The Limits of Error Correction with $l_p$ Decoding

Meng Wang  Weiyu Xu  Ao Tang
School of ECE, Cornell University, Ithaca, NY 14853, USA

Abstract—An unknown vector $f$ in $\mathbb{R}^n$ can be recovered from corrupted measurements $y = Af + e$ where $A^m \times n (m \geq n)$ is the coding matrix if the unknown error vector $e$ is sparse. We investigate the relationship of the fraction of errors and the recovering ability of $l_p$-minimization ($0 < p \leq 1$) which returns a vector $x$ minimizing the “$l_p$-norm” of $y - Ax$. We give sharp thresholds of the fraction of errors that determine the successful recovery of $f$. If $e$ is an arbitrary unknown vector, the threshold strictly decreases from 0.5 to 0.239 as $p$ increases from 0 to 1. If $e$ has fixed support and fixed signs on the support, the threshold is $\frac{2}{3}$ for all $p$ in $(0, 1)$, while the threshold is 1 for $l_1$-minimization.

I. INTRODUCTION

We consider recovering a vector $f$ in $\mathbb{R}^n$ from corrupted measurements $y = Af + e$, where $A^m \times n (m \geq n)$ is the coding matrix and $e$ is an arbitrary and unknown vector of errors. Obviously, if the fraction of the corrupted entries is too large, there is no hope of recovering $f$ from $Af + e$. However, if the fraction of corrupted measurements is small enough, one can actually recover $f$ from $y = Af + e$. As the sparsity of $e$ is represented by the $l_0$ norm, $\|e\|_0 := \{i : e_i \neq 0\}$, one natural way is to find a vector $x$ such that the number of terms where $y$ and $Ax$ differ is minimized. Mathematically, we solve the following $l_0$-minimization problem:

$$\min_{x \in \mathbb{R}^n} \|y - Ax\|_0.$$  \hspace{1cm} (1)

However, (1) is combinatorial and computationally intractable, and one commonly used approach is to solve a closely related $l_1$-minimization problem:

$$\min_{x \in \mathbb{R}^n} \|y - Ax\|_1$$ \hspace{1cm} (2)

where $\|x\|_1 := \sum_i |x_i|$. (2) can be recast as a linear program, thus can be solved efficiently. Conditions under which (2) can successfully recover $f$ have been extensively studied in the literature of compressed sensing ([11]–[6]). For example, (3) gives a sufficient condition known as the Restricted Isometry Property (RIP).

Recently, there has been great research interest in recovering $f$ by $l_p$-minimization for $p < 1$ ([7]–[11]) as follows,

$$\min_{x \in \mathbb{R}^n} \|y - Ax\|_p.$$ \hspace{1cm} (3)

Recall that $\|x\|_p := \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$ for $p > 0$. We say $f$ can be recovered by $l_p$-minimization if and only if it is the unique solution to (3). Then the question is what is the relationship between the sparsity of the error vector and the successful recovery with $l_p$-minimization? (3) is non-convex, and thus it is generally hard to compute the global minimum. However, [7] shows numerically that we can recover $f$ by finding a local minimum of (3), and $l_p$-minimization outperforms $l_1$-minimization in terms of the sparsity restriction for $e$. [9] extends RIP to $l_p$-minimization and analyzes the ability of $l_p$-minimization to recover signals from noisy measurements. [11] also provides a condition for the success recovery via $l_p$-minimization, which can be generalized to $L_1$ case. Both conditions are sufficient but not necessary, and thus are too restrictive in general.

Let $e \in \mathbb{R}^m$ be an arbitrary and unknown vector of errors on support $T = \{i : e_i \neq 0\}$. We say $e$ is $\rho m$-sparse if $|T| \leq \rho m$ for some $\rho < 1$ where $|T|$ is the cardinality of set $T$. Our main contribution is a sharp threshold $\rho^*(p)$ for all $p \leq 1$ such that for $\rho < \rho^*(p)$, if $m \geq Cn$ for some constant $C$ and the entries of $A$ are i.i.d. Gaussian, then $l_p$-minimization can recover $f$ with overwhelming probability. We provide two thresholds: one ($\rho^*$) is for the case when $e$ is an arbitrary unknown vector, and the other ($\rho^*_w$) assumes that $e$ has fixed support and fixed signs. In the latter case, the condition of successful recovery with $l_1$-minimization from any possible error vector is the same, while the condition of successful recovery with $l_p$-minimization ($p < 1$) from different error vectors differs. Using worst-case performance as criterion, we prove that though $l_p$ outperforms $l_1$ in the former case, it is not comparable to $l_1$ in the latter case. Both bounds $\rho^*$ and $\rho^*_w$ are tight in the sense that once the fraction of errors exceeds $\rho^*$ (or $\rho^*_w$), $l_p$-minimization can be made to fail with overwhelming probability. Our technique stems from [12], which only focuses on $l_1$-minimization and the case that $e$ is arbitrary.

