Compressed Sensing*

Michael Wakin

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 2.0†

Abstract

This collection reviews fundamental concepts underlying the use of concise models for signal processing. Topics are presented from a geometric perspective and include low-dimensional linear, sparse, and manifold-based signal models, approximation, compression, dimensionality reduction, and Compressed Sensing.

A new theory known as Compressed Sensing (CS) has recently emerged that can also be categorized as a type of dimensionality reduction. Like manifold learning, CS is strongly model-based (relying on sparsity in particular). However, unlike many of the standard techniques in dimensionality reduction (such as manifold learning or the JL lemma), the goal of CS is to maintain a low-dimensional representation of a signal x from which a faithful approximation to x can be recovered. In a sense, this more closely resembles the traditional problem of data compression (see Compression[1]). In CS, however, the encoder requires no a priori knowledge of the signal structure. Only the **decoder** uses the model (sparsity) to recover the signal. We justify such an approach again using geometric arguments.

1 Motivation

Consider a signal $x \in \mathbb{R}^N$, and suppose that the basis Ψ provides a K-sparse representation of x

$$x = \Psi\alpha, \tag{1}$$

with $\| \alpha \|_0 = K$. (In this section, we focus on exactly K-sparse signals, though many of the key ideas translate to compressible signals [11], [16]. In addition, we note that the CS concepts are also extendable to tight frames.)

As we discussed in Compression[2], the standard procedure for compressing sparse signals, known as transform coding, is to (i) acquire the full N-sample signal x; (ii) compute the complete set of transform coefficients α; (iii) locate the K largest, significant coefficients and discard the (many) small coefficients; (iv) encode the **values and locations** of the largest coefficients.

This procedure has three inherent inefficiencies: First, for a high-dimensional signal, we must start with a large number of samples N. Second, the encoder must compute **all**N of the transform coefficients α, even though it will discard all but K of them. Third, the encoder must encode the locations of the large coefficients, which requires increasing the coding rate since the locations change with each signal.

*Version 1.5: Sep 22, 2009 1:37 pm -0500

†http://creativecommons.org/licenses/by/2.0/

[1] "Compression" <http://cnx.org/content/m18729/latest/>

[2] "Compression" <http://cnx.org/content/m18729/latest/>

2 Incoherent projections

This raises a simple question: For a given signal, is it possible to directly estimate the set of large $\alpha(n)$'s that will not be discarded? While this seems improbable, Candès, Romberg, and Tao [7], [11] and Donoho [16] have shown that a reduced set of projections can contain enough information to reconstruct sparse signals. An offshoot of this work, often referred to as **Compressed Sensing** (CS) [6], [11], [8], [9], [5], [16], [18], has emerged that builds on this principle.

In CS, we do not measure or encode the K significant $\alpha(n)$ directly. Rather, we measure and encode $M < N$ projections $y(m) =< x, \phi_m{}^T >$ of the signal onto a **second set** of functions $\{\phi_m\}, m = 1, 2, ..., M$. In matrix notation, we measure

$$y = \Phi x, \tag{2}$$

where y is an $M \times 1$ column vector and the **measurement basis** matrix Φ is $M \times N$ with each row a basis vector ϕ_m. Since $M < N$, recovery of the signal x from the measurements y is ill-posed in general; however the additional assumption of signal **sparsity** makes recovery possible and practical.

The CS theory tells us that when certain conditions hold, namely that the functions $\{\phi_m\}$ cannot sparsely represent the elements of the basis $\{\psi_n\}$ (a condition known as **incoherence** of the two dictionaries [11], [7], [16], [28]) and the number of measurements M is large enough, then it is indeed possible to recover the set of large $\{\alpha(n)\}$ (and thus the signal x) from a similarly sized set of measurements y. This incoherence property holds for many pairs of bases, including for example, delta spikes and the sine waves of a Fourier basis, or the Fourier basis and wavelets. Significantly, this incoherence also holds with high probability between an arbitrary fixed basis and a randomly generated one.

