Extensions of Compressed Sensing

Yaakov Tsaig
David L. Donoho

October 22, 2004

Abstract

We study the notion of Compressed Sensing (CS) as put forward in [14] and related work [20, 3, 4]. The basic idea behind CS is that a signal or image, unknown but supposed to be compressible by a known transform, (eg. wavelet or Fourier), can be subjected to fewer measurements than the nominal number of pixels, and yet be accurately reconstructed. The samples are nonadaptive and measure 'random' linear combinations of the transform coefficients. Approximate reconstruction is obtained by solving for the transform coefficients consistent with measured data and having the smallest possible ℓ^1 norm.

We perform a series of numerical experiments which validate in general terms the basic idea proposed in [14, 3, 5], in the favorable case where the transform coefficients are sparse in the strong sense that the vast majority are zero. We then consider a range of less-favorable cases, in which the object has all coefficients nonzero, but the coefficients obey an ℓ^p bound, for some $p \in (0, 1]$. These experiments show that the basic inequalities behind the CS method seem to involve reasonable constants.

We next consider synthetic examples modelling problems in spectroscopy and image processing, and note that reconstructions from CS are often visually "noisy" . We post-process using translation-invariant de-noising, and find the visual appearance considerably improved.

We also consider a multiscale deployment of compressed sensing, in which various scales are segregated and CS applied separately to each; this gives much better quality reconstructions than a literal deployment of the CS methodology.

We also report that several workable families of 'random' linear combinations all behave equivalently, including random spherical, random signs, partial Fourier and partial Hadamard.

These results show that, when appropriately deployed in a favorable setting, the CS framework is able to save significantly over traditional sampling, and there are many useful extensions of the basic idea.

Key Words and Phrases: Basis Pursuit. Underdetermined Systems of Linear Equations. Linear Programming. Random Matrix Theory.

Acknowledgements. Partial support from NSF DMS 00-77261, and 01-40698 (FRG) and ONR. Thanks to Michael Saunders for optimization advice, and Emmanuel Candès for discussions of his own related work with J. Romberg and T. Tao. Raphy Coifman asked us insistently about the question of the constants in (1.2) which inspired this report.

1 Introduction

In the modern multimedia world, 'everyone' knows that all humanly-intelligible data are highly compressible. In exploiting this fact, the dominant approach is to first sample the data, and then eliminate redundancy using various compression schemes. This raises the question: why is it

necessary to sample the data in a pedantic way and then later to compress it? Can't one directly acquire a compressed representation? Clearly, if this were possible, there would be implications in a range of different fields, extending from faster data acquisition, to higher effective sampling rates, and lower communications burden.

Several recent papers [20, 3, 14, 5], have shown that, under various assumptions, it may be possible to directly acquire a form of compressed representation. In this paper, we put such ideas to the test by making a series of empirical studies of the effectiveness of compressed sensing schemes.

1.1 The Approach

We adopt language and notation from [14]; the approach there is rather abstract, but has the advantage of factoring across many potential applications areas.

We assume the object of interest is a vector $x_0 \in \mathbf{R}^m$ – this could represent the m sampled values of a digital signal or image. We assume the object is *a priori* compressible by transform coding of the type used in e.g. JPEG or JPEG-2000. Mathematically, we let Ψ denote the matrix of the orthogonal transform in question, having columns ψ_i, $i = 1, \ldots, m$. By 'compressible' we mean that for some $p \leq 1$ and moderate $R > 0$, $\|\Psi^T x_0\|_p \leq R$; see [14] for further explanation of this condition, which imposes sparsity on the transform coefficients $\Psi^T x_0$.

To make our compressed measurements, we start from an n by m matrix Φ with $n < m$ satisfying certain conditions called CS1-CS3 in [14]. We form the matrix $\Xi = \Phi\Psi^T$, also $n \times m$. We take the n-pieces of measured information $y = (y(i) : 1 \leq i \leq n)$ according to the linear system $y = \Xi x_0$. Since $n < m$ this amounts to having fewer measurements than degrees of freedom for the signal x_0.

The individual measured values are of the form $y(i) = \langle \xi_i, x_0 \rangle$, i.e. each can be obtained by integrating the object x_0 against a measurement kernel ξ_i. Here ξ_i denotes the i-th row of Ξ; from the formula $\Xi = \Phi\Psi^T$ and the known properties of CS-matrices Φ, we may view each measurement kernel as a kind of 'random' linear combination of the basis elements ψ_j.

To reconstruct an approximation to x_0 we solve

$$(L_1) \quad \min_x \|\Psi^T x\|_1 \text{ subject to } y = \Xi x. \tag{1.1}$$

Call the result $\hat{x}_{1,n}$. In words, $\hat{x}_{1,n}$ is, among all objects generating the same measured data, the one having transform coefficients with the smallest ℓ^1 norm. In [14] it was mentioned that this reconstruction procedure can be implemented by linear programming, and so may be considered computationally tractable. Also the needed matrices Φ satisfying CS1-CS3 were shown to be constructible by random sampling from a uniform distribution on the columns of Φ.

In short the approach involves linear, nonadaptive measurement, followed by nonlinear approximate reconstruction.

The paper [14] proved error bounds showing that despite the apparent undersampling ($n < m$), good accuracy reconstruction is possible for compressible objects. Such bounds take the form

$$\|\hat{x}_{1,n} - x_0\|_2 \leq C_p \cdot R \cdot (n/\log(m))^{1/2-1/p}, \qquad n, m > n_0. \tag{1.2}$$

As $p < 2$, this bound guarantees that reconstruction is accurate for large n, with a very weak dependence on m. Such bounds were interpreted in [14] to say that n measurements with $n = O(N \log(m))$ are just as good as knowing the N biggest transform coefficients. Examples were sketched for model problems caricaturing imaging and spectroscopy.

In related prior work, classical literature in approximation theory [23, 19, 27] (developing the theory of Gel'fand n-widths) deals with closely related problems from an even more abstract

viewpoint; see the discussion in [14]. More recently, Gilbert et al. [20] considered n-by-m matrices Φ made of n special rows out of the m-by-m Fourier matrix, while Candès, Romberg and Tao considered matrices Φ made of n randomly chosen rows from the Fourier matrix. Candès, Romberg, and Tao [3] considered also the use of ℓ^1 minimization, just as here while Gilbert et al. [20] considered a different nonlinear procedure. Candès has informed us of work in preparation considering also random matrices.

Finally we note that at some level of generality, we are discussing here the idea of getting sparse solutions to underdetermined systems of equations using ℓ^1 methods, which forms part of a now-extensive body of work: [6, 9, 11, 12, 13, 16, 17, 21, 22, 25, 28, 29]. We expect that many of the authors of the just-cited papers will be contributing to this special issue

1.2 Questions ...

Readers may want to ask numerous questions about the result (1.2), starting with:

- *How large do n and m have to be?* Is (1.2) meaningless for practical problem sizes?
- *How large is the constant C?* Even if (1.2) applies, perhaps the constant is miserable?
- *When the object is not perfectly reconstructed, what sorts of errors occur?* Perhaps the error term, though controlled in size by (1.2), has an objectionable structure?

They might continue with

- *How should the CS framework be applied to realistic signals?* In [14] models of spectroscopy and imaging were considered; for such models, it was proposed to deploy CS in a hybrid strategy, with coarse-scale measurements obtained by classical linear sampling, and the bulk obtained at fine scales by the CS strategy. Are such ideas derived for the purpose of simplifying mathematical proofs, actually helpful in a concrete setting?
- *Are there artifacts caused by CS which should be suppressed?* Is any post-processing of CS needed to 'clean up' the reconstructions?
- *What happens if there is noise in the measurements?* Perhaps the framework falls apart if there's any noise in the observations – even just the small errors of floating point representation.

