Compressive Sensing in Wireless Networks

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

March 29

Abstract

Conventional wisdom and common practice in acquisition and reconstruction of images from frequency data follow the basic principle of the Nyquist density sampling theory.The theorem specifies that to avoid losing information when capturing a signal, one must sample at least two times faster than the signal bandwidth.

Compressive sampling or compressed sensing tells us it is feasible to reconstruct signal at a rate significantly below the Nyquist rate.

Nowadays compressive sensing has implications in many fields such as statistics, information theory, coding theory, and theoretical computer science.

In this short report, firstly, we introduce the mathematical insights underlying this new theory. Then explain some of the interactions between compressive sensing and applications in wireless networking.

1 Introduction

As is said in those paper we have read , that most of the data we acquire from signals and images can be thrown away with almost no perceptual loss most of the data we acquire c. That means a lot of resources and time are wasted when we process signal in old ways. In fact, compressive sampling suggests ways to economically and efficiently translate analog data into already compressed digital form .

As is known to all, because typical signals have some structure, they can be compressed efficiently without much perceptual loss. The way we used before to compress a signal is to acquires the full signal firstly,then computes the complete set of transform coefficients, encode the largest coefficients and discard all the others.Take Fourier transform for example:

$$F(w) = [\int_{-\infty}^{+\infty} f(t)e^{-jwt}\, dt.] \tag{1}$$

It is clear that coefficients can be computed only after we get all the samples because of the orthogonal basis e^{jwt}.And then we know which coefficients are important.

The process of massive data acquisition followed by compression is obviously wasteful .In addition , when using the old methods to process signals,very serious problems will appear if we lose some of the most coefficient.

Because of the shortcomings of these existent methods ,we are surely to think : Cant we just directly measure the part that wont end up being thrown away? So in the recent years , the theory of compressive sampling has emerged .

The main purpose of this report is to summarize what we have read and learn and introduce some of the key mathematical insights underlying this new theory, and make a survey to its applications to wireless networking.

2 Overview on compressive sampling

To process a signal $x \in \mathbb{R}^{N\times 1}$ in the way of compressed sensing include three key step. Firstly , we try to find its M Liner Measurements ,$s = \Phi x$, where $s \in \mathbb{R}^{M\times 1}$. We can consider each row of Φ as sensor . Here Φ (Measurement matrix)should meet RIP to make sure that it is incoherence with the orthogonal basis we choose.Gaussian whitenoise matrix is an example . Secondly , we choose a suitable orthogonal basis ,such as Fourier basis and curvelet ,with which we can express vector x as $\Psi^H y$ ($\Psi \in \mathbb{R}^{N\times N}$).So y is Ψx , then we set all elements of y to zero except the K most important (large) ones and get y^. To find new and efficient basis may be a hard work and we only understand the Fourier one well so far . Last ,also the most important step is to recover the signal from y^. This is a convex problem , which is usually solved by minimizing the l-1 norm near-solution . There many accesses to this problem , among which ,OMP(Orthogonal Matching Pursuit) and BP(Basis Pursuit) is widely used . OMP seems more easy to understand and one of us is working on a little program based on OMP .

3 Incoherence And The Sensing Of Sparse Signal

3.1 Sparsity

Consider a vector $f \in \mathbb{R}^N$, which we expand in an orthonormal basis. Using the $N \times N$ basic matrix $\psi = [\psi_1\psi_2...\psi_n]$ f can be expressed as

$$f(t) = \sum_{i=1}^{N} x_i\psi_i(t). \tag{2}$$

where **x** is the the $N \times 1$ column vector of weighting coefficients of f, $x_i = \langle f, \psi_i \rangle$. To keep on using convenient matrix notations, we can write $x = \psi \times f$,which is equal to $f = \psi^* x$. Clearly, f and x are equivalent representations of the signal, with f in the time or space domain and x in the ψ domain.

If signal f is a linear combination of only $K(K << N)$ basis vectors, We say f is sparse. That means only K of the x_i coefficients in (1) are nonzero and $N - K$ are zero. The signal f is compressible if the representation (1) has just a few large coefficients and many small coefficients.

Now,let's talk about the natural signals. Many natural signals have concise representations when expressed in a convenient basis. Consider, for example, the image in Figure 1(a) and its wavelet transform in (b). Although nearly all the image pixels have nonzero values, the wavelet coefficients offer a concise summary: most coefficients are small, and the relatively few large coefficients capture most of the information.

