A Simple Proof for Recoverability of ℓ_1-Minimization: Go Over or Under?

Yin Zhang
Technical Report TR05-09
Department of Computational and Applied Mathematics
Rice University, Houston, Texas, 77005, U.S.A.

August, 2005

Abstract

It is well-known by now that ℓ_1 minimization can help recover sparse solutions to under-determined linear equations or sparsely corrupted solutions to over-determined equations, and the two problems are equivalent under appropriate conditions. So far almost all theoretic results have been obtained through studying the "under-determined side" of the problem. In this note, we take a different approach from the "over-determined side" and show that a recoverability result (with the best available order) follows almost immediately from an inequality of Garnaev and Gluskin. We also connect dots with recoverability conditions obtained from different spaces.

1 Introduction

Let us consider the ℓ_1-norm approximation of an over-determined linear system:

$$(O1): \qquad \min_{x \in \Re^p} \|A^T x - b\|_1 \tag{1}$$

where $A \in \Re^{p \times n}$ with $p < n$ and $b \in \Re^n$, and the minimum ℓ_1-norm solution to an under-determined linear system:

$$(U1): \qquad \min_{y \in \Re^n} \{\|y\|_1 : By = c\} \tag{2}$$

where $B \in \Re^{q \times n}$ with $q < n$ and $c \in \Re^q$. To avoid trivial cases, we always assume that b and c are nonzero vectors. Let h be a sparse vector, and

$$b = A^T \hat{x} + h, \quad c = Bh. \tag{3}$$

The solution recovery problems associated with $(O1)$, also called*error-correction*, is to recover $\hat{x}$ for all sufficiently sparse vectors h. On the other hand, the problem associated with $(U1)$, also called *sparse basis selection* among many names, is to recover the vector h whenever it is sufficiently sparse. These two problems are equivalent under appropriate conditions [4] (see also [14]). We present here a formal statement as given in [16].

Proposition 1 (Equivalence). *Let both $A \in \Re^{p\times n}$ and $B \in \Re^{q\times n}$ be of full-rank with $p+q=n$. Then (O1) and (U1) are equivalent if and only if*

$$AB^T = 0 \quad \textit{and} \quad c = Bb. \tag{4}$$

Under this equivalence, if x^* solves (O1), then $b - A^T x^*$ solves (U1); and if y^* solves (U1), then $(AA^T)^{-1}A(b-y^*)$ solves (O1).

These solution-recovery problems have recently been studied by a number of authors (for example, see [1, 2, 3, 4, 5, 6, 14, 15]), and many intriguing results have been obtained.

1.1 Notations

By a partition (S, Z), we mean a partition of the index set $\{1, 2, \cdots, n\}$ into two disjoint subsets S and Z so that $S \cup Z = \{1, 2, \cdots, n\}$ and $S \cap Z = \emptyset$. In particular, for any $h \in \Re^n$, the partition $(S(h), Z(h))$ refers to the support $S(h)$ of h and its complement – the zero set $Z(h)$; namely,

$$S(h) = \{i : h_i \neq 0,\ 1 \leq i \leq n\}, \quad Z(h) = \{i : h_i = 0,\ i = 1 \leq i \leq n\}. \tag{5}$$

We will occasionally omit the dependence of a partition (S, Z) on h when it is clear from the context.

For any index subset $J \subset \{1, 2, \cdots, n\}$, $|J|$ is the cardinality of J. For any matrix $A \in \Re^{p\times n}$ and any index subset J, $A_J \in \Re^{p\times|J|}$ denotes the sub-matrix of A consisting of those columns of A whose indices are in J. For a vector $v \in \Re^n$, similarly, v_J denotes the sub-vector of v with those components whose indices are in J.

We call a vector a *binary vector* if all its components take value either -1 or $+1$ (not zero or one). We use $\mathcal{B}^k$ to denote the set of all binary vectors in $\Re^k$.

We use range$(\cdot)$ to denote the range space of a matrix and conv$(\cdot)$ the convex hull of a set of points.

