PROCEEDINGS OF SPIE

Signal recovery from random projections

Emmanuel J. Candes, Justin K. Romberg

SPIE.

Signal Recovery from Random Projections

Emmanuel Candès and Justin Romberg

Applied and Computational Mathematics, Caltech, Pasadena, CA, 91125

ABSTRACT

Can we recover a signal $f \in \mathbb{R}^N$ from a small number of linear measurements? A series of recent papers developed a collection of results showing that it is surprisingly possible to reconstruct certain types of signals accurately from limited measurements. In a nutshell, suppose that f is compressible in the sense that it is well-approximated by a linear combination of M vectors taken from a known basis Ψ. Then not knowing anything in advance about the signal, f can (very nearly) be recovered from about $M \log N$ generic nonadaptive measurements only. The recovery procedure is concrete and consists in solving a simple convex optimization program.

In this paper, we show that these ideas are of practical significance. Inspired by theoretical developments, we propose a series of practical recovery procedures and test them on a series of signals and images which are known to be well approximated in wavelet bases. We demonstrate that it is empirically possible to recover an object from about $3M$–$5M$ projections onto *generically chosen* vectors with an accuracy which is as good as that obtained by the ideal M-term wavelet approximation. We briefly discuss possible implications in the areas of data compression and medical imaging.

1. INTRODUCTION

In many fields of science and technology, one is often able to make only a limited number of measurements about an object of interest, i.e. a digital signal, a digital image and so on. Many important problems in medicine or astrophysics such as Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) of course all come to mind. Against this background, several recent works [1–3] have studied in depth the problem of recovering a signal from a highy incomplete set of measurements. Roughly speaking, the main results of these papers state that if a signal f has a parsimonious representation in a basis Ψ, it is possible to reconstruct (via a tractable optimization program) f accurately (or even exactly) from a small number of projections onto *randomly chosen* subspaces. To date, this work has been been for the most part theoretical in nature and mainly concerned with mathematical explanations for this remarkable phenomenon. However, we believe that these results can potentially impact several areas of signal processing. The purpose of this paper is to show that these ideas have practical import, and to demonstrate how they may deployed on real signals and images.

1.1. Exact Recovery of Sparse Signals

Suppose we wish to recover a finite signal $f \in \mathbb{R}^N$ from a set of K linear measurements

$$y_k = \langle f, u_k \rangle \quad k = 0, \ldots, K-1 \qquad \text{or} \qquad y = \mathcal{M}f \tag{1}$$

taken against vectors $u_k \in \mathbb{R}^N$ (u_k is the kth row of $\mathcal{M}$). Of special interest is the vastly underdetermined case, $K << N$, where there are many more unknowns than observations. Clearly, one cannot expect to recover every $f \in \mathbb{R}^N$ from the data y. However, recent work [1–3] has shown that if f is *sparse*, in the sense that f can be written as a superposition of a small number of vectors taken from a basis Ψ, then (1) exact recovery is possible and (2) the 'true' signal f actually is the solution to a simple convex optimization problem.

To make these ideas concrete, suppose that one collects K Fourier coefficients of f so that the *measurement ensemble* $\mathcal{M} = F_\Omega$ is the matrix obtained by sampling K rows of the N by N discrete Fourier transform matrix corresponding to frequencies in a set Ω. In other words, the data y is of the form

$$y_k = \langle f, e^{i\frac{2\pi\omega_k}{N}} \rangle, \quad \omega_k \in \Omega \subset \{0, \ldots, N-1\}, \quad |\Omega| = K. \tag{2}$$

Now suppose that f is sparse in the sense that only a few of its entries are non-zero; that is, we can write f as a superposition of M spikes*

$$f(t) = \sum_{\tau \in T} \alpha_\tau \delta(t - \tau), \tag{3}$$

for some $T \subset \{0, \ldots, N-1\}$, $|T| = M$. We do not know where the locations nor the amplitudes of the spikes. The central theorem of [1] states that for an overwhelming percentage of sets Ω with cardinality obeying

$$|\Omega| = K \geq \text{Const} \cdot M \log N, \tag{4}$$

f is the unique solution to the convex optimization program:

$$\min_{g \in \mathbb{R}^N} \|g\|_{\ell_1} := \sum_t |g(t)| \quad \text{subject to} \quad F_\Omega g = y. \tag{5}$$

That is, it is possible—with high probability—to recover f from the knowledge of its projection onto a randomly selected K-dimensional subspace (spanned by the rows of F_Ω).

The number of observations required in the Fourier domain is proportional–up to a logarithmic factor—to the number of non-zeros components of f and is optimal; if M is small, we can "undersample" the Fourier transform $\hat{f}$ drastically and still be able to recover f *without error*. In addition, the recovery procedure is *tractable*. It simply consists of solving a convex optimization problem (5). In fact, if f is real, (5) can be recast as a linear program (see [4] for a detailed description).

