Appl. Math. J. Chinese Univ.
2022, 37(3): 422-434

On the grouping effect of the l_{1-2} models

SHEN Yi[1] GUO Wan-ling[1] HU Rui-fang[2,*]

Abstract. This paper aims to study the mathematical properties of the l_{1-2} models that employ measurement matrices with correlated columns. We first show that the l_{1-2} model satisfies the grouping effect which ensures that coefficients corresponding to highly correlated columns in a measurement matrix have small differences. Then we provide the stability analysis based on the sparse approximation property. When the entries of the vectors have different signs, we show that the grouping effect also holds for the constraint l_{1+2} minimization model which is implicated by the linearized Bregman iteration.

§1 Introduction

Many high-dimensional signals have inherently low-dimensional structures, i.e., most information can be represented by fewer coefficients. The fundamental problem of compressed sensing is recovering sparse signals from limited measurements. Sparse recovery has gained much attention, with applications in several areas such as signal processing, astronomy, and digital image restoration. Interested readers could consult, for example, [5, 7, 8] for details.

Let $X \in \mathbb{R}^{n\times p}$ denote the measurement matrix and $\boldsymbol{\beta}^* \in \mathbb{R}^p$ denote an unknown true vector to be recovered. The standard model of compressed sensing is formulated as an underdetermined linear system

$$\boldsymbol{y} = X\boldsymbol{\beta}^* + \boldsymbol{z}, \tag{1.1}$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is said to be the measurements and $\boldsymbol{z}$ denotes the noise term. Throughout this paper, the matrix X is assumed to be a surjective map with $n < p$. It follows that the linear system has infinite solutions. The goal of compressed sensing is to recover $\boldsymbol{\beta}^*$ successfully from $\boldsymbol{y}$ and X by taking sparsity into account. The l_1 norm which works as the convex relaxation of l_0 has been widely used in information theory and statistical learning, see e.g. [5, 8, 18, 26]. Another relaxed method is l_q "norm" with $0 < q \leq 1$, interested readers can see the recent work in [6, 10, 17] and references therein.

Received: 2020-09-05. Revised: 2021-03-21.
MR Subject Classification: 90C26, 65K10, 49M29, 68T05.
Keywords: grouping effect, sparsity, linearized Bregman, non-convex, compressed sensing.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-022-4256-5.
Research of Yi Shen was supported by the Zhejiang Provincial Natural Science Foundation of China (LR19A010001), the NSF of China (12022112), Research of Hu Ruifang was supported by the general research project of Jiaxing Nanhu University (62107YL).
*Corresponding author.

To ensure successful recovery, the measurement matrices are usually assumed to be incoherent systems. Since measurement matrices with highly correlated columns often appear in machine learning and statistics. This assumption may have limitations in some situations. For the highly coherence system, the difference between the l_1 and l_2 norms denoted as l_{1-2}, has been shown to have superior performance over the classic l_1 method in [11, 12, 22]. For $q = 1$ and $q = 2$, the l_q norm of the vector $\boldsymbol{\beta}$ is defined by $\|\boldsymbol{\beta}\|_q = \left(\sum_{i=1}^{p} |\beta_i|^q\right)^{1/q}$. The unconstrained l_{1-2} model is formulated as follows,

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \lambda\left(\|\boldsymbol{\beta}\|_1 - \|\boldsymbol{\beta}\|_2\right) + \frac{1}{2}\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2. \tag{1.2}$$

The corresponding constrained version is

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \lambda\{\|\boldsymbol{\beta}\|_1 - \|\boldsymbol{\beta}\|_2\} \quad \text{subject to} \quad X\boldsymbol{\beta} = \boldsymbol{y} \tag{1.3}$$

where λ is a positive parameter. Numerical experiments demonstrate that the l_{1-2} method is always better than the l_1 method to recover sparse vectors, especially for the high coherence sensing matrices [12]. The l_{1-2} penalty term is also successfully used in multichannel blind deconvolution problem where the measurement operator is highly correlated [21]. But currently, the corresponding theory that guarantees the performance of the l_{1-2} model for dependent matrices is lacking. This motivates us to analyze l_{1-2} based models from a statistical point of view. More specifically, we study the grouping effect and provide the stable guarantee for correlated Gaussian random matrices.

The grouping effect property of the elastic net was first introduced to [25, Theorem 1] then further discussed in [3, 15, 16, 24]. Elastic net is a linear combination of l_1 and l_2 penalties, and combines properties of LASSO and ridge regularization. Let $\boldsymbol{x}_i$ denote the i-th column vector of the measurement matrix X. The design matrix is assumed to be standardized, i,e, $\|\boldsymbol{x}_i\|_2 = 1$, $i = 1, \ldots, p$. Then the correlation between the vector $\boldsymbol{x}_i$ and the vector $\boldsymbol{x}_j$ is denoted by

$$\rho := \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{x}_j,$$

where $\boldsymbol{x}_i^{\mathsf{T}}$ denotes the transpose of $\boldsymbol{x}_i$. Note that if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are close, then $\rho \approx 1$. Simple calculation leads to

$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 = 2 - 2\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = 2(1 - \rho). \tag{1.4}$$

Roughly speaking, if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are close, the corresponding difference between the coefficient paths of predictors i and j is expected to be small, i,e, $\hat{\beta}_i \approx \hat{\beta}_j$. If any two columns of the measurement matrix $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are highly relevant, then the corresponding coefficients $\hat{\beta}_i$ and $\hat{\beta}_j$ are expected to be very close. Mathematically, the grouping effect can be expressed as

$$\left|\hat{\beta}_i - \hat{\beta}_j\right| \leq \delta \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2,$$

where δ is a constant. Zou [25] first studied the grouping effect of elastic net and gave the mathematical theorem for the situation of the same sign. Zhou gave the strict mathematical proof process and relaxed the condition of the same sign, such that the conclusion still holds for the elastic net model [24]. We also can find this property in other models such as image restoration and signal processing. For example, the constrained l_{1+2} minimization model was proved to satisfy the grouping effect on condition of the same signs [3]. The balance approach which was first obtained in [2] from the image inpainting problem also satisfies the grouping effect property [16].

