FAST COMPRESSIVE SAMPLING WITH STRUCTURALLY RANDOM MATRICES

*Thong T. Do[†], Trac D. Tran[†] * and Lu Gan[‡]*

[†] Department of Electrical and Computer Engineering
The Johns Hopkins University
[‡]Department of Electrical Engineering and Electronics
The University of Liverpool, UK

ABSTRACT

This paper presents a novel framework of fast and efficient compressive sampling based on the new concept of structurally random matrices. The proposed framework provides four important features. (i) It is universal with a variety of sparse signals. (ii) The number of measurements required for exact reconstruction is nearly optimal. (iii) It has very low complexity and fast computation based on block processing and linear filtering. (iv) It is developed on the provable mathematical model from which we are able to quantify trade-offs among streaming capability, computation/memory requirement and quality of reconstruction. All currently existing methods only have at most three out of these four highly desired features. Simulation results with several interesting structurally random matrices under various practical settings are also presented to verify the validity of the theory as well as to illustrate the promising potential of the proposed framework.

Index Terms— Fast compressive sampling, random projections, nonlinear reconstruction, structurally random matrices

1. INTRODUCTION

In the compressive sampling framework [1], if the signal is compressible, i.e., it has a sparse representation under some linear transformation, a small number of random projections of that signal contains sufficient information for exact reconstruction. The key components of compressive sampling are the sensing matrix at the encoder that must be highly incoherent with the sparsifying transformation of the signal and a non-linear reconstruction algorithm at the decoder such as basis pursuit, orthogonal matching pursuit (OMP), iterative thresholding associated with projection onto convex sets and their variants that attempt to find the sparsest signal from the received measurements.

The first family of sensing matrices for l_1 based reconstruction algorithms consists of random Gaussian/Bernoulli matrices (or more generally, sub-Gaussian random matrices [2]). Their main advantage is that they are universally incoherent with any sparse signal and thus, the number of compressed measurements required for exact reconstruction is almost minimal. However, they inherently have two major drawbacks in practical applications: huge memory buffering for storage of matrix elements and high computational complexity due to their completely unstructured nature [3].

The second family is partial Fourier [3] (or more generally, random rows of any orthonormal matrix). Partial Fourier exploits the fast computational property of Fast Fourier Transform (FFT)

and thus, reduces significantly the complexity of a sampling system. However, partial Fourier matrix is only incoherent with signals which are sparse in the time domain, severely narrowing its scope of applications. Recently, random filtering was proposed empirically in [4] as a potential sampling method for fast low-cost compressed sensing applications. Unfortunately, this method currently lacks a theoretical foundation for quantifying and analyzing its performance.

In this paper, we propose a novel framework of compressive sampling for signals that can be sparse in any domain other than time. Our approach is based on the new concept of structurally random matrices. Here, we define a structurally random matrix as an orthonormal matrix whose columns are permuted randomly or the sign of its entries in each column are reversed simultaneously with the same probability. A structurally random matrix inherently possesses two key features: it is nearly incoherent with almost all other orthonormal matrices (except the identity matrix and extremely sparse matrices); it may be decomposed into elementwise product of a fixed, structured and in many cases, block diagonal matrix with a random permutation or Bernoulli vector.

Our algorithm first pre-randomizes the signal using one of these two random vectors and then applies block transformation (or linear filtering), followed by subsampling to obtain the compressed measurements. At the decoder, the reconstruction algorithm uses corresponding adjoint operators, then proceeds to find the sparsest signal via the conventional l_1-norm minimization decoding approach of solving a linear programming problem or employing greedy algorithms such as basis pursuit.

This approach may be regarded as the efficient hybrid model of two current methods: completely random Gaussian/Bernoulli matrices and partial Fourier. It retains almost all desirable features of these aforementioned methods while simultaneously eliminates or at least minimizes their significant drawbacks. A special case of our method was mentioned of its efficiency in [5, 6] (as the so-called Scrambled/Permuted FFT) but without an analysis of its performance.

The remainder of the paper is organized as follow. Section 2 gives fundamental definitions and theoretical results of incoherence of structurally random matrices. Section 3 presents theoretical results of compressive sampling performance based on the proposed structurally random matrices. Simulation results are presented in Section 4 and conclusions and future works are presented in Section 5. Due to lack of space, only heuristic arguments and proof sketches are provided. Detail proofs of these theorems and associated lemmas are provided in the journal version of this paper [7].