Results analogous to (4) can be derived for other measurement ensembles. In some sense, we can use almost any $\mathcal{M}$ to the same end. Specifically, suppose we generate a *Gaussian ensemble* by choosing each $(\mathcal{M})_{k,n}$ independently from a zero-mean, normal distribution with unit variance

$$(\mathcal{M})_{k,n} \sim N(0,1) \quad k = 0, \ldots, K-1, \quad n = 0, \ldots, N-1, \tag{6}$$

and use it to measure a sparse signal f, $y = \mathcal{M}f$. Again, if $K \geq \text{Const} \cdot M \log N$, then f is the unique solution to

$$\min \|g\|_{\ell_1} \quad \text{subject to} \quad \mathcal{M}g = y \tag{7}$$

with very high probability [3]. Results of this type with similar bounds hold for other random measurement ensembles as in the case where $(\mathcal{M})_{k,n}$ take values $\{-1, 1\}$, each with probability $1/2$.

When we construct the $K \times N$ measurement ensemble as in (6) and use it to measure f, we are essentially choosing a K dimensional subspace uniformly at random from the set of all K dimensional subspaces, and projecting f onto it. The fact that we can recover f means that although K can be much smaller than N, the projection retains enough critical information to specify f uniquely.

The Gaussian measurement ensemble easily allows to extend the results to signals that are sparse in any fixed orthonormal basis Ψ (in (3), f is sparse in the identity basis). To recover the signal, we modify (7) to search over coefficient sequences in the Ψ-domain:

$$\min \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \mathcal{M}\Psi\alpha = y. \tag{8}$$

Because the subspace is chosen uniformly at random, it does not matter which set of axes the signal is aligned with. Mathematically speaking, if $\mathcal{M}$ has i.i.d. Gaussian entries and Ψ is orthonormal, then the distribution of the random matrix $\mathcal{M}\Psi$ is exactly the same as that of $\mathcal{M}$; making measurements of f using $\mathcal{M}$ and solving (8) will recover signals with M-sparse representations in the Ψ domain when $K \geq \text{Const} \cdot M \log N$. This invariance property makes the Gaussian measurement ensemble especially attractive; we can recover sparse signals in *any* fixed basis from randomly sampled K measurement vectors with very high probability.

---

*The sparsity basis Ψ here is the identity matrix.

1.2. Approximate Recovery of Compressible SIgnals

In general, real-world signals are not exactly sparse in any orthogonal basis. Instead, a commonly discussed model [5,6] are *compressible* signals. A compressible signal is such that the reordered entries of its Ψ-coefficients α, $f = \Psi\alpha$, decay like a power-law; that is, when we rearrange the sequence of in decreasing order of magnitude $|\alpha|_{(1)} \geq |\alpha|_{(2)} \geq \cdots \geq |\alpha|_{(N)}$, the n-largest entry obeys

$$|\alpha|_{(n)} \leq \text{Const} \cdot n^{-s} \tag{9}$$

for some $s \geq 1$. Given an M, the M-term linear combination of elements which best approximate f in an L_2-sense is obtained by keeping the M largest terms in the expansion

$$f_M(t) = \sum_{n=0}^{M-1} \alpha_{(n)} \psi_{(n)}(t).$$

If α obeys (9), then the error between f_M and f also obeys a power-law:

$$\|f_M - f\|_2 \leq \text{Const} \cdot M^{-(s-1/2)}.$$

As such, accurate approximations to f can be constructed using a small number of vectors from Ψ.

This notion of compressibility is fundamental to signal processing. Both in theory and in practice, our ability to denoise and compress (among other things) a certain class of signals relies critically on finding a basis in which the expansion coefficients of these signals decay rapidly. Much energy over the past two decades has been devoted to finding bases in which certain types of signals have fast decay. Perhaps the most notable of these, the wavelet transform [7,8] and its later geometric extensions [9], have found widespread use due to their suitability for general classes of piecewise smooth signals, especially photograph-like images (see Figure 1).

No matter which of these bases we decide to use, the recovery via (8) from Gaussian measurements works as well for signals that are compressible as it does for signals that are truly sparse. If the expansion coefficients α obey (9) and we make $K \sim M \log N$ Gaussian measurements, the solution $f_K^\sharp$ to (8) has error on the same order as f_M [3]:

$$\|f_K^\sharp - f\|_2 \ \leq \ \text{Const} \cdot M^{-(s-1/2)}, \qquad K \sim M \log N. \tag{10}$$

It is worth stepping back for a moment and looking at these results in a broader context. Geometrically, the class of signals for which we can construct accurate M-term approximations in a basis Ψ is tightly clustered around a union of M-dimensional subspaces in $\mathbb{R}^N$. We have seen that we can (nearly) recover a signal in this class after projecting it onto a subspace of dimension about $M \log N$; the projection of the cluster onto this subspace is (almost) invertible. In addition, we do not have to search for "special" subspaces for which this is true. Choosing a subspace in a completely random manner will work almost all of the time.

Our framework is a significant departure from the current zeitgeist. In some information theoretic sense, the most efficient way to represent information about a smooth signal with isolated singularities is by way of its M largest wavelet coefficients [5]. However and in the scenario considered in this paper, one would need to have available an *oracle* letting us know which coefficients are large beforehand. What we are suggesting here is radically different. The observations $y = \mathcal{M}f$ are made against vectors that are completely random; they have nothing to do with the structure of the underlying signal, and hence they certainly do not provide a sparse or compressible representation of the signal. Yet using (8) to recover the image from $K \sim M \log N$ random observations works nearly as well as constructing the optimal approximation from M wavelets, both in theory (10), and in practice, as we shall see in the next section. The inconvenience of adapting the measurements to the signal has also been avoided; the same set of K linear functionals can be used to recover every M-compressible signal as (10) holds (with overwhelming probability) for all signals obeying (9) [3].

