Linear Algebra and its Applications 557 (2018) 134–173

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Robust group lasso: Model and recoverability [☆]

Xiaohan Wei [a,*], Qing Ling [b], Zhu Han [c]

[a] Department of Electrical Engineering, University of Southern California, United States of America
[b] School of Data and Computer Science, Sun Yat-Sen University, China
[c] Department of Electrical and Computer Engineering, University of Houston, United States of America

ARTICLE INFO

Article history:
Received 19 November 2016
Accepted 30 June 2018
Available online xxxx
Submitted by H. Rauhut

MSC:
60D05
94A12

Keywords:
Compressed sensing
Group sparsity
Sparse noise
Restricted isometry property
Golfing scheme

ABSTRACT

This paper considers the problem of recovering a group sparse signal matrix $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_L]$ from sparsely corrupted measurements $\mathbf{M} = [\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L] + \mathbf{S}$, where $\mathbf{A}_{(i)}$'s are known sensing matrices and $\mathbf{S}$ is an unknown sparse error matrix. A robust group lasso (RGL) model is proposed to recover $\mathbf{Y}$ and $\mathbf{S}$ through simultaneously minimizing the $\ell_{2,1}$-norm of $\mathbf{Y}$ and the ℓ_1-norm of $\mathbf{S}$ under the measurement constraints. We prove that $\mathbf{Y}$ and $\mathbf{S}$ can be exactly recovered from the RGL model with high probability for a very general class of $\mathbf{A}_{(i)}$'s.

[☆] Qing Ling's research is partially supported by National Science Foundation of China under Grant 61573331 and by the National Science Foundation Anhui under Grant 1608085QF130. Zhu Han's research is partially supported by US MURI, NSF CNS-1717454, CNS-1731424, CNS-1702850, CNS-1646607.

[*] Corresponding author.
E-mail addresses: xiaohanw@usc.edu (X. Wei), qingling@mail.ustc.edu.cn (Q. Ling), zhan2@uh.edu (Z. Han).

https://doi.org/10.1016/j.laa.2018.06.034
0024-3795/© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Consider the problem of recovering a group sparse matrix $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_L] \in \mathbb{R}^{n \times L}$ from sparsely corrupted measurements

$$\mathbf{M} = [\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L] + \mathbf{S}, \tag{1}$$

where $\mathbf{M} = [\mathbf{m}_1, \cdots, \mathbf{m}_L] \in \mathbb{R}^{m \times L}$ is a measurement matrix, $\mathbf{A}_{(i)} \in \mathbb{R}^{m \times n}$ is the i-th sensing matrix, and $\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_L] \in \mathbb{R}^{m \times L}$ is an unknown sparse error matrix. The error matrix $\mathbf{S}$ is sparse as it has only a small number of nonzero entries. The signal matrix $\mathbf{Y}$ is group sparse, meaning that $\mathbf{Y}$ is sparse and its nonzero entries appear in a small number of common rows.

Given $\mathbf{M}$ and $\mathbf{A}_{(i)}$'s, our goal is to recover $\mathbf{Y}$ and $\mathbf{S}$ from the linear measurement equation (1). In this paper, we propose to accomplish the recovery task through solving the following robust group lasso (RGL) model

$$\min_{\mathbf{Y},\mathbf{S}} \quad \|\mathbf{Y}\|_{2,1} + \lambda\|\mathbf{S}\|_1, \tag{2}$$

$$s.t. \quad \mathbf{M} = [\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L] + \mathbf{S}. \tag{3}$$

Denoting y_{ij} and s_{ij} as the (i,j)-th entries of $\mathbf{Y}$ and $\mathbf{S}$, respectively, $\|\mathbf{Y}\|_{2,1} \triangleq \sum_{i=1}^n \sqrt{\sum_{j=1}^L y_{ij}^2}$ is defined as the $\ell_{2,1}$-norm of $\mathbf{Y}$ and $\|\mathbf{S}\|_1 \triangleq \sum_{i=1}^m \sum_{j=1}^L |s_{ij}|$ is defined as the ℓ_1-norm of $\mathbf{S}$. Minimizing the $\ell_{2,1}$-norm term promotes group sparsity of $\mathbf{Y}$ while minimizing the ℓ_1-norm term promotes sparsity of $\mathbf{S}$; λ is a nonnegative parameter to balance the two terms. We prove that solving the RGL model (2)–(3), which is a convex program, enables exact recovery of $\mathbf{Y}$ and $\mathbf{S}$ with high probability, given that $\mathbf{A}_{(i)}$'s satisfy certain conditions.

1.1. *From group lasso to robust group lasso*

Sparse signal recovery has attracted much research interest in the signal processing and optimization communities during the past few years. Various sparsity models have been proposed to better exploit the sparse structures of high-dimensional data, such as sparsity of a vector [1], [2], group sparsity of vectors [3], and low-rankness of a matrix [4]. For more topics related to sparse signal recovery, readers are referred to the recent survey paper [5].

In this paper we are interested in the recovery of group sparse (also known as block sparse [6] or jointly sparse [7]) signals which finds a variety of applications such as direction-of-arrival estimation [8], [9], collaborative spectrum sensing [10–12] and motion detection [13]. A signal matrix $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_L] \in \mathbb{R}^{n \times L}$ is called k-group sparse if k rows of $\mathbf{Y}$ are nonzero. A measurement matrix $\mathbf{M} = [\mathbf{m}_1, \cdots, \mathbf{m}_L] \in \mathbb{R}^{m \times L}$ is taken from linear projections $\mathbf{m}_i = \mathbf{A}_{(i)}\mathbf{y}_i$, $i = 1, \cdots, L$, where $\mathbf{A}_{(i)} \in \mathbb{R}^{m \times n}$ is a sensing matrix. In order to recover $\mathbf{Y}$ from $\mathbf{A}_{(i)}$'s and $\mathbf{M}$, the standard $\ell_{2,1}$-norm minimization formulation proposes to solve a convex program

$$\begin{aligned} &\min_{\mathbf{Y}} \ \ \|\mathbf{Y}\|_{2,1}, \\ &s.t. \quad \mathbf{M} = [\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L]. \end{aligned} \tag{4}$$

This is a straightforward extension from the canonical ℓ_1-norm minimization formulation that recovers a sparse vector. Theoretical guarantee of exact recovery has been developed based on the restricted isometric property (RIP) of $\mathbf{A}_{(i)}$'s [14]. The performance of exploiting more structures of $\mathbf{Y}$ by simultaneously minimizing the $\ell_{2,1}$-norm and the nuclear norm is analyzed in [15].

Consider that in practice the measurements are often corrupted by random noise, resulting in $\mathbf{M} = [\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L] + \mathbf{N}$ where $\mathbf{N} = [\mathbf{n}_1, \cdots, \mathbf{n}_L] \in \mathbb{R}^{m\times L}$ is a noise matrix. To address the noise-corrupted case, the group lasso model in [3] solves

$$\begin{aligned} &\min_{\mathbf{Y},\mathbf{E}} \ \ \|\mathbf{Y}\|_{2,1} + \gamma\|\mathbf{N}\|_F^2, \\ &s.t. \quad \mathbf{M} = [\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L] + \mathbf{N}, \end{aligned} \tag{5}$$

where γ is a nonnegative parameter and $\|\mathbf{N}\|_F$ is the Frobenius norm of $\mathbf{N}$. An alternative to (5) is

$$\begin{aligned} &\min_{\mathbf{Y}} \ \ \|\mathbf{Y}\|_{2,1}, \\ &s.t. \quad \|\mathbf{M} - [\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L]\|_F^2 \le \varepsilon^2, \end{aligned} \tag{6}$$

where ε controls the noise level. It has been shown in [14] that if the sensing matrices $\mathbf{A}_{(i)}$'s satisfy RIP, then the distance between the solution to (6) and the true signal matrix, which is measured by the Frobenius norm, is within a constant multiple of ε.

The exact recovery guarantee for (6) is elegant, but works only if the noise level ε is not too large. However, in many practical applications, some of the measurements may be seriously contaminated or even missing due to uncertainties such as sensor failures and transmission errors. Meanwhile, this kind of measurement errors are often sparse (see [16] for detailed discussions). In this case, the exact recovery guarantee does not hold and the solution of (6) can be far away from the true signal matrix.

The need of handling large but sparse measurement errors in the group sparse signal recovery problem motivates the RGL model (2)–(3), which has found successful applications in, for example, the cognitive network sensing problem [16]. Efficient decentralized algorithms solving (2)–(3) are proposed in [17] and their performances are evaluated via extensive simulations. In (2)–(3), the measurement matrix $\mathbf{M}$ is contaminated by a sparse error matrix $\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_L] \in \mathbb{R}^{m\times L}$ whose nonzero entries might be unbounded. Through simultaneously minimizing the $\ell_{2,1}$-norm of $\mathbf{Y}$ and the ℓ_1 norm of $\mathbf{S}$, we expect to recover the group sparse signal matrix $\mathbf{Y}$ and the sparse error matrix $\mathbf{S}$.

The RGL model (2)–(3) is closely related to robust lasso and robust principle component analysis (RPCA), both of which have been proved effectively in recovering true signal from sparse gross corruptions. The robust lasso model, which has been discussed extensively in [18], [19], [20], minimizes ℓ_1-norm of a sparse signal vector and the ℓ_1-norm of a sparse error vector simultaneously in order to get rid of sparse corruptions. Whereas the RPCA model, which is first proposed in [21] and then extended by [22] and [23], recovers a low rank matrix by minimizing the nuclear norm of signal matrix plus ℓ_1-norm of sparse error matrix.

1.2. Contribution and paper organization

This paper proves that with high probability, the proposed RGL model (2)–(3) exactly recovers the group sparse signal matrix and the sparse error matrix simultaneously under certain restrictions on measurements for a very general class of sample matrices.

The rest of this paper is organized as follows. Section 2 provides the main result (see Theorem 1) on the recoverability of the RGL model (2)–(3) under assumptions on the sensing matrices and the true signal and error matrices (see Assumptions 1–4). Section 2 also introduces several supporting lemmas and corollaries (see Lemmas 1–4 and Corollaries 1–2). Section 3 gives the dual certificate of (2)–(3), which is a sufficient condition guaranteeing the exact recovery from the RGL model with high probability. Their proofs are based on two supporting lemmas (see Lemmas 5–6). Section 4 proves that the inexact dual certificate of (2)–(3) can be satisfied through a constructive manner (see Theorem 3 and Lemma 7). This way, we prove the main result given in Section 2. Section 5 concludes the paper.

1.3. Notations

We introduce several notations that are used in the subsequent sections. Bold uppercase letters denote matrices, whereas bold lowercase letters with subscripts and superscripts stand for column vectors and row vectors, respectively. For a matrix $\mathbf{U}$, we denote $\mathbf{u}_i$ as its i-th column, $\mathbf{u}^i$ as its j-th row, and u_{ij} as its (i,j)-th element. For a given vector $\mathbf{u}$, we denote u_i as its i-th element. The notations $\{\mathbf{U}_{(i)}\}$ and $\{\mathbf{u}_{(i)}\}$ denote the family of matrices and vectors indexed by i, respectively. The notations $\{\mathbf{U}_{(i,j)}\}$ and $\{\mathbf{u}_{(i,j)}\}$ denote the family of matrices and vectors indexed by (i,j), respectively. $\mathrm{vec}(\cdot)$ is the vectorizing operator that stacks the columns of a matrix one after another. $\{\cdot\}'$ denotes the transpose operator. $\mathbf{diag}\{\cdot\}$ represents a diagonal matrix and $\mathbf{BLKdiag}\{\cdot\}$ represents a block diagonal matrix. The notation $\langle\cdot,\cdot\rangle$ denotes the inner product, when applying to two matrices $\mathbf{U}$ and $\mathbf{V}$. $\mathrm{sgn}(\mathbf{u})$ and $\mathrm{sgn}(\mathbf{U})$ are sign vector and sign matrix for $\mathbf{u}$ and $\mathbf{U}$, respectively.

Additionally, we use several standard matrix and vector norms. For a vector $\mathbf{u} \in \mathbb{R}^n$, define

- ℓ_2-norm: $\|\mathbf{u}\|_2 = \sqrt{\sum_{j=1}^n u_j^2}$.
- ℓ_1-norm: $\|\mathbf{u}\|_1 = \sum_{j=1}^n |u_j|$.

For a matrix $\mathbf{U} \in \mathbb{R}^{m\times n}$, define

- $\ell_{2,1}$-norm: $\|\mathbf{U}\|_{2,1} = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{n} u_{ij}^2}$.
- $\ell_{2,\infty}$-norm: $\|\mathbf{U}\|_{2,\infty} = \max_i \sqrt{\sum_{j=1}^{n} u_{ij}^2}$.
- ℓ_1-norm: $\|\mathbf{U}\|_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} |u_{ij}|$.
- Frobenius norm: $\|\mathbf{U}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} u_{ij}^2}$.
- ℓ_∞-norm: $\|\mathbf{U}\|_\infty = \max_{i,j} |u_{ij}|$.

Also, we use the notation $\|\mathbf{U}\|_{(p,q)}$ to denote induced norms, which stands for

$$\|\mathbf{U}\|_{(p,q)} = \max_{\mathbf{x}\in\mathbb{R}^n} \frac{\|\mathbf{U}\mathbf{x}\|_p}{\|\mathbf{x}\|_q}.$$

For the signal matrix $\mathbf{Y} \in \mathbb{R}^{n\times L}$ and noise matrix $\mathbf{S} \in \mathbb{R}^{m\times L}$, we use the following set notations throughout the paper.

- T: The row group support (namely, the set of row coordinates corresponding to the nonzero rows of the signal matrix) whose cardinality is denoted as $k_T = |T|$.
- T^c: The complement of T (namely, $\{1, \cdots, n\} \setminus T$).
- Ω: The support of error matrix (namely, the set of coordinates corresponding to the nonzero elements of the error matrix) whose cardinality is denoted as $k_\Omega = |\Omega|$.
- Ω^c: The complement of Ω (namely, $\{1, \cdots, n\} \times \{1, \cdots, L\} \setminus \Omega$).
- Ω_i: The support of the i-th column the error matrix whose cardinality is denoted as $k_{\Omega_i} = |\Omega_i|$.
- Ω_i^c: The complement of Ω_i (namely, $\{1, \cdots, n\} \setminus \Omega_i$).
- Ω_i^*: An arbitrary fixed subset of Ω_i^c with cardinality $m - k_{\max}$, where $k_{\max} = \max_i k_{\Omega_i}$. Intuitively, Ω_i^* stands for the *maximal non-corrupted set* across different $i \in \{1, \cdots, L\}$.

For any given matrices $\mathbf{U} \in \mathbb{R}^{m\times L}$, $\mathbf{V} \in \mathbb{R}^{n\times L}$ and given vectors $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^n$, define orthogonal projection operators as follows.

- $\mathcal{P}_\Omega\mathbf{U}$: The orthogonal projection of matrix $\mathbf{U}$ onto Ω (namely, set every entry of $\mathbf{U}$ whose coordinate belongs to Ω^c as 0 while keep other entries unchanged).
- $\mathcal{P}_{\Omega_i}\mathbf{u}$, $\mathcal{P}_{\Omega_i^c}\mathbf{u}$, $\mathcal{P}_{\Omega_i^*}\mathbf{u}$: The orthogonal projections of $\mathbf{u}$ onto Ω_i, Ω_i^c, and Ω_i^*, respectively.
- $\mathcal{P}_T\mathbf{v}$: The orthogonal projection of $\mathbf{v}$ onto T.
- $\mathcal{P}_{\Omega_i}\mathbf{U}$, $\mathcal{P}_{\Omega_i^c}\mathbf{U}$, and $\mathcal{P}_{\Omega_i^*}\mathbf{U}$: The orthogonal projections of each column of $\mathbf{U}$ onto Ω_i, Ω_i^c, and Ω_i^*, respectively (namely, $\mathcal{P}_{\Omega_i}\mathbf{U} = [\mathcal{P}_{\Omega_i}\mathbf{u}_1, \cdots, \mathcal{P}_{\Omega_i}\mathbf{u}_L]$, $\mathcal{P}_{\Omega_i^c}\mathbf{U} = [\mathcal{P}_{\Omega_i^c}\mathbf{u}_1, \cdots, \mathcal{P}_{\Omega_i^c}\mathbf{u}_L]$ and $\mathcal{P}_{\Omega_i^*}\mathbf{U} = [\mathcal{P}_{\Omega_i^*}\mathbf{u}_1, \cdots, \mathcal{P}_{\Omega_i^*}\mathbf{u}_L]$).
- $\mathcal{P}_T\mathbf{V}$: The orthogonal projection of each column of $\mathbf{V}$ onto T.

Furthermore, we admit a notational convention that for any projection operator $\mathcal{P}$ and corresponding matrix $\mathbf{U}$ (or vector $\mathbf{u}$), it holds

$$\mathbf{U}'\mathcal{P} = (\mathcal{P}\mathbf{U})' \, (\text{or } \mathbf{u}'\mathcal{P} = (\mathcal{P}\mathbf{u})').$$

Finally, by saying an event occurs with high probability, we mean that the occurring probability of the event is at least $1 - C(nL)^{-1}$ where C is a constant.

2. Main result of exact recovery

This section provides the theoretical performance guarantee of the RGL model (2)–(3). Section 2.1 makes several assumptions under which (2)–(3) recovers the true group sparse signal and sparse error matrices with high probability. The main result is summarized in Theorem 1. Section 2.2 discusses several related results and applications. Section 2.3 gives several probability tools that are useful in the proof of the main result.

2.1. Assumptions and main result

We start from several assumptions on the sensing matrices, as well as the true group sparse signal and sparse error matrices.

Consider L distributions $\{\mathcal{F}_i\}_{i=1}^{L}$ in $\mathbb{R}^n$ and an independently sampled vector $\mathbf{a}_{(i)}$ from each $\mathcal{F}_i$. The correlation matrix is defined as

$$\mathbf{\Sigma}_{(i)} = \mathbb{E}\left[\mathbf{a}_{(i)}\mathbf{a}_{(i)}'\right],$$

and the corresponding condition number is bounded as follows

$$\sqrt{\frac{\lambda_{\max}\{\mathbf{\Sigma}_{(i)}\}}{\lambda_{\min}\{\mathbf{\Sigma}_{(i)}\}}} \leq \kappa, \quad \forall i \in \{1, 2, \cdots, L\},$$

where κ is a positive constant and $\lambda_{\max}\{\cdot\}$, $\lambda_{\min}\{\cdot\}$ denote the largest and smallest eigenvalues of a matrix, respectively. Observe that this condition number is finite if and only if the covariance matrix is invertible, and is larger than or equal to 1 in any case.

Assumption 1. For $i = 1, \cdots, L$, define the i-th sensing matrix as

$$\mathbf{A}_{(i)} \triangleq \frac{1}{\sqrt{m}} \begin{pmatrix} \mathbf{a}_{(i)1}' \\ \vdots \\ \mathbf{a}_{(i)m}' \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

Therein, $\{\mathbf{a}_{(i)1}, \cdots, \mathbf{a}_{(i)m}\}$ is assumed to be a sequence of i.i.d. random vectors drawn from the distribution $\mathcal{F}_i$ in $\mathbb{R}^n$.

By Assumption 1, we suppose that every sensing matrix $\mathbf{A}_{(i)}$ is randomly sampled from a corresponding distribution $\mathcal{F}_i$. We proceed to assume the properties of the distributions $\{\mathcal{F}_i\}_{i=1}^L$.

Assumption 2. For $i = 1, \cdots, L$, every distribution $\mathcal{F}_i$ satisfy the following two properties.

