Measurements vs. Bits: Compressed Sensing meets Information Theory

Shriram Sarvotham, Dror Baron, and Richard G. Baraniuk
Department of Electrical and Computer Engineering
Rice University, Houston, TX 77005, USA

***Abstract*— Compressed sensing is a new framework for acquiring sparse signals based on the revelation that a small number of linear projections (measurements) of the signal contain enough information for its reconstruction. The foundation of Compressed sensing is built on the availability of noise-free measurements. However, measurement noise is unavoidable in analog systems and must be accounted for. We demonstrate that measurement noise is the crucial factor that dictates the number of measurements needed for reconstruction. To establish this result, we evaluate the information contained in the measurements by viewing the measurement system as an information theoretic channel. Combining the capacity of this channel with the rate-distortion function of the sparse signal, we lower bound the rate-distortion performance of a compressed sensing system. Our approach concisely captures the effect of measurement noise on the performance limits of signal reconstruction, thus enabling to benchmark the performance of specific reconstruction algorithms.**

I. Introduction

Consider a discrete-time real-valued signal $\mathbf{X}$ of length n that has only k non-zero coefficients for some $k \ll n$. The core tenet of Compressed Sensing (CS) is that it is unnecessary to measure all n values of the sparse signal; rather, we can recover the signal using a small number of linear projections onto an *incoherent* basis [1, 2]. To measure (encode) $\mathbf{X}$, we compute the ideal (noiseless) measurement vector $\mathbf{Y_0} \in \mathbb{R}^m$ as m linear projections of $\mathbf{X}$ via the matrix-vector multiplication $\mathbf{Y_0} = \mathbf{\Phi X}$. The goal in CS is to reconstruct (decode) $\mathbf{X}$ — either accurately or approximately — given the measurements.

CS reconstruction can be performed with $O(n^3)$ computation via ℓ_1 minimization by applying linear programming techniques.[1] This approach requires approximately $k \log(n/k)$ measurements [1–3], where we use the base-two logarithm.

The CS community has also studied acquisition of signals that are not k-sparse but *compressible*, meaning that their coefficient magnitudes decay according to a power law. In this case too there are polynomial complexity algorithms that achieve $||\widehat{\mathbf{X}} - \mathbf{X}||_2^2 \leq c_1 \sigma_k^2$, where the operator $|| \cdot ||_2^2$ denotes the squared ℓ_2 norm, σ_k^2 is the squared ℓ_2 error in the best k-term approximation to the signal $\mathbf{X}$, and c_1 is a constant. For compressible signals that lie in an ℓ_1 ball, these algorithms require $m = O(k \log(n/k))$ measurements [1]. Additionally, it has been shown that at least $m = \Omega(k \log(n/k))$ measurements are required in this case [4–6].

In order to reduce resource consumption, a problem of considerable interest is to seek practical CS measurement and reconstruction schemes that require fewer measurements. Indeed, as Donoho wrote [2], "Why go to so much effort to acquire all the data when most of what we get will be thrown away?"

Recalling that $m = \Omega(k \log(n/k))$ measurements are required for compressible signals, let us consider more restrictive signal classes. It has been shown that we cannot hope to perform CS reconstruction of a k-sparse signal using fewer than $m = k + 1$ measurements [3, 7]. However, the only approach for CS reconstruction using $m = k+1$ is via ℓ_0 minimization, which is known to be NP-complete and therefore impractical [8]. This raises the question: can we construct practical CS schemes that require $m = o(k \log(n/k))$ measurements for k-sparse signals?

In addition to computational issues and appropriate signal classes, any performance evaluation of CS reconstruction techniques must account for imprecisions in the measurement process. We emphasize that all analog measurement systems are imperfect and add various artifacts. Furthermore, any CS hardware system that relies on analog-to-digital conversion will contain quantization noise, which must also be accounted for in the performance analysis.

In this paper we show that the performance limits of CS reconstruction systems subject to additive white Gaussian noise obey

$$\delta \geq \frac{2R(\mathcal{E}(\mathcal{D}_{\text{X}}))}{\log(1 + \text{SNR})},$$

where $\delta = m/n$ is the measurement rate, $R(\cdot)$ is the rate-distortion function of the signal source, $\mathcal{E}(\mathcal{D}_{\text{X}})$ is the distortion level using decoder $\mathcal{D}_{\text{X}}$ to reconstruct $\mathbf{X}$, and SNR is the measurement signal-to-noise ratio (details in the sequel). We show that each CS measurement is similar to one usage of a communication channel and contributes $\frac{1}{2}\log(1 + \text{SNR})$ bits toward

This work was supported by the grants DARPA/ONR N66001-06-1-2011, N00014-06-1-0610; NSF CCF-0431150, CNS-0520280, CNS-0435425; ONR N00014-06-1-0769; AFOSR FA9550-04-1-0148; and by the Texas Instruments Leadership University Program.
Email: {shri, drorb, richb}@rice.edu; Web: dsp.rice.edu/cs.

