ANALYSIS NON-SPARSE RECOVERY FOR RELAXED ALASSO

HANG XU AND SONG LI

School of Mathematical Sciences, Zhejiang University
Hangzhou 310027, China

ABSTRACT. This paper considers recovery of signals that are corrupted with noise. We focus on a novel model which is called relaxed ALASSO (RALASSO) model introduced by Z. Tan et al. (2014). Compared to the well-known ALASSO, RALASSO can be solved better in practice. Z. Tan et al. (2014) used the D-RIP to characterize the sparse or approximately sparse solutions for RALASSO when the D-RIP constant \( \delta_{2k} < 0.1907 \), where the solution is sparse or approximately sparse in terms of a tight frame \( D \). However, their estimate of error bound for solution heavily depends on the term \( \|D^*D\|_{1,1}^{-1} \). Besides, compared to other works on signals recovering from ALASSO, the condition \( \delta_{2k} < 0.1907 \) is even stronger. Based on the RALASSO model, we use new methods to get a better estimate of error bound and give a weaker sufficient condition in this article for the inadequacies of the results by Z. Tan et al. (2014). One of the result of this paper is to use another method called the robust \( \ell_2 \)-D-Null Space Property to obtain the sparse or non-sparse solution of RALASSO and give the error estimation of RALASSO, where we eliminate the term \( \|D^*D\|_{1,1}^{-1} \) in the constants. Another result of the paper is to utilize the D-RIP to obtain a new condition \( \delta_{2k} < 0.3162 \) which is weaker than the condition \( \delta_{2k} < 0.1907 \). To some extent, RALASSO is equivalent to ALASSO and the condition is also weaker than the similar one \( \delta_{2k} < 0.25 \) by J. Lin, and S. Li (2014) and \( \delta_{2k} < 0.25 \) by Y. Xia, and S. Li (2016).

1. Introduction. For the last fifteen years, the recovery of sparse signals has been a very active area of recent research in mathematics, engineering and computer science[7, 8, 11] and triggered an enormous amount of research activities in radar systems[16] and medical imaging[18].

In compressed sensing, the linear model is given by

\[
b = Ax + w,
\]

where \( A \in \mathbb{R}^{m \times n} \) is the sensing matrix \( m < n \), \( b \in \mathbb{R}^m \) is a vector of noisy observations and \( w \in \mathbb{R}^m \) is a noise vector. The purpose is to reconstruct the unknown signal \( x \) based on \( A \) and \( b \). Generally speaking, we have \( m < n \) and cannot solve the ill-posed inverse problem without any other assumptions.

When \( x \) is a sparse vector, under advisable conditions on \( A \), we can utilize the \( \ell_0 \)-minimization to search for the sparsest vector as the solution of this problem. Based on the fact that handling the \( \ell_0 \)-minimization is NP-hard [20], we turn to
consider the $\ell_1$-minimization as a convex hull of $\ell_0$-minimization. A famous recovery algorithm is the LASSO estimator [24]. However, in practical examples, there are enormous signals of interest which are not sparse in an orthonormal basis. Often, sparsity is expressed not in terms of an orthogonal basis but in terms of a redundant dictionary $D \in \mathbb{R}^{n \times p}$ [13]. After that, the signal can be expressed as

$$x = Df,$$

where $f \in \mathbb{R}^p (p \geq n)$. Then $y = ADf$, where $f$ is a sparse vector [10, 14, 22]. In the remainder of this paper, we assume that $D$ is a tight frame.

**Definition 1.1** (tight frame[26]). $D \in \mathbb{R}^{n \times p} (n \leq p)$ forms a tight frame for $\mathbb{R}^n$, for all $x \in \mathbb{R}^n$, if

$$\sum_k |\langle x, D_k \rangle|^2 = \|x\|_2^2,$$

where $D_k (k = 1, ..., p)$ are the columns of $D$. In other words, if $D$ is a tight frame, we have $DD^* = I$.

When the matrix $AD$ satisfies the restricted isometry property (RIP), the method can stably recover a sparse signal $f$ where the frame $D$ is required to have columns that are extremely uncorrelated[22]. But if the columns of $D$ are highly correlated, $AD$ may not satisfy the RIP generally[22, 9]. It turns out to have good results along with the $\ell_1$-analysis model which finds the estimator $x$ directly by solving an $\ell_1$-minimization problem[9, 19], referred to as analysis basis pursuit (ABP). Combining the LASSO estimator, the recovery problem can be formulated as

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|b - Ax\|_2^2 + \lambda \|D^* x\|_1.$$  \hspace{1cm} (1)

We call the model (1) as analysis LASSO (ALASSO)[12, 25], which is equivalent to ABP [9, 19]. To ensure stable recovery, we need the matrix $A$ to satisfy the $D$-RIP [9].

**Definition 1.2** (D-RIP). Let $D$ be an $n \times p$ matrix. A matrix $A$ is said to obey the restricted isometry property adapted to $D$ of order $k$ with constant $\delta$ if for all $k$-sparse vectors $f \in \mathbb{R}^p$, we have

$$(1 - \delta)\|Df\|_2^2 \leq \|ADf\|_2^2 \leq (1 + \delta)\|Df\|_2^2.$$  \hspace{1cm} (2)

The $D$-RIP constant $\delta_k$ is defined as the smallest $\delta$ such that (2) holds.

In this paper, we pay attention to a novel model:

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^p} \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|z\|_1 + \frac{1}{2} \rho \|z - D^* x\|_2^2,$$  \hspace{1cm} (3)

which we call relaxed ALASSO (RALASSO)[23]. As the name implies, RALASSO is similar to ALASSO. If we choose $\rho \to \infty$, then the third item implies $z = D^* x$ and the problem is equivalent to ALASSO. But in general, RALASSO can be solved better in practice.

So far, ALASSO can be solved by two common methods: interior point methods[4] and alternating direction method of multipliers (ADMM) [5, 1]. Interior point methods work efficiently in the low-dimension case, but when the dimension grows, the rate of convergence of these techniques becomes very slow because they need solutions of linear systems. And the efficiency of ADMM relies highly on nice structure of the matrix $A$. Based on the knowledge of optimization, some scholars turn to
find fast algorithms to deal with high dimensional data when the matrix $A$ has any structure not just nice structure.

The fast iterative shrinkage thresholding algorithm (FISTA)[2] or a monotone version of FISTA (MFISTA)[3] are fast versions of first-order algorithms which are extensively used in dealing with large dimensional data and do not need $A$ to have nice structure. But it is obvious that these fast versions of first-order algorithms are hard to be directly applied to ALASSO because of the non-smooth term $\|D^*x\|_1$.

