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 $\mathbf{R}^n$ 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 $\mathbf{R}^n$ 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:

$$\min_{x\in\mathbf{R}^n} \|y - Ax\|_0. \tag{1}$$

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

$$\min_{x\in\mathbf{R}^n} \|y - Ax\|_1 \tag{2}$$

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 ([1]–[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,

$$\min_{x\in\mathbf{R}^n} \|y - Ax\|_p^p. \tag{3}$$

Recall that $\|x\|_p^p := (\sum_i |x_i|^p)$ 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 $e \in \mathbf{R}^m$ be an arbitrary and unknown vector of errors on support $T = \{i : e_i \neq 0\}$. We say e is ρm-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 (ρ^*) is for the case when e is an arbitrary unknown vector, and the other (ρ_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 ρ^* and ρ_w^* are tight in the sense that once the fraction of errors exceeds ρ^* (or ρ_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 ρm-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 l_p minimization problem $(0 < p \leq 1)$ for every f and for every ρm-sparse e if and only if*

$$\sum_{i\in T} |(Az)_i|^p < \sum_{i\in T^c} |(Az)_i|^p \tag{4}$$

for every $z \in \mathbf{R}^n$, 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 $0 < q \leq p$

([10]). Now we define the threshold of successful recovery ρ^* as a function of p.

Lemma 1. *Let X_1, X_2,...,X_m be i.i.d $N(0,1)$ random variables and let Y_1, Y_2,...,Y_m be the sorted ordering (in non-increasing order) of $|X_1|^p$, $|X_2|^p$,...,$|X_m|^p$ for some $p \in (0,1]$. For a $\rho > 0$, define S_ρ as $\sum_{i=1}^{\lceil \rho m \rceil} Y_i$. 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 $g(t) = \int_t^\infty z^p f(z)dz$. g is continuous and decreasing in $[0,\infty]$, and $g(0) = E[Z^p] = \frac{S}{m}$, $\lim_{t\to\infty} g(t) = 0$. Then there exists z^* such that $g(z^*) = \frac{g(0)}{2}$, we claim that $\rho^* = 1 - F(z^*)$ has the desired property.

Let $T_t = \sum_{i:Y_i \geq t^p} Y_i$. Then $E[T_{z^*}] = mg(z^*)$. Since $E[|T_{z^*} - S_{\rho^*}|]$ 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

$$H(z^*,p) := \int_0^{z^*} x^p f(x)dx - \int_{z^*}^\infty x^p f(x)dx = 0, \qquad (5)$$

and

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

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

From the Implicit Function Theorem,

$$\frac{dz^*}{dp} = -\frac{\frac{\partial H}{\partial p}}{\frac{\partial H}{\partial z^*}} = -\frac{\int_0^{z^*} x^p(\ln x)f(x)dx - \int_{z^*}^\infty x^p(\ln x)f(x)dx}{2z^{*p}f(z^*)}$$

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

$$\frac{d\rho^*}{dp} = \frac{\int_0^{z^*} x^p(\ln x)f(x)dx - \int_{z^*}^\infty x^p(\ln x)f(x)dx}{2z^{*p}} \qquad (6)$$

Note the numerator of (6) is less than 0 from (5), thus $\frac{d\rho^*}{dp} < 0$.

■

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

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

Lemma 2. *For any $p \in (0,1]$, let X_1,...,X_m, Y_1,...,Y_m, S_ρ 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 $1 - 2e^{-c_1 m}$, $|S_\rho - E[S_\rho]| \leq \delta S$.*

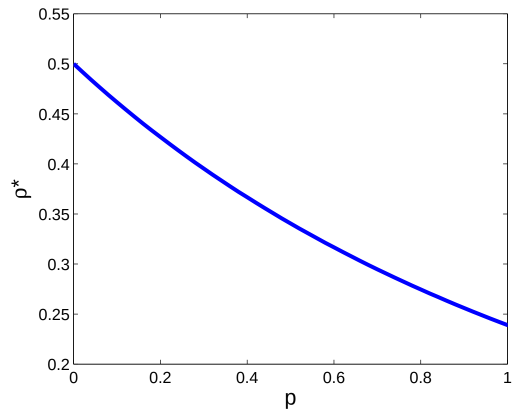

Fig. 1. Threshold ρ^* of successful recovery with l_p-minimization

Proof: Let $X = [X_1, ..., X_m]^T$. 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',

