RECENT ADVANCES IN SPARSITY-DRIVEN SIGNAL RECOVERY

*David L. Donoho**

Department of Statistics
Stanford University
Stanford, CA 94305, USA

Yaakov Tsaig

Institute for Computational and Mathematical Engineering
Stanford University
Stanford, CA 94305, USA

ABSTRACT

We briefly recall previous literature about the interaction between sparsity and ℓ^1 minimization. We then discuss ℓ^1 minimization in geometry separation, in compressed sensing, and in compressed sensing of separable signals.

1. INTRODUCTION

Sparsity has played a role in signal recovery for decades. Recently, however, the term 'sparsity' is appearing more frequently in titles and abstracts. The role of ℓ^1-based penalization for dealing with sparse signal recovery problems has also been known for years, and is also becoming increasingly prominent. Why is this happening now?

1.1. The Early Days

The connection between sparsity and ℓ^1 has been known as a rough empirical matter for a long time. The first author became aware of this theme in the late 1970's while working in seismic exploration; at that time, "Sparse Spike Train" processing became popular. The idea was that the real earth was irregularly layered, making the underlying reflectivity an irregular series of spikes. The seismic experiment gave information missing low frequencies, creating an ill-posed linear inverse problem. Researchers began to use ℓ^1-penalization in deconvolution; this gave good empirical results when applied to real and simulated signals where the solution was truly sparse. Jon Claerbout and his lab at Stanford did much early work on ℓ^1 methods in the 1970's, e.g. [4]. A paper crisply stating many of the ideas that were in the air appeared in *Geophysics*, the flagship exploration journal in early 1979 [20]. In this paper, Taylor, Banks and McCoy applied l^1 penalized deconvolution with adjustable penalty factors. By varying the penalty factor, spiky reconstructions with varying degrees of spikiness were obtained. The sparsest reconstructions looked surprisingly good, and the ability to 'tune' the sparsity of the solution merely by selecting the penalty factor was quite suggestive.

1.2. Rigorous Results

So ℓ^1 penalization in ill-posed deconvolution problems has been available for some time. In the 1990's theoretical validation of its benefits came available, in a simple statement that is so striking that it perhaps could wake some people up to the potential. Donoho and Logan's result [6] concerned bandlimited deconvolution for functions on the real line $t \in \mathbf{R}$. Suppose that one observes $y = k \star f_0$, with k the highpass kernel 'killing' frequencies below π. Suppose that $f_0(t)$ is sparse in the sense that it vanishes outside a set T which obeys the sparsity condition

$$\sup_{t \in R} |T \cap [t, t+1)|$$

Then the minimum L^1 solution

$$\min \|f\|_1 \text{ subject to } y = k \star f$$

is precisely f_0. Here all low frequencies are entirely missing and yet f_0 can be recovered perfectly! As a traditional linear inverse problem, the problem is wildly ill-posed, and yet the recovery is perfect. Moreover, the solution is stable; if $y = k \star f_0 + z$ with $\|z\|_1 \leq \epsilon$, then the minimum L^1 solution is order $O(\epsilon)$ away from f_0. Here sparsity of f_0 and the L^1 norm produce a truly nonlinear and powerful regularization.

1.3. Abstract Generalizations

The 1990's brought new signal processing problems far more abstract than deconvolution. Consider an overcomplete signal representation, where a signal of interest is composed of terms taken from two different basis sets simultaneously, using only a relatively few terms from each basis. Since one can obviously represent any signal using one basis alone, using two bases gives a system of underdetermined equations with more unknowns (coefficients in the two bases) than equations (signal values to reconstruct). Heuristic proposals to use combined representations were by workers in the computational harmonic analysis community in the early-to-mid 1990s [5, 19, 3]. The first author's work on 'Basis Pursuit' was based on his intuition that the problem of combined representation was in some way analogous to problems of highpass deconvolution, and that the Donoho/Logan result on stable recovery would carry over to the more abstract setting.

1.4. Rigorous Results in the Abstract Setting

Two types of results have been proved concerning the success of ℓ^1 methods.

In the first setting, we let Ω and Ψ be orthonormal matrices representing bases of $\mathbf{R}^n$ and let $\Phi = [\Omega\Psi]$ be the nonsquare n by $2n$ matrix built by sideways concatenation. Given a signal y thought to be a sparse superposition of elements from the two bases, solve

$$\min_x \|x\|_1 \text{ subject to } y = \Phi x.$$

*Partial support from NSF DMS 00-77261, and 01-40698 (FRG) and ONR. Thanks to Michael Saunders, Emmanuel Candès.

If the elements of the cross-products matrix $\Omega^T\Psi$ are all less than M in absolute value, then the minimal ℓ^1 norm solution is the sparsest solution, provided the solution has at most $(M^{-1}+1)/2$ nonzeros; see [7, 14, 8, 15, 17, 21] for this and further results.

