ON THE RECOVERY OF NONNEGATIVE SPARSE VECTORS FROM SPARSE MEASUREMENTS INSPIRED BY EXPANDERS

M.Amin Khajehnejad and Babak Hassibi

Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA
Email: {amin, hassibi}@caltech.edu

ABSTRACT

This paper studies compressed sensing for the recovery of *non-negative* sparse vectors from a smaller number of measurements than the ambient dimension of the unknown vector. We focus on measurement matrices that are sparse, i.e., have only a constant number of nonzero (and non-negative) entries in each column. For such measurement matrices we give a simple necessary and sufficient condition for l_1 optimization to successfully recover the unknown vector. Using a simple "perturbation" to the adjacency matrix of an unbalanced expander, we obtain simple closed form expressions for the threshold relating the ambient dimension n, number of measurements m and sparsity level k, for which l_1 optimization is successful with overwhelming probability. Simulation results suggest that the theoretical thresholds are fairly tight and demonstrate that the "perturbations" significantly improve the performance over a direct use of the adjacency matrix of an expander graph.

Index Terms— compressed sensing, expander graph, non-negative vector, l_1 optimization

1. INTRODUCTION

Recently, expander graphs have been studied in the context of compressed sensing [5], [6], [7], [8],[10]. The rationale is that the adjacency matrix of an unbalanced expander graph can be used as the measurement matrix A in compressed sensing. This is for several reasons. First, the resulting measurement matrices are very sparse and simple in contrast to random orthoprojectors or other suggested structures. Therefore in applications where *measurement* is costly (like DNA micro arrays) this plays a crucial role([11]). Second, there are specific algorithms apart from linear programming that can be applied to expander graphs for the recovery of general sparse vectors. One such fast recovery algorithm can be found in [7] and its modified version in [8]. Finally, unlike random measurement

matrices (such as Gaussian or Bernoulli), which only guarantee the recovery of sparse vectors with high probability, expander graphs give deterministic guarantees (see, e.g., [7], which gives a deterministic guarantee for the fast algorithm proposed, and [5], which shows that the adjacency matrix of an expander graph satisfies a so-called Restricted Isometry Property(RIP) which guarantees that linear programming can be used to recover sparse enough vectors).

However, it turns out that the RIP is only a sufficient condition. Linear programming is effective if and only if a certain condition on the null space of A holds. More precisely, if for any vector w in the null space of A, the sum of the absolute values of any k elements of w is less that the sum of the absolute values of the rest of the elements, then the solution to $\min ||x||_0 \ subject \ to \ Ax = y$ can always be obtained by solving $\min ||x||_1 \ subject \ to \ Ax = y$, provided x is k-sparse.[1] This condition is stated in the work of Donoho as the k-neighborly polytope property of A, [1], and in the work of Candes et. al. as the uncertainty principle ,[3]. Donoho et. al. also have been able to show the validity of this condition with overwhelming probability for random i.i.d Gaussian matrices and are therefore able to compute fairly tight thresholds on when linear-programming-based compressed sensing works [2]. The first analysis of the null space property for expender graphs has been done by Indyk. He shows that every $(2k,\epsilon)$ expander graph with $\epsilon \leq \frac{1}{6}$ has this null space property, [6]. In fact, he shows that for every vector in the null space of the adjacency matrix of a $(2k,\epsilon)$ unbalanced expander, the absolute sum of every k elements is less than or equal to $\frac{2\epsilon}{1-4\epsilon}$ the absolute sum of the rest. Using Theorem 1 of [9], which is a generalization of the null space property theorem for the recovery of approximately sparse signals, Indyk's result gives an upper bound on the error when linear programming is used to recover approximately k-sparse vectors using expander graph measurements.

Unlike Gaussian matrices, where reasonably sharp bounds on the thresholds which guarantee linear programming to re-

---

[1] Here $|| \cdot ||_0$ represents the number of non-zero entries in its argument vector and $|| \cdot ||_1$ is the standard l_1-norm.

cover sparse signals have been obtained, such sharp bounds do not exist for expander-graph-based measurements. This is the main focus of the current paper. However, we shall focus on the special case where the k-sparse vector has nonnegative entries. This is an important special case that arises in many applications. It turns out that, due to the additional non-negativity constraint, one can say a lot more and, in fact, one can recover "less sparse" signals in this case. This nonnegative case (albeit for Gaussian measurement matrices) has also been studied in [4].