Given a closed proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, the proximal operator of $f$ is defined as

$$\text{Prox}_f(x) = \arg\min_{u \in \mathbb{R}^n} \{f(u) + \frac{1}{2}\|u - x\|_2^2\}.$$  

The proximal operator is commonly used to solve the following problem:

$$\min_{x \in \mathbb{R}^n} \{F(x) + G(x)\},$$

where $F : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable convex function, $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a closed proper convex function. The core of FISTA or MFISTA is still the proximal gradient method whose updates are

$$x_k = \text{Prox}_{\frac{1}{L}G}(x_{k-1} - \frac{1}{L}\nabla F(x_{k-1})).$$

Here, $L$ is the upper bound of Lipschitz constant of gradient $\nabla F$. Thus, directly applying MFISTA to ALASSO requires computing the proximal operator of $\lambda \|D^*x\|_1$.

However, considering RALASSO, the idea is to introduce an auxiliary variable $z$ so that we can avoid computing the proximal operator of $\lambda \|D^*x\|_1$. Let $F(x,z) = \frac{1}{2}\|Ax-b\|_2^2 + \frac{1}{2}\rho\|z-D^*x\|_2^2, G(z) = \lambda \|z\|_1$ and we only need to compute the proximal operator of $\lambda \|z\|_1$. It is obvious that computing the proximal operator of $\lambda \|D^*x\|_1$ is very difficult while it is easy to know the proximal operator of $\lambda \|z\|_1$. This is the reason why a decomposition-based MFISTA(DFISTA) is proposed to solve this analysis sparse recovery problem in [23].

In this paper, based on the RALASSO model, we use new methods to get a better estimate of error bound and give a weaker sufficient condition for the inadequacies of the results in [23]. One of the result of this paper is to use another method called the robust $\ell_2$ D-Null Space Property to obtain the sparse or non-sparse solution of RALASSO and give the error estimation of RALASSO, where we eliminate the term $\|D^*D\|_{1,1}$ in the constants. Another result of the paper is to utilize the D-RIP to obtain a new condition $\delta_{2k} < 0.3162$ which is weaker than the condition $\delta_{2k} < 0.1907$. It is easy to see that when $\rho \to \infty$, RALASSO is equivalent to ALASSO and the condition is also weaker than the similar one $\delta_{2k} < 0.25$ in [17] and $\delta_{2k} < 0.25$ in [26].

2. Notations and organization. The following notations are used throughout this paper. We use capital italic bold letters to represent matrices and lower-case italic bold letters to represent vectors. Let $D^*$ be the conjugate matrix of $D$. The set of indices of the nonzero entries of a vector $x$ is called the support of $x$ and denoted as $\text{supp}(x)$. Denote $T \subset \{1, 2, ..., n\}$ to be an index set and for a given matrix $D \in \mathbb{R}^{m \times n}$, $T^c$ is the complement of $T$, $D_T$ is the submatrix of $D$ formed from the columns of $D$ indexed by $T$, while setting all other columns to zero. Thus $D_T^*$ is denoted by the matrix that maintains the rows of $D^*$ indexed by $T$. And we use $x_T$ to represent the $i$-th element of $x$ and $A_i(A^T)$ to represent the $i$-th row(column) of $A$. For a vector $x$, the $\ell_1, \ell_2$ norms are denoted by $\|x\|_1, \|x\|_2$. 


\[ \|x\|_2 \text{ and denote } \|x\|_0 = |\text{supp}(x)|, \|x\|_\infty = \max_i |x_i|. \] For a matrix \( A \), \( \|A\|_2 \) denotes the spectral norm of \( A \) and \( \|A\|_{p,q} \) denotes the norm of \( A \) from \( \ell_p \) to \( \ell_q \), i.e., \( \|A\|_{p,q} = \max_x \frac{\|Ax\|_p}{\|x\|_p} \).

The paper is organized as follows. In Section 3, we give the main results and their Gaussian noise forms. Finally, in Section 4, we give the main lemmas used in the proofs and prove the main theorems.

3. Main results. In [23], the authors used the \( D\text{-RIP} \) to obtain the solution of RALASSO when \( \delta_{2k} < 0.1907 \). However, their estimate of error bound for solution heavily depends on \( \ell_1 \)-norm for this tight frame. Therefore, the error bound estimate for the solution of RALASSO is not optimal. Then we use another method called as the robust \( \ell_2 \) \( D\text{-Null Space Property} \) to obtain the sparse or non-sparse solution of RALASSO to eliminate the term \( \|D^*D\|_{1,1} \).

**Definition 3.1 (\( \ell_2\text{-DNSP}[15] \)).** Let \( D \) be an \( n \times p \) matrix. A matrix \( A \in \mathbb{R}^{m \times n} \) is said to obey the robust \( \ell_2 \) \( D\text{-Null Space Property} \) of order \( k \) with constants \( \beta > 0 \) and \( 0 < \gamma < \frac{1}{2} \), if for any \( x \in \mathbb{R}^n \), we have
\[ \|D^*_k x\|_2 \leq \beta \|Ax\|_2 + \gamma \frac{\|D^*x\|_1}{\sqrt{k}}, \]
where \( D^*_k x \) is the best \( k \)-sparse approximation of \( D^*x \).

Now, we consider RALASSO (3) where the matrix \( A \) is required to satisfy the \( \ell_2\text{-DNSP} \) of order \( k \) with constants \((\beta, \gamma)\). Our main result in Theorem 1 shows that the reconstruction can be completed by solving RALASSO.

**Theorem 3.2.** Let \( D \in \mathbb{R}^{n \times p} \) be a tight frame, and let \( A \in \mathbb{R}^{m \times n} \) be a measurement matrix satisfying the \( \ell_2\text{-DNSP} \) of order \( k \) with constants \((\beta, \gamma)\). Let \( y = Ax + w \), where \( w \) is the noise that satisfies \( \|D^*A^*w\|_\infty \leq \frac{1}{2} \). Then the solution \( \hat{x}_{\text{RAL}} \) to RALASSO satisfies
\[ \|\hat{x}_{\text{RAL}} - x\|_2 \leq c_1 \lambda \sqrt{k} + c_2 \frac{\|D^*x - (D^*x)_k\|_1}{\sqrt{k}} + c_3 \frac{\lambda p}{\sqrt{k^2}}, \]
where \( c_1 \) is a constant depending on the \( \beta \) and \( \gamma \), \( c_2, c_3 \) are constants depending only on \( \gamma \), and \( \beta > 0 \) and \( 0 < \gamma < \frac{1}{2} \) are constants depending on \( \delta_{2k} \).

Compared to the Theorem IV.1 in [23], we eliminate the term \( \|D^*D\|_{1,1} \) in the constant \( c_1 \). When the noise \( w \) is Gaussian, we also obtain the error estimation of RALASSO in the following theorem.