- **Completeness**: The correlation matrix $\mathbf{\Sigma}_{(i)}$ is invertible.
- **Incoherence:** Each sensing vector $\mathbf{a}_{(i)}$ sampled from $\mathcal{F}_i$ satisfies

$$\max_{j\in\{1,\cdots,n\}} |\langle \mathbf{a}_{(i)}, \mathbf{e}_k \rangle| \le \sqrt{\mu}, \tag{7}$$

$$\max_{j\in\{1,\cdots,n\}} |\langle \mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)}, \mathbf{e}_k \rangle| \le \sqrt{\mu}, \tag{8}$$

for some fixed constant $\mu \ge 1$, where $\{\mathbf{e}_k\}_{k=1}^n$ is the standard basis in $\mathbb{R}^n$.

We call μ as the incoherence parameter. Note that this incoherence condition is stronger than the one originally presented in [26], which does not require (8). If one wants to get rid of (8), then some other restrictions must be imposed on the sensing matrices (see [27] for related results).

Observe that the bounds (7) and (8) in Assumption 2 are meaningless unless we fix the scale of $\mathbf{a}_{(i)}$. Thus, we have the following assumption.

Assumption 3. The correlation matrix $\mathbf{\Sigma}_{(i)}$ satisfies

$$\lambda_{\max}\{\mathbf{\Sigma}_{(i)}\} = \lambda_{\min}\{\mathbf{\Sigma}_{(i)}\}^{-1}, \tag{9}$$

for any $\mathcal{F}_i,\ i = 1, \cdots, L$.

Given any complete $\mathcal{F}_i$, (9) can always be achieved by scaling $\mathbf{a}_{(i)}$ up or down. This is true because if we scale $\mathbf{a}_{(i)}$ up, then $\lambda_{\max}\{\mathbf{\Sigma}_{(i)}\}$ increases and $\lambda_{\min}\{\mathbf{\Sigma}_{(i)}\}^{-1}$ decreases. Observe that the optimization problem (2)–(3) is invariant under scaling. Thus, Assumption 3 does not pose any extra constraint.

Denote $\overline{\mathbf{Y}}$ and $\overline{\mathbf{S}}$ as the true group sparse signal and sparse error matrices to recover, respectively. The row group support of $\overline{\mathbf{Y}}$ and the support of $\overline{\mathbf{S}}$ are fixed and denoted as T and Ω, respectively.

The assumption on $\overline{\mathbf{Y}}$ and $\overline{\mathbf{S}}$ is given as follows,

Assumption 4. The true signal matrix $\overline{\mathbf{Y}}$ and error matrix $\overline{\mathbf{S}}$ satisfy the following two properties.

- **Random sign:** The signs of the elements of $\overline{\mathbf{Y}}$ and $\overline{\mathbf{S}}$ are i.i.d. and equally likely to be $+1$ or -1.

- **Row incoherence:** Let $\overline{\mathbf{y}}^i$ be the i-th row of $\overline{\mathbf{Y}}$. Then, there exists a fixed constant $\nu \geq 1$ so that

$$\max_{l\in\{1,2,\cdots,L\}} \left| \left\langle \frac{\overline{\mathbf{y}}^i}{\|\overline{\mathbf{y}}^i\|_2}, \mathbf{e}_l \right\rangle \right| \leq \sqrt{\frac{\nu}{L}},$$

for any $i \in T$ and $l \in \{1, 2, \cdots, L\}$, where $\{\mathbf{e}_l\}_{l=1}^L$ is a standard basis in $\mathbb{R}^L$.

Under the assumptions stated above, we have the following main theorem on the recoverability of the RGL model (2)–(3).

Theorem 1. *Let $\theta \in [1, L]$ be a tradeoff parameter. Under Assumptions 1–4, the solution pair ($\hat{\mathbf{Y}}$, $\hat{\mathbf{S}}$) to the optimization problem (2)–(3) is exact and unique with probability at least $1 - (16 + 2e^{\frac{1}{4}})(nL)^{-1}$, provided that $\lambda = \sqrt{\frac{\theta}{L\log(nL)}}$,*

$$k_T \leq \alpha \frac{m}{\max\left\{\kappa, \frac{\nu}{\theta}\right\} \mu \log^2(nL)}, \quad k_\Omega \leq \beta \frac{mL}{\mu\theta}, \quad k_{\max} \leq \gamma \frac{m}{\kappa}, \tag{10}$$

where $k_{\max} \triangleq \max_i k_{\Omega_i}$, and $\alpha \leq \frac{1}{9600}$, $\beta \leq \frac{1}{3136}$, $\gamma \leq \frac{1}{4}$ are all positive constants.[1]

Note that from (10), the constant θ is indeed a trade-off parameter. Choosing θ large relaxes the constraint on row supports but restricts the number of sparse errors, and vice versa.

The incoherence condition of sensing vectors (Assumption 2) is common in lasso and robust lasso literatures, which implies that the columns of $[\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L]$ are not aligned with sparse vectors. On the other hand, the row incoherence (Assumption 4) is unique in the current group lasso context. It indicates that the rows of $[\mathbf{A}_{(1)}\mathbf{y}_1, \cdots, \mathbf{A}_{(L)}\mathbf{y}_L]$ are not aligned with sparse vectors either. Note that this condition is trivial when $\nu = L$, in which cases choosing $\theta = L$ in Theorem 1 gives the measurement bound $mL \geq \mathcal{O}(k_T L \log^2(nL))$ and $m \geq \mathcal{O}(k_\Omega)$. In general when L is large and the rows of the true signal matrix $\overline{Y}$ are dense, we could have $\nu \ll L$, in which case choosing θ close to 1 results in an improved measurement bound $mL \geq \mathcal{O}(k_T L \log^2(nL))$ and $mL \geq \mathcal{O}(k_\Omega)$. For matrix recovery literatures, similar two-way incoherence assumptions are adopted. For example, in RPCA ([4], [21]), the signal matrices are assumed to have both left and right singular vectors being incoherent with sparse vectors.

Finally, notice that $\overline{\mathbf{Y}}$ and $\overline{\mathbf{S}}$ are assumed to have fixed supports but random signs. This assumption is commonly adopted in the RIPless analysis of lasso ([19], [26]) and seems to be crucial in the proof. Alternatively, one could always avoid assuming random signs of $\overline{\mathbf{Y}}$ by taking the sensing matrices as $\mathbf{A}_{(i)}\mathbf{D}_{(i)}$, $i = 1, 2, \cdots, L$, where $\mathbf{A}_{(i)}$ is

[1] The bounds on α, β, γ are chosen such that all the requirements on these constants in the subsequent lemmas and theorems are met.

the same as above and $\mathbf{D}_{(i)} \in \mathbb{R}^{n\times n}$ is a diagonal matrix with each diagonal entry i.i.d. and equally likely to be $+1$ or -1. It is still an open problem whether or not one can completely get rid of this random sign assumption.

2.2. Related works and applications

2.2.1. Examples

In this section, we provide several examples demonstrating that our sensing model covers a wide range of random measurements.

- Random Fourier measurements: An important application for our sensing model is the random subsampled discrete Fourier transform (DFT) matrix. Consider an $n \times n$ matrix $\mathbf{W}$ with each entry

$$W_\omega^t = e^{-i2\pi\omega t/n}, \ w, t \in \{0, 1, \cdots, n-1\}.$$

  Let $\mathbf{A}_{(i)} \in \mathbb{C}^{m\times n}$ be a sensing matrix so that each row is sampled uniformly at random from the rows of $\mathbf{W}$.[2] It can be easily verified that $\mathbb{E}\left[\mathbf{a}_{(i)}\mathbf{a}'_{(i)}\right] = \mathbf{I}$ and the incoherence parameter $\mu = 1$. Such sensing model arises in various applications including magnetic resonance imaging and collaborative spectrum sensing ([11]). More generally, any random row sub-matrix $\mathbf{A}_{(i)}$ of an arbitrary bounded orthogonal matrix $\mathbf{W}$ (with incoherent rows) fits into our model. This includes sampling from the class of Hadamard matrices (orthogonal matrices with all entries $+1$ or -1).
- Random sampling from frames: A frame is a set of vectors $\{\mathbf{u}_k\}_{k=1}^K \subseteq \mathbb{R}^n$ which satisfies the following relation: There exist positive constants C_1 and C_2 such that

$$C_1\|\mathbf{x}\|_2^2 \le \frac{1}{K}\sum_{k=1}^K \langle \mathbf{u}_k, \mathbf{x}\rangle^2 \le C_2\|\mathbf{x}\|_2^2, \ \forall \mathbf{x} \in \mathbb{R}^d.$$

  This can be viewed as a generalization of orthogonal systems without linear independence. Furthermore, assume that each $\mathbf{u}_k$ satisfies the incoherence condition. Let $\mathbf{A}_{(i)} \in \mathbb{R}^{m\times n}$ be a sensing matrix so that each row is sampled uniformly at random from the set $\{\mathbf{u}_k\}_{k=1}^K$. This sampling model also fits into our formulation. Random measurements of this kind arise in Fourier transform with a continuous frequency spectrum ([31]) as well as wavelet frame based reconstruction ([32]).

2.2.2. Related results

We see from Theorem 1 that when the signal matrix $\overline{\mathbf{Y}}$ is sufficiently group sparse and the error matrix $\overline{\mathbf{S}}$ sufficiently sparse, then with high probability we are able to exactly

[2] Our model is developed in $\mathbb{R}$ while this example is over $\mathbb{C}$, all our definitions and results can be generalized to complex numbers with little changes.

recover both of them by solving a convex program. Similar results on structured recovery have been established previously by posing more restrictions on the measurements such as Gaussian ([29]) or orthogonality ([30]). Here, we prove the performance guarantee under a more general sensing model.

Note that the total number of measurements is mL. Thus, Theorem 1 gives the measurement bounds $mL \geq \mathcal{O}(k_T L \log^2(nL))$ and $mL \geq \mathcal{O}(k_\Omega)$. Specifically, by setting $L = 1$, this result meets the measurement bounds in the robust lasso model (for example, [19]), where the number of measurements $m \geq \mathcal{O}(k \log^2 p)$ and $m \geq \mathcal{O}(k_\Omega)$ guarantees the high probability recovery of an p-dimensional k-sparse signal vector and an m-dimensional k_Ω-sparse error vector simultaneously.

Theorem 1 is a result of RIPless analysis, which shares the same limitation as all other RIPless analyses. To be specific, Theorem 1 only holds for arbitrary but *fixed* $\overline{\mathbf{Y}}$ and $\overline{\mathbf{S}}$ (except that the elements of $\overline{\mathbf{Y}}$ and $\overline{\mathbf{S}}$ have uniform random signs by Assumption 4). If we expect to have a uniform recovery guarantee here (namely, considering random sensing matrices as well as signal and error matrices with random supports), then certain stronger assumptions must be made on the sensing matrices such as the RIP condition.

The proof of Theorem 1 is based on the construction of an inexact dual certificate through the golfing scheme. The golfing scheme was first introduced in [24] for low rank matrix recovery. Subsequently, [26] and [27] refined and used the scheme to prove the lasso recovery guarantee. The work [19] generalized it to mix-norm recovery. In this paper, we consider a new mix-norm problem, namely, summation of the $\ell_{2,1}$-norm and the ℓ_1-norm.

2.3. Matrix concentration inequalities

Below we give several probability tools that are useful in the proofs of the paper. We begin with two versions of Bernstein inequalities from [25]. The first one is a matrix Bernstein inequality.

Lemma 1 *(Matrix Bernstein inequality). Consider a finite sequence of independent random matrices $\{\mathbf{M}_{(j)} \in \mathbb{R}^{d\times d}\}$. Assume that every random matrix satisfies $\mathbb{E}\left[\mathbf{M}_{(j)}\right] = 0$ and $\|\mathbf{M}_{(j)}\|_{(2,2)} \leq B$ almost surely. Define*

$$\sigma^2 \triangleq \max\left\{\left\|\sum_j \mathbb{E}\left[\mathbf{M}'_{(j)}\mathbf{M}_{(j)}\right]\right\|_{(2,2)}, \left\|\sum_j \mathbb{E}\left[\mathbf{M}_{(j)}\mathbf{M}'_{(j)}\right]\right\|_{(2,2)}\right\}.$$

Then, for all $t \geq 0$, we have

$$Pr\left\{\left\|\sum_j \mathbf{M}_{(j)}\right\|_{(2,2)} \geq t\right\} \leq 2d \exp\left(-\frac{t^2/2}{\sigma^2 + Bt/3}\right).$$

We also need a vector form of the Bernstein inequality.

Lemma 2 *(Vector Bernstein inequality). Consider a finite sequence of independent random vectors $\{\mathbf{g}_{(j)} \in \mathbb{R}^d\}$. Assume that every random vector satisfies $\mathbb{E}\left[\mathbf{g}_{(j)}\right] = 0$ and $\|\mathbf{g}_{(j)}\|_2 \le B$ almost surely. Define $\sigma^2 \triangleq \sum_k \mathbb{E}\left[\|\mathbf{g}_{(j)}\|_2^2\right]$. Then, for all $0 \le t \le \sigma^2/B$, we have*

$$Pr\left(\left\|\sum_j \mathbf{g}_{(j)}\right\|_2 \ge t\right) \le exp\left(-\frac{t^2}{8\sigma^2} + \frac{1}{4}\right).$$

Next, we use the matrix Bernstein inequality to prove its extension on a block anisotropic matrix.

Lemma 3. *Consider a matrix $\mathbf{A}_{(i)}$ satisfying the model described in Section 2.1, and denote $\tilde{\mathbf{A}}_{(i)} = \mathbf{\Sigma}_{(i)}^{-1}\mathbf{A}'_{(i)}\mathcal{P}_{\Omega_i^*}\mathbf{A}_{(i)}$. For any $\tau > 0$, it holds*

$$Pr\left\{\left\|\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(i)}\mathbf{\Sigma}_{(i)}^{-1} - \mathbf{\Sigma}_{(i)}^{-1}\right)\mathcal{P}_T\right\|_{(2,2)} \ge \tau\right\}$$
$$\le 2k_T \exp\left(-\frac{m-k_{\max}}{\kappa k_T \mu}\frac{\tau^2}{4(\kappa + \frac{2\tau}{3})}\right),$$

and

$$Pr\left\{\left\|\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(i)} - \mathbf{I}\right)\mathcal{P}_T\right\|_{(2,2)} \ge \tau\right\} \le 2k_T \exp\left(-\frac{m-k_{\max}}{\kappa k_T \mu}\frac{\tau^2}{4(1 + \frac{2\tau}{3})}\right).$$

We prove the first part in Appendix A, and the second part can be proved in a similar way. Two consequent corollaries of Lemma 3 show that the restriction of $\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\tilde{\mathbf{A}}_{(1)}, \cdots, \tilde{\mathbf{A}}_{(L)}\right\}$ to the corresponding support T is near isometric.

Corollary 1. *Denote $\tilde{\mathbf{A}}_{(i)} = \mathbf{\Sigma}_{(i)}^{-1}\mathbf{A}'_{(i)}\mathcal{P}_{\Omega_i^*}\mathbf{A}_{(i)}$. Given $k_T \le \alpha\frac{m}{\mu\kappa\log(nL)}$, $k_{\max} \le \gamma m$, and $\frac{1-\gamma}{\alpha} \ge 64$, then with probability at least $1 - 2(nL)^{-2}$, we have*

$$\left\|\boldsymbol{BLKdiag}\left\{\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(1)} - \mathbf{I}\right)\mathcal{P}_T,\ \cdots,\ \mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(L)} - \mathbf{I}\right)\mathcal{P}_T\right\}\right\|_{(2,2)} < \frac{1}{2}. \quad (11)$$

Furthermore, given $k_T \le \alpha\frac{m}{\mu\kappa\log^2(nL)}$, $k_{\max} \le \gamma m$, and $\frac{1-\gamma}{\alpha} \ge 64$, with at least the same probability, we have

$$\left\|\boldsymbol{BLKdiag}\left\{\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(1)}-\mathbf{I}\right)\mathcal{P}_T,\ \cdots,\ \mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(L)}-\mathbf{I}\right)\mathcal{P}_T\right\}\right\|_{(2,2)} < \frac{1}{2\sqrt{\log(nL)}}. \tag{12}$$

Proof. First, following directly from the first part of Lemma 3, for all $i = 1, \cdots, L$, it holds

$$Pr\left\{\left\|\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(i)}-\mathbf{I}\right)\mathcal{P}_T\right\|_{(2,2)} \geq \tau\right\} \leq 2k_T\exp\left\{-\frac{m-k_{\max}}{k_T\mu\kappa}\frac{\tau^2}{4(1+\frac{2\tau}{3})}\right\}. \tag{13}$$

Taking a union bound over all $i = 1, \cdots, L$ yields

$$\begin{aligned}
&Pr\left\{\left\|\mathbf{BLKdiag}\left\{\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(1)}-\mathbf{I}\right)\mathcal{P}_T,\ \cdots,\right.\right.\right.\\
&\qquad\left.\left.\left.\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(L)}-\mathbf{I}\right)\mathcal{P}_T\right\}\right\|_{(2,2)} \geq \tau\right\}\\
&= Pr\left\{\max_i\left\{\left\|\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(i)}-\mathbf{I}\right)\mathcal{P}_T\right\|_{(2,2)}\right\} \geq \tau\right\}\\
&\leq \sum_{i=1}^{L} Pr\left\{\left\|\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(i)}-\mathbf{I}\right)\mathcal{P}_T\right\|_{(2,2)} \geq \tau\right\}\\
&\leq 2k_TL\exp\left\{-\frac{m-k_{\max}}{k_T\mu\kappa}\frac{\tau^2}{4(1+\frac{2\tau}{3})}\right\}.
\end{aligned} \tag{14}$$

Plugging in $\tau = \frac{1}{2}$ and using the fact that $k_T \leq \alpha\frac{m}{\mu\kappa\log(nL)}$ and $k_{\max} \leq \gamma m$, we get

$$\begin{aligned}
\text{The last line of (14)} &= 2k_TL\exp\left\{-\frac{3(1-\gamma)}{64\alpha}\log(nL)\right\}\\
&= 2k_TL(nL)^{-\frac{3(1-\gamma)}{64\alpha}},\\
&\leq 2k_TL(nL)^{-3} \leq 2(nL)^{-2}
\end{aligned}$$

where the first inequality follows from $\frac{1-\gamma}{\alpha} \geq 64$ and the second inequality follows from $k_T \leq n$. Similarly, plugging in $\tau = \frac{1}{2\sqrt{\log(nL)}}$ and using the fact that $k_T \leq \alpha\frac{m}{\mu\kappa\log^2(nL)}$, we prove (12) as long as $\frac{1-\gamma}{\alpha} \geq 64$. □

Corollary 2. *Given that $k_T \leq \alpha\frac{m}{\mu\kappa\log(nL)}$, $k_{\max} \leq \gamma m$, and $\frac{1-\gamma}{\alpha} \geq 64$, then with probability at least $1 - 2(nL)^{-2}$, we have*

$$\left\| \boldsymbol{BLKdiag}\left\{ \mathcal{P}_T \left(\frac{m}{m-k_{\max}} \tilde{\mathbf{A}}_{(1)} \boldsymbol{\Sigma}_{(1)}^{-1} - \boldsymbol{\Sigma}_{(1)}^{-1} \right) \mathcal{P}_T, \right. \right.$$
$$\left. \left. \cdots, \mathcal{P}_T \left(\frac{m}{m-k_{\max}} \tilde{\mathbf{A}}_{(L)} \boldsymbol{\Sigma}_{(L)}^{-1} - \boldsymbol{\Sigma}_{(L)}^{-1} \right) \mathcal{P}_T \right\} \right\|_{(2,2)} < \frac{\kappa}{2}. \tag{15}$$

The proof is the same as proving (11) using Lemma 3. We omit the details for brevity. Finally, we have the following lemma show that if the support of the columns in $\mathbf{A}_{(i)}$ is restricted to Ω_i^*, then no column indexed inside T can be well approximated by the column indexed outside of T. In other words, those columns correspond to the true signal matrix shall be well distinguished.

