3-2010

Shifting Inequality and Recovery of Sparse Signals

T. Tony Cai
University of Pennsylvania

Lie Wang

Guangwu Xu

Recommended Citation

Cai, T., Wang, L., & Xu, G. (2010). Shifting Inequality and Recovery of Sparse Signals. *IEEE Transactions on Signal Processing, 58* (3), 1300-1308. http://dx.doi.org/10.1109/TSP.2009.2034936

Abstract
In this paper, we present a concise and coherent analysis of the constrained ??$_1$ minimization method for stable recovering of high-dimensional sparse signals both in the noiseless case and noisy case. The analysis is surprisingly simple and elementary, while leads to strong results. In particular, it is shown that the sparse recovery problem can be solved via ??$_1$ minimization under weaker conditions than what is known in the literature. A key technical tool is an elementary inequality, called Shifting Inequality, which, for a given nonnegative decreasing sequence, bounds the ??$_2$ norm of a subsequence in terms of the ??$_1$ norm of another subsequence by shifting the elements to the upper end.

Keywords
minimisation, signal reconstruction, constrained minimization method, high-dimensional sparse signals, nonnegative decreasing sequence, shifting inequality, signal processing, sparse signals recovery, restricted isometry property, shifting inequality, sparse recovery

Disciplines
Computer Sciences | Statistics and Probability

Shifting Inequality and Recovery of Sparse Signals

T. Tony Cai* Lie Wang† and Guangwu Xu‡

Abstract

In this paper we present a concise and coherent analysis of the constrained ℓ_1 minimization method for stable recovering of high-dimensional sparse signals both in the noiseless case and noisy case. The analysis is surprisingly simple and elementary, while leads to strong results. In particular, it is shown that the sparse recovery problem can be solved via ℓ_1 minimization under weaker conditions than what is known in the literature. A key technical tool is an elementary inequality, called Shifting Inequality, which, for a given nonnegative decreasing sequence, bounds the ℓ_2 norm of a subsequence in terms of the ℓ_1 norm of another subsequence by shifting the elements to the upper end.

Keywords: ℓ_1 minimization, restricted isometry property, Shifting Inequality, sparse recovery.

1 Introduction

Reconstructing a high-dimensional sparse signal based on a small number of measurements, possibly corrupted by noise, is a fundamental problem in signal processing. This and other related problems in compressed sensing have attracted much recent interest in a number of fields including applied mathematics, electrical engineering, and statistics. Specifically one considers the following model:

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

where the matrix $\Phi \in \mathbb{R}^{n\times p}$ (with $n \ll p$) and $z \in \mathbb{R}^n$ is a vector of measurement errors. The goal is to reconstruct the unknown vector $\beta \in \mathbb{R}^p$. In this paper, our main interest is

*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 Mathematics, Massachusetts Institute of Technology; e-mail: liewang@math.mit.edu

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

the case where the signal β is sparse and the noise z is Gaussian, $z \sim N(0, \sigma^2 I_n)$. We shall approach the problem by considering first the noiseless case and then the bounded noise case, both of significant interest in their own right. The results for the Gaussian case will then follow easily.

It is now well understood that the method of ℓ_1 minimization provides an effective way for reconstructing a sparse signal in many settings. The ℓ_1 minimization method in this context is

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

where $\mathcal{B}$ is a bounded set determined by the noise structure. For example, $\mathcal{B} = \{0\}$ in the noiseless case and $\mathcal{B}$ is the feasible set of the noise in the case of bounded error.

The sparse recovery problem has now been well studied in the framework of the *Restricted Isometry Property (RIP)* introduced by Candes and Tao [7]. A vector $v = (v_i) \in \mathbb{R}^p$ is *k-sparse* if $|\text{supp}(v)| \le k$, where $\text{supp}(v) = \{i : v_i \ne 0\}$ is the support of v. For an $n \times p$ matrix Φ and an integer k, $1 \le k \le p$, the *k-restricted isometry constant* $\delta_k(\Phi)$ is the smallest constant such that

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

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

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

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

It has been shown that ℓ_1 minimization can recover a sparse signal with a small or zero error under various conditions on δ_k and $\theta_{k,k'}$. See, for example, Candes and Tao [7, 8], and Candes, Romberg and Tao [6]. These conditions essentially require that every set of columns of Φ with certain cardinality approximately behaves like an orthonormal system. For example, the condition $\delta_k + \theta_{k,k} + \theta_{k,2k} < 1$ was used in Candes and Tao [7], $\delta_{3k} + 3\delta_{4k} < 2$ in Candes, Romberg and Tao [6], and $\delta_{2k} + \theta_{k,2k} < 1$ in Candes and Tao [8]. Simple conditions involving only δ have also been used in the literature on sparse recovery, for example, $\delta_{2k} < \frac{1}{3}$ was used in Cohen, Dahmen and DeVore [9], and $\delta_{2k} < \sqrt{2} - 1$ was used in Candes [5]. In a recent paper, Cai, Xu and Zhang [4] sharpened the previous results by showing that stable recovery can be achieved under the condition $\delta_{1.5k} + \theta_{k,1.5k} < 1$ (or a stronger but simpler condition $\delta_{1.75k} < \sqrt{2} - 1$).

In the present paper we provide a concise and coherent analysis of the constrained ℓ_1 minimization method for stable recovery of sparse signals. The analysis, which yields strong results, is surprisingly simple and elementary. At the heart of our simplified analysis of the ℓ_1 minimization method for stable recovery is an elementary, yet highly useful, inequality. This inequality, called Shifting Inequality, shows that, given a finite decreasing sequence of nonnegative numbers, the ℓ_2 norm of a subsequence can be bounded in terms of the ℓ_1 norm of another subsequence by "shifting" the terms involved in the ℓ_2 norm to the upper end.