**Theorem 3.3.** Let \( D \in \mathbb{R}^{n \times p} \) be a tight frame, and let \( A \in \mathbb{R}^{m \times n} \) be a measurement matrix satisfying the \( \ell_2\text{-DNSP} \) of order \( k \) with constants \((\beta, \gamma)\). Assume that \( w \sim \mathcal{N}(0, \sigma^2 I) \) and that \( \hat{x}_{\text{RAL}} \) is the solution of RALASSO with \( \lambda = 4\sigma\alpha\sqrt{\log p} \), where \( \alpha = \max_{i \in [p]} \|AD_i\|_2 \). Then we have
\[ \|\hat{x}_{\text{RAL}} - x\|_2 \leq C_1 \alpha \sigma \sqrt{k \log p} + C_2 \frac{\|D^*x - (D^*x)_k\|_1}{\sqrt{k}} + C_3 \frac{\sigma \alpha \sqrt{\log pp}}{\sqrt{k^2}}, \]
with probability at least \( 1 - \frac{1}{\sqrt{2\pi \log pp}} \), where \( C_1 = 4c_1, C_2 = c_2, C_3 = 4c_3 \).

**Remark 1.** In fact, when \( \lambda \geq 4\sigma\alpha\sqrt{\log p} \) is a fixed constant, the above conclusion still holds.

Since the \( \ell_2\text{-DNSP} \) is a relaxation of the \( D\text{-RIP} \), we can build some connection between the \( \ell_2\text{-DNSP} \) and the \( D\text{-RIP} \). Here we give a proposition from [15].
Proposition 1. Given a matrix $D \in \mathbb{R}^{n \times p}$, the $D$-restricted isometry constant $\delta_{2k}$ of a measurement matrix $A \in \mathbb{R}^{m \times n}$ satisfies

$$\delta_{2k} < \frac{\sigma_{\min}}{\sigma_{\max}},$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are the smallest and largest singular values of $D$, then the matrix $A$ satisfies the $\ell_2$-DNSP of order $k$ with constants $\beta = \frac{2\sqrt{\rho} \sigma_{\min}(1 - \delta_k)}{\rho_{\min} (1 - \delta_k)} > 0$ and $\gamma = \frac{2\sigma_{\max}}{\sigma_{\min}} \cdot \frac{\delta_{2k}}{1 - \delta_k} + \frac{1}{4} \in (0, \frac{1}{2})$.

Based on the Proposition 1, we can know that if $\sigma_{\min} = \sigma_{\max}$, then $\delta_{2k} < \frac{1}{9}$ is the largest bound to make the matrix $A$ satisfy the $\ell_2$-DNSP and it is obvious that the condition is very strong. We also utilize the $D$-RIP to solve RALASSO and obtain a new bound $\delta_{2k} < 0.3162$ which is stated in Theorem 3. Compared to the condition $\delta_{2k} < 0.1907$ in [23], our result is better.

Theorem 3.4. Let $D \in \mathbb{R}^{n \times p}$ be a tight frame, and let $A \in \mathbb{R}^{m \times n}$ be a measurement matrix satisfying the $D$-RIP with $\delta_{2k} < 0.3162$. Consider the measurement $y = Ax + w$, where $w$ is the noise that satisfies $\|D^*A^*w\|_{\infty} \leq \frac{1}{2}$. Then the solution $\hat{x}_{R_L}$ to RALASSO satisfies

$$\|\hat{x}_{R_L} - x\|_2 \leq c_4 \lambda \sqrt{k} + c_5 \frac{\|D^*x - (D^*x)_k\|_1}{\sqrt{k}} + c_6 \frac{\lambda p}{\sqrt{k} \rho},$$

where $c_4$ is a constant depending on the $D$-RIP constant $\delta_{2k}$ and $\|D^*D\|_{1,1}$, and $c_5, c_6$ are constants depending only on the $D$-RIP constant $\delta_{2k}$.

Similarly, we provide the error estimation of RALASSO in the following theorem when the noise $w$ is Gaussian and $A$ satisfies the $D$-RIP.

Theorem 3.5. Let $D \in \mathbb{R}^{n \times p}$ be a tight frame, and let $A \in \mathbb{R}^{m \times n}$ be a measurement matrix satisfying the $D$-RIP with $\delta_{2k} < 0.3162$. Assume that $w \sim N(0, \sigma^2 I)$ and that $\hat{x}_{R_L}$ is the solution of RALASSO with $\lambda = 4\sigma \sqrt{2 \log p}$. Then we have

$$\|\hat{x}_{R_L} - x\|_2 \leq C_4 \sigma \sqrt{k} \log p + C_5 \frac{\|D^*x - (D^*x)_k\|_1}{\sqrt{k}} + C_6 \frac{\sigma \sqrt{\log p} p}{\sqrt{k} \rho},$$

with probability at least $1 - \frac{1}{\sqrt{2 \pi \log pp}}$, where $C_4 = 4\sqrt{2} c_4, C_5 = c_5, C_6 = 4\sqrt{2} c_6$.

Remark 2. Similar to Remark 1, when $\lambda \geq 4\sigma \sqrt{2 \log p}$ is a fixed constant, the above conclusion still holds.

When choose $\rho \to \infty$ in Theorem 3, then the third item implies $z = D^*x$ and the problem is equivalent to ALASSO. And the result can be generalized directly to ALASSO. In [17], the similar bound is $\delta_{3k} < 0.25$ which is equivalent to $\delta_{2k} < 0.0833$ by using Corollary 3.4 in [21]. Obviously, our result is better. The condition $\delta_{2k} < 0.3162$ is also weaker than $\delta_{2k} < 0.25$ which is stated in the Remark 11 of [26] for ALASSO.

4. Proofs. Let $T$ be the set of $k$ largest coefficients of $D^*x$ in magnitude. We decompose index set $T^c$ into sets of size $k$. Denote those sets $T_1, T_2, \ldots$, where $T_1$ denotes the $k$ largest coefficients of $D^*x$ in magnitude, $T_2$ denotes the next $k$ largest coefficients and so on. For simplify, we denote $h = \hat{x}_{R_L} - x$ where $\hat{x}_{R_L}$ is the optimal solution of RALASSO.
4.1. Supporting lemmas. Here we give some useful supporting lemmas that can be used when proofing the main theorems:

**Lemma 4.1** ([6]). Any \( v \in T(\alpha, k) = \{ v \in \mathbb{R}^p : \|v\|_\infty \leq \alpha, \|v\|_1 \leq k\alpha \} \) can be expressed as

\[
v = \sum_{i=1}^{N} \lambda_i u_i, \text{and } 0 \leq \lambda_i \leq 1, \sum_{i=1}^{N} \lambda_i = 1, \]

and \( u_i \in U(\alpha, k, v) = \{ u \in \mathbb{R}^p : \text{supp}(u) \subseteq \text{supp}(v), \|u\|_0 \leq k, \|u\|_\infty \leq \alpha, \|u\|_1 = \|v\|_1 \} \).

