Restricted p-isometry property and its application for nonconvex compressive sensing *

Yi Shen[†] and Song Li[‡]

Abstract

Compressed sensing is a new scheme which shows the ability to recover sparse signal from fewer measurements, using l_1 minimization. Recently, Chartrand and Staneva shown in [7] that the l_p minimization with $0 < p < 1$ recovers sparse signals from fewer linear measurements than does the l_1 minimization. They proved that l_p minimization with $0 < p < 1$ recovers S-sparse signals $x \in \mathbb{R}^N$ from fewer Gaussian random measurements for some smaller p with probability exceeding

$$1 - 1/\binom{N}{S}.$$

The first aim of this paper is to show that above result is right for the case of random,Gaussian measurements with probability exceeding $1 - 2e^{-c(p)M}$, where M is the numbers of rows of random, Gaussian measurements and $c(p)$ is a positive constant that guarantees $1-2e^{-c(p)M} > 1-1/\binom{N}{S}$ for p smaller. The second purpose of the paper is to show that under certain weaker conditions, decoders $\triangle_p$ are stable in the sense that they are $(2,p)$ instance optimal for a large class of encoder for $0 < p < 1$.

Keywords. Compressed sensing, l_p minimization, instance optimality, stability

1 Introduction

Compressive sensing is a new type of sampling theory, that predicts that sparse signals and images can be reconstructed from what was previously believed to be incomplete information [2, 3, 10]. The key idea is that the sparsity helps in isolating the original vector. Given a vector $x \in \mathbb{R}^N$ and a matrix $\Psi \in \mathbb{R}^{M\times N}$ with $M \ll N$, we are interested in sparse solutions to the equation $y = \Psi x$, where Ψ is called *measurement matrix*. We also need a decoder $\triangle$ which produces $\triangle(\Psi x) \in \mathbb{R}^N$ and should be an approximation to x. The approach would be to solve the following optimization problems:

$$\triangle_0(y) := \arg\min_x \|x\|_0, \quad \text{subject to } \Psi x = y, \tag{1.1}$$

*This work is supported by NSF of China under grant numbers 10771190, 10971189 and Zhejiang Provincial NSF of China under grants number Y6090091

[†]Center of Mathematical Sciences Zhejiang University Hangzhou 310027, China

[‡]Corresponding author: songli@zju.edu.cn, Department of Mathematics, Zhejiang University Hangzhou, 310027, China

where $\|x\|_0$ denotes the numbers of non-zero elements of x. Thus, if $\|x\|_0 \leq S$, vector x is called S-sparse. Let $\Psi \in \mathbb{R}^{M\times N}$, under certain conditions on the matrix Ψ [12], there is a decoder $\triangle$ which satisfies $\triangle(\Psi x) = x$ for all $\|x\|_0 < S$

$$\text{when } M > 2S \tag{1.2}$$

with probability 1. However, because $\|x\|_0$ is non-convex and combinatorial, solving it directly is NP-hard, (1.1) is impractical for real applications. A practical alternative is

$$\triangle_1(y) := \arg\min_x \|x\|_1, \quad \text{subject to } \Psi x = y. \tag{1.3}$$

It was shown in [4] that, when $\Psi \in \mathbb{R}^{M\times N}$ is an Gaussian random matrix, then every S-sparse vector can be recovered by solving (1.3)

$$\text{when } M \geq c_1 S \log(N/S) \tag{1.4}$$

with probability exceeding $1 - 2e^{-c_2 M}$, where c_1 and c_2 are some positive constants.

The l_0 norm is naturally related to the l_p norms with $0 < p < 1$; all are measures of sparsity and, in fact the l_0 norm is the limit as $p \to 0$ of the l_p norms in the following sense:

$$\|x\|_0 = \lim_{p\to 0} \|x\|_p^p = \lim_{p\to 0} \sum_k |x_k|^p.$$

This motivates us to consider the l_p-minimization problem with $0 < p < 1$:

$$\triangle_p(y) := \arg\min_x \|x\|_p, \quad \text{subject to } \Psi x = y, \tag{1.5}$$

where the l_p norm, is defined by $\|u\|_p = (\sum |u_i|^p)^{1/p}$. It is well known that the l_p norm with $0 < p < 1$ is not actually a norm and is only a quasi-norm. For any fixed $0 < p < 1$, checking the global minimal value of (1.5) is NP-hard; but computing a local minimizer of the problem is polynomial time doable [13]. Therefore, to solve (1.5) is still much faster than (1.1) at least locally. There are some papers to be devoted to studying the l_p-minimization problem with $0 < p < 1$ [14, 15, 16]. Furthermore, the l_p minimization strategy offer some theoretical advantages. The fewer measurements are required for exact reconstruction with $0 < p < 1$ than when $p = 1$ that was demonstrated by numerical experiments in [6] and theorems in [5]. Chartarand and Staneva [7] shown that when there exist Gaussian random matrices $\Psi \in \mathbb{R}^{M\times N}$ with only