II. RECOVERY FROM ARBITRARY ERROR VECTOR

In this section, we shall give a function $\rho^*(p)$ such that for a given $p$, for any $\rho < \rho^*(p)$, when the entries of $A$ are i.i.d. Gaussian, the $l_p$-minimization can recover $f$ with overwhelming probability as long as the error $e$ is $\rho m$-sparse.

The following theorem gives an equivalent condition for the success of $l_p$ minimization ([7], [8]).

Theorem 1 ([7], [8]). $f$ is the unique solution to the $l_p$ minimization problem ($0 < p \leq 1$) for every $f$ and for every $\rho m$-sparse $e$ if and only if

$$\sum_{i \in T} |(Az)_i|^p < \sum_{i \in \hat{T}} |(Az)_i|^p$$ \hspace{1cm} (4)

for every $z \in \mathbb{R}^n$, and every support $T$ with $|T| \leq \rho m$.

One important property is that if the condition (4) is satisfied for some $0 < p \leq 1$, then it is also satisfied for all $0 < q \leq p$.
Now we define the threshold of successful recovery $\rho^*$ as a function of $p$.

**Lemma 1.** Let $X_1, X_2, \ldots, X_m$ be i.i.d $N(0,1)$ random variables and let $Y_1, Y_2, \ldots, Y_m$ be the sorted ordering (in non-increasing order) of $|X_1|^p, |X_2|^p, \ldots, |X_m|^p$ for some $p \in (0,1]$. For a $\rho > 0$, define $S_{\rho}$ as $\sum_{i=1}^{m} Y_i$. Let $S$ denote $E[S_1]$, the expected value of $S_1$. Then there exists a constant $\rho^*(p)$ such that $\lim_{m \to \infty} \frac{E[S_{\rho^*}]}{S} = \frac{1}{2}$.

**Proof:** Let $X \sim N(0,1)$ and let $Z = |X|$. Let $f(z)$ denote the p.d.f. of $Z$ and $F(z)$ be its c.d.f. Define $g(t) = \int_{z}^{\infty} z^p f(z) dz$. $g$ is continuous and decreasing in $[0,\infty]$, and $g(0) = E[Z^p] = \frac{S}{m}, \lim_{m \to \infty} g(t) = 0$. Then there exists $z^*$ such that $g(z^*) = \frac{\rho}{2}$. We claim that $\rho^* = 1 - F(z^*)$ has the desired property.

Let $T_{i} = \sum_{i \neq Y_i} Y_i$. Then $E[T_{i}] = mg(z^*)$. Since $E[|T_{i} - S_{\rho}|]$ is bounded by $O(\sqrt{m})$, and $S = mg(0)$, thus $\lim_{m \to \infty} \frac{E[S_{\rho}]}{S} = \frac{1}{2}$.

**Proposition 1.** The function $\rho^*(p)$ is strictly decreasing in $p$ on $(0,1]$.

**Proof:** From the definition of $z^*$ and $\rho^*(p)$, we have

$$H(z^*, p) := \int_{0}^{z^*} x^p f(x) dx - \int_{z^*}^{\infty} x^p f(x) dx = 0, \quad (5)$$

and

$$\rho^* = 1 - F(z^*),$$

where $f()$ and $F()$ are the p.d.f. and c.d.f. of $|X|$, $X \sim N(0,1)$.

From the Implicit Function Theorem,

$$\frac{dz^*}{dp} = -\frac{\partial H}{\partial x} = -\frac{\int_{0}^{z^*} x^p \ln f(x) dx - \int_{z^*}^{\infty} x^p \ln f(x) dx}{2z^{p}f(z^*)}$$

From the chain rule, we know $\frac{d\rho^*}{dp} = \frac{dp^*}{dz^*} \frac{dz^*}{dp}$, thus

$$\frac{d\rho^*}{dp} = \frac{\int_{0}^{z^*} x^p \ln x f(x) dx - \int_{z^*}^{\infty} x^p \ln x f(x) dx}{2z^{p}} \quad (6)$$

Note the numerator of $\frac{d\rho^*}{dp}$ is less than 0 from $\frac{d\rho^*}{dp} < 0$.

