On Recovery of Sparse Signals Via ℓ_1 Minimization

T. Tony Cai, Guangwu Xu, and Jun Zhang, *Senior Member, IEEE*

***Abstract*—This paper considers constrained ℓ_1 minimization methods in a unified framework for the recovery of high-dimensional sparse signals in three settings: noiseless, bounded error, and Gaussian noise. Both ℓ_1 minimization with an ℓ_∞ constraint (Dantzig selector) and ℓ_1 minimization under an ℓ_2 constraint are considered. 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. In particular, our results illustrate the relationship between ℓ_1 minimization with an ℓ_2 constraint and ℓ_1 minimization with an ℓ_∞ constraint. This paper also establishes connections between restricted isometry property and the mutual incoherence property. Some results of Candes, Romberg, and Tao (2006), Candes and Tao (2007), and Donoho, Elad, and Temlyakov (2006) are extended.**

***Index Terms*—Dantzig selectorℓ_1 minimization, restricted isometry property, sparse recovery, sparsity.**

I. 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 compressed sensing, constructive approximation, model selection in linear regression, and inverse problems.

Suppose we have n observations of the form

$$y = F\beta + z \tag{I.1}$$

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. See, for example, Fuchs [13], Candes and Tao [5], [6], Candes, Romberg, and Tao [4], Tropp [18], and Donoho, Elad, and Temlyakov [10].

A special case of particular interest is when no noise is present in (I.1) and $y = F\beta$. This is then 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. Although there are infinitely many general solutions, under regularity conditions there is a unique sparse solution. Indeed, in many cases the unique sparse solution can be found exactly through ℓ_1 minimization

$$\text{(Exact)} \quad \min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \quad \text{subject to } F\gamma = y. \tag{I.2}$$

This ℓ_1 minimization problem has been studied, for example, in Fuchs [13], Candes and Tao [5], and Donoho [8]. Understanding the noiseless case is not only of significant interest in 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 [5], [6] and Donoho [8], [9].

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

$$(\ell_2\text{-Constraint}) \quad \min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \text{ subject to } \|y - F\gamma\|_2 \le \eta. \tag{I.3}$$

Writing in terms of the Lagrangian function of (ℓ_2-Constraint), this is closely related to finding the solution to the ℓ_1 regularized least squares

$$\min_{\gamma\in\mathbb{R}^p} \left\{ \|y - F\gamma\|_2^2 + \rho\|\gamma\|_1 \right\}. \tag{I.4}$$

The latter is often called the Lasso in the statistics literature (Tibshirani [16]). Tropp [18] gave a detailed treatment of the ℓ_1 regularized least squares problem.

Another method, called the Dantzig selector, was recently proposed by Candes and Tao [6]. 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

$$\text{(DS)} \quad \min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \text{ subject to } \|F^T(y - F\gamma)\|_\infty \le \lambda. \tag{I.5}$$

Candes and Tao [6] showed that the Dantzig selector can be computed by solving a linear program

$$\min \sum_i u_i \quad \text{subject to } -u \le \gamma \le u$$
$$\text{and } -\lambda\sigma\mathbf{1} \le F^T(y - F\gamma) \le \lambda\sigma\mathbf{1}$$

where the optimization variables are $u,\ \gamma \in \mathbb{R}^p$. Candes and Tao [6] also showed that the Dantzig selector mimics the performance of an oracle procedure up to a logarithmic factor $\log p$.

Manuscript received May 01, 2008; revised November 13, 2008. Current version published June 24, 2009. The work of T. Cai was supported in part by the National Science Foundation (NSF) under Grant DMS-0604954, the work of G. Xu was supported in part by the National 973 Project of China (No. 2007CB807903).

T. T. Cai is with the Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: tcai@wharton.upenn.edu).

G. Xu and J. Zhang are with the Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI 53211 USA (e-mail: gxu4uwm@uwm.edu; junzhang@uwm. edu).

Communicated by J. Romberg, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2009.2021377

0018-9448/$25.00 © 2009 IEEE

It is clear that some 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 [5], [6], Candes and Tao considered sparse recovery problems in the RIP framework. They provided beautiful solutions to the problem under some conditions on the so-called restricted isometry constant $\delta_k(F)$ and restricted orthogonality constant $\theta_{k,k'}(F)$ (defined in Section II). These conditions essentially require that every set of columns of F with certain cardinality approximately behaves like an orthonormal system. Several different conditions have been imposed in various settings. For example, the condition $\delta_k(F) + \theta_{k,k}(F) + \theta_{k,2k}(F) < 1$ was used in Candes and Tao [5], $\delta_{3k}(F) + 3\delta_{4k}(F) < 2$ in Candes, Romberg, and Tao [4], $\delta_{2k}(F) + \theta_{k,2k}(F) < 1$ in Candes and Tao [6], and $\delta_{2k}(F) < \sqrt{2} - 1$ in Candes [3], where k is the sparsity index. A natural question is: Can these conditions be weakened in a unified way? Another widely used condition for sparse recovery is the Mutual Incoherence Property (MIP) which requires the pairwise correlations among the column vectors of F to be small. See [10], [11], [13], [14], [18].