$$M \geq C_1(p)S + pC_2(p)S\log(N/S) \tag{1.6}$$

rows such that for all S-sparse vector x, we can recover x exactly from Ψx by (1.5) with probability exceeding

$$1 - 1/\binom{N}{S}.$$

It is important to note that $pC_2(p)$ goes to zero as p goes to zero. The condition in (1.6) with various p fills in the gaps between the (1.2) and (1.4). This gives a wide range of choices of l_p minimization that meets various demands for approximation power, the sparse of x and the measurement matrices Ψ in applications. However, the probability of success is polynomial decay with N and do not match the case $p = 0$ and $p = 1$. This motivated us to give the following results.

Theorem 1.1. *Let Ψ be an $M \times N$ matrix whose entries are i.i.d. random distributed normally with mean zero and variance σ^2, where $M < N$. Then there exist constants $C_1(p)$ and $C_2(p)$ such that whenever $0 < p \leq 1$ and $M \geq C_1(p)S + pC_2(p)S\log(N/S)$, the following is true with probability exceeding $1 - 2e^{-c(p)M}$: for any S-sparse $x \in \mathbb{R}^N$, x is the unique solution of (1.5), where $c(p)$ is a positive constant that guarantees $1 - 2e^{-c(p)M} > 1 - 1/\binom{N}{S}$.*

Remark 1.2. *The constants $C_1(p)$ and $C_2(p)$ are given in (2.10). In section 2, we show that decreasing p for (1.5) to successful S-sparse recovery with increasing probability. The limiting case of Theorem 1.1 as $p \to 0$ is essentially the l_0 case.*

Compressive sensing is based on the empirical observation that many types of real-world signals and images have a sparse expansion in terms of a suitable basis or frame, for instance a wavelet expansion. This means that the expansion has only a small number of significant terms, or in other words, that the coefficient vector can be well-approximated with one having only a small number of nonvanishing entries. It is an important problem to compare the performance of the best S-term approximation with compressive sampling. The best S-term approximation error of a vector $x \in \mathbb{R}^N$ in l_p is defined as

$$\sigma_S(x)_p = \inf_{\|z\|_0 \leq S} \|x - z\|_p.$$

If $\sigma_S(x)_p$ decays quickly in S, then x is called compressible. Indeed, in order to compress x, one may simply store only the S largest entries. When reconstructing x from its compressed version, the reconstruction error is $\sigma_S(x)_p$. We say that an measurement-decoder pair $(\Psi, \triangle)$ is (q, p) *instance optimal* of order S with constant C if

$$\|x - \triangle(\Psi x)\|_q \leq C\frac{\sigma_S(x)_p}{S^{1/p-1/q}}.$$

The instance optimal evaluate the efficiency of given measurement-decoder pair $(\Psi, \triangle)$. In this paper, we also discuss the instance optimal and stability of l_p minimization with respect to the measurement error and how those results depend on the measurement matrix Ψ.

Theorem 1.3. *Let $0 < p \leq 1$ and Ψ be an $M \times N$ matrix whose entries are i.i.d. random distributed normally with mean zero and variance σ^2, where $M < N$. Suppose that $N \geq M[\log(M)]^2$. There exist constants $c_1, c_2 > 0$ such that for all $S \in \mathbb{N}$ with*

$$M \geq C_1(p)S + pC_2(p)S\log(N/S),$$

where the constant $C_1(p)$ and $C_2(p)$ are given in (2.10).

The following are true.

- *There exists Ω_1 with $P(\Omega_1) \geq 1 - 3e^{-c_1 M}$ such that for all $\omega \in \Omega_1$*

$$\|\triangle_p(\Psi x + r) - x\|_2 \leq C\left(\|r\|_2 + \frac{\sigma_S(x)_p}{S^{1/p-1/2}}\right) \tag{1.7}$$

  for any $x \in \mathbb{R}^N$ and for any $r \in \mathbb{R}^M$.

- *For any $x \in \mathbb{R}^N$, there exists Ω_x with $P(\Omega_x) \geq 1 - 4e^{-c_2 M}$ such that for all $\omega \in \Omega_x$*

$$\|\triangle_p(\Psi x + r) - x\|_2 \leq C\left(\|r\|_2 + \sigma_S(x)_2\right)$$

  for any $r \in \mathbb{R}^M$.

