Stable Signal Recovery from Incomplete and Inaccurate Measurements

Emmanuel Candes[†], Justin Romberg[†], and Terence Tao[♯]

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

♯ Department of Mathematics, University of California, Los Angeles, CA 90095

February, 2005

Abstract

Suppose we wish to recover a vector $x_0 \in \mathbb{R}^n$ (e.g. a digital signal or image) from incomplete and contaminated observations $y = Ax_0 + e$; A is a p by n matrix with far fewer rows than columns ($p \ll n$) 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. Second, suppose one observes few Fourier samples of x_0, then stable recovery occurs for almost any set of p coefficients provided that the number of nonzeros is of the order of $p/[\log n]^3$.

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

Keywords. ℓ_1 minimization, basis pursuit, restricted orthonormality, sparsity, singular values of random matrices.

Acknowledgments. E. C. is partially supported by a National Science Foundation grant DMS 01-40698 (FRG) and by an Alfred P. Sloan Fellowship. J. R. is supported by National Science Foundation grants DMS 01-40698 and ITR ACI-0204932. T. T. is supported in part by grants from the Packard Foundation.

1 Introduction

1.1 Sparse recovery

Recent papers [2–5, 7] have developed a series of powerful results about the exact recovery of a finite signal $x_0 \in \mathbb{R}^n$ from a very limited number of observations. As a representative result from this literature, consider the problem of recovering an unknown *sparse* signal $x_0(t) \in \mathbb{R}^n$; that is, 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 p linear measurements of the form

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

where the $a_k \in \mathbb{R}^n$ are known test signals. Of special interest is the vastly underdetermined case, $p \ll n$, 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]

$$(P_1) \qquad \min \|x\|_{\ell_1} \quad \text{subject to} \quad Ax = y, \tag{1}$$

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

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

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

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 [5]) that if S obeys

$$\delta_S + \delta_{2S} + \delta_{3S} < 1, \tag{3}$$

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

1.2 Stable recovery

Everyone would agree that in most practical situations, it is not realistic to assume that one would know Ax_0 with arbitrary precision. A more realistic model would assume that one is 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 the available information is severely incomplete but in addition, 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

$$th(P_2) \qquad \min \|x\|_{\ell_1} \quad \text{subject to} \quad \|Ax - y\|_{\ell_2} \leq \epsilon. \tag{4}$$

[1] (P_1) can even be recast as a linear program [10].

The main result of this paper shows that contrary to the belief expressed above, the solution to (P_2) recovers the 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 *Let S be such that $\delta_{3S} + 3\delta_{4S} < 2$. Then for any signal x_0 supported on T_0 with $|T_0| \le S$ and any perturbation e with $\|e\|_{\ell_2} \le \epsilon$, the solution $x^\sharp$ to (P_2) obeys*

$$\|x^\sharp - x_0\|_{\ell_2} \le C_S \cdot \epsilon, \tag{5}$$

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

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 (3).

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 one had available an *oracle* letting us know, in advance, the support T_0 of x_0. With this additional information, the problem is well-posed and one could reconstruct x_0 by the method of Least-Squares for example,

$$\hat{x} = (A^*_{T_0} A_{T_0})^{-1} A^*_{T_0} y.$$

In the absence of any other information, one could easily argue that no method would exhibit a fundamentally better performance. Now of course,

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

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 one can hope for.

Remarkably, not only can we recover sparse input vectors but one can also stably recover approximately sparse vectors, as we have the following companion theorem.

Theorem 2 *Suppose that x_0 is an arbitary vector in $\mathbb{R}^n$ and let $x_{0,S}$ be the truncated vector corresponding to the S largest values of x_0 (in absolute value). Under the hypothesis of Theorem 2, the solution $x^\sharp$ to (P_2) obeys*

$$\|x^\sharp - x_0\|_{\ell_2} \le C_{1,S} \cdot \epsilon + C_{2,S} \cdot \frac{\|x - x_{0,S}\|_{\ell_1}}{\sqrt{S}}. \tag{6}$$