In this paper, we consider ℓ_1 minimization methods in a single unified framework for sparse recovery in three cases, noiseless, bounded error, and Gaussian noise. Both ℓ_1 minimization with an ℓ_∞ constraint (DS) and ℓ_1 minimization under the ℓ_2 constraint (ℓ_2-Constraint) are considered. Our results improve on the existing results in [3]–[6] by weakening the conditions and tightening the error bounds. In particular, our results clearly illustrate the relationship between ℓ_1 minimization with an ℓ_2 constraint and ℓ_1 minimization with an ℓ_∞ constraint (the Dantzig selector). In addition, we also 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 [10].

In all cases, we solve the problems under the weaker condition

$$\delta_{1.5k}(F) + \theta_{k,1.5k}(F) < 1. \tag{I.6}$$

The improvement on the condition shows that for fixed n and p, 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.

It is sometimes convenient to impose conditions that involve only the restricted isometry constant δ. Efforts have been made in this direction in the literature. In [7], the recovery result was established under the condition $\delta_{2k}(F) < \frac{1}{3}$. In [3], the weaker condition $\delta_{2k}(F) < \sqrt{2} - 1$ was used. Similar conditions have also been used in the construction of (random) compressed sensing matrices. For example, conditions $\delta_{4k}(F) < \frac{1}{2}$ and $\delta_{3k}(F) < \frac{1}{3}$ were used in [15] and [1], respectively. We shall remark that, our results implies that the weaker condition

$$\delta_{1.75k}(F) < \sqrt{2} - 1$$

suffices in sparse signal reconstruction.

The paper is organized as follows. In Section II, after basic notation and definitions are reviewed, we introduce an elementary inequality, which allow us to make finer analysis of the sparse recovery problem. We begin the analysis of ℓ_1 minimization methods for sparse recovery by considering the exact recovery in the noiseless case in Section III. Our result improves the main result in Candes and Tao [5] 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 IV. The connections between the RIP and MIP are also explored. Sparse recovery with Gaussian noise is treated in Section V. Appendices A–D contain the proofs of technical results.

II. Preliminaries

In this section, we first introduce basic notation and definitions, and then present a technical inequality which will be used in proving our main results.

Let $p \in \mathbb{N}$. Let $v = (v_1, v_2, \dots, v_p) \in \mathbb{R}^p$ be a vector. The support of v is the subset of $\{1, 2, \dots, p\}$ 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 $\delta_k(F)$ is defined to be the smallest constant such that

$$\sqrt{1 - \delta_k(F)}\|c\|_2 \leq \|Fc\|_2 \leq \sqrt{1 + \delta_k(F)}\|c\|_2 \tag{II.1}$$

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

$$|\langle Fc, Fc' \rangle| \leq \theta_{k,k'}(F)\|c\|_2\|c'\|_2 \tag{II.2}$$

for all c and c' such that c and c' are k-sparse and k'-sparse, respectively, and have disjoint supports. Roughly speaking, the isometry constant $\delta_k(F)$ and restricted orthogonality constant $\theta_{k,k'}(F)$ measure how close subsets of cardinality k of columns of F are to an orthonormal system.

For notational simplicity we shall write δ_k for $\delta_k(F)$ and $\theta_{k,k'}$ for $\theta_{k,k'}(F)$ hereafter. It is easy to see that δ_k and $\theta_{k,k'}$ are monotone. That is

$$\delta_k \leq \delta_{k_1}, \quad \text{if } k \leq k_1 \leq p \tag{II.3}$$

$$\theta_{k,k'} \leq \theta_{k_1,k_1'}, \quad \text{if } k \leq k_1,\ k' \leq k_1',\ k_1 + k_1' \leq p. \tag{II.4}$$

Candes and Tao [5] showed that the constants δ_k and $\theta_{k,k'}$ are related by the following inequalities

$$\theta_{k,k'} \leq \delta_{k+k'} \leq \theta_{k,k'} + \max(\delta_k, \delta_{k'}). \tag{II.5}$$

As mentioned in the Introduction, different conditions on δ and θ have been used in the literature. It is not always immediately transparent which condition is stronger and which is weaker. We shall present another important property on θ and δ

which can be used to compare the conditions. In addition, it is especially useful in producing simplified recovery conditions.

