Compressed Sensing and Affine Rank Minimization Under Restricted Isometry

T. Tony Cai and Anru Zhang

Abstract—This paper establishes new restricted isometry conditions for compressed sensing and affine rank minimization. It is shown for compressed sensing that $\delta_k^A + \theta_{k,k}^A < 1$ guarantees the exact recovery of all k sparse signals in the noiseless case through the constrained ℓ_1 minimization. Furthermore, the upper bound 1 is sharp in the sense that for any $\epsilon > 0$, the condition $\delta_k^A + \theta_{k,k}^A < 1 + \epsilon$ is not sufficient to guarantee such exact recovery using any recovery method. Similarly, for affine rank minimization, if $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ then all matrices with rank at most r can be reconstructed exactly in the noiseless case via the constrained nuclear norm minimization; and for any $\epsilon > 0$, $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1 + \epsilon$ does not ensure such exact recovery using any method. Moreover, in the noisy case the conditions $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ are also sufficient for the stable recovery of sparse signals and low-rank matrices respectively. Applications and extensions are also discussed.

Index Terms—Affine rank minimization, compressed sensing, Dantzig selector, constrained ℓ_1 minimization, low-rank matrix recovery, constrained nuclear norm minimization, restricted isometry, sparse signal recovery.

I. Introduction

COMPRESSED sensing has received much recent attention in signal processing, applied mathematics and statistics. A closely related problem is affine rank minimization. The central goal in these problems is to accurately reconstruct a high dimensional object of a certain special structure, namely a sparse signal in compressed sensing and a low-rank matrix in affine rank minimization, through a small number of linear measurements. Interesting applications of compressed sensing and affine rank minimization include coding theory [1], [13], magnetic resonance imaging [22], signal acquisition [16], [29], radar system [4], [21], [32] and image compression [27], [30].

In compressed sensing, one wishes to recover a signal $\beta \in \mathbb{R}^p$ based on (A, y) where

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

Here $A \in \mathbb{R}^{n\times p}$ is a given sensing matrix and $z \in \mathbb{R}^n$ is the measurement error. In affine rank minimization, one observes

$$y = \mathcal{M}(X) + z \tag{2}$$

where $\mathcal{M} : \mathbb{R}^{m\times n} \to \mathbb{R}^q$ is a known linear map, $X \in \mathbb{R}^{m\times n}$ is an unknown matrix, and $z \in \mathbb{R}^q$ is an error vector. The goal is to reconstruct X based on y and the linear map $\mathcal{M}$. In these problems, the dimension is typically much larger than the number of measurements, i.e., $p \gg n$ and $\min(m, n) \gg q$. A rather remarkable fact is that, when the signal β is sparse and the matrix X has low rank, they can be reconstructed exactly in the noiseless case and stably in the noisy case using computational efficient algorithms, provided that the sensing matrix A and the linear map $\mathcal{M}$ satisfy certain restricted orthogonality conditions.

For the reconstruction of β and X, the most intuitive approach is to find the sparsest signal or the lowest-rank matrix in the feasible set of possible solutions, i.e.,

$$\begin{aligned} \text{minimize } \|\beta\|_0, &\quad \text{subject to } A\beta - y \in \mathcal{B} \\ \text{minimize rank}(X), &\quad \text{subject to } \mathcal{M}(X) - y \in \mathcal{B} \end{aligned}$$

where $\|\beta\|_0$ denote the ℓ_0 norm of β, which is defined to be the number of nonzero coordinates, and $\mathcal{B}$ is a bounded set determined by the error structure. However, it is well-known that such methods are NP-hard and thus computationally infeasible in the high dimensional settings. Convex relaxations of these methods have been proposed and studied in the literature. Candès and Tao [13] introduced an ℓ_1 minimization method for the sparse signal recovery and Recht, *et al.* [27] proposed a nuclear norm minimization method for the matrix reconstruction,

$$\hat{\beta} = \arg\min_{\beta}\{\|\beta\|_1 \text{ subject to } A\beta - y \in \mathcal{B}\}, \tag{3}$$

$$X_* = \arg\min_{X}\{\|X\|_* \text{ subject to } \mathcal{M}(X) - y \in \mathcal{B}\}, \tag{4}$$

where $\|X\|_*$ is the nuclear norm of X which is defined to be the sum of all singular values of X. Here $\mathcal{B} = \{0\}$ in the noiseless case and $\mathcal{B}$ is the feasible set of the error vector z when z is bounded. These methods have been shown to be effective for the recovery of sparse signals and low-rank matrices in a range of settings. See, e.g., [13], [14], [18], [27], [15].

One of the most commonly used frameworks for compressed sensing is the *Restricted Isometry Property* (RIP) introduced in [13]. The RIP framework was later extended to the affine rank minimization problem by Recht *et al.* in [27]. A vector is said to be *k-sparse* if $|\text{supp}(v)| \le k$, where $\text{supp}(v) = \{i : v_i \ne 0\}$ is the support of v. We shall use the phrase"r-rank matrices" to refer to matrices of rank at most r. For matrices $X = (x_{ij}) \in \mathbb{R}^{m\times n}$, and $Y = (y_{ij}) \in \mathbb{R}^{m\times n}$, define the inner product of X and Y as $\langle X, Y\rangle = \text{trace}(X^TY) = \sum_{i=1}^m \sum_{j=1}^n x_{ij}y_{ij}$. The norm associated with this inner product is the Frobenius norm, $\|X\|_F = \sqrt{\langle X, X\rangle} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2}$. The following definitions are given by [13], [27], [23].

Manuscript received November 19, 2012; revised February 26, 2013; accepted April 11, 2013. Date of publication April 19, 2013; date of current version June 03, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ljubisa Stankovic. The research was supported in part by NSF FRG Grant DMS-0854973 and NIH Grant R01 CA127334.

The authors are with the Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, 19104 USA (e-mail: tcai@wharton.upenn.edu; anrzhang@wharton.upenn.edu).

Digital Object Identifier 10.1109/TSP.2013.2259164

Definition 1.1: Let $A \in \mathbb{R}^{n\times p}$ and let $1 \le k, k_1, k_2 \le p$ be integers. The restricted isometry constant (RIC) of order k is defined to be the smallest non-negative number δ_k^A such that

$$\left(1-\delta_k^A\right)\|\beta\|_2^2 \le \|A\beta\|_2^2 \le \left(1+\delta_k^A\right)\|\beta\|_2^2 \quad (5)$$

for all k-sparse vectors β. The restricted orthogonality constant (ROC) of order (k_1, k_2) is defined to be the smallest non-negative number θ_{k_1,k_2}^A such that

$$|\langle A\beta_1, A\beta_2\rangle| \le \theta_{k_1,k_2}^A \|\beta_1\|_2\|\beta_2\|_2 \quad (6)$$

for all k_1-sparse vector β_1 and k_2-sparse vector β_2 with disjoint supports.

Similarly, let $\mathcal{M} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear map and let $1 \le r, r_1, r_2 \le \min(m,n)$ be integers. The restricted isometry constant (RIC) of order r is defined to be the smallest non-negative number $\delta_r^{\mathcal{M}}$ such that

$$\left(1-\delta_r^{\mathcal{M}}\right)\|X\|_F^2 \le \|\mathcal{M}(X)\|_2^2 \le \left(1+\delta_r^{\mathcal{M}}\right)\|X\|_F^2 \quad (7)$$

for all $m \times n$ matrix X of rank at most r. The restricted orthogonality constant (ROC) of order (r_1, r_2) is defined to be the smallest non-negative number $\theta_{r_1,r_2}^{\mathcal{M}}$ such that

$$|\langle \mathcal{M}(X_1), \mathcal{M}(X_2)\rangle| \le \theta_{k_1,k_2}^{\mathcal{M}} \|X_1\|_F\|X_2\|_F \quad (8)$$

for all matrices X_1 and X_2 which have rank at most r_1 and r_2 respectively, and satisfy $X_1^T X_2 = 0$ and $X_1 X_2^T = 0$.

In addition to RIP, another widely used criterion is the mutual incoherence property (MIP) defined in terms of $\mu = \max_{i\neq j}|\langle A_i, A_j\rangle|$. See, for example, [19], [7]. The MIP is a special case of the restricted orthogonal property as $\mu = \theta_{1,1}$ when the columns of A are normalized.

Roughly speaking, the RIC δ_k^A and ROC θ_{k_1,k_2}^A measure how far subsets of cardinality k of columns of A are to an orthonormal system. It is obvious that δ_k and θ_{k_1,k_2} are increasing in each of their indices. It is noteworthy that our definition of ROC in the matrix case is different from the one given in [23].

Sufficient conditions in terms of the RIC and ROC for the exact recovery of k-sparse signals in the noiseless case include $\delta_k^A + \theta_{k,k}^A + \theta_{k,2k}^A < 1$ [13]; $\delta_{2k}^A + \theta_{k,2k}^A < 1$ [14]; $\delta_{1.5k}^A + \theta_{k,1.5k}^A < 1$ [5], $\delta_{1.25k}^A + \theta_{k,1.25k}^A < 1$ [6], and $\theta_{1,1}^A < \frac{1}{2k-1}$ when $\delta_1^A = 0$ [19], [20], [7]. Sufficient conditions for the exact recovery of r-rank matrices include $\delta_{2r+\alpha r} + \frac{1}{\sqrt{\beta}}\theta_{2r+\alpha r,\beta r} < 1$ where $2\alpha \le \beta \le 4\alpha$ [23]. It is however unclear if any of these conditions can be further improved.

In this paper we establish more relaxed RIP conditions for sparse signal and low-rank matrix recovery. More specifically, we show that the condition

$$\delta_k^A + \theta_{k,k}^A < 1 \quad (9)$$

guarantees the exact recovery of all k-sparse signals in the noiseless case via the constrained ℓ_1 minimization (3) with $\mathcal{B} = \{0\}$. Furthermore, we show that the constant 1 in (9) is sharp in the sense that for any $\epsilon > 0$, the condition $\delta_k^A + \theta_{k,k}^A < 1 + \epsilon$ is not sufficient to guarantee such exact recovery using any method. Similarly it is shown that the condition

$$\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1 \quad (10)$$

is sufficient for the exact reconstruction of all r-rank matrices in the noiseless case through the constrained nuclear norm minimization (4) with $\mathcal{B} = \{0\}$, and that for any $\epsilon > 0$, the condition $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1 + \epsilon$ is not sufficient to guarantee such exact recovery using any method. Moreover, in the noisy case the conditions (9) and (10) also guarantee the stable recovery of sparse signals and low-rank matrices respectively. In addition to the sufficient conditions (9) and (10), extensions to the more general RIP conditions are also considered.

The new RIP conditions are weaker than the known RIP conditions in the literature. The techniques and results developed in the present paper have a number of applications in signal processing, including the design of compressed sensing matrices, signal acquisition, and analysis of compressed sensing based radar system. We discuss these applications in Section IV.