2. PRACTICAL RECOVERY

The theoretical results in the previous section indicate that we can recover the signal just as well from $K \sim M \log N$ observations in a random basis as we could from the M *best* observations in the sparsity basis Ψ. These results are rather asymptotic in nature, and the practical viability of these ideas depends largely upon the constants (4),(10) not being too large. In this section, we will present numerical experiments which suggest that in practice, we can recover an approximation as close to the original signal as the optimal M-term approximation by using about $3M$ to $5M$ random observations.

For concreteness, we will take Ψ to be a discrete wavelet transform for signals of length $N = 2^J$; in 1D the coefficients α are parameterized by a scale j and a shift m

$$f(t) = \Psi\alpha = \sum_{m=0}^{2^L-1} \alpha_{L,m} \varphi_{L,m}(t) + \sum_{j=L+1}^{J} \sum_{m=0}^{2^{j-1}-1} \alpha_{j,m} \psi_{j,m}(t). \tag{11}$$

The $\varphi_{L,m}$ are scaling functions and the $\psi_{j,m}$ are oscillatory wavelets, see [7] for details. We will refer to the collection of coefficients at a given scale $\{\alpha_{j,m}, m = 0, \ldots, 2^{j-1} - 1\}$ as a *subband*. In the separable extension of (11) to 2D, there are three subbands at each scale. All experiments in this paper use the 'Daubechies-8' wavelet [8].

The wavelet transform is tremendously popular in the signal and image processing communities, due in large part to its ability to provide parsimonious representations for signals that are smooth away from isolated discontinuities (examples are shown in Figure 1). The ability to construct low-dimensional, accurate approximations make wavelets particularly useful for image compression [10] and restoration [11]. In the remainder of this section, we will demonstrate that from a projection onto a non-adaptive, randomly selected subspace of similarly low dimension, we can recover equally accurate approximations to real-world signals and images.

2.1. A Practical Recovery Algorithm

The recovery problem (8) is convex, and when f is real, it can be recast as a linear program and solved using modern interior-point methods [4]. As discussed in the last section, in doing so we recover the signal to within near optimal accuracy. In practice, however, there are two drawbacks to recovery via (8).

First, the iterative interior point methods, although tractable and stable, require a $K \times K$ system of equations to be solved at each step. We will be interested in cases where K is less than, but still a significant fraction of, the signal length N. If $N \sim 10^6$ (as for a typical image), the procedure is currently not feasible for typical computing environments.

Second, although (10) guarantees that the recovered signal will be close to the original in Euclidean distance, there are often visually undesirable artifacts in the solution. A 1D example of recovery from a projection onto a random subspace is shown in Figure 2(a); we can see that there are spurious fine-scale oscillations in the reconstruction. Because the ℓ_1 norm functional does not distinguish between coefficients at different scales, energy has been shifted from coarse to fine scale.

To address these shortcomings, we propose a slightly different recovery procedure than (8) that requires a small amount of *a priori* information about the signal we are to recover, but allows us to impose other desirable characteristics on the solution. Our scheme is again iterative, but the cost of each iteration will be much lower, essentially that of applying the measurement ensemble.

To begin, suppose that the condition for exact reconstruction is satisfied; that is, the wavelet coefficients α are non-zero only on a small set T. Let $\mathcal{M}' = \mathcal{M}\Psi$ be the measurement matrix expressed in the wavelet domain. Since α is the unique solution to (8), the ℓ_1-ball $B = \{\beta : \|\beta\|_{\ell_1} \leq \|\alpha\|_{\ell_1}$ and the hyperplane $H = \{\beta : \mathcal{M}'\beta = y\}$ meet at exactly one point; $B \cup H = \{\alpha\}$. Thus, if in addition to the random projection values $y = \mathcal{M}'\alpha$, we are given the ℓ_1-norm $\|\alpha\|_{\ell_1}$, we can recover α (and hence f) by finding the point of intersection. Since both B and H are convex, this can be accomplished with the alternate projections onto convex sets (POCS) algorithm [12].

From a starting point α_0 outside of B, we iterate by alternating projections onto H, then onto B. The algorithm is guaranteed to converge to a point (in this case, α) in $B \cup H$ [12]. Both of the projections are fairly straightforward.

Projection onto H:

To find the closest vector $\hat{\beta}$ in H to an arbitrary $\beta \in \mathbb{R}^N$, we use the formula

$$\hat{\beta} = \beta + \mathcal{M}'^{*}(\mathcal{M}'\mathcal{M}'^{*})^{-1}(y - \mathcal{M}'\beta).$$

The computational cost of a projection onto H is essentially the cost of applying $\mathcal{M}' = \mathcal{M}\Psi$ and its adjoint $\mathcal{M}'^{*}$. (If necessary, $(\mathcal{M}'\mathcal{M}'^{*})^{-1}$ can be computed beforehand; if the rows of $\mathcal{M}$ are orthogonal, then $(\mathcal{M}'\mathcal{M}'^{*})^{-1} = I_K$.) Applying the wavelet transform Ψ is fast — it requires $O(N)$ operations. Applying the $K \times N$ measurement matrix $\mathcal{M}$ will require $K \cdot N$ operations in general, which can be prohibitively slow when N (and K) are large. However, if the measurement matrix is structured (as in the randomly sampled Fourier case (2)), the complexity can be reduced to $O(N \log N)$, say.

