(r_0, 1 - \frac{1}{d})$ expander, i.e. there exists a set of $r \leq r_0$ columns that is adjacent to $r - 1$ (or fewer) rows. Therefore the rank of the submatrix corresponding to these r columns must be strictly smaller than r regardless of what the nonzero entries are. By selecting an adversary sparse signal supported exactly on these r columns, we see it is impossible for any algorithm to recover it, even if the support is known, since there is a rank loss in the corresponding measurement submatrix.

■

5 Recovery Thresholds of Compressive Sensing for Minimal Expanders

In summary of Section 4, if one can construct a $(r_0, 1 - \frac{1}{d})$ expander graph with left degree d , then from the adjacency matrix of this expander, it is possible to obtain a measurement matrix that allows the recovery of *every* $\frac{r_0}{d}$ sparse nonnegative signals using ℓ_1 minimization. The explicit relationship between the ratio $\frac{m}{n}$ and the relative size, $\frac{r_0}{dn}$, of a recoverable set is not yet clear. In this section, by addressing the question of existence of appropriate expander graphs, we find this explicit relationship known as the strong recovery threshold. We further derive a weak recovery threshold that is concerned with the recovery of *almost all* sparse vectors of a certain sparsity.

5.1 Strong Bound

For fixed values of $n > m > r_0$ and d we are interested in the existence of $(r_0, \epsilon = \frac{d-1}{d})$ expanders with constant left degree d . There are a few previous works that address the construction of expanders (random or deterministic), and try to find the relationships between their parameters. In [24] for instance, it has been shown that for any value of $\beta = \frac{m}{n}$ and $0 \leq \epsilon \leq 1$, there exists a left degree d bipartite $(\mu n, \epsilon)$ expander for some $0 \leq \mu \leq 1$ and some constant (not growing with n) d . Also, an explicit construction of constant regular left degree lossless (with $1 - \epsilon$ arbitrarily close to 1) expanders is given in [25]. As a consequence of the results of the latter, for any fixed ϵ it is possible to explicitly construct a (k_{max}, ϵ) expander with $k_{max} = O(\frac{m}{d\epsilon})$, and $d = O(\frac{\log(\beta)}{\epsilon})$. The main reason we cannot use these results directly here is that the relationship between the expansion factor $(1 - \epsilon)$ and the left degree of the graph (d) is not clear in any of the previous works (even order wise). Besides, in most of the previous arguments, the

relationship between the size of expansion set k , the left set size m and the factor ϵ is expressed in terms of order functions, and the explicit constants are omitted. Our attempt here is to derive those constants explicitly.

We use the standard first moment method argument to prove the existence of our expanders of interest, namely $(r_0, \epsilon = \frac{d-1}{d})$ expanders, for appropriate n , m , r_0 and d . The main result is given below, while the complete proof can be found in Appendix A.

Theorem 5.1. *For sufficiently large n , with $m = \beta n$ and $r_0 = \mu n$, there exists a bipartite graph with n left vertices and m right vertices, which is a $(r_0, \frac{d-1}{d})$ expander, if*

$$d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log(\frac{\beta}{\mu})}, \quad (14)$$

where $H(\cdot)$ is the Shannon entropy function defined as $H(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$.

More important is the question of how big the ratio $\frac{\mu}{d}$ can be, since we earlier proved that we can recover vectors with sparsity up to $\frac{r_0}{d} = \frac{\mu}{d}n$. A combination of the previous derivations and Theorem 5.1 directly implies the following theorem, stating the strong sparsity threshold for sparse measurement matrices.

Theorem 5.2 (Strong Threshold). *For a fixed $\beta = \frac{m}{n}$, let d^* and μ^* be the solutions of the following optimization program*

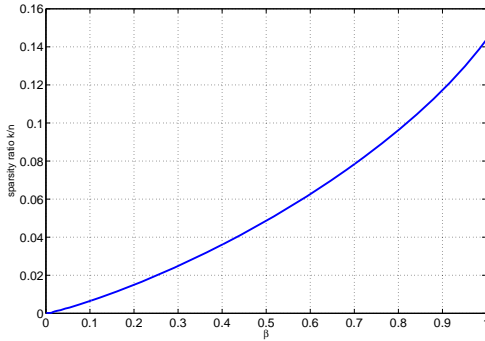
$$\begin{aligned} & \max \frac{\mu}{d} \\ & \text{s.t. } \left\{ \begin{array}{l} 0 \leq \mu \leq 1, \quad d \in \mathbb{Z}^+, \quad d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log(\frac{\beta}{\mu})} \end{array} \right\} \end{aligned}$$

For sufficiently large n , there exists a sparse measurement matrix \mathbf{A} with d^* nonzero entries in each column, so that every $\frac{\mu^* n}{d^*}$ -sparse nonnegative vector \mathbf{x} can be recovered from $\mathbf{A}\mathbf{x}$ using ℓ_1 minimization.

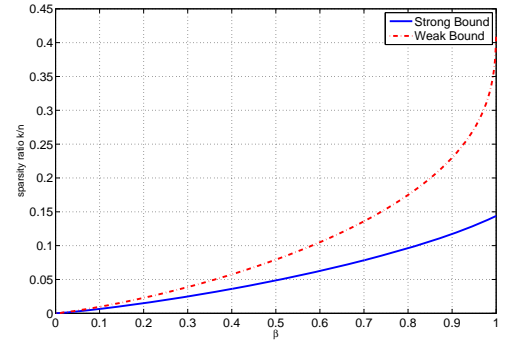
Figure 1a illustrates the strong recovery threshold for different values of β , derived from Theorem 5.2.

5.2 Weak Bound

We are now interested in deriving conditions for recovering a specific support set S of size $k = \alpha n$, rather than obtaining a worst case bound for matrices that work for all support sets. Recall that $m = \beta n$, left degree is d , and define $\gamma_1 := (1 - e^{-d\frac{\alpha}{\beta}})\beta$.



(a) Strong bound.



(b) Weak vs strong bound.

Figure 1: Recoverable sparsity size, weak achievable bound of Section 5.2 and the strong achievable threshold of (14). β is the ratio $\frac{m}{n}$.

