On construction and analysis of sparse random matrices and expander graphs with applications to compressed sensing.

Bubacarr Bah

Laboratory for Information and Inference Systems
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
Email: bubacarr.bah@epfl.ch

Jared Tanner

Mathematics Institute and Exeter College
University of Oxford
Oxford, United Kingdom
Email: tanner@maths.ox.ac.uk

Abstract—We revisit the probabilistic construction of sparse random matrices where each column has a fixed number of nonzeros whose row indices are drawn uniformly at random. These matrices have a one-to-one correspondence with the adjacency matrices of lossless expander graphs. We present tail bounds on the probability that the cardinality of the *set of neighbors* for these graphs will be less than the expected value. The bounds are derived through the analysis of collisions in unions of sets using a *dyadic splitting* technique. This analysis led to the derivation of better constants that allow for quantitative theorems on existence of lossless expander graphs and hence the sparse random matrices we consider and also quantitative compressed sensing sampling theorems when using sparse non mean-zero measurement matrices.

I. Introduction

Sparse matrices are particularly useful in applied and computational mathematics because of their low storage complexity and fast implementation as compared to dense matrices. Of late, significant progress has been made to incorporate sparse matrices in compressed sensing, with [1], [2], [3], [4] giving both theoretical performance guarantees and also exhibiting numerical results that shows sparse matrices coming from expander graphs can be as good sensing matrices as their dense counterparts. In fact, Blanchard and Tanner [5] recently demonstrated in a GPU implementation how well these type of matrices do compared to dense Gaussian and Discrete Cosine Transform matrices even with very small fixed number of nonzeros per column (as considered here).

In this manuscript we consider random sparse matrices that are adjacency matrices of lossless expander graphs. Expander graphs are highly connected graphs with very sparse adjacency matrices, a precise definition of a lossless expander graph is given in Definition 1.

Definition 1: $G(U, V, E)$ is a lossless (k, d, ϵ)-expander if it is a bipartite graph with $|U| = N$ left vertices, $|V| = n$ right vertices and has a regular left degree d, such that any $X \subset U$ with $|X| \leq k$ has a set of neighbors $\Gamma(X) \subset V$ with $|\Gamma(X)| \geq (1 - \epsilon) d |X|$ neighbors.

Note that these graphs are *lossless* because $\epsilon \ll 1$, they are also referred to as *unbalanced expanders* in the literature

because $n \ll N$ and a (k, d, ϵ)-lossless expander graph has an *expansion* of $(1 - \epsilon) d$. Such graphs have been well studied in theoretical computer science and mathematics and have many applications. Probabilistic constructions of such graphs using random left-regular bipartite graphs with optimal parameters exist but deterministic constructions only achieve sub-optimal parameters, see [6] or [7] for a more detailed survey.

Using a novel technique of *dyadic splitting of sets*, this work derives quantitative guarantees on the probabilistic construction of these graphs in the form of a bound on the tail probability of the size of the *set of neighbors*, $\Gamma(X)$ for a given $X \subset U$, of a randomly generated left-degree bipartite graph. Moreover, this tail bound proves a bound on the tail probability of the *expansion* of the graph, $|\Gamma(X)|/|X|$. In addition, we derive the first phase transitions showing regions in parameter space that depicting when a left-regular bipartite graph with a given set of parameters is guaranteed to be a lossless expander with high probability. Similar results in terms of the adjacency matrices of these graphs is also presented. Another contribution of this work is the derivation of sampling theorems comparing performance guarantees for some of the algorithms proposed for compressed sensing using such sparse matrices as well as the more traditional ℓ_1 minimization compressed sensing formulation. It also provides phase transitions of ℓ_1 minimization performance guarantees for such sparse matrices compared to what ℓ_2 restricted isometry constants (RIC$_2$) analysis yields for Gaussian matrices.