The remainder of this paper is organized as follows. In Section 2 we formulate the problem. Then in section 3 we state the main theorem that relates the success of linear programming recovery in terms of a null space characterization. We define the new concept of complete rank and relate it to unbalanced expanders and also to the compressed sensing problem. Section 4 is dedicated to a probabilistic approach based on which we prove the existence of a class of desired expanders and consequently sparse matrices with proportional complete rank. In Section 5 we provide a weak condition and achieve a weak bound below which l_1-optimization is successful with overwhelming probability. Finally we wrap up the paper with simulation results.

2. PROBLEM FORMULATION

The goal in compressed sensing is to recover a sparse vector from a set of (linear) under-determined system of equations. In many real world applications the original data vector is nonnegative. The problem of nonnegative input vectors recovery is as following:

$$\min_{Ax=y, x \geq 0} ||x||_0 \qquad (1)$$

where $A^{m \times n}$ and $y^{m \times 1}$ are respectively the measurement matrix and observation vector. Although (1) is an NP-hard problem, Donoho and Tanner have shown in [4] that for a class of matrices A maintaining a so-called outwardly k-neighborly property and x being at most k-sparse, the solution to (1) is unique and can be recovered via the following linear programming problem:

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

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 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 $0.558m$ sparse vectors x via l_1 optimization for sufficiently large n.

In this paper, we seek the answer to a similar question when our measurement matrix is the adjacency matrix of an unbalanced bipartite graph with constant left degree d. Now 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. NULL SPACE CHARACTERIZATION AND COMPLETE RANK

We first state an equivalent version of the outwardly neighborly condition of [4], which has a much more mundane interpretation for the special case of regular bipartite graphs.

Theorem 3.1. *integer. These two statements are equivalent:*

1. *Whenever x_0 is a solution to (1) with at most k nonzero elements, x_0 is also the unique solution of (2).*

2. *For every vector w in the null space of A , and every index set $S \subset \{1, 2, ..., n\}$ with $|S| = k$ such that w_{S^c} is a non-negative vector, the following holds: $\sum_{i=1}^{n} w_i > 0$.*[2]

let A be a nonnegative $m \times n$ matrix and $k < n/2$, a positive We refer to the latter statement as ***non-negative null space property***.

Proof. Similar to the proof of Theorem 1 of [9] with the additional notice that $x_1 - x_0$ is nonnegative on S^c. □

Now let's assume that A is the adjacency matrix of a bipartite graph with left vertex size n, right vertex size m and left degree d. In other words A is a $0 - 1 \ (m \times n)$ matrix with exactly d ones in each column. The following lemma holds for any such matrix.

Lemma 3.1. *For any vector w in the null space of A, if $w_{\geq 0}$ is the non-negative part of w and $w_{<0}$ is the negative portion of w then $||w_{\geq 0}||_1 = ||w_{<0}||_1$*

Proof. Wlog let's assume $w = \begin{bmatrix} w_{\geq 0} \\ w_{<0} \end{bmatrix}$. The proof comes from the fact that: $\left\| A \begin{bmatrix} w_{\geq 0} \\ 0 \end{bmatrix} \right\|_1 = d \left\| \begin{bmatrix} w_{\geq 0} \\ 0 \end{bmatrix} \right\|_1$ and $\left\| A \begin{bmatrix} 0 \\ w_{<0} \end{bmatrix} \right\|_1 = d \left\| \begin{bmatrix} 0 \\ w_{<0} \end{bmatrix} \right\|_1$. □

Theorem 3.2. *For a 0-1 matrix $A^{m \times n}$ with exactly d 1's in each column, the following two statements are equivalent:*

1. *Whenever x_0 is a solution to (1) with at most k nonzero elements, x_0 is also the unique solution of (2).*

2. *Every vector w in the null space of A has at least k negative elements.*

---

[2] a sufficient condition similar to that of [9] would be $||w_S||_1 \leq ||w_S^c||_1$.

Proof. Readily follows from Theorem 3.1 and Lemma 3.1.
□

Our goal is to prove the existence of bipartite graphs with the above null space property. We begin with two definitions.

Definition 3.1. *For a matrix $A^{m \times n}$ we define the* Complete Rank *of A (denoted by $Cr(A)$) to be the maximum value r_0 for which every r_0 columns of A are linear independent.*

Definition 3.2. *A left d-regular bipartite graph with X and Y the set of left and right vertices($|X| = n, |Y| = m$), is called a (k, ϵ) unbalanced expander if for every $S \subset X$ with $|S| \leq k$, the following holds: $|N(S)| \geq kd(1 - \epsilon)$, where $N(S)$ is the set of neighbors of S in Y.*

The following lemma is almost obvious:

Lemma 3.2. *Every left regular bipartite graph (X,Y,d) with adjacency matrix A is a $(Cr(A), \frac{d-1}{d})$ expander. Consequently, for every $S \subset X : |N(S)| > \min(|S|, Cr(A))$.*

The notion of complete rank (also known as *Kruskal* rank in linear algebra) is tightly related to the expansion property. It is also related to the null space characterization we are shooting for.