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

$$\theta_{k,\sum_{i=1}^{l}k_i} \le \sqrt{\sum_{i=1}^{l}\theta_{k,k_i}^2}. \tag{II.6}$$

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

A proof of the proposition is provided in Appendix A.

Remark: Candes and Tao [6] imposes $\delta_{2k}+\theta_{k,2k}<1$ and in a more recent paper Candes [3] 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 $\theta_{k,2k}\le\sqrt{\delta_{2k}^2+\delta_{2k}^2}=\sqrt{2}\delta_{2k}$ which means that $\delta_{2k}<\sqrt{2}-1$ implies $\delta_{2k}+\theta_{k,2k}<1$.

We now introduce a useful elementary inequality. This inequality allows us to perform finer estimation on ℓ_1, ℓ_2 norms. It will be used in proving our main results.

Proposition 2.2: Let w be a positive integer. Then any descending chain of real numbers

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

satisfies

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

The proof of Proposition 2.2 is given in Appendix B.

III. Signal Recovery in the Noiseless Case

As mentioned in the Introduction, we shall give a unified treatment for the methods of ℓ_1 minimization with an ℓ_2 constraint and ℓ_1 minimization with an ℓ_∞ constraint for recovery of sparse signals in three cases: noiseless, bounded error, and Gaussian noise. We begin in this section by considering the simplest setting: 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 [13], Donoho [8], and Candes and Tao [5]. 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 [5]. 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 [5] 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 [5]): Let $F\in\mathbb{R}^{n\times p}$. Suppose $k\ge 1$ satisfies

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

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

As mentioned in the Introduction, other conditions on δ and θ have also been used in the literature. Candes, Romberg, and Tao [4] uses the condition $\delta_{3k}+3\delta_{4k}\le 2$. Candes and Tao [6] 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 $\delta_{2k}+\theta_{k,2k}<1$. Candes [3] imposes the condition $\delta_{2k}<\sqrt{2}-1$.

We shall show below that these conditions can be uniformly improved by a transparent argument. A direct application of Proposition 2.2 yields the following result which weakens the above conditions to

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

Note that it follows from (II.3) and (II.4) that $\delta_{1.5k}\le\delta_{2k}$, and $\theta_{k,1.5k}\le\theta_{k,2k}$. So the condition $\delta_{1.5k}+\theta_{k,1.5k}<1$ is weaker than $\delta_{2k}+\theta_{k,2k}<1$. It is also easy to see from (II.5) and (II.6) that the condition $\delta_{1.5k}+\theta_{k,1.5k}<1$ is also weaker than $\delta_k+\theta_{k,k}+\theta_{k,2k}<1$ and the other conditions mentioned above.

Theorem 3.2: Let $F\in\mathbb{R}^{n\times p}$. Suppose $k\ge1$ satisfies

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

and $y=F\beta$. Then the minimizer $\hat\beta$ to the problem $(Exact)$ obeys

$$\|\hat\beta-\beta\|_2\le C_0k^{-\frac{1}{2}}\|\beta_{-\max(k)}\|_1$$

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.

This theorem improves the results in [5], [6]. The improvement on the condition shows that for fixed n and p, signals with larger support can be recovered accurately.

Remark: It is sometimes more convenient to use conditions only involving the restricted isometry constant δ. Note that the condition

$$\delta_{1.75k}<\sqrt{2}-1 \tag{III.2}$$

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. Hence, Theorem 3.2 holds under the condition (III.2). The condition $\delta_{1.5k}+\delta_{2.5k}<1$ can also be used.

Proof of Theorem 3.2: The proof relies on Proposition 2.2 and makes use of the ideas from [4]–[6]. In this proof, we shall also identify a vector $v=(v_1,v_2,\ldots,v_p)\in\mathbb{R}^p$ as a function $v:\{1,2,\ldots,p\}\to\mathbb{R}$ by assigning $v(i)=v_i$.

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

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

such that $|h(n_{k+1})| \geq |h(n_{k+2})| \geq |h(n_{k+3})| \geq \cdots$. Let

$$T_1 = \{n_{k+1}, n_{k+2}, \ldots, n_{2k}\}, \quad T_2 = \{n_{2k+1}, \ldots, n_{3k}\}, \ldots.$$

For a subset $E \subset \{1, 2, \ldots, m\}$, 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, $\text{supp}(h_i) \subset T_i$, and $|T_i| = k$ for $i \geq 0$. We first consider the case where k is divisible by 4.