Several sufficient conditions have been obtained to guarantee that sparse vectors can be recovered from (1.2) robustly. Basically, there are two types of sufficient conditions. One is based on the restricted isometry property [5] and the other is based on the mutual coherence [7]. The first sufficient condition for stable recovery was given in [22]. Another sufficient condition via mutual coherence was studied in [19]. Both of them require that the column vectors should be uncorrelated. Since the l_{1-2} model takes good numerical performance of high correlated matrix, we consider stable recovery of the l_{1-2} model with measurement matrix with highly correlated columns. A commonly used statistical model is the covariance matrix which takes columns to be Gaussian vectors with correlated entries.

The rest of this paper is organized as follows. In Section 2, we prove the main results on grouping effect of two l_{1-2} models. In Section 3, we extend the work in [3]. We prove that the grouping effect of the l_{1+2} model holds without the condition that the coefficients have the same signs. In Section 4, the stability of the constraint l_{1-2} model is obtained.

§2 The grouping effect

In this section, we will focus on l_{1-2} models. We study the grouping effect property and give the relevant proof. Since the l_{1-2} models are non convex, the relationship between the model (1.2) and (1.3) is not clear. The grouping effect properties of the model (1.2) and the model (1.3) were differently established. We consider the unconstrained minimization program (1.2) first.

Theorem 1. *Given the response $\boldsymbol{y}$ and the standardized matrix X. Let $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)^{\mathsf{T}} \in \mathbb{R}^p$ be the solution to the minimization program* (1.2). *If $\hat{\beta}_i\hat{\beta}_j > 0$, then*

$$\left|\hat{\beta}_i - \hat{\beta}_j\right| \leq \frac{\sqrt{2(1-\rho)}}{\lambda} \left\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right\|_2 \left\|\hat{\boldsymbol{\beta}}\right\|_2. \tag{2.1}$$

Proof. For any given real number β, $\operatorname{sgn}(\beta)$ denotes its sign. Since the vector $\hat{\boldsymbol{\beta}}$ is the solution to minimization program (1.2), for each i,j, we have

$$0 \in \lambda \operatorname{sgn}\left(\hat{\beta}_i\right) - \lambda \frac{\hat{\beta}_i}{\left\|\hat{\boldsymbol{\beta}}\right\|_2} - \boldsymbol{x}_i^{\mathsf{T}}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right) \tag{2.2}$$

and

$$0 \in \lambda \operatorname{sgn}\left(\hat{\beta}_j\right) - \lambda \frac{\hat{\beta}_j}{\left\|\hat{\boldsymbol{\beta}}\right\|_2} - \boldsymbol{x}_j^{\mathsf{T}}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right). \tag{2.3}$$

Under the assumption that $\hat{\beta}_i\hat{\beta}_j > 0$, both $\operatorname{sgn}\left(\hat{\beta}_i\right)$ and $\operatorname{sgn}\left(\hat{\beta}_j\right)$ contain only one element (1 or -1). Hence, the "$\in$" in (2.2) and (2.3) can be replaced by "=", and $\operatorname{sgn}\left(\hat{\beta}_i\right) = \operatorname{sgn}\left(\hat{\beta}_j\right)$.

Subtracting equation (2.2) from equation (2.3) gives

$$\lambda \frac{\hat{\beta}_i - \hat{\beta}_j}{\left\|\hat{\boldsymbol{\beta}}\right\|_2} = -\left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right). \tag{2.4}$$

Since

$$\left|\left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right)\right| \leq \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \left\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right\|_2$$

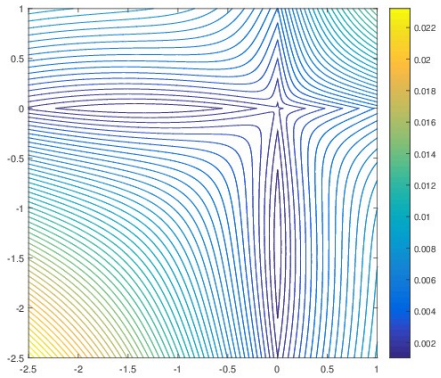

Figure 1. Contour line of the cost function in Example 1.

and rearranging the terms (2.4), we have

$$\begin{aligned}\left|\hat{\beta}_i - \hat{\beta}_j\right| &\leq \frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2}{\lambda} \left\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right\|_2 \left\|\hat{\boldsymbol{\beta}}\right\|_2 \\ &= \frac{\sqrt{2(1-\rho)}}{\lambda} \left\|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right\|_2 \left\|\hat{\boldsymbol{\beta}}\right\|_2.\end{aligned}$$

□

Compared with the results of the elastic net, the condition $\hat{\beta}_i\hat{\beta}_j > 0$ in Theorem 1 seems superfluous. While the following example shows that the condition $\hat{\beta}_i\hat{\beta}_j > 0$ is necessary for the l_{1-2} model.

Example 1. *The linear system $\boldsymbol{y} = X\boldsymbol{\beta}$ is with*

$$\boldsymbol{y} = \begin{pmatrix} -0.01 \\ 0.06 \end{pmatrix} \quad \textit{and} \quad X = \begin{pmatrix} 0.025 & 0.026 \\ -0.026 & -0.024 \end{pmatrix}.$$

The cost function is set to be

$$F(\boldsymbol{\beta}) = \lambda(\|\boldsymbol{\beta}\|_1 - \|\boldsymbol{\beta}\|_2) + \frac{1}{2}\|X\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$$

with $\lambda = 0.01$. The contour of the function $F(\boldsymbol{\beta})$ is presented in Figure 1. The columns vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are highly correlated. It is observed in Figure 1 that there exist two local minimizers. One local minimizer is $(\beta_1, 0)$. We see that the difference between the two coefficients $|\hat{\beta}_1 - \hat{\beta}_2|$ is large. The other local minimizer which is $(0, \beta_2)$ has the same property. Therefore, the assumption $\hat{\beta}_i\hat{\beta}_j > 0$ can not be removed in Theorem 1.

