Sparse Representation of a Polytope and Recovery of Sparse Signals and Low-Rank Matrices

T. Tony Cai and Anru Zhang

Abstract—This paper considers compressed sensing and affine rank minimization in both noiseless and noisy cases and establishes sharp restricted isometry conditions for sparse signal and low-rank matrix recovery. The analysis relies on a key technical tool, which represents points in a polytope by convex combinations of sparse vectors. The technique is elementary while yielding sharp results. It is shown that for any given constant $t \geq 4/3$, in compressed sensing, $\delta_{tk}^A < \sqrt{(t-1)/t}$ guarantees the exact recovery of all k sparse signals in the noiseless case through the constrained ℓ_1 minimization, and similarly, in affine rank minimization, $\delta_{tr}^{\mathcal{M}} < \sqrt{(t-1)/t}$ ensures the exact reconstruction of all matrices with rank at most r in the noiseless case via the constrained nuclear norm minimization. In addition, for any $\epsilon > 0$, $\delta_{tk}^A < \sqrt{\frac{t-1}{t}} + \epsilon$ is not sufficient to guarantee the exact recovery of all k-sparse signals for large k. Similar results also hold for matrix recovery. In addition, the conditions $\delta_{tk}^A < \sqrt{(t-1)/t}$ and $\delta_{tr}^{\mathcal{M}} < \sqrt{(t-1)/t}$ are also shown to be sufficient, respectively, for stable recovery of approximately sparse signals and low-rank matrices in the noisy case.

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

I. INTRODUCTION

EFFICIENT recovery of sparse signals and low-rank matrices has been a very active area of recent research in applied mathematics, statistics, and machine learning, with many important applications, ranging from signal processing [16], [28] to medical imaging [22] to radar systems [3], [21]. A central goal is to develop fast algorithms that can recover sparse signals and low-rank matrices from a relatively small number of linear measurements. Constrained ℓ_1-norm minimization and nuclear norm minimization are among the most well-known algorithms for the recovery of sparse signals and low-rank matrices respectively.

In compressed sensing, one observes

$$y = A\beta + z, \qquad (1)$$

where $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times p}$ with $n \ll p$, $\beta \in \mathbb{R}^p$ is an unknown sparse signal, and $z \in \mathbb{R}^n$ is a vector of measurement errors.

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).

The goal is to recover the unknown signal $\beta \in \mathbb{R}^p$ based on the measurement matrix A and the observed signal y. The constrained ℓ_1 minimization method proposed by Candés and Tao [11] estimates the signal β by

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p}\{\|\beta\|_1 : \text{ subject to } A\beta - y \in \mathcal{B}\}, \qquad (2)$$

where $\mathcal{B}$ is a set determined by the noise structure. In particular, $\mathcal{B}$ is taken to be $\{0\}$ in the noiseless case. This constrained ℓ_1 minimization method has now been well studied and it is understood that the procedure provides an efficient method for sparse signal recovery.

A closely related problem to compressed sensing is the affine rank minimization problem (ARMP) (Recht et al. [26]), which aims to recover an unknown low-rank matrix based on its affine transformation. In ARMP, one observes

$$b = \mathcal{M}(X) + z, \qquad (3)$$

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 low-rank matrix of interest, and $z \in \mathbb{R}^q$ is measurement error. The goal is to recover the low-rank matrix X based on the linear map $\mathcal{M}$ and the observation $b \in \mathbb{R}^q$. Constrained nuclear norm minimization [26], which is analogous to ℓ_1 minimization in compressed sensing, estimates X by

$$X_* = \arg\min_{B \in \mathbb{R}^{m \times n}}\{\|B\|_* : \text{ subject to } \mathcal{M}(B) - b \in \mathcal{B}\}, \qquad (4)$$

where $\|B\|_*$ is the nuclear norm of B, which is defined as the sum of all singular values of B.

One of the most widely used frameworks in compressed sensing is the *restrict isometry property* (RIP) introduced in Candés and Tao [11]. A vector $\beta \in \mathbb{R}^p$ is called s-sparse if $|\text{supp}(\beta)| \leq s$, where $\text{supp}(\beta) = \{i : \beta_i \neq 0\}$ is the support of β.

Definition 1.1: Suppose $A \in \mathbb{R}^{n \times p}$ is a measurement matrix and $1 \leq s \leq p$ is an integer. The restricted isometry constant (RIC) of order s is defined as the smallest number δ_s^A such that for all s-sparse vectors $\beta \in \mathbb{R}^p$,

$$(1 - \delta_s^A)\|\beta\|_2^2 \leq \|A\beta\|_2^2 \leq (1 + \delta_s^A)\|\beta\|_2^2. \qquad (5)$$

When s is not an integer, we define δ_s^A as $\delta_{\lceil s \rceil}^A$.

Different conditions on the RIC for sparse signal recovery have been introduced and studied in the literature. For example, sufficient conditions for the exact recovery in the noiseless case include $\delta_{2k} < \sqrt{2} - 1$ in [14], $\delta_{2k} < 0.472$ in [6], $\delta_{2k} < 0.497$ in [23], $\delta_k < 0.307$ in [8], $\delta_k < 1/3$ and $\delta_{2k} \leq 1/2$ in [9]. There are also other sufficient conditions that involve the RIC of different orders, e.g. $\delta_{3k}^A + 3\delta_{4k}^A < 2$ in [12],

$\delta_k^A + \delta_{2k}^A < 1$ in [10], $\delta_{2k}^A < 0.5746$ jointly with $\delta_{8k}^A < 1$, $\delta_{3k}^A < 0.7731$ jointly with $\delta_{16k}^A < 1$ in [30] and $\delta_{2k}^A < 4/\sqrt{41}$ in [1].

Similar to the RIP for the measurement matrix A in compressed sensing given in Definition 1.1, a restricted isometry property for a linear map $\mathcal{M}$ in ARMP can be given. For two matrices X and Y in $\mathbb{R}^{m \times n}$, define their inner product as $\langle X, Y \rangle = \sum_{i,j} X_{ij} Y_{ij}$ and the Frobenius norm as $\|X\|_F = \sqrt{\langle X, X \rangle} = \sqrt{\sum_{i,j} X_{ij}^2}$.

Definition 1.2: Suppose $\mathcal{M} : \mathbb{R}^{n \times m} \to \mathbb{R}^q$ is a linear map and $1 \leq r \leq \min(m, n)$ is an integer. The restricted isometry constant (RIC) of order r for $\mathcal{M}$ is defined as the smallest number $\delta_r^{\mathcal{M}}$ such that for all matrices X with rank at most r,

$$(1 - \delta_r^{\mathcal{M}})\|X\|_F^2 \leq \|\mathcal{M}(X)\|_2^2 \leq (1 + \delta_r^{\mathcal{M}})\|X\|_F^2. \quad (6)$$

When r is not an integer, we define $\delta_r^{\mathcal{M}}$ as $\delta_{\lceil r \rceil}^{\mathcal{M}}$.

As in compressed sensing, there are many sufficient conditions based on the RIC to guarantee the exact recovery of matrices of rank at most r through the constrained nuclear norm minimization (4). These include $\delta_{4r}^{\mathcal{M}} < \sqrt{2} - 1$ [15], $\delta_{5r}^{\mathcal{M}} < 0.607$, $\delta_{4r}^{\mathcal{M}} < 0.558$, and $\delta_{3r}^{\mathcal{M}} < 0.4721$ [24], $\delta_{2r}^{\mathcal{M}} < 0.4931$ [29], $\delta_r^{\mathcal{M}} < 0.307$ [29], $\delta_r^{\mathcal{M}} < 1/3$ [9], and $\delta_{2r}^{\mathcal{M}} < 1/2$ [9].

Among these sufficient RIP conditions, $\delta_k^A < 1/3$ and $\delta_r^{\mathcal{M}} < 1/3$ have been verified in [9] to be sharp for both sparse signal recovery and low-rank matrix recovery problems. Sharp conditions on the higher order RICs are however still unknown. As pointed out by Blanchard and Thompson [4], higher-order RIC conditions can be satisfied by a significantly larger set of Gaussian random matrices in some settings. It is therefore of both theoretical and practical interests to obtain sharp sufficient conditions on the high order RICs.

In this paper, we develop a new elementary technique for the analysis of the constrained ℓ_1-norm minimization and nuclear norm minimization procedures and establish sharp RIP conditions on the high order RICs for sparse signal and low-rank matrix recovery. The analysis is surprisingly simple, while leads to sharp results. The key technical tool we develop states an elementary geometric fact: Any point in a polytope can be represented as a convex combination of sparse vectors. The following lemma may be of independent interest.

Lemma 1.1 (Sparse Representation of a Polytope): For a positive number α and a positive integer s, define the polytope $T(\alpha, s) \subset \mathbb{R}^p$ by

$$T(\alpha, s) = \{v \in \mathbb{R}^p : \|v\|_\infty \leq \alpha, \ \|v\|_1 \leq s\alpha\}.$$

For any $v \in \mathbb{R}^p$, define the set of sparse vectors $U(\alpha, s, v) \subset \mathbb{R}^p$ by

$$U(\alpha, s, v) = \{u \in \mathbb{R}^p : \text{supp}(u) \subseteq \text{supp}(v), \ \|u\|_0 \leq s,$$
$$\|u\|_1 = \|v\|_1, \ \|u\|_\infty \leq \alpha\}. \quad (7)$$

Then $v \in T(\alpha, s)$ if and only if v is in the convex hull of $U(\alpha, s, v)$. In particular, any $v \in T(\alpha, s)$ can be expressed as

$$v = \sum_{i=1}^N \lambda_i u_i, \text{ and } 0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^N \lambda_i = 1,$$
$$\text{and } u_i \in U(\alpha, s, v).$$

Lemma 1.1 shows that any point $v \in \mathbb{R}^p$ with $\|v\|_\infty \leq \alpha$ and $\|v\|_1 \leq s\alpha$ must lie in a convex polytope whose extremal points are s-sparse vectors u with $\|u\|_1 = \|v\|_1$ and $\|u\|_\infty \leq \alpha$, and vice versa. This geometric fact turns out to be a powerful tool in analyzing constrained ℓ_1-norm minimization for compressed sensing and nuclear norm minimization for ARMP, since it represents a non-sparse vector by the sparse ones, which provides a bridge between general vectors and the RIP conditions. A graphical illustration of Lemma 1.1 is given in Figure 1.

