STABLE SPARSE APPROXIMATIONS VIA NONCONVEX OPTIMIZATION

Rayan Saab *, *Rick Chartrand* †, *Özgür Yılmaz* ‡

ABSTRACT

We present theoretical results pertaining to the ability of ℓ_p minimization to recover sparse and compressible signals from incomplete and noisy measurements. In particular, we extend the results of Candès, Romberg and Tao [1] to the $p < 1$ case. Our results indicate that depending on the restricted isometry constants (see, e.g., [2] and [3]) and the noise level, ℓ_p minimization with certain values of $p < 1$ provides better theoretical guarantees in terms of stability and robustness than ℓ_1 minimization does. This is especially true when the restricted isometry constants are relatively large.

***Index Terms*—** Compressed Sensing, Compressive Sampling, ℓ_p minimization, Sparse Recovery

1. INTRODUCTION

The problem of recovering sparse signals in $\mathbb{R}^N$ from $M < N$ measurements has received a lot of attention lately, especially with the advent of compressive sensing and related applications, e.g., [4, 5, 1, 6]. More precisely, let $x \in \mathbb{R}^N$ be a sparse vector, let $\mathbf{A} \in \mathbb{R}^{M \times N}$ be a *measurement matrix*, and suppose the possibly noisy *observation* vector $\mathbf{b}$ is given by

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}. \tag{1}$$

where $\mathbf{e} \in \mathbb{R}^M$ denotes the noise. The goal of a sparse recovery algorithm is to obtain an estimate of $\mathbf{x}$ given only $\mathbf{b}$ and $\mathbf{A}$. This problem is non-trivial since $\mathbf{A}$ is overcomplete. That is, the linear system of M equations in (1) is underdetermined, and thus admits infinitely many solutions among which the correct one must be chosen. As the original signal $\mathbf{x}$ is sparse, the problem of finding the desired solution can be phrased as some optimization problem where the objective is to maximize an appropriate measure of sparsity while simultaneously satisfying the constraints defined by (1). As the sparsity of $\mathbf{x}$ is reflected by the number of its non-zero entries, equivalently its so-called ℓ_0 norm, in the noise-free case of (1) one would seek to solve the P_0 problem, e.g., [7],

$$P_0 : \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{b} = \mathbf{A}\mathbf{x}. \tag{2}$$

It can be shown that P_0 recovers $\mathbf{x}$ exactly if $\mathbf{x}$ is sufficiently sparse depending on the matrix $\mathbf{A}$ [7]. However, this optimization problem is combinatorial in nature, thus its complexity grows extremely quickly as N becomes much larger than M. Naturally, one then seeks to modify the optimization problem so that it lends itself to solution methods that are more tractable than combinatorial search. In fact, it has been shown (e.g., [2, 1]) that, in the noise-free setting, ℓ_1 minimization, i.e., solving

$$P_1 : \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{b} = \mathbf{A}\mathbf{x}, \tag{3}$$

recovers $\mathbf{x}$ exactly if $\|\mathbf{x}\|_0 \leq S$ and the matrix $\mathbf{A}$ obeys a particular restricted isometry property, e.g., $\delta_{2S} + 2\delta_{3S} < 1$. Here δ_S are the S-restricted isometry constants of $\mathbf{A}$, defined as the smallest constants satisfying

$$(1 - \delta_S)\|\mathbf{c}\|_2^2 \leq \|\mathbf{A}_T\mathbf{c}\|_2^2 \leq (1 + \delta_S)\|\mathbf{c}\|_2^2 \tag{4}$$

for all subsets of columns T with $\#(T) \leq S$ and any vector $\mathbf{c}$. In the general setting, [1] provides error guarantees when the underlying vector is not "exactly" sparse and when the observation is noisy.

Theorem 1 *[1] Assume that $\mathbf{x}$ is arbitrary, $\mathbf{b} = \mathbf{A}\mathbf{x}$ and suppose that $\delta_{3S} + 3\delta_{4S} < 2$. Then the solution $\mathbf{x}^*$ to P_1^ϵ (see (9)) obeys*

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_S\epsilon. \tag{5}$$

For reasonable values of δ_{4S}, the constant C is well behaved; e.g. $C = 8.82$ for $\delta_{4S} = 1/5$.