II. Tail Bound

Our main result is the presentation of formulae for the expected cardinality of the *set of neighbors* of (k, d, ϵ)-lossless expander graphs and the sparse non-mean zero matrices from these graphs. Based on this, we present a tail bound on the probability that this cardinality will be less than the expected value. We start by defining the class of matrices we consider and a key concept of a *set of neighbors* used in our derivation.

Definition 2: Let A be an $n \times N$ matrix with d nonzeros in each column. We refer to A as a random a) sparse expander

(SE) if every nonzero has value 1 and b) sparse signed expander (SSE) if every nonzero has value from $\{-1, 1\}$.

The support set of the d nonzeros per column of these matrices are drawn uniformly at random and independently for each column. An SE matrix is an adjacency matrix of (k, d, ϵ)-lossless expander graph while an SSE matrix have random sign patterns in the nonzeros of an adjacency matrix of a (k, d, ϵ)-lossless expander graph. If A is either an SE or SSE it will have only d nonzeros per column and since we fix $d \ll n$, A is therefore extremely sparse.

We formally define the *set of neighbors* in both graph theory and linear algebra notation to aid translation between the terminology of the two communities. Denote A_S as a submatrix of A composed of columns of A indexed by the set S with $|S| = s$.

Definition 3: Consider a bipartite graph $G(U, V, E)$ where E is the set of edges and $e_{ij} = (x_i, y_j)$ is the edge that connects vertex x_i to vertex y_j. For a given subset of left vertices $S \subset U$ its set of neighbors $\Gamma(S) \subset V$ is defined as $\Gamma(S) := \{y_j | x_i \in S \text{ and } e_{ij} \in E\}$. In terms of the adjacency matrix, A, of $G(U, V, E)$ the set of neighbors of A_S denoted by A_s, is the set of rows with at least one nonzero.

Henceforth, we will only use the linear algebra notation A_s which is equivalent to $\Gamma(S)$. Note that $|A_s|$ is a random variable depending on the draw of the set of columns, S, for each fixed A. Therefore, we can ask what is the probability that $|A_s|$ is not greater than a_s, in particular where a_s is smaller than the expected value of $|A_s|$. This is the question that Theorem 4 attempts to answers.

Theorem 4 (Theorem 1.6, [8]): For fixed s, n, N, d and $d \leq a_s < \infty$, let an $n \times N$ matrix, A be drawn from either of the classes of matrices defined in Definition 2, then

$$\text{Prob}\left(|A_s| \leq a_s\right) < p_{max}(s, d) \cdot e^{[n \cdot \Psi(a_s, \ldots, a_1)]} \quad (1)$$

where $p_{max}(s, d) = \frac{2}{25\sqrt{2\pi s^3 d^3}}$, and for random variables $a_s, \ldots, a_2$ and $a_1 := d$, $\Psi(a_s, \ldots, a_1)$ is given by

$$\frac{1}{n}\left[3s \log(5d) + \sum_{i=1}^{\lceil s/2 \rceil} \frac{s}{2i}\left((n - a_i) \cdot \text{H}\left(\frac{a_{2i} - a_i}{n - a_i}\right)\right.\right.$$
$$\left.\left. + a_i \cdot \text{H}\left(\frac{a_{2i} - a_i}{a_i}\right) - n \cdot \text{H}\left(\frac{a_i}{n}\right)\right)\right],$$

where $\text{H}(\cdot)$ is the Shannon entropy function of base e logarithm. Consequently:

1) if no restriction is imposed on a_s then the a_i for $i > 1$ take on the expected values of $|A_s|$, which are given by $\hat{a}_{2i} = \hat{a}_i\left(2 - \frac{\hat{a}_i}{n}\right)$ for $i = 1, 2, 4, \ldots, \lceil s/2 \rceil$;
2) else if a_s is restricted to be less than $\hat{a}_s$, then the a_i for $i > 1$ are the unique solutions to the following polynomial system $a_{2i}^3 - 2a_i a_{2i}^2 + 2a_i^2 a_{2i} - a_i^2 a_{4i} = 0$ for $i = 1, 2, \ldots, \lceil s/4 \rceil$ with $a_{2i} \geq a_i$ for each i.