Combining the results developed in Sections II and III, we establish the following sharp sufficient RIP conditions for the exact recovery of all k-sparse signals and low-rank matrices in the noiseless case. We focus here on the exact sparse and noiseless case; the general approximately sparse (low-rank) and noisy case is considered in Sections II and III.

Theorem 1.1: Let $y = A\beta$ where $\beta \in \mathbb{R}^p$ is a k-sparse vector. If

$$\delta_{tk}^A < \sqrt{\frac{t-1}{t}} \quad (8)$$

for some $t \geq 4/3$, then the ℓ_1 norm minimizer $\hat{\beta}$ of (2) with $\mathcal{B} = \{0\}$ recovers β exactly.

Similarly, suppose $b = \mathcal{M}(X)$ where the matrix $X \in \mathbb{R}^{m \times n}$ is of rank at most r. If

$$\delta_{tr}^{\mathcal{M}} < \sqrt{\frac{t-1}{t}} \quad (9)$$

for some $t \geq 4/3$, then the nuclear norm minimizer X_* of (4) with $\mathcal{B} = \{0\}$ recovers X exactly.

Moreover, it will be shown that for any $\epsilon > 0$, $\delta_{tk}^A < \sqrt{\frac{t-1}{t}} + \epsilon$ is not sufficient to guarantee the exact recovery of all k-sparse signals for large k. Similar result also holds for matrix recovery. For the more general approximately sparse (low-rank) and noisy cases considered in Sections II and III, it is shown that Conditions (8) and (9) are also sufficient respectively for stable recovery of (approximately) k-sparse signals and (approximately) rank-r matrices in the noisy case. An oracle inequality is also given in the case of compressed sensing with Gaussian noise under the condition $\delta_{tk}^A < \sqrt{(t-1)/t}$ when $t \geq 4/3$.

The rest of the paper is organized as follows. Section II considers sparse signal recovery and Section III focuses on low-rank matrix recovery. Discussions on the case $t < 4/3$ and some related issues are given in Section IV. The proofs of the key technical result Lemma 1.1 and the main theorems are contained in Section V.

II. COMPRESSED SENSING

We consider compressed sensing in this section and establish the sufficient RIP condition $\delta_{tk}^A < \sqrt{(t-1)/t}$ in the noisy case which implies immediately the results in the noiseless case given in Theorem 1.1. For $v \in \mathbb{R}^p$, we denote $v_{\max(k)}$ as v with all but the largest k entries in absolute value set to zero, and $v_{-\max(k)} = v - v_{\max(k)}$.

Let us consider the signal recovery model (1) in the setting where the observations contain noise and the signal is not
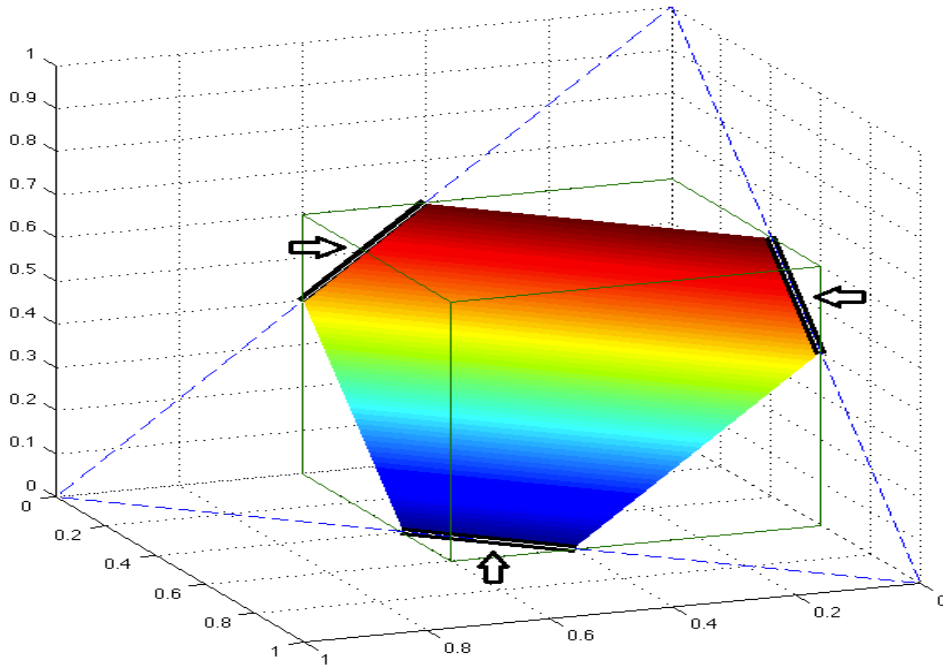
Fig. 1.   A graphical illustration of sparse representation of a polytope in one orthant with $p = 3$ and $s = 2$. All the points in the colored area can be expressed as convex combinations of the sparse vectors represented by the three pointed black line segments on the edges.

exactly k-sparse. This is of significant interest for many applications. Two types of bounded noise settings,

$$z \in \mathcal{B}^{\ell_2}(\varepsilon) \triangleq \{z : \|z\|_2 \leq \varepsilon\}$$
$$\text{and} \quad z \in \mathcal{B}^{DS}(\varepsilon) \triangleq \{z : \|Az\|_\infty \leq \varepsilon\},$$

are of particular interest. The first bounded noise case was considered for example in [18]. The second case is motivated by the *Dantzig Selector* procedure proposed in [13]. Results on the Gaussian noise case, which is commonly studied in statistics, follow immediately. For notational convenience, we write δ for δ_{tk}^A.

Theorem 2.1: Consider the signal recovery model (1) with $\|z\|_2 \leq \varepsilon$. Suppose $\hat{\beta}^{\ell_2}$ is the minimizer of (2) with $\mathcal{B} = \mathcal{B}^{\ell_2}(\eta) = \{z : \|z\|_2 \leq \eta\}$ for some $\eta \geq \varepsilon$. If $\delta = \delta_{tk}^A < \sqrt{(t-1)/t}$ for some $t \geq 4/3$, then

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \leq \frac{\sqrt{2(1+\delta)}}{1 - \sqrt{t/(t-1)}\delta}(\varepsilon + \eta)$$
$$+ \left(\frac{\sqrt{2}\delta + \sqrt{t(\sqrt{(t-1)/t} - \delta)\delta}}{t(\sqrt{(t-1)/t} - \delta)} + 1\right)\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}. \quad (10)$$

Now consider the signal recovery model (1) with $\|A^T z\|_\infty \leq \varepsilon$. Suppose $\hat{\beta}^{DS}$ is the minimizer of (2) with $\mathcal{B} = \mathcal{B}^{DS}(\eta) = \{z : \|A^T z\|_\infty \leq \eta\}$ for some $\eta \geq \varepsilon$. If $\delta = \delta_{tk}^A < \sqrt{(t-1)/t}$ for some $t \geq 4/3$, then

$$\|\hat{\beta}^{DS} - \beta\|_2 \leq \frac{\sqrt{2tk}}{1 - \sqrt{t/(t-1)}\delta}(\varepsilon + \eta)$$
$$+ \left(\frac{\sqrt{2}\delta + \sqrt{t(\sqrt{(t-1)/t} - \delta)\delta}}{t(\sqrt{(t-1)/t} - \delta)} + 1\right)\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}. \quad (11)$$

Remark 2.1: The result for the noiseless case follows directly from Theorem 2.1. When β is exactly k-sparse and there is no noise, by setting $\eta = \epsilon = 0$ and by noting

$\beta_{-\max(k)} = 0$, we have $\hat{\beta} = \beta$ from (10), where $\hat{\beta}$ is the minimizer of (2) with $\mathcal{B} = \{0\}$.

Remark 2.2: It should be noted that Theorems 1.1 and 2.1 also hold for $1 < t < 4/3$ with exactly the same proof. However the bound $\sqrt{(t-1)/t}$ is not sharp for $1 < t < 4/3$. See Section IV for further discussions. The condition $t \geq 4/3$ is crucial for the "sharpness" results given in Theorem 2.2 at the end of this section.

The signal recovery model (1) with Gaussian noise is of particular interest in statistics and signal processing. The following results on the i.i.d. Gaussian noise case are immediate consequences of the above results on the bounded noise cases using the same argument as that in [5], [6], since the Gaussian random variables are essentially bounded.