We plot $\rho^*$ against $p$ numerically in Fig. 1. $\rho^*(p)$ goes to $\frac{1}{2}$ as $p$ tends to zero. Note that $\rho^*(1) = 0.239\ldots$, which coincides with the result in (12).

Now we proceed to prove that $\rho^*$ is the threshold of successful recovery with $l_p$ minimization for $p$ in $(0,1]$. First we state the concentration property of $S_{\rho}$ in the following lemma.

**Lemma 2.** For any $p \in (0,1]$, let $X_1, \ldots, X_m$, $Y_1, \ldots, Y_m$, $S_{\rho}$ and $S$ be as above. For any $\rho > 0$ and any $\delta > 0$, there exists a constant $c_1 > 0$ such that when $m$ is large enough, with probability at least $1 - 2e^{-c_1 m}$, $|S_{\rho} - E[S_{\rho}]| \leq 3\delta S$.

**Fig. 1.** Threshold $\rho^*$ of successful recovery with $l_p$-minimization

**Proof:** Let $X = [X_1, \ldots, X_m]^T$. If two vectors $X$ and $X'$ only differ in co-ordinate $i$, then for any $p$, $|S_{\rho}(X) - S_{\rho}(X')| \leq ||X_i||^p - ||X'_i||^p$. Thus for any $X$ and $X'$,

$$|S_{\rho}(X) - S_{\rho}(X')| \leq \sum_{i: X_i \neq X'_i} ||X_i||^p - ||X'_i||^p = \sum_{i} ||X_i||^p - ||X'_i||^p.$$  

Since $||X_i||^p - ||X'_i||^p \leq ||X_i - X'_i||^p$ for all $p \in (0,1]$,

$$|S_{\rho}(X) - S_{\rho}(X')| \leq \sum_{i} ||X_i - X'_i||^p. \quad (7)$$

From the isoperimetric inequality for the Gaussian measure $[13]$, for any set $A$ with measure at least a half, the set $A_t = \{x \in \mathbb{R}^m : d(x,A) \leq t\}$ has measure at least $1 - e^{-t^2/2}$, where $d(x,A) = \inf_{y \in A} ||x-y||_2$. Let $M_{\rho}$ be the median value of $S_{\rho} = S_{\rho}(X)$. Define set $A = \{x \in \mathbb{R}^m : S_{\rho}(x) \leq M_{\rho}\}$, then

$$Pr[d(x,A) \leq t] \geq 1 - e^{-t^2/2}.$$  

We claim that $d(x,A) \leq t$ implies that $S_{\rho}(x) \leq M_{\rho} + m^{(-1-p/2)t}p$. If $x \in A$, then $S_{\rho}(x) \leq M_{\rho}$, thus the claim holds as $m^{(-1-p/2)t}p$ is non-negative. If $x \notin A$, then there exists $x' \in A$ such that $||x - x'||_2 \leq t$. Let $v_i = 1$ for all $i$ and let $v_i = |x_i - x'_{i}|^p$. From H"older’s inequality

$$\sum_{i} ||x_i - x'_{i}||^p \leq \left( \sum_{i} |u_i|^{2/(2-p)} \right)^{1-p/2} \left( \sum_{i} |v_i|^{2/p} \right)^{p/2} \leq m^{(1-p/2)(t^2)^{p/2}} = m^{(1-p/2)t}p \quad (8)$$

From (7) and (8), $|S_{\rho}(x) - S_{\rho}(x')| \leq m^{(1-p/2)t}p$. Since $x \notin A$ and $x' \in A$, then $S_{\rho}(x) > M_{\rho} \geq S_{\rho}(x')$. Thus $S_{\rho}(x) \leq M_{\rho} + m^{(1-p/2)t}p$, which verifies our claim. Then

$$Pr[S_{\rho}(x) \leq M_{\rho} + m^{(-1-p/2)t}p] \geq Pr[d(x,A) \leq t] \geq 1 - e^{-t^2/2}.$$     \quad (9)

Similarly,

$$Pr[S_{\rho}(x) \geq M_{\rho} - m^{(-1-p/2)t}p] \geq 1 - e^{-t^2/2}. \quad (10)$$

Combining (9) and (10),

$$Pr[|S_{\rho}(x) - M_{\rho}| \leq m^{(1-p/2)t}p] \leq 2e^{-t^2/2}. \quad (11)$$
The difference of $E[S_p] - M_p$ can be bounded as follows,

$$
|E[S_p] - M_p| \leq E[|S_p - M_p|] = \int_0^\infty Pr(|S_p(x) - M_p| \geq y)dy \\
\leq \int_0^\infty 2e^{-\frac{y^2}{2}m(1-\frac{1}{2})}dy = m(1-\frac{1}{2}) \int_0^\infty 2e^{-\frac{y^2}{2}ds}
$$

Note that $c := \int_0^\infty 2e^{-\frac{1}{2}y^{2/(2+\rho)}}ds$ is a finite constant for all $p \in (0,1]$. As $p > 0$ and $m = EM[|x_i|^p]$, thus for any $\delta > 0$, $cm(1-\frac{1}{2}) < \frac{1}{2}S$ when $m$ is large enough.