Theorem 3.3. *If $A^{m \times n}$ is the adjacency matrix of a left regular bipartite graph, and w is a vector in its null space, then the number of negative elements of w is at least $\frac{Cr(A)}{d}$.*

Proof. Let S^+ be the set of vertices in X corresponding to the positive elements of w, and likewise S^- be set of vertices corresponding to the negative elements. Let $S = S^+ \cup S^-$. Clearly $N(S^+) = N(S^-) = N(S)$, since otherwise there is a vertex in Y connected to exactly one of S^+ and S^- and this contradicts the fact that $Aw = 0$. Besides, $|S| \geq Cr(A)$. Therefore

$$d|S^-| \geq |N(S^-)| = |N(S)| > Cr(A).$$

□

The question is whether for n and $m = \beta n$ large there exist matrices $A^{m \times n}$ with d (constant) 1's in each column such that $Cr(A)$ is proportional to n? Furthermore, what is the maximum achievable value of $\frac{Cr(A)}{nd}$? 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 A, as shown in the next lemma (whose proof we omit for brevity).

Lemma 3.3. *For a matrix $A \in \mathbb{R}^{m \times n}$ which is the adjacency matrix of a bipartite left d-regular graph, if the submatrix formed by any r_0 columns of A has at least r_0 nonzero rows, then it is possible to add a slight perturbation to each nonzero entry of A and get $\tilde{A}$ with $Cr(\tilde{A}) \geq r_0$. Furthermore, perturbations can be done in a way that the sum of each column remains a constant d.*
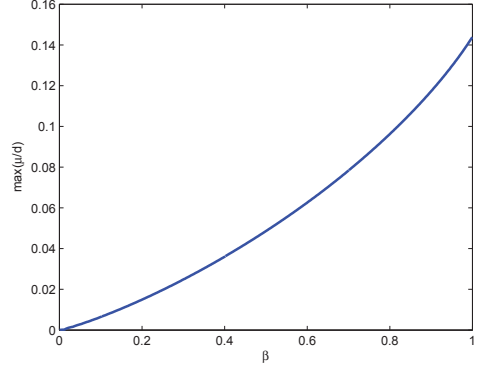
Fig. 1: maximum achievable ratio of $\frac{\mu}{d}$ for different values of β.

After modifying A based on perturbations described above, Theorem 3.1, Lemma 3.1 and Theorems 3.2 and 3.3 all continue to hold for this class of matrices $\tilde{A}$. Note that matrix A in Lemma 3.3 is nothing but the adjacency matrix of a $(r_0, \frac{d-1}{d})$ unbalanced expander graph. Therefore what we really care about is to construct a $(r_0, \frac{d-1}{d})$ expander with $\frac{r_0}{nd}$ as large as possible. We now use the probabilistic method to show that the desired $(r_0, \frac{d-1}{d})$ expanders with r_0 a fraction of n exist.

4. SPARSE MATRICES WITH COMPLETE RANK PROPORTIONAL TO DIMENSION

For fixed values of $n > m > r_0$ and d we are interested in the following question: Does there exist a $(r_0, \frac{d-1}{d})$ expander with constant left degree d? Using the probabilistic method, we are able to show the following result.
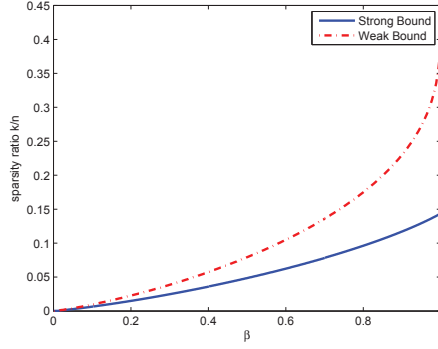
Theorem 4.1. *For sufficiently large n, with $m = \beta n$ and $r_0 = \mu n$, there exists a bipartite graph with left vertex size n and right size m which is a $(r_0, \frac{d-1}{d})$ expander, if $d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log(\frac{\beta}{\mu})}$, where $H(x)$ is the entropy function defined as $H(x) = x \log_2 \frac{1}{x} + (1 - x) \log_2 \frac{1}{1-x}$*

Figure 1 illustrates the maximum achievable ratio $\frac{\mu}{d}$ for different values of β based on Theorem 4.1

5. WEAK BOUND

Another very important question is how large can k be so that *almost* all k sparse vectors that are the solutions to (1) will be recovered via (2). This will be addressed in this section.