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

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

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

$$T_{i1} = \{n_{ik+1}, n_{ik+2}, \ldots, n_{ik+\frac{k}{2}}\}$$

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

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

$$\begin{aligned} T_{11} &= \{n_{k+1}, n_{k+2}, \ldots, n_{k+\frac{k}{4}}\} \\ T_{12} &= \{n_{k+\frac{k}{4}+1}, \ldots, n_{k+\frac{k}{2}}\} \\ T_{13} &= \{n_{k+\frac{k}{2}+1}, \ldots, n_{k+\frac{3k}{4}}\} \quad \text{and} \\ T_{14} &= \{n_{k+\frac{3k}{4}+1}, \ldots, n_{2k}\}. \end{aligned}$$

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

Note that

$$\sum_{i \geq 1} \|h_i\|_1 \leq \|h_0\|_1 + 2\|\beta_{-\max(k)}\|_1. \quad \text{(III.3)}$$

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

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

Since $\|\beta\|_1 = \|\beta_{\max(k)}\|_1 + \|\beta_{-\max(k)}\|_1$, this yields

$$\sum_{i \geq 1} \|h_i\|_1 \leq \|h_0\|_1 + 2\|\beta_{-\max(k)}\|_1.$$

The following claim follows from our Proposition 2.2.

Claim:

$$\begin{aligned} \|h_{13} + h_{14}\|_2 + \sum_{i \geq 2} \|h_i\|_2 &\leq \frac{\sum_{i \geq 1} \|h_i\|_1}{\sqrt{k}} \\ &\leq \|h_0\|_2 + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}. \quad \text{(III.4)} \end{aligned}$$

In fact, from Proposition 2.2 and the fact that $\|h_{11}\|_1 \geq \|h_{12}\|_1 \geq \|h_{13}\|_1 \geq \|h_{14}\|_1$, we have

$$\begin{aligned} &\|h_{12}\|_1 + 2\|h_{13}\|_1 + \|h_{14}\|_1 \\ &\quad \leq \frac{2}{3} \left(2\|h_{11}\|_1 + 2\|h_{12}\|_1 + \|h_{13}\|_1 + \|h_{14}\|_1\right). \end{aligned}$$

It then follows from Proposition 2.2 that

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

Proposition 2.2 also yields

$$\|h_2\|_2 \leq \frac{\|h_{13} + h_{14}\|_1 + 2\|h_{21}\|_1 + \|h_{22}\|_1}{2\sqrt{k}}$$
and
$$\|h_i\|_2 \leq \frac{\|h_{(i-1)2}\|_1 + 2\|h_{i1}\|_1 + \|h_{i2}\|_1}{2\sqrt{k}}$$

for any $i > 2$. Therefore

$$\begin{aligned} &\|h_{13} + h_{14}\|_2 + \sum_{i \geq 2} \|h_i\|_2 \\ &\leq \frac{2\|h_{11}\|_1 + 2\|h_{12}\|_1 + \|h_{13}\|_1 + \|h_{14}\|_1}{2\sqrt{k}} \\ &\quad + \frac{\|h_{13} + h_{14}\|_1 + 2\|h_{21}\|_1 + \|h_{22}\|_1}{2\sqrt{k}} \\ &\quad + \frac{\|h_{22}\|_1 + 2\|h_{31}\|_1 + \|h_{32}\|_1}{2\sqrt{k}} + \cdots \\ &\leq \frac{2\|h_1\|_1 + 2\|h_2\|_1 + 2\|h_3\|_1 + \cdots}{2\sqrt{k}} \\ &= \frac{\sum_{i \geq 1} \|h_i\|_1}{\sqrt{k}} \overset{\text{by (III.3)}}{\leq} \frac{\|h_0\|_1 + 2\|\beta_{-\max(k)}\|_1}{\sqrt{k}} \\ &\leq \|h_0\|_2 + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}. \end{aligned}$$

In the rest of our proof we write $h_{11} + h_{12} = h_1'$. Note that $Fh = F\hat{\beta} - F\beta = 0$. So we get the equation at the top of the following page. This yields

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

$$
\begin{aligned}
0 &= |\langle Fh, F(h_0 + h_1')\rangle| \\
&= \left| \langle F(h_0 + h_1'), F(h_0 + h_1')\rangle + \langle F(h_{13} + h_{14}), F(h_0 + h_1')\rangle + \sum_{i\ge 2} \langle Fh_i, F(h_0 + h_1')\rangle \right| \\
&\overset{\text{(II.1,II.2)}}{\ge} (1-\delta_{1.5k})\|h_0 + h_1'\|_2^2 - \theta_{0.5k,1.5k}\|h_{13}+h_{14}\|_2\|h_0+h_1'\|_2 - \sum_{i\ge 2}\theta_{k,1.5k}\|h_i\|_2\|h_0+h_1'\|_2 \\
&\ge \|h_0+h_1'\|_2\left((1-\delta_{1.5k})\|h_0+h_1'\|_2 - \theta_{k,1.5k}\left(\|h_{13}+h_{14}\|_2 + \sum_{i\ge 2}\|h_i\|_2\right)\right) \\
&\overset{\text{(III.4)}}{\ge} \|h_0+h_1'\|_2\left((1-\delta_{1.5k})\|h_0+h_1'\|_2 - \theta_{k,1.5k}\|h_0\|_2 - \theta_{k,1.5k}\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\right).
\end{aligned}
$$