Projection onto B:

To project a vector β onto the ℓ_1-ball B we simply apply a soft-thresholding operation

$$\hat{\beta}(t) = \begin{cases} \beta(t) - \gamma & \beta(t) > \gamma \\ 0 & |\beta(t)| \leq \gamma \\ \beta(t) + \gamma & \beta(t) < -\gamma \end{cases}.$$

To determine the threshold γ such that $\|\hat{\beta}\|_{\ell_1} \leq \|\alpha\|_{\ell_1}$, we sort the coefficients by magnitude and perform a linear search, a procedure requiring $O(N \log N)$ operations.

Of course, when f can only be approximated — instead of represented exactly — by a superposition of a small number of vectors, $B \cup H$ will in general contain more than just α. From these, we will choose the coefficients corresponding to the signal that has smallest *total variation* (TV) by solving

$$\min \|\Psi\beta\|_{TV} \quad \text{such that} \quad \mathcal{M}'\beta = y, \quad \|\beta\|_{\ell_1} \leq \|\alpha\|_{\ell_1}, \tag{12}$$

where

$$\|g\|_{TV} = \sum_{s,t} \sqrt{(g(s+1,t) - g(s,t))^2 + (g(s,t+1) - g(s,t))^2}$$

is the sum of the magnitudes of the (discretized) gradient of $g(s,t)$. The TV functional is well-known in signal and image processing for its tendency to suppress spurious high-frequency features [13,14]; roughly speaking, (12) will choose the signal in $B \cup H$ with the least amount of oscillation. Again, (12) can be solved with an iterative scheme with two simple stages: a step in the direction of the (negative) TV gradient, followed by a projection onto $B \cup H$ using POCS. If we start the procedure from a feasible point (by running POCS beforehand), it takes very few POCS iterations to re-project onto $B \cup H$. Note that in practice, finding an exact solution to (12) is not absolutely necessary. After finding a feasible initial point, each descent iteration can be viewed as an improvement on the last, and the process can be terminated at the behest of the user. In our experience, only a few iterations are needed to remove high-frequency artifacts.

We can further restrict the feasible set by imposing a few additional constraints. In some ways, the condition (9) is a bit lax in that it does not place any restrictions whatsoever on where the large wavelet coefficients can appear. For signals and images such as those in Figure 1, we expect the large wavelet coefficients to appear at coarse scales (small values of j in (11)). To account for this, we will constrain the ℓ_1-norm of each *subband* $\alpha_j = \{\alpha_{j,m}, m = 0, \ldots, 2^{j-1} - 1\}$ of the wavelet transform. In the experiments which follow, we will assume that the values $\|\alpha_j\|_{\ell_1}$ are known. Admittedly, this falls slightly outside of our framework of making only random linear measurements of the signal f. However, we emphasize that our primary purpose here is to explore the potential performance of recovery from random projections and in this direction, establishing a few benchmarks may be of benefit. In addition, there is little doubt that future research will provide compeititive alternatives to these additional assumptions. Equipped with the $\|\alpha_j\|_{\ell_1}$'s , our program (12) becomes

$$\min \|\Psi\beta\|_{TV} \quad \text{such that} \quad \mathcal{M}'\beta = y, \quad \|\beta_j\|_{\ell_1} \leq \|\alpha_j\|_{\ell_1}, \; j = L, \ldots, J \tag{13}$$
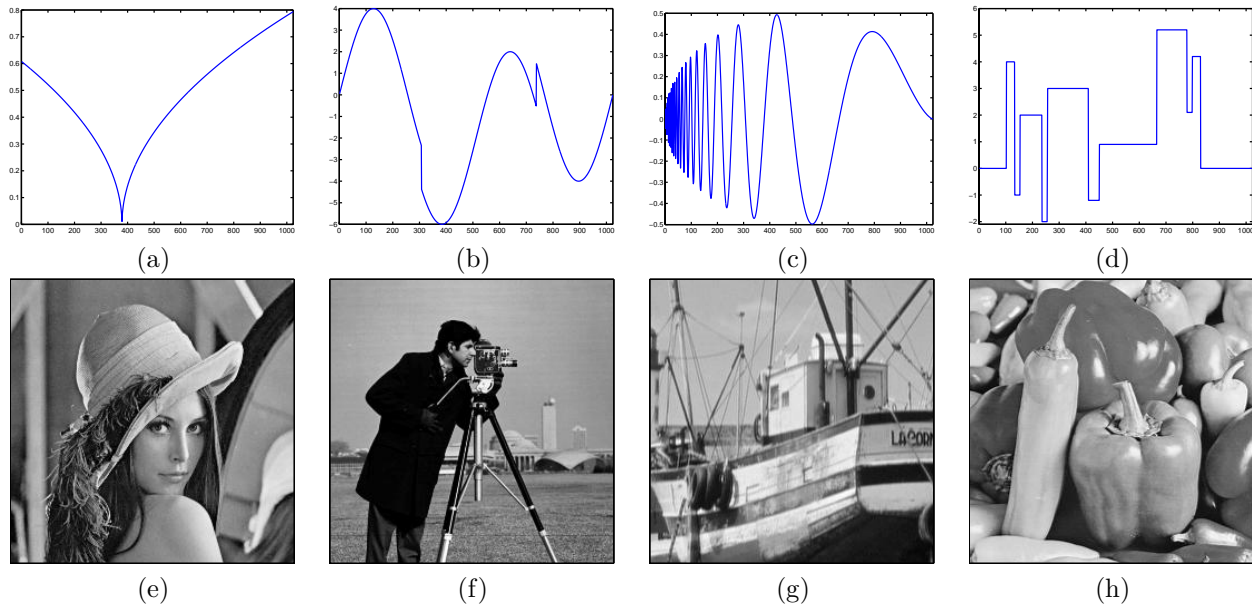
Figure 1. *Test signals and images. (a) 'Cusp', (b) 'Heavisine', (c) 'Doppler', (d) 'Blocks', (e) 'Lena', (f) 'Camera', (g) 'Boats', (h) 'Peppers'*