Proposition 2.1: Suppose the error vector $z \sim N_n(0, \sigma^2 I)$ in (1). $\delta_{tk}^A < \sqrt{(t-1)/t}$ for some $t \geq 4/3$. Let $\hat{\beta}^{\ell_2}$ be the minimizer of (2) with $\mathcal{B} = \{z : \|z\|_2 \leq \sigma\sqrt{n + 2\sqrt{n \log n}}\}$ and let $\hat{\beta}^{DS}$ be the minimizer of (2) with $\mathcal{B} = \{z : \|A^T z\|_\infty \leq 2\sigma\sqrt{\log p}\}$. Then with probability at least $1 - 1/n$,

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

and with probability at least $1 - 1/\sqrt{\pi \log p}$,

$$\|\hat{\beta}^{DS} - \beta\|_2 \leq \frac{4\sqrt{2t}}{1 - \sqrt{t/(t-1)}\delta}\sigma\sqrt{k \log p}$$
$$+ \left(\frac{\sqrt{2}\delta + \sqrt{t(\sqrt{(t-1)/t} - \delta)\delta}}{t(\sqrt{(t-1)/t} - \delta)} + 1\right)\frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}.$$

The oracle inequality approach was introduced by Donoho and Johnstone [20] in the context of wavelet thresholding

for signal denoising. It provides an effective way to study the performance of an estimation procedure by comparing it to that of an ideal estimator. In the context of compressed sensing, oracle inequalities have been given in [7], [9], [13], [15] under various settings. Proposition 2.2 below provides an oracle inequality for compressed sensing with Gaussian noise under the condition $\delta_{tk}^A < \sqrt{(t-1)/t}$ when $t \geq 4/3$.

Proposition 2.2: Given (1), suppose the error vector $z \sim N_n(0, \sigma^2 I)$, β is k-sparse. Let $\hat{\beta}^{DS}$ be the minimizer of (2) with $\mathcal{B} = \{z : \|A^T z\|_\infty \leq 4\sigma\sqrt{\log p}\}$. If $\delta_{tk}^A < \sqrt{(t-1)/t}$ for some $t \geq 4/3$, then with probability at least $1 - 1/\sqrt{\pi \log p}$,

$$\|\hat{\beta}^{DS} - \beta\|_2^2 \leq \frac{256t}{(1 - \sqrt{t/(t-1)}\delta_{tk}^A)^2} \log p \sum_i \min(\beta_i^2, \sigma^2). \tag{12}$$

We now turn to show the sharpness of the condition $\delta_{tk}^A < \sqrt{(t-1)/t}$ for the exact recovery in the noiseless case and stable recovery in the noisy case. It should be noted that the result in the special case $t = 2$ was shown in [17].

Theorem 2.2: Let $t \geq 4/3$. For all $\varepsilon > 0$ and $k \geq 5/\varepsilon$, there exists a matrix A satisfying $\delta_{tk} < \sqrt{\frac{t-1}{t}} + \varepsilon$ and some k-sparse vector β_0 such that

- in the noiseless case, i.e. $y = A\beta_0$, the ℓ_1 minimization method (2) with $\mathcal{B} = \{0\}$ fail to exactly recover the k-sparse vector β_0, i.e. $\hat{\beta} \neq \beta_0$, where $\hat{\beta}$ is the solution to (2).
- in the noisy case, i.e. $y = A\beta_0 + z$, for all constraints $\mathcal{B}_z$ (may depends on z), the ℓ_1 minimization method (2) fails to stably recover the k-sparse vector β_0, i.e. $\hat{\beta} \nrightarrow \beta$ as $z \to 0$, where $\hat{\beta}$ is the solution to (2).

III. AFFINE RANK MINIMIZATION

We consider the affine rank minimization problem (3) in this section. As mentioned in the introduction, this problem is closely related to compressed sensing. The close connections between compressed sensing and ARMP have been studied in Oymak, et al. [25]. We shall present here the analogous results on affine rank minimization without detailed proofs.

For a matrix $X \in \mathbb{R}^{m \times n}$ (without loss of generality, assume that $m \leq n$) 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, we define $X_{\max(r)} = \sum_{i=1}^r a_i u_i v_i^T$ and $X_{-\max(r)} = \sum_{i=r+1}^m a_i u_i v_i^T$. 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 $\mathcal{M} : \mathbb{R}^{m \times n} \to \mathbb{R}^q$, its dual operator is denoted by $\mathcal{M}^* : \mathbb{R}^q \to \mathbb{R}^{m \times n}$.

Similarly as in compressed sensing, we first consider the matrix recovery model (3) in the case where the error vector z is in bounded sets: $\|z\|_2 \leq \epsilon$ and $\|\mathcal{M}^*(z)\| \leq \varepsilon$. The corresponding nuclear norm minimization methods are given

by (4) with $\mathcal{B} = \mathcal{B}^{\ell_2}(\eta)$ and $\mathcal{B} = \mathcal{B}^{DS}(\eta)$ respectively, where

$$\mathcal{B}^{\ell_2}(\eta) = \{z : \|z\|_2 \leq \eta\}, \tag{13}$$
$$\mathcal{B}^{DS}(\eta) = \{z : \|\mathcal{M}^*(z)\| \leq \eta\}. \tag{14}$$

Proposition 3.1: Consider ARMP (3) with $\|z\|_2 \leq \varepsilon$. Let $X_*^{\ell_2}$ be the minimizer of (4) with $\mathcal{B} = \mathcal{B}^{\ell_2}(\eta)$ defined in (13) for some $\eta \geq \epsilon$. If $\delta_r^{\mathcal{M}} < \sqrt{(t-1)/t}$ with $t \geq 4/3$, then

$$\|X_*^{\ell_2} - X\|_F \leq \frac{\sqrt{2(1+\delta)}}{1 - \sqrt{t/(t-1)}\delta}(\varepsilon + \eta)$$
$$+ \left(\frac{\sqrt{2}\delta + \sqrt{t(\sqrt{(t-1)/t} - \delta)}\delta}{t(\sqrt{(t-1)/t} - \delta)} + 1\right)\frac{2\|X_{-\max(r)}\|_1}{\sqrt{r}}. \tag{15}$$

Similarly, consider ARMP (3) with z satisfying $\|\mathcal{M}^*(z)\| \leq \varepsilon$. Let X_*^{DS} be the minimizer of (4) with $\mathcal{M} = \mathcal{B}^{DS}(\eta)$ defined in (14), then

$$\|X_*^{DS} - X\|_F \leq \frac{\sqrt{2tr}}{1 - \sqrt{t/(t-1)}\delta}(\varepsilon + \eta)$$
$$+ \left(\frac{\sqrt{2}\delta + \sqrt{t(\sqrt{(t-1)/t} - \delta)}\delta}{t(\sqrt{(t-1)/t} - \delta)} + 1\right)\frac{2\|X_{-\max(r)}\|_1}{\sqrt{r}}. \tag{16}$$

In the special noiseless case where $z = 0$, it can be seen from either of these two inequalities above that all matrices X with rank at most r can be exactly recovered provided that $\delta_{tr}^{\mathcal{M}} < \sqrt{(t-1)/t}$, for some $t \geq 4/3$.

The following result shows that the condition $\delta_{tr}^{\mathcal{M}} < \sqrt{(t-1)/t}$ with $t \geq 4/3$ is sharp. These results together establish the optimal bound on $\delta_{tr}^{\mathcal{M}}$ ($t \geq 4/3$) for the exact recovery in the noiseless case.

Proposition 3.2: Suppose $t \geq 4/3$. For all $\varepsilon > 0$ and $r \geq 5/\varepsilon$, there exists a linear map $\mathcal{M}$ with $\delta_{tr}^{\mathcal{M}} < \sqrt{(t-1)/t} + \varepsilon$ and some matrix X_0 of rank at most r such that

- in the noiseless case, i.e. $b = \mathcal{M}(X_0)$, the nuclear norm minimization method (4) with $\mathcal{B} = \{0\}$ fails to exactly recover X_0, i.e. $X_* \neq X_0$, where X_* is the solution to (4).
- in the noisy case, i.e. $b = \mathcal{M}(X_0) + z$, for all constraints $\mathcal{B}_z$ (may depends on z), the nuclear norm minimization method (4) fails to stably recover X_0, i.e. $X_* \nrightarrow X_0$ as $z \to 0$, where X_* is the solution to (4) with $\mathcal{B} = \mathcal{B}_z$.

IV. DISCUSSION

We shall focus the discussions in this section exclusively on compressed sensing as the results on affine rank minimization is analogous. In Section II, we have established the sharp RIP condition on the high-order RICs,

$$\delta_{tk}^A < \sqrt{\frac{t-1}{t}} \quad \text{for some } t \geq \frac{4}{3},$$

for the recovery of k-sparse signals in compressed sensing. In addition, it is known from [9] that $\delta_k^A < 1/3$ is also a sharp RIP condition. For a general $t > 0$, denote the sharp bound for δ_{tk}^A as $\delta_*(t)$. Then

$$\delta_*(1) = 1/3 \quad \text{and} \quad \delta_*(t) = \sqrt{(t-1)/t}, \quad t \geq 4/3.$$

A natural question is: What is the value of $\delta_*(t)$ for $t < 4/3$ and $t \neq 1$? That is, what is the sharp bound for δ_{tk}^A when
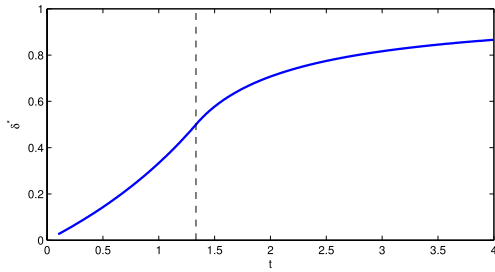
Fig. 2.   Plot of δ_* as a function of t. The dotted line is $t = 4/3$.

Fig. 3.   Plot of n_* as a function of t.

$t < 4/3$ and $t \neq 1$? We have the following partial answer to the question

Proposition 4.1: Let $y = A\beta$ where $\beta \in \mathbb{R}^p$ is k-sparse. Suppose $0 < t < 1$ and $tk \geq 0$ to be an integer

- When tk is even and $\delta_{tk}^A < \frac{t}{4-t}$, the ℓ_1 minimization (2) with $\mathcal{B} = \{0\}$ recovers β exactly.
- When tk is odd and $\delta_{tk}^A < \frac{\sqrt{t^2 - 1/k^2}}{4 - 2t + \sqrt{t^2 - 1/k^2}}$, the ℓ_1 minimization (2) with $\mathcal{B} = \{0\}$ recovers β exactly.

