Compressive Sensing

By Lin Yu ,Jing Lifu ,Wang Yu,Zhao Qianfu

May 2

Abstract

Compressed Sensing or Compressive Sensing is about acquiring a sparse signal in a most efficient way (subsampling) with the help of an incoherent projecting basis.

Three main tasks in compressive sensing are as follows :

1. search for bases or dictionaries in which sets of signals can be decomposed in a sparse manner.

2. find and quantify specific measurements tools that are incoherent with said dictionaries.

3. reconstruct the original signal as accurate as possible.

In other words, once a signal is known to be sparse in a specific basis, one of the main challenge is to find a set of measurement tools (producing the compressed measurements) and the attendant nonlinear solver that reconstructs the original full signal. There are theoretical results yielding the minimum number of required measurements needed to produce the original signal given a specific pair of measurement matrices and nonlinear solvers. In all cases, the expected number of compressed measurements is expected to be low relative to traditional Nyquist sampling constraints.

Our attention focuse on the second topic which is mentioned above. Different incoherent matrixes are involved and we also make comparison between them to point out the advantages and disadvantages of each method.

1 Introduction

In the compressive sampling framework, 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 nonlinear 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 ?nd the sparsest signal from the received measurements.

The first family of sensing matrices for L1 based reconstruction algorithms consists of **random Gaussian/Bernoulli matrices** (or more generally, **sub-Gaussian random matrices** [2]). Their main advan-

tage 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.

Also, a novel framework of compressive sampling for signals that can be sparse in any domain other than time were propose by Thong T. Doy, Trac D. Trany and Lu Gan in their paper. This method is based on the new concept of **structurally random matrices**,which can be defined 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.The algorithm retains almost all desirable features of these aforementioned methods while simultaneously eliminates or at least minimizes their significant drawbacks.

We first discuss the basic constraint i.e. RIP which all measurements matrix should obey .By the way, many new restrict prerequisites based on RIP are proposed already. Then we talk about some measurements matrix and compare their performance with the help of codes which is downloaded from *http://www.dsp.ece.rice.edu/cs* . Among them ,we emphasis on random filter and SRM(structurally random matrices).

2 The Restricted Isometry Property (RIP)

In this section, we introduce a key notion that has proved to be very useful to study CS; the so-called restricted isometry property (RIP) [5].

Definition 2.1

For each integer S = 1, 2, . . . , define the isometry constant δ_S of a matrix A as the smallest number such that

$$(1-\delta_S) \parallel x \parallel_{\ell_2}^2 \leq \parallel Ax \parallel_{\ell_2}^2 \leq (1+\delta_S) \parallel x \parallel_{\ell_2}^2 \qquad (1)$$

holds for all S-sparse vectors x. We will loosely say that a matrix A obeys the RIP of order S if δ_S is not too close to one. When this property holds, A approximately preserves the Euclidean length of S-sparse sig-

nals, which in turn implies that S-sparse vectors cannot be in the null space of A. (This is useful as otherwise there would be no hope of reconstructing these vectors.) An equivalent description of the RIP is to say that all subsets of S columns taken from A are in fact nearly orthogonal (the columns of A cannot be exactly orthogonal since we have more columns than rows).To see the connection between the RIP and CS, imagine we wish to acquire S-sparse signals with A. Suppose that $\delta_2 S$ is sufficiently less than one. This implies that all pairwise distances between S-sparse signals must be well preserved in the measurement space. That is,$(1-\delta_{2S}) \parallel x_1-x_2 \parallel_{\ell_2}^2 \leq \parallel Ax_1-Ax_2 \parallel_{\ell_2}^2 \leq (1+\delta_{2S}) \parallel x_1-x_2 \parallel_{\ell_2}^2$ holds for all S-sparse vectors x_1, x_2. As demonstrated in the next section, this encouraging fact guarantees the existence of efficient and robust algorithms for discriminating S-sparse signals based on their compressive measurements.

3 Several Sensing Matrices

Since sub-Gaussian random matrices has been referred in last report,and this method is not an efficient choice,it will not be introduced any more in this section.

3.1 Random Filtering

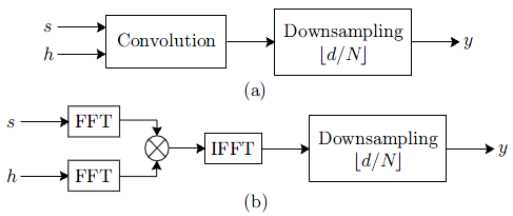

Figure 1: Block diagrams for signal acquisition through random filtering: (a) using convolution; (b) using FFT/IFFT. The FIR filter h has random taps, which must be known in order to recover the signal s from the compressed data y.