The rest of the paper is organized as follows. In Section II, we first introduce the basic notations and definitions and then present the main results for both sparse signal recovery and low-rank matrix recovery. Extensions of the results $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ to the more general RIP conditions are also considered. Section III discusses the relationship between our results and other known RIP conditions. Section IV illustrates some applications of the results in signal processing. The proofs of the main results are given in Section V.

II. New RIP Conditions

We present the main results in this section. It will be first shown that the conditions $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ are sharp for the exact recovery in the noiseless case and stable recovery in the noisy case. The more general RIP conditions will be considered at the end of this section.

Let us begin with basic notation. For $v \in \mathbb{R}^p$, $v_{\max(k)}$ is defined as the vector v with all but the largest k entries in absolute value set to zero, and $v_{-\max(k)} = v - v_{\max(k)}$. For a matrix $X \in \mathbb{R}^{m\times n}$ (without loss of generality, assume that $m \le n$) with the singular value decomposition $X = \sum_{i=1}^m a_i u_i v_i^T$ where the singular values a_i are in descending order $a_1 \ge a_2 \ge \cdots \ge a_m \ge 0$, we define $X_{\max(r)} = \sum_{i=1}^r a_i u_i v_i^T$ and $X_{-\max(r)} = X - X_{\max(r)}$. We should also note that the nuclear norm $\|\cdot\|_*$ of a matrix equals the sum of the singular values, and the spectral norm $\|\cdot\|$ of a matrix equals its largest singular value. Their roles are similar to those of ℓ_1 norm and ℓ_∞ norm in the vector case, respectively. For a linear operator $\mathcal{M} : \mathbb{R}^{m\times n} \to \mathbb{R}^q$, we denote its dual operator by $\mathcal{M}^* : \mathbb{R}^q \to \mathbb{R}^{m\times n}$.

It follows from [25] that the results for the low-rank matrix recovery are parallel to those for the sparse signal recovery. So we shall present the results for the two problems together in this section. The following theorem shows that the conditions (9) and (10) guarantee the exact recovery of all k-sparse signals and r-rank matrices through the constrained ℓ_1 minimization and constrained nuclear norm minimization respectively.

Theorem 2.1: Let $\beta \in \mathbb{R}^p$ be a k-sparse vector and $y = A\beta$. If $\delta_k^A + \theta_{k,k}^A < 1$, then $\hat{\beta} = \beta$, where $\hat{\beta}$ is the minimizer of (3) with $\mathcal{B} = \{0\}$. Similarly, let X be an r-rank matrix and $y = \mathcal{M}(X)$. If $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$, then $X_* = X$, where X_* is the minimizer of (4) with $\mathcal{B} = \{0\}$.

We now turn to the noisy case. Although our main focus is on the recovery of sparse signals and low-rank matrices, we shall state the results for general signals and matrices that are not necessarily sparse or low-rank.

We consider two bounded noise settings: $\|z\|_2 \le \epsilon$, and $\|A^T z\|_\infty \le \epsilon$ (signal case) and $\|\mathcal{M}^*(z)\| \le \epsilon$ (matrix case). The case of Gaussian noise, which is of significant interest in statistics, can be essentially reduced to the bounded noise case. See, for example, Section 4 in [6] for more discussions. In the theorems below, we shall write δ for δ_k^A and $\delta_k^{\mathcal{M}}$ and write θ for $\theta_{k,k}^A$ and $\theta_{k,k}^{\mathcal{M}}$. We first consider the case where the ℓ_2 norm of the error vector z is bounded.

Theorem 2.2: Consider the signal recovery model (1) with $\|z\|_2 \le \epsilon$. Let $\hat{\beta}$ be the minimizer of (3) with $\mathcal{B} = \{z \in \mathbb{R}^n : \|z\|_2 \le \eta\}$ for some $\eta \ge \epsilon$. If $\delta_k^A + \theta_{k,k}^A < 1$ for some $k \ge 1$, then

$$\|\hat{\beta} - \beta\|_2 \le \frac{\sqrt{2(1+\delta)}}{1-\delta-\theta}(\epsilon+\eta) + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\left(\frac{\sqrt{2}\theta}{1-\delta-\theta}+1\right). \quad (11)$$

Similarly, consider the matrix recovery model (2) with $\|z\|_2 \le \epsilon$. Let X_* be the minimizer of (4) with $\mathcal{B} = \{z \in \mathbb{R}^q : \|z\|_2 \le \eta\}$ for some $\eta \ge \epsilon$. If $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ for some $r \ge 1$, then

$$\|X_* - X\|_F \le \frac{\sqrt{2(1+\delta)}}{1-\delta-\theta}(\epsilon+\eta) + \frac{2\|X_{-\max(r)}\|_*}{\sqrt{r}}\left(\frac{\sqrt{2}\theta}{1-\delta-\theta}+1\right). \quad (12)$$

We now consider the case where the error vector z is in a polytope defined by $\|A^T z\|_\infty \le \epsilon$ and $\|\mathcal{M}^*(z)\| \le \epsilon$. This case is motivated by the Dantzig Selector method considered in [14] for the Gaussian noise case.

Theorem 2.3: Consider the signal recovery model (1) with $\|A^T z\|_\infty \le \epsilon$. Let $\hat{\beta}$ be the minimizer of (3) with $\mathcal{B} = \{z \in \mathbb{R}^n : \|A^T z\|_\infty \le \eta\}$ for some $\eta \ge \epsilon$. If $\delta_k^A + \theta_{k,k}^A < 1$ for some $k \ge 1$, then

$$\|\hat{\beta} - \beta\|_2 \le \frac{\sqrt{2k}}{1-\delta-\theta}(\epsilon+\eta) + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\left(\frac{\sqrt{2}\theta}{1-\delta-\theta}+1\right). \quad (13)$$

Similarly, suppose we have the signal and matrix recovery model (2) with $\|\mathcal{M}^*(z)\| \le \epsilon$. Let $\hat{\beta}$, X_* be the minimizer of (4) with $\mathcal{B} = \{z \in \mathbb{R}^q : \|\mathcal{M}^*(z)\| \le \eta\}$ for some $\eta \ge \epsilon$. If $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ for some $r \ge 1$, then

$$\|X_* - X\|_F \le \frac{\sqrt{2r}}{1-\delta-\theta}(\epsilon+\eta) + \frac{2\|X_{-\max(r)}\|_*}{\sqrt{r}}\left(\frac{\sqrt{2}\theta}{1-\delta-\theta}+1\right). \quad (14)$$

Theorems 2.1, 2.2, and 2.3 shows that the conditions $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ are respectively sufficient for the exact and stable reconstruction of sparse signals and low-rank matrices via the constrained ℓ_1 minimization and nuclear norm minimization. The following theorem shows that the upper bound 1 in these conditions is in fact sharp.

Theorem 2.4: Let $1 \le k \le p/2$. There exists a sensing matrix $A \in \mathbb{R}^{n\times p}$ such that $\delta_k^A + \theta_{k,k}^A = 1$ and for some k-sparse signals $u, v \in \mathbb{R}^p$ with $u \ne v$, $Au = Av$. Consequently, there does not exist any method that can exactly recover all k-sparse signals β based on (A, y) with $y = A\beta$.

Let $1 \le r \le \min(m,n)/2$. There exists a linear map $\mathcal{M}$ such that $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} = 1$ and for some matrices $U, V \in \mathbb{R}^{m\times n}$ with $\mathrm{rank}(U),\ \mathrm{rank}(V) \le r$, and $\mathcal{M}(U) = \mathcal{M}(V)$. Therefore, it is impossible for any method to recover all r-rank matrices exactly based on $(\mathcal{M}, y)$ with $y = \mathcal{M}(X)$.

Remark 2.1: Theorem 2.4 implies that for any $\epsilon > 0$, $\delta_k^A + \theta_{k,k}^A < 1+\epsilon$ fails to guarantee the exact recovery of all k-sparse signals. These results immediately show that for any $\epsilon > 0$, the condition $\delta_k^A + \theta_{k,k}^A < 1+\epsilon$ or $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1+\epsilon$ is not sufficient to ensure in the noisy case stably recovery of all k-sparse signals and all r-rank matrices.

Remark 2.2: The results on the bounded noise case can be applied to immediately yield the corresponding results for the Gaussian noise case by using the same argument as in [5], [6]. We illustrate this point for the signal recovery. Suppose $z \sim \mathcal{N}_n(0, \sigma^2)$ in (1). Define $\mathcal{B}^{DS} = \{z : \|\Phi^T z\|_\infty \le \sigma\sqrt{2\log p}\}$ and $\mathcal{B}^{\ell_2} = \{z : \|z\|_2 \le \sigma\sqrt{n + 2\sqrt{n\log n}}\}$. Then, with probability at least $1 - \frac{1}{\sqrt{\pi \log p}}$, the Dantzig selector $\hat{\beta}^{DS}$ given by (3) with $\mathcal{B} = \mathcal{B}^{DS}$ satisfies

$$\|\hat{\beta}^{DS} - \beta\|_2 \le \frac{2\sqrt{2}}{1-\delta-\theta}\sigma\sqrt{2k\log p} + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\left(\frac{\sqrt{2}\theta}{1-\delta-\theta}+1\right), \quad (15)$$

and the ℓ_2 constraint minimizer $\hat{\beta}^{\ell_2}$ defined in (3) with $\mathcal{B} = \mathcal{B}^{\ell_2}$ satisfies

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \le \frac{2\sqrt{2(1+\delta)}}{1-\delta-\theta}\sigma\sqrt{n + 2\sqrt{n\log n}} + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\left(\frac{\sqrt{2}\theta}{1-\delta-\theta}+1\right), \quad (16)$$

with probability at least $1 - 1/n$. We refer readers to [5], [6] for further details.

A. Extensions to More General RIP Conditions

We have shown that the conditions $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ are sufficient respectively for sparse signal recovery and for low-rank matrix recovery. The same techniques can be used to extend the results to a more general form,

$$\delta_a^A + C_{a,b,k}\theta_{a,b}^A < 1, \quad \text{where } C_{a,b,k} = \max\left\{\frac{2k-a}{\sqrt{ab}}, \sqrt{\frac{2k-a}{a}}\right\}, 1 \le a \le k, \quad (17)$$

$$\delta_a^{\mathcal{M}} + C_{a,b,r}\theta_{a,b}^{\mathcal{M}} < 1, \quad \text{where } C_{a,b,r} = \max\left\{\frac{2r-a}{\sqrt{ab}}, \sqrt{\frac{2r-a}{a}}\right\}, 1 \le a \le r. \quad (18)$$

Theorem 2.5: In the noiseless case, Theorem 2.1 holds with the conditions $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ replaced by (17) and (18) respectively.

In the noisy case, we have the following two theorems parallel to Theorems 2.2 and 2.3.