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

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

$$|S_\rho(X) - S_\rho(X')| \leq \sum_i |X_i - X_i'|^p. \qquad (7)$$

From the isoperimetric inequality for the Gaussian measure ([13]), for any set A with measure at least a half, the set $A_t = \{x \in \mathbf{R}^m : d(x,A) \leq t\}$ has measure at least $1 - e^{-t^2/2}$, where $d(x,A) = \inf_{y\in A} \|x-y\|_2$. Let M_ρ be the median value of $S_\rho = S_\rho(X)$. Define set $A = \{x \in \mathbf{R}^m : S_\rho(x) \leq M_\rho\}$, 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 $u_i = 1$ for all i and let $v_i = |x_i - x_i'|^p$. From Hölder's inequality

$$\begin{aligned}\sum_i |x_i - x_i'|^p &\leq \left(\sum_i |u_i|^{2/(2-p)}\right)^{1-p/2} \left(\sum_i |v_i|^{2/p}\right)^{p/2} \\ &\leq m^{(1-p/2)}(t^2)^{p/2} = m^{(1-p/2)}t^p \qquad (8)\end{aligned}$$

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

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

Similarly,

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

Combining (9) and (10),

$$Pr[|S_\rho(x) - M_\rho| \geq m^{(1-p/2)}t^p] \leq 2e^{-t^2/2}. \qquad (11)$$

The difference of $E[S_\rho]$ and M_ρ can be bounded as follows,

$$\begin{aligned}|E[S_\rho]-M_\rho| &\le E[|S_\rho - M_\rho|]\\ &= \int_0^\infty Pr[|S_\rho(x)-M_\rho|\ge y]dy\\ &\le \int_0^\infty 2e^{-\frac{1}{2}y^{\frac{2}{p}}m^{(1-\frac{2}{p})}}dy\\ &= m^{(1-\frac{p}{2})}\int_0^\infty 2e^{-\frac{1}{2}s^{\frac{2}{p}}}ds\end{aligned}$$

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

Let $t=\left(\frac{1}{2}\delta S m^{(\frac{p}{2}-1)}\right)^{\frac{1}{p}}=(\frac{1}{2}\delta E[|x_i|^p])^{\frac{1}{p}}\sqrt{m}$, from (11) with probability at least $(1-2e^{-\frac{1}{2}(\frac{1}{2}\delta E[|x_i|^p])^{\frac{2}{p}}m})$, $|S_\rho - M_\rho|<\frac{1}{2}\delta S$. Thus $|S_\rho - E[S_\rho]|\le |S_\rho - M_\rho|+|M_\rho - E[S_\rho]|<\delta S$ with probability at least $1-2e^{-c_1 m}$ 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 $1-2e^{-c_2 m}$, $S_\rho\le(\frac{1}{2}-\delta)S$.*

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

$$\begin{aligned}E[S_\rho] &= E[S_{\rho^*}]-\sum_{i=\lceil\rho m\rceil+1}^{\lceil\rho^* m\rceil}E[|X_i|^p]\\ &\le E[S_{\rho^*}]-(\lceil\rho^* m\rceil-\lceil\rho m\rceil)E[|X_i|^p]\end{aligned}$$

Then $E[S_\rho]/S\le\frac{1}{2}-2\delta$ for a suitable δ 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 $1-2e^{-c_3 m}$, it holds that $(1-\epsilon)S\le S_1\le(1+\epsilon)S$.*

The above two corollaries indicate that with overwhelming probability the sum of the largest $\lceil\rho m\rceil$ 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 Az where matrix $A^{m\times n}$ has i.i.d. Gaussian entries and z is any vector in $\mathbf{R}^n$.

