Journal of Computational Mathematics
Vol.xx, No.x, 2024, 1–32.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2310-m2022-0282

ORACLE INEQUALITIES FOR CORRUPTED COMPRESSED SENSING*

Liping Yin and Peng Li[1)]
School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China
Email: yinlp21@lzu.edu.cn, lp@lzu.edu.cn

Abstract

In this paper, we establish the oracle inequalities of highly corrupted linear observations $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e} \in \mathbb{R}^m$. Here the vector $\mathbf{x}_0 \in \mathbb{R}^n$ with $n \gg m$ is a (approximately) sparse signal and $\mathbf{f}_0 \in \mathbb{R}^m$ is a sparse error vector with nonzero entries that can be possible infinitely large, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$ represents the Gaussian random noise vector. We extend the oracle inequality $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 \lesssim \sum_i \min\{|x_0(i)|^2, \sigma^2\}$ for Dantzig selector and Lasso models in [E.J. Candès and T. Tao, Ann. Statist., 35 (2007), 2313–2351] and [T.T. Cai, L. Wang, and G. Xu, IEEE Trans. Inf. Theory, 56 (2010), 3516–3522] to $\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2+\|\hat{\mathbf{f}}-\mathbf{f}_0\|_2^2 \lesssim \sum_i \min\{|x_0(i)|^2, \sigma^2\} + \sum_j \min\{|\lambda f_0(j)|^2, \sigma^2\}$ for the extended Dantzig selector and Lasso models. Here $(\hat{\mathbf{x}}, \hat{\mathbf{f}})$ is the solution of the extended model, and $\lambda > 0$ is the balance parameter between $\|\mathbf{x}\|_1$ and $\|\mathbf{f}\|_1$, i.e. $\|\mathbf{x}\|_1 + \lambda\|\mathbf{f}\|_1$.

Mathematics subject classification: Primary 94A12, 62G05, Secondary 90C25.
Key words: Corrupted compressed sensing, Oracle inequality, Extended Dantzig selector, Extended Lasso, Generalized restricted isometry property.

1. Introduction

1.1. Corrupted compressed sensing problem

Over the past twenty years, the idea of compressed sensing has received extensive attention and has been employed in several potential technologies [8, 10]. It offers an excellent strategy for reconstructing a (approximately) sparse signal from a few observations. In particular, an s-sparse signal $\mathbf{x}_0 \in \mathbb{R}^n$ is evaluated by

$$\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}, \tag{1.1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \ll n$ is the sensing matrix, $\mathbf{b} \in \mathbb{R}^m$ denotes the observation vector and $\mathbf{e} \in \mathbb{R}^m$ is the possible noise vector.

The following optimization problem:

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_0 \\ &\text{s.t.} \quad \mathbf{A}\mathbf{x} - \mathbf{b} \in \mathcal{C}(\eta) \end{aligned}$$

provides a good estimator for the reconstruction of $\mathbf{x}_0$. Here $\|\mathbf{x}\|_0 = |\{i : x_i \neq 0\}|$ expresses the sparsity of $\mathbf{x}$, $\mathcal{C}(\eta)$ is a bounded set with the parameter $\eta > 0$ determined by the error

* Received December 15, 2022 / Revised version received March 30, 2023 / Accepted October 27, 2023 / Published online March 1, 2024 /
1) Corresponding author

structure, for example, $\mathcal{C}(\eta) = \{\mathbf{z} : \|\mathbf{z}\|_2 \leq \eta\}$ or $\mathcal{C}(\eta) = \{\mathbf{z} : \|\mathbf{A}^\top\mathbf{z}\|_\infty \leq \eta\}$ [11]. Here and following, we use the notation $\mathbf{A}^\top \in \mathbb{R}^{n\times m}$ denotes the transposition of the matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$. There exist some convex methods to solve this optimization problem. The method of basis pursuit [13, 14] transformed the ℓ_0-minimization $\|\mathbf{x}\|_0$ to its relative convex ℓ_1-minimization $\|\mathbf{x}\|_1$ ($\|\mathbf{x}\|_1 = \sum_i |x_i|$), solved the non-deterministic polynomial (NP) hard problem. Candès and Tao [9] proved that the original signal $\mathbf{x}_0$ can be exactly recovered by solving that ℓ_1-minimization problem. Based on this, a number of methods for different noise types have been proposed, such as Lasso [41], quadratically constrained basis pursuit [18], Dantzig selector [11], and RLAD [44,47]. Extensive studies appear under different frameworks, such as the null space property [17, 22, 39], the restricted isometry property (RIP) [5, 15, 16, 49], and the coherence [4, 19, 28, 29, 42], solving this problem.

When certain unknown items of the observation vector are badly distorted, we can get a novel method inspired by the above classic compressed sensing issue. In mathematics, we have

$$\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}. \tag{1.2}$$

Here $\mathbf{f}_0 \in \mathbb{R}^m$ is a corrupted error, which is unidentified and cannot be disregarded. Corrupted compressed sensing is the issue of reconstructing the sparse signal $\mathbf{x}_0$ and sparse error $\mathbf{f}_0$ from the observations (1.2). Laska *et al.* [25] first considered recovering the signal and the corruption from corrupted measurements and designed an algorithm dubbed Justice Pursuit. They extended the classical RIP to the generalized restricted isometry property (GRIP) as follows.

Definition 1.1. *For any matrix $\mathbf{\Phi} = [\mathbf{A}, \mathbf{I}] \in \mathbb{R}^{m\times(n+m)}$, the (s,t)-GRIP-constant $\delta_{s,t}$ is defined as the infimum of δ such that*

$$(1-\delta)\left(\|\mathbf{x}\|_2^2 + \|\mathbf{f}\|_2^2\right) \leq \left\|\mathbf{\Phi}\begin{bmatrix}\mathbf{x}\\ \mathbf{f}\end{bmatrix}\right\|_2^2 \leq (1+\delta)\left(\|\mathbf{x}\|_2^2 + \|\mathbf{f}\|_2^2\right)$$

holds for any $\mathbf{x} \in \mathbb{R}^n$ with $|\mathrm{supp}(\mathbf{x})| \leq s$ and $\mathbf{f} \in \mathbb{R}^m$ with $|\mathrm{supp}(\mathbf{f})| \leq t$.

As a nontrivial extension of compressed sensing, the corrupted compressed sensing problem has been used in various practical fields, such as super-resolution and inpainting [33], signal recovery from the impulsive observations [36], signal separation [21].

In recent years, many breakthroughs have been obtained in the research of the corrupted compressed sensing problem. In the absence of the noise $\mathbf{e}$, Wright and Ma [45] proposed to recover the signal $\mathbf{x}_0$ and the corruption $\mathbf{f}_0$ from the observations $\mathbf{b}$ in (1.2) by solving the following problem:

$$\begin{aligned}&\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{f}\in\mathbb{R}^m} \|\mathbf{x}\|_1 + \|\mathbf{f}\|_1\\ &\text{s.t.} \quad \mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{f}.\end{aligned}$$

Considering the general situation with random noise $\mathbf{e}$, being tiny, Lin and Li [31] proposed to recover the sparse signal from the corrupted observations (1.2) with coherent tight frames via separation analysis Dantzig selector (SADS)

$$\begin{aligned}&\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{f}\in\mathbb{R}^m} \|\mathbf{D}^\top\mathbf{x}\|_1 + \|\mathbf{\Omega}^\top\mathbf{f}\|_1\\ &\text{s.t.} \quad \left\|\mathbf{W}^\top[\mathbf{A},\mathbf{I}]^\top(\mathbf{A}\mathbf{x} + \mathbf{f} - \mathbf{b})\right\|_\infty \leq \eta,\end{aligned}$$

and separation analysis Lasso (SALasso)

$$\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{f}\in\mathbb{R}^m} \frac{1}{2\rho}\|\mathbf{A}\mathbf{x}+\mathbf{f}-\mathbf{b}\|_2^2+\|\mathbf{D}^\top\mathbf{x}\|_1+\|\mathbf{\Omega}^\top\mathbf{f}\|_1.$$

Here $\eta \geq 0$ is the noise boundedness parameter, ρ represents a regularized parameter, the matrices $\mathbf{D}\in\mathbb{R}^{n\times d}, \mathbf{\Omega}\in\mathbb{R}^{m\times M}$ and $\mathbf{W}=[\mathbf{D},\mathbf{0};\mathbf{0},\mathbf{\Omega}]\in\mathbb{R}^{(n+m)\times(M+d)}$ are the tight frames for $\mathbb{R}^n, \mathbb{R}^m$ and $\mathbb{R}^{n+m}$, respectively.

To improve the robustness of the model, some authors proposed to add a balance parameter between $\|\mathbf{x}\|_1$ and $\|\mathbf{f}\|_1$. For instance, Nguyen and Tran [35] demonstrated that by choosing an appropriate balance parameter $\lambda > 0$, the linear programming

$$\begin{aligned}&\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{f}\in\mathbb{R}^m} \|\mathbf{x}\|_1+\lambda\|\mathbf{f}\|_1\\ &\text{s.t.}\quad \mathbf{A}\mathbf{x}+\mathbf{f}=\mathbf{b}\end{aligned}$$

can accurately recover both signal $\mathbf{x}_0$ and error $\mathbf{f}_0$, even when the sparsity of $\mathbf{x}_0$ grows almost linearly in the dimension of signal and the errors in $\mathbf{f}_0$ are up to a constant fraction of all the entries. What we should point out is that, when $\|\mathbf{x}\|_1/\|\mathbf{f}\|_1 \ll 1$ or $\gg 1$, the choice of the value of λ is vital to recover both signal $\mathbf{x}_0$ and error $\mathbf{f}_0$.

Soon afterwards, inspired by the traditional Lasso model, Nguyen and Tran [34] established the extended Lasso model for corrupted compressed sensing on the noisy case as follows:

$$\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{f}\in\mathbb{R}^m} \frac{1}{2\rho}\|\mathbf{b}-\mathbf{A}\mathbf{x}-\mathbf{f}\|_2^2+\|\mathbf{x}\|_1+\lambda\|\mathbf{f}\|_1.$$

Li [30] developed a sufficient condition for signal's stable recovery in the framework of GRIP. He required that the balance parameter λ is in the interval $[\sqrt{s/t}/c, c\sqrt{s/t}]$ with $c \geq 1$, where s and t are the sparsity of $\mathbf{x}_0$ and $\mathbf{f}_0$, respectively. Especially, when $s \in \lfloor \alpha m/(1+\log(n/m))\rfloor$ and $t \in \lfloor \alpha m \rfloor$, the author in [30] took the parameter $\lambda = 1/\sqrt{1+\log(n/m)}$.

Later, Li *et al.* [27] proposed the extended Dantzig selector model as follows:

$$\begin{aligned}&\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{f}\in\mathbb{R}^m} \|\mathbf{x}\|_1+\lambda\|\mathbf{f}\|_1\\ &\text{s.t.}\quad \left\|[\mathbf{A},\mathbf{I}]^\top(\mathbf{A}\mathbf{x}+\mathbf{f}-\mathbf{b})\right\|_\infty \leq \eta.\end{aligned} \tag{1.3}$$

The balance parameter λ is limited within the range of $[\sqrt{s/t}/c_1, c_2\sqrt{s/t}]$ with $c_1, c_2 \geq 1$.

In recent years, Wan *et al.* [43] introduced a novel Bayesian technique for robust corrupted compressed sensing. Adcock *et al.* [1] offered a novel theoretical argument for the extended ℓ_1-minimization method that seeks to recover sparse expansion coefficients in the presence of corrupted measurements. For more works on corrupted compressed sensing, readers can refer to [26, 37, 38, 48].

1.2. Oracle inequalities

In the technology of wavelet thresholding for signal de-noising, Donoho and Johnstone [20] first introduced the oracle inequality conception. By comparing the performance of the real estimator with that of the hypothetical estimator, it offers an useful tool for determining how accurately the estimation process performs. Later, it has been used to inverse problems [12], statistical estimation [6] and so on.

Candès [6] gave a detailed explanation of the statistical implications of the oracle inequality. In short, it can be interpreted as a bridge between the performance of the actual estimator and the ideal estimator, which is achieved with perfect information supplied by the oracle, and which is not available in reality. There is a simple example offered by Candès [6], which helps us to comprehend the oracle inequality. Suppose $\mathbf{y} \sim \mathcal{N}(\theta, \mathbf{I}_m)$. Now we need to estimate the parameter $\theta \in \mathbb{R}^m$ through the observations $\mathbf{y} \in \mathbb{R}^m$. There is a family of estimators $\hat{\theta}^c = c \cdot \mathbf{y}$. And the mean-squared error (MSE) of them are written as $\mathrm{MSE}(\hat{\theta}^c, \theta) = \mathbb{E}\|\hat{\theta}^c - \theta\|_2^2$, which is used to measure the performance of the estimators $\hat{\theta}^c$. Calculating the minimum MSE leads to $c^* = \arg\min_c \mathrm{MSE}(\hat{\theta}^c, \theta)$, i.e. $c^* = g_1(\theta) = \|\theta\|_2^2/(\|\theta\|_2^2 + m)$. This implies that we can obtain the best estimator with the help of an oracle that tells us the true parameters. Let $\hat{\theta} = g_2(\mathbf{y}) \cdot \mathbf{y}$ be a practical estimator. It is clear that $\mathrm{MSE}(\hat{\theta}, \theta) \geq \inf_c \mathrm{MSE}(\hat{\theta}^c, \theta)$. Suppose that the parameter $\hat{\theta}$ obeys $\mathrm{MSE}(\hat{\theta}, \theta) \leq \nu + \inf_c \mathrm{MSE}(\hat{\theta}^c, \theta)$, where ν is a constant. It means that the estimator $\hat{\theta}$ nearly has the performance as good as if we could know the best model estimator with the help of an oracle. So we call it an oracle inequality.

Candès and Tao [11] applied the oracle inequality approach to study the compressed sensing. The using of the oracle inequality is extremely significant for compressed sensing. Suppose that $\mathbf{x}_0$ is highly small so that $\mathbf{x}_0$ falls considerably below the noise level, i.e. $|x_0(i)| \ll \sigma$ for all i. Setting $\hat{\mathbf{x}} = \mathbf{0}$ in this case would result in a squared error loss of only $\sum_{i=1} |x_0(i)|^2$, which might be considerably less than σ^2 times the sparsity of $\mathbf{x}_0$. And they also considered the observations $\mathbf{b}$ in (1.1) with $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$. Under the framework of RIP, they developed an oracle inequality for the Dantzig selector model

$$\begin{aligned} &\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 \\ &\text{s.t.} \quad \|\mathbf{A}^\top(\mathbf{A}\mathbf{x} - \mathbf{b})\|_\infty \leq \eta_*, \end{aligned}$$

where $\eta_* = (1 + t^{-1})\sqrt{2\log m} \cdot \sigma$ and t is a positive scalar. Suppose that $\mathbf{A}$ is a matrix with ℓ_2-unit-norm columns, and $\mathbf{x}_0$ is sufficiently sparse. Then Candès and Tao showed that the estimator error

$$\left\|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\right\|_2^2 \lesssim \left(\sigma^2 + \sum_{i=1}^m \min\left\{|x_0(i)|^2, \sigma^2\right\}\right) \tag{1.4}$$

holds with high probability. Here and follows, we use $A \lesssim B$ to denote $A \leq C_0 B$ for any $A, B \in \mathbb{R}$, where $C_0 \in \mathbb{R}^+$ is an absolute constant and the value varies with the constant. The notation $\gtrsim$ can be defined similarly. Later, Cai *et al.* [3] established the oracle inequality for the Dantzig selector under the condition of the mutual incoherence property. Setting

$$\eta_* = \sigma\left(\sqrt{2\log m} + \frac{3}{2}\right),$$

they came to the conclusion (1.4) with high probability. Recently, based on the Lasso model, Li and Chen [29] established the oracle inequalities via Lasso and Dantzig selector for (approximately) sparse signal recovery under the framework of the mutual incoherence property.

Candès and Plan [7] proposed the matrix Lasso and the matrix Dantzig selector models and established the oracle inequalities for (approximately) low-rank matrix recovery. Consider the model $\mathbf{b} = \mathcal{A}(\mathbf{X}_0) + \mathbf{e}$. Here $\mathbf{X}_0 \in \mathbb{R}^{n_1 \times n_2}$ is the (approximately) low-rank matrix, $\mathcal{A}$ is a linear mapping from $\mathbb{R}^{n_1 \times n_2}$ to $\mathbb{R}^m$ with $\mathcal{A}(\mathbf{X}_0) = <\mathbf{B}_j, \mathbf{X}_0>$ and $\mathbf{B}_j \in \mathbb{R}^{n_1 \times n_2}$, and $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$. They showed that the estimation

$$\|\hat{\mathbf{X}} - \mathbf{X}_0\|_F^2 \lesssim \sum_i \min\left\{\sigma_i^2(\mathbf{X}_0), n\sigma^2\right\}$$

holds with high probability, where $n = \max\{n_1, n_2\}$ and $\sigma_i^2(\mathbf{X}_0)$ is the singular value of $\mathbf{X}_0$. Here $\hat{\mathbf{X}}$ is the solution of the matrix Dantzig selector as follows:

$$\begin{aligned} &\min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} \|\mathbf{X}\|_* \\ &\text{s.t.} \quad \left\|\mathcal{A}^*\left(\mathcal{A}(\mathbf{X})-\mathbf{b}\right)\right\| \le \eta, \end{aligned}$$

where $\|\cdot\|$ is the spectral norm ($\|\mathbf{X}\| = \max_j \sigma_j(\mathbf{X})$) and $\|\cdot\|_*$ is the nuclear norm ($\|\mathbf{X}\|_* = \sum_j \sigma_j(\mathbf{X})$), and $\mathcal{A}^*$ is the adjoint mapping of $\mathcal{A}$ with $\mathcal{A}^*(\mathbf{y}) = \sum_{j=1}^m y_j \mathbf{B}_j$.