In addition, the following result shows that $\delta_*(t) \leq \frac{t}{4-t}$ for all $0 < t < 4/3$. In particular, when $t = 1$, the upper bound $t/(4 - t)$ coincides with the true sharp bound $1/3$.

Proposition 4.2: For $0 < t < 4/3$, $\varepsilon > 0$ and any integer $k \geq 1$, $\delta_{tk}^A < \frac{t}{4-t} + \varepsilon$ is not sufficient for the exact recovery. Specifically, there exists a matrix A with $\delta_{tk}^A = \frac{t}{4-t}$ and a k-sparse vector β_0 such that $\hat{\beta} \neq \beta_0$, where $\hat{\beta}$ is the minimizer of (2) with $\mathcal{B} = \{0\}$.

Propositions 4.1 and 4.2 together show that $\delta_*(t) = \frac{t}{4-t}$ when tk is even and $0 < t < 1$. We are not able to provide a complete answer for $\delta_*(t)$ when $0 < t < 4/3$. We conjecture that $\delta_*(t) = \frac{t}{4-t}$ for all $0 < t < 4/3$. The following figure plots $\delta_*(t)$ as a function of t based on this conjecture for the interval $(0, 4/3)$.

Our results show that exact recovery of k-sparse signals in the noiseless case is guaranteed if $\delta_{tk}^A < \sqrt{(t-1)/t}$ for some $t \geq 4/3$. It is then natural to ask the question: Among all these RIP conditions $\delta_{tk}^A < \delta_*(t)$, which one is easiest to be satisfied? There is no general answer to this question as no condition is strictly weaker or stronger than the others. It is however interesting to consider special random measurement matrices $A = (A_{ij})_{n \times p}$ where

$$A_{ij} \sim \mathcal{N}(0, 1/n), \quad A_{ij} \sim \begin{cases} 1/\sqrt{n} & \text{w.p.} 1/2 \\ -1/\sqrt{n} & \text{w.p.} 1/2 \end{cases},$$

$$\text{or } A_{ij} \sim \begin{cases} \sqrt{3/n} & \text{w.p.} 1/6 \\ 0 & \text{w.p.} 1/2 \\ -\sqrt{3/n} & \text{w.p.} 1/6 \end{cases}.$$

Baraniuk et al [2] provides a bound on RICs for a set of random matrices from concentration of measure. For these random measurement matrices, Theorem 5.2 of [2] shows that for positive integer $m < n$ and $0 < \lambda < 1$,

$$P(\delta_m^A < \lambda) \geq 1 - 2 \left(\frac{12ep}{m\lambda} \right)^m \exp\left(-n(\lambda^2/16 - \lambda^3/48) \right). \quad (17)$$
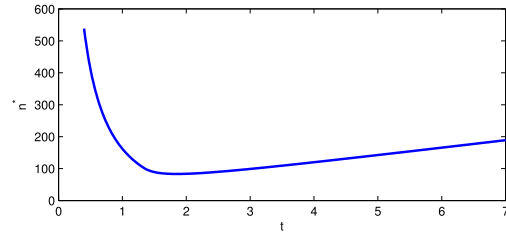
Hence, for $t \geq 4/3$,

$$P(\delta_{tk}^A < \sqrt{(t-1)/t})$$
$$\geq 1 - 2 \exp\left(tk \left(\log(12e/\sqrt{t(t-1)}) + \log(p/k) \right) \right.$$
$$\left. -n \left(\frac{t-1}{16t} - \frac{(t-1)^{3/2}}{48t^{3/2}} \right) \right).$$

For $0 < t < 4/3$, using the conjectured value $\delta_*(t) = \frac{t}{4-t}$, we have

$$P(\delta_{tk}^A < t/(4-t))$$
$$\geq 1 - 2 \exp\left(tk(\log(12(4-t)e/t^2) + \log(p/k)) \right.$$
$$\left. -n \left(\frac{t^2}{16(4-t)^2} - \frac{t^3}{48(4-t)^3} \right) \right).$$

It is easy to see when p, k, and $p/k \to \infty$, the lower bound of n to ensure $\delta_{tk}^A < t/(4-t)$ or $\delta_{tk}^A < \sqrt{(t-1)/t}$ to hold in high probability is $n \geq k \log(p/k)n^*(t)$, where

$$n^* \triangleq \begin{cases} t/\left(\frac{t^2}{16(4-t)^2} - \frac{t^3}{48(4-t)^3} \right) & t < 4/3; \\ t/\left(\frac{t-1}{16t} - \frac{(t-1)^{3/2}}{48t^{3/2}} \right), & t \geq 4/3. \end{cases}$$

For the plot of $n^*(t)$, see Figure 1. $n^*(t)$ has minimum 83.2 when $t = 1.85$. Moreover, among integer t, $t = 2$ can also provide a near-optimal minimum: $n^*(2) = 83.7$.

We should note that the above analysis is based on the bound given in (17) which itself can be possibly improved.

V. PROOFS

We shall first establish the technical result, Lemma 1.1, and then prove the main results.

Proof of Lemma 1.1: First, suppose $v \in T(\alpha, s)$. We can prove v is in the convex hull of $U(\alpha, s, v)$ by induction. If v is s-sparse, v itself is in $U(\alpha, s, v)$.

Suppose the statement is true for all $(l-1)$-sparse vectors v ($l-1 \geq s$). Then for any l-sparse vector v such that $\|v\|_\infty \leq \alpha$, $\|v\|_1 \leq s\alpha$, without loss of generality we assume that v is not $(l-1)$-sparse (otherwise the result holds by assumption of $l-1$). Hence we can express v as $v = \sum_{i=1}^l a_i e_i$, where e_i's are different unit vectors with one entry of ± 1 and other entries of zeros; $a_1 \geq a_2 \geq \cdots \geq a_l > 0$. Since $\sum_{i=1}^l a_i = \|v\|_1 \leq s\alpha$, so

$$1 \in D \triangleq \{1 \leq j \leq l-1 : a_j + a_{j+1} + \cdots + a_l \leq (l - j)\alpha\},$$

which means D is not empty. Take the largest element in D as j, which implies

$$a_j + a_{j+1} + \cdots + a_l \le (l - j)\alpha,$$
$$a_{j+1} + a_{j+2} + \cdots + a_l > (l - j - 1)\alpha. \quad (18)$$

(It is noteworthy that even if the largest j in D is $l - 1$, (18) still holds). Define

$$b_w \triangleq \frac{\sum_{i=j}^{l} a_i}{l - j} - a_w, \quad j \le w \le l, \quad (19)$$

which satisfies $\sum_{i=j}^{l} a_i = (l - j) \sum_{i=j}^{l} b_i$. By (18), for all $j \le w \le l$,

$$b_w \ge b_j = \frac{\sum_{i=j+1}^{l} a_i}{l - j} - \frac{l - j - 1}{l - j} a_j$$
$$\ge \frac{\sum_{i=j+1}^{l} a_i - (l - j - 1)\alpha}{l - j} > 0.$$

In addition, we define

$$v_w \triangleq \sum_{i=1}^{j-1} a_i e_i + (\sum_{i=j}^{l} b_i) \sum_{i=j, i \ne w}^{l} e_i \in R^p,$$
$$\lambda_w \triangleq \frac{b_w}{\sum_{i=j}^{l} b_i}, \quad j \le w \le l, \quad (20)$$

then $0 \le \lambda_w \le 1$, $\sum_{w=j}^{l} \lambda_w = 1$, $\sum_{w=j}^{l} \lambda_w v_w = v$, $\mathrm{supp}(v_w) \subseteq \mathrm{supp}(v)$. We also have

$$\|v_w\|_1 = \sum_{i=1}^{j-1} a_i + (l - j) \sum_{w=j}^{l} b_w = \sum_{i=1}^{j-1} a_i + \sum_{i=j}^{l} a_i = \|v\|_1,$$
$$\|v_w\|_\infty = \max\{a_1, \ldots, a_{j-1}, \sum_{i=j}^{l} b_i\}$$
$$\le \max\{\alpha, \frac{\sum_{i=j}^{l} a_i}{l - j}\} \le \alpha.$$

The last inequality is due to the first part of (18). Finally, note that v_w is $(l-1)$-sparse, we can use the induction assumption to find $\{u_{i,w} \in \mathbb{R}^p, \lambda_{i,w} \in \mathbb{R} : 1 \le i \le N_w, j \le w \le l\}$ such that

$$u_{i,w} \text{ is } s\text{-sparse}, \quad \mathrm{supp}(u_{i,w}) \subseteq \mathrm{supp}(v_i) \subseteq \mathrm{supp}(v),$$
$$\|u_{i,w}\|_1 = \|v_i\|_1 = \|v\|_1, \quad \|u_{i,w}\|_\infty \le \alpha;$$

In addition, $v_i = \sum_{i=1}^{N_w} \lambda_{i,w} u_{i,w}$, so $v = \sum_{w=j}^{l} \sum_{i=1}^{N_w} \lambda_w \lambda_{i,w} u_{i,w}$, which proves the result for l.