We propose random filters as a new paradigm for compressive signal acquisition. Our approach captures a signal s by convolving it with a random-tap FIR filter h and then downsampling the filtered signal to obtain a compressed representation y. Figure 1 illustrates the measurement process. Reconstruction of s involves a nonlinear algorithm.

This method has several benefits:

1.measurements are time-invariant and nonadaptive;

2.measurement operator is stored and applied efficiently;

3.we can trade longer filters for fewer measurements;

4.it is easily implementable in software or hardware;

5.it generalizes to streaming or continuous-time signals.

Encoding Draw a random filter h of length B. And the filter requires just $O(B)$ storage. To take N measurements of a signal s of length d, we must calculate:

$$y = \mathcal{D} \downarrow (h * s) \tag{2}$$

where $\mathcal{D} \downarrow$ downsamples by a factor of $\lfloor d/N \rfloor$. Note that, be cause this process is linear, the map from the signal s to th summary y can be viewed as y Φ s, where Φ is an $N \times d$ matrix. This matrix is banded and quasi-Toeplitz: each row has B nonzero entries, and each row of Φ is a copy of the row above, shifted right by $\lfloor d/N \rfloor$ places.

Method 1: The first method for calculating the measurements, illustrated in Figure 1(a), performs linear convolution and downsampling simultaneously. For n , ,...,N-1 , the n-th measurement is calculated as

$$y(n) = \Sigma_{j=0}^{B-1} s(n\lfloor d/N \rfloor + j)h(B-j) \qquad (3)$$

Computing N measurements requires $O(BN)$ arithmetic operations. This method can be applied in systems where the input s is streaming, since the measurements are localized in time and also time-invariant.

Method 2: The second method, illustrated in Figure 1(b), uses FFTs to calculate the convolution. In this case, we compute

$$\mathbf{y} = \mathcal{D}_{\downarrow}\mathcal{F}^{-1}H(\omega)S(\omega) \qquad (4)$$

which is equivalent to using a circulant Φ matrix. The cost of computing the measurements is $O(d\log(d))$, independent of the filter length or the number of measurements. Compared to Method 1, this calculation may be faster if the filter has many taps. Note, however, that the entire signal must be presented at once.

It appears that these two encoding methods are at least as efficient as anything described in the CS literature. We also note that filtering can be performed with other standard methods, such as overlapCadd, but we omit this discussion.

One expects that signals sparse in the time domain, i.e., $\Psi = I$, are the most difficult to acquire with random filters because of high coherence. Yet we present empirical evidence that random filters are effective for recovering timesparse signals: a random filter of length d performs as well as a fully Gaussian matrix. When the filter length decreases, the number of measurements increases somewhat. For signals sparse in the frequency domain, the number of measurements depends weakly on the filter length; a four-tap filter already yields good reconstruction probability.

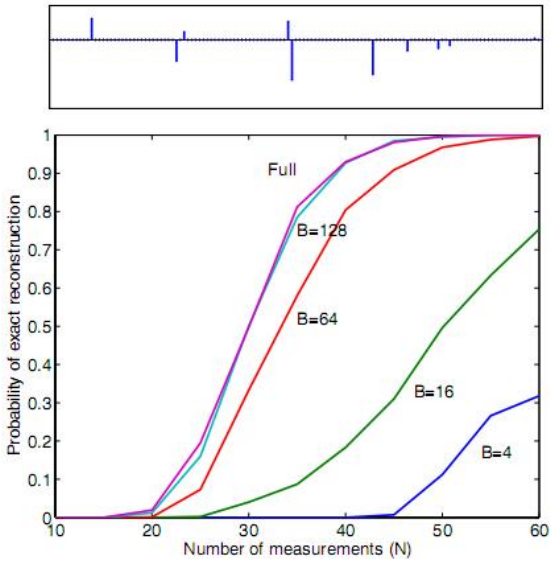

Figure 2: Timesparse signals.

NUMERICAL RESULTS

We begin with signals that are sparse in the time domain, i.e., $\Psi = I$. Recall that this case is challenging due to high coherence. We choose the signal length $d = 128$ and sparsity $m = 10$. Figure 2 displays the probability of success for several filter lengths, in comparison with fully random measurements. Observe that the two longest filters ($B = 64$ and 128) succeed almost as well as the fully Gaussian matrix, despite having far fewer degrees of freedom.