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

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

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

As commonly discussed in the literature of mathematical signal processing, suppose for example that one is interested in recovering objects obeying

$$\|x\|_{\ell_p} := \sum_{i=1}^{p} |x_i|^p \leq 1, \tag{7}$$

with $p \leq 1$, say. Then

$$\frac{\|x - x_{0,S}\|_{\ell_1}}{\sqrt{S}} \leq C_p \cdot S^{1/2-1/p},$$

for some constant C_p only depending on p. Observe now that in this case

$$\|x - x_{0,S}\|_{\ell_2} \leq C'_p \cdot S^{1/2-1/p},$$

and for generic elements obeying (7), there are no fundamentally better estimates available. Hence, we see that with p measurements only, we achieve an approximation error which is almost as good as that one would obtain by knowing everything about the signal x and selecting its S-largest entries.

We would like to point out that in the noiseless case, Theorem 2 improves upon an earlier result from Candès and Tao, see also [9]; 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, 4, 7] give several examples of matrices such that (3) holds for S on the order of p to within log factors. Examples include (proofs and additional discussion can be found in [4]):

- *Random matrices with i.i.d. entries.* Suppose the entries of A are i.i.d. Gaussian with mean zero and variance $1/p$, then [4, 7, 16] show that the condition for Theorem 1 holds with overwhelming probability when

  $$S \leq C \cdot p / \log(n/p).$$

  In fact, [5] gives numerical values for the constant C as a function of the ratio p/n. The same conclusion applies to binary matrices with independent entries taking values $\pm 1/\sqrt{p}$ with equal probability.

- *Fourier ensemble.* Suppose now that A is obtained by selecting p rows from the $n \times n$ discrete Fourier transform and renormalizing the columns so that they are unit-normed. If the rows are selected at random, the condition for Theorem 1 holds with overwhelming probability for $S \leq C \cdot p/(\log n)^3$ [4].

  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 tomography), Astrophysics (interferometric imaging), and geophysical exploration.

- *General orthogonal measurement ensembles.* Suppose A is obtained by selecting p rows from an n by n orthonormal matrix U and renormalizing the columns so that they are unit-normed. Then [6] shows that if the rows are selected at random, the condition for Theorem 1 holds with overwhelming probability provided $S \leq C \cdot (n/\mu^2) \cdot p/(\log n)^3$, where $\mu := \max_{i,j} |U_{i,j}|$.

  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; e.g. in a nice wavelet basis. Suppose that we use as test signals a set of p vectors taken from a second orthonormal basis Φ. We then solve (P_1) in the coefficient domain

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

  where A is obtained by extracting p rows from the orthonormal matrix $U = \Phi\Psi^*$. The recovery condition then depends on the *mutual incoherence* μ between the measurement basis Φ and the sparsity basis Ψ which measures the similarity between Φ and Ψ; $\mu = \max \ |\langle \phi_k, \psi_j \rangle|$, $\phi_k \in \Phi$, $\psi_j \in \Psi$.

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* (finding sparse representations in overcomplete dictionaries). For a full discussion, we refer the reader to [11, 13] and the references therein.

We would especially like to note two works by Donoho, Elad, and Temlyakov [12], and Tropp [17] 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 $M(A) = \max_{k,m:k\neq m} |(A^*A)|_{k,m}$. Stable recovery occurs when the number of nonzeros is at most $(M^{-1}+1)/4$. For instance, when A is a Fourier ensemble and p is on the order of n, we will have M at least of the order $1/\sqrt{p}$ (with high probability), meaning that stable recovery is known to occur when the number of nonzeros is about at most $O(\sqrt{p})$. In contrast, the condition for Theorem 1 will hold when this number is about $p/(\log n)^3$, due to the range of support sizes for which the uniform uncertainty principle holds. In [17], 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 $\sqrt{p}$ in the situation where p is comparable to n. In other words, *whereas these results require at least $O(\sqrt{n})$ observations per unknown, our results show that only $O(1)$ or $O((\log n)^3)$ depending on the setup are, in general, sufficient.*

