New Bounds for Restricted Isometry Constants

T. Tony Cai, Lie Wang, and Guangwu Xu

Abstract—**This paper discusses new bounds for restricted isometry constants in compressed sensing. Let Φ be an $n \times p$ real matrix and k be a positive integer with $k \leq n$. One of the main results of this paper shows that if the restricted isometry constant δ_k of Φ satisfies**

$$\delta_k < 0.307$$

then k-sparse signals are guaranteed to be recovered exactly via ℓ_1 minimization when no noise is present and k-sparse signals can be estimated stably in the noisy case. It is also shown that the bound cannot be substantially improved. An explicit example is constructed in which $\delta_k = \frac{k-1}{2k-1} < 0.5$, but it is impossible to recover certain k-sparse signals.

Index Terms—**Compressed sensing, ℓ_1 minimization, restricted isometry property, sparse signal recovery.**

I. INTRODUCTION

COMPRESSED sensing aims to recover high-dimensional sparse signals based on considerably fewer linear measurements. Formally one considers the following model:

$$y = \Phi\beta + z \tag{1}$$

where the matrix $\Phi \in \mathbb{R}^{n \times p}$ (with $n \ll p$) and $z \in \mathbb{R}^n$ is a vector of measurement errors. The goal is to reconstruct the unknown signal $\beta \in \mathbb{R}^p$ based on y and Φ. A remarkable fact is that β can be recovered exactly in the noiseless case under suitable conditions, provided that the signal is sparse.

A naïve approach for solving this problem is to consider ℓ_0 minimization where the goal is to find the sparsest solution in the feasible set of possible solutions. However, this is NP hard and thus is computationally infeasible. It is then natural to consider the method of ℓ_1 minimization which can be viewed as a convex relaxation of ℓ_0 minimization. The ℓ_1 minimization method in this context is

$$(P_\mathcal{B}) \quad \hat{\beta} = \arg\min_{\gamma \in \mathbb{R}^p} \{\|\gamma\|_1 \text{ subject to } y - \Phi\gamma \in \mathcal{B}\} \tag{2}$$

where $\mathcal{B}$ is a bounded set determined by the noise structure. For example, $\mathcal{B} = \{0\}$ in the noiseless case and $\mathcal{B}$ is the feasible set of the noise in the case of bounded error. This method has been successfully used as an effective way for reconstructing a sparse signal in many settings. See, e.g., [6]–[9], [11], [13], [2], [3].

One of the most commonly used frameworks for sparse recovery via ℓ_1 minimization is the *Restricted Isometry Property* (RIP) introduced by Candès and Tao [7]. The RIP essentially requires that every subset of columns of Φ with certain cardinality approximately behaves like an orthonormal system. A vector $v = (v_i) \in \mathbb{R}^p$ is k-*sparse* if $|\text{supp}(v)| \leq k$, where $\text{supp}(v) = \{i : v_i \neq 0\}$ is the support of v. For an $n \times p$ matrix Φ and an integer k, $1 \leq k \leq p$, the k-*restricted isometry constant* $\delta_k(\Phi)$ is the smallest constant such that

$$\sqrt{1 - \delta_k(\Phi)}\|c\|_2 \leq \|\Phi c\|_2 \leq \sqrt{1 + \delta_k(\Phi)}\|c\|_2 \tag{3}$$

for every k-sparse vector c. If $k + k' \leq p$, the k, k'-*restricted orthogonality constant* $\theta_{k,k'}(\Phi)$, is the smallest number that satisfies

$$|\langle \Phi c, \Phi c' \rangle| \leq \theta_{k,k'}(\Phi)\|c\|_2\|c'\|_2 \tag{4}$$

for all c and c' such that c and c' are k-sparse and k'-sparse respectively, and have disjoint supports. For notational simplicity, we shall write δ_k for $\delta_k(\Phi)$ and $\theta_{k,k'}$ for $\theta_{k,k'}(\Phi)$ hereafter.

It has been shown that ℓ_1 minimization can recover a sparse signal with a small or zero error under various conditions on δ_k and $\theta_{k,k'}$. For example, the condition $\delta_k + \theta_{k,k} + \theta_{k,2k} < 1$ was used in Candès and Tao [7], $\delta_{3k} + 3\delta_{4k} < 2$ in Candès *et al.* [6], and $\delta_{2k} + \theta_{k,2k} < 1$ in Candès and Tao [9]. In [4], Cai *et al.* proved that stable recovery can be achieved when $\delta_{1.5k} + \theta_{k,1.5k} < 1$.[1] In a recent paper, Cai *et al.* [3] further improve the condition to $\delta_{1.25k} + \theta_{k,1.25k} < 1$.

It is important to note that the RIP conditions are difficult to verify for a given matrix Φ. A widely used technique for avoiding checking the RIP directly is to generate the matrix Φ randomly and to show that the resulting random matrix satisfies the RIP with high probability using the well-known Johnson–Lindenstrauss Lemma. See, for example, Baraniuk, *et al.* [1]. This is typically done for conditions involving only the restricted isometry constant δ. Attention has been focused on δ_{2k} as it is obviously necessary to have $\delta_{2k} < 1$ for model identifiability. In a recent paper, Davies and Gribonval [10] constructed examples which showed that if $\delta_{2k} \geq \frac{1}{\sqrt{2}}$, exact recovery of certain k-sparse signal can fail in the noiseless case. On the other hand, sufficient conditions on δ_{2k} has been given. For example, $\delta_{2k} < \sqrt{2} - 1$ is used by Candès [5] and $\delta_{2k} < 0.4531$ by Foucart and Lai [14]. The results given in

---

[1]For a positive real number α, $\delta_{\alpha k}$ and $\theta_{k,\alpha k}$ are understood as $\delta_{\lceil \alpha k \rceil}$ and $\theta_{k,\lceil \alpha k \rceil}$.