2. COHERENCE OF STRUCTURALLY RANDOM MATRICES

2.1. Basic Definitions

Definition 2.1.1: Given a unit-length vector $x \in \mathcal{R}^n$ and a *random seed vector* $\pi \in \mathcal{R}^n$, define a new random vector y as $y = \pi(x)$. We consider the following two models of π.

(i) *Global randomization model*: π is a uniformly random permutation of the set $\{1, 2, ..., n\}$, assign $y(\pi(i)) = x(i)$ for all $i = 1, .., n$.

(ii) *Local randomization model*: π is a vector of i.i.d Bernoulli random variables ($p = 1/2$), assign $y = x \circ \pi$, where $\circ$ is the element-wise product.

Definition 2.1.2: Given a *fixed orthonormal seed matrix* $A \in \mathcal{R}^{n \times n}$ and a *random seed vector* $\pi \in \mathcal{R}^n$, a (row-based) structurally random matrix is generated by applying one of two randomization models in Definition 2.1.1 to all rows of the matrix A. Denote this random matrix as $\pi(A)$.

Lemma 2.1.1 Given a structurally random matrix $\pi(A) \in \mathcal{R}^{n \times n}$ and a fixed vector $x \in \mathcal{R}^n$, $\pi(A)x = A\pi(x)$.

The lemma above simply states that we can implement a fast computation of a product of a structurally random matrix with a signal by first randomizing the signal using the *random seed vector* and then applying fast transformation of the *fixed seed matrix* to the randomized signal. This feature is, indeed, the spirit of our work.

2.2. Problem Formulation and Main Results

Given a structurally random matrix $\Phi \in \mathcal{R}^{n \times n}$ (whose subset of rows is a sensing matrix) and some fixed orthonormal matrix $\Psi \in \mathcal{R}^{n \times n}$ (i.e. the sparsifying matrix) and assume that the average support of rows of Φ is s, i.e. each row of Φ has s nonzero entries on average. We are interested in the coherence of Φ and Ψ [3] w.r.t. parameters n and s. The relationship of this coherence with minimal number of measurements required for exact reconstruction in the compressive sampling framework [3] is provided in Section 3.

Assumption 2.2.1: Our ultimate goal is to design the sensing matrix Φ to be both simple and efficient. Thus, we would like to consider the case that absolute nonzero entries of Φ are roughly equal, i.e. they are in the order of $O(1/\sqrt{s})$. For the sake of simplicity, these absolute values may be set freely to be $1/\sqrt{s}$ when necessary. Note that this assumption does not violate the orthonormality of Φ because there exists families of orthonormal matrices whose all absolute nonzero entries are $1/\sqrt{s}$, for example, a Kronecker product of a Hadamard matrix and an identity matrix.

Assumption 2.2.2: To prevent the degenerate case, i.e. Φ and Ψ become identity matrices or extremely sparse matrices, we need another reasonable assumption that the average row's and column's supports of these matrices is at least $\log n$ – a quite realistic range with known sparsifying matrices.

With the aforementioned assumptions, the following theorems hold for structurally random matrices generated by the *local randomization model*.

Theorem 2.2.1: The coherence of Φ and Ψ is not larger than $O(\sqrt{\log n / s})$ with probability at least $1 - O(1/n)$.

Theorem 2.2.2: The 2-Babel cumulative coherence [8] of Φ and a uniformly random set of k columns of Ψ is not larger than $O(\sqrt{k/n + \sqrt{k} \log n^{3/2}/s})$ with probability at least $1 - O(1/n)$.

In the case that the sensing matrix Φ is generated by a *global randomization model*, the results are weaker because we need an additional assumption on Φ and Ψ below. This is mainly due to our method of approximating a random permutation vector by a weakly dependent random vector.

Assumption 2.2.3: If Φ is generated by a *global randomization model*, every column of Ψ has sum of its entries equals to zero. In addition, we limit our consideration to the case when Ψ (Φ) is dense and Φ (Ψ) has average row and column supports s to be in the order of $o(\sqrt{n})$ (i.e. $s/\sqrt{n}$ goes to zero when n goes to infinity).

Theorem 2.2.3: The theorems 2.2.1 and 2.2.2 also hold when Φ is generated by a *global randomization model* and Assumption 2.2.3 is satisfied.

Proof sketch:

The main technical tools are large deviation inequalities of sum of independent random variables. In particular, the Bernstein's and Hoeffding's concentration inequalities of sum of independent random variables [9] are used very frequently. Key arguments are as follows.

(i) Of two models, the global randomization is harder to analyze due to its combinatorial nature. We approximate it by the following proposition which is proved by using *Mutual Information* to compute the asymptotical distance between join probability and product of marginal probability functions.