Remark 1.4. *In Theorem 1.3, we prove that the condition $M \geq c_1 S \log(N/S)$ of Theorem 2.14 in [16] can be replaced by the weak condition in (1.6). In fact, we give a positive answer to the open question on the l_p minimization proposed in [16]: the decoder $\triangle_p$ possess stability and instance optimality properties similar to those of $\triangle_1$, and these are obtained under weaker conditions on the measurement matrices that the analogous ones with $p = 1$.*

The paper is organized as follows. In section 2, we determine how many Gaussian random matrix are sufficient for the l_p minimization with high probability for $0 < p \leq 1$. We also show that the restricted p-isometry property guarantee $(2, p)$ instance optimal. In section 3, following the line of [16] and [17], we prove that for the fewer range of dimensions as for decoding by l_1 minimization, l_p minimization is also $(2, p)$ instance optimal in probability for $0 < p \leq 1$.

2 Instance Optimality

In [11], Candès and Tao introduced the notion of *restricted isometry constants* (RIC(S, δ)) of a matrix, which was shown to play a fundamental role in compressive sensing. Let Ψ be an $M \times N$ matrix, where $M < N$ and S is a positive integer, we define the constant δ_S to be the smallest positive number such that

$$(1 - \delta_S)\|x\|_2^2 \leq \|\Psi x\|_2^2 \leq (1 + \delta_S)\|x\|_2^2, \quad \forall\ \|x\|_0 \leq S. \tag{2.1}$$

Chartrand and Staneva generalized *restricted isometry constants* of a matrix and considered the *restricted p-isometry constant* (p-RIC(S, δ)) of a matrix Ψ. Given $0 < p \leq 1$, let S be a positive integer. Then the *p-restricted isometry constants* δ_S is denoted to be the smallest positive number such that

$$(1 - \delta_S)\|x\|_2^p \leq \|\Psi x\|_p^p \leq (1 + \delta_S)\|x\|_2^p, \quad \forall\ \|x\|_0 \leq S. \tag{2.2}$$

A similar definition in the case of $p = 1$ appeared in [11], and is related to the Banach-Mazur distance of Banach space theory. The restricted isometry property implies instance optimal in l_1 [8], as well as by l_p minimization for $0 < p < 1$ [14, 16]. Our next theorem shows that p-restricted isometry property implies also a bound on the reconstruction error in l_2 and l_p, respectively.

Theorem 2.1. *Let $0 < p \leq 1$. Suppose the matrix Ψ satisfies the p-RIC$(S + L, \delta)$ with $\gamma := \frac{1+\delta_L}{1-\delta_{L+S}}(S/L)^{1-p/2} < 1$. Then*

$$\|x - \triangle_p \Psi(x)\|_p^p \leq \frac{2 + 2\gamma}{1 - \gamma}\sigma_S(x)_p^p$$

and

$$\|x - \triangle_p \Psi(x)\|_2^p \leq S^{p/2-1} \frac{2+2\gamma}{1-\gamma} \sigma_S(x)_p^p$$

for all $x \in \mathbb{R}^N$*.*

Proof. Let $x^* \in \mathbb{R}^N$ be a solution of (1.5). Denote $h = x - x^*$. Let T_0 denote the set of indices of the largest S entries of h, T_1 the next L largest, and so on. For $j > 0$, consider $i \in T_j$ and $i' \in T_{j+1}$. Since $|h_i| \leq |h_{i'}|$, we derive that $|h_i|^p \leq L^{-1} \|h_{T_{j+1}}\|_p^p$. It follows that

$$\|h_{T_{j+1}}\|_2^p \leq \|h_{T_j}\|_2^2 \leq L^{p/2-1} \|h_{T_j}\|_p^p.$$

The above inequality easily implies

$$\sum_{j>1} \|h_{T_j}\|_2^p \leq L^{p/2-1} \left(\sum_{j>0} \|h_{T_j}\|_p^p \right) = L^{p/2-1} \|h_{T_0^c}\|_p^p, \tag{2.3}$$

where T_0^c denoting the complement of T_0 in $\{1, \dots, N\}$.

Note that $\Psi h = 0$, we have

$$0 = \|\Psi h\|_p^p = \left\| \Psi h_{T_{01}} + \sum_{j>1} \Psi h_{T_j} \right\|_p^p \geq \|\Psi h_{T_{01}}\|_p^p - \sum_{j>1} \|\Psi h_{T_j}\|_p^p,$$

where $T_{01} := T_0 \cup T_1$. From p-RIP, and the above inequality, the following sequence of inequalities is deduced,

$$(1 - \delta_{L+S}) \|h_{T_{01}}\|_2^p \leq (1 + \delta_L) \sum_{j>1} \|h_{T_j}\|_2^p. \tag{2.4}$$