The main contribution of the present paper is two-fold: firstly it is shown that the sparse recovery problem can be solved under weaker conditions and secondly the analysis of the ℓ_1 minimization method can be very elementary and much simplified. In particular, we show that stable recovery of k-sparse signals can be achieved if

$$\delta_{1.25k} + \theta_{k,1.25k} < 1.$$

This condition is weaker than the ones known in the literature. In particular, the results in Candes and Tao [7, 8], Cai, Xu and Zhang [4] and Candes [5] are extended. In fact, our general treatment of this problem produces a family of sparse recovery conditions. Interesting conditions include

$$\delta_{1.625k} < \sqrt{2} - 1 \text{ and } \delta_{3k} < 4 - 2\sqrt{3} \approx 0.535.$$

In the case of Gaussian noise, one of the main results is the following.

Theorem 1 *Consider the model (1) with $z \sim N(0, \sigma^2 I_n)$. If β is k-sparse and*

$$\delta_{1.25k} + \theta_{k,1.25k} < 1,$$

then, the ℓ_1 minimizer $\hat{\beta}^{DS} = \arg\min\{\|\gamma\|_1 : \|\Phi^T(y - \Phi\gamma)\|_\infty \le \sigma\sqrt{2\log p}\}$ satisfies, with high probability,

$$\|\hat{\beta}^{DS} - \beta\|_2 \le \frac{\sqrt{10}}{1 - \delta_{1.25k} - \theta_{k,1.25k}} \sqrt{k}\sigma\sqrt{2\log p}, \tag{5}$$

and the ℓ_1 minimizer $\hat{\beta}^{\ell_2} = \arg\min\{\|\gamma\|_1 : \|y - \Phi\gamma\|_2 \le \sigma\sqrt{n + 2\sqrt{n\log n}}\}$ satisfies, with high probability,

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \le \frac{2\sqrt{2(1 + \delta_{1.25k})}}{1 - \delta_{1.25} - \theta_{k,1.25k}} \sigma\sqrt{n + 2\sqrt{n\log n}}. \tag{6}$$

In comparison to Theorem 1.1 in Candes and Tao (2007), the result given in (5) for $\hat{\beta}^{DS}$ weakens the condition from $\delta_{2k} + \theta_{k,2k} < 1$ to $\delta_{1.25k} + \theta_{k,1.25k} < 1$ and improves the constant in the bound from $4/(1 - \delta_{2k} - \theta_{k,2k})$ to $\sqrt{10}/(1 - \delta_{1.25k} - \theta_{k,1.25k})$. Although our primary interest in this paper is to recovery sparse signals, all the main results in the subsequent sections are given for general signals that are not necessarily k-sparse.

Weakening the RIP condition also has direct implications to the construction of compressed sensing (CS) matrices. It is important to note that it is computationally difficult to verify the RIP for a given design matrix Φ when p is large and the sparsity k is not too small. It is required to bound the condition numbers of $\binom{p}{k}$ submatrices. The spectral norm of a matrix is often difficult to compute and the combinatorial complexity makes it infeasible to check the RIP for reasonable values of p and k. A general technique for avoiding checking the RIP directly is to generate the entries of the matrix Φ randomly and to show that the resulting random matrix satisfies the RIP with high probability using the well-known Johnson-Lindenstrauss Lemma. See, for example, Baraniuk, et al. [2]. Weakening the RIP condition makes it easier to prove that the resulting random matrix satisfies the CS properties.

The paper is organized as follows. After Section 2, in which basic notation and definitions are reviewed, we introduce in Section 3.1 the elementary Shifting Inequality, which enables us to make finer analysis of the sparse recovery problem. We then consider the problem of exact recovery in the noiseless case in Section 3.2 and stable recovery of sparse signals in Section 3.3. The Gaussian noise case is treated in Section 4. Section 5 discusses various conditions on RIP and effects of the improvement of the RIP condition on the construction of CS matrices. The proofs of some technical results are relegated to the Appendix.

2 Preliminaries

We begin by introducing basic notation and definitions related to the RIP. We also collect a few elementary results needed for the later sections.

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

For a subset T of $\{1, \cdots, p\}$, we use Φ_T to denote the submatrix obtained by extracting the columns of Φ according to the indices in T. Let $\mathcal{SSV}_T = \{\lambda : \lambda \text{ an eigenvalue of } \Phi_T'\Phi_T\}$, and $\Lambda_{\min}(k) = \min\{\cup_{|T|\le k}\mathcal{SSV}_T\}$, $\Lambda_{\max}(k) = \max\{\cup_{|T|\le k}\mathcal{SSV}_T\}$. It can be seen that

$$1 - \delta_k \le \Lambda_{\min}(k) \le \Lambda_{\max}(k) \le 1 + \delta_k.$$

Hence the condition (3) can be viewed as a condition on $\Lambda_{\min}(k)$ and $\Lambda_{\max}(k)$.

The following relations can be easily checked.

$$\delta_k \le \delta_{k_1}, \text{ if } k \le k_1 \le p \tag{7}$$

$$\theta_{k,k'} \le \theta_{k_1,k_1'}, \text{ if } k \le k_1, k' \le k_1', \text{ and } k_1 + k_1' \le p. \tag{8}$$

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

$$\theta_{k,k'} \le \delta_{k+k'} \le \theta_{k,k'} + \max(\delta_k, \delta_{k'}). \tag{9}$$

Cai, Xu and Zhang obtained the following properties for δ and θ in [4], which are especially useful in producing simplified recovery conditions:

$$\theta_{k,\sum_{i=1}^l k_i} \le \sqrt{\sum_{i=1}^l \theta_{k,k_i}^2} \le \sqrt{\sum_{i=1}^l \delta_{k+k_i}^2}. \tag{10}$$

Consider the ℓ_1 minimization problem $(P_\mathcal{B})$. Let β be a feasible solution to $(P_\mathcal{B})$, i.e., $y - \Phi\beta \in \mathcal{B}$. Without loss of generality we assume that $\text{supp}(\beta_{\max(k)}) = \{1, 2, \cdots, k\}$. Let $\hat\beta$ be a solution to the minimization problem $(P_\mathcal{B})$. Then it is clear that $\|\hat\beta\|_1 \le \|\beta\|_1$. Let $h = \hat\beta - \beta$ and $h_0 = h\mathbb{I}_{\{1,2,\cdots,k\}}$ for some positive integer $k \le p$. Here $\mathbb{I}_A$ denotes the indicator function of a set $A \subseteq \{1, 2, \cdots, p\}$, i.e., $\mathbb{I}_A(j) = 1$ if $j \in A$ and 0 if $j \notin A$.

