A Simple Proof for Recoverability of ℓ_1-Minimization (II): the Nonnegativity Case*

Yin Zhang
Technical Report TR05-10
Department of Computational and Applied Mathematics
Rice University, Houston, Texas, 77005, U.S.A.

September, 2005

Abstract

When using ℓ_1 minimization to recover a sparse, *nonnegative* solution to a underdetermined linear system of equations, what is the highest sparsity level at which recovery can still be guaranteed? Recently, Donoho and Tanner [10] discovered, by invoking classic results from the theory of convex polytopes [11, 12], that the highest sparsity level equals half of the number of equations. In this paper, we connect dots for different recoverability conditions obtained from different spaces, and provide an alternative, self-contained and elementary proof for this remarkable result.

1 Introduction

This paper is a sequel to [17] which concerns solution recovery problems associated with the following two problems:

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

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

where $A \in \Re^{p \times n}$ and $B \in \Re^{q \times n}$ are of full-rank with $p + q = n$, $b \in \Re^n$ and $c \in \Re^q$. The former is to find an approximate solution to an over-determined systems via ℓ_1-minimization,

*Research supported in part by NSF Grants DMS-0442065 and DMS-0405831

and the latter is to find the least ℓ_1-norm solution to an under-determined system. Under the conditions

$$AB^T = 0 \quad \text{and} \quad c = Bb. \tag{3}$$

problems $(O1)$ and $(U1)$ are equivalent. The solution recovery problems associated with them are pertinent to the question that given a sparse vector $h \in \Re^n$ and

$$b = A^T \hat{x} + h, \quad c = Bh, \tag{4}$$

will $\hat{x}$ and h uniquely solve $(O1)$ and $(U1)$, respectively?

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 The Nonnegativity Case

In the case $h \geq 0$, $(O1)$ and $(U1)$ have the following two counterparts, respectively,

$$(O1+): \qquad \min\{e^T(b - A^T x) : \ A^T x \leq b\}, \tag{5}$$

$$(U1+): \qquad \min\{e^T y : \ By = c, \ y \geq 0\}. \tag{6}$$

where $e \in \Re^n$ is the vector of all ones. Under the conditions in (3), these two problems are again equivalent.

Proposition 1 (Equivalence). *Let both $A \in \Re^{p \times n}$ and $B \in \Re^{q \times n}$ be of full-rank with $p + q = n$. If conditions in (3) hold, then (O1+) and (U1+) are equivalent. Specifically, 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+).*

Throughout of the paper, we will assume that the conditions in Proposition 1, in particular (3), always hold.

1.2 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,

$$S(h) = \{i : h_i \neq 0, \ 1 \leq i \leq n\}, \ \ 0, \ i = 1 \leq i \leq n\}. \tag{7}$$

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 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

One of the main purposes of this paper is to provide a self-contained and elementary proof for the following recoverability result. By recoverability, we mean that given $b = A^T\hat{x} + h$ with $h \geq 0$, the vectors $\hat{x}$ and h uniquely solve $(O1+)$ and $(U1+)$, respectively.

Theorem 1. *For any natural numbers p and n with $p < n$, there exists a set of p-dimensional subspaces of $\Re^n$ that has the following property. Let $A \in \Re^{p\times n}$, $\mathrm{range}(A^T) \subset \Re^n$ be a p-dimensional subspace in this set and $b = A^T\hat{x} + h$ with $h \geq 0$ and $|S(h)| = k$. Then $\hat{x}$ and h uniquely solve $(O1+)$ and $(U1+)$, respectively, whenever*

$$k \leq \frac{n-p}{2} \equiv \frac{q}{2}. \tag{8}$$

An equivalent form of the above result has been discovered by Donoho and Tanner [10] in connection to some classic results in the theory of convex polytopes [11, 12] (see also [13, 18]). This theorem is not a mere existence result because a number of qualifying p-dimensional subspaces can be explicitly constructed that possess the property of the theorem.

Normally, results have been obtained and stated in terms of the null space of range(A^T), i.e., range(B^T), which is q-dimensional (recall that $p + q = n$). In particular, to prove Theorem 1 it suffices to consider a q-dimensional subspace range(B^T) where B is defined as follows. Given any n non-zero real numbers $t_1 < t_2 < \cdots < t_n$, define

$$B = \begin{bmatrix} t_1 & t_2 & \cdots & t_n \\ t_1^2 & t_2^2 & \cdots & t_n^2 \\ \vdots & \vdots & \cdots & \vdots \\ t_1^q & t_2^q & \cdots & t_n^q \end{bmatrix} \in \Re^{q\times n}. \tag{9}$$

The convex hull spanned by the columns of B forms a *cyclic polytope* in $\Re^q$ which was shown [12], along with other examples, to possess a property called $q/2$-neighborliness (see the next section for a definition). It is this $q/2$-neighborliness property that enables Donoho and Tanner [10] to derive a $q/2$-recoverability result for the under-determined problem $(U1+)$. (In fact, the bound $q/2$ is known to be tight.)