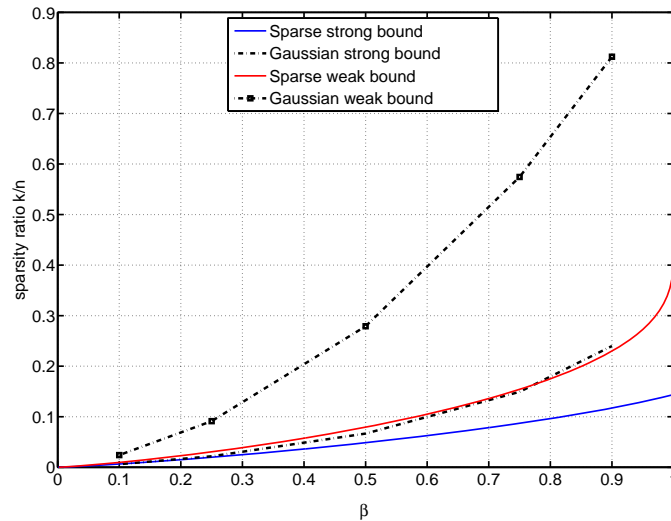


Figure 2: Comparison of weak and strong bounds for dense i.i.d. Gaussian matrices (and nonnegative signals) from [4] with those of the current paper for sparse matrices. β here is equal to $\frac{m}{n}$

Theorem 5.3 (Weak Threshold). *Define the function*

$$F(\rho_1, \rho_2) := \alpha H\left(\frac{\rho_1}{\alpha}\right) + (1 - \alpha)H\left(\frac{\rho_2}{1 - \alpha}\right) + \beta H\left(\frac{\rho_1 + \rho_2}{\beta}\right) + d(\rho_1 + \rho_2) \log\left(\frac{\rho_1 + \rho_2}{\beta}\right). \quad (15)$$

For every α such that $F(\rho_1, \rho_2) < 0$ for every ρ_1, ρ_2 that satisfy $\rho_1 < \alpha, \rho_2 < 1 - \alpha, \rho_1 + \rho_2 < \gamma_1$, a randomly selected subset of size $k = \alpha n$ is recoverable using ℓ_1 minimization of (2) from a random perturbed matrix $\tilde{\mathbf{A}}$ with probability $1 - o(1)$.

The bound that results from Theorem 5.3 is plotted in Figure 1b and has been compared to the strong threshold previously achieved. Also, the comparison of these bounds with those of dense Gaussian

i.i.d. matrices for the nonnegative case that were obtained in the paper [4] is illustrated in Figure 2. Note that compared with the performance bounds achievable for Gaussian dense matrices, the provable theoretical performance bounds from our derivations is generally smaller, but they are getting closer to the performance bounds achievable for Gaussian dense matrices (See Figure 2) and represent a step towards closing the gap between the provable bounds for dense and sparse measurement matrices. Finally, to highlight the importance of our achieved bounds, we have compared the strong threshold of this paper with the thresholds of [5] that also uses expander graphs with $\epsilon \leq 1/6$ in Figure 3. Ostensibly, the use of minimal expanders has increased the theoretical recovery thresholds by almost three orders of magnitude.

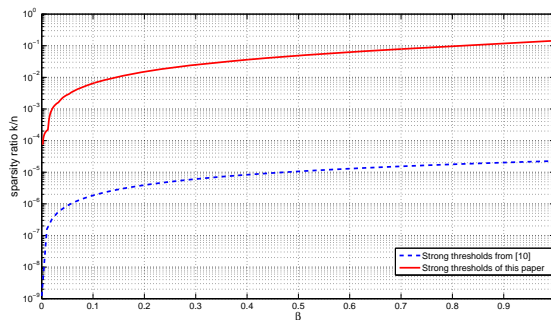


Figure 3: Comparison of size of recoverable sparsity (strong bound) of this paper with those from [5]. $\beta = \frac{m}{n}$.

The full proof of Theorem 5.3 is given in Appendix B. The key argument is a matching condition for the recoverability of vectors supported on a specific subset S . The condition involves looking at the two-hop graph from S and checking if all sets of size up to $|\Gamma(S)| + 1$ is saturated by a partial perfect matching:

Lemma 5.1. *Given a set S consider $\Gamma(S)$ and denote $S_2 = \Gamma(\Gamma(S)) \setminus S$. Let the bipartite two-hop graph of S be denoted by $B_S = (S \cup S_2, \Gamma(S \cup S_2))$. If every subset $S' \subset S \cup S_2$ of size $|S'| \leq |\Gamma(S)| + 1$ has minimal expansion: $|\Gamma(S')| \geq |S'|$ then there is a perturbation of nonzero entries of \mathbf{A} resulting in the matrix $\tilde{\mathbf{A}}$ so that any nonnegative vector \mathbf{x}_0 supported on S can be recovered from $\mathbf{y} = \tilde{\mathbf{A}}\mathbf{x}_0$ using the optimization formulation (2).*

Proof. Consider the two-hop bipartite graph of S and let $C = (S \cup S_2)^c$ denote the remainder of the nodes in X . Further let \mathbf{A}_S denote the submatrix of \mathbf{A} corresponding to B_S . By Hall's theorem since every subset of $S \cup S_2$ of size up to $|\Gamma(S)| + 1$ has expansion equal to its size, it must also be saturated by a partial perfect matching. In other words, B_S corresponds to a $(|\Gamma(S) + 1|, 1 - \frac{1}{d})$ expander. Therefore by Lemma 4.2 it is possible to perturb its nonzero entries to obtain a matrix $\tilde{\mathbf{A}}_S$ with $\mathcal{H}(\tilde{\mathbf{A}}_S) \geq |\Gamma(S)|$.

To show that a set S can be recovered, it suffices to show that every nonzero vector \mathbf{w} in the nullspace of $\tilde{\mathbf{A}}$ cannot have all its negative components in S . Assume otherwise: that some \mathbf{w} has indeed all its negative support $S_{\mathbf{w}}^- \subseteq S$. Observe now that C cannot contain any of the positive support of \mathbf{w} , because every equation that is adjacent to a positive element must also be adjacent to a negative elements (since the matrix coefficients are nonnegative) and $\Gamma(S_{\mathbf{w}}^-)$ does not intersect $\Gamma(C)$. Therefore the whole support of \mathbf{w} must be contained in $S \cup S_2$.

Now we can show that $|S_{\mathbf{w}}| \leq |\Gamma(S)|$. Assume otherwise, that $|S_{\mathbf{w}}| > |\Gamma(S)|$. Then we could select a subset of $K \subseteq S_{\mathbf{w}}$ such that $|K| = |\Gamma(S)| + 1$. This set K satisfies our assumption and is contained in B_S and therefore must have the minimal expansion $|\Gamma(K)| \geq |K| = |\Gamma(S)| + 1$. But since $\Gamma(S_{\mathbf{w}}) = \Gamma(S_{\mathbf{w}}^-) \subseteq \Gamma(S)$ and $K \subseteq S_{\mathbf{w}}$ (recall that $S_{\mathbf{w}}^- \subseteq S$ by assumption), it must hold that $|\Gamma(K)| \leq |\Gamma(S)|$, which contradicts the minimal expansion inequality.

Therefore, $|S_{\mathbf{w}}|$ must be saturated by a partial perfect matching which means that we can find a full rank submatrix $\mathbf{A}_{\mathbf{w}}$ (corresponding to that partial matching) such that $\mathbf{A}_{\mathbf{w}}\mathbf{w}_S = 0$ (where \mathbf{w}_S means the vector \mathbf{w} restricted to its support). Since $\mathbf{A}_{\mathbf{w}}$ is full rank, \mathbf{w} must be the all zeros vector which contradicts the assumption that $S_{\mathbf{w}}^-$ can be contained in S . ■

6 Fast Algorithm

We now describe a fast algorithm for the recovery of sparse nonnegative vectors from noiseless or noisy measurements. This algorithm relies on the minimal expansion we described in section 4.2. We employ a $(kd + 1, 1 - \frac{1}{d})$ expander and perturb it as Lemma (4.2) to obtain a sparse nonnegative matrix $\tilde{\mathbf{A}}$ with $\mathcal{H}(\tilde{\mathbf{A}}) \geq kd + 1$. The algorithm has two variations, one that works specifically for k -sparse signals with no measurement noise, and one that also accounts for additive noise. When measurement noise is present, the observation vector is given by the following equation.

$$\mathbf{y} = \tilde{\mathbf{A}}\mathbf{x} + \mathbf{v}. \tag{16}$$

In either case, the sparsity k of the unknown signal is given to the algorithm.