By definition, $f_S := \Psi x_S$, where here and below, x_S is the vector of coefficients (x_i) with all but the largest S set to zero. This vector is sparse in a strict sense since all but a few of its entries are zero; we will call S-sparse such objects with at most S nonzero entries. Since Ψ is an orthonormal basis (or orthobasis), we have $\| f - f_S \|_{\ell_1}$ = S$\| x - x_S \|_{\ell_1}$, and if x is sparse or compressible in the sense that the sorted magnitudes of the (xi) decay quickly, then x is well approximated by x_S and, therefore, the error $\| f - f_S \|_{\ell_1}$ is small. In plain terms, one can throw away a large fraction of the coefficients without much loss.

3.2 Incoherence

Now let's explain the definition of incoherence. The coherence between the sensing basis ϕ and the representation basis ψ is

$$\mu(\phi, \psi) = \sqrt{n} \cdot \max_{1 \le k, j \le n} | \langle \varphi_k, \psi_j \rangle | \tag{3}$$

So, the coherence measures the largest correlation between any two elements of ϕ and ψ. If ϕ and ψ contain correlated elements, the coherence is large. Otherwise, it is small. As for how large and how small, it follows from linear algebra that $\mu(\phi, \psi) \in [1, \sqrt{n}]$. Compressive sampling is mainly concerned with low coherence pairs, and we now give examples of such pairs. In our first example, ϕ is the canonical or spike basis $\varphi_k = \delta(t - k)$ and ψ is the Fourier basis, $\psi_j(t) = \frac{1}{\sqrt{n}} e^{i2\pi jt/n}$. Since ϕ is the sensing matrix, this corresponds to the classical sampling scheme in time or space. The time-frequency pair obeys $\mu(\phi, \psi) = 1$ and, therefore, we have maximal incoherence. Further, spikes and sinusoids are maximally incoherent not just in one dimension but in any dimension.

4 Signal recovery

4.1 sparse signal recovery

Recent advances in signal theory have demonstrated that the class of sparse signals is a good signal model for several kinds of interesting signals, often encountered in communications, radar, and image processing applications. The assumption in such a model is that the signal, when expressed in some sparsifying basis or dictionary, has very few significant coefficients, and the remaining ones are zero or approximately zero. For example, communications signals are often sparse in the short-time Fourier domain, and radar signals are sparse in the chirplet domain.

Specifically, a vector x is sparse in some basis if its basis expansion has a small ℓ_p−norm, for $p \leq 1$. The ℓ_p−norm (technically a pseudonorm if $p < 1$) of a vector is defined as:

$$\| x \|_p = (\sum_i | (x)_i |)^{\frac{1}{p}}. \tag{4}$$

For p = 0, the norm is defined as the number of nonzero coefficients, and is equal to the limit of $\| x \|_p^p$ as p tends to 0.

Consider the general problem of reconstructing a vector $x \in R^N$ from linear measurements y about x of the form

$$y_k = \langle x, \psi_k \rangle, k = 1, 2, 3....K. \tag{5}$$

Suppose here that one collects an incomplete set of frequency samples of a discrete signal x of length N. The goal is to reconstruct the full signal f given only K samples in the Fourier domain

$$y_k = \frac{1}{\sqrt{n}} \sum_{t=0}^{N-1} x_t e^{i2\pi\omega_k t/n} \tag{6}$$

With this information, we decide to recover the signal by ℓ_1-norm minimization; the proposed reconstruction f^* is given by $f^* = \psi x^*$, where x^* is the solution to the convex optimization program
($\| x \|_{\ell_1} = \sum_{i=1}^N | x_i |$).

$$\min_{\tilde{x} \in \mathbb{R}^n} \| x \|_{\ell_1} \; subject\ to\ y_k = \langle x, \psi_k \rangle \tag{7}$$

That is, among all objects $\tilde{f} = \psi\tilde{x}$ consistent with the data, we pick that whose coefficient sequence has minimal ℓ_1 -norm.