We will present a simple proof for Theorem 1 in Section 4 that requires no prior results of any kind other than some basic facts about polynomials.

3 Connecting the Dots in Different Spaces

In this section, we give a "global picture" for three equivalent recoverability conditions listed in Table 1. We start with defining necessary terminologies.

Definition 1 (Half k-balancedness, k-thickness and k-neighborliness).
(1) A subspace $\mathbb{A} \in \Re^n$ is half k-balanced (in ℓ_1-norm) if for any partition (S, Z) with $|S| = k$

$$e_S^T v_S \le \|v_Z\|_1, \ \ \forall\, v \in \mathbb{A} \text{ satisfying } v_Z \le 0.$$

It is strictly half k-balanced if the strict inequality holds.
(2) A subspace of $\Re^n$ is half k-thick if it intersects with all the $(n-k)$-faces of the set $\{v \in \Re^n : \ v \le 1\}$. It is strictly half k-thick if all the intersections lie in the relative interiors of the $(n-k)$-faces.
(3) Let $B := [b_1 \ \cdots \ b_n] \in \Re^{q \times n}$ $(q < n)$ be of full rank. The polytope

$$P(B) := \mathrm{conv}(\{b_1, b_2, \cdots, b_n\}) \subset \Re^q \tag{10}$$

is called k-neighborly if every set of k vertices of $P(B)$ is the vertex set for a face of $P(B)$.

We note that the "half" k-balancedness and thickness are weaker notions than and are implied by their "full"-counterparts (see [17]). Moreover, it is clear that half k-balancedness (thickness) implies half $(k-1)$-balancedness (thickness).

Table 1: Equivalent Recoverability Conditions for Different Spaces

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

Recall that recoverability means necessary and sufficient conditions for being able to uniquely recover $\hat{x}$ via solving $(O1+)$ or h via solving $(U1+)$. Precise statements on the equivalence of the recoverability conditions in Table 1 will be presented in Theorem 2, as a consequence of the following lemmas.

Lemma 1. *The vectors $\hat{x}$ and h solve (O1+) and (U1+), respectively, for all $h \geq 0$ with $|S(h)| = k$ if and only if* $\text{range}(A^T)$ *is half k-balanced. The solutions are unique if and only* $\text{range}(A^T)$ *is strictly half k-balanced.*

Proof. For ant $x \in \Re^p$, let

$$v = A^T(x - \hat{x}).$$

Then x is feasible (i.e., $A^T x \leq b \equiv A^T \hat{x} + h$) if and only if $v_Z \leq 0$ and $v_S \leq h_S$. On the other hand, for any v with $v_Z \leq 0$, there exists a $\gamma > 0$ satisfying $\gamma v \leq h$ so that $x = \hat{x} + \gamma v$ is feasible. Now we examine the objective of $(O1+)$:

$$e^T(b - A^T x) = e^T(b - A^T \hat{x}) - e^T v,$$

which implies that $\hat{x}$ is a minimizer if and only if $e^T v \leq 0$ for all v such that $v_Z \leq 0$; namely, $e_S^T v_S \leq -e_Z^T v_Z = \|v_Z\|_1$ for all v such that $v_Z \leq 0$. Therefore, $\hat{x}$ solves $(O1+)$ (or h solves $(U1+)$) for all $h \geq 0$ with $|S(h)| = k$ if and only if $\text{range}(A^T)$ is half k-balanced. The uniqueness results also follows immediately. □

Lemma 2. *The vectors $\hat{x}$ and h solve (O1+) and (U1+), respectively, for all $h \geq 0$ with $|S(h)| = k$ if and only if* $\text{range}(B^T)$ *is half k-thick.*

Proof. The dual of $(U1+)$ is: $\max\{c^T x : B^T x \leq e\}$, where $c = Bh$. By strong duality, $y = h$ solves $(U1+)$ if and only if there exists a dual feasible x such that

$$c^T x = (Bh)^T x = h^T B^T x = h_S^T (B_S^T x) = h_S^T e_S = e^T y,$$

which occurs if and only if x satisfies $B_S^T x = e_S$ and $B_Z^T x \leq 1$; namely, $B^T x$ intersects with the $(n-k)$-face of $\{v \in \Re^n : v \leq 1\}$ defined by $v_S = e_S$. Therefore, $\hat{x}$ solves $(O1+)$ (or h solves $(U1+)$) for all h with $|S(h)| = k$ if and only if $\text{range}(B^T)$ is half k-thick. □