2 A Recoverability Result

By recoverability, we mean that for $b = A^T\hat{x} + h$, $\hat{x}$ and h uniquely solve $(O1)$ and $(U1)$, respectively. For notational convenience, let us define

$$\phi(t) := \frac{t}{1 - log(t)}, \quad t \in (0, 1]. \tag{6}$$

Obviously, $\phi(t) \leq 1$ for $t \in (0, 1]$, $\lim_{t \to 0} \phi(t) = 0$, $\phi(1) = 1$, and the function is monotone.

Theorem 1. *For any natural numbers p and n with $p < n$, there exists a set of p-dimensional subspaces of $\Re^n$ (with a positive measure in Grassmanian) that has the following property. Let the rows of $A \in \Re^{p \times n}$ span a p-dimensional subspace in this set and $b = A^T\hat{x} + h$ with $|S(h)| = k$. Then $\hat{x}$ and h uniquely solve $(O1)$ and $(U1)$, respectively, whenever*

$$k < \alpha\phi\left(\frac{n-p}{n}\right)n, \tag{7}$$

where $\alpha > 0$ is an absolute constant independent of p and n. In the case of $n = 2p$, the recoverability holds whenever

$$k < \frac{p}{384}. \tag{8}$$

In Theorem 1, the terms on the right-hand sides of the inequalities serve as lower bounds on the maximal sparsity levels of h that allow recoverability. Unlike previously obtained bounds, the bounds in Theorem 1 are completely deterministic. These results are existence results, though, that do not give any concrete construction of a desirable subspace. The general bound in (7) unfortunately involves a unknown constant.

The bounds in Theorem 1 are not tighter than the existing probabilistic bounds as far as the order is concerned. A number of probabilistic results have been obtained [3, 4, 6, 14] stating that recoverability holds in high probability for random matrices if k does not exceed a fixed fraction of n and p/n remains a constant. Though, the bound in (8) for the case of $n = 2p$ appears tighter than currently available probabilistic ones. For instance, a bound of $k < p/3000$ was given by Candes and Tao [4, Theorem 1.3] for this case which holds in an overwhelming probability.

In the next section, we show that Theorem 1 is an immediate consequence of the following specialization of a result by Garnaev and Gluskin [11] (which was an improvement to an earlier result of Kasin [12]).

Lemma 1 (Garnaev-Gluskin). *For any natural numbers p and n with $p < n$, there exists a set of p-dimensional subspaces (with a positive measure in Grassmanian) of $\Re^n$ that has* *the following property. For any p-dimensional subspace $\mathbb{A}$ in this set,*

$$\|v\|_2 \leq \beta \sqrt{\frac{1+\log(n/(n-p))}{n-p}} \|v\|_1, \quad \forall v \in \mathbb{A}, \tag{9}$$

where $\beta > 0$ is an absolute constant independent of p and n.

3 A Simple Proof

We first need to make a crucial, yet extremely elementary, observation in Lemma 2 below. Then Theorem 1 will follow immediately from (9) via some simple arguments (also used by Linial and Novik [13] to prove a seemingly unrelated, yet in fact closed connected, result, as explained in the next section).

Lemma 2. *Given $b = A^T\hat{x} + h$, the vector $\hat{x}$ solves (O1) for all h with $|S(h)| \leq k$ if and only if for all partitions (S, Z) with $|S| = k$,*

$$\|v_S\|_1 \leq \|v_Z\|_1, \quad \forall\, v \in \text{range}\left(A^T\right). \tag{10}$$

In addition, $\hat{x}$ is always the unique solution if and only if the strict inequality holds.

Proof. Let $|S(h)| \leq k$ and $v = A^T(x - \hat{x})$ for an arbitrary $x \in \Re^p$. Then

$$\begin{aligned}
\|A^T x - b\|_1 &= \|A^T(x-\hat{x}) - h\|_1 = \|v - h\|_1 \\
&= \|h_S - v_S\|_1 + \|v_Z\|_1 \\
&\geq \|h_S\|_1 - \|v_S\|_1 + \|v_Z\|_1 \\
&= \|h\|_1 + (\|v_Z\|_1 - \|v_S\|_1) \\
&\equiv \|A^T\hat{x} - b\|_1 + (\|v_Z\|_1 - \|v_S\|_1).
\end{aligned}$$

Therefore, $\hat{x}$ is a minimizer if $\|v_Z\|_1 \geq \|v_S\|_1$ and a unique minimizer if the strict inequality holds. This establishes the sufficiency.

