Stable Sparse Recovery with Three Unconstrained Analysis Based Approaches

Huanmin Ge, Jinming Wen, Wengu Chen, Jian Weng and Ming-Jun Lai

Abstract

Efficient recovery of sparse signals from underdetermined linear systems has received lots of attentions in recent decade. This paper considers the stable recovery of a signal $\boldsymbol{x}$, which is not k-sparse itself but is k-sparse in terms of a tight frame $\boldsymbol{D}$, from the observations $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$. Three unconstrained minimization approaches will be studied. We show that if the sensing matrix $\boldsymbol{A}$ satisfies the restricted isometry property adapted to a tight frame $\boldsymbol{D}$ with $\delta_{tk} < \sqrt{\frac{t-1}{t+8}}$ for any $t > 1$, then these three approaches can stably recover $\boldsymbol{x}$. As a consequence, when $t = 2$ and 3, our result significantly improves existing best sufficient conditions in terms of δ_{2k} and δ_{3k}. Moreover, we theoretically characterize the recovery error of these methods.

Index Terms

Compressed sensing, unconstrained analysis based approaches, restricted isometry property, signal recovery, tight frame.

I. Introduction

ONE of the central aims of compressed sensing is to efficiently recover sparse signals from underdetermined linear systems. Mathematically, it recasts to recover an unknown sparse signal $\boldsymbol{x} \in \mathbb{R}^n$ from the following linear measurements:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ is a measurement vector, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ $(m \ll n)$ is a known sensing matrix and $\boldsymbol{e} \in \mathbb{R}^m$ is a vector of measurement errors.

If the signal $\boldsymbol{x}$ is k-sparse, i.e., it has at most k nonzero entries, then under some suitable conditions on the sensing matrix $\boldsymbol{A}$, $\boldsymbol{x}$ can be stably recovered via some efficient sparse recovery algorithms, see, e.g., [1]– [17]. The theory for the sparse recovery has been extended to the setting of sparse dictionary recovery motivated by numerous practical applications, e.g. [18], [19]. That is, $\boldsymbol{x}$ in (1) is no longer sparse itself, but it is sparse in terms of an overcomplete dictionary $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ with $d >> n$, i.e. the signal $\boldsymbol{x}$ is expressed as $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{v}$, where $\boldsymbol{v} \in \mathbb{R}^d$ is sparse. Many sparse recovery results for (1) have been extended to the sparse dictionary recovery setting. See, e.g., [20]–[23] and references therein.

In this paper, we assume $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ $(d \geq n)$ is a tight frame, i.e., $\|\boldsymbol{D}_i\|_2 = 1$ for $1 \leq i \leq n$, and for any $\boldsymbol{w} \in \mathbb{R}^n$, it holds that $\boldsymbol{w} = \sum_{i=1}^d (\boldsymbol{w}^\top \boldsymbol{D}_i)\boldsymbol{D}_i$, where $\boldsymbol{D}_i$ is the i-th column of $\boldsymbol{D}$. We refer to [24] for basic theory on tight frame. We say $\boldsymbol{x}$ is sparse in terms of an overcomplete dictionary if $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{v}$ and $\boldsymbol{v} \in \mathbb{R}^d$ is sparse. One of the most commonly used frameworks to recovery sparse dictionary is the restricted isometry property adapted to $\boldsymbol{D}$ (D-RIP) [22] which is defined as follows:

Definition 1. *A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is said to obey the restricted isometry property adapted to $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ (abbreviated D-RIP) of order k with constant δ_k if*

$$(1-\delta_k)\|\boldsymbol{D}\boldsymbol{v}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{D}\boldsymbol{v}\|_2^2 \leq (1+\delta_k)\|\boldsymbol{D}\boldsymbol{v}\|_2^2 \tag{2}$$

H. Ge is with Graduate School, China Academy of Engineering Physics, Beijing, 100088, China (e-mail: gehuanmin@163.com).

J. Wen is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, M5S 3G4, Canada (e-mail: jinming.wen@utoronto.ca).

W. Chen is with Institute of Applied Physics and Computational Mathematics, Beijing, 100088, China (e-mail: chenwg@iapcm.ac.cn).

J. Weng is with the School of Information Technology, Jinan University, Guangzhou 510632, China (e-mail:cryptjweng@gmail.com).

M. Lai is with Department of Mathematics, University of Georgia, Athens, GA 30602, U.S.A., (e-mail: mjlai@uga.edu).

holds for all k-sparse vectors $\boldsymbol{v} \in \mathbb{R}^d$. The smallest constant δ_k is called as the the restricted isometry constant adapted to $\boldsymbol{D}$ (abbreviated D-RIC).

From Definition 1, one can easily see that the D-RIP reduces to the standard RIP when $\boldsymbol{D}$ is the $n \times n$ identity matrix. Thus, the D-RIP is an extension of the restricted isometry property (RIP). Similar to the definition of the standard RIP of order k in [25], we define δ_k as $\delta_{\lceil k \rceil}$ when k is not an integer, where $\lceil k \rceil$ is an integer satisfying $k < \lceil k \rceil < k+1$. In the rest of the paper, we use δ_k to denote the D-RIC of $\boldsymbol{A}$ with order of k.

The following constrained analysis basis pursuit has been proposed in [22] to reconstruct the signal $\boldsymbol{x}$ from (1), where $\boldsymbol{e}$ satisfies $\|\boldsymbol{e}\|_2 \leq \varepsilon$ for a positive constant ε:

$$\min_{\tilde{\boldsymbol{x}} \in \mathbb{R}^n} \|\boldsymbol{D}^\top \tilde{\boldsymbol{x}}\|_1 \quad \text{subject to} \quad \|\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{y}\|_2 \leq \varepsilon. \tag{3}$$

This method has been well studied and a number of sufficient conditions of stably recovering $\boldsymbol{x}$ have been proposed. These include $\delta_{2k} < 0.08$ [22], $\delta_{2k} < 0.2$ [26], $\delta_{2k} < 0.472$ [23], $\delta_{2k} < 0.4931$ [27] and $\delta_{2k} < \frac{\sqrt{2}}{2}$ [28].

In addition, the following unconstrained analysis based approach, which is widely used in image processing and computer vision [29]–[31], has been considered:

$$\min_{\tilde{\boldsymbol{x}} \in \mathbb{R}^n} \lambda \|\boldsymbol{D}^\top \tilde{\boldsymbol{x}}\|_1 + \frac{1}{2}\|\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{y}\|_2^2, \tag{4}$$

where λ is a positive parameter. The unconstrained analysis approach (4) is equivalent to the analysis-based method (3) in the sense that for any $\varepsilon > 0$, there exists a λ such that the solution of (4) is that of (3), and vice versa. It is well known that since (4) is an unconstrained minimization, the computation of (4) is more convenient than that for (3). In addition, if the noise level is not given or cannot be accurately estimated, we use (4) instead of (3). See [32] for the more detailed relation between (3) and (4). An overview of (4) can be found in [33].

It has been shown in [34] and [35] that $\delta_{3k} < \frac{1}{4}$ and $\delta_{2k} < 0.1907$, are sufficient for the stable recovery of the signal $\boldsymbol{x}$ via the approach (4) if

$$\|\boldsymbol{D}^\top \boldsymbol{A}^\top \boldsymbol{e}\|_\infty \leq \frac{\lambda}{2}. \tag{5}$$

The best existing sufficient condition, which is based on δ_{2k} for stably recovering $\boldsymbol{x}$ via (4) is $\delta_{2k} < 0.2$ that was proposed by Shen et al. in [36].

Many efficient algorithms have been proposed to solve the unconstrained analysis problem (4), see, e.g., [35], [37]–[39]. In particular, Zhao et al. [35] considered the monotone version of the fast iterative shrinkage-thresholding algorithm to solve the minimization problem (4). But the proximal operator of this algorithm for $\|\boldsymbol{D}^\top \boldsymbol{x}\|_1$ does not have a closed-form solution, so based on smoothing and decomposition transformations, the following general nonsmooth convex optimization problem was introduced to recover $\boldsymbol{x}$ in [35]:

$$\min_{\tilde{\boldsymbol{x}} \in \mathbb{R}^n, \boldsymbol{z} \in \mathbb{R}^d} \lambda \|\boldsymbol{z}\|_1 + \frac{1}{2}\|\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{y}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{z} - \boldsymbol{D}^\top \tilde{\boldsymbol{x}}\|_2^2. \tag{6}$$

More detailed discussions on the relation between (6) and (4) can be found in [35]. It has been respectively shown in [35] and [36] that $\delta_{2k} < 0.1907$ and $\delta_{2k} < 0.2$ is a sufficient condition of stably recovering the signal $\boldsymbol{x}$ from (1) via the method (6) when $\boldsymbol{e}$ satisfies (5).

Although (4) and (6) can stably recover $\boldsymbol{x}$ from (1) when $\boldsymbol{A}$ satisfies the D-RIP with certain D-RIC if $\boldsymbol{e}$ satisfies (5), they may not be able to stably recover $\boldsymbol{x}$ based on $\boldsymbol{y}$ and $\boldsymbol{A}$ if the measurements $\boldsymbol{y} \in \mathbb{R}^m$ are corrupted by both the bounded noise $\boldsymbol{e_1} \in \mathbb{R}^m$ and the sparse noise $\boldsymbol{e_2} \in \mathbb{R}^m$ in term of certain tight frame $\boldsymbol{D_2} \in \mathbb{R}^{m \times d}$ (i.e., $\|\boldsymbol{D_2}\boldsymbol{e_2} - (\boldsymbol{D_2}^\top \boldsymbol{e_2})_{\max(s)}\|_2 = 0$ for some integer s) [40], that is

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e_1} + \boldsymbol{e_2}. \tag{7}$$

In order to recover signals $\boldsymbol{x}$ in (7), the following separation unconstrained analysis has been introduced in [34]:

$$\min_{\tilde{\boldsymbol{\nu}}\in\mathbb{R}^{n+m}} \lambda\|\boldsymbol{W}^\top\tilde{\boldsymbol{\nu}}\|_1 + \frac{1}{2}\|\boldsymbol{\Phi}\tilde{\boldsymbol{\nu}} - \boldsymbol{y}\|_2^2, \tag{8}$$

where

$$\boldsymbol{\Phi} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{I}_m\end{bmatrix}, \ \boldsymbol{\nu} = \begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{e_2}\end{bmatrix}, \ \boldsymbol{W} = \begin{bmatrix}\boldsymbol{D_1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D_2}\end{bmatrix}. \tag{9}$$

It has been shown that the conditions $\delta_{3(k+s)} < \frac{1}{4}$ [34] and $\delta_{2(k+s)} < 0.2$ [36] can guarantee the stable recovery of the signal $\boldsymbol{x}$ in (7) via the approach (8) if $\boldsymbol{e_1}$ satisfies

$$\|\boldsymbol{W}^\top\boldsymbol{\Phi}^\top\boldsymbol{e_1}\|_\infty \le \frac{\lambda}{2}, \tag{10}$$

where $\boldsymbol{W}$ is defined in (9).

Since there are efficient algorithms for solving (4), (6) and (8), we shall focus on theoretically investigating sufficient conditions of stable recovering $\boldsymbol{x}$ with these three approaches in this paper. From the above introduction, we can see that sufficient conditions can be characterized by using δ_{2k} and δ_{3k}. Since it is usually difficult to check which condition is better than the other one for any two conditions which are based on two different orders of D-RIC of $\boldsymbol{A}$, our paper will use δ_{tk} $(t > 1)$ to characterize new sufficient conditions. One of the main advantages is that the proposed conditions are valid for any order of D-RIP with $t > 1$, and hence the new conditions are more general. The main contributions of this paper are summarized as follows:

- We will show that if $\boldsymbol{A}$ in (1) satisfies the D-RIP of order tk with

$$\delta_{tk} < \sqrt{\frac{t-1}{t+8}}, \qquad t > 1, \tag{11}$$

  then signals $\boldsymbol{x}$ in (1) can be stably recovered via solving (4) or (6) if $\boldsymbol{e}$ satisfies (5) (see Theorem 1 and Theorem 3 below). It is not hard to see that, when $t = 2$, (11) significantly improves the best existing δ_{2k}-based sufficient condition which is $\delta_{2k} < 0.2$ given by [36, Theorem 3] for (4) and [36, Theorem 5] for (6). Moreover, (11) with $t = 3$ greatly improves the best existing sufficient condition based on δ_{3k}, which is $\delta_{3k} < \frac{1}{4}$ given by [34, Theorem 3.1] for (4).
- We will show that if $\boldsymbol{\Phi}$ (see (9)) satisfies the D-RIP with

$$\delta_{t(k+s)} < \sqrt{\frac{t-1}{t+8}}, \qquad t > 1 \tag{12}$$

  and the noise vector $\boldsymbol{e_1}$ in (7) obeys (10), then signals $\boldsymbol{x}$ in (7) can be stably recovered via solving (8) (see Theorem 4 below). It can be easily check that (12) with $t = 2$ and $t = 3$ significantly improves existing best sufficient conditions which are $\delta_{2(k+s)} < 0.2$ given by [36, Theorem 6] and $\delta_{3(k+s)} < \frac{1}{4}$ given by [34, Theorem 4.1], respectively.