Algorithm 1. *Reverse Expansion Recovery*

1. Sort elements of \mathbf{y} in terms of absolute value, pick the smallest $m - kd$ of them, and denote them by \mathbf{y}_1 . Also denote by T_1 the index set of the elements of \mathbf{y}_1 in \mathbf{y} , and by T_2 its complement set.

Without loss of generality, assume that $\mathbf{y} = (\mathbf{y}_1 \ \mathbf{y}_2)^T$.

2. Locate in X the neighbors of the set of nodes in Y corresponding to T_1 , name the set S_1 and name the set of their complement nodes in X by S_2 .
3. Identify the sub-matrix of $\tilde{\mathbf{A}}$ that represents the edges emanating from S_2 to T_2 . Call this sub-matrix $\tilde{\mathbf{A}}_2$. Columns of $\tilde{\mathbf{A}}_2$ correspond to nodes in S_2 , and its rows correspond to the nodes in T_2 .
4. Set $\hat{\mathbf{x}}_{S_1} = 0$, and

$$\hat{\mathbf{x}}_{S_2} = \begin{cases} \tilde{\mathbf{A}}_2^\dagger \mathbf{y}_2 & \text{If noiseless} \\ \operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^{|S_2| \times 1}} \|\mathbf{A}_2 \mathbf{z} - \mathbf{y}_2\|_p & \text{Otherwise} \end{cases} \quad (17)$$

where $\mathbf{A}^\dagger = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t$ and $p \geq 1$. Declare $\hat{\mathbf{x}}$ as the output.

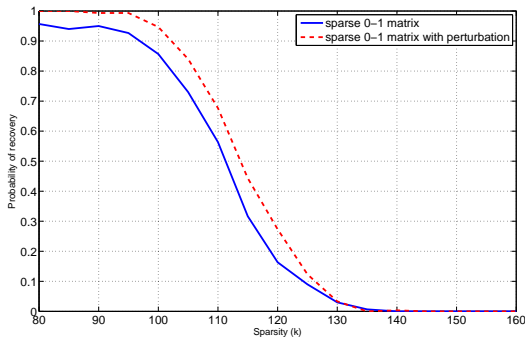
Note that in the noiseless case, Algorithm 1 transforms the original under-determined system of equations to an over-determined subset of equations. Appendix C provides theoretical justification for the effectiveness of this algorithm. Also, in next theorem we claim that Algorithm 1 is robust to measurement noise.

Theorem 6.1. *If \mathbf{A} is the adjacency matrix of a $(2k, \epsilon)$ expander with $\epsilon < 0.5$, \mathbf{x} is a k -sparse nonnegative vector and $\hat{\mathbf{x}}$ is the output of Algorithm 1 with $p = 1$ for $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$, then $\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{7-4\epsilon}{1-2\epsilon} \|\mathbf{v}\|_1$.*

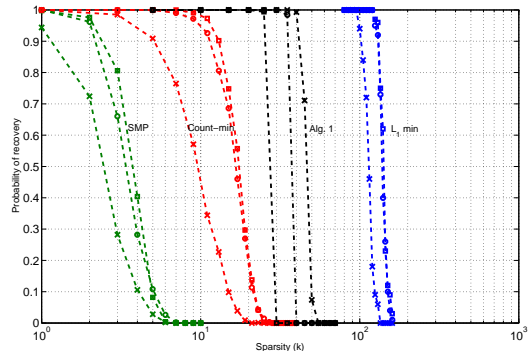
Proof. Given in Appendix E. ■

7 Experimental Evaluation

We generated a random $m \times n$ matrix \mathbf{A} with $n = 2m = 500$, and $d = 3$ 1's in each column. We then multiplied random sparse vectors with different sparsity levels by \mathbf{A} , and tried recovering them via the linear program (2). Next, we added a random set of perturbations to the nonzero entries of \mathbf{A} while keeping it nonnegative and constant column sum (one way to do that is by adding a uniformly random number in $[-1, 1]$ to each 1 of \mathbf{A} , and then normalizing each column) and applied the same sparse vectors to compare the recovery percentages in the two cases. This process was repeated for a few generations of \mathbf{A} and the best of the improvements we obtained is illustrated in Figure 4a.



(a) ℓ_1 minimization for sparse 0-1 matrices with and without perturbations.



(b) Comparison of algorithms.

Figure 4: (a) Probability of successful recovery of ℓ_1 minimization for a random 0 – 1 sparse matrix of size 250×500 with $d = 3$ ones in each column, and the same probability when the matrix is randomly perturbed in the nonzero entries. (b) Comparison of ℓ_1 minimization nonnegative recovery, Algorithm 1, count-min algorithm of [7] and SMP algorithm of [8] for sparse 0 – 1 measurement matrices with d ones in each column. \times : $d = 3$, \circ : $d = 6$, \square : $d = 9$. Blue: ℓ_1 minimization, Black: Algorithm 1, Red: Count-min, Green: SMP.

In Figure 4b we have plotted the recovery percentage of Algorithm 1 for a random 0 – 1 sparse matrix \mathbf{A} of size 250×500 with d ones in each column. We have compared the performance with the ℓ_1 minimization method, the count-min algorithm of [7] and the Sparse Matching Pursuit (SMP) method of [8], all specific to positive signals. Note that for the count-min algorithm, the measurement matrix must have other properties in addition to the constant number of 1s in each column. Please see [8] for more details. Although the deterministic theoretical bounds for the proposed algorithm and the ℓ_1 minimization are the same, as observed in Figure 4b, in practice ℓ_1 minimization is more effective for less sparse signals. However Algorithm 1 is considerably faster than linear programming and easier to implement.