Cai *et al.* [3] implies that $\delta_{2k} < 0.472$ is a sufficient condition for sparse signal recovery.

Among the conditions of the form $\delta_{ck} < \alpha$, the most natural and desirable condition for recovering a k-sparse signal is arguably

$$\delta_k < \alpha$$

for some quantity α.

The purpose of this paper is to establish, to the best of our knowledge, the first such condition on δ_k. To be more specific, we show that under the condition

$$\delta_k < 0.307 \tag{5}$$

k-sparse signals are guaranteed to be recovered exactly via ℓ_1 minimization when no noise is present and k-sparse signals can be estimated stably in the noisy case. Although we are mainly interested in recovering sparse signals, the results can be extended to the general setting where the true signal is not necessarily k-sparse.

It is also shown in the present paper that the bound (5) cannot be substantially improved. An upper bound for δ_k is also given. An explicit example is constructed in which $\delta_k = \frac{k-1}{2k-1} < 0.5$, but it is impossible to recover certain k-sparse signals.

Our analysis is simple and elementary. The main ingredients in proving the new RIP conditions are the *norm inequality for ℓ_1 and ℓ_2*, and the *square root lifting inequality* for the restricted orthogonality constant $\theta_{k,k'}$. Let $x \in \mathbb{R}^n$. A direct consequence of the Cauchy-Schwartz inequality is that $0 \leq \|x\|_2 - \frac{\|x\|_1}{\sqrt{n}}$. Our norm inequality for ℓ_1 and ℓ_2 gives an upper bound for the quantity $\|x\|_2 - \frac{\|x\|_1}{\sqrt{n}}$, namely

$$\|x\|_2 - \frac{\|x\|_1}{\sqrt{n}} \leq \frac{\sqrt{n}}{4} \left(\max_{1 \leq i \leq n} |x_i| - \min_{1 \leq i \leq n} |x_i| \right). \tag{6}$$

This is an inequality of its own interest. The square root lifting inequality is a result we developed in [3] which states that if $a \geq 1$ and k', ak' are positive integers, then

$$\theta_{k,ak'} \leq \sqrt{a}\,\theta_{k,k'}. \tag{7}$$

Indeed we derive a more general result on the RIP and obtain (5) as a special case.

The rest of the paper is organized as follows. In Section II, after some basic notations and definitions are introduced, the square root lifting inequality and the norm inequality for ℓ_1 and ℓ_2 are discussed. These inequalities are the main technical tools which enable us to make finer analysis of the sparse recovery problem. Our new RIP bounds are presented in Section III. In Section IV, upper bounds for the RIP constants are given.

II. PRELIMINARIES AND SOME INEQUALITIES

We begin by introducing basic notations and definitions related to the RIP. We also include two important inequalities needed for the later sections.

For a vector $v = (v_i) \in \mathbb{R}^p$, we shall denote by $v_{\max(k)}$ the vector v with all but the k largest entries (in absolute value) set to zero and define $v_{-\max(k)} = v - v_{\max(k)}$, the vector v with the k largest entries (in absolute value) set to zero. We use the standard notation $\|v\|_q = (\sum_{i=1}^{p} |v_i|^q)^{1/q}$ to denote the ℓ_q-norm of the vector v. We shall also treat a vector $v = (v_i)$ as a function $v : \{1, 2, \ldots, p\} \to \mathbb{R}$ by assigning $v(i) = v_i$.

For a subset T of $\{1, \ldots, p\}$, we use Φ_T to denote the submatrix obtained by taking the columns of Φ according to the indices in T. If $c \in \mathbb{R}^p$ with support T, then

$$\sqrt{1 - \delta_{|T|}}\|c\|_2 \leq \|\Phi_T c\|_2 \leq \sqrt{1 + \delta_{|T|}}\|c\|_2.$$

The following monotone properties can be easily checked

$$\delta_k \leq \delta_{k_1}, \text{ if } k \leq k_1 \leq p \tag{8}$$
$$\theta_{k,k'} \leq \theta_{k_1,k_1'}, \text{ if } k \leq k_1, k' \leq k_1', \text{ and } k_1 + k_1' \leq p. \tag{9}$$

Candès and Tao [7] showed that the constants δ_k and $\theta_{k,k'}$ are related by the following inequalities

$$\theta_{k,k'} \leq \delta_{k+k'} \leq \theta_{k,k'} + \max(\delta_k, \delta_{k'}). \tag{10}$$

A. Square Root Lifting Inequality

The following properties for δ and θ, developed by Cai, Xu and Zhang in [4], have been especially useful in producing simplified recovery conditions

$$\theta_{k,\sum_{i=1}^{l} k_i} \leq \sqrt{\sum_{i=1}^{l} \theta_{k,k_i}^2} \leq \sqrt{\sum_{i=1}^{l} \delta_{k+k_i}^2}. \tag{11}$$

It follows from (11) that for any positive integer a, we have $\theta_{k,ak'} \leq \sqrt{a}\,\theta_{k,k'}$. This fact was further generalized by Cai, Wang and Xu in [3] to the following square root lifting inequality.