Theorem 2.6: Consider the signal recovery model (1) with $\|z\|_2 \leq \epsilon$. Let $\hat{\beta}$ be the minimizer of (3) with $\mathcal{B} = \{z \in \mathbb{R}^n : \|z\|_2 \leq \eta\}$ for some $\eta \geq \epsilon$. If $\delta_a^A + C_{a,b,k}\theta_{a,b}^A < 1$ for some positive integers a and b with $1 \leq a \leq k$, then

$$\|\hat{\beta} - \beta\|_2 \leq \frac{\sqrt{2(1+\delta)k/a}}{1-\delta-C_{a,b,k}\theta}(\epsilon+\eta) + 2\|\beta_{-\max(k)}\|_1 \times \left(\frac{\sqrt{2k}C_{a,b,k}\theta}{(1-\delta-C_{a,b,k}\theta)(2k-a)} + \frac{1}{\sqrt{k}}\right). \quad (19)$$

Similarly, consider the matrix recovery model (2) with $\|z\|_2 \leq \epsilon$. Let X_* be the minimizer of (4) with $\mathcal{B} = \{z \in \mathbb{R}^q : \|z\|_2 \leq \eta\}$ for some $\eta \geq \epsilon$. If $\delta_a^{\mathcal{M}} + C_{a,b,r}\theta_{a,b}^{\mathcal{M}} < 1$ for some positive integers a and b with $1 \leq a \leq r$, then

$$\|X_* - X\|_F \leq \frac{\sqrt{2(1+\delta)r/a}}{1-\delta-C_{a,b,r}\theta}(\epsilon+\eta) + 2\|X_{-\max(r)}\|_* \times \left(\frac{\sqrt{2r}C_{a,b,r}\theta}{(1-\delta-C_{a,b,r}\theta)(2r-a)} + \frac{1}{\sqrt{r}}\right). \quad (20)$$

Theorem 2.7: Consider the signal recovery model (1) with $\|A^Tz\|_\infty \leq \epsilon$. Let $\hat{\beta}$ be the minimizer of (3) with $\mathcal{B} = \{z \in \mathbb{R}^n : \|A^Tz\|_\infty \leq \eta\}$ for some $\eta \geq \epsilon$. If $\delta_a^A + C_{a,b,k}\theta_{a,b}^A < 1$ for some positive integers a and b with $1 \leq a \leq k$, then

$$\|\hat{\beta} - \beta\|_2 \leq \frac{\sqrt{2k}}{1-\delta-C_{a,b,k}\theta}(\epsilon+\eta) + 2\|\beta_{-\max(k)}\|_1 \times \left(\frac{\sqrt{2k}C_{a,b,k}\theta}{(1-\delta-C_{a,b,k}\theta)(2k-a)} + \frac{1}{\sqrt{k}}\right). \quad (21)$$

Similarly, suppose we have the signal and matrix recovery model (2) with $\|\mathcal{M}^*(z)\| \leq \epsilon$. Let $\hat{\beta}$, X_* be the minimizer of (4) with $\mathcal{B} = \{z \in \mathbb{R}^q : \|\mathcal{M}^*(z)\| \leq \eta\}$ for some $\eta \geq \epsilon$. If $\delta_a^{\mathcal{M}} + C_{a,b,r}\theta_{a,b}^{\mathcal{M}} < 1$ for some integers a and b with $1 \leq a \leq r$, then

$$\|X_* - X\|_F \leq \frac{\sqrt{2r}}{1-\delta-C_{a,b,r}\theta}(\epsilon+\eta) + 2\|X_{-\max(r)}\|_* \times \left(\frac{\sqrt{2r}C_{a,b,r}\theta}{(1-\delta-C_{a,b,r}\theta)(2r-a)} + \frac{1}{\sqrt{r}}\right). \quad (22)$$

The next theorem shows that the upper bound 1 in the conditions $\delta_a^A + C_{a,b,k}\theta_{a,b}^A < 1$ and $\delta_a^{\mathcal{M}} + C_{a,b,r}\theta_{a,b} < 1$ cannot be further improved.

Theorem 2.8: Let $1 \leq k \leq p/2$, $1 \leq a \leq k$, and $b \geq 1$. Let $C_{a,b,k}$ be defined as (17). Then there exists a sensing matrix $A \in \mathbb{R}^{n\times p}$ such that $\delta_a^A + C_{a,b,k}\theta_{a,b}^A = 1$ and for some k-sparse signals $u, v \in \mathbb{R}^p$ with $u \neq v$, $Au = Av$. Consequently, there does not exist any method that can exactly recover all k-sparse signals β based on (A, y) with $y = A\beta$.

Similarly, let $1 \leq r \leq \min(m,n)/2$, $1 \leq a \leq k$ and $b \geq 1$. Let $C_{a,b,r}$ be defined as (18). Then there exists a linear map $\mathcal{M}$ such that $\delta_a^{\mathcal{M}} + C_{a,b,r}\theta_{a,b}^{\mathcal{M}} = 1$ and for some matrices $U, V \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(U),\ \operatorname{rank}(V) \leq r$, and $\mathcal{M}(U) = \mathcal{M}(V)$. Consequently, it is impossible for any method to exactly recover all r-rank matrices based on $(\mathcal{M}, y)$ with $y = \mathcal{M}(X)$.

Same as Theorem 2.4, Theorem 2.8 implies that in the noisy case stably recovery of all k-sparse signals and all r-rank matrices cannot be guaranteed by $\delta_a^A + C_{a,b,k}\theta_{a,b}^A < 1 + \epsilon$ or $\delta_a^{\mathcal{M}} + C_{a,b,r}\theta_{a,b}^{\mathcal{M}} < 1 + \epsilon$ for any $\epsilon > 0$.

Remark 2.3: We established the more general RIP conditions $\delta_a^A + C_{a,b,r}\theta_{a,b}^A < 1$ and $\delta_a^{\mathcal{M}} + C_{a,b,r}\theta_{a,b,r}^{\mathcal{M}} < 1$. For fixed a, among these conditions, the one with $b = 2k - a$ or $b = 2r - a$ is the weakest. We shall illustrate this for the signal case. By Lemma 5.4,

$$\begin{aligned}
&\delta_a + C_{a,2k-a,k}\theta_{a,2k-a} \\
&\quad \leq \delta_a + C_{a,2k-a,k}\sqrt{\frac{2k-a}{\min\{b,2k-a\}}}\theta_{a,\min\{b,2k-a\}} \\
&\quad = \delta_a + \sqrt{\frac{2k-a}{a}} \cdot \sqrt{\frac{2k-a}{\min\{b,2k-a\}}}\theta_{a,\min\{b,2k-a\}} \\
&\quad = \delta_a + C_{a,b,k}\theta_{a,\min\{b,2k-a\}} \leq \delta_a + C_{a,b,k}\theta_{a,b}
\end{aligned}$$

Hence, for all $b \geq 1$, $\delta_a + C_{a,b,k}\theta_{a,b} < 1$ implies $\delta_a + C_{a,2k-a,k}\theta_{a,2k-a} < 1$.

III. Relationship to Other Restricted Isometry Conditions

In the last section, we have established the sufficient conditions $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ for the exact recovery in the noiseless case and stable recovery in the noisy case. We discuss in this section the relationships between these conditions and other restricted isometry conditions introduced in the literature.

By the simple fact that for $k_1 \leq k_2$ and $k_1' \leq k_2'$, $\delta_{k_1}^A \leq \delta_{k_2}^A$ and $\theta_{k_1,k_1'}^A \leq \theta_{k_2,k_2'}^A$, it is easy to see that the condition $\delta_k^A + \theta_{k,k}^A < 1$ is weaker than $\delta_k^A + \theta_{k,k}^A + \theta_{k,2k}^A < 1$, $\delta_{2k}^A + \theta_{2k,k}^A < 1$, $\delta_{1.5k}^A + \theta_{1.5k,k}^A < 1$ and $\delta_{1.25k}^A + \theta_{1.25k,k}^A < 1$, which were mentioned in the introduction. Note that setting $a = b = 1$ in the condition $\delta_a + C_{a,b,k}\theta_{a,b} < 1$ yields a sufficient condition $\delta_1^A + (2k-1)\theta_{1,1} < 1$ which is more general than the MIP condition $\theta_{1,1} < \frac{1}{2k-1}$ when $\delta_1^A = 0$ given in [19] and [20] for the noiseless case and [7] for the noisy case.

There are also several sufficient conditions in the literature that are based on the RIC δ alone, such as $\delta_{3k}^A + 3\delta_{4k}^A < 2$ [10], $\delta_{2k}^A < \sqrt{2} - 1$ [11]; $\delta_{2k}^A < 0.472$ [6]; $\delta_k^A < 0.307$ [8]; $\delta_{2k}^A < 0.493$ [26] and $\delta_k^A < 1/3$ and $\delta_{2k}^A < 1/2$ [9]. For the matrix recovery, sufficient conditions include $\delta_{4r}^{\mathcal{M}} < \sqrt{2} - 1$ [15]; $\delta_{5r}^{\mathcal{M}} < 0.607$, $\delta_{4r}^{\mathcal{M}} < 0.558$, and $\delta_{3r}^{\mathcal{M}} < 0.4721$ [23]; $\delta_{2r}^{\mathcal{M}} < 0.4931$ and $\delta_r^{\mathcal{M}} < 0.307$ [31], and $\delta_r^{\mathcal{M}} < 1/3$ and $\delta_{2r}^{\mathcal{M}} < 1/2$ [9]. In particular, Cai and Zhang [9] showed that $\delta_k^A < 1/3$ and $\delta_r^{\mathcal{M}} < 1/3$ are sharp RIP conditions for the exact recovery. It is interesting to compare these results on δ_k^A, δ_{2k}^A, $\delta_r^{\mathcal{M}}$, and $\delta_{2r}^{\mathcal{M}}$ with $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$.

The following lemma provides a bound for the ROC θ in terms of the RIC δ and can be used to compare different RIP conditions.

Lemma 3.1: Let $A \in \mathbb{R}^{n\times p}$. Then we have

$$\theta_{k,k}^A \le \begin{cases} 2\delta_k^A, & \text{when } k \text{ is even}, k\ge 2;\\ \dfrac{2k}{\sqrt{k^2-1}}\delta_k^A, & \text{when} k \text{ is odd}, k\ge 3.\end{cases} \tag{23}$$

In addition, both coefficients, 2 in the even case and $\frac{2k}{\sqrt{k^2-1}}$ in the odd case, cannot be further improved.

In the matrix case, suppose $\mathcal{M}: \mathbb{R}^{m\times n} \to \mathbb{R}^q$ is a linear map, then

$$\theta_{r,r}^{\mathcal{M}} \le \begin{cases} 2\delta_r^{\mathcal{M}}, & \text{when } r \text{ is even}, r\ge 2;\\ \dfrac{2r}{\sqrt{r^2-1}}\delta_r^{\mathcal{M}}, & \text{when } r \text{ is odd}, r\ge 3.\end{cases} \tag{24}$$

In addition, the coefficient 2 in the even case cannot be further improved.

With Lemma 3.1, we can naturally obtain the following result which shows that the conditions $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ are mostly weaker than the RIP conditions $\delta_k^A < 1/3$ and $\delta_r^{\mathcal{M}} < 1/3$ respectively.

