On Recovery of Sparse Signals via ℓ_1 Minimization

T. Tony Cai* Guangwu Xu† and Jun Zhang‡

Abstract

This article considers constrained ℓ_1 minimization methods for the recovery of high dimensional sparse signals in three settings: noiseless, bounded error and Gaussian noise. A unified and elementary treatment is given in these noise settings for two ℓ_1 minimization methods: the Dantzig selector and ℓ_1 minimization with an ℓ_2 constraint. The results of this paper improve the existing results in the literature by weakening the conditions and tightening the error bounds. The improvement on the conditions shows that signals with larger support can be recovered accurately. This paper also establishes connections between restricted isometry property and the mutual incoherence property. Some results of Candes, Romberg and Tao (2006) and Donoho, Elad, and Temlyakov (2006) are extended.

Keywords: Dantzig selector, ℓ_1 minimization, Lasso, overcomplete representation, sparse recovery, sparsity.

1 Introduction

The problem of recovering a high-dimensional sparse signal based on a small number of measurements, possibly corrupted by noise, has attracted much recent attention. This problem arises in many different settings, including model selection in linear regression, constructive approximation, inverse problems, and compressive sensing.

Suppose we have n observations of the form

$$y = F\beta + z \tag{1.1}$$

*Department of Statistics, The Wharton School, University of Pennsylvania, PA, USA; e-mail: tcai@wharton.upenn.edu. Research supported in part by NSF Grant DMS-0604954.

†Department of EE & CS, University of Wisconsin-Milwaukee, WI, USA; e-mail: gxu4uwm@uwm.edu

‡Department of EE & CS, University of Wisconsin-Milwaukee, WI, USA; e-mail: junzhang@uwm.edu

where the matrix $F \in \mathbb{R}^{n \times p}$ with $n \ll p$ is given and $z \in \mathbb{R}^n$ is a vector of measurement errors. The goal is to reconstruct the unknown vector $\beta \in \mathbb{R}^p$. Depending on settings, the error vector z can either be zero (in the noiseless case), bounded, or Gaussian where $z \sim N(0, \sigma^2 I_n)$. It is now well understood that ℓ_1 minimization provides an effective way for reconstructing a sparse signal in all three settings.

A special case of particular interest is when no noise is present in (1.1) and $y = F\beta$. This is an underdetermined system of linear equations with more variables than the number of equations. It is clear that the problem is ill-posed and there are generally infinite many solutions. However, in many applications the vector β is known to be sparse or nearly sparse in the sense that it contains only a small number of nonzero entries. This sparsity assumption fundamentally changes the problem, making unique solution possible. Indeed in many cases the unique sparse solution can be found exactly through ℓ_1 minimization:

$$(P) \quad \min \|\gamma\|_1 \quad \text{subject to} \quad F\gamma = y. \tag{1.2}$$

This ℓ_1 minimization problem has been studied, for example, in Fuchs [11], Candes and Tao [4] and Donoho [6]. Understanding the noiseless case is not only of significant interest on its own right, it also provides deep insight into the problem of reconstructing sparse signals in the noisy case. See, for example, Candes and Tao [4, 5] and Donoho [6, 7].

When noise is present, there are two well known ℓ_1 minimization methods. One is ℓ_1 minimization under the ℓ_2 constraint on the residuals:

$$(P_1) \quad \min \|\gamma\|_1 \quad \text{subject to} \quad \|y - F\gamma\|_2 \leq \epsilon. \tag{1.3}$$

Writing in terms of the Lagrangian function of (P_1), this is closely related to finding the solution to the ℓ_1 regularized least squares:

$$\min_{\gamma} \left\{ \|y - F\gamma\|_2^2 + \rho \|\gamma\|_1 \right\}. \tag{1.4}$$

The latter is often called the Lasso in the statistics literature (Tibshirani [13]). Tropp [14] gave a detailed treatment of the ℓ_1 regularized least squares problem.

Another method, called the Dantzig selector, is recently proposed by Candes and Tao [5]. The Dantzig selector solves the sparse recovery problem through ℓ_1-minimization with a constraint on the correlation between the residuals and the column vectors of F:

$$(DS) \quad \min_{\gamma} \|\gamma\|_1 \quad \text{subject to} \quad \|F^T(y - F\gamma)\|_\infty \leq \lambda. \tag{1.5}$$

Candes and Tao [5] showed that the Dantzig selector can be computed by solving a linear program and it mimics the performance of an oracle procedure up to a logarithmic factor $\log p$.

It is clear that regularity conditions are needed in order for these problems to be well behaved. Over the last few years, many interesting results for recovering sparse signals have been obtained in the framework of the *Restricted Isometry Property* (RIP). In their seminal work [4, 5], Candes and Tao considered sparse recovery problems in the RIP framework. They provided beautiful solutions to the problem under some conditions on the restricted isometry constant and restricted orthogonality constant (defined in Section 2). Several different conditions have been imposed in various settings.

In this paper, we consider ℓ_1 minimization methods for the sparse recovery problem in three cases: noiseless, bounded error and Gaussian noise. Both the Dantzig selector (DS) and ℓ_1 minimization under the ℓ_2 constraint (P_1) are considered. We give a unified and elementary treatment for the two methods under the three noise settings. Our results improve on the existing results in [2, 3, 4, 5] by weakening the conditions and tightening the error bounds. In all cases we solve the problems under the weaker condition

$$\delta_{1.5k} + \theta_{k,1.5k} < 1$$

where k is the sparsity index and δ and θ are respectively the restricted isometry constant and restricted orthogonality constant defined in Section 2. The improvement on the condition shows that signals with larger support can be recovered. Although our main interest is on recovering sparse signals, we state the results in the general setting of reconstructing an arbitrary signal.

Another widely used condition for sparse recovery is the so called *Mutual Incoherence Property* (MIP) which requires the pairwise correlations among the column vectors of F to be small. See [8, 9, 11, 12, 14]. We establish connections between the concepts of RIP and MIP. As an application, we present an improvement to a recent result of Donoho, Elad, and Temlyakov [8].

The paper is organized as follows. In Section 2, after basic notation and definitions are reviewed, two elementary inequalities, which allow us to make finer analysis of the sparse recovery problem, are introduced. We begin the analysis of ℓ_1 minimization methods for sparse recovery by considering the exact recovery in the noiseless case in Section 3. Our result improves the main result in Candes and Tao [4] by using weaker conditions and providing tighter error bounds. The analysis of the noiseless case provides insight to the case when the observations are contaminated by noise. We then consider the case of bounded error in Section 4. The connections between the RIP and MIP are also explored. The case of Gaussian noise is treated in Section 5. The Appendix contains the proofs of some technical results.

2 Preliminaries

In this section we first introduce basic notation and definitions, and then develop some technical inequalities which will be used in proving our main results.

Let $p \in \mathbb{N}$. Let $v = (v_1, v_2, \cdots, v_p) \in \mathbb{R}^p$ be a vector. The support of v is the subset of $\{1, 2, \cdots, p\}$ defined by

$$\text{supp}(v) = \{i : v_i \neq 0\}.$$

For an integer $k \in \mathbb{N}$, a vector v is said to be *k-sparse* if $|\text{supp}(v)| \leq k$. For a given vector v 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 shall use the standard notation $\|v\|_q$ to denote the ℓ_q-norm of the vector v.