Theorem 4 gives a bound on the probability that the cardinality of a union of k sets each with d elements is less than a_k. Figure 1 shows plots of values of $|A_k|$ (size of set of neighbors) for different k taken over 500 realizations (in blue),

superimposed on these plots is the empirical mean values of $|A_k|$ over the 500 runs (in red) and the $\hat{a}_k$ in green.
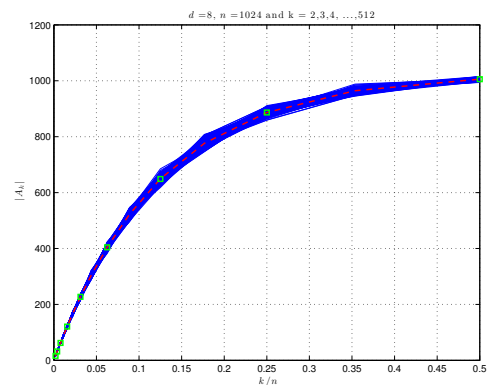

Fig. 1. For fixed $d = 8$ and $n = 2^{10}$, over 500 realizations, plots (in blue) of the cardinalities of the index sets of nonzeros in a given number of set sizes, k. The dotted red curve is mean of the simulations and the green squares are the $\hat{a}_k$.

Furthermore, simulations illustrate that the $\hat{a}_k$ are the expected values of the cardinalities of the union of k sets, $|A_k|$, as shown in Figure 2, where we show the relative error between $\hat{a}_k$ and the empirical mean values of the $|A_k|$, denoted by $\bar{a}_k$, realized over 500 runs, to be less than 10^{-3}.
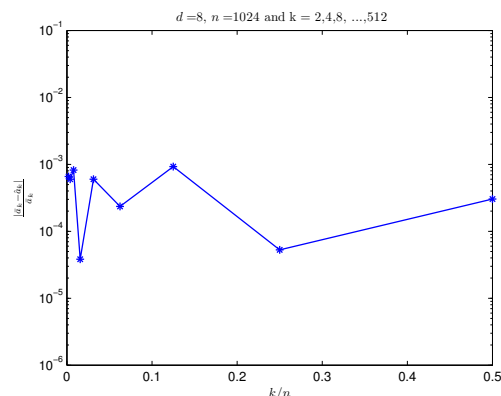

Fig. 2. For fixed $d = 8$ and $n = 2^{10}$, over 500 realizations, plots of the relative error between the mean values of a_k (referred to as $\bar{a}_k$) and the $\hat{a}_k$.

III. SAMPLING THEOREMS

We now use Theorem 4 with the ℓ_1-norm restricted isometry property (RIP-1), introduced by Berinde et. al. in [1], to deduce the corollaries that follow which are about the probabilistic construction of expander graphs, the matrices we consider, and sampling theorems of some selected compressed sensing algorithms. Firstly, using only the expansion property of these graphs we can draw the following corollary from Theorem 4.

Corollary 5: For fixed s, n, N, d and $0 < \epsilon < 1/2$, let an $n \times N$ matrix, A be drawn from the class of matrices defined in Definition 2, then

$$\text{Prob}\left(\|A_S x\|_1 \leq (1 - 2\epsilon)d\|x\|_1\right) < p_{max}(s, d) \cdot e^{[n \cdot \Psi(s, d, \epsilon)]},$$

where $\Psi(s, d, \epsilon) = \Psi(a_s, \ldots, a_1)$ with $a_s = (1 - \epsilon)ds$.

Theorem 4 and Corollary 5 allow us to calculate s, n, N, d, ϵ where the probability of the probabilistic constructions in Definition 2 not being a (s, d, ϵ)-lossless expander is exponentially small. Using Corollary 5 and the RIP-1 results in [1] we derived a bound for the probability that a random draw of a matrix with d 1s or ± 1s in each column fails to satisfy the lower bound of the RIP-1 constant (RIC_1) and hence fails to come from the class of matrices given in Definition 2, for details see [8]. From this bound we deduce the following corollary which is a sampling theorem on the existence of lossless expander graphs.