It then follows from (III.4) that

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

We now turn to the case that k is not divisible by 4. Let $l = \lfloor \frac{k}{2} \rfloor$. Note that in this case $\delta_{1.5k}$ and $\theta_{k,1.5k}$ are understood as δ_{k+l} and $\theta_{k,k+l}$, respectively. So the proof for the previous case works if we set

$$
\begin{aligned}
T_{i1} &= \{ik+1, ik+2, \ldots, ik+l\} \text{ and} \\
T_{i2} &= \{ik+l+1, ik+l+2, \ldots, (i+1)k\}, \quad \text{for } i > 1
\end{aligned}
$$

and

$$
\begin{aligned}
T_{11} &= \left\{k+1, \ldots, k+\left\lfloor \frac{l}{2} \right\rfloor\right\} \\
T_{12} &= \left\{k+\left\lfloor \frac{l}{2} \right\rfloor + 1, \ldots, k+l\right\} \\
T_{13} &= \left\{k+l+1, \ldots, k+l+\left\lfloor \frac{l}{2} \right\rfloor\right\} \text{ and} \\
T_{14} &= \left\{k+l+\left\lfloor \frac{l}{2} \right\rfloor + 1, \ldots, 2k\right\}.
\end{aligned}
$$

In this case, we need to use the following inequality whose proof is essentially the same as Proposition 2.2: For any descending chain of real numbers $a_1 \ge \cdots \ge a_{w+1} \ge b_1 \ge \cdots \ge b_w \ge c_1 \ge \cdots \ge c_{w+1} \ge 0$, we have

$$
\sqrt{b_1^2 + \cdots + b_w^2 + c_1^2 + \cdots + c_{w+1}^2} \le \frac{a_1 + \cdots + a_{w+1}}{2\sqrt{2w+1}} + \frac{2(b_1 + \cdots + b_w) + c_1 + \cdots + c_{w+1}}{2\sqrt{2w+1}}. \quad \square
$$

IV. 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 V.

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

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

We shall specifically consider two cases: $\mathcal{B} = \{z : \|F^T z\|_\infty \le \lambda\}$ and $\mathcal{B} = \{z : \|z\|_2 \le \eta\}$. The first case is closely connected to the Dantzig selector in the Gaussian noise setting which will be discussed in more detail in Section V. Our results improve the results in Candes, Romberg, and Tao [4], Candes and Tao [6], and Donoho, Elad, and Temlyakov [10].

We shall first consider

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

Let $\hat{\beta}$ be the solution to the (DS) problem given in (I.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 $\|F^T z\|_\infty \le \lambda$. If

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

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

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

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 $\|\hat{\beta} - \beta\|_2 \le C_1 k^{\frac{1}{2}}\lambda$.

Remark: Theorem 4.1 is comparable to Theorem 1.1 of Candes and Tao [6], but the result here is a deterministic one instead of a probabilistic one as bounded errors are considered. The proof in Candes and Tao [6] can be adapted to yield a similar result for bounded errors under the stronger condition $\delta_{2k} + \theta_{k,2k} < 1$.

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

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

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

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

To finish the proof, we observe the following.

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

$$\begin{aligned}
&|\langle Fh, F(h_0 + h_1')\rangle| \\
&\quad = |\langle (F\hat{\beta} - y) + z, F_{T_0 \cup T_{10} \cup T_{11}}(h_0 + h_1')\rangle| \\
&\quad = |\langle F_{T_0 \cup T_{10} \cup T_{11}}^T \left((F\hat{\beta} - y) + z \right), h_0 + h_1' \rangle| \\
&\quad \leq \|F_{T_0 \cup T_{10} \cup T_{11}}^T \left((F\hat{\beta} - y) + z \right)\|_2 \|h_0 + h_1'\|_2 \\
&\quad \leq \sqrt{1.5k} 2\lambda \|h_0 + h_1'\|_2.
\end{aligned}$$

