The Restricted Isometry Property and Its Implications for Compressed Sensing

Emmanuel J. Candès [1]

Applied & Computational Mathematics, California Institute of Technology, Pasadena, CA 91125-5000.

Received *****; accepted after revision +++++

Presented by ?????

Abstract

It is now well-known that one can reconstruct sparse or compressible signals accurately from a very limited number of measurements, possibly contaminated with noise. This technique known as "compressed sensing" or "compressive sampling" relies on properties of the sensing matrix such as the *restricted isometry property.* In this note, we establish new results about the accuracy of the reconstruction from undersampled measurements which improve on earlier estimates, and have the advantage of being more elegant.

Résumé

La propiété d'isométrie restreinte et ses conséquences pour le compressed sensing. Il est maintenant bien connu que l'on peut reconstruire des signaux compressibles de manière précise à partir d'un nombre étonnamment petit de mesures, peut-être même bruitées. Cette technique appelée le "compressed sensing" ou "compressive sampling" utilise des propriétés de la matrice d'échantillonage comme la propriété d'isométrie restreinte. Dans cette note, nous présentons de nouveaux résultats sur la reconstruction de signaux à partir de données incomplètes qui améliorent des travaux précedents et qui, en outre, ont l'avantage d'être plus élégants.

1. Introduction

Suppose that we observe

$$y = \Phi x, \tag{1}$$

where $x \in \mathbb{R}^n$ is an object we wish to reconstruct, $y \in \mathbb{R}^m$ are available measurements, and Φ is a known $m \times n$ matrix. Here, we are interested in the underdetermined case with fewer equations than unknowns,

Email address: `emmanuel@acm.caltech.edu` (Emmanuel J. Candès).

URL: `http ://www.acm.caltech.edu/∼emmanuel` (Emmanuel J. Candès).

[1] This work was supported in part by a National Science Foundation grant CCF-515362 and by the 2006 Waterman Award.

i.e. $m < n$, and ask whether it is possible to reconstruct x with good accuracy. As such, the problem is of course ill-posed but suppose now that x is known to be sparse or nearly sparse in the sense that it depends on a smaller number of unknown parameters. This premise radically changes the problem, making the search for solutions feasible. In fact, it has been shown that the solution $x^\star$ to

$$\min_{\tilde{x}\in\mathbb{R}^n} \|\tilde{x}\|_{\ell_1} \qquad \text{subject to} \qquad \Phi\tilde{x} = y, \tag{2}$$

recovers x exactly provided that 1) x is sufficiently sparse and 2) the matrix Φ obeys a condition known as the *restricted isometry property* introduced below [2]. The purpose of this note is to introduce sharper and perhaps more elegant results in this direction.

To state our main results, we first recall the concept of restricted isometry constants.

Définition 1.1 *For each integer $s = 1, 2, \ldots$, define the isometry constant δ_s of a matrix Φ as the smallest number such that*

$$(1-\delta_s)\|x\|_{\ell_2}^2 \le \|\Phi x\|_{\ell_2}^2 \le (1+\delta_s)\|x\|_{\ell_2}^2 \tag{3}$$

holds for all s-sparse vectors. A vector is said to be s-sparse *if it has at most s nonzero entries.*

Our claims concern not only sparse vectors but *all* vectors $x \in \mathbb{R}^n$ and to measure the quality of the reconstruction, we will compare the reconstruction $x^\star$ with the *best sparse approximation* one could obtain if one knew exactly the locations and amplitudes of the s-largest entries of x; here and below, we denote this approximation by x_s, i.e. the vector x with all but the s-largest entries set to zero.

Theorem 1.1 (Noiseless recovery) *Assume that $\delta_{2s} < \sqrt{2} - 1$. Then the solution $x^\star$ to* (2) *obeys*

$$\|x^\star - x\|_{\ell_1} \le C_0\, \|x - x_s\|_{\ell_1} \tag{4}$$

and

$$\|x^\star - x\|_{\ell_2} \le C_0\, s^{-1/2}\, \|x - x_s\|_{\ell_1} \tag{5}$$

for some constant C_0 given explicitly below. In particular, if x is s-sparse, the recovery is exact.