The following is a widely used fact. See, for example, [4, 6, 8, 13].

Lemma 1

$$\|h - h_0\|_1 \le \|h_0\|_1 + 2\|\beta_{-max(k)}\|_1.$$

This follows from the fact that $\|\beta\|_1 \ge \|\hat\beta\|_1 = \|\beta_{\max(k)} + h_0\|_1 + \|h - h_0 + \beta_{-\max(k)}\|_1 \ge \|\beta_{\max(k)}\|_1 - \|h_0\|_1 + \sum_{i\ge 1}\|h_i\|_1 - \|\beta_{-\max(k)}\|_1$.

Note also that the Cauchy-Schwarz Inequality yields that for $u \in \mathbb{R}^k$

$$\|u\|_1 \le \sqrt{k}\,\|u\|_2. \tag{11}$$

3 Shifting Inequality, Exact and Stable Recovery

In this section, we consider exact recovery of high-dimensional sparse signals in the noiseless case and stable recovery in the bounded noise case. Recovery of sparse signals with Gaussian noise will be discussed in Section 4. We begin by introducing an elementary inequality which we call the *Shifting Inequality*. This useful inequality plays a key role in our analysis of the properties of the solution to the ℓ_1 minimization problem.

3.1 The Shifting Inequality

The following elementary inequality enables us to perform finer estimation involving ℓ_1 and ℓ_2 norms as can be seen from the proofs of Theorem 2 in Section 3.2 and other main results.

Lemma 2 (Shifting Inequality) *Let q, r be positive integers satisfying $r \leq q \leq 3r$. Then any descending chain of real numbers*

$$a_1 \geq a_2 \geq \cdots a_r \geq b_1 \geq \cdots b_q \geq c_1 \geq \cdots \geq c_r \geq 0$$

satisfies

$$\sqrt{\sum_{i=1}^{q} b_i^2 + \sum_{i=1}^{r} c_i^2} \leq \frac{\sum_{i=1}^{r} a_i + \sum_{i=1}^{q} b_i}{\sqrt{q+r}}. \tag{12}$$

In particular, any descending chain of real numbers

$$b_1 \geq \cdots \geq b_q \geq 0$$

satisfies

$$\sqrt{\sum_{i=1}^{q} b_i^2 + r b_q^2} \leq \frac{r b_1 + \sum_{i=1}^{q} b_i}{\sqrt{q+r}}.$$

Proof of this lemma is presented in the Appendix.

Remark 1 A particularly useful case is $q = 3r$. Let

$$d_1 \geq \cdots \geq d_r \geq d_{r+1} \geq \cdots \geq d_{4r} \geq \cdots \geq d_{5r} \geq 0.$$

Then Lemma 2 yields

$$\sum_{j=r+1}^{5r} d_j^2 \leq \frac{1}{4r} \left(\sum_{j=1}^{4r} d_j \right)^2.$$

We will see that the Shifting Inequality, albeit very elementary, not only simplifies the analysis of ℓ_1 minimization method but also weakens the required condition on the RIP.

3.2 Exact Recovery of Sparse Signals

We shall start with the simple setting where no noise is present. In this case the goal is to recover the signal β exactly when it is sparse. This case is of significant interest in its own right as it is also closely connected to the problem of decoding of linear codes. See, for example, Candes and Tao [7]. The ideas used in treating this special case can be easily extended to treat the general case where noise is present.

Suppose $y = \Phi\beta$. Based on (Φ, y), we wish to reconstruct the vector β exactly when it is sparse. Equivalently, we wish to find the sparest representation of the signal y in the dictionary consisting of the columns of the matrix Φ. Let $\hat{\beta}$ be the minimizer to the problem

$$\text{(Exact)} \qquad \min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \quad \text{subject to} \quad \Phi\gamma = y. \tag{13}$$

Note that this is a special case of the ℓ_1 minimization problem $(P_{\mathcal{B}})$ with $\mathcal{B} = \{0\}$. We have the following result.

Theorem 2 *Suppose that β is k-sparse and that*

$$\delta_{k+a} + \sqrt{\frac{k}{b}}\theta_{k+a,b} < 1 \tag{14}$$

holds for some positive integers a and b satisfying $2a \le b \le 4a$. Then the solution $\hat{\beta}$ to the ℓ_1 minimization problem (Exact) *recovers β exactly. In general, if (14) holds, then $\hat{\beta}$ satisfies*

$$\|\hat{\beta} - \beta\|_2 \le \frac{1 - \delta_{k+a} + \theta_{k+a,b}}{1 - \delta_{k+a} - \sqrt{\frac{k}{b}}\theta_{k+a,b}} \frac{2}{\sqrt{b}} \|\beta_{-max(k)}\|_1.$$

Remark 2 We should note that in this and following main theorems, we use the general condition $\delta_{k+a} + \sqrt{\frac{k}{b}}\theta_{k+a,b} < 1$, which involves two positive intergers a and b, in addition to the sparsity parameter k. The flexibility in the choice of a and b in the condition allows one to derive interesting conditions for compressed sensing matrices. More discussions on special cases and comparisons with the existing conditions used in the current literature are given in Section 5.