Corollary 6: Consider $0 < \epsilon < 1/2$ and d fixed. If A is drawn from the class of matrices in Definition 2 and any k-sparse x with $(k, n, N) \to \infty$ while $k/n \to \rho \in (0, 1)$ and $n/N \to \delta \in (0, 1)$ then for $\rho < (1 - \gamma)\rho^{exp}(\delta; d, \epsilon)$ and $\gamma > 0$

$$\text{Prob}\left(\|Ax\|_1 \geq (1 - 2\epsilon)d\|x\|_1\right) \to 1 \qquad (2)$$

exponentially in n, where $\rho^{exp}(\delta; d, \epsilon)$ is the largest limiting value of k/n for which $\text{H}\left(\frac{k}{N}\right) + \frac{n}{N}\Psi(k, d, \epsilon) = 0$.

For each fixed $0 < \epsilon < 1/2$ and each fixed d, $\rho^{exp}(\delta; d, \epsilon)$ in Corollary 6 is a function of δ and a phase transition function in the (δ, ρ) plane. Below the curve of $\rho^{exp}(\delta; d, \epsilon)$ the probability in (2) goes to one exponentially in n as the problem size grows. That is if A is drawn at random with d 1s or $d \pm 1$s in each column and having parameters (k, n, N) that fall below the curve of $\rho^{exp}(\delta; d, \epsilon)$ then we say it is from the class of matrices in Definition 2 with probability approaching one exponentially in n. In terms of $|\Gamma(X)|$ for $X \subset U$ and $|X| \leq k$, Corollary 6 say that the probability $|\Gamma(X)| \geq (1 - \epsilon)dk$ goes to one exponentially in n if the parameters of our graph lies in the region below $\rho^{exp}(\delta; d, \epsilon)$. This implies that if we draw a random bipartite graphs that has parameters in the region below the curve of $\rho^{exp}(\delta; d, \epsilon)$ then with probability approaching one exponentially in n that graph is a (k, d, ϵ)-lossless expander.
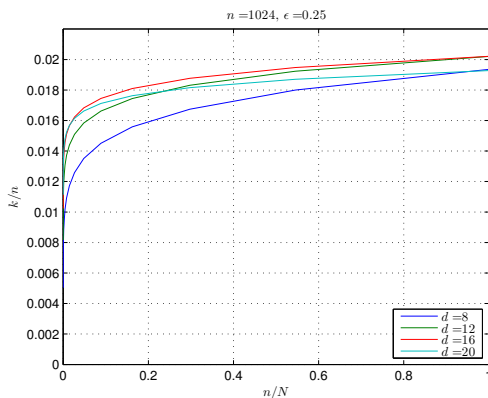

Fig. 3. Phase transition plots of $\rho^{exp}(\delta; d, \epsilon)$ for fixed $\epsilon = 1/6$ and $n = 2^{10}$ with d varied.

Figure 3 shows a plot of what $\rho^{exp}(\delta; d, \epsilon)$ converge to for different values of d with ϵ and n fixed. It is interesting to note how increasing d increases the phase transition up

to a point then it decreases the phase transition. Essentially beyond $d = 16$ there is inconsequential gain in increasing d. This vindicates the use of small d in most of the numerical simulations involving the class of matrices considered here. Note the vanishing sparsity as the problem size (k, n, N) grows while d is fixed to a small value of 8.