Lemma 4 *(Off-support incoherence). Denote $\tilde{\mathbf{A}}_{(i)} = \boldsymbol{\Sigma}_{(i)}^{-1} \mathbf{A}'_{(i)} \mathcal{P}_{\Omega_i^*} \mathbf{A}_{(i)}$. Given $k_T \leq \alpha \frac{m}{\mu\kappa \log(nL)}$ and $\alpha < \frac{1}{24}$, with probability at least $1 - e^{\frac{1}{4}}(nL)^{-2}$, we have*

$$\max_{i \in \{1,\cdots,L\}, k \in T^c} \left\| \mathcal{P}_T \tilde{\mathbf{A}}_{(i)} \mathbf{e}_k \right\|_2 \leq 1, \tag{16}$$

where $\{\mathbf{e}_k\}_{k=1}^n$ is a standard basis in $\mathbb{R}^n$.

The proof of Lemma 4 is given in Appendix B.

With particular note, in the above lemmas and corollaries, all the requirements on the constants α, β and γ satisfy the bounds in Theorem 1.

3. Inexact dual certificates

This section gives the dual certificates of the RGL model, namely, the sufficient conditions under which the optimal solution pair of the convex program (2)–(3) is unique and equal to the pair of the true signal and error matrices. First we have two preliminary lemmas.

Lemma 5. *Suppose that $\overline{\mathbf{Y}} \in \mathbb{R}^{n\times L}$ and $\overline{\mathbf{S}} \in \mathbb{R}^{m\times L}$ are the true group sparse signal and sparse error matrices, respectively. If $(\overline{\mathbf{Y}} + \mathbf{H}, \overline{\mathbf{S}} - \mathbf{F})$ is an optimal solution pair to (2)–(3), where $\mathbf{H} \in \mathbb{R}^{n\times L}$ and $\mathbf{F} \in \mathbb{R}^{m\times L}$, then the following results hold:*

i) $\left[\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathbf{A}_{(L)}\mathbf{h}_L\right] = \mathbf{F}$;

ii) $\|\overline{\mathbf{Y}} + \mathbf{H}\|_{2,1} + \lambda\|\overline{\mathbf{S}} - \mathbf{F}\|_1 \geq \|\overline{\mathbf{Y}}\|_{2,1} + \lambda\|\overline{\mathbf{S}}\|_1 + \|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1} + \lambda\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1 + \langle \overline{\mathbf{V}}, \mathbf{H}\rangle - \lambda\langle sgn(\overline{\mathbf{S}}), \mathbf{F}\rangle$,

where $\overline{\mathbf{V}} \in \mathbb{R}^{n\times L}$ satisfies $(\mathcal{P}_T\overline{\mathbf{V}})^i = \frac{\bar{\mathbf{y}}^i}{\|\bar{\mathbf{y}}^i\|_2}$ and $(\mathcal{P}_{T^c}\overline{\mathbf{V}})^i = \mathbf{0}$, $\forall i = 1, \cdots, n$. Here $(\mathcal{P}_T\overline{\mathbf{V}})^i$ denotes the i-th row of $\mathcal{P}_T\overline{\mathbf{V}}$ and $\bar{\mathbf{y}}^i$ denotes the i-th row of $\overline{\mathbf{Y}}$.

The proof of Lemma 5 is given in Appendix C.

Lemma 6. *For any two matrices $\mathbf{H} \in \mathbb{R}^{n\times L}$ and $\mathbf{F} \in \mathbb{R}^{m\times L}$, with probability at least $1 - 2n^{-2}$, $\mathbf{H} = \mathbf{0}$ and $\mathbf{F} = \mathbf{0}$ if the following conditions are satisfied:*

i) $k_T \le \alpha \frac{m}{\mu\kappa \log^2(nL)}$, $k_{\max} \le \gamma m$, *and* $\frac{1-\gamma}{\alpha} \ge 64$;
ii) $\left[\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathbf{A}_{(L)}\mathbf{h}_L\right] = \mathbf{F}$;
iii) $\mathcal{P}_{T^c}\mathbf{H} = \mathbf{0}$ *and* $\mathcal{P}_{\Omega^c}\mathbf{F} = \mathbf{0}$.

The proof of Lemma 6 is given in Appendix D.

Below we show that the optimal solution pair $(\hat{\mathbf{Y}}, \hat{\mathbf{S}})$ of the convex program (2)–(3) is equal to the true signal and noise pair $(\overline{\mathbf{Y}}, \overline{\mathbf{S}})$ when certain certificate conditions hold.

Theorem 2 *(Inexact duality). Suppose that $\overline{\mathbf{Y}} \in \mathbb{R}^{n\times L}$ and $\overline{\mathbf{S}} \in \mathbb{R}^{m\times L}$ are the true group sparse signal and sparse error matrices satisfying the assumptions in Theorem 1. The pair $(\overline{\mathbf{Y}}, \overline{\mathbf{S}})$ is the unique solution to the RGL model (2)–(3) with probability at least $1 - (2 + e^{\frac{1}{4}})(nL)^{-1}$, if the parameter $\lambda < 1$ and there exists a dual certificate $(\mathbf{W}, \mathbf{V}) \in \mathbb{R}^{m\times L} \times \mathbb{R}^{n\times L}$ such that*

$$\|\mathcal{P}_T\mathbf{V} - \overline{\mathbf{V}}\|_F \le \frac{\lambda}{4\sqrt{\kappa}}, \tag{17}$$

$$\|\mathcal{P}_{T^c}\mathbf{V}\|_{2,\infty} \le \frac{1}{4}, \tag{18}$$

$$\|\mathcal{P}_{\Omega^c}\mathbf{W}\|_\infty \le \frac{\lambda}{4}, \tag{19}$$

and

$$\mathbf{V} = \left[\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^c}\mathbf{w}_1, \cdots, \mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^c}\mathbf{w}_L\right] + \lambda\left[\mathbf{A}'_{(1)} sgn(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)} sgn(\bar{\mathbf{s}}_L)\right], \tag{20}$$

where $\overline{\mathbf{V}} \in \mathbb{R}^{n\times L}$ satisfies $(\mathcal{P}_T\overline{\mathbf{V}})^i = \frac{\bar{\mathbf{y}}^i}{\|\bar{\mathbf{y}}^i\|_2}$ and $(\mathcal{P}_{T^c}\overline{\mathbf{V}})^i = \mathbf{0}$.

Proof. Suppose that $(\overline{\mathbf{Y}}+\mathbf{H}, \overline{\mathbf{S}}-\mathbf{F})$ is an optimal solution pair to (2)–(3). Therefore, the two results in Lemma 5 hold true. Proving that the pair $(\overline{\mathbf{Y}}, \overline{\mathbf{S}})$ is the unique solution to (2)–(3) is equivalent to showing that $\mathbf{H} = \mathbf{0}$ and $\mathbf{F} = \mathbf{0}$. Hence, the proof resorts to verifying the three conditions in Lemma 6.

Since $\overline{\mathbf{Y}}$ and $\overline{\mathbf{S}}$ satisfy the assumptions in Theorem 1, we have

$$k_T \le \alpha \frac{m}{\mu\kappa \log^2(nL)}, \quad k_{\max} \le \gamma \frac{m}{\kappa}, \quad \alpha \le \frac{1}{9600}, \quad \gamma \le \frac{1}{4}.$$

Considering $\kappa \ge 1$, we know that condition i) in Lemma 6 holds. By result i) of Lemma 5, condition ii) also holds. Therefore, it remains to verify condition iii), namely, $\mathcal{P}_{T^c}\mathbf{H} = \mathbf{0}$ and $\mathcal{P}_{\Omega^c}\mathbf{F} = \mathbf{0}$.

Consider the term $\langle \overline{\mathbf{V}}, \mathbf{H}\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle$ in result ii) of Lemma 5. Using the equation $\mathbf{V} = \mathcal{P}_T\mathbf{V} + \mathcal{P}_{T^c}\mathbf{V}$, we rewrite the term as

$$\langle \overline{\mathbf{V}}, \mathbf{H}\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle = \langle \overline{\mathbf{V}} - \mathcal{P}_T\mathbf{V}, \mathbf{H}\rangle - \langle \mathcal{P}_{T^c}\mathbf{V}, \mathbf{H}\rangle + \langle \mathbf{V}, \mathbf{H}\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle. \quad (21)$$

Consider the term $\langle \mathbf{V}, \mathbf{H}\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle$ on the right hand side of (21). By (20), we have

$$\begin{aligned} &\langle \mathbf{V}, \mathbf{H}\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle \\ = &\left\langle \left[\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^c}\mathbf{w}_1, \cdots, \mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^c}\mathbf{w}_L\right], \mathbf{H}\right\rangle \\ &+ \lambda\left\langle \left[\mathbf{A}'_{(1)}\mathrm{sgn}(\overline{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}\mathrm{sgn}(\overline{\mathbf{s}}_L)\right], \mathbf{H}\right\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle. \end{aligned}$$

By adjoint relations of inner products, we have

$$\begin{aligned} &\left\langle \left[\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^c}\mathbf{w}_1, \cdots, \mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^c}\mathbf{w}_L\right], \mathbf{H}\right\rangle \\ = &\left\langle \left[\mathcal{P}_{\Omega_1^c}\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathcal{P}_{\Omega_L^c}\mathbf{A}_{(L)}\mathbf{h}_L\right], \mathbf{W}\right\rangle, \end{aligned}$$

and

$$\begin{aligned} &\langle \left[\mathbf{A}'_{(1)}\mathrm{sgn}(\overline{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}\mathrm{sgn}(\overline{\mathbf{s}}_L)\right], \mathbf{H}\rangle \\ = &\langle \mathrm{sgn}(\overline{\mathbf{S}}), \left[\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathbf{A}_{(L)}\mathbf{h}_L\right]\rangle. \end{aligned}$$

Thus, it holds

$$\begin{aligned} &\langle \mathbf{V}, \mathbf{H}\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle \\ = &\left\langle \left[\mathcal{P}_{\Omega_1^c}\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathcal{P}_{\Omega_L^c}\mathbf{A}_{(L)}\mathbf{h}_L\right], \mathbf{W}\right\rangle \\ &+ \lambda\left\langle \mathrm{sgn}(\overline{\mathbf{S}}), \left[\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathbf{A}_{(L)}\mathbf{h}_L\right]\right\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle. \end{aligned}$$

According to conclusion i) in Lemma 5, which is $\left[\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathbf{A}_{(L)}\mathbf{h}_L\right] = \mathbf{F}$, it follows

$$\langle \mathbf{V}, \mathbf{H}\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}, \mathbf{F})\rangle = \langle \mathcal{P}_{\Omega^c}\mathbf{F}, \mathbf{W}\rangle.$$

Combining (21) and above equality gives

$$\langle \overline{\mathbf{V}}, \mathbf{H}\rangle - \lambda\langle \mathrm{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle = \langle \overline{\mathbf{V}} - \mathcal{P}_T\mathbf{V}, \mathbf{H}\rangle - \langle \mathcal{P}_{T^c}\mathbf{V}, \mathbf{H}\rangle + \langle \mathcal{P}_{\Omega^c}\mathbf{F}, \mathbf{W}\rangle. \quad (22)$$

Next, we manage to find out a lower bound for the right-hand side of the equality (22). First, by (17),

$$\langle \overline{\mathbf{V}} - \mathcal{P}_T\mathbf{V}, \mathbf{H}\rangle \geq -\|\mathcal{P}_T\mathbf{V} - \overline{\mathbf{V}}\|_F\|\mathcal{P}_T\mathbf{H}\|_F \geq -\frac{\lambda}{4\sqrt{\kappa}}\|\mathcal{P}_T\mathbf{H}\|_F.$$

Then, by (18),

$$-\langle \mathcal{P}_{T^c}\mathbf{V}, \mathbf{H}\rangle \geq -\|\mathcal{P}_{T^c}\mathbf{V}\|_{2,\infty}\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1} \geq -\frac{1}{4}\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1}.$$

Finally, by (19),

$$\langle \mathcal{P}_{\Omega^c}\mathbf{F}, \mathbf{W}\rangle \geq -\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1\|\mathcal{P}_{\Omega^c}\mathbf{W}\|_\infty \geq -\frac{\lambda}{4}\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1.$$

Therefore, (22) gives

$$\langle \overline{\mathbf{V}}, \mathbf{H}\rangle - \lambda\langle \operatorname{sgn}(\overline{\mathbf{S}}), \mathbf{F}\rangle \geq -\frac{\lambda}{4\sqrt{\kappa}}\|\mathcal{P}_T\mathbf{H}\|_F - \frac{1}{4}\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1} - \frac{\lambda}{4}\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1.$$

Substitute the above inequality into conclusion ii) of Lemma 5 gives

$$\begin{aligned}&\|\overline{\mathbf{Y}} + \mathbf{H}\|_{2,1} + \lambda\|\overline{\mathbf{S}} - \mathbf{F}\|_1\\ &\geq \|\overline{\mathbf{Y}}\|_{2,1} + \lambda\|\overline{\mathbf{S}}\|_1 + \frac{3}{4}\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1} + \frac{3\lambda}{4}\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1 - \frac{\lambda}{4\sqrt{\kappa}}\|\mathcal{P}_T\mathbf{H}\|_F.\end{aligned}$$

Since $(\overline{\mathbf{Y}} + \mathbf{H}, \overline{\mathbf{S}} - \mathbf{F})$ is an optimal solution pair to (2)–(3), it follows that $\|\overline{\mathbf{Y}} + \mathbf{H}\|_{2,1} + \lambda\|\overline{\mathbf{S}} - \mathbf{F}\|_1 \leq \|\overline{\mathbf{Y}}\|_{2,1} + \lambda\|\overline{\mathbf{S}}\|_1$. Hence, we have

$$\frac{3}{4}\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1} + \frac{3\lambda}{4}\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1 - \frac{\lambda}{4\sqrt{\kappa}}\|\mathcal{P}_T\mathbf{H}\|_F \leq 0. \tag{23}$$

To complete the proof, we need to show that inequality (23) implies $\mathcal{P}_{T^c}\mathbf{H} = \mathbf{0}$ and $\mathcal{P}_{\Omega^c}\mathbf{F} = \mathbf{0}$. To do so, we first derive an upper bound for $\|\mathcal{P}_T\mathbf{H}\|_F$, expressed as a linear combination of $\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1}$ and $\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1$.

Using (11), it follows

$$\begin{aligned}&\left\|\mathbf{BLKdiag}\left\{\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(1)} - \mathbf{I}\right)\mathcal{P}_T, \cdots, \mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(L)} - \mathbf{I}\right)\mathcal{P}_T\right\}\operatorname{vec}(\mathbf{H})\right\|_2\\ &\qquad \leq \frac{1}{2}\|\mathcal{P}_T\mathbf{H}\|_F.\end{aligned}$$

Since $\|\mathbf{BLKdiag}\{\mathcal{P}_T, \cdots, \mathcal{P}_T\}\operatorname{vec}(\mathbf{H})\|_2 = \|\mathcal{P}_T\mathbf{H}\|_F$, applying the triangle inequality yields

$$\|\mathcal{P}_T\mathbf{H}\|_F \leq 2\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\tilde{\mathbf{A}}_{(1)}\mathcal{P}_T, \cdots, \mathcal{P}_T\tilde{\mathbf{A}}_{(L)}\mathcal{P}_T\right\}\operatorname{vec}(\mathbf{H})\right\|_2.$$

Observing $\operatorname{vec}(\mathcal{P}_T\mathbf{H}) = \operatorname{vec}(\mathbf{H}) - \operatorname{vec}(\mathcal{P}_{T^c}\mathbf{H})$ and using the triangle inequality again, we have

$$\begin{aligned}\|\mathcal{P}_T\mathbf{H}\|_F \leq & 2\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\tilde{\mathbf{A}}_{(1)},\cdots,\mathcal{P}_T\tilde{\mathbf{A}}_{(L)}\right\}\mathrm{vec}(\mathbf{H})\right\|_2 \\ & + 2\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\tilde{\mathbf{A}}_{(1)},\cdots,\mathcal{P}_T\tilde{\mathbf{A}}_{(L)}\right\}\mathrm{vec}(\mathcal{P}_{T^c}\mathbf{H})\right\|_2. \qquad (24)\end{aligned}$$

Below, we upper bound the two terms at the right-hand side of (24), respectively.

① **Bounding the first term of (24):** By definitions $\tilde{\mathbf{A}}_{(i)} = \mathbf{\Sigma}_{(i)}^{-1}\mathbf{A}'_{(i)}\mathcal{P}_{\Omega_i^*}\mathbf{A}_{(i)}$, it follows

$$\begin{aligned}&\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\tilde{\mathbf{A}}_{(1)},\cdots,\mathcal{P}_T\tilde{\mathbf{A}}_{(L)}\right\}\mathrm{vec}(\mathbf{H})\right\|_2 \\ =&\left\|\frac{m}{m-k_{\max}}\mathrm{vec}\left(\left[\mathcal{P}_T\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^*}\mathbf{A}_{(1)}\mathbf{h}_1,\cdots,\mathcal{P}_T\mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^*}\mathbf{A}_{(L)}\mathbf{h}_L\right]\right)\right\|_2 \\ =&\left\|\frac{m}{m-k_{\max}}\mathrm{vec}\left(\left[\mathcal{P}_T\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^*}\mathbf{f}_1,\cdots,\mathcal{P}_T\mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^*}\mathbf{f}_L\right]\right)\right\|_2. \qquad (25)\end{aligned}$$

Here the second equality comes from result i) in Lemma 5, namely, $\mathbf{F} = [\mathbf{A}_{(1)}\mathbf{h}_1,\cdots,\mathbf{A}_{(L)}\mathbf{h}_L]$. Recalling that Ω_i^* is a subset of Ω_i^c for any $i=1,\cdots,L$, we have

$$\left\|\mathrm{vec}\left(\left[\mathcal{P}_{\Omega_1^*}\mathbf{f}_1,\cdots,\mathcal{P}_{\Omega_L^*}\mathbf{f}_L\right]\right)\right\|_2 \leq \|\mathrm{vec}(\mathcal{P}_{\Omega^c}\mathbf{F})\|_2.$$

Based on this inequality, we upper bound (25) using the induced norm property

$$\begin{aligned}&\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\tilde{\mathbf{A}}_{(1)},\cdots,\mathcal{P}_T\tilde{\mathbf{A}}_{(L)}\right\}\mathrm{vec}(\mathbf{H})\right\|_2 \\ \leq&\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^*},\cdots,\mathcal{P}_T\mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^*}\right\}\right\|_{(2,2)} \\ &\cdot\left\|\mathrm{vec}\left(\left[\mathcal{P}_{\Omega_1^*}\mathbf{f}_1,\cdots,\mathcal{P}_{\Omega_L^*}\mathbf{f}_L\right]\right)\right\|_2 \\ \leq&\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^*},\cdots,\mathcal{P}_T\mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^*}\right\}\right\|_{(2,2)}\cdot\|\mathrm{vec}(\mathcal{P}_{\Omega^c}\mathbf{F})\|_2 \\ \leq&\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^*},\cdots,\mathcal{P}_T\mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^*}\right\}\right\|_{(2,2)}\cdot\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1. \qquad (26)\end{aligned}$$