We now consider signals that are sparse in the Fourier domain, i.e., $\Psi = \mathcal{F}$. As above, the sig-

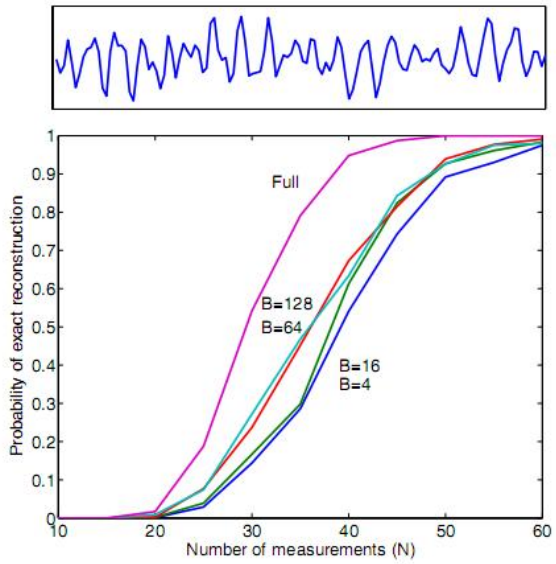

Figure 3: Fourier-sparse signals.

nal length $d = 128$ and sparsity $m = 10$. Figure 3 displays the probability of success for several filter lengths. Note that all four filters yield similar probabilities of success, which are slightly worse than the Gaussian matrix. The filter length has a limited impact since the Fourier basis is incoherent with the random filter.

3.2 Structurally Random Matrices(SRM)

Before we talk about SRM, I will introduce some basic definitions as follows:

Definition 3.3.1:

Given a unit-length vector $x \in R^n$ and a random seed vector $\pi \in R^n$, 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 3.3.2:

Given a fixed orthonormal seed matrix $A \in R^n$£n and a random seed vector $\pi \in R^n$, a (row-based) structurally random matrix is generated by applying one of two randomization models in Definition 3.3.1 to all rows of the matrix A. Denote this random matrix as $\pi(A)$.

Lemma 3.3.1 Given a structurally random matrix ($A \in R^n$ and a fixed vector $x \in R^n$, $\pi(A)x = A\pi(x)$.

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.

Then,given a structurally random matrix . $\Phi \in R^{n \times n}$ (whose subset of rows is a sensing matrix) and some fixed orthonormal matrix $\Psi \in R^{n \times n}$ (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.

In [1] the 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.

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 rows and columns supports of these matrices is at least $\log n - a$ quite realistic range with known sparsifying matrices.

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 $\Psi(\Phi)$ is dense and $\Phi(\Psi)$ 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).

3.3 Complexity

Results from the CS literature provide a benchmark for studying the performance of random filters. Taking N linear measurements of the signal δ can be viewed as multiplication $y = \Phi s$ by a $N \times d$ measurement matrix Φ. Several distributions for Φ have been proposed. If Φ is fully i.i.d. Gaussian, then several different algorithms can recover msparse signals from $N = O(m \log d)$ measurements. If Φ is a Rademacher (i.e., 1) matrix or a random row submatrix of the DFT matrix, then a similar number of measurements suffice in practice.

The costs for CS encoding and decoding depend significantly on the type of measurement matrix. Gaussian and Rademacher matrices require storage and computation $O(dN)$ for encoding. Fourier measurement matrices improve storage to $O(d)$ and encoding times to $O(d \log d)$. Two different algorithms, ℓ_1 minimization [1, 2] and OMP [3], are commonly used for signal reconstruction. The ℓ_1 minimization approach uses linear programming to solve the problem . Reconstruction costs via ℓ_1 minimization have not been reported, but one expects them to be $O(d3.5)$ in general. Greedy pursuit methods such as OMP attempt to build up an approximation to θ based on correlations between y and the columns of the matrix $\Phi\Psi$. OMP requires $O(mNd)$ arithmetic operations in general, but it can be improved to $O(md \log d)$ with Fourier measurements.

4 references

[1]Thong T. Do, Trac D. Tran, and Lu Gan, Fast compressive sampling with structurally random matrices. (Preprint, 2007)

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

[3]E. Cand‘es and J. Romberg, Sparsity and incoherence in compressive sampling, Inverse Problems, vol. 23(3), pp. 969C985, 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. 872C875, Toulouse, May 2006.

[5]E. Cands and T.Tao, Decoding by linear programming, IEEE Trans. Inform. Theory, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.