1.3. Our contributions

In this subsection, we aim to extend the oracle inequalities to the corrupted compressed sensing. We consider the Gaussian noise model as follows:

$$\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e} =: \mathbf{\Phi}\mathbf{h}_0 + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m), \tag{1.5}$$

where

$$\mathbf{\Phi} = [\mathbf{A}, \mathbf{I}] \in \mathbb{R}^{m\times(n+m)}, \quad \mathbf{h}_0 = [\mathbf{x}_0; \mathbf{f}_0] \in \mathbb{R}^{n+m}.$$

Suppose that the standard deviation σ is known, $\|\mathbf{x}_0\|_0 \le s$ and $\|\mathbf{f}_0\|_0 \le t$. We denote index sets $S \subset \text{supp}(\mathbf{x}_0)$, $T \subset \text{supp}(\mathbf{f}_0)$ and $I = \{S, T\}$ with $|I| \le s + t < m$. For a fixed index set $I_* \subset \mathbb{R}^{(m+n)}$, we use the least squares (LS) method to estimate the following problem:

$$\hat{\mathbf{h}}_{I_*}^{LS} = \arg\min_{\mathbf{h}} \|\mathbf{b} - \mathbf{\Phi}_{I_*}\mathbf{h}\|_2^2, \tag{1.6}$$

where $\mathbf{\Phi}_{I_*} \in \mathbb{R}^{m\times(n+m)}$ with a restriction on the columns of it as: $(\mathbf{\Phi}_{I_*})_i = \mathbf{\Phi}_i$ for $i \in I_*$, $(\mathbf{\Phi}_{I_*})_i = \mathbf{0}$ otherwise. We can rewrite the LS estimator as follows:

$$\hat{\mathbf{h}}_{I_*}^{LS} = \left(\mathbf{\Phi}_{I_*}^\top \mathbf{\Phi}_{I_*}\right)^{-1} \mathbf{\Phi}_{I_*}^\top \mathbf{b}.$$

Once an oracle is aware of the true vector $\mathbf{h}_0$, it will choose the optimal index set I and reduce the MSE, which can be expressed as follows:

$$I_{\text{oracle}} = \arg\min \left\{ \mathbb{E}\left\|\hat{\mathbf{h}}_I^{LS} - \mathbf{h}_0\right\|_2^2 : I \subset \text{supp}(\mathbf{h}_0) \right\}.$$

The minimum MSE that may be obtained across all index sets is known as the oracle risk $\mathbb{E}\|\hat{\mathbf{h}}_I^{LS} - \mathbf{h}_0\|_2^2$. Notice that for each given index set I, the MSE of $\hat{\mathbf{h}}_I^{LS}$ can be determined as follows:

$$\mathbb{E}\left\|\hat{\mathbf{h}}_I^{LS} - \mathbf{h}_0\right\|_2^2 = \mathbb{E}\left\|\hat{\mathbf{h}}_I^{LS} - \mathbb{E}\hat{\mathbf{h}}_I^{LS}\right\|_2^2 + \left\|\mathbb{E}\hat{\mathbf{h}}_I^{LS} - \mathbf{h}_0\right\|_2^2.$$

It is well known that the LS estimator is an unbiased estimator, namely, $\mathbb{E}\hat{\mathbf{h}}_I^{LS} = (\mathbf{h}_0)_I$. Thus, we can get

$$\begin{aligned} \mathbb{E}\left\|\hat{\mathbf{h}}_I^{LS} - \mathbf{h}_0\right\|_2^2 &= \mathbb{E}\left\|\hat{\mathbf{h}}_I^{LS} - (\mathbf{h}_0)_I\right\|_2^2 + \|(\mathbf{h}_0)_{I^c}\|_2^2 \\ &= \left\|\left(\mathbf{\Phi}_I^\top \mathbf{\Phi}_I\right)^{-1} \mathbf{\Phi}_I^\top \mathbf{\Phi}(\mathbf{h}_0)_{I^c}\right\|_2^2 + \sigma^2 \text{Tr}\left(\left(\mathbf{\Phi}_I^\top \mathbf{\Phi}_I\right)^{-1}\right) + \|(\mathbf{h}_0)_{I^c}\|_2^2 \\ &\ge \sigma^2 \text{Tr}\left(\left(\mathbf{\Phi}_I^\top \mathbf{\Phi}_I\right)^{-1}\right) + \|(\mathbf{h}_0)_{I^c}\|_2^2. \end{aligned}$$

Notice that all eigenvalues of the linear operator $\mathbf{\Phi}_I^{\top}\mathbf{\Phi}_I$ belong to the interval $[1-\delta_{|I|}, 1+\delta_{|I|}]$ and $0<\delta_{|I|}<1$ (see [25, Theorem 1]). Therefore, we can get

$$\begin{aligned}
\mathbb{E}\big\|\hat{\mathbf{h}}_I^{LS}-\mathbf{h}_0\big\|_2^2 &\geq \frac{|I|\sigma^2}{1+\delta_{|I|}}+\|(\mathbf{h_0})_{I^c}\|_2^2\\
&\geq \frac{1}{2}\left(\|(\mathbf{x}_0)_{S^c}\|_2^2+\sigma^2|S|+\frac{1}{\lambda^2}\|\lambda(\mathbf{f}_0)_{T^c}\|_2^2+\sigma^2|T|\right)\\
&\geq \frac{1}{2}\sum_i \min\left\{|x_0(i)|^2,\sigma^2\right\}+\frac{t}{2c_2^2 s}\sum_j \min\left\{|\lambda f_0(j)|^2,\sigma^2\right\},
\end{aligned}$$

where the last inequality is established because of $\lambda\in[\sqrt{s/t}/c_1, c_2\sqrt{s/t}]$ with $c_1,c_2\geq 1$ (see [27]). In conclusion, the oracle bound of corrupted compressed sensing obeys

$$\begin{aligned}
&\inf_{S,T}\mathbb{E}\left[\big\|\hat{\mathbf{x}}_S^{LS}-\mathbf{x}_0\big\|_2^2+\big\|\hat{\mathbf{f}}_T^{LS}-\mathbf{f}_0\big\|_2^2\right]\\
&\gtrsim \sum_i \min\left\{|x_0(i)|^2,\sigma^2\right\}+\sum_j \min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}.
\end{aligned}\tag{1.7}$$

Now a fundamental question: Given the date $\mathbf{b}$ and the model (1.5), without knowing the support sets of $\mathbf{x}_0$ and $\mathbf{f}_0$, can we design an estimator which achieve (1.7)? In this paper, we try to solve this problem and give a positive answer. So, we mainly analyze the oracle inequalities of corrupted compressed sensing for the extended Dantzig selector model

$$\begin{aligned}
&\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{f}\in\mathbb{R}^m} \|\mathbf{x}\|_1+\lambda\|\mathbf{f}\|_1\\
&\text{s.t.}\quad \big\|[\mathbf{A},\mathbf{I}]^{\top}(\mathbf{A}\mathbf{x}+\mathbf{f}-\mathbf{b})\big\|_\infty\leq\eta_*,
\end{aligned}\tag{1.8}$$

and the extended Lasso model

$$\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{f}\in\mathbb{R}^m}\frac{1}{2\rho_*}\|\mathbf{b}-\mathbf{A}\mathbf{x}-\mathbf{f}\|_2^2+\|\mathbf{x}\|_1+\lambda\|\mathbf{f}\|_1.\tag{1.9}$$

Here we set

$$\eta_*=2\kappa\sigma\sqrt{\log(n+m)},\quad \rho_*=\frac{2\kappa\sigma\sqrt{\log(n+m)}}{\max\{1,\lambda\}},\quad \sqrt{2}<\kappa<2.$$

Firstly, we discuss the sparse signal recovery with $\|\mathbf{x}_0\|_0\leq s$ and $\|\mathbf{f}_0\|_0\leq t$. We consider the Gaussian noise model (1.5), and assume that the measurement matrix $\mathbf{A}=(a_{ij})_{i,j=1}^{m,n}$ is sampled from the Gaussian measurement ensemble, and $a_{ij}\sim\mathcal{N}(0,1/m)$. Suppose that $m\gtrsim(s+t)\log((n+m)/(s+t))$ and $\lambda\in[\sqrt{s/t}/c_1, c_2\sqrt{s/t}]$ with $c_1,c_2\geq 1$. Let $(\hat{\mathbf{x}},\hat{\mathbf{f}})$ be the optimal solution of the extended Dantzig selector (1.8) or the extended Lasso (1.9). Then we get the recovery error

$$\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2^2+\|\hat{\mathbf{f}}-\mathbf{f}_0\|_2^2\lesssim\sum_{i=1}^n \min\left\{|x_0(i)|^2,\sigma^2\right\}+\sum_{j=1}^m \min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}$$

with probability at least $1-1/\sqrt{\pi\log(n+m)}-1/n-3e^{-K_2 m}$. Please refer to Theorems 2.1 and 2.2 below.

Later, we discuss the approximately sparse signal recovery. Under the same hypothesis, we get an estimation

$$\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 + \|\hat{\mathbf{f}} - \mathbf{f}_0\|_2^2 \lesssim \left[\sum_{i \in S_*} \min \left\{ |x_0(i)|^2, \sigma^2 \right\} + \|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2 \right] + \left[\sum_{j \in T_*} \min \left\{ |\lambda f_0(j)|^2, \sigma^2 \right\} + \|(\mathbf{f}_0)_{-\max(t_*)}\|_2^2 \right]$$

with probability at least $1 - 1/\sqrt{\pi \log(n+m)} - 1/n - 3e^{-K_2 m} - e^{-m/100}$ provided that $m \gtrsim (s+t)\log((n+m)/(s+t))$. Here $\mathrm{supp}((\mathbf{x}_0)_{\max(s_*)}) \subset S_*$ with $|S_*| = s_*$, $\mathrm{supp}((\mathbf{f}_0)_{\max(t_*)}) \subset T_*$ with $|T_*| = t_*$ and $s_* + t_* = m/\log(e(n+m)/m)$, and $\lambda \in [\sqrt{s_*/t_*}/c_1, c_2\sqrt{s_*/t_*}]$ with $c_1, c_2 \geq 1$. Please refer to Theorems 3.1 and 3.2 below.

It must be pointed out that our results for corrupted compressed sensing are NOT trivial extensions of the compressed sensing (non-corrupted) with $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}$.

1.4. Organization and notations

The article is organized as follows. In Section 2, we establish the oracle inequalities of sparse signal recovery based on the extended Dantzig selector and the extended Lasso models. In Section 3, we discuss the oracle inequalities for approximately sparse signal recovery. In Section 4, we show numerical experiments to demonstrate the effect of the extended Dantzig selector and extended Lasso models and explain the significance of the oracle inequalities. In Section 5, we present the conclusions and comments.

We employ the following critical notation through the article. $\mathbf{x}_S$ denotes a vector that all elements equal to $\mathbf{x}$ if indices of it in set S, otherwise, equal to zero, and $\mathbf{x}_{S^c} = \mathbf{x} - \mathbf{x}_S$. For $\mathbf{x} \in \mathbb{R}^n$, we denote $\mathbf{x}_{\max(s)}$ as the vector $\mathbf{x}$ with all but the largest s entries in absolute value set to zero. Let $\mathbf{I} \in \mathbb{R}^{m \times m}$ be an $m \times m$ dimensional identity matrix. The matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times (m+n)}$ denotes the joint matrix of measurement matrix $\mathbf{A}$ and identity matrix $\mathbf{I}$, namely, $\mathbf{\Phi} = [\mathbf{A}, \mathbf{I}]$. To state conveniently, we use $A \lesssim B$ to denote $A \leq C_0 B$ for any $A, B \in \mathbb{R}$, where $C_0 \in \mathbb{R}^+$ is an absolute constant. The notion $\gtrsim$ can be defined similarly. We also use the notation $\mathcal{O}(n)$ to denote the number Cn with the universal constant C.

2. Oracle Inequalities for Sparse Signal

Based on the extended Dantzig selector and extended Lasso models, we establish the oracle inequalities for the Gaussian noise model (1.5) in this section.

Theorem 2.1. *Consider the Gaussian noise model* (1.5)*. Suppose that* $\mathbf{x}_0$ *is s-sparse,* $\mathbf{f}_0$ *is t-sparse and measurement matrix* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *is sampled from the Gaussian measurement ensemble, and* $a_{ij} \sim \mathcal{N}(0, 1/m)$*. Set*

$$\eta_* = 2\eta = 2\kappa\sigma\sqrt{\log(n+m)}, \quad \sqrt{2} < \kappa < 2, \quad m \geq K_1(s+t)\log\frac{n+m}{s+t}$$

with constant K_1 *depending on GRIP constant* δ*, and* $\lambda \in [\sqrt{s/t}/c_1, c_2\sqrt{s/t}]$ *with the constants* $c_1, c_2 \geq 1$*. Then the solution* $(\hat{\mathbf{x}}^{DS}, \hat{\mathbf{f}}^{DS})$ *of the model*

$$\min_{\mathbf{x} \in \mathbb{R}^n, \mathbf{f} \in \mathbb{R}^m} \|\mathbf{x}\|_1 + \lambda \|\mathbf{f}\|_1$$
$$s.t. \quad \left\|[\mathbf{A}, \mathbf{I}]^\top (\mathbf{A}\mathbf{x} + \mathbf{f} - \mathbf{b})\right\|_\infty \leq \eta_*$$

with $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}$ *satisfies*

$$\left\|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\right\|_2^2 + \left\|\hat{\mathbf{f}}^{DS} - \mathbf{f}_0\right\|_2^2 \leq C_1 \left[\sum_{i=1}^{n} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \sum_{j=1}^{m} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\}\right]$$

with probability at least $1 - 1/\sqrt{\pi \log(n+m)} - 1/n - 3e^{-K_2 m}$. *Here* $C_1 = \mathcal{O}(\log(m+n))$ *is a constant.*

Theorem 2.2. *Consider the Gaussian noise model* (1.5)*. Suppose that* $\mathbf{x}_0$ *is* s*-sparse,* $\mathbf{f}_0$ *is* t*-sparse and measurement matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *is sampled from the Gaussian measurement ensemble, and* $a_{ij} \sim \mathcal{N}(0, 1/m)$*. Set*

$$\rho_* = 2\kappa\sigma \frac{\sqrt{\log(n+m)}}{\max\{1, \lambda\}}, \quad \sqrt{2} < \kappa < 2, \quad m \geq K_1(s+t)\log\frac{n+m}{s+t}$$

with constant K_1 *depending on GRIP constant* δ*, and* $\lambda \in [\sqrt{s/t}/c_1, c_2\sqrt{s/t}]$ *with the constants* $c_1, c_2 \geq 1$*. Then the solution* $(\hat{\mathbf{x}}^L, \hat{\mathbf{f}}^L)$ *of the model*

$$\min_{\mathbf{x}\in\mathbb{R}^n, \mathbf{f}\in\mathbb{R}^m} \frac{1}{2\rho_*}\|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2 + \|\mathbf{x}\|_1 + \lambda\|\mathbf{f}\|_1$$

with $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}$ *satisfies*

$$\left\|\hat{\mathbf{x}}^{L} - \mathbf{x}_0\right\|_2^2 + \left\|\hat{\mathbf{f}}^{L} - \mathbf{f}_0\right\|_2^2 \leq C_1^{'} \left[\sum_{i=1}^{n} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \sum_{j=1}^{m} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\}\right]$$

with probability at least $1 - 1/\sqrt{\pi \log(n+m)} - 1/n - 3e^{-K_2 m}$. *Here* $C_1^{'}=\mathcal{O}(\log(m+n))$ *is a constant.*

Based on the oracle inequality introduced above, we also can design a solution $(\hat{\mathbf{x}}^{GDS}, \hat{\mathbf{f}}^{GDS})$ based on the bia-removing two-stage procedure in [11], and improve the performance of the solution $(\hat{\mathbf{x}}^{DS}, \hat{\mathbf{f}}^{DS})$ (see Example 4.3). We call it an extended Gaussian Dantzig selector (GDS). The two-stage procedure is as follows:

(i) Solve the extended DS model (1.8) and obtain a solution $(\hat{\mathbf{x}}^{DS}, \hat{\mathbf{f}}^{DS})$. Then we can obtain the support sets $\hat{S} = \{i \in \{1,\ldots,n\} : |\hat{x}_i| > \alpha_1\sigma\}$ and $\hat{T} = \{j \in \{1,\ldots,m\} : |\lambda\hat{f}_j| > \alpha_2\sigma\}$ for some parameters $\alpha_1 > 0$ and $\alpha_2 > 0$.

(ii) Generate the LS solution $(\hat{\mathbf{x}}^{DS}, \hat{\mathbf{f}}^{DS})$ via the estimator (1.6).

In the following part, we explain the significance of λ in our results.