Using the definitions $\tilde{\mathbf{A}}_{(i)} = \mathbf{\Sigma}_{(i)}^{-1}\mathbf{A}'_{(i)}\mathcal{P}_{\Omega_i^*}\mathbf{A}_{(i)}$ and applying the triangle inequality as well as Corollary 2, with probability at least $1-2(nL)^{-2}$ it holds

$$\begin{aligned}&\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^*}\mathbf{A}_{(1)}\mathbf{\Sigma}_{(1)}^{-1}\mathcal{P}_T,\cdots,\right.\right. \\ &\left.\left.\mathcal{P}_T\mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^*}\mathbf{A}_{(L)}\mathbf{\Sigma}_{(L)}^{-1}\mathcal{P}_T\right\}\right\|_{(2,2)} \\ =&\left\|\frac{m}{m-k_{\max}}\mathbf{BLKdiag}\left\{\mathcal{P}_T\tilde{\mathbf{A}}_{(1)}\mathbf{\Sigma}_{(1)}^{-1}\mathcal{P}_T,\cdots,\mathcal{P}_T\tilde{\mathbf{A}}_{(L)}\mathbf{\Sigma}_{(L)}^{-1}\mathcal{P}_T\right\}\right\|_{(2,2)}\end{aligned}$$

$$
\begin{aligned}
&\leq \left\| \mathbf{BLKdiag}\left\{ \mathcal{P}_T \left(\frac{m}{m-k_{\max}} \tilde{\mathbf{A}}_{(1)} \boldsymbol{\Sigma}_{(1)}^{-1} - \boldsymbol{\Sigma}_{(1)}^{-1} \right) \mathcal{P}_T, \ \cdots, \right.\right. \\
&\qquad \left.\left. \mathcal{P}_T \left(\frac{m}{m-k_{\max}} \tilde{\mathbf{A}}_{(L)} \boldsymbol{\Sigma}_{(L)}^{-1} - \boldsymbol{\Sigma}_{(L)}^{-1} \right) \mathcal{P}_T \right\} \right\|_{(2,2)} \\
&\quad + \left\| \mathbf{BLKdiag}\left\{ \mathcal{P}_T \boldsymbol{\Sigma}_{(1)}^{-1} \mathcal{P}_T, \cdots, \mathcal{P}_T \boldsymbol{\Sigma}_{(L)}^{-1} \mathcal{P}_T \right\} \right\|_{(2,2)} \\
&\leq \frac{\kappa}{2} + \left\| \mathbf{BLKdiag}\left\{ \mathcal{P}_T \boldsymbol{\Sigma}_{(1)}^{-1} \mathcal{P}_T, \cdots, \mathcal{P}_T \boldsymbol{\Sigma}_{(L)}^{-1} \mathcal{P}_T \right\} \right\|_{(2,2)} \leq \frac{3}{2}\kappa.
\end{aligned}
$$

Consequently,

$$
\left\| \sqrt{\frac{m}{m-k_{\max}}} \mathbf{BLKdiag}\left\{ \mathcal{P}_T \boldsymbol{\Sigma}_{(1)}^{-1} \mathbf{A}'_{(1)} \mathcal{P}_{\Omega_1^*}, \ \cdots, \ \mathcal{P}_T \boldsymbol{\Sigma}_{(L)}^{-1} \mathbf{A}'_{(L)} \mathcal{P}_{\Omega_L^*} \right\} \right\|_{(2,2)} \leq \sqrt{\frac{3}{2}\kappa}.
$$

Combining (26), this gives

$$
\left\| \frac{m}{m-k_{\max}} \mathbf{BLKdiag}\left\{ \mathcal{P}_T \tilde{\mathbf{A}}_{(1)}, \cdots, \mathcal{P}_T \tilde{\mathbf{A}}_{(L)} \right\} \operatorname{vec}(\mathbf{H}) \right\|_2 \leq \sqrt{\frac{3}{2} \frac{\kappa m}{m-k_{\max}}} \left\| \mathcal{P}_{\Omega^c} \mathbf{F} \right\|_1 .
$$

② **Bounding the second term of (24):** The following chains of equalities and inequalities hold:

$$
\begin{aligned}
&\left\| \mathbf{BLKdiag}\left\{ \mathcal{P}_T \tilde{\mathbf{A}}_{(1)}, \cdots, \mathcal{P}_T \tilde{\mathbf{A}}_{(L)} \right\} \operatorname{vec}(\mathcal{P}_{T^c}\mathbf{H}) \right\|_2 \\
&= \left\| \left[\mathcal{P}_T \tilde{\mathbf{A}}_{(1)} \mathcal{P}_{T^c} \mathbf{h}_1, \cdots, \mathcal{P}_T \tilde{\mathbf{A}}_{(L)} \mathcal{P}_{T^c} \mathbf{h}_L \right] \right\|_F \\
&= \left\| \sum_{k \in T^c} \left[h_{1k} \mathcal{P}_T \tilde{\mathbf{A}}_{(1)} \mathbf{e}_k, \cdots, h_{Lk} \mathcal{P}_T \tilde{\mathbf{A}}_{(L)} \mathbf{e}_k \right] \right\|_F \\
&\leq \sum_{k \in T^c} \left\| \left[h_{1k} \mathcal{P}_T \tilde{\mathbf{A}}_{(1)} \mathbf{e}_k, \cdots, h_{Lk} \mathcal{P}_T \tilde{\mathbf{A}}_{(L)} \mathbf{e}_k \right] \right\|_F \\
&= \sum_{k \in T^c} \sqrt{ \sum_{i=1}^{L} \left\| \mathcal{P}_T \tilde{\mathbf{A}}_{(i)} \mathbf{e}_k \right\|_2^2 \cdot |h_{ik}|^2 } \\
&\leq \sum_{k \in T^c} \left(\max_{i \in \{1,\cdots,L\}} \left\{ \left\| \mathcal{P}_T \tilde{\mathbf{A}}_{(i)} \mathbf{e}_k \right\|_2 \right\} \right) \cdot \left\| \mathbf{h}^k \right\|_2 \\
&\leq \left(\max_{i \in \{1,\cdots,L\}, \ k \in T^c} \left\{ \left\| \mathcal{P}_T \tilde{\mathbf{A}}_{(i)} \mathbf{e}_k \right\|_2 \right\} \right) \cdot \left\| \mathcal{P}_{T^c} \mathbf{H} \right\|_{2,1} , \qquad (27)
\end{aligned}
$$

where $\mathbf{h}^k$ denotes the k-th row of matrix $\mathbf{H}$ and h_{ik} denotes the (i,k)-th element of $\mathbf{H}$. In (27), the last inequality follows from the definition of the $\ell_{2,1}$-norm. According to Lemma 4, with probability at least $1 - e^{\frac{1}{4}}(nL)^{-2}$, (27) implies

$$
\left\| \mathbf{BLKdiag}\left\{ \mathcal{P}_T \tilde{\mathbf{A}}_{(1)}, \cdots, \mathcal{P}_T \tilde{\mathbf{A}}_{(L)} \right\} \operatorname{vec}(\mathcal{P}_{T^c}\mathbf{H}) \right\|_2 \leq \left\| \mathcal{P}_{T^c} \mathbf{H} \right\|_{2,1}.
$$

Summarizing the results above, we have an upper bound for the right-hand side of (24):

$$\|\mathcal{P}_T\mathbf{H}\|_F \le \sqrt{\frac{6\kappa m}{m-k_{\max}}}\,\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1 + \frac{2m}{m-k_{\max}}\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1}. \tag{28}$$

Finally, substituting (28) into (23) gives

$$\left(\frac{3}{4} - \frac{1}{2\sqrt{\kappa}}\frac{m}{m-k_{\max}}\lambda\right)\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1} + \left(\frac{3}{4} - \frac{\sqrt{6}}{4}\sqrt{\frac{m}{m-k_{\max}}}\right)\lambda\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1 \le 0.$$

In the above inequality, $\frac{3}{4} - \frac{\sqrt{6}}{4}\sqrt{\frac{m}{m-k_{\max}}}$ and $\frac{3}{4} - \frac{\lambda}{2\sqrt{\kappa}}\frac{m}{m-k_{\max}}$ are both larger than zero provided that $\lambda < 1$ and $\frac{k_{\max}}{m} \le \frac{\gamma}{\kappa} < \frac{1}{3}$. Thus, we have $\|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1} = 0$ and $\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1 = 0$, which prove $\mathcal{P}_{T^c}\mathbf{H} = \mathbf{0}$ and $\mathcal{P}_{\Omega^c}\mathbf{F} = \mathbf{0}$. □

4. Construction of dual certificate

By explicitly constructing a pair of dual certificate, this section proves the following theorem, which is sufficient for proving our main result given by Theorem 1.

Theorem 3. *Under the assumptions in Theorem 1, with high probability, there exists a pair of dual certificate* $(\mathbf{U}, \mathbf{W})$ *such that*

$$\mathbf{U} = \left[\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^c}\mathbf{w}_1, \cdots, \mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^c}\mathbf{w}_L\right],$$

and

$$\left\|\lambda\mathcal{P}_{T^c}\left[\mathbf{A}'_{(1)}sgn(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}sgn(\bar{\mathbf{s}}_L)\right]\right\|_{2,\infty} \le \frac{1}{8}, \tag{29}$$

$$\left\|\mathcal{P}_T\mathbf{U} + \lambda\mathcal{P}_T\left[\mathbf{A}'_{(1)}sgn(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}sgn(\bar{\mathbf{s}}_L)\right] - \overline{\mathbf{V}}\right\|_F \le \frac{\lambda}{4\sqrt{\kappa}}, \tag{30}$$

$$\|\mathcal{P}_{T^c}\mathbf{U}\|_{2,\infty} \le \frac{1}{8}, \tag{31}$$

$$\|\mathcal{P}_{\Omega^c}\mathbf{W}\|_\infty \le \frac{\lambda}{4}, \tag{32}$$

where $\overline{\mathbf{V}} \in \mathbb{R}^{n\times L}$ *satisfies* $(\mathcal{P}_T\overline{\mathbf{V}})^i = \frac{\bar{\mathbf{y}}^i}{\|\bar{\mathbf{y}}^i\|_2}$ *and* $(\mathcal{P}_{T^c}\overline{\mathbf{V}})^i = \mathbf{0}$, $\forall i = 1, \cdots, n$.

Comparing to Theorem 2, the above theorem breaks $\|\mathcal{P}_{T^c}\mathbf{V}\|_{2,\infty} \le \frac{1}{4}$ in (18) into two constraints (29) and (31). Thus, Theorem 3 implies that an inexact dual certificate exists with high probability. Therefore, Theorem 3 implies our main result Theorem 1.

We start with the procedure of constructing $\mathbf{U}$ and $\mathbf{W}$. This procedure stems from the classical golfing scheme (see [19], [21], and [26]). Basically, it constructs a sequence of matrices $\{\mathbf{Q}_{(j)}\}_{j=0}^{l}$ via l sampled batches of row vectors in each $\mathcal{P}_{\Omega_i^*}\mathbf{A}_{(i)}$, $i \in \{1, \cdots, L\}$, so that different batches are not overlapped and the sequence $\left\{\|\mathbf{Q}_{(j)}\|_F\right\}_{j=0}^{l}$ shrinks exponentially fast in finite steps with high probability. We then write $\mathbf{W}$ and subsequently $\mathbf{U}$ as functions of $\{\mathbf{Q}_{(j)}\}_{j=0}^{l}$ so that they meet the constraints (29)–(32).

Define the initial value of the sequence $\{\mathbf{Q}_{(j)}\}_{j=0}^{l}$ as

$$\mathbf{Q}_{(0)} = \overline{\mathbf{V}} - \lambda\mathcal{P}_T[\mathbf{A}'_{(1)}\mathrm{sgn}(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}\mathrm{sgn}(\bar{\mathbf{s}}_L)]. \tag{33}$$

For each $i = 1, \cdots, L$, we split the maximal non-corrupted set Ω_i^* into l disjoint batch sets, namely, $\Omega_i^* \supseteq K_{i1} \bigcup \cdots \bigcup K_{il}$, so that for any $j = 1, \cdots, l$, the cardinalities of the sets $|K_{ij}|$ satisfy $|K_{1j}| = \cdots = |K_{Lj}| \triangleq m_j$. Notice that it is possible to split Ω_i^* in this way since we enforce $|\Omega_1^*| = \cdots = |\Omega_L^*| = m - k_{\max}$.

Define $\mathcal{P}_{K_{ij}}$ as the orthogonal projection of a vector in $\mathbb{R}^m$ onto coordinates in K_{ij} and define $\tilde{\mathbf{A}}_{(i,j)} = \mathbf{\Sigma}_{(i)}^{-1}\mathbf{A}'_{(i)}\mathcal{P}_{K_{ij}}\mathbf{A}_{(i)}$ as the total number of batches $l \triangleq \lfloor \log(nL) + 1 \rfloor$. For each $j = 1, \cdots, l$, recursively define

$$\begin{aligned}\mathbf{Q}_{(j)} &= \left[\mathcal{P}_T\left(\mathbf{I} - \frac{m}{m_j}\tilde{\mathbf{A}}_{(1,j)}\right)\mathcal{P}_T\mathbf{q}_{(j-1)1}, \cdots, \mathcal{P}_T\left(\mathbf{I} - \frac{m}{m_j}\tilde{\mathbf{A}}_{(L,j)}\right)\mathcal{P}_T\mathbf{q}_{(j-1)L}\right] \\ &= \left[\left(\prod_{r=1}^{j}\mathcal{P}_T\left(\mathbf{I} - \frac{m}{m_r}\tilde{\mathbf{A}}_{(1,r)}\right)\mathcal{P}_T\right)\mathbf{q}_{(0)1}, \cdots, \left(\prod_{r=1}^{j}\mathcal{P}_T\left(\mathbf{I} - \frac{m}{m_r}\tilde{\mathbf{A}}_{(L,r)}\right)\mathcal{P}_T\right)\mathbf{q}_{(0)L}\right].\end{aligned} \tag{34}$$

We choose

$$m_1 = m_2 = \left\lceil \frac{m}{4} \right\rceil, \quad m_j = \left\lceil \frac{m}{4\log(nL)} \right\rceil, \forall j \geq 3.$$

Finally, we set $\mathbf{W}$ so that

$$\mathcal{P}_{\Omega^c}\mathbf{W} = \left[\sum_{j=1}^{l}\frac{m}{m_j}\mathcal{P}_{K_{1j}}\mathbf{A}_{(1)}\mathcal{P}_T\mathbf{q}_{(j-1)1}, \cdots, \sum_{j=1}^{l}\frac{m}{m_j}\mathcal{P}_{K_{Lj}}\mathbf{A}_{(L)}\mathcal{P}_T\mathbf{q}_{(j-1)L}\right], \tag{35}$$

and $\mathcal{P}_\Omega\mathbf{W} = \mathbf{0}$. Also, set $\mathbf{U}$ to be

$$\begin{aligned}\mathbf{U} &= \left[\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^c}\mathbf{w}_1, \cdots, \mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^c}\mathbf{w}_L\right] \\ &= \left[\sum_{j=1}^{l}\frac{m}{m_j}\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{K_{1j}}\mathbf{A}_{(1)}\mathcal{P}_T\mathbf{q}_{(j-1)1}, \cdots, \sum_{j=1}^{l}\frac{m}{m_j}\mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{K_{Lj}}\mathbf{A}_{(L)}\mathcal{P}_T\mathbf{q}_{(j-1)L}\right] \\ &= \sum_{j=1}^{l}\frac{m}{m_j}\left[\tilde{\mathbf{A}}_{(1,j)}\mathcal{P}_T\mathbf{q}_{(j-1)1}, \cdots, \tilde{\mathbf{A}}_{(L,j)}\mathcal{P}_T\mathbf{q}_{(j-1)L}\right].\end{aligned} \tag{36}$$

Our construction is justified by the following lemma, which shows the sequence $\{\|\mathbf{Q}_{(j)}\|_F\}_{j=0}^{l}$ shrinks exponentially fast with high probability.

Lemma 7. *Given $k_T \leq \alpha \frac{m}{\mu\kappa \log^2(nL)}$ with $\alpha \leq \frac{1}{256}$, then, with probability at least $1 - 2(nL)^{-1}$, the following set of inequalities hold simultaneously*

$$\left\| \textbf{\textit{BLKdiag}} \left\{ \mathcal{P}_T \left(\frac{m}{m_j} \tilde{\mathbf{A}}_{(1,j)} - \mathbf{I} \right) \mathcal{P}_T, \ \cdots, \ \mathcal{P}_T \left(\frac{m}{m_j} \tilde{\mathbf{A}}_{(L,j)} - \mathbf{I} \right) \mathcal{P}_T \right\} \right\|_{(2,2)} \leq c_j, \quad (37)$$

where $c_1 = c_2 = \frac{1}{2\sqrt{\log(nL)}}$ and $c_j = \frac{1}{2}$, $j \geq 3$.

This lemma is proved in Appendix E. From Lemma 7, the following chains of contractions hold with probability at least $1 - 2(nL)^{-1}$:

$$\|\mathbf{Q}_{(1)}\|_F \leq \frac{1}{2\sqrt{\log(nL)}} \|\mathbf{Q}_{(0)}\|_F, \quad (38)$$

$$\|\mathbf{Q}_{(2)}\|_F \leq \frac{1}{4\log(nL)} \|\mathbf{Q}_{(0)}\|_F,$$

$$\vdots$$

$$\|\mathbf{Q}_{(l)}\|_F \leq \prod_{j=1}^{l} c_j \|\mathbf{Q}_{(0)}\|_F \leq \frac{1}{\log(nL)} \frac{1}{2^l} \|\mathbf{Q}_{(0)}\|_F. \quad (39)$$

The next key observation behind this construction is that $\mathcal{P}_T\mathbf{U} = \mathbf{Q}_{(0)} - \mathbf{Q}_{(l)}$, and thus by definition of $Q_{(0)}$ in (33), the in-support difference

$$\left\| \mathcal{P}_T \mathbf{U} + \lambda \mathcal{P}_T \left[\mathbf{A}'_{(1)} \mathrm{sgn}(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)} \mathrm{sgn}(\bar{\mathbf{s}}_L) \right] - \overline{\mathbf{V}} \right\|_F = \|\mathbf{Q}_{(l)}\|_F,$$

which is exponentially small, while the size of the off-support terms $\|\mathcal{P}_{T^c}\mathbf{U}\|_{2,\infty}$ and $\|\mathcal{P}_{\Omega^c}\mathbf{W}\|_\infty$ are roughly sum of geometric sequence and thus bounded. The details of the proof of Theorem 3 is given in Appendix F.

5. Conclusion

This paper proposes the robust group lasso (RGL) model that recovers a group sparse signal matrix for sparsely corrupted measurements. The RGL model minimizes the mixed $\ell_{2,1}/\ell_1$-norm under linear measurement constraints, and hence is convex. We establish the recoverability of the RGL model, showing that the true group sparse signal matrix and the sparse error matrix can be exactly recovered with high probability under certain conditions. Our theoretical analysis provides a solid performance guarantee to the RGL model.