Proposition 3.1: If $\delta_k^A < 1/3$ for some integer $k \ge 2$, then

$$\begin{aligned} &\delta_k^A + \theta_{k,k}^A < 1, && \text{when } k \text{ is even};\\ &\delta_k^A + \theta_{k,k}^A < \frac{1}{3} + \frac{2k}{3\sqrt{k^2-1}} \approx 1 + \frac{1}{3k^2}, && \text{when } k \text{ is odd}.\end{aligned} \tag{25}$$

Similarly in the matrix case, if $\delta_r^{\mathcal{M}} < 1/3$ for some integer $r \ge 2$, then

$$\begin{aligned} &\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1, && \text{when } k \text{ is even};\\ &\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < \frac{1}{3} + \frac{2r}{3\sqrt{r^2-1}} \approx 1 + \frac{1}{3r^2}, && \text{when } k \text{ is odd}.\end{aligned} \tag{26}$$

Sufficient conditions in terms of δ_{2k}^A and $\delta_{2r}^{\mathcal{M}}$ are also commonly used in the literature. To the best of our knowledge, the weakest bounds on δ_{2k}^A and $\delta_{2r}^{\mathcal{M}}$ for the exact recovery are $\delta_{2k}^A \le 1/2$ and $\delta_{2r}^{\mathcal{M}} \le 1/2$ given by Cai and Zhang [9]. It is easy to see that the conditions $\delta_k^A + \theta_{k,k}^A < 1$ and $\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1$ given in the present paper are strictly weaker than these conditions respectively.

Proposition 3.2: If $\delta_{2k}^A < 1/2$ for some integer $k \ge 1$, then

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

Similarly, if $\delta_{2r}^{\mathcal{M}} < 1/2$ for some integer $r \ge 1$, then

$$\delta_r^{\mathcal{M}} + \theta_{r,r}^{\mathcal{M}} < 1.$$

This is an immediate consequence of the results given in Section II and the following lemma given in [15].

Lemma 3.2: Suppose $A \in \mathbb{R}^{n\times p}$ and $\mathcal{M}$ is a linear map from $\mathbb{R}^{m\times n}$ to $\mathbb{R}^q$, then

$$\theta_{k,k}^A \le \delta_{2k}^A, \quad \theta_{r,r}^{\mathcal{M}} \le \delta_{2r}^{\mathcal{M}}. \tag{27}$$

IV. Applications

As mentioned earlier, compressed sensing and affine rank minimization have a wide range of applications. The techniques and results developed in this paper naturally have a number of applications in signal processing, including the design of compressed sensing matrices, signal acquisition, and analysis of compressed sensing based radar system. We discuss some of these applications in this section.

An important problem in compressed sensing is the design of sensing matrices that guarantee the exact recovery in the noiseless case and stable recovery in the noisy case. Different types of matrices have been shown to satisfy the previously known sufficient RIP or MIP conditions with high probability. Examples include i.i.d. Gaussian matrices [13], [14], general random matrix satisfying concentration inequality [3], Toeplitz-structured matrices [2], structurally random matrices [17] and the matrices from transmission waveform optimization [32]. These matrices are thus provably suitable for compressed sensing. A direct consequence of the weaker RIP condition obtained in this paper is that a smaller number of measurements are required to guarantee the exact or stable recovery of sparse signals.

Take for example i.i.d. Gaussian or Bernoulli random matrices. Theorem 5.2 in [3] shows that if a random sensing matrix $A = (a_{ij}) \in \mathbb{R}^{n\times p}$ satisfies

$$a_{ij} \overset{iid}{\sim} \mathcal{N}(0, 1/n), \quad \text{or} \quad a_{ij} \overset{iid}{\sim} \begin{cases} 1/\sqrt{n} & \text{w.p. } 1/2\\ -1/\sqrt{n} & \text{w.p. } 1/2\end{cases},$$
$$\text{or} \quad a_{ij} \overset{iid}{\sim} \begin{cases} \sqrt{3/n} & \text{w.p. } 1/6\\ 0 & \text{w.p. } 1/2\\ -\sqrt{3/n} & \text{w.p. } 1/6\end{cases},$$

then for any positive integer $m < n$ and $0 < t < 1$, the RIC δ_m^A of the matrix A satisfies

$$P\left(\delta_m^A < t\right) \ge 1 - 2\left(\frac{12ep}{mt}\right)^m \exp\left(-n\left(\frac{t^2}{16} - \frac{t^3}{48}\right)\right). \tag{28}$$

It is helpful to compare the condition $\delta_k^A + \theta_{k,k}^A < 1$ in terms of these random sensing matrices to the best known RIP conditions in the literature: $\delta_k < 1/3$ and $\delta_{2k} < 1/2$ [9]. Suppose for some given $0 < \epsilon < 1$ one wishes the sensing matrix A to satisfy the RIP condition $\delta_k^A < 1/3$ or $\delta_{2k}^A < 1/2$ with probability at least $1 - \epsilon$. Then, based on (28), for given k and p the number of measurements n must satisfy respectively

$$n \ge 162\left[k(\log(p/k) + 4.6) - \log(\epsilon/2)\right]$$

and

$$n \ge 153.6\left[k(\log(p/k) + 3.5) - \frac{\log(\epsilon/2)}{2}\right].$$

On the other hand, it is easy to see that $\delta_k^A + \theta_{k,k}^A < 1$ is implied by $\delta_k^A + \delta_{2k}^A < 1$ which is in turn implied by the condition $\delta_k^A < 0.4$ and $\delta_{2k}^A < 0.6$. Note that for given k and p, $n \ge n_1$ with

$$n_1 = 115.4[k(\log(p/k) + 4.4) - \log(\epsilon/4)]$$

guarantees $\delta_k^A < 0.4$ with probability at least $1 - \epsilon/2$, and $n \ge n_2$ with

$$n_2 = 111.1\left[k(\log(p/k) + 3.3) - \frac{\log(\epsilon/4)}{2}\right]$$

ensures $\delta_{2k}^A < 0.6$ with probability at least $1-\epsilon/2$. Hence, $\delta_k^A+\delta_{2k}^A<1$ holds with probability at least $1-\epsilon$ if the number of measurements n satisfies

$$n \geq \max\{n_1, n_2\}. \tag{29}$$

Therefore, for large k and p, the required number of measurements to ensure $\delta_k^A + \theta_{k,k}^A < 1$ is less than 71.2% (115.4/162) and 75.1% (115.4/153.6) of the corresponding required number of measurements to ensure $\delta_k^A < 1/3$ and $\delta_{2k}^A < 1/2$, respectively.

The results given in this paper can also be used for certain theoretical analysis in signal processing. One example is the signal acquisition problem studied in [16]. Davenport *et al.* [16] considered acquiring a finite window of a band-limited signal $x(t)$ given by

$$x(t) = \Psi(\alpha) = \sum_{j=0}^{p-1} \alpha_t \psi_j(t),$$

where $\psi_j(t) = e^{i2\pi jt}$ (i is the imaginary unit) are the Fourier basis functions, and $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_{p-1}]$ is k sparse. Suppose the measurements $y_1, \dots, y_n$ are acquired as

$$\begin{aligned} y_j &= \langle \phi_j(t), x(t)\rangle + z = \left\langle \phi_j(t), \sum_{l=0}^{p-1} \alpha_l \psi_l(t) \right\rangle + z \\ &= \sum_{l=0}^{p-1} \alpha_l \langle \phi_j(t), \psi_l(t)\rangle + z \triangleq \sum_{l=0}^{p-1} r_{jl}\alpha_l + z \end{aligned}$$

where z is measurement error. Then it can be written as

$$y = R\alpha + z, \tag{30}$$

which is exactly (1). When $R = (r_{ij})$ with r_{ij} i.i.d. Gaussian or Bernoulli, as discussed above, the measurement matrix R satisfies the RIP condition of order k or $2k$ with high probability provided that

$$n \gtrsim \kappa_0 k \log(p/k), \tag{31}$$

in which case stable recovery of the signal $x(t)$ can be achieved through ℓ_1 minimization.

The lower bound of κ_0 in (31) is typically computed through simulations [16], [29]. Our results yield a theoretical lower bound for κ_0, namely $\kappa_0 \geq 115.4$ based on (29). It is also helpful to provide an upper bound for the error of recovery. Suppose that $z \sim N_n(0,\sigma^2)$ and Condition (31) is satisfied. Then (15) and (16) yield that the Dantzig selector and ℓ_2 constraint minimizer given in Remark 2.2 satisfy, with high probability,

$$\begin{aligned} \|\hat{x}(t)^{DS} - x(t)\|_2 &= \|\hat{\alpha}^{DS} - \alpha\|_2 \\ &\leq C_1\sigma\sqrt{k\log p} + C_2 \frac{\|\alpha_{-\max(k)}\|_1}{\sqrt{k}} \\ \|\hat{x}(t)^{\ell_2} - x(t)\|_2 &= \|\hat{\alpha}^{\ell_2} - \alpha\|_2 \\ &\leq C_3\sigma\sqrt{n} + C_2 \frac{\|\alpha_{-\max(k)}\|_1}{\sqrt{k}} \end{aligned}$$

where C_1, C_2, C_3 are constants specified in Remark 2.2.

In addition, the results obtained in this paper are also useful in the analysis of compressed sensing based radar system [4]. Suppose the object of interest is represented by $u(t)$ and the transmitted radar pulse for detecting the object is $s_T(t)$. Then the received radar signal is $s_R(t) = c\int s_T(t-\tau)u(\tau)\,dt$. Baraniuk and Steeghs [4] discretizes this equation and the compressed sensing based radar model then becomes

$$s_R(mD\Delta) = c\sum_{n=1}^{N} p(mD-n)u(n\Delta), \quad m = 1, \dots, M$$

which is the same as the compressed sensing model (1) in the noiseless case. Whether it is possible to recover the signal $u(t)$ with accuracy requires checking the condition on the matrix $A = (a_{mn})_{M\times N}$ with $a_{mn} = p(mD-n)$. Weaker RIP condition makes it easier to guarantee the recovery of the signal $u(t)$.

V. Proofs

We now prove the main results of the paper. Throughout this section, we shall call a vector an "indicator vector" if it has only one non-zero entry and the value of this entry is either 1 or -1.

We first state and prove a key technical tool used in the proof of the main results. It provides a way to estimate the inner product $\langle \alpha, \beta\rangle$ and $\langle X_1, X_2\rangle$ by the ROC when only one component is sparse or low-rank.

Lemma 5.1: Let $k_1, k_2 \leq p$ and $\lambda \geq 0$. Suppose $\alpha, \beta \in \mathbb{R}^p$ have disjoint supports and α is k_1-sparse. If $\|\beta\|_1 \leq \lambda k_2$ and $\|\beta\|_\infty \leq \lambda$, then

$$|\langle A\alpha, A\beta\rangle| \leq \theta_{k_1,k_2}^A \|\alpha\|_2 \cdot \lambda\sqrt{k_2}. \tag{32}$$

Let $r_1, r_2 \leq \min\{m,n\}$ and $\lambda \geq 0$. Suppose $X_1, X_2 \in \mathbb{R}^{m\times n}$ satisfy $X_1^T X_2 = 0$, $X_1 X_2^T = 0$, and $\text{rank}(X_1) \leq r_1$. If $\|X_2\|_* \leq \lambda r_2$ and $\|X_2\| \leq \lambda$, then

$$|\langle \mathcal{M}(X_1), \mathcal{M}(X_2)\rangle| \leq \theta_{k_1,k_2}^{\mathcal{M}} \|X_2\|_F \cdot \lambda\sqrt{r_2}. \tag{33}$$

Proof of Lemma 5.1: We first state the following result which characterizes the property of X and Y when $X^TY = 0$ and $XY^T = 0$. The result follows directly from Lemma 2.3 in [27] and we thus omit the proof here.