To solve the constrained l_{1-2} model (1.3), an iterative scheme based on the difference of convex algorithm (DCA) was proposed in [12]. The Lagrange multiplier formulation of the constrained l_{1-2} model (1.3) is

$$L(\boldsymbol{\beta}, \boldsymbol{\omega}) = \lambda(\|\boldsymbol{\beta}\|_1 - \|\boldsymbol{\beta}\|_2) + \boldsymbol{\omega}^{\mathsf{T}}(\boldsymbol{y} - X\boldsymbol{\beta}), \tag{2.5}$$

where $\boldsymbol{\omega}$ is the Lagrange multiplier. It was proved in [12] that the sequence generated by DCA

iterations converges to a stationary point which satisfies the first-order optimality condition:

$$\begin{cases} \mathbf{0} \in \lambda \operatorname{sgn}\left(\hat{\boldsymbol{\beta}}\right) - \lambda \dfrac{\hat{\boldsymbol{\beta}}}{\left\|\hat{\boldsymbol{\beta}}\right\|_2} - X^{\mathsf{T}}\boldsymbol{\omega}, \\ X\hat{\boldsymbol{\beta}} = \boldsymbol{y}. \end{cases} \tag{2.6}$$

For any given matrix X, let $\|X\|_2$ denote its spectral norm. The following results show that any stationary point which satisfies (2.6) has the grouping effect property.

Theorem 2. *Suppose that the response $\boldsymbol{y}$ is given and the measurement matrix X is standardized, Let $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)^{\mathsf{T}}$ satisfy* (2.6). *If $\hat{\beta}_i\hat{\beta}_j > 0$, then*

$$\left|\hat{\beta}_i - \hat{\beta}_j\right| \leq \sqrt{2(1-\rho)}\left(\left\|\hat{\boldsymbol{\beta}}\right\|_2 \sqrt{p\left\|\left(XX^{\mathsf{T}}\right)^{-1}\right\|_2} + \left\|\left(XX^{\mathsf{T}}\right)^{-1}\right\|_2 \|\boldsymbol{y}\|_2\right). \tag{2.7}$$

Proof. For every $i \neq j \in \{1, \ldots, p\}$ such that $\hat{\beta}_i\hat{\beta}_j > 0$, the first equation in (2.6) leads to

$$\lambda \operatorname{sgn}\left(\hat{\beta}_i\right) - \lambda \frac{\hat{\beta}_i}{\left\|\hat{\boldsymbol{\beta}}\right\|_2} - \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\omega} = 0, \tag{2.8}$$

$$\lambda \operatorname{sgn}\left(\hat{\beta}_j\right) - \lambda \frac{\hat{\beta}_j}{\left\|\hat{\boldsymbol{\beta}}\right\|_2} - \boldsymbol{x}_j^{\mathsf{T}}\boldsymbol{\omega} = 0. \tag{2.9}$$

Subtracting equation (2.8) from equation (2.9) gives

$$\lambda \frac{\hat{\beta}_i - \hat{\beta}_j}{\left\|\hat{\boldsymbol{\beta}}\right\|_2} = -\left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\boldsymbol{\omega}.$$

It follows that

$$\left|\hat{\beta}_i - \hat{\beta}_j\right| \leq \frac{\sqrt{2(1-\rho)}}{\lambda}\|\boldsymbol{\omega}\|_2 \left\|\hat{\boldsymbol{\beta}}\right\|_2. \tag{2.10}$$

Next we estimate $\|\boldsymbol{\omega}\|_2$. For the stationary point, there exists a vector $\boldsymbol{p} \in \partial\left(\left\|\hat{\boldsymbol{\beta}}\right\|_1\right)$ such that

$$X^{\mathsf{T}}\boldsymbol{\omega} = \lambda\left(\boldsymbol{p} - \frac{\hat{\boldsymbol{\beta}}}{\left\|\hat{\boldsymbol{\beta}}\right\|_2}\right).$$

Multiplying both sides by X, we have that

$$\boldsymbol{\omega} = \lambda\left((XX^{\mathsf{T}})^{-1}X\boldsymbol{p} - \frac{(XX^{\mathsf{T}})^{-1}X\hat{\boldsymbol{\beta}}}{\left\|\hat{\boldsymbol{\beta}}\right\|_2}\right) = \lambda\left((XX^{\mathsf{T}})^{-1}X\boldsymbol{p} - \frac{(XX^{\mathsf{T}})^{-1}\boldsymbol{y}}{\left\|\hat{\boldsymbol{\beta}}\right\|_2}\right).$$

Therefore,

$$\begin{aligned} \|\boldsymbol{\omega}\|_2 &\leq \lambda\left(\left\|(XX^{\mathsf{T}})^{-1}X\right\|_2 \|\boldsymbol{p}\|_2 + \frac{\left\|(XX^{\mathsf{T}})^{-1}\right\|_2 \|\boldsymbol{y}\|_2}{\left\|\hat{\boldsymbol{\beta}}\right\|_2}\right) \\ &\leq \lambda\left(\sqrt{p\,\|(XX^{\mathsf{T}})^{-1}\|_2} + \frac{\left\|(XX^{\mathsf{T}})^{-1}\right\|_2 \|\boldsymbol{y}\|_2}{\left\|\hat{\boldsymbol{\beta}}\right\|_2}\right). \end{aligned} \tag{2.11}$$

Note that $\|\boldsymbol{p}\|_2 \leq \sqrt{p}$ and

$$\begin{aligned}\left\|(XX^T)^{-1}X\right\|_2 &= \sqrt{\left\|\left((XX^T)^{-1}X\right)^T(XX^T)^{-1}X\right\|_2}\\ &= \sqrt{\left\|X^T(XX^T)^{-2}X\right\|_2}\\ &= \sqrt{\left\|(XX^T)^{-1}\right\|_2}. \end{aligned} \tag{2.12}$$

It follows from (2.10) and (2.11) that (2.7) holds. □

§3 Linearized Bregman Iteration

This section considers the l_{1+2} model that is implicated by the iterative scheme in [3]. A simple and fast iteration scheme based on linearized Bregman iteration was proposed to [23] find the sparse solution to the linear system (1.1). The convergence analysis for the linearized Bregman iteration was given in [3,4]. The limit of sequence generated by the linearized Bregman iteration was proved to be the unique solution of the convex minimization problem

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\left\{\lambda_1\|\boldsymbol{\beta}\|_1+\lambda_2\|\boldsymbol{\beta}\|_2^2\right\} \quad \text{subject to} \quad X\boldsymbol{\beta}=\boldsymbol{y}, \tag{3.1}$$

where λ_1 and λ_2 are positive parameters. The linearized Bregman iteration has led to a frame based deblurring algorithm in [3] and a low rank matrix completion algorithm in [1]. The following result says if $\hat{\beta}_i\hat{\beta}_j > 0$, then the solution of minimization problem (3.1) satisfies the grouping effect property.