In order to present our results, we need some notation and definitions. *Notation*: For a positive integer d, $[d]$ denotes the index set $\{1, 2, \cdots, d\}$. For $S \subseteq [d]$, let $S^c \subseteq [d]$ be the complement of S and $|S|$ be the number of elements in S. For a vector $\boldsymbol{x} \in \mathbb{R}^d$, define the ℓ_p $(1 \le p \le \infty)$ norm of $\boldsymbol{x}$ by $\|\boldsymbol{x}\|_p = (\sum_{i=1}^d |\boldsymbol{x}_i|^p)^{\frac{1}{p}}$ with $1 \le p < \infty$ and $\|\boldsymbol{x}\|_\infty = \max_{i\in[d]} |\boldsymbol{x}_i|$. The norm $\|\boldsymbol{x}\|_0$ denotes the number of nonzero entries of $\boldsymbol{x}$. The support of $\boldsymbol{x}$ is defined by $\mathrm{supp}(\boldsymbol{x}) = \{i \in [d] : \boldsymbol{x}_i \neq 0\}$. For a set $S \subseteq [d]$, denote $\boldsymbol{D}_S$ as $\boldsymbol{D}$ with all but the columns indexed by S set to zero vector. Denote $\boldsymbol{x}_S$ as $\boldsymbol{x}$ with all but the entries indexed by S set to zero. Let $\boldsymbol{x}_{\max(k)}$ as $\boldsymbol{x}$ with all but the largest k entries in absolute value set to zero, and $\boldsymbol{x}_{-\max(k)} = \boldsymbol{x} - \boldsymbol{x}_{\max(k)}$.

The rest of the paper is organized as follows. In Section II, we first introduce some technical lemmas. We present our main results in Section III and show them in Section IV. Finally, we summarize the results in this paper in Section V.

II. Technical lemmas

To prove our main results, let us start with some useful technical lemmas. We will use the following results established in [10, Lemma 1.1].

Lemma 1. *For a positive number α and a positive integer s, define the polytope $T(\alpha, s) \subset \mathbb{R}^d$ by*

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

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

$$U(\alpha, s, \boldsymbol{v}) = \{\boldsymbol{u} \in \mathbb{R}^d : \mathrm{supp}(\boldsymbol{u}) \subseteq \mathrm{supp}(\boldsymbol{v}), \|\boldsymbol{u}\|_0 \le s, \|\boldsymbol{u}\|_1 = \|\boldsymbol{v}\|_1, \|\boldsymbol{u}\|_\infty \le \alpha\}. \tag{13}$$

Then any $\boldsymbol{v} \in T(\alpha, s)$ can be expressed as

$$\boldsymbol{v} = \sum_{i=1}^{N} \lambda_i \boldsymbol{u_i}$$

for some positive integer N, where $\boldsymbol{u_i} \in U(\alpha, s, \boldsymbol{v})$ and

$$\sum_{i=1}^{N} \lambda_i = 1 \text{ with } 0 \le \lambda_i \le 1. \tag{14}$$

Next we need a null space property. That is, the matrix $\boldsymbol{A}$ has the ℓ_2-robust null space property of order k if it satisfies the D-RIP. See [41, Definition 3].

Lemma 2. *Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ satisfy $\boldsymbol{D}\boldsymbol{D}^\top = \boldsymbol{I}$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ satisfy the D-RIP of order tk ($t > 1$) with $\delta_{tk} \in (0, 1)$. Let $T \subseteq [d]$ be an index set with $|T| \le k$. Then for any vector $\boldsymbol{h} \in \mathbb{R}^n$, it holds that*

$$\|\boldsymbol{D}_T^\top \boldsymbol{h}\|_2 \le \beta_1 \|\boldsymbol{A}\boldsymbol{h}\|_2 + \beta_2 \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{\sqrt{k}}, \tag{15}$$

where

$$\beta_1 = \frac{2}{(1-\delta_{tk})\sqrt{1+\delta_{tk}}}, \qquad \beta_2 = \frac{\delta_{tk}}{\sqrt{(1-\delta_{tk}^2)(t-1)}}. \tag{16}$$

Proof. This is a critical results. We include a proof in this paper. See Appendix A. □

In fact, Lemma 2 is a nontrivial improvement of [36, Lemma 1] as the proof is based on Lemma 1. More precisely, [36, Lemma 1] indicates that if $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ satisfies the D-RIP of order $2k$ with $\delta_{tk} \in (0, 0.5)$, then

$$\|\boldsymbol{D}_T^\top \boldsymbol{h}\|_2 \le \bar{\beta}_1 \|\boldsymbol{A}\boldsymbol{h}\|_2 + \bar{\beta}_2 \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{\sqrt{k}},$$

where

$$\overline{\beta}_1 = \frac{\sqrt{1+\delta_{2k}}}{1-2\delta_{2k}}, \qquad \overline{\beta}_2 = \frac{\delta_{2k}}{1-2\delta_{2k}}. \tag{17}$$

From Lemma 2, one can see that (15) holds if $\boldsymbol{A}$ satisfy the D-RIP of order tk for any $t > 1$, thus it is an extension of [36, Lemma 1]. In the following, we will show that although $\beta_1 \ge \overline{\beta}_1$ when $0 < \delta_{2k} \le 2-\sqrt{3}$, $\beta_2 < \overline{\beta}_2$ holds for all $\delta_{2k} \in (0, 0.5)$ when $t = 2$. Thus, Lemma 2 improves [36, Lemma 1] when $t = 2$.

By (16), when $t = 2$

$$\beta_1 = \frac{2}{(1-\delta_{2k})\sqrt{1+\delta_{2k}}}, \qquad \beta_2 = \frac{\delta_{2k}}{\sqrt{1-\delta_{2k}^2}}. \tag{18}$$

Fig. 1. The rates $\beta_1/\overline{\beta}_1$ and $\beta_2/\overline{\beta}_2$ versus δ_{2k} show that β_2 is always smaller than $\overline{\beta}_2$ and β_1 is smaller than $\overline{\beta}_1$ when δ_{2k} is larger enough.

In the following, we compare β_1 with $\bar{\beta}_1$ and β_2 with $\bar{\beta}_2$. By (17), (18) and the fact that $\delta_{2k} \in (0, 0.5)$, it is not hard to show that

$$0 < \frac{\beta_1}{\overline{\beta}_1} = \frac{2}{(1-\delta_{2k})\sqrt{1+\delta_{2k}}} \cdot \frac{1-2\delta_{2k}}{\sqrt{1+\delta_{2k}}} = \frac{2(1-2\delta_{2k})}{1-\delta_{2k}^2} < 2,$$
$$0 < \frac{\beta_2}{\overline{\beta}_2} = \frac{\delta_{2k}}{\sqrt{1-\delta_{2k}^2}} \cdot \frac{1-2\delta_{2k}}{\delta_{2k}} = \frac{1-2\delta_{2k}}{\sqrt{1-\delta_{2k}^2}} < 1.$$

To clearly see the relationships between β_1 and $\bar{\beta}_1$, and between β_2 and $\bar{\beta}_2$, we show them in Figure 1. From Figure 1 one can see that although $\beta_1/\overline{\beta}_1 > 1$ when $0 < \delta_{2k} \leq 2-\sqrt{3}$, $\beta_2/\overline{\beta}_2 < 1$ holds for all $\delta_{2k} \in (0, 0.5)$.

To show our main results in the next section, the following lemma is also necessary.

Lemma 3. *Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ satisfy $\boldsymbol{D}\boldsymbol{D}^\top = \boldsymbol{I}_n$ and $S,\ T \in [d]$ be two index sets with $|S| \leq s$ and $|T| \leq k$. Suppose that for any given vectors $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{h} \in \mathbb{R}^n$, it holds that*

$$\|\boldsymbol{D}_{S^c}^\top \boldsymbol{h}\|_1 \leq a\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_1 + b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta, \tag{19}$$

where constants a, b and β satisfy $a > 0$, $b \geq 0$ and $\beta \geq 0$. If the matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ satisfies the D-RIP with

$$\delta_{s+c} < \sqrt{\frac{c}{c+a^2 s}} \tag{20}$$

for certain constant $c > 0$, then

$$\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2 \leq \frac{1}{\Delta}\Big[\frac{(2as\delta_{s+c} + \sqrt{2\delta_{s+c}s\Delta})(b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta)}{2\sqrt{s}} + \sqrt{c(c+a^2 s)(1+\delta_{s+c})}\|\boldsymbol{A}\boldsymbol{h}\|_2\Big], \tag{21}$$

where

$$\Delta = (c+a^2 s)\Big(\sqrt{\frac{c}{c+a^2 s}} - \delta_{s+c}\Big). \tag{22}$$

Proof. Please see Appendix B. □

We remark that to the best of the authors' knowledge, such kind of result has not been reported in the literature before.

III. Main results

In this section, we will show that if $\boldsymbol{A}$ in (1) satisfy the D-RIP of order tk with (11), then signals $\boldsymbol{x}$ in (1) can be stably recovered via solving (4) or (6) if $\boldsymbol{e}$ satisfies (5). We will also prove that if $\boldsymbol{\Phi}$ (see (9)) satisfies the D-RIP with (12), then signals $\boldsymbol{x}$ in (7) can be stably recovered via solving (8) if the noise vector $\boldsymbol{e_1}$ in (7) obeys (10). These sufficient conditions significantly improves existing ones. In addition to give the sufficient conditions, upper bounds on the recovery errors of these three methods will also be presented in this section.

A. A sufficient condition of stably recovering $\boldsymbol{x}$ with solving (4)

This subsection focuses on investigating the recovery performance of (4), and our main results are stated as in Theorem 1 below which not only provides a sufficient condition of stably recovering $\boldsymbol{x}$ based on solving (4), but also characterizes the recover errors of this method.

Theorem 1. *Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a tight frame and $\boldsymbol{A}$ in* (1) *satisfy the D-RIP of order tk with* (11)*. Suppose that $\boldsymbol{e}$ in* (1) *obeys* (5)*. Then the solution $\hat{\boldsymbol{x}}$ of* (4) *satisfies*

$$\|\boldsymbol{A}(\hat{\boldsymbol{x}}-\boldsymbol{x})\|_2 \le C_1\sqrt{k}\lambda + C_2\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}}, \tag{23}$$

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \le C_3\sqrt{k}\lambda + C_4\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}}, \tag{24}$$

where the constants C_1–C_4 are given precisely as follows:

$$C_1 = \frac{6}{(1-\delta_{tk})\sqrt{1+\delta_{tk}}}, \tag{25}$$

$$C_2 = \frac{2(1-\delta_{tk})\sqrt{1+\delta_{tk}}}{3}, \tag{26}$$

$$C_3 = \frac{12\sqrt{t-1}}{(1-\delta_{tk})\sqrt{t+8}\left(\sqrt{\frac{t-1}{t+8}}-\delta_{tk}\right)}, \tag{27}$$

$$C_4 = \frac{\frac{4(1-\delta_{tk}^2)\sqrt{(t+8)(t-1)}}{3} + 24\delta_{tk} + 4\sqrt{2\delta_{tk}(t+8)\left(\sqrt{\frac{t-1}{t+8}}-\delta_{tk}\right)}}{(t+8)\left(\sqrt{\frac{t-1}{t+8}}-\delta_{tk}\right)} + 1. \tag{28}$$

Proof. Please see Section IV-A. □

In the noise-free case, i.e., when $\boldsymbol{e}=\boldsymbol{0}$, we can choose a λ which is close to 0 such that $C_3\sqrt{k}\lambda$ is close to 0. Then Theorem 1 indicates that $\boldsymbol{x}$ can be approximately estimated by solving (4) if $\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1$ is small. Furthermore, from Theorem 1 one can see that $\boldsymbol{x}$ can be accurately recovered by solving (4) when $\boldsymbol{D}^\top\boldsymbol{x}$ is k-sparse and $\boldsymbol{e}=\boldsymbol{0}$.

To compare our sufficient condition for stably recovery of signals $\boldsymbol{x}$ via solving (4) with existing ones, we give the following result for $t=2$ which can be easily obtained from Theorem 1.