On the other hand, for any given $x \in \Re^p$, there are many $h \in \Re^n$ with $|S(h)| = k$ such that the inequality

$$\|h_S - v_S\|_1 \geq \|h_S\|_1 - \|v_S\|_1,$$

holds as an equality, as long as the components of h_S are chosen to have the same signs as and the same or larger magnitudes than the corresponding components of v_S. This establishes the necessity, and completes the proof. □

In view of Lemma 2, to prove Theorem 1 it suffices to show for some matrix A that corresponding to any partition (S, Z) with $|S| = k$,

$$\|v_S\|_1 < \frac{1}{2}\|v\|_1, \quad \forall\, v \in \text{range}\left(A^T\right),$$

which is clearly equivalent to the strict inequality in (10). Let A be a matrix so that A^T spans a subspace satisfying (9) in Lemma 1. Then

$$\|v_S\|_1 \le \sqrt{k}\|v_S\|_2 \le \sqrt{k}\|v\|_2 \le \sqrt{k}\beta\sqrt{\frac{1+\log(n/(n-p))}{n-p}}\|v\|_1.$$

The right-most side in the above is less than $(1/2)\|v\|_1$ if

$$\sqrt{k} < \frac{1}{2\beta}\sqrt{\frac{n-p}{1+\log(n/(n-p))}},$$

which is equivalent to (7) with $\alpha = 1/(4\beta^2) > 0$.

To prove (8), we again follow an argument used in [13]. A result by Kasin [12] implies the existence of an orthogonal transformation U in $\Re^p$ such that (see equations (2) and (3) in [13])

$$\|x\|_2 \le \frac{8e/\pi}{\sqrt{p}}(\|U^Tx\|_1 + \|x\|_1). \tag{11}$$

Let $A = [U\ I] \in \Re^{p\times 2p}$ and $v = A^Tx$. Then by (11)

$$\|v\|_2 = \sqrt{2}\|x\|_2 \le \sqrt{2}\,\frac{8e/\pi}{\sqrt{p}}(\|U^Tx\|_1 + \|x\|_1) = \frac{2^{3.5}e/\pi}{\sqrt{p}}\|v\|_1. \tag{12}$$

Now we can verify (8) in the same way as we did with (7) except that the inequality (12) is used in place of (9). The number 384 comes from the fact that $2^9e^2/\pi^2 < 384$. This completes the proof.

4 Recoverability Conditions in Different Spaces

Much has been learned so far about the necessary and sufficient conditions for recoverability. In this section, we will try to give a "global picture" for these conditions by showing the equivalence of three recoverability conditions as listed in Table 1. We first need to define the terminologies used in the table.

Definition 1 (k-balancedness, k-thickness and k-neighborliness).

(1) A subspace $\mathbb{A} \subset \Re^n$ is k-balanced (in ℓ_1-norm) if for any partition (S, Z) with $|S| = k$

$$\|v_S\|_1 \le \|v_Z\|_1, \ \ \forall\, v \in \mathbb{A}.$$

It is strictly k-balanced if the strict inequality holds.

(2) A subspace of $\Re^n$ *is k-thick if it intersects with all the* $(n-k)$*-dimensional faces (or simply* $(n-k)$*-faces) of the unit cube* $\{v \in \Re^n : \|v\|_\infty \leq 1\}$*. It is strictly k-thick if all the intersections lie in the relative interiors of the the* $(n-k)$*-faces.*

(3) Let $B := [b_1 \ \cdots \ b_n] \in \Re^{q \times n}$ *be of full rank* $q < n$*. The polytope*

$$P_c(B) := \mathrm{conv}(\{\pm b_j : j = 1, 2, \cdots, n\}) \subset \Re^q \tag{13}$$

is called (centrally) k-neighborly if every set of k vertices of $P_c(B)$ *not including any antipodal pair is the vertex set for a face of* $P_c(B)$*.*

A precise statement for the equivalence of recoverability conditions in Table 1 will be presented in Theorem 2. Recall that by recoverability, we mean a necessary and sufficient condition for uniquely recovering $\hat{x}$ by solving $(O1)$ or h by solving $(U1)$.