This means that given $\mathbf{b}$, solving P_1^ϵ recovers the underlying sparse signal within the noise level (thus, perfectly if $\epsilon = 0$).

Theorem 2 *[1] Assume that $\mathbf{x}$ is arbitrary, $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}$, $\|\mathbf{e}\|_2 < \epsilon$ and suppose that $\delta_{3S} + 3\delta_{4S} < 2$. Then the solution $\mathbf{x}^*$ to P_1^ϵ, where*

$$P_1^\epsilon\text{: } \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon, \tag{6}$$

obeys

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_{1,S}\epsilon + C_{2,S}\frac{\|\mathbf{x} - \mathbf{x}_S\|_1}{\sqrt{S}}. \tag{7}$$

For reasonable values of δ_{4S}, the constants are well behaved; e.g. $C_{1,S} = 8.77$ and $C_{2,S} = 12.04$ for $\delta_{4S} = 1/5$.

More recently, [8] showed that in the noise-free setting, a sufficiently sparse signal can be recovered perfectly via ℓ_p minimization, $0 < p < 1$, under less restrictive isometry conditions than those needed for ℓ_1 minimization.

Theorem 3 *[8] Let $0 < p \leq 1$. Assume that $\mathbf{x}$ is S-sparse, $\mathbf{b} = \mathbf{A}\mathbf{x}$ and suppose that $\delta_{kS} + k^{\frac{2-p}{p}}\delta_{(k+1)S} < k^{\frac{2-p}{p}} - 1$, for some $k > 1$. Then the minimizer $\mathbf{x}^*$ to P_p, where*

$$P_p\text{: } \min_{\mathbf{x}} \|\mathbf{x}\|_p \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{8}$$

is exactly $\mathbf{x}$.

Note that, for example, when $p = 0.5$ and $k = 3$, the above theorem only requires $\delta_{3S} + 27\delta_{4S} < 26$ to guarantee perfect reconstruction with $\ell_{0.5}$ minimization, a less restrictive condition than the one needed to guarantee perfect reconstruction by ℓ_1 minimization.

In what follows, we present generalizations of the above theorems, giving stability and robustness guarantees for ℓ_p minimization.

*Rayan Saab is with the Department of Electrical and Computer Engineering, The University of British Columbia. He was partially supported by an NSERC Discovery Grant and an NSERC CRD Grant

†Rick Chartrand is with the Theoretical Division of Los Alamos National Laboratory.

‡Özgür Yılmaz is with the Department of Mathematics, The University of British Columbia. He was partially supported by an NSERC Discovery Grant and an NSERC CRD Grant.

These are of the same nature as those provided above for ℓ_1 minimization in the general (noisy and non-sparse) setting while being less restrictive. We also present simulation results, further illustrating the possible benefits of using ℓ_p minimization.

2. STABLE RECOVERY IN THE PRESENCE OF NOISE WITH ℓ_P MINIMIZATION

In this section, we present our main theoretical results pertaining to the ability of ℓ_p minimization to recover sparse and compressible signals in presence of noise. To that end, we define

$$P_p^\epsilon : \min_{\mathbf{x}} \|\mathbf{x}\|_p^p \text{ subject to } \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon. \tag{9}$$

Theorem 4 (Sparse Case) *Assume that* $\mathbf{x}$ *is* S*-sparse and suppose that for some* $k > 1$, $kS \in \mathbb{Z}^+$

$$\delta_{kS} + k^{\frac{2}{p}-1}\delta_{(k+1)S} < k^{\frac{2}{p}-1} - 1. \tag{10}$$

Let $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}$ *where* $\|\mathbf{e}\|_2 \leq \epsilon$. *Then the minimizer* $\mathbf{x}^*$ *of* P_p^ϵ *obeys*

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_{S,k,p}\epsilon, \quad \textit{where}$$

$$C_{S,k,p} = \frac{2\sqrt{1 + \frac{1}{k^{2/p-1}(2/p-1)}}}{\left((1-\delta_{(k+1)S})^{p/2} - (1+\delta_{kS})^{p/2}k^{p/2-1}\right)^{1/p}}.$$

In summary, the theorem states that if (10) is satisfied then we can recover S-sparse signals stably by solving P_p^ϵ. Note that by setting $p = 1$ and $k = 3$ in Theorem 4, we obtain Theorem 1 as a special case. By setting $\epsilon = 0$, i.e.,assuming no noise, we obtain Theorem 3 as a corollary. An important question that arises next is how well one can recover a signal that is "just nearly sparse" [9]. In this context, let $\mathbf{x}$ be arbitrary and let $\mathbf{x}_S$ be the vector obtained by retaining the S coefficients of $\mathbf{x}$ with the highest magnitudes and setting the rest to zero.