Combining the estimates (2.3) and (2.4), we obtain

$$\|h_{T_0}\|_p^p \leq S^{1-p/2} \|h_{T_0}\|_2^p \leq \frac{1+\delta_L}{1-\delta_{L+S}} S^{1-p/2} L^{p/2-1} \|h_{T_0^c}\|_p^p = \gamma \|h_{T_0^c}\|_p^p. \tag{2.5}$$

Because x^* is a minimizer of (1.5), we have $\|x^*\|_p^p \leq \|x\|_p^p$. This, together with $h = x - x^*$, implies that

$$\|x_{T_0}\|_p^p - \|h_{T_0}\|_p^p + \|h_{T_0^c}\|_p^p - \|x_{T_0^c}\|_p^p \leq \|x^*_{T_0}\|_p^p + \|x^*_{T_0^c}\|_p^p \leq \|x_{T_0}\|_p^p + \|x_{T_0^c}\|_p^p. \tag{2.6}$$

Since $\|x_{T_0}\|_q^q = \sigma_S(x)_p^p$, (2.6) yields

$$\|h_{T_0^c}\|_p^p \leq 2\sigma_S(x)_p^p + \|h_{T_0}\|_p^p.$$

The use of this estimate, combined with the inequality (2.5), gives

$$\|h_{T_0^c}\|_p^p \leq \frac{2}{1-\gamma} \sigma_S(x)_p^p. \tag{2.7}$$

Consequently,

$$\|x-x^*\|_p^p = \|h\|_p^p = \|h_{T_0}\|_p^p + \|h_{T_0^c}\|_p^p \le (1+\gamma)\|h_{T_0^c}\|_p^p = \frac{2+2\gamma}{1-\gamma}\sigma_S(x)_p^p.$$

Finally, we use (2.3), (2.5) and (2.7) to conclude that

$$\begin{aligned}\|x-x^*\|_2^p = \|h\|_2^p &\le \|h_{T_0}\|_2^p + \|h_{T_0^c}\|_2^p \\ &\le L^{p/2-1}\|h_{T_0^c}\|_p^p + S^{p/2-1}\gamma\|h_{T_0^c}\|_p^p \\ &\le S^{p/2-1}\frac{2+2\gamma}{1-\gamma}\sigma_S(x)_p^p.\end{aligned}$$

□

Let Ψ be an $M\times N$ matrix whose entries are i.i.d Gaussian random variables with expectation 0 and variance σ^2. To prove Theorem 1.1, we need to determine how large M must be for $\gamma<1$ to hold with high probability. For a given p, let $\mu_p := \sigma^p 2^{p/2}\Gamma(\frac{p+1}{2})/\sqrt{\pi}$. The following result in [7] will be useful:

Lemma 2.2. *[7, Lemma 3.3.] Let $0<p\le 1$, Φ an $M\times L$ submatrix of Ψ. Let $\delta>0$. Choose $\eta,\epsilon>0$ such that $\frac{\eta+\epsilon^p}{1-\epsilon^p}\le\delta$. Then*

$$M\mu_p(1-\delta)\|x\|_2^p \le \|\Phi x\|_p^p \le M\mu_p(1+\delta)\|x\|_2^p \tag{2.8}$$

holds uniformly for $x\in\mathbb{R}^L$ with probability exceeding $1-2(1+2/\epsilon)^L e^{-\frac{\eta^2 M}{2p\beta_p^2}}$, where

$$\beta_p = (31/40)^{1/4}\left[1.13+\sqrt{p}\left(\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}\right)^{-1/p}\right]. \tag{2.9}$$

Proof of Theorem 1.1. Let $L=(a+1)S=(\lceil b^{\frac{2}{2-p}}\rceil+1)S$, $b>1$. Theorem 2.1 states that for $\delta_{(a+1)S}<(b-1)/(b+1)$, there exits unique solution of (1.5). We choose $\eta=r(b-1)/(b+1)$ for $r\in(0,1)$ and $\epsilon^p=(1+r)(b-1)/2b$ to satisfy

$$\frac{\eta+\epsilon^p}{1-\epsilon^p}\le\delta_{(a+1)S}\le\frac{b-1}{b+1}.$$

By Lemma 2.2, the matrix Ψ will fail to satisfy p-restricted isometry property with probability

$$\begin{aligned}\le\binom{N}{L}(1+2/\epsilon)^L e^{-\frac{\eta^2 M}{2p\beta_p^2}} &= (eN/L)^L(1+2/\epsilon)^L e^{-\frac{\eta^2 M}{2p\beta_p^2}} \\ &= e^{L\ln(eN/L)+L\ln(1+2/\epsilon)-\frac{\eta^2 M}{2p\beta_p^2}},\end{aligned}$$

where β_p is given in (2.9). It is suffices to show that right hand side of above quantity can be bounded by $e^{-\frac{\eta^2 M}{4p\beta_p^2}}$, i.e