Theorem 3. *[3, Theorem 4.6.] Given the response $\boldsymbol{y}$ and the standardized matrix X. Let $\hat{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{\beta}}_1,\ldots,\hat{\boldsymbol{\beta}}_p\right)^{\mathsf{T}}$ be the solution to the minimization program (3.1). If $\hat{\beta}_i\hat{\beta}_j > 0$, we have*

$$\left|\hat{\beta}_i-\hat{\beta}_j\right| \leq \frac{\sqrt{2(1-\rho)}}{2\lambda_2}\left(\lambda_1\sqrt{p\left\|\left(XX^{\mathsf{T}}\right)^{-1}\right\|_2}+2\lambda_2\left\|\left(XX^{\mathsf{T}}\right)^{-1}\right\|_2\|\boldsymbol{y}\|_2\right). \tag{3.2}$$

Motivated by the work in [24], we prove in this section that the conclusion of Theorem 3 still holds without the condition that $\hat{\beta}_i\hat{\beta}_j > 0$. We start with the following lemma on the non zero entries of the true vector. The Lagrange multiplier formulation of the model (3.1) is

$$L(\boldsymbol{\beta},\boldsymbol{\omega}) = \lambda_1\|\boldsymbol{\beta}\|_1+\lambda_2\|\boldsymbol{\beta}\|_2^2+\boldsymbol{\omega}^{\mathsf{T}}(\boldsymbol{y}-X\boldsymbol{\beta}), \tag{3.3}$$

where $\boldsymbol{\omega}$ is the Lagrange multiplier.

Lemma 4. *Suppose that $\hat{\boldsymbol{\beta}} \in \mathbb{R}^p$ is a solution of the program (3.1). Let $i \in \{1,\ldots,p\}$, if $\hat{\beta}_i \neq 0$, then*

$$\left|\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\omega}\right| = \lambda_1+2\lambda_2\left|\hat{\beta}_i\right| > \lambda_1,$$

and if $\hat{\beta}_i = 0$, then $\left|\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\omega}\right| \leq \lambda_1$.

Proof. Let the function F denote as the regularized empirical error in (3.1) by

$$F(\boldsymbol{\beta}) = \lambda_1\|\boldsymbol{\beta}\|_1+\lambda_2\|\boldsymbol{\beta}\|_2^2+\boldsymbol{\omega}^{\mathsf{T}}(\boldsymbol{y}-X\boldsymbol{\beta}).$$

Denote $\boldsymbol{e}_i \in \mathbb{R}^p$ the vector with the i-th component 1 and all the other components 0. Here

$X\boldsymbol{e}_i = \boldsymbol{x}_i$. Then for $\epsilon \in \mathbb{R}$ we have

$$F\left(\hat{\boldsymbol{\beta}} + \epsilon \boldsymbol{e}_i\right) - F\left(\hat{\boldsymbol{\beta}}\right) = \lambda_1 \left(\left|\hat{\beta}_i + \epsilon\right| - \left|\hat{\beta}_i\right|\right) + \lambda_2 \left(2\hat{\beta}_i \epsilon + \epsilon^2\right) - \epsilon \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega}. \tag{3.4}$$

We can consider the two different cases, for $i \in \{1, \ldots, p\}$. If $\hat{\beta}_i \neq 0$, we restrict

$$\epsilon \in \left(-\left|\hat{\beta}_i\right|, \left|\hat{\beta}_i\right|\right)$$

and see that both $\hat{\beta}_i + \epsilon$ and $\hat{\beta}_i$ have the same signs and

$$\left|\hat{\beta}_i + \epsilon\right| - \left|\hat{\beta}_i\right| = \operatorname{sgn}\left(\hat{\beta}_i\right)\left(\hat{\beta}_i + \epsilon - \hat{\beta}_i\right) = \operatorname{sgn}\left(\hat{\beta}_i\right)\epsilon. \tag{3.5}$$

Combining (3.4) and (3.5), we have

$$F\left(\hat{\boldsymbol{\beta}} + \epsilon \boldsymbol{e}_i\right) - F\left(\hat{\boldsymbol{\beta}}\right) = \left(\lambda_1 \operatorname{sgn}\left(\hat{\beta}_i\right) + 2\lambda_2 \hat{\beta}_i - \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega}\right)\epsilon + \lambda_2 \epsilon^2 \geq 0.$$

For $|\epsilon|$ is sufficiently small, we can derive $\lambda_1 \operatorname{sgn}\left(\hat{\beta}_i\right) + 2\lambda_2 \hat{\beta}_i - \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega} = 0$. Hence

$$\begin{aligned} \left|\boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega}\right| &= \left|\lambda_1 \operatorname{sgn}\left(\hat{\beta}_i\right) + 2\lambda_2 \hat{\beta}_i\right| \\ &= \left|\left(\lambda_1 + 2\lambda_2 \left|\hat{\beta}_i\right|\right) \operatorname{sgn}\left(\hat{\beta}_i\right)\right| \\ &= \lambda_1 + 2\lambda_2 \left|\hat{\beta}_i\right| \\ &> \lambda_1. \end{aligned} \tag{3.6}$$

If $\hat{\beta}_i = 0$, we see from (3.4) that

$$F\left(\hat{\boldsymbol{\beta}} + \epsilon \boldsymbol{e}_i\right) - F\left(\hat{\boldsymbol{\beta}}\right) = \lambda_1 |\epsilon| + \lambda_2 \epsilon^2 - \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega} \epsilon \geq 0.$$

When $\boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega} \neq 0$ and ϵ has the same sign, we have

$$\left(\lambda_1 - \left|\boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega}\right|\right)|\epsilon| + \lambda_2 \epsilon^2 \geq 0.$$

Let $|\epsilon|$ be sufficiently small, there have

$$\left|\boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega}\right| \leq \lambda_1. \tag{3.7}$$

□

Now we state the main result in this section. The proof can be viewed as a further analysis of the proof of [3, Theorem 4.6].