3 Methods for signal recovery

Although the problem of recovering x from y is ill-posed in general (because $x \in \mathbb{R}^N$, $y \in \mathbb{R}^M$, and $M < N$), it is indeed possible to recover **sparse** signals from CS measurements. Given the measurements $y = \Phi x$, there exist an infinite number of candidate signals in the shifted nullspace $\mathcal{N}(\Phi) + x$ that could generate the same measurements y (see Linear Models from Low-Dimensional Signal Models[3]). Recovery of the correct signal x can be accomplished by seeking a **sparse** solution among these candidates.

3.1 Recovery via combinatorial optimization

Supposing that x is exactly K-sparse in the dictionary Ψ, then recovery of x from y can be formulated as the ℓ_0 minimization

$$\hat{\alpha} = argmin \| \alpha \|_0 \quad \text{s.t. } y = \Phi\Psi\alpha. \tag{3}$$

Given some technical conditions on Φ and Ψ (see Theorem Section 3.1 (Recovery via combinatorial optimization)below), then with high probability this optimization problem returns the proper K-sparse solution α, from which the true x may be constructed. (Thanks to the incoherence between the two bases, if the original signal is sparse in the α coefficients, then no other set of sparse signal coefficients $\alpha^{'}$ can yield the same projections y.) We note that the recovery program (3) can be interpreted as finding a K-term approximation to y from the columns of the dictionary $\Phi\Psi$, which we call the **holographic basis** because of the complex pattern in which it encodes the sparse signal coefficients [16].

In principle, remarkably few incoherent measurements are required to recover a K-sparse signal via ℓ_0 minimization. Clearly, more than K measurements must be taken to avoid ambiguity; the following theorem (which is proved in [3]) establishes that $K + 1$ random measurements will suffice. (Similar results were established by Venkataramani and Bresler [30].)

Theorem 1:
Let Ψ be an orthonormal basis for $\mathbb{R}^N$, and let $1 \leq K < N$. Then the following statements hold:

[3] "Low-Dimensional Signal Models": Section Linear models <http://cnx.org/content/m18726/latest/#uid1>

1. Let Φ be an $M \times N$ measurement matrix with i.i.d. Gaussian entries with $M \geq 2K$. Then with probability one the following statement holds: all signals $x = \Psi\alpha$ having expansion coefficients $\alpha \in \mathbb{R}^N$ that satisfy $\| \alpha \|_0 = K$ can be recovered uniquely from the M-dimensional measurement vector $y = \Phi x$ via the ℓ_0 optimization (3).
2. Let $x = \Psi\alpha$ such that $\| \alpha \|_0 = K$. Let Φ be an $M \times N$ measurement matrix with i.i.d. Gaussian entries (notably, independent of x) with $M \geq K+1$. Then with probability one the following statement holds: x can be recovered uniquely from the M-dimensional measurement vector $y = \Phi x$ via the ℓ_0 optimization (3).
3. Let Φ be an $M \times N$ measurement matrix, where $M \leq K$. Then, aside from pathological cases (specified in the proof), no signal $x = \Psi\alpha$ with $\| \alpha \|_0 = K$ can be uniquely recovered from the M-dimensional measurement vector $y = \Phi x$.

The second statement of the theorem differs from the first in the following respect: when $K < M < 2K$, there will necessarily exist K-sparse signals x that cannot be uniquely recovered from the M-dimensional measurement vector $y = \Phi x$. However, these signals form a set of measure zero within the set of **all**K-sparse signals and can safely be avoided if Φ is randomly generated independently of x.

Unfortunately, as discussed in Nonlinear Approximation from Approximation[4], solving this ℓ_0 optimization problem is prohibitively complex. Yet another challenge is robustness; in the setting of Theorem "Recovery via ℓ 0 optimization" (Section 3.1: Recovery via combinatorial optimization), the recovery may be very poorly conditioned. In fact, **both** of these considerations (computational complexity and robustness) can be addressed, but at the expense of slightly more measurements.

