Stable Signal Recovery
from Incomplete and Inaccurate Measurements

EMMANUEL J. CANDÈS
California Institute of Technology

JUSTIN K. ROMBERG
California Institute of Technology

AND

TERENCE TAO
University of California at Los Angeles

Abstract

Suppose we wish to recover a vector $x_0 \in \mathbb{R}^m$ (e.g., a digital signal or image) from incomplete and contaminated observations $y = Ax_0 + e$; A is an $n \times m$ matrix with far fewer rows than columns ($n \ll m$) and e is an error term. Is it possible to recover x_0 accurately based on the data y?

To recover x_0, we consider the solution $x^\sharp$ to the ℓ_1-regularization problem

$$\min \|x\|_{\ell_1} \quad \text{subject to} \quad \|Ax - y\|_{\ell_2} \leq \epsilon,$$

where ϵ is the size of the error term e. We show that if A obeys a uniform uncertainty principle (with unit-normed columns) and if the vector x_0 is sufficiently sparse, then the solution is within the noise level

$$\|x^\sharp - x_0\|_{\ell_2} \leq C \cdot \epsilon.$$

As a first example, suppose that A is a Gaussian random matrix; then stable recovery occurs for almost all such A's provided that the number of nonzeros of x_0 is of about the same order as the number of observations. As a second instance, suppose one observes few Fourier samples of x_0; then stable recovery occurs for almost any set of n coefficients provided that the number of nonzeros is of the order of $n/(\log m)^6$.

In the case where the error term vanishes, the recovery is of course exact, and this work actually provides novel insights into the exact recovery phenomenon discussed in earlier papers. The methodology also explains why one can also very nearly recover approximately sparse signals. © 2006 Wiley Periodicals, Inc.

1 Introduction

1.1 Exact Recovery of Sparse Signals

Recent papers [2, 3, 4, 5, 10] have developed a series of powerful results about the exact recovery of a finite signal $x_0 \in \mathbb{R}^m$ from a very limited number of observations. As a representative result from this literature, consider the problem of

recovering an unknown *sparse* signal $x_0(t) \in \mathbb{R}^m$, i.e., a signal x_0 whose support $T_0 = \{t : x_0(t) \neq 0\}$ is assumed to have small cardinality. All we know about x_0 are n linear measurements of the form

$$y_k = \langle x_0, a_k \rangle, \quad k = 1, \ldots, n \ \text{ or } \ y = Ax_0,$$

where the $a_k \in \mathbb{R}^m$ are known test signals. Of special interest is the vastly under-determined case, $n \ll m$, where there are many more unknowns than observations. At first glance, this may seem impossible. However, it turns out that one can actually recover x_0 exactly by solving the convex program[1]

$$(\text{P}_1) \qquad\qquad\qquad \min \|x\|_{\ell_1} \ \text{ subject to } Ax = y$$

provided that the matrix $A \in \mathbb{R}^{n \times m}$ obeys a *uniform uncertainty principle*.

The uniform uncertainty principle, introduced in [5] and refined in [4], essentially states that the $n \times m$ measurement matrix A obeys a "restricted isometry hypothesis." To introduce this notion, let A_T, $T \subset \{1, \ldots, m\}$, be the $n \times |T|$ sub-matrix obtained by extracting the columns of A corresponding to the indices in T. Then [4] defines the S-restricted isometry constant δ_S of A, which is the smallest quantity such that

$$(1.1) \qquad\qquad (1 - \delta_S)\|c\|_{\ell_2}^2 \leq \|A_T c\|_{\ell_2}^2 \leq (1 + \delta_S)\|c\|_{\ell_2}^2$$

for all subsets T with $|T| \leq S$ and coefficient sequences $(c_j)_{j \in T}$. This property essentially requires that every set of columns with cardinality less than S approximately behaves like an orthonormal system. It was shown (also in [4]) that if S verifies

$$(1.2) \qquad\qquad\qquad \delta_S + \delta_{2S} + \delta_{3S} < 1,$$

then solving (P_1) recovers *any* sparse signal x_0 with support size obeying $|T_0| \leq S$.

1.2 Stable Recovery from Imperfect Measurements

This paper develops results for the "imperfect" (and far more realistic) scenarios where the measurements are noisy and the signal is not exactly sparse. Everyone would agree that in most practical situations, we cannot assume that Ax_0 is known with arbitrary precision. More appropriately, we will assume instead that we are given "noisy" data $y = Ax_0 + e$, where e is some unknown perturbation bounded by a known amount $\|e\|_{\ell_2} \leq \epsilon$. To be broadly applicable, our recovery procedure must be *stable*: small changes in the observations should result in small changes in the recovery. This wish, however, may be quite hopeless. How can we possibly hope to recover our signal when not only is the available information severely incomplete, but the few available observations are also inaccurate?

Consider nevertheless (as in [12], for example) the convex program searching, among all signals consistent with the data y, for that with minimum ℓ_1-norm

$$(\text{P}_2) \qquad\qquad\qquad \min \|x\|_{\ell_1} \ \text{ subject to } \|Ax - y\|_{\ell_2} \leq \epsilon.$$

---

[1] (P_1) can even be recast as a linear program [6].

The first result of this paper shows that contrary to the belief expressed above, solving (P$_2$) will recover an unknown sparse object with an error at most proportional to the noise level. Our condition for stable recovery again involves the restricted isometry constants.

THEOREM 1.1 *Let S be such that $\delta_{3S} + 3\delta_{4S} < 2$. Then for any signal x_0 supported on T_0 with $|T_0| \leq S$ and any perturbation e with $\|e\|_{\ell_2} \leq \epsilon$, the solution $x^\sharp$ to (P$_2$) obeys*

$$(1.3) \qquad \|x^\sharp - x_0\|_{\ell_2} \leq C_S \cdot \epsilon,$$

where the constant C_S depends only on δ_{4S}. For reasonable values of δ_{4S}, C_S is well behaved; for example, $C_S \approx 8.82$ for $\delta_{4S} = \frac{1}{5}$ and $C_S \approx 10.47$ for $\delta_{4S} = \frac{1}{4}$.