Let $t = \left(\frac{1}{2}Sm\left(\frac{1}{2}-1\right)\right)^{\frac{1}{2}} = \left(\frac{1}{2}\delta E[|x_i|^p]\right)^{\frac{1}{2}}m$, from (11) with probability at least $(1-2e^{-\frac{1}{2}(\delta E[|x_i|^p])^{\frac{1}{2}}m})$, $|S_p - M_p| < \frac{1}{2}\delta S$. Thus $|S_p - E[S_p]| \leq |S_p - M_p| + |M_p - E[S_p]| < \delta S$ with probability at least $1-2e^{-c_m\delta}$ for some constant $c_1$. ■

**Corollary 1.** For any $\rho < \rho^*$, there exists a $\delta > 0$ and a constant $c_2 > 0$ such that when $m$ is large enough, with probability $1 - 2e^{-c_2m}$, $S_p \leq (\frac{1}{2} - \delta)S$.

**Proof:** When $\rho < \rho^*$,

$$
E[S_p] = E[S_{p^*}] - \sum_{i=|pm|+1}^{\rho^*m} E[|X_i|^p] \\
\leq E[S_{p^*}] - (\rho^*m - |pm|)E[|X_i|^p]
$$

Then $E[S_p]/S \leq \frac{1}{2} - 2\delta$ for a suitable $\delta$ as $S = mE[|X_i|^p]$. The result follows by combining the above with Lemma 2. ■

**Corollary 2.** For any $\epsilon > 0$, there exists a constant $c_3 > 0$ such that when $m$ is large enough, with probability $1 - 2e^{-c_3m}$, it holds that $(1-\epsilon)S \leq S_{1} \leq (1 + \epsilon)S$.

The above two corollaries indicate that with overwhelming probability the sum of the largest $|pm|$ terms of $Y_i$’s is less than half of the total sum $S_1$ if $\rho < \rho^*$. The following lemma extends the result to every vector $Ax$ where matrix $A \in \mathbb{R}^{m \times n}$ has i.i.d. Gaussian entries and $z$ is any vector in $\mathbb{R}^n$.

**Lemma 3.** For any $0 < p \leq 1$, given any $\rho < \rho^*(p)$, there exist constants $c_4, c_5 > 0$ such that when $m \geq c_4n$ and $n$ is large enough, with probability $1 - e^{-c_5n}$, $m \times n$ matrix $A$ with i.i.d. $N(0,1)$ entries has the following property: for every $z \in \mathbb{R}^n$ and every subset $T \subseteq \{1,\ldots,m\}$ with $|T| \leq \rho m$, $\sum_{i \in T} |(Az)_i|^p \geq \delta S \|z\|^p_2$.

**Proof:** For any given $\gamma > 0$, there exists a $\gamma$-net $K$ of cardinality less than $(1 + \frac{2}{\gamma})^{2n}$ (13). A $\gamma$-net $K$ is a set of points such that $\gamma^{v_k}2 = 1$ for all $v_k$ in $K$ and for any $z$ with $\|z\|_2 = 1$, there exists some $v^k$ such that $\|z - v^k\|_2 \leq \gamma$.

Since $A$ has i.i.d. $N(0,1)$ entries, then $Ax$ has $m$ i.i.d. $N(0,1)$ entries. Applying a union bound to Corollary 1 and 2 we know that for some $\delta > 0$ and for every $\epsilon > 0$, with probability $1 - 2e^{-c\delta}$ for some $c > 0$, we have

$$
S_p(Av_k) \leq \left(\frac{1}{2} - \delta\right)S
$$

and

$$
(1 - \epsilon)S \leq S_1(Av_k) \leq (1 + \epsilon)S
$$

hold for a vector $v^k$ in $K$. Taking $m = c_4n$ for large enough $c_4$, from union bound we get that (12) and (13) hold for all the points in $K$ at the same time with probability at least $1 - e^{-c_5n}$ for some $c_5 > 0$.