Corollary 6 can also be arrived at based on similar probabilistic constructions of expander graphs first proven by Pinsker in [9] with more recent proofs in [10], [6]. To put our results in perspective, we compare them to the phase transitions derived from the constants from the construction in [10], shown in Figure 4.
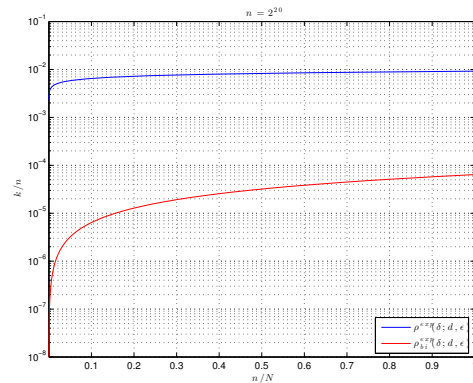

Fig. 4. A comparison of ρ^{exp} in Corollary 6 to ρ_{bi}^{exp} derived from the alternative construction proven in [10].

Furthermore, for moderate values of ϵ this allows us to make quantitative sampling theorems for some compressed sensing reconstruction algorithms. As usual in compressed sensing, in addition to ℓ_1-minimization quite a few *combinatorial* greedy algorithms have been proposed for these sparse non-mean zero matrices. These algorithms iteratively locates and eliminate large (in magnitude) components of the vector, [1]. They include Sequential Sparse Matching Pursuit (SSMP), see [11]; and Expander Recovery (ER), see [3]. Besides, theoretical guarantees have been given for ℓ_1 recovery and some of the greedy algorithms including SSMP and ER. Base on these theoretical guarantees, we derived sampling theorems and present here phase transition curves which are plots of phase transition functions $\rho^{alg}(\delta; d, \epsilon)$ of algorithms such that for $k/n \to \rho < (1 - \gamma)\rho^{alg}(\delta; d, \epsilon)$, $\gamma > 0$, a given algorithm is guaranteed to recovery all k-sparse signals with overwhelming probability approaching one exponentially in n.

Figure 5 compares the phase transition of thee above mentioned algorithms. Remarkably, for ER recovery is guaranteed for a larger portion of the (δ, ρ) plane than is guaranteed by the theory for ℓ_1-minimization using sparse matrices; however, ℓ_1-minimization has a larger recovery region than does SSMP. Figure 6 shows a comparison of the phase transition of ℓ_1-minimization as presented by Blanchard et. al. in [12] for dense Gaussian matrices based on RIC_2 analysis and the phase transition we derived here for the sparse binary matrices coming from lossless expander based on RIC_1 analysis. This
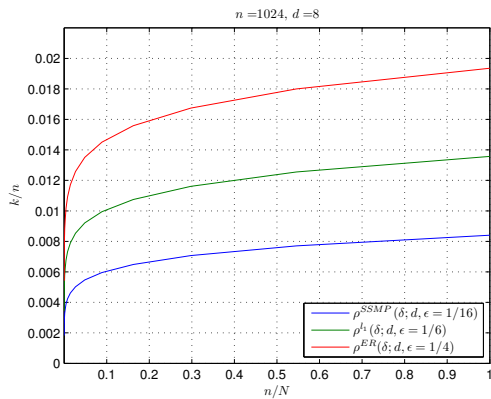
Fig. 5. Phase transition curves $\rho^{alg}(\delta; d, \epsilon)$ computed over finite values of $\delta \in (0, 1)$ with d fixed and the different ϵ values for each algorithm - 1/4, 1/6 and 1/16 for ER, ℓ_1 and SSMP respectively.

shows a significant difference between the two with sparse matrices having better performance guarantees. However, these improved recovery guarantees are likely more due to the closer match of the method of analysis than to the efficacy of sparse matrices over dense matrices.
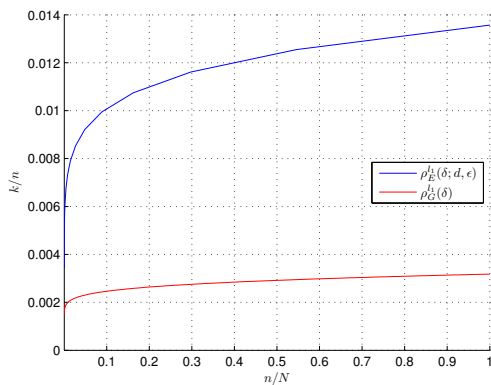

Fig. 6. Phase transition plots of ℓ_1, $\rho_G^{\ell_1}(\delta)$, for Gaussian matrices derived using RIC_2 and $\rho_E^{\ell_1}(\delta; d, \epsilon)$ for adjacency matrices of expander graphs with $n = 1024$, $d = 8$, and $\epsilon = 1/6$.