**Lemma 4.2.** If \( \sum_{i=1}^{N} \lambda_i = 1 \), then the following equation holds,

\[
\sum_{i=1}^{N} \lambda_i \|AD(\sum_{j=1}^{N} \lambda_j \beta_j - c\beta_j)\|_2^2 + (1 - 2c) \sum_{1 \leq i < j \leq N} \lambda_i \lambda_j \|AD(\beta_i - \beta_j)\|_2^2 = \sum_{i=1}^{N} \lambda_i (1 - c)^2 \|AD\beta_i\|_2^2.
\]

**Lemma 4.3.** Let \( D \in \mathbb{R}^{n \times p} \) be a tight frame, for any \( x \in \mathbb{R}^p \), we have

\[
\|Dx\|_2 \leq \|x\|_2.
\]

**Lemma 4.4.** The optimal solution \( \hat{x}_{\text{RAL}} \) for RALASSO satisfies

\[
\|Ah\|_2^2 + \lambda \|D_T^+ h\|_1 \leq 3\lambda \|D_T^+ h\|_1 + 4\lambda \|D_T^+ x\|_1 + \frac{\lambda^2}{\rho}.
\]

**Lemma 4.5** ([23]). The optimal solution \( \hat{x}_{\text{RAL}} \) for RALASSO satisfies

\[
\|D^* A^* Ah\|_\infty \leq \frac{1}{2} + \|D^* D\|_{1,1} \lambda.
\]

4.2. Proof of main theorems and corollaries.

4.2.1. A proof of Theorem 1.

**Proof.** From the definition of \( T_j \) and some properties of norms, the following inequality holds,

\[
\|D_T^+ h\|_2 \leq \sum_{j \geq 1} \|D_T^+ h\|_2 \leq \sum_{j \geq 1} \sqrt{k} \|D_T^+ h\|_\infty \\
\leq \sum_{j \geq 1} k^{-\frac{1}{2}} \|D_{T_{j-1}} h\|_1 = k^{-\frac{1}{2}} \|D^* h\|_1 \\
= k^{-\frac{1}{2}}(\|D_T^+ h\|_1 + \|D_T^+ h\|_1). \tag{4}
\]

Based on Lemma 4, we have

\[
\|D_T^+ h\|_1 \leq 3\|D_T^+ h\|_1 + 4\|D_T^+ x\|_1 + \frac{\lambda}{\rho}. \tag{5}
\]

And together with (4), we have

\[
\|D_T^+ h\|_2 \leq k^{-\frac{1}{2}}(4\|D_T^+ h\|_1 + 4\|D_T^+ x\|_1 + \frac{\lambda}{\rho}) \\
\leq k^{-\frac{1}{2}}(4\sqrt{k}\|D_T^+ h\|_2 + 4\|D_T^+ x\|_1 + \frac{\lambda}{\rho}) \\
= 4\|D_T^+ h\|_2 + \frac{4\|D_T^+ x\|_1}{\sqrt{k}} + \frac{\lambda p}{\sqrt{k} \rho}. \tag{6}
\]
Thus, applying (6), we obtain
\[ \|h\|_2 = \|D^* h\|_2 = \sqrt{\|D^*_T h\|_2^2 + \|D^*_T h\|_2^2} \]
\[ \leq \sqrt{\|D^*_T h\|_2^2 + (4\|D^*_T h\|_2 + 4\frac{\|D^*_T \cdot x\|_1}{\sqrt{k}} + \frac{\lambda p}{\sqrt{k\rho}})^2} \]
\[ \leq \sqrt{17\|D^*_T h\|_2^2 + 4\frac{\|D^*_T \cdot x\|_1}{\sqrt{k}} + \frac{\lambda p}{\sqrt{k\rho}}} \tag{7} \]

So the error estimation is further translated to estimate the bound of \(\|D^*_T h\|_2\). Applying the \(\ell_2\)-DNSP, we have
\[ \|D^*_T h\|_2 \leq \beta\|Ah\|_2 + \gamma \|D^*_T h\|_1 \]
\[ \leq \beta\|Ah\|_2 + \frac{\gamma}{\sqrt{k}} (4\|D^*_T h\|_1 + 4\|D^*_T \cdot x\|_1) + \lambda p \tag{8} \]
\[ \leq \beta\|Ah\|_2 + 4\gamma\|D^*_T h\|_2 + \frac{4\gamma}{\sqrt{k}} \|D^*_T \cdot x\|_1 + \lambda p \tag{9} \]

The inequality (8) holds because of (5), and (9) utilizes the Cauchy-Schwarz Inequality. Thus,
\[ \|D^*_T h\|_2 \leq \frac{1}{1 - 4\gamma} \left( \beta\|Ah\|_2 + \frac{4\gamma}{\sqrt{k}} \|D^*_T \cdot x\|_1 + \frac{\lambda p}{\sqrt{k\rho}} \right) \tag{10} \]

Based on Lemma 4 again, we have
\[ \|Ah\|_2^2 \leq 3\lambda\|D^*_T h\|_1 + 4\lambda\|D^*_T \cdot x\|_1 + \frac{\lambda^2}{\rho} \]
\[ \leq 3\lambda\sqrt{k}\|D^*_T h\|_2 + 4\lambda\|D^*_T \cdot x\|_1 + \frac{\lambda^2}{\rho} \]
\[ \leq \frac{3\lambda\sqrt{k}}{1 - 4\gamma}\|Ah\|_2 + \frac{4\lambda - 4\lambda\gamma}{1 - 4\gamma}\|D^*_T \cdot x\|_1 + \frac{1 - \gamma}{1 - 4\gamma}\lambda^2 \rho \tag{11} \]

Therefore
\[ \|Ah\|_2 \leq \frac{3\lambda\sqrt{k}}{1 - 4\gamma} \left( \frac{1 - \gamma}{3\sqrt{k}} + \frac{4\lambda - 4\lambda\gamma}{3\sqrt{k}} \right) \|D^*_T \cdot x\|_1 + \frac{\lambda p}{\rho} \tag{12} \]

Applying (10) and (11) to (7), we derive
\[ \|h\|_2 \leq \frac{3\sqrt{17}\beta^2}{(1 - 4\gamma)^2} \lambda\sqrt{k} + \frac{12 + 4\sqrt{17} + (8\sqrt{17} - 48)\gamma}{3 - 12\gamma} \|D^*_T \cdot x\|_1 \]
\[ + \frac{\sqrt{17} + (2\sqrt{17} - 12)\gamma}{3 - 12\gamma} \frac{\lambda p}{\sqrt{k\rho}} \]
which finishes the proof.