Corollary 1. *Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a tight frame and $\boldsymbol{A}$ in* (4) *satisfy*

$$\delta_{2k} < \frac{\sqrt{10}}{10}. \tag{29}$$

If $\boldsymbol{e}$ in (1) *satisfies* (5)*, then the solution $\hat{\boldsymbol{x}}$ of* (4) *satisfies*

$$\|\boldsymbol{A}(\hat{\boldsymbol{x}}-\boldsymbol{x})\|_2 \le \widetilde{C}_1\sqrt{k}\lambda + \widetilde{C}_2\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}}, \text{ and } \|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \le \widetilde{C}_3\sqrt{k}\lambda + \widetilde{C}_4\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}},$$

where

$$\widetilde{C}_1 = \frac{6}{(1-\delta_{2k})\sqrt{1+\delta_{2k}}}, \quad \widetilde{C}_2 = \frac{2(1-\delta_{2k})\sqrt{1+\delta_{2k}}}{3}, \tag{30}$$

$$\widetilde{C}_3 = \frac{12}{(1-\delta_{2k})(1-\sqrt{10}\delta_{2k})}, \tag{31}$$

$$\widetilde{C}_4 = \frac{\frac{4\sqrt{10}(1-\delta_{2k}^2)}{3} + 24\delta_{2k} + 4\sqrt{2\delta_{2k}\left(\sqrt{10}-10\delta_{2k}\right)}}{\sqrt{10}-10\delta_{2k}} + 1. \tag{32}$$

(a) $\widetilde{C}_1/\overline{C}_1$ and $\widetilde{C}_2/\overline{C}_2$

(b) $\widetilde{C}_3/\overline{C}_3$ and $\widetilde{C}_4/\overline{C}_4$

Fig. 2. The rates $\widetilde{C}_1/\overline{C}_1$, $\widetilde{C}_2/\overline{C}_2$, $\widetilde{C}_3/\overline{C}_3$ and $\widetilde{C}_4/\overline{C}_4$ versus δ_{2k} which shows that these coefficients are similar

In the following, we compare our new sufficient condition of stably recovery of $\boldsymbol{x}$ via solving (4) with existing ones. To save space, instead of comparing our sufficient condition with all of those in [34]–[36], we only compare it with the best existing one ([36, Theorem 3]) which is stated as follows:

Theorem 2. *Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a tight frame and $\boldsymbol{A}$ in* (4) *satisfy*

$$\delta_{2k} < 0.2. \tag{33}$$

If $\boldsymbol{e}$ in (1) *satisfies* (5)*, then the solution $\hat{\boldsymbol{x}}$ of* (4) *satisfies*

$$\|\boldsymbol{A}(\hat{\boldsymbol{x}}-\boldsymbol{x})\|_2 \le \overline{C}_1\sqrt{k}\lambda + \overline{C}_2\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}} \text{ and } \|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \le \overline{C}_3\sqrt{k}\lambda + \overline{C}_4\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}},$$

where

$$\overline{C}_1 = \frac{3\sqrt{1+\delta_{2k}}}{1-2\delta_{2k}}, \quad \overline{C}_2 = \frac{4(1-2\delta_{2k})}{3\sqrt{1+\delta_{2k}}}, \tag{34}$$

$$\overline{C}_3 = \frac{3(1+\delta_{2k})(7-12\delta_{2k})}{2(1-2\delta_{2k})^2(1-5\delta_{2k})}, \tag{35}$$

$$\overline{C}_4 = \frac{46-62\delta_{2k}}{15(1-5\delta_{2k})(1-2\delta_{2k})} + \frac{8}{5}. \tag{36}$$

Since $\sqrt{10}/10 \approx 0.3162$, by (29) and (33), one can easily see that our new D-RIP condition on $\boldsymbol{A}$ is less restrictive than that given by (33). To clearly see the upper bounds on the recovery error, we draw Figure 2 to show $\widetilde{C}_1/\overline{C}_1$, $\widetilde{C}_2/\overline{C}_2$, $\widetilde{C}_3/\overline{C}_3$ and $\widetilde{C}_4/\overline{C}_4$. From Figure 2 (a), we can see that although $\widetilde{C}_1/\overline{C}_1 > 1$ for $0 \le \delta_{2k} < 0.2$, $\widetilde{C}_2/\overline{C}_2 < 1$ for $0 \le \delta_{2k} < 0.2$. Figure 2 (b) shows that both $\widetilde{C}_3/\overline{C}_3 < 1$ and $\widetilde{C}_4/\overline{C}_4 < 1$ hold when $0.05 \le \delta_{2k} < 0.2$. Since $\widetilde{C}_1$ is always larger than $\widetilde{C}_2$ when $0 \le \delta_{2k} < 0.2$, how to improve the upper bound on $\|\boldsymbol{A}(\hat{\boldsymbol{x}}-\boldsymbol{x})\|_2$ by decreasing $\widetilde{C}_1$ will be investigated in the future.

B. A sufficient condition of stably recovering $\boldsymbol{x}$ with solving (6)

The recovery performance of (6) for recovering signals $\boldsymbol{x}$ can be characterized by the following theorem.

Theorem 3. *Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a tight frame and the matrix $\boldsymbol{A}$ in* (1) *satisfy* (11)*. If $\boldsymbol{v}$ in* (1) *obeys* (5)*, then the solution $\hat{\boldsymbol{x}}$ of* (6) *satisfies*

$$\|\boldsymbol{A}(\hat{\boldsymbol{x}}-\boldsymbol{x})\|_2 \le C_1\sqrt{k}\lambda + C_2\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}} + \frac{1}{C_1}\frac{\lambda d}{\sqrt{k}\rho}, \tag{37}$$

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \leq C_3\sqrt{k}\lambda + C_4\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}} + \frac{C_4}{4}\frac{\lambda d}{\sqrt{k}\rho}, \tag{38}$$

where d is the number of columns of $\boldsymbol{D}$ and C_1, C_2, C_3 and C_4 are defined in (25)-(28).

Proof. Please see Section IV-B. □

Clearly, if $t=2$, our sufficient condition given by Theorem 3 reduces to $\delta_{2k} < \frac{\sqrt{10}}{10} \approx 0.3162$ for the method (6). This sufficient condition is less restrictive than the best existing sufficient condition which is $\delta_{2k} < 0.2$ that was given by [36, Theorem 5]. In the following, we look into the upper bound on the recovery error of method (6). By [36, Theorem 5], we have

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \leq \overline{C}_3\sqrt{k}\lambda + \overline{C}_4\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}} + \frac{\overline{C}_4}{4}\frac{\lambda d}{\sqrt{k}\rho}$$

if $\boldsymbol{A}$ in (1) satisfies the D-RIP with $\delta_{2k} < 0.2$, where the constants $\overline{C}_3$ and $\overline{C}_4$ are defined in (35) and (36). Our bound on the estimation error of the method (6) is given by (38), where $\widetilde{C}_3$ and $\widetilde{C}_4$ are respectively defined in (31) and (32). By Figure 2 (b), one can see that both $\frac{\widetilde{C}_3}{\overline{C}_3}$ and $\frac{\widetilde{C}_4}{\overline{C}_4}$ are less than 1 when $\delta_{2k} \in (0.05, 0.2)$, thus our bound on the estimation error is sharper than that given by [36, Theorem 5] in this case.

C. A sufficient condition of stably recovering $\boldsymbol{x}$ with solving (8)

In this subsection, we analyze the recovery performance of (8) for recovering signals $\boldsymbol{x}$ from (7). Specifically, we have the following result.

Theorem 4. *Let frames $\boldsymbol{D_1} \in \mathbb{R}^{n\times d}$ and $\boldsymbol{D_2} \in \mathbb{R}^{m\times d}$ be tight frames such that $\boldsymbol{D_1}^\top\boldsymbol{x}$ is sparse and*

$$\|\boldsymbol{D_2}^\top\boldsymbol{e_2} - (\boldsymbol{D_2}^\top\boldsymbol{e_2})_s\|_1 = 0. \tag{39}$$

Suppose that $\boldsymbol{\Phi}$ (see (9)*) satisfies the D-RIP with* (12) *and $\boldsymbol{e_1}$ in* (7) *obeys* (10)*. Then the solution $\hat{\boldsymbol{\nu}} = \begin{bmatrix}\hat{\boldsymbol{x}}\\ \hat{\boldsymbol{e}}_{\boldsymbol{2}}\end{bmatrix}$ of* (8) *satisfies*

$$\|\boldsymbol{\Phi}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\|_2 \leq C_5\sqrt{k+s}\lambda + C_6\frac{\|(\boldsymbol{D_1}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}}, \tag{40}$$

$$\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \leq C_7\sqrt{k+s}\lambda + C_8\frac{\|(\boldsymbol{D_1}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k+s}}, \tag{41}$$

where

$$C_5 = \frac{6}{(1-\delta_{t(k+s)})\sqrt{1+\delta_{t(k+s)}}}, \tag{42}$$

$$C_6 = \frac{2(1-\delta_{t(k+s)})\sqrt{1+\delta_{t(k+s)}}}{3}, \tag{43}$$

$$C_7 = \frac{12\sqrt{(t-1)}}{(1-\delta_{t(k+s)})\sqrt{t+8}\left(\sqrt{\frac{t-1}{t+8}}-\delta_{t(k+s)}\right)}, \tag{44}$$

$$C_8 = \frac{\frac{4(1-\delta^2_{t(k+s)})\sqrt{(t+8)(t-1)}}{3} + 24\delta_{t(k+s)} + 4\sqrt{2\delta_{t(k+s)}(t+8)\left(\sqrt{\frac{t-1}{t+8}}-\delta_{t(k+s)}\right)}}{(t+8)\left(\sqrt{\frac{t-1}{t+8}}-\delta_{t(k+s)}\right)} + 1. \tag{45}$$

Proof. Please see Section IV-C. □

Obviously, if $t = 2$, our sufficient condition given by Theorem 4 is $\delta_{2(k+s)} < \frac{\sqrt{10}}{10} \approx 0.3162$ for the method (8). This sufficient condition is less restrictive than the best existing one which is $\delta_{2(k+s)} < 0.2$ that was given by [36, Theorem 6]. In the following, we look into the upper bound on the estimation error of method (8). By (41), our bound on the estimation error of the method (8) is

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \widetilde{C}_7\sqrt{k+s}\lambda + \widetilde{C}_8\frac{\|(\boldsymbol{D}_{\boldsymbol{1}}^\top \boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k+s}},$$

where

$$\begin{aligned}\widetilde{C}_7 =& \frac{12}{(1-\delta_{2(k+s)})(1-\sqrt{10}\delta_{2(k+s)})},\\ \widetilde{C}_8 =& \frac{\frac{4\sqrt{10}(1-\delta_{2(k+s)}^2)}{3} + 24\delta_{2(k+s)} + 4\sqrt{2\delta_{2(k+s)}\left(\sqrt{10}-10\delta_{2(k+s)}\right)}}{\sqrt{10}-10\delta_{2(k+s)}} + 1.\end{aligned}$$

From [36, Theorem 6], it follows that

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \overline{C}_7\sqrt{k+s}\lambda + \overline{C}_8\frac{\|(\boldsymbol{D}_{\boldsymbol{1}}^\top \boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k+s}}$$

where

$$\overline{C}_7 = \frac{3(1+\delta_{2(k+s)})(7-12\delta_{2(k+s)})}{2(1-2\delta_{2(k+s)})^2(1-5\delta_{2(k+s)})} \text{ and } \overline{C}_8 = \frac{46-62\delta_{2(k+s)}}{15(1-5\delta_{2(k+s)})(1-2\delta_{2(k+s)})} + \frac{8}{5}.$$

Notice that the ratios $\frac{\widetilde{C}_7}{\overline{C}_7}$ and $\frac{\widetilde{C}_8}{\overline{C}_8}$ versus $\delta_{2(k+s)} \in (0, 0.2)$ are the same as $\frac{\widetilde{C}_3}{\overline{C}_3}$ and $\frac{\widetilde{C}_4}{\overline{C}_4}$ versus $\delta_{2k} \in (0, 0.2)$. Therefore, by Figure 2 (b), one can see that both $\frac{\widetilde{C}_7}{\overline{C}_7}$ and $\frac{\widetilde{C}_8}{\overline{C}_8}$ are less than 1 when $\delta_{2(k+s)} \in (0.05, 0.2)$, which means that our bound on the estimation error is sharper than that given by [36, Theorem 5] in this case.

IV. Proofs

In this section, we prove our main results given in Section III.