2.2. Examples

We start with four 1D examples with $N = 1024$ taken from the well-known Donoho-Johnstone [11] collection of synthetic test signals. The 'Cusp', 'Heavisine', 'Doppler', and 'Blocks' signals shown in Figure 1(a)–(d) are all smooth away from isolated singularities, and hence are compressible in the wavelet domain. Since N is of modest size here, applying $\mathcal{M}$ directly is not a computational burden, and select $\mathcal{M}$ by sampling a matrix with i.i.d. Gaussian entries which we then orthonormalize. The results for several values of K are shown in Figure 2 and Table 1; using (13) reconstructs the signal from $3M$–$5M$ random projections with the same accuracy as the best M-term wavelet approximation.

Recovering 2D images from random projections via (13) works equally well. In two dimensions, however, applying the $K \times N$ measurement matrix is unwieldy, and we therefore sample the random projection in a different way. We consider applying random Fourier matrices to *in the wavelet domain* (where f becomes a decaying spike train). In theory and for coefficient sequences behaving as in (9), we can recover f from $F_\Omega \alpha$ essentially as well as from $\mathcal{M}'\alpha$ [3]. In addition, there is of course a fast algorithm (the FFT) to apply F_Ω and F_Ω^*, so that the projections onto H can be rapidly computed. We will be careful to count each measured Fourier coefficient of α as two observations: $y_{2k} = \text{Re}\langle \alpha, e^{i2\pi\omega_k/N}\rangle, y_{2k+1} = \text{Im}\langle \alpha, e^{i2\pi\omega_k/N}\rangle$.

The venerable 'Lena', 'Cameraman', 'Boats', and 'Peppers' images are shown in Figure 1(e)–(h). On each image f, we computed the the solution $f^\sharp$ to (13) for various values of K and different measurement matrices $F_\Omega\Psi$. The empirical recovery errors $\|f - f^\sharp\|_2$ are tabulated in Table 2 and compared against those for the optimal M-term wavelet approximations. Again, the recovery from $3M$–$5M$ random projections is comparable to the best M-term wavelet partial reconstruction.

As our final example, we will consider the common task in medical imaging of reconstructing an image from a limited number of Fourier samples[†]. The sampling domains can follow any one of a number of patterns, one of the most prevalent is the 'star-shaped' domain shown in Figure 4(a), where we are given high-resolution samples along a relatively small number of equally spaced lines through the origin. As in (2), our measurement ensemble is a partial Fourier transform $\mathcal{M} = F_\Omega$.

[†]In most scenarios, these are samples of a continuous-space image that do not lie on a cartesian grid, making the situation somewhat more complicated. We are considering a simplified version here, where we are given samples of digital images that lie on the usual discrete Fourier grid.

Table 1. *Recovery results for 1D synthetic test signals of length $N = 1024$. Column K denotes the number of observations, E_K is the average recovery error over 10 experiments, and K_{nla} is the number of coefficients in the lowest-order wavelet approximation that achieves error E_K. The 'Blocks' signal is recovered perfectly for each value of K shown.*

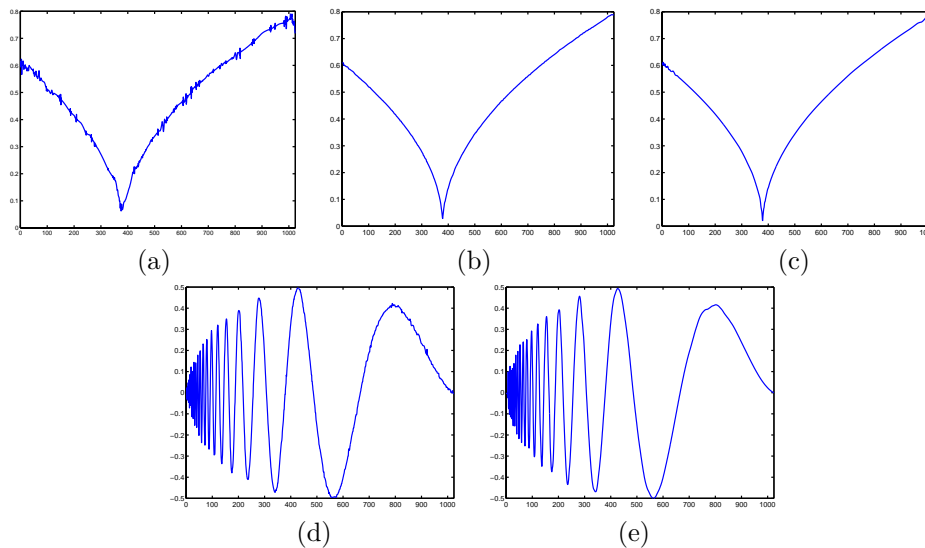|	Cusp		Heavisine		Doppler		Blocks	
K	E_K	K_{nla}	E_K	K_{nla}	E_K	K_{nla}	E_K	K_{nla}
100	0.1559	31	4.0211	23	3.0732	16	0	67
150	0.0390	45	1.2661	40	1.7101	29	0	67
200	0.0129	57	0.2734	61	1.0186	40	0	67
250	0.0057	63	0.0239	89	0.6613	51	0	67
300	0.0021	69	0.0084	98	0.4222	67	0	67
350	0.0009	75	0.0038	103	0.2379	85	0	67