[1] For two functions $f(n)$ and $g(n)$, $f(n) = O(g(n))$ if $\exists c, n_0 \in \mathbb{R}^+$, $0 \leq f(n) \leq cg(n)$ for all $n > n_0$. Similarly, $f(n) = o(g(n))$ if for all $c > 0$, $\exists n_0 \in \mathbb{R}^+$, $0 \leq f(n) < cg(n)$ for all $n > n_0$, and $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$. Finally, $f(n) = \Theta(g(n))$ if $\exists c_1, c_2, n_0 \in \mathbb{R}^+$, $0 \leq c_1 g(n) \leq f(n) \leq c_2 g(n)$ for all $n > n_0$.

the resolution of the signal. The revelation here is the direct relationship in this noisy setting between the measurement rate required and the rate-distortion function of the signal source being acquired. We apply information theoretic tools such as channel coding and rate-distortion to illuminate these points. These information theoretic tools shed new light on CS techniques, enabling us to realize the crucial significance of rate-distortion tradeoffs and measurement noise.

To illustrate this principle, consider an observed measurement vector $\mathbf{Y}$ that incorporates additive white Gaussian noise $\mathbf{Z}$ such that $\mathbf{Y} = \mathbf{Y_0} + \mathbf{Z}$. We will derive a lower bound (a converse result) on the CS measurement rate as a function of the squared reconstruction error and the measurement signal-to-noise ratio. This bound dictates the minimum reconstruction distortion that the user must contend with in the face of Gaussian measurement noise. To derive our result, we model the measurement process as ideal measurements passing through a noisy stochastic channel. Because the capacity of this measurement channel is finite, perfect signal recovery is impossible. The lower bound is obtained by relating the capacity of the measurement channel to the rate-distortion function of the input signal.

One of the grand successes of information theory has been the rise of low-complexity capacity-achieving channel codes [9, 10]. Our broader vision is to develop similar CS reconstruction algorithms. Indeed, our recent work on sparse signal acquisition and recovery using LDPC measurement matrices [11], which borrows ideas from the channel coding community [9, 10, 12], indicates that reconstruction of k-sparse signals, which have $\Theta(k \log(n/k))$ rate-distortion content, can be performed using $m = O(k \log(n/k))$ noisy measurements with modest $O(n \log(n))$ complexity.

The remainder of the paper is organized as follows. Section II describes the mathematical setting formally, followed by a derivation of our lower bound on the measurement rate in Section III. Examples are provided in Section IV. We discuss our work on reconstruction via LDPC measurement matrices in Section V, and conclude the paper in Section VI.

II. Set up

A. Notation and preliminaries

We denote *random variables* by upper-case letters and realizations of random variables by their corresponding lower-case letters, e.g., x is a realization of the random variable X. We also use lower-case letters to represent deterministic variables, e.g., n to denote the signal length. Vectors are denoted in boldface in order to distinguish them from scalars, e.g., the vector $\mathbf{x}$ is a realization of the random vector $\mathbf{X}$. We refer to the entries of a vector using the subscript notation, and so $\mathbf{X} = [X_1, X_2, ..., X_n]^T$ is a length n random vector, where $[\cdot]^T$ is the transpose operator. We assume that all vectors are column vectors. We denote the probability density function (pdf) of a random variable X as $p_X(x)$ and the cumulative distribution function (cdf) as $P_X(x)$. We represent estimators of variables using the "hat" notation, e.g., the random variable $\widehat{X}$ is an estimator for X.

A *source* associated with a random variable X (called "source X") produces a random vector $\mathbf{X} = [X_1, X_2, ..., X_n]^T$ of a specified length n, where the entries X_i, $i = 1, 2, ..., n$ are independent and identically distributed (i.i.d.) and $X_i \sim p_X(x)$.

Finally, we use standard notation to refer to information theoretic measures such as the channel capacity (C) and rate-distortion function $(R(\cdot))$ [13].