The proof of the other part of the lemma is easier. When v is in the convex hull of $U(\alpha, s, v)$, then we have

$$\|v\|_\infty = \|\sum_{i=1}^{N} \lambda_i u_i\|_\infty \le \sum_{i=1}^{N} \lambda_i \|u_i\|_\infty \le \alpha,$$
$$\|v\|_1 = \|\sum_{i=1}^{N} \lambda_i u_i\|_1 \le \sum_{i=1}^{N} \lambda_i \|u_i\|_1$$
$$\le \sum_{i=1}^{N} \lambda_i \|u_i\|_0 \|u_i\|_\infty \le s\alpha,$$

which finished the proof of the lemma. $\square$

Proof of Theorem 1.1: First, we assume that tk is an integer. By the well-known Null Space Property (Theorem 1 in [27]), we only need to check for all $h \in \mathcal{N}(A) \setminus \{0\}$, $\|h_{\max(k)}\|_1 < \|h_{-\max(k)}\|_1$. Suppose there exists $h \in \mathcal{N}(A) \setminus \{0\}$, such that $\|h_{\max(k)}\|_1 \ge \|h_{-\max(k)}\|_1$. Set $\alpha = \|h_{\max(k)}\|_1 / k$. We divide $h_{-\max(k)}$ into two parts, $h_{-\max(k)} = h^{(1)} + h^{(2)}$, where

$$h^{(1)} = h_{-\max(k)} \cdot 1_{\{i \,|\, |h_{-\max(k)}(i)| > \alpha/(t-1)\}},$$
$$h^{(2)} = h_{-\max(k)} \cdot 1_{\{i \,|\, |h_{-\max(k)}(i)| \le \alpha/(t-1)\}}.$$

Then $\|h^{(1)}\|_1 \le \|h_{-\max(k)}\|_1 \le \alpha k$. Denote $|\mathrm{supp}(h^{(1)})| = \|h^{(1)}\|_0 = m$. Since all non-zero entries of $h^{(1)}$ have magnitude larger than $\alpha/(t-1)$, we have

$$\alpha k \ge \|h^{(1)}\|_1$$
$$= \sum_{i \in \mathrm{supp}(h^{(1)})} |h^{(1)}(i)| \ge \sum_{i \in \mathrm{supp}(h^{(1)})} \alpha/(t-1) = m\alpha/(t-1).$$

Namely $m \le k(t - 1)$. In addition we have

$$\|h^{(2)}\|_1 = \|h_{-\max(k)}\|_1 - \|h^{(1)}\|_1$$
$$\le k\alpha - \frac{m\alpha}{t-1} = (k(t-1) - m) \cdot \frac{\alpha}{t-1},$$
$$\|h^{(2)}\|_\infty \le \frac{\alpha}{t-1}. \quad (21)$$

We now apply Lemma 1.1 with $s = k(t - 1) - m$. Then $h^{(2)}$ can be expressed as a convex combination of sparse vectors: $h^{(2)} = \sum_{i=1}^{N} \lambda_i u_i$, where u_i is $(k(t - 1) - m)$-sparse and

$$\|u_i\|_1 = \|h^{(2)}\|_1, \quad \|u_i\|_\infty \le \frac{\alpha}{(t-1)},$$
$$\mathrm{supp}(u_i) \subseteq \mathrm{supp}(h^{(2)}). \quad (22)$$

Hence,

$$\|u_i\|_2 \le \sqrt{\|u_i\|_0} \|u_i\|_\infty \le \sqrt{k(t-1) - m} \|u_i\|_\infty$$
$$\le \sqrt{k(t-1)} \|u_i\|_\infty \le \sqrt{k/(t-1)}\alpha. \quad (23)$$

Now we suppose $\mu \ge 0, c \ge 0$ are to be determined. Denote $\beta_i = h_{\max(k)} + h^{(1)} + \mu u_i$, then

$$\sum_{j=1}^{N} \lambda_j \beta_j - c\beta_i = h_{max(k)} + h^{(1)} + \mu h^{(2)} - c\beta_i$$
$$= (1 - \mu - c)(h_{\max(k)} + h^{(1)}) - c\mu u_i + \mu h. \quad (24)$$

Since $h_{\max(k)}$, $h^{(1)}$, u_i are k-, m-, $(k(t - 1) - m)$-sparse respectively, $\beta_i = h_{\max(k)} + h^{(1)} + \mu u_i$, $\sum_{j=1}^{N} \lambda_j \beta_j - c\beta_i - \mu h = (1 - \mu - c)(h_{\max(k)} + h^{(1)}) - c\mu u_i$ are all tk-sparse vectors.

We can check the following identity in ℓ_2 norm,

$$\sum_{i=1}^{N} \lambda_i \|A(\sum_{j=1}^{N} \lambda_j \beta_j - c\beta_i)\|_2^2$$
$$+ (1 - 2c) \sum_{1 \le i < j \le N} \lambda_i \lambda_j \|A(\beta_i - \beta_j)\|_2^2$$
$$= \sum_{i=1}^{N} \lambda_i (1 - c)^2 \|A\beta_i\|_2^2. \quad (25)$$

Since $Ah = 0$ and (24), we have $A(\sum_{j=1}^{N} \lambda_j \beta_j - c\beta_i) = A((1 - \mu - c)(h_{\max(k)} + h^{(1)}) - c\mu u_i)$. Set $c = 1/2$,

$\mu = \sqrt{t(t-1)} - (t-1)$, let the left hand side of (25) minus the right hand side, we get

$$
\begin{aligned}
0 \leq{} & (1 + \delta_{tk}^A) \\
& \cdot \sum_{i=1}^{N} \lambda_i \Big((1 - \mu - c)^2 \| h_{\max(k)} + h^{(1)} \|_2^2 + c^2 \mu^2 \| u_i \|_2^2 \Big) \\
& - (1 - \delta_{tk}^A) \\
& \cdot \sum_{i=1}^{N} \lambda_i (1 - c)^2 \Big(\| h_{\max(k)} + h^{(1)} \|_2^2 + \mu^2 \| u_i \|_2^2 \Big) \\
={} & \sum_{i=1}^{N} \lambda_i \Big[\Big((1 + \delta_{tk}^A)(\tfrac{1}{2} - \mu)^2 - (1 - \delta_{tk}^A) \cdot \tfrac{1}{4} \Big) \\
& \cdot \| h_{\max(k)} + h^{(1)} \|_2^2 + \tfrac{1}{2} \delta_{tk}^A \mu^2 \| u_i \|_2^2 \Big] \\
\leq{} & \sum_{i=1}^{N} \lambda_i \| h_{\max(k)} + h^{(1)} \|_2^2 \\
& \cdot \Big[(\mu^2 - \mu) + \delta_{tk}^A \Big(\tfrac{1}{2} - \mu + (1 + \tfrac{1}{2(t-1)}) \mu^2 \Big) \Big] \\
={} & \| h_{\max(k)} + h^{(1)} \|_2^2 \cdot \Big[\delta_{tk}^A \Big((2t-1)t - 2t\sqrt{t(t-1)} \Big) \\
& - \Big((2t-1)\sqrt{t(t-1)} - 2t(t-1) \Big) \Big] < 0.
\end{aligned}
$$

We used the fact that

$$
\delta_{tk}^A < \sqrt{(t-1)/t},
$$
$$
\| u_i \|_2 \leq \sqrt{k/(t-1)} \alpha
$$
$$
\leq \frac{\| h_{\max(k)} \|_2}{\sqrt{(t-1)}} \leq \frac{\| h_{\max(k)} + h^{(1)} \|_2}{\sqrt{t-1}}
$$

above. This is a contradiction.

When tk is not an integer, note $t' = \lceil tk \rceil / k$, then $t' > t$, $t'k$ is an integer,

$$
\delta_{t'k} = \delta_{tk} < \sqrt{\frac{t-1}{t}} < \sqrt{\frac{t'k-1}{t'k}},
$$

which can be deduced to the former case. Hence we finished the proof. $\quad\square$

Proof of Theorem 2.1: We first prove the inequality on $\hat{\beta}^{\ell_2}$ (10). Again, we assume that tk is an integer at first. Suppose $h = \hat{\beta}^{\ell_2} - \beta$, we shall use a widely known result (see, e.g., [5], [12], [13], [19]),

$$
\| h_{-\max(k)} \|_1 \leq \| h_{\max(k)} \|_1 + 2 \| \beta_{-\max(k)} \|_1.
$$

Besides,

$$
\| Ah \|_2 \leq \| y - A\beta \|_2 + \| A\hat{\beta}^{\ell_2} - y \|_2 \leq \varepsilon + \eta. \tag{26}
$$

Define $\alpha = (\| h_{\max(k)} \|_1 + 2 \| \beta_{-\max(k)} \|_1) / k$. Similarly as the proof of Theorem 1.1, we divide $h_{-\max(k)}$ into two parts, $h_{-\max(k)} = h^{(1)} + h^{(2)}$, where

$$
h^{(1)} = h_{-\max(k)} \cdot 1_{\{ i \,\mid\, |h_{-\max(k)}(i)| > \alpha/(t-1) \}},
$$
$$
h^{(2)} = h_{-\max(k)} \cdot 1_{\{ i \,\mid\, |h_{-\max(k)}(i)| \leq \alpha/(t-1) \}}.
$$

Then $\| h^{(1)} \|_1 \leq \| h_{-\max(k)} \|_1 \leq \alpha k$. Denote $| \mathrm{supp}(h^{(1)}) | = \| h^{(1)} \|_0 = m$. Since all non-zero entries of $h^{(1)}$ have magnitude larger than $\alpha/(t-1)$, we have

$$
\begin{aligned}
\alpha k \geq \| h^{(1)} \|_1 ={} & \sum_{i \in \mathrm{supp}(h^{(1)})} |h^{(1)}(i)| \\
\geq{} & \sum_{i \in \mathrm{supp}(h^{(1)})} \alpha/(t-1) = m\alpha/(t-1).
\end{aligned}
$$

Namely $m \leq k(t-1)$. Hence, (21) still holds. Besides, $\| h_{\max(k)} + h^{(1)} \|_0 = k + m \leq tk$, we have

$$
\begin{aligned}
\langle A(h_{\max(k)} + h^{(1)}), Ah \rangle \\
\leq \| A(h_{\max(k)} + h^{(1)}) \|_2 \| Ah \|_2 \\
\leq \sqrt{1 + \delta} \| h_{\max(k)} + h^{(1)} \|_2 (\varepsilon + \eta). \tag{27}
\end{aligned}
$$