Lemma 3. *For any $0<p\le 1$, given any $\rho<\rho^*(p)$, there exist constants c_4, c_5, $\delta>0$ such that when $m\ge c_4 n$ and n is large enough, with probability $1-e^{-c_5 n}$, an $m\times n$ matrix A with i.i.d. $N(0,1)$ entries has the following property: for every $z\in\mathbf{R}^n$ and every subset $T\subseteq\{1,...,m\}$ with $|T|\le\rho m$, $\sum_{i\in T^c}|(Az)_i|^p-\sum_{i\in T}|(Az)_i|^p\ge\delta S\|z\|_2^p$.*

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

Since A has i.i.d $N(0,1)$ entries, then Av^k 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 $1-2e^{-cm}$ for some $c>0$, we have

$$S_\rho(Av^k)\le(\frac{1}{2}-\delta)S \tag{12}$$

and

$$(1-\epsilon)S\le S_1(Av^k)\le(1+\epsilon)S \tag{13}$$

hold for a vector v^k in K. Taking $m=c_4 n$ 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 $1-e^{-c_5 n}$ for some $c_5>0$.

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

$$z=\sum_{j\ge 0}\gamma_j v_j$$

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

Thus for any $z\in\mathbf{R}^n$, we have $z=\|z\|_2\sum_{j\ge0}\gamma_j v_j$.

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

$$\begin{aligned}\sum_{i\in T}|(Az)_i|^p &= \|z\|_2^p\sum_{i\in T}|(\sum_{j\ge0}\gamma_j Av_j)_i|^p\\ &\le \|z\|_2^p\sum_{i\in T}\sum_{j\ge0}\gamma^{jp}|(Av_j)_i|^p\\ &= \|z\|_2^p\sum_{j\ge0}\gamma^{jp}\sum_{i\in T}|(Av_j)_i|^p\\ &\le S\|z\|_2^p\frac{1-2\delta}{2(1-\gamma^p)}\end{aligned}$$

$$\begin{aligned}\sum_i|(Az)_i|^p &= \|z\|_2^p\sum_i|(\sum_{j\ge0}\gamma_j Av_j)_i|^p\\ &\ge \|z\|_2^p\sum_i(|(Av_0)_i|^p-\sum_{j\ge1}\gamma_j^p|(Av_j)_i|^p)\\ &\ge \|z\|_2^p(\sum_i|(Av_0)_i|^p-\sum_{j\ge1}\gamma^{jp}\sum_i|(Av_j)_i|^p)\\ &\ge \|z\|_2^p((1-\epsilon)S-\sum_{j\ge1}\gamma^{jp}(1+\epsilon)S)\\ &\ge S\|z\|_2^p\frac{1-2\gamma^p-\epsilon}{1-\gamma^p}\end{aligned}$$

Thus $\sum_{i\in T^c}|(Az)_i|^p-\sum_{i\in T}|(Az)_i|^p\ge S\|z\|_2^p\frac{2\delta-2\gamma^p-\epsilon}{1-\gamma^p}$. For a given δ, we can pick γ and ϵ small enough such that $\sum_{i\in T^c}|(Az)_i|^p-\sum_{i\in T}|(Az)_i|^p\ge\delta S\|z\|_2^p$. ■

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

Theorem 2. *For any $0<p\le1$, given any $\rho<\rho^*(p)$, there exist constants c_4, $c_5>0$ such that when $m\ge c_4 n$ and n is large enough, with probability $1-e^{-c_5 n}$, an $m\times n$ matrix A with i.i.d. $N(0,1)$ entries has the following property: for every $f\in\mathbf{R}^n$ and every error e with its support T satisfying $|T|\le\rho m$, f is the unique solution to the l_p-minimization problem (3).*

Proof: Lemma 3 indicates that $\sum_{i\in T^c}|(Az)_i|^p-\sum_{i\in T}|(Az)_i|^p\ge\delta S\|z\|_2^p>0$ 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 ρ^* is a sharp bound for successful recovery. For any $\rho > \rho^*$, from Lemma 2, with overwhelming probability the sum of the largest $\lceil \rho m \rceil$ terms of $|(Az)_i|^p$'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 $\lceil \rho m \rceil$ terms of $|(Az)_i|^p$'s. If the error vector e agrees with $|(Az)_i|^p$ on the support T and is zero elsewhere, then with large probability $\|e - Az\|_p^p$ is no greater than that of $\|e\|_p^p$, 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_1}-minimization is better than l_{p_2}-minimization 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 II, 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 ρm. 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 ρ_w^* 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_1-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 $f \in \mathbf{R}^n$ 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 l_p-minimization problem (3), then*

$$\sum_{i \in T^-} |(Az)_i|^p \leq \sum_{i \in T^c} |(Az)_i|^p$$

for all $z \in \mathbf{R}^n$ where $T^- = \{i \in T : (Az)_i e_i < 0\}$.