Proposition 2.2.1: If entries of a vector $\mathbf{x} \in \mathcal{R}^n$ are distinct and an integer s of the order of $o(\sqrt{n})$, the randomized vector $y = \pi(x)$ may be asymptotically approximated by the s-independent random vector, i.e. entries of y are identical distributed random variables and entries in every subgroup of size s are mutually independent.

(ii) The asymptotical behavior of the above theorems are described by the following proposition.

Proposition 2.2.2: The normalized inner product of a dense row of Φ and a dense column of Ψ is a random variable with zero-mean and unit variance. In addition, it is asymptotically normal when n goes to infinity.

(iii) To quantitatively measure the tail probability of coherence, we use the following proposition which is directly derived from Bernstein's deviation inequality and a union bound for the supremum of a random process.

Proposition 2.2.3 Let $x_1, x_2, ...x_n$ be a sequence of independent, bounded, discrete random variables with zero-mean. Let $s = \sum_i x_i \in S$ and its variance denoted by σ^2. Also, define $M = \sup_i |x_i|$ and $K = \max(O(M^2 \log n), O(\sigma^2))$. If the cardinality of S is n^2, then $\lambda = O(\sqrt{K \log n})$ satisfies $P(\sup_{s \in S} |s| > \lambda) < O(1/n)$.

Theorem 2.2.1 is then the direct corollary of this proposition under the assumptions 2.2.1 and 2.2.2. Notice that in this case, s is the inner product of a row of Φ and a column of Ψ and assumption 2.2.2 implies the maximum absolute entries of Φ and Ψ are not larger than $\sqrt{1/\log n}$.

Theorem 2.2.2 uses the main assumption that the subset of k columns of Ψ are chosen uniformly randomly. Notice that inner products of a row of Φ and these k random columns of Ψ is the sequence of k independent random variables. In addition, these random variables are upper-bounded as the result of Theorem 2.2.1. Define the probabilistic event that *all* these random variables are upper-bounded by $O(\sqrt{\log n/s})$. Applying the Hoeffding's concentration inequality for this sequence of independent random variables and then the union bound for the supremum of a random process and the conditional probability inequality to remove the conditioned event results in Theorem 2.2.2.

Theorem 2.2.3 uses the similar set of arguments of sum of independent random variables with the notice to Proposition 2.2.1 and the additional assumption 2.2.3.

3. COMPRESSIVE SAMPLING USING STRUCTURALLY RANDOM MATRICES

In the compressive sampling framework, the number of measurements required for exact reconstruction are directly proportional to the coherence of the sensing matrix and the sparsifying matrix [3]. If this coherence was large (i.e. these matrices were not incoherent), compressive sampling loses its effectiveness and can even become useless. However, the incoherence between a sensing matrix and a sparsifying matrix is not sufficient to guarantee exact reconstruction. The other important condition is the stochastic independence of compressed measurements. In subGaussian matrix framework - a generalization of Gaussian/Bernoulli matrices, rows of the random matrix are required to be stochastically independent. In partial Fourier framework, a random subset of rows is used to generate stochastic independence among these deterministic rows.

If Φ is a structurally random matrix, its rows are not stochasticaly independent because they are randomized from the same *random seed vector* and thus are correlated. This is the main difference between a structurally random matrix and a subGaussian matrix. Relaxing the independence among its rows enables a structurally random matrix to have some particular structure with fast computation. The independence of compressed measurements is then generated by using the same method of partial Fourier - a random subset of rows of a structurally random matrix.

Assumption 3.2.1: Suppose that matrices Φ and Ψ satisfy our assumptions 2.2.1, 2.2.2. If the global randomization model is used, the additional weaker version of the assumption 2.2.3, i.e. every column of Ψ, except at most one, has sum of its entries equals to zero, is also required. Also, assume that the signal is k-sparse in the domain Ψ and satisfies the condition of uniformly random sign as in [3]. The concept of *non-uniformly* exact reconstruction as defined in [3] is also used.

With the above assumption, the following theorems are about the number of measurements required for exact reconstruction when a structurally random matrix Φ is used to acquire compressed measurements

Theorem 3.1: (*Non-uniform exact reconstruction*) A random subset of compressed measurements of size m guarantees exact reconstruction with probability at least $min\{(1-\delta), 1-O(1/n)\}$, provided $m = O((kn/s) \log n \log n/\delta)$.