In general, the complexity of Algorithm 1 is $O(nk^2)$ which, when k is proportional to n , is similar to linear programming's $O(n^3)$. However the constants are much smaller, which is of practical advantage. Furthermore, taking advantage of fast matrix inversion algorithms for very sparse matrices, Algorithm 1 can be performed in dramatically fewer operations. Figure 5 shows the Signal to Error Ratio as a function of Signal to Noise Ratio when Algorithm 1 with $p = 2$ has been used to recover noisy observations. Assuming that the output of the algorithm is $\hat{\mathbf{x}}$, Signal to Noise Ratio (SNR) and Signal to Error Ratio (SER) functions are defined as

$$SNR = 10 \log \frac{\|\mathbf{Ax}\|_2^2}{\|\mathbf{v}\|_2^2}, \quad SER = 10 \log \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}$$

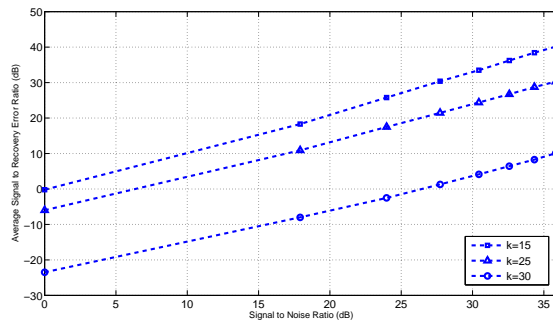


Figure 5: Simulation results for Algorithm 1, noisy case; Signal to Error Ratio Vs. SNR.

Measurement matrices are the same as before.

8 Conclusion

In this paper we considered the recovery of a nonnegative sparse vector using a sparse measurement matrix in the compressed sensing framework. We used the perturbed adjacency matrix of a bipartite expander graph to construct the sparse measurement matrix and proposed a novel fast algorithm. We computed recovery thresholds and showed that for measurement matrices with nonnegative entries and constant column sum the constraint set $\{\mathbf{x} | \mathbf{x} \geq 0, \mathbf{A}\mathbf{x} = \mathbf{y}\}$ is a singleton set, whenever ℓ_1 optimization is successful (which also means that any other nontrivial optimization scheme which can examine the feasible set would be successful). Finally, determining whether the matrices constructed satisfy an RIP-2 property, and constructing 0-1 matrices that have complete rank proportional to n are open problems that may be worthy of further scrutiny.

References

- [1] D. Donoho, “High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension,” *Discrete and Computational Geometry*, 102(27), pp. 617-652 2006, Springer.
- [2] D. Donoho and J. Tanner, “Neighborliness of randomly-projected simplices in high dimensions,” *Proc. National Academy of Sciences*, 102(27), pp. 9452-9457, 2005.
- [3] E. Candès, I. Romberg and T. Tao, “Robust Uncertainty Principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. on Info. Theory*, 52(2) pp. 489-509, Feb. 2006.

- [4] D. Donoho and J. Tanner, “Sparse nonnegative solution of underdetermined linear equations by linear programming,” *Proceedings of the National Academy of Sciences*, Vol. 102(27) (2005) 9446-9451.
- [5] R. Berinde, A. Gilbert, P. Indyk, H. Karloff and M. Strauss, “Combining geometry and combinatorics: a unified approach to sparse signal recovery,” 46th Annual Allerton Conference, September 23-26, 2008.
- [6] P. Indyk and M. Ruzic, “Near-optimal sparse recovery in the L_1 norm”, *Proceedings of 49th Symposium on Foundations of Computer Science (FOCS)*, 2008.
- [7] G. Cormode and S. Muthukrishnan, “Improved data stream summaries: The count-min sketch and its applications,” *FSTTCS*, 2004.
- [8] R. Berinde, P. Indyk and M. Ruzic, “Practical near-optimal sparse recovery in the L_1 norm,” *Proceedings of 46th Annual Allerton Conference 2008*.
- [9] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. on Info. Theory*, 53(12) pp. 4655-4666, December 2007.
- [10] R. Berinde and P. Indyk, “Sparse recovery using sparse matrices,” *Computer Science and Artificial Intelligence Laboratory Technical Report, MIT-CSAIL-TR-2008-001*, January 10, 2008.
- [11] A. M. Bruckstein, M. Elad, and M. Zibulevsky, “A nonnegative and sparse enough solution of an underdetermined linear system of equations is unique”, *IEEE Transactions on Information Theory*, Vol. 54, No. 11, Pages 4813-4820, November 2008.
- [12] W. Xu and B. Hassibi, “Efficient compressive sensing with deterministic guarantees using expander graphs”, *IEEE Information Theory Workshop, Lake Tahoe*, September 2007.
- [13] W. Xu and B. Hassibi, “Further results on performance analysis for compressive sensing using expander graphs,” *Proceedings of the Forty-First Asilomar Conference on Signals, Systems and Computers*, pp. 621-625, 2007 .
- [14] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, “Efficient compressed sensing using high-quality expander graphs,” *IEEE Trans. on Info. Theory*, Vol. 55, 9, pp.4299-4308, 2009.
- [15] P. Indyk, “Explicit constructions for compressed sensing of sparse signals,” *Proceedings of SODA*, 2008.
- [16] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, “Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays,” *IEEE Journal of Selected Topics In Signal Proc.*, Vol.2, Issue 3, pp. 275-285 06-2008.

- [17] M.Stojnic, W. Xu, and B. Hassibi, "Compressed sensing - probabilistic analysis of a null-space characterization," IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP 2008
- [18] V. Guruswami, J.R. Lee, A. Wigderson, "Euclidean sections of L_1^n with sublinear randomness and error-correction over the reals," Proceedings of RANDOM 2008.
- [19] V. Guruswami, J.R. Lee, A. Razborov, "Almost Euclidean subspaces of L_1^n via expander codes," Proceedings of SODA, 2008.
- [20] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases" IEEE Transactions on Information Theory 49(6): 1579-1581 (2003).
- [21] N. Linial and I. Novik, "How neighborly can a centrally symmetric polytope be?" Discrete and Computational Geometry,, pp. 273-281, 2006.
- [22] M. Sipser and D.A. Spielman, "Expander codes," IEEE Transactions on Information Theory, Vol 42, No. 6, pp. 1710-1722. 1996.
- [23] Y. Zhang, "When is missing data recoverable?" Rice CAAM Department Technical Report TR05-15, 2005, available online at <http://www.dsp.ece.rice.edu/cs/>.
- [24] D. Burshtein and G. Miller, "Expander graph arguments for message passing algorithms," IEEE Transactions on Information Theory, volume 47, pp. 782-790, Feb. 2001.
- [25] M. Capalbo, O. Reingold, S. Vadham and A. Wigderson, "Randomness conductors and constant-degree expanders," Proceeding of the 17th IEEE Annual Conference on Computational Complexity (CCC'02).
- [26] J.B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," Linear Algebra Appl. 18 (1977).
- [27] W. Dai, M. Sheikh, O. Milenkovic, and R. Baraniuk, "Compressive sensing DNA microarrays," EURASIP Journal on Bioinformatics and Systems Biology, 2009, avaiialble online at <http://www-ece.rice.edu>.
- [28] F. Zhang and H. D. Pfister, "On the iterative decoding of high-rate LDPC codes with applications in compressed sensing," Submitted to IEEE Trans. on Inform. Theory, 2009. available on Arxiv.
- [29] Y. Lu, A. Montanari and B. Parabhakar, "Counter Braids: Asymptotic optimality of the message passing decoding algorithm," Proceedings of Allerton Conference, Sep. 2008.
- [30] J. Wolf, "Redundancy, the discrete Fourier transform and impulse noise cancellation," IEEE Transactions on Communications, Vol. 31, No. 3, (1983)

[31] S. Jokar and Volker Mehrmann, “Sparse representation of solutions of Kronecker product of systems,” available online on Arxiv.