It is interesting to note that for S obeying the condition of the theorem, the reconstruction from noiseless data is exact. It is quite possible that for some matrices A, this condition tolerates larger values of S than (1.2).

We would like to offer two comments. First, the matrix A is rectangular with many more columns than rows. As such, most of its singular values are zero. As emphasized earlier, the fact that the severely ill-posed matrix inversion keeps the perturbation from "blowing up" is rather remarkable and perhaps unexpected.

Second, no recovery method can perform fundamentally better for arbitrary perturbations of size ϵ. To see why this is true, suppose we had available an oracle that knew, in advance, the support T_0 of x_0. With this additional information, the problem is well-posed and we could reconstruct x_0 by the method of least squares:

$$\hat{x} = \begin{cases} (A_{T_0}^* A_{T_0})^{-1} A_{T_0}^* y & \text{on } T_0 \\ 0 & \text{elsewhere.} \end{cases}$$

Lacking any other information, we could easily argue that no method would exhibit a fundamentally better performance. Now of course, $\hat{x} - x_0 = 0$ on the complement of T_0, while on T_0

$$\hat{x} - x_0 = (A_{T_0}^* A_{T_0})^{-1} A_{T_0}^* e,$$

and since by hypothesis, the eigenvalues of $A_{T_0}^* A_{T_0}$ are well behaved,[2]

$$\|\hat{x} - x_0\|_{\ell_2} \approx \|A_{T_0}^* e\|_{\ell_2} \approx \epsilon,$$

at least for perturbations concentrated in the row space of A_{T_0}. In short, obtaining a reconstruction with an error term whose size is guaranteed to be proportional to the noise level is the best we can hope for.

Remarkably, stable recovery is also guaranteed for signals that are approximately sparse because we have the following companion theorem:

---

[2] Observe the role played by the singular values of A_{T_0} in the analysis of the oracle error.

THEOREM 1.2 *Suppose that x_0 is an arbitrary vector in $\mathbb{R}^m$, and let $x_{0,S}$ be the truncated vector corresponding to the S largest values of x_0 (in absolute value). Under the hypotheses of Theorem 1.1, the solution $x^\sharp$ to (P$_2$) obeys*

$$(1.4) \qquad \|x^\sharp - x_0\|_{\ell_2} \leq C_{1,S} \cdot \epsilon + C_{2,S} \cdot \frac{\|x_0 - x_{0,S}\|_{\ell_1}}{\sqrt{S}}.$$

For reasonable values of δ_{4S}, the constants in (1.4) are well behaved; for example, $C_{1,S} \approx 12.04$ and $C_{2,S} \approx 8.77$ for $\delta_{4S} = \frac{1}{5}$.

Roughly speaking, the theorem says that minimizing ℓ_1 stably recovers the S largest entries of an m-dimensional unknown vector x from only n measurements.

We now specialize this result to a commonly discussed model in mathematical signal processing, namely, the class of *compressible* signals. We say that x_0 is compressible if its entries obey a power law

$$(1.5) \qquad |x_0|_{(k)} \leq C_r \cdot k^{-r},$$

where $|x_0|_{(k)}$ is the kth largest value of x_0 ($|x_0|_{(1)} \geq \cdots \geq |x_0|_{(m)}$), $r \geq 1$, and C_r is a constant that depends only on r. Such a model is appropriate for the wavelet coefficients of a piecewise smooth signal, for example. If x_0 obeys (1.5), then

$$\frac{\|x_0 - x_{0,S}\|_{\ell_1}}{\sqrt{S}} \leq C_r' \cdot S^{-r+1/2}.$$

Observe now that in this case

$$\|x_0 - x_{0,S}\|_{\ell_2} \leq C_r'' \cdot S^{-r+1/2},$$

and for generic elements obeying (1.5), there are no fundamentally better estimates available. Hence, we see that from only n measurements, we achieve an approximation error that is almost as good as we would have obtained had we known everything about the signal x_0 and selected its S largest entries.

As a last remark, we would like to point out that in the noiseless case, Theorem 1.2 improves upon an earlier result from Candès and Tao; see also [8]. It is sharper in the sense that (1) this is a deterministic statement and there is no probability of failure, (2) it is universal in that it holds for all signals, (3) it gives upper estimates with better bounds and constants, and (4) it holds for a wider range of values of S.

1.3 Examples

It is of course of interest to know which matrices obey the uniform uncertainty principle with good isometry constants. Using tools from random matrix theory, [3, 5, 10] give several examples of matrices such that (1.2) holds for S on the order of n to within log factors. Examples include (proofs and additional discussion can be found in [5]):

(1) *Random matrices with i.i.d. entries.* Suppose the entries of A are independent and identically distributed (i.i.d.) Gaussian random variables with mean zero and variance $\frac{1}{n}$; then [5, 10, 17] show that the condition for Theorem 1.1 holds with overwhelming probability when

$$S \leq C \cdot \frac{n}{\log(m/n)}.$$

In fact, [4] gives numerical values for the constant C as a function of the ratio n/m. The same conclusion applies to binary matrices with independent entries taking values $\pm 1/\sqrt{n}$ with equal probability.

(2) *Fourier ensemble.* Suppose now that A is obtained by selecting n rows from the $m \times m$ discrete Fourier transform matrix and renormalizing the columns so that they are unit-normed. If the rows are selected at random, the condition for Theorem 1.1 holds with overwhelming probability for $S \leq C \cdot n/(\log m)^6$ [5]. (For simplicity, we have assumed that A takes on real-valued entries although our theory clearly accommodates complex-valued matrices so that our discussion holds for both complex- and real-valued Fourier transforms.)

This case is of special interest, as reconstructing a digital signal or image from incomplete Fourier data is an important inverse problem with applications in biomedical imaging (MRI and computed tomography), astrophysics (interferometric imaging), and geophysical exploration.