Focusing on the case where x is sparse, one would in fact want to find the sparsest solution of $\Phi\tilde{x} = y$ and solve

$$\min_{\tilde{x}\in\mathbb{R}^n} \|\tilde{x}\|_{\ell_0} \qquad \text{subject to} \qquad \Phi\tilde{x} = y.$$

This is a hard combinatorial problem. However, Theorem 1.1 asserts that the ℓ_0 and ℓ_1 problems are in fact formally equivalent in the following sense:

- *if $\delta_{2s} < 1$, the ℓ_0 problem has a unique s-sparse solution;*
- *if $\delta_{2s} < \sqrt{2} - 1$, the solution to the ℓ_1 problem is that of the ℓ_0 problem. In other words, the convex relaxation is exact.*

Indeed, if $\delta_{2s} < 1$, any s-sparse solution is unique and we omit the standard details. In the other direction, suppose that $\delta_{2s} = 1$. Then $2s$ columns of Φ may be linearly dependent in which case there is a $2s$-sparse vector h obeying $\Phi h = 0$. One can then decompose h as $x - x'$ where both x and x' are s-sparse. This gives $\Phi x = \Phi x'$ which means that one cannot reconstruct all s-sparse vectors by any method whatsoever.

Our goal now is to extend the analysis to a more applicable situation in which the measurements are corrupted with noise. We observe

$$y = \Phi x + z, \tag{6}$$

where z is an unknown noise term. In this context, we propose reconstructing x as the solution to the convex optimization problem

$$\min_{\tilde{x}\in\mathbb{R}^n} \|\tilde{x}\|_{\ell_1} \qquad \text{subject to} \qquad \|y - \Phi\tilde{x}\|_{\ell_2} \le \varepsilon, \tag{7}$$

where ε is an upper bound on the size of the noisy contribution. Our next statement shows that one can stably reconstruct x under the same hypotheses as in Theorem 1.1.

Theorem 1.2 (Noisy recovery) *Assume that $\delta_{2s} < \sqrt{2} - 1$ and $\|z\|_{\ell_2} \le \varepsilon$. Then the solution to (7) obeys*

$$\|x^\star - x\|_{\ell_2} \le C_0 \, s^{-1/2} \, \|x - x_s\|_{\ell_1} + C_1 \, \varepsilon \tag{8}$$

with the same constant C_0 as before and some C_1 given explicitly below.

We emphasize that the constants C_0 and C_1 are rather small. For instance, when $\delta_{2s} = 0.2$, we have that the error in (8) is less than $4.2 \, s^{-1/2} \|x - x_s\|_{\ell_1} + 8.5\epsilon$.

2. Proofs

Lemma 2.1 *We have*

$$|\langle \Phi x, \Phi x' \rangle| \le \delta_{s+s'} \, \|x\|_{\ell_2} \, \|x'\|_{\ell_2}$$

for all x, x' supported on disjoint subsets $T, T' \subseteq \{1, \ldots, n\}$ with $|T| \le s$, $|T'| \le s'$.

Proof. This is a very simple application of the parallelogram identity. Suppose x and x' are unit vectors with disjoint support as above. Then

$$2(1 - \delta_{s+s'}) \le \|\Phi x \pm \Phi x'\|_{\ell_2}^2 \le 2(1 + \delta_{s+s'}).$$

Now the parallelogram identity asserts that

$$|\langle \Phi x, \Phi x' \rangle| = \frac{1}{4} \left| \|\Phi x + \Phi x'\|_{\ell_2}^2 - \|\Phi x - \Phi x'\|_{\ell_2}^2 \right| \le \delta_{s+s'},$$

which concludes the proof. □

The proofs of Theorems 1.1 and 1.2 parallel that in [1]. We begin by proving the latter which, in turn, gives the first part of the former, namely, (5). A modification of the argument then gives the second part, i.e. (4). Throughout the paper, x_T is the vector equal to x on an index set T and zero elsewhere.