Remark 3 A particularly interesting choice is $b = k$ and $a = k/4$. Theorem 2 shows that if β is k-sparse and

$$\delta_{1.25k} + \theta_{k,1.25k} < 1, \tag{15}$$

then the ℓ_1 minimization method recovers β exactly. This condition is weaker than other conditions on RIP currently available in the literature. Compare, for example, Candes and

Tao [7, 8], Candes, Romberg and Tao [6], Candes [5], and Cai, Xu and Zhang [4]. See more discussions in Section 5.

Proof. The proof of Theorem 2 is elementary. The key to the proof is the Shifting Inequality. Again, set $h = \hat{\beta} - \beta$. We shall cut the error vector h into pieces and then apply the Shifting Inequality to subvectors.

Without loss of generality, we assume the first k coordinates of β are the largest in magnitude. Making rearrangement if necessary, we may also assume that

$$|h(k+1)| \geq |h(k+2)| \geq \cdots .$$

Set $T_0 = \{1, 2, \cdots, k\}$, $T_* = \{k+1, k+2, \cdots, k+a\}$ and $T_i = \{k+a+(i-1)b+1, \cdots, k+a+ib\}$, $i = 1, 2, ...$, with the last subset of size less than or equal to b. Let $h_0 = h\mathbb{I}_{T_0}$, $h_* = h\mathbb{I}_{T_*}$ and $h_i = h\mathbb{I}_{T_i}$ for $i \geq 1$.

$h = \hat{\beta} - \beta$:	h_0	h_*	h_1	h_2	h_3	$\cdots$
Support Size:	k	a	b	b	b	$\cdots$

To apply the Shifting Inequality, we shall first divide each vector h_i into two pieces. Set $T_{i1} = \{k+a+(i-1)b+1, \cdots, k+ib\}$ and $T_{i2} = T_i \setminus T_{i1} = \{k+1+ib, \cdots, k+a+ib\}$. We note that $|T_{i1}| = b-a$ and $|T_{i2}| = a$ for all $i \geq 1$. Let $h_{i1} = h_i\mathbb{I}_{T_{i1}}$ and $h_{i2} = h_i\mathbb{I}_{T_{i2}}$.

	$\cdots$	$h_{(i-1)1}$	$h_{(i-1)2}$	h_{i1}	h_{i2}	$\cdots$
length			a	$b-a$	a	$\cdots$

Note that $a \leq b - a \leq 3a$. Applying the Shifting Inequality (12) to the vectors $\{h_*, h_{11}, h_{12}\}$ and $\{h_{(i-1)2}, h_{i1}, h_{i2}\}$ for $i = 2,\ 3, ...$ yields

$$\|h_1\|_2 \leq \frac{\|h_*\|_1 + \|h_{11}\|_1}{\sqrt{b}}, \|h_2\|_2 \leq \frac{\|h_{12}\|_1 + \|h_{21}\|_1}{\sqrt{b}}, \cdots, \|h_i\|_2 \leq \frac{\|h_{(i-1)2}\|_1 + \|h_{i1}\|_1}{\sqrt{b}}, \cdots .$$

It then follows from Lemma 1 and the inequality (11) that

$$\begin{aligned}
\sum_{i\geq 1} \|h_i\|_2 &\leq \frac{\|h_*\|_1 + \sum_{i\geq 1}\|h_i\|_1}{\sqrt{b}} = \frac{\|h - h_0\|_1}{\sqrt{b}} \\
&\leq \frac{\|h_0\|_1 + 2\|\beta_{-max(k)}\|_1}{\sqrt{b}} \leq \sqrt{\frac{k}{b}}\|h_0\|_2 + \frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}} \\
&\leq \sqrt{\frac{k}{b}}\|h_0 + h_*\|_2 + \frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}}.
\end{aligned}$$

Now the fact that $\Phi h = 0$ yields

$$\begin{aligned} 0 &= |\langle \Phi h, \Phi(h_0 + h_*)\rangle| = |\langle \Phi(h_0+h_*), \Phi(h_0+h_*)\rangle + \sum_{i\geq 1} \langle \Phi h_i, \Phi(h_0+h_*)\rangle| \\ &\geq (1-\delta_{k+a})\|h_0+h_*\|_2^2 - \sum_{i\geq 1} \theta_{k+a,b}\|h_0+h_*\|_2\|h_i\|_2 \\ &\geq \|h_0+h_*\|_2 \left(\big(1-\delta_{k+a} - \sqrt{\frac{k}{b}}\theta_{k+a,b}\big)\|h_0+h_*\|_2 - \theta_{k+a,b}\frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}} \right) \end{aligned}$$

This implies

$$\|h_0+h_*\|_2 \leq \frac{\theta_{k+a,b}}{1-\delta_{k+a}-\sqrt{\frac{k}{b}}\theta_{k+a,b}} \frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}}.$$

Therefore,

$$\begin{aligned} \|h\|_2 &\leq \|h_0+h_*\|_2 + \sum_{i\geq 1}\|h_i\|_2 \leq \big(1+\sqrt{\frac{k}{b}}\big)\|h_0+h_*\|_2 + \frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}} \\ &\leq \frac{1-\delta_{k+a}+\theta_{k+a,b}}{1-\delta_{k+a}-\sqrt{\frac{k}{b}}\theta_{k+a,b}} \frac{2}{\sqrt{b}}\|\beta_{-max(k)}\|_1. \end{aligned}$$

If β is k-sparse, then $\beta_{-max(k)} = 0$, which implies $\beta = \hat{\beta}$. ■

The key argument used in the proof of Theorem 2 is the Shifting Inequality. This simply analysis requires a condition on the RIP that is weaker than other conditions on the RIP used in the literature.

In addition to Theorem 2, we also have the following result under s simpler condition.

Theorem 3 *Let k be a positive integer and suppose*

$$\delta_k + \sqrt{k}\theta_{k,1} < 1. \tag{16}$$

Then $\hat{\beta}$ satisfies

$$\|\hat{\beta} - \beta\|_2 \leq \frac{2(1-\delta_k+\theta_{k,1})}{1-\delta_k-\sqrt{k}\theta_{k,1}}\|\beta_{-max(k)}\|_1.$$

In particular, if β is k-sparse, the ℓ_1 minimization recovers β exactly.