2) $\|\hat{\beta} - \beta\|_2 \leq \sqrt{2} \left(\|h_0 + h_1'\|_2 + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}} \right)$
   In fact

$$\begin{aligned}
&\|\hat{\beta} - \beta\|_2^2 \\
&= \|h\|_2^2 = \|h_0 + h_1'\|_2^2 + \|h_{13} + h_{14}\|_2^2 + \sum_{i \geq 2} \|h_i\|_2^2 \\
&\leq \|h_0 + h_1'\|_2^2 + \left(\|h_{13} + h_{14}\|_2 + \sum_{i \geq 2} \|h_i\|_2 \right)^2 \\
&\overset{\text{(III.4)}}{\leq} \|h_0 + h_1'\|_2^2 + \left(\|h_0\|_2 + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}} \right)^2 \\
&\leq 2 \left(\|h_0 + h_1'\|_2 + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}} \right)^2.
\end{aligned}$$

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 $\|z\|_2 \leq \epsilon$. Once again, this problem can be solved through constrained ℓ_1 minimization

$$\min \|\gamma\|_1 \text{ subject to } \|y - F\gamma\|_2 \leq \eta. \tag{IV.3}$$

An alternative to the constrained ℓ_1 minimization approach is the so-called Lasso given in (I.4). The Lasso recovers a sparse signal via ℓ_1 regularized least squares. It is closely connected to the ℓ_2-constrained ℓ_1 minimization. The Lasso is a popular method in statistics literature (Tibshirani [16]). See Tropp [18] for a detailed treatment of the ℓ_1 regularized least squares problem.

By using a similar argument, we have the following result on the solution of the minimization (IV.3).

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 $\|z\|_2 \leq \epsilon$. If

$$\delta_{1.5k} + \theta_{k,1.5k} < 1 \tag{IV.4}$$

then for any $\eta \geq \epsilon$, the minimizer $\hat{\beta}$ to the problem (IV.3) obeys

$$\|\hat{\beta} - \beta\|_2 \leq C(\eta + \epsilon) \tag{IV.5}$$

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 $\eta \geq \epsilon$ implies that $\|\hat{\beta}\|_1 \leq \|\beta\|_1$, 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

$$\|h_0\|_1 \geq \sum_{i \geq 1} \|h_i\|_1$$

and

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

Note that

$$\|Fh\|_2 = \|F(\beta - \hat{\beta})\|_2 \leq \|F\beta - y\|_2 + \|F\hat{\beta} - y\|_2 \leq \eta + \epsilon.$$

So

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

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

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

(The η was set to be ϵ in [4].) 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 [4] by enlarging the support of β by 60%.

Remark: 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

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

A. Connections Between RIP and MIP

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

$$M = \max_{1\le i,j\le p, i\ne j} |\langle f_i, f_j\rangle| \tag{IV.6}$$

be small, where $f_1, f_2, \ldots, f_p$ 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 [10], [11], [13], [14], [18]. For example, a recent paper, Donoho, Elad, and Temlyakov [10] 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 (ℓ_2-Constraint) 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 [10] can be improved under the RIP framework, by using Theorem 4.2.

The following is a simple result that gives RIP constants from MIP. The proof can be found in Appendix C. It is remarked that the first inequality in the next proposition can be found in [17].

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

$$\delta_k \le (k-1)M \quad \text{and} \quad \theta_{k,k'} \le \sqrt{kk'}M. \tag{IV.7}$$

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 (ℓ_2-Constraint) obeys

$$\|\hat{\beta} - \beta\|_2 \le C(\eta + \epsilon) \tag{IV.8}$$

with $C = \frac{\sqrt{8+12t-8M}}{2+2M-(3+\sqrt{6})t}$.

Proof of Theorem 4.3: It follows from Proposition 4.1 that

$$\begin{aligned}\delta_{1.5k} + \theta_{k,1.5k} &\le (1.5k + \sqrt{1.5}k - 1)M \\ &= (1.5 + \sqrt{1.5})t - M.\end{aligned}$$

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

$$\begin{aligned}\|\hat{\beta} - \beta\|_2 &\le \frac{\sqrt{2(1+\delta_{1.5k})}}{1-\delta_{1.5k}-\theta_{k,1.5k}}(\eta+\epsilon) \\ &\le \frac{\sqrt{2(1+(1.5k-1)M)}}{1+M-(1.5+\sqrt{1.5})t}(\eta+\epsilon) \\ &= \frac{\sqrt{8+12t-8M}}{2+2M-(3+\sqrt{6})t}(\eta+\epsilon).\end{aligned}$$

□

Remarks: In this theorem, the result of Donoho, Elad, and Temlyakov [10] 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 [10] 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 $C = \frac{\sqrt{8+12t-8M}}{2+2M-(3+\sqrt{6})t}$, as t is close to $1/4$.