(3) *General orthogonal measurement ensembles.* Suppose A is obtained by selecting n rows from an $m \times m$ orthonormal matrix U and renormalizing the columns so that they are unit normed. Then [5] shows that if the rows are selected at random, the condition for Theorem 1.1 holds with overwhelming probability provided

$$(1.6) \qquad S \leq C \cdot \frac{1}{\mu^2} \cdot \frac{n}{(\log m)^6},$$

where $\mu := \sqrt{m} \max_{i,j} |U_{i,j}|$. Observe that for the Fourier matrix, $\mu = 1$, and thus (1.6) is an extension of the result for the Fourier ensemble.

This fact is of significant practical relevance because in many situations, signals of interest may not be sparse in the time domain but rather may be (approximately) decomposed as a sparse superposition of waveforms in a fixed orthonormal basis Ψ; for instance, piecewise smooth signals have approximately sparse wavelet expansions. Suppose that we use as test signals a set of n vectors taken from a second orthonormal basis Φ. We then solve (P_1) in the coefficient domain

$$(P_1') \qquad \min \|\alpha\|_{\ell_1} \quad \text{subject to } A\alpha = y,$$

where A is obtained by extracting n rows from the orthonormal matrix $U = \Phi\Psi^*$. The recovery condition then depends on the *mutual coherence* μ between the measurement basis Φ and the sparsity basis Ψ that measures the similarity between Φ and Ψ; $\mu(\Phi, \Psi) = \sqrt{m} \max |\langle \phi_k, \psi_j \rangle|$, $\phi_k \in \Phi$, $\psi_j \in \Psi$.

1.4 Prior Work and Innovations

The problem of recovering a sparse vector by minimizing ℓ_1 under linear equality constraints has recently received much attention, mostly in the context of *basis pursuit*, where the goal is to uncover sparse signal decompositions in overcomplete dictionaries. We refer the reader to [11, 13] and the references therein for a full discussion.

We would especially like to note two works by Donoho, Elad, and Temlyakov [12] and Tropp [18] that also study the recovery of sparse signals from noisy observations by solving (P_2) (and other closely related optimization programs), and give conditions for stable recovery. In [12], the sparsity constraint on the underlying signal x_0 depends on the magnitude of the maximum entry of the Gram matrix $M(A) = \max_{i,j:i \neq j} |(A^*A)|_{i,j}$. Stable recovery occurs when the number of nonzeros is at most $(M^{-1} + 1)/4$. For instance, when A is a Fourier ensemble and n is on the order of m, we will have M at least of the order $1/\sqrt{n}$ (with high probability), meaning that stable recovery is known to occur when the number of nonzeros is about at most $O(\sqrt{n})$. In contrast, the condition for Theorem 1.1 will hold when this number is about $n/(\log m)^6$ due to the range of support sizes for which the uniform uncertainty principle holds. In [18], a more general condition for stable recovery is derived. For the measurement ensembles listed in the previous section, however, the sparsity required is still on the order of $\sqrt{n}$ in the situation where n is comparable to m. In other words, *whereas these results require at least $O(\sqrt{m})$ observations per unknown, our results show that—ignoring loglike factors—only $O(1)$ are, in general, sufficient.*

More closely related is the very recent work of Donoho [9], who shows a version of (1.3) in the case where $A \in \mathbb{R}^{n \times m}$ is a Gaussian matrix with n proportional to m, with unspecified constants for both the support size and that appearing in (1.3). Our main claim is on a very different level since it is (1) deterministic (it can of course be specialized to random matrices) and (2) widely applicable since it extends to any matrix obeying the condition $\delta_{3S} + 3\delta_{4S} < 2$. In addition, the argument underlying Theorem 1.1 is short and simple, giving precise and sharper numerical values.

Finally, we would like to point out connections with fascinating ongoing work on fast randomized algorithms for sparse Fourier transforms [14, 19]. Suppose x_0 is a fixed vector with $|T_0|$ nonzero terms, for example. Then [14] shows that it is possible to randomly sample the frequency domain $|T_0| \operatorname{poly}(\log m)$ times ($\operatorname{poly}(\log m)$ denotes a polynomial term in $\log m$) and to reconstruct x_0 from this frequency data with positive probability. We do not know whether these algorithms are stable in the sense described in this paper or whether they can be modified to be universal; here, an algorithm is said to be universal if it reconstructs exactly all signals of small support.

2 Proofs

2.1 Proof of Theorem 1.1: Sparse Case

The proof of the theorem makes use of two geometrical facts about the solution $x^\sharp$ to (P$_2$).

(1) *Tube constraint.* First, $Ax^\sharp$ is within 2ϵ of the "noisefree" observations Ax_0 thanks to the triangle inequality

$$(2.1) \qquad \|Ax^\sharp - Ax_0\|_{\ell_2} \leq \|Ax^\sharp - y\|_{\ell_2} + \|Ax_0 - y\|_{\ell_2} \leq 2\epsilon.$$

Geometrically, this says that $x^\sharp$ is known to be in a cylinder around the n-dimensional plane Ax_0.

(2) *Cone constraint.* Since x_0 is feasible, we must have $\|x^\sharp\|_{\ell_1} \leq \|x_0\|_{\ell_1}$. Decompose $x^\sharp$ as $x^\sharp = x_0 + h$. As observed in [13],

$$\|x_0\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} \leq \|x_0 + h\|_{\ell_1} \leq \|x_0\|_{\ell_1},$$

where T_0 is the support of x_0, and $h_{T_0}(t) = h(t)$ for $t \in T_0$ and zero elsewhere (similarly for $h_{T_0^c}$). Hence, h obeys the cone constraint

$$(2.2) \qquad\qquad\qquad \|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1},$$

which expresses the geometric idea that h must lie in the cone of descent of the ℓ_1-norm at x_0.

Figure 2.1 illustrates both of these geometrical constraints. Stability follows from the fact that the intersection between (2.1) ($\|Ah\|_{\ell_2} \leq 2\epsilon$) and (2.2) is a set with small radius. This holds because every vector h in the ℓ_1-cone (2.2) is approximately orthogonal to the null space of A. We shall prove that $\|Ah\|_{\ell_2} \approx \|h\|_{\ell_2}$ and together with (2.1), this establishes the theorem.