Supposing that such questions do not have devastating answers, there are also natural questions about extending the method by extending the kind of matrices Φ which are in use:

- *More Concrete CS matrices Φ.* It would be useful to have CS-matrices which can be explicitly described.
- *CS matrices Φ which can rapidly applied.* This is connected with the previous question. For applying the linear programming formulation of (1.1) it is very convenient [6] that Ψ and Φ each can be rapidly applied – along with their transposes.

We agree that these are all important questions to answer.

1.3 Experiments...

In this paper we approach the above questions through computational experiments. We consider each question, describe experiments to study it, create synthetic signals, and interpret the results of our experiments.

In Section 2 we consider the application of CS to signals which have only a small number of truly nonzero coefficients, getting empirical results which mirror the theory in [14]. Section 3 considers signals which have all coefficients nonzero, but which still have sparse coefficients as measured by ℓ^p norms, $0 < p \leq 1$. The results in this setting are in some sense 'noisier' than they were in Section 2. To alleviate that, we consider an extension of CS by post-processing to remove the 'noise' in CS reconstructions. Section 4 considers a different attack on 'noise' based on a noise-aware reconstruction method.

In Section 5, we consider an extension of CS mentioned in [14]: a hybrid scheme, using conventional linear sampling and reconstruction for coarse-scale information and compressed sampling on fine scale information. The 'noise' in reconstructions can be dramatically lower for a given number of samples. Section 6 pushes this approach farther, deploying CS in a multiscale fashion. Different scales are segregated and CS is applied separately to each one. Again, the 'noise' can be dramatically lower.

In Section 7, we explore the freedom available in the choice of CS-matrices, describing several different matrix ensembles which seem to give equivalent results. While the approaches in [14] involved random matrices, we have found that several other ensembles work well, including the partial Fourier ensembles of [20, 3, 5].

In a final section we summarize our results.

2 Simple Examples: Controlling the Number of Nonzeros

The first basic question one could ask is: suppose we have an object x_0 with a very high degree of transform sparsity – only k nonzeros out of m coefficients. How big does n have to be for the CS scheme to work well?

To answer this question, we conduct an experiment. We construct a signal with a fixed number k of nonzeros, located in random positions in the transform domain, all of equal amplitude. We then apply the CS framework, while varying n, the number of samples used in the sensing scheme. The CS matrix Φ is constructed so that columns are drawn independently at random from a uniform distribution on the unit sphere $\mathbf{S}^{n-1}$ in Euclidean n-space.

Figure 1, panels (a)-(c) have plots of the ℓ^2 error for increasing values of n, with $m = 1024$, $k = 20, 50$ and 100. In each case we see that as n increases beyond about $2k$, the error starts to drop, and for $n > 4k$ the error is negligible. This indicates that, for objects with only k nonzeros at randomly-chosen sites, something like $n = 3k$ measurements are typically needed. This seems much smaller than the $n = ck\log(m)$ one might expect based on (1.2).

As a more intuitive representation of this phenomenon, we considered the object *Blocks* from the Wavelab package [1]. As Figure 2 shows, the object is piecewise constant, and its Haar wavelet transform has relatively few nonzero coefficients. In fact, *Blocks* has 77 nonzero coefficients for signal length $m = 2048$. Figures 3(a),(b) show the reconstruction results with $n = 340$, and 256 compressed samples, respectively. Clearly, the results are better for larger n, and somewhere about $n = 4k$ the method works well, while for $n = 3k$ the results are somewhat 'noisy'.

At this point, we remark that measuring sparsity by counting nonzeros is inherently limited. While it is a seemingly ideal approach to measuring sparsity, in practice, real signals will not

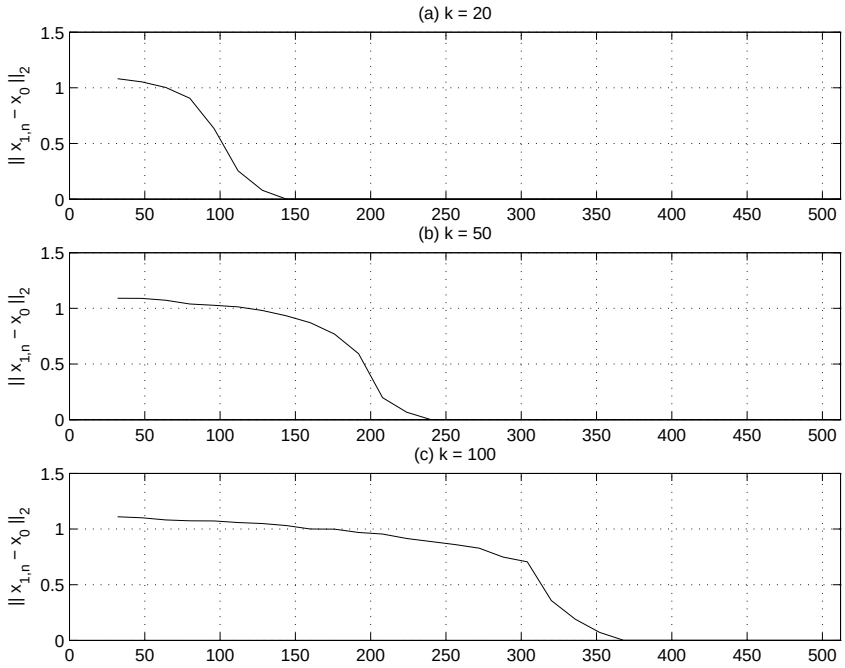

Figure 1: Error of reconstruction versus number of samples for (a) k = 20 nonzeros; (b) k = 50; (c) k = 100.

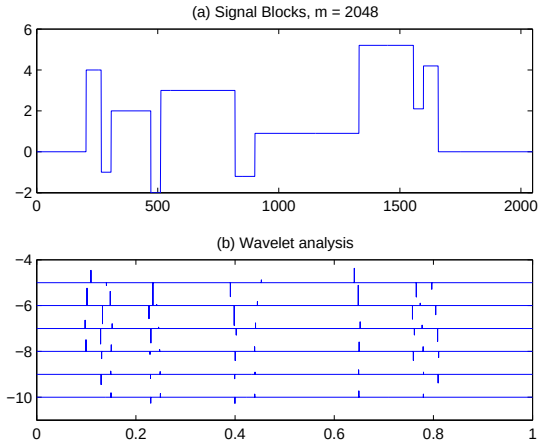

Figure 2: (a) Signal *Blocks*; (b) its expansion in a Haar wavelet basis. There are 77 nonzero coefficients.

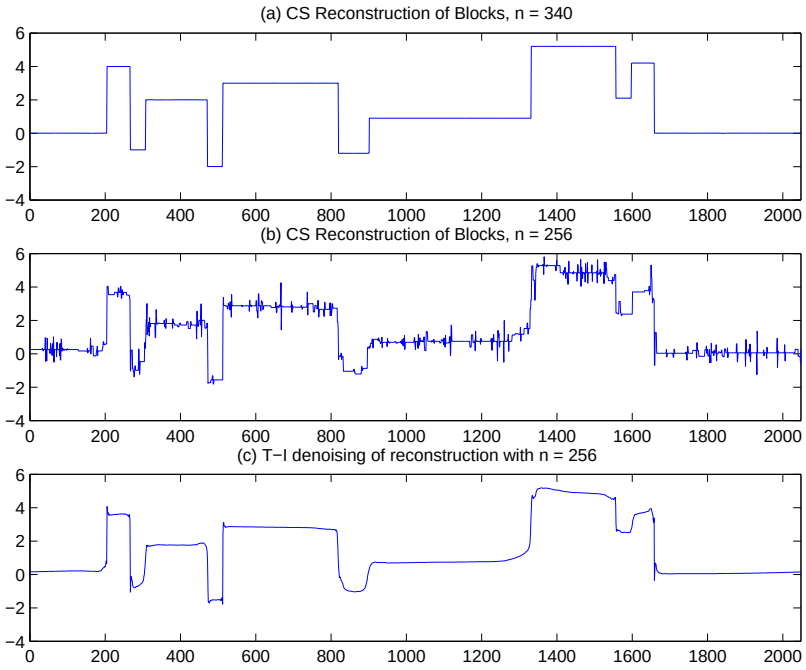

Figure 3: CS reconstructions of *Blocks* from (a) $n = 340$ and (b) $n = 256$ measurements. (c) Translation-invariant denoising of (b).

typically have exact zeros anywhere in the transform. Hence, results of the type just shown, while instructive, are of limited practical interest.