B. Measurement and reconstruction

We acquire m real-valued *measurements* of a signal $\mathbf{X}$ by multiplying $\mathbf{X}$ with a measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$. We denote the ideal (uncontaminated) measurement vector by $\mathbf{Y_0}$ so that $\mathbf{Y_0} = \mathbf{\Phi X}$. The observed measurements are obtained by corrupting the ideal measurements with *additive noise*. The noise vector $\mathbf{Z} \in \mathbb{R}^m$ consists of m i.i.d. $\mathcal{N}(0, 1)$ random variables. We denote the observed measurement vector by $\mathbf{Y}$ so that

$$\mathbf{Y} = \mathbf{Y_0} + \mathbf{Z} = \mathbf{\Phi X} + \mathbf{Z}.$$

Therefore, the statistics of the measurement process are fully characterized by specifying the measurement matrix $\mathbf{\Phi}$.

A *reconstruction* scheme uses a *decoder* to estimate the signal $\mathbf{X}$ using the observed measurements $\mathbf{Y}$. The decoder $\mathcal{D}_{\mathrm{X}}$ is a mapping $\mathcal{D}_{\mathrm{X}} : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ that takes the measurement vector $\mathbf{Y}$ as an input and produces an estimate $\widehat{\mathbf{X}}$ of $\mathbf{X}$. It is assumed that the signal statistics and the measurement matrix $\mathbf{\Phi}$ are known to the decoder.

C. Measurement and reconstruction quality

We define the *measurement signal-to-noise ratio* SNR as the ratio between the expected noiseless measurement energy and the expected noise energy:

$$\text{SNR} \triangleq \frac{E[||\mathbf{Y_0}||_2^2]}{E[||\mathbf{Z}||_2^2]} = \frac{E[||\mathbf{Y_0}||_2^2]}{m}.$$

We also define a metric to gauge the quality of reconstruction for a *given* pair of CS measurement matrix $\mathbf{\Phi}$ and reconstruction scheme $\mathcal{D}_{\mathrm{X}}$. We use the *normalized squared reconstruction error*

$$\mathcal{E}(\mathcal{D}_{\mathrm{X}}) \triangleq \frac{E[\|\mathbf{X} - \widehat{\mathbf{X}}\|_2^2]}{E[\|\mathbf{X}\|_2^2]},$$

where the expectation in the numerator is over the joint distribution of $\mathbf{X}$ and $\widehat{\mathbf{X}}$. Finally, we define the measurement rate as

$$\delta \triangleq \frac{m}{n}.$$

III. Lower bound on reconstruction error

The goal of this section is to find the minimum measurement rate needed to reconstruct the signal $\mathbf{X}$ to achieve a given fidelity using CS measurement and reconstruction schemes as described above. Specifically, we seek to determine an asymptotic *lower bound* on δ in order to achieve a reconstruction quality $\mathcal{E}(\mathcal{D}_{\mathrm{X}})$, *irrespective* of $\mathbf{\Phi}$ and the decoding scheme $\mathcal{D}_{\mathrm{X}}$.

A. Approach

To probe the performance limits of CS, we draw inspiration from information theory. Our motivation stems from the following insights. The source X is a discrete-time continuous amplitude source. The measurements $\mathbf{Y}$ are modeled as outputs of a Gaussian channel, where the channel inputs are the ideal CS measurements $\mathbf{Y_0}$ (the precise nature of the channel is described in Section III-C). This channel has a finite capacity and so each measurement only extracts a finite amount of information. Therefore, perfect signal recovery is impossible [14]. We seek to find a lower bound on the measurement rate δ in terms of the distortion $\mathcal{E}(\mathcal{D}_{\mathrm{X}}) > 0$ and measurement signal-to-noise ratio SNR.

In order to apply information theoretic insights to solve this problem, we investigate the amount of information that can be extracted from the CS measurements. This quantity is determined by the capacity of the measurement channel.

Having upper bounded the information contained in the measurements, we investigate the minimum information (in bits) needed to reconstruct the signal with distortion $\mathcal{E}(\mathcal{D}_{\mathrm{X}})$. This result can be obtained from the rate-distortion function for the source X using the mean squared error distortion measure. We have thus characterized δ and $\mathcal{E}(\mathcal{D}_{\mathrm{X}})$ in terms of a common currency, namely bits. Invoking the source-channel separation theorem [15], we obtain a lower bound on δ as a function of $\mathcal{E}(\mathcal{D}_{\mathrm{X}})$ and SNR.