Lemma 5.2: For $X, Y \in \mathbb{R}^{m\times n}$, $X^TY = 0$, $XY^T = 0$ if and only if there exist orthonormal bases $\{u_i \in \mathbb{R}^m : 1 \leq i \leq m\}$ and $\{v_i \in \mathbb{R}^n 1 \leq i \leq n\}$ such that the singular value decompositions of X and Y have the form

$$X = \sum_{i\in T_1} a_i u_i v_i^T \quad \text{and} \quad Y = \sum_{i\in T_2} b_i u_i v_i^T$$

where T_1 and T_2 are disjoint subsets of $\{1, \dots, \min(m,n)\}$, $a_i, b_j \geq 0$.

We shall only prove the signal case in Lemma 5.1 as the proof for the matrix case is essentially the same. Suppose $\|\beta\|_0 = l$, then β is an l-sparse vector. When $l \leq k_2$, by the definition of δ_{k_1,k_2}^A,

$$|\langle A(\alpha), A(\beta)\rangle| \leq \theta_{k_1,k_2}^A \|\alpha\|_2 \|\beta\|_2 \leq \theta_{k_1,k_2}^A \|\alpha\|_2 \sqrt{k_2}\lambda$$

since $\|\beta\|_\infty \leq \lambda$. Thus (32) holds for $l \leq k_2$.

Now consider the case $l > k_2$. We shall prove by induction. Assume that (32) holds for $l-1$. For l, suppose β can be written as $X_2 = \sum_{i=1}^{l} c_i u_i$, where $c_1 \geq c_2 \geq \cdots \geq c_l > 0$, $\{u_i\}_{i=1}^{l}$

are indicator vectors (defined in the beginning of this section) with different supports. Notice that $\sum_{i=1}^{l} c_i \le \lambda k_2 \le (l-1)\lambda$, so

$$1 \in D \triangleq \{1 \le j \le l-1 : c_j + c_{j+1} + \cdots + c_l \le (l-j)\lambda\},$$

which means D is not empty. We can pick the largest element $j \in D$, which implies

$$\begin{aligned} &c_j + c_{j+1} + \cdots + c_l \le (l-j)\lambda, \\ &c_{j+1} + c_{j+2} + \cdots + c_l > (l-j-1)\lambda. \end{aligned} \tag{34}$$

(It is noteworthy that even if the largest j in D is $l-1$, (34) still holds). Define

$$d_w = \frac{\sum_{i=j}^{l} c_i}{l-j} - c_w, \quad j \le w \le l \tag{35}$$

and

$$\gamma_w = \frac{d_w}{\sum_{i=j}^{l} d_i} \sum_{i=1}^{j-1} c_i u_i + \sum_{i=j, i\neq w}^{l} d_w u_i \in \mathbb{R}^p, \quad j \le i \le l. \tag{36}$$

It is easy to check that $\sum_{w=j}^{l} \gamma_w = \beta$, $\sum_{i=j}^{l} c_i = (l-j)\sum_{i=j}^{l} d_i$. By (34), for all $j \le w \le l$,

$$\begin{aligned} d_w \ge d_j &= \frac{\sum_{i=j+1}^{l} c_i}{l-j} - \frac{l-j-1}{l-j} c_j \\ &\ge \frac{\sum_{i=j+1}^{l} c_i - (l-j-1)\lambda}{l-j} > 0. \end{aligned}$$

We also have

$$\begin{aligned} \|\gamma_w\|_1 &= \frac{d_w}{\sum_{i=j}^{l} d_i} \sum_{i=1}^{j-1} c_i + (l-j) d_w \\ &= \frac{d_w}{\sum_{i=j}^{l} d_i} \left(\sum_{i=1}^{j-1} c_i + \sum_{i=j}^{l} c_i \right) \\ &= \frac{d_w}{\sum_{i=j}^{l} d_i} \|\beta\|_1 \le \frac{d_w}{\sum_{i=j}^{l} d_i} \lambda k_2, \end{aligned}$$

and

$$\begin{aligned} \|\gamma_w\|_\infty &= \max\left\{ \frac{d_w}{\sum_{i=j}^{l} d_i} c_1, \ldots, \frac{d_w}{\sum_{i=j}^{l} d_i} c_{j-1}, d_w \right\} \\ &\le \max\left\{ \frac{d_w}{\sum_{i=j}^{l} d_i} \lambda, \frac{d_w(\sum_{i=j}^{l} c_i)}{(l-j)\left(\sum_{i=j}^{l} d_i\right)} \right\} \le \frac{d_w}{\sum_{i=j}^{l} d_i} \lambda. \end{aligned}$$

The last inequality follows from the first part of (34). Finally since γ_w is $(l-1)$-sparse, we can use the induction assumption,

$$\begin{aligned} |\langle A\alpha, A\beta\rangle| &\le \sum_{w=j}^{l} |\langle A\alpha, A\gamma_w\rangle| \\ &\le \theta_{k_1,k_2}^{A} \|\alpha\|_2 \sum_{w=j}^{l} \frac{d_w}{\sum_{i=j}^{l} d_i} \lambda \sqrt{k_2} \\ &= \theta_{k_1,k_2}^{A} \|\alpha\|_2 \lambda \sqrt{k_2} \end{aligned}$$

which gives (32) for l. □

Proof of Theorems 2.1 and 2.5: It suffices to prove Theorem 2.5 as Theorem 2.1 is a spacial case of Theorem 2.5. We first state two lemmas. Lemma 5.3, which characterizes the null space properties, is from [28] and [24]. Lemma 5.4, which reveals the relationship between ROC's of different orders, is from [6].

Lemma 5.3: In the noiseless case, using (3) with $\mathcal{B} = \{0\}$ one can recover all k-sparse signal β if and only if for all $h \in \mathcal{N}(A)\backslash\{0\}$,

$$2\|h_{\max(k)}\|_1 < \|h\|_1.$$

Similarly in the noiseless case, using (4) with $\mathcal{B} = \{0\}$ one can recover all matrices X of rank at most r if and only if for all $R \in \mathcal{N}(\mathcal{M})\backslash\{0\}$,

$$2\|R_{\max(r)}\|_* < \|R\|_*.$$

Lemma 5.4: For any $\mu \ge 1$ and positive integers k_1, k_2 such that μk_2 is an integer, then

$$\theta_{k_1,\mu k_2} \le \sqrt{\mu}\theta_{k_1,k_2}$$

As mentioned before, by [25], the results for the sparse signal recovery imply the corresponding results for the low-rank matrix recovery. So we will only prove the signal case. By Lemma 5.3, it suffices to show that for all vectors $h \in \mathcal{N}(A) \setminus \{0\}$, $\|h_{\max(k)}\|_1 < \|h_{-\max(k)}\|_1$.

Suppose there exists $h \in \mathcal{N}(A)\backslash\{0\}$ such that $\|h_{\max(k)}\|_1 \ge \|h_{-\max(k)}\|_1$. Let $h = \sum_{i=1}^{p} c_i u_i$, where $\{c_i\}_{i=1}^{p}$ is a non-negative and non-increasing sequence; $\{u_i\}_{i=1}^{p}$ are indicator vectors (defined at the beginning of this section) with different supports in $\mathbb{R}^p$. Then we have $\sum_{i=1}^{k} c_i \ge \sum_{i=k+1}^{p} c_i$. Hence,

$$\|h_{-\max(a)}\|_\infty = c_{a+1} \le \frac{\sum_{i=1}^{a} c_i}{a} = \frac{\|h_{\max(a)}\|_1}{a}$$

and

$$\begin{aligned} \|h_{-\max(a)}\|_1 &= \sum_{i=a+1}^{k} c_i + \sum_{i=k+1}^{p} c_i \le \frac{k-a}{k} \sum_{i=1}^{k} c_i + \sum_{i=1}^{k} c_i \\ &\le \frac{k-a}{a} \sum_{i=1}^{a} c_i + \frac{k}{a} \sum_{i=1}^{a} c_i \\ &= \frac{2k-a}{a} \|h_{\max(a)}\|_1. \end{aligned}$$

We set $\lambda = \frac{\|h_{\max(a)}\|_1}{a}$, $k_1 = a$, $k_2 = 2k-a$, It then follows from Lemma 5.1 that

$$\begin{aligned} &|\langle A(h_{\max(a)}), A(h_{-\max(a)})\rangle| \\ &\quad \le \theta_{a,2k-a}^{A} \sqrt{2k-a} \|h_{\max(a)}\|_2 \cdot \frac{\|h_{\max(a)}\|_1}{a} \\ &\quad \le \theta_{a,2k-a}^{A} \sqrt{\frac{2k-a}{a}} \|h_{\max(a)}\|_2^2. \end{aligned}$$

On the other hand, Lemma 5.4 yields

$$\theta_{a,2k-a} \le \sqrt{\frac{2k-a}{\min\{b,2k-a\}}}\theta_{a,\min\{b,2k-a\}} \le \max\left\{\sqrt{\frac{2k-a}{b}},1\right\}\theta_{a,b}.$$

Hence,

$$\begin{aligned}0 &= |\langle A(h_{\max(a)}), A(h)\rangle| \\ &\ge |\langle A(h_{\max(a)}), A(h_{\max(a)})\rangle| - |\langle A(h_{\max(a)}), A(h_{-\max(a)})\rangle| \\ &\ge \left(1-\delta_a^A\right)\|h_{\max(a)}\|_2^2 - \theta^A_{a,2k-a}\sqrt{\frac{2k-a}{a}}\|h_{\max(a)}\|_2^2 \\ &\ge \left(1-\delta_a^A - \max\left\{\frac{2k-a}{\sqrt{ab}}, \sqrt{\frac{2k-a}{a}}\right\}\theta^A_{a,b}\right)\|h_{\max(a)}\|_2^2 \\ &= \left(1-\delta_a^A - C_{a,b,k}\theta^A_{a,b}\right)\|h_{\max(a)}\|_2^2\end{aligned}$$

which contradicts the fact that $h \neq 0$ and $\delta_a^A + C_{a,b,k}\theta^A_{a,b} < 1$. □

Proof of Theorems 2.2, 2.3, 2.6 and 2.7: Again, it suffices to prove Theorems 2.6 and 2.7. We need the following Lemma 5.5 from [9] which provides an inequality between the sums of the ρth power of two sequences of nonnegative numbers based on the inequality of their sums.