Remark 2.1. In the compressed sensing problem $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}$, Candès and Tao [11] argued that the advantage of the oracle inequality upper bound $\sum_i \min\{|x_0(i)|, \sigma^2\}$ mainly reflects in the part where the signal elements are below the standard deviation of the random error $\mathbf{e}$, namely, $|x_0(i)| < \sigma$. Therefore, we will only consider the situation where the standard deviation σ is larger than all the elements of the signal $\mathbf{x}_0$ and the error $\mathbf{f}_0$, and $\|\mathbf{f}_0\|_2^2 = \mathcal{O}(10^k)\|\mathbf{x_0}\|_2^2$ with some $k \geq 1$. If $\lambda = 1$, the error bound can be expressed as

$$\begin{aligned}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 + \|\hat{\mathbf{f}} - \mathbf{f}_0\|_2^2 &\lesssim \sum_{i=1}^{n} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \sum_{j=1}^{m} \min\left\{|f_0(j)|^2, \sigma^2\right\} \\ &= \|\mathbf{x}_0\|_2^2 + \|\mathbf{f}_0\|_2^2 = \mathcal{O}(10^k)\|\mathbf{x}_0\|_2^2 = \mathcal{O}(1)\|\mathbf{f}_0\|_2^2.\end{aligned}$$

It is obvious that this upper bound is meaningless, and we can not obtain the stable solutions of $\hat{\mathbf{x}}$ and $\hat{\mathbf{f}}$. However, for a small $\lambda = 10^{-k/2} < 1$, we obtain the error bound

$$\begin{aligned}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 + \|\hat{\mathbf{f}} - \mathbf{f}\|_2^2 &\lesssim \sum_i |x_0(i)|^2 + \sum_j |\lambda f_0(j)|^2 \\ &= \|\mathbf{x}_0\|_2^2 + \lambda^2 \|\mathbf{f}_0\|_2^2 = \mathcal{O}(10^{-k})\|\mathbf{f}_0\|_2^2 = \mathcal{O}(1)\|\mathbf{x}_0\|_2^2,\end{aligned}$$

which gives a stable solution for $\mathbf{f}_0$. Although $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2/\|\mathbf{x}_0\|_2^2 = \mathcal{O}(1)$ does not imply a stable estimation of $\mathbf{x}_0$, it also greatly improves the error bound of $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2$ from $\mathcal{O}(10^k)$ to $\mathcal{O}(1)$.

Next we give an interpretation of the parameter $\rho_* = 2\kappa\sigma\sqrt{2\log(n+m)}/\max\{1,\lambda\}$ with constant $\sqrt{2} < \kappa < 2$ in Theorem 2.2 above.

Remark 2.2. What we should point out is that the parameters λ and ρ_* in the extended Lasso model satisfy $\|\mathbf{A}^\top \mathbf{e}\|_\infty \le \rho_*/2$ and $\|\mathbf{e}\|_\infty \le (\rho_*\lambda)/2$ (see [27, Theorem 2]). However, the parameter $\eta_* = 2\eta$ in the extended Dantzig selector model satisfies $\|[\mathbf{A},\mathbf{I}]^\top \mathbf{e}\|_\infty \le \eta$ with high probability and $\eta = \kappa\sigma\sqrt{2\log(n+m)}$ with the constant $\kappa \in (\sqrt{2}, 2)$ (see Proposition 2.1 below). Therefore, it follows from $\|[\mathbf{A},\mathbf{I}]^\top \mathbf{e}\|_\infty = \max\{\|\mathbf{A}^\top \mathbf{e}\|_\infty, \|\mathbf{e}\|_\infty\} \le \rho_* \max\{1,\lambda\}/2$ that $\rho_* = 2\eta/\max\{1,\lambda\} = \eta_*/\max\{1,\lambda\}$.

Remark 2.3. (i) Our results are not trivial extensions of the [3, Theorem 4.1] and [28, Theorem 3.2]. It is clear to see that if we take $\|\mathbf{z}\|_1$ with $\mathbf{z} = [\mathbf{x}; \lambda\mathbf{f}]$, we can get $\mathbf{b} = \mathbf{C}\mathbf{z} + \mathbf{e}$ with $\mathbf{C} = [\mathbf{A}, \mathbf{I}/\lambda]$. For sparse signal recovery, if the parameter $\lambda = 1$, the matrix $\mathbf{C} = [\mathbf{A}, \mathbf{I}/\lambda]$ satisfies the RIP (see Lemma 2.1 below), and meets the condition of [3, Theorem 4.1] and [28, Theorem 3.2]. And it leads to our conclusions of Theorems 2.1 and 2.2. However, when the parameter $\lambda \neq 1$, it is not sure whether the matrix $\mathbf{C} = [\mathbf{A}, \mathbf{I}/\lambda]$ satisfies the RIP or not. Thus our results above can not be reduced by that of [3, Theorem 4.1] and [28, Theorem 3.2].

(ii) If we take the setup $s \in \lfloor \alpha m/(1+\log(n/m)) \rfloor$ and $t \in \lfloor \alpha m \rfloor$ as that in [30], then the parameter λ can be taken as $\lambda = 1/\sqrt{1+\log(n/m)}$, which depends on the length of the signal and the number of the measurements.

Lastly, we give a remark regarding robust PCA.

Remark 2.4. We notice that Tanner and Vary [40] proposed a low-rank plus sparse model

$$\begin{aligned}&\min_{\mathbf{L},\mathbf{S}\in\mathbb{R}^{n_1\times n_2}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \\ &\text{s.t.} \quad \|\mathcal{A}(\mathbf{L}+\mathbf{S}) - \mathbf{b}\|_2 \le \epsilon\end{aligned}$$

for robust PCA, where $\mathbf{b} = \mathcal{A}(\mathbf{S}_0 + \mathbf{L}_0) + \mathbf{e}$. They developed an novel RIP for low-rank matrix plus sparse matrix $\mathbf{L} + \mathbf{S}$ and showed the recovery error as follows:

$$\left\|(\hat{\mathbf{L}} + \hat{\mathbf{S}}) - (\mathbf{L}_0 + \mathbf{S}_0)\right\|_F \le C\epsilon.$$

The oracle inequalities for corrupted compressed sensing in this paper maybe can be extended to this problem. We conjecture that the corresponding results has the form

$$\left\|\hat{\mathbf{L}} - \mathbf{L}_0\right\|_F^2 + \left\|\hat{\mathbf{S}} - \mathbf{S}_0\right\|_F^2 \lesssim \sum_{(i,j)} \min\left\{|S_0(i,j)|, \sigma^2\right\} + \sum_j \min\left\{\sigma_j^2(\mathbf{L}_0), n\sigma^2\right\}$$

with $n = \min\{n_1, n_2\}$ for the s-sparse matrix $\mathbf{S}_0$ and the r-rank matrix $\mathbf{L}_0$, and the Gaussian noise $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$.

Many scholars, for example the authors in [2, 7, 29], demonstrated that the Dantzig selector and the Lasso almost have equivalent features, resulting in similar error bounds for the sparse regression problem. In light of this, we only provide a thorough justification for the extended Dantzig selector model.

2.1. Auxiliary results for sparse signal recovery

The proof of Theorem 2.1 requires a few auxiliary results. Firstly, we recall the generalized restricted isometry property of the matrix $\mathbf{\Phi} = [\mathbf{A}, \mathbf{I}] \in \mathbb{R}^{m\times(n+m)}$.

Lemma 2.1 ([25]). *Suppose the matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *with elements* a_{ij} *drawn according to* $\mathcal{N}(0, 1/m)$ *and* $m \geq K_1(s+t)\log((n+m)/(s+t))$. *Then the matrix* $\mathbf{\Phi} = [\mathbf{A}, \mathbf{I}] \in \mathbb{R}^{m\times(n+m)}$ *satisfies* (s,t)*-GRIP and*

$$\left| \left\| \mathbf{\Phi} \begin{bmatrix} \mathbf{x} \\ \mathbf{f} \end{bmatrix} \right\|_2^2 - \left\| \begin{bmatrix} \mathbf{x} \\ \mathbf{f} \end{bmatrix} \right\|_2^2 \right| > 2\tau \left\| \begin{bmatrix} \mathbf{x} \\ \mathbf{f} \end{bmatrix} \right\|_2^2 \tag{2.1}$$

with probability at least $1 - 3e^{-K_2 m}$, *where* $\tau \in (0,1)$ *is given, and* K_1 *and* K_2 *are constants depending only on the relative GRIP constant* δ.

Next, we define an auxiliary notation K and show some properties about it. Define

$$K\big((\mathbf{x}, \mathbf{f}); (\mathbf{x_0}, \mathbf{f_0})\big) = \gamma(\|\mathbf{x}\|_0 + \|\mathbf{f}\|_0) + \left\| \mathbf{\Phi} \begin{bmatrix} \mathbf{x} \\ \mathbf{f} \end{bmatrix} - \mathbf{\Phi} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2, \tag{2.2}$$

where $\gamma = \eta^2/(1+\delta_{1,1})$ with $\eta = \kappa\sigma\sqrt{\log(n+m)}$, $\sqrt{2} < \kappa < 2$. Based on this notation, we can define an intermediate estimator $(\bar{\mathbf{x}}, \bar{\mathbf{f}})$ as follows:

$$\begin{cases} \bar{\mathbf{x}} = \arg\min_{\mathbf{x}} K\big((\mathbf{x}, \mathbf{f}); (\mathbf{x}_0, \mathbf{f}_0)\big), \\ \bar{\mathbf{f}} = \arg\min_{\mathbf{f}} K\big((\mathbf{x}, \mathbf{f}); (\mathbf{x}_0, \mathbf{f}_0)\big). \end{cases} \tag{2.3}$$

Then we have the following two properties.

Lemma 2.2. *(i) The intermediate estimator* $(\bar{\mathbf{x}}, \bar{\mathbf{f}})$ *defined in the Eq.* (2.3) *above satisfies* $\|\bar{\mathbf{x}}\|_0 \leq s$ *and* $\|\bar{\mathbf{f}}\|_0 \leq t$.

(ii) Suppose the matrix $\mathbf{\Phi}$ *satisfies GRIP. Then the intermediate variable* $(\bar{\mathbf{x}}, \bar{\mathbf{f}})$ *satisfies*

$$\left\| \mathbf{\Phi}^\top \mathbf{\Phi} \left(\begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{f}} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right) \right\|_\infty \leq \eta,$$

where $\eta = \kappa\sigma\sqrt{\log(n+m)}$, $\sqrt{2} < \kappa < 2$.

Proof. Firstly, we show the item (i). By the definition of the $\bar{\mathbf{x}}$, we can get

$$\begin{aligned} K\big((\bar{\mathbf{x}}, \mathbf{f}_0); (\mathbf{x}_0, \mathbf{f}_0)\big) &= \gamma(\|\bar{\mathbf{x}}\|_0 + \|\mathbf{f}_0\|_0) + \left\| \mathbf{\Phi} \begin{bmatrix} \bar{\mathbf{x}} \\ \mathbf{f}_0 \end{bmatrix} - \mathbf{\Phi} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \\ &\leq K\big((\mathbf{x}_0, \mathbf{f}_0); (\mathbf{x}_0, \mathbf{f}_0)\big) \\ &= \gamma(\|\mathbf{x}_0\|_0 + \|\mathbf{f}_0\|_0) + \left\| \mathbf{\Phi} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} - \mathbf{\Phi} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2, \end{aligned}$$

i.e.

$$\gamma\|\bar{\mathbf{x}}\|_0 + \left\|\mathbf{\Phi}\begin{bmatrix}\bar{\mathbf{x}}\\ \mathbf{f}_0\end{bmatrix} - \mathbf{\Phi}\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2 \leq \gamma\|\mathbf{x}_0\|_0.$$

So we can draw a conclusion that $\|\bar{\mathbf{x}}\|_0 \leq \|\mathbf{x}_0\|_0 \leq s$. Similarly, we can show $\|\bar{\mathbf{f}}\|_0 \leq \|\mathbf{f}_0\|_0 \leq t$.

Next, we give the proof of item (ii) by contradiction. Suppose there exists a vector $\mathbf{u} = \mathbf{e}_i$ or $\mathbf{u} = -\mathbf{e}_i$ ($\mathbf{e}_i$ is a standard orthogonal basis vector in $\mathbb{R}^{n+m}$), satisfying

$$\left\langle \mathbf{u}, \mathbf{\Phi}^\top\mathbf{\Phi}\left(\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\rangle > \eta.$$

We construct a perturbation

$$\begin{bmatrix}\check{\mathbf{x}}\\ \check{\mathbf{f}}\end{bmatrix} = \begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix} - \alpha\mathbf{u},$$

and let

$$\alpha = \frac{1}{\|\mathbf{\Phi}(\mathbf{u})\|_2^2}\left\langle \mathbf{u}, \mathbf{\Phi}^\top\mathbf{\Phi}\left(\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\rangle$$

Then, we have

$$\begin{aligned}
\left\|\mathbf{\Phi}\left(\begin{bmatrix}\check{\mathbf{x}}\\ \check{\mathbf{f}}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\|_2^2 &= \left\|\mathbf{\Phi}\left(\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix} - \alpha\mathbf{u} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\|_2^2 \\
&= \left\|\mathbf{\Phi}\left(\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\|_2^2 \\
&\quad - 2\alpha\left\langle \mathbf{\Phi}(\mathbf{u}), \mathbf{\Phi}\left(\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\rangle + \alpha^2\|\mathbf{\Phi}(\mathbf{u})\|_2^2 \\
&= \left\|\mathbf{\Phi}\left(\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\|_2^2 - \alpha^2\|\mathbf{\Phi}(\mathbf{u})\|_2^2.
\end{aligned}$$

Thus, it is evident that

$$\begin{aligned}
K\big((\check{\mathbf{x}}, \check{\mathbf{f}}); (\mathbf{x}_0, \mathbf{f}_0)\big) &= \gamma\big[\|\check{\mathbf{x}}\|_0 + \|\check{\mathbf{f}}\|_0\big] + \left\|\mathbf{\Phi}\left(\begin{bmatrix}\check{\mathbf{x}}\\ \check{\mathbf{f}}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\|_2^2 \\
&\leq \gamma\big[\|\bar{\mathbf{x}}\|_0 + \|\bar{\mathbf{f}}\|_0\big] + \gamma + \left\|\mathbf{\Phi}\left(\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\|_2^2 - \alpha^2\|\mathbf{\Phi}(\mathbf{u})\|_2^2 \\
&= K\big((\bar{\mathbf{x}}, \bar{\mathbf{f}}); (\mathbf{x}_0, \mathbf{f}_0)\big) + \gamma - \alpha^2\|\mathbf{\Phi}(\mathbf{u})\|_2^2. \qquad (2.4)
\end{aligned}$$

By the GRIP condition, we can conclude that

$$\|\mathbf{\Phi}(\mathbf{u})\|_2^2 \leq (1 + \delta_{1,1})\|\mathbf{u}\|_2^2 = 1 + \delta_{1,1}.$$

By the definition of the parameter α, and putting into $\gamma = \eta^2/(1 + \delta_{1,1})$, one has

$$\alpha^2\|\mathbf{\Phi}(\mathbf{u})\|_2^2 > \frac{\eta^2}{\|\mathbf{\Phi}(\mathbf{u})\|_2^4}\|\mathbf{\Phi}(\mathbf{u})\|_2^2 = \frac{\eta^2}{\|\mathbf{\Phi}(\mathbf{u})\|_2^2} \geq \frac{\eta^2}{(1 + \delta_{1,1})} = \gamma.$$

Substituting the estimation above into the inequality (2.4), we get

$$K\big((\check{\mathbf{x}}, \check{\mathbf{f}}); (\mathbf{x}_0, \mathbf{f}_0)\big) < K\big((\bar{\mathbf{x}}, \bar{\mathbf{f}}); (\mathbf{x}_0, \mathbf{f}_0)\big),$$

which provides the contradiction. □

Next, we recall an error estimation of the extended Dantzig selector model as follows.

Lemma 2.3 ([27, Corollary 1]). *Let* $\lambda \in [\sqrt{s/t}/c_1, c_2\sqrt{s/t}]$ *with* $c_1, c_2 \geq 1$. *Assume that the matrix* $\mathbf{\Phi}$ *satisfies the GRIP with* $\hat{\delta} = \delta_{2s,2t} + 2c_1c_2\delta_{2s,2t} < 1, \mathbf{x}_0$ *is s-sparse,* $\mathbf{f}_0$ *is t-sparse, and* $\|\mathbf{\Phi}^\top \mathbf{e}\|_\infty \leq \eta$. *Then the solution* $(\hat{\mathbf{x}}^{DS}, \hat{\mathbf{f}}^{DS})$ *of the model* (1.3) *with* $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}$ *satisfies*

$$\left\|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\right\|_2^2 + \left\|\hat{\mathbf{f}}^{DS} - \mathbf{f}_0\right\|_2^2 \leq \left(\frac{2\sqrt{2}\sqrt{s+t}(1+\sqrt{2}c_1c_2)}{1-\hat{\delta}}\right)^2 \eta^2 = \widetilde{C}\eta^2(s+t),$$

where $\widetilde{C} > 32$.

To end this subsection, we give an upper bound of $\|\mathbf{\Phi}^\top \mathbf{e}\|_\infty$ in probability.

Proposition 2.1. *Taking* $\eta = \kappa\sigma\sqrt{\log(n+m)}$ *with the constant* $\sqrt{2} < \kappa < 2$, *the event* $E = \{\|[\mathbf{A}, \mathbf{I}]^\top \mathbf{e}\|_\infty \leq \eta\}$ *occurs with probability at least* $1 - 1/\sqrt{\pi \log(n+m)} - 1/n$. *Especially, the event* $F = \{\|\mathbf{A}^\top \mathbf{e}\|_\infty \leq \kappa\sigma\sqrt{\log n}\}$ *occurs with probability as least* $1 - 1/\sqrt{\pi \log n} - 1/n$.

Proof. Our proof follows ideas from that [4, Lemma 5.1]. However, it has vital differences with that of [4, Lemma 5.1]. We do not assume that the columns of measurement matrix $\mathbf{A}$ are normalized to have unit norm $\|\mathbf{A}_i\|_2^2 = 1$ for $i = 1, \ldots, n$.