B. Main result

Theorem 1: For a signal source with rate-distortion function $R(\cdot)$ and measurement scheme specified above, the lower bound on the CS measurement rate required to obtain normalized reconstruction error $\mathcal{E}(\mathcal{D}_{\mathrm{X}})$ subject to a fixed SNR is given by

$$\delta \geq \frac{2R(\mathcal{E}(\mathcal{D}_{\mathrm{X}}))}{\log(1+\mathrm{SNR})}$$

as $n \to \infty$.

Before providing the proof for Theorem 1, let us pause to reflect on the theorem statement. First, the theorem is valid for *any* i.i.d. source for which the rate-distortion function can be characterized. In particular, the scope of the theorem extends beyond signals that are exactly sparse. Second, the theorem is valid in the asymptotic regime when $n \to \infty$. Finally, the theorem reveals the crucial dependence of δ on the measurement SNR.

C. Proof of Theorem 1

In the first part of the proof, we derive the capacity of the measurement channel. This result enables us to compute the maximum information that can be extracted from m noisy real valued measurements $\mathbf{Y}$. The second part of the proof applies the source-channel separation theorem for discrete-time continuous-amplitude ergodic sources to the rate-distortion function of the source and the aforementioned capacity.

1) Capacity of the measurement channel: We consider noisy measurements $\mathbf{Y}$ obtained by passing the noise free measurements $\mathbf{Y_0}$ through an additive Gaussian noise channel with i.i.d. noise components $\mathcal{N}(0,1)$. We call this the *measurement channel*, and it is characterized by the input-output relationship

$$\mathbf{Y} = \mathbf{Y_0} + \mathbf{Z}.$$

In order to calculate the capacity of the measurement channel, we consider the covariance matrices $\Sigma_{\mathbf{Y_0}}$ and $\Sigma_{\mathbf{Z}} = I_{m\times m}$ of $\mathbf{Y_0}$ and $\mathbf{Z}$ respectively. Note that we can write the SNR in terms of $\Sigma_{\mathbf{Y_0}}$ as $\mathrm{SNR} = (1/m)\mathrm{tr}\left(\Sigma_{\mathbf{Y_0}}\right)$, where $\mathrm{tr}(\cdot)$ refers to the trace of a matrix.

To compute the capacity of this channel, we use the well-known result [13]

$$C = \max_{\mathrm{tr}(\Sigma_{\mathbf{Y_0}})\leq m\mathrm{SNR}} \frac{1}{2m}\log\frac{|\Sigma_{\mathbf{Y_0}}+\Sigma_{\mathbf{Z}}|}{|\Sigma_{\mathbf{Z}}|},$$

where C is the channel capacity in bits per measurement, and $|\cdot|$ denotes the determinant. Because $\Sigma_{\mathbf{Z}} = I_{m\times m}$ we have $|\Sigma_{\mathbf{Z}}| = 1$, and the above equation reduces to

$$C = \max_{\mathrm{tr}(\Sigma_{\mathbf{Y_0}})\leq m\mathrm{SNR}} \frac{1}{2m}\log|\Sigma_{\mathbf{Y_0}}+I_{m\times m}|.$$

To maximize the channel capacity, we seek to determine the correlation matrix $\Sigma_{\mathbf{Y_0}}$ that maximizes $|\Sigma_{\mathbf{Y_0}}+I_{m\times m}|$ subject to the constraint $\mathrm{tr}(\Sigma_{\mathbf{Y_0}}) \leq m\mathrm{SNR}$. For this, we apply Hadamard's inequality [13] which states that the determinant of any positive definite matrix Λ is less than or equal to the product of the diagonal elements, i.e., $|\Lambda| \leq \prod_i \Lambda(i,i)$, with equality if and only if the matrix is diagonal. Since $(\Sigma_{\mathbf{Y_0}}+I_{m\times m})$ is a positive definite matrix (it is a sum of two covariance matrices that are each positive definite), we have $|\Sigma_{\mathbf{Y_0}}+I_{m\times m}| \leq \prod_i\left(1+\Sigma_{\mathbf{Y_0}}(i,i)\right)$. Finally, the maximum value for the product $\prod_i\left(1+\Sigma_{\mathbf{Y_0}}(i,i)\right)$ under the constraint $\mathrm{tr}(\Sigma_{\mathbf{Y_0}}) \leq m\mathrm{SNR}$ is attained when the diagonal entries $\Sigma_{\mathbf{Y_0}}(i,i)$ all equal SNR. Using the above arguments, we have