Lemma 2.1 (Square Root Lifting Inequality): For any $a \geq 1$ and positive integers k, k' such that ak' is an integer

$$\theta_{k,ak'} \leq \sqrt{a}\,\theta_{k,k'}. \tag{12}$$

It is interesting to note that some useful properties of the restricted isometry constant can be established by the square root lifting inequality. For example, for any positive integer $m > 1$

$$\begin{aligned}
\delta_{2m} &\leq \theta_{m,m} + \delta_m \\
&= \theta_{\frac{m}{\lceil \frac{m}{2} \rceil}\lceil \frac{m}{2} \rceil, \frac{m}{\lfloor \frac{m}{2} \rfloor}\lfloor \frac{m}{2} \rfloor} + \delta_m \\
&\leq \sqrt{\frac{m}{\lceil \frac{m}{2} \rceil} \frac{m}{\lfloor \frac{m}{2} \rfloor}}\, \theta_{\lceil \frac{m}{2} \rceil, \lfloor \frac{m}{2} \rfloor} + \delta_m \\
&\leq \frac{m}{\sqrt{\lceil \frac{m}{2} \rceil \lfloor \frac{m}{2} \rfloor}}\, \delta_m + \delta_m \\
&= \begin{cases} 3\delta_m, & \text{if } m \text{ is even} \\ \frac{2 + \sqrt{1 - \frac{1}{m^2}}}{\sqrt{1 - \frac{1}{m^2}}}\delta_m, & \text{if } m \text{ is odd.} \end{cases}
\end{aligned}$$

B. Norm Inequality for ℓ_1 and ℓ_2

In the last part of this section, we will develop a useful inequality for achieving finer conversion between ℓ_1-norm and ℓ_2-norm.

Let $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. A direct consequence of the Cauchy-Schwartz inequality is that

$$0 \leq \|x\|_2 - \frac{\|x\|_1}{\sqrt{n}}$$

and the equality hold if and only if $|x_1| = |x_2| = \cdots = |x_n|$. The next result provides a sharp upper bound for the quantity $\|x\|_2 - \frac{\|x\|_1}{\sqrt{n}}$. This norm inequality plays a major role in the subsequent analysis of the RIP conditions.

Proposition 2.1: For any $x \in \mathbb{R}^n$

$$\|x\|_2 - \frac{\|x\|_1}{\sqrt{n}} \leq \frac{\sqrt{n}}{4} \left(\max_{1 \leq i \leq n} |x_i| - \min_{1 \leq i \leq n} |x_i| \right).$$

The equality is attained if and only if $|x_1| = |x_2| = \cdots = |x_n|$, or $n = 4m$ for some positive integer m and x satisfies $|x_{i_1}| = |x_{i_2}| = \cdots = |x_{i_m}|$ for some $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ and $x_k = 0$ for $k \notin \{i_1, i_2, \ldots, i_m\}$.

Proof: It is obvious that the result holds when the absolute values of all coordinates are equal. Without loss of generality, we now assume that $x_1 \geq x_2 \geq \cdots \geq x_n \geq 0$ and not all x_i are equal. Let

$$f(x) = \|x\|_2 - \frac{\|x\|_1}{\sqrt{n}}.$$

Note that for any $i \in \{2, 3, \ldots, n-1\}$

$$\frac{\partial f}{\partial x_i} = \frac{x_i}{\|x\|_2} - \frac{1}{\sqrt{n}}.$$

This implies that when $x_i \leq \frac{\|x\|_2}{\sqrt{n}}$, $f(x)$ is decreasing in x_i; otherwise $f(x)$ is increasing in x_i. Therefore, if we fix x_1 and x_n, when $f(x)$ achieves its maximum, x must be of the form that $x_1 = x_2 = \cdots = x_k$ and $x_{k+1} = \cdots = x_n$ for some $1 \leq k < n$. Now

$$f(x) = \sqrt{k(x_1^2 - x_n^2) + nx_n^2} - \frac{k}{\sqrt{n}}(x_1 - x_n) - \sqrt{n}x_n.$$

Treat this as a function of k for $k \in (0, n)$

$$g(k) = \sqrt{k(x_1^2 - x_n^2) + nx_n^2} - \frac{k}{\sqrt{n}}(x_1 - x_n) - \sqrt{n}x_n.$$

By taking the derivative, it is easy to see that

$$g(k) \leq g\left(n \frac{\left(\frac{x_1 + x_n}{2}\right)^2 - x_n^2}{x_1^2 - x_n^2} \right)$$

$$= \sqrt{n}(x_1 - x_n)\left(\frac{1}{2} - \frac{x_1 + 3x_n}{4(x_1 + x_n)} \right).$$

Now, since $\frac{x_1 + 3x_n}{4(x_1 + x_n)} \geq 1/4$, we have

$$\|x\|_2 \leq \frac{\|x\|_1}{\sqrt{n}} + \frac{\sqrt{n}}{4}(x_1 - x_n).$$

We can also see that the above inequality becomes an equality if and only if $x_{k+1} = \cdots = x_n = 0$ and $k = n/4$. ∎

Remark 2.1: A direct consequence of Proposition 2.1 is that for any $x \in \mathbb{R}^n$

$$\|x\|_2 \leq \frac{\|x\|_1}{\sqrt{n}} + \frac{\sqrt{n}\|x\|_\infty}{4}.$$

III. NEW RIP BOUNDS OF COMPRESSED SENSING MATRICES

In this section, we consider new RIP conditions for sparse signal recovery. However, the results can be easily extended to general signals β with error bounds involved with $\beta_{-\max(k)}$, as discussed in [3], [4].