$$\frac{\eta^2 M}{4p\beta_p^2}\ge L\ln(eN/L)+L\ln(1+3/\epsilon).$$

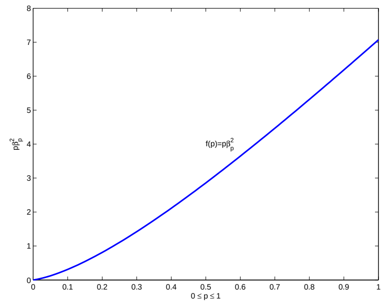

Figure 1: The graph of function $f(p) = p\beta_p^2$

It is equal to show that

$$\begin{aligned} M &\geq \frac{4p\beta_p^2}{\eta^2}\left[L\ln(eN/L) + L\ln(1+3/\epsilon)\right] \\ &= \frac{4p\beta_p^2(b+1)^2}{r^2(b-1)^2}\left[S(\lceil b^{\frac{2}{2-p}}\rceil + 1)\left(\ln\frac{N}{S} + \ln 3 + \ln\frac{1}{\lceil b^{\frac{2}{2-p}}\rceil + 1} + \frac{1}{p}\ln\frac{2b}{(1-r)(b-1)}\right)\right] \\ &= \frac{4\beta_p^2(b+1)^2}{r^2(b-1)^2}S(\lceil b^{\frac{2}{2-p}}\rceil + 1)\ln\frac{2b}{(1-r)(b-1)} \\ &\quad + \frac{4\beta_p^2(b+1)^2}{r^2(b-1)^2}\left[S(\lceil b^{\frac{2}{2-p}}\rceil + 1)\left(\ln\frac{N}{S} + \ln 3 + \ln\frac{1}{\lceil b^{\frac{2}{2-p}}\rceil + 1}\right)\right]. \end{aligned}$$

With the same arguments as in [7], we choose $r = 0.849$ and $b = 5$. Then the following inequality holds:

$$\begin{aligned} M &\geq 35.2\beta_p^2(\lceil 5^{\frac{2}{2-p}}\rceil + 1)S \\ &\quad + 12.5p\beta_p^2\left[(\lceil 5^{\frac{2}{2-p}}\rceil + 1)(\ln 3 - \ln(\lceil 5^{\frac{2}{2-p}}\rceil + 1) + 1)S + (\lceil 5^{\frac{2}{2-p}}\rceil + 1)S\ln\frac{N}{S}\right]. \end{aligned} \tag{2.10}$$

From the above discussion, we conclude that Theorem 1.1 holds with probability exceeding

$$1 - 2e^{-\frac{\eta^2 M}{4p\beta_p^2}} = 1 - 2e^{-\frac{r^2(b-1)^2}{4(b+1)^2}\frac{M}{p\beta_p^2}} = 1 - 2e^{-0.7049\frac{M}{p\beta_p^2}} \geq 1 - 2e^{-c(p)M}$$

where $c(p) = \frac{1}{2p\beta_p^2}$. □

In Figure 1, the graph of the function $f(p) = p\beta_p^2$ shows that $f(p)$ is an increasing function in $[0,1]$ and $\lim_{p\to 0} c(p) = +\infty$. Therefore, decreasing p for (1.5) to successful S-sparse recovery with increasing probability. The limiting case of Theorem 1.1 as $p \to 0$ is essentially the l_0 case. The following propositions show that our results improve the Theorem 3.1 in [7].

Proposition 2.3. *Let* $c(p) = \frac{1}{2p\beta_p^2}$. *Suppose* $M > c(p)(S + S\ln(N/S) + \ln 2)$. *Then*

$$1 - 2e^{-c(p)M} > 1 - \frac{1}{\binom{N}{S}},$$

where $\beta_p = (31/40)^{1/4}\left[1.13 + \sqrt{p}\left(\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}\right)^{-1/p}\right]$.

Proof. For $M > 2p\beta_p^2(S + S\ln(N/S) + \ln 2)$, we have

$$\frac{M}{2p\beta_p^2} > S + S\ln(N/S) + \ln 2.$$

By Stirling approximation, we deduce

$$2e^{c(p)M} > \frac{N^S e^S}{S^S} \geq \binom{N}{S}.$$

Therefore, $1 - 2e^{-c(p)M} > 1 - 1/\binom{N}{S}$. □

Proposition 2.4. *Let* $c(p) = \frac{1}{2p\beta_p^2}$. *Suppose that* M *satisfies (2.10) with constants* C_1', C_2' *such that* $M \geq C_1'\beta_p^2 S + C_2' p\beta_p^2 S\ln(N/S)$. *Then*

$$\frac{2e^{-c(p)M}}{1/\binom{N}{S}} \leq 2e^{-\frac{C_1'S}{p}} \quad \text{and} \quad \lim_{p\to 0}\frac{2e^{-c(p)M}}{1/\binom{N}{S}} = 0.$$