We define $(\mathbf{A}_i)_{i=1}^n$ as the columns of the matrix $\mathbf{A}$, $(\mathbf{I}_i)_{i=1}^m$ as the columns of the unit matrix $\mathbf{I}$, namely, $\mathbf{\Phi} = [\mathbf{A}, \mathbf{I}] = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_n, \mathbf{I}_1, \mathbf{I}_2, \cdots, \mathbf{I}_m] \in \mathbb{R}^{m\times(n+m)}$. Let the variable w_i as follows:

$$w_i = \langle \mathbf{\Phi}_i, \mathbf{e}\rangle = \begin{cases} \langle \mathbf{A}_i, \mathbf{e}\rangle, & i = 1, 2, \ldots, n, \\ \langle \mathbf{I}_{i-n}, \mathbf{e}\rangle, & i = n+1, \ldots, n+m. \end{cases}$$

Here, the random noise $\mathbf{e}$ follows a Gaussian distribution, denoted as $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$. It is clear that the variables $\{\langle \mathbf{A}_i, \mathbf{e}\rangle\}_{i=1}^n$ are independent Gaussian random variables with mean zero and variance $\{\|\mathbf{A}_i\|_2^2\sigma^2\}_{i=1}^n$, and the variables $\{\langle \mathbf{I}_j, \mathbf{e}\rangle\}_{j=1}^m$ are independent mean zero and variance σ^2 Gaussian random variables. Therefore, the distribution of the variables w_i can be expressed as

$$w_i \sim \begin{cases} \mathcal{N}\big(0, \|\mathbf{A}_i\|_2^2\sigma^2\big), & i = 1, 2, \ldots, n, \\ \mathcal{N}(0, \sigma^2), & i = n+1, n+2, \ldots, n+m. \end{cases}$$

By the definition of the infinity norm, we can get the below probability formula

$$\begin{aligned} \mathbb{P}\left(\|\mathbf{\Phi}^\top \mathbf{e}\|_\infty \leq \eta\right) &= \mathbb{P}\left(\cap_{i=1}^{n+m}\{|w_i| \leq \eta\}\right) \\ &= 1 - \mathbb{P}\left(\cup_{i=1}^{n+m}\{|w_i| > \eta\}\right) \geq 1 - \sum_{i=1}^{n+m} \mathbb{P}(|w_i| > \eta). \end{aligned} \tag{2.5}$$

Recall that the elements of the measurement matrix $\mathbf{A}$ obey the Gaussian distribution $\mathcal{N}(0, 1/m)$. Then, it is easy to calculate that $\mathbb{E}\|\mathbf{A}_i\|_2^2 = 1$ for $i = 1, \ldots, n$. According to the concentration inequality for norms of Gaussian random vectors [46, Lemma E.3], we can deduce that

$$\mathbb{P}\Big(\big|\|\mathbf{A}_i\|_2^2 - \mathbb{E}\|\mathbf{A}_i\|_2^2\big| > \epsilon\Big) = \mathbb{P}\Big(\big|\|\mathbf{A}_i\|_2^2 - 1\big| > \epsilon\Big) \leq 2\exp\left(-\frac{m\epsilon^2}{8}\right),$$

where $\epsilon \in (0, 1)$. Thus one has

$$\mathbb{P}\big(\|\mathbf{A}_i\|_2^2 > 1 + \epsilon\big) \leq \exp\left(-\frac{m\epsilon^2}{8}\right).$$

Once the event $\{\|\mathbf{A}_i\|_2^2 \leq 1 + \epsilon\}$ occurs, we can obtain the probability inequality as follows:

$$\mathbb{P}\left(|w_i| > \eta \,\middle|\, \|\mathbf{A}_i\|_2^2 \le 1+\epsilon\right) \le \frac{2}{\sqrt{2\pi}\left(\eta/(\sigma\sqrt{1+\epsilon})\right)} \exp\left(-\frac{\left(\eta/(\sigma\sqrt{1+\epsilon})\right)^2}{2}\right), \quad i = 1,\dots,n,$$

where the inequality comes from the standard tail bound for Gaussian random variables. According to the conditional probability formula $\mathbb{P}(AB) = \mathbb{P}(A|B)\mathbb{P}(B)$, we have

$$\begin{aligned}\mathbb{P}(|w_i| \le \eta) &\ge \mathbb{P}\left(\{|w_i| \le \eta\} \cap \left\{\|\mathbf{A}_i\|_2^2 \le 1+\epsilon\right\}\right)\\ &= \mathbb{P}\left(|w_i| \le \eta \,\middle|\, \|\mathbf{A}_i\|_2^2 \le 1+\epsilon\right)\mathbb{P}\left(\|\mathbf{A}_i\|_2^2 \le 1+\epsilon\right)\\ &\ge \left(1 - \frac{2}{\sqrt{2\pi}\left(\eta/(\sigma\sqrt{1+\epsilon})\right)}\exp\left(-\frac{\left(\eta/(\sigma\sqrt{1+\epsilon})\right)^2}{2}\right)\right)\left(1 - \exp\left(-\frac{m\epsilon^2}{8}\right)\right),\end{aligned}$$

where the first inequality comes from the probability formula $\mathbb{P}(A) \ge \mathbb{P}(AB)$ (there A and B all represent the probability event). Therefore,

$$\begin{aligned}\mathbb{P}(|w_i| > \eta) \le& \frac{2}{\sqrt{2\pi}\left(\eta/(\sigma\sqrt{1+\epsilon})\right)}\exp\left(-\frac{\left(\eta/(\sigma\sqrt{1+\epsilon})\right)^2}{2}\right)\\ &+ \exp\left(-\frac{m\epsilon^2}{8}\right), \quad i = 1,\dots,n.\end{aligned} \tag{2.6}$$

On the other hand, by a standard tail bound for Gaussian random variables, we obtain that

$$\mathbb{P}\left(|w_i| > \eta\right) = \mathbb{P}\left(\left|\frac{w_i}{\sigma}\right| > \frac{\eta}{\sigma}\right) \le \frac{2}{\sqrt{2\pi}(\eta/\sigma)}\exp\left(-\frac{(\eta/\sigma)^2}{2}\right), \quad i = n+1,\dots,n+m. \tag{2.7}$$

Thus, by combining the inequalities (2.6) with (2.7), we can derive that

$$\begin{aligned}\sum_{i=1}^{n+m}\mathbb{P}(|w_i| > \eta) &= \sum_{i=1}^{n}\mathbb{P}(|w_i| > \eta) + \sum_{i=n+1}^{n+m}\mathbb{P}(|w_i| > \eta)\\ &\le \frac{2n}{\sqrt{2\pi}\left(\eta/(\sigma\sqrt{1+\epsilon})\right)}\exp\left(-\frac{\left(\eta/(\sigma\sqrt{1+\epsilon})\right)^2}{2}\right)\\ &\quad + n\exp\left(-\frac{m\epsilon^2}{8}\right) + \frac{2m}{\sqrt{2\pi}(\eta/\sigma)}\exp\left(-\frac{(\eta/\sigma)^2}{2}\right).\end{aligned}$$

Bring into the values $\eta = \sigma\sqrt{2(1+\epsilon)\log(n+m)}$ and $\epsilon = 4\sqrt{(\log n)/m}$, we can come to the following estimation:

$$\sum_{i=1}^{n+m}\mathbb{P}(|w_i| > \eta) \le \frac{1}{\sqrt{\pi\log(n+m)}} + \frac{1}{n}.$$

Substituting the above estimation into the inequality (2.5), we finish the demonstration. □

2.2. Proof of Theorem 2.1

Proof. Firstly, we divide $\|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\|_2^2 + \|\hat{\mathbf{f}}^{DS} - \mathbf{f}_0\|_2^2$ into two parts as follows:

$$\left\|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\right\|_2^2 + \left\|\hat{\mathbf{f}}^{DS} - \mathbf{f}_0\right\|_2^2 = \left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2$$

$$\leq 2\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}\right\|_2^2+2\left\|\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2$$

$$=: 2F_1+2F_2. \tag{2.8}$$

Next, we deal with the item F_1. The combination of Lemma 2.2 and Proposition 2.1 leads to

$$\begin{aligned}\left\|[\mathbf{A},\mathbf{I}]^\top(\mathbf{A}\bar{\mathbf{x}}+\bar{\mathbf{f}}-\mathbf{b})\right\|_\infty &= \left\|\mathbf{\Phi}^\top\left(\mathbf{\Phi}\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}-\mathbf{b}\right)\right\|_\infty\\ &= \left\|\mathbf{\Phi}^\top\left[\left(\mathbf{\Phi}\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}-\mathbf{\Phi}\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)+\left(\mathbf{\Phi}\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}-\mathbf{b}\right)\right]\right\|_\infty\\ &\leq \left\|\mathbf{\Phi}^\top\mathbf{\Phi}\left(\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right)\right\|_\infty+\left\|\mathbf{\Phi}^\top\mathbf{e}\right\|_\infty\\ &\leq \eta+\eta=\eta_*.\end{aligned}$$

Therefore, if the matrix $\mathbf{\Phi}$ satisfies the GRIP, then $(\bar{\mathbf{x}},\bar{\mathbf{f}})$ is a feasible solution of the extended Dantzig selector (1.8) with the probability at least $1-1/\sqrt{\pi\log(n+m)}-1/n$. Combining this with Lemma 2.3, one has

$$F_1=\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}\right\|_2^2=\left\|\hat{\mathbf{x}}^{DS}-\bar{\mathbf{x}}\right\|_2^2+\left\|\hat{\mathbf{f}}^{DS}-\bar{\mathbf{f}}\right\|_2^2\leq \widetilde{C}\eta_*^2\left(\|\bar{\mathbf{x}}\|_0+\|\bar{\mathbf{f}}\|_0\right), \tag{2.9}$$

which gives an estimation of the item F_1. To estimate the item F_2, we need to get the distance between $[\mathbf{x}_0;\mathbf{f}_0]$ and $[\bar{\mathbf{x}};\bar{\mathbf{f}}]$ as follows:

$$F_2=\left\|\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}-\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}\right\|_2^2\leq\frac{1}{1-\delta_{2s,2t}}\left\|\mathbf{\Phi}\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}-\mathbf{\Phi}\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}\right\|_2^2. \tag{2.10}$$

Here we use

$$\|\mathbf{x}_0-\bar{\mathbf{x}}\|_0\leq\|\mathbf{x}_0\|_0+\|\bar{\mathbf{x}}\|_0\leq 2s,\quad \|\mathbf{f}_0-\bar{\mathbf{f}}\|_0\leq\|\mathbf{f}_0\|_0+\|\bar{\mathbf{f}}\|_0\leq 2t,$$

which come from Lemma 2.2.

Plugging the estimations (2.9) and (2.10) into the inequality (2.8) gives

$$\begin{aligned}\left\|\hat{\mathbf{x}}^{DS}-\mathbf{x}_0\right\|_2^2+\left\|\hat{\mathbf{f}}^{DS}-\mathbf{f}_0\right\|_2^2&\leq 2\widetilde{C}\eta_*^2\big(\|\bar{\mathbf{x}}\|_0+\|\bar{\mathbf{f}}\|_0\big)+\frac{2}{1-\delta_{2s,2t}}\left\|\mathbf{\Phi}\begin{bmatrix}\bar{\mathbf{x}}\\ \bar{\mathbf{f}}\end{bmatrix}-\mathbf{\Phi}\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2\\ &\leq C_2K\big((\bar{\mathbf{x}},\bar{\mathbf{f}});(\mathbf{x}_0,\mathbf{f}_0)\big),\end{aligned} \tag{2.11}$$

where $C_2=\max\{8\widetilde{C}(1+\delta_{1,1}),2/(1-\delta_{2s,2t})\}$.

Next, we estimate $K((\bar{\mathbf{x}},\bar{\mathbf{f}});(\mathbf{x}_0,\mathbf{f}_0))$. By the definition of $(\bar{\mathbf{x}},\bar{\mathbf{f}})$, we know

$$K\big((\bar{\mathbf{x}},\bar{\mathbf{f}});(\mathbf{x}_0,\mathbf{f}_0)\big)\leq K\big(((\mathbf{x}_0)_{S_1},(\mathbf{f}_0)_{S_2});(\mathbf{x}_0,\mathbf{f}_0)\big).$$

Therefore, to give an upper bound of $\|\hat{\mathbf{x}}^{DS}-\mathbf{x}_0\|_2^2+\|\hat{\mathbf{f}}^{DS}-\mathbf{f}_0\|_2^2$, we only need to estimate an upper bound of $K(((\mathbf{x}_0)_{S_1},(\mathbf{f}_0)_{S_2});(\mathbf{x}_0,\mathbf{f}_0))$. Define two index subsets as follows:

$$\begin{aligned}S_1&=\big\{i\in\{1,2,\ldots,n\}:|x_0(i)|>\sigma\big\},\\ S_2&=\big\{j\in\{1,2,\ldots,m\}:|\lambda f_0(j)|>\sigma\big\}.\end{aligned}$$

By a simple calculation, one has

$$
\begin{aligned}
&K\big(((\mathbf{x}_0)_{s_1},(\mathbf{f}_0)_{s_2});(\mathbf{x}_0,\mathbf{f}_0)\big)\\
&=\gamma\big[\|(\mathbf{x}_0)_{s_1}\|_0+\|(\mathbf{f}_0)_{s_2}\|_0\big]+\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{s_1}\\(\mathbf{f}_0)_{s_2}\end{bmatrix}-\mathbf{\Phi}\begin{bmatrix}\mathbf{x}_0\\\mathbf{f}_0\end{bmatrix}\right\|_2^2\\
&\le\frac{\eta^2}{(1+\delta_{1,1})}\big[\|(\mathbf{x}_0)_{s_1}\|_0+\|(\mathbf{f}_0)_{s_2}\|_0\big]+(1+\delta_{s,t})\left\|\begin{bmatrix}(\mathbf{x}_0)_{s_1}\\(\mathbf{f}_0)_{s_2}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\\mathbf{f}_0\end{bmatrix}\right\|_2^2\\
&\le\frac{4\log(n+m)}{(1+\delta_{1,1})}\big[\sigma^2(\|(\mathbf{x}_0)_{s_1}\|_0+\|(\mathbf{f}_0)_{s_2}\|_0)\big]+(1+\delta_{s,t})\left\|\begin{bmatrix}(\mathbf{x}_0)_{s_1}\\(\mathbf{f}_0)_{s_2}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\\mathbf{f}_0\end{bmatrix}\right\|_2^2\\
&\le\max\left\{\frac{4\log(n+m)}{(1+\delta_{1,1})},(1+\delta_{s,t})\right\}\left[\sum_{i=1}^n\sigma^2\cdot 1_{\{|x_0(i)|>\sigma\}}+\sum_{i=1}^n|\mathbf{x}_0(i)|^2\cdot 1_{\{|x_0(i)|\le\sigma\}}\right]\\
&\quad+\max\left\{\frac{4\log(n+m)}{(1+\delta_{1,1})},\frac{(1+\delta_{s,t})}{\lambda^2}\right\}\left[\sum_{j=1}^m\sigma^2\cdot 1_{\{|\lambda f_0(j)|>\sigma\}}+\sum_{j=1}^m|\lambda f_0(j)|^2\cdot 1_{\{|\lambda f_0(j)|\le\sigma\}}\right]\\
&\le\frac{4\log(n+m)}{(1+\delta_{1,1})}\left[\sum_{i=1}^n\min\left\{|x_0(i)|^2,\sigma^2\right\}+\sum_{j=1}^m\min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}\right].
\end{aligned}
$$

Combining the above estimation with the estimation (2.11), we come to the conclusion that

$$
\begin{aligned}
&\left\|\hat{\mathbf{x}}^{DS}-\mathbf{x}_0\right\|_2^2+\left\|\hat{\mathbf{f}}^{DS}-\mathbf{f}_0\right\|_2^2\\
&\le C_2K\big((\bar{\mathbf{x}},\bar{\mathbf{f}});(\mathbf{x},\mathbf{f}_0)\big)\le C_2K\big(((\mathbf{x}_0)_{s_1},(\mathbf{f}_0)_{s_2});(\mathbf{x}_0,\mathbf{f}_0)\big)\\
&\le C_2\frac{4\log(n+m)}{(1+\delta_{1,1})}\left[\sum_{i=1}^n\min\left\{|x_0(i)|^2,\sigma^2\right\}+\sum_{j=1}^m\min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}\right]\\
&=C_1\left[\sum_{i=1}^n\min\left\{|x_0(i)|^2,\sigma^2\right\}+\sum_{j=1}^m\min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}\right],
\end{aligned}
$$

where the constant $C_1=\mathcal{O}(\log(n+m))$. The second inequality is drawn from the definition of middle vectors $\bar{\mathbf{x}}$ and $\bar{\mathbf{f}}$.

From Lemma 2.1 and Proposition 2.1, we can get that the matrix $\mathbf{\Phi}$ satisfies the GRIP and the event E occur at the same time with probability at least

$$1-\frac{1}{\sqrt{\pi\log(n+m)}}-\frac{1}{n}-3e^{-K_2m}=:1-\varepsilon.$$

Therefore, we can get our conclusion with probability at least $1-\varepsilon$. □

3. Oracle Inequalities for Approximately Sparse Signal Recovery

In this section, we discuss the approximately sparse signal recovery and establish the corresponding oracle inequalities for Gaussian noise model (1.5).

Theorem 3.1. *Consider the Gaussian noise model* (1.5). *Suppose that* $\mathbf{A}$ *is sampled from the Gaussian measurement ensemble, and* $a_{ij}\sim\mathcal{N}(0,1/m)$. *Assume that* $\mathrm{supp}((\mathbf{x}_0)_{\max(s_*)})\subset S_*$ *with* $|S_*|=s_*$, $\mathrm{supp}((\mathbf{f}_0)_{\max(t_*)})\subset T_*$ *with* $|T_*|=t_*$, *and* $s_*+t_*=m/\log(e(n+m)/m)$.