Lemma 5.5: Suppose $m \ge r$, $a_1 \ge a_2 \ge \cdots \ge a_m \ge 0$, and $\sum_{i=1}^r a_i \ge \sum_{i=r+1}^m a_i$. Then for all $\rho \ge 1$,

$$\sum_{j=r+1}^m a_j^\rho \le \sum_{i=1}^r a_i^\rho. \tag{37}$$

More generally, suppose $\lambda \ge 0$, $a_1 \ge a_2 \ge \cdots \ge a_m \ge 0$, and $\sum_{i=1}^r a_i + \lambda \ge \sum_{i=r+1}^m a_i$, then for all $\rho \ge 1$,

$$\sum_{j=r+1}^m a_j^\rho \le r\left(\sqrt[\rho]{\frac{\sum_{i=1}^r a_i^\rho}{r}} + \frac{\lambda}{r}\right)^\rho. \tag{38}$$

We first prove Theorem 2.2. Set $h = \hat{\beta} - \beta$ and $R = X_* - X$. The following inequalities are well known,

$$\|h_{-\max(k)}\|_1 \le \|h_{\max(k)}\|_1 + 2\|\beta_{-\max(k)}\|_1$$
$$\text{and} \quad \|R_{-\max(r)}\|_* \le \|R_{\max(r)}\|_* + 2\|X_{-\max(r)}\|_*.$$

See, for example, [18] (signal case) and [31] (matrix case). Again, we show the signal case only. By the boundedness of z and the definition of the feasible set for $\hat{\beta}$,

$$\|Ah\|_2 \le \|Ah - y\|_2 + \|y - A\hat{\beta}\|_2 \le \varepsilon + \eta. \tag{39}$$

On the other hand, suppose $h = \sum_{i=1}^p c_i u_i$, where $\{c_i\}_{i=1}^p$ are non-negative and non-decreasing, $\{u_i\}_{i=1}^p$ are indicator vectors with different supports. Then

$$\sum_{i=k+1}^m c_i \le \sum_{i=1}^k c_i + 2\|\beta_{-\max(k)}\|_1. \tag{40}$$

Hence,

$$\|h_{-\max(a)}\|_\infty = c_{a+1} \le \frac{\sum_{i=1}^a c_i}{a} = \frac{\|h_{\max(a)}\|_1}{a} \le \frac{\|h_{\max(a)}\|_1}{a} + \frac{2\|\beta_{-\max(k)}\|_1}{2k-a}$$

and

$$\begin{aligned}\|h_{-\max(a)}\|_1 &= \sum_{i=a+1}^k c_i + \sum_{i=k+1}^p c_i \\ &\le \frac{k-a}{k}\sum_{i=1}^k c_i + \sum_{i=1}^k c_i + 2\|\beta_{-\max(k)}\|_1 \\ &\le \frac{k-a}{a}\sum_{i=1}^a c_i + \frac{k}{a}\sum_{i=1}^a c_i + 2\|\beta_{-\max(k)}\|_1 \\ &= \frac{2k-a}{a}\|h_{\max(a)}\|_1 + 2\|\beta_{-\max(k)}\|_1.\end{aligned}$$

Now set $\lambda = \frac{\|h_{\max(a)}\|_1}{a} + \frac{2\|\beta_{-\max(k)}\|_1}{2k-a}$, $k_1 = a$, and $k_2 = 2k-a$. Lemma 5.1 then yields

$$\begin{aligned}&|\langle A(h_{\max(a)}), A(h_{-\max(a)})\rangle| \\ &\le \theta^A_{a,2k-a}\sqrt{2k-a}\|h_{\max(a)}\|_2 \\ &\quad\times\left(\frac{\|h_{\max(a)}\|_1}{a} + \frac{2\|\beta_{-\max(k)}\|_1}{2k-a}\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}|\langle Ah, Ah_{\max(a)}\rangle| &\le \|Ah\|_2\|Ah_{\max(a)}\|_2 \\ &\le (\varepsilon+\eta)\sqrt{1+\delta}\|h_{\max(a)}\|_2.\end{aligned} \tag{41}$$

Now we denote $\theta_{a,2k-a}$ as $\tilde{\theta}$, then

$$\begin{aligned}&(\varepsilon+\eta)\sqrt{1+\delta}\|h_{\max(a)}\|_2 \\ &\ge |\langle Ah, Ah_{\max(a)}\rangle| \\ &\ge \|Ah_{\max(a)}\|_2^2 - |\langle Ah_{-\max(a)}, Ah_{\max(a)}\rangle| \\ &\ge (1-\delta)\|h_{\max(a)}\|_2^2 - \tilde{\theta}\|h_{\max(a)}\|_2 \\ &\quad\cdot\sqrt{2k-a}\left(\frac{\|h_{\max(a)}\|_1}{a} + \frac{2\|\beta_{-\max(k)}\|_1}{2k-a}\right) \\ &\ge \left(1-\delta-\sqrt{\frac{2k-a}{a}}\tilde{\theta}\right)\|h_{\max(a)}\|_2^2 \\ &\quad - \tilde{\theta}\|h_{\max(a)}\|_2\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{2k-a}}.\end{aligned}$$

Hence

$$\begin{aligned}\|h_{\max(a)}\|_2 &\le \frac{\sqrt{1+\delta}(\varepsilon+\eta)}{1-\delta-\sqrt{(2k-a)/a}\tilde{\theta}} \\ &\quad + \frac{\tilde{\theta}}{1-\delta-\sqrt{(2k-a)/a}\tilde{\theta}}\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{2k-a}}.\end{aligned} \tag{42}$$

Applying Lemma 5.5 with $\rho = 2$ and $\lambda = 2\|h_{-\max(k)}\|_1$ yields

$$\begin{aligned}\|h\|_2 &= \sqrt{\sum_{i=1}^{k} c_i^2 + \sum_{i=k+1}^{p} c_i^2} \\ &\le \sqrt{\sum_{i=1}^{k} c_i^2 + \left(\sqrt{\sum_{i=1}^{k} c_i^2} + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}\right)^2} \\ &\le \sqrt{2\sum_{i=1}^{k} c_i^2} + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}} \\ &\le \sqrt{\frac{2k}{a}\sum_{i=1}^{a} c_i^2} + \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}} \\ &\le \frac{\sqrt{2(1+\delta)k/a}(\varepsilon+\eta)}{1-\delta-\sqrt{(2k-a)/a}\tilde{\theta}} \\ &\quad + \left(\frac{\sqrt{2k/a}\tilde{\theta}}{1-\delta-\sqrt{(2k-a)/a}\tilde{\theta}}\frac{2}{\sqrt{2k-a}} + \frac{2}{\sqrt{k}}\right) \\ &\quad \times \|\beta_{-\max(k)}\|_1.\end{aligned}$$

Finally, it follows from Lemma 5.4 that

$$\begin{aligned}\tilde{\theta} = \theta_{a,2k-a} &\le \sqrt{\frac{2k-a}{\min\{b,2k-a\}}}\theta_{a,\min\{b,2k-a\}} \\ &\le \max\left\{\sqrt{\frac{2k-a}{b}}, 1\right\}\theta_{a,b} \\ &= \sqrt{\frac{a}{2k-a}}C_{a,b,k}\theta_{a,b}.\end{aligned}$$

So

$$\begin{aligned}\|h\|_2 \le \;& \frac{\sqrt{2(1+\delta)k/a}(\varepsilon+\eta)}{1-\delta-C_{a,b,k}\theta} \\ &+ 2\|\beta_{-\max(k)}\|_1 \\ &\times \left(\frac{\sqrt{2k}C_{a,b,k}\theta}{(1-\delta-C_{a,b,k}\theta)(2k-a)} + \frac{1}{\sqrt{k}}\right),\end{aligned}$$

which finishes the proof of Theorem 2.2.

The proof of Theorem 2.7 is basically the same, where we only need to use the inequalities

$$\begin{aligned}\|A^TAh\|_\infty &\le \|A^T(A\beta - y)\|_\infty + \|A^T(y - A\hat{\beta})\|_\infty \\ &\le (\varepsilon + \eta)\end{aligned}$$

and

$$\begin{aligned}|\langle Ah, Ah_{\max(a)}\rangle| &= |h_{\max(a)}^T A^T Ah| \\ &\le \|h_{\max(a)}\|_1 \|A^TAh\|_\infty \\ &\le \sqrt{a}\|h_{\max(a)}\|_2(\varepsilon+\eta)\end{aligned}$$

instead of (39) and (41). □

Proof of Theorem 2.4 and 2.8: Again, it suffices to prove Theorem 2.8. We first prove the signal case. Set

$$h_1 = \operatorname{diag}\left(\overbrace{\frac{1}{\sqrt{2k}}, \dots, \frac{1}{\sqrt{2k}}}^{2k}, 0, \dots, 0\right) \in \mathbb{R}^p.$$

Since $\|h_1\|_2 = 1$, we can extend h_1 into an orthonormal basis $\{h_1, \dots, h_p\}$ of $\mathbb{R}^p$. Define the linear map $A : \mathbb{R}^p \to \mathbb{R}^p$ by

$$Ax = \sqrt{\frac{2}{2-a/(2k)}}\sum_{i=2}^{p} c_i h_i \tag{43}$$

for all $x = \sum_{i=1}^{p} c_i h_i$. The Cauchy-Schwarz Inequality yields that for all a-sparse vector x

$$|\langle x, h_1\rangle| \le \|h_1 \cdot 1_{\mathrm{supp}(x)}\|_2 \|x\|_2 \le \sqrt{\frac{a}{2k}}\|x\|_2.$$

Note that

$$\begin{aligned}\|Ax\|_2^2 = \sum_{i=2}^{p} c_i^2 &= \frac{2}{2-a/(2k)}\left(\|x\|_2^2 - c_1^2\right) \\ &= \frac{2}{2-a/(2k)}\left(\|x\|_2^2 - |\langle x, h_1\rangle|^2\right).\end{aligned}$$

So

$$\begin{aligned}\left(1 - \frac{a/(2k)}{2-a/(2k)}\right)\|x\|_2^2 &\le \|Ax\|_2^2 \\ &\le \left(1 + \frac{a/(2k)}{2-a/(2k)}\right)\|x\|_2^2\end{aligned}$$

which implies $\delta_a^A \le \frac{a/(2k)}{2-a/(2k)}$.

Now we estimate $\theta_{a,b}^A$. For any a-sparse vector x_1 and b-sparse vector $x_2 \in \mathbb{R}^p$ with disjoint supports, write $x_1 = \sum_{i=1}^{p} c_i h_i$ and $x_2 = \sum_{i=1}^{p} d_i h_i$, we have $\frac{a/(2k)}{2-a/(2k)}\sum_{i=1}^{p} c_i d_i = \langle x_1, x_2\rangle = 0$.