In the second setting, much more delicate structure is needed. We have a large underdetermined system of linear equations $y = \Phi x$, with Φ an n by m matrix, and n large, whose columns look 'like noise' (e.g. we can have the entries of Φ be i.i.d. Gaussians). If there is a solution $y = \Phi x_0$ with x_0 sufficiently sparse, the sparsest solution is also the ℓ^1 minimizer; here sufficiently sparse means that the number of nonzeros $< \rho n$ [10, 2]. Also, when noise is present, solving

$$\min_x \|x\|_1 \text{ subject to } \|y - \Phi x\|_2 \le \epsilon$$

gives an approximation to x_0 again provided that the underlying signal is sufficiently sparse [11]. Similar but slightly weaker results are possible if Φ is a randomly-selected set of n rows from an $m \times m$ Fourier matrix; see [1, 16]

In the remainder of this brief note, we give three quick examples of recent results in this direction.

2. GEOMETRIC SEPARATION

One original motivation for studying overcomplete representations was to separate pointlike from linelike and curvelike features in an image; e.g. see Xiaoming Huo's Stanford thesis for an early attempt [18].

Jean-Luc Starck has recently had success in solving geometric separation problems applying the ℓ^1 norm using overcomplete representations. His implementation approximately solves

$$\min_{u,v} \|Wu\|_1 + \|Cv\|_2 + \lambda\|y - (u+v)\|_2^2$$

Here u is supposed to be the part of the image sparsified by the wavelet transform W, v is supposed to be the part sparsified by the curvelet transform C, and λ is a noise tolerance parameter.

Recently, theoretical results became available validating this approach [13]. If the underlying 'image' $f(x, y)$ is a generalized function on the continuum domain $f = \pi + \gamma$ where π is a superposition of point singularities and γ is a curvilinear singularity, it has been shown that a multiscale version of the above problem, with P_j the projection on wavelet scale j and W_j and C_j wavelet and curvelet transforms at scale j

$$\min_{u,v} \|W_j u\|_1 + \|C_j v\|_1 \text{ subject to } P_j f = u + v,$$

satisfies $u \approx P_j\pi$, $v \approx P_j\gamma$; i.e. the two components u and v carry the two different geometric types. Here the sense of approximation improves as the scale refines. An interesting point: the notion of sparsity is not simply the number of nonzero coefficients, but involves also the arrangement of the nonzeros (wavelet coefficients clustering near points and curvelet coefficients clustering near curves, respectively).

3. COMPRESSED SENSING AND EXTENSIONS

In [1, 12, 2], the idea of making reduced numbers of measurements about a compressible object was proposed. If f_0 is an unknown vector in R^m which is compressible in basis Ψ, we take a random n by m matrix Φ, getting n *compressed sensing* (CS) measurements $y = \Phi\Psi^T f$, with $n < m$. To reconstruct, we solve the ℓ^1 problem

$$\min_f \|\Psi^T f\|_1 \text{ subject to } y = \Phi\Psi^T f.$$

The paper [23] extended compressed sensing to accomodate for multiscale phenomena. In multiscale CS, a small number of samples is allocated at each scale, and compressed sensing is applied at that scale. At the reconstruction stage, at each scale j an ℓ^1 minimization problem is solved. See [23] for details. Figure 1 shows the result of such a multiscale CS scheme, deployed using a Curvelet frame. Panel (a) displays the well-known Shepp-Logan phantom, discretized on a 512×512 grid. Panel (b) shows the result of reconstructing from $n = 480256$ linear measurements of Curvelet coefficients. Panel (c) has the result of reconstructing from $n = 103218$ multiscale compressed samples. We observe that a decrease by a factor of 5 in the number of samples gives comparable results in terms of measured ℓ^2 error. The point here is that the Curvelet system is highly overcomplete, but this approach allows us to record and use many fewer coefficients than the nominal number of Curvelet elements.

4. CS FOR SEPARABLE SIGNALS

Combing the last two sections, we can apply compressed sensing to signals which are not sparse in a single basis, but instead sparse in a combined representation. Numerical experiments have verified this idea. For n, p given, we generate an $n \times p$ matrix Φ, with columns iid uniform on $\mathbf{S}^{n-1}$. We create a matrix Θ by concatenating the identity and hadamard matrices, i,.e. $\Theta = [IH]$. The resulting matrix Θ is of dimensions $p \times m$, with $m = 2p$. Hence $f = \Theta\alpha$ generates a vector $f \in \mathbf{R}^p$ which is made up possibly from terms in two different bases. In our experiment, we select a sparse coefficient vector α_0 consisting of k nonzeros, so that $f_0 = \Theta\alpha_0$ is the underlying signal vector; we apply compressed sensing: $y = \Phi f_0 = y = \Phi\Theta\alpha_0$. We then solve

$$\min \|\alpha\|_1 \text{ subject to } \Phi\Theta\alpha = y, \tag{4.1}$$

for α. If the ℓ^1 solution $\alpha_1 = \alpha_0$ to machine precision, we say the method allows 'perfect' reconstruction of $f_0 = \Theta\alpha_0$.