Set $\eta_* = 2\eta = 2\kappa\sigma\sqrt{\log(n+m)}, \sqrt{2} < \kappa < 2, \lambda \in [\sqrt{s_*/t_*}/c_1, c_2\sqrt{s_*/t_*}]$ *with the constants* $c_1, c_2 \geq 1$. *Let* $(\hat{\mathbf{x}}^{DS}, \hat{\mathbf{f}}^{DS})$ *be the optimal solution of the model as follows:*

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n, \mathbf{f}\in\mathbb{R}^m} \|\mathbf{x}\|_1 + \lambda\|\mathbf{f}\|_1 \\ &s.t. \quad \left\|[\mathbf{A}, \mathbf{I}]^\top (\mathbf{A}\mathbf{x} + \mathbf{f} - \mathbf{b})\right\|_\infty \leq \eta_* \end{aligned}$$

with $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}$. *Then with the probability at least* $1 - 1/\sqrt{\pi\log(n+m)} - 1/n - 3e^{-K_2 m} - e^{-m/100}$, *the solution* $(\hat{\mathbf{x}}^{DS}, \hat{\mathbf{f}}^{DS})$ *satisfies*

$$\begin{aligned} &\left\|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\right\|_2^2 + \left\|\hat{\mathbf{f}}^{DS} - \mathbf{f}_0\right\|_2^2 \\ &\leq L_1\left[\sum_{i\in S_*} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2\right] \\ &\quad + L_2\left[\sum_{j\in T_*} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\} + \|(\mathbf{f}_0)_{-\max(t_*)}\|_2^2\right], \end{aligned}$$

where the constants $L_1 = \mathcal{O}(\log(n+m))$ *and* $L_2 = \mathcal{O}(\log(n+m))$.

Theorem 3.2. *Consider the Gaussian noise model* (1.5). *Suppose that* $\mathbf{A}$ *is sampled from the Gaussian measurement ensemble, and* $a_{ij} \sim \mathcal{N}(0, 1/m)$. *Assume that* $\mathrm{supp}((\mathbf{x}_0)_{\max(s_*)}) \subset S_*$ *with* $|S_*| = s_*, \mathrm{supp}((\mathbf{f}_0)_{\max(t_*)}) \subset T_*$ *with* $|T_*| = t_*$, *and* $s_* + t_* = m/\log(e(n+m)/m)$. *Set* $\rho_* = 2\kappa\sigma\sqrt{\log(n+m)}/\max\{1, \lambda\}, \sqrt{2} < \kappa < 2, \lambda \in [\sqrt{s_*/t_*}/c_1, c_2\sqrt{s_*/t_*}]$ *with the constants* $c_1, c_2 \geq 1$. *Let* $(\hat{\mathbf{x}}^L, \hat{\mathbf{f}}^L)$ *be the optimal solution of the model as follows:*

$$\min_{\mathbf{x}\in\mathbb{R}^n, \mathbf{f}\in\mathbb{R}^m} \frac{1}{2\rho_*}\|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{f}\|_2^2 + \|\mathbf{x}\|_1 + \lambda\|\mathbf{f}\|_1$$

with $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}$. *Then with the probability at least* $1 - 1/\sqrt{\pi\log(n+m)} - 1/n - 3e^{-K_2 m} - e^{-m/100}$, *the solution* $(\hat{\mathbf{x}}^L, \hat{\mathbf{f}}^L)$ *satisfies*

$$\begin{aligned} &\left\|\hat{\mathbf{x}}^L - \mathbf{x}_0\right\|_2^2 + \left\|\hat{\mathbf{f}}^L - \mathbf{f}_0\right\|_2^2 \\ &\leq L_1^{'}\left[\sum_{i\in S_*} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2\right] \\ &\quad + L_2^{'}\left[\sum_{j\in T_*} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\} + \|(\mathbf{f}_0)_{-\max(t_*)}\|_2^2\right], \end{aligned}$$

where the constants $L_1^{'} = \mathcal{O}(\log(n+m))$ *and* $L_2^{'} = \mathcal{O}(\log(n+m))$.

Remark 3.1. It must be mentioned that our results above are non-trivial extensions of the oracle inequality

$$\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 \lesssim \sum_{i\in S_*} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2$$

in [28, Theorem 3.4]. Note that $[(\mathbf{x}_0)_{\max(s_*)}; (\mathbf{f}_0)_{\max(t_*)}] \neq [\mathbf{x}_0; \mathbf{f}_0]_{\max(s_*+t_*)}$. Therefore, we can't directly obtain the support sets $S_* = \mathrm{supp}((\mathbf{x}_0)_{\max(s_*)})$ and $T_* = \mathrm{supp}((\mathbf{f}_0)_{\max(t_*)})$ via $I_* = \mathrm{supp}((\mathbf{z}_0)_{\max(s_*+t_*)})$ with $\mathbf{z}_0 = [\mathbf{x}_0; \mathbf{f}_0]$. Moreover, our conclusions of Theorems 3.1 and 3.2 can not be reduced by that of compressed sensing (non-corrupted) with $\mathbf{b} = C\mathbf{z}_0 + \mathbf{e}$.

3.1. Auxiliary results for approximately sparse signal recovery

In this subsection, we provide several auxiliary results for approximately sparse signal recovery. Firstly, we define two index subsets as follows:

$$\begin{aligned} S_1 &= \big\{ i \in \{1, 2, \dots, n\} : |x_0(i)| > \sigma \big\}, \\ S_2 &= \big\{ j \in \{1, 2, \dots, m\} : |\lambda f_0(j)| > \sigma \big\}. \end{aligned} \tag{3.1}$$

If the values of $\mathbf{x}_0$ and $\mathbf{f}_0$ are much larger than σ and σ/λ, respectively, then the restricted vectors $(\mathbf{x}_0)_{S_1}$ and $(\mathbf{f}_0)_{S_2}$ are sparse. Otherwise, the sparsity of $(\mathbf{x}_0)_{S_1}$ (or $(\mathbf{f}_0)_{S_2}$) approximates the dimension of $\mathbf{x}_0$ (or $\mathbf{f}_0$). We discuss the two cases in following two subsections.

3.1.1. Sparse restricted vectors $(\mathbf{x}_0)_{S_1}$ and $(\mathbf{f}_0)_{S_2}$

Firstly, we establish a proposition that provides an error bound for this special case.

Proposition 3.1. *Suppose the matrix* $\mathbf{\Phi}$ *satisfies GRIP. Let the intermediate estimator* $(\bar{\mathbf{x}}, \bar{\mathbf{f}})$ *be as defined in* (2.3)*, the sets* S_1, S_2 *be as defined in* (3.1) *with*

$$\bar{s} = \max\big\{\|\bar{\mathbf{x}}\|_0, |S_1|\big\} < n, \quad \bar{t} = \max\big\{\|\bar{\mathbf{f}}\|_0, |S_2|\big\} < m.$$

Then the solution $(\hat{\mathbf{x}}^{DS}, \hat{\mathbf{f}}^{DS})$ *of the extended Dantzig selector* (1.8) *with* $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}$ *and* $\|\mathbf{\Phi}^\top \mathbf{e}\|_\infty \le \eta$ *satisfies*

$$\left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \le M_1 \left[\sum_{i=1}^{n} \min\big\{|x_0(i)|^2, \sigma^2\big\} + \sum_{j=1}^{m} \min\big\{|\lambda f_0(j)|^2, \sigma^2\big\} \right] + M_2 \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2,$$

where $M_1 = \mathcal{O}(\log(n+m))$ *and* M_2 *is a universal constant depending on* $\delta_{2\bar{s}, 2\bar{t}}$.

Proof. The proof is similar to that of Theorem 2.1, therefore we briefly review the primary stages. Firstly, we divide $\|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\|_2^2 + \|\hat{\mathbf{f}}^{DS} - \mathbf{f}_0\|_2^2$ into two parts as follows:

$$\left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \le 2 \left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{f}} \end{bmatrix} \right\|_2^2 + 2 \left\| \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{f}} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 =: 2G_1 + 2G_2.$$

For the part G_2, we have the following estimation:

$$\begin{aligned} G_2 &\le 2 \left\| \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{f}} \end{bmatrix} - \begin{bmatrix} (\mathbf{x}_0)_{S_1} \\ (\mathbf{f}_0)_{S_2} \end{bmatrix} \right\|_2^2 + 2 \left\| \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 \\ &\le \frac{2}{1 - \delta_{2\bar{s}, 2\bar{t}}} \left\| \mathbf{\Phi} \left(\begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{f}} \end{bmatrix} - \begin{bmatrix} (\mathbf{x}_0)_{S_1} \\ (\mathbf{f}_0)_{S_2} \end{bmatrix} \right) \right\|_2^2 + 2 \left\| \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 \\ &\le \frac{4}{1 - \delta_{2\bar{s}, 2\bar{t}}} \left\| \mathbf{\Phi} \left(\begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{f}} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right) \right\|_2^2 + \frac{4}{1 - \delta_{2\bar{s}, 2\bar{t}}} \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 + 2 \left\| \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2, \end{aligned} \tag{3.2}$$

where the second inequality comes from the fact

$$\begin{aligned} &\|\bar{\mathbf{x}} - (\mathbf{x}_0)_{S_1}\|_0 \le \|\bar{\mathbf{x}}\|_0 + \|(\mathbf{x}_0)_{S_1}\|_0 \le 2\bar{s}, \\ &\|\bar{\mathbf{f}} - (\mathbf{f}_0)_{S_2}\|_0 \le \|\bar{\mathbf{f}}\|_0 + \|(\mathbf{f}_0)_{S_2}\|_0 \le 2\bar{t} \end{aligned}$$

and the GRIP condition. We also can gain the bound of G_1 as follows:

$$G_1 = \left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{f}} \end{bmatrix} \right\|_2^2 \leq \widetilde{C}\eta_*^2\big(\|\bar{\mathbf{x}}\|_0 + \|\bar{\mathbf{f}}\|_0\big), \tag{3.3}$$

where the inequality comes from Lemma 2.3. Hence, the combination of inequalities (3.2) and (3.3) gives

$$\left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \leq 2\widetilde{C}\eta_*^2\big(\|\bar{\mathbf{x}}\|_0 + \|\bar{\mathbf{f}}\|_0\big) + \frac{8}{1-\delta_{2\bar{s},2\bar{t}}} \left\| \mathbf{\Phi} \left(\begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{f}} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right) \right\|_2^2$$
$$+ \frac{8}{1-\delta_{2\bar{s},2\bar{t}}} \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 + 4 \left\| \begin{bmatrix} (\mathbf{x}_0 t)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 .$$

By the definitions of $K((\mathbf{x},\mathbf{f});(\mathbf{x}_0,\mathbf{f}_0))$ in (2.2) and the intermediate estimator $(\bar{\mathbf{x}},\bar{\mathbf{f}})$ in (2.3), we get

$$\left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \leq \widetilde{M}K\big((\bar{\mathbf{x}},\bar{\mathbf{f}});(\mathbf{x}_0,\mathbf{f}_0)\big) + 4\left\| \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 + \frac{8}{1-\delta_{2\bar{s},2\bar{t}}} \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2$$
$$\leq \widetilde{M}K\big(((\mathbf{x}_0)_{S_1},(\mathbf{f}_0)_{S_2});(\mathbf{x}_0,\mathbf{f}_0)\big) + 4\left\| \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 + \frac{8}{1-\delta_{2\bar{s},2\bar{t}}} \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2$$
$$\leq M_1 \left[\sum_{i=1}^{n} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \sum_{j=1}^{m} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\} \right] + M_2 \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2,$$

where the last inequality holds because of $\eta = \kappa\sigma\sqrt{\log(n+m)}, \kappa \in (\sqrt{2}, 2)$. These constants are defined as

$$\widetilde{M} = \max\left\{ \frac{8}{1-\delta_{2\bar{s},2\bar{t}}}, 8\widetilde{C}(1+\delta_{1,1}) \right\}, \quad M_1 = 4\widetilde{M}\frac{\log(n+m)}{1+\delta_{1,1}}, \quad M_2 = \frac{8}{1-\delta_{2\bar{s},2\bar{t}}} + \widetilde{M}.$$

Thus finishes the proof. □

3.1.2. General restricted vectors $(\mathbf{x}_0)_{S_1}$ and $(\mathbf{f}_0)_{S_2}$

The above result is established under the assumption that $(\mathbf{x_0})_{S_1}$ and $(\mathbf{f_0})_{S_2}$ are sparse. To obtain ideal error boundness when the sparsity of $(\mathbf{x_0})_{S_1}$ (or $(\mathbf{f_0})_{S_2}$) approaches the dimension of $\mathbf{x}_0$ (or $\mathbf{f}_0$), we recall the ℓ_1-quotient (LQ) property, and display some results about it.

Definition 3.1 ([22, Definition 11.11]). *We say that the measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is satisfied to own the $LQ(\beta)$, also known as ℓ_1-quotient property with constant $\beta > 0$, if there exists a vector $\widetilde{\mathbf{x}} \in \mathbb{R}^n$ satisfying*

$$\mathbf{A}\mathbf{x} = \mathbf{A}\widetilde{\mathbf{x}}, \quad \|\widetilde{\mathbf{x}}\|_1 \leq \frac{1}{\beta}\|\mathbf{A}\mathbf{x}\|_2 .$$

The following lemma shows that the matrix $\mathbf{\Phi} = [\mathbf{A},\mathbf{I}] \in \mathbb{R}^{m\times(n+m)}$ satisfies the $LQ(\beta)$ with high probability.

Lemma 3.1. *If the matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *is drawn from Gaussian measurement ensemble with* $n \geq 2m$*, and* $a_{ij} \sim \mathcal{N}(0, 1/m)$*, then the matrix* $\mathbf{\Phi} = [\mathbf{A}, \mathbf{I}] \in \mathbb{R}^{m\times(n+m)}$ *satisfies the LQ*(β) *with probability at least* $1 - e^{-m/100}$*, where* $\beta = 1/(38\sqrt{s_* + t_*})$ *and* $s_* + t_* = m/\log(e(n+m)/m)$*.*

Proof. According to [22, Lemma 11.17, Remark 11.18], we need to prove

$$\mathbb{P}\left(\|\mathbf{v}\|_2 \leq 38\sqrt{s_* + t_*}\,\|\mathbf{\Phi}^\top \mathbf{v}\|_\infty\right) \geq 1 - \exp\left(-\frac{100}{m}\right), \quad \forall\, \mathbf{v} \in \mathbb{R}^m,$$

where

$$\|\mathbf{\Phi}^\top \mathbf{v}\|_\infty = \max\left\{\|\mathbf{A}^\top \mathbf{v}\|_\infty, \|\mathbf{v}\|_\infty\right\}.$$

(1) If $\|\mathbf{v}\|_2 \leq 38\sqrt{s_* + t_*}\,\|\mathbf{v}\|_\infty$, we have

$$\mathbb{P}\left(\|\mathbf{v}\|_2 \leq 38\sqrt{s_* + t_*}\,\|\mathbf{\Phi}^\top \mathbf{v}\|_\infty\right) = 1.$$

(2) Else $\|\mathbf{v}\|_2 > 38\sqrt{s_* + t_*}\,\|\mathbf{v}\|_\infty$, one has

$$\mathbb{P}\left(\|\mathbf{v}\|_2 \leq 38\sqrt{s_* + t_*}\,\|\mathbf{\Phi}^\top \mathbf{v}\|_\infty\right) = \mathbb{P}\left(\|\mathbf{v}\|_2 \leq 38\sqrt{s_* + t_*}\,\|\mathbf{A}^\top \mathbf{v}\|_\infty\right).$$

Set $\tilde{s}_* := m/\log(en/m)$. We claim that

$$38\sqrt{s_* + t_*} = 38\sqrt{\frac{m}{\log\left(e(n+m)/m\right)}} \geq 34\sqrt{\frac{m}{\log(en/m)}} = 34\sqrt{\tilde{s}_*}. \tag{3.4}$$

Therefore, one get

$$\begin{aligned}\mathbb{P}\left(\|\mathbf{v}\|_2 \leq 38\sqrt{s_* + t_*}\|\mathbf{A}^\top \mathbf{v}\|_\infty\right) &\geq \mathbb{P}\left(\|\mathbf{v}\|_2 \leq 34\sqrt{\tilde{s}_*}\|\mathbf{A}^\top \mathbf{v}\|_\infty\right)\\ &\geq 1 - \exp\left(-\frac{100}{m}\right),\end{aligned}$$

where the last inequality comes from [22, Theorem 11.19].

Next, we show the inequality (3.4). In fact, we take a constant d such that

$$d\sqrt{\frac{m}{\log(e(n+m)/m)}} = 34\sqrt{\frac{m}{\log(en/m)}},$$

i.e.

$$d = 34\sqrt{\frac{1 + \log(1 + n/m)}{1 + \log(n/m)}}, \quad n \geq 2m.$$

It is evident that the function

$$g(x) = \frac{1 + \log(1 + x)}{1 + \log(x)}$$

is monotonically decreasing for all $x > 0$. Then

$$d = 34\sqrt{g\left(\frac{n}{m}\right)} \leq 34\sqrt{g(2)} = 34\sqrt{\frac{1 + \log 3}{1 + \log 2}} \leq 38,$$

which proves the inequality (3.4).

Combining the two cases above, one has

$$\mathbb{P}\left(\|\mathbf{v}\|_2 \leq 38\sqrt{s_*+t_*}\|\mathbf{\Phi}^\top \mathbf{v}\|_\infty\right) \geq 1-\exp\left(-\frac{100}{m}\right).$$

Thus we finish the proof of Lemma 3.1. □

Before giving the proof of Theorem 3.1, we also recall an error estimation of the extended Dantzig selector as follows.