$$\begin{aligned} C &\leq \max_{\mathrm{tr}(\Sigma_{\mathbf{Y_0}})\leq m\mathrm{SNR}} \frac{1}{2m}\log\prod_i\left(1+\Sigma_{\mathbf{Y_0}}(i,i)\right) \\ &\leq \frac{1}{2}\log(1+\mathrm{SNR}), \end{aligned} \tag{1}$$

where equality is attained when $\mathbf{Y_0}$ is diagonal and the diagonal entries are all equal to SNR. Therefore, the best CS measurement system has statistically independent measurements $\mathbf{Y_0}$, with all measurements having the same variance. This revelation can be used as a practical guiding principle in order to construct good compressed sensing matrices $\mathbf{\Phi}$. We summarize the important conclusions in the form of a Lemma.

Lemma 1: The upper bound on the capacity of the CS measurement channel is given by

$$C \leq \frac{1}{2}\log(1 + \text{SNR})$$

bits per measurement. Equality in the above equation requires that the measurements in $\mathbf{Y_0}$ are statistically independent and all measurements have the same variance equal to SNR.

2) Computing the error bound using the source-channel separation theorem: So far, we have described the information provided by the measurements $\mathbf{Y}$. We can also characterize the information content of the requisite signal reconstruction quality using rate-distortion formulae.

The source-channel separation theorem for discrete-time continuous amplitude stationary ergodic signals [15] states that a source X can be communicated up to distortion quality D via m channel uses if and only if the information content mC that can be extracted from the channel exceeds the information content $nR(D)$ of the quantized source.

We complete the proof of Theorem 1 by applying the converse portion of the separation theorem with Lemma 1 and the rate-distortion function $R(D)$. $\square$

IV. Examples

Consider a k-sparse signal $\mathbf{X}$ where the spikes have uniform amplitude. In this case, it is well known that precise description of $\mathbf{X}$ would require $\log\binom{n}{k} \approx k\log(n/k)$ bits, where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

This problem can be extended to a lossy description of $\mathbf{X}$ using a rate-distortion approach for a binary asymmetric source [13] or recent results on the rate-distortion of spike processes [16, 17]. For small distortion values, the rate-distortion content remains roughly $k\log(n/k)$ bits. Combining this result with our lower bound, we obtain the following condition on the number of measurements

$$m \gtrsim \frac{2k\log(n/k)}{\log(1 + \text{SNR})},$$

where the approximation is due to the asymptotic nature of the information theoretic tools that were used in our proof.

Let us now examine specific numerical examples. First, suppose that the signal is of length $n = 10^7$ and contains $k = 10^3$ spikes. If the measurement signal-to-noise ratio satisfies SNR $= 10$ dB, then the number of measurements must satisfy

$$m \gtrsim \frac{2 \cdot 10^3 \log(10^7/10^3)}{\log(1 + 10^1)} = 7,682.$$

Therefore, a "reasonable" SNR requires a modest number of measurements. In contrast, if we choose SNR $= -20$ dB, then m must exceed $1.85 \cdot 10^6$. Although in this case m is still much smaller than n, a poor signal-to-noise ratio prevents a drastic reduction in the number of measurements.

Our lower bound can also be used to provide results for denoising sparse signals from a reduced number of measurements. Fletcher et al. [18] provided bounds on denoising sparse signals using redundant frames. Their work considered several specific sparse signal models. Although our work does not consider the extension to redundant frames, our result is more general in the sense than any i.i.d. signal model can be used.

V. Practical reconstruction scheme

We have extended the success of LDPC codes [10, 12] to the problem of CS measurement and reconstruction. The crucial principle is the use of low density structure for the CS matrix $\mathbf{\Phi}$. This special structure for $\mathbf{\Phi}$ can be leveraged in three ways. First, the LDPC structure enables low-complexity computation of measurements, because each measurement depends only on a small set of coefficients. Second, the sparse structure of $\mathbf{\Phi}$ can be used to provide low-complexity reconstruction schemes by deploying message passing algorithms [9, 10, 12]. Third, these LDPC-based CS reconstruction schemes can operate close to the theoretical limits of Theorem 1.

