Missing Data Recovery for High-dimensional Signals with Nonlinear Low-dimensional Structures

Pengzhi Gao, *Student Member, IEEE*, Meng Wang, *Member, IEEE*, Joe H. Chow, *Fellow, IEEE*, Matthew Berger, and Lee M. Seversky

***Abstract*—Motivated by missing data recovery in power system monitoring, we study the problem of recovering missing entries of high-dimensional signals that exhibit low-dimensional nonlinear structures. We propose a novel model, termed as "union and sums of subspaces", to characterize practical nonlinear datasets. In this model, each data point belongs to either one of a few low-dimensional subspaces or the sum of a subset of subspaces. We propose convex-optimization-based methods to recover missing entries under this model. We theoretically analyze the recovery guarantee of our proposed methods with both noiseless and noisy measurements. Numerical experiments on synthetic data and simulated power system data are conducted to verify the effectiveness of the proposed methods.**

***Index Terms*—Missing data recovery, low-dimensional structure, union of subspaces, sum of subspaces**

I. Introduction

MANY practical datasets contain missing points. The problem of missing data recovery finds applications in collaborative filtering [1], computer vision [13], [52], machine learning [2], [3], remote sensing [46], system identification [33], power system monitoring [21], etc. The missing data recovery problem becomes extremely challenging when the signal of interest is high-dimensional. Due to the correlations in practical datasets, a collection of data points often exhibits low-dimensional structures. The low-dimensional structures have been utilized for various applications, such as dimensionality reduction and data compression [8], [16], [45], classification [34], [51], [54], [59], motion segmentation [43], geology [25], [27], and Internet topology estimation [18].

If the low-dimensional structure is linear, then all data points belong to one r-dimensional subspace in $\mathbb{R}^n$ with r much less than n. A data matrix that collects N ($N > n$) points as column vectors is rank r. The missing data recovery problem is reduced to a low-rank matrix completion problem [19], which has been intensively studied in recent years. A low-rank matrix can be correctly recovered as long as $O(rN\log^2 N)$[1] randomly selected entries are observed [10], [12], [23], [44].

Nonlinear structures have also been analyzed under the models such as the nonlinear manifold model [8], [9], [45], [47] and the union-of-subspaces (UoS) model [17], [57], [60], both of which have been studied in a variety of applications [17], [29], [32], [35], [36], [50], [53], [55], [58]. The manifold model can characterize highly nonlinear data but requires a large dimension to fit the entire dataset.

The UoS model directly extends the linear model and assumes that the data points belong to a union of k r-dimensional subspaces in $\mathbb{R}^n$. It is mostly studied in the subspace clustering problem [29], [32], [49], [50], where the objective is to estimate each of these k subspaces. Points not in one of the subspaces are treated as outliers. Although subspace clustering methods that can handle outliers exist [24], [49], [50], the corresponding theoretical analysis is based on the assumption that outliers are randomly distributed in the ambient space $\mathbb{R}^n$ [24], [49] and thus high-dimensional. This assumption might not hold for practical datasets, in which the corruptions might be near some subspaces and still exhibit low-dimensional structures.

The missing data recovery in datasets with low-dimensional nonlinear structures has received much less attention. Although convincing numerical results were reported for data recovery under the UoS model [11], [18], [60], [61], only Ref. [18] analyzed the theoretical limit of their recovery method in terms of the percentage of missing entries it successfully recovers. When the measurements are noiseless, Ref. [18] proves that one can recover the matrix from $O(rN\log^2 n)$ randomly observed entries even when the matrix is full rank.

One contribution of this paper is the development of a novel nonlinear model, termed as "union and sums of subspaces". Some data points belong to one of k r-dimensional subspaces, while the rest belong to the sums of a few subspaces. The additional sums of the subspaces model the corruptions that are low-dimensional and close to the subspaces of interest. Motivating applications include multiple event localization in environmental monitoring using wireless sensor networks [7] and multiple disturbance detection in power system monitoring [30], [56]. Each data point represents the observations of a device (e.g., sensors and phasor measurement units (PMUs) [39]) across time. Data points in the same subspace characterize the same local event, while data points in the sum of subspaces are influenced by multiple events.

This paper considers missing data recovery under this new model. Our major contributions include the development of data recovery methods and the theoretical characterization of

P. Gao, M. Wang and J. H. Chow are with the Dept. of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, 12180. Email: {gaop, wangm7, chowj}@rpi.edu. Phone: 518-276-3842. Fax: 518-276-6261.

M. Berger is with the University of Arizona, P.O. Box 210077, Tucson, AZ 85721. Email: matthewberger@email.arizona.edu. Phone: 520-621-0308.

L. M. Seversky is with Air Force Research Laboratory, Information Directorate, 26 Electronic Parkway, Rome, NY, 13441. Email: Lee.Seversky@us.af.mil. Phone: 315-330-2846. Fax: 315-330-4547.

Partial and preliminary results have appeared in [20].

[1]We use the notations $g(n) \in O(h(n))$, $g(n) \in \Omega(h(n))$, or $g(n) = \Theta(h(n))$ if as n goes to infinity, $g(n) \leq c \cdot h(n)$, $g(n) \geq c \cdot h(n)$ or $c_1 \cdot h(n) \leq g(n) \leq c_2 \cdot h(n)$ eventually holds for some positive constants c, c_1 and c_2 respectively.

the recovery limits. With noiseless measurements, we prove that $O(r^2 N \log^3 n)$ observations are sufficient to determine the missing points of an $n \times N$ matrix with columns belonging to the "union and sums" of k r-dimensional subspaces. This number is much less than the matrix dimension nN when r is small. We also analytically quantify the recovery error with respect to the noise level in the noisy measurements. Numerical experiments are conducted on synthetic data and simulated PMU data to verify the proposed methods.