Lemma 3.2 ([27, Theorem 1]). *Suppose that the matrix $\mathbf{\Phi}=[\mathbf{A},\mathbf{I}]$ satisfies $(2s,2t)$-GRIP with $\hat{\delta}=\delta_{2s,2t+2c_1c_2}\delta_{2s,2t}<1, \lambda\in[\sqrt{s/t}/c_1, c_2\sqrt{s/t}]$ with $c_1,c_2\geq 1$ and $\|\mathbf{\Phi}^\top\mathbf{e}\|_\infty\leq\eta$. Let $(\hat{\mathbf{x}}^{DS},\hat{\mathbf{f}}^{DS})$ be the optimal solution of (1.3). Then*

$$\begin{aligned}\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2 &\leq \frac{2\sqrt{2}\sqrt{s+t}(1+\sqrt{2}c_1c_2)}{1-\hat{\delta}}\eta+\left(\frac{2\sqrt{2}c_1c_2(1+\sqrt{2}c_1c_2)\delta_{2s,2t}}{1-\hat{\delta}}+2c_1c_2\right)\\ &\quad\times\left(\frac{\|(\mathbf{x}_0)_{-\max(s)}\|_1}{\sqrt{s}}+\frac{\|(\mathbf{f}_0)_{-\max(t)}\|_1}{\sqrt{t}}\right)\\ &=: Q_1'\sqrt{s+t}\eta+Q_2'\left(\frac{\|(\mathbf{x}_0)_{-\max(s)}\|_1}{\sqrt{s}}+\frac{\|(\mathbf{f}_0)_{-\max(t)}\|_1}{\sqrt{t}}\right)\\ &\leq Q_1'\sqrt{s+t}\eta+\frac{Q_3'}{\sqrt{s+t}}\left\|\begin{bmatrix}(\mathbf{x}_0)_{-\max(s)}\\ (\mathbf{f}_0)_{-\max(t)}\end{bmatrix}\right\|_1,\end{aligned}$$

where $Q_3'=Q_2'\max\{\sqrt{1+s/t},\sqrt{1+t/s}\}$.

With those in hand, we can estimate the recovery error, which corresponds to that of Proposition 3.1.

Proposition 3.2. *Suppose that the matrix $\mathbf{\Phi}$ satisfies $LQ(\beta)$ with $\beta=1/(38\sqrt{s_*+t_*})$, and $\|\mathbf{\Phi}^\top\mathbf{e}\|_\infty\leq\eta$. Let $(\hat{\mathbf{x}}^{DS},\hat{\mathbf{f}}^{DS})$ be the solution of the model (1.8). Then one has*

$$\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2\leq Q\left(\eta\sqrt{s_*+t_*}+\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2\right)+\left\|\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2,$$

where the constant Q depends on δ_{s_*,t_*}.

Proof. By the triangle inequality, one has

$$\begin{aligned}\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2 &\leq \left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\left(\begin{bmatrix}(\mathbf{x}_0)_{\max(s_*)}\\ (\mathbf{f}_0)_{\max(t_*)}\end{bmatrix}+\begin{bmatrix}\widetilde{\mathbf{x}}\\ \widetilde{\mathbf{f}}\end{bmatrix}\right)\right\|_2\\ &\quad+\left\|\begin{bmatrix}\widetilde{\mathbf{x}}\\ \widetilde{\mathbf{f}}\end{bmatrix}\right\|_2+\left\|\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2.\end{aligned}\tag{3.5}$$

Firstly, we give the bound of $\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\left(\begin{bmatrix}(\mathbf{x}_0)_{\max(s_*)}\\ (\mathbf{f}_0)_{\max(t_*)}\end{bmatrix}+\begin{bmatrix}\widetilde{\mathbf{x}}\\ \widetilde{\mathbf{f}}\end{bmatrix}\right)\right\|_2$. Setting $\beta=1/(38\sqrt{s_*+t_*})$, then the $LQ(\beta)$ property results in

$$\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}=\mathbf{\Phi}\begin{bmatrix}\widetilde{\mathbf{x}}\\ \widetilde{\mathbf{f}}\end{bmatrix}$$

with some $[\widetilde{\mathbf{x}};\widetilde{\mathbf{f}}]$ satisfying

$$\left\| \begin{bmatrix} \widetilde{\mathbf{x}} \\ \widetilde{\mathbf{f}} \end{bmatrix} \right\|_1 \leq \frac{1}{\beta} \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{-\max(s_*)} \\ (\mathbf{f}_0)_{-\max(t_*)} \end{bmatrix} \right\|_2 . \tag{3.6}$$

Additionally, we recall the identity

$$\mathbf{\Phi} \left(\begin{bmatrix} (\mathbf{x}_0)_{\max(s_*)} \\ (\mathbf{f}_0)_{\max(t_*)} \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{x}} \\ \widetilde{\mathbf{f}} \end{bmatrix} \right) = \mathbf{\Phi} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix},$$

which implies that

$$\left\| \mathbf{\Phi}^\top \left[\mathbf{\Phi} \left(\begin{bmatrix} (\mathbf{x}_0)_{\max(s_*)} \\ (\mathbf{f}_0)_{\max(t_*)} \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{x}} \\ \widetilde{\mathbf{f}} \end{bmatrix} \right) - \mathbf{b} \right] \right\|_\infty = \|\mathbf{\Phi}^\top \mathbf{e}\|_\infty \leq \eta.$$

It follows from Lemma 3.2 that

$$\left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \left(\begin{bmatrix} (\mathbf{x}_0)_{\max(s_*)} \\ (\mathbf{f}_0)_{\max(t_*)} \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{x}} \\ \widetilde{\mathbf{f}} \end{bmatrix} \right) \right\|_2 \leq Q_1 \sqrt{s_* + t_*}\eta + \frac{Q_3}{\sqrt{s_* + t_*}} \left\| \begin{bmatrix} \widetilde{\mathbf{x}} \\ \widetilde{\mathbf{f}} \end{bmatrix} \right\|_1 .$$

Plugging in the inequality (3.6), we obtain

$$\begin{aligned} &\left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \left(\begin{bmatrix} (\mathbf{x}_0)_{\max(s_*)} \\ (\mathbf{f}_0)_{\max(t_*)} \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{x}} \\ \widetilde{\mathbf{f}} \end{bmatrix} \right) \right\|_2 \\ &\leq Q_1 \sqrt{s_* + t_*}\eta + \frac{Q_3}{\beta\sqrt{s_* + t_*}} \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{-\max(s_*)} \\ (\mathbf{f}_0)_{-\max(t_*)} \end{bmatrix} \right\|_2 . \end{aligned} \tag{3.7}$$

Next, we give the bound of $\|[\tilde{\mathbf{x}};\tilde{\mathbf{f}}]\|_2^2$. Assume $\text{supp}(\tilde{\mathbf{x}}) \subseteq V$ and $\text{supp}(\tilde{\mathbf{f}}) \subseteq U$. We divide index sets V and U as $V = \sum_{j=1}^{l_1} V_j$ and $U = \sum_{j=1}^{l_2} U_j$. Here V_1 is the index set of the s_* largest entries in absolute value of $\tilde{\mathbf{x}}, V_2$ is the index set of the next s_* largest entries in absolute value of $\tilde{\mathbf{x}}$, and so on, where the cardinality of the last set V_{l_1} maybe smaller than s_*. Similarly, U_1 is the index set of the s_* largest entries in absolute value of $\widetilde{\mathbf{f}}, U_2$ is the index set of the next t_* largest entries in absolute value of $\widetilde{\mathbf{f}}$, and so on. We set $l = \max\{l_1, l_2\}$, then

$$\widetilde{\mathbf{x}} = \begin{cases} \widetilde{\mathbf{x}}_{V_j}, & \text{if} \quad j = 1, 2, \ldots, l_1, \\ 0, & \text{if} \quad j = l_1 + 1, \ldots, l, \end{cases}$$

$$\widetilde{\mathbf{f}} = \begin{cases} \widetilde{\mathbf{f}}_{U_j}, & \text{if} \quad j = 1, 2, \ldots, l_2, \\ 0, & \text{if} \quad j = l_2 + 1, \ldots, l. \end{cases}$$

Let $\widetilde{\mathbf{h}} := [\widetilde{\mathbf{x}};\widetilde{\mathbf{f}}]$ and $\widetilde{\mathbf{h}}_j = [\widetilde{\mathbf{x}}_{V_j};\widetilde{\mathbf{f}}_{U_j}]$. Then $\widetilde{\mathbf{h}} = \sum_{j=1}^{l} \widetilde{\mathbf{h}}_j$. For any $j \geq 2$, we have

$$\|\widetilde{\mathbf{h}}_j\|_2 \leq \frac{1}{\sqrt{s_* + t_*}} \|\widetilde{\mathbf{h}}_{j-1}\|_1,$$

and thus

$$\|\widetilde{\mathbf{h}}\|_2 \leq \|\widetilde{\mathbf{h}}_1\|_2 + \sum_{j\geq 2} \|\widetilde{\mathbf{h}}_j\|_2 \leq \|\widetilde{\mathbf{h}}_1\|_2 + \frac{1}{\sqrt{s_* + t_*}} \|\widetilde{\mathbf{h}}\|_1. \tag{3.8}$$

By the GRIP, we can give the bound of $\|\widetilde{\mathbf{h}}_1\|_2$ as follows:

$$\|\widetilde{\mathbf{h}}_1\|_2 \leq \frac{1}{\sqrt{1-\delta_{s_*,t_*}}}\|\mathbf{\Phi}(\widetilde{\mathbf{h}}_1)\|_2 \leq \frac{1}{\sqrt{1-\delta_{s_*,t_*}}}\left(\|\mathbf{\Phi}(\widetilde{\mathbf{h}})\|_2 + \sum_{i\geq 2}\|\mathbf{\Phi}(\widetilde{\mathbf{h}}_i)\|_2\right).$$

It follows from $\|\mathbf{\Phi}(\widetilde{\mathbf{h}}_i)\|_2 \leq \sqrt{1+\delta_{s_*,t_*}}\|\widetilde{\mathbf{h}}_i\|_2$ that

$$\sum_{i\geq 2}\|\mathbf{\Phi}(\widetilde{\mathbf{h}}_i)\|_2 \leq \sqrt{1+\delta_{s_*,t_*}}\sum_{i\geq 2}\|\widetilde{\mathbf{h}}_i\|_2 \leq \frac{\sqrt{1+\delta_{s_*,t_*}}}{\sqrt{s_*+t_*}}\|\widetilde{\mathbf{h}}\|_1.$$

Combining this with

$$\mathbf{\Phi}(\widetilde{\mathbf{h}}) = \mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ \mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}$$

gives

$$\|\widetilde{\mathbf{h}}_1\|_2 \leq \frac{1}{\sqrt{1-\delta_{s_*,t_*}}}\left(\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2 + \frac{\sqrt{1+\delta_{s_*,t_*}}}{\sqrt{s_*+t_*}}\|\widetilde{\mathbf{h}}\|_1\right). \tag{3.9}$$

Substituting the estimations (3.9) and (3.6) into the inequality (3.8), we have

$$\begin{aligned}\left\|\begin{bmatrix}\widetilde{\mathbf{x}}\\ \widetilde{\mathbf{f}}\end{bmatrix}\right\|_2 &\leq \frac{1}{\sqrt{1-\delta_{s_*,t_*}}}\left(\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2 + \frac{\sqrt{1+\delta_{s_*,t_*}}}{\sqrt{s_*+t_*}}\|\widetilde{\mathbf{h}}\|_1\right) + \frac{\|\widetilde{\mathbf{h}}\|_1}{\sqrt{s_*+t_*}}\\ &\leq \frac{1}{\sqrt{1-\delta_{s_*,t_*}}}\left(\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2 + \frac{\sqrt{1+\delta_{s_*,t_*}}}{\beta\sqrt{s_*+t_*}}\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2\right)\\ &\quad + \frac{1}{\beta\sqrt{s_*+t_*}}\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2 =: Q_4\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2.\end{aligned} \tag{3.10}$$

Lastly, substituting the inequalities (3.7) and (3.10) into the division (3.5), we get

$$\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix} - \begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2 \leq Q\left(\sqrt{s_*+t_*}\eta + \left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2\right) + \left\|\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2,$$

where the universal constant Q depends only on δ_{s_*,t_*}. □

3.2. The proof of Theorem 3.1

Proof. Our proof follows the outline of the proof in [7, Theorem 2.8]. However, there are vital differences. According to the sparsity of $\mathbf{x}_0$ and $\mathbf{f}_0$, we discuss the nine cases (see Fig. 3.1 below). Next, we give the proof in details.

By Lemma 2.1, the event $E = \{\|\mathbf{\Phi}^\top \mathbf{e}\|_\infty \leq \eta\}$ occurs with probability at least $1 - 1/n - 1/\sqrt{\pi\log(n+m)}$. By Lemmas 2.1 and 3.1, the matrix $\mathbf{\Phi} = [\mathbf{A}, \mathbf{I}]$ satisfies the (s,t)-GRIP with probability exceeding $1-3e^{-K_2 m}$ and the $LQ(\alpha)$ property with probability at least $1-e^{-m/100}$ with $\alpha = 1/(38\sqrt{s_*+t_*})$. We shall assume that the event E occurs, the matrix $\mathbf{\Phi}$ satisfies the GRIP and $LQ(\alpha)$ properties.

1. Case $(\mathbf{a_1}, \mathbf{b_1})$. Suppose

$$\begin{cases}K\big(((\mathbf{x}_0)_{S_1}, \mathbf{0}); (\mathbf{x}_0, \mathbf{0})\big) \leq \gamma s_*,\\ K\big((\mathbf{0}, (\mathbf{f}_0)_{S_2}); (\mathbf{0}, \mathbf{f}_0)\big) \leq \gamma t_*.\end{cases}$$

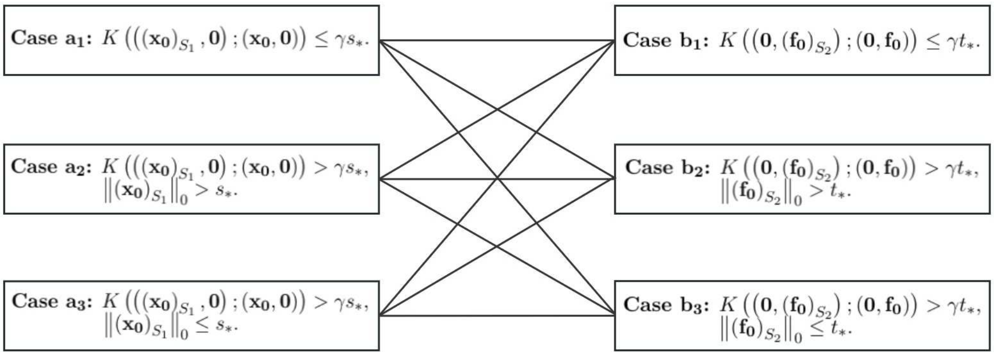

Fig. 3.1. The architecture diagram of nine case.

Then $\|(\mathbf{x}_0)_{S_1}\|_0 \le s_*, \|\bar{\mathbf{x}}\|_0 \le s_*, \|(\mathbf{f}_0)_{S_2}\|_0 \le t_*, \|\bar{\mathbf{f}}\|_0 \le t_*$. Therefore, by Proposition 3.1, we get

$$\left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \le M_1^{'} \left[\sum_{i=1}^{n} \min\left\{ |x_0(i)|^2, \sigma^2 \right\} + \sum_{j=1}^{m} \min\left\{ |\lambda f_0(j)|^2, \sigma^2 \right\} \right] + M_2^{'} \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2. \tag{3.11}$$

Notice that

$$\begin{aligned} \left\| \mathbf{\Phi} \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 &\le (1+2\tau) \left\| \begin{bmatrix} (\mathbf{x}_0)_{S_1^c} \\ (\mathbf{f}_0)_{S_2^c} \end{bmatrix} \right\|_2^2 \\ &\le (1+2\tau) \sum_i \min\left\{ |x_0(i)|^2, \sigma^2 \right\} + \frac{1+2\tau}{\lambda^2} \sum_j \min\left\{ |\lambda f_0(j)|^2, \sigma^2 \right\} \\ &\le (1+2\tau) \sum_i \min\left\{ |x_0(i)|^2, \sigma^2 \right\} + \frac{c_1^2 t(1+2\tau)}{s} \sum_j \min\left\{ |\lambda f_0(j)|^2, \sigma^2 \right\}, \end{aligned}$$

where the constant $c_1 \ge 1$ and the first inequality comes from Lemma 2.1. Substituting the inequality above into the estimation (3.11), we can get

$$\left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \le M_1^{''} \sum_{i=1}^{n} \min\left\{ |x_0(i)|^2, \sigma^2 \right\} + M_2^{''} \sum_{i=1}^{m} \min\left\{ |\lambda f_0(i)|^2, \sigma^2 \right\}, \tag{3.12}$$

where

$$M_1^{''} = \mathcal{O}\big(\log(n+m)\big), \quad M_2^{''} = \mathcal{O}\big(\log(n+m)\big).$$

In this situation, we need that the event E occurs and the matrix $\mathbf{\Phi}$ satisfies the GRIP. So the estimation (3.12) is established with probability at least $1 - 1/\sqrt{\pi \log(n+m)} - 1/n - 3e^{-K_2 m}$.