We begin by dividing T_0^c into subsets of size M (we will choose M later) and enumerate T_0^c as $n_1, \ldots, n_{N-|T_0|}$ in decreasing order of magnitude of $h_{T_0^c}$. Set $T_j = \{n_\ell, (j-1)M + 1 \leq \ell \leq jM\}$. That is, T_1 contains the indices of the M largest coefficients of $h_{T_0^c}$, T_2 contains the indices of the next M largest coefficients, and so on.

With this decomposition, the ℓ_2-norm of h is concentrated on $T_{01} = T_0 \cup T_1$. Indeed, the kth-largest value of $h_{T_0^c}$ obeys

$$|h_{T_0^c}|_{(k)} \leq \frac{\|h_{T_0^c}\|_{\ell_1}}{k},$$

and therefore

$$\|h_{T_{01}^c}\|_{\ell_2}^2 \leq \|h_{T_0^c}\|_{\ell_1}^2 \sum_{k=M+1}^{m} \frac{1}{k^2} \leq \frac{\|h_{T_0^c}\|_{\ell_1}^2}{M}.$$

Further, the ℓ_1-cone constraint gives

$$\|h_{T_{01}^c}\|_{\ell_2}^2 \leq \frac{\|h_{T_0}\|_{\ell_1}^2}{M} \leq \frac{\|h_{T_0}\|_{\ell_2}^2 \cdot |T_0|}{M},$$
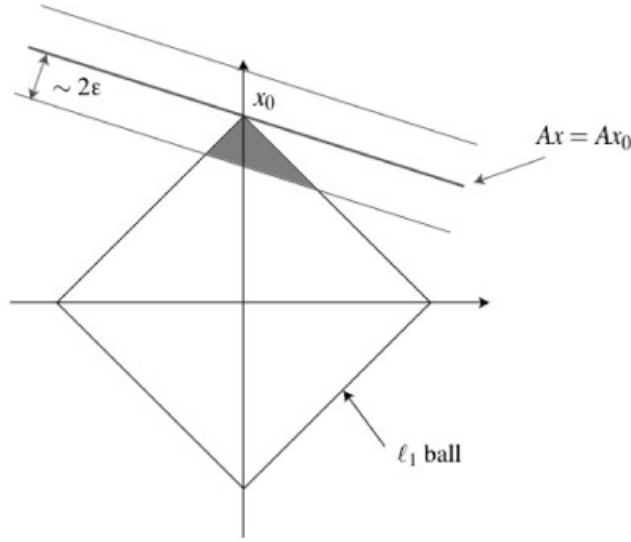
FIGURE 2.1. Geometry in $\mathbb{R}^2$. Here, the point x_0 is a vertex of the ℓ_1-ball, and the shaded area represents the set of points obeying both the tube and the cone constraints. By showing that every vector in the cone of descent at x_0 is approximately orthogonal to the null space of A, we will ensure that $x^\sharp$ is not too far from x_0.

and thus

$$(2.3) \qquad \|h\|_{\ell_2}^2 = \|h_{T_{01}}\|_{\ell_2}^2 + \|h_{T_{01}^c}\|_{\ell_2}^2 \leq \left(1 + \frac{|T_0|}{M}\right) \cdot \|h_{T_{01}}\|_{\ell_2}^2.$$

Observe now that

$$\|Ah\|_{\ell_2} = \left\| A_{T_{01}} h_{T_{01}} + \sum_{j \geq 2} A_{T_j} h_{T_j} \right\|_{\ell_2} \geq \|A_{T_{01}} h_{T_{01}}\|_{\ell_2} - \left\| \sum_{j \geq 2} A_{T_j} h_{T_j} \right\|_{\ell_2}$$

$$\geq \|A_{T_{01}} h_{T_{01}}\|_{\ell_2} - \sum_{j \geq 2} \|A_{T_j} h_{T_j}\|_{\ell_2},$$

and thus

$$\|Ah\|_{\ell_2} \geq \sqrt{1 - \delta_{M+|T_0|}} \, \|h_{T_{01}}\|_{\ell_2} - \sqrt{1 + \delta_M} \sum_{j \geq 2} \|h_{T_j}\|_{\ell_2}.$$

Set $\rho = |T_0|/M$. As we shall see later,

$$(2.4) \qquad \sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \leq \sqrt{\rho} \cdot \|h_{T_0}\|_{\ell_2},$$

which gives

$$(2.5) \quad \|Ah\|_{\ell_2} \geq C_{|T_0|,M} \cdot \|h_{T_{01}}\|_{\ell_2}, \quad C_{|T_0|,M} := \sqrt{1 - \delta_{M+|T_0|}} - \sqrt{\rho}\sqrt{1 + \delta_M}.$$

It then follows from (2.3) and $\|Ah\|_{\ell_2} \leq 2\epsilon$ that

$$(2.6) \qquad \|h\|_{\ell_2} \leq \sqrt{1+\rho} \cdot \|h_{T_{01}}\|_{\ell_2} \leq \frac{\sqrt{1+\rho}}{C_{|T_0|,M}} \cdot \|Ah\|_{\ell_2} \leq \frac{2\sqrt{1+\rho}}{C_{|T_0|,M}} \cdot \epsilon,$$

provided that the denominator is of course positive.

We may specialize (2.6) and take $M = 3|T_0|$. The denominator is positive if $\delta_{3|T_0|} + 3\delta_{4|T_0|} < 2$ (this is true if $\delta_{4|T_0|} < \frac{1}{2}$, say) which proves the theorem. Note that if δ_{4S} is a little smaller, the constant in (1.3) is not large. For $\delta_{4S} \leq \frac{1}{5}$, $C_S \approx 8.82$, while for $\delta_{4S} \leq \frac{1}{4}$, $C_S \approx 10.47$ as claimed.