A. Measurement process

We compute the CS measurements using a *sparse* CS matrix $\mathbf{\Phi}$, where the entries of $\mathbf{\Phi}$ are drawn from the set $\{0, 1, -1\}$. Note that in addition to using the elements 0 and 1 as in LDPC codes, we also include the element -1 in order to ensure that the expected values of the elements in each row (and column) of $\mathbf{\Phi}$ are zero. In this setting, (uncorrupted) measurements $\mathbf{Y_0}(i)$ are just sums and differences of small subsets of the coefficients of the signal $\mathbf{X}$. The design of $\mathbf{\Phi}$ (such as fixing the row weight, column weight, and so on) is based on the properties of the sparse signal $\mathbf{X}$ as well as the accompanying decoding algorithm. The goal is to imbibe the reconstruction algorithm with low-complexity yet require a modest number of measurements. We have argued that choosing the row weight to be inversely proportional to the sparsity rate of the input signal yields good performance [11].

B. Reconstruction via belief propagation

The use of sparse CS matrix facilitates the application of *message passing* algorithms for signal reconstruction. The key property that enables us to use message passing algorithms is that the sparse structure of $\mathbf{\Phi}$ can be represented as a sparse bipartite graph. Signal reconstruction can be viewed as a Bayesian inference problem, and can be solved by iteratively exchanging messages over the edges of the said graph using the well known *belief propagation* algorithm. The stochastic signal model (where we model the coefficients of the input signal as i.i.d. outcomes of a distribution $p_X(\cdot)$) can be used as a *prior* to model the input signal. Belief propagation allows us to estimate the signal that explains the measurements and best matches the signal prior. We have shown that this technique exhibits $O(n \log(n))$ complexity and empirically observed that $m = O(k \log(n))$ measurements are required.

C. "One shot" reconstruction

We have also considered a simplified "one shot" reconstruction algorithm for sparse signals. The algorithm is based on a median filter approach. The complexity of this approach is also $O(n \log(n))$. Surprisingly, this algorithm achieves decent reconstruction fidelity for k-sparse signals using $m = O(k \log(n/k))$ measurements. Again, because the rate-distortion content of a k-sparse signal is $\Theta(k \log(n/k))$, the combination of this result with our lower bound (Theorem 1) indicates the potential to develop low-complexity "capacity approaching" CS schemes.

VI. Conclusions

In this paper, we presented information theoretic arguments to lower bound the number of noisy compressed sensing (CS) measurements required for signal reconstruction. The key idea is to model the noisy signal acquisition process as a communication channel. The capacity of this channel allows us to express the information contained in the measurements in terms of bits. Using this result along with the rate-distortion function of the source yields a converse result on the measurement rates required. We considered the example of using a spike process as input, and presented numerical results given by the theoretical bound.

This work further strengthens the connections between information theory and CS. As part of our ongoing work, we are investigating the best achievable CS reconstruction scheme. In particular, we are striving to provide converse and achievable bounds that are tight. The resolution of this problem will determine the effectiveness of CS in encoding discrete-time real-valued signals.

References

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Nov. 2005, Submitted.

[4] E. D. Gluskin, "On some finite-dimensional problems in the theory of widths," *Vestnik Leningrad Univ. Math*, vol. 14, pp. 163–170, 1982.

[5] E. D. Gluskin, "Norms of random matrices and widths of finite dimensional sets," *Math USSR Sbornik*, vol. 48, pp. 173–182, 1984.

[6] B. Kashin, "The widths of certain finite dimensional sets and classes of smooth functions," *Izvestia*, vol. 41, pp. 334–351, 1977.

[7] R. Venkataramani and Y. Bresler, "Further results on spectrum blind sampling of 2D signals," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Chicago, Oct. 1998, vol. 2.

[8] E. Candès and T. Tao, "Error correction via linear programming," *Found. of Comp. Math.*, 2005, Submitted.

[9] C. Berrou, A. Glavieux, and P.Thitimajshima, "Near Shannon limit error correcting coding and decoding: Turbo codes," *Proc. 1993 IEEE Int. Conf. Communications*, pp. 1064–1070, May 1993.

[10] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[11] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Compressed sensing reconstruction via belief propagation," Technical report ECE-06-01, Department of Electrical and Computer Engineering, Rice University, Jul. 2006.

[12] R. G. Gallager, "Low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 8, pp. 21–28, Jan. 1962.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[14] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. on Pure and Applied Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[15] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[16] C. Weidmann and M. Vetterli, "Rate distortion behavior of sparse sources," in *Technical report, EPFL*, 2001.

[17] R. Castro, M. Wakin, and M. Orchard, "On the problem of simultaneous encoding of magnitude and location information," in *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002.

[18] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Denoising by sparse approximation: Error bounds based on rate-distortion theory," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. Article ID 26318, 19 pages, 2006.