More closely related is the very recent work of Donoho [8] who shows a version of (5) in the case where $A \in \mathbb{R}^{p\times n}$ is a Gaussian matrix with p proportional to n, with unspecified constants for both the support size and that appearing in (5). 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 is short and simple, giving precise and much sharper numerical values. Finally, we would like to point out connections with fascinating ongoing work which develops fast randomized algorithms for sparse Fourier transforms [14, 18]. 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 $|T_0|\text{poly}(\log N)$ times ($\text{poly}(\log N)$ denotes a polynomial term in $\log N$), and reconstruct x_0 from these frequency data with positive probability. We do not know whether these algorithms are stable in the sense described in this paper, and whether they can be modified to be universal, i.e. reconstruct all signals of small support.

2 Proofs

2.1 Proof of Theorem 1: sparse case

The proof of the theorem makes use of two geometrical special facts about the solution $x^\sharp$ to (P_2).

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

$$\|Ax^\sharp - Ax_0\|_{\ell_2} \le \|Ax^\sharp - y\|_{\ell_2} + \|Ax_0 - y\|_{\ell_2} \le 2\epsilon. \tag{8}$$

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

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

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

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

$$\|h_{T_0^c}\|_{\ell_1} \le \|h_{T_0}\|_{\ell_1} \tag{9}$$

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

Figure 1 illustrates both these geometrical constraints. Stability follows from the fact that the intersection between (8) ($\|Ah\|_{\ell_2} \le 2\epsilon$) and (9) is a set with small radius, which holds because every vector h in the ℓ_1-cone (9) is approximately orthogonal to the nullspace of A. We shall prove that $\|Ah\|_{\ell_2} \approx \|h\|_{\ell_2}$ and together with (8), 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 $n_1, n_2, \ldots, n_{N-|T_0|}$ in decreasing order of magnitude of $h_{T_0^c}$. Set $T_j = \{n_\ell, (j-1)M + 1 \le \ell \le jM\}$. 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

$$|h_{T_0^c}|_{(k)} \;\le\; \|h_{T_0^c}\|_{\ell_1}/k$$

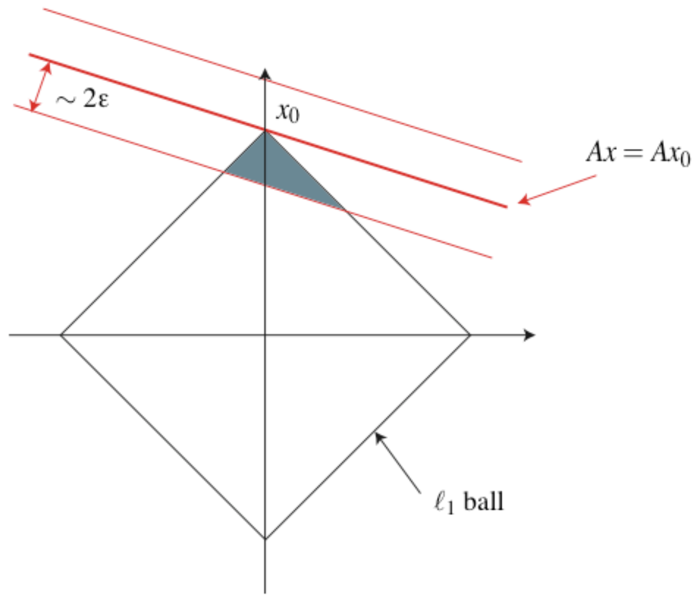

Figure 1: *Geometry in $\mathbb{R}^2$. Here, the point x_0 is a vertex of the ℓ_1 ball and the shaded area represents the set of points obeying both the tube and the cone constraints. By showing that every vector in the cone of descent at x_0 is approximately orthogonal to the nullspace of A, we will ensure that $x^\sharp$ is not too far from x_0.*

and, therefore,

$$\|h_{T_{01}^c}\|_{\ell_2}^2 \;\le\; \|h_{T_0^c}\|_{\ell_1}^2 \sum_{k=M+1}^{N} 1/k^2 \;\le\; \|h_{T_0^c}\|_{\ell_1}^2/M.$$