IV. SKETCH OF MAIN PROOF

Due to space constraints the details of the proofs are skipped and the interested reader is referred to [8]. It is however important to briefly describe the key innovations in the derivation of the main result, Theorem 4.

For one fixed set of columns of A, denoted A_S, the probability in (1) can be understood as the cardinality of the unions of nonzeros in the columns. Our analysis of this probability follows from a nested unions of subsets using a *dyadic splitting* technique. Given a starting set of columns we recursively split the number of columns from this set and the resulting sets into two sets of cardinality of the ceiling and floor of the cardinality of their union until a level when the cardinalities are at most two. Resulting from this type of

splitting is a regular binary tree where the size of each child is either the ceiling or the floor of the size of it's parent set. The probability of interest becomes a product of the probabilities involving all the children from the dyadic splitting of A_s. The proof therefore reduces to upper bounding this product.

Furthermore, in the binary tree resulting from our dyadic splitting scheme the number of columns in the two children of a parent node is the ceiling and the floor of half of the number of columns of the parent node. At each level of the split the number of columns of the children of that level differ by one. The enumeration of these two quantities at each level of the splitting process is necessary in the computation of the bound in (1). This led to another novel technical result in our derivation, i.e. *dyadic splitting lemma* (Lemma 2.5 in [8]).

V. CONCLUSIONS

This work derived bounds on the tail probability of the cardinality of the *set of neighbours* of expander graphs resulting into better order constants than the standard probabilistic construction. Using this bound and RIC_1 analysis, we deduce sampling theorems for the existence of expander graphs and their adjacency matrices. The derivation of the tail bound used a novel technique of *dyadic set splitting*. We also compared quantitatively, performance guarantees of compressed sensing algorithms which show greater phase transitions for ER than ℓ_1-minimization which in turn is greater than SSMP. A comparison of ℓ_1-minimization for dense and sparse matrices shows a higher phase transition for sparse matrices.

REFERENCES

[1] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*. IEEE, 2008, pp. 798–805.

[2] R. Berinde and P. Indyk, "Sparse recovery using sparse random matrices," *preprint*, 2008.

[3] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *Information Theory, IEEE Transactions on*, vol. 55, no. 9, pp. 4299–4308, 2009.

[4] W. Xu and B. Hassibi, "Further results on performance analysis for compressive sensing using expander graphs," in *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*. IEEE, 2007, pp. 621–625.

[5] J. Blanchard and J. Tanner, "GPU accelerated greedy algorithms for compressed sensing," *Preprint*, 2012.

[6] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson, "Randomness conductors and constant-degree lossless expanders," in *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*. ACM, 2002, pp. 659–668.

[7] S. Hoory, N. Linial, and A. Wigderson, "Expander graphs and their applications," *Bulletin of the American Mathematical Society*, vol. 43, no. 4, pp. 439–562, 2006.

[8] B. Bah and J. Tanner, "Vanishingly sparse matrices and expander graphs, with application to compressed sensing," *arXiv preprint arXiv:1207.3094*, 2012.

[9] M. Pinsker, "On the complexity of a concentrator," in *7th annual teletraffic conference*, 1973, p. 318.

[10] R. Berinde, "Advances in sparse signal recovery methods," Master's thesis, Massachusetts Institute of Technology, 2009.

[11] R. Berinde and P. Indyk, "Sequential sparse matching pursuit," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, 2009, pp. 36–43.

[12] J. Blanchard, C. Cartis, J. Tanner, and A. Thompson, "Phase transitions for greedy sparse approximation algorithms," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 188–203, 2011.