Theorem 5. *Suppose that the response $\boldsymbol{y}$ is given and the measurement matrix X is standardized, then a minimization solution $\hat{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_p\right)^{\mathsf{T}}$ of (3.1) satisfies (3.2).*

Proof. From the Lagrange multiplier formulation (3.3), solving (3.1) is equivalent to solving

$$\begin{cases} \mathbf{0} \in \lambda_1 \operatorname{sgn}\left(\hat{\boldsymbol{\beta}}\right) + 2\lambda_2 \hat{\boldsymbol{\beta}} - X^{\mathsf{T}} \boldsymbol{\omega}, \\ \boldsymbol{y} = X\hat{\boldsymbol{\beta}}. \end{cases} \tag{3.8}$$

Since the vector $\hat{\boldsymbol{\beta}}$ is the solution of (3.1), for each $i \neq j \in \{1, \ldots, p\}$, the first equation in (3.8) can be written as

$$0 \in \lambda_1 \operatorname{sgn}\left(\hat{\beta}_i\right) + 2\lambda_2 \hat{\beta}_i - \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\omega} \tag{3.9}$$

and

$$0 \in \lambda_1 \operatorname{sgn}\left(\hat{\beta}_j\right) + 2\lambda_2 \hat{\beta}_j - \boldsymbol{x}_j^{\mathsf{T}} \boldsymbol{\omega}. \tag{3.10}$$

Note that for all $\hat{\beta}_i \neq 0$ ($i \in \{1, \ldots, p\}$), the $\operatorname{sgn}\left(\hat{\beta}_i\right)$ contains only one element. So in the

equations of (3.9) and (3.10), we can use "=" instead of "$\in$". To prove the inequality (3.2), we consider the four different cases as follows.

Case 1. $\hat{\beta}_i$ and $\hat{\beta}_j$ are both non zero and have the same signs. Namely, $\operatorname{sgn}\left(\hat{\beta}_i\right) = \operatorname{sgn}\left(\hat{\beta}_j\right)$. This part has been shown in [3]. We include it for completeness. Subtracting equation (3.9) from equation (3.10) gives

$$0 = \lambda_1 \left(\operatorname{sgn}\left(\hat{\beta}_i\right) - \operatorname{sgn}\left(\hat{\beta}_j\right)\right) + 2\lambda_2 \left(\hat{\beta}_i - \hat{\beta}_j\right) - \left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\boldsymbol{\omega}.$$

Since $\operatorname{sgn}\left(\hat{\beta}_i\right) - \operatorname{sgn}\left(\hat{\beta}_j\right) = 0$, we find that

$$2\lambda_2 \left(\hat{\beta}_i - \hat{\beta}_j\right) - \left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\boldsymbol{\omega} = 0.$$

It is equal to

$$\hat{\beta}_i - \hat{\beta}_j = \frac{\left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\boldsymbol{\omega}}{2\lambda_2}.$$

Combining with (1.4), we conclude that

$$\left|\hat{\beta}_i - \hat{\beta}_j\right| = \left|\frac{\left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\boldsymbol{\omega}}{2\lambda_2}\right| \leq \frac{\sqrt{2\left(1-\rho\right)}\|\boldsymbol{\omega}\|_2}{2\lambda_2}. \tag{3.11}$$

Case 2. $\hat{\beta}_i$ and $\hat{\beta}_j$ are both non zero and have different signs. Then $\operatorname{sgn}\left(\hat{\beta}_i\right) = -\operatorname{sgn}\left(\hat{\beta}_j\right)$ and by (3.9) and (3.10), we see that

$$0 = 2\lambda_1 \operatorname{sgn}\left(\hat{\beta}_i\right) + 2\lambda_2 \left(\hat{\beta}_i - \hat{\beta}_j\right) - \left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\boldsymbol{\omega}.$$

Since $\operatorname{sgn}\left(\hat{\beta}_i - \hat{\beta}_j\right) = \operatorname{sgn}\left(\hat{\beta}_i\right)$, we have

$$\left(2\lambda_1 + 2\lambda_2 \left|\hat{\beta}_i - \hat{\beta}_j\right|\right)\operatorname{sgn}(\hat{\beta}_i) - \left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\boldsymbol{\omega} = 0.$$

Hence,

$$\left|\hat{\beta}_i - \hat{\beta}_j\right| = \left|\frac{\left(\boldsymbol{x}_i^{\mathsf{T}} - \boldsymbol{x}_j^{\mathsf{T}}\right)\boldsymbol{\omega}}{2\lambda_2}\right| - \frac{\lambda_1}{\lambda_2} \leq \frac{\sqrt{2(1-\rho)}\|\boldsymbol{\omega}\|_2}{2\lambda_2}. \tag{3.12}$$

Case 3. $\hat{\beta}_i \neq 0$ and $\hat{\beta}_j = 0$. Now we infer from (3.6) and (3.7) and get

$$\left|\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\omega}\right| = \lambda_1 + 2\lambda_2 \left|\hat{\beta}_i\right| \geq \left|\boldsymbol{x}_j^{\mathsf{T}}\boldsymbol{\omega}\right| + 2\lambda_2 \left|\hat{\beta}_i\right|.$$

After rearranging, we derive

$$\left|\hat{\beta}_i\right| \leq \frac{\left|\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\omega}\right| - \left|\boldsymbol{x}_j^{\mathsf{T}}\boldsymbol{\omega}\right|}{2\lambda_2}.$$

Since $\hat{\beta}_j = 0$, we have

$$\left|\hat{\beta}_i - \hat{\beta}_j\right| \leq \frac{\left|\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\omega}\right| - \left|\boldsymbol{x}_j^{\mathsf{T}}\boldsymbol{\omega}\right|}{2\lambda_2} \leq \frac{\sqrt{2(1-\rho)}}{2\lambda_2}\|\boldsymbol{\omega}\|_2. \tag{3.13}$$

Case 4. If $\hat{\beta}_i = \hat{\beta}_j = 0$, then (3.2) is trivial.