Further, the ℓ_1-cone constraint gives

$$\|h_{T_{01}^c}\|_{\ell_2}^2 \;\le\; \|h_{T_0}\|_{\ell_1}^2/M \;\le\; \|h_{T_0}\|_{\ell_2}^2 \cdot |T_0|/M$$

and thus

$$\|h\|_{\ell_2}^2 \;=\; \|h_{T_{01}}\|_{\ell_2}^2 + \|h_{T_{01}^c}\|_{\ell_2}^2 \;\le\; (1 + |T_0|/M) \cdot \|h_{T_{01}}\|_{\ell_2}^2. \qquad (10)$$

Observe now that

$$\begin{aligned}
\|Ah\|_{\ell_2} = \|A_{T_{01}} h_{T_{01}} + \sum_{j\ge 2} A_{T_j} h_{T_j}\|_{\ell_2} \;&\ge\; \|A_{T_{01}} h_{T_{01}}\|_{\ell_2} - \|\sum_{j\ge 2} A_{T_j} h_{T_j}\|_{\ell_2} \\
&\ge\; \|A_{T_{01}} h_{T_{01}}\|_{\ell_2} - \sum_{j\ge 2} \|A_{T_j} h_{T_j}\|_{\ell_2} \\
&\ge\; \sqrt{1 - \delta_{M+|T_0|}}\, \|h_{T_{01}}\|_{\ell_2} - \sqrt{1+\delta_M} \sum_{j\ge 2} \|h_{T_j}\|_{\ell_2}.
\end{aligned}$$

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

$$\sum_{j\ge 2} \|h_{T_j}\|_{\ell_2} \le \sqrt{\rho_M} \cdot \|h_{T_0}\|_{\ell_2} \qquad (11)$$

which gives

$$\|Ah\|_{\ell_2} \geq C_{|T_0|,M} \cdot \|h_{T_{01}}\|_{\ell_2}, \quad C_{|T_0|,M} := \sqrt{1-\delta_{M+|T_0|}} - \sqrt{\rho_M}\,\sqrt{1+\delta_M}. \tag{12}$$

It then follows from (10) and $\|Ah\|_{\ell_2} \leq 2\epsilon$ that

$$\|h\|_{\ell_2} \leq \sqrt{1+\rho_M} \cdot \|h_{T_{01}}\|_{\ell_2} \leq \frac{\sqrt{1+\rho_M}}{C_{|T_0|,M}} \cdot \|Ah\|_{\ell_2} \leq \frac{2\sqrt{1+\rho_M}}{C_{|T_0|,M}} \cdot \epsilon, \tag{13}$$

provided that the denominator is of course positive.

We may specialize (13) 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 $\delta_{4|T_0|} < 1/2$, say) which proves the theorem. Note that if δ_{4S} is a little smaller, the constant in (5) is not large. For $\delta_{4S} \leq 1/5$, $C_S \approx 8.82$, while for $\delta_{4S} \leq 1/4$, $C_S \approx 10.47$ as claimed.

It remains to argue about (11). Observe that by construction, the magnitude of each coefficient in T_{j+1} is less than the average of the magnitudes in T_j:

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

Then

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

and (11) follows from

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

2.2 Proof of Theorem 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 (9) 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

$$\|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} + 2\|x_{0,T_0^c}\|_{\ell_1}. \tag{14}$$

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

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

which in turn implies

$$\|h\|_{\ell_2} \leq (1+\sqrt{\rho_M})\|h_{T_{01}}\|_{\ell_2} + 2\sqrt{\rho_M}\delta_0, \quad \delta_0 := \|x_{0,T_0^c}\|_{\ell_1}/\sqrt{|T_0|}. \tag{15}$$