Figure 2. *Top: 'Cusp' recovery example with $K = 150$. (a) Solution to (8). (b) Solution to (12). (c) Wavelet approximation with same error, $K_{nla} = 46$. Bottom: 'Doppler recovery example with $K = 320$. (c) Solution to (12). (d) Wavelet approximation, $K_{nla} = 82$.*

Table 2. *Recovery results for 256×256 images. $PSNR_K$ is the peak signal to noise ratio between the original image f and the recovered image $f^\sharp$, $PSNR_K(f, f^\sharp) := 20 \log_{10}(\|f - f^\sharp\|_2/255)$.*

	Lenna		Camera		Boats		Peppers	
K	$PSNR_K$	K_{nla}	$PSNR_K$	K_{nla}	$PSNR_K$	K_{nla}	$PSNR_K$	K_{nla}
10000	26.5	1769	26.2	2024	26.7	2050	21.6	1966
15000	28.7	3125	28.7	3562	29.8	3896	25.3	4005
20000	30.4	4501	30.9	5415	31.8	5474	27.5	5638
25000	32.1	6148	33.0	7539	33.7	7207	29.4	7366

Figure 3. 256×256 *image recovery examples. Left: 'Lena' with $K = 25000$. (a) Recovered from random projections via (12). (b) Optimal 6148-term wavelet approximation with same PSNR as (b). Right: 'Boats' with $K = 20000$. (c) Recovered from random projections via (12). (d) Optimal 7207-term wavelet approximation with same PSNR as (b).*

The reconstruction framework is slightly different here; the samples are not randomly selected, rather they are dictated by the imaging device. (The fact that the results from the previous section hold with high probability means that solving (7) will be effective for most sampling configurations.) But in the end, the basic problem is the same: we wish to recover an image from its projection onto a low dimensional subspace.

We also do not have access to the ℓ_1 norms of the wavelet coefficients of the underlying image. Because of this, we will recover the image by solving (12) in Lagrange form

$$\min \|\Psi\beta\|_{TV} + \lambda \cdot \|\beta\|_{\ell_1} \quad \text{such that} \quad \mathcal{M}'\beta = y \tag{14}$$

for a user specified value of $\lambda \geq 0$. The choice of λ prescribes the importance of the solution having small ℓ_1 norm in the wavelet domain versus having small TV norm in the spatial domain. Note that (12) and (14) are essentially the same in that the ℓ_1 constraint in (12) will dictate the value of λ used in (14). We can also extend (14) (as we did with (13)) to accommodate a different parameter λ_j for each scale.

The top row of Figure 4(b)–(d) shows the reconstruction result for the classic Shepp-Logan phantom from a sampling domain consisting of 22 radial lines. The recovery is perfect. In light of our previous results, this is to be expected, since the phantom is piecewise-constant: not only are most of its wavelet coefficients exactly zero (it is strictly sparse), but also that it has very small total variation. For comparison, column (b) shows the 'minimum energy' reconstruction obtained by setting the unobserved Fourier coefficients to zero (this is basically what a simple backprojection algorithm would do). The bottom row of Figure 4 shows a similar result. The image is again very simple; it is simply a superposition of 10 ellipses. However, notice that even even the fine-scale features (such as the small ellipse in the upper left) are recovered perfectly by solving (12).

A more realistic example is shown in Figure 5. Here we take a high-resolution angiogram (image of blood vessels in the brain), project it onto a Fourier star-domain with 80 lines, and reconstructed using (12). By searching for an image that matches the observations while having rapidly decaying wavelet coefficients and small TV norm, we are able to retain much of the detail of the original, while eliminating many of the artifacts that appear in the backprojection.