Let the matrix $F \in \mathbb{R}^{n \times p}$ and $1 \leq k \leq p$, the *k-restricted isometry constant* δ_k of F is defined to be the smallest constant such that

$$\sqrt{1-\delta_k}\|c\|_2 \leq \|Fc\|_2 \leq \sqrt{1+\delta_k}\|c\|_2 \tag{2.1}$$

for every vector c which is k-sparse. If $k + k' \leq p$, we can define another quantity, the *k, k'-restricted orthogonality constant* $\theta_{k,k'}$, as the smallest number that satisfies

$$|\langle Fc, Fc' \rangle| \leq \theta_{k,k'}\|c\|_2\|c'\|_2, \tag{2.2}$$

for all c and c' such that c and c' are k-sparse and k'-sparse respectively, and have disjoint supports. Candes and Tao [4] showed that the constants δ_k and $\theta_{k,k'}$ are related by the following inequalities,

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

Another useful property is as follows.

Proposition 2.1 *If $k + \sum_{i=1}^{l} k_i \leq p$, then*

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

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

Proof of Proposition 2.1. Let c be k-sparse and c' be $(\sum_{i=1}^{l} k_i)$-sparse. Suppose their supports are disjoint. Decompose c' as

$$c' = c'_1 + c'_2 + \cdots + c'_l$$

such that c'_i is k_i-sparse for $i = 1, \cdots, j$ and $\text{supp}(c')_i \cap \text{supp}(c')_j = \emptyset$ for $i \neq j$. We have

$$\begin{aligned} |\langle Fc, Fc' \rangle| &= |\langle Fc, \sum_{i=1}^{l} Fc'_i \rangle| \leq \sum_{i=1}^{l} |\langle Fc, Fc'_i \rangle| \\ &\leq \sum_{i=1}^{l} \theta_{k,k_i} \|c\|_2 \|c'_i\|_2 = \|c\|_2 \sqrt{\sum_{i=1}^{l} \theta_{k,k_i}^2} \sqrt{\sum_{i=1}^{l} \|c'_i\|_2^2} \\ &= \sqrt{\sum_{i=1}^{l} \theta_{k,k_i}^2} \|c\|_2 \|c'\|_2. \end{aligned}$$

This yields $\theta_{k,\sum_{i=1}^{l} k_i} \leq \sqrt{\sum_{i=1}^{l} \theta_{k,k_i}^2}$. Since $\theta_{k,k'} \leq \delta_{k+k'}$, we also have $\theta_{k,\sum_{i=1}^{l} k_i} \leq \sqrt{\sum_{i=1}^{l} \delta_{k+k_i}^2}$. ∎

Remark: Different conditions on δ and θ have been used in the literature. For example, Candes and Tao [5] imposes $\delta_{2k} + \theta_{k,2k} < 1$ and Candes [2] uses $\delta_{2k} < \sqrt{2} - 1$. A direct consequence of Proposition 2.1 is that $\delta_{2k} < \sqrt{2} - 1$ is in fact a strictly stronger condition than $\delta_{2k} + \theta_{k,2k} < 1$ since Proposition 2.1 yields $\theta_{k,2k} \leq \sqrt{\delta_{2k}^2 + \delta_{2k}^2} = \sqrt{2}\delta_{2k}$ which means that $\delta_{2k} < \sqrt{2} - 1$ implies $\delta_{2k} + \theta_{k,2k} < 1$.

We now introduce two useful elementary inequalities. These inequalities allow us to perform finer estimation on ℓ_1, l_2 norms.

Proposition 2.2 *Let w be a positive integer. For any descending chain of real numbers*

$$a_1 \geq a_2 \geq \cdots \geq a_w \geq a_{w+1} \geq \cdots \geq a_{2w} \geq 0,$$

we have

$$\sqrt{a_{w+1}^2 + a_{w+2}^2 + \cdots + a_{2w}^2} \leq \frac{a_1 + a_2 + \cdots + a_w + a_{w+1} + \cdots + a_{2w}}{2\sqrt{w}}.$$

Proof of Proposition 2.2. Since $a_i \geq a_j$ for $i < j$, we have

$$\begin{aligned} (a_1 + a_2 + \cdots + a_{2w})^2 &= a_1^2 + a_2^2 + \cdots a_{2w}^2 + 2\sum_{i<j} a_i a_j \\ &\geq a_1^2 + a_2^2 + \cdots a_{2w}^2 + 2\sum_{i<j} a_j^2 \\ &= a_1^2 + 3a_2^2 + \cdots + (2w-1)a_w^2 + \\ &\quad +(2w+1)a_{w+1}^2 + \cdots + (4w-3)a_{2w-1}^2 + (4w-1)a_{2w}^2 \\ &= \big(a_1^2 + (4w-1)a_{2w}^2\big) + \big(3a_2^2 + (4w-3)a_{2w-1}^2\big) + \cdots \\ &\quad +\big((2w-1)a_w^2 + (2w+1)a_{w+1}^2\big) \\ &\geq 4wa_{2w}^2 + 4wa_{2w-1}^2 + \cdots 4wa_{w+1}^2. \end{aligned}$$ ∎

Proposition 2.2 can be used to improve the main result in Candes and Tao [5] by weakening the condition to $\delta_{1.75k} + \theta_{k,1.75k} < 1$. However, the next proposition, which we will use in proving our main results, is more powerful for our applications.

Proposition 2.3 *Let w be a positive integer. Then any descending chain of real numbers*

$$a_1 \geq a_2 \geq \cdots \geq a_w \geq a_{w+1} \geq \cdots \geq a_{3w} \geq 0$$

satisfies

$$\sqrt{a_{w+1}^2 + a_{w+2}^2 + \cdots + a_{3w}^2} \leq \frac{a_1 + \cdots + a_w + 2(a_{w+1} + \cdots + a_{2w}) + a_{2w+1} + \cdots + a_{3w}}{2\sqrt{2w}}.$$

The proof of Proposition 2.3 is given in the Appendix.

3 Signal Recovery in the Noiseless Case

As mentioned in the introduction we shall consider recovery of sparse signals in three cases: noiseless, bounded error, and Gaussian noise. We begin in this section by considering the problem of exact recovery of sparse signals when no noise is present. This is an interesting problem by itself and has been considered in a number of papers. See, for example, Fuchs [11], Donoho [6], and Candes and Tao [4]. More importantly, the solutions to this "clean" problem shed light on the noisy case. Our result improves the main result given in Candes and Tao [4]. The improvement is obtained by using the technical inequalities we developed in previous section. Although the focus is on recovering sparse signals, our results are stated in the general setting of reconstructing an arbitrary signal.

Let $F \in \mathbb{R}^{n\times p}$ with $n < p$ and suppose we are given F and y where $y = F\beta$ for some unknown vector β. The goal is to recover β exactly when it is sparse. Candes and Tao [4] showed that a sparse solution can be obtained by ℓ_1 minimization which is then solved via linear programming.