3 More Challenging Examples: Controlling the ℓ^p Norm

Motivated by the results of the previous section, we now consider a range of more challenging examples, in which sparsity, rather than being measured by the number of nonzeros, is measured by the ℓ^p norm, with $p < 1$. There is an intimate connection between ℓ^p norms, $0 < p \leq 1$ and the number of nonzeros (a.k.a. the ℓ^0 norm); for example, as $p \to 0$ the $\|x\|_p^p \to \|x\|_0 \equiv \#\{i : x(i) \neq 0\}$. Measuring sparsity by ℓ^p greatly expands the class of signals we may consider. See [14] and citations there for examples of ℓ^p constraints obeyed for natural classes of signals.

For the purpose of this experiment, we created random signals with all coefficients nonzero, coefficients having random signs, and the coefficients having an ℓ^p norm 1. In our construction, the amplitude, $|\theta|_{(k)}$, of the k-th largest-sized coefficient obeys

$$|\theta|_{(k)} = C_{p,m} \cdot \left(\frac{k}{\log m}\right)^{-1/p}, \quad k = 1, 2, \ldots. \tag{3.1}$$

Figure 4 panel (a) shows such a spiky object, and panel (b) shows the reconstruction. Note that the largest few spikes are well-recovered, but not the many small ones.

As a more intuitive example, we considered the object *Bumps* from the Wavelab package [1], with $m = 2048$. As Figure 5 panel (a) shows, the object is a superposition of smooth bumps. Panel (b) shows that the large coefficients at each scale happen near the bump locations, and panel (c) shows the decreasing rearrangement of the wavelet coefficients on a log scale. We applied the CS framework with $n = 256$ and got the result in panel (a) of Figure 6. Panel

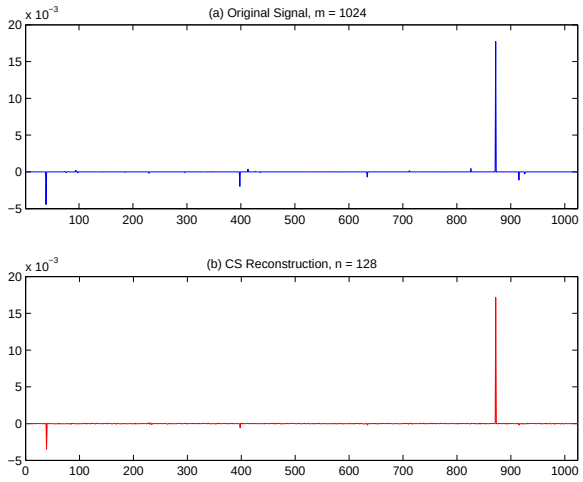

Figure 4: CS reconstruction of a signal with controlled ℓ^p norm, $p = 1/2$. (a) Original signal, $m = 1024$; (b) CS reconstruction with $n = 128$ samples.

(b) shows the result with $n = 512$. Clearly both results are 'noisy', with, understandably, the 'noisier' one coming at the lower sampling rate.

Indeed, the results in Figures 3 and 6 show that 'noise' sometimes appears in reconstructions as the number of measurements decreases (even though the data are not noisy in our examples). To alleviate this phenomenon, we considered the test cases shown earlier, namely *Blocks* and *Bumps*, and applied translation-invariant wavelet de-noising [7] to the reconstructed 'noisy' signals. Results are shown in panel (c) of Figure 3 and panels (b) and (d) of Figure 6. At least visually, there is a great deal of improvement.

4 Noise-Aware Reconstruction

So far we have not allowed for the possibility of measurement noise, digitization errors, etc. We remark that the theory allows for accommodating to a small amount of noise already, through the ℓ^1-stability property proved in [14].

To make further accommodation for noise, our primary adjustment would be to use *Basis Pursuit de-noising* (BPDN) rather than Basis Pursuit. For a given noise level $\epsilon > 0$, define the optimization problem

$$(L_{1,\epsilon}) \qquad \min \|\Psi^T x\|_1 \text{ subject to } \|y_n - \Phi\Psi^T x\|_2 \leq \epsilon.$$

This can be written as a linearly constrained convex quadratic program, and is considered practical to solve. In [6] BPDN was successfully used in cases where $n < m$ and both are quite large: $n = 8192$ and $m = 262144$. Our proposal for dealing with noisy data is simply to measure $y_n = \Phi\Psi^T x + noise$ and then use $(L_{1,\epsilon})$ with an appropriate noise tolerance $\epsilon > 0$.

Extending the theory in [14] suggests that we should expect bounds of the form:

$$\|\hat{x}_{1,\epsilon}(x) - x\|_2 \leq C_1 \cdot \epsilon + C_p \cdot R \cdot (n/\log(m))^{1/2-1/p},$$

where R stands for the ℓ^1 norm of the transform coefficients.

To validate this claim, we performed the following experiment. We took the test signals *Blocks* and *Bumps*, shown in Figures 2 and 5, respectively, and added zero-mean white gaussian

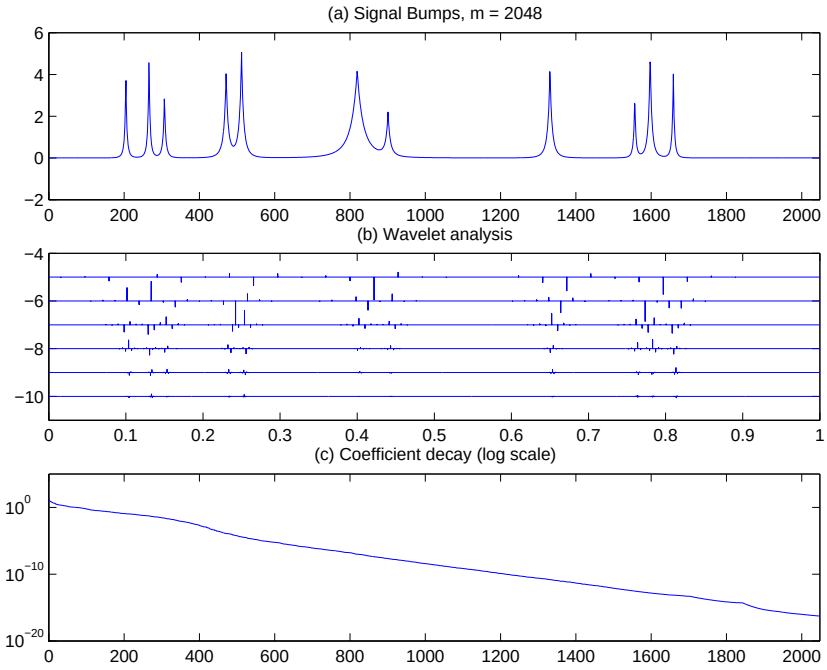

Figure 5: (a) Signal *Bumps*, $m = 2048$; (b) its wavelet coefficients; (c) Decay of coefficient magnitudes on a log scale.