For any $z$ such that $\|z\|_2 = 1$, there exists $v_0$ in $K$ such that $\|z - v_0\|_2 \leq \gamma_1 \leq \gamma$. Let $z_1$ denote $z - v_0$, then $\|z_1 - \gamma_1v_1\|_2 \leq \gamma_2 \leq \gamma_1 \leq \gamma^2$ for some $v_1$ in $K$. Repeating this process, we have

$$
z = \sum_{j=0}^\infty \gamma jv_j
$$

where $\gamma_0 = 1, \gamma_j \leq \gamma^j$ and $v_j \in K$.

Thus for any $z \in \mathbb{R}^n$, we have $z = \|z\|_2 \sum_{j=0}^\infty \gamma jv_j$.

For any index set $T$ with $|T| \leq \rho m$,

$$
\sum_{i \in T} |(Az)_i|^p = \|z\|^p_2 \sum_{i \in T} |(\gamma_j Av_j)_i|^p \\
\geq \|z\|^p_2 \sum_{i \in T} \|\gamma_j Av_j\|_p^p - \gamma^p \|Av_j\|_p \\
\geq \|z\|^p_2 \sum_{i \in T} |(Av_j)_i|^p - \gamma^p \sum_{i \in T} |(Av_j)_i|^p \\
\geq \|z\|^p_2 (1 - \epsilon)S - \sum_{j \geq 1} \gamma^p (1 + \epsilon)S \\
\geq S\|z\|^p_2 \frac{1 - 2\epsilon\rho - \epsilon}{1 - \gamma^p}
$$

Thus $\sum_{i \in T} |(Az)_i|^p - \sum_{i \in T} |(Az)_i|^p \geq \delta S \|z\|^p_2$. For a given $\epsilon$, we can pick $\gamma$ and $\epsilon$ small enough such that for $\sum_{i \in T} |(Az)_i|^p - \sum_{i \in T} |(Az)_i|^p \geq \delta S \|z\|^p_2$. ■

We can now establish one main result regarding the threshold of successful recovery with $l_p$-minimization.

**Theorem 2.** For any $0 < p \leq 1$, given any $\rho < \rho^*(p)$, there exist constants $c_4, c_5 > 0$ such that when $m \geq c_4n$ and $n$ is large enough, with probability $1 - e^{-c_5n}$, $m \times n$ matrix $A$ with i.i.d. $N(0,1)$ entries has the following property: for every $f \in \mathbb{R}^n$ and every error $e$ with its support $T$ satisfying $|T| \leq \rho m$, $f$ is the unique solution to the $l_p$-minimization problem (3).

**Proof:** Lemma 3 indicates that $\sum_{i \in T^c} |(Az)_i|^p - \sum_{i \in T} |(Az)_i|^p \geq \delta S \|z\|^p_2 > 0$ for every non-zero $z$, then from
Theorem 1. \( f \) is the unique solution to the \( l_p \)-minimization problem (3).

We remark here that \( \rho^* \) is a sharp bound for successful recovery. For any \( \rho > \rho^* \), from Lemma 2 with overwhelming probability the sum of the largest \( \rho m \) terms of \( |(A z)_i|^p \)'s is more than the half of the total sum \( S_1 \), then Theorem 1 indicates that the \( l_p \)-recovery fails in this case. In fact, for any vector \( f' \neq f \), let \( z = f' - f \), and let \( T \) be the support of the largest \( \rho m \) terms of \( |(A z)_i|^p \)'s. If the error vector \( e \) agrees with \( |(A z)_i|^p \) on the support \( T \) and is zero elsewhere, then with large probability \( \|e - A z\|^p \) is no greater than that of \( \|e \|^p \) which implies that \( l_p \)-minimization cannot correctly return \( f \). Proposition 1 thus implies that the threshold strictly decreases as \( p \) increases. The performance of \( l_p \)-minimization is better than \( l_p \) for \( p_1 < p_2 \leq 1 \) in the sense that the sparsity requirement for the arbitrary error vector is less strict for smaller \( p \).

III. RECOVERY FROM ERROR VECTOR WITH FIXED SUPPORT AND SIGNS

In Section III for some \( \rho > 0 \), we call \( l_p \)-minimization successful if and only if it can recover \( f \) from any error \( e \) whose support size is at most \( \rho m \). Here we only require \( l_p \)-minimization to recover \( f \) from errors with fixed but unknown support and signs. We will provide a sharp threshold \( \rho^*_p \) of the proportion of errors below which \( l_p \)-minimization is successful.