Theorem 3.1 (Candes and Tao [4]) *Let $F \in \mathbb{R}^{n\times p}$. Suppose $k \geq 1$ satisfies*

$$\delta_k + \theta_{k,k} + \theta_{k,2k} < 1. \tag{3.1}$$

Let β be a k-sparse vector and $y := F\beta$. Then β is the unique minimizer to the problem

$$(P) \quad \min \|\gamma\|_1 \quad \textit{subject to} \quad F\gamma = y.$$

We shall show that this result can be further improved by a transparent argument. A direct application of Proposition 2.3 yields the following result which improves Theorem 3.1. by weakening the condition from

$$\delta_k + \theta_{k,k} + \theta_{k,2k} < 1,$$

to

$$\delta_{1.5k} + \theta_{k,1.5k} < 1.$$

Theorem 3.2 *Let $F \in \mathbb{R}^{n\times p}$. Suppose $k \geq 1$ satisfies*

$$\delta_{1.5k} + \theta_{k,1.5k} < 1$$

and $y = F\beta$. Then the minimizer $\hat{\beta}$ to the problem

$$(P) \quad \min \|\gamma\|_1 \quad \textit{subject to} \quad F\gamma = y$$

obeys

$$\|\hat{\beta} - \beta\|_2 \leq C_0 k^{-\frac{1}{2}} \|\beta_{-\max(k)}\|_1$$

where $C_0 = \frac{2\sqrt{2}(1-\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}}$.

In particular, if β is a k-sparse vector, then $\hat{\beta} = \beta$, i.e., the ℓ_1 minimization recovers β exactly.

Proof of Theorem 3.2: The proof relies on Proposition 2.3 and makes use of the ideas from [3, 4, 5]. In this proof, we shall also identify a vector $v = (v_1, v_2, \cdots, v_p) \in \mathbb{R}^p$ as a function $v : \{1, 2, \cdots, p\} \to \mathbb{R}$ by assigning $v(i) = v_i$.

Let $\hat{\beta}$ be a solution to the ℓ_1 minimization problem (P). Let $T_0 = \{n_1, n_2, \cdots, n_k\} \subset \{1, 2, \cdots, p\}$ be the support of $\beta_{\max(k)}$ and let $h = \hat{\beta} - \beta$. Write

$$\{1, 2, \cdots, p\} \setminus \{n_1, n_2, \cdots, n_k\} = \{n_{k+1}, n_{k+2}, \cdots, n_p\}$$

such that $|h(n_{k+1})| \geq |h(n_{k+2})| \geq |h(n_{k+3})| \geq \cdots$. Fix an integer $t > 0$ and let

$$T_1 = \{n_{k+1}, n_{k+2}, \cdots, n_{(t+1)k}\},\ T_2 = \{n_{(t+1)k+1}, n_{(t+1)k+2}, \cdots, n_{(2t+1)k}\}, \cdots.$$

For a subset $E \subset \{1, 2, \cdots, m\}$, we use I_E to denote the characteristic function of E, i.e.,

$$I_E(j) = \begin{cases} 1 & \text{if } j \in E, \\ 0 & \text{if } j \notin E. \end{cases}$$

For each i, let $h_i = hI_{T_i}$. Then h is decomposed to $h = h_0 + h_1 + h_2 + \cdots$. Note that T_i's are pairwise disjoint, $\mathrm{supp}(h_i) \subset T_i$, and $|T_0| = k, |T_i| = tk$ for $i > 0$. Without loss of generality, we assume k is divisible by 4.

For each $i > 1$, we divide h_i into two halves in the following manner

$$h_i = h_{i1} + h_{i2} \text{ with } h_{i1} = h_i I_{T_{i1}}, \text{ and } h_{i2} = h_i I_{T_{i2}},$$

where T_{i1} is the first half of T_i, i.e.,

$$T_{i1} = \{n_{((i-1)t+1)k+1}, n_{((i-1)t+1)k+2}, \cdots, n_{((i-1)t+1)k+\frac{k}{2}}\},$$

and $T_{i2} = T_i \setminus T_{i1}$.

We shall treat h_1 as a sum of four functions and divide T_1 into 4 equal parts $T_1 = T_{11} \cup T_{12} \cup T_{13} \cup T_{14}$ with

$$T_{11} = \{n_{k+1}, n_{k+2}, \cdots, n_{k+t\frac{k}{4}}\},\ T_{12} = \{n_{k+t\frac{k}{4}+1}, \cdots, n_{k+t\frac{k}{2}}\},$$

$$T_{13} = \{n_{k+t\frac{k}{2}+1}, \cdots, n_{k+t\frac{3k}{4}}\} \text{ and } T_{14} = \{n_{k+t\frac{3k}{4}+1}, \cdots, n_{k+tk}\}.$$

We then define h_{1i} for $1 \le i \le 4$ by $h_{1i}(j) = h_1 I_{T_{1i}}$. It is clear that $h_1 = \sum_{i=1}^{4} h_{1i}$.

Note that

$$\sum_{i\ge 1} \|h_i\|_1 \le \|h_0\|_1 + 2\|\beta_{-\max(k)}\|_1. \tag{3.2}$$

In fact, since $\|\beta\|_1 \ge \|\hat{\beta}\|_1$, we have

$$\begin{aligned}\|\beta\|_1 &\ge \|\hat{\beta}\|_1 = \|\beta + h\|_1 = \|\beta_{\max(k)} + h_0\|_1 + \|h - h_0 + \beta_{-\max(k)}\|_1 \\ &\ge \|\beta_{\max(k)}\|_1 - \|h_0\|_1 + \sum_{i\ge 1}\|h_i\|_1 - \|\beta_{-\max(k)}\|_1.\end{aligned}$$

Since $\|\beta\|_1 = \|\beta_{\max(k)}\|_1 + \|\beta_{-\max(k)}\|_1$, this yields $\sum_{i\ge 1}\|h_i\|_1 \le \|h_0\|_1 + 2\|\beta_{-\max(k)}\|_1$.

The following claim follows from our Proposition 2.3.

Claim

$$\|h_{13} + h_{14}\|_2 + \sum_{i\ge 2}\|h_i\|_2 \le \frac{\sum_{i\ge 1}\|h_i\|_1}{\sqrt{tk}} \le \frac{\|h_0\|_2}{\sqrt{t}} + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{tk}}. \tag{3.3}$$

In fact, from Proposition 2.3 and the fact that $\|h_{11}\|_1 \ge \|h_{12}\|_1 \ge \|h_{13}\|_1 \ge \|h_{14}\|_1$, we have

$$\|h_{12}\|_1 + 2\|h_{13}\|_1 + \|h_{14}\|_1 \le \frac{2}{3}\big(2\|h_{11}\|_1 + 2\|h_{12}\|_1 + \|h_{13}\|_1 + \|h_{14}\|_1\big).$$

It then follows from Proposition 2.3 that

$$\begin{aligned}
\|h_{13}+h_{14}\|_2 &\le \frac{\|h_{12}\|_1+2\|h_{13}\|_1+\|h_{14}\|_1}{2\sqrt{\frac{tk}{2}}} \\
&\le \frac{2}{3}\frac{2\|h_{11}\|_1+2\|h_{12}\|_1+\|h_{13}\|_1+\|h_{14}\|_1}{2\sqrt{\frac{tk}{2}}} \\
&\le \frac{2\|h_{11}\|_1+2\|h_{12}\|_1+\|h_{13}\|_1+\|h_{14}\|_1}{2\sqrt{tk}}.
\end{aligned}$$