It is not difficult to see that the null space of A is half k-thick if and only if for any partition (S, Z) with $|S| = k$ and $e \in \Re^k$ (the vector of k ones this time)

$$\min_{v \in \Re^n} \left\{ \max_{i \in Z} v_i : Av = 0, \; v_S = e \right\} \leq 1,$$

and is strictly half k-thick if and only if the strict inequality holds.

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

$$\min_{v\in\Re^n}\left\{\max_{i\in Z} v_i : Av = 0,\ v_S = e\right\} = \max_{0\neq x\in\Re^p}\left\{\frac{e^T v_S}{\|v_Z\|_1} : v = A^T x,\ v_Z \leq 0\right\}. \tag{11}$$

As a result, range(A^T) *is (strictly) half k-balanced if and only if* range(B^T) *is half (strictly) k-thick.*

Proof. The equality (11) follows from the following calculation,

$$\begin{aligned}
& \min_{w\in\Re^{n-k}}\{\max(w) : A_Z w + A_S e = 0\} \\
= \;& \min_{w}\{\xi : A_Z w = -A_S e,\ w \leq \xi\} \\
= \;& \max_{x,u}\{(-A_S e)^T x : u = A_Z^T x,\ e^T u = 1,\ u \geq 0\} \\
= \;& \max_{x}\{(-A_S^T x)^T e : \|A_Z^T x\|_1 = 1, A_Z^T x \geq 0\} \\
= \;& \max_{x}\{(A_S^T x)^T e : \|A_Z^T x\|_1 = 1, A_Z^T x \leq 0\} \\
= \;& \max_{x\neq 0}\{e^T (A_S^T x)/\|A_Z^T x\|_1 : A_Z^T x \leq 0\},
\end{aligned}$$

where we have used an equivalent linear program in the first equality and its dual in the second. Finally, the last statement of the lemma follows readily from (11). □

Donoho and Tanner [10, Theorem 1] have shown that, for any $c = Bh$, solving $(U1+)$ recovers all $h \geq 0$ with $|S(h)| = k$ if and only if $P(B)$ is k-neighborly (we refer to their paper for more discussions on the topic of neighborliness of polytopes). The lemma below gives an equivalence result based on a straightforward argument. (A similar argument can also be applied to show the equivalence between strictly (full) k-thickness and the k-neighborliness of centrally symmetric polytopes.)

Lemma 4. *The subspace* range(B^T) *is strictly half k-thick if and only if the polytope $P(B) \subset \Re^q$ is k-neighborly.*

Proof. Let $B = [b_1\ b_2\ \cdots\ b_n] \in \Re^{q\times n}$, range$(B^T)$ be strictly half k-thick and (S, Z) be any partition with $|S| = k$. Then there is a vector $a \in \Re^q$ such that

$$b_i^T a \begin{cases} = 1, & i \in S \\ < 1, & i \in Z. \end{cases}$$

Clearly, $a^T x = 1$ is a supporting hyperplane for $P(B)$, and $\{b_i : i \in S\}$ is a vertex set that spans the face $P(B) \cap \{x \in \Re^q : a^T x = 1\}$, implying that $P(B)$ is k-neighborly. This argument can be easily reversed. □

Using above four lemmas, we have established the following equivalence recoverability results.

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 $h \geq 0$ with $|S(h)| \leq k$, $\hat{x}$ and h uniquely solve (O1+) and (U1+), respectively, if and only if one of the following three equivalent conditions holds: (1) $\text{range}(A^T) \subset \Re^n$ is strictly half k-balanced; or (2) $\text{range}(B^T) \subset \Re^n$ is strictly half k-thick; or (3) $P(B) \subset \Re^q$ is a k-neighborly polytope.*

The equivalence between recoverability and the k-neighborliness of polytopes allows one to utilize results from one side to obtain results for the other, as has been done in [10] where an equivalent result to Theorem 1 has been established by invoking the classic results on k-neighborliness of polytopes [11, 12]. In the following section, we give a proof based on the notion of strictly half k-thickness of a space rather than the k-neighborliness of a polytope. Our proof provides an alternative to Gale's charmingly simple proof [12] for the $q/2$-neighborliness of the cyclic polytope $P(B)$, slightly longer but equally elementary.