Theorem 5 (General Case) *Assume that* $\mathbf{x}$ *is arbitrary and suppose that (10) holds for some* $k > 1$, $kS \in \mathbb{Z}^+$. *Then the solution* $\mathbf{x}^*$ *to* P_p^ϵ *obeys*

$$\|\mathbf{x}^* - \mathbf{x}\|_2^p \leq C_{S,k,p}^{(1)}\epsilon^p + C_{S,k,p}^{(2)}\frac{\|\mathbf{x} - \mathbf{x}_S\|_p^p}{S^{1-p/2}}, \quad \textit{where}$$

$$C_{S,k,p}^{(1)} = 2^p\frac{1 + k^{p/2-1}(2/p-1)^{-p/2}}{(1-\delta_{(k+1)S})^{p/2} - (1+\delta_{kS})^{p/2}k^{p/2-1}}, \quad \textit{and}$$

$$C_{S,k,p}^{(2)} = \frac{2(\frac{p}{2-p})^{p/2}}{k^{1-p/2}}\left[1 + \frac{(1+k^{p/2-1})(1+\delta_{kS})^{p/2}}{(1-\delta_{(k+1)S})^{p/2} - \frac{(1+\delta_{kS})^{p/2}}{k^{1-p/2}}}\right].$$

Thus, the reconstruction error (to the p^{th} power) is bounded by the sum of two terms; the first is proportional to the observation error, while the second is proportional to the best S-term approximation error of the signal. Note here that by setting $p = 1$ and $k = 3$ in Theorem 5, we obtain Theorem 2, with precisely the same constants.

Remarks

In Theorems 4 and 5, we provide sufficient conditions for recoverability of sparse or compressible signals from noisy and incomplete measurements via ℓ^p minimization where $0 < p < 1$. The constants $C_{S,k,p}$ and $C_{S,k,p}^{(i)}$ determine upper bounds on the recovery error in the sparse and general settings, respectively. These constants depend on S, which reflects the sparsity or the degree of compressibility of the signal to be recovered, on p, determined by the recovery algorithm we use, and on k, which is a free parameter provided (10) holds. Our actual goal is to obtain the smallest possible constants for given S and p, which can be done by finding the value of k that minimizes the corresponding constant in each case. In summary, given S and p, we can replace $C_{S,k,p}$ and $C_{S,k,p}^{(i)}$ with $C_{S,k^*,p}$ and $C_{S,k*,p}^{(i)}$ where $k^*(S,p)$ minimizes the constants in each case.

An extensive analysis of this last minimization step depends on the behavior of the restricted isometry constants of the matrix $\mathbf{A}$ and is beyond the scope of this paper. In Section 4 we present empirical behavior of these constants when A is a Gaussian matrix.

Finally, note that (10) is less restrictive for smaller values of p. For example, when S is large so that (10) does not hold for $p = 1$, there may still exist some $p < 1$ for which (10) holds for some k.

3. PROOF OUTLINES

Due to lack of space, we only present the outlines of the proofs, which mainly follow those of [1] with modifications to account for the fact $p < 1$.

Proof outline for Theorem 4. Let $\mathbf{x}$ be the original signal with its S nonzero coefficients supported on T_0 and let $\mathbf{x}^*$ the solution to P_p^ϵ. Let $\mathbf{h} = \mathbf{x}^* - \mathbf{x} = \mathbf{h}_{T_0} + \mathbf{h}_{T_0^c}$ be the difference between the original and recovered signal, divided into two parts $\mathbf{h_{T_0}}$ with nonzero coefficients on T_0 and $\mathbf{h_{T_0^c}}$ similarly supported on T_0^c. It can easily be shown that $\|\mathbf{h}_{T_0^c}\|_p^p \leq \|\mathbf{h}_{T_0}\|_p^p$.