It remains to argue about (2.4). Observe that by construction, the magnitude of each coefficient in T_{j+1} is less than the average of the magnitudes in T_j:

$$|h_{T_{j+1}}(t)| \leq \frac{\|h_{T_j}\|_{\ell_1}}{M}.$$

Then

$$\|h_{T_{j+1}}\|_{\ell_2}^2 \leq \frac{\|h_{T_j}\|_{\ell_1}^2}{M}$$

and (2.4) follows from

$$\sum_{j\geq 2} \|h_{T_j}\|_{\ell_2} \leq \sum_{j\geq 1} \frac{\|h_{T_j}\|_{\ell_1}}{\sqrt{M}} \leq \frac{\|h_{T_0}\|_{\ell_1}}{\sqrt{M}} \leq \sqrt{\frac{|T_0|}{M}} \cdot \|h_{T_0}\|_{\ell_2}.$$

2.2 Proof of Theorem 1.2: General Case

Suppose now that x_0 is arbitrary. We let T_0 be the indices of the largest $|T_0|$ coefficients of x_0 (the value $|T_0|$ will be decided later) and just as before, we divide up T_0^c into sets $T_1, \ldots, T_J$ of equal size $|T_j| = M$, $j \geq 1$, by decreasing order of magnitude. The cone constraint (2.2) may not hold but a variation does. Indeed, $x = x_0 + h$ is feasible and therefore

$$\|x_{0,T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} - \|x_{0,T_0^c}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} \leq \|x_{0,T_0} + h_{T_0}\|_{\ell_1} + \|x_{0,T_0^c} + h_{T_0^c}\|_{\ell_1}$$
$$\leq \|x_0\|_{\ell_1},$$

which gives

$$(2.7) \qquad \|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} + 2\|x_{0,T_0^c}\|_{\ell_1}.$$

The rest of the argument now proceeds essentially as before. First, h is in some sense concentrated on $T_{01} = T_0 \cup T_1$ since with the same notation

$$\|h_{T_{01}^c}\|_{\ell_2} \leq \frac{\|h_{T_0}\|_{\ell_1} + 2\|x_{0,T_0^c}\|_{\ell_1}}{\sqrt{M}} \leq \sqrt{\rho} \cdot \left(\|h_{T_0}\|_{\ell_2} + \frac{2\|x_{0,T_0^c}\|_{\ell_1}}{\sqrt{|T_0|}} \right),$$

which in turn implies

$$(2.8) \qquad \|h\|_{\ell_2} \leq (1 + \sqrt{\rho})\|h_{T_{01}}\|_{\ell_2} + 2\sqrt{\rho} \cdot \eta, \qquad \eta := \frac{\|x_{0,T_0^c}\|_{\ell_1}}{\sqrt{|T_0|}}.$$

TABLE 3.1. Recovery results for sparse one-dimensional signals. Gaussian white noise of variance σ^2 was added to each of the $n = 300$ measurements, and (P$_2$) was solved with ϵ chosen such that $\|e\|_2 \leq \epsilon$ with high probability (see (3.1)).

σ	0.01	0.02	0.05	0.1	0.2	0.5
ϵ	0.19	0.37	0.93	1.87	3.74	9.34
$\|x^\sharp - x_0\|_2$	0.25	0.49	1.33	2.55	4.67	6.61

TABLE 3.2. Recovery results for compressible one-dimensional signals. Gaussian white noise of variance σ^2 was added to each measurement, and (P$_2$) was solved with ϵ as in (3.1).

σ	0.01	0.02	0.05	0.1	0.2	0.5
ϵ	0.19	0.37	0.93	1.87	3.74	9.34
$\|x^\sharp - x_0\|_2$	0.69	0.76	1.03	1.36	2.03	3.20

Better estimates via Pythagoras' formula are of course possible (see (2.3)), but we ignore such refinements in order to keep the argument as simple as possible. Second, the same reasoning as before gives

$$\sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \leq \frac{\|h_{T_0^c}\|_{\ell_1}}{\sqrt{M}} \leq \sqrt{\rho} \cdot (\|h_{T_0}\|_{\ell_2} + 2\eta)$$

and thus

$$\|Ah\|_{\ell_2} \geq C_{|T_0|,M} \cdot \|h_{T_{01}}\|_{\ell_2} - 2\sqrt{\rho}\sqrt{1 + \delta_M} \cdot \eta,$$

where $C_{|T_0|,M}$ is the same as in (2.5). Since $\|Ah\| \leq 2\epsilon$, we again conclude that

$$\|h_{T_{01}}\|_{\ell_2} \leq \frac{2}{C_{|T_0|,M}} \cdot \left(\epsilon + \sqrt{\rho}\sqrt{1 + \delta_M}\,\eta\right)$$

(note that the constant in front of the ϵ-factor is the same as in the truly sparse case), and the claim (1.4) follows from (2.8). Specializing the bound to $M = 3|T_0|$ and assuming that $\delta_S \leq \frac{1}{5}$ gives the numerical values reported in the statement of the theorem.

3 Numerical Examples

This section illustrates the effectiveness of the recovery by means of a few simple numerical experiments. Our simulations demonstrate that in practice, the constants in (1.3) and (1.4) seem to be quite low.