4 A Simple Proof for Theorem 1

Without loss of generality, let us assume that $q \geq 2$ is an even number (the odd-number case can be similarly treated). For any $a \in \Re^q$, define $f_q(\cdot; a) : \Re \to \Re$ to be the q-th degree polynomial (associated with a) of the form:

$$f_q(\tau; a) = \sum_{i=1}^{q} a_i \tau^i. \tag{12}$$

In light of Lemma 3, it suffices to show that for the matrix B defined in (9), $\text{range}(B^T)$ is strictly half $q/2$-thick. Since any $v \in \text{range}(B^T)$ can be written as

$$v = B^T a = [f_q(t_1; a)\ f_q(t_2; a)\ \cdots\ f_q(t_n; a)]^T$$

for some $a \in \Re^q$, $\text{range}(B^T)$ is strictly half $q/2$-thick if and only if for any partition (S, Z) with $|S| = q/2$, one can find some $a \in \Re^q$ so that

$$f_q(t_i; a) \begin{cases} = 1, & i \in S, \\ < 1, & i \in Z. \end{cases} \tag{13}$$

To determine the q unknowns in $a \in \Re^q$, we impose the following q linear equations,

$$f_q(t_i; a) = 1, \;\; f_q'(t_i; a) = 0, \;\; i \in S. \tag{14}$$

Together with the implicit condition $f_q(0; a) = 0$, these are well-known Hermite interpolation conditions that uniquely determines a polynomial of degree q of the form (12) corresponding to the coefficient vector $a = a(S)$ associated with the given partition (S, Z). With the substitution $a = a(S)$, the equality conditions in (13) are all satisfied. So we need only to verify the inequality ones. We claim that the points in $\{t_i : i \in S\}$ are all maximizers of $f_q(\tau; a(S))$. Sort these points into $t_{i_1} < t_{i_1} < \cdots < t_{i_{q/2}}$. Since at all these points the function takes the unit value 1, by the mean-value theorem, there exists a point in each interval $(t_{i_m}, t_{i_{m+1}})$, $m = 1, \cdots, q/2 - 1$, where the derivative vanishes. The total number of such in-between stationary points of $f_q(\tau; a(S))$ is $q/2 - 1$. Together with the $q/2$ points in $\{t_i : i \in S\}$, we already have all $q - 1$ stationary points for $f_q(\tau; a(S))$ — a polynomial of degree q. Moreover, each of these stationary points must have multiplicity one as a root of $f_q'(\tau; a(S))$. This implies that all points in $\{t_i : i \in S\}$ are either all minimizers or all maximizers. However, they cannot be all minimizers because

$$0 = f_q(0; a(S)) < f_q(t_i; a(S)) = 1, \ \forall i \in S.$$

Clearly, the set $\{t_i : i \in S\}$ comprises all maximizers of $f_q(\tau; a(S))$. Hence, outside of this set, $f_q(\tau; a(S)) < 1$ and the inequality in (13) holds. This establishes that range(B^T) is indeed strictly half $q/2$-thick, and completes the proof.

As has been pointed out in [10], the nonnegativity in h helps tremendously in terms how many non-zeros can be allowed in h while still guaranteeing exact recoverability. We believe that this fact will find applications in a number of computational areas.

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. Proceedings of the National Academy of Sciences (USA), 2005 v102(27): 9452-9457.

[10] David L. Donoho and Jared Tanner. Sparse Nonnegative Solutions of Underdetermined Linear Equations by Linear Programming. Proceedings of the National Academy ofv-Sciences (USA), 2005 v102(27): 9446-9451.

[11] David Gale. Neighboring vertices on a convex polyhedron. In Linear inequalities and related system, Annals of Mathematics Studies, no. 38, pages 255263. Princeton University Press, Princeton, N.J., 1956.

[12] David Gale. Neighborly and cyclic polytopes. In Proc. Sympos. Pure Math., Vol. VII, pages 225232. Amer. Math. Soc., Providence, R.I., 1963.

[13] Branko Grunbaum. Convex polytopes. Volume 221 of Graduate Texts in Mathematics. Springer-Verlag, New York, second edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Gunter M. Ziegler.

[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.

[17] Y. Zhang. A Simple Proof for Recoverability of ℓ_1-Minimization: Go Over or Under? Technical report TR05-09, Department of computational and Applied Mathematics, Rice University, Houston, TX, 2005.

[18] Gunter M. Ziegler. Lectures on polytopes. Volume 152 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1995.