Once the support and the signs of an error vector is fixed, the condition of successful recovery with \( l_p \)-minimization from any such error vector is the same, however, the condition of successful recovery with \( l_p \)-minimization from different error vectors differs even the support and the signs of the error is fixed. Here we consider the worst case scenario in the sense that the recovery with \( l_p \)-minimization is defined to be “successful” if \( f \) can be recovered from any such error \( e \).

We characterize this case in Theorem 3. Note that if there is further constraint on \( e \), then the condition of successful recovery with \( l_p \)-minimization may be different from the one stated in Theorem 3.

Theorem 3. Given any \( p \in (0, 1) \), for every \( f \in \mathbb{R}^n \) and every error \( e \) with fixed support \( T \) and fixed sign for each entry \( e_i, i \in T \), if \( f \) is always the unique solution to the \( l_p \)-minimization problem (3), then

\[
\sum_{i \in T^c} |(A z)_i|^p \geq \sum_{i \in T} |(A z)_i|^p
\]

for all \( z \in \mathbb{R}^n \) where \( T^c = \{ i \in T : (A z)_i e_i < 0 \} \).

Conversely, if \( f \) is always the unique solution to the \( l_p \)-minimization problem (2) provided that

\[
\sum_{i \in T^c} |(A z)_i|^p < \sum_{i \in T} |(A z)_i|^p
\]

for all non-zero \( z \in \mathbb{R}^n \).

Proof: First part. Suppose there exists \( z \) such that

\[
\sum_{i \in T^c} |(A z)_i|^p \geq \sum_{i \in T} |(A z)_i|^p, \text{ let } \delta = \sum_{i \in T^c} |(A z)_i|^p - \sum_{i \in T} |(A z)_i|^p > 0.
\]

Let \( e_i = 0 \) for every \( i \) in \( T^c \), let \( e_i = -(A z)_i \) for every \( i \) in \( T^c \). For every \( i \) in \( T^c = T - T^c \), let \( e_i \) satisfy \( (A z)_i e_i \geq 0 \). As \( p \in (0, 1) \), we can pick \( e_i \) (in \( T^c \)) with \( |e_i| \) large enough such that \( \sum_{i \in T^c} |e_i + (A z)_i|^p - \sum_{i \in T^c} |e_i|^p < \frac{\delta}{2} \). Then

\[
\|e + A z\|^p_p = \sum_{i \in T^c} |e_i + (A z)_i|^p + \sum_{i \in T} |(A z)_i|^p < \sum_{i \in T^c} |e_i|^p + \frac{\delta}{2} + \sum_{i \in T} |(A z)_i|^p < \sum_{i \in T^c} |e_i|^p + \frac{\delta}{2} + \sum_{i \in T} |(A z)_i|^p - \delta = \|e\|^p_p - \frac{\delta}{2}.
\]

Thus \( \|y - A f(z)\|^p_p = \|e + A z\|^p_p < \|e\|^p_p = \|y - A f\|^p_p \), \( f \) is not a solution to (3), which is a contradiction.

Second part. For any \( e \) on support \( T \) with fixed signs and for any \( f \), let \( y = A f + e \). For any \( x \neq f \), let \( z = f - x \), and so

\[
\|y - A z\|^p_p = \|(y - A f) + A z\|^p_p = \sum_{i \in T^c} |e_i + (A z)_i|^p + \sum_{i \in T} |e_i + (A z)_i|^p + \sum_{i \in T} |(A z)_i|^p \geq \sum_{i \in T^c} |e_i|^p + \sum_{i \in T^c} (|e_i|^p - |(A z)_i|^p) + \sum_{i \in T} |(A z)_i|^p > \|e\|^p_p.
\]

The first inequality holds as for each \( i \) in \( T^c \), \( (A z)_i \) has the same sign as that of \( e_i \) if not zero; and for \( p \in (0, 1) \), \( |e_i + (A z)_i|^p \geq |e_i|^p - |(A z)_i|^p \) holds. The second inequality comes from the assumption that \( \sum_{i \in T^c} |(A z)_i|^p < \sum_{i \in T^c} |(A z)_i|^p \). Thus

\[
\|y - A z\|^p_p > \|y - A f\|^p_p \text{ for all } x \neq f. \]

Lemma 4. Let \( X_1, X_2, \ldots, X_m \) be i.i.d. \( N(0, 1) \) random variables and \( T \) be a set of indices with size \([T] = pm \) for some \( \rho > 0 \). Let \( e \in \mathbb{R}^m \) be any vector on support \( T \) with fixed signs for each entry. If \( \rho < \rho^*_p = \frac{1}{2} \), for every \( e > 0 \), when \( m \) is large enough, with probability \( 1 - e^{-c_0 m} \) for some constant \( c_0 > 0 \), the following two properties hold:

- \( \frac{1}{2} m (\mu - \epsilon) < \sum_{i \in T, X_i < 0} |X_i|^p < \frac{1}{2} m (\mu + \epsilon) \)
- \( (1 - \rho) m (\mu - \epsilon) < \sum_{i \in T} |X_i|^p < (1 - \rho) m (\mu + \epsilon) \)

where \( \mu = E[|X|^p], X \sim N(0, 1) \).