Theorem 5.1. *let $x_0 \in (\mathbb{R}^+)^n$ be fixed and $y = Ax_0$. Then the solution x of (2) will be identical to x_0 if and only if there exists no w in the null space of A such that w_{S^c} is nonnegative and $||w_{S^c}||_1 \leq ||w_S||_1$, where S is the support set of x_0.*

Fig. 2: Comparison of weak and strong bound on $\frac{k}{n}$.

Fig. 3: Recovery percentage Vs. sparsity for perturbed expanders and its modified perturbed version.

Proof. Very similar to the proof of Theorem 3.1 ☐

Corollary 5.1. *Let A be the adjacency matrix of a bipartite graph with left constant degree, x_0 be a fixed nonnegative vector and $y = Ax_0$. Then the solution x of (2) will be identical to x_0 if and only if there exists no w in the null space of A such that w_{S^c} is a nonnegative vector, where S is the support set of x_0.*

Proof. Directly from Theorem 5.1 and Lemma 3.1. ☐

Using these results, an application of the probabilistic method now allows the following.

Theorem 5.2. *Let $k = \gamma m$. For n large enough, there exists a nonnegative sparse matrix $A^{m \times n}$ with each column having d nonzero entries that sum to d, such that for almost all $S \subset \{1, 2, ..., n\}$, $|S| = k$, there exists no w in the null space of A for which w_{S^c} is a nonnegative vector, provided $d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log(\frac{\beta}{\mu})}$ and $1 - exp(-\gamma d) \leq \frac{\mu}{\beta}$.*
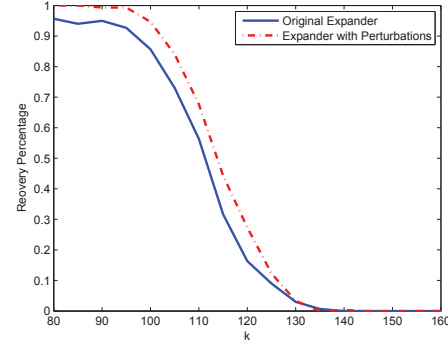
Figure 2 illustrates the weak bound on $\frac{k}{n}$ as a function of $\beta = \frac{m}{n}$ from Theorem 5.2.

6. SIMULATIONS

We generated random $m \times n$ matrices A with $n = 2m = 500$, and $d = 3$ 1's in each column. We then multiplied random sparse vectors with different sparsity levels by A, and tried recovering them via the linear programming of (2). Next we added the perturbations described in section 3 to A and applied the same sparse vectors to compare the recovery percentage in the two cases. This process was repeated for a few generations of A and the improvements we obtained is illustrated in Figure 6. 100 samples were tried for each sparsity level.

7. REFERENCES

[1] David Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete and Computational Geometry*, 102(27),pp. 617-652 2006, Springer.

[2] David Donoho and Jared Tanner "Neighborlyness of randomly-projected simplices in high dimensions," *Proc. National Academy of Sciences*, 102(27), pp. 9452-9457, 2005.

[3] E.Candes, 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] David L.Donoho and Jared 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] A. Gilbert, P. Indyk, H. Karloff, and M. Strauss "Combining geometry and combinatorics: a unified approach to sparse signal recovery". Manuscript, 2007.

[6] Radu Berinde and Piotr Indyk "Sparse recovery using sparse matrices", *Computer Science and Artificial Intelligence Laboratory Technical Report*, MIT-CSAIL-TR-2008-001,January 10, 2008.

[7] Weiyu Xu and Babak Hassibi "Efficient compressive sensing with determinstic guarantees using expander graphs", *IEEE Information Theory Workshop*, Lake Tahoe, September 2007.

[8] Sina Jafarpour, Weiyu Xu, Babak Hassibi, and Robert Calderbank "Efficient compressed sensing using high-quality expander graphs". Preprint, 2008).

[9] Weiyu Xu and Babak Hassibi "Compressed sensing over the Grassmann manifold: A unified analytical framework". Sixth Annual Allerton Conference, September 23-26, 2008, University of Illinois at Urbana-Champaign, IL, USA

[10] P. Indyk Explicit constructions for compressed sensing of sparse signals, SODA, 2008.

[11] Farzad Parvaresh, Haris Vikalo, Sidhant Misra, and Babak 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.