Suppose

$$y = \Phi\beta + z$$

with $\|z\|_2 \leq \varepsilon$. Denote $\hat{\beta}$ the solution of the following ℓ_1 minimization problem:

$$\hat{\beta} = \arg\min\{\|\gamma\|_1 : \quad \text{subject to} \quad \|\Phi\gamma - y\|_2 \leq \varepsilon\}. \quad (13)$$

Theorem 3.1: Suppose β is k-sparse. Let k_1, k_2 be positive integers such that $k_1 \geq k$ and $8(k_1 - k) \leq k_2$. Let

$$t = \sqrt{\frac{k_1}{k_2}} + \frac{1}{4}\sqrt{\frac{k_2}{k_1}} - \frac{2(k_1 - k)}{\sqrt{k_1 k_2}}.$$

Then under the condition

$$\delta_{k_1} + t\theta_{k_1, k_2} < 1$$

the ℓ_1 minimizer $\hat{\beta}$ defined in (13) satisfies

$$\|\beta - \hat{\beta}\|_2 \leq \frac{2\sqrt{2}\sqrt{1 + \delta_{k_1}}}{1 - \delta_{k_1} - t\theta_{k_1, k_2}}\varepsilon.$$

In particular, in the noiseless case where $y = \Phi\beta$, ℓ_1 minimization recovers β exactly.

Proof: Let $h = \hat{\beta} - \beta$. For any subset $Q \subset \{1, 2, \ldots, p\}$, we define

$$h_Q = h\mathbb{1}_Q$$

where $\mathbb{1}_Q$ denotes the indicator function of the set Q, i.e., $\mathbb{1}_Q(j) = 1$ if $j \in Q$ and 0 if $j \notin Q$. Rearranging the indices if necessary, we assume $|h(1)| \geq |h(2)| \geq \cdots \geq |h(k+1)| \geq |h(k+2)| \geq \cdots$.

Let $T = \{1, 2, \ldots, k\}$ and let Ω be the support of β. The following fact, which is based on the minimality of $\hat{\beta}$, has been widely used, see [3], [6], [13].

$$\|h_\Omega\|_1 \geq \|h_{\Omega^c}\|_1.$$

Note that $\Omega^c \cap T$ and $\Omega \cap T^c$ both have $k - |\Omega \cap T|$ elements, so we have

$$\|h_{\Omega^c \cap T}\|_1 \geq \|h_{\Omega \cap T^c}\|_1.$$

We shall show that this implies that

$$\|h_T\|_1 \geq \|h_{T^c}\|_1.$$

In fact

$$
\begin{aligned}
\|h_T\|_1 &= \|h_{\Omega \cap T}\|_1 + \|h_{\Omega^c \cap T}\|_1 \\
&= \|h_\Omega\|_1 + \|h_{\Omega^c \cap T}\|_1 - \|h_{\Omega \cap T^c}\|_1 \\
&\geq \|h_{\Omega^c}\|_1 + \|h_{\Omega^c \cap T}\|_1 - \|h_{\Omega \cap T^c}\|_1 \\
&= \|h_{\Omega^c \cap T^c}\|_1 + 2\|h_{\Omega^c \cap T}\|_1 - \|h_{\Omega \cap T^c}\|_1 \\
&= \|h_{T^c}\|_1 + 2\|h_{\Omega^c \cap T}\|_1 - 2\|h_{\Omega \cap T^c}\|_1 \\
&\geq \|h_{T^c}\|_1.
\end{aligned}
$$

Partition $\{1, 2, \ldots, p\}$ into the following sets:

$$
\begin{aligned}
S_0 &= \{1, 2, \ldots, k_1\}, S_1 = \{k_1 + 1, \ldots, k_1 + k_2\} \\
S_2 &= \{k_1 + k_2 + 1, \ldots, k_1 + 2k_2\}, \ldots.
\end{aligned}
$$