Theorem 1

Fix $f \in R^n$ and suppose that the coefficient sequence x of f in the basis Ψ is S-sparse. Select m measurements in the Φ domain uniformly at random. Then if

$$m \geqslant C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot \log n \tag{8}$$

for some positive constant C, the solution to (6) is exact with overwhelming probability. (It is shown that the probability of success exceeds $1 - \delta$ if $m \geqslant C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot \log(n/)$. In addition, the result is only guaranteed for nearly all sign sequences x with a fixed support.

We wish to make three comments: 1) The role of the coherence is completely transparent; the smaller the coherence, the fewer samples are needed, hence our emphasis on low coherence systems in the previous section. 2) One suffers no information loss by measuring just about any set of m coefficients which may be far less than the signal size apparently demands. If $\mu(\Phi, \Psi)$ is equal or close to one, then on the order of Slog n samples suffice instead of n. 3) The signal f can be exactly recovered from our condensed data set by minimizing a convex functional which does not assume any knowledge about the number of nonzero coordinates of x, their locations, or their amplitudes which we assume are all completely unknown a priori. We just run the algorithm and if the signal happens to be sufficiently sparse, exact recovery occurs. The theorem indeed suggests a very concrete acquisition protocol: sample nonadaptively in an incoherent domain and invoke linear programming after the acquisition step. Following this protocol would essentially acquire the signal in a compressed form. All that is needed is a decoder to decompress this data; this is the role of l_1 minimization.

Now let's introduce the notion of uniform uncertainty principle (UUP) . The UUP essentially states that the $K \times N$ sensing matrix Φ obeys a "restricted isometry hypothesis".Let $\Phi_T, T \subset 1, ..., N$ be the $K \times \mid T \mid$ submatrix obtained by extracting the columns of Φ corresponding to the indices in T .Then the S-restricted isometry constant δ_s of Φ which is the smallest quantity such that

$$(1 - \delta_s) \parallel c \parallel^2_{\ell_2} \leq \parallel \Phi_T c \parallel^2_{\ell_2} \leq (1 + \delta_s) \parallel c \parallel^2_{\ell_2} \tag{9}$$

for all subsets T with $\mid T \mid \leq S$ and coefficient sequences $(c_j)_{j \in T}$. This property essentially requires that every set of columns with cardinality less than S approximately behaves like an orthonormal system. An important result is that if the columns of the sensing matrix Φ are approximately orthogonal, then the exact recovery phenomenon occurs.Assume that x is S-sparse and suppose that $\delta_{2S} + \delta_{3S} < 1$or, better, $\delta_{2S} + \theta_{S,2S} < 1$. Then the solution x^* to (8) is $x^* = x$.

4.2 Rubust signal recovery from noisy data

We have shown that one could recover sparse signals from just a few measurements but in order to be really powerful, CS needs to be able to deal with both nearly sparse signals and with noise. First, general objects of interest are not exactly sparse but approximately sparse. The issue here is whether or not it is possible to obtain accurate reconstructions of such objects from highly undersampled measurements. Second, in any real application measured data will invariably be corrupted by at least a small amount of noise as sensing devices do not have infinite precision. It is therefore imperative that CS be robust vis a vis such nonidealities. At the very least, small perturbations in the data should cause small perturbations in the reconstruction. This section examines these two issues simultaneously. Before we begin, however, it will ease the exposition to consider the abstract problem of recovering a vector $x \in R_n$ from data

$$y = Ax + z \tag{10}$$

where A is an $m \times n$ sensing matrix giving us information about x, and z is a stochastic or deterministic unknown error term. The setup of the last section is of this form since with $f = \Psi x$ and $y = R\Phi f$ (R is the $m \times n$ matrix extracting the sampled coordinates in M), one can write $y = Ax$, where $A = R\Phi\Psi$. Hence, one can work with the abstract model (9) bearing in mind that x may be the coefficient sequence of the object in a proper basis.

We are given noisy data as in (9) and use $\ell 1$ minimization with relaxed constraints for reconstruction:

$$min \parallel x \parallel_{\ell_1} \ subjectto \parallel Ax - y \parallel_{\ell_2} \leqslant \epsilon \tag{11}$$

where ϵ bounds the amount of noise in the data. (One could also consider recovery programs such as the Dantzig selector or a combinatorial optimization program proposed by Haupt and Nowak ; both algorithms have provable results in the case where the noise is Gaussian with bounded variance.) Problem (10) is often called the LASSO . To the best of our knowledge, it was first proposed in [8]. This is again a convex problem (a second-order cone program) and can be solved efficiently.