Next we estimate $\|\boldsymbol{\omega}\|_2$. The first equation in (3.8) gives

$$X^{\mathsf{T}}\boldsymbol{\omega} = \lambda_1 \operatorname{sgn}\left(\hat{\boldsymbol{\beta}}\right) + 2\lambda_2\hat{\boldsymbol{\beta}}.$$

Multiplying both sides by X, we can derive that

$$\boldsymbol{\omega} = \lambda_1 \left(XX^{\mathsf{T}}\right)^{-1} X \operatorname{sgn}\left(\hat{\boldsymbol{\beta}}\right) + 2\lambda_2 \left(XX^{\mathsf{T}}\right)^{-1} X\hat{\boldsymbol{\beta}}.$$

Therefore,

$$\begin{aligned}\|\boldsymbol{\omega}\|_2 &= \left\|\lambda_1 \left(XX^T\right)^{-1} X \operatorname{sgn}\left(\hat{\boldsymbol{\beta}}\right) + 2\lambda_2 \left(XX^{\mathsf{T}}\right)^{-1} \boldsymbol{y}\right\|_2 \\ &\leq \lambda_1 \left\|\left(XX^{\mathsf{T}}\right)^{-1} X\right\|_2 \left\|\operatorname{sgn}\left(\hat{\boldsymbol{\beta}}\right)\right\|_2 + 2\lambda_2 \left\|\left(XX^{\mathsf{T}}\right)^{-1}\right\|_2 \|\boldsymbol{y}\|_2. \end{aligned} \tag{3.14}$$

Because each entry of $\operatorname{sgn}\left(\hat{\boldsymbol{\beta}}\right)$ is in $[-1,1]$, we have $\left\|\operatorname{sgn}\left(\hat{\boldsymbol{\beta}}\right)\right\|_2 \leq \sqrt{p}$. By combining (3.11) ((3.12) or (3.13)), (3.14) and (2.12) together, we finally obtain

$$\left|\hat{\beta}_i - \hat{\beta}_j\right| \leq \left(\lambda_1 \sqrt{p \left\|\left(XX^T\right)^{-1}\right\|_2} + 2\lambda_2 \left\|\left(XX^T\right)^{-1}\right\|_2 \|\boldsymbol{y}\|_2\right) \frac{\sqrt{2(1-\rho)}}{2\lambda_2}.$$

□

§4 Sparse Approximation Property

The stability of convex model with correlated measurements was established in [13–15, 20]. This section establishes the stability of the l_{1-2} model with the correlated Gaussian random matrices. The *correlated Gaussian random matrix* X_Σ is with independent and identically distributed rows obeying the distribution $\mathcal{N}(0,\Sigma)$ with zero mean and a covariance matrix Σ. The square root of the smallest eigenvalue of Σ is denoted by $\rho_1(\Sigma)$ and its maximal variance is denoted by $\rho_2^2(\Sigma)$, respectively. Throughout this section, the measurement matrix that is normalized by

$$X = \frac{1}{\sqrt{n}} X_\Sigma$$

with Σ being a positive definite matrix. It was proved in [14, Theorem 1] that the following inequlality

$$\|X\boldsymbol{\beta}\|_2 \geq \kappa_1 \|\boldsymbol{\beta}\|_2 - \kappa_2 \sqrt{\frac{\log p}{n}} \|\boldsymbol{\beta}\|_1 \quad \text{for all} \quad \boldsymbol{\beta} \in \mathbb{R}^p \tag{4.1}$$

holds with high probability with two positive constants $\kappa_1 = \frac{\rho_1(\Sigma)}{4}$ and $\kappa_2 = 9\rho_2(\Sigma)$. It was illustrated in [14] that many types of matrices satisfy (4.1) such as Toeplitz matrices, spiked identity models.

Let $\boldsymbol{\beta}_S$ denote the vector whose s-th entry is equal to s-th entry of $\boldsymbol{\beta}$ for s in S and equal to zero otherwise. The sparse approximation property was used to provide sufficient conditions for the stable recovery of the basis pursuit problem in [9, 17]. The relationship between the sparse approximation property in [9, 17] and the inequality (4.1) was discussed in [15].

Proposition 6. *[15, Proposition 4.] If the number of measurements n satisfies*

$$n > 2(\kappa_2/\kappa_1)^2 s \log p, \tag{4.2}$$

then with probability at least $1 - c_3 \exp(-c_4 n)$, the normalized measurement matrix satisfies the sparse approximation property,

$$\|\boldsymbol{\beta}_S\|_2 \leq c_1 \|X\boldsymbol{\beta}\|_2 + c_2 \frac{\|\boldsymbol{\beta} - \boldsymbol{\beta}_S\|_1}{\sqrt{2s}}, \qquad \textit{for all } \boldsymbol{\beta} \in \mathbb{R}^p \textit{ and } \#S \leq 2s. \tag{4.3}$$

with

$$c_1 = \frac{1}{\kappa_1 - \kappa_2 \sqrt{\frac{2s\log p}{n}}}, \quad c_2 = \frac{\kappa_2 \sqrt{\frac{2s\log p}{n}}}{\kappa_1 - \kappa_2 \sqrt{\frac{2s\log p}{n}}}, \tag{4.4}$$

c_3 and c_4 are positive constants.

We denote

$$a(s)=\frac{\sqrt{s}-1}{\sqrt{s}+1},\quad b(s)=1+\frac{\sqrt{2}}{\sqrt{s}-1}.$$

The performance guarantees of the l_{1-2} model are stated as follows.

Theorem 7. *Consider the linear regression model* (1.1) *with*

$$n>72(\kappa_2/\kappa_1)^2 s\log p. \tag{4.5}$$

Let $\boldsymbol{\beta}_s^*$ be the best s-sparse approximation of $\boldsymbol{\beta}^*$ with $s\geq 2$. If the noise vector $\boldsymbol{z}$ is bounded by $\|\boldsymbol{z}\|_2\leq\varepsilon$, then the minimizer $\hat{\boldsymbol{\beta}}$ of the following programming:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\{\|\boldsymbol{\beta}\|_1-\|\boldsymbol{\beta}\|_2\}\quad \textit{subject to}\quad \|X\boldsymbol{\beta}-\boldsymbol{y}\|_2\leq\varepsilon, \tag{4.6}$$

satisfies

$$\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_2\leq C_5\varepsilon+C_6\frac{\|\boldsymbol{\beta}^*-\boldsymbol{\beta}_s^*\|_1}{\sqrt{s}} \tag{4.7}$$

with probability at least $1-c_3\exp(-c_4 n)$, where positive constants

$$C_5=\left(1+\frac{1}{a(s)}\right)\frac{2c_1}{1-c_2 b(s)}$$

and

$$C_6=\left(1+\frac{1}{a(s)}\right)\left(\frac{\sqrt{s}}{\sqrt{s}-1}\right)\frac{\sqrt{2}c_2}{1-c_2 b(s)}+\frac{2\sqrt{s}}{\sqrt{s}-1}$$

depend on κ_1, κ_2, s and p.