V. Recovery of Sparse Signals in Gaussian Noise

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

$$y = F\beta + z, \quad z \sim N\left(0, \sigma^2 I_n\right) \tag{V.1}$$

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 interest, in particular in statistics. Many methods, including the Lasso (Tibshirani [16]), LARS (Efron, Hastie, Johnstone, and Tibshirani [12]) and Dantzig selector (Candes and Tao [6]), have been introduced and studied.

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

Lemma 5.1: The Gaussian error $z \sim N(0, \sigma^2 I_n)$ satisfies

$$P\left(\|F^T z\|_\infty \le \sigma\sqrt{2\log p}\right) \ge 1 - \frac{1}{2\sqrt{\pi \log p}} \tag{V.2}$$

and

$$P\left(\|z\|_2 \le \sigma\sqrt{n + 2\sqrt{n\log n}}\right) \ge 1 - \frac{1}{n}. \tag{V.3}$$

Inequality (V.2) follows from standard probability calculations and inequality (V.3) is proved in Appendix D.

Lemma 5.1 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 [6] introduced the Dantzig selector for sparse recovery in the Gaussian noise setting. Given the observations in (V.1), the Dantzig selector $\hat{\beta}^{\mathrm{DS}}$ is the minimizer of

$$(DS) \quad \min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \text{ subject to } \|F^T(y - F\gamma)\|_\infty \le \lambda_p \tag{V.4}$$

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

$$F^T y = F^T F\beta. \tag{V.5}$$

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 (V.3). And similar to the noiseless case, ℓ_1 minimization leads to the "sparsest" solution over the space of all feasible solutions.

Candes and Tao [6] showed the following result.

Theorem 5.1 (Candes and Tao [6]): Suppose $\beta \in \mathbb{R}^p$ is a k-sparse vector. Let F be such that

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

Choose $\lambda_p = \sigma\sqrt{2\log p}$ in (I.1). Then with large probability, the Dantzig selector $\hat\beta^{\rm DS}$ obeys

$$\|\hat\beta^{\rm DS} - \beta\|_2 \le C\sigma\sqrt{k}\sqrt{2\log p} \tag{V.6}$$

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

As mentioned earlier, the Lasso is another commonly used method in statistics. The Lasso solves the ℓ_1 regularized least squares problem (I.4) and is closely related to the ℓ_2-constrained ℓ_1 minimization problem (ℓ_2-Constraint). In the Gaussian error case, we shall consider a particular setting. Let $\hat\beta^{\ell_2}$ be the minimizer of

$$\min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \text{ subject to } \|y - F\gamma\|_2 \le \epsilon_n \tag{V.7}$$

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

Combining our results from the last section together with Lemma 5.1, we have the following results on the Dantzig selector $\hat\beta^{\rm DS}$ 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 $P \ge 1 - \frac{1}{2\sqrt{\pi\log p}}$, the Dantzig selector $\hat\beta^{\rm DS}$ obeys

$$\|\hat\beta^{\rm DS} - \beta\|_2 \le C_1\sigma\sqrt{k}\sqrt{2\log p} \tag{V.8}$$

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

$$\|\hat\beta^{\ell_2} - \beta\|_2 \le D_1\sigma\sqrt{n + 2\sqrt{n\log n}} \tag{V.9}$$

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

Remark: In comparison to Theorem 5.1, our result in Theorem 5.2 weakens the condition from $\delta_{2k} + \theta_{k,2k} < 1$ to $\delta_{1.5k} + \theta_{k,1.5k} < 1$ and improves the constant in the bound from $C = 4/(1-\delta_{2k}-\theta_{k,2k})$ to $C_1 = 2\sqrt{3}/(1-\delta_{1.5k}-\theta_{k,1.5k})$. Note that $C_1 < C$ since $\delta_{1.5k} + \theta_{k,1.5k} \le \delta_{2k} + \theta_{k,2k}$. The improvement on the error bound is minor. The improvement on the condition is more significant as it shows signals with larger support can be recovered accurately for fixed n and p.

[1]In Candes and Tao [6], the constant C was stated as $C = \frac{4}{1-\delta_k-\theta_{k,2k}}$. It appears that there was a typo and the constant C should be $C = 4/(1-\delta_{2k}-\theta_{k,2k})$.

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

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

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 $P \ge 1 - \frac{1}{n}$

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

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

Remark: Candes and Tao [6] also proved an Oracle Inequality for the Dantzig selector $\hat\beta^{\rm DS}$ in the Gaussian noise setting under the condition $\delta_{2k} + \theta_{k,2k} < 1$. With some additional work, our approach can be used to improve [6, Theorems 1.2 and 1.3] by weakening the condition to $\delta_{1.5k} + \theta_{k,1.5k} < 1$.