3. DISCUSSION

This paper introduced promising early experiments suggesting that it is possible to reconstruct an object accurately from a few random projections. Indeed, our experiments show consistently that one can recover an object from the knowledge of about $3K$ to $5K$ random coefficients, with the same accuracy as if one had knowledge of the K most significant wavelet coefficients. Although, this seems like an unexpected feat, we would like to emphasize that this is only the tip of the iceberg as there are many ideas in the literature that would certainly further enhance our reconstruction algorithms. We mention a few such possibilities.
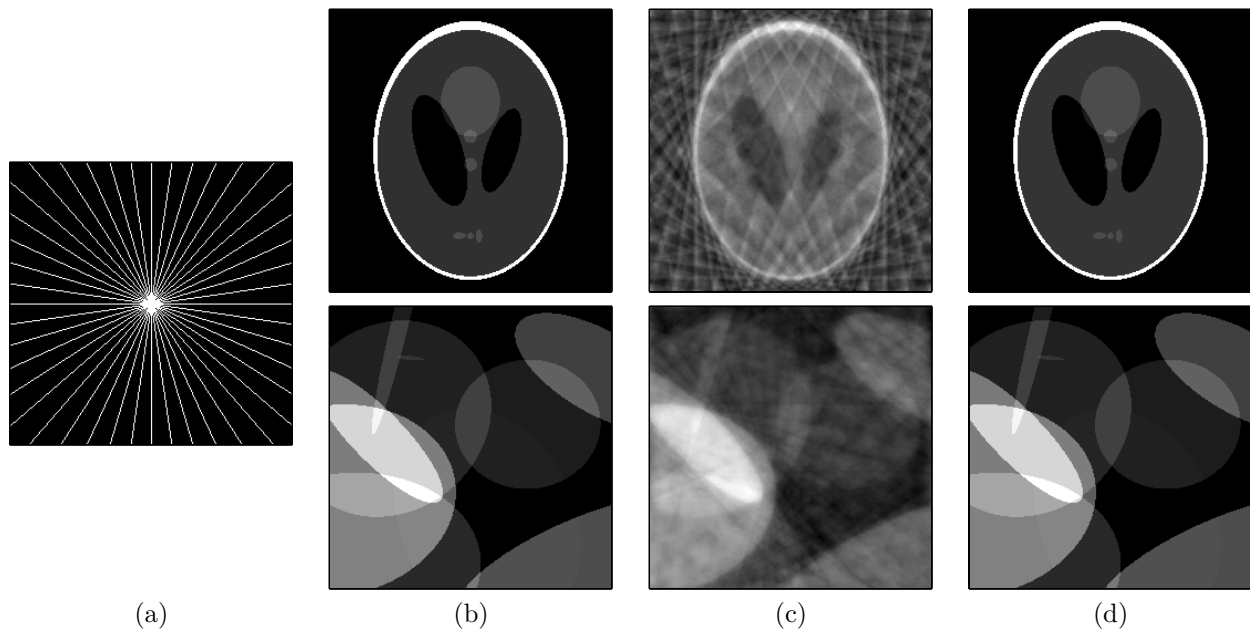
Figure 4. *Exact reconstruction from limited Fourier domain observations. Top row: Shepp-Logan phantom, Bottom row: Superposition of 11 ellipses. (a) Observation locations in the Fourier domain. Fourier coefficients along 22 radial lines are observed, $K = 5481$ in all. (b) Original images. (c) Minimum energy reconstruction. (d) Recovered via (12). Both reconstructions match the originals exactly.*
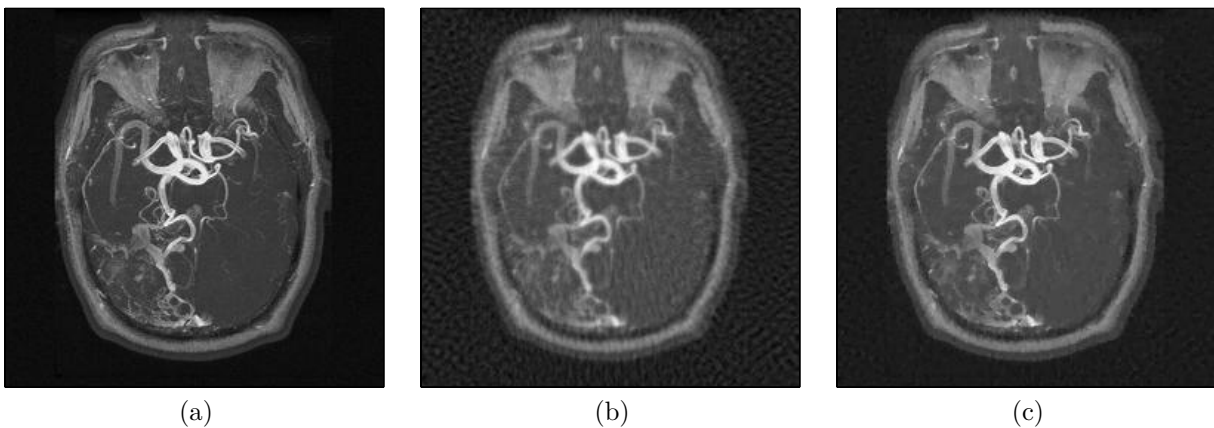

Figure 5. *Angiogram example, with observations along 80 lines in the Fourier domain, $K = 16129$. (a) Original. (b) Backprojection. (c) Recovered via (12).*

- *Undecimated Wavelet Transforms.* It is known that in certain areas of signal processing such as "de-noising," it is best to use undecimated wavelet transforms as opposed to orthogonal bases of wavelets. Therefore, it is likely that searching for the sparsest decomposition in an undecimated wavelet tree while fitting the data would improve upon the results presented in this paper.

- *Combined representations.* We limited our exposition to wavelet bases although it is clear that other representations might offer enhanced sparsity [9, 14]. In addition, a popular approach nowadays consists in combining several representations as to provide the flexibility of efficiently representing different feature types with different basis functions [15]. It is clear that our methodology is compatible with these ideas as one could easily search for the sparsest decomposition in a combined dictionary of waveforms, e.g. a dictionary including for example local cosines for textures , wavelets for point-like singularities, and curvelets for edges.