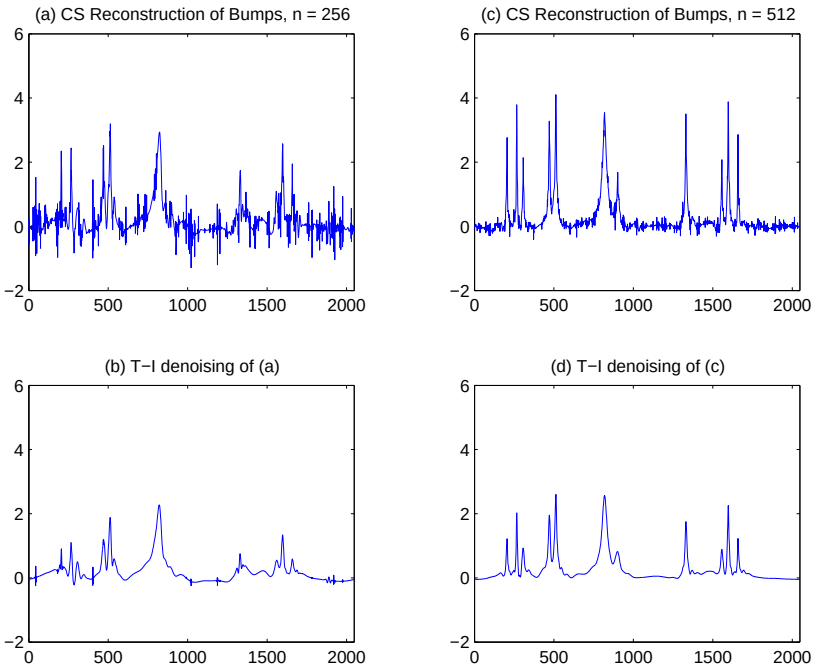

Figure 6: CS reconstruction of *Bumps* from (a) $n = 256$ and (c) $n = 512$ measurements. Translation-invariant denoising in (b) and (d).

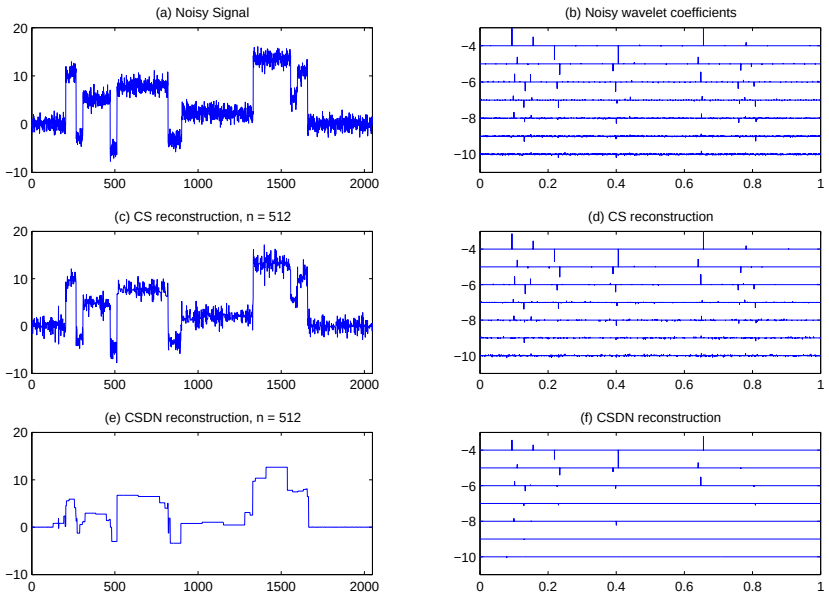

Figure 7: CSDN reconstruction of *Blocks*, $m = 2048, n = 512$. Signal and reconstructions are shown on the left panel, with corresponding wavelet expansions on the right panel.

noise to them. The noise was rescaled to enforce a specific noise level $\epsilon = 0.2$. We applied the compressed sensing scheme with denoising (CSDN) to the noisy wavelet expansions. For comparison, we also attempted regular CS reconstruction. Results are shown in Figures 7 and 8. We used signal length $m = 2048$, and attempted reconstructions with $n = 512$. Indeed, the reconstruction achieved with CSDN is far superior to the CS reconstruction in both cases.

5 Two-Gender Hybrid CS

In [14] model spectroscopy and imaging problems were considered from a theoretical perspective. CS was deployed differently there than so far in this paper – in particular, it was not proposed that CS alone 'carry all the load'. In that deployment, CS was applied to measuring only *fine scale* properties of the signal, while ordinary linear measurement and reconstruction was used to obtain the coarse-scale properties of the signal.

In more detail, the proposal was as follows; we spell out the ideas for dimension 1 only. Expand the object x_0 in the wavelet basis

$$x_0 = \sum_k \beta_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1}\sum_k \alpha_{j,k}\psi_{j,k}$$

where j_0 is some specified coarse scale, j_1 is the finest scale, $\phi_{j_0,k}$ are male wavelets at coarse scale and $\psi_{j,k}$ are fine scale female wavelets. Let $\alpha = (\alpha_{j,k} : j_0 \le j \le j_1, 0 \le k < 2^j)$ denote the grouping together of all wavelet coefficients, and let $\beta = (\beta_{j_0,k} : 0 \le k < 2^{j_0})$ denote the male coefficients. Now consider a scheme where different strategies are used for the two genders α and β. For the male coarse-scale coefficients, we simply take direct measurements

$$\hat{\beta} = (\langle \phi_{j_0,k}, x_0 \rangle : 0 \le k < 2^{j_0}).$$

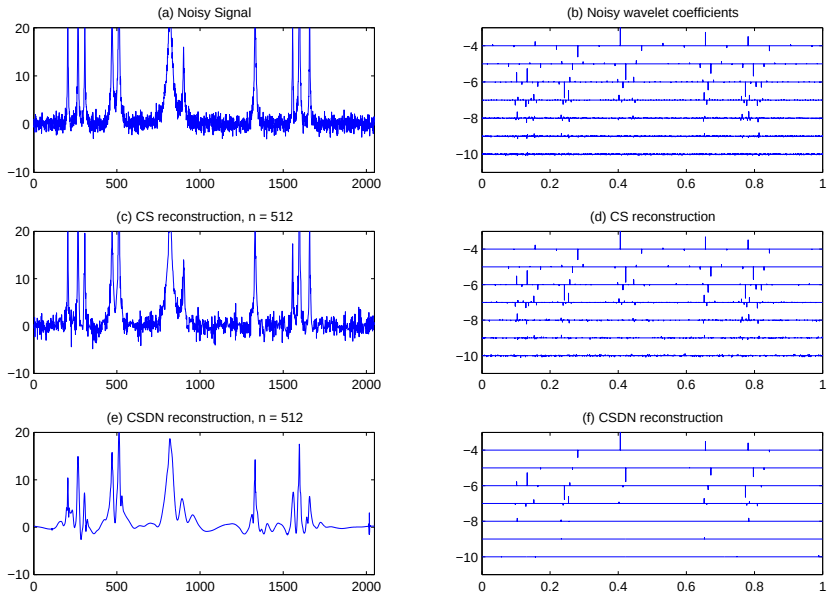

Figure 8: CSDN reconstruction of *Bumps*, $m = 2048, n = 512$. Signal and reconstructions are shown on the left panel, with corresponding wavelet expansions on the right panel.