Proof. Let $\boldsymbol{h}=\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*$. It follows that

$$\begin{aligned}\|X\boldsymbol{h}\|_2&=\|X(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\|_2\\&=\|(X\hat{\boldsymbol{\beta}}-\boldsymbol{y})-(X\boldsymbol{\beta}^*-\boldsymbol{y})\|_2\\&\leq\|X\boldsymbol{\beta}^*-\boldsymbol{y}\|_2+\|X\hat{\boldsymbol{\beta}}-\boldsymbol{y}\|_2\\&\leq 2\varepsilon.\end{aligned}$$

For any $S\subset\{1,\ldots,n\}$, we have

$$\begin{aligned}0&\geq\left(\|\hat{\boldsymbol{\beta}}\|_1-\|\boldsymbol{\beta}^*\|_1\right)-\left(\|\hat{\boldsymbol{\beta}}\|_2-\|\boldsymbol{\beta}^*\|_2\right)\\&\geq\|(\boldsymbol{\beta}^*+\boldsymbol{h})_S\|_1+\|(\boldsymbol{\beta}^*+\boldsymbol{h})_{S^c}\|_1-(\|\boldsymbol{\beta}_S^*\|_1+\|\boldsymbol{\beta}_{S^c}^*\|_1)-\|\boldsymbol{h}\|_2\\&\geq\|\boldsymbol{h}_{S^c}\|_1-\|\boldsymbol{h}_S\|_1-2\|\boldsymbol{\beta}_{S^c}^*\|_1-\|\boldsymbol{h}\|_2.\end{aligned}$$

It follows that

$$\|\boldsymbol{h}_{S^c}\|_1\leq\|\boldsymbol{h}_S\|_1+\|\boldsymbol{h}_S\|_2+\|\boldsymbol{h}_{S^c}\|_2+2\|\boldsymbol{\beta}_{S^c}^*\|_1. \tag{4.8}$$

By the index set Λ, we denote the locations of the s largest entries of the vector $\boldsymbol{\beta}^*$ in magnitude. Using the standard decomposition method in compressed sensing, we let Λ_1 denote locations of the s largest entries in magnitude of $\boldsymbol{h}$ in Λ^c where Λ^c is the complement set of Λ. Then Λ_2 denotes the locations of the s largest entries into magnitude of $\boldsymbol{h}$ in $(\Lambda\cup\Lambda_1)^c$, and so on. A key inequality proved in [22] is

$$\sum_{i\geq 2}\|\boldsymbol{h}_{\Lambda_i}\|_2\leq\frac{\|\boldsymbol{h}_{\Lambda^c}\|_1-\|\boldsymbol{h}_{\Lambda^c}\|_2}{\sqrt{s}-1}. \tag{4.9}$$

Let $T = \Lambda \cup \Lambda_1$, then $T^c = \cup_{i\geq 2}\Lambda_i$. It follows from (4.8) and (4.9) that

$$\begin{aligned}\|\boldsymbol{h}_{T^c}\|_2 \leq \sum_{i\geq 2}\|\boldsymbol{h}_{\Lambda_i}\|_2 &\leq \frac{\|\boldsymbol{h}_{\Lambda^c}\|_1 - \|\boldsymbol{h}_{\Lambda^c}\|_2}{\sqrt{s}-1}\\ &\leq \frac{\|\boldsymbol{h}_{\Lambda}\|_1 + \|\boldsymbol{h}_{\Lambda}\|_2 + 2\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{s}-1}\\ &\leq \frac{(\sqrt{s}+1)\|\boldsymbol{h}_{\Lambda}\|_2 + 2\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{s}-1}.\end{aligned} \tag{4.10}$$

It follows from (4.9) and (4.10) that

$$\begin{aligned}\|\boldsymbol{h}\|_2 &= \|\boldsymbol{h}_T + \boldsymbol{h}_{T^c}\|_2\\ &\leq \|\boldsymbol{h}_T\|_2 + \sum_{i\geq 2}\|\boldsymbol{h}_{\Lambda_i}\|_2\\ &\leq \|\boldsymbol{h}_T\|_2 + \frac{(\sqrt{s}+1)\|\boldsymbol{h}_{\Lambda}\|_2 + 2\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{s}-1}\\ &\leq \|\boldsymbol{h}_T\|_2 + \frac{(\sqrt{s}+1)\|\boldsymbol{h}_{T}\|_2 + 2\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{s}-1}\\ &= \left(1+\frac{\sqrt{s}+1}{\sqrt{s}-1}\right)\|\boldsymbol{h}_T\|_2 + \frac{2}{\sqrt{s}-1}\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1.\end{aligned} \tag{4.11}$$

Together with sparse approximation property we have

$$\begin{aligned}\|\boldsymbol{h}_T\|_2 &\leq c_1\|X\boldsymbol{h}\|_2 + c_2\frac{\|\boldsymbol{h}_{T^c}\|_1}{\sqrt{2s}},\\ &\leq 2c_1\varepsilon + c_2\frac{\|\boldsymbol{h}_T\|_1 + \|\boldsymbol{h}_T\|_2 + \|\boldsymbol{h}_{T^c}\|_2 + 2\|\boldsymbol{\beta}^*_{T^c}\|_1}{\sqrt{2s}}\\ &\leq 2c_1\varepsilon + c_2\frac{\|\boldsymbol{h}_T\|_1 + \|\boldsymbol{h}_T\|_2 + \|\boldsymbol{h}_{T^c}\|_2 + 2\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{2s}}\\ &\leq 2c_1\varepsilon + c_2\left[\left(1+\frac{1}{\sqrt{2s}}\right)\|\boldsymbol{h}_T\|_2 + \frac{1}{\sqrt{2s}}\|\boldsymbol{h}_{T^c}\|_2 + 2\frac{\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{2s}}\right]\\ &\leq 2c_1\varepsilon + c_2\left[\left(1+\frac{1}{\sqrt{2s}}\right)\|\boldsymbol{h}_T\|_2 + \frac{1}{\sqrt{2s}}\frac{(\sqrt{s}+1)\|\boldsymbol{h}_{\Lambda}\|_2 + 2\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{s}-1} + 2\frac{\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{2s}}\right]\\ &\leq 2c_1\varepsilon + c_2\left[\left(1+\frac{1}{\sqrt{2s}}\right)\|\boldsymbol{h}_T\|_2 + \frac{1}{\sqrt{2s}}\frac{(\sqrt{s}+1)\|\boldsymbol{h}_{T}\|_2 + 2\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{s}-1} + 2\frac{\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{2s}}\right].\end{aligned} \tag{4.12}$$