APPENDIX A
PROOF OF PROPOSITION 2.1

Let c be k-sparse and c' be $\left(\sum_{i=1}^l k_i\right)$-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 $i = 1,\ldots,j$ and $\mathrm{supp}(c'_i) \cap \mathrm{supp}(c'_j) = \emptyset$ for $i \ne j$. Using the Cauchy–Schwartz inequality, we have

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

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

APPENDIX B
PROOF OF PROPOSITION 2.2

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

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

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

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

where

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

and

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

Now

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

and

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

Therefore

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

and the inequality is proved. □

Appendix C
Proof of Proposition 4.1

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

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

Now let us bound the second term. Note that

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

These give us

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

and hence

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

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

$$\theta_{k,k'} \leq \sqrt{k'}\theta_{k,1} \leq \sqrt{kk'}\theta_{1,1} = \sqrt{kk'}M.$$ □

Appendix D
Proof of Lemma 5.1

The first inequality is standard. For completeness, we give a short proof here. Let $w_i = \langle f_i, z \rangle$. Then $w_1, \ldots, w_p$ each marginally has Gaussian distribution $N(0, \sigma^2)$. Hence

$$\begin{aligned}
&P\left(\|F^T z\|_\infty \leq \sigma\sqrt{2\log p}\right) \\
&\quad= 1 - P\left(\bigcup_{i=1}^{p} \left\{w_i > \sigma\sqrt{2\log p}\right\}\right)
\end{aligned}$$

$$
\begin{aligned}
&\geq 1-\sum_{i=1}^{p} P\left(w_i>\sigma\sqrt{2\log p}\right)\\
&=1-p\cdot P\left(w_1>\sigma\sqrt{2\log p}\right)\\
&\geq 1-\frac{1}{2\sqrt{\pi\log p}}
\end{aligned}
$$

where the last step follows from the Gaussian tail probability bound that for a standard Gaussian variable V and any constant $\lambda>0$, $P(V>\lambda)\leq\lambda^{-1}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\lambda^2}$.

We now prove inequality (V.3). Note that $X=\|z\|_2^2/\sigma^2$ is a χ_n^2 random variable. It follows from Lemma 4 in Cai [2] that for any $\lambda>0$

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

Hence

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

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

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

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

Acknowledgment

The authors wish to thank the referees for thorough and useful comments which have helped to improve the presentation of the paper.

References

[1] W. Bajwa, J. Haupt, J. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. IEEE SSP Workshop*, Madison, WI, Aug. 2007, pp. 294–298.

[2] T. Cai, "On block thresholding in wavelet regression: Adaptivity, block size and threshold level," *Statist. Sinica*, vol. 12, pp. 1241–1273, 2002.

[3] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l' Academie des Sciences Paris*, ser. I, vol. 346, pp. 589–592, 2008.

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

[5] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[6] E. J. Candes 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.

[7] A. Cohen, W. Dahmen, and R. Devore, "Compressed Sensing and Best k-Term Approximation" 2006, preprint.

[8] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, pp. 797–829, 2006.

[9] D. L. Donoho, "For most large underdetermined systems of equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution," *Commun. Pure Appl. Math.*, vol. 59, pp. 907–934, 2006.

[10] 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, no. 1, pp. 6–18, Jan. 2006.

[11] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[12] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression (with discussion," *Ann. Statist.*, vol. 32, pp. 407–451, 2004.

[13] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, Jun. 2004.

[14] J.-J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601–3608, Oct. 2005.

[15] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proc. 40th Annu. Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 2006, pp. 207–212.

[16] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, pp. 267–288, 1996.

[17] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[18] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

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

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

Prof. Cai is the recipient of the 2008 COPSS Presidents' Award and a fellow of the Institute of Mathematical Statistics.

Guangwu Xu received the Ph.D. degree in mathematics from the State University of New York (SUNY), Buffalo.

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

Jun Zhang (S'85–M'88–SM'01) received the B.S. degree in electrical and computer engineering from Harbin Shipbuilding Engineering Institute, Harbin, China, in 1982 and was admitted to the graduate program of the Radio Electronic Department of Tsinghua University. After a brief stay at Tsinghua, he came to the U.S. for graduate study on a scholarship from the Li Foundation, Glen Cover, New York. He received the M.S. and Ph.D. degrees, both in electrical engineering, from Rensselaer Polytechnic Institute, Troy, NY, in 1985 and 1988, respectively.

He joined the faculty of the Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, and currently is a Professor. His research interests include image processing and computer vision, signal processing and digital communications.

Prof. Zhang has been an Associate Editor of IEEE Transactions on Image Processing.