- *Other variational functionals.* This paper focussed on the Total-Variation norm while obviously, there are many other ways of regularizing the object we wish to reconstruct, see the vast literature on inverse problems [16].

- *Additional constraints.* Finally, in many application of interest, we often have a priori information that can easily be incorporated in the reconstruction strategy. In medical imaging or astronomy for example, it is often the case that the object under study is nonnegative. It is our experience that adding seemingly uninformative constraints of this type may in fact boost significantly the performance of the reconstruction algorithm.

Our results may have the potential for impact in several areas of signal processing. To conclude the paper, consider the compression problem for example. In [3], it was proposed to use random projections as a universal coding strategy. The *encoder* would encode correlations between the signal and randomly sampled waveforms, send them to the *decoder*, which would reconstruct the signal by searching for that sparse decomposition matching the transmitted information. Obviously, the decoder would need to know the random waveforms, which in practice can be accomplished by synchronizing the random number generators at either end by communicating a seed (or otherwise).

The performance of this type compression scheme depends, of course, on the number of measurements needed to recover a signal to within a prescribed error, the experimental focus of this paper. We have seen that we require about four or five times as many random measurements as (optimally selected) wavelets coefficients to get the same error. Against this background, several comments are in order.

- *No position information is encoded.* Wavelet based coders (and other transform coders) spend bits not only encoding important coefficients, but also specifying *which* coefficients are important. Some wavelet coefficients are far more important than others, and which coefficients these are vary from signal to signal. The cost of encoding (either implicitly or explicitly) this auxiliary information is significant [10]; much of the research in image compression in centered on lowering this cost [17–19].

  In contrast, every observation in our random measurement compression scheme is equally (un)important. Indeed, the subspace onto which the encoder projects the signal is completely arbitrary, and can again be determined by the decoder with careful synchronization.

- *Quantization is straightforward.* The projection coefficients are, by construction, random variables. In fact, if we use the Gaussian ensemble, the observations are independent, identically distributed normal random variables with zero mean and variance $\|f\|_2^2$. Thus, standard quantization schemes can be directly applied.

- *Robustness.* A fundamental problem with most existing coding strategies is their fragility vis a vis bit-loss. Take JPEG 2000, the current digital still-picture compression standard, for example. All the bits in JPEG 2000 do not have the same value and if important bits are missing (e.g. because of packet loss), then there is simply no way the information can be retrieved accurately.

  The situation is very different when one is using the scheme suggested above. Suppose that in each packet of information, we have both encoded the (quantized) value of the coefficients y_k but also the label of the

corresponding coefficients k. Consider now a situation in which half of the information is lost in the sense that only half of the coefficients are actually received. Then one would still be able to reconstruct an object from $K/2$ coefficients which is nearly as good as the best $K/2$ wavelet approximation!

- *Flexibility.* The encoding strategy is, of course, independent of the methodology deployed at the decoder's end. This is an appealing feature since it is likely that future research will develop better image representations, hence offering the opportunity to provide lower distortions without having to re-encode the signal.

For all these reasons and others, it seems worthwhile to pursue the potential of these ideas for compression. Actually, the design of a full-fledged encoding scheme, and the corresponding rate-distortion analysis, is a topic of current research. We hope to report on our progress in a future paper.

REFERENCES

1. E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *submitted to IEEE Trans. Inform. Theory* , June 2004.
2. E. Candès and J. Romberg, "The role of sparsity and incoherence for exactly reconstructing a signal from limited measurements," *Technical Report, California Institute of Technology* , 2004.
3. E. Candès and T. Tao, "Near-optimal signal recovery from random projections and universal encoding strategies," *submitted to IEEE Trans. Inform. Theory* , November 2004. Available on the ArXiV preprint server: `math.CA/0410542`.
4. S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing* **20**, pp. 33–61, 1999.
5. D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inform. Theory* **44**, pp. 2435–2476, 1998.
6. D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *J. Comp. and Appl. Harmonic Analysis* **1**, pp. 100–115, Dec. 1993.
7. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, second ed., 1999.
8. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, New York, 1992.
9. E. Candès and D. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise c^2 singularities," *Comm. on Pure and Applied Math.* **57**, pp. 219–266, 2004.
10. B. E. Usevich, "A tutorial on modern lossy image compression: Foundations of JPEG 2000," *IEEE Signal Processing Mag.* , pp. 22–35, September 2001.
11. D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika* **81**, pp. 425–455, 1994.
12. L. M. Bregman, "The method of successive projection for finding a common point of convex sets," *Soviet Math. Dokl.* **6**, pp. 688–92, 1965.
13. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation noise removal algorithm," *Physica D* **60**, pp. 259–68, 1992.
14. E. Candès and F. Guo, "New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction," *Signal Processing* **82**, pp. 1519–43, 2002.
15. J. L. Starck, D. L. Donoho, and E. J. Candès, "Astronomical image representation by the curvelet transform," *Astronomy and Astrophysics* **398**, pp. 785–800, 2003.
16. M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing Ltd., 1998.
17. J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing* **41**, pp. 3445–3462, Dec. 1993.
18. Z. Xiong, K. Ramchandran, and M. T. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Trans. Image Processing* **6**(5), pp. 677–693, 1997.
19. A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.* **6**, pp. 243–250, June 1996.