Conversely, f is always the unique solution to l_p-minimization problem (3) provided that

$$\sum_{i \in T^-} |(Az)_i|^p < \sum_{i \in T^c} |(Az)_i|^p$$

for all non-zero $z \in \mathbf{R}^n$.

Proof: First part. Suppose there exists z such that $\sum_{i \in T^-} |(Az)_i|^p > \sum_{i \in T^c} |(Az)_i|^p$, let $\delta = \sum_{i \in T^-} |(Az)_i|^p - \sum_{i \in T^c} |(Az)_i|^p > 0$.

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

$$\begin{aligned}
\|e + Az\|_p^p &= \sum_{i \in T^-} 0 + \sum_{i \in T^+} |e_i + (Az)_i|^p + \sum_{i \in T^c} |(Az)_i|^p \\
&< \sum_{i \in T^+} |e_i|^p + \frac{\delta}{2} + \sum_{i \in T^c} |(Az)_i|^p \\
&= \sum_{i \in T^+} |e_i|^p + \frac{\delta}{2} + \sum_{i \in T^-} |(Az)_i|^p - \delta \\
&= \|e\|_p^p - \frac{\delta}{2}.
\end{aligned}$$

Thus $\|y - A(f-z)\|_p^p = \|e + Az\|_p^p < \|e\|_p^p = \|y - Af\|_p^p$, 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 $y = Af + e$. For any $x \neq f$, let $z = f - x$, and so

$$\begin{aligned}
& \|y - Ax\|_p^p = \|(y - Af) + Az\|_p^p \\
= & \sum_{i \in T^+} |e_i + (Az)_i|^p + \sum_{i \in T^-} |e_i + (Az)_i|^p + \sum_{i \in T^c} |(Az)_i|^p \\
\geq & \sum_{i \in T^+} |e_i|^p + \sum_{i \in T^-} (|e_i|^p - |(Az)_i|^p) + \sum_{i \in T^c} |(Az)_i|^p \\
> & \|e\|_p^p.
\end{aligned}$$

The first inequality holds as for each i in T^+, $(Az)_i$ has the same sign as that of e_i if not zero; and for $p \in (0,1)$, $|e_i + (Az)_i|^p \geq |e_i|^p - |(Az)_i|^p$ holds. The second inequality comes from the assumption that $\sum\limits_{i \in T^-} |(Az)_i|^p < \sum\limits_{i \in T^c} |(Az)_i|^p$. Thus $\|y - Ax\|_p^p > \|y - Af\|_p^p$ for all $x \neq f$. ∎

Lemma 4. *Let X_1, X_2,...,X_m be i.i.d. $N(0,1)$ random variables and T be a set of indices with size $|T| = \rho m$ for some $\rho > 0$. Let $e \in \mathbf{R}^m$ be any vector on support T with fixed signs for each entry. If $\rho < \rho_w^* = \frac{2}{3}$, for every $\epsilon > 0$, when m is large enough, with probability $1 - e^{-c_6 m}$ for some constant $c_6 > 0$, the following two properties hold:*

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

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 $\sum_{i \in T: X_i e_i < 0} |X_i|^p = \sum_{i \in T} |X_i|^p s_i$. $E[|X_i|^p s_i] = \frac{1}{2}\mu$ for every i in T as $X_i \sim N(0,1)$. From Chernoff bound, for any $\epsilon > 0$, there exist $d_1 > 0$ and $d_2 > 0$ such that

$$Pr[\textstyle\sum_{i \in T} |X_i|^p s_i \leq \frac{1}{2}\rho m(\mu - \epsilon)] \leq e^{-d_1 m},$$
$$Pr[\textstyle\sum_{i \in T} |X_i|^p s_i \geq \frac{1}{2}\rho m(\mu + \epsilon)] \leq e^{-d_2 m}.$$