4.2.2. A proof of Theorem 3. Before giving the proof of Theorem 2, we firstly complete the proof of Theorem 3. The proof mainly follows from the ideas in [6] and [27]. But we still adopt some tricks to get over the obstacles in the proof.

Proof. Based on Lemma 4, we have
\[ \|D^*_T h\|_1 \leq 3\|D^*_T h\|_1 + 4\|D^*_T \cdot x\|_1 + \frac{\lambda}{\rho} \tag{12} \]
and we define \( \alpha = \frac{3\|D^*_T h\|_1 + 4\|D^*_T h\|_1 + k^2 p}{k} \). Then we divide \( D^*_T h \) into two parts:

\[
D^*_{T1} h = D^*_T h \cdot 1_{\{i | (D^*_T h)(i) > \alpha \}}, \quad D^*_{T2} h = D^*_T h \cdot 1_{\{i | (D^*_T h)(i) \leq \alpha \}}.
\]

Thus, \( \|D^*_T h\|_1 \leq k \alpha \). Denote \( \|D^*_T h\|_0 = m \), we have

\[
\alpha m \leq \|D^*_{T1} h\|_1 \leq \|D^*_T h\|_1 \leq k \alpha,
\]

i.e. \( m \leq k \). As for \( D^*_{T2} h \), we have

\[
\|D^*_{T2} h\|_1 = \|D^*_T h\|_1 - \|D^*_{T1} h\|_1 \leq (k - m) \alpha, \quad \|D^*_{T2} h\|_\infty \leq \alpha.
\]

Applying Lemma 1, \( D^*_{T2} h \) can be expressed as

\[
D^*_{T2} h = \sum_{i=1}^N \lambda_i u_i, \tag{13}
\]

where \( u_i \in \mathbb{R}^n \) is \((k - m)\)-sparse, and \( \sum_{i=1}^N \lambda_i = 1 \), \( \|u_i\|_\infty \leq \alpha \), \( \|u_i\|_1 \leq (k - m) \alpha \). Thus,

\[
\|u_i\|_2 \leq \sqrt{\|u_i\|_0 \cdot \|u_i\|_\infty} \leq \sqrt{k} \alpha. \tag{14}
\]

Denote \( \beta_i = D^*_T h + D^*_{T1} h + \mu u_i \). Obviously, \( \beta_i \) is 2\( k \)-sparse, then

\[
\sum_{j=1}^N (\lambda_j \beta_j) - c \beta_i - \mu D^* h = (1 - c - \mu)(D^*_T h + D^*_{T1} h) - c \mu u_i,
\]

c, \( \mu \) are determined constants. And \( \sum_{j=1}^N (\lambda_j \beta_j) - c \beta_i - \mu D^* h \) is also 2\( k \)-sparse.

Based on Lemma 2, we set \( c = \frac{1}{2} \) and get

\[
0 = \sum_{i=1}^N \lambda_i \|AD\left(\sum_{j=1}^N \lambda_j \beta_j - \frac{1}{2} \beta_i\right)\|_2^2 - \sum_{i=1}^N \frac{\lambda_i}{4} \|AD \beta_i\|_2^2
\]

\[
= \sum_{i=1}^N \lambda_i \|AD\left(\frac{1}{2} - \mu\right)(D^*_T h + D^*_{T1} h) - \frac{\mu}{2} u_i\|_2^2 + \mu Ah\|_2^2 - \sum_{i=1}^N \frac{\lambda_i}{4} \|AD \beta_i\|_2^2
\]

\[
= \sum_{i=1}^N \lambda_i \|AD\left(\frac{1}{2} - \mu\right)(D^*_T h + D^*_{T1} h) - \frac{\mu}{2} u_i\|_2^2 + \mu(1 - \mu)(AD(D^*_T h + D^*_{T1} h), Ah) - \sum_{i=1}^N \frac{\lambda_i}{4} \|AD \beta_i\|_2^2. \tag{15}
\]

Depending on Lemma 5 and \( m < k \), we can get

\[
\langle AD(D^*_T h + D^*_{T1} h), Ah \rangle = \langle D^*_T h + D^*_{T1} h, A^* Ah \rangle \\
\leq \|D^*_T h + D^*_{T1} h\|_1 \|A^* Ah\|_\infty \\
\leq \sqrt{k} + m\|D^*_T h + D^*_{T1} h\|_2 \|A^* Ah\|_\infty \\
\leq \sqrt{2k} \lambda \left(\frac{1}{2} + \|D^* D\|_{1,1}\right) \|D^*_T h + D^*_{T1} h\|_2. \tag{16}
\]
Considering (21), the following equation holds,

$$
0 \leq (1 + \delta_{2k}) \sum_{i=1}^{N} \lambda_i \| D \left( \frac{1}{2} - \mu \right) (D_T^* h + D_T_{1}^* h) - \frac{\mu}{2} u_i \|^2
$$

(17)

$$
+ \mu (1 - \mu) \sqrt{2k} \lambda \left( \frac{1}{2} + \| D^* D \|_{1,1} \right) \| D_T^* h + D_T_{1}^* h \|_2 - (1 - \delta_{2k}) \sum_{i=1}^{N} \lambda_i \| D \beta_i \|^2.
$$

(18)

By the equations (13) and $\beta_i = D_T^* h + D_T_{1}^* h + \mu u_i$, we have

$$
I_1 = (1 + \delta_{2k}) \left( \frac{\mu^2}{2} \sum_{i=1}^{N} \lambda_i \| Du_i \|^2 + \left( \frac{1}{2} - \mu \right)^2 \| D (D_T^* h + D_T_{1}^* h) \|^2
$$

$$
- \mu \left( \frac{1}{2} - \mu \right) \langle D (D_T^* h + D_T_{1}^* h), DD_T^* h \rangle],
$$

(19)

$$
I_2 = \frac{1 - \delta_{2k}}{4} \left( \| D (D_T^* h + D_T_{1}^* h) \|^2 + \mu^2 \sum_{i=1}^{N} \lambda_i \| Du_i \|^2
$$

$$
+ 2\mu \langle D (D_T^* h + D_T_{1}^* h), DD_T^* h \rangle].
$$

(20)