Figures 2,3 give results from performing this experiment numerous times for different values of k, n, m; we calculated the percentage of perfect reconstructions. Panels (a),(b) of Figure 2 displays this percentage versus k, for $n = 500$, $m = 1024$ and 1536, respectively. Panels (a),(b) of Figure 3 have plots of the ℓ^2 error in the reconstruction, for $n = 500$, $m = 1024$ and 1536, respectively.

In [24], a heuristic was proposed to predict the breakdown point of local equivalence between the solutions of the ℓ^1 problem and the ℓ^0 problem. At the core of this heuristic is measurement of a function $\nu_0(A)$, followed by some simple algebra. To study this, we conducted the following experiment. For n, m given, we drew a dictionary $\Phi_{n,m}$ at random from a uniform ensemble, and computed $V_{n,m}$, the value of (4.1), with y uniformly random on $\mathbf{S}^{n-1}$. We repeated this experiment numerous times at each specific (n, m) pair, taking the median at each instance. To our empirical results we fitted a decaying power law of the form $\nu_0(A) = C \cdot A^{-\gamma}$. The results are illustrated in Figure 4. In our experiment, a least-squares line fit on the logarithmic scale resulted in the estimate

$$\nu_0(A) \approx 2 \cdot A^{-0.586}.$$

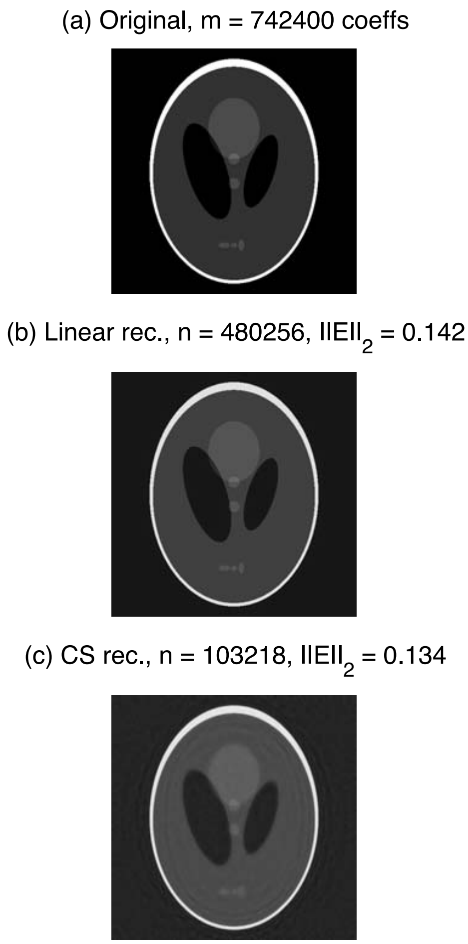

Fig. 1. (a) Shepp-Logan phantom, discretized on a 512×512 grid; (b) Reconstruction from $n = 480256$ linear measurements, $\|\hat{x}_{lin} - x_0\|_2 = 0.142$; (c) Reconstruction from $n = 103218$ Multiscale Compressed Samples using the Curvelets Frame, $\|\hat{x}_{ms} - x_0\|_2 = 0.134$.

With this estimate for $\nu_0(A)$, we may readily apply the heuristic described in [24] to predict the breakdown of local equivalence. Figure 5 compares the heuristic prediction with the empirical breakdown point versus A. Panels (a),(b) show plots for $n = 300$ and 500. We see that the heuristic gives a fair estimate for the empirical behavior of the ℓ^1 solution. This suggests an interesting possibility: that we can perform compressed sensing of geometrically-separable objects, not needing to separate before compression.

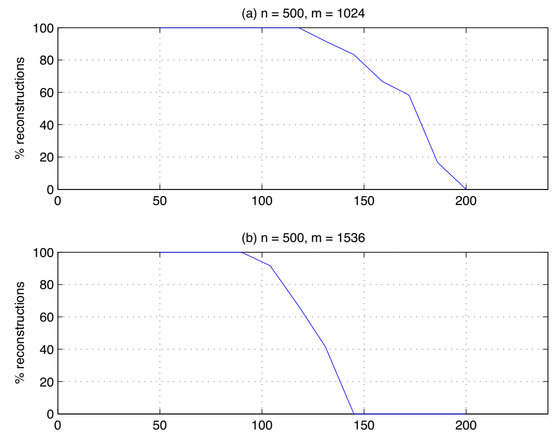

Fig. 2. Compressed Sensing from an Overcomplete System. Panels (a),(b) show percentage of perfect reconstructions versus number of nonzeros for $n = 500$, $m = 1024$ and 1536, respectively.