Proof. The proof is similar to that of Theorem 2. For each $i \geq 1$, let $T_i = \{k+i\}$ and

$h_i = h\mathbb{I}_{T_i}$. We note

$$\begin{aligned}
|\langle \Phi h, \Phi h_0\rangle| &= |\langle \Phi h_0, \Phi h_0\rangle + \sum_{i\geq 1}\langle \Phi h_i, \Phi h_0\rangle| \\
&\geq (1-\delta_k)\|h_0\|_2^2 - \sum_{i\geq 1}\theta_{k,1}\|h_0\|_2\|h_i\|_2 \\
&= \|h_0\|_2\left((1-\delta_k)\|h_0\|_2 - \theta_{k,1}\sum_{i\geq 1}\|h_i\|_2\right) \\
&= \|h_0\|_2\left((1-\delta_k)\|h_0\|_2 - \theta_{k,1}\sum_{i\geq 1}\|h_i\|_1\right) \\
&\geq \|h_0\|_2\left((1-\delta_k)\|h_0\|_2 - \theta_{k,1}(\|h_0\|_1 + 2\|\beta_{-max(k)}\|_1)\right) \\
&\geq \|h_0\|_2\left((1-\delta_k)\|h_0\|_2 - \theta_{k,1}\left(\sqrt{k}\|h_0\|_2 + 2\|\beta_{-max(k)}\|_1\right)\right) \\
&\geq \|h_0\|_2\left(\left(1-\delta_k-\sqrt{k}\theta_{k,1}\right)\|h_0\|_2 - 2\theta_{k,1}\|\beta_{-max(k)}\|_1\right).
\end{aligned}$$

The remaining steps are the same as those in the proof of Theorem 2. ■

3.3 Recovery in the Presence of Errors

We now consider reconstruction of high dimensional sparse signals in the presence of *bounded* noise. Let $\mathcal{B} \subset I\!R^n$ be a bounded set. Suppose we observe (Φ, y) where $y = \Phi\beta + z$ with the error vector $z \in \mathcal{B}$, and we wish to reconstruct β by solving the ℓ_1 minimization problem $(P_{\mathcal{B}})$. Specifically, we consider two types of bounded errors: $\mathcal{B}_1(\eta) = \{z : \|\Phi' z\|_\infty \leq \eta\}$ and $\mathcal{B}_2(\eta) = \{z : \|z\|_2 \leq \eta\}$. We shall use $\hat{\beta}^{DS}$ to denote the solution of the ℓ_1 minimization problem $(P_{\mathcal{B}})$ with $\mathcal{B} = \mathcal{B}_1(\eta)$ and use $\hat{\beta}^{\ell_2}$ to denote the solution of $(P_{\mathcal{B}})$ with $\mathcal{B} = \mathcal{B}_2(\eta)$.

The Shifting Inequality again plays a key role in our analysis in this case. In addition, the analysis of the Gaussian noise case follows easily from that of the bounded noise case.

Theorem 4 *Suppose*

$$\delta_{k+a} + \sqrt{\frac{k}{b}}\theta_{k+a,b} < 1 \tag{17}$$

holds for positive integers k, a and b where $2a \leq b \leq 4a$. Then the minimizers $\hat{\beta}^{DS}$ and $\hat{\beta}^{\ell_2}$ satisfy

$$\|\hat{\beta}^{DS} - \beta\|_2 \leq A\eta + B\|\beta_{-max(k)}\|_1,$$

and

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \leq C\eta + B\|\beta_{-max(k)}\|_1$$

where

$$A = 2\sqrt{1+\frac{k}{b}}\frac{\sqrt{k+a}}{1-\delta_{k+a}-\sqrt{\frac{k}{b}}\theta_{k+a,b}}, \tag{18}$$

$$B = \frac{2}{\sqrt{b}}\left(1+\frac{\theta_{k+a,b}\sqrt{1+k/b}}{1-\delta_{k+a}-\sqrt{\frac{k}{b}}\theta_{k+a,b}}\right), \tag{19}$$

$$C = 2\sqrt{1+\frac{k}{b}}\frac{\sqrt{1+\delta_{k+a}}}{1-\delta_{k+a}-\sqrt{\frac{k}{b}}\theta_{k+a,b}}. \tag{20}$$

A proof of Theorem 4 based on the ideas of that for Theorem 2 is given in the appendix.

Remark 4 As in the noiseless setting, an especially interesting case is $b = k$ and $a = k/4$. In this case, Theorem 4 yields that if β is k-sparse and

$$\delta_{1.25k} + \theta_{k,1.25k} < 1 \tag{21}$$

holds, then the ℓ_1 minimizers $\hat{\beta}^{DS}$ and $\hat{\beta}^{\ell_2}$ satisfy

$$\|\hat{\beta}^{DS} - \beta\|_2 \le \frac{\sqrt{10}}{1-\delta_{1.25k}-\theta_{k,1.25k}} \cdot \sqrt{k}\eta, \tag{22}$$

and

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \le \frac{2\sqrt{2(1+\delta_{1.25k})}}{1-\delta_{1.25k}-\theta_{k,1.25k}} \cdot \eta. \tag{23}$$

Again, the condition (21) for stable recovery in the noisy case is weaker than the existing RIP conditions in the literature. See, for example, Candes and Tao [7, 8], Candes, Romberg and Tao [6], Candes [5]), and Cai, Xu and Zhang [4].