[32] T. M. Cover and J. A. Thomas, “Elements of Information Theory,” 2006, Wiley.

A Proof of Theorem 5.1

Assuming that we generate a random matrix \mathbf{A} by placing d ones in each column uniformly at random and the rest of the entries zero, it suffices to show that the probability that \mathbf{A} has the desired expansion property is positive. For $1 \leq i_1 < i_2 < \dots < i_r \leq m$, we denote by E_{i_1, i_2, \dots, i_r} the event that the columns of \mathbf{A} corresponding to the numbers i_1, i_2, \dots, i_r have at least $m - r + 1$ entire 0 rows (rows that do not have a single nonzero element in the columns $\mathbf{A}_{i_1}, \mathbf{A}_{i_2}, \dots, \mathbf{A}_{i_r}$). In other words E_{i_1, i_2, \dots, i_r} is the event that the set of nodes $\{i_1, i_2, \dots, i_r\}$ in X contracts in Y .

$$\mathbb{P}[\mathbf{A} \text{ is a } (r_0, \frac{d-1}{d})\text{-Exp.}] = 1 - \mathbb{P}\left[\bigcup_{d \leq r \leq r_0, 1 \leq i_1 < i_2 < \dots < i_r} E_{i_1, i_2, \dots, i_r}\right] \geq 1 - \sum_{r=d}^{r_0} \binom{n}{r} \mathbb{P}[E_{1,2,\dots,r}].$$

We use the following combinatorial analysis to bound $\mathbb{P}[E_{1,2,\dots,r}]$. The total number of 0-1 matrices of size $m \times r$ that have exactly d 1s in every column is $\binom{m}{d}^r$. In addition, if a matrix is to have at least $m - r + 1$ entirely zero columns, all of its nonzero entries must be included in a submatrix of size $r \times r$. There are $\binom{m}{r}$ choices for such a square submatrix, and for each selection, the number of matrices that can be constructed is by the same token as before equal to $\binom{r}{d}^r$ (It should be clear that we are multiple-counting many cases). This yields the following:

$$\mathbb{P}[E_{1,2,\dots,r}] \leq \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r},$$

Hence

$$\mathbb{P}[A \text{ is a } (r_0, \frac{d-1}{d})\text{-Exp.}] \geq 1 - \sum_{r=d}^{r_0} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r}. \quad (\text{A.18})$$

We show that for certain regimes of β , μ and d , the right hand side of (A.18) becomes arbitrarily close to 1. To see this, we prove the following two lemmas.

Lemma A.1. If $0 < \alpha < e^{\frac{d}{2-d}} \beta^{\frac{1-d}{2-d}}$ and $d \geq 3$ then $\sum_{r=d}^{\alpha n} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} = O(n^{1-d(d-2)})$.

Proof. We can write:

$$\sum_{r=d}^{\alpha n} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} \leq \sum_{r=d}^{\alpha n} \binom{n}{r} \binom{m}{r} \left(\frac{r}{m}\right)^{rd} \leq \sum_{r=d}^{\alpha n} \left(\frac{ne}{r}\right)^r \left(\frac{me}{r}\right)^r \left(\frac{r}{m}\right)^{rd} = \sum_{r=d}^{\alpha n} \left(\frac{cr}{n}\right)^{r(d-2)}, \quad (\text{A.19})$$

where $c = e^{\frac{2}{d-2}} \beta^{\frac{1-d}{d-2}}$, and we have used the bounds $\frac{\binom{r}{d}}{\binom{m}{d}} \leq \left(\frac{r}{m}\right)^d$ for $r < m$, and $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$. It is easy to show that when $\alpha < \frac{1}{ec}$, $\left(\frac{cr}{n}\right)^r$ is decreasing in r , and thus replacing all the terms in (A.19) by the first term will only increase the sum. The whole term is thus smaller than $\alpha n \left(\frac{cd}{n}\right)^{d(d-2)} = \lambda n^{1-d(d-2)}$ for some positive constant λ . ■

Lemma A.2. For $m = \beta n$ and $r_0 = \mu n$, if $d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log_2(\frac{\beta}{\mu})}$, then for any $0 < \alpha < \mu$ the sum $\sum_{r=\alpha n+1}^{\mu n} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r}$ decays exponentially as $n \rightarrow \infty$.

Proof. Using the standard bounds of (F.46) on binomial coefficients and the fact that $\frac{\binom{r}{d}}{\binom{m}{d}} \leq \left(\frac{r}{m}\right)^d$ for $r < m$, we can write:

$$\sum_{r=\alpha n+1}^{\mu n} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} \leq n^2 \sum_{r=\alpha n+1}^{\mu n} 2^{nH(\frac{r}{n}) + mH(\frac{r}{m}) + rd \log_2 \frac{r}{m} + \log_2 \{(n+1)(m+1)\}}, \quad (\text{A.20})$$

where $H(x) = x \log(\frac{1}{x}) + (1-x) \log(\frac{1}{1-x})$ is the entropy function. Assuming that $\mu \leq \frac{\beta}{2}$, the largest term on the right hand side of (A.20) is the one corresponding to $r = \mu n$ (since $H(x)$ and $x \log_2 x$ are both increasing for $0 \leq x \leq \frac{1}{2}$), and therefore we can write

$$\sum_{r=\alpha n+1}^{\mu n} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} \leq 4n^5 2^{n(H(\mu) + \beta H(\frac{\mu}{\beta}) + \mu d \log(\frac{\mu}{\beta}))}. \quad (\text{A.21})$$

Note that we have used the fact that $(n+1)(m+1) < 4n^2$. The right hand side of (A.21) vanishes as $n \rightarrow \infty$, if $d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log_2(\frac{\beta}{\mu})}$. ■

B Derivation of the Weak Bound

For a random set S that has a linear-size cardinality (i.e. $\frac{|S|}{n}$ is not asymptotically zero), and $\frac{|S|}{m}$ small enough, $\mathbb{E}|\Gamma(S)| = (1 - e^{-d \frac{|S|}{m}}) \beta n =: \gamma_1 n$ when $n \rightarrow \infty$. We first need to ensure that $|\Gamma(S)|$ is concentrated.