Appendix A. Proof of Lemma 3

Proof. Here we prove the second part of Lemma 3. By definitions $\tilde{\mathbf{A}}_{(i)} = \mathbf{\Sigma}_{(i)}^{-1}\mathbf{A}'_{(i)}\mathcal{P}_{\Omega_i^*} \times \mathbf{A}_{(i)}$, it holds

$$\begin{aligned}&\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(i)}\mathbf{\Sigma}_{(i)}^{-1} - \mathbf{\Sigma}_{(i)}^{-1}\right)\mathcal{P}_T\\ =&\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\mathbf{\Sigma}_{(i)}^{-1}\mathbf{A}'_{(i)}\mathcal{P}_{\Omega_i^*}\mathbf{A}_{(i)}\mathbf{\Sigma}_{(i)}^{-1} - \mathbf{\Sigma}_{(i)}^{-1}\right)\mathcal{P}_T\\ =&\sum_{j\in\Omega_i^*}\mathbf{M}_{(j)},\end{aligned}$$

where

$$\mathbf{M}_{(j)} \triangleq \frac{1}{m-k_{\max}}\mathcal{P}_T\left(\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\mathbf{a}'_{(i)j}\mathbf{\Sigma}_{(i)}^{-1} - \mathbf{\Sigma}_{(i)}^{-1}\right)\mathcal{P}_T.$$

Since $\mathbb{E}\left[\mathbf{a}_{(i)j}\mathbf{a}'_{(i)j}\right] = \mathbf{\Sigma}_{(i)}$, it is obvious that $\mathbb{E}\left[\mathbf{M}_{(j)}\right] = 0$. We estimate the induced $\ell_{(2,2)}$-norm of $\mathbf{M}_{(j)}$ in order to implement the matrix Bernstein inequality later. It holds

$$\begin{aligned}\|\mathbf{M}_{(j)}\|_{(2,2)} &= \left\|\frac{1}{m-k_{\max}}\mathcal{P}_T\left(\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\mathbf{a}'_{(i)j}\mathbf{\Sigma}_{(i)}^{-1} - \mathbf{\Sigma}_{(i)}^{-1}\right)\mathcal{P}_T\right\|_{(2,2)}\\ &\le \frac{1}{m-k_{\max}}\left(\left\|\mathcal{P}_T\left(\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\mathbf{a}'_{(i)j}\mathbf{\Sigma}_{(i)}^{-1}\right)\mathcal{P}_T\right\|_{(2,2)} + \left\|\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\mathcal{P}_T\right\|_{(2,2)}\right)\\ &\le \frac{1}{m-k_{\max}}\left(\left\|\mathcal{P}_T\left(\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\mathbf{a}'_{(i)j}\mathbf{\Sigma}_{(i)}^{-1}\right)\mathcal{P}_T\right\|_{(2,2)} + \kappa\right)\\ &= \frac{1}{m-k_{\max}}\left(\left\|\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\right\|_2^2 + \kappa\right) \le \frac{1}{m-k_{\max}}(\mu k_T + \kappa),\end{aligned}$$

where the first inequality follows from the triangle inequality and the last inequality follows from Assumption (8). Since $\kappa \ge 1$ and $\mu \ge 1$, the above bound on $\|\mathbf{M}_{(j)}\|_{(2,2)}$ can be further relaxed as

$$\|\mathbf{M}_{(j)}\|_{(2,2)} \le \frac{2\kappa\mu k_T}{m-k_{\max}} \triangleq B.$$

Meanwhile, since $\mathbf{M}'_{(j)}\mathbf{M}_{(j)} = \mathbf{M}_{(j)}\mathbf{M}'_{(j)}$, we only need to consider one of them.

$$\begin{aligned}&\left\|\mathbb{E}\left[\mathbf{M}'_{(j)}\mathbf{M}_{(j)}\right]\right\|_{(2,2)}\\ =&\frac{1}{(m-k_{\max})^2}\left\|\mathbb{E}\left[\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\left(\mathbf{a}'_{(i)j}\mathbf{\Sigma}_{(i)}^{-1}\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\right)\mathbf{a}'_{(i)j}\mathbf{\Sigma}_{(i)}^{-1}\mathcal{P}_T\right.\right.\\ &\left.\left. - \left(\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\mathcal{P}_T\right)^2\right]\right\|_{(2,2)}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(m-k_{\max})^2} \left\| \mathbb{E}\left[\left\| \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i)j} \right\|_2^2 \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i,j)} \mathbf{a}'_{(i)j} \mathbf{\Sigma}_{(i)}^{-1} \mathcal{P}_T \right] - \left(\mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathcal{P}_T \right)^2 \right\|_{(2,2)} \\
&\le \frac{1}{(m-k_{\max})^2} \left(\left\| \mathbb{E}\left[\left\| \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i)j} \right\|_2^2 \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i,j)} \mathbf{a}'_{(i)j} \mathbf{\Sigma}_{(i)}^{-1} \mathcal{P}_T \right] \right\|_{(2,2)} + \kappa^2 \right) \\
&\le \frac{1}{(m-k_{\max})^2} \left(\mu k_T \left\| \mathbb{E}\left[\mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i)j} \mathbf{a}'_{(i)j} \mathbf{\Sigma}_{(i)}^{-1} \mathcal{P}_T \right] \right\|_{(2,2)} + \kappa^2 \right) \\
&\le \frac{\kappa \mu k_T + \kappa^2}{(m-k_{\max})^2} \le \frac{\kappa^2(\mu k_T + 1)}{(m-k_{\max})^2} \le \frac{2\kappa^2 \mu k_T}{(m-k_{\max})^2},
\end{aligned}
$$

where the first equality follows from straight-up calculation using $\mathbb{E}\left[\mathbf{a}_{(i)j}\mathbf{a}'_{(i)j}\right] = \mathbf{\Sigma}_{(i)}$. The first inequality follows from triangle inequality, the second inequality follows from the definition of incoherence (8), and the rest of the inequalities uses the fact that $\kappa \ge 1$ and $\mu \ge 1$. Thus, by triangle inequality,

$$
\left\| \mathbb{E}\left[\sum_{j \in \Omega_i^*} \mathbf{M}'_{(j)} \mathbf{M}_{(j)} \right] \right\|_{(2,2)} \le \frac{2\kappa^2 \mu k_T}{(m-k_{\max})^2} \cdot (m - k_{\max}) = \frac{2\kappa^2 \mu k_T}{m - k_{\max}} \triangleq \sigma^2.
$$

Plugging B and σ^2 into Matrix Bernstein inequality, we finish the proof of Lemma 3. □

Appendix B. Proof of Lemma 4

Proof. We use the vector Bernstein inequality to prove the lemma. Picking any $k \in T^c$ and any $i \in \{1, \cdots, L\}$, we have

$$
\tilde{\mathbf{A}}_{(i)} \mathbf{e}_k = \frac{1}{m} \sum_{j \in \Omega_i^*} \langle \mathbf{a}_{(i)j}, \mathbf{e}_k \rangle \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i)j}.
$$

Letting

$$
\mathbf{g}_{(i,j)} = \frac{1}{m} \langle \mathbf{a}_{(i)j}, \mathbf{e}_k \rangle \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i)j},
$$

then it holds

$$
\mathcal{P}_T \tilde{\mathbf{A}}_{(i)} \mathbf{e}_k = \sum_{j \in \Omega_i^*} \mathbf{g}_{(i,j)}. \tag{B.1}
$$

Since $\{\mathbf{a}_{(i)j}\}_{j \in \Omega_i^*}$ are i.i.d. samples from $\mathcal{F}_i$, the sequence of vectors $\left\{\mathbf{g}_{(i,j)}\right\}_{j \in \Omega_i^*}$ are i.i.d. random variables. In order to apply the vector Bernstein inequality, we first need to show that $\mathbb{E}\left[\mathbf{g}_{(i,j)}\right] = 0$ for any $j \in \Omega_i^*$:

$$
\mathbb{E}\left[\mathbf{g}_{(i,j)}\right] = \frac{1}{m} \mathbb{E}\left[\langle \mathbf{a}_{(i)j}, \mathbf{e}_k \rangle \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i)j} \right] = \frac{1}{m} \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1} \mathbb{E}\left[\mathbf{a}_{(i)j} \mathbf{a}'_{(i)j} \right] \mathbf{e}_k = \frac{1}{m} \mathcal{P}_T \mathbf{e}_k = 0.
$$

The last equality is true since $k \in T^c$. Second, we calculate the bound B for any single $\left\|\mathbf{g}_{(i,j)}\right\|_2$:

$$\|\mathbf{g}_{(i,j)}\|_2^2 = \frac{1}{m^2}|\langle \mathbf{a}_{(i)j}, \mathbf{e}_k\rangle|^2 \|\mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\|_2^2 \leq \frac{\mu\|\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\|_2^2}{m^2} \leq \frac{\mu^2 k_T}{m^2},$$

where the first inequality follows from the incoherence condition (7) and the second inequality follows from (8). Furthermore, we have

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{g}_{(i,j)}\right\|_2^2\right] &= \frac{1}{m^2}\mathbb{E}\left[\left(\langle \mathbf{a}_{(i)j}, \mathbf{e}_k\rangle \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\right)'\left(\langle \mathbf{a}_{(i)j}, \mathbf{e}_k\rangle \mathcal{P}_T \mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\right)\right]\\ &\leq \frac{1}{m^2}\mu\mathbb{E}\left[\mathbf{a}_{(i)j}'\mathbf{\Sigma}_{(i)}^{-1}\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)j}\right]\\ &= \frac{1}{m^2}\mu \cdot \mathrm{Tr}\left(\mathbb{E}\left[\mathbf{a}_{(i)j}\mathbf{a}_{(i)j}'\right]\mathbf{\Sigma}_{(i)}^{-1}\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\right)\\ &= \frac{1}{m^2}\mu \cdot \mathrm{Tr}\left(\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}\right) \leq \frac{\mu k_T \kappa}{m^2},\end{aligned}$$

where Tr(·) denotes the trace of a matrix. The first inequality follows from the incoherence property (7). The last inequality follows from the fact that $\mathcal{P}_T\mathbf{\Sigma}_{(i)}^{-1}$ is of rank at most k_T so that its trace is upper bounded by $k_T\kappa$. Thus, it holds

$$\sum_{j\in\Omega_i^*}\mathbb{E}\left[\left\|\mathbf{g}_{(i,j)}\right\|_2^2\right] \leq \sum_{j\in\Omega_i^*}\frac{\mu\kappa k_T}{m^2} \leq \frac{\mu\kappa k_T}{m} \triangleq \sigma^2. \tag{B.2}$$

Substituting the above bound to the vector Bernstein inequality yields

$$Pr\left(\left\|\sum_{j\in\Omega_i^*}\mathbf{g}_{(i,j)}\right\|_2 \geq t\right) \leq \exp\left(-\frac{t^2}{\frac{8\mu\kappa k_T}{m}} + \frac{1}{4}\right),$$

given $\sigma^2/B = \sqrt{k_T}\kappa \geq 1$. Let $t = \sqrt{C\log(nL)\frac{\mu\kappa k_T}{m}}$. Using the fact that $k_T \leq \alpha\frac{m}{\mu\kappa\log(nL)}$, it holds $t \leq \sqrt{C\alpha}$ when $\alpha \leq \frac{1}{24}$. We can choose $C = 24$ such that $C\alpha \leq 1$, which guarantees $t \leq 1$ and gives

$$Pr\left(\left\|\sum_{j\in\Omega_i^*}\mathbf{g}_{(i,j)}\right\|_2 \geq 1\right) \leq e^{\frac{1}{4}}(nL)^{-3}.$$

Recalling (B.1) and taking a union bound over all $k \in T^c$ and $i \in \{1, \cdots, L\}$, we have

$$Pr\left(\max_{i\in\{1,\cdots,L\},k\in T^c}\left\|\mathcal{P}_T\tilde{\mathbf{A}}_{(i)}\mathbf{e}_k\right\|_2 \geq 1\right)$$

$$\leq \sum_{i=1}^{L} \sum_{k \in T^c} Pr\left(\left\| \sum_{j \in \Omega_i^*} \mathbf{g}_{(i,j)} \right\|_2 \geq 1 \right)$$
$$\leq (n - k_T) L e^{\frac{1}{4}} (nL)^{-3} \leq e^{\frac{1}{4}} (nL)^{-2}.$$

This completes the proof. □

Appendix C. Proof of Lemma 5

Proof. Since $\overline{\mathbf{Y}}$ and $\overline{\mathbf{S}}$ are the true group sparse signal and sparse error matrices, respectively, they satisfy the measurement equation

$$\mathbf{M} = [\mathbf{A}_{(1)}\bar{\mathbf{y}}_1, \cdots, \mathbf{A}_{(L)}\bar{\mathbf{y}}_L] + \overline{\mathbf{S}}.$$

Furthermore, since $(\overline{\mathbf{Y}} + \mathbf{H}, \overline{\mathbf{S}} - \mathbf{F})$ is an optimal solution to the RGL model (2)–(3), they must also satisfy the constraint

$$\mathbf{M} = [\mathbf{A}_{(1)}(\bar{\mathbf{y}}_1 + \mathbf{h}_1), \cdots, \mathbf{A}_{(L)}(\bar{\mathbf{y}}_L + \mathbf{h}_L)] + \overline{\mathbf{S}} - \mathbf{F}.$$

Subtracting these two equations yields result i) of Lemma 5.

Since the objective function of (2)–(3) is convex, we obtain an inequality

$$\|\overline{\mathbf{Y}} + \mathbf{H}\|_{2,1} + \lambda\|\overline{\mathbf{S}} - \mathbf{F}\|_1 \geq \|\overline{\mathbf{Y}}\|_{2,1} + \lambda\|\overline{\mathbf{S}}\|_1 + \langle \partial\|\overline{\mathbf{Y}}\|_{2,1}, \mathbf{H}\rangle - \lambda\langle \partial\|\overline{\mathbf{S}}\|_1, \mathbf{F}\rangle, \quad \text{(C.1)}$$

where $\partial\|\overline{\mathbf{Y}}\|_{2,1}$ denotes a subgradient of the $\ell_{2,1}$-norm at $\overline{\mathbf{Y}}$ and $\partial\|\overline{\mathbf{S}}\|_1$ denotes a subgradient of the ℓ_1-norm at $\overline{\mathbf{S}}$. Furthermore, the corresponding subgradients can be written as

$$\partial\|\overline{\mathbf{Y}}\|_{2,1} = \overline{\mathbf{V}} + \overline{\mathbf{R}},$$
$$\partial\|\overline{\mathbf{S}}\|_1 = \operatorname{sgn}(\overline{\mathbf{S}}) + \overline{\mathbf{Q}},$$

where $\overline{\mathbf{V}} \in \mathbb{R}^{n\times L}$ satisfies $(\mathcal{P}_T\overline{\mathbf{V}})^i = \frac{\bar{\mathbf{y}}^i}{\|\bar{\mathbf{y}}^i\|_2}$ and $(\mathcal{P}_{T^c}\overline{\mathbf{V}})^i = \mathbf{0}$, $\forall i = 1, \cdots, n$; $\overline{\mathbf{R}} \in \mathbb{R}^{n\times L}$ satisfies $\mathcal{P}_T\overline{\mathbf{R}} = \mathbf{0}$ and $\|\mathcal{P}_{T^c}\overline{\mathbf{R}}\|_{2,\infty} \leq 1$; $\overline{\mathbf{Q}} \in \mathbb{R}^{m\times L}$ satisfies $\mathcal{P}_\Omega\overline{\mathbf{Q}} = \mathbf{0}$ and $\|\mathcal{P}_{\Omega^c}\overline{\mathbf{Q}}\|_\infty \leq 1$. Therefore, we have

$$\|\overline{\mathbf{Y}} + \mathbf{H}\|_{2,1} + \lambda\|\overline{\mathbf{S}} - \mathbf{F}\|_1 \geq \|\overline{\mathbf{Y}}\|_{2,1} + \lambda\|\overline{\mathbf{S}}\|_1 + \langle \overline{\mathbf{V}} + \overline{\mathbf{R}}, \mathbf{H}\rangle - \lambda\langle sgn(\overline{\mathbf{S}}) + \overline{\mathbf{Q}}, \mathbf{F}\rangle, \quad \text{(C.2)}$$

for any $\overline{\mathbf{R}}$ and $\overline{\mathbf{Q}}$ satisfying the conditions mentioned above.

We construct a specific pair of $\overline{\mathbf{R}}$ and $\overline{\mathbf{Q}}$ in the following way. Let

$$\bar{\mathbf{r}}^i = \begin{cases} \frac{\mathbf{h}^i}{\|\mathbf{h}^i\|_2}, & \text{if } \mathbf{h}^i \neq \mathbf{0}' \text{ and } i \in T^c; \\ \mathbf{0}', & \text{otherwise,} \end{cases}$$

where $\mathbf{h}^i$ and $\bar{\mathbf{r}}^i$ are the i-th row of $\mathbf{H}$ and $\overline{\mathbf{R}}$, respectively. Meanwhile, let $\overline{\mathbf{Q}} = -\mathrm{sgn}(\mathcal{P}_{\Omega^c}\mathbf{F})$. It follows that

$$\begin{aligned}\langle \overline{\mathbf{R}}, \mathbf{H}\rangle &= \|\mathcal{P}_{T^c}\mathbf{H}\|_{2,1},\\ \langle \overline{\mathbf{Q}}, \mathbf{F}\rangle &= -\|\mathcal{P}_{\Omega^c}\mathbf{F}\|_1.\end{aligned}$$

Substituting the above equalities into (C.2) gives result ii) of Lemma 5. □

Appendix D. Proof of Lemma 6

Proof. We first show that $\mathcal{P}_T\mathbf{H} = \mathbf{0}$. Since $\left[\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathbf{A}_{(L)}\mathbf{h}_L\right] = \mathbf{F}$ and $\mathcal{P}_{\Omega^c}\mathbf{F} = \mathbf{0}$, it holds

$$[\mathcal{P}_{\Omega_1^c}\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathcal{P}_{\Omega_L^c}\mathbf{A}_{(L)}\mathbf{h}_L] = 0.$$

Meanwhile, $\mathcal{P}_{T^c}\mathbf{H} = \mathbf{0}$ implies $[\mathcal{P}_{\Omega_1^c}\mathbf{A}_{(1)}\mathcal{P}_{T^c}\mathbf{h}_1, \cdots, \mathcal{P}_{\Omega_L^c}\mathbf{A}_{(L)}\mathcal{P}_{T^c}\mathbf{h}_L] = 0$. Therefore, it holds

$$\begin{aligned}&[\mathcal{P}_{\Omega_1^c}\mathbf{A}_{(1)}\mathcal{P}_T\mathbf{h}_1, \cdots, \mathcal{P}_{\Omega_L^c}\mathbf{A}_{(L)}\mathcal{P}_T\mathbf{h}_L]\\ =&[\mathcal{P}_{\Omega_1^c}\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathcal{P}_{\Omega_L^c}\mathbf{A}_{(L)}\mathbf{h}_L] - [\mathcal{P}_{\Omega_1^c}\mathbf{A}_{(1)}\mathcal{P}_{T^c}\mathbf{h}_1, \cdots, \mathcal{P}_{\Omega_L^c}\mathbf{A}_{(L)}\mathcal{P}_{T^c}\mathbf{h}_L] = \mathbf{0}.\end{aligned}$$

Since for any $i = 1, \cdots, L$, Ω_i^* is a subset of Ω_i^c, it follows

$$\left[\mathcal{P}_{\Omega_1^*}\mathbf{A}_{(1)}\mathcal{P}_T\mathbf{h}_1, \cdots, \mathcal{P}_{\Omega_L^*}\mathbf{A}_{(L)}\mathcal{P}_T\mathbf{h}_L\right] = \mathbf{0},$$

and consequently

$$\begin{aligned}&\mathbf{Blkdiag}\left\{\mathcal{P}_T\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(1)}\mathcal{P}_T, \cdots, \mathcal{P}_T\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(L)}\mathcal{P}_T\right\}\cdot \mathrm{vec}(\mathbf{H})\\ =&\frac{m}{m-k_{\max}}\mathrm{vec}\left(\left[\mathcal{P}_T\mathbf{\Sigma}_{(1)}^{-1}\mathbf{A}'_{(1)}\mathcal{P}_{\Omega_1^*}\mathbf{A}_{(1)}\mathcal{P}_T\mathbf{h}_1, \cdots, \mathcal{P}_T\mathbf{\Sigma}_{(L)}^{-1}\mathbf{A}'_{(L)}\mathcal{P}_{\Omega_L^*}\mathbf{A}_{(L)}\mathcal{P}_T\mathbf{h}_L\right]\right)\\ =&\mathbf{0}.\end{aligned}$$

This equality implies

$$\begin{aligned}&\left\|\mathbf{Blkdiag}\left\{\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(1)} - \mathbf{I}\right)\mathcal{P}_T, \cdots, \mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(L)} - \mathbf{I}\right)\mathcal{P}_T\right\}\cdot \mathrm{vec}(\mathbf{H})\right\|_2\\ &= \|\mathcal{P}_T\mathbf{H}\|_F.\end{aligned}$$