Theorem 3.2: (*Non-uniform exact reconstruction for uniformly sparse signals*) If the signal is *uniformly sparse*, i.e. its nonzero entries are uniformly randomly distributed in its sparse domain, a random subset of compressed measurements of size m guarantees exact reconstruction with probability at least $min\{(1-\delta), 1-O(1/n)\}$, provided that $m = O((k + (n/s)\sqrt{k}\sqrt{log^3 n}) \log n/\delta)$.

Theorem 3.2.1 implies that if the sensing matrix is dense (i.e. $s = n$), the number of measurements is nearly minimal (except for the $\log n$ factor), regardless of the sparsifying matrix Ψ. In general, the number of measurements is linearly inverse-proportional to the average sparsity of the sensing matrix. Theorem 3.2.2 shows a significant improvement that this inverse-proportional relationship is, indeed, sub-linear when k and s are in the order of $O(\sqrt{n})$, which is usually the case.

Proof sketch:

The proof of these theorems follows a similar set of arguments in [3] and the above results of coherence of a structurally random matrix. First notice that the structurally random matrix is always orthonormal. Its orthonormality is untouched by the global or local randomization operators. Thus, the structually random matrix may be considered as the *deterministic* orthonormal matrix under the set of arguments in [3]. The innovating part is that the coherence of this sensing matrix is *explicitly* specified by Theorems 2.2.1, 2.2.2 and Theorem 2.2.3.

For Theorem 3.1: The main components of the proof are Theorems 2.2.1, 2.2.2 and Theorem 1.1 in [3]. The technical tool used is the conditional probability inequality. Define the probabilistic event that the coherence of Φ and Ψ is not larger than $\sqrt{logn/s}$. Theorem 2.2.1 says that this event occurs with probability at least $1 - O(1/n)$. After conditioning on this event, we apply the Theorem 1.1 in [3] because all matrices are orthogonal. Applying the conditional probability inequality to remove the conditioned event results in Theorem 3.1.

For Theorem 3.2: This theorem exploits the concept of 2-Babel cumulative coherence [8] to give a tighter bound of number of measurements required for exact reconstruction when signals are uniformly sparse. Let T be the index set (of size k) of nonzero entries of the sparse signal and U_T be the matrix of size $n \times k$ corresponding to columns of the product matrix of Φ and Ψ indexed by T. Let u^i be row vectors of U_T. As arguments in [3], the number of measurements required for exact reconstruction is proportional to $max_{1 \leq i \leq n} \| u^i \|$. A normal treatment is to approximate this term by the coherence of Φ and Ψ and the Theorem 3.1 follows. However, notice that if the nonzero entries of the sparse signal is uniformly randomly distributed, this term may be better approximated by the 2-Babel cumulative coherence between Φ and Ψ. Applying Theorem 2.2.2 for the upper bound of this cumulative coherence and using the same technical tools of the conditional probability inequality and a probabilistic event of upper bound of the cumulative coherence results in the Theorem 3.2

4. SIMULATION RESULTS

In this section, we illustrate the performance of the proposed framework with other existing ones in large scale applications. Natural images of size 512×512 are used as input signals of length 2^{18}. The sparsifying operator Ψ is chosen as the popular Daubechies $9/7$ wavelet transform. The l_1 based linear programming solver is based on the GPSR algorithm [10]. Figure 1 and Figure 2 illustrate the PSNR results of reconstructed *Lena* and *Boat* images, respectively.

Our main interest here is the performance of highly sparse sensing matrices. In this experiment, we choose block-diagonal Hadamard and DCT matrices. Two sizes are considered: 32×32 and 512×512. Notice that the density rate of those sensing matrices are only 2^{-13} and 2^{-9}, which are highly sparse. The global and local randomization models are used for 32×32 and 512×512 matrices, respectively. The method of random subset of measurements is used to acquire compressed measurements. In Figure 1 and Figure 2, DCT32 and Ha32 correspond to the results of 32×32 block-diagonal DCT and Hadamard matrices, respectively. Likewise, DCT512 and Ha512 represent the results of 512×512 DCT and and Hadamard matrices, respectively. For comparison purposes, we also implement other popular sampling operators. The first one is the "brute force (BF)" method, which transforms the signal into its sparse representation and uses partial FFT (PFFT) to sample in the transform domain. Since the FFT and the identity matrix are
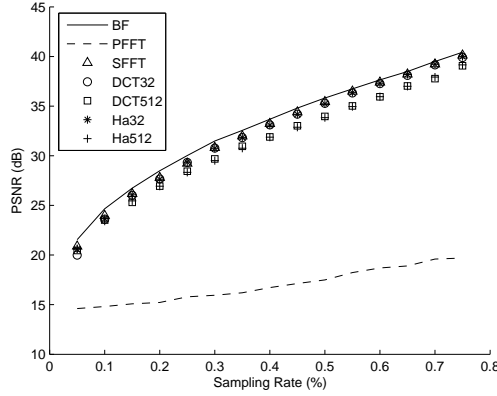
5. CONCLUSIONS AND FUTURE WORKS