Then, substituting (19) and (20) into (17) and (18), we have

$$
0 \leq \mu (1 - \mu) \sqrt{2k} \lambda \left( \frac{1}{2} + \| D^* D \|_{1,1} \right) \| D_T^* h + D_T_{1}^* h \|_2
$$

$$
+ \left[ (1 + \delta_{2k}) \left( \frac{1}{2} - \mu \right)^2 - \frac{1 - \delta_{2k}}{4} \| D (D_T^* h + D_T_{1}^* h) \|^2
$$

$$
+ \left[ (1 + \delta_{2k}) \mu \left( \frac{1 - \mu}{2} \right) - \frac{2\mu (1 - \delta_{2k})}{4} \| D (D_T^* h + D_T_{1}^* h), DD_T^* h \rangle
$$

$$
+ \frac{\delta_{2k}}{2} \mu^2 \sum_{i=1}^{N} \lambda_i \| Du_i \|^2.
$$

Considering (21), the following equation holds,

$$
\| D (D_T^* h + D_T_{1}^* h) \|^2 = \| D (D^* h - DD_T^* h) \|^2 = \| h \|^2 + \| DD_T^* h \|^2 - 2 \langle h, DD_T^* h \rangle
$$

$$
= \| D_T^* h + D_T_{1}^* h \|^2 + \| DD_T^* h \|^2 - \| D_T^* h \|^2.
$$

(23)

Considering (22), the following equation holds,

$$
\langle D (D_T^* h + D_T_{1}^* h), DD_T^* h \rangle = \langle D (D^* h - DD_T^* h), DD_T^* h \rangle
$$

$$
= \| D_T^* h \|^2 - \| DD_T^* h \|^2.
$$

(24)

Further, we derive

$$
0 \leq \mu (1 - \mu) \sqrt{2k} \lambda \left( \frac{1}{2} + \| D^* D \|_{1,1} \right) \| D_T^* h + D_T_{1}^* h \|_2
$$

$$
+ \left[ \mu^2 - \mu + \frac{1}{2} \delta_{2k} + \mu^2 - \mu \| D_T^* h + D_T_{1}^* h \|^2.
$$
By solving this inequality, we get
\[ \|u_i\|_2 \leq \sqrt{k} \alpha = \frac{3\|D_T^x h\|_1 + 4\|D_T x\|_1 + \frac{\lambda p}{\sqrt{k} p}}{\sqrt{k}} \leq \frac{4\|D_T^x x\|_1}{\sqrt{k}} + 3\|D_T h\|_2 + \frac{\lambda p}{\sqrt{k} p} \leq 2P + 3X + \frac{\lambda p}{\sqrt{k} p}. \]

Then the inequality (25) can be translated into
\begin{align*}
0 & \leq \left[ \frac{1}{2} \mu^2 - \mu + \frac{1}{2} \right] \delta_{2k} + \mu^2 - \mu |X|^2 + \frac{\delta_{2k}}{2} \mu^2 \left( \frac{\lambda p}{\sqrt{k} p} + 2P \right)^2 \\
& \quad + \left[ 3\delta_{2k} \mu^2 \left( \frac{\lambda p}{\sqrt{k} p} + 2P \right) + \mu(1 - \mu) \left( \frac{1}{2} + \|D^* D\|_{1,1} \right) \sqrt{2k} \lambda \right] |X|, \tag{26}
\end{align*}

which is a second-order inequality for \( X \). In order to get the higher bounds of \( X \) from the inequality (26), we denote that the second-order item coefficient is negative. Then,
\[ \delta_{2k} < \max_{\mu \leq \frac{1}{2}} \frac{-\mu^2 + \mu}{\frac{1}{2} \mu^2 - \mu + \frac{1}{2}} = \frac{\sqrt{10}}{10} (\approx 0.3162), \]
where the restriction comes from \( \frac{1}{2} - \mu \delta_{2k} \geq 0 \), and when \( \mu = \frac{\sqrt{10} - 1}{9} \), the function \( \frac{-\mu^2 + \mu}{\frac{1}{2} \mu^2 - \mu + \frac{1}{2}} \) reaches the maximum. Thus, substituting \( \mu = \frac{\sqrt{10} - 1}{9} \) into (26), we obtain
\[ 0 \leq (10 \delta_{2k} - \sqrt{10}) X^2 + \left[ 3\left( \frac{\lambda p}{\sqrt{k} p} + 2P \right) \delta_{2k} + 2\sqrt{5} \left( \frac{1}{2} + \|D^* D\|_{1,1} \right) \sqrt{k} \lambda \right] |X| \\
+ \frac{\delta_{2k}}{2} \left( \frac{\lambda p}{\sqrt{k} p} + 2P \right)^2. \]

By solving this inequality, we get
\begin{align*}
X & \leq \frac{2\sqrt{5} \left( \frac{1}{2} + \|D^* D\|_{1,1} \right) \sqrt{k} \lambda}{\sqrt{10} - 10 \delta_{2k}} + \frac{6 \delta_{2k} + \sqrt{2(\sqrt{10} - 10 \delta_{2k}) \delta_{2k}} P}{\sqrt{10} - 10 \delta_{2k}} \\
& \quad + \frac{6 \delta_{2k} + \sqrt{2(\sqrt{10} - 10 \delta_{2k}) \delta_{2k}} \lambda p}{2(\sqrt{10} - 10 \delta_{2k}) \sqrt{k} p}. \tag{27}
\end{align*}

Finally, applying (7) together with (27), we have
\begin{align*}
\|h\|_2 & \leq \sqrt{17} \|D_T^x h\|_2 + \frac{\|D_T^x x\|_1}{\sqrt{k}} + \frac{\lambda p}{\sqrt{k} p} \leq \sqrt{17} X + 2P + \frac{\lambda p}{\sqrt{k} p} \\
& \leq \frac{2\sqrt{5} \left( \frac{1}{2} + \|D^* D\|_{1,1} \right) \sqrt{k} \lambda}{\sqrt{10} - 10 \delta_{2k}}.
\end{align*}
ANALYSIS NON-SPARSE RECOVERY FOR RELAXED ALASSO  

\[
\begin{align*}
+ \left\{ \frac{2\sqrt{17}}{\sqrt{10} - 10\delta_{2k}} \left[ 6\delta_{2k} + \sqrt{2(\sqrt{10} - 10\delta_{2k})\delta_{2k}} \right] \right\} \frac{\|D^*_x\|_1}{\sqrt{k}} \\
+ \left\{ \frac{\sqrt{17}}{2(\sqrt{10} - 10\delta_{2k})} \left[ \right] \right\} \frac{\lambda_p}{\sqrt{\rho}}
\end{align*}
\]

which finishes the proof. \(\square\)

4.2.3. Proofs of Theorem 2 and Theorem 4. We firstly give the proof of Theorem 2.

Proof. Based on the result of Theorem 1, we need only to compute the probability \(\mathbb{P}(\|D^*A^*w\|_\infty \leq \frac{\lambda}{2})\). Let \(z_i = \langle AD\beta_i, w \rangle \), then \(z_i\) has Gaussian distribution \(\mathcal{N}(0, 1)\). By using the union bound and \(\alpha = \max_{i \in [p]} \|AD\|_2\), we have

\[
\mathbb{P}(\|D^*A^*w\|_\infty > \frac{\lambda}{2}) = \mathbb{P}(\max_{i \in [p]} |z_i| > 2\sqrt{\log p}) \leq \frac{1}{\sqrt{2\pi \log p}}.
\]