Then it follows from Proposition 2.1 that

$$
\begin{aligned}
\sum_{i \geq 1} \|h_{S_i}\|_2 &\leq \sum_{i \geq 1} \frac{\|h_{S_i}\|_1}{\sqrt{k_2}} + \frac{\sqrt{k_2}}{4}(|h(k_1 + 1)| \\
&\quad - |h(k_1 + k_2)| + |h(k_1 + k_2 + 1)| \\
&\quad - |h(k_1 + 2k_2)| + \cdots) \\
&\leq \frac{1}{\sqrt{k_2}}(\|h_{T^c}\|_1 - (k_1 - k)|h(k_1 + 1)|) \\
&\quad + \frac{\sqrt{k_2}}{4}|h(k_1 + 1)| \\
&\leq \frac{1}{\sqrt{k_2}}(\|h_T\|_1 - (k_1 - k)|h(k_1 + 1)|) \\
&\quad + \frac{\sqrt{k_2}}{4}|h(k_1 + 1)| \\
&\leq \frac{1}{\sqrt{k_2}}(\|h_{S_0}\|_1 - 2(k_1 - k)|h(k_1 + 1)|) \\
&\quad + \frac{\sqrt{k_2}}{4}|h(k_1 + 1)| \\
&\leq \sqrt{\frac{k_1}{k_2}}\|h_{S_0}\|_2 \\
&\quad + \left(\frac{\sqrt{k_2}}{4} - \frac{2(k_1 - k)}{\sqrt{k_2}}\right)|h(k_1 + 1)| \\
&\leq \sqrt{\frac{k_1}{k_2}}\|h_{S_0}\|_2 \\
&\quad + \left(\frac{\sqrt{k_2}}{4\sqrt{k_1}} - \frac{2(k_1 - k)}{\sqrt{k_1 k_2}}\right)\|h_{S_0}\|_2 \\
&= t\|h_{S_0}\|_2.
\end{aligned}
$$

Now

$$
\begin{aligned}
|\langle \Phi h, \Phi h_{S_0}\rangle| &= \left| \langle \Phi h_{S_0}, \Phi h_{S_0}\rangle + \sum_{i \geq 1} \langle \Phi_{S_i} h_{S_i}, \Phi h_{S_0}\rangle \right| \\
&\geq (1 - \delta_{k_1})\|h_{S_0}\|_2^2 \\
&\quad - \theta_{k_1, k_2}\|h_{S_0}\|_2 \sum_{i \geq 1} \|h_{S_i}\|_2 \\
&\geq (1 - \delta_{k_1} - t\theta_{k_1, k_2})\|h_{S_0}\|_2^2
\end{aligned}
$$

Note that

$$
\begin{aligned}
\|\Phi h\|_2 &= \|\Phi(\beta - \hat{\beta})\|_2 \\
&\leq \|\Phi\beta - y\|_2 + \|\Phi\hat{\beta} - y\|_2 \leq 2\varepsilon.
\end{aligned}
$$

Also the next relation

$$\left\|h_{S_0^c}\right\|_2^2 \leq \left\|h_{S_0^c}\right\|_1 \frac{\|h_{S_0}\|_1}{k_1} \leq \frac{\|h_{S_0}\|_1^2}{k_1} \leq \|h_{S_0}\|_2^2$$

implies

$$\|h\|_2^2 = \|h_{S_0}\|_2^2 + \left\|h_{S_0^c}\right\|_2^2 \leq 2\|h_{S_0}\|_2^2.$$

Putting them together we get[2]

$$
\begin{aligned}
\|h\|_2 &\leq \sqrt{2}\|h_{S_0}\|_2 \\
&\leq \frac{\sqrt{2}|\langle \Phi h, \Phi h_{S_0}\rangle|}{(1 - \delta_{k_1} - t\theta_{k_1, k_2})\|h_{S_0}\|_2} \\
&\leq \frac{\sqrt{2}\|\Phi h\|_2 \|\Phi h_{S_0}\|_2}{(1 - \delta_{k_1} - t\theta_{k_1, k_2})\|h_{S_0}\|_2} \\
&\leq \frac{2\sqrt{2}\varepsilon\sqrt{1 + \delta_{k_1}}\|h_{S_0}\|_2}{(1 - \delta_{k_1} - t\theta_{k_1, k_2})\|h_{S_0}\|_2} \\
&\leq \frac{2\sqrt{2}\sqrt{1 + \delta_{k_1}}}{1 - \delta_{k_1} - t\theta_{k_1, k_2}}\varepsilon.
\end{aligned}
$$

∎

Remark 3.1: Different choices of k_1 and k_2 can result in different conditions. Here we list several of them which are of certain interest.[3]

k_1	k_2	Recovery condition
k	k	$\delta_k + 1.25\theta_{k,k} < 1$
k	$\frac{4}{9}k$	$\delta_k + \frac{5}{3}\theta_{k, \frac{4k}{9}} < 1$
$\frac{9}{8}k$	k	$\delta_{\frac{9k}{8}} + \sqrt{\frac{9}{8}}\theta_{k, \frac{9k}{8}} < 1$
$\frac{8}{7}k$	$\frac{8}{7}k$	$\delta_{\frac{8k}{7}} + \theta_{\frac{8k}{7}, \frac{8k}{7}} < 1$

Remark 3.2: It should be noted that when $k_2 = 1$, we can set $t = \sqrt{k}$. In fact, this can be observed by taking $k_1 = k$ and noting $h(k_1 + (i - 1)k_2 + 1) = h(k_1 + ik_2)$ for $i > 0$, in the proof of Theorem 3.1.

---

[2]If $h_{S_0} = 0$, then the theorem is trivially true. So here, we assume that $h_{S_0} \neq 0$.

[3]Here we assume that the the fraction multiple of k are integers. Otherwise, we have to use the ceiling notation.

The following is our main result of the paper. It is the consequence of Theorem 3.1 and the square root lifting inequality.

Theorem 3.2: Let $y = \Phi\beta + z$ with $\|z\|_2 \leq \varepsilon$. Suppose β is k-sparse with $k > 1$. Then under the condition

$$\delta_k < 0.307$$

the constrained ℓ_1 minimizer $\hat{\beta}$ given in (13) satisfies

$$\|\beta - \hat{\beta}\|_2 \leq \frac{\varepsilon}{0.307 - \delta_k}.$$

In particular, in the noiseless case $\hat{\beta}$ recovers β exactly.