We can show that:

$$\mathbb{P}[|\Gamma(S)| \leq \mathbb{E}|\Gamma(S)| + \epsilon_1] > 1 - \frac{1}{n},$$

This concentration bound can be obtained by the standard martingale concentration arguments if $|S| \geq c_1 n$, for $\epsilon_1 = c_2 \sqrt{n \log n}$, see [22, 24]. Therefore we define the event $E_1 = \{|\Gamma(S)| \leq \gamma_1 n + \epsilon_1\}$. Consider the random graph created from placing d nonzero entries (with repetition) in every column of $\tilde{\mathbf{A}}$. From the set S , form $\Gamma(S)$, the corresponding S_2 , and finally the bipartite graph $B_S = (S \cup S_2, \Gamma(S \cup S_2))$. Using the given combinatorial condition in Lemma 5.1, we can recover a signal supported on S if every subset $S_i \subset S \cup S_2$ of size $|S_i| \leq |\Gamma(S)| + 1$ has sufficient expansion: $|\Gamma(S_i)| \geq |S_i|$ (note that subsequently we drop the $+1$ term since it is negligible for large n). First we condition on the concentration of $|\Gamma(S)|$:

$$\mathbb{P}[S \text{ not recoverable}] = \mathbb{P}[S \text{ not recoverable}|E_1]\mathbb{P}[E_1] + \mathbb{P}[S \text{ not recoverable}|E_1^c]\mathbb{P}[E_1^c] \quad (\text{B.22})$$

$$\leq \mathbb{P}[S \text{ not recoverable}|E_1]\left(1 - \frac{1}{n}\right) + \frac{1}{n}, \quad (\text{B.23})$$

Therefore it suffices to bound the probability conditioned on $|\Gamma(S)|$ concentrated. We are going to do a union bound over all possible selections R_1 of r_1 nodes in S and R_2 of r_2 nodes in S_2 so that $r_1 + r_2 \leq |\Gamma(S)|$. Since we are conditioning on E_1 , it suffices to have $r_1 + r_2 \leq \gamma_1 n + \epsilon_1$. The second problem is that the set S_2 is random and dependent on $\Gamma(S)$. We are going to avoid this conditioning by allowing the choice of R_2 to range over all the $n - k$ nodes in S^c .

$$\mathbb{P}[S \text{ not recoverable}|E_1] \leq \sum_{r_1+r_2 \leq \gamma_1 n + \epsilon_1} \binom{k}{r_1} \binom{n-k}{r_2} \mathbb{P}(R_1 \cup R_2 \text{ contracts}|E_1). \quad (\text{B.24})$$

Now the problem is that conditioning on E_1 implies that the set R_1 does not expand too much, so it is actually increasing the probability of the bad contraction event. We can however easily show that this increase is at most a factor of 2:

$$\mathbb{P}(R_1 \cup R_2 \text{ contracts}|E_1) = \frac{\mathbb{P}(R_1 \cup R_2 \text{ contracts} \cap E_1)}{\mathbb{P}(E_1)} \leq \frac{\mathbb{P}(R_1 \cup R_2 \text{ contracts})}{\mathbb{P}(E_1)}. \quad (\text{B.25})$$

Now since $\mathbb{P}(E_1) \geq 1 - 1/n$, for sufficiently large n , $1/\mathbb{P}(E_1) \leq 2$, so

$$\mathbb{P}(R_1 \cup R_2 \text{ contracts}|E_1) \leq 2\mathbb{P}(R_1 \cup R_2 \text{ contracts}). \quad (\text{B.26})$$

The probability that the set $R_1 \cup R_2$ contracts can be further bounded by assuming $|\Gamma(R_1 \cup R_2)| = r_1 + r_2$ (any smaller neighborhood will have smaller probability) so

$$\mathbb{P}(R_1 \cup R_2 \text{ contracts}) \leq \binom{m}{r_1 + r_2} \left(\frac{r_1 + r_2}{m} \right)^{d(r_1 + r_2)}.$$

Putting everything together we obtain the bound

$$\mathbb{P}[S \text{ not recoverable} | E_1] \leq 2 \sum_{r_1 + r_2 \leq \gamma_1 n + \epsilon_1} \binom{k}{r_1} \binom{n-k}{r_2} \binom{m}{r_1 + r_2} \left(\frac{r_1 + r_2}{m} \right)^{d(r_1 + r_2)}. \quad (\text{B.27})$$

We move everything to the exponent, assume $\epsilon_1/n \rightarrow 0$, and use standard binomial approximations to obtain

$$\mathbb{P}[S \text{ not recoverable} | E_1] \leq 2n^3 \sum_{r_1 + r_2 \leq \gamma_1 n} 2^{kH(\frac{r_1}{k}) + (n-k)H(\frac{r_2}{n-k}) + mH(\frac{r_1+r_2}{m}) + d(r_1+r_2) \log(\frac{r_1+r_2}{m})}. \quad (\text{B.28})$$

Recall that the recoverable fraction is $k = \alpha n$, $m = \beta n$, and denote $\rho_1 = r_1/n$, $\rho_2 = r_2/n$. Define the function

$$F(\rho_1, \rho_2) := \alpha H\left(\frac{\rho_1}{\alpha}\right) + (1 - \alpha)H\left(\frac{\rho_2}{1 - \alpha}\right) + \beta H\left(\frac{\rho_1 + \rho_2}{\beta}\right) + d(\rho_1 + \rho_2) \log\left(\frac{\rho_1 + \rho_2}{\beta}\right), \quad (\text{B.29})$$

and observe that the bound on the probability of failure (B.28) becomes

$$\mathbb{P}[S \text{ not recoverable} | E_1] \leq 2n^3 \sum_{r_1 + r_2 \leq \gamma_1 n} 2^{nF(\rho_1, \rho_2)}.$$

Therefore for fixed β and γ_1 , we are trying to find the largest α^* that makes $F(\rho_1, \rho_2)$ negative for every ρ_1, ρ_2 for which $\rho_1 + \rho_2 \leq \gamma_1$. For this α^* , we can recover sparse signals with sparsity no bigger than $\alpha^* n$ with polynomially high probability, conditioned on the fact that the sublinear sets do not contract (which has already been established).

C Proof of the Validity of Algorithm 1

Algorithm 1 identifies a big zero portion of the output and eliminates two large sets of nodes from X and Y . Having done this, a smaller system of linear equation remains, which turns out to be an

over-determined system and can be uniquely solved using matrix inversions. This procedure is therefore nothing but a block diagonalization (after rearranging the rows and columns) of $\tilde{\mathbf{A}}$ into a lower triangular matrix:

$$\tilde{\mathbf{A}} = \begin{pmatrix} \tilde{\mathbf{A}}_{11} & 0 \\ \tilde{\mathbf{A}}_{12} & \tilde{\mathbf{A}}_2 \end{pmatrix} \quad (\text{C.30})$$

where $\tilde{\mathbf{A}}_2$ is a square or tall full rank matrix. The following theorem certifies that Algorithm 1 is indeed valid and it recovers any k -sparse vector without error.

Theorem C.1. (*Validity of Algorithm 1*) *If \mathbf{x} is a k -sparse nonnegative vector and $\tilde{\mathbf{A}}$ is a perturbed $(kd + 1, 1 - \frac{1}{d})$ expander with $\mathcal{H}(\tilde{\mathbf{A}}) \geq kd + 1$, then $\hat{\mathbf{x}} = \mathbf{x}$.*

Proof. First of all, note that \mathbf{y} is kd -sparse, since every column of $\tilde{\mathbf{A}}$ has exactly d nonzero entries. We also show that $|S_2| \leq |T_2|$ and therefore $\tilde{\mathbf{A}}_2$ is a full rank matrix. Suppose $|S_2| > |T_2|$. We also know that $|T_2| \leq kd$. Select an arbitrary subset $S_2' \subseteq S_2$ of size $|T_2| + 1$. Because of the fact that $|S_2| \leq kd + 1$ and the expansion property: $|\Gamma(S_2')| \geq |S_2'| > |T_2|$. But $\Gamma(S_2')$ is a subset of T_2 and this is a contradiction. Diagonalization of (C.30) and the fact that $\tilde{\mathbf{A}}_2$ is a tall matrix and $\mathcal{H}(\tilde{\mathbf{A}}) \geq kd + 1$ together imply that $\tilde{\mathbf{A}}_2$ has full column rank.