Proof. Since $C_1' > 1$ and $C_2' > 1$, by Stirling approximation we deduce that

$$\begin{aligned}\frac{2e^{-c(p)M}}{1/\binom{N}{S}} &\leq \frac{2e^{-\frac{1}{2p\beta_p^2}(C_1'\beta_p^2 S + C_2' p\beta_p^2 S\ln(N/S))}}{1/\binom{N}{S}} \\ &= 2e^{-\frac{C_1'S}{2p}}\frac{e^{-C_2'S\ln(N/S)}}{1/\binom{N}{S}} \leq 2e^{-\frac{C_1'S}{2p}}\frac{e^{-S\ln(N/S)}}{1/\binom{N}{S}} \leq 2e^{-\frac{C_1'S}{2p}}.\end{aligned}$$

□

3 Stability results

In this section, we mainly consider the case when the measurement is corrupted by noise i.e. we apply the decoder to the vector $\Psi(x) + r$. The stability results for l^p minimization have been considered in [17] for $p = 1$ and in [16] for $0 < p < 1$. We suppose that entries of $\Psi = \Psi_\omega$ are chosen as i.i.d. Gaussian random variables with expectation 0 and variance $1/\sqrt{M}$. (Ω, P) denotes the associated probability space. The proof of Theorem 1.3 relies heavily on the geometrical property of the Gaussian random matrix. There are some well known results for measurement matrices which satisfies quotient property. A matrix Ψ has l_p *quotient property* ($LQ_p(\alpha)$ for short) if

$$\alpha B_2^M \subset \Psi(B_p^N), \quad 0 < p \leq 1,$$

where B_p^N denotes the l_p unit ball in $\mathbb{R}^N$. The $LQ_p(\alpha)$ is satisfied with high probability by random Gaussian matrices [16, 17]. Let $0 < p \leq 1$, and let Ψ be an $M \times N$ Gaussian random matrix. For

$0 < \mu < 1/\sqrt{2}$. Suppose that $K_1 M(\log M)^\xi \le N \le e^{K_2 M}$ for some $\xi > (1-2\mu^2)^{-1}$ and some constants $K_1, K_2 > 0$. Then, there exist a constant $c_3 = c_3(\mu, \xi, K_1, K_2) > 0$, independent of p, M and N, and a set

$$\Omega_\mu = \left\{ \omega \in \Omega : \frac{1}{C_3(p)} \left(\mu^2 \frac{\log N/M}{M} \right)^{1/p-1/2} B_2^M \subset \Psi(B_p^N) \right\}$$

such that $P(\Omega_\mu) \ge 1 - e^{-c_3 M}$. Here $C_3(p)$ is a positive constant that depends only on p.

Lemma 3.1. *Let $0 < p \le 1$, and suppose that the matrix $\Psi \in \mathbb{R}^{M \times N}$ satisfies p-RIP(S, δ) and $LQ_p(\gamma_p / S^{1/p-1/2})$ with $\gamma_p := \mu^{2/p-1}/C_3(p)$. Then for every $x \in \mathbb{R}^N$, there exists $\tilde{x} \in \mathbb{R}^N$ such that*

$$\Psi x = \Psi \tilde{x}, \tag{3.1}$$

$$\|\tilde{x}\|_p \le \frac{S^{1/p-1/2}}{\gamma_p} \|\Psi x\|_2, \tag{3.2}$$

$$\|\tilde{x}\|_2 \le C_4(S, p, \delta_S, \gamma_p) \|\Psi x\|_2. \tag{3.3}$$

Proof. Vector $\tilde{x}$ satisfies (3.1) and (3.2) we get directly from LQ condition. Let T_0 denote the set of indices of the largest S entries of $\tilde{x}$, T_1 the next S largest, and so on. Now for any $i \in T_{j+1}$ and $i' \in T_j$, we have $|\tilde{x}_i| \le |\tilde{x}_{i'}|$. It follows that $\|\tilde{x}_{T_{j+1}}\|_2^p \le S^{\frac{p}{2}-1} \|\tilde{x}_{T_j}\|_p^p$, and consequently,

$$\|\tilde{x}_{T_0^c}\|_2^p \le \sum_{j>0} \|\tilde{x}_{T_j}\|_2^p \le S^{\frac{p}{2}-1} \sum_{j>0} \|\tilde{x}_{T_j}\|_p^p \le S^{\frac{p}{2}-1} \|\tilde{x}\|_p^p \le \frac{1}{\gamma_p} \|\Psi x\|_2^p. \tag{3.4}$$