A. *Proof of Theorem* 1

Proof. To simplify notation, denote

$$T = \text{supp}((\boldsymbol{D}^\top \boldsymbol{x})_{\max(k)}), \tag{46}$$

then by the definition of $(\boldsymbol{D}^\top \boldsymbol{x})_{\max(k)}$, we obtain $|T| \leq k$. Furthermore, denote

$$\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}, \tag{47}$$

then by [36, eq. (3.1)], we have

$$\|\boldsymbol{A}\boldsymbol{h}\|_2^2 \leq 3\lambda\|\boldsymbol{D}_T^\top \boldsymbol{h}\|_1 + 4\lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 - \lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1. \tag{48}$$

Therefore,

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{h}\|_2^2 \overset{(a)}{\leq}& 3\lambda\sqrt{k}\|\boldsymbol{D}_T^\top \boldsymbol{h}\|_2 + 4\lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 - \lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_2\\ \overset{(b)}{\leq}& 3\lambda\sqrt{k}\Big(\beta_1\|\boldsymbol{A}\boldsymbol{h}\|_2 + \beta_2\frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{\sqrt{k}}\Big) + 4\lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 - \lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1\\ =& 3\lambda\sqrt{k}\beta_1\|\boldsymbol{A}\boldsymbol{h}\|_2 + 4\lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 - \lambda(1-3\beta_2)\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1\end{aligned}$$

$$\overset{(c)}{\leq} 3\lambda\sqrt{k}\beta_1\|\boldsymbol{A}\boldsymbol{h}\|_2 + 4\lambda\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1, \tag{49}$$

where β_1 and β_2 are defined in (16), (a) is from the Cauchy-Schwartz inequality and the fact that $|T| \leq k$, (b) follows from (15) and (c) is because $1 - 3\beta_2 > 0$. Indeed, by (11) and the second equality in (16), we have

$$\begin{aligned} 1 - 3\beta_2 &= 1 - \frac{3\delta_{tk}}{\sqrt{(1-\delta_{tk}^2)(t-1)}} > 1 - \frac{3\sqrt{\frac{t-1}{t+8}}}{\sqrt{(1-\frac{t-1}{t+8})(t-1)}} \\ &= \frac{1}{\sqrt{1-\frac{t-1}{t+8}}}\Big(\sqrt{1-\frac{t-1}{t+8}} - \sqrt{\frac{9}{t+8}}\Big) = 0. \end{aligned}$$

By (49), we obtain

$$\begin{aligned} \|\boldsymbol{A}\boldsymbol{h}\|_2 &\leq \frac{3\lambda\sqrt{k}\beta_1 + \sqrt{9k\beta_1^2\lambda^2 + 16\lambda\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1}}{2} \leq \frac{3\lambda\sqrt{k}\beta_1 + \sqrt{\Big(3\sqrt{k}\beta_1\lambda + \frac{8\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1}{3\sqrt{k}\beta_1}\Big)^2}}{2} \\ &= 3\lambda\sqrt{k}\beta_1 + \frac{4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1}{3\sqrt{k}\beta_1} = 3\lambda\sqrt{k}\beta_1 + \frac{4\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{3\sqrt{k}\beta_1}, \end{aligned} \tag{50}$$

where the last equality is from (46), and the definitions of $(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}$ and $(\boldsymbol{D}^\top\boldsymbol{x})_{\max(k)}$. Then, by the first equality in (16), (25), (26), (47) and (50), one can see that (23) holds.

Since $\boldsymbol{D}$ is a tight frame, by [24, Chapter 3], we have $\boldsymbol{D}\boldsymbol{D}^T = \boldsymbol{I}$. Then

$$\|\boldsymbol{h}\|_2 = \|\boldsymbol{D}^\top\boldsymbol{h}\|_2 = \sqrt{\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2^2 + \|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_2^2}, \tag{51}$$

where we denote

$$S = \mathrm{supp}((\boldsymbol{D}^\top\boldsymbol{h})_{\max(k)}). \tag{52}$$

By (46), (52), and the definitions of $(\boldsymbol{D}^\top\boldsymbol{h})_{\max(k)}$ and $(\boldsymbol{D}^\top\boldsymbol{x})_{\max(k)}$, one can see that

$$\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_1 \geq \|\boldsymbol{D}_T^\top\boldsymbol{h}\|_1. \tag{53}$$

Thus

$$\begin{aligned} \|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_1 &= \|\boldsymbol{D}^\top\boldsymbol{h}\|_1 - \|\boldsymbol{D}_S^\top\boldsymbol{h}\|_1 \leq \|\boldsymbol{D}^\top\boldsymbol{h}\|_1 - \|\boldsymbol{D}_T^\top\boldsymbol{h}\|_1 = \|\boldsymbol{D}_{T^c}^\top\boldsymbol{h}\|_1 \\ &\leq 3\|\boldsymbol{D}_T^\top\boldsymbol{h}\|_1 + 4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1 \leq 3\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_1 + 4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1, \end{aligned} \tag{54}$$

where the second to the last inequality is from (48) and the last inequality is due to (53).

We show (24) by giving an upper bound on $\|\boldsymbol{h}\|_2$. Towards this aim, by (51), we give upper bounds on $\|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_2$ and $\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2$. We first give an upper bound on $\|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_2$. Specifically, we have

$$\begin{aligned} \|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_2 &\leq \sqrt{\|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_1\|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_\infty} \\ &\overset{(a)}{\leq} \sqrt{(3\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_1 + 4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1)\frac{\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_1}{k}} \\ &\overset{(b)}{\leq} \sqrt{(3\sqrt{k}\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2 + 4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1)\frac{\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2}{\sqrt{k}}} = \sqrt{3\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2^2 + \frac{4\|\boldsymbol{D}_S^\top\boldsymbol{x}\|_1\|\boldsymbol{D}_{T^c}^\top\boldsymbol{h}\|_2}{\sqrt{k}}}, \end{aligned} \tag{55}$$

where (a) follows from (54) and

$$\|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_\infty \leq \frac{\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_1}{k}$$

which is based on the facts that the absolute values of the entries of $\boldsymbol{D}_S^\top \boldsymbol{h}$ are not smaller than those of $\boldsymbol{D}_{S^c}^\top \boldsymbol{h}$ and $|S| \leq k$, and (b) is from the Cauchy-Schwarz inequality.

In the following, we give an upper bound on $\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2$ by using Lemma 3 with assuming $a = 3$, $b = 4$, $\beta = 0$, $c = (t-1)k, s = k$. By (46) and (52), we have $|T| \leq k$ and $|S| \leq s$. Moreover, by (11) and (54), one can check that (19) and (20) hold. Then by (22), we obtain

$$\Delta = ((t-1)k + 9k)\Big(\sqrt{\frac{(t-1)k}{(t-1)k+9k}} - \delta_{tk}\Big) = k(t+8)\Big(\sqrt{\frac{t-1}{t+8}} - \delta_{tk}\Big). \tag{56}$$

Therefore, according to (21), we obtain

$$\begin{aligned}\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2 \leq & \frac{1}{\Delta}\Big[\Big(12k\delta_{tk} + 2\sqrt{2k\delta_{tk}\Delta}\Big)\frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1}{\sqrt{k}} + k\sqrt{(t-1)(t+8)(1+\delta_{tk})}\|\boldsymbol{A}\boldsymbol{h}\|_2\Big] \\ \leq & \frac{1}{\Delta}\Big[\Big(12k\delta_{tk} + 2\sqrt{2k\delta_{tk}\Delta} + \frac{2k(1-\delta_{tk}^2)\sqrt{(t+8)(t-1)}}{3}\Big)\frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1}{\sqrt{k}} \\ & + \frac{6k\sqrt{k(t+8)(t-1)}}{1-\delta_{tk}}\lambda\Big],\end{aligned} \tag{57}$$

where Δ is defined in (56), and the second inequality follows from (23), (46) and (47). Then, we obtain

$$\begin{aligned}\|\boldsymbol{h}\|_2 \overset{(a)}{\leq} & \Big(\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2^2 + 3\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2^2 + \frac{4\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1}{\sqrt{k}}\Big)^{\frac{1}{2}} \leq 2\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2 + \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1}{\sqrt{k}} \\ \overset{(b)}{\leq} & \frac{1}{\Delta}\Big[\Big(24k\delta_{tk} + 4\sqrt{2k\delta_{tk}\Delta} + \frac{4k(1-\delta_{tk}^2)\sqrt{(t+8)(t-1)}}{3} + \Delta\Big)\frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1}{\sqrt{k}} \\ & + \frac{12k\sqrt{k(t+8)(t-1)}}{1-\delta_{tk}}\lambda\Big],\end{aligned} \tag{58}$$

where (a) is from (51) and (55), and (b) follows from (57).

By (27), (28), (47), (56) and (58), we can see that (24) holds.

□

B. Proof of Theorem 3

Proof. Let T and $\boldsymbol{h}$ be defined as in (46) and (47), then $|T| \leq k$ and by the proof of [35, Lemma IV.4], one has

$$\|\boldsymbol{A}\boldsymbol{h}\|_2^2 + \lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1 \leq 3\lambda\|\boldsymbol{D}_T^\top \boldsymbol{h}\|_1 + 4\lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \frac{\lambda^2 d}{\rho}. \tag{59}$$

Hence, using the techniques for deriving (49), we obtain

$$\|\boldsymbol{A}\boldsymbol{h}\|_2^2 \leq 3\lambda\sqrt{k}\beta_1\|\boldsymbol{A}\boldsymbol{h}\|_2 + 4\lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \frac{\lambda^2 d}{\rho}, \tag{60}$$

where β_1 is defined in (16). From the above inequality, it follows that

$$\begin{aligned}& \|\boldsymbol{A}(\hat{\boldsymbol{x}} - \boldsymbol{x})\|_2 \overset{(a)}{=} \|\boldsymbol{A}\boldsymbol{h}\|_2 \\ \overset{(b)}{\leq} & \frac{3\lambda\sqrt{k}\beta_1 + \sqrt{9k\beta_1^2\lambda^2 + 16\lambda\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \frac{4\lambda^2 d}{\rho}}}{2} \\ \leq & \frac{3\lambda\sqrt{k}\beta_1 + \sqrt{\Big(3\sqrt{k}\beta_1\lambda + \frac{8\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \frac{2\lambda d}{\rho}}{3\sqrt{k}\beta_1}\Big)^2}}{2}\end{aligned}$$

$$
\begin{aligned}
&=3\lambda\sqrt{k}\beta_1+\frac{1}{3\sqrt{k}\beta_1}\Big(4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1+\frac{\lambda d}{\rho}\Big)\\
&\overset{(c)}{=}3\lambda\sqrt{k}\beta_1+\frac{1}{3\sqrt{k}\beta_1}\Big(4\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1+\frac{\lambda d}{\rho}\Big)\\
&\overset{(d)}{=}\frac{6}{(1-\delta_{tk})\sqrt{1+\delta_{tk}}}\sqrt{k}\lambda+\frac{2(1-\delta_{tk})\sqrt{1+\delta_{tk}}}{3}\cdot\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}}\\
&\quad+\frac{(1-\delta_{tk})\sqrt{1+\delta_{tk}}}{6}\frac{\lambda d}{\sqrt{k}\rho}\\
&\overset{(e)}{=}C_1\sqrt{k}\lambda+C_2\frac{\|(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k}}+\frac{1}{C_1}\frac{\lambda d}{\sqrt{k}\rho},
\end{aligned}
$$

where (a), (b) and (d) follows respectively from (47), (60) and the first equality in (16), (c) is due to (46) and the definitions of $(\boldsymbol{D}^\top\boldsymbol{x})_{-\max(k)}$ and $(\boldsymbol{D}^\top\boldsymbol{x})_{\max(k)}$, and (e) is from (25), (26). Thus, (37) is proved.