Table 1: Equivalent Recoverability Conditions for Different Spaces

Space	Condition
$\mathrm{range}(A^T) \subset \Re^n$	being strictly k-balanced
$\mathrm{range}(B^T) \subset \Re^n$	being strictly k-thick
$\mathrm{range}(B) = \Re^q$	$P_c(B)$ being k-neighborly

It is not difficult to verify that the null space of A is k-thick if and only if for any $u \in \mathcal{B}^k$

$$\min_{v \in \Re^n} \{\|v\|_\infty : Av = 0, \ v_S = u\} \leq 1,$$

and is strictly k-thick if and only if the strict inequality holds (see [16]).

Lemma 3. *Let* $A \in \Re^{p \times n}$ $(p < n)$ *and* (S, Z) *be a partition with* $|S| = k$*, then*

$$\max_{u \in \mathcal{B}^k} \min_{v \in \Re^n} \{\|v\|_\infty : Av = 0, \ v_S = u\} = \max_{x \in \Re^p} \left\{ \frac{\|v_S\|_1}{\|v_Z\|_1} : v = A^T x \right\}. \tag{14}$$

Consequently, let $\mathbb{A} = \mathrm{range}(A^T)$ *and* $\mathbb{A}^\perp$ *be its orthogonal complement, then*

$$\mathbb{A} \textit{ is (strictly) } k\textit{-balanced} \iff \mathbb{A}^\perp \textit{ is (strictly) } k\textit{-thick.} \tag{15}$$

Proof. First observe that for any binary vector $u \in \mathcal{B}^k$,

$$\begin{aligned}
\min_w \{\|w\|_\infty : A_Z w + A_S u = 0\} &= \min_w \{\xi : A_Z w = -A_S u,\ \|w\|_\infty \le \xi\} \\
&= \max_x \{(-A_S u)^T x : \|A_Z^T x\|_1 \le 1\} \\
&= \max_x \{(-A_S^T x)^T u : \|A_Z^T x\|_1 = 1\} \\
&\le \max_x \{\|A_S^T x\|_1 : \|A_Z^T x\|_1 = 1\} \\
&= \max_{x \ne 0} \{\|A_S^T x\|_1 / \|A_Z^T x\|_1\},
\end{aligned}$$

where we have used an equivalent conic program in the first equality and its dual in the second. This shows that the left-hand side of (14) is less than or equal to the right-hand side. On the other hand, if x^* is a solution to the right-hand side of (14), then $u = \mathrm{sign}(-A_S^T x^*)$, which is in $\mathcal{B}^k$, attains the equality. Therefore, (14) holds. Finally, (15) follows readily from (14). □

The polytope $P_c(B)$ defined in (13) is centrally symmetric, i.e., $z \in P_c(B)$ implies $-z \in P_c(B)$ as well. Recently, Donoho *et al* [7, 8, 9, 10] have extensively studied the connections between solution recovery for the under-determined problem ($U1$) and the k-neighborliness of $P_c(B)$. In particular, Donoho [7, Theorem 1] has shown that, for any $c = Bh$, solving ($U1$) recovers all h with $|S(h)| \le k$ if and only if $P_c(B)$ is a k-neighborly polytope of $2n$ vertices. We refer to their works (also that of Linial and Novik [13]) for detailed discussions and further references on the topic of neighborliness of centrally symmetric polytopes.