Better estimates via Pythagoras' formula are of course possible (see (10)) but we ignore such refinements in order to keep the argument as simple as possible. Second, the same reasoning as before gives

$$\sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \leq \frac{\|h_{T_{01}^c}\|_{\ell_1}}{\sqrt{M}} \leq \sqrt{\rho_M} \cdot (\|h_{T_0}\|_{\ell_2} + 2\delta_0)$$

and thus

$$\|Ah\|_{\ell_2} \geq C_{|T_0|,M} \cdot \|h_{T_{01}}\|_{\ell_2} - 2\sqrt{\rho_M}\,\sqrt{1+\delta_M}\delta_0,$$

where $C_{|T_0|,M}$ is the same as in (12). Since $\|Ah\| \leq 2\epsilon$, we again conclude that

$$\|h_{T_{01}}\|_{\ell_2} \leq \frac{2}{C_{|T_0|,M}} \cdot (\epsilon + \sqrt{\rho_M}\,\sqrt{1+\delta_M}\delta_0),$$

(note that the constant in front of the ϵ factor is the same as in the truly sparse case) and the claim (6) follows from (15). Specializing the bound to $M = 3|T_0|$ and assuming that $\delta_S \leq 1/5$ gives the numerical values reported in the statement of the theorem.

3 Conclusions

The convex program (P_2) is a simple instance of a class of problems known as second-order cone programs which are nowadays solved very efficiently by interior-point methods [1]. Our model assumes approximate knowledge of the noise level, i.e. $\|e\|_{\ell_2}$, which is the case in many practical situations. For example, suppose that e is white noise, a vector of i.i.d. Gaussian components with zero-mean and variance σ^2. Then $\|e\|_{\ell_2}^2/\sigma^2$ is distributed as a χ_p^2 with p degrees of freedom, and standard argument show that $\|e\|_{\ell_2}^2/\sigma^2 \leq p + O(\sqrt{p})$ with overwhelming probability. Another possible application concerns the case where e arises because of quantization errors. Knowledge of the quantization step would automatically translate into appropriate estimates for $\|e\|_{\ell_2}$. When the noise-level is completely unknown, one would have to find ways of selecting reasonable regularization parameters, perhaps by making some additional assumptions. This is, of course, a very different subject.

References

[1] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[2] E. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. Submitted to *Foundations of Comput. Math.*, November 2004. Available on the ArXiv preprint server:`math.CA/0411273`.

[3] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. Submitted *IEEE Trans. Inform. Theory*, June 2004. Available on theArXiV preprint server: `math.GM/0409186`.

[4] E. Candès and T. Tao. Near-optimal signal recovery from random projections and universal encoding strategies. Submitted to *IEEE Trans. Inform. Theory*, November 2004. Available on the ArXiV preprint server: `math.CA/0410542`.

[5] E. J. Candès and T. Tao. Decoding by linear programming. Submitted, December 2004. Available on the ArXiV preprint server: `math.MG/0502327`.

[6] E. Candès and J. Romberg. Sparsity, incoherence, and exact signal reconstruction from limited data. Manuscript, February 2005.

[7] D. L. Donoho. For most large undetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. Manuscript, September 2004.

[8] D. L. Donoho. For most large undetermined systems of linear equations the minimal ℓ_1-norm near-solution is also the sparsest near-solution. Manuscript, September 2004.

[9] D. L. Donoho. Compressed sensing. Manuscript, September 2004.

[10] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33–61, 1999.

[11] D. L. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization. *Proc. Natl. Acad. Sci. USA*, 100:2197–2202, 2003.

[12] D. L. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. Submitted to *IEEE Trans. Inform. Theory*, February 2004.

[13] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47:2845–2862, 2001.

[14] A. C. Gilbert, S. Muthukrishnan, and M. Strauss, Beating the B^2 bottleneck in estimating B-term Fourier representations, unpublished manuscript, May 2004.

[15] R. Gribonval and M. Nielsen, Sparse representations in unions of bases. *IEEE Trans. Inform. Theory* **49** (2003), 3320–3325.

[16] S. J. Szarek. Condition numbers of random matrices. *J. Complexity* **7** (1991), 131–149.

[17] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. Submitted to *IEEE Trans. Inform. Theory.* February 2005.

[18] J. Zou, A. C. Gilbert, M. Strauss, and I. Daubechies. Theoretical and Experimental Analysis of a Randomized Algorithm for Sparse Fourier Transform Analysis. *Submitted to Journal of Computational Physics.* November 2004.