To prove Theorem 1.2, we start with the basic observation

$$\|\Phi(x^\star - x)\|_{\ell_2} \le \|\Phi x^\star - y\|_{\ell_2} + \|y - \Phi x\|_{\ell_2} \le 2\varepsilon, \tag{9}$$

which follows from the triangle inequality and the fact that x is feasible for the problem (7). Set $x^\star = x + h$ and decompose h into a sum of vectors $h_{T_0}, h_{T_1}, h_{T_2}, \ldots$, each of sparsity at most s. Here, T_0 corresponds to the locations of the s largest coefficients of x; T_1 to the locations of the s largest coefficients of $h_{T_0^c}$; T_2 to the locations of the next s largest coefficients of $h_{T_0^c}$, and so on. The proof proceeds in two steps: the first shows that the size of h outside of $T_0 \cup T_1$ is essentially bounded by that of h on $T_0 \cup T_1$. The second shows that $\|h_{(T_0 \cup T_1)^c}\|_{\ell_2}$ is appropriately small.

For the first step, we note that for each $j \ge 2$,

$$\|h_{T_j}\|_{\ell_2} \le s^{1/2} \|h_{T_j}\|_{\ell_\infty} \le s^{-1/2} \|h_{T_{j-1}}\|_{\ell_1}$$

and thus

$$\sum_{j \ge 2} \|h_{T_j}\|_{\ell_2} \le s^{-1/2} (\|h_{T_1}\|_{\ell_1} + \|h_{T_2}\|_{\ell_1} + \ldots) \le s^{-1/2} \, \|h_{T_0^c}\|_{\ell_1}. \tag{10}$$

In particular, this gives the useful estimate

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} = \|\sum_{j \ge 2} h_{T_j}\|_{\ell_2} \le \sum_{j \ge 2} \|h_{T_j}\|_{\ell_2} \le s^{-1/2} \, \|h_{T_0^c}\|_{\ell_1}. \tag{11}$$

The key point is that $\|h_{T_0^c}\|_{\ell_1}$ cannot be very large for $\|x + h\|_{\ell_1}$ is minimum. Indeed,

$$\|x\|_{\ell_1} \ge \|x + h\|_{\ell_1} = \sum_{i \in T_0} |x_i + h_i| + \sum_{i \in T_0^c} |x_i + h_i| \ge \|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|$$

($\|x_{T_0^c}\|_{\ell_1} = \|x - x_s\|_{\ell_1}$ by definition), which gives

$$\|h_{T_0^c}\|_{\ell_1} \le \|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}. \tag{12}$$

Applying (11), then (12) and the Cauchy-Schwarz inequality to bound $\|h_{T_0}\|_{\ell_1}$ by $s^{1/2}\,\|h_{T_0}\|_{\ell_2}$ gives

$$\|h_{(T_0\cup T_1)^c}\|_{\ell_2} \le \|h_{T_0}\|_{\ell_2} + 2e_0, \quad e_0 \equiv s^{-1/2}\,\|x - x_s\|_{\ell_1}. \tag{13}$$

The second step bounds $\|h_{T_0\cup T_1}\|_{\ell_2}$. To do this, observe that $\Phi h_{T_0\cup T_1} = \Phi h - \sum_{j\ge 2}\Phi h_{T_j}$ and, therefore,

$$\|\Phi h_{T_0\cup T_1}\|_{\ell_2}^2 = \langle \Phi h_{T_0\cup T_1}, \Phi h\rangle - \langle \Phi h_{T_0\cup T_1}, \sum_{j\ge 2}\Phi h_{T_j}\rangle.$$