To the best of our knowledge, this seems to be the first result for sparse recovery with conditions that only involve δ_k.

Proof: We will present the proof for the case $k \equiv 0 (\mathrm{mod}\, 9)$ in this section. This is the case that can be treated in a concise way and for which the proof also conveys the main ideas. A complete proof for the general case is given in the Appendix.

In Theorem 3.1, set $k_1 = k$ and $k_2 = \frac{4}{9}k$. Let

$$t = \sqrt{\frac{k}{k_2}} + \frac{1}{4}\sqrt{\frac{k_2}{k}} = \frac{5}{3}.$$

Then under the condition

$$\delta_k + \frac{5}{3}\theta_{k, \frac{4}{9}k} < 1$$

we have

$$\|\beta - \hat{\beta}\|_2 \leq \frac{2\sqrt{2}\sqrt{1+\delta_k}}{1 - \delta_k - \frac{5}{3}\theta_{k,\frac{4}{9}k}}\epsilon.$$

Using the square root lifting inequality, we get

$$\delta_k + \frac{5}{3}\theta_{k,\frac{4}{9}k} = \delta_k + \frac{5}{3}\theta_{\frac{9}{5}\frac{5k}{9},\frac{4k}{9}}$$
$$\leq \delta_k + \frac{5}{3}\sqrt{\frac{9}{5}}\theta_{\frac{5k}{9},\frac{4k}{9}} \leq (1+\sqrt{5})\delta_k$$
$$< 1.$$

In this case

$$\|\beta - \hat{\beta}\|_2 \leq \frac{2\sqrt{2}\sqrt{1+\delta_k}}{1 - \delta_k - t\theta_{k,\frac{4k}{9}}}\varepsilon \leq \frac{2\sqrt{2}\sqrt{1+\delta_k}}{1 - (1+\sqrt{5})\delta_k}\varepsilon$$
$$\leq \frac{3.256}{1 - 3.256\delta_k}\varepsilon \leq \frac{\varepsilon}{0.307 - \delta_k}.$$

∎

Remark 3.3: In the proof of Theorem 3.2, we used a weaker form of estimation in the last line. The purpose is to make the result consistent with the general case which will be treated in the Appendix.

For the special case of $k \equiv 0 (\mathrm{mod}\, 9)$, we actually have a slightly better error bound, that is

$$\|\beta - \hat{\beta}\|_2 \leq \frac{2\sqrt{2}\sqrt{1+\delta_k}}{1 - (1+\sqrt{5}) \cdot \delta_k}\varepsilon \leq \frac{1}{0.309 - \delta_k}\varepsilon.$$

For simplicity, we have focused on recovering k-sparse signals in the present paper. When β is not k-sparse, ℓ_1 minimization can also recover β with accuracy if β has good k-term approximation, i.e., $\|\beta_{-\max(k)}\|_1$ is small. Similar to [2], [4], this result can be extended to the general setting.

Theorem 3.3: Let $y = \Phi\beta + z$ with $\|z\|_2 \leq \varepsilon$. Suppose

$$\delta_k < 0.307$$

for some $k \geq 1$. Then the constrained ℓ_1 minimizer $\hat{\beta}$ given in (13) satisfies

$$\|\hat{\beta} - \beta\|_2 \leq \frac{\varepsilon}{0.307 - \delta_k} + \frac{1}{0.307 - \delta_k}\frac{\|\beta_{-\max(k)}\|_1}{\sqrt{k}}.$$

We now consider stable recovery of k-sparse signals with error in a different bounded set. Candès and Tao [9] treated the sparse signal recovery in the Gaussian noise case by solving $(P_{\mathcal{B}})$ with $\mathcal{B} = \mathcal{B}^{DS} = \{z : \|\Phi'z\|_\infty \leq \eta\}$ and referred the solution as the *Dantzig Selector*. The following result shows that the condition $\delta_k < 0.307$ is also sufficient when the error is in the bounded set $\mathcal{B}^{DS} = \{z : \|\Phi'z\|_\infty \leq \lambda\}$.

Theorem 3.4: Consider the model (1) with z satisfying $\|\Phi'z\|_\infty \leq \lambda$. Suppose β is k-sparse and $\hat{\beta}$ is the minimizer

$$\hat{\beta} = \arg\min_{\gamma \in \mathbb{R}^p}\{\|\gamma\|_1 : \|\Phi'(y - \Phi\gamma)\|_\infty \leq \lambda\}.$$

Then

$$\|\hat{\beta} - \beta\|_2 \leq \frac{\sqrt{k}}{0.307 - \delta_k}\lambda$$

provided $\delta_k < 0.307$.

The proof of this theorem can be easily obtained based on a minor modification of the proof of Theorem 3.1.

IV. UPPER BOUNDS OF δ_k

We have established the sufficient condition

$$\delta_k < 0.307$$

for sparse recovery in the previous section. It is interesting to know the limit of possible improvement within this framework. In this section, we shall show that this bound cannot be substantially improved. An explicitly example is constructed in which $\delta_k = \frac{k-1}{2k-1} < 0.5$, but it is impossible to recover certain k-sparse signals. Therefore, the bound for δ_k cannot go beyond 0.5 in general in order to guarantee stable recovery of k-sparse signals. In the special case of $k = 2$, $\frac{k-1}{2k-1} = \frac{1}{3} \approx 0.333$ and so the upper and lower bounds on the RIP are very close in this case.