For the female fine-scale coefficients, we apply the CS scheme. Let $m = 2^{j_1} - 2^{j_0}$, and let the $2^{j_1} \times m$ matrix Ψ have, for columns, the vectors $\psi_{j,k}$ in some standard order. Given an n by m CS matrix Φ, define $\Xi = \Phi\Psi^T$, so that, in some sense, the columns of of Ξ are 'noisy' linear combinations of columns of Ψ. Now make n measurements

$$y = \Xi x_0.$$

To reconstruct from these observations, define $\Omega = \Xi\Psi$ and consider the basis-pursuit optimization problem

$$(BP) \quad \min_a \|a\|_1 \text{ subject to } y_n = \Omega a, \tag{5.1}$$

a minor relabelling of the (L_1) problem. Call the answer $\hat{\alpha}$. The overall reconstruction is

$$\hat{x}_{hy} = \sum_k \hat{\beta}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1}\sum_k \hat{\alpha}_{j,k}\psi_{j,k},$$

a combination of linear reconstruction at coarse scales and nonlinear reconstruction based on undersampling at fine scales. When we speak of this scheme, of course, the total number of samples $n_{hy} = 2^{j_0} + n$ is the number of linear samples plus the number of compressed samples. The theory derived in [14] shows that this hybrid scheme, when applied to objects obeying certain constraints (e.g. bounded variation) gets accuracy comparable to linear sampling at scale 2^{-j_1}, only using many fewer total samples.

The gender-segregated deployment of CS was suggested in [14] for mathematical convenience, but in experiments, it substantially outperforms straightforward gender-blind deployment. To see this, consider Figure 9. Panel (a) shows a blocky signal of original length $m = 16384$ reconstructed from $n = 256$ linear samples, and panel (b) shows reconstruction with $n_{hy} = 208$

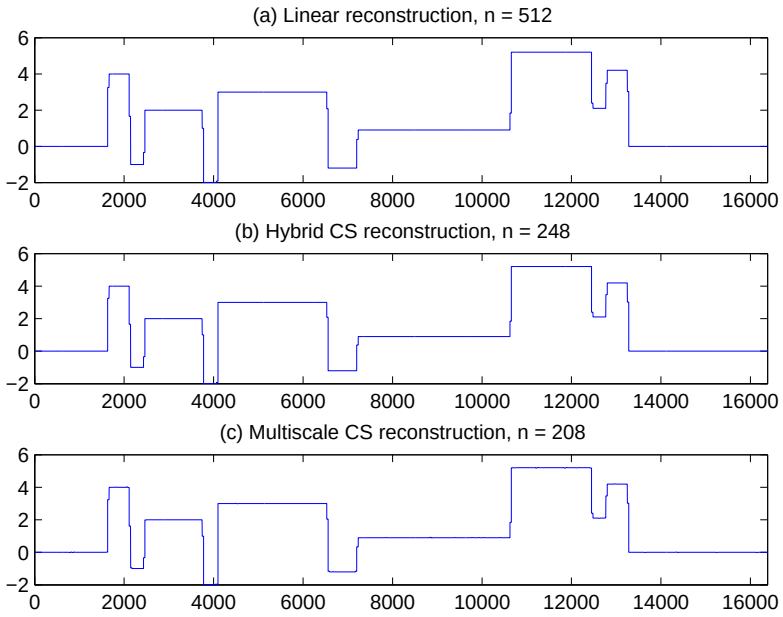

Figure 9: Reconstruction of Signal *Blocks*, $m = 16384$. (a) Linear reconstruction from 512 samples, $\|\hat{x}_{lin} - x_0\|_2 = 0.091$; (b) Hybrid CS reconstruction from 248 samples (Gender-Segregated), $\|\hat{x}_{hy} - x_0\|_2 = 0.091$; (c) Multiscale CS reconstruction from 208 samples, $\|\hat{x}_{ms} - x_0\|_2 = 0.091$.

hybrid compressed samples (32 male samples, 176 compressed female samples). The accuracy is evidently comparable.

Now consider Figure 10. Panel (a) shows a bumpy signal of original length $m = 16384$ reconstructed from $n = 1024$ linear samples, and panel (b) shows reconstruction from $n_{hy} = 640$ hybrid compressed samples (32 male samples, 608 compressed female samples). Again the reconstruction accuracy is comparable.

In [14] the idea of gender-segregated sampling was extended to higher dimensions in considering the class of images of bounded variation. The ideas are a straightforward extension of the 1-D case, and we shall not repeat them here. We investigate the performance of hybrid CS sampling applied to image data in the following experiment. Figure 11 shows a *Mondrian* image of size 1024×1024, so that $m = 2^{20}$. Figure 12 has reconstruction results with $n = 4096$ linear samples, and also with $n_{hy} = 1152$ hybrid compressed samples (128 male samples, 1024 compressed female samples). The reconstruction accuracy is evidently comparable.

6 Multiscale Compressed Sensing

Encouraged by the success of Hybrid CS, we now consider a fully multiscale deployment of CS. The simplest way to explain the concept is to use for our multiscale system a standard 1-D orthogonal wavelet system. However, as we will see, the same idea can be applied with other multiscale systems and in higher dimensions.

Expand the object x_0 in the wavelet basis

$$x_0 = \sum_k \beta_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1}\sum_k \alpha_{j,k}\psi_{j,k}$$

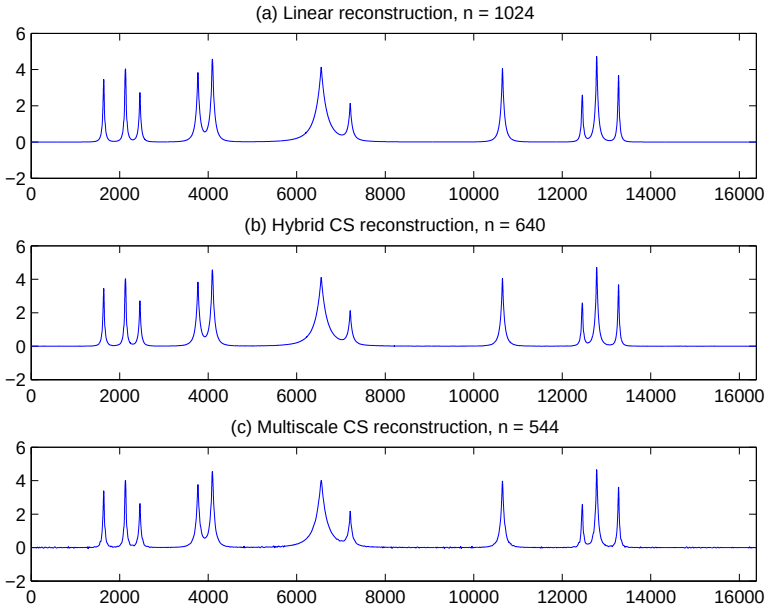

Figure 10: Reconstruction of Signal *Bumps*, $m = 16384$. (a) Linear reconstruction from 1024 samples, $\|\hat{x}_{lin} - x_0\|_2 = 0.0404$; (b) Hybrid CS reconstruction from 640 samples (Gender-Segregated), $\|\hat{x}_{hy} - x_0\|_2 = 0.0411$; (c) Multiscale CS reconstruction from 544 samples, $\|\hat{x}_{ms} - x_0\|_2 = 0.0425$.

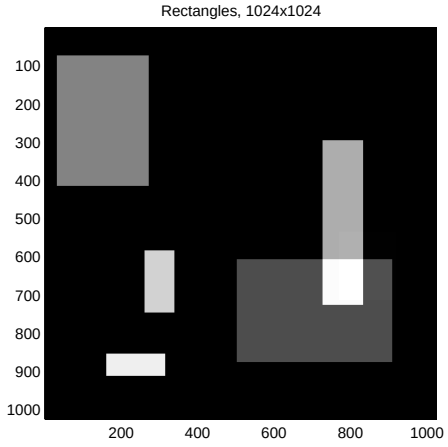

Figure 11: 'Mondrian' image, 1024×1024.