The inequality (4.12) implies that

$$\|\boldsymbol{h}_T\|_2 \leq \frac{2c_1\varepsilon}{1-c_2b(s)} + \frac{\sqrt{2}c_2}{1-c_2b(s)}\left(\frac{\sqrt{s}}{\sqrt{s}-1}\right)\frac{\|\boldsymbol{\beta}^*_{\Lambda^c}\|_1}{\sqrt{s}}. \tag{4.13}$$

The inequality (4.7) follows from (4.11) and (4.13). A sufficient condition of (4.6) is

$$1 - c_2b(s) > 0,$$

which is implied by

$$c_2 < 0.2. \tag{4.14}$$

Using (4.4), we verify that (4.5) is a sufficient condition for (4.14) holds. □

References

[1] J F Cai, E J Candès, Z W Shen. *A Singular Value Thresholding Algorithm for Matrix Completion*, Siam Journal on Optimization, 2010, 20(4): 1956-1982.

[2] J F Cai, R H Chan, Z W Shen. *A framelet-based image inpainting algorithm*, Applied and Computational Harmonic Analysis, 2008, 24(2): 131-149.

[3] J F Cai, S Osher, Z W Shen. *Linearized Bregman iterations for compressed sensing*, Mathematics of Computation, 2009, 78(267): 1515-1536.

[4] J F Cai, S Osher, Z W Shen. *Convergence of the Linearized Bregman Iteration for ℓ_1-norm Minimization*, Mathematics of Computation, 2009, 78(268): 2127-2136.

[5] E J Candès, T Tao. *Decoding by linear programming*, IEEE Transactions on information theory, 2005, 51(12): 4203-4215.

[6] W Chen, P Li. *Truncated sparse approximation property and truncated q-norm minimization*, Applied Mathematics-A Journal of Chinese Universities Series B, 2019, 34(3): 261-283.

[7] D L Donoho, X Huo. *Uncertainty principles and ideal atomic decomposition*, IEEE Transactions on information theory, 2001, 47(7): 2845-2862.

[8] D L Donoho. *Compressed sensing*, IEEE Transactions on information theory, 2006, 52(4): 1289-1306.

[9] S Foucart. *Stability and robustness of ℓ_1-minimizations with weibull matrices and redundant dictionaries*, Linear Algebra and its Applications, 2014, 441: 4-21.

[10] Y Gao, J G Peng, S G Yue. *Sparse recovery in probability via l_q-minimization with Weibull random matrices for $0 < q \leq 1$*, Applied Mathematics-A Journal of Chinese Universities, 2018, 33(1): 1-24.

[11] Y F Lou, S Osher, J Xin. *Computational Aspects of Constrained L_1-L_2 Minimization for Compressive Sensing*, Advances in Intelligent Systems and Computing, 2015, 359: 169-180.

[12] Y F Lou, P h Yin, Q He, J Xin. *Computing sparse representation in a highly coherent dictionary based on difference of L_1 and L_2*, Journal of Scientific Computing, 2015, 64(1): 178-196.

[13] S N Negahban, P Ravikumar, M J Wainwright, B Yu. *A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizer*, Statistical Science, 2012, 27(4): 538-557.

[14] G Raskutti, M J Wainwright, B Yu. *Restricted eigenvalue properties for correlated gaussian designs*, The Journal of Machine Learning Research, 2010, 11: 2241-2259.

[15] Y Shen, B Han, E Braverman. *Stability of the Elastic Net Estimator*, Journal of Complexity, 2016, 32(1): 20-39.

[16] Y Shen, B Han, E Braverman. *Removal of Mixed Gaussian and Impulse Noise Using Directional Tensor Product Complex Tight Framelets*, Journal of Mathematical Imaging and Vision, 2016, 54(1): 64-77.

[17] Q Sun. *Sparse approximation property and stable recovery of sparse signals from noisy measurements*, IEEE Transactions on Signal Processing, 2011, 59(10): 5086-5090.

[18] R Tibshirani. *Regression shrinkage and selection via the lasso*, Journal of the Royal Statal Society: Series B(Methodological), 1996, 58(1): 267-288.

[19] J Wen, J Weng, Tong C. *Sparse Signal Recovery With Minimization of 1-Norm Minus 2-Norm*, IEEE Transactions on Vehicular Technology, 2019, 68(7): 6847-6854.

[20] Y Xia, S Li. *Analysis Recovery With Coherent Frames and Correlated Measurements*, IEEE Transactions on Information Theory, 2016, 62(11): 6493-6507.

[21] Y Xia, S Li. *Identifiability of Multichannel Blind Deconvolution and Nonconvex Regularization Algorithm*, IEEE Transactions on Signal Processing, 2018, 66(20): 5299-5312.

[22] P Yin, Y Lou, Q He, J Xin. *Minimization of ℓ_{1-2} for compressed sensing*, Siam Journal on Scientific Computing, 2015, 37(1): 536-563.

[23] W Yin, S Osher, D Goldfarb, J Darbon. *Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing*, SIAM J Imaging Sci, 2008, 1(1): 143-168.

[24] D X Zhou. *On grouping effect of elastic net*, Statistics & Probability Letters, 2013, 83(9): 2108-2112.

[25] H Zou, T Hastie. *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society, Series B: Statistical Methodology, 2005, 67(2): 301-320.

[26] H Zou, T Hastie, R Tibshirani. *On the degrees of freedom of the lasso*, The Annals of Statistics, 2007, 35(5): 2173-2192.

[1]The Department of Mathematics, Zhejiang Sci-Tech University, Hangzhou 310018, China.
E-mail: yshen@zstu.edu.cn, wl-guo@foxmail.com

[2]School of Information Engineering, Jiaxing Nanhu University, Jiaxing 314001, China.
E-mail: ruifanghu@qq.com