Proposition 2.3 also yields

$$\|h_2\|_2 \le \frac{\|h_{13}+h_{14}\|_1+2\|h_{21}\|_1+\|h_{22}\|_1}{2\sqrt{tk}}$$

and

$$\|h_i\|_2 \le \frac{\|h_{(i-1)2}\|_1+2\|h_{i1}\|_1+\|h_{i2}\|_1}{2\sqrt{tk}}$$

for any $i > 2$. Therefore,

$$\begin{aligned}
\|h_{13}+h_{14}\|_2+\sum_{i\ge 2}\|h_i\|_2 &\le \frac{2\|h_{11}\|_1+2\|h_{12}\|_1+\|h_{13}\|_1+\|h_{14}\|_1}{2\sqrt{tk}} \\
&\quad +\frac{\|h_{13}+h_{14}\|_1+2\|h_{21}\|_1+\|h_{22}\|_1}{2\sqrt{tk}} \\
&\quad +\frac{\|h_{22}\|_1+2\|h_{31}\|_1+\|h_{32}\|_1}{2\sqrt{tk}}+\cdots \\
&\le \frac{2\|h_1\|_1+2\|h_2\|_1+2\|h_3\|_1+\cdots}{2\sqrt{tk}} \\
&= \frac{\sum_{i\ge 1}\|h_i\|_1}{\sqrt{tk}} \\
&\overset{\text{by (3.2)}}{\le} \frac{\|h_0\|_1+2\|\beta_{-\max(k)}\|_1}{\sqrt{tk}} \le \frac{\|h_0\|_2}{\sqrt{t}}+\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{tk}}.
\end{aligned}$$

In the rest of our proof we write $h_{11}+h_{12}=h_1'$. Note that $Fh=F\hat\beta-F\beta=0$. So

$$\begin{aligned}
0 &= |\langle Fh, F(h_0+h_1')\rangle| \\
&= |\langle F(h_0+h_1'), F(h_0+h_1')\rangle + \langle F(h_{13}+h_{14}), F(h_0+h_1')\rangle + \sum_{i\ge 2}\langle Fh_i, F(h_0+h_1')\rangle| \\
&\overset{(2.1,2.2)}{\ge} (1-\delta_{(\frac{1}{2}t+1)k})\|h_0+h_1'\|_2^2 - \theta_{\frac{1}{2}tk,(\frac{1}{2}t+1)k}\|h_{13}+h_{14}\|_2\|h_0+h_1'\|_2 \\
&\quad -\sum_{i\ge 2}\theta_{tk,(\frac{1}{2}t+1)k}\|h_i\|_2\|h_0+h_1'\|_2 \\
&\ge \|h_0+h_1'\|_2\Big((1-\delta_{(\frac{1}{2}t+1)k})\|h_0+h_1'\|_2 - \theta_{tk,(\frac{1}{2}t+1)k}\big(\|h_{13}+h_{14}\|_2+\sum_{i\ge 2}\|h_i\|_2\big)\Big) \\
&\overset{(3.3)}{\ge} \|h_0+h_1'\|_2\Big((1-\delta_{(\frac{1}{2}t+1)k})\|h_0+h_1'\|_2 - \theta_{tk,(\frac{1}{2}t+1)k}\frac{\|h_0\|_2}{\sqrt{t}} - \theta_{tk,(\frac{1}{2}t+1)k}\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{tk}}\Big) \\
&\ge \|h_0+h_1'\|_2\left\{\left(1-\delta_{(\frac{1}{2}t+1)k}-\frac{\theta_{tk,(\frac{1}{2}t+1)k}}{\sqrt{t}}\right)\|h_0+h_1'\|_2 - \theta_{tk,(\frac{1}{2}t+1)k}\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{tk}}\right\}.
\end{aligned}$$

Take $t=1$. Then

$$\|h_0+h_1'\|_2 \le \frac{2\theta_{k,1.5k}}{1-\delta_{1.5k}-\theta_{k,1.5k}}k^{-\frac{1}{2}}\|\beta_{-\max(k)}\|_1$$

It then follows from (3.3) that

$$\begin{aligned}
\|h\|_2^2 &= \|h_0+h_1'\|_2^2+\|h_{13}+h_{14}\|_2^2+\sum_{i\ge 2}\|h_i\|_2^2 \le \|h_0+h_1'\|_2^2 + (\|h_{13}+h_{14}\|_2+\sum_{i\ge 2}\|h_i\|_2)^2 \\
&\le 2(\|h_0+h_1'\|_2+2k^{-\frac{1}{2}}\|\beta_{-\max(k)}\|_1)^2 \le 2\left(\frac{2(1-\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}}k^{-\frac{1}{2}}\|\beta_{-\max(k)}\|_1\right)^2. \quad \blacksquare
\end{aligned}$$

Remarks.

1. Candes and Tao [5] considers the Gaussian noise case. A special case with noise level $\sigma=0$ of Theorem 1.1 in that paper improves Theorem 3.1 by weakening the condition from $\delta_k+\theta_{k,k}+\theta_{k,2k}<1$ to $\delta_{2k}+\theta_{k,2k}<1$.

2. This theorem improves the results in [4, 5]. The condition $\delta_{1.5k}+\theta_{k,1.5k}<1$ is weaker than $\delta_k+\theta_{k,k}+\theta_{k,2k}<1$ and $\delta_{2k}+\theta_{k,2k}<1$.

3. Note that the condition $\delta_{1.75k}<\sqrt{2}-1$ implies $\delta_{1.5k}+\theta_{k,1.5k}<1$. This is due to the fact $\delta_{1.5k}+\theta_{k,1.5k}\le\delta_{1.5k}+\sqrt{\delta_{1.75k}^2+\delta_{1.75k}^2}\le(\sqrt{2}+1)\delta_{1.75k}$ by Proposition 2.1. The condition $\delta_{1.5k}+\delta_{2.5k}<1$, which involves only δ, can also be used.

4. The quantity t in the proof can be any number such that $tk\in\mathbb{N}$. As pointed out in [4, 5], other values of t may be used for obtaining some interesting results.

4 Recovery of Sparse Signals in Bounded Error

We now turn to the case of bounded error. The results obtained in this setting have direct implication for the case of Gaussian noise which will be discussed in Section 5.

Let $F \in \mathbb{R}^{n \times p}$ and let

$$y = F\beta + z$$

where the noise z is bounded, i.e., $z \in \mathcal{B}$ for some bounded set $\mathcal{B}$. In this case the noise z can either be stochastic or deterministic. The ℓ_1 minimization approach is to estimate β by the minimizer $\hat{\beta}$ of

$$\min \|\gamma\|_1 \quad \text{subject to} \quad y - F\gamma \in \mathcal{B}.$$

We shall specifically consider two cases: $\mathcal{B} = \{z : \|F^T z\|_\infty \leq \lambda\}$ and $\mathcal{B} = \{z : \|z\|_2 \leq \epsilon\}$. Our results improve the results in Candes and Tao [4, 5] and Donoho, Elad and Temlyakov [8].

We shall first consider

$$y = F\beta + z \quad \text{where } z \text{ satisfies} \quad \|F^T z\|_\infty \leq \lambda.$$

Let $\hat{\beta}$ be the solution to the (DS) problem, i.e., $\hat{\beta}$ is obtained by solving

$$\min_{\gamma \in \mathbb{R}^p} \|\gamma\|_1 \qquad \text{subject to} \qquad \|F^T\big(y - F\gamma\big)\|_\infty \leq \lambda. \tag{4.1}$$

The Dantzig selector $\hat{\beta}$ has the following property.