Theorem 2. *Let $A \in \Re^{p\times n}$ and $B \in \Re^{q\times n}$ be full rank such that $p + q = n$ and $AB^T = 0$. Let $b = A^T\hat{x} + h$ and $c = Bb$. Then for any h with $|S(h)| \le k$, $\hat{x}$ and h uniquely solve ($O1$) and ($U1$), respectively, if and only if one of the following three equivalent conditions holds: (1) $\mathrm{range}(A^T) \subset \Re^n$ is strictly k-balanced; or (2) $\mathrm{range}(B^T) \subset \Re^n$ is strictly k-thick; or (3) $P_c(B) \subset \Re^q$ is a k-neighborly polytope of $2n$ vertices.*

Proof. By Proposition 1, the recoverability of ($O1$) is equivalent to that of ($U1$), which in turn is equivalent to the k-neighborliness of $P_c(B)$ [7, Theorem 1]. On the other hand, by Lemma 2 the recoverability of ($O1$) is equivalent to the k-balancedness of $\mathrm{range}(A^T)$, which in turn is equivalent to the k-thickness of $\mathrm{range}(B^T)$ by Lemma 3. □

The equivalence between recoverability and the k-neighborliness of centrally symmetric polytopes allows one to utilize results from one side to obtain results for the other side, as was done in [7]. Theorem 1 in this paper could have been established by invoking the recent result of Linial and Novik [13] on k-neighborliness of centrally symmetric polytopes and its equivalence to recoverability. However, our proof is more self-contained and elementary, requiring only the Garnvaer-Gluskin inequality. Moreover, it has led to additional equivalence relationships and possibly opened up opportunities to study the problem at hand from different subspaces.

5 Concluding Remarks

The solution recovery problems have rich geometric, algebraic and combinatoric structures, and can be approached from different perspectives and spaces. Unlike most existing treatment, our derivations demonstrates the benefit of studying recoverability through examining the over-determined system in addition to the under-determined one.

References

[1] E. Candès, J. Romberg, and T. Tao. Exact signal reconstruction from highly incomplete frequency information. Submitted for publication, June 2004.

[2] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? Submitted for publication, Nov. 2004.

[3] E. J. Candès and T. Tao. Decoding by Linear Programming. Available from arXiv:math.MG/0502327, Feb. 2005.

[4] E. J. Candès and T. Tao. Error Correction via Linear Programming. Submitted to FOCS 2005.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by Basis Pursuit. SIAM J. Sci. Comput., 20(1):3361, 1999.

[6] David L. Donoho. For Most Large Underdetermined Systems of Linear Equations the Minimal ℓ^1-norm Solution is also the Sparsest Solution. Manuscript, Department of Statistics, Stanford University. September, 2004.

[7] David L. Donoho. High-Dimensional Centrosymmetric Polytopes with Neighborliness Proportional to Dimension. Manuscript, Department of Statistics, Stanford University, 2005.

[8] David L. Donoho. Neighborly Polytopes and Sparse Solutions of Underdetermined Linear Equations. Manuscript, Department of Statistics, Stanford University, 2005.

[9] David L. Donoho and Jared Tanner. Neighborliness of Randomly-Projected Simplices in High Dimensions. Manuscript, Department of Statistics, Stanford University, 2005.

[10] David L. Donoho and Jared Tanner. Sparse Nonnegative Solutions of Underdetermined Linear Equations by Linear Programming. Manuscript, Department of Statistics, Stanford University, 2005.

[11] A. Yu. Garnaev and E. D. Gluskin. The widths of a Euclidean ball. (Russian), Dokl. Akad. Nauk SSSR 277 (1984), no. 5, 10481052. (English translation: Soviet Math. Dokl. 30 (1984), no. 1, 200204.)

[12] B.S. Kasin. The widths of certain finite-dimensional sets and classes of smooth functions. (Russian), Izv. Akad. Nauk SSSR Ser. Mat. 41, 334-351 (1977). (English translation: Math. USSR-Izv. 11 (1977), no. 2, 317-333 (1978)).

[13] Nati Linial and Isabella Novik. How neighborly can a centrally symmetric polytope be? Manuscript, 2005.

[14] M. Rudelson and R. Veshynin. Geometric approach to error correcting codes and reconstruction of signals. Available from arXiv:math.MG/0502299, Feb. 2005.

[15] Joel A. Tropp and Anna C. Gilbert. Signal recovery from partial information via orthogonal matching pursuit. Manuscript, 2005.

[16] Y. Zhang. Solution-recovery in L1-norm for non-square linear systems: deterministic conditions and open questions. Technical report TR05-06, Department of computational and Applied Mathematics, Rice University, Houston, TX, 2005.