Remark 5 A generalization of Theorem 3 can also be obtained for the bounded noise case. Suppose β is k-sparse and

$$\delta_k + \sqrt{k}\theta_{k,1} < 1.$$

holds. Then the ℓ_1 minimizers $\hat{\beta}^{DS}$ and $\hat{\beta}^{\ell_2}$ satisfy

$$\|\hat{\beta}^{DS} - \beta\|_2 \le \frac{2\sqrt{k^2+k}}{1-\delta_k-\sqrt{k}\theta_{k,1}} \cdot \eta \quad \text{and} \quad \|\hat{\beta}^{\ell_2} - \beta\|_2 \le \frac{2\sqrt{1+\delta_k}\sqrt{1+k}}{1-\delta_k-\sqrt{k}\theta_{k,1}} \cdot \eta.$$

4 Gaussian Noise

The Gaussian noise case is of particular interest in statistics and several methods have been developed. See, for example, Tibshirani [19], Efron, et al. [14], and Candes and Tao [8]. The results presented in Section 3.3 on the bounded noise case are directly applicable to the case where the noise is Gaussian. This is due to the fact that Gaussian noise is "essentially bounded". Suppose we observe

$$y = \Phi\beta + z, \quad z \sim N(0, \sigma^2 I_n) \tag{24}$$

and wish to recover the signal β based on $(\Phi,\ y)$. We assume that σ is known and that the columns of Φ are standardized to have unit ℓ_2 norm. Define two bounded sets

$$\mathcal{B}_3 = \{z:\ \|\Phi^T z\|_\infty \le \sigma\sqrt{2\log p}\} \quad \text{and} \quad \mathcal{B}_4 = \{z:\ \|z\|_2 \le \sigma\sqrt{n + 2\sqrt{n\log n}}\} \tag{25}$$

The following result, which follows from standard probability calculations, shows that the Gaussian noise z is essentially bounded. The readers are referred to Cai, Xu and Zhang [4] for a proof.

Lemma 3 *The Gaussian error $z \sim N(0, \sigma^2 I_n)$ satisfies*

$$P\left(z \in \mathcal{B}_3\right) \ge 1 - \frac{1}{2\sqrt{\pi\log p}} \quad \textit{and} \quad P\left(z \in \mathcal{B}_4\right) \ge 1 - \frac{1}{n}. \tag{26}$$

Lemma 3 indicates that the Gaussian variable z is in the bounded sets $\mathcal{B}_3$ and $\mathcal{B}_4$ with high probability. The results obtained in the previous sections for bounded errors can thus be applied directly to treat Gaussian noise. In this case, we shall consider two particular constrained ℓ_1 minimization problems. Let $\hat{\beta}^{DS}$ be the minimizer of

$$\min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \quad \text{subject to} \quad y - \Phi\gamma \in \mathcal{B}_3 \tag{27}$$

and let $\hat{\beta}^{\ell_2}$ be the minimizer of

$$\min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \quad \text{subject to} \quad y - \Phi\gamma \in \mathcal{B}_4. \tag{28}$$

The following theorem is a direct consequence of Lemma 3 and Theorem 4.

Theorem 5 *Suppose*

$$\delta_{k+a} + \sqrt{\frac{k}{b}}\theta_{k+a,b} < 1 \tag{29}$$

holds for some positive integers k, a and b with $2a \le b \le 4a$. Then with probability at least $1 - \frac{1}{2\sqrt{\pi \log p}}$ the minimizer $\hat{\beta}^{DS}$ satisfies

$$\|\hat{\beta}^{DS} - \beta\|_2 \le A\sigma\sqrt{2\log p} + B\|\beta_{-max(k)}\|_1,$$

and with probability at least $1 - \frac{1}{n}$, the minimizer $\hat{\beta}^{\ell_2}$ satisfies

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \le C\sigma\sqrt{n + 2\sqrt{n\log n}} + B\|\beta_{-max(k)}\|_1,$$

where the constants A, B and C are given as in Theorem 4.

Remark: Again, a special case is $b = k$ and $a = k/4$. In this case, if β is k-sparse and

$$\delta_{1.25k} + \theta_{k,1.25k} < 1,$$

then, with high probability, the ℓ_1 minimizers $\hat{\beta}^{DS}$ and $\hat{\beta}^{\ell_2}$ satisfy

$$\|\hat{\beta}^{DS} - \beta\|_2 \le \frac{\sqrt{10}}{1 - \delta_{1.25k} - \theta_{k,1.25k}}\sqrt{k}\sigma\sqrt{2\log p} \quad (30)$$

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \le \frac{2\sqrt{2(1 + \delta_{1.25k})}}{1 - \delta_{1.25} - \theta_{k,1.25k}}\sigma\sqrt{n + 2\sqrt{n\log n}}. \quad (31)$$

The result given in (30) for $\hat{\beta}^{DS}$ improves Theorem 1.1 of Candes and Tao [8] by weakening the condition from $\delta_{2k} + \theta_{k,2k} < 1$ to $\delta_{1.25k} + \theta_{k,1.25k} < 1$ and reducing the constant in the bound from $4/(1 - \delta_{2k} - \theta_{k,2k})$ to $\sqrt{10}/(1 - \delta_{1.25k} - \theta_{k,1.25k})$. 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.

Candes and Tao [8] also derived an oracle inequality for $\hat{\beta}^{DS}$ in the Gaussian noise setting under the condition $\delta_{2k} + \theta_{k,2k} < 1$. Our method can also be used to improve Theorems 1.2 and 1.3 in Candes and Tao [8] by weakening the condition to $\delta_{1.25k} + \theta_{k,1.25k} < 1$.