Again from Chernoff bound, there exist some constants $d_3 > 0$, $d_4 > 0$ such that

$$Pr[\textstyle\sum_{i \in T^c} |X_i|^p \leq (1-\rho)m(\mu - \epsilon)] \leq e^{-d_3 m},$$
$$Pr[\textstyle\sum_{i \in T^c} |X_i|^p \geq (1-\rho)m(\mu + \epsilon)] \leq e^{-d_4 m}.$$

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 4 implies that $\sum_{i \in T: X_i e_i < 0} |X_i|^p < \sum_{i \in T^c} |X_i|^p$ holds with large probability when $|T| = \rho m < \frac{2}{3}m$. 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 $\mathbf{R}^n$. 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 $\rho < \frac{2}{3}$, there exist constants c_7, $c_8 > 0$ such that when $m \geq c_7 n$ and n is large enough, with probability $1 - e^{-c_8 n}$, an $m \times n$ matrix A with i.i.d. $N(0,1)$ entries has the following property: for every $f \in \mathbf{R}^n$ 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 $c_8 > 0$,

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

hold for all the vectors v in a γ-net K at the same time. Moreover, for any $z \in \mathbf{R}^n$, we have $z = \|z\|_2 \sum_{j \geq 0} \gamma_j v_j$, where $\gamma_0 = 1$, $v_j \in K$ for all j and $\gamma_j \leq \gamma^j$.

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

$$\begin{aligned} |(Az)_i|^p &= \|z\|_2^p \big|(\sum_{j \geq 0} \gamma_j A v_j)_i\big|^p \\ &\leq \|z\|_2^p \big|(\sum_{j:(Av_j)_i e_i < 0} \gamma_j A v_j)_i\big|^p \\ &\leq \|z\|_2^p \sum_{j:(Av_j)_i e_i < 0} \gamma^{jp} |(Av_j)_i|^p \end{aligned}$$

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

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

$$\begin{aligned} &\sum_{i \in T^c} |(Az)_i|^p = \|z\|_2^p \sum_{i \in T^c} |(\sum_{j \geq 0} \gamma_j A v_j)_i|^p \\ \geq\ & \|z\|_2^p \big(\sum_{i \in T^c} |(Av_0)_i|^p - \sum_{j \geq 1} \gamma^{jp} \sum_{i \in T^c} |(Av_j)_i|^p \big) \\ >\ & \|z\|_2^p \big((1-\rho)m(\mu - \epsilon) - \sum_{j \geq 1} \gamma^{jp} (1-\rho)m(\mu + \epsilon) \big) \\ \geq\ & \|z\|_2^p (1-\rho)m \frac{\mu - 2\mu\gamma^p - \epsilon}{1 - \gamma^p} \end{aligned}$$

Thus $\sum_{i \in T^c} |(Az)_i|^p - \sum_{i \in T^-} |(Az)_i|^p > \|z\|_2^p \frac{m\mu}{1-\gamma^p} \big(1 - \frac{3}{2}\rho - 2\gamma^p(1-\rho) - \frac{\epsilon}{\mu}(1 - \frac{\rho}{2})\big)$. For any $\rho < \frac{2}{3}$, we can pick γ and ϵ small enough such that the righthand side is positive. The result follows by applying Theorem 3.

■

We remark here that ρ_w^* is a sharp bound for successful recovery in this setup. For any $\rho > \rho_w^*$, from Lemma 4, with overwhelming probability that $\sum_{i \in T: X_i e_i < 0} |X_i|^p > \sum_{i \in T^c} |X_i|^p$, then Theorem 3 indicates that the l_p-recovery fails for some error vector e in this case.

Surprisingly, the successful recovery threshold ρ^* when fixing the support and the signs of an error vector is $\frac{2}{3}$ for all p in $(0,1)$ and is strictly less than the threshold for $p = 1$, which is 1 ([14]). 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. 5406–5425, 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 via nonconvex optimization," in *Proc. ICASSP*, 2008.

[10] M. E. Davies and R. Gribonval, "Restricted isometry constants where l_p sparse recovery can fail for $0 < p \leq 1$," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2203–2214, 2009.

[11] S. Foucart and M.-J. Lai, "Sparsest 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 lp 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.