On the other hand, according to (12), it follows with high probability

$$\left\|\mathbf{Blkdiag}\left\{\mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(1)} - \mathbf{I}\right)\mathcal{P}_T, \cdots, \mathcal{P}_T\left(\frac{m}{m-k_{\max}}\tilde{\mathbf{A}}_{(L)} - \mathbf{I}\right)\mathcal{P}_T\right\}\cdot \mathrm{vec}(\mathbf{H})\right\|_2$$

$$\leq \left\| \mathbf{Blkdiag} \left\{ \mathcal{P}_T \left(\frac{m}{m - k_{\max}} \tilde{\mathbf{A}}_{(1)} - \mathbf{I} \right) \mathcal{P}_T, \cdots, \mathcal{P}_T \left(\frac{m}{m - k_{\max}} \tilde{\mathbf{A}}_{(L)} - \mathbf{I} \right) \mathcal{P}_T \right\} \right\|_{(2,2)}$$

$$\cdot \|\mathcal{P}_T \mathbf{H}\|_F$$

$$\leq \frac{1}{2\sqrt{\log(nL)}} \|\mathcal{P}_T \mathbf{H}\|_F.$$

Thus,

$$\|\mathcal{P}_T \mathbf{H}\|_F \leq \frac{1}{2\sqrt{\log(nL)}} \|\mathcal{P}_T \mathbf{H}\|_F,$$

which implies $\mathcal{P}_T \mathbf{H} = \mathbf{0}$. Because $\mathcal{P}_{T^c} \mathbf{H} = \mathbf{0}$, we have $\mathbf{H} = \mathbf{0}$. Since $\mathbf{F} = \left[\mathbf{A}_{(1)}\mathbf{h}_1, \cdots, \mathbf{A}_{(L)}\mathbf{h}_L\right]$, it follows that $\mathbf{F} = 0$. □

Appendix E. Proof of Lemma 7

Proof. Following the proof of Lemma 3, for any $i = 1, \cdots, L$ and $j = 1, \cdots, l$ we have

$$Pr\left\{ \left\| \mathcal{P}_T \left(\frac{m}{m_j} \tilde{\mathbf{A}}_{(i,j)} - \mathbf{I} \right) \mathcal{P}_T \right\|_{(2,2)} \geq \tau \right\} \leq 2k_T \exp\left(-\frac{m_j}{\kappa k_T \mu} \frac{\tau^2}{4(1 + \frac{2\tau}{3})} \right).$$

Next, same as the proof of (11) and (12), for each $j = 1, \cdots, l$, taking a union bound over all $i = 1, \cdots, L$, which gives

$$Pr\left\{ \left\| \mathbf{BLKdiag} \left\{ \mathcal{P}_T \left(\frac{m}{m_j} \tilde{\mathbf{A}}_{(1,j)} - \mathbf{I} \right) \mathcal{P}_T, \ \cdots, \ \mathcal{P}_T \left(\frac{m}{m_j} \tilde{\mathbf{A}}_{(L,j)} - \mathbf{I} \right) \mathcal{P}_T \right\} \right\|_{(2,2)} \geq \tau \right\}$$

$$\leq 2k_T L \exp\left\{ -\frac{m_j}{k_T \mu \kappa} \frac{\tau^2}{4(1 + \frac{2\tau}{3})} \right\}. \tag{E.1}$$

If $j \geq 3$, then substituting $\tau = \frac{1}{2}$ and $m_j = \frac{m}{4\log(nL)}$ into above inequality gives

$$Pr\left\{ \left\| \mathbf{BLKdiag} \left\{ \mathcal{P}_T \left(\frac{m}{m_j} \tilde{\mathbf{A}}_{(1,j)} - \mathbf{I} \right) \mathcal{P}_T, \ \cdots, \ \mathcal{P}_T \left(\frac{m}{m_j} \tilde{\mathbf{A}}_{(L,j)} - \mathbf{I} \right) \mathcal{P}_T \right\} \right\|_{(2,2)} \geq \tau \right\}$$

$$\leq 2k_T L \exp\left\{ -\frac{3}{256} \frac{m}{k_T \mu \kappa \log(nL)} \right\}$$

$$\leq 2k_T L \exp\left\{ -3\log(nL) \right\} \leq 2(nL)^{-2},$$

where the second inequality follows from $k_T \leq \alpha \frac{m}{\mu\kappa \log^2(nL)}$ and $\alpha \leq \frac{1}{256}$. If $j \leq 2$, then substituting $\tau = \frac{1}{2\sqrt{\log(nL)}}$ and $m_j = \frac{m}{4}$ into (E.1) gives

$$
\begin{aligned}
&Pr\left\{\left\|\mathbf{BLKdiag}\left\{\mathcal{P}_T\left(\frac{m}{m_j}\tilde{\mathbf{A}}_{(1,j)}-\mathbf{I}\right)\mathcal{P}_T,\ \cdots,\ \mathcal{P}_T\left(\frac{m}{m_j}\tilde{\mathbf{A}}_{(L,j)}-\mathbf{I}\right)\mathcal{P}_T\right\}\right\|_{(2,2)}\geq\tau\right\}\\
&\leq 2k_T L\exp\left\{-\frac{3}{64}\frac{m}{k_T\mu\kappa}\frac{\sqrt{\log(nL)}}{\log(nL)(3\sqrt{\log(nL)}+1)}\right\}\\
&\leq 2k_T L\exp\left\{-\frac{3}{256}\frac{m}{k_T\mu\kappa}\frac{1}{\log(nL)}\right\}\\
&\leq 2k_T L\exp\left\{-3\log(nL)\right\}\leq 2(nL)^{-2}.
\end{aligned}
$$

Now taking a union bound over all $j=1,\cdots,l$ gives

$$
\begin{aligned}
Pr\left\{(37)\text{ holds for all } j=1,\cdots,l\right\}&\geq 1-2(nL)^{-2}l\geq 1-2(nL)^{-2}(\log(nL)+1)\\
&\geq 1-2(nL)^{-1},
\end{aligned}
$$

which finishes the proof. □

Appendix F. Proof of Theorem 3: existence of inexact dual certificate

① **Bounding the initial value:** $\left\|\lambda\mathcal{P}_{T^c}\left[\mathbf{A}'_{(1)}\mathrm{sgn}(\bar{\mathbf{s}}_1),\cdots,\mathbf{A}'_{(L)}\mathrm{sgn}(\bar{\mathbf{s}}_L)\right]\right\|_{2,\infty}\leq\frac{1}{8}$.

Proof. It is sufficient to prove

$$
\left\|\lambda\left[\mathbf{A}'_{(1)}\mathrm{sgn}(\bar{\mathbf{s}}_1),\cdots,\mathbf{A}'_{(L)}\mathrm{sgn}(\bar{\mathbf{s}}_L)\right]\right\|_{2,\infty}\leq\frac{1}{8}. \tag{F.1}
$$

Let $\mathbf{a}^r_{(i)}$ be the r-th row of $\sqrt{m}\mathbf{A}'_{(i)}$ and $a_{(i)rj}$ be the (r,j)-th element in $\sqrt{m}\mathbf{A}'_{(i)}$. Since sgn $(\bar{\mathbf{S}})$ is an i.i.d. Rademacher random matrix (because of i.i.d. signs), for any $r=1,\cdots,n$, we claim the following probability bound for the row ℓ_2-norm holds:

$$
\begin{aligned}
&Pr\left\{\sqrt{\sum_{i=1}^{L}|\mathbf{a}^r_{(i)}\mathrm{sgn}(\bar{\mathbf{s}}_i)|^2}-\sqrt{\sum_{i=1}^{L}\|\mathbf{a}^r_{(i)}\mathcal{P}_{\Omega_i}\|_2^2}\geq t\right\}\\
&\leq 4\exp\left\{-t^2\Big/\left(16\sum_{i=1}^{L}\|\mathbf{a}^r_{(i)}\mathcal{P}_{\Omega_i}\|_2^2\right)\right\}.
\end{aligned}
\tag{F.2}
$$

The proof of (F.2) follows from Corollary 4.10 in [28]. The details are given below.

According to Corollary 4.10 in [28], if $\mathbf{Z}\in\mathbb{R}^{m\times L}$ is distributed according to some product measure on $[-1,1]^{m\times L}$ and there exists a function $f:\ \mathbb{R}^{m\times L}\to\mathbb{R}$ which is convex and K-Lipschitz, then it holds

$$
Pr\{|f(\mathbf{Z})-\mathbb{E}\left[f(\mathbf{Z})\right]|\geq t\}\leq 4\exp\left\{-\frac{t^2}{16K^2}\right\}.
$$

Here we take $\mathbf{Z} = \operatorname{sgn}\left(\overline{\mathbf{S}}\right)$ and $f(\cdot) = \sqrt{\sum_{i=1}^{L} |\mathbf{a}_{(i)}^{r} \mathcal{P}_{\Omega_i}(\cdot)|^2}$. Notice that $\operatorname{sgn}\left(\overline{\mathbf{S}}\right)$ is entry-wise Bernoulli and the function f we choose is convex with the Lipschitz constant $K \leq \sqrt{\sum_{i=1}^{L} \left\|\mathbf{a}_{(i)}^{r} \mathcal{P}_{\Omega_i}\right\|_2^2}$, the requirements in above proposition are satisfied. In order to bound $\mathbb{E}\left[f(\mathbf{Z})\right]$ from above, we first compute $\mathbb{E}\left[f(\mathbf{Z})^2\right]$ and then use the property that $\mathbb{E}\left[f(\mathbf{Z})\right] \leq \sqrt{\mathbb{E}\left[f(\mathbf{Z})^2\right]}$. We have

$$\begin{aligned}
\mathbb{E}\left[f(\mathbf{Z})^2\right] &= \mathbb{E}\left[\sum_{i=1}^{L} \left|\mathbf{a}_{(i)}^{r} \operatorname{sgn}(\overline{\mathbf{s}}_i)\right|^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{L} \left|\sum_{j=1}^{m} a_{(i)rj} \operatorname{sgn}(\bar{s}_{ij})\right|^2\right] \\
&= \sum_{i=1}^{L} \mathbb{E}\left[\left|\sum_{j=1}^{m} a_{(i)rj} \operatorname{sgn}(\bar{s}_{ij})\right|^2\right] \\
&= \sum_{i=1}^{L} \sum_{j=1}^{m} \sum_{k=1}^{m} \mathbb{E}\left[a_{(i)rj} a_{(i)rk} \operatorname{sgn}(\bar{s}_{ij}) \operatorname{sgn}(\bar{s}_{ik})\right] \\
&= \sum_{i=1}^{L} \left\|\mathbf{a}_{(i)}^{r} \mathcal{P}_{\Omega_i}\right\|_2^2,
\end{aligned}$$

where the last step follows from the fact that for each $i = 1, \cdots, L$, $\operatorname{sgn}(\mathbf{s}_i)$ is a random vector with nonzero entries i.i.d. sot that all cross terms vanish. Thus, $\mathbb{E}\left[f(\mathbf{Z})\right] \leq \sqrt{\sum_{i=1}^{L} \left\|\mathbf{a}_{(i)}^{r} \mathcal{P}_{\Omega_i}\right\|_2^2}$. Hence,

$$\begin{aligned}
&Pr\left\{\sqrt{\sum_{i=1}^{L} \left|\mathbf{a}_{(i)}^{r} \operatorname{sgn}(\overline{\mathbf{s}}_i)\right|^2} - \sqrt{\sum_{i=1}^{L} \left\|\mathbf{a}_{(i)}^{r} \mathcal{P}_{\Omega_i}\right\|_2^2} \geq t\right\} \\
&\leq Pr\left\{\sqrt{\sum_{i=1}^{L} \left|\mathbf{a}_{(i)}^{r} \operatorname{sgn}(\overline{\mathbf{s}}_i)\right|^2} - \mathbb{E}\left[f(\mathbf{Z})\right] \geq t\right\} \\
&\leq Pr\left\{\left|\sqrt{\sum_{i=1}^{L} \left|\mathbf{a}_{(i)}^{r} \operatorname{sgn}(\overline{\mathbf{s}}_i)\right|^2} - \mathbb{E}\left[f(\mathbf{Z})\right]\right| \geq t\right\} \\
&\leq 4 \exp\left\{-\frac{t^2}{16K^2}\right\} \leq 4 \exp\left\{t^2 \Big/ \left(16 \sum_{i=1}^{L} \left\|\mathbf{a}_{(i)}^{r} \mathcal{P}_{\Omega_i}\right\|_2^2\right)\right\},
\end{aligned}$$

which proves (F.2).

Next, choose $t = 6\sqrt{\log(nL)}\sqrt{\sum_{i=1}^{L}\left\|\mathbf{a}_{(i)}^{r}\mathcal{P}_{\Omega_i}\right\|_2^2}$. Then with probability exceeding $1-4\exp\left\{-\frac{9}{4}\log(nL)\right\} = 1-4(nL)^{-\frac{9}{4}}$, it holds

$$\lambda\sqrt{\sum_{i=1}^{L}\left|\mathbf{a}_{(i)}^{r}\operatorname{sgn}(\bar{\mathbf{s}}_i)\right|^2} \le \lambda\left(6\sqrt{\log(nL)}+1\right)\sqrt{\sum_{i=1}^{L}\left\|\mathbf{a}_{(i)}^{r}\mathcal{P}_{\Omega_i}\right\|_2^2}$$

$$\le 7\sqrt{\frac{\theta}{L}\sum_{i=1}^{L}\left\|\mathbf{a}_{(i)}^{r}\mathcal{P}_{\Omega_i}\right\|_2^2} \le 7\sqrt{\frac{\mu k_\Omega \theta}{L}},$$

where the second from the last inequality follows from $\lambda = \sqrt{\frac{\theta}{L\log(nL)}}$, and the last inequality follows from the definition of incoherence parameter in (7) and the fact that $|\Omega| = k_\Omega$. Recall that $\mathbf{a}_{(i)}^{r}$ be the r-th row of $\sqrt{m}\mathbf{A}'_{(i)}$, taking a union bound over all $r = 1,\cdots,n$ gives

$$Pr\left\{\left\|\lambda\mathcal{P}_{T^c}\left[\mathbf{A}'_{(1)}\operatorname{sgn}(\bar{\mathbf{s}}_1)\ \cdots\ \mathbf{A}'_{(L)}\operatorname{sgn}(\bar{\mathbf{s}}_L)\right]\right\|_{2,\infty} \ge 7\sqrt{\frac{\mu k_\Omega\theta}{mL}}\right\}$$

$$= Pr\left\{\max_{r\in\{1,2,\cdots,n\}}\left\{\lambda\sqrt{\sum_{i=1}^{L}\left|\mathbf{a}_{(i)}^{r}\operatorname{sgn}(\bar{\mathbf{s}}_i)\right|^2}\right\} \ge 7\sqrt{\frac{\mu k_\Omega\theta}{L}}\right\}$$

$$\le \sum_{r=1}^{n} Pr\left\{\lambda\sqrt{\sum_{i=1}^{L}\left|\mathbf{a}_{(i)}^{r}\operatorname{sgn}(\bar{\mathbf{s}}_i)\right|^2} \ge 7\sqrt{\frac{\mu k_\Omega\theta}{L}}\right\}$$

$$\le 4(nL)^{-\frac{9}{4}}\cdot n \le 4(nL)^{-\frac{5}{4}} \le 4(nL)^{-1}.$$

Substituting the bounds $k_\Omega \le \beta\frac{mL}{\mu\theta}$ and $\beta \le \frac{1}{3136}$ into the above inequality finally gives

$$Pr\left\{\left\|\lambda\left[\mathbf{A}'_{(1)}\operatorname{sgn}(\bar{\mathbf{s}}_1),\cdots,\mathbf{A}'_{(L)}\operatorname{sgn}(\bar{\mathbf{s}}_L)\right]\right\|_{2,\infty} \ge \frac{1}{8}\right\} \le 4(nL)^{-1},$$

which finishes the proof. □

② **Bounding the term:** $\left\|\mathcal{P}_T\mathbf{U} + \lambda\mathcal{P}_T\left[\mathbf{A}'_{(1)}\operatorname{sgn}(\bar{\mathbf{s}}_1),\cdots,\mathbf{A}'_{(L)}\operatorname{sgn}(\bar{\mathbf{s}}_L)\right] - \overline{\mathbf{V}}\right\|_F \le \frac{\lambda}{4\sqrt{\kappa}}$.

Proof. Recalling the definition of $\mathbf{U}$ in (36), we have

$$\mathcal{P}_T\mathbf{U} = \mathcal{P}_T\left[\sum_{j=1}^{l}\frac{m}{m_j}\tilde{\mathbf{A}}_{(1,j)}\mathcal{P}_T\mathbf{q}_{(j-1)1},\cdots,\sum_{j=1}^{l}\frac{m}{m_j}\tilde{\mathbf{A}}_{(L,j)}\mathcal{P}_T\mathbf{q}_{(j-1)L}\right].$$

According to the definition of $\mathbf{Q}_{(0)}$ in (33), $\mathcal{P}_T\mathbf{Q}_{(0)} = \mathbf{Q}_{(0)}$. Since each subsequent mapping from $\mathbf{Q}_{(j-1)}$ to $\mathbf{Q}_{(j)}$ defined in (34) is a mapping from T to T, it follows that $\mathcal{P}_T\mathbf{Q}_{(j)} = \mathbf{Q}_{(j)}$ for any $j = 1, \cdots, l$. Therefore, it holds

$$\begin{aligned}
\mathcal{P}_T\mathbf{U} = &\sum_{j=1}^{l} \left(\mathbf{Q}_{(j-1)} - \mathcal{P}_T\mathbf{Q}_{(j-1)}\right) \\
&+ \mathcal{P}_T\left[\sum_{j=1}^{l}\frac{m}{m_j}\tilde{\mathbf{A}}_{(1,j)}\mathcal{P}_T\mathbf{q}_{(j-1)1}, \cdots, \sum_{j=1}^{l}\frac{m}{m_j}\tilde{\mathbf{A}}_{(L,j)}\mathcal{P}_T\mathbf{q}_{(j-1)L}\right] \\
= &\sum_{j=1}^{l}\left(\mathbf{Q}_{(j-1)} - \left[\mathcal{P}_T\left(\mathbf{I} - \frac{m}{m_j}\tilde{\mathbf{A}}_{(1,j)}\right)\mathcal{P}_T\mathbf{q}_{(j-1)1}, \cdots,\right.\right. \\
&\left.\left.\mathcal{P}_T\left(\mathbf{I} - \frac{m}{m_j}\tilde{\mathbf{A}}_{(L,j)}\right)\mathcal{P}_T\mathbf{q}_{(j-1)L}\right]\right) \\
= &\sum_{j=1}^{l}\left(\mathbf{Q}_{(j-1)} - \mathbf{Q}_{(j)}\right) = \mathbf{Q}_{(0)} - \mathbf{Q}_{(l)},
\end{aligned}$$

where the second last equality follows from the definition of $\mathbf{Q}_{(j)}$. Thus, substituting the definition of $\mathbf{Q}_{(0)}$ in (33) yields

$$\mathbf{Q}_{(l)} = \mathbf{Q}_{(0)} - \mathcal{P}_T\mathbf{U} = \overline{\mathbf{V}} - \lambda\mathcal{P}_T\left[\mathbf{A}'_{(1)}\mathrm{sgn}(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}\mathrm{sgn}(\bar{\mathbf{s}}_L)\right] - \mathcal{P}_T\mathbf{U},$$

which further implies

$$\left\|\mathcal{P}_T\mathbf{U} + \lambda\mathcal{P}_T\left[\mathbf{A}'_{(1)}\mathrm{sgn}(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}s\bar{g}n(\bar{\mathbf{s}}_L)\right] - \overline{\mathbf{V}}\right\|_F = \|\mathbf{Q}_{(l)}\|_F.$$

Thus, we are able to bound the target function on the left-hand side by bounding $\|\mathbf{Q}_{(l)}\|_F$ instead. It is enough to obtain an upper bound for $\|\mathbf{Q}_{(0)}\|_F$ and apply contractions (38)–(39). From (F.1), it follows

$$\left\|\lambda\mathcal{P}_T\left[\mathbf{A}'_{(1)}\mathrm{sgn}(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}\mathrm{sgn}(\bar{\mathbf{s}}_L)\right]\right\|_F \le \frac{\sqrt{k_T}}{8}.$$

Since $\|\overline{\mathbf{V}}\|_F = \sqrt{k_T}$, by triangle inequality, we have

$$\left\|\mathbf{Q}_{(0)}\right\| = \left\|\lambda\mathcal{P}_T\left[\mathbf{A}'_{(1)}\mathrm{sgn}(\bar{\mathbf{s}}_1), \cdots, \mathbf{A}'_{(L)}\mathrm{sgn}(\bar{\mathbf{s}}_L)\right] - \overline{\mathbf{V}}\right\|_F \le \frac{9\sqrt{k_T}}{8} \tag{F.3}$$

Thus, by contractions of $\{\mathbf{Q}_{(j)}\}_{j=1}^{l}$ in (38)–(39), we have

$$\|\mathbf{Q}_{(l)}\|_F \le \frac{1}{\log n}\frac{1}{2^l}\|\mathbf{Q}_{(0)}\|_F \le \frac{1}{\log(nL)}\frac{1}{2^l}\frac{9\sqrt{k_T}}{8} \le \frac{1}{\log(nL)}\frac{1}{nL}\frac{9\sqrt{k_T}}{8} \le \frac{\lambda}{4\sqrt{\kappa}},$$

provided that $\alpha \leq \frac{4}{81}$, where the last inequality follows from the fact $k_T \leq \alpha \frac{m}{\mu\kappa \log(nL)}$, and $\sqrt{m} \leq \sqrt{\mu} n \log(nL)$. Furthermore, from the proof, as long as (37) and (F.1) hold, this bound is guaranteed. □

③ **Bounding the term:** $\|\mathcal{P}_{T^c}\mathbf{U}\|_{2,\infty} \leq \frac{1}{8}$.