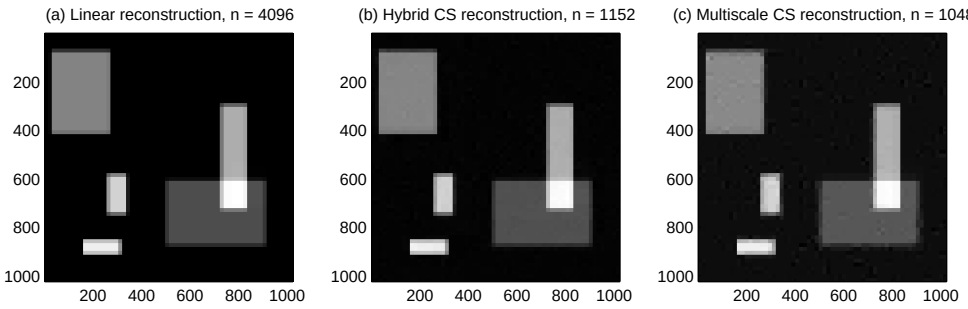

Figure 12: Reconstruction of *Mondrian* image, $m = 2^{20}$. (a) Linear reconstruction from 4096 samples, $\|\hat{x}_{lin} - x_0\|_2 = 0.227$; (b) Hybrid CS reconstruction from 1152 samples (Gender-Segregated), $\|\hat{x}_{hy} - x_0\|_2 = 0.228$; (c) Multiscale CS reconstruction from 1048 samples, $\|\hat{x}_{ms} - x_0\|_2 = 0.236$.

Consider now a multilevel stratification of the object in question, partitioning the coefficient vector as

$$[(\beta_{j_0,\cdot}), (\alpha_{j_0,\cdot}), (\alpha_{j_0+1,\cdot}), \ldots, (\alpha_{j_1-1,\cdot})]$$

We then apply ordinary linear sampling to measure the coefficients $(\beta_{j_0,\cdot})$ directly, and then separately apply compressed sensing scale-by-scale, sampling data y_j about the coefficients $(\alpha_{j,\cdot})$ at level j using an $n_j \times 2^j$ CS matrix Φ_j. We obtain thereby a total of

$$n_{ms} = 2^{j_0} + n_{j_0} + n_{j_0+1} + \cdot + n_{j_1-1}$$

samples, compared to

$$m = 2^{j_0} + 2^{j_0} + 2^{j_0+1} + \cdot + 2^{j_1-1} = 2^{j_1}$$

coefficients in total. To obtain a reconstruction, we then solve the sequence of problems

$$(BP_j) \qquad \min_a \|a\|_1 \text{ subject to } y_j = \Phi_j a, \qquad j = j_0, \ldots, j_1 - 1,$$

calling the obtained solutions $\hat{\alpha}^{(j)}$; to reconstruct, we set

$$\hat{x}_{ms} = \sum_k \beta_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1-1} \sum_k \hat{\alpha}_k^{(j)} \psi_{j,k}.$$

(Of course, variations are possible; we might group together several coarse scales $j_0, j_0 + 1, \ldots, j_0 + \ell$ to get a larger value of m.)

For an example of results obtained using Multiscale CS, consider Figure 9(c). It shows the signal *Blocks* reconstructed from $n_{ms} = 208$ compressed samples (32 coarse-scale samples, 48 compressed samples at each scale). Indeed, the reconstruction accuracy is comparable to the linear and gender-segregated results. Similarly, Figure 10(c) has multiscale reconstruction of *Bumps* from $n_{ms} = 544$ compressed samples (32 coarse-scale samples, 128 compressed samples at each scale). Again the reconstruction accuracy is comparable to that achieved by the other

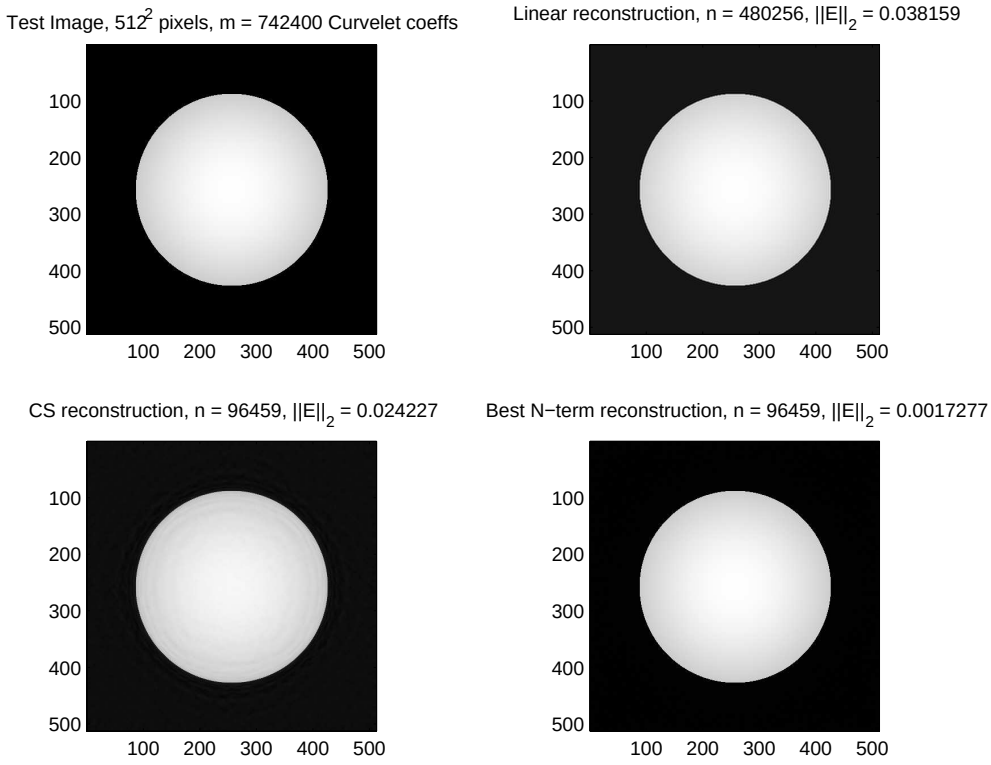

Figure 13: (a) *Disk* image; (b) Reconstruction from 480256 Linear Samples, $\|\hat{x}_{lin} - x_0\|_2 = 0.0381$; (c) Reconstruction from 96459 Multiscale Compressed Samples using the Curvelets Frame, $\|\hat{x}_{ms} - x_0\|_2 = 0.0242$; (d) Reconstruction from best N-term approximation with $N = 96459$, $\|\hat{x}_{N-term} - x_0\|_2 = 0.00173$.

methods. Finally, Figure 12(c) has multiscale CS reconstruction of the *Mondrian* image from 1048 compressed samples (64 coarse-scale samples, 328 compressed samples at each scale).

Consider now an example working with a frame rather than an orthobasis, in this case the Curvelets frame [2]. Theory supporting the possible benefits of using this frame for cartoon-like images was developed in [14].

Like the wavelet basis, there is a scale parameter j which specifies the size of the curvelet frame element; we considered a deployment of multiscale compressed sensing which used different n_j at each level. In Figure 13 we give an example deriving from a very simple black-and-white image depicting a disk. For comparison, we also include the result of best N-term approximation based on (hypothetical) direct measurement of the N most important curvelet coefficients.

To help the reader gain more insight into the level-by-level performance, we compare in Figure 14 the disk coefficients and the reconstructed ones. Clearly, there is additional noise in the reconstruction. Nonetheless, this noise is not evident in the overall CS reconstruction.