2. Case $(\mathbf{a_2}, \mathbf{b_2})$. Suppose

$$\begin{cases} K\big(\big((\mathbf{x}_0)_{S_1}, \mathbf{0}\big); (\mathbf{x}_0, \mathbf{0})\big) > \gamma s_*, \\ K\big(\big(\mathbf{0}, (\mathbf{f}_0)_{S_2}\big); (\mathbf{0}, \mathbf{f}_0)\big) > \gamma t_*, \end{cases}$$

and $\|(\mathbf{x}_0)_{S_1}\|_0 > s_*, \|(\mathbf{f}_0)_{S_2}\|_0 > t_*$, namely, $S_1 \supseteq S_*$ and $S_2 \supseteq T_*$. By combining Proposition 3.2 and Lemma 2.1, we can get

$$\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2 \le Q\left(\eta\sqrt{s_*+t_*}+\left\|\mathbf{\Phi}\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2\right)+\left\|\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2$$
$$\le Q_1'\left(\eta\sqrt{s_*+t_*}+\left\|\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2\right). \tag{3.13}$$

The estimation leads to

$$\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2 \le 2Q_2'\left(\eta^2(s_*+t_*)+\left\|\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2^2\right)$$
$$\le 8Q_2'\log(n+m)\Bigg[\sum_{i\in S_*}\min\left\{|x_0(i)|^2,\sigma^2\right\}+\|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2$$
$$+\sum_{i\in T_*}\min\left\{|\lambda f_0(i)|^2,\sigma^2\right\}+\|(\mathbf{f}_0)_{-\max(t_*)}\|_2^2\Bigg], \tag{3.14}$$

where the last inequality comes from the $\eta=\kappa\sigma\sqrt{\log(n+m)}$, $\sqrt{2}<\kappa<2$, and Q_2' is a universal constant depending on δ_{s_*,t_*}. In this situation, we need the assumptions that the event E occurs and the matrix $\mathbf{\Phi}$ satisfies the GRIP and $LQ(\beta)$ property. So the estimation (3.14) is established with probability at least $1-1/\sqrt{\pi\log(n+m)}-1/n-3e^{-K_2m}-e^{-m/100}$.

3. Case $(\mathbf{a_3},\mathbf{b_3})$. Suppose

$$\begin{cases}K\big(((\mathbf{x}_0)_{S_1},\mathbf{0});(\mathbf{x}_0,\mathbf{0})\big)>\gamma s_*,\\ K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)>\gamma t_*,\end{cases}$$

and $\|(\mathbf{x}_0)_{S_1}\|_0 \le s_*, \|(\mathbf{f}_0)_{S_2}\|_0 \le t_*$, namely, $S_1^c \supseteq S_*^c$ and $S_2^c \supseteq T_*^c$. By the inequalities above, we can get

$$(1+\delta_{1,1})K\big(((\mathbf{x}_0)_{S_1},\mathbf{0});(\mathbf{x}_0,\mathbf{0})\big)>\eta^2 s_*,$$
$$(1+\delta_{1,1})K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)>\eta^2 t_*.$$

Combining this with the estimation (3.13) above, we have

$$\left\|\begin{bmatrix}c\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2 \le 2Q_2'\left(\eta^2(s_*+t_*)+\left\|\begin{bmatrix}(\mathbf{x}_0)_{-\max(s_*)}\\ (\mathbf{f}_0)_{-\max(t_*)}\end{bmatrix}\right\|_2^2\right)$$
$$\le 2Q_2'\Big[(1+\delta_{1,1})K\big(((\mathbf{x}_0)_{S_1},\mathbf{0});(\mathbf{x}_0,\mathbf{0})\big)+\|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2$$
$$+(1+\delta_{1,1})K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)+\|(\mathbf{f}_0)_{-\max(t_*)}\|_2^2\Big]$$
$$=2Q_2'\Bigg\{\eta^2\|(\mathbf{x}_0)_{S_1}\|_0+(1+\delta_{1,1})\left\|\mathbf{\Phi}\left(\begin{bmatrix}(\mathbf{x}_0)_{S_1}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{0}\end{bmatrix}\right)\right\|_2^2$$
$$+\|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2+\eta^2\|(\mathbf{f}_0)_{S_2}\|_0$$
$$+(1+\delta_{1,1})\left\|\mathbf{\Phi}\left(\begin{bmatrix}\mathbf{0}\\ (\mathbf{f}_0)_{S_2}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \mathbf{f}_0\end{bmatrix}\right)\right\|_2^2+\|(\mathbf{f}_0)_{-\max(t^*)}\|_2^2\Bigg\}.$$

Applying Lemma 2.1 to the estimation above, we get

$$\begin{aligned}\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2 &\leq 2Q_2^{'}\Big\{\eta^2\|(\mathbf{x}_0)_{S_1}\|_0+(1+\delta_{1,1})(1+2\tau)\|(\mathbf{x}_0)_{S_1^c}\|_2^2+\|(\mathbf{x}_0)_{S_1^c}\|_2^2\\ &\qquad +\eta^2\|(\mathbf{f}_0)_{S_2}\|_0+(1+\delta_{1,1})\|(\mathbf{f}_0)_{S_2^c}\|_2^2+\|(\mathbf{f}_0)_{S_2^c}\|_2^2\Big\}\\ &\leq 2Q_2^{'}\max\left\{\big((1+\delta_{1,1})(1+2\tau)+1\big),4\log(n+m)\right\}\sum_{i=1}^{n}\min\left\{|x_0(i)|^2,\sigma^2\right\}\\ &\quad +2Q_2^{'}\max\left\{\frac{c_1^2t\big((1+\delta_{1,1})+1\big)}{s},4\log(n+m)\right\}\sum_{j=1}^{m}\min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}\\ &=8Q_2^{'}\log(n{+}m)\left[\sum_{i=1}^{n}\min\left\{|x_0(i)|^2,\sigma^2\right\}+\sum_{j=1}^{m}\min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}\right],\end{aligned}\tag{3.15}$$

where the second inequality comes from the $\eta=\kappa\sigma\sqrt{\log(n+m)}$, $\sqrt{2}<\kappa<2$, and the universal constant $Q_2^{'}$ depends on δ_{s_*,t_*}. In this situation, we need that the event E occurs, the matrix $\mathbf{\Phi}$ satisfies the GRIP and $LQ(\beta)$ property. So the above conclusion is established with probability at least $1-1/\sqrt{\pi\log(n+m)}-1/n-3e^{-K_2m}-e^{-m/100}$.

4. Case $(\mathbf{a_2},\mathbf{b_3})$. Suppose

$$\begin{cases}K\big(((\mathbf{x}_0)_{S_1},\mathbf{0});(\mathbf{x}_0,\mathbf{0})\big)>\gamma s_*,\\ K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)>\gamma t_*,\end{cases}$$

and $\|(\mathbf{x}_0)_{S_1}\|_0>s_*$, $\|(\mathbf{f}_0)_{S_2}\|_0\leq t_*$, namely, $S_1\supseteq S_*$ and $S_2^c\supseteq T_*^c$. By the inequalities above, we can get

$$(1+\delta_{1,1})K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)>\eta^2t_*.$$

Note that $\mathbf{x}_0$ is similar to Case $(\mathbf{a_2},\mathbf{b_2})$ and $\mathbf{f}_0$ is similar to Case $(\mathbf{a_3},\mathbf{b_3})$. It follows from the estimation (3.14) and (3.15) that

$$\begin{aligned}\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2 &=\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{0}\end{bmatrix}\right\|_2^2+\left\|\begin{bmatrix}\mathbf{0}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2\\ &=8Q_2^{'}\log(n+m)\Bigg[\sum_{i\in S_*}\min\left\{|x_0(i)|^2,\sigma^2\right\}+\|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2\\ &\qquad\qquad +\sum_{j=1}^{m}\min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}\Bigg]\end{aligned}$$

with probability at least $1-1/\sqrt{\pi\log(n+m)}-1/n-3e^{-K_2m}-e^{-m/100}$.

5. Case $(\mathbf{a_3},\mathbf{b_2})$. Suppose

$$\begin{cases}K\big(((\mathbf{x}_0)_{S_1},\mathbf{0});(\mathbf{x}_0,\mathbf{0})\big)>\gamma s_*,\\ K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)>\gamma t_*,\end{cases}$$

and $\|(\mathbf{x}_0)_{S_1}\|_0\leq s_*$, $\|(\mathbf{f}_0)_{S_2}\|_0>t_*$, namely, $S_1^c\supseteq S_*^c$ and $S_2\supseteq T_*$. By the assumption above, we can get

$$(1+\delta_{1,1})K\big(((\mathbf{x}_0)_{S_1},\mathbf{0});(\mathbf{x}_0,\mathbf{0})\big)>\eta^2s_*.$$

By dealing with $\mathbf{x}_0$ and $\mathbf{f}_0$ in a similar way as in Case $(\mathbf{a_3}, \mathbf{b_3})$ and Case $(\mathbf{a_2}, \mathbf{b_2})$, respectively, we have

$$\begin{aligned}\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2 &= \left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{0}\end{bmatrix}\right\|_2^2+\left\|\begin{bmatrix}\mathbf{0}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2\\ &= 8Q_2^{'}\log(n+m)\left[\sum_{i=1}^{n}\min\left\{|x_0(i)|^2,\sigma^2\right\}\right.\\ &\qquad\left.+\sum_{j\in T_*}\min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}+\|(\mathbf{f}_0)_{-\max(t_*)}\|_2^2\right]\end{aligned}$$

with probability at least $1-1/\sqrt{\pi\log(n+m)}-1/n-3e^{-K_2m}-e^{-m/100}$.

6. Case $(\mathbf{a_1}, \mathbf{b_2})$. Suppose

$$\begin{cases}K\big(((\mathbf{x}_0)_{S_1},\mathbf{0});(\mathbf{x}_0,\mathbf{0})\big)\le\gamma s_*,\\ K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)>\gamma t_*,\end{cases}$$

and $\|(\mathbf{f}_0)_{S_2}\|_0>t_*$, namely, $S_2\supseteq T_*$. By the assumption above, we get $\|(\mathbf{x}_0)_{S_1}\|_0\le s_*$ and $\|\bar{\mathbf{x}}\|_0\le s_*$. Here $\mathbf{x}_0$ is similar to Case $(\mathbf{a_1}, \mathbf{b_1})$, and $\mathbf{f}_0$ is similar to Case $(\mathbf{a_2}, \mathbf{b_2})$. Therefore we get the estimation as follows:

$$\begin{aligned}\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2 &= \left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{0}\end{bmatrix}\right\|_2^2+\left\|\begin{bmatrix}\mathbf{0}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2\\ &\le M_1^{''}\sum_{i=1}^{n}\min\left\{|x_0(i)|^2,\sigma^2\right\}\\ &\quad+8Q_2^{'}\log(n+m)\left[\sum_{i\in T_*}\min\left\{|\lambda f_0(i)|^2,\sigma^2\right\}+\|(\mathbf{f}_0)_{-\max(t_*)}\|_2^2\right]\end{aligned}$$

with probability at least $1-1/\sqrt{\pi\log(n+m)}-1/n-3e^{-K_2m}-e^{-m/100}$.

7. Case $(\mathbf{a_1}, \mathbf{b_3})$. We discuss the case

$$\begin{cases}K\big(((\mathbf{x}_0)_{S_1},\mathbf{0});(\mathbf{x}_0,\mathbf{0})\big)\le\gamma s_*,\\ K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)>\gamma t_*,\end{cases}$$

and $\|(\mathbf{f}_0)_{S_2}\|_0\le t_*$, namely, $S_2^c\supseteq T_*^c$. According to the inequalities above, we get

$$(1+\delta_{1,1})K\big((\mathbf{0},(\mathbf{f}_0)_{S_2});(\mathbf{0},\mathbf{f}_0)\big)>\eta^2t_*,$$

and $\|(\mathbf{x}_0)_{S_1}\|_0\le s_*$, $\|\bar{\mathbf{x}}\|_0\le s_*$. By combining the discussion of Case $(\mathbf{a_1}, \mathbf{b_1})$ and Case $(\mathbf{a_3}, \mathbf{b_3})$, we obtain

$$\begin{aligned}\left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2 &= \left\|\begin{bmatrix}\hat{\mathbf{x}}^{DS}\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{x}_0\\ \mathbf{0}\end{bmatrix}\right\|_2^2+\left\|\begin{bmatrix}\mathbf{0}\\ \hat{\mathbf{f}}^{DS}\end{bmatrix}-\begin{bmatrix}\mathbf{0}\\ \mathbf{f}_0\end{bmatrix}\right\|_2^2\\ &\le M_1^{''}\sum_{i=1}^{n}\min\left\{|x_0(i)|^2,\sigma^2\right\}+8Q_2^{'}\log(n+m)\sum_{j=1}^{m}\min\left\{|\lambda f_0(j)|^2,\sigma^2\right\}\end{aligned}$$

with probability at least $1-1/\sqrt{\pi\log(n+m)}-1/n-3e^{-K_2m}-e^{-m/100}$.

8. Case $(\mathbf{a_2}, \mathbf{b_1})$. Suppose

$$\begin{cases} K\big(((\mathbf{x}_0)_{S_1}, \mathbf{0}); (\mathbf{x}_0, \mathbf{0})\big) > \gamma s_*, \\ K\big((\mathbf{0}, (\mathbf{f}_0)_{S_2}); (\mathbf{0}, \mathbf{f}_0)\big) \leq \gamma t_*, \end{cases}$$

and $\|(\mathbf{x}_0)_{S_1}\|_0 > s_*$, namely, $S_1 \supseteq S_*$. It follows from the inequalities (3.12) and (3.14) above that

$$\begin{aligned} \left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 &= \left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{0} \end{bmatrix} \right\|_2^2 + \left\| \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \\ &\leq 8Q_2^{'} \log(n+m) \left[\sum_{i \in S_*} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2 \right] \\ &\quad + M_2^{''} \sum_{j=1}^{m} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\} \end{aligned}$$

with probability at least $1 - 1/\sqrt{\pi \log(n+m)} - 1/n - 3e^{-K_2 m} - e^{-m/100}$. Here the estimations $\|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\|_2^2$ and $\|\hat{\mathbf{f}}^{DS} - \mathbf{f}_0\|_2^2$ come from the Case $(\mathbf{a_2}, \mathbf{b_2})$ and Case $(\mathbf{a_1}, \mathbf{b_1})$, respectively.

9. Case $(\mathbf{a_3}, \mathbf{b_3})$. Lastly, we discuss the case

$$\begin{cases} K\big(((\mathbf{x}_0)_{S_1}, \mathbf{0}); (\mathbf{x}_0, \mathbf{0})\big) > \gamma s_*, \\ K\big((\mathbf{0}, (\mathbf{f}_0)_{S_2}); (\mathbf{0}, \mathbf{f}_0)\big) \leq \gamma t_*, \end{cases}$$

and $\|(\mathbf{x}_0)_{S_1}\|_0 \leq s_*$, namely, $S_1^c \supseteq S_*^c$ and $S_2^c \supseteq T_*^c$. Then, by the inequalities above, we get

$$(1 + \delta_{1,1}) K\big(((\mathbf{x}_0)_{S_1}, \mathbf{0}); (\mathbf{x}_0, \mathbf{0})\big) > \eta^2 s_*,$$

$\|(\mathbf{f}_0)_{S_2}\|_0 \leq t_*$ and $\|\bar{\mathbf{f}}\|_0 \leq t_*$. Note that $\mathbf{x}_0$ is similar to Case $(\mathbf{a_3}, \mathbf{b_3})$, and $\mathbf{f}_0$ is similar to Case $(\mathbf{a_1}, \mathbf{b_1})$. Therefore, one has

$$\begin{aligned} \left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 &= \left\| \begin{bmatrix} \hat{\mathbf{x}}^{DS} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{0} \end{bmatrix} \right\|_2^2 + \left\| \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{f}}^{DS} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_0 \end{bmatrix} \right\|_2^2 \\ &\leq 8Q_2^{'} \log(n+m) \sum_{i=1}^{n} \min\left\{|x_0(i)|^2, \sigma^2\right\} + M_2^{''} \sum_{j=1}^{m} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\} \end{aligned}$$

with probability at least $1 - 1/\sqrt{\pi \log(n+m)} - 1/n - 3e^{-K_2 m} - e^{-m/100}$.

Finally, by the discussions above, we can obtain the estimation

$$\begin{aligned} \|\hat{\mathbf{x}}^{DS} - \mathbf{x}_0\|_2^2 + \|\hat{\mathbf{f}}^{DS} - \mathbf{f}_0\|_2^2 \lesssim &\sum_{i \in \mathrm{supp}((\mathbf{x}_0)_{\max(s^*)})} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \|(\mathbf{x}_0)_{-\max(s_*)}\|_2^2 \\ &+ \sum_{j \in \mathrm{supp}((\mathbf{f}_0)_{\max(t^*)})} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\} + \|(\mathbf{f}_0)_{-\max(t_*)}\|_2^2 \end{aligned}$$

with high probability. This completes the proof. □

4. Numerical Experiments

In this section, we present several relative numerical experiments on corrupted compressed sensing. They are focused on three areas: (i) the effects of the balance parameter λ for the extended Dantzig selector (1.8) and the extended Lasso (1.9), (ii) the performance of the two models with $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}$ for approximately sparse signal recovery, (iii) the significance of the oracle inequality on sparse signal recovery. We solve these models via CVX [24]. We do 500 independent repeated tests and record those results. Here we take the approximately sparse (or compressive) signal $\mathbf{x}_0 \in \mathbb{R}^n$ with $\mathbf{x}_0(i_k) = 2^{-k}$ for $k = 1, \ldots, n$ and $n = 120$, which meets the definition for approximately sparse signal in [22]. The measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a Gaussian matrix with $a_{ij} \sim \mathcal{N}(0, 1/m)$, and the noise $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$ has the standard deviation $\sigma = 10^{-3}$. The relative error (RE) of the reconstructed signal $\hat{\mathbf{x}}$ and the original signal $\mathbf{x}_0$ is defined by $RE(\hat{\mathbf{x}}, \mathbf{x}_0) = \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2/\|\mathbf{x}_0\|_2$. The signal to noise ratio (SNR) is defined by $SNR(\hat{\mathbf{x}}, \mathbf{x}_0) = -20\log_{10} RE(\hat{\mathbf{x}}, \mathbf{x}_0)$. Take the outliers $\mathbf{f}_0 \in \mathbb{R}^m$ with $\text{supp}(\mathbf{f}_0) = T$ and $|T| = \lceil 0.05m \rceil$ (refer to [43]). If $i \in T$, $f_0(i) \sim \text{U}(-\iota, \iota)$ with the uniform distribution parameter $\iota = 100$ and $f_0(i) = 0$, otherwise. Therefore, the observations $\mathbf{b}$ can be written as follows:

$$b(j) = \begin{cases} \langle \mathbf{A}_i, \mathbf{x}_0\rangle + f_0(j) + e_j, & j \in T, \\ \langle \mathbf{A}_i, \mathbf{x}_0\rangle + e_j, & j \in T^c. \end{cases}$$

Note that we give those examples in details.