Divide T_0^c into sets $T_1, T_2, ...$ such that $\cup_{i\geq 1} T_i = T_0^c$, where T_1 supports the kS largest coefficients of $\mathbf{h}_{T_0^c}$, T_2 supports the *second* kS largest coefficients of $\mathbf{h}_{T_0^c}$, and so on. Let $T_{01} = T_0 \cup T_1$. Note that $\mathbf{A}\mathbf{h} = \mathbf{A}_{T_{01}}\mathbf{h}_{T_{01}} + \sum_{i\geq 2}\mathbf{A}_{T_i}\mathbf{h}_{T_i}$. Since both $\mathbf{x}$ and $\mathbf{x}^*$ are feasible then $\|\mathbf{A}\mathbf{h}\|_2 \leq 2\epsilon$. This leads to the following inequality

$$(2\epsilon)^p \geq \|\mathbf{A}\mathbf{h}\|_2^p \geq \|\mathbf{A}_{T_{01}}\mathbf{h}_{T_{01}}\|_2^p - \sum_{i\geq 2}\|\mathbf{A}_{T_i}\mathbf{h}_{T_i}\|_2^p. \tag{11}$$

Since $\#(T_{01}) = (k+1)S$ and $\#(T_i) = kS$, then

$$(2\epsilon)^p \geq \left(1 - \delta_{(k+1)S}\right)^{p/2}\|\mathbf{h}_{T_{01}}\|_2^p - \left(1 + \delta_{(k)S}\right)^{p/2}\sum_{i\geq 2}\|\mathbf{h}_{T_i}\|_2^p. \tag{12}$$

What remains now is to bound $\sum_{i\geq 2}\|\mathbf{h}_{T_i}\|_2^p$ and $\|\mathbf{h}_{T_{01}}\|_2^p$ in terms of $\|\mathbf{h}\|_2$. Observe that $|\mathbf{h}_{T_0^c}|_{(l)}^p \leq \frac{\sum_i |\mathbf{h}_{T_0^c}|_{(i)}^p}{l} = \frac{\|\mathbf{h}_{T_0^c}\|_p^p}{l}$, where $|\mathbf{h}_{T_0^c}|_{(l)}$ is the l^{th} largest element of $|\mathbf{h}_{T_0^c}|$. Thus, taking the p^{th} root, squaring, and summing over $l \in T_{01}^c$ we get

$$\|\mathbf{h}_{T_{01}^c}\|_2^2 \leq \frac{\|\mathbf{h}_{T_0^c}\|_p^2}{\frac{2-p}{p}(kS)^{2/p-1}} \leq \frac{\|\mathbf{h}_{T_0}\|_p^2}{\frac{2-p}{p}(kS)^{2/p-1}} \tag{13}$$

Now, note that $|\mathbf{h}_{T_{i+1}(u)}|^p \leq \sum_{t\in T_i}|\mathbf{h}_{T_i(t)}|^p/(kS) = \|\mathbf{h}_{T_j}\|_p^p/(kS)$. Taking the p^{th} root, squaring, and summing over $u \in T_{i+1}$, we get $\|\mathbf{h}_{T_{i+1}}\|_2^2 \leq (kS)^{1-2/p}\|\mathbf{h}_{T_i}\|_p^2$. Thus,

$$\sum_{i\geq 2}\|\mathbf{h}_{T_i}\|_2^p \leq (kS)^{p/2-1}\|\mathbf{h}_{T_0}\|_p^p. \tag{14}$$

Noting that $\|\mathbf{h}_{T_0}\|_p \leq S^{1/p-1/2}\|\mathbf{h}_{T_0}\|_2$, so $\|\mathbf{h}_{T_0}\|_p^p \leq S^{1-p/2}\|\mathbf{h}_{T_{01}}\|_2^p$

we can now substitute in (12) to get

$$(2\epsilon)^p \geq \left(1-\delta_{(k+1)S}\right)^{p/2} \|\mathbf{h}_{T_{01}}\|_2^p - \left(1+\delta_{kS}\right)^{p/2} \frac{\|\mathbf{h}_{T_{01}}\|_2^p}{k^{1-p/2}}. \quad (15)$$

Using (13),

$$\|\mathbf{h}\|_2^2 = \|\mathbf{h}_{T_{01}}\|_2^2 + \|\mathbf{h}_{T_{01}^c}\|_2^2 \leq \|\mathbf{h}_{T_{01}}\|_2^2 (1 + \frac{1}{k^{2/p-1}(2/p-1)}),$$

which when substituted in (15) yields the desired result.