7 Different Ensembles of CS Matrices

In the examples so far, we have constructed CS matrices Φ from the uniform spherical ensemble described in [14]. Numerous other possibilities exist for construction of CS matrices. To clarify what we mean, we define four specific ensembles of random matrices

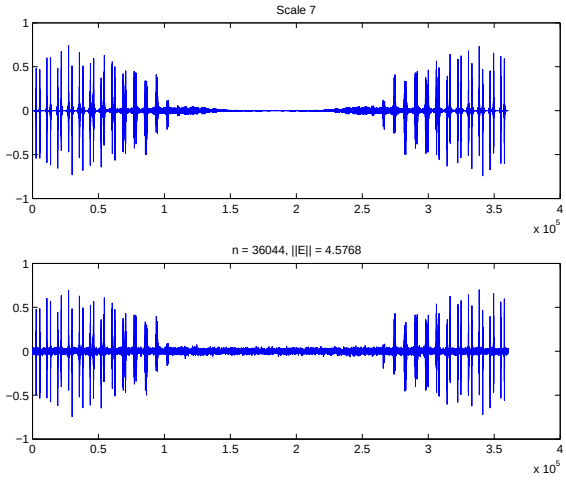

Figure 14: Exact Curvelet coefficients of *Disk* at scale 7 (top), and reconstructed coefficients in CS framework (bottom)

- *Random Signs Ensemble.* Here Φ_{ij} has entries $\pm 1/\sqrt{n}$ with signs chosen independently and both signs equally likely.
- *Uniform Spherical Ensemble.* The columns of Φ are iid random uniform on the sphere S^{n-1}.
- *Partial Fourier Ensemble.* We select at random n rows out of the $m \times m$ Fourier matrix, getting an n-by-m Partial Fourier matrix.
- *Partial Hadamard Ensemble.* We select at random n rows out of the $m \times m$ Hadamard matrix, getting an n-by-m Partial Hadamard matrix. (For his purpose we consider only $m = 2^k$.)

These choices are inspired by the following earlier work:

- The work of Kashin [23], followed by Garnaev and Gluskin [19], implicitly considered the random signs ensemble in the dual problem of Kolmogorov n-widths. Owing to a duality relationship between Gel'fand and Kolmogorov n-widths ([26]), and a relationship between Gel'fand n-widths and compressed sensing [14, 27] these matrices are suitable for use in the case $p = 1$.
- Donoho [12, 13, 14] considered the uniform Spherical ensemble.
- Candès, Romberg, and Tao [3] recently have generated a great deal of excitement by showing several interesting properties of Random Partial Fourier matrices and making claims about their possible use in Compressed Sensing [4].
- Partial Hadamard matrices are known to generate near optimal subspaces in certain special cases for the related problem of determining Kolmogorov n-widths of the octahedron $b_{1,m}$ with respect to ℓ_m^∞ norm; see Pinkus' book [26].

We also remark that there are numerous very interesting practical applications where Partial Fourier and Partial Hadamard matrices are of direct interest, for example in Fourier transform imaging and Hadamard transform spectroscopy.

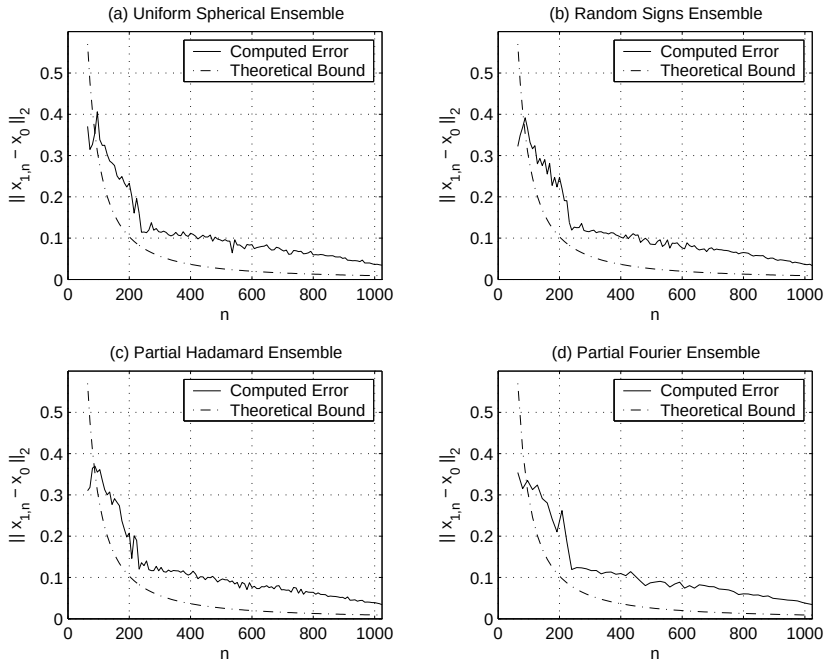

Figure 15: Error versus number of measurements n: (a) Uniform Spherical Ensemble (b) Random Signs Ensemble (c) Partial Hadamard Ensemble; (d) Partial Fourier Ensemble.

In Figure 15, we compare the theoretical predictions in [14] with actual errors in the different matrix ensembles just defined. We do so as follows. For each ensemble, we consider an object defined as in Section 3, i.e. an m-vector with unit p-norm, whose k-th largest amplitude coefficient $|\theta|_{(k)}$ obeys (3.1). A typical example is shown in Figure 4(a). We consider families of experiments where n, the number of measurements, varies, and for each n, we apply the CS framework and measure the ℓ^2 reconstruction error. (In fact, we average over a number of instances). Figure 15 depicts error versus sampling n for the different ensembles, with $p = 1/2$. To validate theoretical predictions, we also plot the error bound (1.2). For the purpose of comparison, we set the bound constant C_p in Eq. (1.2) to $C_{1/2} = 8$. As we can see, the simulation results are in line with the predicted error bound. Further, we observe that the different ensembles show similar behavior. This suggests that all such ensembles are equally good in practice.

8 Conclusions

In this paper, we have reviewed the basic Compressed Sensing framework, in which we gather n pieces of information about an object which, nominally, has m degrees of freedom, $n < m$. When the object is compressible in the sense that the ℓ^p norm of its transform coefficients is well-controlled, then by measuring n essentially 'random' linear functionals of the object and reconstructing by ℓ^1 minimization on the transform coefficients, we get, in theory, an accurate reconstruction; see (1.2).

We raised a number of questions about its application. The first concerns the strength of the theoretical bound (1.2). This bound contains unspecified constants, and so the practical relevance of the framework depends heavily on the precise values of the constants. We conducted a number of numerical experiments to test whether the constants in the basic theoretical inequality are small enough to have practical implications, even at moderate n and m, m in the

low thousands, n in the few hundreds. We concluded that this is so.

We note however, that the method comes off much more impressively when the underlying object has relatively few nonzeros than it does when the object has a large number of small nonzeros (as in the ℓ^p model). In effect, when the number of samples exceeds the number of nonzeros by a factor of 3 or so, the method often performs exceptionally well. Unfortunately, signals made of only a few nonzero terms are rather special.

At the same time, the work we have done is consistent with a value of $C_{1/2} = 10$ or smaller, so the constants in (1.2) do not seem catastrophically large.

We conducted numerical experiments on objects *Blocks* and *Bumps*, caricaturing spectra and scan lines in images; for such objects, whose visual appearance can be gauged, we found that below a certain threshold on n, the CS reconstructions look visually noisy – even though there is no noise in the observations.

Our main effort in this article has been to go beyond these initial observations by considering several extensions of the CS framework. These extensions include,

- *Postprocessing noise removal.* In order to defeat the appearance of visual noise in the CS reconstructions, we applied post-processing to the CS reconstruction by translation invariant de-noising with a level-dependent threshold. We found the appearance of visual noise dramatically reduced (for objects *Blocks* and *Bumps*).

- *Allowing noise in the measurements.* The basic CS framework does not allow for noise in the observations. To handle the case where noise is present, we suggested the use of a noise-tolerant ℓ^1 minimization, essentially using 'Basis-Pursuit Denoising' in place of Basis Pursuit, on which CS is based. We presented examples showing that this can in our examples substantially reduce the effects of noise on the reconstruction.