This question was considered for the case of δ_{2k}. In [3], among a family of recovery conditions, it is shown that

$$\delta_{2k} < 0.472$$

is sufficient for reconstructing k-sparse signals. On the other hand, the results of Davies and Gribonval [10] indicate that $\frac{1}{\sqrt{2}} \approx 0.707$ is likely the upper bound for δ_{2k}.

Theorem 4.1: Let k be a positive integer. Then there exists a $(2k-1) \times 2k$ matrix Φ with the restricted isometry constant $\delta_k = \frac{k-1}{2k-1}$, and two nonzero k-sparse vectors β_1 and β_2 with disjoint supports such that

$$\Phi\beta_1 = \Phi\beta_2.$$

Remark 4.1: This result implies that the model (1) is not identifiable in general under the condition $\delta_k = \frac{k-1}{2k-1}$ and, therefore, not all k-sparse signals can be recovered exactly in the noiseless case. In the noisy case, it is easy to see that Theorem 3.2 fails because no estimator $\hat{\beta}$ can be close to both β_1 and β_2 when the noisy level ε is sufficiently small.

Proof: Let Γ be a $2k \times 2k$ matrix such that each diagonal element of Γ is 1 and each off diagonal element equals $-\frac{1}{2k-1}$. Then it is easy to see that Γ is a positive-semidefinite matrix with rank $2k-1$.

Note that the symmetric matrix Γ can be decomposed as $\Gamma = \Phi'\Phi$ where Φ is a $(2k-1) \times 2k$ matrix with rank $2k-1$. More precisely, since Γ has two distinct eigenvalues $\frac{2k}{2k-1}$ and 0, with the multiplicities of $2k-1$ and 1 respectively, there is an orthogonal matrix U such that

$$\Gamma = U\mathrm{Diag}\left\{\underbrace{\frac{2k}{2k-1}, \frac{2k}{2k-1}, \ldots, \frac{2k}{2k-1}}_{2k-1}, 0\right\} U'.$$

Define Φ as

$$\Phi = \begin{pmatrix} \sqrt{\frac{2k}{2k-1}} & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{\frac{2k}{2k-1}} & \cdots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \cdots & \sqrt{\frac{2k}{2k-1}} & 0 \end{pmatrix} U'.$$

Let $T \subset \{1, 2, \ldots, 2k\}$ with $|T| = k$. Then it can be verified that

$$\Phi_T'\Phi_T = \begin{pmatrix} 1 & -\frac{1}{2k-1} & \cdots & -\frac{1}{2k-1} \\ -\frac{1}{2k-1} & 1 & \cdots & -\frac{1}{2k-1} \\ & & \ddots & \\ -\frac{1}{2k-1} & -\frac{1}{2k-1} & \cdots & 1 \end{pmatrix}_{k \times k}.$$

The characteristic polynomial of $\Phi_T'\Phi_T$ is

$$p(\lambda) = \left(\lambda - \frac{k}{2k-1}\right)\left(\lambda - \frac{2k}{2k-1}\right)^{k-1}.$$

This shows that for Φ,

$$\delta_k(\Phi) = 1 - \frac{k}{2k-1} = \frac{k-1}{2k-1}.$$

Since the rank of Φ is $2k-1$, there exists some $\gamma \in \mathbb{R}^{2k}$ such that $\gamma \neq 0$ and $\Phi\gamma = 0$. Suppose β_1, $\beta_2 \in \mathbb{R}^{2k}$ are given by

$$\beta_1 = (\gamma(1), \gamma(2), \ldots, \gamma(k), 0, \ldots, 0)'$$

and

$$\beta_2 = \left(\underbrace{0, 0, \ldots, 0}_{k}, -\gamma(k+1), -\gamma(k+2), \ldots, -\gamma(2k)\right)'.$$

Then both β_1 and β_2 are k-sparse vectors but $\Phi\beta_1 = \Phi\beta_2$. This means the model is not identifiable within the class of k-sparse signals. ∎

APPENDIX A
COMPLETION OF THE PROOF OF THEOREM 3.2

In Theorem 3.1, let $k_1 = k, 1 \leq k_2 < k$, and

$$t = \sqrt{\frac{k}{k_2}} + \frac{1}{4}\sqrt{\frac{k_2}{k}}.$$

Then under the condition $\delta_k + t\theta_{k,k_2} < 1$, we have

$$\|\beta - \hat{\beta}\|_2 \leq \frac{2\sqrt{2}\sqrt{1+\delta_k}}{1 - \delta_k - t\theta_{k,k_2}}\epsilon.$$

By the square root lifting inequality

$$\begin{aligned} \delta_k + t\theta_{k,k_2} &\leq \delta_k + t\theta_{\frac{k}{k-k_2}(k-k_2),k_2} \\ &\leq \delta_k + t\sqrt{\frac{k}{k-k_2}}\theta_{k-k_2,k_2} \\ &\leq \left(1 + t\sqrt{\frac{k}{k-k_2}}\right)\delta_k. \end{aligned}$$

Denote $A_k = 1 + t\sqrt{\frac{k}{k-k_2}}$ and let

$$f(x) = 1 + \frac{1}{\sqrt{1-x}}\left(\frac{1}{\sqrt{x}} + \frac{1}{4}\sqrt{x}\right) \quad x \in (0, 1)$$

then

$$A_k = f\left(\frac{k_2}{k}\right).$$

Since $f'(x) = \frac{9x-4}{8(x-x^2)^{\frac{3}{2}}}$, f is increasing when $\frac{4}{9} \leq x < 1$ and decreasing when $0 < x < \frac{4}{9}$.