Let S be defined as in (52), then $|S|\le k$. Moreover, (51), (53) and (54) hold. Then, using the techniques for deriving (54) and (59), one can obtain

$$
\|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_1\le 3\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_1+4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1+\frac{\lambda d}{\rho}. \tag{61}
$$

Similar to the proof of (24), in order to show (38), we first give an upper bound on $\|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_2$. Specifically, by the techniques for deriving (55) and (61), one obtains

$$
\|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}\|_2\le\sqrt{3\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2^2+\Big(4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1+\frac{\lambda d}{\rho}\Big)\frac{\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2}{\sqrt{k}}}. \tag{62}
$$

In the following, we give an upper bound on $\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2$ by applying Lemma 3 with $a=3$, $b=4$, $\beta=\frac{\lambda d}{\rho}$ and $c=(t-1)k$. By (46) and (52), we have $|T|\le k$ and $|S|\le s$. Moreover, by (11) and (61), one also checks that (19) and (20) hold. Since Δ in (22) is only depend on a, b and s, (56) holds. Hence, from (21), it follows that

$$
\begin{aligned}
\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2\le&\frac{1}{\Delta}\Big[(12k\delta_{tk}+2\sqrt{2k\delta_{tk}\Delta})\frac{\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1}{\sqrt{k}}+(6k\delta_{tk}+\sqrt{2k\delta_{tk}\Delta})\frac{\lambda d}{2\sqrt{k}\rho}\\
&+k\sqrt{(t-1)(t+8)(1+\delta_{tk})}\|\boldsymbol{A}\boldsymbol{h}\|_2\Big]\\
\le&\frac{1}{\Delta}\Big[\Big(12k\delta_{tk}+2\sqrt{2k\delta_{tk}\Delta}+\frac{2k(1-\delta_{tk}^2)\sqrt{(t+8)(t-1)}}{3}\Big)\frac{\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1}{\sqrt{k}}\\
&+\Big(6k\delta_{tk}+\sqrt{2k\delta_{tk}\Delta}+\frac{k(1-\delta_{tk}^2)\sqrt{(t+8)(t-1)}}{3}\Big)\frac{\lambda d}{2\sqrt{k}\rho}+\frac{6k\sqrt{k(t+8)(t-1)}}{1-\delta_{tk}}\lambda\Big],
\end{aligned} \tag{63}
$$

where Δ is defined in (56), and the last inequality is from (37), (47) and (46). Then, one obtains that

$$
\begin{aligned}
\|\boldsymbol{h}\|_2\overset{(a)}{\le}&\Big[\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2^2+3\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2^2+\Big(4\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1+\frac{\lambda d}{\rho}\Big)\frac{\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2}{\sqrt{k}}\Big]^{\frac12}\\
\le&2\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2+\Big(\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1+\frac{\lambda d}{4\rho}\Big)\frac{1}{\sqrt{k}}\\
\overset{(b)}{\le}&\frac{1}{\Delta}\Big[\Delta+\Big(24k\delta_{tk}+4\sqrt{2k\delta_{tk}\Delta}+\frac{4k(1-\delta_{tk}^2)\sqrt{(t+8)(t-1)}}{3}\Big)\frac{\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1}{\sqrt{k}}\\
&+\frac{12k\sqrt{k(t+8)(t-1)}}{(1-\delta_{tk})}\lambda+\Big(\frac{\Delta}{4}+6k\delta_{tk}+\sqrt{2k\delta_{tk}\Delta}+\frac{k(1-\delta_{tk}^2)\sqrt{(t+8)(t-1)}}{3}\Big)\frac{\lambda d}{\sqrt{k}\rho}\Big]
\end{aligned}
$$

$$\overset{(c)}{=} C_3\sqrt{k}\lambda + C_4\frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1}{\sqrt{k}} + \frac{C_4}{4}\frac{\lambda d}{\sqrt{k}\rho},$$

where (a) is from (51) and (62), (b) is due to (63) and (c) follows from (56), (27) and (28). Therefore, by the above inequality, (47), (52) and the definitions of $(\boldsymbol{D}^\top \boldsymbol{x})_{-\max(k)}$ and $(\boldsymbol{D}^\top \boldsymbol{x})_{\max(k)}$, (37) holds. We complete the proof of the theorem. □

C. Proof of Theorem 4

Proof. Since $\boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{e_2} \end{bmatrix}, \tilde{\boldsymbol{\nu}} = \begin{bmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{e}}_{\boldsymbol{2}} \end{bmatrix}, \hat{\boldsymbol{\nu}} = \begin{bmatrix} \hat{\boldsymbol{x}} \\ \hat{\boldsymbol{e}}_{\boldsymbol{2}} \end{bmatrix}$, $\boldsymbol{W}$ and $\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{I}_m \end{bmatrix}$ in (8) can be respectively viewed as $\boldsymbol{x}, \tilde{\boldsymbol{x}}, \hat{\boldsymbol{x}}$, $\boldsymbol{D}$ and $\boldsymbol{A}$ in (4), this theorem can be proved by utilizing Theorem 1. Specifically, by (12) and (10), (11) and (5) hold, then according to Theorem 1, we obtain

$$\begin{aligned}
\|\boldsymbol{\Phi}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu})\|_2 \overset{(a)}{\leq} & C_5\sqrt{k+s}\lambda + C_6\frac{\left\|\left(\boldsymbol{W}^\top \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{e_2} \end{bmatrix}\right)_{-\max(k+s)}\right\|_1}{\sqrt{k+s}} \\
\overset{(b)}{=} & C_5\sqrt{k+s}\lambda + C_6\frac{\left\|\left(\begin{bmatrix} \boldsymbol{D_1}^\top & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D_2}^\top \end{bmatrix}\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{e_2} \end{bmatrix}\right)_{-\max(k+s)}\right\|_1}{\sqrt{k+s}} \\
= & C_5\sqrt{k+s}\lambda + C_6\frac{\left\|\begin{bmatrix} \boldsymbol{D_1}^\top \boldsymbol{x} \\ \boldsymbol{D_2}^\top \boldsymbol{e_2} \end{bmatrix}_{-\max(k+s)}\right\|_1}{\sqrt{k+s}} \\
\leq & C_5\sqrt{k+s}\lambda + C_6\frac{\left\|\begin{bmatrix} (\boldsymbol{D_1}^\top \boldsymbol{x})_{-\max(k)} \\ (\boldsymbol{D_2}^\top \boldsymbol{e_2})_{-\max(s)} \end{bmatrix}\right\|_1}{\sqrt{k+s}} \\
\overset{(c)}{=} & C_5\sqrt{k+s}\lambda + C_6\frac{\|(\boldsymbol{D_1}^\top \boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k+s}},
\end{aligned}$$

where the constants C_5 and C_6 are respectively defined in (42) and (43), (a) is from (23), (b) follows from (9) and (c) is due to (39).

Furthermore, by using the techniques for the above inequality and (24), we can show that

$$\|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\|_2 \leq C_7\sqrt{k+s}\lambda + C_8\frac{\left\|\left(\boldsymbol{W}^\top \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{e_2} \end{bmatrix}\right)_{-\max(k+s)}\right\|_1}{\sqrt{k+s}} = C_7\sqrt{k+s}\lambda + C_8\frac{\|(\boldsymbol{D_1}^\top \boldsymbol{x})_{-\max(k)}\|_1}{\sqrt{k+s}}, \quad (64)$$

where the constants C_7 and C_8 are respectively defined in (44) and (45). Since (64) and

$$\|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\|_2 = \left\| \begin{bmatrix} \hat{\boldsymbol{x}} \\ \hat{\boldsymbol{e}}_{\boldsymbol{2}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{e_2} \end{bmatrix} \right\|_2 \geq \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2,$$

(41) holds. We complete the proof of the theorem. □

V. Conclusion

In this paper, we have studied sufficient conditions for stable recovery of a signal $\boldsymbol{x}$, which is not sparse itself but it is sparse in terms of a tight frame, from noisy measurements by three unconstrained analysis based approaches. We have proved that if the sensing matrix $\boldsymbol{A}$ satisfies $\delta_{tk} < \sqrt{\frac{t-1}{t+8}}$ with $t > 1$, then any signal $\boldsymbol{x}$ can be stably recovered from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$ by solving (4) or (6) under the noise level $\|\boldsymbol{D}^\top \boldsymbol{A}^\top \boldsymbol{e}\|_\infty \leq \frac{\lambda}{2}$. In addition, we have also proved that $\delta_{tk} < \sqrt{\frac{t-1}{t+8}}$ is a sufficient condition of stable recovering any signal $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}_1 + \boldsymbol{e}_2$ by solving (8) under some constrains on the noises $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$.

APPENDIX A
PROOF OF LEMMA 2

Proof. We first show that (15) holds when tk is an integer. For a given $t > 1$, define

$$T_1 = \left\{ i \in T^c \big| |(\boldsymbol{D}_{T^c}^\top \boldsymbol{h})_i)| > \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{(t-1)k} \right\}, \tag{65}$$

$$T_2 = \left\{ i \in T^c \big| |(\boldsymbol{D}_{T^c}^\top \boldsymbol{h})_i)| \leq \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{(t-1)k} \right\}. \tag{66}$$

Then clearly $T^c = T_1 \cup T_2$ which implies that

$$\boldsymbol{D}^\top \boldsymbol{h} = \boldsymbol{D}_T^\top \boldsymbol{h} + \boldsymbol{D}_{T^c}^\top \boldsymbol{h} = \boldsymbol{D}_T^\top \boldsymbol{h} + \boldsymbol{D}_{T_1}^\top \boldsymbol{h} + \boldsymbol{D}_{T_2}^\top \boldsymbol{h}. \tag{67}$$

Furthermore, by (65), $T_1 \cap T = \emptyset$ which implies that $\|\boldsymbol{D}_T^\top \boldsymbol{h}\|_2 \leq \|\boldsymbol{D}_{T \cup T_1}^\top \boldsymbol{h}\|_2$. Thus, to show (15), it suffices to show that

$$\|\boldsymbol{D}_{T \cup T_1}^\top \boldsymbol{h}\|_2 \leq \beta_1 \|\boldsymbol{A}\boldsymbol{h}\|_2 + \beta_2 \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{\sqrt{k}}, \tag{68}$$

where β_1 and β_2 are defined in (16). Let us use Lemma 1 to prove (68). We first show that

$$|T_1| < (t-1)k. \tag{69}$$

If $T_1 = \emptyset$, then $|T_1| = 0$ and hence (69) holds. Otherwise, by (65), one yields that

$$\|\boldsymbol{D}_{T_1}^\top \boldsymbol{h}\|_1 > |T_1| \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{(t-1)k} = |T_1| \frac{\|\boldsymbol{D}_{T_1 \cup T_2}^\top \boldsymbol{h}\|_1}{(t-1)k} \geq |T_1| \frac{\|\boldsymbol{D}_{T_1}^\top \boldsymbol{h}\|_1}{(t-1)k}.$$

Thus a quick simplification of the above inequality yields the (69).

By (65) and (66), $T^c = T_1 \cup T_2$ and $T_1 \cap T_2 = \emptyset$. The above inequality implies

$$\|\boldsymbol{D}_{T_2}^\top \boldsymbol{h}\|_1 = \|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1 - \|\boldsymbol{D}_{T_1}^\top \boldsymbol{h}\|_1 \leq \|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1 - |T_1| \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{(t-1)k} = ((t-1)k - |T_1|) \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{(t-1)k}$$

and we use (66) to have

$$\|\boldsymbol{D}_{T_2}^\top \boldsymbol{h}\|_\infty \leq \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{(t-1)k}.$$

Since we assumed tk is an integer, $(t-1)k - |T_1|$ is a positive integer. Then by Lemma 1 with

$$\alpha = \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{(t-1)k}, \;\; s = (t-1)k - |T_1|, \;\; \boldsymbol{v} = \boldsymbol{D}_{T_2}^\top \boldsymbol{h},$$

one can see that $\boldsymbol{D}_{T_2}^\top \boldsymbol{h}$ can be expressed as

$$\boldsymbol{D}_{T_2}^\top \boldsymbol{h} = \sum_{i=1}^{N} \lambda_i \boldsymbol{u_i} \tag{70}$$

for certain positive integer N, where $1 \leq i \leq N$, λ_i satisfies (14) and

$$\boldsymbol{u_i} \in U\Big(\frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_2}{(t-1)k}, (t-1)k - |T_1|, \boldsymbol{D}_{T_2}^\top \boldsymbol{h} \Big)$$

with the set U being defined in (13). Moreover, by (13), one can see that

$$\|\boldsymbol{u_i}\|_2^2 \leq \|\boldsymbol{u_i}\|_0 \|\boldsymbol{u_i}\|_\infty^2 \leq ((t-1)k - |T_1|) \Big(\frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1}{(t-1)k} \Big)^2 \leq \frac{\|\boldsymbol{D}_{T^c}^\top \boldsymbol{h}\|_1^2}{(t-1)k}. \tag{71}$$

To simplify notation, for $1 \leq i \leq N$, we define

$$\boldsymbol{v_i} = (1+\delta_{tk})\boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h} + \delta_{tk}\boldsymbol{u_i}, \tag{72}$$

$$\bar{\boldsymbol{v}}_{\boldsymbol{i}} = (1-\delta_{tk})\boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h} - \delta_{tk}\boldsymbol{u_i}. \tag{73}$$