1) When $b \le 2k - a$, The Cauchy-Schwarz Inequality yields that

$$\begin{aligned}|c_1| = |\langle h_1, x_1\rangle| &\le \sqrt{\frac{a}{2k}}\|x_1\|_2 \\ \text{and}\quad |d_1| = |\langle h_1, x_2\rangle| &\le \sqrt{\frac{b}{2k}}\|x_1\|_2.\end{aligned}$$

So

$$\begin{aligned}&\frac{2-a/(2k)}{2}|\langle Ax_1, Ax_2\rangle| \\ &\quad = \left|\sum_{i=2}^{p} c_i d_i\right| = |-c_1 d_1| \le \frac{\sqrt{ab}}{2k}\|x_1\|_2\|x_2\|_2\end{aligned}$$

and consequently $\theta_{a,b} \le \frac{2}{2-a/(2k)} \cdot \frac{\sqrt{ab}}{2k}$. Hence

$$\begin{aligned}\delta_a^A + C_{a,b,k}\theta_{a,b}^A \le \frac{a/(2k)}{2-a/(2k)} + \max\left\{\frac{2k-a}{\sqrt{ab}}, \sqrt{\frac{2k-a}{a}}\right\} \\ \times \frac{2}{2-a/(2k)}\frac{\sqrt{ab}}{2k} \le 1.\end{aligned}$$

2) When $b > 2k - a$, if $x_1 = 0$ or $x_2 = 0$, it is clear that $\langle Ax_1, Ax_2\rangle = 0 \le C\|x_1\|_2\|x_2\|_2$ for any $C \ge 0$. Without loss of generality, we assume that x_1 and x_2 are non-zero and are normalized so that $\|x_1\|_2 = \|x_2\|_2 = 1$. Since x_1 and x_2 are a, b-sparse respectively and x_1 and x_2 have disjoint supports, it follows from the Cauchy-Schwarz Inequality that for all $\lambda \ge 0$,

$$|c_1| = |\langle h_1, x_1\rangle| \le \sqrt{\frac{a}{2k}}\|x_1\|_2 = \sqrt{\frac{a}{2k}}$$

and

$$\begin{aligned}&\left|d_1 \pm \sqrt{\frac{a}{2k-a}}c_1\right|\\ &\quad= \left|\left\langle h_1, x_2 \pm \sqrt{\frac{a}{2k-a}}x_1\right\rangle\right| \le \left\|x_2 \pm \sqrt{\frac{a}{2k-a}}x_1\right\|_2\\ &\quad= \sqrt{\|x_1\|_2^2 + \frac{a}{2k-a}\|x_1\|_2^2} = \sqrt{\frac{2k}{2k-a}}.\end{aligned}$$

Hence,

$$\begin{aligned}&\frac{2-a/(2k)}{2}|\langle Ax_1, Ax_2\rangle|\\ &\quad= \left|\sum_{i=2}^{mn} c_i d_i\right| = |-c_1 d_1|\\ &\quad= \left(\max\left\{\left|d_1 + \sqrt{\frac{a}{2k-a}}c_1\right|, \left|d_1 - \sqrt{\frac{a}{2k-a}}c_1\right|\right\}\right.\\ &\qquad\left. - \left|\sqrt{\frac{a}{2k-a}}c_1\right|\right)\cdot|c_1|\\ &\quad\le |c_1|\cdot\left(\sqrt{\frac{2k}{2k-a}} - \sqrt{\frac{a}{2k-a}}|c_1|\right)\\ &\quad= -\sqrt{\frac{a}{2k-a}}\left(\sqrt{\frac{k}{2a}} - |c_1|\right)^2 + \frac{k}{2\sqrt{a(2k-a)}}\\ &\quad\le \frac{\sqrt{a(2k-a)}}{2k}\end{aligned}$$

where the last inequality is due to the facts that $|c_1| \le \sqrt{a/(2k)}$ and $a \le k$. So

$$\theta_{a,b}^A \le \frac{2}{2-a/(2k)}\cdot\frac{\sqrt{a(2k-a)}}{2k}$$

and

$$\begin{aligned}\delta_a^A + C_{a,b,k}\theta_{a,b}^A \le& \frac{a/(2k)}{2-a/(2k)} + \max\left\{\frac{2k-a}{\sqrt{ab}}, \sqrt{\frac{2k-a}{a}}\right\}\\ &\cdot\frac{2}{2-a/(2k)}\frac{\sqrt{a(2k-a)}}{2k} \le 1.\end{aligned}$$

To sum up, we have shown $\delta_a^A + C_{a,b,k}\theta_{a,b}^A \le 1$. Furthermore, let

$$u = (\overbrace{1,\ldots,1}^{k}, 0, \ldots) \quad\text{and}$$
$$v = (\overbrace{0,\ldots,0}^{k}, \overbrace{-1,\ldots,-1}^{k}, 0, \ldots),$$

so u and v are both k-sparse and $Au = Av$, since $A(u-v) = 0$. Suppose $y = Ax_1 = Ax_2$, then the k-sparse signals u and v are not distinguishable based on (y, A). Finally, $\delta_a^A + C_{a,b,k}\theta_{a,b}^A < 1$ is impossible by Theorem 2.5, we must have $\delta_a^A + C_{a,b,k}\theta_{a,b}^A = 1$. This finishes the proof for the signal case.

For the matrix case, the proof is essentially the same as the signal case. First we present the following lemma which can be regarded as an extension of the Cauchy-Schwarz Inequality $\langle B, X\rangle \le \|B\|_F\|X\|_F$ with a constraint on $\mathrm{rank}(B)$.

Lemma 5.6: Let $X \in \mathbb{R}^{m\times n}(m \le n)$ be a matrix with singular values $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m$, then for all $B \in \mathbb{R}^{m\times n}$ with rank at most r,

$$|\langle B, X\rangle| \le \|B\|_F\sqrt{\sum_{i=1}^{r}\lambda_i^2}.$$

Then the matrix case can be proved by replacing the notations of vectors in the above proof by matrices and by using Lemma 5.6 instead of the Cauchy-Schwarz's Inequality in the proof of the signal case. □

Proof of Lemma 3.1: For k-sparse vectors $\beta, \gamma \in \mathbb{R}^p$ with disjoint supports, we can write them as

$$\beta = \sum_{i\in T_1} a_i e_i, \quad \gamma = \sum_{i\in T_2} b_i e_i$$

where $a_i > 0, b_i > 0$, T_1 is the support of β, T_2 is the support of γ, and e_i is the vector with ith entry equals to ± 1 and all others entries equal to zero.

Correspondingly, suppose $X, Y \in \mathbb{R}^{m\times n}$ with rank at most r, which satisfies $X^TY = XY^T = 0$. Lemma 5.2 shows that they have singular value decompositions

$$X = \sum_{i\in T_1} a_i u_i v_i^T \quad\text{and}\quad Y = \sum_{i\in T_2} b_i u_i v_i^T,$$

where the disjoint subsets T_1 and T_2 satisfy $|T_1|, |T_2| \le r$. We now consider the even and odd cases separately.

Case 1. $k, r \ge 2$ Is Even: We only consider the matrix case. The proof of the signal case is similar. Without loss of generality, we assume that X and Y are normalized so $\|X\|_F = \|Y\|_F = 1$. Divide T_1 and T_2 into two parts, $T_1 = T_{11} \cup T_{12}$, $T_2 = T_{21} \cup T_{22}$, such that $T_{11}, T_{12}, T_{21}, T_{22}$ are disjoint and $|T_{ij}| \le r/2$ for $i, j \in \{1, 2\}$. Denote

$$X_i = \sum_{i\in T_{1i}} a_i u_i v_i^T \quad\text{and}\quad Y_i = \sum_{i\in T_{2i}} b_i u_i v_i^T, \qquad i = 1, 2.$$

Then

$$\begin{aligned}&|\langle \mathcal{M}(X), \mathcal{M}(Y)\rangle|\\ &\quad\le \sum_{i,j=1}^{2}|\langle \mathcal{M}(X_i), \mathcal{M}(Y_j)\rangle|\\ &\quad= \frac{1}{4}\sum_{i,j=1}^{2}\left|\|\mathcal{M}(X_i+Y_j)\|_F^2 - \|\mathcal{M}(X_i-Y_j)\|_F^2\right|\\ &\quad\le \frac{1}{4}\sum_{i,j=1}^{2}\left[\left(1+\delta_r^{\mathcal{M}}\right)\sum_{i\in T_{ij}\cup T_{ij}} a_i^2 - \left(1-\delta_r^{\mathcal{M}}\right)\sum_{i\in T_{ij}\cup T_{ij}} a_i^2\right]\\ &\quad= \delta_r^{\mathcal{M}}\left(\|X\|_F^2 + \|Y\|_F^2\right) = 2\delta_r^{\mathcal{M}}\end{aligned}$$

and consequently $\theta_{r,r}^{\mathcal{M}} \le 2\delta_r^{\mathcal{M}}$. Now in the example provided in the proof of Theorem 2.4, if $a = b = k$, we have $\delta_r^A = 1/3$,

$\theta_{r,r}^{\mathcal{M}} = 2/3$, which means the coefficient "2" in the inequalities of the even case in (24) cannot be improved.

Case 2. $k, r \geq 3$ Is Odd: For the proof of (23) and (24), we only show the matrix case as the signal case is similar. Since we can set $a_i = 0$ or $b_i = 0$ for $i \notin T_1$ or $i \notin T_2$, Without loss of generality, we assume that $|T_1| = r, |T_2| = r$, a_i, b_i might be 0 for $i \in T_1 \cup T_2$. Also without loss of generality, we can assume X and Y are normalized so

$$\|X\|_F^2 = \sum_{i \in T_1} a_i^2 = \sqrt{\frac{r-1}{r+1}} \quad \text{and}$$
$$\|Y\|_F^2 = \sum_{i \in T_2} b_i^2 = \sqrt{\frac{r+1}{r-1}}.$$

Then we have

$$\begin{aligned}
&\left|4\binom{r-1}{(r-1)/2}\binom{r-1}{(r-3)/2}\langle \mathcal{M}(X), \mathcal{M}(Y)\rangle\right| \\
&= \left|4\binom{r-1}{(r-1)/2}\binom{r-1}{(r-3)/2} \cdot \left\langle \mathcal{M}\left(\sum_{i\in T_1} a_i u_i v_i^T\right), \mathcal{M}\left(\sum_{i\in T_2} b_i u_i v_i^T\right)\right\rangle\right| \\
&= \left|\sum_{\substack{A\subseteq T_1, |A|=(r+1)/2,\\ B\subseteq T_2, |B|=(r-1)/2}} \left[\left\|\mathcal{M}\left(\sum_{i\in A} a_i u_i v_i^T + \sum_{i\in B} b_i u_i v_i^T\right)\right\|^2 - \left\|\mathcal{M}\left(\sum_{i\in A} a_i u_i v_i^T - \sum_{i\in B} b_i u_i v_i^T\right)\right\|^2\right]\right| \\
&\leq \sum_{\substack{A\subseteq T_1, |A|=(r+1)/2,\\ B\subseteq T_2, |B|=(r-1)/2}} 2\delta_r^{\mathcal{M}}\left[\sum_{i\in A} a_i^2 + \sum_{i\in B} b_i^2\right] \\
&= 2\delta_r^{\mathcal{M}}\left[\binom{r-1}{(r-1)/2}\binom{r}{(r-1)/2}\sum_{i\in T_1} a_i^2 + \binom{r-1}{(r-3)/2}\binom{r}{(r+1)/2}\sum_{i\in T_2} b_i^2\right] \\
&= 2\delta_r^{\mathcal{M}}\binom{r-1}{(r-1)/2}\binom{r-1}{(r-3)/2} \cdot \left[\frac{r}{(r-1)/2}\sum_{i\in T_1} a_i^2 + \frac{r}{(r+1)/2}\sum_{i\in T_2} b_i^2\right] \\
&= 8\delta_r^{\mathcal{M}}\binom{r-1}{(r-1)/2}\binom{r-1}{(r-3)/2}\frac{r}{\sqrt{r^2-1}} \\
&= 4\binom{r-1}{(r-1)/2}\binom{r-1}{(r-3)/2}\frac{2r}{\sqrt{r^2-1}}\delta_r^{\mathcal{M}}\|X\|_F\|Y\|_F
\end{aligned}$$

which implies

$$\theta_{r,r}^{\mathcal{M}} \leq \frac{2r}{\sqrt{r^2-1}}\delta_r^{\mathcal{M}}.$$