Our first series of experiments is summarized in Tables 3.1 and 3.2. In ach experiment, a signal of length 1024 was measured with the same 300×1024 Gaussian measurement ensemble. The measurements were then corrupted by additive white Gaussian noise: $y_k = \langle x_0, a_k \rangle + e_k$ with $e_k \sim \mathcal{N}(0, \sigma^2)$ for various noise levels
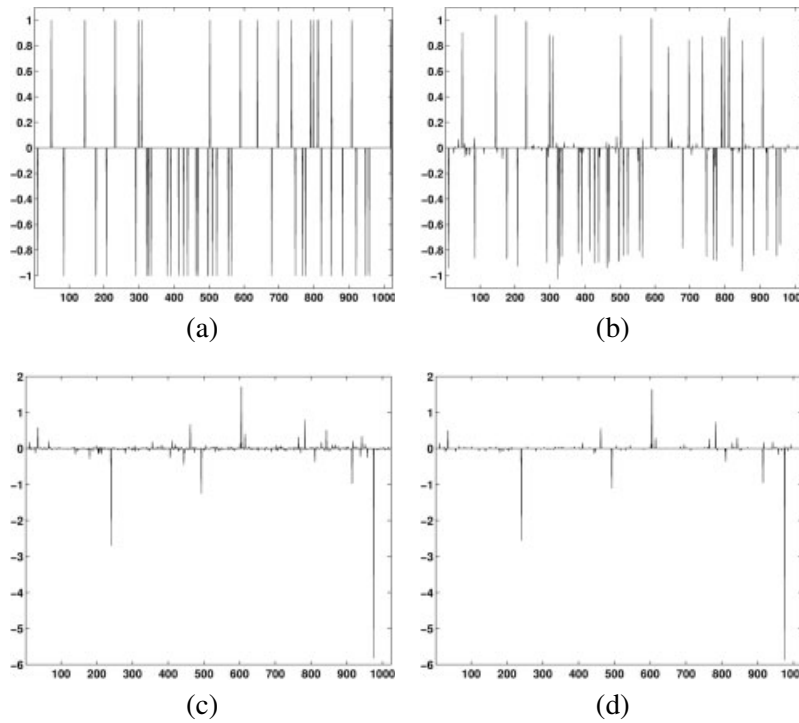
FIGURE 3.1. (a) Example of a sparse signal used in the one-dimensional experiments. There are 50 nonzero coefficients taking values ± 1. (b) Sparse signal recovered from noisy measurements with $\sigma = 0.05$. (c) Example of a compressible signal used in the one-dimensional experiments. (d) Compressible signal recovered from noisy measurements with $\sigma = 0.05$.

σ. The squared norm of the error $\|e\|_{\ell_2}^2$ is a chi-square random variable with mean $\sigma^2 n$ and standard deviation $\sigma^2\sqrt{2n}$; owing to well-known concentration inequalities, the probability that $\|e\|_{\ell_2}^2$ exceeds its mean plus 2 or 3 standard deviations is small. We then solve (P$_2$) with

$$(3.1) \qquad\qquad \epsilon^2 = \sigma^2(n + \lambda\sqrt{2n})$$

and select $\lambda = 2$ although other choices are of course possible.

Table 3.1 charts the results for sparse signals with 50 nonzero components. Ten signals were generated by choosing 50 indices uniformly at random and then selecting either -1 or 1 at each location with equal probability. An example of such a signal is shown in Figure 3.1(a). Previous experiments [3] demonstrated that we are empirically able to recover such signals perfectly from 300 noiseless Gaussian measurements, which is indeed the case for each of the 10 signals considered here. The average value of the recovery error (taken over the 10 signals) is recorded in

the bottom row of Table 3.1. In this situation, the constant in (1.3) appears to be less than 2.

Table 3.2 charts the results for 10 compressible signals whose components are all nonzero but decay as in (1.5). The signals were generated by taking a fixed sequence

$$(3.2) \qquad\qquad x_{\text{sort}}(t) = (5.819) \cdot t^{-10/9},$$

randomly permuting it, and multiplying by a random sign sequence (the constant in (3.2) was chosen so that the norm of the compressible signals is the same—$\sqrt{50}$— as the sparse signals in the previous set of experiments). An example of such a signal is shown in Figure 3.1(c). Again, 10 such signals were generated, and the average recovery error recorded in the bottom row of Table 3.2. For small values of σ, the recovery error is dominated by the approximation error—the second term on the right-hand side of (1.4). As a reference, the 50-term nonlinear approximation errors of these compressible signals is around 0.47; at low signal-to-noise ratios our recovery error is about 1.5 times this quantity. As the noise power gets large, the recovery error becomes less than ϵ, just as in the sparse case.

Finally, we apply our recovery procedure to realistic imagery. Photograph-like images, such as the 256×256 pixel *Boats* image shown in Figure 3.2(a), have wavelet coefficient sequences that are compressible (see [7]). The image is a 65 536-dimensional vector, making the standard Gaussian ensemble too unwieldy.[3] Instead, we make 25 000 measurements of the image using a *scrambled* real Fourier ensemble; i.e., the test functions $a_k(t)$ are real-valued sines and cosines (with randomly selected frequencies) that are temporally scrambled by randomly permuting the m time points. This ensemble is obtained from the (real-valued) Fourier ensemble by a random permutation of the columns. For our purposes here, the test functions behave like a Gaussian ensemble in the sense that from n measurements, one can recover signals with about $n/5$ nonzero components exactly from noise-less data. There is a computational advantage as well, since we can apply A and its adjoint A^* to an arbitrary vector by means of an m-point fast Fourier transform. To recover the wavelet coefficients of the object, we simply solve

$$(\text{P}'_2) \qquad\qquad \min \|\alpha\|_{\ell_1} \quad \text{subject to } \|AW^*\alpha - y\|_{\ell_2} \leq \epsilon,$$

where A is the scrambled Fourier ensemble and W is the discrete Daubechies-8 orthogonal wavelet transform.

We will attempt to recover the image from measurements perturbed in two different manners. First, as in the one-dimensional experiments, the measurements were corrupted by additive white Gaussian noise with $\sigma = 5 \cdot 10^{-4}$ so that $\sigma \cdot \sqrt{n} = 0.0791$. As shown in Figure 3.3, the noise level is significant; the signal-to-noise ratio is $\|Ax_0\|_{\ell_2}/\|e\|_{\ell_2} = 4.5$. With $\epsilon = 0.0798$ as in (3.1), the recovery error is $\|\alpha^\sharp - \alpha_0\|_{\ell_2} = 0.1303$ (the original image has unit norm).