Proof outline for Theorem 5. This proof is similar to the analogous proof in [1] and differs from the previous one by defining T_0 as the support set of the S largest coefficients of $\mathbf{x}$, which is now no longer assumed sparse. This leads to $\|\mathbf{h}_{T_0^c}\|_p^p \leq \|\mathbf{h}_{T_0}\|_p^p + 2\|\mathbf{x}_{T_0^c}\|_p^p$. Using this inequality instead of the analogous one from the previous proof, the rest proceeds similarly with minor modifications to lead to the desired result.

4. NUMERICAL EXPERIMENTS

In this section we present numerical experiments, to illustrate the behavior of the constants in Theorems 4 and 5 and to empirically investigate the solution of (9) in the presence of noise, and how it depends on p.

To that end, a 256×1024 matrix $\mathbf{A}$ is randomly generated from a mean-zero Gaussian and held fixed. We estimate the restricted isometry constants of $\mathbf{A}$ (see Figure 1(a)) by computing singular values of 1000 randomly selected $M \times S$ submatrices (thus providing lower bounds on the true values). We then estimate the values of $C_{S,k,p}$ from Theorem 4 and since k is a free parameter, we compute the minimum value of the constant over all admissible $k > 1$, independently for each p. As shown in Figure 1(b) (with the optimal k values in Figure 1(c)), the resulting bound is much tighter than that obtained by fixing k, e.g., $k = 3$ as in Candès and Tao [3]. Note that the constants $C_{S,k,p}$ provide upper bounds which are possibly rather pessimistic: numerical experiments where we solve (9) yield lower errors, particularly when p is close to 0, see Figures 2 and 3. Nevertheless, there is a wide range of p values for which the constants are well behaved, and they guarantee a stable recovery.

We now describe the experiments where we generate $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for sparse and compressible $\mathbf{x}$ with various noise levels and solve (9). Note that the nonconvexity of P_p^ϵ when $p < 1$ means that our solutions may only be local minima. However, in the noise-free setting, the observation that local ℓ_p minimization can recover signals exactly [8], together with theoretical results providing circumstances under which the global ℓ_p minimizer is exact [10], suggest that local minimization algorithms may give global solutions in this context.

Our approach to solving (9) is to first solve the simpler, unconstrained formulation:

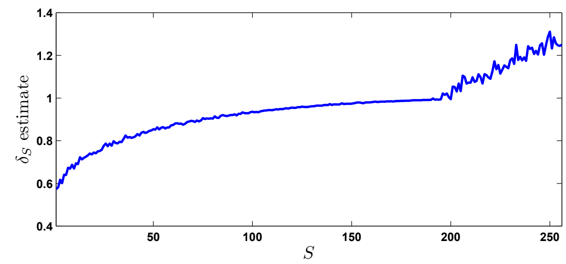
$$P_{\mu,p} : \min_{\mathbf{x}} \|\mathbf{x}\|_p^p + \mu \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2. \quad (16)$$

The parameter μ is adjusted manually and the minimization repeated, until the constraint in (9) is active. For each μ, the problem (16) is solved using an iteratively-reweighted least squares approach [11]. Code for this was contributed by Wotao Yin, for which we are grateful. The previous iterate $\mathbf{x}_{n-1}$ is substituted into the Euler-Lagrange equation, leading to a linear equation to solve for the next iterate $\mathbf{x}_n$:

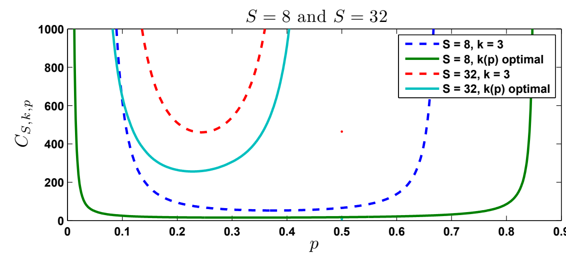
$$\left[\mathbf{x}_{n-1}^2 + \epsilon\right]^{\frac{p-2}{2}} \mathbf{x}_n + \mu A^T (A\mathbf{x}_n - b) = 0, \quad (17)$$

where the operations in the first term are to be understood componentwise, and the ϵ is added to avoid division by zero (as $p - 2$ is negative). We begin the iteration with the minimum-norm solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$, and use the strategy found effective in [8] (see also [12]) of using a moderately large $\epsilon = 1$, then iteratively decrementing ϵ by a factor of 10 after convergence and then repeating. Convergence was deemed complete when $\|\mathbf{x}_n - \mathbf{x}_{n-1}\|_2 < 10^{-8}$.

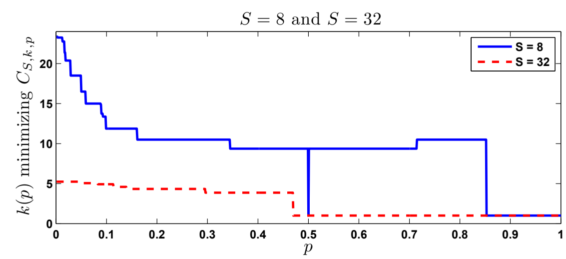
Sample results are shown in Figures 2 and 3 . The signals $\mathbf{x}$ were randomly generated from a mean-zero Gaussian distribution ($\sigma = 1$) on the support of $\mathbf{x}$. Solutions were computed for $p = 0.01, 0.1, 0.2, \ldots, 1$, for a very sparse signal ($S = 8$), a not so sparse signal ($S = 90$), and a compressible but not sparse signal



(a) Estimates of δ_S



(b) $C_{S,k,p}$ for $k = 3$ and for an optimal k with $S = 8$ and 32



(c) The value for the optimal k as a function of p

Fig. 1. **(a)** Estimates of δ_S for a specific Gaussian matrix $\mathbf{A}$ of size 256×1024. **(b)** The resulting constants $C_{S,k,p}$ at $S = 8$ and 32 computed both for $k = 3$ and for $k^*(S,p)$ which minimizes the constant for a given p and S. Choosing the best k reduces the upper bound on the reconstruction error. **(c)** The values of $k^*(S,p)$. The dip at $p = \frac{1}{2}$ for $S = 8$ is likely due to small numerical differences arising from $1/p$ being an integer. For both $S = 8$ and $S = 32$, once p is large enough, there is no $k > 1$ for which (10) is satisfied.

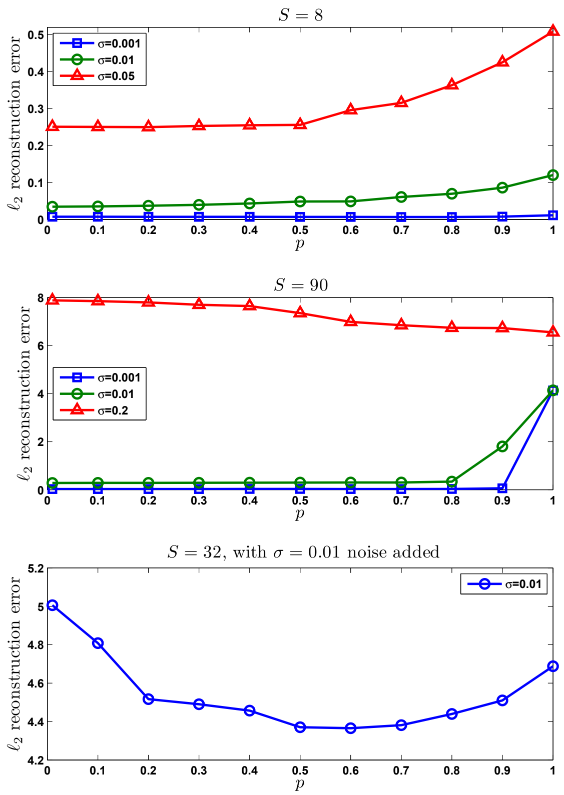


Fig. 2. Plots of reconstruction error versus p for solutions of (9). Top: for a very sparse signal, the reconstruction error rises with p. Middle: for a less sparse signal, when the noise level is not too large, the error changes little, except for large p where the reconstruction is poor. For stronger noise, the reconstruction is uniformly poor. Bottom: for a signal that is compressible but not sparse, the error rises for small or large p, being least for p about $1/2$.