- *Multiscale Deployment.* More significantly, we considered the use of CS *not* as the only way to gather information about an object, but as a tool to be used in conjunction with classical linear sampling at coarse scales. We considered a hybrid method, already discussed in [14] from a theoretical perspective, in which very coarse-scale linear sampling was combined with CS applied at all finer scales. We also introduced a fully multiscale method in which CS was applied at each scale separately of all others, and linear sampling only at the coarser scale. Both approaches can significantly outperform linear sampling, giving comparable accuracy of reconstruction with far fewer samples in our examples. This tends to support the theoretical claim in [14] that compressed sampling, deployed in a Hybrid fashion, can, for certain kinds of objects, produce an order-of-magnitude improvement over classical linear sampling.

- *Alternate CS Matrices.* Perhaps of most significance is our empirical finding that several apparently different random matrix ensembles all perform similarly when used in the CS framework. Such a finding, if true in general, is important for algorithmic reasons. To apply interior-point methods to solve the linear program (1.1) requires many matrix vector products Φu and $\Phi^T v$ for strategically chosen vectors u and v [6]. It is *much* faster to apply matrices Φ and Φ^T defined by the Partial Fourier and Partial Hadamard ensembles, where the cost will be $O(m \log(m))$ flops. Clearly, for large m and n one would much prefer to use such matrices over the Random Uniform Ensemble or the Random Signs ensemble, which cost $O(nm)$ flops to apply. It appears that we have many choices of matrices yielding adequate performance in the CS framework, consistent with the Theorems in [14], and that among those matrices are some which can be applied rapidly.

In short, it appears that Compressed Sensing is a fairly promising strategy for sampling certain kinds of signals characterized by large numbers of nominal samples and yet high compressibility in terms of transforms like the wavelet transform or Fourier transform. Our practical results already show gains by factors of 2, 4, or even more in moderate-size problems sizes, and these can be enhanced by deployment in a multiscale fashion and by applying de-noising to reconstructions.

Of course the theoretical implications of the bound (1.2) are even stronger than what we observe here, in the sense that at least for n and m large, really dramatic benefits ought to be possible for certain classes of compressible objects. We know that several groups are actively exploring applications of these ideas, first of course being Candès, Romberg, and Tao, who are actively pursuing the implications of [3, 4, 5], which have proved so inspiring. In addition, R.R. Coifman at Yale has expressed interest in experimental work on this topic, and announced his intention to conduct experiments in the CS framework. Clearly we can expect far more experimentation and theoretical development of such ideas in the near future, and the interested reader should find it profitable to pursue all the above-mentioned threads.

References

[1] J. Buckheit and D. L. Donoho (1995) WaveLab and reproducible research, in A. Antoniadis, Editor, *Wavelets and Statistics*, Springer, 1995.

[2] E. J. Candès and DL Donoho (2004) New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure and Applied Mathematics* **LVII** 219-266.

[3] E.J. Candès, J. Romberg and T. Tao. (2004) Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. Manuscript.

[4] Candès, E.J. and Romberg, J. (2004) Presentations at Second International Conference on Computational Harmonic Anaysis, Nashville Tenn. May 2004.

[5] Candès, E.J. and Tao, T. (2004) Estimates for Fourier Minors, with Applications. Manuscript.

[6] Chen, S., Donoho, D.L., and Saunders, M.A. (1999) Atomic Decomposition by Basis Pursuit. *SIAM J. Sci Comp.*, **20**, 1, 33-61.

[7] Coifman, R.R. and Donoho, D.L. (1995) Translation-invariant de-noising. In *Wavelets and Statistics*, Antoniadis, A. and Oppenheim, G. (Eds.), Lect. Notes Statist., 103, pp. 125-150, New York: Springer-Verlag.

[8] R.R. Coifman, Y. Meyer, S. Quake, and M.V. Wickerhauser (1990) Signal Processing and Compression with Wavelet Packets. in *Wavelets and Their Applications*, J.S. Byrnes, J. L. Byrnes, K. A. Hargreaves and K. Berry, eds. 1994,

[9] Donoho, D.L. and Huo, Xiaoming (2001) Uncertainty Principles and Ideal Atomic Decomposition. *IEEE Trans. Info. Thry.* **47** (no.7), Nov. 2001, pp. 2845-62.

[10] Donoho, D.L. and Elad, Michael (2002) Optimally Sparse Representation from Overcomplete Dictionaries via ℓ^1 norm minimization. *Proc. Natl. Acad. Sci. USA* March 4, 2003 **100** 5, 2197-2002.

[11] Donoho, D., Elad, M., and Temlyakov, V. (2004) Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. Submitted. URL: http://www-stat.stanford.edu/~donoho/Reports/2004.

[12] Donoho, D.L. (2004) For most large underdetermined systems of linear equations, the minimal ℓ^1 solution is also the sparsest solution. Manuscript. Submitted. URL: http://www-stat.stanford.edu/~donoho/Reports/2004

[13] Donoho, D.L. (2004) For most underdetermined systems of linear equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution. Manuscript. Submitted. URL: http://www-stat.stanford.edu/~donoho/Reports/2004

[14] Donoho, D.L. (2004) Compressed Sensing. Manuscript. Submitted. URL: http://www-stat.stanford.edu/~donoho/Reports/2004

[15] A. Dvoretsky (1961) Some results on convex bodies and Banach Spaces. *Proc. Symp. on Linear Spaces.* Jerusalem, 123-160.

[16] M. Elad and A.M. Bruckstein (2002) A generalized uncertainty principle and sparse representations in pairs of bases. *IEEE Trans. Info. Thry.* **49** 2558-2567.

[17] J.J. Fuchs (2002) On Sparse Representations in Arbitrary Redundant Bases. *IEEE Trans. Info. Thry* **50** (no.6), June 2004, pp. 1341-44.

[18] T. Figiel, J. Lindenstrauss and V.D. Milman (1977) The dimension of almost-spherical sections of convex bodies. *Acta Math.* **139** 53-94.

[19] Garnaev, A.Y. and Gluskin, E.D. (1984) On widths of the Euclidean Ball. *Soviet Mathematics – Doklady* **30** (in English) 200-203.

[20] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan and M. Strauss, (2002) Near-optimal sparse fourier representations via sampling, in *Proc 34th ACM symposium on Theory of Computing*, pp. 152–161, ACM Press.

[21] R. Gribonval and M. Nielsen. Sparse Representations in Unions of Bases. *IEEE Trans. Info. Thry* **49** (no.12), Dec. 2003, pp. 1320-25.

[22] R. Gribonval and M. Nielsen. Highly Sparse Representations from Dictionaries are Unique and Independent of the Sparseness Measure. Manuscript.

[23] Boris S. Kashin (1977) Diameters of certain finite-dimensional sets in classes of smooth functions. *Izv. Akad. Nauk SSSR, Ser. Mat.* **41** (2) 334-351.

[24] S. Mallat, Z. Zhang, (1993). Matching Pursuits with Time-Frequency Dictionaries. *IEEE Trans. Sig. Proc.*, **41** (no.12), pp. 3397-3415.

[25] B.K. Natarajan (1995) Sparse Approximate Solutions to Linear Systems. *SIAM J. Comput.* **24**: 227-234.

[26] Pinkus, A. (1985) *n-widths in Approximation Theory.* Springer-Verlag.

[27] Pinkus, A. (1986) n-widths and Optimal Recovery in *Approximation Theory*, Proceeding of Symposia in Applied Mathematics, **36**, Carl de Boor, Editor. American Mathematical Society, Providence, RI.

[28] J.A. Tropp (2003) Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Trans Info. Thry.* **50** (no.11), Oct. 2004, pp. 2231-42.

[29] J.A. Tropp (2004) Just Relax: Convex programming methods for Subset Sleection and Sparse Approximation. Manuscript.