3.2 Recovery via convex optimization

The practical revelation that supports the new CS theory is that it is not necessary to solve the ℓ_0-minimization problem to recover α. In fact, a much easier problem yields an equivalent solution (thanks again to the incoherency of the bases); we need only solve for the ℓ_1-sparsest coefficients α that agree with the measurements y[7], [6], [11], [8], [9], [5], [16], [18]

$$\hat{\alpha} = argmin \| \alpha \|_1 \quad \text{s.t. } y = \Phi\Psi\alpha. \tag{4}$$

As discussed in Nonlinear Approximation from Approximation[5], this optimization problem, also known as **Basis Pursuit**[12], is significantly more approachable and can be solved with traditional linear programming techniques whose computational complexities are polynomial in N.

There is no free lunch, however; according to the theory, more than $K+1$ measurements are required in order to recover sparse signals via Basis Pursuit. Instead, one typically requires $M \geq cK$ measurements, where $c > 1$ is an **oversampling factor**. As an example, we quote a result asymptotic in N. For simplicity, we assume that the sparsity scales linearly with N; that is, $K = SN$, where we call S the **sparsity rate**.

Theorem 2:
[10], [17], [15] Set $K = SN$ with $0 < S \ll 1$. Then there exists an oversampling factor $c(S) = O(log(1/S))$, $c(S) > 1$, such that, for a K-sparse signal x in the basis Ψ, the following statements hold:

1. The probability of recovering x via Basis Pursuit from $(c(S)+\varepsilon)K$ random projections, $\varepsilon > 0$, converges to one as $N \to \infty$.
2. The probability of recovering x via Basis Pursuit from $(c(S)-\varepsilon)K$ random projections, $\varepsilon > 0$, converges to zero as $N \to \infty$.

[4] "Approximation": Section Nonlinear approximation <http://cnx.org/content/m18727/latest/#uid4>
[5] "Approximation": Section Nonlinear approximation <http://cnx.org/content/m18727/latest/#uid4>

In an illuminating series of recent papers, Donoho and Tanner [15], [17], [19] have characterized the oversampling factor $c(S)$ precisely (see also "The geometry of Compressed Sensing" (Section 5: The geometry of Compressed Sensing)). With appropriate oversampling, reconstruction via Basis Pursuit is also provably robust to measurement noise and quantization error [7].

We often use the abbreviated notation c to describe the oversampling factor required in various settings even though $c(S)$ depends on the sparsity K and signal length N.

A CS recovery example on the Cameraman test image is shown in Figure 1. In this case, with $M = 4K$ we achieve near-perfect recovery of the sparse measured image.

Figure 1: Compressive sensing reconstruction of the nonlinear approximation Cameraman image from here[6](b). Using $M = 16384$ random measurements of the K-term nonlinear approximation image (where $K = 4096$), we solve an ℓ_1-minimization problem to obtain the reconstruction shown above. The MSE with respect to the measured image is 0.08, so the reconstruction is virtually perfect.

3.3 Recovery via greedy pursuit

At the expense of slightly more measurements, iterative greedy algorithms such as Orthogonal Matching Pursuit (OMP) [28], Matching Pursuit (MP) [26], and Tree Matching Pursuit (TMP) [21], [24] have also been proposed to recover the signal x from the measurements y (see Nonlinear Approximation from Approximation[7]). In CS applications, OMP requires $c \approx 2ln(N)$[28] to succeed with high probability. OMP is also guaranteed to converge within M iterations. We note that Tropp and Gilbert require the OMP algorithm to succeed in the first K iterations [28]; however, in our simulations, we allow the algorithm to run up to the maximum of M possible iterations. The choice of an appropriate practical stopping criterion (likely somewhere between K and M iterations) is a subject of current research in the CS community.