Next we will construct an example for the signal recovery in the odd case where $\theta_{k,k}^A = \frac{2k}{\sqrt{k^2-1}}\delta_k^A \neq 0$. Suppose $k \geq 3$ is odd and $2k \leq p$, denote

$$\beta_1 = \frac{1}{\sqrt{2k}}(\overbrace{1, 1, \ldots, 1}^{2k}, 0, \ldots) \in \mathbb{R}^p$$

and

$$\beta_2 = \frac{1}{\sqrt{2k}}(\overbrace{1, 1, \ldots, 1}^{k}, \overbrace{-1, \ldots, -1}^{k}, 0, \ldots) \in \mathbb{R}^p. \quad (44)$$

Similarly as in the proof of Theorem 2.4, we can extend β_1 and β_2 to an orthonormal basis of $\mathbb{R}^p$ as $\{\beta_1, \beta_2, \ldots, \beta_p\}$. Then for $0 < \lambda < 1$, we define $A : \mathbb{R}^p \to \mathbb{R}^p$ by

$$A\beta = \sqrt{1+\lambda}a_1\beta_1 + \sqrt{1-\lambda}a_2\beta_2 + \sum_{i=3}^p a_i\beta_i$$

for $\beta = \sum_{i=1}^p a_i\beta_i$. Then it is clear that for all $\beta \in \mathbb{R}^p$,

$$(1-\lambda)\|\beta\|_2^2 \leq \|A\beta\|_2^2 \leq (1+\lambda)\|\beta\|_2^2.$$

Let β and γ be k-sparse vectors with disjoint supports and $\|\beta\|_2 = \|\gamma\|_2 = 1$. Then

$$\begin{aligned}
&|\langle A\beta, A\gamma\rangle| \\
&= \frac{1}{4}\left|\|A(\beta+\gamma)\|_2^2 - \|A(\beta-\gamma)\|_2^2\right| \\
&\leq \max\left\{\frac{1+\lambda}{4}\|\beta+\gamma\|_2^2 - \frac{1-\lambda}{4}\|\beta-\gamma\|_2^2, \frac{1+\lambda}{4}\|\beta-\gamma\|_2^2 - \frac{1-\lambda}{4}\|\beta+\gamma\|_2^2\right\} \\
&= \frac{2\lambda}{4}\left(\|\beta\|_2^2 + \|\gamma\|_2^2\right) = \lambda\|\beta\|_2\|\gamma\|_2
\end{aligned}$$

which implies $\theta_{k,k}^A \leq \lambda$. It can be easily verified that when

$$\beta = (\overbrace{1, 1 \ldots, 1}^{k}, 0, \ldots),$$
$$\gamma = (\overbrace{0, 0, \ldots, 0}^{k}, \overbrace{1, 1, \ldots, 1}^{k}, 0, \ldots),$$

we have $|\langle A\beta, A\gamma\rangle| = \lambda\|\beta\|_2\|\gamma\|_2$. These together imply $\theta_{k,k}^A = \lambda$.

Denote $\beta(i)$ as the ith entry of β. Now let us estimate δ_k^A. For all k-sparse $\beta \in \mathbb{R}^p$, suppose $\beta = \sum_{i=1}^p c_i\beta_i$, then

$$\begin{aligned}
&\|A\beta\|_2^2 \\
&= (1+\lambda)|\langle\beta, \beta_1\rangle|^2 + (1-\lambda)|\langle\beta, \beta_2\rangle|^2 c + \sum_{i=3}^p |\langle\beta, \beta_i\rangle|^2 \\
&= \|\beta\|_2^2 + \lambda\left(|\langle\beta, \beta_1\rangle|^2 - |\langle\beta, \beta_2\rangle|^2\right) \\
&= \|\beta\|_2^2 + \lambda\left(\left(\sum_{i=1}^{2k}\beta(i)\right)^2 - \left(\sum_{i=1}^{k}\beta(i) - \sum_{i=k+1}^{2k}\beta(i)\right)^2\right)\Bigg/ 2k \\
&= \|\beta\|_2^2 + \frac{4}{2k}\lambda\left(\sum_{i=1}^{k}\beta(i)\right)\left(\sum_{i=j+1}^{2k}\beta(i)\right).
\end{aligned}$$

Suppose $T_1 = \operatorname{supp}(\beta) \cap \{1,\ldots,k\}$ and $T_2 = \operatorname{supp}(\beta) \cap \{k+1,\ldots,2k\}$, then $|T_1| + |T_2| \leq k$ and

$$\begin{aligned}
&\left|\left(\sum_{i=1}^{k}\beta(i)\right)\left(\sum_{i=k+1}^{2k}\beta(i)\right)\right| \\
&= \left|\left(\sum_{i\in T_1}\beta(i)\right)\left(\sum_{i\in T_2}\beta(i)\right)\right| \\
&\leq \sqrt{|T_1|\sum_{i\in T_1}\beta(i)^2\cdot|T_2|\sum_{i\in T_2}\beta(i)^2} \\
&\leq \frac{\sqrt{|T_1|\cdot|T_2|}}{2}\sum_{i\in T_1\cup T_2}\beta(i)^2 \\
&\leq \frac{\sqrt{|T_1|(k-|T_1|)}}{2}\|\beta\|_2^2 \\
&\leq \frac{\sqrt{\frac{k-1}{2}\frac{k+1}{2}}}{2}\|\beta\|_2^2,
\end{aligned}$$

where the last inequality is due to the facts that $|T_1|$ is a nonnegative integer and k is odd. It then follows that for all k-sparse vector $\beta \in \mathbb{R}^p$,

$$\left(1-\frac{\sqrt{k^2-1}}{2k}\lambda\right)\|\beta\|_2^2 \leq \|A\beta\|_2^2 \leq \left(1+\frac{\sqrt{k^2-1}}{2k}\lambda\right)\|\beta\|_2^2.$$

It can also be easily verified that the equality above can be achieved for

$$\beta = (\overbrace{1,\ldots,1}^{(k+1)/2},\overbrace{0,\ldots,0}^{(k-1)/2},\overbrace{1,\ldots,1}^{(k-1)/2},0,\ldots)$$

Hence $\delta_k^A = \lambda\frac{\sqrt{k^2-1}}{2k}$. In summary, $\theta_{k,k}^A = \frac{2k}{\sqrt{k^2-1}}\delta_k^A$ in our setting, which implies that the constant $\frac{2k}{\sqrt{k^2-1}}$ in (23) is not improvable. □

References

[1] M. Akcakaya and V. Tarokh, "A frame construction and a universal distortion bound for sparse representations," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2443–2450, 2008.

[2] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. 14th Workshop on Statist. Signal Process.*, 2007, pp. 294–298.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, pp. 253–263, 2008.

[4] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, 2007, pp. 128–133.

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

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

[7] T. Cai, L. Wang, and G. Xu, "Stable recovery of sparse signals and an oracle inequality," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3516–3522, 2010.

[8] T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4388–4394, 2010.

[9] T. Cai and A. Zhang, "Sharp RIP bound for sparse signal and low-rank matrix recovery," *Appl. Comput. Harmon. Anal.*, vol. 35, pp. 74–93, 2013.

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

[11] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l' Academie des Sci.*, vol. 346, pp. 589–592, 2008, Serie I.

[12] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, pp. 717–772, 2009.

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

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

[15] E. Candès and Y. Plan, "Tight oracle for low-rank matrix recovery from a minimal number of random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, 2011.

[16] M. Davenport, J. Laska, J. Treichler, and R. Baraniuk, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4628–4642, 2012.

[17] T. Do, L. Gan, N. Nguyen, and T. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 139–154, 2012.

[18] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

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

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

[21] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, 2009.

[22] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

[23] K. Mohan and M. Fazel, "New restricted isometry results for noisy low-rank recovery," in *Proc. Intl. Symp. Inf. Theory (ISIT)*, 2010, pp. 1573–1577.

[24] S. Oymak and B. Hassibi, "New null space results and recovery thresholds for matrix rank minimization," 2010 [Online]. Available: http://arxiv.org/pdf/1011.6326.pdf, arXiv.

[25] S. Oymak, K. Mohan, M. Fazel, and B. Hassibi, "A simplified approach to recovery conditions for low-rank matrices," in *Proc. Intl. Symp. Inf. Theory (ISIT)*, 2011, pp. 2318–2322.

[26] Q. Mo and S. Li, "New bounds on the restricted isometry constant δ_{2k}," *Appl. Comput. Harmon. Anal.*, vol. 31, pp. 460–468, 2011.

[27] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, pp. 471–501, 2010.

[28] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing—Probabilistic analysis of a null-space characterization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2008, pp. 3377–3380.

[29] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse, bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[30] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, "An architecture for compressive imaging," in *Proc. Int. Conf. Image Process. (ICIP)*, 2006, pp. 1273–1276.

[31] H. Wang and S. Li, "The bounds of restricted isometry constants for low rank matrices recovery," *Sci. China, Ser. A*, 2013, to be published.

[32] J. Zhang, D. Zhu, and G. Zhang, "Adaptive compressed sensing radar oriented toward cognitive detection in dynamic sparse target scene," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1718–1729, 2012.

T. Tony Cai received the Ph.D. degree from Cornell University, Ithaca, NY, in 1996. His research interests include high-dimensional inference, large-scale multiple testing, nonparametric function estimation, functional data analysis and statistical decision theory. He is the Dorothy Silberberg Professor of Statistics at the Wharton School of the University of Pennsylvania, Philadelphia, PA, USA. Dr. Cai is the recipient of the 2008 COPSS Presidents' Award and a fellow of the Institute of Mathematical Statistics. He is the past editor of the *Annals of Statistics*.

Anru Zhang received the B.S. degree from Peking University, Beijing, China, in 2010. He is currently a Ph.D. student at the University of Pennsylvania, Philadelphia, PA. His research interests include high-dimensional inference, nonparametric function estimation and large-scale multiple testing.