Use the fact $\Psi x = \Psi \tilde{x}$, we have

$$\begin{aligned} \|\tilde{x}_{T_0}\|_2^p &\le \frac{1}{1-\delta_S} \|\Psi \tilde{x}_{T_0}\|_p^p \\ &\le \frac{1}{1-\delta_S} \|\Psi \tilde{x}\|_p^p = \frac{1}{1-\delta_S} \|\Psi x\|_p^p \\ &\le \frac{S^{1-\frac{p}{2}}}{1-\delta_S} \|\Psi x\|_2^p. \end{aligned} \tag{3.5}$$

Together with (3.4), we obtain

$$\|\tilde{x}\|_2^p \le \|\tilde{x}_{T_0^c}\|_2^p + \|\tilde{x}_{T_0}\|_2^p \le \left(\frac{S^{1-\frac{p}{2}}}{1-\delta_S} + \frac{1}{\gamma_p} \right) \|\Psi x\|_2^p.$$

This leads to the estimate (3.2) with $C_4(S, p, \delta_S, \gamma_p) := \frac{S^{1-\frac{p}{2}}}{1-\delta_S} + \frac{1}{\gamma_p}$. ◻

Using Lemma 3.1, we prove the following results.

Lemma 3.2. *Let $0 < p \le 1$. Suppose that $\Psi \in \mathbb{R}^{M \times N}$ satisfies p-RIP(S, δ) and $LQ_p(\gamma_p)$. If the pair $(\Psi, \triangle)$ is $(2, p)$ instance optimal with constant $C_{2,p}$. Then there exists C_5, C_6, C_7 such that for every $x \in \mathbb{R}^N$ and $r \in \mathbb{R}^M$, all of the following hold.*

(i)

$$\|x - \triangle(\Psi(x) + r)\|_2 \le C_5 \left(\|r\|_2 + \frac{\sigma_S(x)_p}{S^{1/p-1/2}} \right) \tag{3.6}$$

(ii)

$$\|x - \triangle(\Psi(x))\|_2 \le C_6(\|\Psi x_{T_0^c}\|_2 + \sigma_S(x)_2) \tag{3.7}$$

(iii)

$$\|x - \triangle(\Psi(x) + r)\|_2 \le C_7(\|r\|_2 + \sigma_S(x)_2 + \|\Psi x_{T_0^c}\|_2). \tag{3.8}$$

Proof. Since Ψ satisfies LQ, there exists $z \in \mathbb{R}^N$ such that $\Psi(z) = r$. It follows that

$$\Psi(x + z) = \Psi(x) + r.$$

By Lemma 3.1, we have that

$$\|z\|_p \le \frac{S^{1/p-1/2}}{\gamma_p} \gamma_p \|r\|_2, \quad \|z\|_2 \le C_4 \|r\|_2.$$

The instance optimal of Ψ implies that

$$\|(x + z) - \triangle(\Psi x + r)\|_2 \le C_{2,p} \frac{\sigma_S(x + z)_p}{S^{1/p-1/2}}.$$

This shows that

$$\begin{aligned}
\|x - \triangle(\Psi x + r)\|_2 &\le \|z\|_2 + C_{2,p} \frac{\sigma_S(x + z)_p}{S^{1/p-1/2}} \\
&\le C_4 \|r\|_2 + C_{2,p} \frac{\sigma_S(x + z)_p}{S^{1/p-1/2}} \\
&\le C_4 \|r\|_2 + 2^{1/p-1} C_{2,p} \frac{\sigma_S(x)_p + \|z\|_p}{S^{1/p-1/2}} \\
&\le C_4 \|r\|_2 + 2^{1/p-1} C_{2,p} \frac{\sigma_S(x)_p}{S^{1/p-1/2}} + 2^{1/p-1} C_4 C_{2,p} \|r\|_2.
\end{aligned}$$

Taking $C_5 = C_4 + 2^{1/p-1} C_4 C_{2,p}$, we prove (3.6).