THEOREM 3[11]

Assume that $\delta_2 S < \sqrt{2} - 1$. Then the solution x^* to (10) obeys

$$\parallel x^* - x \parallel_{\ell_2} \leqslant \parallel x - x_S \parallel_{\ell_1} / \sqrt{S} + C_1 \cdot \epsilon \tag{12}$$

for some constants C_0 and C_1. (Again, this theorem is unpublished as stated and is a variation on the result found in [11].) This can hardly be simpler. The reconstruction error is bounded by the sum of two terms. The first is the error which would occur if one had noiseless data. The second is just proportional to the noise level. Further, the constants C_0 and C_1 are typically small. With $\delta_2 S = 1/4$ for example, $C_0 \div 5.5 and C_1 \div 6$. This last result establishes CS as a practical and robust sensing mechanism. It works with all kinds of not necessarily sparse signals, and it handles noise gracefully. What remains to be done is to design efficient sensing matrices obeying the RIP. RANDOM SENSING Returning to the RIP, we would like to find sensing matrices with the property that column vectors taken from arbitrary subsets are nearly orthogonal. The larger these subsets, the better. This is where randomness re-enters the picture. Consider the following sensing matrices: i) form A by sampling n column vectors uniformly at random on the unit sphere of R_m; ii) form A by sampling i.i.d. entries from the normal distribution with mean 0 and variance $1/m$; iii) form A by sampling a random projection P as in Incoherent Sampling and normalize: $A = \sqrt{n/m}P$; and iv) form A by sampling i.i.d. entries from a symmetric Bernoulli distribution ($P(A_{i,j} = \pm 1/\sqrt{m} = 1/2$ or other sub-gaussian distribution. With overwhelming probability, all these matrices obey the RIP (i.e. the condition of our theorem) provided that

$$m \geqslant C \cdot S \log(n/S) \tag{13}$$

where C is some constant depending on each instance. The claims for i).iii) use fairly standard results in probability theory; arguments for iv) are more subtle; see [12] and the work of Pajor and his coworkers, e.g., [13]. In all these examples, the probability of sampling a matrix not obeying the RIP when (14) holds is exponentially small in m.One can also establish the RIP for pairs of orthobases as in Incoherence and the Sensing of Sparse Signals. With $A = R\Phi\Psi$ where R extracts m coordinates uniformly at random, it is sufficient to have

$$m \geqslant C \cdot S (\log n)^4 \tag{14}$$

for the property to hold with large probability; see [15] and [16]. If one wants a probability of failure no larger than $O(n^{-\beta})$ for some $\beta > 0$, then the best known exponent in (14) is five instead of four (it is believed that (13) holds with just log n).

This proves that one can stably and accurately reconstruct nearly sparse signals from dramatically undersampled data in an incoherent domain.

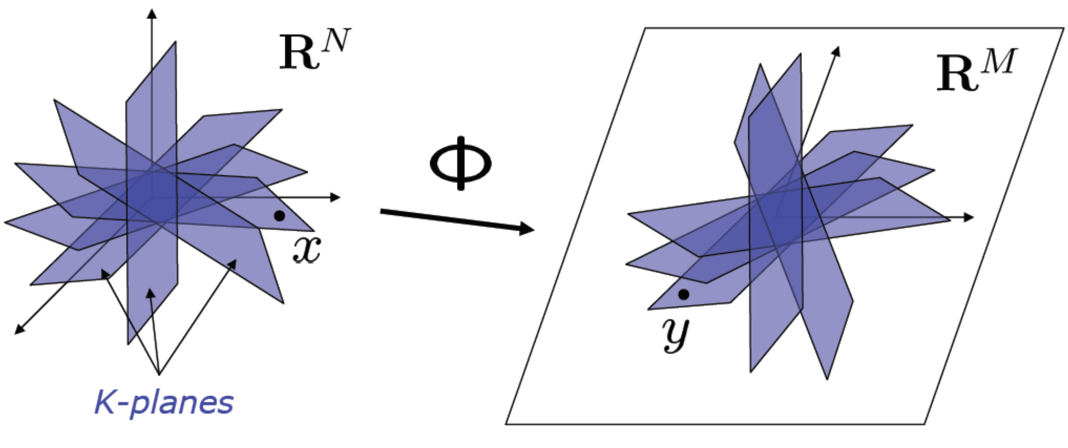

Figure 1: $M \geqslant 2K$ linear measurements - necessary for injectivity

- sufficient for injectivity when Gaussian

But not enough for efficient, robust recovery

5 Applications in wireless networking

5.1 Compressed Sensing of Channels

Consider a MIMO channel corresponding to uniform linear arrays of N_T transmit antennas and N_R receive antennas. Throughout the paper, we implicitly consider signaling over this channel using packets of duration T and (two-sided) bandwidth W. In the absence of noise, the corresponding baseband transmitted and received signal are related as:

$$x(t) = \int_{-\frac{W}{2}}^{\frac{W}{2}} H(t, f)S(f)e^{j2\pi ft}df, 0 \leq t \leq T \tag{15}$$

One of the most salient characteristics of multipath wireless channels is signal propagation over multiple spatially distributed paths. A MIMO channel can be accurately modeled in terms of these physical paths as

$$H(t, f) = \sum_{n=1}^{N_p} \beta_n \alpha_R(\theta_{R,n})\alpha_T^H(\theta_{T,n})e^{j2\pi\nu_n t}e^{-j2\pi\tau_n f} \tag{16}$$

The key idea behind virtual channel modeling is to provide a low-dimensional approximation of (9) by uniformly sampling the multipath environment in the angle-delay-Doppler domain at a resolution commensurate with the signal space parameters:
$(\Delta\theta_R, \Delta\theta_T, \Delta\tau, \Delta\nu) = (1/N_R, 1/N_T, 1/W, 1/T)$,That is

$$H(t, f) \approx \sum_{i=1}^{N_R} \sum_{k=1}^{N_T} \sum_{\ell=0}^{L-1} \sum_{m=-M}^{M} H_v(i, k, \ell, m)$$

$$\alpha_R(\frac{i}{N_R})\alpha_T^H(\frac{a}{N_T})e^{j2\pi\frac{m}{T}t}e^{-j2\pi\frac{\ell}{w}f} \tag{17}$$

$$H_v(i, k, \ell, m) \approx \sum_{n\in S_{R,i}\cap S_{T,k}\cap S_{\tau,\ell}\cap S_{\nu,m}} \beta_n \tag{18}$$

In the case of a narrowband MIMO channel , the physical channel model (9) and its virtual representation (10) reduce to

$$H(t, f) = \sum_{n=1} \beta_n \alpha_R(\theta_{R,n})\alpha_T^H(\theta_{T,n}) \approx A_R H_v A_T^H \tag{19}$$

To learn the $N_R \times N_T$ (antenna domain) matrix H, training-based channel estimation methods dedicate part of the packet duration T to transmit known signals to the receiver. Assuming this training duration to be T_{tr}, many traditional training-based receivers stack the $M_{tr} = T_{tr}W$ received (vector-valued) training signals $x(n), n = 1, ..., M_{tr}$ into an $M_{tr} \times N_R$ matrix X to yield the following system of equations:

$$X = \sqrt{\frac{\varepsilon}{M_{tr}}} S H^T + W \tag{20}$$

recover H from X:$\hat{H} = \sqrt{\frac{M_{tr}}{\varepsilon}}(S^H S)^{-1} S^H X$ Then,we get:

$$\mathbb{E}[\| \hat{H} - H \|_F^2] \geq \frac{N_T(N_R N_T)}{\varepsilon} = \frac{N_T D_{max}}{\varepsilon} \tag{21}$$

5.2 Noise reduction

we consider a test word y to be a linear combination of exemplar words ω_n, where the index n denotes a specific exemplar word ($1 \leq n \leq N$) and N the total number of exemplar words in the training corpus. We write:

$$y = \sum_{n=1}^{N} \alpha_n \omega_n, \alpha_n \in \mathbb{R} \tag{22}$$

Denoting the k^{th} vector element of ω_n by ω_n^k, and recalling that each word in the example set is represented by a vector with dimensionality K, we write our set of N exemplar words as a matrix A with dimensionality $K \times N$:

$$A = \begin{bmatrix} w_1^1 & w_2^1 & \cdots & w_{N-1}^1 & w_N^1 \\ w_1^2 & w_2^2 & \cdots & w_{N-1}^2 & w_N^2 \\ \\ \vdots & \vdots & & \vdots & \vdots \\ w_1^K & w_2^K & \cdots & w_{N-1}^K & w_N^K \end{bmatrix}.$$

We can now express any word y as $y = Ax$. With $x = [\alpha_1 \alpha_2 ... \alpha_{N-1} \alpha_N]^T$ an N-dimensional vector that will be sparsely represented in A.