We now show that $\hat{\mathbf{x}} = \mathbf{x}$. If any entry in \mathbf{x}_{S_1} is greater than zero, then there is at least one entry in \mathbf{y} which is indexed in T_1 and is nonzero, since $S_1 = \Gamma(T_1)$. This is in contradiction with the choice of T_1 . Therefore $\mathbf{x}_{S_1} = 0 = \hat{\mathbf{x}}_{S_1}$. Also since $\tilde{\mathbf{A}}_2 \mathbf{x}_{S_2} = \mathbf{y}_2$ and $\tilde{\mathbf{A}}_2 \hat{\mathbf{x}}_{S_2} = \mathbf{y}_2$ and $\tilde{\mathbf{A}}_2$ is full rank we conclude that $\mathbf{x}_{S_2} = \hat{\mathbf{x}}_{S_2}$.

■

D RIP-1 for Expanders

We present a simple argument to show that the adjacency matrix \mathbf{A} of a (k, ϵ) expander graph with $\epsilon < 1/2$ has RIP-1 property, which means that for every k -sparse vector \mathbf{x} and suitable constants c_1, c_2 , the ℓ_1 norm of $\|\mathbf{Ax}\|_1$ is close to the norm of \mathbf{x} :

$$(1 - c_1)\|\mathbf{x}\|_1 \leq \|\mathbf{Ax}\|_1 \leq (1 + c_2)\|\mathbf{x}\|_1, \quad (\text{D.31})$$

where c_1 and c_2 are two constants depending on the expansion coefficient.

It should be acknowledged that Berinde et al. [5] already prove this property, generally for p norms where $p \leq 1 - 1/\log n$. The argument we present here is arguably simpler and easily extends to the case where the matrix is perturbed in the nonzero entries.

Consider \mathbf{A} to be the adjacency matrix of a (k, ϵ) unbalanced expander for $\epsilon < 1/2$. Consider S , the support set of a k -sparse \mathbf{x} . By Hall's theorem, since every set S of size up to k has $d(1 - \epsilon)|S|$ neighbors, there must exist a partial $d(1 - \epsilon)$ -matching, i.e. one in which every node in S can be matched to $d(1 - \epsilon)$ unique neighbors. Decompose the measurement matrix

$$\mathbf{A} = \mathbf{A}_M + \mathbf{A}_C. \quad (\text{D.32})$$

Where \mathbf{A}_M is supported on the partial $d(1 - \epsilon)$ -matching (i.e every row has one non-zero entry and every column has $d(1 - \epsilon)$ non-zero entries). The remainder matrix \mathbf{A}_C has ϵd non-zero entries in each column, and notice that the decomposition is adapted to the support of the vector \mathbf{x} . By the triangle inequality:

$$\|\mathbf{A}\mathbf{x}\|_1 \geq \|\mathbf{A}_M\mathbf{x}\|_1 - \|\mathbf{A}_C\mathbf{x}\|_1. \quad (\text{D.33})$$

It is easy to see that

$$\|\mathbf{A}_M\mathbf{x}\|_1 \geq d(1 - \epsilon)\|\mathbf{x}\|_1, \quad (\text{D.34})$$

since $\mathbf{A}_M\mathbf{x}$ is a vector that contains $d(1 - \epsilon)$ copies of each entry of \mathbf{x} . Also since each column of \mathbf{A}_C contains ϵd non-zero entries,

$$\|\mathbf{A}_C\mathbf{x}\|_1 \leq \epsilon d\|\mathbf{x}\|_1, \quad (\text{D.35})$$

since each entry of $\mathbf{A}_C\mathbf{x}$ is a summation of terms of \mathbf{x} and $\|\mathbf{A}_M\mathbf{x}\|_1$ is also a summation in which each entry of \mathbf{x} appears $d\epsilon$ times. A similar argument implies the upper bound:

$$\|\mathbf{A}\mathbf{x}\|_1 \leq d\|\mathbf{x}\|_1 \quad (\text{D.36})$$

Therefore, putting these together we obtain:

$$(1 - 2\epsilon)\|\mathbf{x}\|_1 \leq \|\mathbf{A}\mathbf{x}\|_1 \leq d\|\mathbf{x}\|_1. \quad (\text{D.37})$$

E Proof of Robustness of Algorithm 1

We first state the following lemma from [5]:

Lemma E.1. *Consequence of Lemma 9 of [5]: If \mathbf{A} is the adjacency matrix of a (k, ϵ) expander with $\epsilon < 0.5$ and \mathbf{u} is a k -sparse vector, then $d(1 - 2\epsilon)\|\mathbf{u}\|_1 \leq \|\mathbf{A}\mathbf{u}\| \leq d\|\mathbf{u}\|_1$.*

If the original vector \mathbf{x} is k -sparse, and each column of \mathbf{A} has d 1s, then $\mathbf{A}\mathbf{x}$ has at most kd 1s, i.e. $\mathbf{A}\mathbf{x}$ is kd -sparse. By rearranging the rows and columns of \mathbf{A} , we may assume $\mathbf{x} = (\mathbf{x}_1 \ \mathbf{x}_2)^T$, $\mathbf{y} = (\mathbf{y}_1 \ \mathbf{y}_2)^T$, $\mathbf{v} = (\mathbf{v}_1 \ \mathbf{v}_2)^T$ and $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & 0 \\ \mathbf{A}_{12} & \mathbf{A}_2 \end{pmatrix}$, where \mathbf{y}_1 and \mathbf{y}_2 are those obtained by the algorithm, $\mathbf{x}_1 = \mathbf{x}_{S_1}$ and $\mathbf{x}_2 = \mathbf{x}_{S_2}$. Also let $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ be the reconstruction error vector. By (16) we then have

$$\mathbf{y}_1 = \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{v}_1, \quad \mathbf{y}_2 = \mathbf{A}_{12}\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 + \mathbf{v}_2, \quad \mathbf{e} = (\mathbf{x}_1 \ \mathbf{x}_2 - \hat{\mathbf{x}}_2)^T. \quad (\text{E.38})$$

Hence we have:

$$\|\mathbf{x}_1\|_1 \leq \|\mathbf{A}_{11}\mathbf{x}_1\|_1 = \|\mathbf{y}_1 - \mathbf{v}_1\|_1 \leq \|\mathbf{y}_1\|_1 + \|\mathbf{v}_1\|_1 \leq 2\|\mathbf{v}\|_1. \quad (\text{E.39})$$