4 Impact and applications

CS appears to be promising for a number of applications in signal acquisition and compression. Instead of sampling a K-sparse signal N times, only cK incoherent measurements suffice, where K can be orders of magnitude less than N. Therefore, a sensor can transmit far fewer measurements to a receiver, which can

[6] "Approximation", Figure 2 <http://cnx.org/content/m18727/latest/#fs-id1169211320066>
[7] "Approximation": Section Nonlinear approximation <http://cnx.org/content/m18727/latest/#uid4>

reconstruct the signal and then process it in any manner. Moreover, the cK measurements need not be manipulated in any way before being transmitted, except possibly for some quantization. Finally, independent and identically distributed (i.i.d.) Gaussian or Bernoulli/Rademacher (random ± 1) vectors provide a useful **universal** basis that is incoherent with all others. Hence, when using a random basis, CS is universal in the sense that the sensor can apply the same measurement mechanism no matter what basis the signal is sparse in (and thus the coding algorithm is independent of the sparsity-inducing basis) [11], [16], [1].

These features of CS make it particularly intriguing for applications in remote sensing environments that might involve low-cost battery operated wireless sensors, which have limited computational and communication capabilities. Indeed, in many such environments one may be interested in sensing a **collection** of signals using a network of low-cost signals.

Other possible application areas of CS include imaging [27], medical imaging [7], [25], and RF environments (where high-bandwidth signals may contain low-dimensional structures such as radar chirps) [20]. As research continues into practical methods for signal recovery (see Section 3 (Methods for signal recovery)), additional work has focused on developing physical devices for acquiring random projections. Our group has developed, for example, a prototype digital CS camera based on a digital micromirror design [27]. Additional work suggests that standard components such as filters (with randomized impulse responses) could be useful in CS hardware devices [29].

5 The geometry of Compressed Sensing

It is important to note that the core theory of CS draws from a number of deep geometric arguments. For example, when viewed together, the CS encoding/decoding process can be interpreted as a linear projection $\Phi : \mathbb{R}^N \mapsto \mathbb{R}^M$ followed by a nonlinear mapping $\Delta : \mathbb{R}^M \mapsto \mathbb{R}^N$. In a very general sense, one may naturally ask for a given class of signals $\mathcal{F} \in \mathbb{R}^N$ (such as the set of K-sparse signals or the set of signals with coefficients $\| \alpha \|_{\ell_p} \leq 1$), what encoder/decoder pair Φ, Δ will ensure the best reconstruction (minimax distortion) of all signals in $\mathcal{F}$. This best-case performance is proportional to what is known as the Gluskin n-width [23], [22] of $\mathcal{F}$ (in our setting $n = M$), which in turn has a geometric interpretation. Roughly speaking, the Gluskin n-width seeks the $(N - n)$-dimensional slice through $\mathcal{F}$ that yields signals of greatest energy. This n-width bounds the best-case performance of CS on classes of compressible signals, and one of the hallmarks of CS is that, given a sufficient number of measurements this optimal performance is achieved (to within a constant) [16], [14].

Additionally, one may view the ℓ_0/ℓ_1 equivalence problem geometrically. In particular, given the measurements $y = \Phi x$, we have an $(N - M)$-dimensional hyperplane $\mathcal{H}_y = \{x^{'} \in \mathbb{R}^N : y = \Phi x^{'}\} = \mathcal{N}(\Phi) + x$ of feasible signals that could account for the measurements y. Supposing the original signal x is K-sparse, the ℓ_1 recovery program will recover the correct solution x if and only if $\| x^{'} \|_1 > \| x \|_1$ for every other signal $x^{'} \in \mathcal{H}_y$ on the hyperplane. This happens only if the hyperplane $\mathcal{H}_y$ (which passes through x) does not "cut into" the ℓ_1-ball of radius $\| x \|_1$. This ℓ_1-ball is a polytope, on which x belongs to a $(K - 1)$-dimensional "face." If Φ is a random matrix with i.i.d. Gaussian entries, then the hyperplane $\mathcal{H}_y$ will have random orientation. To answer the question of how M must relate to K in order to ensure reliable recovery, it helps to observe that a randomly generated hyperplane $\mathcal{H}$ will have greater chance to slice into the ℓ_1 ball as $\dim(\mathcal{H}) = N - M$ grows (or as M shrinks) or as the dimension $K - 1$ of the face on which x lives grows. Such geometric arguments have been made precise by Donoho and Tanner [15], [17], [19] and used to establish a series of sharp bounds on CS recovery.