By (13), $\|\boldsymbol{u_i}\|_0 \leq (t-1)k - |T_1|$ which combing with (69) and the assumption that $|T| \leq k$, we can see that both $\boldsymbol{v_i}$ and $\bar{\boldsymbol{v}}_{\boldsymbol{i}}$ are tk-sparse for $1 \leq i \leq N$. Moreover, by (72) and (73), we have

$$\begin{aligned}
\sum_{i=1}^N \lambda_i(\|\boldsymbol{AD}\boldsymbol{v_i}\|_2^2 - \|\boldsymbol{AD}\bar{\boldsymbol{v}}_{\boldsymbol{i}}\|_2^2) =& \sum_{i=1}^N \lambda_i \Big(\|\boldsymbol{ADD}_{T\cup T_1}^\top \boldsymbol{h} + \delta_{tk}\boldsymbol{AD}(\boldsymbol{u_i} + \boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h})\|_2^2 \\
&- \|\boldsymbol{ADD}_{T\cup T_1}^\top \boldsymbol{h} - \delta_{tk}\boldsymbol{AD}(\boldsymbol{u_i} + \boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h})\|_2^2 \Big) \\
=& \sum_{i=1}^N \lambda_i \Big(4\delta_{tk}(\boldsymbol{ADD}_{T\cup T_1}^\top \boldsymbol{h})^\top \boldsymbol{AD}(\boldsymbol{u_i} + \boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h}) \Big) \\
\overset{(a)}{=}& 4\delta_{tk}(\boldsymbol{ADD}_{T\cup T_1}^\top \boldsymbol{h})^\top \boldsymbol{AD}(\sum_{i=1}^N \lambda_i \boldsymbol{u_i} + \boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h}) \\
\overset{(b)}{=}& 4\delta_{tk}(\boldsymbol{ADD}_{T\cup T_1}^\top \boldsymbol{h})^\top \boldsymbol{ADD}^\top \boldsymbol{h} \\
\overset{(c)}{=}& 4\delta_{tk}(\boldsymbol{ADD}_{T\cup T_1}^\top \boldsymbol{h})^\top \boldsymbol{Ah} \\
\overset{(d)}{\leq}& 4\delta_{tk}\sqrt{1+\delta_{tk}}\|\boldsymbol{DD}_{T\cup T_1}^\top \boldsymbol{h}\|_2 \|\boldsymbol{Ah}\|_2 \\
\overset{(e)}{\leq}& 4\delta_{tk}\sqrt{1+\delta_{tk}}\|\boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h}\|_2 \|\boldsymbol{Ah}\|_2,
\end{aligned} \tag{74}$$

where (a) follows from (14), (b) is because of (67) and (70), (c) is from the assumption that $\boldsymbol{DD}^\top = \boldsymbol{I}$, (d) follows from (2), (69) and the Cauchy-Schwarz inequality and (e) is from the fact that the maximal singular value of $\boldsymbol{D}$ is not larger than 1 which can be obtained from $\boldsymbol{DD}^\top = \boldsymbol{I}$.

In the following, we give a lower bound on the left-hand side of (74) which combing with (74) can show (68). Toward this goal, we assume $\bar{\boldsymbol{D}} \in \mathbb{R}^{(d-n)\times d}$ is the orthonormal complement of $\boldsymbol{D}$. Then

$$\bar{\boldsymbol{D}}\boldsymbol{D}^\top = \boldsymbol{0}, \ \ \|\boldsymbol{v}\|_2^2 = \|\boldsymbol{Dv}\|_2^2 + \|\bar{\boldsymbol{D}}\boldsymbol{v}\|_2^2, \tag{75}$$

for any $\boldsymbol{v} \in \mathbb{R}^d$. Moreover, by (2), for any k-sparse signal $\boldsymbol{v} \in \mathbb{R}^d$, we have

$$(1-\delta_k)\|\boldsymbol{v}\|_2^2 \leq \|\boldsymbol{ADv}\|_2^2 + \|\bar{\boldsymbol{D}}\boldsymbol{v}\|_2^2 \leq (1+\delta_k)\|\boldsymbol{v}\|_2^2. \tag{76}$$

Since $\bar{\boldsymbol{D}}\boldsymbol{D}^\top = \boldsymbol{0}$ (see (75)), using the techniques for deriving (74), one can obtain

$$\sum_{i=1}^N \lambda_i(\|\bar{\boldsymbol{D}}\boldsymbol{v_i}\|_2^2 - \|\bar{\boldsymbol{D}}\bar{\boldsymbol{v}}_{\boldsymbol{i}}\|_2^2) = 4\delta_{tk}(\bar{\boldsymbol{D}}\boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h})^\top \bar{\boldsymbol{D}}\boldsymbol{D}^\top \boldsymbol{h} = 0.$$

Then,

$$\begin{aligned}
\sum_{i=1}^N \lambda_i(\|\boldsymbol{AD}\boldsymbol{v_i}\|_2^2 - \|\boldsymbol{AD}\bar{\boldsymbol{v}}_{\boldsymbol{i}}\|_2^2) =& \sum_{i=1}^N \lambda_i(\|\boldsymbol{AD}\boldsymbol{v_i}\|_2^2 + \|\bar{\boldsymbol{D}}\boldsymbol{v_i}\|_2^2) - \sum_{i=1}^N \lambda_i(\|\boldsymbol{AD}\bar{\boldsymbol{v}}_{\boldsymbol{i}}\|_2^2 + \|\bar{\boldsymbol{D}}\bar{\boldsymbol{v}}_{\boldsymbol{i}}\|_2^2) \\
\overset{(a)}{\geq}& \sum_{i=1}^N \lambda_i \Big((1-\delta_{tk})\|\boldsymbol{v_i}\|_2^2 - (1+\delta_{tk})\|\bar{\boldsymbol{v}}_{\boldsymbol{i}}\|_2^2 \Big) \\
\overset{(b)}{=}& \sum_{i=1}^N \lambda_i \Big(2\delta_{tk}(1-\delta_{tk}^2)\|\boldsymbol{D}_{T\cup T_1}^\top \boldsymbol{h}\|_2^2 - 2\delta_{tk}^3\|\boldsymbol{u}_i\|_2^2 \Big)
\end{aligned}$$

$$\overset{(c)}{=}2\delta_{tk}(1-\delta_{tk}^2)\|\boldsymbol{D}_{T\cup T_1}^\top\boldsymbol{h}\|_2^2-2\delta_{tk}^3\sum_{i=1}^N(\lambda_i\|\boldsymbol{u}_i\|_2^2)$$
$$\overset{(d)}{\geq}2\delta_{tk}(1-\delta_{tk}^2)\|\boldsymbol{D}_{T\cup T_1}^\top\boldsymbol{h}\|_2^2-2\delta_{tk}^3\frac{\|\boldsymbol{D}_{T^c}^\top\boldsymbol{h}\|_1^2}{(t-1)k},$$

where (a) is from (76) and the fact that both $\boldsymbol{v_i}$ and $\bar{\boldsymbol{v}}_{\boldsymbol{i}}$ are tk-sparse for $1\leq i\leq N$ (see the first sentence under (73)), (b) follows from (72) and (73), (c) is because of (14) and (d) is from (14) and (71).

Then, by (74), we have

$$2\delta_{tk}(1-\delta_{tk}^2)\|\boldsymbol{D}_{T\cup T_1}^\top\boldsymbol{h}\|_2^2-2\delta_{tk}^3\frac{\|\boldsymbol{D}_{T^c}^\top\boldsymbol{h}\|_1^2}{(t-1)k}\leq 4\delta_{tk}\sqrt{1+\delta_{tk}}\|\boldsymbol{D}_{T\cup T_1}^\top\boldsymbol{h}\|_2\|\boldsymbol{A}\boldsymbol{h}\|_2,$$

which is equivalent to

$$(1-\delta_{tk}^2)\|\boldsymbol{D}_{T\cup T_1}^\top\boldsymbol{h}\|_2^2-(2\sqrt{1+\delta_{tk}}\|\boldsymbol{A}\boldsymbol{h}\|_2)\|\boldsymbol{D}_{T\cup T_1}^\top\boldsymbol{h}\|_2-\frac{\delta_{tk}^2}{(t-1)k}\|\boldsymbol{D}_{T^c}^\top\boldsymbol{h}\|_1^2\leq 0.$$

Therefore,

$$\begin{aligned}\|\boldsymbol{D}_{T\cup T_1}^\top\boldsymbol{h}\|_2\leq&\Big[2\sqrt{1+\delta_{tk}}\|\boldsymbol{A}\boldsymbol{h}\|_2+\Big((2\sqrt{1+\delta_{tk}}\|\boldsymbol{A}\boldsymbol{h}\|_2)^2+4\delta_{tk}^2(1-\delta_{tk}^2)\frac{\|\boldsymbol{D}_{T^c}^\top\boldsymbol{h}\|_1^2}{(t-1)k}\Big)^{\frac{1}{2}}\Big]\frac{1}{2(1-\delta_{tk}^2)}\\
\leq&\frac{\delta_{tk}}{\sqrt{(1-\delta_{tk}^2)(t-1)}}\frac{\|\boldsymbol{D}_{T^c}^\top\boldsymbol{h}\|_1}{\sqrt{k}}+\frac{2}{(1-\delta_{tk})\sqrt{1+\delta_{tk}}}\|\boldsymbol{A}\boldsymbol{h}\|_2,\end{aligned}$$

where the last inequality is based on the fact that $\sqrt{f^2+g^2}\leq f+g$ for $f,g\geq 0$. Then, by (16), (68) holds which implies (15) holds.

If tk is not an integer, we define $t'=\frac{\lceil tk\rceil}{k}$, then $t'k$ is an integer and $\delta_{t'k}=\delta_{tk}<1$. From the above proof and Definition 1, one can see that (15) holds in this case, thus it holds no matter whether tk is an integer or not. □

Appendix B
Proof of Lemma 3

Proof. To simplify notation, the right-hand side of (19) is denoted by η, i.e.,

$$\eta=a\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_1+b\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1+\beta. \tag{77}$$

We first show that (21) holds when $c>0$ is an integer. Define the index sets

$$S_1=\left\{i\in S^c:|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}(i)|>\frac{\eta}{c}\right\}, \tag{78}$$
$$S_2=\left\{i\in S^c:|\boldsymbol{D}_{S^c}^\top\boldsymbol{h}(i)|\leq\frac{\eta}{c}\right\}. \tag{79}$$

Then it is easy to see

$$S_1\cap S_2=\emptyset\text{ and }S^c=S_1\cup S_2, \tag{80}$$

which imply

$$\boldsymbol{D}^\top\boldsymbol{h}=\boldsymbol{D}_S^\top\boldsymbol{h}+\boldsymbol{D}_{S^c}^\top\boldsymbol{h}=\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}+\boldsymbol{D}_{S_2}^\top\boldsymbol{h}. \tag{81}$$

Moreover, by the definition of S_1, $S_1\cap S=\emptyset$ which implies that $\|\boldsymbol{D}_S^\top\boldsymbol{h}\|_2\leq\|\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}\|_2$. Thus to show (21), it is suffices to show that

$$\|\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}\|_2\leq\Big[\frac{(2as\delta_{s+c}+\sqrt{2\delta_{s+c}s\Delta})(b\|\boldsymbol{D}_{T^c}^\top\boldsymbol{x}\|_1+\beta)}{2\sqrt{s}}+\sqrt{c(c+a^2s)(1+\delta_{s+c})}\|\boldsymbol{A}\boldsymbol{h}\|_2\Big]\frac{1}{\Delta}, \tag{82}$$

where Δ is defined in (22). In order to use Lemma 1 to show (82), we first prove that

$$|S_1| < c. \tag{83}$$

If $S_1 = \emptyset$, then $|S_1| = 0$ and hence (83) holds. Otherwise, from (78), it follows that

$$\|\boldsymbol{D}_{S_1}^\top \boldsymbol{h}\|_1 \overset{(a)}{>} |S_1|\frac{\eta}{c} \overset{(b)}{\geq} |S_1|\frac{\|\boldsymbol{D}_{S^c}^\top \boldsymbol{h}\|_1}{c} \overset{(c)}{=} |S_1|\frac{\|\boldsymbol{D}_{S_1\cup S_2}^\top \boldsymbol{h}\|_1}{c} \overset{(d)}{\geq} |S_1|\frac{\|\boldsymbol{D}_{S_1}^\top \boldsymbol{h}\|_1}{c},$$

where (a) is from (78), (b) is due to the assumption (19) and (77), (c) and (d) follow from (80). Thus (83) holds.

By (79) and (80), we have

$$\|\boldsymbol{D}_{S_2}^\top \boldsymbol{h}\|_\infty \leq \frac{\eta}{c}$$

and

$$\|\boldsymbol{D}_{S_2}^\top \boldsymbol{h}\|_1 = \|\boldsymbol{D}_{S^c}^\top \boldsymbol{h}\|_1 - \|\boldsymbol{D}_{S_1}^\top \boldsymbol{h}\|_1 \leq \eta - \frac{\eta|S_1|}{c} = (c - |S_1|)\frac{\eta}{c},$$

where the inequality is from (19), (77) and (78).