It follows from (9) and the restricted isometry property that

$$|\langle \Phi h_{T_0\cup T_1}, \Phi h\rangle| \le \|\Phi h_{T_0\cup T_1}\|_{\ell_2}\|\Phi h\|_{\ell_2} \le 2\varepsilon\sqrt{1+\delta_{2s}}\|h_{T_0\cup T_1}\|_{\ell_2}.$$

Moreover, it follows from Lemma 2.1 that

$$|\langle \Phi h_{T_0}, \Phi h_{T_j}\rangle| \le \delta_{2s}\|h_{T_0}\|_{\ell_2}\|h_{T_j}\|_{\ell_2}$$

and likewise for T_1 in place of T_0. Since $\|h_{T_0}\|_{\ell_2} + \|h_{T_1}\|_{\ell_2} \le \sqrt{2}\|h_{T_0\cup T_1}\|_{\ell_2}$ for T_0 and T_1 are disjoint,

$$(1-\delta_{2s})\|h_{T_0\cup T_1}\|_{\ell_2}^2 \le \|\Phi h_{T_0\cup T_1}\|_{\ell_2}^2 \le \|h_{T_0\cup T_1}\|_{\ell_2}\,(2\varepsilon\sqrt{1+\delta_{2s}} + \sqrt{2}\,\delta_{2s}\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2}).$$

Therefore, it follows from (10) that

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le \alpha\,\varepsilon + \rho\,s^{-1/2}\,\|h_{T_0^c}\|_{\ell_1}, \qquad \alpha \equiv \frac{2\sqrt{1+\delta_{2s}}}{1-\delta_{2s}}, \rho \equiv \frac{\sqrt{2}\,\delta_{2s}}{1-\delta_{2s}}. \tag{14}$$

We now conclude from this last inequality and (12) that

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le \alpha\varepsilon + \rho\|h_{T_0\cup T_1}\|_{\ell_2} + 2\rho e_0 \quad \Rightarrow \quad \|h_{T_0\cup T_1}\|_{\ell_2} \le (1-\rho)^{-1}(\alpha\varepsilon + 2\rho e_0).$$

And finally,

$$\|h\|_{\ell_2} \le \|h_{T_0\cup T_1}\|_{\ell_2} + \|h_{(T_0\cup T_1)^c}\|_{\ell_2} \le 2\|h_{(T_0\cup T_1)}\|_{\ell_2} + 2e_0 \le 2(1-\rho)^{-1}(\alpha\varepsilon + (1+\rho)e_0),$$

which is what we needed to show.

In the course of the proof, we established a useful fact:

Lemma 2.2 *Let h be any vector in the nullspace of Φ and let T_0 be any set of cardinality s. Then*

$$\|h_{T_0}\|_{\ell_1} \le \rho\,\|h_{T_0^c}\|_{\ell_1}, \qquad \rho = \sqrt{2}\,\delta_{2s}(1-\delta_{2s})^{-1}. \tag{15}$$

Indeed, this follows from $\|h_{T_0}\|_{\ell_1} \le s^{1/2}\|h_{T_0}\|_{\ell_2} \le s^{1/2}\|h_{T_0\cup T_1}\|_{\ell_2}$ and (14) with $\varepsilon = 0$.

To derive the last inequality (4), we use (12) and (15) to obtain

$$\|h_{T_0^c}\|_{\ell_1} \le \rho\|h_{T_0^c}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1} \qquad \Rightarrow \qquad \|h_{T_0^c}\|_{\ell_1} \le 2(1-\rho)^{-1}\|x_{T_0^c}\|_{\ell_1}.$$

Therefore, in the noiseless case, $h = x^* - x$ obeys

$$\|h\|_{\ell_1} = \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} \le 2(1+\rho)(1-\rho)^{-1}\|x_{T_0^c}\|_{\ell_1},$$

which is what we wanted.

References

[1] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, August 2006.

[2] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, December 2005.