5 Discussions

The flexibility in the choice of a and b in the condition $\delta_{k+a} + \sqrt{\frac{k}{b}}\theta_{k+a,b} < 1$ used in Theorems 2, 4 and 5 enables us to deduce interesting conditions for compressed sensing matrices. We shall highlight several of them here and compare with the existing conditions used in the current literature. As mentioned in the introduction, it is sometimes more convenient to use conditions only involving the restricted isometry constant δ and for this reason we shall mainly focus on δ. By choosing different values of a and b and using equation (10), it is easy to show that each of the following conditions is sufficient for the exact recovery of k-sparse signals in the noiseless case and stable recovery in the noisy case:

1. $\delta_{1.25k} + \theta_{k,1.25k} < 1$

2. $\delta_{1.625k} < \sqrt{2} - 1 \approx 0.414$

3. $\delta_{2k} < \sqrt{6} - 2 \approx 0.449$

4. $\delta_{3k} < 2(2 - \sqrt{3}) \approx 0.535$

5. $\delta_{4k} < 2 - \sqrt{2} \approx 0.585$

For instance, Condition 2 follows from Condition 1 and equation (10). In fact, if $\delta_{1.625k} < \sqrt{2} - 1$, then

$$\delta_{1.25k} + \theta_{k,1.25k} \leq \delta_{1.625k} + \sqrt{\delta_{k+0.625k} + \delta_{k+0.625k}} < 1.$$

These conditions for stable recovery improve the conditions used in the literature, e.g., the conditions $\delta_{3k} + 3\delta_{4k} < 2$ in [6], $\delta_{2k} + \theta_{k,2k} < 1$ in [8], $\delta_{1.5k} + \theta_{1.5k,k} < 1$ in [4], $\delta_{2k} < \sqrt{2} - 1$ in [5], and $\delta_{1.75k} < \sqrt{2} - 1$ in [4]. It is also interesting to note that Condition 4 allows δ_{3k} to be large than 0.5.

The flexibility in the condition $\delta_{k+a} + \sqrt{\frac{k}{b}}\theta_{k+a,b} < 1$ also enables us to discuss the asymptotic properties of the RIP conditions. Letting $a = tk, b = 4tk$ and using the equations (7), (8), and (10), it is easy to see that each of the following conditions is sufficient for stable recovery of k-sparse signals:

1. $\delta_{(3t+1)k} < \frac{\sqrt{2t}}{1+\sqrt{2t}}, \quad t \geq \frac{1}{3}$

2. $\delta_{\frac{9t+1}{2}k} < \frac{\sqrt{2t}}{1+\sqrt{2t}}, \;\; t \in (\frac{1}{7}, \frac{1}{3})$

3. $\delta_{(1+t)k} < \frac{\sqrt{2t}}{1+\sqrt{2t}}, \;\; t \in (0, \frac{1}{7}]$

These conditions reveal two asymptotic properties of the restricted isometry constant δ. The first is that δ can be close to 1 (as t gets large), provided that checking the RIP for Φ must be done for sets of columns whose cardinality is much bigger than k, the sparsity for recovery. The second is that if δ is allowed to be small, then checking the RIP for Φ can be done for sets of columns whose cardinality is close to k (as t gets small).

It is clear that with weaker RIP conditions, more matrices can be verified to be compressed sensing matrices. As mentioned in the introduction, for a given $n \times p$ matrix, it is computationally difficult to check its restrictive isometry property. However, it has been very successful in constructing random compressed sensing matrices using δ_r and $\theta_{r,r'}$, see [1, 2, 6, 7, 8, 18].

For example, Baraniuk et al [2] showed that if Φ is an $n \times p$ matrix whose entries are drawn independently according to Gaussian or Bernoulli, then Φ fails to have RIP

$$\delta_r < \alpha$$

with probability less than

$$\tau_\alpha = 2e^{-\frac{\alpha^2(3-\alpha)}{48}n}\left(\frac{cp}{r\alpha}\right)^r,$$

where c is a constant.

It is not hard to see that the probability of failing drops at a considerable rate as the bound α increases and/or the index r decreases. In fact, with a weaker condition $\delta_r < \alpha + \epsilon < 1$, this rate is

$$\frac{\tau_\alpha}{\tau_{\alpha+\epsilon}} = \frac{2e^{-\frac{\alpha^2(3-\alpha)}{48}n}\left(\frac{cp}{r\alpha}\right)^r}{2e^{-\frac{(\alpha+\epsilon)^2(3-\alpha-\epsilon)}{48}n}\left(\frac{cp}{r(\alpha+\epsilon)}\right)^r} \geq e^{\frac{\alpha\epsilon}{12}n}\left(1+\frac{\epsilon}{\alpha}\right)^r.$$

This rate is very large if n is large.

On the other hand, the improvement of RIP conditions can be interpreted as enlarging the sparsity of the signals to be recovered. For example, one of the previous results showed that the condition $\delta_{2k} < \sqrt{2}-1$ ensures the recovery of a k-sparse signal. Replacing the condition by $\delta_{1.625k} < \sqrt{2}-1$, we see that the sparsity of the signals to be recovered is relaxed $\frac{2}{1.625} \approx 1.23$ times.

References

[1] , W. Bajwa, J. Haupt, G. Raz, S. Wright and R. Nowak, Toeplitz-Structured Compressed Sensing Matrices, *14th Workshop on Statistical Signal Processing, 2007*, 294-298.

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

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

[4] T. Cai, G. Xu, and Zhang, On Recovery of Sparse Signals via ℓ_1 Minimization, *IEEE Trans. Inf. Theory*, (2008), to appear.

[5] E. J. Candes, The restricted isometry property and its implications for compressed sensing, *Compte Rendus de l' Academie des Sciences*, Paris, Serie I, **346** 589-592.

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

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

[8] 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.

[9] A. Cohen, W. Dahmen, R. DeVore, Compressed sensing and best k-term approximation, *preprint*, 2006.

[10] 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.

[11] 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.

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

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

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

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

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

[17] R. Gribonval and M. Nielsen, Sparse representations in unions of bases, *IEEE Trans. Inf. Theory*, 49(2003), 3320-3325.

[18] M. Rudelson and R. Vershynin, Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements, *CISS 2006 (40th Annual Conference on Information Sciences and Systems)*, 2006.

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

[20] J. Tropp, Greed is good: algorithmic results for sparse approximation, *IEEE Trans. Inf. Theory*, 50(2004), 2231-2242.