In this paper, a novel approach of fast compressive sampling is proposed based on the new concept of structurally random matrices. The essence of a structurally random sensing matrix is that it decouples the properties of sub-Gaussian and stochastic independence of rows of a sub-Gaussian matrix. The sub-Gaussian property of rows is realized by random seed vector with appropriate distributions. The stochastic independence property of rows is realized by a random subset of rows as in the partial Fourier approach. Via decoupling, the sensing ensemble can be implemented very efficiently with fast integer computation and streaming capability.

The proposed framework may be regarded as an innovating mixture of partial Fourier and sub-Gaussian matrices. It provides the *universality* feature for partial Fourier by pre-randomizing signals and the *fast and efficient computation* for sub-Gaussian matrices by realizing the sub-Gaussian and the independence properties of rows separately.

While the stochastic independence of compressed measurements is one of sufficient conditions for exact reconstruction, the highly incoherence between sensing matrices and signals is to guarantee the minimal number of measurements required. However, it is still unknown whether or not stochastic independence of measurements is necessary for optimal performance. In addition, how to quantify this relationship between independence, incoherence and performance is another interesting open problem. Our future works will be in this direction.

6. REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, pp. 489 – 509, Feb. 2006.

[2] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for bernoulli and subgaussian ensembles.," *Preprint*, Aug. 2006.

[3] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23(3), pp. 969–985, 2007.

[4] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," *Proc. IEEE ICASSP*, vol. 3, pp. 872–875, Toulouse, May 2006.

[5] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, Aug. 2006.

[6] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Fast reconstruction of piecewise smooth signals from incoherent projections," *SPARS'05*, Rennes, France, Nov 2005.

[7] T. Do, T. D. Tran, and L. Gan, "Fast compressive sampling using structurally random matrices," *to be submitted to IEEE Trans. on Information Theory*, 2007.

[8] K. Schnass and P. Vandergheynst, "Average performance analysis for thresholding," *IEEE Signal Processing Letters*, 2007.

[9] G. Lugosi, "Concentration-of-measure inequalities," *Lecture notes*, Feb. 2006.

[10] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction," *to appear in IEEE Journal of Selected Topics in Signal Processing*, 2007.

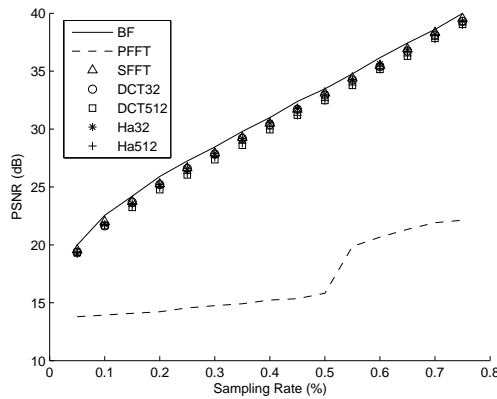**Fig. 1**. Reconstruction results for 512×512 *Lena* image.

Fig. 2. Reconstruction results for 512×512 *Boat* image.

perfectly incoherent, the brute force method seems to be powerful in theory. However, it is often prohibited from practice because its computation is too costly and it requires knowing the sparsifying matrix at the encoder. In general, it is required that the system to be able to sample signals *without* prior knowledge of their sparsifying matrices. Other methods are also considered including the partial Fourier and scrambled FFT (SFFT) [5, 6], a special case of our proposed framework. The family of sub-Gaussian matrices such as the Gaussian/Bernoulli is not studied here as it is impossible to implement them in such a large scale experiment like this.

From Figure 1 and Figure 2, one can observe that the partial Fourier operator is not the appropriate matrix for sensing smooth signals as it is highly coherent with such signals. These figures also depict that the performances of four proposed sensing matrices, which are very sparse and require much lower computation and memory buffering than all other methods, are roughly equal to those of the most complicated, "brute force" method. The difference is only about 1dB which is a reasonable sacrifice. They are much simpler and require much lower computation than the Scrambled FFT. In addition, thanks to the streaming feature of the local randomization model, two matrices DCT512 and Ha512 can provide the streaming capability for the sampling system effectively. These simulation results are, indeed, perfectly matched with our theoretical results above.