Again by (21), we apply Lemma 1.1 by setting $s = k(t-1) - m$, we can express $h^{(2)}$ as a weighted mean: $h^{(2)} = \sum_{i=1}^{N} \lambda_i u_i$, where u_i is $(k(t-1) - m)$-sparse and (22) still holds. Hence,

$$
\begin{aligned}
\| u_i \|_2 \leq{} & \sqrt{\| u_i \|_0} \| u_i \|_\infty \\
\leq{} & \sqrt{k(t-1) - m} \| u_i \|_\infty \\
\leq{} & \sqrt{k(t-1)} \| u_i \|_\infty \leq \sqrt{k/(t-1)} \alpha.
\end{aligned}
$$

Now we suppose $1 \geq \mu \geq 0, c \geq 0$ are to be determined. Denote $\beta_i = h_{\max(k)} + h^{(1)} + \mu u_i$, then we still have (24). Similarly to the proof of Theorem 1.1, since $h_{\max(k)}, h^{(1)}, u_i$ are k-, m-, $(k(t-1) - m)$-sparse vectors, respectively, we know $\beta_i = h_{\max(k)} + h^{(1)} + \mu u_i$, $\sum_{j=1}^{N} \lambda_j \beta_j - c\beta_i - \mu h = (1 - \mu - c)(h_{\max(k)} + h^{(1)}) - c\mu u_i$ are all tk sparse vectors.

Suppose $x = \| h_{\max(k)} + h^{(1)} \|_2$, $P = \frac{2 \| \beta_{-\max(k)} \|_1}{\sqrt{k}}$, then

$$
\begin{aligned}
\| u_i \|_2 \leq{} & \sqrt{k/(t-1)} \alpha \\
\leq{} & \frac{\| h_{\max(k)} \|_2}{\sqrt{(t-1)}} + \frac{2 \| \beta_{-\max(k)} \|_1}{\sqrt{k(t-1)}} \\
\leq{} & \frac{\| h_{\max(k)} + h^{(1)} \|_2}{\sqrt{t-1}} + \frac{2 \| \beta_{-\max(k)} \|_1}{\sqrt{k(t-1)}} \\
={} & \frac{x + P}{\sqrt{t-1}}.
\end{aligned}
$$

We still use the ℓ_2 identity (25). Set $c = 1/2$, $\mu = \sqrt{t(t-1)} - (t-1)$ and take the difference of the left- and right-hand sides of (25), we get

$$
\begin{aligned}
0 ={} & \sum_{i=1}^{N} \lambda_i \Big\| A\Big((h_{\max(k)} + h^{(1)} + \mu h^{(2)}) \\
& - \tfrac{1}{2}(h_{\max(k)} + h^{(1)} + \mu u_i) \Big) \Big\|_2^2 - \sum_{i=1}^{N} \frac{\lambda_i}{4} \| A\beta_i \|_2^2 \\
={} & \sum_{i=1}^{N} \lambda_i \Big\| A\Big((\tfrac{1}{2} - \mu)(h_{\max(k)} + h^{(1)}) - \tfrac{\mu}{2} u_i + \mu h \Big) \Big\|_2^2 \\
& - \sum_{i=1}^{N} \frac{\lambda_i}{4} \| A\beta_i \|_2^2
\end{aligned}
$$

$$= \sum_{i=1}^{N} \lambda_i \left\| A \left((\frac{1}{2} - \mu)(h_{\max(k)} + h^{(1)}) - \frac{\mu}{2} u_i \right) \right\|_2^2$$

$$+ 2 \left\langle A \left((\frac{1}{2} - \mu)(h_{\max(k)} + h^{(1)}) - \frac{\mu}{2} h^{(2)} \right), \mu A h \right\rangle$$

$$+ \mu^2 \|Ah\|_2^2 - \sum_{i=1}^{N} \frac{\lambda_i}{4} \|A\beta_i\|_2^2$$

$$= \sum_{i=1}^{N} \lambda_i \left\| A \left((\frac{1}{2} - \mu)(h_{\max(k)} + h^{(1)}) - \frac{\mu}{2} u_i \right) \right\|_2^2$$

$$+ \mu(1 - \mu) \left\langle A(h_{\max(k)} + h^{(1)}), Ah \right\rangle - \sum_{i=1}^{N} \frac{\lambda_i}{4} \|A\beta_i\|_2^2.$$

Now since β_i, $(\frac{1}{2} - \mu)(h_{\max(k)} + h^{(1)}) - \frac{\mu}{2} u_i$ are all tk-sparse vectors, we apply the definition of δ_{tk}^A and also (27) to get

$$0 \le (1 + \delta) \sum_{i=1}^{N} \lambda_i \left((\frac{1}{2} - \mu)^2 \|h_{\max(k)} + h^{(1)}\|_2^2 + \frac{\mu^2}{4} \|u_i\|_2^2 \right)$$

$$+ \mu(1 - \mu)\sqrt{1+\delta} \|h_{\max(k)} + h^{(1)}\|_2 (\varepsilon + \eta)$$

$$- (1 - \delta) \sum_{i=1}^{N} \frac{\lambda_i}{4} \left(\|h_{\max(k)} + h^{(1)}\|_2^2 + \mu^2 \|u_i\|_2^2 \right)$$

$$= \sum_{i=1}^{N} \lambda_i \left\{ \left((1+\delta)(\frac{1}{2} - \mu)^2 - (1 - \delta) \cdot \frac{1}{4} \right) \right.$$

$$\left. \cdot \left\| h_{\max(k)} + h^{(1)} \right\|_2^2 + \frac{1}{2} \delta \mu^2 \|u_i\|_2^2 \right\}$$

$$+ \mu(1 - \mu)\sqrt{1+\delta} \left\| h_{\max(k)} + h^{(1)} \right\|_2 (\varepsilon + \eta)$$

$$\le \left[(\mu^2 - \mu) + \delta \left(\frac{1}{2} - \mu + (1 + \frac{1}{2(t-1)})\mu^2 \right) \right] x^2$$

$$+ \left[\mu(1 - \mu)\sqrt{1+\delta}(\varepsilon + \eta) + \frac{\delta \mu^2 P}{t - 1} \right] x + \frac{\delta \mu^2 P^2}{2(t-1)}$$

$$= -t \left((2t - 1) - 2\sqrt{t(t-1)} \right) \left(\sqrt{\frac{t-1}{t}} - \delta \right) x^2$$

$$+ \left[\mu^2 \sqrt{\frac{t}{t-1}} \cdot \sqrt{1+\delta}(\varepsilon + \eta) + \frac{\delta \mu^2 P}{t - 1} \right] x + \frac{\delta \mu^2 P^2}{2(t-1)}$$

$$= \frac{\mu^2}{t-1} \left[-t \left(\sqrt{\frac{t-1}{t}} - \delta \right) x^2 \right.$$

$$\left. + \left(\sqrt{t(t-1)(1+\delta)}(\varepsilon + \eta) + \delta P \right) x + \frac{\delta P^2}{2} \right], \quad (28)$$

which is an second-order inequality for x. By solving this inequality we get

$$x \le \left\{ \left(\sqrt{t(t-1)(1+\delta)}(\varepsilon + \eta) + \delta P \right) \right.$$

$$+ \left[\left(\sqrt{t(t-1)(1+\delta)}(\varepsilon + \eta) + \delta P \right)^2 \right.$$

$$\left. + 2t(\sqrt{(t-1)/t} - \delta)\delta P^2 \right]^{1/2} \right\} \cdot \left(2t(\sqrt{(t-1)/t} - \delta) \right)^{-1}$$

$$\le \frac{\sqrt{t(t-1)(1+\delta)}}{t(\sqrt{(t-1)/t} - \delta)} (\varepsilon + \eta)$$

$$+ \frac{2\delta + \sqrt{2t(\sqrt{(t-1)/t} - \delta)\delta}}{2t(\sqrt{(t-1)/t} - \delta)} P.$$

Finally, note that $\|h_{-\max(k)}\|_1 \le \|h_{\max(k)}\|_1 + P\sqrt{k}$, by Lemma 5.3 in [9], we obtain $\|h_{-\max(k)}\|_2 \le \|h_{\max(k)}\|_2 + P$, so

$$\|h\|_2 = \sqrt{\|h_{\max(k)}\|_2^2 + \|h_{-\max(k)}\|_2^2}$$

$$\le \sqrt{\|h_{\max(k)}\|_2^2 + (\|h_{\max(k)}\|_2 + P)^2}$$

$$\le \sqrt{2\|h_{\max(k)}\|_2^2} + P$$

$$\le \sqrt{2} x + P$$

$$\le \frac{\sqrt{2t(t-1)(1+\delta)}}{t(\sqrt{(t-1)/t} - \delta)} (\varepsilon + \eta)$$

$$+ \left(\frac{\sqrt{2}\delta + \sqrt{t(\sqrt{(t-1)/t} - \delta)\delta}}{t(\sqrt{(t-1)/t} - \delta)} + 1 \right) \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}}$$

$$= \frac{\sqrt{2(1+\delta)}}{1 - \sqrt{t/(t-1)}\delta} (\varepsilon + \eta)$$

$$+ \left(\frac{\sqrt{2}\delta + \sqrt{t(\sqrt{(t-1)/t} - \delta)\delta}}{t(\sqrt{(t-1)/t} - \delta)} + 1 \right) \frac{2\|\beta_{-\max(k)}\|_1}{\sqrt{k}},$$

which finished the proof.