Theorem 4.1 *Suppose $\beta \in \mathbb{R}^p$ and $y = F\beta + z$ with z satisfying $\|F^T z\|_\infty \leq \lambda$. If*

$$\delta_{1.5k} + \theta_{k,1.5k} < 1, \tag{4.2}$$

then the solution $\hat{\beta}$ to (4.1) obeys

$$\|\hat{\beta} - \beta\|_2 \leq C_1 k^{\frac{1}{2}} \lambda + C_2 k^{-\frac{1}{2}} \|\beta_{-\max(k)}\|_1 \tag{4.3}$$

with $C_1 = \frac{2\sqrt{3}}{1-\delta_{1.5k}-\theta_{k,1.5k}}$, and $C_2 = \frac{2\sqrt{2}(1-\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}}$.

In particular, if β is a k-sparse vector, then $\|\hat{\beta} - \beta\|_2 \leq C_1 k^{\frac{1}{2}} \lambda$.

Proof of Theorem 4.1 . We shall use the same notation as in the proof of Theorem 3.2. Since $\|\beta\|_1 \geq \|\hat{\beta}\|_1$, letting $h = \hat{\beta} - \beta$ and following essentially the same steps as in the first part of the proof of Theorem 3.2, we get

$$|\langle Fh, F(h_0 + h_1')\rangle| \geq \|h_0 + h_1'\|_2 \left\{ \Big(1 - \delta_{1.5k} - \theta_{k,1.5k}\Big) \|h_0 + h_1'\|_2 - \theta_{k,1.5k} \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}} \right\}.$$

If $\|h_0 + h_1'\|_2 = 0$, then $h_0 = 0$ and $h_1' = 0$. The latter forces that $h_j = 0$ for every $j > 1$, and we have $\hat{\beta} - \beta = 0$. Otherwise

$$\|h_0 + h_1'\|_2 \le \frac{|\langle Fh, F(h_0 + h_1')\rangle|}{\big(1 - \delta_{1.5k} - \theta_{k,1.5k}\big)\|h_0 + h_1'\|_2} + \frac{2\theta_{k,1.5k}\|\beta_{-\max(k)}\|_1}{\big(1 - \delta_{1.5k} - \theta_{k,1.5k}\big)\sqrt{k}}.$$

To finish the proof, we observe the following.

1. $|\langle Fh, F(h_0 + h_1')\rangle| \le \sqrt{1.5k}\, 2\lambda \|h_0 + h_1'\|_2$.

   In fact, let $F_{T_0 \cup T_{10} \cup T_{11}}$ be the $n \times (1.5k)$ submatrix obtained by extracting the columns of F according to the indices in $T_0 \cup T_{10} \cup T_{11}$, as in [5]. Then

$$\begin{aligned} |\langle Fh, F(h_0 + h_1')\rangle| &= |\langle (F\hat{\beta} - y) + z, F_{T_0 \cup T_{10} \cup T_{11}}(h_0 + h_1')\rangle| \\ &= |\langle F^T_{T_0 \cup T_{10} \cup T_{11}}\big((F\hat{\beta} - y) + z\big), h_0 + h_1'\rangle| \\ &\le \|F^T_{T_0 \cup T_{10} \cup T_{11}}\big((F\hat{\beta} - y) + z\big)\|_2 \|h_0 + h_1'\|_2 \\ &\le \sqrt{1.5k}\, 2\lambda \|h_0 + h_1'\|_2. \end{aligned}$$

2. $\|\hat{\beta} - \beta\|_2 \le \sqrt{2}\big(\|h_0 + h_1'\|_2 + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\big)$.

   In fact,

$$\begin{aligned} \|\hat{\beta} - \beta\|_2^2 &= \|h\|_2^2 = \|h_0 + h_1'\|_2^2 + \|h_{13} + h_{14}\|_2^2 + \sum_{i \ge 2} \|h_i\|_2^2 \\ &\le \|h_0 + h_1'\|_2^2 + \Big(\|h_{13} + h_{14}\|_2 + \sum_{i \ge 2} \|h_i\|_2\Big)^2 \\ &\overset{\text{by (3.3)}}{\le} \|h_0 + h_1'\|_2^2 + \left(\|h_0\|_2 + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\right)^2 \\ &\le 2\left(\|h_0 + h_1'\|_2 + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\right)^2. \end{aligned}$$

We get the result by combining 1 and 2. This completes the proof. ∎

We now turn to the second case where the noise z is bounded in ℓ_2-norm. Let $F \in \mathbb{R}^{n \times p}$ with $n < p$. The problem is to recover the sparse signal $\beta \in \mathbb{R}^p$ from

$$y = F\beta + z$$

where the noise satisfies $\|z\|_2 \le \epsilon$. We shall again consider constrained ℓ_1 minimization:

$$\min \|\gamma\|_1 \quad \text{subject to} \quad \|y - F\gamma\|_2 \le \eta.$$

By using a similar argument, we have the following result.

Theorem 4.2 *Let $F \in \mathbb{R}^{n\times p}$. Suppose $\beta \in \mathbb{R}^p$ is a k-sparse vector and $y = F\beta + z$ with $\|z\|_2 \le \epsilon$. If*

$$\delta_{1.5k} + \theta_{k,1.5k} < 1, \tag{4.4}$$

then for any $\eta \ge \epsilon$, the minimizer $\hat\beta$ to the problem

$$\min \|\gamma\|_1 \quad \textit{subject to} \quad \|y - F\gamma\|_2 \le \eta$$

obeys

$$\|\hat\beta - \beta\|_2 \le C(\eta + \epsilon) \tag{4.5}$$

with $C = \frac{\sqrt{2}(1+\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}}$.

Proof of Theorem 4.2 . Notice that the condition $\eta \ge \epsilon$ implies that $\|\hat\beta\|_1 \le \|\beta\|_1$, so we can use the first part of the proof of Theorem 3.2. The notation used here is the same as that in the proof of Theorem 3.2.

First, we have

$$\|h_0\|_1 \ge \sum_{i\ge 1} \|h_i\|_1,$$

and

$$\|h_0 + h_1'\|_2 \le \frac{|\langle Fh, F(h_0 + h_1')\rangle|}{\|h_0 + h_1'\|_2\big(1 - \delta_{1.5k} - \theta_{k,1.5k}\big)}.$$

Note that $\|Fh\|_2 = \|F(\beta - \hat\beta)\|_2 \le \|F\beta - y\|_2 + \|F\hat\beta - y\|_2 \le \eta + \epsilon$.

So

$$\begin{aligned}
\|\hat\beta - \beta\|_2 &\le \sqrt{2}\|h_0 + h_1'\|_2 \\
&\le \sqrt{2}\frac{\|Fh\|_2\|F(h_0 + h_1')\|_2}{\|h_0 + h_1'\|_2\big(1 - \delta_{1.5k} - \theta_{k,1.5k}\big)} \\
&\le \sqrt{2}\frac{(\eta+\epsilon)(1+\delta_{1.5k})\|h_0 + h_1'\|_2}{\|h_0 + h_1'\|_2\big(1 - \delta_{1.5k} - \theta_{k,1.5k}\big)} \\
&\le \frac{\sqrt{2}(\eta+\epsilon)(1+\delta_{1.5k})}{1 - \delta_{1.5k} - \theta_{k,1.5k}}. \quad \blacksquare
\end{aligned}$$

Remarks:

1. Candes, Romberg and Tao [3] showed that, if $\delta_{3k} + 3\delta_{4k} < 2$, then

$$\|\hat\beta - \beta\|_2 \le \frac{4}{\sqrt{3 - 3\delta_{4k}} - \sqrt{1 + \delta_{3k}}}\epsilon.$$

(The η was set to be ϵ in [3].) Now suppose $\delta_{3k} + 3\delta_{4k} < 2$. This implies $\delta_{3k} + \delta_{4k} < 1$ which yields $\delta_{2.4k} + \theta_{1.6k,2.4k} < 1$, since $\delta_{2.4k} \le \delta_{3k}$ and $\theta_{1.6k,2.4k} \le \delta_{4k}$. It then follows from Theorem 4.2 that, with $\eta = \epsilon$,

$$\|\hat{\beta} - \beta\|_2 \le \frac{2\sqrt{2}(1+\delta_{1.5k'})}{1-\delta_{1.5k'}-\theta_{k',1.5k'}}\epsilon$$

for all k'-sparse vector β where $k' = 1.6k$. Therefore Theorem 4.2 improves the above result in Candes, Romberg and Tao [3] by enlarging the support of β by 60%.