Note that (3.7) is a special case of (3.8) with $r = 0$, we only need to prove (3.8). Given $x \in \mathbb{R}^N$, let T_0 be the set of S-largest entries of x in absolute value. Then $\|x_{T_0^c}\|_2 = \sigma_S(x)_p$. Consider $r \in \mathbb{R}^M$ and $x_{T_0^c} \in \mathbb{R}^N$. By Lemma 3.1, there exist $v, z \in \mathbb{R}^N$ such that

$$\Psi(v) = r, \quad \|v\|_p \le \frac{S^{1/p-1/2}}{\gamma_p} \|r\|_2, \quad \|v\|_2 \le C_4 \|r\|_2.$$

$$\Psi(z) = \Psi(x_{T_0^c}), \quad \|z\|_p \le \frac{S^{1/p-1/2}}{\gamma_p} \|\Psi x_{T_0^c}\|_2, \quad \|z\|_2 \le C_4 \|\Psi(x_{T_0^c})\|_2.$$

Note the fact $\Psi(x_{T_0} + z + v) = \Psi x + r$, it follows that

$$\|(x_{T_0} + z + v) - \triangle(\Psi x + r)\|_2 \le C_{2,p} \frac{\sigma_S(x_{T_0} + z + v)_p}{S^{1/p-1/2}}.$$

Since $\sigma_S(x_{T_0}) = 0$, we have

$$\sigma_S(x_{T_0} + z + v)_p \leq 2^{1/p-1}(\|z\|_p + \|v\|_p).$$

Hence,

$$\begin{aligned}
\|x - \triangle(\Psi x + r)\|_2 &= \|x_{T_0} + x_{T_0^c} - \triangle(\Psi x + r)\|_2 \\
&\leq \|x_{T_0^c} - z - v\|_2 + \|(x_{T_0} + z + v) - \triangle(\Psi x + r)\|_2 \\
&\leq \|x_{T_0^c} - z - v\|_2 + C_{2,p}\frac{\sigma_S(x_{T_0} + z + v)_p}{S^{1/p-1/2}} \\
&\leq \sigma_S(x)_2 + \|z\|_2 + \|v\|_2 + 2^{1/p-1}C_{2,p}\frac{\|z\|_p + \|v\|_p}{S^{1/p-1/2}} \\
&\leq \sigma_S(x)_2 + C_4\|\Psi(x_{T_0^c})\|_2 + C_4\|r\|_2 + 2^{1/p-1}\frac{C_{2,p}}{\gamma_p}(\|\Psi x_{T_0^c}\|_2 + \|r\|_2) \\
&\leq \sigma_S(x)_2 + \left(C_4 + 2^{1/p-1}\frac{C_{2,p}}{\gamma_p}\right)(\Psi(x_{T_0^c}) + \|r\|_2).
\end{aligned}$$

That is (3.8) holds with $C_7 := C_4 + 2^{1/p-1}\frac{C_{2,p}}{\gamma_p}$. Let $r = 0$ in (3.8), we get (3.7) with constants $C_6 = C_7$. □

Proof of Theorem 1.3. By Theorem 1.1 and Lemma 3.2, the proof of Theorem 1.3 follows with little modification from the analogous proof of Theorem 2.14 in [16]. □

References

[1] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, Constr. Approx. 28 (2008) 253-263.

[2] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2006) 489-509.

[3] E.J. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59 (2006) 1207-1223.

[4] E.J. Candès, T. Tao, Near optimal signal recovery from random projections: universal encoding strategies. IEEE Trans. Inform. Theory 52(1) (2006) 5406-5425.

[5] E.J. Candès, T. Tao, Decoding by linear programming IEEE Trans. Inf. Theory 51 (2005) 4203-4215.

[6] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett. 14 (2007) 707-710.

[7] R. Chartrand, V. Staneva, Restricted isometry porperties and nonconvex compressive sensing, Inverse Problems 24 (2009) 1-14.

[8] A. Cohen, W. Dahmen, R. DeVore, Compressed sensing and best k-term approximation, J. Amer. Math. Soc. 22 (2009) 211-245.

[9] R. DeVore, G. Petrova, P. Wojtaczszyk, Instance-optimality in probability with an l_1-minimization decoder, Appl. Comput. Harmon. Anal. 27 (2009) 275-288.

[10] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289-1306.

[11] D. Donoho, For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution, Comm. Pure Appl. Math 59 (2006) 797-829.

[12] D. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. Proc. Natl. Acad. Sci. USA 100 (5) (2003) 2197-2202.

[13] D.D. Ge, X.Y. Jiang, Y.Y. Ye, A Note on Complexity of l_p Minimization, manuscript.

[14] S. Foucart, M.J. Lai, Sparsest Solutions of Underdetermined Linear Systems via l_q-minimization for $0 < q \leq 1$, Appl. Comput. Harmon. Anal. 26 (2009) 395-407.

[15] R. Gribonval, M. Nielsen, Sparse Representations in Unions of Bases, IEEE Trans. Info. Theory 49 (2003) 3320-3325.

[16] R. Saab, Ö. Yılmaz, Sparse revovery by non-convex optimization-instance optimality, Appl. Comput. Harmon. Anal. (2009) doi:10.1016/j.acha.2009.08.002.

[17] P. Wojtaszczyk, Stability and instance optimality for Gaussian measurements in compressed sensing, Found. Comput. Math. 10 (2010) 1-13.