When tk is not an integer, again we define $t' = \lceil tk \rceil / k$, then $t' > t$ and $\delta_{t'k}^A = \delta_{tk}^A < \sqrt{\frac{t-1}{t}} < \sqrt{\frac{t'-1}{t'}}$. We can prove the result by working on $\delta_{t'k}^A$.

For the inequality on $\hat{\beta}^{DS}$ (11), the proof is similar. Define $h = \hat{\beta}^{DS} - \beta$. We have the following inequalities

$$\|A^T Ah\|_\infty \le \|A^T(A\hat{\beta}^{\ell_2} - y)\|_\infty$$

$$+ \|A^T(y - A\beta)\|_\infty \le \eta + \varepsilon,$$

$$\langle A(h_{\max(k)} + h^{(1)}), Ah \rangle = \langle h_{\max(k)} + h^{(1)}, A^T Ah \rangle$$

$$\le \|h_{\max(k)} + h^{(1)}\|_1 (\varepsilon + \eta)$$

$$\le \sqrt{tk}(\varepsilon + \eta)\|h_{\max(k)+h^{(1)}}\|_2, \quad (29)$$

instead of (26) and (27). We can prove (11) basically the same as the proof above except that we use (29) instead of (27) when we go from the third term to the fourth term in (28). □

Proof of Proposition 2.1: By a small extension of Lemma 5.1 in [5], we have $\|z\|_2 \le \sigma\sqrt{n + 2\sqrt{n \log n}}$ with probability at least $1 - 1/n$; $\|A^T z\|_\infty \le \sigma\sqrt{2(1 + \delta_1^A)\log p} \le 2\sigma\sqrt{\log p}$ with probability at least $1 - 1/\sqrt{\pi \log p}$. Then the Proposition is immediately implied by Theorem 2.1. □

Proof of Proposition 2.2: The proof of Proposition (2.2) is similar to that of Theorem 4.1 in [9] and Theorem 2.7 in [15].

First, as in the proof of Proposition 2.1, we have $\|A^T z\|_\infty \le \lambda/2$ with probability at least $1/\sqrt{\pi \log n}$. In the rest proof, we will prove (12) in the event that $\|A^T z\|_\infty \le \lambda/2$. Define

$$K(\xi, \beta) = \gamma \|\xi\|_0 + \|A\beta - A\xi\|_2^2, \quad \gamma = \frac{\lambda^2}{8} = 2\sigma^2 \log p.$$

Let $\bar{\beta} = \arg\min_\xi K(\xi, \beta)$. Since $K(\bar{\beta}, \beta) \le K(\beta, \beta)$, we have $\gamma \|\bar{\beta}\|_0 \le \gamma \|\beta\|_0$, which means $\bar{\beta}$ is k-sparse.

Now we introduce the following lemma which can be regarded as an extension of Lemma 4.1 in [9].

Lemma 5.1: Suppose $A \in \mathbb{R}^{n \times p}$, $k \ge 2$ is an integer, $s > 1$ is real and sk is integer. Then we have $\delta_{sk}^A \le (2s - 1)\delta_k^A$. Similarly, suppose $\mathcal{M} : \mathbb{R}^{m \times n} \to \mathbb{R}^q$ is a linear map, $r \ge 2$

is an integer, $s > 1$ is real and sr is integer. Then we have $\delta_{sr}^{\mathcal{M}} \leq (2s-1)\delta_r^{\mathcal{M}}$.

We omit the proof here as the proof of Lemma 4.1 in [9] can still apply to this lemma.

By Lemma 5.1, we can see when $1 < t < 2$,

$$\delta_{2k}^A \leq (2\frac{2k}{\lceil tk \rceil} - 1)\delta_{\lceil tk \rceil}^A \leq (4/t - 1)\delta_{tk}^A \leq \sqrt{t/(t-1)}\delta_{tk}^A.$$

When $t \geq 2$, $\delta_{2k}^A \leq \delta_{tk}^A$, which means

$$\delta_{2k}^A \leq \sqrt{t/(t-1)}\delta_{tk}^A, \tag{30}$$

whenever $t \geq 4/3$.

Next, we have

$$\|\bar{\beta} - \beta\|_2^2 \leq \frac{1}{1 - \delta_{2k}^A}\|A\bar{\beta} - A\beta\|_2^2$$
$$\leq \frac{1}{1 - \sqrt{t/(t-1)}\delta_{tk}^A}\|A\bar{\beta} - A\beta\|_2^2.$$

With a small edition on Lemma 5.4 in [9] and Lemma 3.5 in [15], we have

$$\|A^T(y - A\bar{\beta})\|_\infty \leq \|A^T(y - A\beta)\|_\infty + \|A^T A(\beta - \bar{\beta})\|_\infty \leq \lambda.$$

Since $\bar{\beta}$ is k-sparse, we can apply Theorem 2.1 by plugging β by $\bar{\beta}$ and get

$$\|\hat{\beta} - \bar{\beta}\|_2 \leq \frac{\sqrt{2t\|\bar{\beta}\|_0}}{1 - \sqrt{t/(t-1)}\delta_{tk}^A}2\lambda.$$

Hence,

$$\|\hat{\beta} - \beta\|_2^2 \leq 2\|\hat{\beta} - \bar{\beta}\|_2^2 + 2\|\bar{\beta} - \beta\|_2^2$$
$$\leq \frac{16t\|\bar{\beta}\|_0\lambda^2}{(1 - \sqrt{t/(t-1)}\delta_{tk}^A)^2} + \frac{2}{1 - \sqrt{t/(t-1)}\delta_{tk}^A}\|A\bar{\beta} - A\beta\|_2^2$$
$$\leq \frac{128t}{(1 - \sqrt{t/(t-1)}\delta_{tk}^A)^2}K(\bar{\beta}, \beta).$$

Suppose $\beta' = \sum_{i=1}^p \beta \cdot 1_{\{|\beta_i| > \mu\}}$, where $\mu = \sqrt{\frac{\gamma}{1+\delta_k^A}}$. Then

$$K(\bar{\beta}, \beta) \leq K(\beta', \beta) \leq \gamma \sum_{i=1}^p 1_{\{|\beta_i| > \mu\}} + \|A\beta' - A\beta\|_2^2$$
$$\leq \gamma \sum_{i=1}^p 1_{\{|\beta_i| > \mu\}} + (1 + \delta_k^A)\sum_{i=1}^p 1_{\{|\beta_i| \leq \mu\}}|\beta_i|^2$$
$$\leq \sum_{i=1}^p \min(\gamma, (1 + \delta_k^A)|\beta_i|^2)$$
$$\leq 2\log p \sum_{i=1}^p \min(\sigma^2, |\beta_i|^2).$$

Therefore, we have proved (12) in the event that $\|A^T z\|_\infty \leq \lambda/2$. $\square$

Proof of Theorem 2.2: For any $\varepsilon > 0$ and $k \geq 5/\varepsilon$, suppose $p \geq 2tk$, $m' = ((t-1) + \sqrt{t(t-1)})k$, m is the largest integer strictly smaller than m'. Then $m < m'$ and $m' - m \leq 1$. Since $t \geq 4/3$, we have $m' \geq k$. Define

$$\beta_1 = \sqrt{k + \frac{mk^2}{m'^2}}^{-1} (\overbrace{1, \cdots, 1}^{k}, \overbrace{-\frac{k}{m'}, \cdots, -\frac{k}{m'}}^{m}, 0, \cdots, 0) \in \mathbb{R}^p,$$

then $\|\beta_1\|_2 = 1$. We define linear map $A : \mathbb{R}^p \to \mathbb{R}^p$, such that for all $\beta \in \mathbb{R}^p$,

$$A\beta = \sqrt{1 + \sqrt{\frac{t-1}{t}}}(\beta - \langle\beta_1, \beta\rangle\beta_1).$$

Now for all $\lceil tk \rceil$-sparse vector β,

$$\|A\beta\|_2^2 = \left(1 + \sqrt{\frac{t-1}{t}}\right)(\beta - \langle\beta_1, \beta\rangle\beta_1)^T(\beta - \langle\beta_1, \beta\rangle\beta_1)$$
$$= \left(1 + \sqrt{\frac{t-1}{t}}\right)\left(\|\beta\|_2^2 - |\langle\beta_1, \beta\rangle|^2\right).$$

Since β is $\lceil tk \rceil$-sparse, by Cauchy-Schwarz Inequality,

$$0 \leq |\langle\beta_1, \beta\rangle|^2$$
$$\leq \|\beta\|_2^2 \cdot \|\beta_1 \cdot 1_{\text{supp}(\beta)}\|_2^2$$
$$\leq \|\beta\|_2^2\|\beta_{1,\max(\lceil tk \rceil)}\|_2^2 = \|\beta\|_2^2 \cdot \frac{m'^2 + k(\lceil tk \rceil - k)}{m'^2 + mk}$$
$$\leq \frac{m'^2 + k^2(t-1) + k}{m'^2 + m'k} \cdot \frac{1}{1 - \frac{k(m'-m)}{m'^2+m'k}}\|\beta\|_2^2$$
$$= \frac{m'^2 + k^2(t-1)}{m'^2 + m'k} \cdot \frac{m'^2 + k^2(t-1) + k}{m'^2 + k^2(t-1)} \cdot \frac{1}{1 - \frac{k(m'-m)}{m'^2+m'k}}\|\beta\|_2^2$$
$$= 2\sqrt{t-1}(\sqrt{t} - \sqrt{t-1}) \cdot (1 + \frac{1}{tk}) \cdot \frac{1}{1 - \frac{1}{2k}}\|\beta\|_2^2$$
$$\leq \left(2\sqrt{t(t-1)} - 2(t-1)\right) \cdot (1 + \frac{5}{2k})\|\beta\|_2^2$$
$$\leq \left(2\sqrt{t(t-1)} - 2(t-1) + \frac{5}{2k}\right)\|\beta\|_2^2.$$

We used the fact that $m' \geq k$, $0 < m' - m \leq 1$ and

$$\frac{m'^2 + k^2(t-1)}{m'^2 + m'k}$$
$$= \frac{((t-1) + \sqrt{t(t-1)})^2 + t - 1}{((t-1) + \sqrt{t(t-1)})^2 + ((t-1) + \sqrt{t(t-1)})}$$
$$= \frac{(t-1)(t - 1 + t + 2\sqrt{t(t-1)} + 1)}{\sqrt{t(t-1)}(\sqrt{t} + \sqrt{(t-1)})^2}$$
$$= \frac{2\sqrt{t-1}}{\sqrt{t} + \sqrt{t-1}} = 2\sqrt{t-1}\left(\sqrt{t} - \sqrt{t-1}\right)$$

above. Hence,

$$\left(1 + \sqrt{\frac{t-1}{t}}\right)\|\beta\|_2^2 \geq \|A\beta\|_2^2$$
$$\geq \left(1 - \sqrt{\frac{t-1}{t}} - \left(1 + \sqrt{\frac{t-1}{t}}\right)\frac{5}{2k}\right)\|\beta\|_2^2$$
$$\geq \left(1 - \sqrt{\frac{t-1}{t}} - \varepsilon\right)\|\beta\|_2^2,$$

which implies $\delta_{tk}^A \leq \sqrt{(t-1)/t} + \varepsilon$.