2. Similar to Theorems 3.2 and 4.1, we can have the estimation without assuming that $\hat{\beta}$ is k-sparse. In the general case, we have

$$\|\hat{\beta} - \beta\|_2 \le C(\eta+\epsilon) + \frac{2\sqrt{2}\theta_{k,1.5k}(1-\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}} k^{-\frac{1}{2}} \|\beta_{-\max(k)}\|_1.$$

Connections between RIP and MIP

In addition to the restricted isometry property (RIP), another commonly used condition in the sparse recovery literature is the so-called mutual incoherence property (MIP). The mutual incoherence property of F requires that the *coherence bound*

$$M = \max_{1\le i,j\le p, i\ne j} |\langle f_i, f_j\rangle| \tag{4.6}$$

be small, where $f_1, f_2, \cdots, f_p$ are the columns of F (f_i's are also assumed to be of length 1 in ℓ_2-norm). Many interesting results on sparse recovery have been obtained by imposing conditions on the coherence bound M and the sparsity k, see [8, 9, 11, 12, 14]. For example, a recent paper, Donoho, Elad, and Temlyakov [8], proved that if $\beta \in \mathbb{R}^p$ is a k-sparse vector and $y = F\beta + z$ with $\|z\|_2 \le \epsilon$, then for any $\eta \ge \epsilon$, the minimizer $\hat{\beta}$ to the problem

$$\min \|\gamma\|_1 \quad \text{subject to} \quad \|y - F\gamma\|_2 \le \eta$$

satisfies

$$\|\hat{\beta} - \beta\|_2 \le C(\eta+\epsilon).$$

with $C = \frac{1}{\sqrt{1-M(4k-1)}}$, provided $k \le \frac{1+M}{4M}$.

We shall now establish some connections between the RIP and MIP and show that the result of Donoho, Elad, and Temlyakov [8] can be improved under the RIP framework, by using Theorem 4.2.

The following is a simple result that gives RIP constants from MIP.

Proposition 4.1 *Let M be the coherence bound for F. Then*

$$\delta_k \le (k-1)M, \quad \text{and} \quad \theta_{k,k'} \le \sqrt{kk'}M. \tag{4.7}$$

Proof of Proposition 4.1 . Let c be a k-sparse vector. Without loss of generality, we assume that $\mathrm{supp}(c) = \{1, 2, \cdots, k\}$. A direct calculation shows that

$$\|Fc\|_2^2 = \sum_{i,j=1}^{k} \langle f_i, f_j\rangle c_i c_j = \|c\|_2^2 + \sum_{1\le i,j\le k, i\ne j} \langle f_i, f_j\rangle c_i c_j.$$

Now let us bound the second term. Note that

$$\begin{aligned} \Big|\sum_{1\le i,j\le k, i\ne j} \langle f_i, f_j\rangle c_i c_j\Big| &\le M \sum_{1\le i,j\le k, i\ne j} |c_i c_j| \\ &\le M(k-1)\sum_{i=1}^{k} |c_i|^2 = M(k-1)\|c\|_2^2. \end{aligned}$$

These give us

$$(1-(k-1)M)\|c\|_2^2 \le \|Fc\|_2^2 \le (1+(k-1)M)\|c\|_2^2,$$

and hence

$$\delta_k \le (k-1)M.$$

For the second inequality, we notice that $M = \theta_{1,1}$. It then follows from Proposition 2.1 that

$$\theta_{k,k'} \le \sqrt{k'}\theta_{k,1} \le \sqrt{kk'}\theta_{1,1} = \sqrt{kk'}M. \quad \blacksquare$$

Now we are able to show the following result.

Theorem 4.3 *Suppose $\beta \in \mathbb{R}^p$ is a k-sparse vector and $y = F\beta + z$ with z satisfying $\|z\|_2 \le \epsilon$. Let $kM = t$. If $t < \frac{2+2M}{3+\sqrt{6}}$ (or, equivalently, $k < \frac{2+2M}{(3+\sqrt{6})M}$), then for any $\eta \ge \epsilon$, the minimizer $\hat\beta$ to the problem*

$$\min \|\gamma\|_1 \quad \text{subject to} \quad \|y - F\gamma\|_2 \le \eta$$

obeys

$$\|\hat\beta - \beta\|_2 \le C(\eta + \epsilon). \tag{4.8}$$

with $C = \frac{\sqrt{2}(2+3t-2M)}{2+2M-(3+\sqrt{6})t}$.

Proof of Theorem 4.3 . It follows from Proposition 4.1 that

$$\delta_{1.5k} + \theta_{k,1.5k} \leq (1.5k + \sqrt{1.5}k - 1)M = (1.5 + \sqrt{1.5})t - M.$$

Since $t < \frac{2+2M}{3+\sqrt{6}}$, the condition $\delta_{1.5k} + \theta_{k,1.5k} < 1$ holds. By Theorem 4.2,

$$\begin{aligned} \|\hat{\beta} - \beta\|_2 &\leq \frac{\sqrt{2}(1 + \delta_{1.5k})}{1 - \delta_{1.5k} - \theta_{k,1.5k}}(\eta + \epsilon) \\ &\leq \frac{\sqrt{2}(1 + (1.5k - 1)M)}{1 + M - (1.5 + \sqrt{1.5})t}(\eta + \epsilon) \\ &= \frac{\sqrt{2}(2 + 3t - 2M)}{2 + 2M - (3 + \sqrt{6})t}(\eta + \epsilon). \quad \blacksquare \end{aligned}$$

Remarks. In this theorem, the result of Donoho, Elad and Temlyakov [8] is improved in the following ways.

1. The sparsity k is relaxed from $k < \frac{1+M}{4M}$ to $k < \frac{2+2M}{3+\sqrt{6}M} \approx 1.47\frac{1+M}{4M}$. So roughly speaking, Theorem 4.3 improves the result in Donoho, Elad and Temlyakov [8] by enlarging the support of β by 47%.

2. It is clear that larger t is preferred. Since M is usually very small, the bound C is tightened from $C = \frac{1}{\sqrt{1+M-4t}}$ to $C = \frac{\sqrt{2}(2+3t-2M)}{2+2M-(3+\sqrt{6})t}$, as t is close to $\frac{1}{4}$.