Proof: Define a random variable \( s_i \) for each \( i \) in \( T \) that is equal to 1 if \( X_i e_i < 0 \) and equal to 0 otherwise. Then

\[
\sum_{i \in T : X_i e_i < 0} |X_i|^p = \sum_{i \in T} |X_i|^p s_i, \text{ E}[|X|^p s_i] = 0.5 \mu \text{ for every } i \text{ in } T \text{ as } X_i \sim N(0, 1). \text{ From Chernoff bound, for any } \epsilon > 0, \text{ there exist } d_1 > 0 \text{ and } d_2 > 0 \text{ such that}
\]

\[
Pr[\sum_{i \in T} |X_i|^p s_i \leq \frac{1}{2} m (\mu - \epsilon)] \leq e^{-d_2 m},
\]

\[
Pr[\sum_{i \in T} |X_i|^p s_i \geq \frac{1}{2} m (\mu + \epsilon)] \leq e^{-d_2 m}.
\]

Again from Chernoff bound, there exist some constants \( d_3 > 0, d_4 > 0 \) such that

\[
Pr[\sum_{i \in T} |X_i|^p \leq (1 - \rho) m (\mu - \epsilon)] \leq e^{-d_3 m},
\]

\[
Pr[\sum_{i \in T} |X_i|^p \geq (1 - \rho) m (\mu + \epsilon)] \leq e^{-d_4 m}.
\]
By union bound, there exists some constant $c_6 > 0$ such that the two properties stated in the lemma hold with probability at least $1 - e^{-c_6 m}$.

Lemma 3 implies that $\sum_{i \in T : X_i e_i < 0} |X_i|^p < \sum_{i \in T} |X_i|^p$ holds with large probability when $|T| = \rho m < \frac{2}{\alpha} m$. Applying the similar net argument in Section II, we can extend the result to every vector $Az$ where matrix $A^{m \times n}$ has i.i.d. Gaussian entries and $z$ is any vector in $\mathbb{R}^n$. Then we can establish the main result regarding the threshold of successful recovery with $l_p$-minimization from errors with fixed support and signs.

Theorem 4. For any $p \in (0, 1)$, given any $\rho < \frac{2}{\alpha}$, there exist constants $c_7, c_8 > 0$ such that when $m \geq c_7 n$ and $n$ is large enough, with probability $1 - e^{-c_8 n}$, $m \times n$ matrix $A$ with i.i.d. $N(0, 1)$ entries has the following property: for every $f \in \mathbb{R}^n$ and every error $e$ with fixed support $T$ satisfying $|T| \leq \rho m$ and fixed signs on $T$, $f$ is the unique solution to the $l_p$-minimization problem (3).

Proof: From lemma 4 applying similar arguments in the proof of lemma 3 we get that when $m \geq c_7 n$ and $n$ is large enough, with probability $1 - e^{-c_8 n}$ for some $c_8 > 0$,

- $\frac{1}{2} \rho m (\mu - \epsilon) < \sum_{i : X_i e_i < 0} |(Av_i)|^p < \frac{1}{2} \rho m (\mu + \epsilon)$
- $(1 - \rho) m (\mu - \epsilon) < \sum_{i \in T} |(Av_i)|^p < (1 - \rho) m (\mu + \epsilon)$

hold for all the vectors $v$ in a $\gamma$-net $K$ at the same time. For any $z \in \mathbb{R}^n$, we have $z = \|z\|_2 \sum_{j \geq 0} \gamma_j v_j$, where $\gamma_j = 1, v_j \in K$ for all $j$ and $\gamma_j \leq \gamma$.