Now we consider

$$\beta_0 = (\overbrace{1, \cdots, 1}^{k}, 0, \cdots, 0) \in \mathbb{R}^p,$$

$$\gamma_0 = (\overbrace{0, \cdots, 0}^{k}, \overbrace{\frac{k}{m'}, \cdots, \frac{k}{m'}}^{m}, 0, \cdots, 0).$$

Note that $A\beta_1 = 0$, so $A\beta_0 = A\gamma_0$. Besides, β_0 is k-sparse and $\|\gamma_0\|_1 < \|\beta_0\|_1$.

- In the noiseless case, i.e. $y = A\beta_0$, the ℓ_1 minimization method (2) fails to exactly recover β_0 through y since $y = A\gamma_0$, but $\|\gamma_0\|_1 < \|\beta_0\|_1$.
- In the noisy case, i.e. $y = A\beta_0 + z$, assume that ℓ_1 minimization method (2) can stably recover β_0 with constraint $\mathcal{B}_z$. Suppose $\hat{\beta}_z$ is the solution of ℓ_1 minimization, then $\lim_{z\to 0}\hat{\beta}_z = \beta_0$. Note that $y - A(\hat{\beta}_z - \beta_0 + \gamma_0) = y - A\hat{\beta}_z \in \mathcal{B}_z$, by the definition of $\hat{\beta}_z$, we have $\|\hat{\beta}_z - \beta_0 + \gamma_0\|_1 \geq \|\hat{\beta}_z\|_1$. Let $z \to 0$, it contradicts that $\|\gamma_0\|_1 < \|\beta_0\|_1$. Therefore, ℓ_1 minimization method (2) fails to stably recover β_0. $\square$

Proof of Proposition 4.1: We use the technical tools developed in Cai and Zhang [10] to prove this result. We begin by introducing another important concept in the RIP framework - restricted orthogonal constants (ROC) proposed in [11].

Definition 5.1: Suppose $A \in \mathbb{R}^{n\times p}$, define the restricted orthogonal constants (ROC) of order k_1, k_2 as the smallest non-negative number $\theta^A_{k_1,k_2}$ such that

$$|\langle A\beta_1, A\beta_2\rangle| \leq \theta^A_{k_1,k_2}\|\beta_1\|_2\|\beta_2\|_2,$$

for all k_1-sparse vector $\beta_1 \in \mathbb{R}^p$ and k_2-sparse vector $\beta_2 \in \mathbb{R}^p$ with disjoint supports.

Based on Theorem 2.5 in [10],

$$\delta^A_{tk} + \frac{2k - tk}{tk}\theta^A_{tk,tk} < 1 \qquad (31)$$

is a sufficient condition for exact recovery of all k-sparse vectors. By Lemma 3.1 in [10], $\theta^A_{tk,tk} \leq 2\delta^A_{tk}$ when tk is even; $\theta^A_{tk,tk} \leq \frac{2tk}{\sqrt{(tk)^2-1}}\delta^A_{tk}$ when tk is odd. Hence,

$$\delta^A_{tk} + \frac{2k - tk}{tk}\theta^A_{tk,tk} \leq \frac{4-t}{t}\delta^A_{tk}, \quad tk \text{ is even;}$$

$$\delta^A_{tk} + \frac{2k - tk}{tk}\theta^A_{tk,tk} \leq \left(1 + \frac{4k - 2tk}{\sqrt{(tk)^2 - 1}}\right)\delta^A_{tk}, \quad tk \text{ is odd.}$$

The proposition is implied by the inequalities above and (31). $\square$

Proof of Proposition 4.2: The idea of the proof is quite similar to Theorem 3.2 by Cai and Zhang [9]. Define

$$\gamma = \frac{1}{\sqrt{2k}}(\overbrace{1,\cdots,1}^{2k},0,\cdots,0),$$

$$A : \mathbb{R}^p \to \mathbb{R}^p$$

$$\beta \mapsto \frac{2}{\sqrt{4-t}}\left(\beta - \langle\beta,\gamma\rangle\gamma\right).$$

Now for all non-zero $\lceil tk\rceil$-sparse vector $\beta \in \mathbb{R}^p$,

$$\|A\beta\|_2^2 = \frac{4}{4-t}\langle\beta - \langle\beta,\gamma\rangle\gamma, \beta - \langle\beta,\gamma\rangle\gamma\rangle$$

$$= \frac{4}{4-t}(\|\beta\|_2^2 - \langle\beta,\gamma\rangle^2).$$

We can immediately see $\|A\beta\|_2^2 \leq (1 + t/(4-t))\|\beta\|_2^2$. On the other hand by Cauchy-Schwarz's inequality,

$$\langle\beta,\gamma\rangle^2 = \langle\beta, \gamma \cdot 1_{\{supp(\beta)\}}\rangle^2$$

$$\leq \|\beta\|_2^2(\sum_{i\in supp(\beta)}\gamma_i^2) \leq \|\beta\|_2^2 \cdot \frac{\lceil tk\rceil}{2k}.$$

For $k > 1/\varepsilon$, we have

$$\|A\beta\|_2^2 \geq \frac{4}{4-t}(1 - \frac{\lceil tk\rceil}{2k})\|\beta\|_2^2$$

$$\geq \frac{4}{4-t}(1 - \frac{tk}{2k} - \varepsilon/2)\|\beta\|_2^2 > (1 - \frac{t}{4-t} - \varepsilon)\|\beta\|_2^2.$$

Therefore, we must have $\delta^A_{tk} = \delta^A_{\lceil tk\rceil} < t/(4-t) + \varepsilon$.

Finally, we define

$$\beta_0 = (\overbrace{1,\cdots,1}^{k},0,\cdots,0),$$

$$\beta'_0 = (\overbrace{0,\cdots,0}^{k},\overbrace{-1,\cdots,-1}^{k},0,\cdots,0).$$

Then β_0, β'_0 are both k-sparse, and $y = A\beta_0 = A\beta'_0$. There's no way to recover both β_0, β'_0 only from (y, A). $\square$

REFERENCES

[1] J. Andersson and J. Stromberg, "On the theorem of uniform recovery of random sampling matrices," Dept. Math., R. Inst. Technol., Stockholm, Sweden, Jun. 2013.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

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

[4] J. Blanchard and A. Thompson, "On support sizes of restricted isometry constants," *Appl. Comput. Harmon. Anal.*, vol. 29, pp. 382–390, Mar. 2010.

[5] T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signal via l_1 minimization," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3388–3397, Jul. 2009.

[6] T. Cai, L. Wang, and G. Xu, "Shifting inequality and recovery of sparse signals," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1300–1308, Mar. 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, Jul. 2010.

[8] T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Trans. Inf. Theory*, vol. 56, no. 9, pp. 4388–4394, Sep. 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, Jan. 2013.

[10] T. Cai and A. Zhang "Compressed sensing and affine rank minimization under restricted isometry," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3279–3290, Apr. 2013.

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

[12] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[13] E. Candès and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n," *Ann. Stat.*, vol. 35, no. 6, pp. 2313–2351, 2007.

[14] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus Math.*, vol. 346, nos. 9–10, pp. 589–592, 2008.

[15] E. Candès and Y. Plan "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2009.

[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, Sep. 2012.

[17] M. Davies and R. Gribonval, "Restricted isometry constants where L_p sparse recovery can fail for $0 < p \leq 1$," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2203–2214, Jan. 2009.

[18] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

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

[20] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[21] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, Jun. 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, Mar. 2008.

[23] Q. Mo and S. Li, "New bounds on the restricted isometry constant δ_{2k}," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 3, pp. 460–468, 2011.

[24] K. Mohan and M. Fazel, "New restricted isometry results for noisy low-rank recovery," in *Proc. ISIT*, Jun. 2010, pp. 1573–1577.

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

[26] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[27] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing—Probabilistic analysis of a null-space characterization," in *Proc. ICASSP*, Apr. 2008, pp. 3377–3380.

[28] 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, Jan. 2010.

[29] H. Wang and S. Li, "The bounds of restricted isometry constants for low rank matrices recovery," *Sci. China, Math.*, vol. 56, no. 6, pp. 1117–1127, 2012.

[30] S. Zhou, L. Kong, and N. Xiu, "New bounds for RIC in compressed sensing," *J. Operations Res. Soc. China*, vol. 1, no. 2, pp. 227–237, 2013.

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.