Example 4.1. We show the performances of the different values of the balance parameter λ for the extended Dantzig selector and extended Lasso models. Take the balance parameter λ with $\lambda < 1, \lambda = 1$ and $\lambda > 1$, respectively. Let the measurement $m \in \{46, 48, \ldots, 58\}$. These results in Fig. 4.1 show that the balance parameter has a large significance for the extended

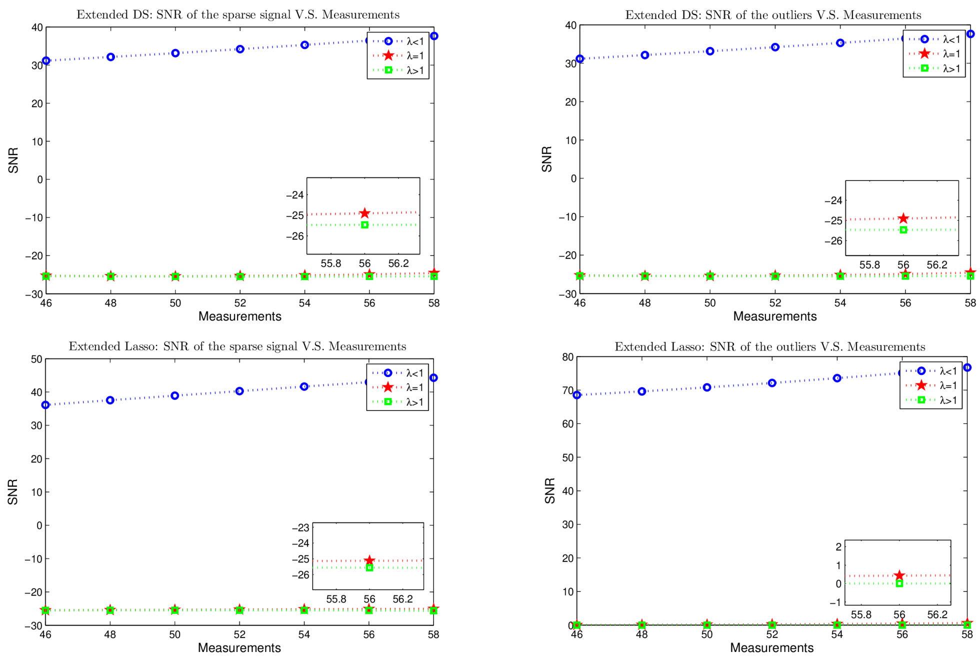

Fig. 4.1. The SNR versus the number m of measurements for different values of the balance parameter λ.

Dantzig selector model or the extended Lasso model. We find that the smaller parameter λ with $\lambda < 1$ has a better numerical performance. What should be pointed out is that numerical performances all are very poor for different values of the parameters $\lambda \geq 1$. Therefore, we only display the performance for $\lambda = 1$, and $\lambda = 10$.

Example 4.2. We display the numerical performance of the extended Dantzig selector and extended Lasso models. We vary the measurement $m \in \{46, 48, \ldots, 58\}$. The results in Fig. 4.2 show that both the extended Dantzig selector and the extended Lasso can stably recover the sparse signal $\mathbf{x}_0$ and the outlier $\mathbf{f}_0$.

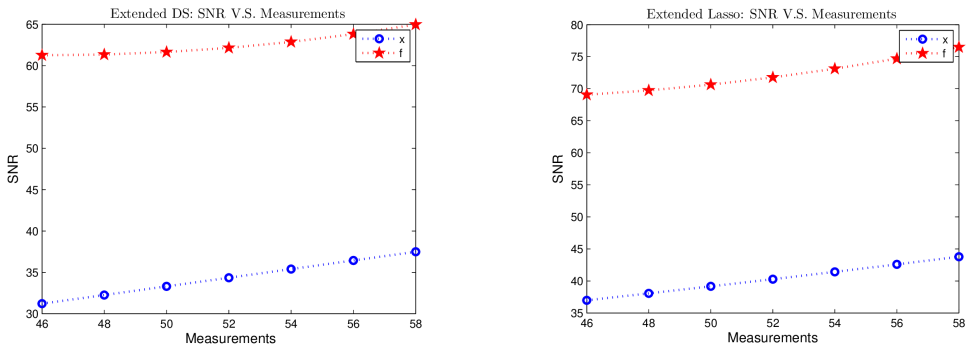

Fig. 4.2. The SNR versus the number of measurements m for the extended Dantzig selector and the extended Lasso models.

Finally, we design an experiment and discuss the significance of oracle inequalities for corrupted compressed sensing.

Example 4.3. We take the setup similar to that of [11, 23]. We randomly choose a subset $S \subset \mathbb{R}^n$ with $|S| = s$, and a subset $T \subset \mathbb{R}^m$ with $|T| = t$. Then, we generate the original signal $\mathbf{x}_0$ and $\mathbf{f}_0$ as follows:

$$x_0(i) = \begin{cases} \xi_i(1 + |d_i|), & i \in S, \\ 0, & \text{otherwise}, \end{cases}$$

$$f_0(j) = \begin{cases} \zeta_j(1 + |d_j|), & j \in T, \\ 0, & \text{otherwise}, \end{cases}$$

where ξ_i and ζ_j all obey the uniform distribution, $\xi_i \sim \mathcal{U}(-1, 1)$ and $\zeta_j \sim \mathcal{U}(-\iota, \iota)$ with $\iota = 100$, and $d_i \sim \mathcal{N}(0, 1)$. And we obtain the observations $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e} \in \mathbb{R}^m$ with noise $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$.

In this numerical experiment, we discuss two noise levels with $\sigma \in \{0.01, 0.05\}$, and various dimensions with $(m, n, s, t) = (72i, 256i, 8i, 2i)$ with $i = 1, 2, 3$. Here $(\hat{\mathbf{x}}, \hat{\mathbf{f}})$ is the solution of the extended Dantzig selector (1.8). We use the bias-removing two-stage procedure after Theorem 2.2 to obtain a refined solution $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{f}})$. The quality of $(\hat{\mathbf{x}}, \hat{\mathbf{f}})$ and $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{f}})$ are measured by

$$\begin{cases} \rho^2_{orign} = \dfrac{\sum_i |\hat{x}_j - x_0(i)|^2 + \sum_j |\hat{f}_j - f_0(j)|^2}{\sum_i \min\left\{|x_0(i)|^2, \sigma^2\right\} + \sum_j \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\}}, \\ \rho^2 = \dfrac{\sum_i |\widetilde{x}_j - x_0(i)|^2 + \sum_j |\widetilde{f}_j - f_0(j)|^2}{\sum_i \min\left\{|x_0(i)|^2, \sigma^2\right\} + \sum_j \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\}}. \end{cases}$$

Note that ρ^2_{orign} and ρ^2 are two indices on the performance of the extended Dantzig selector, which are called the preprocessing errors and the postprocessing errors in [32]. Of course, we are aiming for better selectors with smaller valves. These results in Table 4.1 show that the refined extended Dantzig selector performs better than the extended Dantzig selector.

Table 4.1: The performances of the extended Dantzig selector and the refined extended Dantzig selector.

i	$\sigma = 0.01$		$\sigma = 0.05$	
	ρ^2_{orign}	ρ^2	ρ^2_{orign}	ρ^2
$i = 1$	91.7136	18.1995	38.0129	10.4935
$i = 2$	18.9427	1.6273	16.8622	2.4910
$i = 3$	17.0844	1.9455	15.6463	2.3939

5. Conclusions and Comments

In this paper, we extended oracle inequalities of compressed sensing to corrupted compressed sensing for the Gaussian noises model $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{f}_0 + \mathbf{e}$ with $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$. Here $\mathbf{f} \in \mathbb{R}^m$ is a sparse vector with possible infinity large nonzero entries. And we got the oracle inequalities for the extended Dantzig selector and the extended Lasso for both sparse signal and approximately sparse signal $\mathbf{x}_0 \in \mathbb{R}^n$ recovery from the highly corrupted measurements. Suppose that the sensing matrix is sampled from the Gaussian measurement ensemble without the normalized columns. We discussed the probability of oracle inequalities

$$\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 + \|\hat{\mathbf{f}} - \mathbf{f}_0\|_2^2 \lesssim \sum_i \min\left\{|x_0(i)|^2, \sigma^2\right\} + \sum_j \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\}$$

for sparse signal recovery (see Theorems 2.1 and 2.2), and

$$\begin{aligned}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 + \|\hat{\mathbf{f}} - \mathbf{f}_0\|_2^2 \lesssim & \sum_{i \in \operatorname{supp}((\mathbf{x}_0)_{\max(s^*)})} \min\left\{|x_0(i)|^2, \sigma^2\right\} + \|(\mathbf{x}_0)_{-\max(s^*)}\|_2^2 \\ & + \sum_{j \in \operatorname{supp}((\mathbf{f}_0)_{\max(t^*)})} \min\left\{|\lambda f_0(j)|^2, \sigma^2\right\} + \|(\mathbf{f}_0)_{-\max(t^*)}\|_2^2\end{aligned}$$

for approximately sparse signal recovery (see Theorems 3.1 and 3.2), respectively. Here $\lambda > 0$ is the balance parameter between $\|\mathbf{x}\|_1$ and $\|\mathbf{f}\|_1$, i.e. $\|\mathbf{x}\|_1 + \lambda\|\mathbf{f}\|_1$. What should be pointed out is that our results are not trivial extensions of the compressed sensing (non-corrupted) case $\mathbf{b} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}$.

Acknowledgments. We thanks for Xiang-Yu Zheng for his discussion on Proposition 2.1.

Peng Li is supported by the Natural Science Foundation of China (Grant No. 12201268) and by the Science and Technology Program of Gansu Province of China (Grant No. 21JR7RA511).

References

[1] B. Adcock, A. Bao, J.D. Jakeman, and A. Narayan, Compressed sensing with sparse corruptions: Fault-tolerant sparse collocation approximations, *SIAM-ASA J. Uncertain.*, **6** (2018), 1424–1453.

[2] P.J. Bickel, Y. Ritov, and A.B. Tsybakov, Simultaneous analysis of Lasso and Dantzig selector, *Ann. Stat.*, **37** (2009), 1705–1732.

[3] T.T. Cai, L. Wang, and G. Xu, Stable recovery of sparse signals and an oracle inequality, *IEEE Trans. Inform. Theory*, **56** (2010), 3516–3522.
[4] T.T. Cai, G. Xu, and J. Zhang, On recovery of sparse signals via ℓ_1 minimization, *IEEE Trans. Inform. Theory*, **55**:7 (2009), 3388–3397.
[5] T.T. Cai and A. Zhang, Compressed sensing and affine rank minimization under restricted isometry, *IEEE Trans. Signal Process.*, **61** (2013), 3279–3290.
[6] E.J. Candès, Modern statistical estimation via oracle inequalities, *Acta Numer.*, **15** (2006), 257–325.
[7] E.J. Candès and Y. Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, *IEEE Trans. Inform. Theory*, **57** (2011), 2342–2359.
[8] 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.
[9] E.J. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, **51** (2005), 4203–4215.
[10] E.J. Candès and T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, **52** (2006), 5406–5425.
[11] E.J. Candès and T. Tao, The Dantzig selector: Statistical estimation when p is much larger than n, *Ann. Stat.*, **35** (2007), 2313–2351.
[12] L. Cavalier, G. Golubev, D. Picard, and A. Tsybakov, Oracle inequalities for inverse problems, *Ann. Stat.*, **30**:3 (2002), 843–874.
[13] S.S. Chen, *Basis Pursuit*, PhD Thesis, Stanford University, 1995.
[14] S.S. Chen, D.L. Donoho, and M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.*, **43**:1 (2001), 129–159.
[15] A. Cohen, W. Dahmen, and R. Devore, Compressed sensing and best k-term approximation, *J. Amer. Math. Soc.*, **22** (2009), 211–231.
[16] D.L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, **52** (2006), 1289–1306.
[17] D.L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, *Proc. Natl. Acad. Sci. U.S.A.*, (2003), 2197–2202.
[18] D.L. Donoho, M. Elad, and V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inform. Theory*, **52**:1 (2005), 6–18.
[19] D.L. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inform. Theory*, **47**:7 (2001), 2845–2862.
[20] D.L. Donoho and J.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81** (1994), 425–455.
[21] D.L. Donoho and G. Kutyniok, Microlocal analysis of the geometric separation problem, *Commun. Pure Appl. Math.*, **66** (2013), 1–47.
[22] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, in: *Applied and Numerical Harmonic Analysis Series*, Birkhäuser, 2013.
[23] H. Ge and P. Li, The Dantzig selector: Recovery of signal via ℓ_1-$\alpha\ell_2$ minimization, *Inverse Probl.*, **38**:1 (2021), 015006.
[24] M. Grant, S. Boyd, and Y. Ye, *CVX: Matlab Software for Disciplined Convex Programming*, 2011, `http://cvxr.com/cvx`.
[25] J.N. Laska, M.A. Davenport, and R.G. Baraniuk, Exact signal recovery from sparsely corrupted measurements through the pursuit of justice, in: *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, (2009), 1556–1560.
[26] N. Li, *Signal Recovery and Separation Based on Corrupted Observations*, PhD Thesis, China Academy of Engineering Physics, 2022.
[27] N. Li, W. Chen, and P. Li, Stable recovery of signals from highly corrupted measurements, *IEEE Access*, **6** (2018), 62865–62873.
[28] P. Li and W. Chen, Signal recovery under mutual incoherence property and oracle inequalities,

Front. Math. China, **13** (2018), 1369–1396.

[29] P. Li and W. Chen, Signal recovery under cumulative coherence, *J. Comput. Appl. Math.*,**346** (2019), 399–417.

[30] X. Li, Compressed sensing and matrix completion with constant proportion of corruptions, *Constr. Approx.*, **37** (2013), 73–99.

[31] J. Lin and S. Li, Sparse recovery with coherent tight frames via analysis Dantzig selector and analysis LASSO, *Appl. Comput. Harmon. Anal.*, **37**:1 (2014), 126–139.

[32] Z. Lu, T.K. Pong, and Y. Zhang, An alternating direction method for finding Dantzig selectors, *Comput. Statist. Data Anal.*, **56**:12 (2012), 4037–4046.

[33] S.G. Mallat and G. Yu, Super-resolution with sparse mixing estimators, *IEEE Trans. Image Process.*, **19** (2010), 2889–2900.

[34] N.H. Nguyen and T.D. Tran, Robust Lasso with missing and grossly corrupted observations, *IEEE Trans. Inform. Theory*, **59** (2012), 2036–2058.

[35] N.H. Nguyen and T.D. Tran, Exact recoverability from dense corrupted observations via ℓ_1-minimization, *IEEE Trans. Inform. Theory*, **59**:4 (2013), 2017–2035.

[36] C. Novak, C. Studer, A. Burg, and G. Matz, The effect of unreliable LLR storage on the performance of MIMO-BICM, in: *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, (2010), 736–740.

[37] C. Studer and R.G. Baraniuk, Stable restoration and separation of approximately sparse signals, *Appl. Comput. Harmon. Anal.*, **37** (2014), 12–35.

[38] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, Recovery of sparsely corrupted signals, *IEEE Trans. Inform. Theory*, **58** (2011), 3115–3130.

[39] Q. Sun, Sparse approximation property and stable recovery of sparse signals from noisy measurements, *IEEE Trans. Signal Proces.*, **59** (2011), 5086–5090.

[40] J. Tanner and S. Vary, Compressed sensing of low-rank plus sparse matrices, *Appl. Comput. Harmon. Anal.*, **64** (2023), 254–293.

[41] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc., Ser. B., Methodol.*, **58** (1996), 267–288.

[42] J.A. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory*, **50**:10 (2004), 2231–2242.

[43] Q. Wan, H. Duan, J. Fang, H. Li, and Z. Xing, Robust Bayesian compressed sensing with outliers, *Signal Process.*, **140** (2017), 104–109.

[44] L. Wang, The L_1 penalized LAD estimator for high dimensional linear regression, *J. Multivariate Anal.*, **120** (2013), 135–151.

[45] J. Wright and Y. Ma, Dense error correction via ℓ^1-minimization, *IEEE Trans. Inform. Theory*, **56**:7 (2010), 3540–3560.

[46] J. Wright and Y. Ma, *High-dimensional Data Analysis with Low-Dimensional Models: Principles, Computation, and Applications*, Cambridge University Press, 2022.

[47] J. Yang and Y. Zhang, Alternating direction algorithms for ℓ_1-problems in compressive sensing, *SIAM J. Sci. Comput.*, **33**:1 (2011), 250–278.

[48] P. Zhang, L. Gan, C. Ling, and S. Sun, Uniform recovery bounds for structured random matrices in corrupted compressed sensing, *IEEE Trans. Signal Proces.*, **66** (2018), 2086–2097.

[49] R. Zhang and S. Li, A proof of conjecture on restricted isometry property constants δ_{tk} $(0<t<4/3)$, *IEEE Trans. Inform. Theory*, **64**:3 (2017), 1699–1705.