Let $T^- = \{ i \in T : (Az)_i e_i < 0 \}$. For any $i$ in $T^-$,

$$|(Az)_i|^p = \|z\|_2^p \left( \sum_{j \geq 0} \gamma_j (Av_j)_i \right)^p \leq \|z\|_2^p \left( \sum_{j : (Av_j)_i e_i < 0} \gamma_j (Av_j)_i \right)^p \leq \|z\|_2^p \sum_{j : (Av_j)_i e_i < 0} \gamma_j^p |(Av_j)_i|^p$$

where the first inequality holds as $(Az)_i e_i < 0$. Then

$$\sum_{i \in T^-} |(Az)_i|^p \leq \|z\|_2^p \sum_{i \in T^-} \sum_{j : (Av_j)_i e_i < 0} \gamma_j^p |(Av_j)_i|^p \leq \|z\|_2^p \sum_{i \in T} \sum_{j : (Av_j)_i e_i < 0} \gamma_j^p |(Av_j)_i|^p \leq \|z\|_2^p \sum_{j \geq 0} \gamma_j^p \sum_{i \in T} |(Av_j)_i|^p < \|z\|_2^p \frac{1}{2 (1 - \gamma^p)} \rho m (\mu + \epsilon)$$

Thus $\sum_{i \in T} |(Az)_i|^p - \sum_{i \in T^-} |(Az)_i|^p > \|z\|_2^p \frac{m}{2 (1 - \gamma^p)} (1 - \frac{\gamma^p}{2} - 2 \gamma^p (1 - \rho) - \frac{\gamma}{2} (1 - \frac{\gamma}{2}))$. For any $\rho < \frac{2}{\alpha}$, we can pick $\gamma$ and $\epsilon$ small enough such that the right-hand side is positive.

The result follows by applying Theorem 3.

We remark here that $\rho^*_{m,c_8}$ is a sharp bound for successful recovery in this setup. For any $\rho > \rho^*_{m,c_8}$, from Lemma 3 with overwhelming probability that $\sum_{i \in T : X_i e_i < 0} |X_i|^p > \sum_{i \in T} |X_i|^p$, then Theorem 3 indicates that the $l_p$-recovery fails for some error vector $e$ in this case.

Surprisingly, the successful recovery threshold $\rho^*$ when fixing the support and the signs of an error vector is $\frac{2}{\alpha}$ for all $p \in (0, 1)$ and is strictly less than the threshold for $p = 1$, which is $1$ (4). Thus in this case, $l_1$-minimization has better recovery performance than that of $l_p$-minimization ($p < 1$) in terms of the sparsity requirement for the error vector. The result seems counterintuitive, however, it largely depends on the definition of successful recovery in terms of worse case performance. The condition of successful recovery via $l_1$-minimization from any error vector on the fixed support with fixed signs is the same, while the condition of $l_p$-minimization from different error vectors differs.

Acknowledgments: The authors thank anonymous reviewers for helpful comments. The research is supported by NSF under CCF-0835706.

REFERENCES

[1] D. L. Donoho and J. Tanner, “Sparse nonnegative solution of underdetermined linear equations by linear programming,” in Proc. Natl. Acad. Sci. U.S.A., vol. 102, no. 27, 2005, pp. 9446–9451.
[2] D. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, April 2006.
[3] E. Candès and T. Tao, “Decoding by linear programming,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[4] ——, “Near-optimal signal recovery from random projections: Universal encoding strategies?” IEEE Trans. Inf. Theory, vol. 52, no. 12, pp. 5400–5425, Dec. 2006.
[5] M. Stojnic, W. Xu, and B. Hassibi, “Compressed sensing - probabilistic analysis of a null-space characterization,” in Proc. ICASSP, 2008, pp. 3377–3380.
[6] J. Wright and Y. Ma, “Dense error correction via $l^1$ minimization,” Preprint, 2008.
[7] R. Chartrand, “Exact reconstruction of sparse signals via nonconvex minimization,” Signal Process. Lett., vol. 14, no. 10, pp. 707–710, 2007.
[8] ——, “Nonconvex compressed sensing and error correction,” in Proc. ICASSP, 2007.
[9] R. Saab, R. Chartrand, and O. Yilmaz, “Stable sparse approximations can fail for $l_p$ decoding,” in Proc. ICASSP, 2008.
[10] M. E. Davies and R. Gribonval, “Restricted isometry properties where $l_p$-sparse recovery can fail for $0 < p < 1$,” IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2203–2214, 2009.
[11] S. Foucart and M.-J. Lai, “Sparest solutions of underdetermined linear systems via $l_q$-minimization for $0 < q \leq 1$.” Applied and Computational Harmonic Analysis, vol. 26, no. 3, pp. 395–407, 2009.
[12] C. Dwork, F. McSherry, and K. Talwar, “The price of privacy and the limits of $l_p$ decoding,” in Proc. STOC, 2007, pp. 85–94.
[13] M. Ledoux, Ed., The Concentration of Measure Phenomenon. American Mathematical Society.
[14] D. Donoho, “High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension,” Discrete Comput. Geom., 2006.