Proof. We claim that the following inequality is true with high probability:

$$\|\mathcal{P}_{T^c}\mathbf{U}\|_{2,\infty} \leq \sum_{j=1}^{l} \frac{1}{10\sqrt{k_T}} \|\mathbf{Q}_{(j-1)}\|_F. \tag{F.4}$$

According to the definition of $\mathbf{U}$ in (36), it holds

$$\begin{aligned} \mathcal{P}_{T^c}\mathbf{U} &= \left[\sum_{j=1}^{l} \frac{m}{m_j} \mathcal{P}_{T^c} \tilde{\mathbf{A}}_{(1,j)} \mathcal{P}_T \mathbf{q}_{(j-1)1}, \cdots, \sum_{j=1}^{l} \frac{m}{m_j} \mathcal{P}_{T^c} \tilde{\mathbf{A}}_{(L,j)} \mathcal{P}_T \mathbf{q}_{(j-1)L} \right] \\ &= \sum_{j=1}^{l} \left[\frac{m}{m_j} \mathcal{P}_{T^c} \tilde{\mathbf{A}}_{(1,j)} \mathcal{P}_T \mathbf{q}_{(j-1)1}, \cdots, \frac{m}{m_j} \mathcal{P}_{T^c} \tilde{\mathbf{A}}_{(L,j)} \mathcal{P}_T \mathbf{q}_{(j-1)L} \right]. \end{aligned}$$

Thus, it is enough to show that for any $k \in T^c$, it holds

$$\left\| \sum_{j=1}^{l} \left[\frac{m}{m_j} \mathbf{e}_k' \tilde{\mathbf{A}}_{(1,j)} \mathcal{P}_T \mathbf{q}_{(j-1)1}, \cdots, \frac{m}{m_j} \mathbf{e}_k' \tilde{\mathbf{A}}_{(L,j)} \mathcal{P}_T \mathbf{q}_{(j-1)L} \right] \right\|_2 \leq \sum_{j=1}^{l} \frac{1}{10\sqrt{k_T}} \|\mathbf{Q}_{(j-1)}\|_F, \tag{F.5}$$

with high probability, where $\{\mathbf{e}_k\}_{k=1}^n$ is a standard basis in $\mathbb{R}^n$. By the triangle inequality, a sufficient condition for (F.5) to satisfy is

$$\sum_{j=1}^{l} \left\| \left[\frac{m}{m_j} \mathbf{e}_k' \tilde{\mathbf{A}}_{(1,j)} \mathcal{P}_T \mathbf{q}_{(j-1)1} \quad \cdots \quad \frac{m}{m_j} \mathbf{e}_k' \tilde{\mathbf{A}}_{(L,j)} \mathcal{P}_T \mathbf{q}_{(j-1)L} \right] \right\|_2 \leq \sum_{j=1}^{l} \frac{1}{10\sqrt{k_T}} \|\mathbf{Q}_{(j-1)}\|_F.$$

Therefore, it resorts to proving a one-step-further sufficient condition that with high probability, for any $j = 1, \cdots, l$ and $k \in T^c$, it holds

$$\left\| \left[\frac{m}{m_j} \mathbf{e}_k' \tilde{\mathbf{A}}_{(1j)} \mathcal{P}_T \mathbf{q}_{(j-1)1} \quad \cdots \quad \frac{m}{m_j} \mathbf{e}_k' \tilde{\mathbf{A}}_{(L,j)} \mathcal{P}_T \mathbf{q}_{(j-1)L} \right] \right\|_2 \leq \frac{1}{10\sqrt{k_T}} \|\mathbf{Q}_{(j-1)}\|_F. \tag{F.6}$$

We apply the vector Bernstein inequality to prove (F.6). First, for any $i = 1, \cdots, L$, and any $r \in K_{ij}$, let

$$g_{(i,r)} = \frac{1}{m_j} \mathbf{e}_k' \mathbf{\Sigma}_{(i)}^{-1} \mathbf{a}_{(i)r} \mathbf{a}_{(i)r}' \mathcal{P}_T \mathbf{q}_{(j-1)i}.$$

Observe the fact that $\{\mathbf{a}_{(i)j}\}_{j\in K_{ij}}$ is the set of column vectors in $\sqrt{m}\mathbf{A}'_{(i)}P_{K_{ij}}$, which are nonzero. Also, recall the definition $\tilde{\mathbf{A}}_{(i,j)} = \boldsymbol{\Sigma}_{(i)}^{-1}\mathbf{A}'_{(i)}\mathcal{P}_{K_{ij}}\mathbf{A}_{(i)}$. For any $i = 1, \cdots, L$, it follows

$$\sum_{r\in K_{ij}} g_{(i,r)} = \frac{m}{m_j}\mathbf{e}'_k\tilde{\mathbf{A}}_{(i,j)}\mathcal{P}_T\mathbf{q}_{(j-1)i}. \tag{F.7}$$

For notation convenience, without loss of generality, suppose $K_{ij} = \{1, \cdots, m_j\}$, $\forall i = 1, \cdots, L$. For any $r = 1, \cdots, m_j$, we align the scalars $g_{(i,r)}$, $i = 1, \cdots, L$ into a single vector as $\left[g_{(1,r)}, \cdots, g_{(L,r)}\right]$. According to (F.7), this vector satisfies

$$\sum_{r=1}^{m_j}\left[g_{(1,r)}, \cdots, g_{(L,r)}\right] = \left[\frac{m}{m_j}\mathbf{e}'_k\tilde{\mathbf{A}}_{(1,j)}\mathcal{P}_T\mathbf{q}_{(j-1)1}, \cdots, \frac{m}{m_j}\mathbf{e}'_k\tilde{\mathbf{A}}_{(L,j)}\mathcal{P}_T\mathbf{q}_{(j-1)L}\right].$$

Notice that $\mathbf{Q}_{(j-1)}$ is also a random variable. In the following proof, we apply the vector Bernstein inequality conditioned on $\mathbf{Q}_{(j-1)}$. It is obvious that given $\mathbf{Q}_{(j-1)}$, $\left[g_{(1,r)}, \cdots, g_{(L,r)}\right]$ are i.i.d. for different r and

$$\mathbb{E}\left[\left[g_{(1,r)}, \cdots, g_{(L,r)}\right]\middle|\mathbf{Q}_{(j-1)}\right] = \frac{m}{m_j}\left[\mathbf{e}'_k\mathcal{P}_T\mathbf{q}_{(j-1)1}, \cdots, \mathbf{e}'_k\mathcal{P}_T\mathbf{q}_{(j-1)L}\right] = 0,$$

since $k \in T^c$. Next, we compute

$$\begin{aligned}
\mathbb{E}\left[\left(g_{(i,r)}\right)^2\middle|\mathbf{Q}_{(j-1)}\right] &= \frac{1}{m_j^2}\mathbb{E}\left[\left(\mathbf{e}'_k\boldsymbol{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)r}\mathbf{a}'_{(i)r}\mathcal{P}_T\mathbf{q}_{(j-1)i}\right)^2\middle|\mathbf{Q}_{(j-1)}\right]\\
&= \frac{1}{m_j^2}\mathbb{E}\left[(\mathbf{e}'_k\boldsymbol{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)r})^2(\mathbf{a}'_{(i)r}\mathcal{P}_T\mathbf{q}_{(j-1)i})^2\middle|\mathbf{Q}_{(j-1)}\right]\\
&\le \frac{\mu}{m_j^2}\mathbb{E}\left[\mathbf{q}'_{(j-1)i}\mathcal{P}_T\mathbf{a}_{(i)r}\mathbf{a}'_{(i)r}\mathcal{P}_T\mathbf{q}_{(j-1)i}\middle|\mathbf{Q}_{(j-1)}\right]\\
&\le \frac{\mu\kappa}{m_j^2}\|\mathbf{q}_{(j-1)i}\|_2^2.
\end{aligned}$$

Therein, the first inequality follows from the definition of the incoherence parameter (8). The second inequality follows from the fact that for each $i = 1, \cdots, L$, the sampled batches of vectors $\mathcal{P}_{K_{ij}}\mathbf{A}_{(i)}$ are not overlapped for different batches $j = 1, \cdots, l$ such that $\mathbf{q}_{(j-1)i}$ and $\mathbf{a}_{(i,r)}$ are independent. Thus, we have

$$\sum_{r=1}^{m_j}\mathbb{E}\left[\left\|\left[g_{(1,r)}, \cdots, g_{(L,r)}\right]\right\|_2^2\middle|\mathbf{Q}_{(j-1)}\right] \le \sum_{r=1}^{m_j}\sum_{i=1}^{L}\frac{\mu\kappa}{m_j^2}\|\mathbf{q}_{(j-1)i}\|_2^2 \le \frac{\mu\kappa}{m_j}\|\mathbf{Q}_{(j-1)}\|_F^2 \triangleq \sigma^2.$$

Moreover,

$$|g_{(i,r)}| = \frac{1}{m_j}\left|\mathbf{e}'_k\boldsymbol{\Sigma}_{(i)}^{-1}\mathbf{a}_{(i)r}\mathbf{a}'_{(i)r}\mathcal{P}_T\mathbf{q}_{(j-1)i}\right| \le \frac{1}{m_j}\sqrt{\mu}\left|\mathbf{a}'_{(i)r}\mathcal{P}_T\mathbf{q}_{(j-1)i}\right| \le \frac{\mu\sqrt{k_T}}{m_j}\|\mathbf{q}_{(j-1)i}\|_2,$$

where the first inequality follows from the incoherence assumption (8) and the second inequality follows from the incoherence condition (7). Thus, it holds

$$\left\| \left[g_{(1,r)}, \cdots, g_{(L,r)} \right] \right\|_2 \leq \frac{\mu\sqrt{k_T}}{m_j} \|\mathbf{Q}_{(j-1)}\|_F \triangleq B.$$

Substituting the above bounds into the vector Bernstein inequality conditioned on $\mathbf{Q}_{(j-1)}$ gives

$$Pr\left(\left\| \sum_{r=1}^{m_j} \left[g_{(1,r)}, \cdots, g_{(L,r)} \right] \right\|_2 \geq t \,\middle|\, \mathbf{Q}_{(j-1)} \right) \leq \exp\left(-\frac{t^2}{\frac{8\mu\kappa}{m_j}\|\mathbf{Q}_{(j-1)}\|_F^2} + \frac{1}{4} \right). \tag{F.8}$$

We choose $t = \sqrt{\frac{24}{m_j}\mu\kappa\log(nL)}\|\mathbf{Q}_{(j-1)}\|_F$. First, we need to verify that such a choice satisfies $t \leq \frac{\sigma^2}{B}$. Recall that for any $j = 1, \cdots, l$, $m_j \geq \frac{m}{4\log(nL)}$. Since $k_T \leq \alpha\frac{m}{\mu\kappa\log^2(nL)}$ and $\alpha \leq \frac{1}{9600}$, it holds

$$t \leq \sqrt{\frac{24\mu\kappa\log^2(nL)}{m}}\|\mathbf{Q}_{(j-1)}\|_F \leq \frac{1}{10\sqrt{k_T}}\|\mathbf{Q}_{(j-1)}\|_F \leq \kappa\|\mathbf{Q}_{(j-1)}\|_F = \frac{\sigma^2}{B}.$$

Thus, the choice of t is indeed valid. Substituting this t into (F.8) gives

$$Pr\left(\left\| \sum_{r=1}^{m_j} \left[g_{(1,r)}, \cdots, g_{(L,r)} \right] \right\|_2 \geq \sqrt{\frac{24}{m_j}\mu\kappa\log(nL)}\|\mathbf{Q}_{(j-1)}\|_F \,\middle|\, \mathbf{Q}_{(j-1)} \right) \leq e^{\frac{1}{4}}(nL)^{-3},$$

which implies

$$Pr\left(\left\| \sum_{r=1}^{m_j} \left[g_{(1,r)}, \cdots, g_{(L,r)} \right] \right\|_2 \geq \frac{1}{10\sqrt{k_T}}\|\mathbf{Q}_{(j-1)}\|_F \,\middle|\, \mathbf{Q}_{(j-1)} \right) \leq e^{\frac{1}{4}}(nL)^{-3}.$$

Since the right-hand side does not depend on $\mathbf{Q}_{(j-1)}$, taking expectation from both sides regarding $\mathbf{Q}_{(j-1)}$ gives

$$Pr\left(\left\| \sum_{r=1}^{m_j} \left[g_{(1,r)}, \cdots, g_{(L,r)} \right] \right\|_2 \geq \frac{1}{10\sqrt{k_T}}\|\mathbf{Q}_{(j-1)}\|_F \right) \leq e^{\frac{1}{4}}(nL)^{-3}.$$

Take a union bound over all $j = 1, \cdots, l$ and $k \in T^c$ gives

$$Pr\left(\max_{j\in\{1,\cdots,l\},k\in T^c} \left\{ \left\| \sum_{r=1}^{m_j} \left[g_{(1,r)}, \cdots, g_{(L,r)} \right] \right\|_2 \right\} \geq \frac{1}{10\sqrt{k_T}}\|\mathbf{Q}_{(j-1)}\|_F \right)$$
$$\leq \sum_{j=1}^{l}\sum_{k\in T^c} Pr\left(\left\| \sum_{r=1}^{m_j} \left[g_{(1,r)}, \cdots, g_{(L,r)} \right] \right\|_2 \geq \frac{1}{10\sqrt{k_T}}\|\mathbf{Q}_{(j-1)}\|_F \right)$$

$$\leq \sum_{j=1}^{l} \sum_{k \in T^c} e^{-\frac{1}{4}} (nL)^{-3} \leq (\log(nL)+1) \cdot n \cdot e^{\frac{1}{4}} (nL)^{-3} \leq e^{\frac{1}{4}} (nL)^{-1}.$$

This proves (F.6) and further implies that (F.4) holds. Finally, applying the contractions (38)–(39) gives

$$\|\mathcal{P}_{T^c}\mathbf{U}\|_{2,\infty} \leq \sum_{j=1}^{l} \frac{1}{10\sqrt{k_T}} \|\mathbf{Q}_{(j-1)}\|_F \leq \sum_{j=1}^{l} \frac{1}{10\sqrt{k_T}} \frac{1}{2^j} \|\mathbf{Q}_{(0)}\|_F \leq \frac{1}{10\sqrt{k_T}} \|\mathbf{Q}_{(0)}\|_F.$$

Substituting the bound on $\|\mathbf{Q}_{(0)}\|_F$ in (F.3) gives the desired result. Notice that the inequality $\|\mathcal{P}_{T^c}\mathbf{U}\|_{2,\infty} \leq \frac{1}{8}$ requires (37), (F.1) and (F.4) to hold simultaneously. □

④ **Bounding the term:** $\|\mathcal{P}_{\Omega^c}\mathbf{W}\|_\infty \leq \frac{\lambda}{4}$.

Proof. According to the definition of $\mathbf{W}$ in (35), we aim to prove

$$\left\| \left[\sum_{j=1}^{l} \frac{m}{m_j} \mathcal{P}_{K_{1j}} \mathbf{A}_{(1)} \mathcal{P}_T \mathbf{q}_{(j-1)1}, \quad \cdots, \quad \sum_{j=1}^{l} \frac{m}{m_j} \mathcal{P}_{K_{Lj}} \mathbf{A}_{(L)} \mathcal{P}_T \mathbf{q}_{(j-1)L} \right] \right\|_\infty \leq \frac{\lambda}{4}.$$

Notice that the batch sets K_{ij}, $j = 1, \cdots, l$ are not overlapped. Therefore, it is enough to show with high probability, for any $i = 1, \cdots, L$, any $j = 1, \cdots, l$, and any vector $\mathbf{a}_{(i)r}$ with $r \in K_{ij}$, it holds

$$\left| \frac{\sqrt{m}}{m_j} \mathbf{a}'_{(i)r} \mathcal{P}_T \mathbf{q}_{(j-1)i} \right| \leq \frac{\lambda}{4}.$$

Equivalently, according to the definition of $\mathbf{Q}_{(j)}$ in (34), it is enough to prove for $j \geq 2$ it holds

$$\left| \frac{\sqrt{m}}{m_j} \mathbf{a}'_{(i)r} \left(\prod_{k=1}^{j-1} \mathcal{P}_T \left(\mathbf{I} - \tilde{\mathbf{A}}_{(1,k)} \right) \mathcal{P}_T \right) \mathbf{q}_{(0)i} \right| \leq \frac{\lambda}{4},$$

and for $j = 1$ it holds

$$\left| \frac{\sqrt{m}}{m_j} \mathbf{a}'_{(i)r} \mathcal{P}_T \mathbf{q}_{(0)i} \right| \leq \frac{\lambda}{4}.$$

In order to further simplify the notation, for any vector $\mathbf{a}_{(i)r}$ such that $r \in K_{ij}$, let

$$\mathbf{g}'_{(i,r)} \triangleq \begin{cases} \mathbf{a}'_{(i)r} \left(\prod_{k=1}^{j-1} \mathcal{P}_T \left(\mathbf{I} - \tilde{\mathbf{A}}_{(1,k)} \right) \mathcal{P}_T \right), & \text{if } j \geq 2; \\ \mathbf{a}'_{(i)r}, & \text{if } j = 1. \end{cases}$$

Our goal is to prove that for any $i = 1, \cdots, L$, any $j = 1, \cdots, l$, and any vector $\mathbf{a}_{(i,r)}$ in the j-th batch vectors $\mathcal{P}_{K_{ij}} A_{(i)}$, with high probability it holds

$$\left|\frac{\sqrt{m}}{m_j}\mathbf{g}'_{(i,r)}\mathbf{q}_{(0)i}\right| \leq \frac{\lambda}{4}. \tag{F.9}$$

Since both $\mathbf{g}_{(i,r)}$ and $\mathbf{q}_{(0)i}$ are random variables, it is easier to first bound the left-hand side of (F.9) conditioned on $\mathbf{g}_{(i,r)}$. Recall the definition of $\mathbf{Q}_{(0)}$ in (33), for any $i = 1, \cdots, L$, it holds

$$\mathbf{q}_{(0)i} = \bar{\mathbf{v}}_i - \mathcal{P}_T\mathbf{A}'_{(i)}\text{sgn}(\mathbf{s}_i).$$

By the triangle inequality, we bound $\left|\mathbf{g}'_{(i,r)}\bar{\mathbf{v}}_i\right|$ and $\left|\lambda\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{A}'_{(i)}\text{sgn}(\mathbf{s}_i)\right|$ respectively.