Using Lemma 1 with $\alpha = \frac{\eta}{c}$, $s = c - |S_1|$ (note that this s is different from the s which gives an upper bound on $|S|$ as assumed in this lemma) and $\boldsymbol{v} = \boldsymbol{D}_{S_2}^\top \boldsymbol{h}$, then $\boldsymbol{D}_{S_2}^\top \boldsymbol{h}$ can be expressed as

$$\boldsymbol{D}_{S_2}^\top \boldsymbol{h} = \sum_{i=1}^{N} \lambda_i \boldsymbol{u_i} \tag{84}$$

for certain positive integer N, where λ_i satisfies (14) and

$$\boldsymbol{u_i} \in U\left(\frac{\eta}{c}, c - |S_1|, \boldsymbol{D}_{S_2}^\top \boldsymbol{h}\right) \tag{85}$$

for $1 \leq i \leq N$. Moreover, for $1 \leq i \leq N$, we have

$$\begin{aligned}
\|\boldsymbol{u_i}\|_2^2 \leq \|\boldsymbol{u_i}\|_0 \|\boldsymbol{u_i}\|_\infty^2 &\overset{(a)}{\leq} (c - |S_1|)\left(\frac{\eta}{c}\right)^2 \leq \frac{\eta^2}{c} \\
&\overset{(b)}{=} \frac{1}{c}\Big[a^2\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_1^2 + (b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta)^2 + 2a\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_1 (b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta)\Big] \\
&\overset{(c)}{\leq} \frac{1}{c}\Big[a^2 s\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2^2 + (b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta)^2 + 2a\sqrt{s}\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2 (b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta)\Big] \\
&\overset{(d)}{\leq} \frac{1}{c}\Big[a^2 s\|\boldsymbol{D}_{S\cup S_1}^\top \boldsymbol{h}\|_2^2 + (b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta)^2 + 2a\sqrt{s}\|\boldsymbol{D}_{S\cup S_1}^\top \boldsymbol{h}\|_2 (b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta)\Big] \\
&\overset{(e)}{=} \frac{1}{c}\Big(a^2 sX^2 + sP^2 + 2asXP\Big),
\end{aligned} \tag{86}$$

where (a) is from (13) and (85), (b) follows from (77), (c) is from the Cauchy-Schwarz inequality, (d) is due to the fact that $\|\boldsymbol{D}_S^\top \boldsymbol{h}\|_2 \leq \|\boldsymbol{D}_{S\cup S_1}^\top \boldsymbol{h}\|_2$ which can be seen from (80) and (e) is because we denote

$$X = \|\boldsymbol{D}_{S\cup S_1}^\top \boldsymbol{h}\|_2, \qquad P = \frac{b\|\boldsymbol{D}_{T^c}^\top \boldsymbol{x}\|_1 + \beta}{\sqrt{s}}. \tag{87}$$

To simplify notation, define the vectors

$$\boldsymbol{\gamma_i} = \boldsymbol{D}_{S\cup S_1}^\top \boldsymbol{h} + \mu \boldsymbol{u_i}, \qquad 1 \leq i \leq N, \tag{88}$$

$$\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}} = (\frac{1}{2} - \mu)\boldsymbol{D}_{S\cup S_1}^\top \boldsymbol{h} - \frac{1}{2}\mu \boldsymbol{u_i}, \qquad 1 \leq i \leq N, \tag{89}$$

$$\boldsymbol{\omega} = (1 - \mu)\boldsymbol{D}_{S\cup S_1}^\top \boldsymbol{h} + \mu \boldsymbol{D}^\top \boldsymbol{h}, \tag{90}$$

where the constant $\mu = \frac{\sqrt{c^2+a^2sc}-c}{a^2s}$. Thus

$$\begin{aligned}\mu^2-\mu &= \mu^2\Big(1-\frac{1}{\mu}\Big) = \mu^2\Big(1-\frac{a^2s}{\sqrt{c^2+a^2sc}-c}\Big)\\ &= \mu^2\Big(1-\frac{a^2s(\sqrt{c^2+a^2sc}+c)}{a^2sc}\Big) = -\frac{\mu^2}{c}\sqrt{c^2+a^2sc}\end{aligned} \tag{91}$$

and

$$\frac{1}{2}-\mu = \frac{\mu^2}{2}\Big[\Big(\frac{1}{\mu}-1\Big)^2-1\Big] = \frac{\mu^2}{2}\Big[\Big(\frac{\sqrt{c^2+a^2sc}}{c}\Big)^2-1\Big] = \frac{\mu^2a^2s}{2c}. \tag{92}$$

Then, by (88)-(90), one has the following identities

$$\begin{aligned}\boldsymbol{\omega}-\frac{1}{2}\boldsymbol{\gamma_i} &= (1-\mu)\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}+\mu\boldsymbol{D}^\top\boldsymbol{h}-\frac{1}{2}(\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}+\mu\boldsymbol{u_i})\\ &= (\frac{1}{2}-\mu)\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}-\frac{1}{2}\mu\boldsymbol{u_i}+\mu\boldsymbol{D}^\top\boldsymbol{h} = \bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}+\mu\boldsymbol{D}^\top\boldsymbol{h}.\end{aligned} \tag{93}$$

Moreover,

$$\begin{aligned}\sum_{i=1}^N\lambda_i\boldsymbol{\gamma_i} &= \sum_{i=1}^N\lambda_i(\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}+\mu\boldsymbol{u_i}) = \boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}+\mu\boldsymbol{D}_{S_2}^\top\boldsymbol{h}\\ &= \boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}+\mu(\boldsymbol{D}^\top\boldsymbol{h}-\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}) = \boldsymbol{\omega},\end{aligned} \tag{94}$$

where the first equality is from (88), the second equality is from (84) and $\sum_{i=1}^N\lambda_i=1$ in (14), and the last two equalities are respectively from (81) and (90).

By (13) and (85), $\|\boldsymbol{u_i}\|_0\le c-|S_1|$ which combing with (83) and the assumption that $|S|\le s$, one can see that $\boldsymbol{\gamma_i}$ and $\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}$ are $(s+c)$-sparse for $1\le i\le N$. Thus

$$\begin{aligned}&2\sum_{i=1}^N\lambda_i\langle\boldsymbol{AD}\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}},\mu\boldsymbol{ADD}^\top\boldsymbol{h}\rangle\\ \overset{(a)}{=}&2\sum_{i=1}^N\lambda_i\langle\boldsymbol{AD}[(\frac{1}{2}-\mu)\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}-\frac{1}{2}\mu\boldsymbol{u_i}],\mu\boldsymbol{ADD}^\top\boldsymbol{h}\rangle\\ \overset{(b)}{=}&2(\frac{1}{2}-\mu)\mu\langle\boldsymbol{ADD}_{S\cup S_1}^\top\boldsymbol{h},\boldsymbol{Ah}\rangle-\mu^2\langle\boldsymbol{AD}(\sum_{i=1}^N\lambda_i\boldsymbol{u_i}),\boldsymbol{Ah}\rangle\\ \overset{(c)}{=}&(1-2\mu)\mu\langle\boldsymbol{ADD}_{S\cup S_1}^\top\boldsymbol{h},\boldsymbol{Ah}\rangle-\mu^2\langle\boldsymbol{ADD}_{S_2}^\top\boldsymbol{h},\boldsymbol{Ah}\rangle\\ \overset{(d)}{=}&(1-2\mu)\mu\langle\boldsymbol{ADD}_{S\cup S_1}^\top\boldsymbol{h},\boldsymbol{Ah}\rangle-\mu^2\langle\boldsymbol{AD}(\boldsymbol{D}^\top\boldsymbol{h}-\boldsymbol{D}_{S\cup S_1}^\top\boldsymbol{h}),\boldsymbol{Ah}\rangle\\ =&(1-2\mu)\mu\langle\boldsymbol{ADD}_{S\cup S_1}^\top\boldsymbol{h},\boldsymbol{Ah}\rangle-\mu^2(\|\boldsymbol{Ah}\|_2^2-\langle\boldsymbol{ADD}_{S\cup S_1}^\top\boldsymbol{h},\boldsymbol{Ah}\rangle)\\ =&(1-\mu)\mu\langle\boldsymbol{ADD}_{S\cup S_1}^\top\boldsymbol{h},\boldsymbol{Ah}\rangle-\mu^2\|\boldsymbol{Ah}\|_2^2,\end{aligned} \tag{95}$$

where (a) is from (89), (b) is due to $\sum_{i=1}^N\lambda_i=1$ in (14) and the assumption that $\boldsymbol{DD}^T=\boldsymbol{I}$, (c) and (d) follow from (84) and (81), respectively.

Let $\bar{\boldsymbol{D}}\in\mathbb{R}^{(d-n)\times d}$ be the orthogonal compliment of $\boldsymbol{D}$, then

$$\begin{aligned}&\sum_{i=1}^N\lambda_i\Big(\Big\|\boldsymbol{AD}\Big(\boldsymbol{\omega}-\frac{1}{2}\boldsymbol{\gamma_i}\Big)\Big\|_2^2+\Big\|\bar{\boldsymbol{D}}\Big(\boldsymbol{\omega}-\frac{1}{2}\boldsymbol{\gamma_i}\Big)\Big\|_2^2\Big)\\ \overset{(a)}{=}&\sum_{i=1}^N\lambda_i\Big(\|\boldsymbol{AD}(\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}+\mu\boldsymbol{D}^\top\boldsymbol{h})\|_2^2+\|\bar{\boldsymbol{D}}(\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}+\mu\boldsymbol{D}^\top\boldsymbol{h})\|_2^2\Big)\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{=}\sum_{i=1}^{N}\lambda_i\Big(\|\boldsymbol{AD}\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}\|_2^2+\|\bar{\boldsymbol{D}}\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}\|_2^2\Big)+(\sum_{i=1}^{N}\lambda_i)\mu^2\|\boldsymbol{Ah}\|_2^2+\mu(1-\mu)\langle\boldsymbol{ADD}_{S\cup S_1}^{\top}\boldsymbol{h},\boldsymbol{Ah}\rangle-\mu^2\|\boldsymbol{Ah}\|_2^2\\
&\overset{(c)}{\leq}(1+\delta_{s+c})\sum_{i=1}^{N}\lambda_i\|\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}\|_2^2+\mu(1-\mu)\sqrt{1+\delta_{s+c}}\|\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}\|_2\|\boldsymbol{Ah}\|_2\\
&\overset{(d)}{=}(1+\delta_{s+c})\sum_{i=1}^{N}\lambda_i\Big\|(\frac{1}{2}-\mu)\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}-\frac{1}{2}\mu\mathbf{u}_i\Big\|_2^2+\mu(1-\mu)\sqrt{1+\delta_{s+c}}\|\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}\|_2\|\boldsymbol{Ah}\|_2\\
&\overset{(e)}{=}(1+\delta_{s+c})\Big((\frac{1}{2}-\mu)^2\|\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}\|_2^2+\frac{\mu^2}{4}\sum_{i=1}^{N}\lambda_i\|\boldsymbol{u}_{\boldsymbol{i}}\|_2^2\Big)+\mu(1-\mu)\sqrt{1+\delta_{s+c}}\|\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}\|_2\|\boldsymbol{Ah}\|_2\\
&\overset{(f)}{=}(1+\delta_{s+c})\Big((\frac{1}{2}-\mu)^2X^2+\frac{\mu^2}{4}\sum_{i=1}^{N}\lambda_i\|\boldsymbol{u}_{\boldsymbol{i}}\|_2^2\Big)+\mu(1-\mu)\sqrt{1+\delta_{s+c}}\|\boldsymbol{Ah}\|_2X,
\end{aligned}
\tag{96}
$$

where we use (93) in (a), (b) is from (95) and the first equality in (75), (c) is due to that $\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}$ is $(s+c)$-sparse, the Cauchy-Schwarz inequality and (76), (d) follows from (89), (e) is from $\sum_{i=1}^{N}\lambda_i=1$ in (14), (13), (80), (84) and (85) and (f) follows from (87). Moreover, we have

$$
\begin{aligned}
\sum_{i=1}^{N}\frac{\lambda_i}{4}\Big(\|\boldsymbol{AD}\boldsymbol{\gamma}_{\boldsymbol{i}}\|_2^2+\|\bar{\boldsymbol{D}}\boldsymbol{\gamma}_{\boldsymbol{i}}\|_2^2\Big)=&\sum_{i=1}^{N}\frac{\lambda_i}{4}\Big(\|\boldsymbol{AD}(\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}+\mu\boldsymbol{u}_{\boldsymbol{i}})\|_2^2+\|\bar{\boldsymbol{D}}(\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}+\mu\boldsymbol{u}_{\boldsymbol{i}})\|_2^2\Big)\\
\geq&(1-\delta_{s+c})\sum_{i=1}^{N}\frac{\lambda_i}{4}\|\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}+\mu\boldsymbol{u}_{\boldsymbol{i}}\|_2^2\\
=&\frac{1-\delta_{s+c}}{4}\Big(\|\boldsymbol{D}_{S\cup S_1}^{\top}\boldsymbol{h}\|_2^2+\mu^2\sum_{i=1}^{N}\lambda_i\|\boldsymbol{u}_{\boldsymbol{i}}\|_2^2\Big)\\
=&\frac{1-\delta_{s+c}}{4}\Big(X^2+\mu^2\sum_{i=1}^{N}\lambda_i\|\boldsymbol{u}_{\boldsymbol{i}}\|_2^2\Big),
\end{aligned}
\tag{97}
$$

where the inequality is from the fact that $\boldsymbol{\gamma}_{\boldsymbol{i}}$ is $(s+c)$-sparse and (76), the second equality follows from (14), (13), (80) and (85) and the last equality is by (87).