6 Connections with dimensionality reduction

We have also identified [1] a fundamental connection between the CS and the JL lemma. In order to make this connection, we considered the **Restricted Isometry Property** (RIP), which has been identified as a key property of the CS projection operator Φ to ensure stable signal recovery. We say Φ has RIP of order

K if for every K-sparse signal x,

$$(1-\varepsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x\|_2}{\|x\|_2} \leq (1+\varepsilon)\sqrt{\frac{M}{N}}. \tag{5}$$

A random $M \times N$ matrix with i.i.d. Gaussian entries can be shown to have this property with high probability if $M = O\left(Klog\left(N/K\right)\right)$.

While the JL lemma concerns pairwise distances within a finite cloud of points, the RIP concerns isometric embedding of an **infinite** number of points (comprising a union of K-dimensional subspaces in $\mathbb{R}^N$). However, the RIP can in fact be derived by constructing an effective **sampling** of K-sparse signals in $\mathbb{R}^N$, using the JL lemma to ensure isometric embeddings for each of these points, and then arguing that the RIP must hold true for **all**K-sparse signals. (See [1] for the full details.)

7 Stable embeddings of manifolds

Finally, we have also shown that the JL lemma can also lead to extensions of CS to other concise signal models. In particular, while conventional CS theory concerns sparse signal models, it is also possible to consider manifold-based signal models. Just as random projections can preserve the low- dimensional geometry (the union of hyperplanes) that corresponds to a sparse signal family, random projections can also guarantee a stable embedding of a low-dimensional signal manifold. We have the following result, which states that an RIP-like property holds for families of manifold-modeled signals.

Theorem 3:
Let $\mathcal{M}$ be a compact K-dimensional Riemannian submanifold of $\mathbb{R}^N$ having condition number $\frac{1}{\tau}$, volume V, and geodesic covering regularity R. Fix $0 < \varepsilon < 1$and$0 < \rho < 1$. Let Φ be a random M $\times$ N orthoprojector with

$$M = O\left(\frac{K\log\left(NVR\tau^{-1}\varepsilon^{-1}\right)\log\left(\frac{1}{\rho}\right)}{\varepsilon^2}\right) \tag{6}$$

If $M \leq N$, then with probability at least $1-\rho$ the following statement holds: For every pair of points x_1, $x_2 \in \mathcal{M}$,

$$(1-\varepsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\varepsilon)\sqrt{\frac{M}{N}} \tag{7}$$

The proof of this theorem appears in [2] and again involves the JL lemma. Due to the limited complexity of a manifold model, it is possible to adequately characterize the geometry using a sufficiently fine sampling of points drawn from the manifold and its tangent spaces. In essence, manifolds with higher volume or with greater curvature have more complexity and require a more dense covering for application of the JL lemma; this leads to an increased number of measurements. The theorem also indicates that the requisite number of measurements depends on the geodesic covering regularity of the manifold, a minor technical concept which is also discussed in [2].

This theorem establishes that, like the class of K-sparse signals, a collection of signals described by a K-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$ can have a stable embedding in an M-dimensional measurement space. Moreover, the requisite number of random measurements M is once again linearly proportional to the information level (or number of degrees of freedom) K in the concise model. This has a number of possible implications for manifold-based signal processing. Manifold-modeled signals can be recovered from compressive measurements (using a customized recovery algorithm adapted to the manifold model, in contrast with sparsity-based recovery algorithms) [13], [31]; unknown parameters in parametric models can be estimated from compressive measurements; multi-class estimation/classification problems can be addressed [13] by considering multiple manifold models; and manifold learning algorithms may be efficiently executed by applying

them simply to the projection of a manifold-modeled data set to a low-dimensional measurement space [4]. (As an example, here[8](d) shows the result of applying the ISOMAP algorithm on a random projection of a data set from $\mathbb{R}^{4096}$ down to $\mathbb{R}^{15}$; the underlying parameterization of the manifold is extracted with little sacrifice in accuracy.) In all of this it is not necessary to adapt the sensing protocol to the model; the only change from sparsity-based CS would be the methods for processing or decoding the measurements. In the future, more sophisticated concise models will likely lead to further improvements in signal understanding from compressive measurements.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. The johnson-lindenstrauss lemma meets compressed sensing. 2006. Preprint.