Let us first bound $\left|\mathbf{g}'_{(i,r)}\bar{\mathbf{v}}_i\right|$ conditioned on $\mathbf{g}_{(i,r)}$. From our assumption, the vector $\bar{\mathbf{v}}_i$ is fixed except for the i.i.d. signs. Denote $|\bar{\mathbf{v}}_i|$ as the *entry-wise absolute value vector* of $\bar{\mathbf{v}}_i$, which is not random. Then,

$$\mathbf{g}'_{(i,r)}\bar{\mathbf{v}}_i = \left(\mathbf{g}_{(i,r)} \odot |\bar{\mathbf{v}}_i|\right)' \cdot \text{sgn}(\bar{\mathbf{v}}_i)$$

where $\odot$ denotes the entry-wise Hadamand product. Notice that $\text{sgn}(\bar{\mathbf{v}}_i)$ and $\mathbf{g}_{(i,r)}$ are mutually independent. Applying Hoeffding inequality conditioned on $\mathbf{g}_{(i,r)}$ gives

$$Pr\left\{\left|\left(\mathbf{g}_{(i,r)} \odot |\bar{\mathbf{v}}_i|\right)' \cdot \text{sgn}(\bar{\mathbf{v}}_i)\right| \geq t \;\middle|\; \mathbf{g}_{(i,r)}\right\} \leq 2\exp\left\{-\frac{t^2}{2\left\|\mathbf{g}_{(i,r)} \odot |\bar{\mathbf{v}}_i|\right\|_2^2}\right\}.$$

By the row incoherence condition (Assumption 4) each entry of $\bar{\mathbf{v}}_i$ is within $\left[-\sqrt{\frac{\nu}{L}}, \sqrt{\frac{\nu}{L}}\right]$. Taking $t = 2\sqrt{\log(nL)}\sqrt{\frac{\nu}{L}}\|\mathbf{g}_{(i,r)}\|_2$, it follows

$$Pr\left\{\left|\left(\mathbf{g}_{(i,r)} \odot |\bar{\mathbf{v}}_i|\right)' \cdot \text{sgn}(\bar{\mathbf{v}}_i)\right| \geq 2\sqrt{\log(nL)}\sqrt{\frac{\nu}{L}}\|\mathbf{g}_{(i,r)}\|_2 \;\middle|\; \mathbf{g}_{(i,r)}\right\} \leq 2(nL)^{-2}. \tag{F.10}$$

Second, we bound $\left|\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{A}'_{(i)}\text{sgn}(\bar{\mathbf{s}}_i)\right|$ conditioned on $\mathbf{g}_{(i,r)}$. The key is to prove the argument that $\mathcal{P}_T\mathbf{g}_{(i,r)}$ is independent of $\mathcal{P}_T\mathbf{A}'_{(i)}\text{sgn}(\bar{\mathbf{s}}_i)$. Notice that by definition, $\mathbf{g}_{(i,r)}$ is generated by the column vectors in $\mathbf{A}'_{(i)}$ with column indices from the batch sets K_{ij}, $j = 1, \cdots, l$. Recall the definition of these batch sets under (33), $\cup_{j=1}^{l}K_{ij} \subseteq \Omega_i^* \subseteq \Omega_i^c$. On the other hand, $\mathbf{A}'_{(i)}\text{sgn}(\mathbf{s}_i)$ picks out those column vectors in $\mathbf{A}_{(i)}$ with the column indices from Ω_i. Since different columns of $\mathbf{A}'_{(i)}$ are i.i.d. samples from the distribution $\mathcal{F}_i$, the argument holds true.

Moreover, since the noise support Ω_i are assumed to be fixed and the signs of noise matrix are i.i.d., $\mathbf{A}_{(i)}$ and $\text{sgn}(\bar{\mathbf{s}}_i)$ are also independent. We write

$$\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{A}'_{(i)}\text{sgn}(\bar{\mathbf{s}}_i) = \frac{1}{\sqrt{m}}\sum_{x\in\Omega_i}\mathbf{g}'_{(i,r)}\mathbf{a}_{(i)x} \cdot \text{sgn}(\bar{s}_{ix}).$$

Then, for any $x \in \Omega_i$, we have

$$\mathbb{E}\left[\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{a}_{(i)x}\mathrm{sgn}(\bar{s}_{ix})\middle|\,\mathbf{g}_{(i,r)}\right] = \mathbf{g}'_{(i,r)}\mathbb{E}\left[\mathcal{P}_T\mathbf{a}'_{(i)x}\right]\mathbb{E}\left[\mathrm{sgn}(\bar{s}_{ix})\right] = 0,$$

$$\left|\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{a}_{(i)x}\mathrm{sgn}(\bar{s}_{ix})\right| \leq \sqrt{\mu k_T}\|\mathbf{g}_{(i,r)}\|_2,$$

$$\mathbb{E}\left[\left|\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{a}_{(i)x}\mathrm{sgn}(\bar{s}_{ix})\right|^2\middle|\,\mathbf{g}_{(i,r)}\right] = \mathbf{g}'_{(i,r)}\mathbb{E}\left[\mathcal{P}_T\mathbf{a}_{(i)x}\mathbf{a}'_{(i)x}\mathcal{P}_T\right]\mathbf{g}_{(i,r)} \leq \kappa\|\mathbf{g}_{(i,r)}\|_2^2.$$

Thus, using the one dimensional Bernstein inequality (which can also be viewed as a special case of the matrix Bernstein's inequality), we have

$$\begin{aligned}
&Pr\left\{\left|\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{A}'_{(i)}\mathrm{sgn}(\bar{\mathbf{s}}_i)\right| > \frac{t}{\sqrt{m}}\middle|\,\mathbf{g}_{(i,r)}\right\} \\
&= Pr\left\{\left|\sum_{x\in\Omega_i}\mathbf{g}'_{(i,r)}\mathbf{a}_{(i)x}\cdot\mathrm{sgn}(\bar{s}_{ix})\right| > t\,\middle|\,\mathbf{g}_{(i,r)}\right\} \\
&\leq 2\exp\left(-\frac{\frac{1}{2}t^2}{k_{\Omega_i}\kappa\|\mathbf{g}_{(i,r)}\|_2^2 + \sqrt{k_T}\mu\frac{\|\mathbf{g}_{(i,r)}\|_2 t}{3}}\right)
\end{aligned}$$

Since $k_{\max} \leq \gamma\frac{m}{\kappa}$ with $\gamma \leq \frac{1}{4}$ and $k_T \leq \alpha\frac{m}{\mu\kappa\log^2(nL)}$ with $\alpha \leq \frac{1}{9600}$, choosing $t = 2\sqrt{m\log(nL)}\|\mathbf{g}_{(i,r)}\|_2$ gives

$$\begin{aligned}
&Pr\left\{\left|\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{A}'_{(i)}\mathrm{sgn}(\bar{\mathbf{s}}_i)\right| > 2\sqrt{\log(nL)}\|\mathbf{g}_{(i,r)}\|_2\,\middle|\,\mathbf{g}_{(i,r)}\right\} \\
&\leq 2\exp\left\{-\frac{2m\log(nL)}{\frac{m}{4} + \frac{1}{60\sqrt{6\log(nL)}}}\right\} \leq 2(nL)^{-2},
\end{aligned}$$

which implies

$$Pr\left\{\left|\lambda\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{A}'_{(i)}\mathrm{sgn}(\bar{\mathbf{s}}_i)\right| > 2\sqrt{\frac{\log(nL)}{L}}\|\mathbf{g}_{(i,r)}\|_2\,\middle|\,\mathbf{g}_{(i,r)}\right\} \leq 2(nL)^{-2}. \tag{F.11}$$

Combining (F.10) and (F.11) gives

$$\begin{aligned}
&Pr\left\{\left|\mathbf{g}'_{(i,r)}\mathbf{q}_{(0)i}\right| > 4\sqrt{\frac{\nu\log(nL)}{L}}\|\mathbf{g}_{(i,r)}\|_2\,\middle|\,\mathbf{g}_{(i,r)}\right\} \\
&\leq Pr\left\{\left|\mathbf{g}'_{(i,r)}\mathcal{P}_T\mathbf{A}'_{(i)}\mathrm{sgn}(\bar{\mathbf{s}}_i)\right| > 2\sqrt{\frac{\log(nL)}{L}}\|\mathbf{g}_{(i,r)}\|_2\,\middle|\,\mathbf{g}_{(i,r)}\right\} \\
&\quad + Pr\left\{\left|\mathbf{g}'_{(i,r)}\bar{\mathbf{v}}_i\right| \geq 2\sqrt{\frac{\nu\log(nL)}{L}}\|\mathbf{g}_{(i,r)}\|_2\,\middle|\,\mathbf{g}_{(i,r)}\right\} \leq 4(nL)^{-2}.
\end{aligned}$$

Notice that because we bound the probability conditioned on $\mathbf{g}_{(i,r)}$, the bound hold for any $j = 1, \cdots, l$ and any $r \in K_{ij}$. Now take a union bound over all $i = 1, \cdots, L$,

$$\begin{aligned} &Pr\left\{\bigcup_{i=1}^{L}\left\{\left|\mathbf{g}'_{(i,r)}\mathbf{q}_{(0)i}\right| > 4\sqrt{\frac{\nu\log(nL)}{L}}\|\mathbf{g}_{(i,r)}\|_2\right\}\middle|\,\mathbf{g}_{(i,r)}\right\} \\ \le &\sum_{i=1}^{L} Pr\left\{\left|\mathbf{g}'_{(i,r)}\mathbf{q}_{(0)i}\right| > 4\sqrt{\frac{\nu\log(nL)}{L}}\|\mathbf{g}_{(i,r)}\|_2\middle|\,\mathbf{g}_{(i,r)}\right\} \\ \le & L\cdot 4(nL)^{-2} \le 4(nL)^{-1}. \end{aligned}$$

Since the probability on the right-hand side does not depend on $\mathbf{g}_{(i,r)}$ and the inequality holds for any $j = 1, \cdots, l$, any $r \in K_{ij}$, and any $i = 1, \cdots, L$, with probability at least $1 - 4(nL)^{-1}$ it follows

$$\left|\frac{\sqrt{m}}{m_j}\mathbf{g}'_{(i,r)}\mathbf{q}_{(0)i}\right| \le 4\frac{\sqrt{m}}{m_j}\sqrt{\frac{\nu\log(nL)}{L}}\|\mathbf{g}_{(i,r)}\|_2. \tag{F.12}$$

Next, we bound $\|\mathbf{g}_{(i,r)}\|_2$ using contractions (38)–(39). According to Lemma 7, with probability at least $1 - 2(nL)^{-1}$, (38)–(39) hold simultaneously. Thus, with probability at least $1 - 2(nL)^{-1}$, for any $j \ge 3$, any $r \in K_{ij}$, and any $i = 1, \cdots, L$, it holds

$$\begin{aligned} \|\mathbf{g}_{(i,r)}\|_2 \le \|\mathbf{a}_{(i)r}\|_2 &\left\|\left(\prod_{k=1}^{j-1}\mathcal{P}_T\left(\mathbf{I} - \tilde{\mathbf{A}}_{(1,k)}\right)\mathcal{P}_T\right)\right\|_{(2,2)} \\ &\le \frac{1}{\log(nL)}\frac{1}{2^{j-1}}\sqrt{k_T\mu\theta} \le \frac{1}{\log^2(nL)}\sqrt{\frac{\alpha m\theta}{\nu}}, \end{aligned}$$

given $k_T \le \alpha\frac{m\theta}{\mu\nu\log^2 n}$. Thus, combining with (F.12) gives

$$\begin{aligned} \left|\frac{\sqrt{m}}{m_j}\mathbf{g}'_{(i,r)}\mathbf{q}_{(0)i}\right| &\le \frac{m}{m_j}4\log^{-\frac{3}{2}}(nL)\sqrt{\frac{\alpha\theta}{L}} \\ &\le \frac{16}{\sqrt{9600}}\sqrt{\frac{\theta}{L}}\log^{-\frac{1}{2}}(nL) \\ &= \frac{2}{5\sqrt{6}}\lambda \le \frac{\lambda}{4}, \end{aligned}$$

given $\alpha \le \frac{1}{9600}$.

On the other hand, for any $j \le 2$, any $r \in K_{ij}$, and any $i = 1, \cdots, L$, it holds

$$\|\mathbf{g}_{(i,r)}\|_2 \le \|\mathbf{a}_{(i)r}\|_2 \le \sqrt{k_T\mu} \le \frac{1}{\log n}\sqrt{\frac{\alpha m\theta}{\nu}},$$

given $k_T \le \alpha \frac{m\theta}{\mu\nu \log^2(nL)}$. Thus, combining with (F.12) again gives

$$\left| \frac{\sqrt{m}}{m_j} \mathbf{g}'_{(i,r)} \mathbf{q}_{(0)i} \right| \le \frac{m}{m_j} 4 \log^{-\frac{1}{2}}(nL) \sqrt{\frac{\alpha\theta}{L}} \le \frac{2}{5\sqrt{6}} \lambda \le \frac{\lambda}{4}, \tag{F.13}$$

given $\alpha \le \frac{1}{9600}$. Hence, we finish the proof. Notice that this bound requires (37) and (F.12) to hold simultaneously. □

⑤ Estimation of the total success probability.

So far, we have proved that ①, ②, ③, ④ hold with high probability, respectively. We want a success probability in recovering the true signal, which not only requires ①, ②, ③, ④ to hold simultaneously, but also requires (11), (12), Corollary 2, and Lemma 4 to succeed. From the above proofs, we have

- The bound ① is implied by (F.1) (holds with probability $1 - 4(nL)^{-1}$).
- The bound ② is implied by (37) (holds with probability $1 - 2(nL)^{-1}$) and (F.1).
- The bound ③ is implied by (37), (F.1) and (F.4) (holds with probability $1 - e^{\frac{1}{4}}(nL)^{-1}$)
- The bound ④ is implied by (37) and (F.12) (holds with probability $1 - 4(nL)^{-1}$).

Thus, we take a union bound to get

$$\begin{aligned} Pr\{① \cup ② \cup ③ \cup ④\} &\ge 1 - 4(nL)^{-1} - 2(nL)^{-1} - e^{-\frac{1}{4}}(nL)^{-1} - 4(nL)^{-1} \\ &= 1 - \left(10 + e^{\frac{1}{4}}\right)(nL)^{-1}. \end{aligned}$$

On the other hand, taking a union bound over (11), (12), Corollary 2, and Lemma 4 to find that they hold simultaneously with probability at least $1 - \left(6 + e^{\frac{1}{4}}\right)(nL)^{-2}$. Summarizing the above results, we know that the success probability in recovering the true signal and error matrices is at least $1 - (16 + 2e^{\frac{1}{4}})(nL)^{-1}$.

References

[1] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (Apr. 2006) 1289–1306.

[2] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2) (Feb. 2006) 5406–5425.

[3] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. Roy. Statist. Soc. Ser. B 68 (1) (Feb. 2007) 49–67.

[4] E.J. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (6) (Dec. 2008) 717–772.

[5] E. Candes, Mathematics of sparsity (and a few other things), in: Proceedings of the International Congress of Mathematicians, Seoul, South Korea, 2014.

[6] Y. Eldar, P. Kuppinger, H. Bölcskei, Block-sparse signals: uncertainty relations and efficient recovery, IEEE Trans. Signal Process. 58 (6) (June 2010) 3042–3054.

[7] M.E. Davis, Y.C. Eldar, Rank awareness in joint sparse recovery, IEEE Trans. Inform. Theory 58 (2) (Feb. 2012) 1135–1146.

[8] D. Malioutov, M. Çetin, A.S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (Aug. 2005) 3010–3022.

[9] X. Wei, Y. Yuan, Q. Ling, DOA estimation using a greedy block coordinate descent algorithm, IEEE Trans. Signal Process. 60 (12) (Dec. 2012) 6382–6394.

[10] F. Zeng, C. Li, Z. Tian, Distributed compressive spectrum sensing in cooperative multihop cognitive networks, IEEE J. Sel. Top. Signal Process. 5 (2) (Feb. 2011) 37–48.

[11] J. Meng, W. Yin, H. Li, E. Hossain, Z. Han, Collaborative spectrum sensing from sparse observations in cognitive radio networks, IEEE J. Sel. Areas Commun. 29 (2) (Feb. 2011) 327–337.

[12] J.A. Bazerque, G. Mateos, G.B. Giannakis, Group-lasso on splines for spectrum cartography, IEEE Trans. Signal Process. 59 (10) (Oct. 2011) 4648–4663.

[13] Z. Gao, L.F. Cheong, Y.X. Wang, Block-sparse RPCA for salient motion detection, IEEE Trans. Pattern Anal. Mach. Intell. 36 (10) (2014) 1975–1987.

[14] Y.C. Eldar, M. Mishali, Robust recovery of signals from a structured union of subspaces, IEEE Trans. Inform. Theory 55 (11) (Nov. 2009) 5302–5316.

[15] S. Oymak, A. Jalali, M. Fazel, Y.C. Eldar, B. Hassibi, Simultaneously structured models with application to sparse and low-rank matrices, preprint at http://arxiv.org/abs/1212.3753, 2014.

[16] E. Dall'Anese, J.A. Bazerque, G.B. Giannakis, Group sparse lasso for cognitive network sensing robust to model uncertainties and outliers, Phys. Commun. 5 (2) (June 2012) 161–172.

[17] M. Wang, Y. Li, X. Wei, Q. Ling, Robust group LASSO over decentralized networks, in: IEEE Proceedings GlobalSIP, Washington, DC, 2016.

[18] J. Wright, Y. Ma, Dense error correction via ℓ_1-minimization, IEEE Trans. Inform. Theory 56 (7) (July 2010) 3540–3560.

[19] X. Li, Compressed sensing and matrix completion with constant proportion of corruptions, Constr. Approx. 37 (1) (Feb. 2013) 73–99.

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

[21] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, J. ACM 58 (3) (May 2011) 11.

[22] A. Ganesh, K. Min, J. Wright, Y. Ma, Robust matrix decomposition with sparse corruptions, in: Proceedings of IEEE International Symposium on Information Theory, ISIT, Cambridge, USA, July 2012, pp. 1281–1285.

[23] Y. Chen, A. Jalali, S. Sanghavi, C. Caramanis, Low-rank matrix recovery from errors and erasures, IEEE Trans. Inform. Theory 59 (7) (July 2013).

[24] D. Gross, Recovering low-rank matrices from few coefficients in any basis, IEEE Trans. Inform. Theory 57 (3) (Mar. 2009) 1548–1566.

[25] J.A. Tropp, User-friendly tail bounds for matrix martingales, Found. Comput. Math. 12 (4) (Aug. 2012) 389–434.

[26] E.J. Candès, Y. Plan, A probabilistic and RIPless theory of compressed sensing, IEEE Trans. Inform. Theory 57 (2010) 7235–7254.

[27] R. Kueng, D. Gross, RIPless compressed sensing from anisotropic measurements, preprint at http://arxiv.org/abs/1205.1423.

[28] M. Ledoux, The Concentration of Measure Phenomenon, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.

[29] F. Rina, L. Mackey, Corrupted sensing: novel guarantees for separating structured signals, IEEE Trans. Inform. Theory 60 (2) (2014) 1223–1247.

[30] M.B. McCoy, J.A. Tropp, The achievable performance of convex demixing, arXiv preprint arXiv:1309.7478, 2013.

[31] R. Holger, Random sampling of sparse trigonometric polynomials, Appl. Comput. Harmon. Anal. 22 (1) (2007) 16–42.

[32] J. Li, et al., Robust frame based X-ray CT reconstruction, J. Comput. Math. 34 (6) (2016).