---

[3] Storing a double-precision $25\,000 \times 65\,536$ matrix would use around 13.1 gigabytes of memory, about the capacity of three standard DVDs.
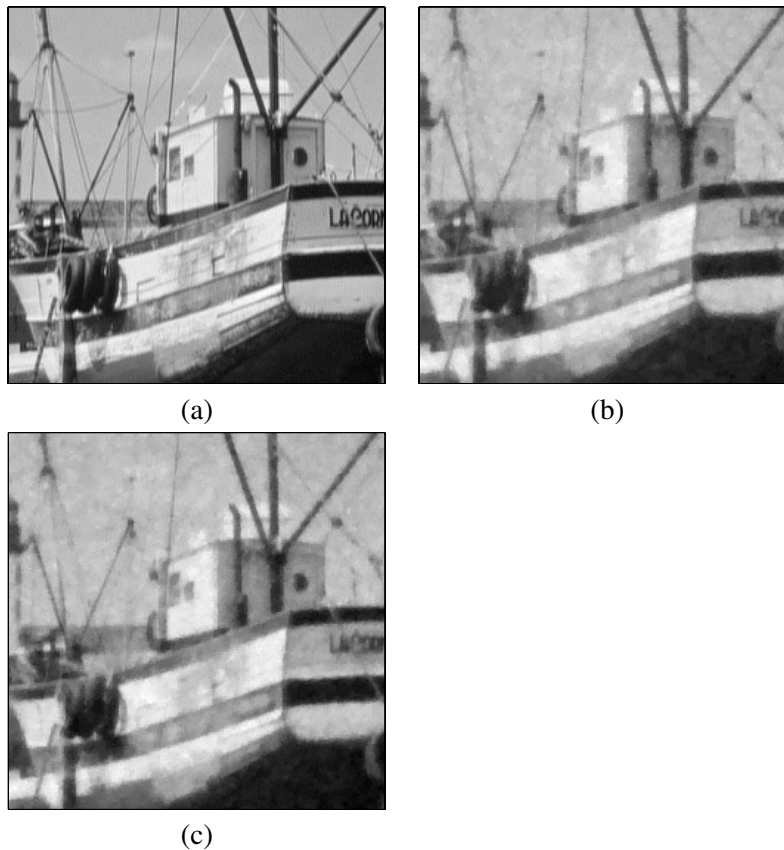
(a)   (b)

(c)

FIGURE 3.2. (a) Original 256×256 *Boats* image. (b) Recovery via
(TV) from $n = 25\,000$ measurements corrupted with Gaussian noise.
(c) Recovery via (TV) from $n = 25\,000$ measurements corrupted by
roundoff error. In both cases, the reconstruction error is less than the
sum of the nonlinear approximation and measurement errors.

For comparison, the 5000-term nonlinear approximation error for the image is
$\|\alpha_{0,5000} - \alpha_0\|_{\ell_2} = 0.050$. Hence the recovery error is very close to the sum of
the approximation error and the size of the perturbation.

Another type of perturbation of practical interest is *roundoff* or *quantization*
error. In general, the measurements cannot be taken with arbitrary precision, ei-
ther because of limitations inherent to the measuring device, or because we wish to
communicate the measurements using some small number of them. Unlike addi-
tive white Gaussian noise, roundoff error is deterministic and signal dependent—a
situation our methodology deals with easily.

We conducted the roundoff error experiment as follows: Using the same scram-
bled Fourier ensemble, we took $25\,000$ measurements of *Boats* and rounded (quan-
tized) them to one digit (we restricted the values of the measurements to be one

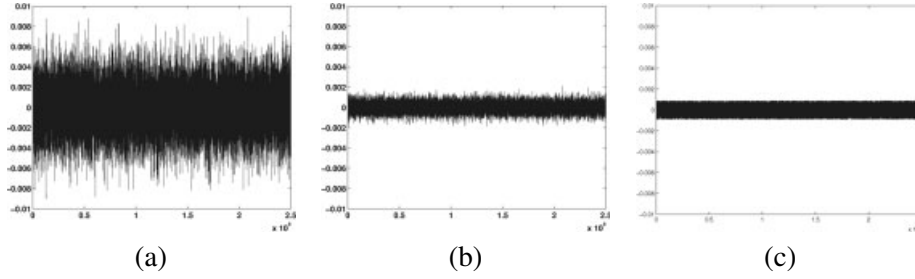(a)   (b)   (c)

FIGURE 3.3. (a) Noiseless measurements Ax_0 of the *Boats* image.
(b) Gaussian measurement error with $\sigma = 5 \cdot 10^{-4}$ in the recovery exper-
iment summarized in the left column of Table 3.3. The signal-to-noise
ratio is $\|Ax_0\|_{\ell_2}/\|e\|_{\ell_2} = 4.5$. (c) Roundoff error in the recovery exper-
iment summarized in the right column of Table 3.3. The signal-to-noise
ratio is $\|Ax_0\|_{\ell_2}/\|e\|_{\ell_2} = 4.3$.

of ten preset values, equally spaced). The measurement error is shown in Fig-
ure 3.3(c), and the signal-to-noise ratio is $\|Ax_0\|_{\ell_2}/\|e\|_{\ell_2} = 4.3$. To choose ϵ, we
used a rough model for the size of the perturbation. To a first approximation, the
roundoff error for each measurement behaves like a uniformly distributed random
variable on $(-\frac{q}{2}, \frac{q}{2})$, where q is the distance between quantization levels. Under
this assumption, the size of the perturbation $\|e\|_{\ell_2}^2$ behaves like a sum of squares of
uniform random variables

$$Y = \sum_{k=1}^{n} X_k^2, \quad X_k \sim \text{Uniform}\left(-\frac{q}{2}, \frac{q}{2}\right).$$

Here, $\text{mean}(Y) = nq^2/12$ and $\text{std}(Y) = \sqrt{n}q^2/(6\sqrt{5})$. Again, Y is no larger than
$\text{mean}(Y) + \lambda \, \text{std}(Y)$ with high probability, and we select

$$\epsilon^2 = n\frac{q^2}{12} + \lambda\sqrt{n}\frac{q^2}{6\sqrt{5}},$$

where as before, $\lambda = 2$. The results are summarized in the second column of
Table 3.3. As in the previous case, the recovery error is very close to the sum of
the approximation and measurement errors. Also note that despite the crude nature
of the perturbation model, an accurate value of ϵ is chosen.