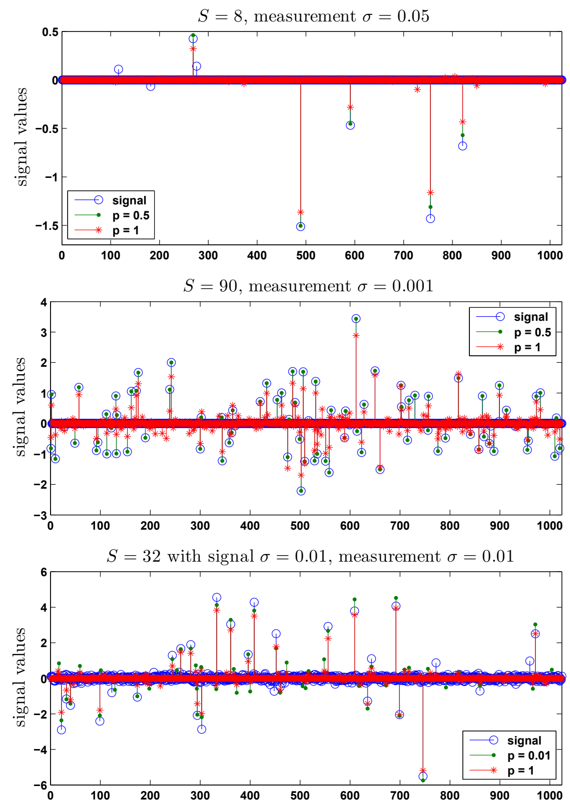


Fig. 3. Reconstructed signals obtained by solving (9). Top: for a very sparse signal with relatively strong noise in the measurements, the smallest components are not recovered. The $p = 1$ reconstruction also has spurious components. Middle: a less sparse signal, with a little noise. For p much less than 1, the reconstruction is nearly perfect, while for $p = 1$ the reconstruction errs in magnitude and has many spurious elements. Bottom: a compressible signal, with moderate measurement noise. The small p solution is much sparser, leading to a greater ℓ_2 distance from the not-sparse signal. The $p = 1$ solution overly penalizes the magnitude of components. The ℓ_2-optimal p is in between.

obtained by adding small Gaussian noise ($\sigma = 0.01$) to a randomly generated ($\sigma = 2$) sparse signal ($S = 32$). Gaussian noise of different levels was added to $\mathbf{A}\mathbf{x}$ to obtain the noisy measurements $\mathbf{b}$.

5. REFERENCES

[1] E. J. Candes, J. Romberg, and T. Tao, "Signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59 (8), pp. 1207–1223, 2005.

[2] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections and universal encoding strategies.," *IEEE Trans. Inform. Theory, 2004*, vol. 52, pp. 5406–5425, 2006.

[3] E. J. Candes and T. Tao, "Decoding by linear programming.," *IEEE Trans. Inform. Theory*, vol. 51, pp. 489–509, 2005.

[4] D. Donoho, "Compressed sensing.," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] Y. Tsaig and D. Donoho, "Extensions of compressed sensing," *Signal Process.*, vol. 86, no. 3, pp. 549–571, 2006.

[6] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, 2006.

[7] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.

[8] Rick Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, 2007.

[9] E. J. Candes, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians, Madrid, Spain*, 2006.

[10] Rick Chartrand and Valentina Staneva, "Restricted isometry properties and nonconvex compressive sensing," Preprint.

[11] J. A. Scales and A. Gersztenkorn, "Robust methods in inverse theory," *Inverse Problems*, vol. 4, no. 4, pp. 1071–1091, 1988.

[12] Rick Chartrand and Wotao Yin, "Iteratively reweighted algorithms for compressive sensing," Preprint.