By (94), we can get the following equations

$$
\begin{aligned}
&\sum_{i=1}^{N}\lambda_i\Big(\Big\|\boldsymbol{AD}\Big(\boldsymbol{\omega}-\frac{1}{2}\boldsymbol{\gamma}_{\boldsymbol{i}}\Big)\Big\|_2^2-\frac{1}{4}\|\boldsymbol{AD}\boldsymbol{\gamma}_{\boldsymbol{i}}\|_2^2\Big)\\
=&\sum_{i=1}^{N}\lambda_i\Big(\|\boldsymbol{AD}\boldsymbol{\omega}\|_2^2-\langle\boldsymbol{AD}\boldsymbol{\omega},\boldsymbol{AD}\boldsymbol{\gamma}_{\boldsymbol{i}}\rangle\Big)=\|\boldsymbol{AD}\boldsymbol{\omega}\|_2^2-\langle\boldsymbol{AD}\boldsymbol{\omega},\boldsymbol{AD}(\sum_{i=1}^{N}\lambda_i\boldsymbol{\gamma}_{\boldsymbol{i}})\rangle=0,
\end{aligned}
\tag{98}
$$

where the last two equalities are from $\sum_{i=1}^{N}\lambda_i=1$ in (14) and (94), respectively.

Furthermore, by using the techniques for deriving (98), we can show that

$$
\sum_{i=1}^{N}\lambda_i\Big(\Big\|\bar{\boldsymbol{D}}\Big(\boldsymbol{\omega}-\frac{1}{2}\boldsymbol{\gamma}_{\boldsymbol{i}}\Big)\Big\|_2^2-\frac{1}{4}\|\bar{\boldsymbol{D}}\boldsymbol{\gamma}_{\boldsymbol{i}}\|_2^2\Big)=0.
\tag{99}
$$

By (97)-(99), one has

$$
\sum_{i=1}^{N}\lambda_i\Big(\Big\|\boldsymbol{AD}\Big(\boldsymbol{\omega}-\frac{1}{2}\boldsymbol{\gamma}_{\boldsymbol{i}}\Big)\Big\|_2^2+\Big\|\bar{\boldsymbol{D}}\Big(\boldsymbol{\omega}-\frac{1}{2}\boldsymbol{\gamma}_{\boldsymbol{i}}\Big)\Big\|_2^2\Big)\geq\frac{1-\delta_{s+c}}{4}\Big(X^2+\mu^2\sum_{i=1}^{N}\lambda_i\|\boldsymbol{u}_{\boldsymbol{i}}\|_2^2\Big).
\tag{100}
$$

Therefore, one obtains

$$
\begin{aligned}
0 \overset{(a)}{\leq} & (1+\delta_{s+c})\Big((\frac{1}{2}-\mu)^2X^2+\frac{\mu^2}{4}\sum_{i=1}^{N}\lambda_i\|\boldsymbol{u_i}\|_2^2\Big)+\mu(1-\mu)\sqrt{1+\delta_{s+c}}\|\boldsymbol{Ah}\|_2X \\
& -\frac{1-\delta_{s+c}}{4}\Big(X^2+\mu^2\sum_{i=1}^{N}\lambda_i\|\boldsymbol{u_i}\|_2^2\Big) \\
= & \Big((\mu^2-\mu)+(\frac{1}{2}-\mu+\mu^2)\delta_{s+c}\Big)X^2+\mu(1-\mu)\sqrt{1+\delta_{s+c}}\|\boldsymbol{Ah}\|_2X+\frac{\mu^2\delta_{s+c}}{2}\sum_{i=1}^{N}\lambda_i\|\boldsymbol{u_i}\|_2^2 \\
\overset{(b)}{\leq} & \Big((\mu^2-\mu)+(\frac{1}{2}-\mu+(1+\frac{a^2s}{2c})\mu^2)\delta_{s+c}\Big)X^2+\Big(\mu(1-\mu)\sqrt{1+\delta_{s+c}}\|\boldsymbol{Ah}\|_2+\frac{as\mu^2\delta_{s+c}}{c}P\Big)X \\
& +\frac{\mu^2\delta_{s+c}s}{2c}P^2 \\
\overset{(c)}{=} & -\frac{\mu^2}{c}\Big[(c+a^2s)\Big(\sqrt{\frac{c}{c+a^2s}}-\delta_{s+c}\Big)X^2-\Big(\sqrt{c^2+a^2sc}\sqrt{1+\delta_{s+c}}\|\boldsymbol{Ah}\|_2+as\delta_{s+c}P\Big)X-\frac{\delta_{s+c}s}{2}P^2\Big] \\
\overset{(d)}{=} & -\frac{\mu^2}{c}\Big[\Delta X^2-\Big(\sqrt{(c^2+a^2sc)(1+\delta_{s+c})}\|\boldsymbol{Ah}\|_2+as\delta_{s+c}P\Big)X-\frac{\delta_{s+c}s}{2}P^2\Big],
\end{aligned} \tag{101}
$$

where (a) is from (96) and (100), (b) is due to (86), (c) is from (91) and (92) and (d) follows from (22). Using (20) and (101), one obtains

$$
\begin{aligned}
X \leq & \frac{1}{2\Delta}\Big[\sqrt{(c^2+a^2sc)(1+\delta_{s+c})}\|\boldsymbol{Ah}\|_2+as\delta_{s+c}P \\
& +\Big(\Big(\sqrt{(c^2+a^2sc)(1+\delta_{s+c})}\|\boldsymbol{Ah}\|_2+as\delta_{s+c}P\Big)^2+2\delta_{s+c}s\Delta P^2\Big)^{\frac{1}{2}}\Big] \\
\leq & \frac{2as\delta_{s+c}+\sqrt{2\delta_{s+c}s\Delta}}{2\Delta}P+\frac{\sqrt{(c^2+a^2sc)(1+\delta_{s+c})}}{\Delta}\|\boldsymbol{Ah}\|_2,
\end{aligned}
$$

where the last inequality uses $\sqrt{f^2+g^2}\leq f+g$ with $f,g\geq 0$. Since $S_1\subseteq S^c$, Then, by (87), (82) holds which implies (21) holds when $c>0$ is an integer.

When the constant c is not an integer, taking $c^{'}=\lceil c\rceil$, then $c^{'}>c$ and $(s+c^{'})$ is an integer, thus

$$
\delta_{s+c^{'}}=\delta_{s+c}<\sqrt{\frac{c}{c+a^2s}}<\sqrt{\frac{c^{'}}{c^{'}+a^2s}}.
$$

Hence, by using the above proof by working on $\delta_{s+c^{'}}$, one can show that (21) holds. □

REFERENCES

[1] E. J. 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.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489-509, 2006.

[3] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877-905, 2007.

[4] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus Math.*, vol. 346, no. 9-10, pp. 589-592, 2008.

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

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

[7] D. L. Donoho, I. Drori, Y. Tsaig, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory.*, vol. 58, no. 2, pp. 1094-1121, 2012.

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

[9] T. Cai and A. Zhang, "Sharp RIP bound for sparse signal and low-rank matrix recovery," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 1, pp. 74-93, 2013.

[10] T. Cai and A. Zhang, "Spares representation of a polytope and recovery of sparse signals and low-rank matrices," *IEEE Trans. Inform. Theory*, vol. 60, no. 1, pp. 122-132, 2014.

[11] 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.

[12] J. Wen, D. Li, and F. Zhu, "Stable recovery of sparse signals via ℓ_p-minimization," *Appl. Comput. Harmon. Anal.*, vol. 38, no. 1, pp. 161-176, 2015.

[13] J. F. Determe, J. Louveaux, L. Jacques, et al., "On the noise robustness of simultaneous orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 864-875, 2017.

[14] L. Yu, , G. Zheng and J.-P. Barbot, "Dynamical sparse recovery with finite-time convergence," *IEEE Trans. Signal Process.*, vol. 65, no. 23, pp. 6146-6157, 2017.

[15] J. Wen, Z. Zhou, J. Wang, et al., "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1370-1382, 2017.

[16] J. Wen, J. Wang, and Q. Zhang, "Nearly Optimal Bounds for Orthogonal Least Squares," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5347-5356, 2017.

[17] R. Zhang and S. Li, "A proof of conjecture on restricted isometry property constants δ_{tk} $(0 < t < \frac{4}{3})$," to appear in *IEEE Trans. Inform. Theory*, 2017.

[18] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861-899, 2006.

[19] E. J. Candès, and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Comm. Pure Appl. Math.*, vol. 57, pp. 219-266, 2004.

[20] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, pp. 34-81, 2009.

[21] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, pp. 129-159, 2001.

[22] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59-73, 2011.

[23] J. Lin, S. Li, and Y. Shen, "New bounds for resticted isometry constants with coherent tight frames," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 611-621, 2013.

[24] L. Daubechies, "Ten lectures on wavelets//CBMS-NSF Regional Conf Ser Appl Math," *Philadelphia: SIAM,*1992.

[25] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203-4215, 2005.

[26] Y. Liu, T. Mi, and S. Li, "Compressed sensing with general frames via optimal-dual-based ℓ_1-analysis," *IEEE Trans. Inform. Theory*, vol. 58, no. 7, pp. 4201-4214, 2012.

[27] S. Li and J. Lin, "Compressed sensing with coherent tight frame via ℓ_q-minimination for $0 < q \leq 1$," *Inverse Probl. Imaging*, vol. 8, no. 3, pp. 761-777, 2014.

[28] R. Zhang and S. Li, "Optimal D-RIP bounds in compressed sensing," *Acta Mathematica Sinica(English Series)*, vol. 31, no. 5, pp. 755-766, 2015.

[29] D. M. Higdon, J. E. Bowsher, V. E. Johnson, et al., "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Trans. Med. Imaging*, vol. 16, no. 5 pp. 516-526, 1997.

[30] M. Elad, J. L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis," *Appl. Comput. Harmon. Anal.*, vol. 19, pp. 340-358, 2005.

[31] S. Farsiu, M. Elad, and P. Milanfar "Multi-frame demosaicing and super-resolution of color images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 141-159, 2006.

[32] H. Zhang, M. Yan, and W. Yin, "One condition for solution uniqueness and robustness of both ℓ_1-synthesis and ℓ_1-analysis minimizations," *Adv. Comput. Math.*, vol. 42, pp. 1381-1399, 2016.

[33] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947-986, 2007.

[34] J. Lin and S. Li, "Sparse recovery with coherent tight frame via analysis dantzig selector and analysis lasso," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 126-139, 2014.

[35] T. Zhao, Y. C. Eldar, A. Beck, et al., "Smoothing and decomposition for analysis sparse recovery," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1762-1774, 2014.

[36] Y. Shen, B. Han, and E. Braverman, "Stable recovery of analysis based approaches," *Appl. Comput. Harmon. Anal.*, vol. 39, no. 1, pp. 161-172, 2015.

[37] S. Boyd and L. Vandenberghe, "Convex optimization," *U.K.: Cambridge Univ. Press,* 2004.

[38] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345-2356, 2010.

[39] S. Boyd, N. Parikh, E. Chu, et al., "Distributed optimization and statistical learning via alternating direction method of multipliers," *in Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1-122, 2010.

[40] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "Exact signal recovery from corrupted measurements through the pursuit of justice," *Proc. Asilomar Conf. on Signals, Systems and Computers*, pp. 1556-1560, 2009.

[41] S. Foucart, "Stability and robustness of ℓ_1-minimizations with Weibull matrices and redundant dictionaries," *Linear Algebra and Its Applications*, vol. 441, pp. 4-21, 2014.