Let $0 \leq r_k \leq 8$ be the integer such that $r_k \equiv 4k \pmod 9$. Now we choose k_2 specifically as follows:

$$k_2 = \begin{cases} \lfloor \frac{4}{9}k \rfloor, & \text{if } r_k \leq 4 \\ \lceil \frac{4}{9}k \rceil, & \text{if } r_k > 4. \end{cases}$$

By the definition of k_2 we get immediately that

$$A_k \leq \max\left(f\left(\frac{4}{9} + \frac{4}{9k}\right), f\left(\frac{4}{9} - \frac{4}{9k}\right)\right).$$

It can be easily verified that $f(\frac{8}{21}) = f(\frac{4}{9} - \frac{4}{9\cdot7}) > f(\frac{4}{9} + \frac{4}{9\cdot7})$. From this fact we get for $k \geq 7$

$$
\begin{aligned}
A_k &\leq \max\left(f\left(\frac{4}{9} + \frac{4}{9\cdot7}\right), f\left(\frac{4}{9} - \frac{4}{9\cdot7}\right) \right) \\
&= f\left(\frac{8}{21}\right) = 1 + \frac{23}{2\sqrt{26}} \\
&< 3.256.
\end{aligned}
$$

A direct calculation shows that

$$
A_4 = A_6 = f(0.5) = 3.25, \text{ and } A_5 = f(0.4) < 3.246.
$$

In order to estimate A_k for $k = 2, 3$, we note from the remark of Theorem 3.1 that in these cases $k_2 = 1$ and $t = \sqrt{k}$. So

$$
A_2 = 1 + \sqrt{2}\sqrt{2} = 3, A_3 = 1 + \sqrt{3}\sqrt{\frac{3}{2}} = 3.122.
$$

These yield

$$
\begin{aligned}
\delta_k + t\theta_{k,\lceil\frac{4}{9}k\rceil} &\leq A_k\delta_k \\
&\leq 3.256 \cdot \delta_k < 1.
\end{aligned}
$$

With the above relation, we can also get

$$
\begin{aligned}
\|\beta - \hat{\beta}\|_2 &\leq \frac{2\sqrt{2}\sqrt{1+\delta_k}}{1 - \delta_k - t\theta_{k,k_2}}\varepsilon \leq \frac{2\sqrt{2}\sqrt{1+\delta_k}}{1 - 3.256 \cdot \delta_k}\varepsilon \\
&\leq \frac{3.256}{1 - 3.256\delta_k}\varepsilon \leq \frac{\varepsilon}{0.307 - \delta_k}.
\end{aligned}
$$

The theorem is proved.   ■

ACKNOWLEDGMENT

The authors would like to thank the referees and the Associate Editor for thorough and useful comments which have helped to improve the presentation of the paper.

REFERENCES

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, 2008.

[2] T. Cai, L. Wang, and G. Xu, *IEEE Trans. Inf. Theory*, vol. 56, pp. 3516–3522, 2010.

[3] T. Cai, L. Wang, and G. Xu, "Shifting inequality and recovery of sparse signals," *IEEE Trans. Signal Process.*, vol. 58, pp. 1300–1308, 2010.

[4] T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via ℓ_1 minimization," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3388–3397, 2009.

[5] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," Compte Rendus de l' Academie des Sciences. Paris, France, ser. I, vol. 346, pp. 589–592.

[6] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006.

[7] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, 2005.

[8] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, 2006.

[9] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n (with discussion)," *Ann. Statist.*, vol. 35, pp. 2313–2351, 2007.

[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*, to be published.

[11] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[12] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, pp. 6–18, 2006.

[13] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2845–2862, 2001.

[14] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26/3, pp. 395–407, 2009.

[15] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," presented at the 40th Annu. Conf. Information Sciences and Systems, 2006.

T. Tony Cai received the Ph.D. degree from Cornell University, Ithaca, NY, in 1996.

His research interests include high-dimensional inference, large-scale multiple testing, nonparametric function estimation, functional data analysis, and statistical decision theory. He is the Dorothy Silberberg Professor of Statistics at the Wharton School of the University of Pennsylvania, Philadelphia.

Dr. Cai is the recipient of the 2008 COPSS Presidents' Award and a fellow of the Institute of Mathematical Statistics. He is also the current editor of the *Annals of Statistics*

Lie Wang received the Ph.D. degree in statistics from the University of Pennsylvania, Philadelphia, in 2008.

He is now with the Department of Mathematics, Massachusetts Institute of Technology, Cambridge. His research interests include nonparametric function estimation, high-dimensional sparse regression, semi-parametric models, and functional data regression.

Guangwu Xu received the Ph.D. degree in mathematics from the State University of New York at Buffalo.

He is now with the Department of Electrical Engineering and Computer Science, University of Wisconsin, Milwaukee. His research interests include compressed sensing, cryptography and information security, computational number theory, algorithms, and functional analysis.