5 Recovery of Sparse Signals in Gaussian Noise

We now turn to the case where the noise is Gaussian. Suppose we observe

$$y = F\beta + z, \quad z \sim N(0, \sigma^2 I_n) \tag{5.1}$$

and wish to recover β from y and F. We assume that σ is known and that the columns of F are standardized to have unit ℓ_2 norm. This is a case of significant interesting, in particular in statistics. Many methods, including the Lasso (Tibshirani [13]), LARS (Efron, Hastie, Johnstone and Tibshirani [10]) and Dantzig selector (Candes and Tao [5]), have been introduced and studied.

The following results show that, with large probability, the Gaussian noise z belongs to bounded sets.

Lemma 1 *The Gaussian error $z \sim N(0, \sigma^2 I_n)$ satisfies*

$$P\left(\|F^T z\|_\infty \leq \sigma\sqrt{2\log p}\right) \geq 1 - \frac{1}{2\sqrt{\pi \log p}} \tag{5.2}$$

and

$$P\left(\|z\|_2 \le \sigma\sqrt{n+2\sqrt{n\log n}}\right) \ge 1-\frac{1}{n}. \tag{5.3}$$

Inequality (5.2) follows from standard probability calculations and inequality (5.3) is proved in the Appendix.

Lemma 1 suggests that one can apply the results obtained in the previous section for the bounded error case to solve the Gaussian noise problem. Candes and Tao [5] introduced the Dantzig selector for sparse recovery in the Gaussian noise setting. Given the observations in (5.1), the Dantzig selector $\hat\beta^{DS}$ is the minimizer of

$$(DS) \quad \min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \quad \text{subject to} \quad \|F^T(y-F\gamma)\|_\infty \le \lambda_p \tag{5.4}$$

where $\lambda_p = \sigma\sqrt{2\log p}$.

In the classical linear regression problem when $p \le n$ the least squares estimator is the solution to the normal equation

$$F^T y = F^T F\beta. \tag{5.5}$$

The constraint $\|F^T(y-F\beta)\|_\infty \le \lambda_p$ in the convex program (DS) can thus be viewed as a relaxation of the normal equation (5.5). And similar to the noiseless case ℓ_1 minimization leads to the "sparsest" solution over the space of all feasible solutions.

Candes and Tao [5] showed the following result.

Theorem 5.1 (Candes and Tao [5]) *Suppose $\beta\in\mathbb{R}^p$ is a k-sparse vector obeying*

$$\delta_{2k} + \theta_{k,2k} < 1.$$

Choose $\lambda_p = \sigma\sqrt{2\log p}$ in (1.5). Then with large probability, the Dantzig selector $\hat\beta$ obeys

$$\|\hat\beta-\beta\|_2 \le C_1\sigma\sqrt{k}\sqrt{2\log p}, \tag{5.6}$$

with $C_1 = \frac{4}{1-\delta_k-\theta_{k,2k}}$[1].

Another commonly used method in statistics is the Lasso which solves the ℓ_1 regularized least squares problem (1.4). This is equivalent to the ℓ_2-constrained ℓ_1 minimization problem (P_1). In the Gaussian error case, we shall consider a particular setting. Let $\hat\beta^{\ell_2}$ be the minimizer of

$$\min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \quad \text{subject to} \quad \|y-F\gamma\|_2 \le \epsilon_n \tag{5.7}$$

[1] It appears that the constant C_1 in Candes and Tao [5] should be $C_1 = 4/(1-\delta_{2k}-\theta_{k,2k})$.

where $\epsilon_n = \sigma\sqrt{n + 2\sqrt{n\log n}}$.

Combining our results from the last section together with Lemma 1, we have the following results on the Dantzig selector $\hat{\beta}^{DS}$ and the estimator $\hat{\beta}^{\ell_2}$ obtained from ℓ_1 minimization under the ℓ_2 constraint. Again, these results improve the previous results in the literature by weakening the conditions and providing more precise bounds.

Theorem 5.2 *Suppose $\beta \in \mathbb{R}^p$ is a k-sparse vector and the matrix F satisfies*

$$\delta_{1.5k} + \theta_{k,1.5k} < 1.$$

Then with probability $P \geq 1 - \frac{1}{2\sqrt{\pi \log p}}$, the Dantzig selector $\hat{\beta}^{DS}$ obeys

$$\|\hat{\beta}^{DS} - \beta\|_2 \leq C_1 \sigma \sqrt{k}\sqrt{2\log p}, \tag{5.8}$$

with $C_1 = \frac{2\sqrt{3}}{1-\delta_{1.5k}-\theta_{k,1.5k}}$, and with probability at least $1 - \frac{1}{n}$, $\hat{\beta}^{\ell_2}$ obeys

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \leq D_1 \sigma \sqrt{n + 2\sqrt{n\log n}} \tag{5.9}$$

with $D_1 = \frac{2\sqrt{2}(1+\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}}$.

Remark: Similar to the results obtained in the previous sections, if β is not necessarily k-sparse, in general we have, with probability $P \geq 1 - \frac{1}{2\sqrt{\pi \log p}}$,

$$\|\hat{\beta}^{DS} - \beta\|_2 \leq C_1 \sigma \sqrt{k}\sqrt{2\log p} + C_2 k^{-\frac{1}{2}} \|\beta_{-\max(k)}\|_1.$$

where $C_1 = \frac{2\sqrt{3}}{1-\delta_{1.5k}-\theta_{k,1.5k}}$ and $C_2 = \frac{2\sqrt{2}(1-\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}}$, and with probability $P \geq 1 - \frac{1}{n}$,

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \leq D_1 \sigma \sqrt{n + 2\sqrt{n\log n}} + D_2 k^{-\frac{1}{2}} \|\beta_{-\max(k)}\|_1$$

where $D_1 = \frac{2\sqrt{2}(1+\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}}$ and $D_2 = \frac{2\sqrt{2}\theta_{k,1.5k}(1-\delta_{1.5k})}{1-\delta_{1.5k}-\theta_{k,1.5k}}$.

6 Appendix

Proof of Proposition 2.3. Let

$$\begin{aligned} \Lambda &= \big((a_1 + \cdots + a_w) + 2(a_{w+1} + \cdots + a_{2w}) + (a_{2w+1} + \cdots + a_{3w})\big)^2 \\ &= \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4 + \Lambda_5 + \Lambda_6. \end{aligned}$$

Where each Λ_i is given (and bounded) by

$$
\begin{aligned}
\Lambda_1 &= \big(a_1 + a_2 + \cdots + a_w\big)^2 \\
&\geq a_1^2 + 3a_2^2 + \cdots + (2w-1)a_w^2 \\
\Lambda_2 &= 4\big(a_{w+1} + a_{w+2} + \cdots + a_{2w}\big)^2 \\
&\geq 4\big(a_{w+1}^2 + 3a_{w+2}^2 + \cdots + (2w-1)a_{2w}^2\big) \\
\Lambda_3 &= \big(a_{2w+1} + a_{2w+2} + \cdots + a_{3w}\big)^2 \\
&\geq a_{2w+1}^2 + 3a_{2w+2}^2 + \cdots + (2w-1)a_{3w}^2 \\
\Lambda_4 &= 4\big(a_1 + a_2 + \cdots + a_w\big)\big(a_{w+1} + a_{w+2} + \cdots + a_{2w}\big) \\
&\geq 4w\big(a_{w+1}^2 + a_{w+2}^2 + \cdots + a_{2w}^2\big) \\
\Lambda_5 &= 2\big(a_1 + a_2 + \cdots + a_w\big)\big(a_{2w+1} + a_{2w+2} + \cdots + a_{3w}\big) \\
&\geq 2w\big(a_{2w+1}^2 + a_{2w+2}^2 + \cdots + a_{3w}^2\big) \\
\Lambda_6 &= 4\big(a_{w+1} + a_{w+2} + \cdots + a_{2w}\big)\big(a_{2w+1} + a_{2w+2} + \cdots + a_{3w}\big) \\
&\geq 4w\big(a_{2w+1}^2 + a_{2w+2}^2 + \cdots + a_{3w}^2\big).
\end{aligned}
$$