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

APPENDIX

A-1 Proof of the Shifting Inequality (Lemma 2)

Let $b_i = b_{i+1} + d_i$ for $i = 1, 2, \cdots, q-1$. Then

$$\begin{aligned}(q+r)(\sum_{i=1}^{q} b_i^2 + rb_q^2) &= (q+r)\left((q+r)b_q^2 + 2b_q\sum_{i=1}^{q-1}\sum_{j=i}^{q-1} d_j + \sum_{i=1}^{q-1}(\sum_{j=i}^{q-1} d_j)^2\right) \\ &= (q+r)^2 b_q^2 + 2(q+r)b_q\sum_{i=1}^{q-1} i d_i + (q+r)\sum_{i=1}^{q-1}(\sum_{j=i}^{q-1} d_j)^2\end{aligned}$$

And

$$\begin{aligned}(rb_1 + \sum_{i=1}^{q} b_i)^2 &= \left((q+r)b_q + \sum_{i=1}^{q-1}(r+i)d_i\right)^2 \\ &= (q+r)^2 b_q^2 + 2(q+r)b_q\sum_{i=1}^{q-1}(r+i)d_i + \left(\sum_{i=1}^{q-1}(r+i)d_i\right)^2\end{aligned}$$

Note that d_i is nonnegative for all i, so $2(q+r)\sum_{i=1}^{q-1}(r+i)d_i \geq 2(q+r)\sum_{i=1}^{q-1} i d_i$. Also, it can be seen that for any $1 \leq i \leq j \leq q-1$, the coefficient of $d_i d_j$ in $\left(\sum_{i=1}^{q-1}(r+i)d_i\right)^2$ is $(1 + I(i \neq j))(r+i)(r+j)$[1]. And the coefficient of $d_i d_j$ in $(q+r)\sum_{i=1}^{q-1}(\sum_{j=i}^{q-1} d_j)^2$ is $(1 + I(i \neq j))(q+r)i$. Since $q \leq 3r$, we know that

$$(r+i)(r+j) \geq (r+i)^2 = (r-i)^2 + 4ri \geq (q+r)i.$$

This means

$$\left(\sum_{i=1}^{q-1}(r+i)d_i\right)^2 \geq (q+r)\sum_{i=1}^{q-1}(\sum_{j=i}^{q-1} d_j)^2.$$

Hence the inequality is proved.

[1]The number $I(i \neq j)$ is 1 unless $i = j$, in which case it is 0

A-2 Proof of Theorem 4

Similar to the proof of theorem 2, we have

$$|\langle \Phi h, \Phi(h_0 + h_*)\rangle| \geq \|h_0 + h_*\|_2 \left(\big(1 - \delta_{k+a} - \sqrt{\frac{k}{b}}\theta_{k+a,b}\big)\|h_0 + h_*\|_2 - \theta_{k+a,b}\frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}} \right)$$

Case I. $\mathcal{B} = \{z : \|z\|_2 \leq \eta\}$. It is easy to see that

$$|\langle \Phi h, \Phi(h_0 + h_*)\rangle| \leq \|\Phi h\|_2 \|\Phi(h_0 + h_*)\|_2 \leq 2\eta\sqrt{1 + \delta_{k+a}}\|h_0 + h_*\|_2$$

Therefore,

$$\|h_0 + h_*\|_2 \leq \frac{2\eta\sqrt{1 + \delta_{k+a}} + \frac{2\theta_{k+a,b}}{\sqrt{b}}\|\beta_{-max(k)}\|_1}{1 - \delta_{k+a} - \sqrt{\frac{k}{b}}\theta_{k+a,b}}$$

Now

$$\begin{aligned} \|h\|_2^2 &= \|h_0 + h_*\|_2^2 + \sum_{i\geq 1}\|h_i\|_2^2 \leq \|h_0 + h_*\|_2^2 + \big(\sum_{i \geq 1}\|h_i\|_2\big)^2 \\ &\leq \|h_0 + h_*\|_2^2 + \left(\sqrt{\frac{k}{b}}\|h_0 + h_*\|_2 + \frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}}\right)^2 \\ &= \left(\sqrt{1 + \frac{k}{b}}\|h_0 + h_*\|_2 + \frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}}\right)^2 \leq \big(C\eta + B\|\beta_{-max(k)}\|_1\big)^2. \end{aligned}$$

Case II. $\mathcal{B} = \{z : \|\Phi' z\|_\infty \leq \eta\}$.
By assumption, there is a $z \in \mathcal{B}$ such that $\Phi\beta = y - z$. So

$$\begin{aligned} |\langle \Phi h, \Phi(h_0 + h_*)\rangle| &= |\langle \Phi\hat{\beta} - y + z, \Phi_{T_0 \cup T_*}(h_0 + h_*)\rangle| \\ &= |\langle \Phi'_{T_0\cup T_*}\big(\Phi\hat{\beta} - y + z\big), (h_0 + h_*)\rangle| \\ &\leq \|\Phi'_{T_0\cup T_*}\big(\Phi\hat{\beta} - y + z\big)\|_2 \|h_0 + h_*\|_2 \\ &\leq 2\sqrt{k + a}\,\eta\|h_0 + h_*\|_2. \end{aligned}$$

This implies

$$\|h_0 + h_*\|_2 \leq \frac{2\eta\sqrt{k + a} + \frac{2\theta_{k+a,b}}{\sqrt{b}}\|\beta_{-max(k)}\|_1}{1 - \delta_{k+a} - \sqrt{\frac{k}{b}}\theta_{k+a,b}}$$

Similar to Case I, we have

$$\begin{aligned} \|h\|_2^2 &\leq \left(\sqrt{1 + \frac{k}{b}}\|h_0 + h_*\|_2 + \frac{2\|\beta_{-max(k)}\|_1}{\sqrt{b}}\right)^2 \\ &\leq \big(A\eta + B\|\beta_{-max(k)}\|_1\big)^2. \quad \blacksquare \end{aligned}$$