5. REFERENCES

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

[2] Candes, EJ. (2004) Presentation at IPAM Workshop, UCLA, Sept. 2004.

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

[4] Jon F. Claerbout and Francis Muir (1973), Robust modeling with erratic data, *Geophysics*, **38**, 826-844.

[5] 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,

[6] Donoho, D. L. and Logan, B. F. (1992) Signal recovery and the large sieve, *SIAM J. Appl. Math.*, **52**, 577–591.

[7] 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.

[8] 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.

[9] Donoho, D., Elad, M., and Temlyakov, V. (2004) Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. Submitted.

[10] Donoho, D.L. (2004) For most large underdetermined systems of linear equations, the minimal ℓ^1 solution is also the sparsest solution. Manuscript.

[11] Donoho, D.L. (2004) For most underdetermined systems of linear equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution. Manuscript.

[12] Donoho, D.L. (2004) Compressed Sensing. Manuscript.

[13] Donoho, D.L. (2004) ‘New Uncertainty Principles and Geometric Separation’. Presentation at Second International Conference of Computational Harmonic Analysis, Nashville, May 2004.

[14] 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.

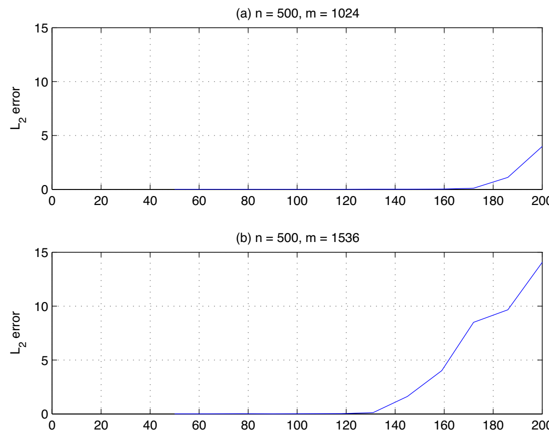

Fig. 3. Panels (a),(b) show the behavior of the ℓ^2 error versus the number of nonzeros for $n = 500$, $m = 1024$ and 1536, respectively.

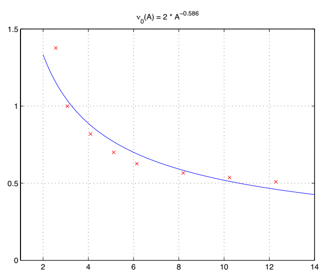

Fig. 4. Empirical Estimate of the limit function $\nu_0(A)$ for the two-bases case.

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

[16] 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.

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

[18] Huo, Xiaoming (1999) Sparse Image Representation by Combined Transforms. Ph.D. Thesis, Stanford

[19] S. Mallat, Z. Zhang, (1993). “Matching Pursuits with Time-Frequency Dictionaries,” IEEE Transactions on Signal Processing, 41(12):3397–3415.

[20] Howard L. Taylor and Stephen C. Banks and John F. McCoy (1979) Deconvolution with the ℓ^1 norm, *Geophysics*, volume = **44**, number = 1, pages = 39-52.

[21] J.A. Tropp (2003) Greed is Good: Algorithmic Results for Sparse Approximation To appear, *IEEE Trans Info. Thry.*

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

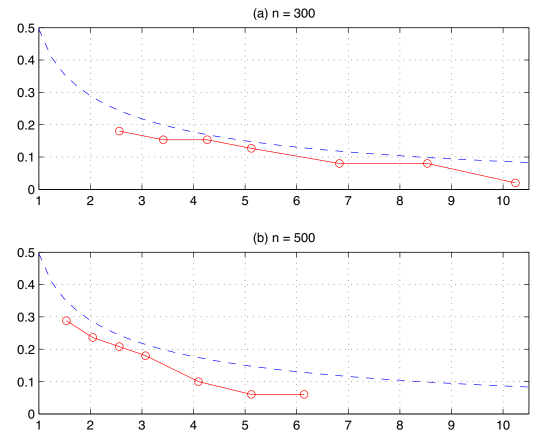

Fig. 5. Heuristic vs. empirical breakdown of local equivalence. Panels (a),(b) show the heuristic breakdown point (dashed line) and the empirical breakdown point as a function of A, for $n = 300$ and 500, respectively.

[23] Y. Tsaig and D.L. Donoho, “Extensions of compressed sensing,” *Submitted*, 2004,

[24] Y. Tsaig and D.L. Donoho, “Breakdown of equivalence between the minimal ℓ^1-norm solution and the sparsest solution,” *Submitted*, 2004,