Although the errors in the recovery experiments summarized in the third row
of Table 3.3 are as we had hoped, the recovered images tend to contain visually
displeasing high-frequency oscillatory artifacts. To address this problem, we can
solve a slightly different optimization problem to recover the image from the same
corrupted measurements. In place of (P$'_2$), we solve

(TV) $\min \|x\|_{TV} \quad \text{subject to } \|Ax - y\|_{\ell_2} \leq \epsilon$

TABLE 3.3. Image recovery results. Measurements of the *Boats* image were corrupted in two different ways: by adding white noise (left column) with $\sigma = 5 \cdot 10^{-4}$ and by rounding off to one digit (right column). In each case, the image was recovered in two different ways: by solving (P$_2'$) (third row) and solving (TV) (fourth row). The (TV) images are shown in Figure 3.2.

	White noise	Round-off
$\|e\|_{\ell_2}$	0.0789	0.0824
ϵ	0.0798	0.0827
$\|\alpha^\sharp - \alpha_0\|_{\ell_2}$	0.1303	0.1323
$\|\alpha^\sharp_{TV} - \alpha_0\|_{\ell_2}$	0.0837	0.0843

where

$$\|x\|_{TV} = \sum_{i,j} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2} = \sum_{i,j} |(\nabla x)_{i,j}|$$

is the *total variation* [16] of the image x: the sum of the magnitudes of the (discretized) gradient. By substituting (TV) for (P$_2'$), we are essentially changing our model for photographlike images. Instead of looking for an image with a sparse wavelet transform that explains the observations, program (TV) searches for an image with a sparse gradient (i.e., without spurious high-frequency oscillations). In fact, it is shown in [3] that just as signals which are exactly sparse can be recovered perfectly from a small number of measurements by solving (P$_2$) with $\epsilon = 0$, signals with gradients that are exactly sparse can be recovered by solving (TV) (again with $\epsilon = 0$).

Figures 3.2(b) and (c) and the fourth row of Table 3.3 show the (TV) recovery results. The reconstructions have smaller error and do not contain visually displeasing artifacts.

4 Discussion

The convex programs (P$_2$) and (TV) are simple instances of a class of problems known as second-order cone programs (SOCPs). As an example, one can recast (TV) as

$$(4.1) \quad \min \sum_{i,j} u_{i,j} \quad \text{subject to} \quad -u_{i,j} \le \|G_{i,j}x\|_{\ell_2} \le u_{i,j}, \quad \|Ax - y\|_{\ell_2} \le \epsilon,$$

where $G_{i,j}x = (x_{i+1,j} - x_{i,j}, x_{i,j+1} - x_{i,j})$ [15]. SOCPs can be solved efficiently by interior-point methods [1] and hence our approach is computationally tractable.

From a certain viewpoint, recovering via (P$_2$) is using a priori information about the nature of the underlying image, i.e., that it is sparse in some known orthobasis, to overcome the shortage of data. In practice, we could of course use far more sophisticated models to perform the recovery. Obvious extensions include looking

for signals that are sparse in overcomplete wavelet or curvelet bases or for images that have certain geometrical structure. The numerical experiments in Section 3 show how changing the model can result in a higher-quality recovery *from the same set of measurements*.

Bibliography

[1] Boyd, S.; Vandenberghe, L. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[2] Candès, E. J.; Romberg, J. Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. Comput. Math.*, in press. arXiv: math.CA/0411273, 2004.

[3] Candès, E. J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52** (2006), no. 2, 489–509.

[4] Candès, E. J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inform. Theory*, **51** (2005), no. 12, 4203–4215.

[5] Candès, E. J.; Tao, T. Near-optimal signal recovery from random projections and universal encoding strategies. *IEEE Trans. Inform. Theory*, submitted. arXiv: math.CA/0410542.

[6] Chen, S. S.; Donoho, D. L.; Saunders, M. A. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** (1998), no. 1, 33–61. *SIAM Rev.* **43** (2001), no. 1, 129–159 (electronic).

[7] DeVore, R. A.; Jawerth, B.; Lucier, B. Image compression through wavelet transform coding. *IEEE Trans. Inform. Theory* **38** (1992), no. 2, part 2, 719–746.

[8] Donoho, D. L. Compressed sensing. *IEEE Trans. Inform. Theory*, submitted.

[9] Donoho, D. L. For most large undetermined systems of linear equations the minimal ℓ_1-norm near-solution is also the sparsest near-solution. *Comm. Pure Appl. Math*, in press.

[10] Donoho, D. L. For most large undetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math*, in press.

[11] Donoho, D. L.; Elad, M. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proc. Natl. Acad. Sci. USA* **100** (2003), no. 5, 2197–2202 (electronic).

[12] Donoho, D. L.; Elad, M.; Temlyakov, V. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, **52** (2006), no. 1, 6–18.

[13] Donoho, D. L.; Huo, X. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory* **47** (2001), no. 7, 2845–2862.

[14] Gilbert, A. C.; Muthukrishnan, S.; Strauss, M. Improved time bounds for near-optimal sparse Fourier representations. DIMACS Technical Report, 2004-49, October 2004.

[15] Goldfarb, D.; Yin, W. Second-order cone programming based methods for total variation image restoration. Technical report, Columbia University, 2004. *SIAM J. Sci. Comput.*, submitted.

[16] Rudin, L. I.; Osher, S.; Fatemi, E. Nonlinear total variation noise removal algorithm. *Phys. D* **60** (1992), 259–68.

[17] Szarek, S. Condition numbers of random matrices. *J. Complexity* **7** (1991), no. 2, 131–149.

[18] Tropp, J. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inform. Theory*, in press.

[19] Zou, J.; Gilbert, A. C.; Strauss, M. J.; Daubechies, I. Theoretical and experimental analysis of a randomized algorithm for sparse Fourier transform analysis. *J. Computational Phys.*, to appear. Preprint. arXiv: math.NA/0411102, 2004.

EMMANUEL J. CANDÈS
Applied and Computational
    Mathematics
California Institute of Technology
Pasadena, CA 91125
E-mail: emmanuel@
    acm.caltech.edu

JUSTIN K. ROMBERG
Applied and Computational
    Mathematics
California Institute of Technology
Pasadena, CA 91125
E-mail: jrom@
    acm.caltech.edu

TERENCE TAO
Department of Mathematics
University of California at Los Angeles
Los Angeles, CA 90095
E-mail: tao@math.ucla.edu