In order to determine the sparse vector x representing a word y, we need to solve the system of linear equations. Typically, the number of exemplar words will be much larger than the dimensionality of the feature representation of the vowels ($K << N$). Thus, the system of linear equations is under-determined and has, generally speaking, no unique solution. Research in the section2, has shown that if x is sparse, x can be determined by solving:

$$\min_{\tilde{x} \in \mathbb{R}^n} \| x \|_{\ell_1} \; subject \; to \; y_k = \langle x, \psi_k \rangle \tag{23}$$

6 conclusions

Signal acquisition based on compressive sensing can be more efficient than traditional sampling for sparse or compressible signals. In compressive sensing, the familiar least squares optimization is inadequate for signal reconstruction, and other types of convex optimization must be invoked.

From a general viewpoint, sparsity and, more generally, compressibility has played and continues to play a fundamental role in many fields of science. Sparsity leads to efficient estimations; for example, the quality of estimation by thresholding or shrinkage algorithms depends on the sparsity of the signal we wish to estimate. Sparsity leads to dimensionality reduction and efficient modeling. The novelty here is that sparsity has bearings on the data acquisition process itself, and leads to efficient data acquisition protocols. The key word here is economically. Everybody knows that because typical signals have some structure, they can be compressed efficiently without much perceptual loss. This raises a fundamental question: because most signals are compressible, why spend so much effort acquiring all the data when we know that most of it will be discarded? Compressive sampling also known as compressed sensing shows that this is indeed possible.

7 references

[1]Emmanuel Cands, Compressive sampling. (Int. Congress of Mathematics, 3, pp. 1433-1452, Madrid, Spain, 2006)

[2]Richard Baraniuk, Compressive sensing. (IEEE Signal Processing Magazine, 24(4), pp. 118-121, July 2007)

[3]Emmanuel Cands and Michael Wakin, An introduction to compressive sampling. (IEEE Signal Processing Magazine, 25(2), pp. 21 - 30, March 2008) [High-resolution version]

[4]Justin Romberg, Imaging via compressive sampling. (IEEE Signal Processing Magazine, 25(2), pp. 14 - 20, March 2008)

[5]David Donoho, Compressed sensing. (IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, April 2006)

[6]David Donoho and Yaakov Tsaig, Extensions of compressed sensing. (Signal Processing, 86(3), pp. 533-548, March 2006)

[7]Waheed Bajwa, Jarvis Haupt, Akbar Sayeed, and Rob Nowak, Compressive wireless sensing. (Int. Conf. on Information Processing in Sensor Networks (IPSN), Nashville, Tennessee, April 2006)

[8] F. Santosa and W.W. Symes,Linear inversion of band-limited reflection seismograms, (SIAM J. Sci. Statist. Comput., vol. 7, no. 4, pp. 1307C1330, 1986)

[9]Emmanuel Cands and Terence Tao, Decoding by linear programming. (IEEE Trans. on Information Theory, 51(12), pp. 4203 - 4215, December 2005)

[10]Waheed U. Bajwa, Akbar Sayeed, and Robert Nowak, Compressed sensing of wireless channels in time, frequency, and space. (Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, California, October 2008)

[11] E. Cands, J. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, (Comm. Pure Appl. Math., vol. 59, no. 8, pp. 1207C1223, Aug. 2006)

[12] R.Baraniuk, M. Davenport, R. DeVore, and M. Wakin, A simple proof of the restricted isometry property for random matrices, (submitted for publication).

[13] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, Uniform uncertainty principle for Bernoulli and subgaussian ensembles,(2006, Preprint).

[14] A. Gilbert, S. Muthukrishnan, and M. Strauss, Improved time bounds for near-optimal sparse Fourier representation, in Proc. Wavelets XI SPIE Optics Photonics, (San Diego, CA, 2005).

[15] M. Rudelson and R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements,(submit-

ted for publication).
[16]E. Cands and T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies, IEEE Trans. Inform. Theory,(vol. 52, no. 12, pp. 5406C5425, Dec. 2006).