The first inequality holds as a result of nonnegativity of \mathbf{x}_1 and \mathbf{A}_{11} , and the fact that every column of \mathbf{A}_{11} has at least one 1. The last inequality holds for the following reason. We know that $\mathbf{A}\mathbf{x}$ is kd -sparse. Let T_0 be a set of $m - kd$ zeros of $\mathbf{A}\mathbf{x}$. We then have

$$\|\mathbf{y}_1\|_1 \leq \|\mathbf{y}_{T_0}\|_1 = \|\mathbf{v}_{T_0}\|_1 \leq \|\mathbf{v}\|_1. \quad (\text{E.40})$$

Let us assume $\mathbf{y}_2 = \mathbf{A}_2\hat{\mathbf{x}}_2 + \delta_2$. From the way $\hat{\mathbf{x}}_2$ is driven in step 4 of the algorithm, it follows that:

$$\|\delta_2\|_1 \leq \|\mathbf{A}_{12}\mathbf{x}_1 + \mathbf{v}_2\|_1. \quad (\text{E.41})$$

And thus

$$\|\mathbf{A}_2(\mathbf{x}_2 - \hat{\mathbf{x}}_2)\|_1 = \|\delta_2 - \mathbf{A}_{12}\mathbf{x}_1 - \mathbf{v}_2\|_1 \leq 2\|\mathbf{A}_{12}\mathbf{x}_1 + \mathbf{v}_2\|_1 \leq 2d\|\mathbf{x}_1\|_1 + 2\|\mathbf{v}_2\|_1. \quad (\text{E.42})$$

We may note that $|S_2| < 2k$. Otherwise, since \mathbf{A} corresponds to a $(2k, \epsilon)$ expander, we can choose an

arbitrary subset $S_2' \subset S_2$ of size $2k$, and conclude that

$$|\Gamma(S_2)| \geq |\Gamma(S_2')| \geq 2k(1 - \epsilon)d > kd. \quad (\text{E.43})$$

However, from the construction of S_2 , we know that $\Gamma(S_2) = T_2$ and $|T_2| = kd$. Therefore the vector $\mathbf{u} = (0 \ \mathbf{x}_2 - \hat{\mathbf{x}}_2)^T$ is $2k$ -sparse (because the size of \mathbf{x}_2 is at most $2k$), and we can apply the RIP-1 condition of Lemma E.1 to it, which yields

$$c_1 \|\mathbf{x}_2 - \hat{\mathbf{x}}_2\|_1 \leq 2d\|\mathbf{x}_1\|_1 + 2\|\mathbf{v}_2\|_1, \quad (\text{E.44})$$

where $c_1 = (1 - 2\epsilon)d$. Equations (E.39) and (E.44) result in (assuming $d \geq 2$):

$$\|\mathbf{e}\|_1 \leq \left(2 + \frac{4d}{c_1}\right)\|\mathbf{v}\|_1 + \frac{2}{c_1}\|\mathbf{v}_2\|_1 \leq \frac{7 - 4\epsilon}{1 - 2\epsilon}\|\mathbf{v}\|_1. \quad (\text{E.45})$$

F Elementary bounds on binomial coefficients

For each $\beta \in (0, 1)$, define the binomial entropy $H(\beta) = -\beta \log_2 \beta - (1 - \beta) \log_2(1 - \beta)$ (and $H(0) = H(1) = 0$ by continuity). We make use of the following standard bounds on the binomial coefficients from [32]:

$$n \left[H\left(\frac{k}{n}\right) - \frac{\log_2(n+1)}{n} \right] \leq \log_2 \binom{n}{k} \leq n \left[H\left(\frac{k}{n}\right) + \frac{\log_2(n+1)}{n} \right]. \quad (\text{F.46})$$



Amin Khajehnejad received his Bachelors degree in Electrical Engineering from University of Tehran in 2007, and his Masters degree in Electrical Engineering from California Institute of Technology in 2009. He is currently a PhD candidate at Caltech. His general interests are in signal processing, coding, information theory and optimization. He has been with Lyric semiconductors Inc., and NEC laboratories America, Inc. summers of 2009 and 2010 respectively.



Alex Dimakis is an assistant Professor at the Viterbi School of Engineering, University of Southern California. He has been a faculty member in the Department of Electrical Engineering - Systems since 2009. He received his Ph.D. in 2008 and M.S. degree in 2005 in electrical engineering and computer science, both from the University of California, Berkeley. Prior to that, he obtained the Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens in 2003. He received the Eli Jury award in 2008 for his thesis work on codes for distributed storage, two outstanding paper awards, the UC Berkeley Regents Fellowship and the Microsoft Research Fellowship. He was a postdoctoral scholar at the Center for the Mathematics of Information (CMI) at Caltech in 2008. His research interests include communications, coding theory, signal processing, and networking, with a current focus on distributed storage, network coding, large-scale inference and message passing algorithms.



Weiyu Xu received the Bachelor degree in Information Engineering from Beijing University of Posts and Telecommunications in 2002. After getting a Master degree in Electronic Engineering from Tsinghua University, Beijing, he went on to pursue his graduate study in the Electrical Engineering Department at California Institute of Technology, Pasadena, California, beginning in September 2005. In August 2009, he obtained his Ph.D. degree in Electrical Engineering at California Institute of Technology, with a minor in Applied and Computational Mathematics. Since September 2009, he has been a postdoctoral associate at Cornell University, Ithaca, New York. His research interests are in networking science and systems, communication, control and signal processing, with an emphasis on detection and estimation, coding and information Theory He received the Charles and Ellen Wilts best dissertation prize in 2010.



Babak Hassibi was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran in 1989, and the M.S. and Ph.D. degrees from Stanford University in 1993 and 1996, respectively, all in electrical engineering.

He has been with the California Institute of Technology since January 2001, where he is currently Professor and Executive Officer of Electrical Engineering. From October 1996 to October 1998 he was a research associate at the Information Systems Laboratory, Stanford University, and from November 1998 to December 2000 he was a Member of the Technical Staff in the Mathematical Sciences Research Center at Bell Laboratories, Murray Hill, NJ. He has also held short-term appointments at Ricoh California Research Center, the Indian Institute of Science, and Linkoping University, Sweden. His research interests include wireless communications and networks, robust estimation and control, adaptive signal processing and linear algebra. He is the coauthor of the books (both with A.H. Sayed and T. Kailath) *Indefinite Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories* (New York: SIAM, 1999) and *Linear Estimation* (Englewood Cliffs, NJ: Prentice Hall, 2000). He is a recipient of an Alborz Foundation Fellowship, the 1999 O. Hugo Schuck best paper award of the American Automatic Control Council (with H. Hindi and S.P. Boyd), the 2002 National Science Foundation Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering and the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE), and was a participant in the 2004 National Academy of Engineering “Frontiers in Engineering” program.

He has been a Guest Editor for the IEEE Transactions on Information Theory special issue on “space-time transmission, reception, coding and signal processing” was an Associate Editor for Communications of the IEEE Transactions on Information Theory during 2004-2006, and is currently an Editor for the Journal “Foundations and Trends in Information and Communication”.