The last inequality follows from the Gaussian tail probability bound. Thus, the probability \(\mathbb{P}(\|D^*A^*w\|_\infty \leq \frac{\lambda}{2}) \geq 1 - \frac{1}{\sqrt{2\pi \log p}}\), which finishes the proof. \(\square\)

Theorem 4 can be obtained similarly, but the difference is the bound of \(\|AD^i\|_2\). From the definition of \(D\)-RIP, we have

\[
\|AD^i\|_2 \leq \sqrt{1 + \delta_{2k}}\|D^i\|_2 \leq \sqrt{1 + \delta_{2k}} < 2, \forall i \in [p].
\]

Repeat the steps in the proof of Theorem 2, we can derive the result.

4.3. Proof of lemmas.

4.3.1. A proof of Lemma 2.

Proof. Firstly, we expand the right-hand side and left-hand side of the equation of Lemma 2:

\[
RHS = \sum_{i=1}^N \lambda_i\|AD\beta_i\|_2^2 - 2c \sum_{i=1}^N \lambda_i\|AD\beta_i\|_2^2 + c^2 \sum_{i=1}^N \lambda_i\|AD\beta_i\|_2^2,
\]

\[
LHS = c^2 \sum_{i=1}^N \lambda_i\|AD\beta_i\|_2^2 + \sum_{i=1}^N \sum_{j=1}^N \lambda_i\lambda_j \langle AD\beta_i, AD\beta_j \rangle
\]

\[
- 2c \sum_{i=1}^N \sum_{j=1}^N \lambda_i\lambda_j \langle AD\beta_i, AD\beta_j \rangle
\]

\[
+ \sum_{1 \leq i < j \leq N} \lambda_i\lambda_j\|AD(\beta_i - \beta_j)\|_2^2 - 2c \sum_{1 \leq i < j \leq N} \lambda_i\lambda_j\|AD(\beta_i - \beta_j)\|_2^2.
\]

(28)
where (28) is the first item of the left-hand side and (29) is the second one. Therefore, we only need to prove

\[
\sum_{i=1}^{N} \lambda_i \|AD\beta_i\|_2^2 = \sum_{1 \leq i < j \leq N} \lambda_i \lambda_j \|AD(\beta_i - \beta_j)\|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \langle AD\beta_i, AD\beta_j \rangle.
\]

And after calculation, we get

\[
\begin{align*}
I_3 &= \sum_{1 \leq i < j \leq N} \lambda_i \lambda_j \left( \|AD\beta_i\|_2^2 - 2\langle AD\beta_i, AD\beta_j \rangle + \|AD\beta_j\|_2^2 \right) \\
I_4 &= 2 \sum_{1 \leq i < j \leq N} \lambda_i \lambda_j \langle AD\beta_i, AD\beta_j \rangle + \sum_{i=1}^{N} \lambda_i^2 \|AD\beta_i\|_2^2.
\end{align*}
\]

Thus,

\[
I_3 + I_4 = \sum_{1 \leq i < j \leq N} \lambda_i \lambda_j \left( \|AD\beta_i\|_2^2 + \|AD\beta_j\|_2^2 \right) + \sum_{i=1}^{N} \lambda_i^2 \|AD\beta_i\|_2^2
= \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \|AD\beta_i\|_2^2 = \sum_{j=1}^{N} \lambda_j \left( \sum_{i=1}^{N} \lambda_i \|AD\beta_i\|_2^2 \right)
= \sum_{i=1}^{N} \lambda_i \|AD\beta_i\|_2^2.
\]

Then the proof is completed. \(\square\)

4.3.2. A proof of Lemma 3.

**Proof.** Based on the fact that \(I = D^*D + D^*_\perp D\perp\), where \(D^*_\perp\) is the complement of \(D^*\), we can get

\[
\frac{1}{2} \|Ax\|_2^2 - 2\langle x, x \rangle = \langle x, x \rangle - 2\langle Dx, x \rangle = \langle x, x \rangle - 2\langle D^*Dx, x \rangle
= \langle x, x \rangle - 2\langle D^*Dx, x \rangle = \langle x, x \rangle - 2\langle D^*_\perp D\perp x, x \rangle = \|D\perp x\|_2^2 \geq 0.
\]

Then, the lemma is proved. \(\square\)

4.3.3. A proof of Lemma 4.

**Proof.** Since \((\hat{x}_{\text{RLA}}, \hat{z}_{\text{RLA}})\) is the optimal solution of the optimization problem RALASSO, for any pair \((x, z)\) satisfying \(z = D^*x\), we have

\[
\frac{1}{2} \|A\hat{x}_{\text{RLA}} - b\|_2^2 + \lambda \|\hat{z}_{\text{RLA}}\|_1 + \frac{\rho}{2} \|\hat{z}_{\text{RLA}} - D^*\hat{x}_{\text{RLA}}\|_2^2 \leq \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|D^*x\|_1.
\]

Since \(b = Ax + w\) and \(h = \hat{x}_{\text{RLA}} - x\), we can get

\[
\frac{1}{2} \|Ah\|_2^2 - \langle Ah, w \rangle \leq \lambda \|D^*x\|_1 - \lambda \|\hat{z}_{\text{RLA}}\|_1 - \frac{\rho}{2} \|\hat{z}_{\text{RLA}} - D^*\hat{x}_{\text{RLA}}\|_2^2.
\]

The first-order optimality condition of RALASSO on \(z\) can be stated as

\[
\lambda v + \rho(\hat{z}_{\text{RLA}} - D^*\hat{x}_{\text{RLA}}) = 0.
\]
where $v$ is a subgradient of the function $\|z\|_1$ and obviously $\|v\|_\infty \leq 1$. Then, $\tilde{z}_{\text{RLA}}$ can be substituted for $\tilde{z}_{\text{RLA}} = -\frac{\lambda}{\rho} v + D^* \tilde{x}_{\text{RLA}}$.

$$\frac{1}{2} \|Ah\|_2^2 - \frac{\lambda}{2} \|D^*v\|_1 \leq \frac{1}{2} \|Ah\|_2^2 - \langle Ah, w \rangle$$