Without loss of generality, we assume that w is even. Write

$$\Lambda_2 = \Lambda_{21} + \Lambda_{22},$$

where

$$\Lambda_{21} = 4\big(a_{w+1}^2 + 3a_{w+2}^2 + \cdots + (w-1)a_{w+\frac{w}{2}}^2 + wa_{w+\frac{w}{2}+1}^2 + wa_{w+\frac{w}{2}+2}^2 + \cdots + wa_{2w}^2\big),$$

and

$$
\begin{aligned}
\Lambda_{22} &= 4\big(a_{w+\frac{w}{2}+1}^2 + 3a_{w+\frac{w}{2}+2}^2 \cdots + (w-1)a_{2w}^2\big) \geq w^2 a_{2w}^2 \\
&= (2w-1)a_{2w}^2 + (2w-3)a_{2w}^2 + \cdots + 3a_{2w}^2 + \cdots + a_{2w}^2.
\end{aligned}
$$

Now

$$
\begin{aligned}
\Lambda_3 + \Lambda_5 + \Lambda_6 + \Lambda_{22} &\geq 6(w+1)a_{2w+1}^2 + (6w+3)a_{2w+2}^2 + \cdots + (8w-1)a_{3w}^2 \\
&\quad +(2w-1)a_{2w}^2 + (2w-3)a_{2w}^2 + \cdots + 3a_{2w}^2 + \cdots + a_{2w}^2 \\
&\geq 6(w+1)a_{2w+1}^2 + (6w+3)a_{2w+2}^2 + \cdots + (8w-1)a_{3w}^2 \\
&\quad +(2w-1)a_{2w+1}^2 + (2w-3)a_{2w+2}^2 + \cdots + 3a_{3w-1}^2 + a_{3w}^2 \\
&\geq 8w\big(a_{2w+1}^2 + a_{2w+3}^2 + \cdots + a_{3w-1}^2 + a_{3w}^2\big)
\end{aligned}
$$

and

$$
\begin{aligned}
\Lambda_1+\Lambda_{21}+\Lambda_4 \;\geq\;& a_1^2+3a_2^2+\cdots+(2w-1)a_w^2\\
&+4\big(a_{w+1}^2+3a_{w+2}^2+\cdots+(w-1)a_{w+\frac{w}{2}}\\
&+wa_{w+\frac{w}{2}+1}^2+wa_{w+\frac{w}{2}+2}^2+\cdots+wa_{2w}^2\big)\\
&+4w\big(a_{w+1}^2+a_{w+2}^2+\cdots+a_{2w}^2\big)\\
\geq\;& w^2a_w^2+4(w+1)a_{w+1}^2+4(w+3)a_{w+2}^2+\cdots 4(2w-1)a_{w+\frac{w}{2}}^2\\
&+8wa_{w+\frac{w}{2}+1}^2+8wa_{w+\frac{w}{2}+2}^2+\cdots+8wa_{2w}^2\\
\geq\;& \overbrace{4(w-1)a_w^2+4(w-3)a_w^2+\cdots+4a_w^2}^{\frac{w}{2}\text{ terms}}\\
&+4(w+1)a_{w+1}^2+4(w+3)a_{w+2}^2+\cdots 4(2w-1)a_{w+\frac{w}{2}}^2\\
&+8wa_{w+\frac{w}{2}+1}^2+8wa_{w+\frac{w}{2}+2}^2+\cdots+8wa_{2w}^2\\
\geq\;& 8w\big(a_{w+1}^2+a_{w+3}^2+\cdots+a_{2w-1}^2+a_{2w}^2\big).
\end{aligned}
$$

Therefore

$$\Lambda \geq 8w\big(a_{w+1}^2+a_{w+3}^2+\cdots++a_{2w}^2+a_{2w+1}^2+\cdots++a_{3w}^2\big),$$

and the inequality is proved. ∎

Proof of Lemma 1. The first inequality is standard. We now prove inequality (5.3). Note that $X=\|z\|_2^2/\sigma^2$ is a χ_n^2 random variable. It follows from Lemma 4 in Cai [1] that for any $\lambda>0$

$$P(X>(1+\lambda)n)\leq \frac{1}{\lambda\sqrt{\pi n}}\exp\{-\frac{n}{2}(\lambda-\log(1+\lambda))\}.$$

Hence,

$$P\left(\|z\|_2\leq\sigma\sqrt{n+2\sqrt{n\log n}}\right)=1-P(X>(1+\lambda)n)\geq 1-\frac{1}{\lambda\sqrt{\pi n}}\exp\{-\frac{n}{2}(\lambda-\log(1+\lambda))\}$$

where $\lambda=2\sqrt{n^{-1}\log n}$. It now follows from the fact $\log(1+\lambda)\leq\lambda-\frac{1}{2}\lambda^2+\frac{1}{3}\lambda^3$ that

$$P\left(\|z\|_2\leq\sigma\sqrt{n+2\sqrt{n\log n}}\right)\geq 1-\frac{1}{n}\cdot\frac{1}{2\sqrt{\pi\log n}}\exp\{\frac{4(\log n)^{3/2}}{3\sqrt{n}}\}.$$

Inequality (5.3) now follows by verifying directly that $\frac{1}{2\sqrt{\pi\log n}}\exp(\frac{4(\log n)^{3/2}}{3\sqrt{n}})\leq 1$ for all $n\geq 2$. ∎

References

[1] T. Cai, On block thresholding in wavelet regression: Adaptivity, block size and threshold level, *Statist. Sinica*, 12 (2002), 1241-1273.

[2] E. J. Candes, The restricted isometry property and its implications for compressed sensing, (2008), technical report.

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

[4] E. J. Candes and T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory*, 51(2005) 4203-4215.

[5] E. J. Candes and T. Tao, The Dantzig selector: statistical estimation when p is much larger than n (with discussion), *Ann. Statist.*, 35(2007), 2313-2351.

[6] D. L. 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.

[7] D. L. Donoho, For most large underdetermined systems of equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution, *Comm. Pure Appl. Math.*, 59(2006), 907-934.

[8] D.L. Donoho, M. Elad, and V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inf. Theory*, 52 (2006), 6-18.

[9] D. L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inf. Theory*, 47(2001), 2845-2862.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, Least angle regression (with discussion). *Ann. Statist.* 32(2004), 407-451.

[11] J.-J. Fuchs, On sparse representations in arbitrary redundant bases, *IEEE Trans. Inf. Theory*, 50(2004), 1341-1344.

[12] J.-J. Fuchs, Recovery of exact sparse representations in the presence of bounded noise, *IEEE Trans. Inf. Theory*, 51(2005), 3601-3608.

[13] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B*, 58(1996), 267-288.

[14] J. Tropp, Just relax: convex programming methods for identifying sparse signals in noise, *IEEE Trans. Inf. Theory*, 52(2006), 1030-1051.