[2] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 2008. To Appear.

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

[4] M.B. Wakin C. Hegde and R.G. Baraniuk. Random projections for manifold learning. In *In Proc. Neural Information Processing Systems (NIPS)*, December 2007.

[5] E. Cand[U+FFFD]nd J. Romberg. Practical signal recovery from random projections. 2005. Preprint.

[6] E. Cand[U+FFFD]nd J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. of Comp. Math.*, 2006. To appear.

[7] E. Cand[U+FFFD] J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2), February 2006.

[8] E. Cand[U+FFFD] J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 2006. To appear.

[9] E. Cand[U+FFFD]nd T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12), December 2005.

[10] E. Cand[U+FFFD]nd T. Tao. Error correction via linear programming. *Found. of Comp. Math.*, 2005. Preprint.

[11] E. Cand[U+FFFD]nd T. Tao. Near optimal signal recovery from random projections and universal encoding strategies. *IEEE Trans. Inform. Theory*, 2006. To appear.

[12] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. on Sci. Comp.*, 20(1):33–61, 1998.

[13] M.A. Davenport, M.F. Duarte, M.B. Wakin, J.N. Laska, D. Takhar, K.F. Kelly, and R.G. Baraniuk. The smashed filter for compressive classification and target recognition. In *Proc. Computational Imaging V at SPIE Electronic Imaging*, January 2007.

[14] R. A. DeVore. Lecture notes on compressed sensing. *Rice University ELEC 631 Course Notes*, Spring 2006.

[15] D. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. January 2005. Preprint.

[8] "Dimensionality Reduction", Figure 1 <http://cnx.org/content/m18732/latest/#fs-id4895761>

[16] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4), April 2006.

[17] D. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. 2005. Preprint.

[18] D. Donoho and Y. Tsaig. Extensions of compressed sensing. 2004. Preprint.

[19] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when then projection radically lowers dimension. Technical report 2006-11, Stanford University Department of Statistics, 2006.

[20] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk. Sparse signal detection from incoherent projections. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 2006.

[21] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk. Fast reconstruction of piecewise smooth signals from random projections. In *Proc. SPARS05*, Rennes, France, Nov. 2005.

[22] A. Garnaev and E. D. Gluskin. The widths of euclidean balls. *Doklady An. SSSR.*, 277:1048–1052, 1984.

[23] B. Kashin. The widths of certain finite dimensional sets and classes of smooth functions. *Izvestia*, (41):334–351, 1977.

[24] C. La and M. N. Do. Signal reconstruction using sparse tree representation. In *Proc. Wavelets XI at SPIE Optics and Photonics*, San Diego, August 2005. SPIE.

[25] M. Lustig, D. L. Donoho, and J. M. Pauly. Rapid mr imaging with compressed sensing and randomly under-sampled 3dft trajectories. In *Proc. 14th Ann. Mtg. ISMRM*, May 2006.

[26] S. Mallat. *A wavelet tour of signal processing.* Academic Press, San Diego, CA, USA, 1999.

[27] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, K. F. Kelly, and R. G. Baraniuk. A compressed sensing camera: New theory and an implementation using digital micromirrors. In *Proc. Computational Imaging IV at SPIE Electronic Imaging*, San Jose, January 2006. SPIE.

[28] J. Tropp and A. C. Gilbert. Signal recovery from partial information via orthogonal matching pursuit. April 2005. Preprint.

[29] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk. Random filters for compressive sampling and reconstruction. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 2006.

[30] R. Venkataramani and Y. Bresler. Further results on spectrum blind sampling of 2d signals. In *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, volume 2, Chicago, Oct. 1998.

[31] M. B. Wakin. *The Geometry of Low-Dimensional Signal Models.* Ph. d. thesis, department of electrical and computer engineering, Rice University, Houston, Tx, August 2006.