Then, we obtain that

$$\|Ah\|_2^2 \leq \lambda(\|D^*h\|_1 + 2\|D^*x\|_1 - 2\|D^*\tilde{x}_{\text{RLA}}\|_1) + \frac{\lambda^2}{\rho} p$$

$$= \lambda(\|D^*_T h\|_1 + \|D^*_T x\|_1 + 2\|D^*_T \tilde{x}_{\text{RLA}}\|_1 - 2\|D^*_T \tilde{x}_{\text{RLA}}\|_1) + \frac{\lambda^2}{\rho} p$$

$$\leq \lambda(3\|D^*_T h\|_1 + 3\|D^*_T x\|_1 - \|D^*_T \tilde{x}_{\text{RLA}}\|_1) + \frac{\lambda^2}{\rho} p$$

$$\leq \lambda(3\|D^*_T h\|_1 + 4\|D^*_T x\|_1 - \|D^*_T h\|_1) + \frac{\lambda^2}{\rho} p.$$ 

The first inequality holds because $\langle Ah, w \rangle = h^* D D^* \tilde{x}_{\text{RLA}} - \|D^* \tilde{x}_{\text{RLA}}\|_1 \leq \|D^* x\|_1$. And the last inequality holds because $\frac{\lambda^2}{\rho} |v_i| - \frac{\lambda^2}{\rho^2} |v_i| = 1$. Then, we obtain that

$$\|Ah\|_2^2 \leq \lambda(\|D^*h\|_1 + 2\|D^*x\|_1 - 2\|D^*\tilde{x}_{\text{RLA}}\|_1) + \frac{\lambda^2}{\rho} p$$

$$\leq \lambda(\|D^*_T h\|_1 + \|D^*_T x\|_1 + 2\|D^*_T \tilde{x}_{\text{RLA}}\|_1) + \frac{\lambda^2}{\rho} p$$

$$\leq \lambda(3\|D^*_T h\|_1 + 3\|D^*_T x\|_1 - \|D^*_T \tilde{x}_{\text{RLA}}\|_1) + \frac{\lambda^2}{\rho} p$$

$$\leq \lambda(3\|D^*_T h\|_1 + 4\|D^*_T x\|_1 - \|D^*_T h\|_1) + \frac{\lambda^2}{\rho} p.$$ 

The inequality (31) utilizes the fact that $\|D^*_T h\|_1 \leq \|D^*_T x\|_1 + \|D^*_T \tilde{x}_{\text{RLA}}\|_1$ and $\|D^*_T x\|_1 - \|D^*_T \tilde{x}_{\text{RLA}}\|_1 \leq \|D^*_T h\|_1$. And (32) utilizes the fact that $\|D^*_T h\|_1 - \|D^*_T x\|_1 \leq \|D^*_T \tilde{x}_{\text{RLA}}\|_1$. Then after transposition, we complete the proof.

**Acknowledgments.** The authors would like to thank the referees for their insightful comments and suggestions.

**REFERENCES**

[1] 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.*, 9 (2010), 2345–2356.

[2] A. Beck, and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Imag. Sci.*, 2 (2009), 183–202.

[3] A. Beck and M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *IEEE Trans. Image Process.*, 18 (2009), 2419–2434.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, U.K. Cambridge Univ. Press, 2004.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via alternating direction method of multipliers, *Found. Trends Mach. Learn.*, 3 (2010), 1–122.

[6] T. Cai and A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, *IEEE Trans. Inform. Theory*, 60 (2014), 122–132.

[7] E. J. Candès, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory*, 52 (2006), 489–509.

[8] E. J. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pure Appl. Math.*, 59 (2006), 1207–1223.
[9] E. J. Candès, Y. C. Eldar, D. Needell and P. Randall, Compressed sensing with coherent and redundant dictionaries, *Appl. Comput. Harmon. Anal.*, 31 (2011), 59–73.
[10] S. S. Chen, D. L. Donoho and M. A. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.*, 43 (2001), 129–159.
[11] D. L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, 52 (2006), 1289–1306.
[12] D. L. Donoho, Y. Tsaig, I. Drori and J. L. Starck, Sparse solution of underdetermined linear equations by stagewise Orthogonal Matching Pursuit, *IEEE Trans. Inform. Theory*, 58 (2012), 1094–1121.
[13] I. Drori, Fast $\ell_1$-minimization by iterative thresholding for multidimensional NMR spectroscopy, *EURASIP J. Adv. Signal Process.*, 2007 (2007), 1–10.
[14] M. Elad, P. Milanfar and R. Rubinstein, Analysis versus synthesis in signal priors, *Inverse Probl.*, 23 (2007), 947–968.
[15] S. Fourcart, Stability and robustness of $\ell_1$-minimization with Weibull matrices and redundant dictionaries, *Linear Alg. Appl.*, 441 (2014), 4–21.
[16] M. Herman and T. Strohmer, High-resolution radar via compressed sensing, *IEEE Trans. Signal Process.*, 57 (2009), 2275–2284.
[17] J. Lin and S. Li, Sparse recovery with coherent tight frame via analysis Dantzig selector and analysis lasso, *Appl. Comput. Harmon. Anal.*, 37 (2014), 126–139.
[18] M. Lustig, D. L. Donoho, J. M. Santos and J. M. Pauly, Compressed sensing MRI, *IEEE Signal Process Mag.*, 27 (2008), 72–82.
[19] S. Nam, M. E. Davies, M. Elad and R. Gribonval, The cosparse analysis model and algorithms, *Appl. Comput. Harmon. Anal.*, 34 (2013), 30–56.
[20] B. K. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Comput.*, 24 (1995), 227–234.
[21] D. Needell and J. A. Tropp, Cosamp: Iterative signal recovery from noisy samples, *Appl. Comput. Harmon. Anal.*, 26 (2008), 301–321.
[22] H. Rauhut, K. Schnass and P. Vandergheynst, Compressed sensing and redundant dictionaries, *IEEE Trans. Inform. Theory*, 54 (2008), 2210–2219.
[23] Z. Tan, Y. C. Eldar, A. Beck and A. Nehorai, Smoothing and decomposition for analysis sparse recovery, *IEEE Trans. Signal Process.*, 62 (2014), 1762–1774.
[24] R. Tibshirani, Regression shrinkage and selection via lasso, *J. R. Statist. Soc. B. Stat. Meth.*, 58 (1996), 267–288.
[25] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu and K. Knight, Sparsity and smoothness via the fused Lasso, *J. R. Statist. Soc. B. Stat. Meth.*, 67 (2005), 91–108.
[26] Y. Xia and S. Li, Analysis recovery with coherent frames and correlated measurements, *IEEE Trans. Inform. Theory*, 62 (2016), 6493–6507.
[27] R. Zhang and S. Li, Optimal RIP bounds for sparse signals recovery via $\ell_p$ minimization, *Appl. Comput. Harmon. Anal.*, 47 (2019), 566–584.

Received August 2019; revised November 2019.

E-mail address: hangxu@zju.edu.cn
E-mail address: songli@zju.edu.cn