The paper is organized as follows. We introduce the problem formulation in Section II and describe our methods and theoretical guarantees in Section III. All the proofs are deferred to the Appendix. Section IV records our numerical experiment. We conclude the paper in Section V.

II. Problem Formulation

A. Problem formulation and assumptions

We assume that k subspaces in $\mathbb{R}^n$, denoted by $S^1, ..., S^k$, are all at most r-dimensional ($r < n$) and pairwise linearly independent. Subspaces S^i and S^j are linearly independent if and only if $S^i \cap S^j = \{0\}$. Let S^{ij} denote the sum of two subspaces S^i and S^j ($i \neq j$). A vector $\mathbf{x}$ belongs to S^{ij} if and only if there exist $\mathbf{y}$ in S^i and $\mathbf{z}$ in S^j such that $\mathbf{x} = \mathbf{y} + \mathbf{z}$. We assume no subspace is embedded in the sum of other two subspaces, i.e., $S^i \subsetneq S^{i'j'}$, $\forall i', j' \neq i$. An example of union and sums of subspaces is shown in Fig. 1[2].

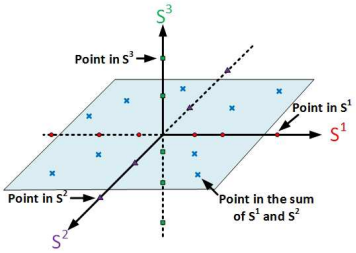

Fig. 1: Lines S^1, S^2, and S^3 are three one-dimensional subspaces in $\mathbb{R}^3$. x-marks stand for the points in S^{12}[2].

Consider a real-valued $n \times N$ matrix X . We say X is in the "union and sums" of $\{S^1, ..., S^k\}$ if its columns belong to either S^i or S^{ij} for $i, j \in \{1, ..., k\}$. Let U^i denote the orthonormal column span of S^i. Then for every column $\mathbf{x}$ of X, either $\mathbf{x} = U^i \boldsymbol{\alpha}$ or $\mathbf{x} = U^i \boldsymbol{\beta}_1 + U^j \boldsymbol{\beta}_2$, for some $i, j \in \{1, ..., k\}$, and $\boldsymbol{\alpha}, \boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2 \in \mathbb{R}^{r \times 1}$. We assume throughout the paper that $\|\boldsymbol{\alpha}\|_2 \leq 1$ and $\|\boldsymbol{\beta}_1\|_2^2 + \|\boldsymbol{\beta}_2\|_2^2 \leq 1$, i.e., every column of X has bounded norm.

The number of columns in S^i is $p_i N$, and the number of columns in S^{ij} but not S^i or S^j is $p_{ij} N$. Define $q_i = \sum_{j \neq i} p_{ij}$, $p_* = \min_i (p_i)$, and $q_* = \min_i (q_i)$. The total percentage of the columns in the sums of the subspaces is $p_c = \sum_{ij: i \neq j} p_{ij}$. The UoS model is a special case when $p_c = 0$.

[2]We consider the sum of two subspaces for notational simplicity here. The analysis can be easily extended to a sum of a constant number of subspaces. Further, the results in this paper can be directly extended to complex vectors.

Let $\xi \in \mathbb{R}^{n \times N}$ denote the measurement noise. We assume that the entries of ξ are uniformly distributed in $[-\epsilon_1/\sqrt{n}, \epsilon_1/\sqrt{n}]$ for some constant ϵ_1. Let

$$\tilde{X} = X + \xi \tag{1}$$

denote the measurement matrix. For a column $\tilde{\mathbf{x}}$ in $\tilde{X}$, we call the corresponding column $\mathbf{x}$ in X as its "noiseless counterpart". Each entry of $\tilde{X}$ is observed independently with probability p_0. Let Ω denote the set of observed entries. Let

$$(\tilde{X}_\Omega)_{ij} = \begin{cases} \tilde{X}_{ij}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{if } (i,j) \notin \Omega. \end{cases} \tag{2}$$

The problem this paper addresses is: given $\tilde{X}_\Omega$, how shall we determine X and S^i's accurately?

Some notations and assumptions are introduced to facilitate the discussion. Given a vector $\mathbf{x}_0 \in \mathbb{R}^n$ and a positive constant ϵ, the ball centered at $\mathbf{x}_0$ with radius ϵ is denoted as

$$B_{\mathbf{x}_0, \epsilon} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}_0\|_2 \leq \epsilon\}. \tag{3}$$

We define S^i_ϵ as the subset of points in S^i that are at least ϵ distance away from all $S^{i'j'}$'s, $\forall i', j' \neq i$. The coherence of an s-dimensional subspace S is defined as

$$\mu(S) := \frac{n}{s} \max_j \|\mathcal{P}_S \mathbf{e}_j\|_2^2, \tag{4}$$

where $\mathbf{e}_j$'s are the canonical unit vectors for $\mathbb{R}^n$, and $\mathcal{P}_S$ is the projection operator onto S. Let ω denote the indices of observed entries in $\mathbf{x}$. $\mathbf{x}_\omega \in \mathbb{R}^{|\omega|}$ denotes the subvector of $\mathbf{x}$. The projection of a partially observed column $\mathbf{x}_\omega$ to S (with orthonormal matrix denoted by U) is defined as

$$\mathcal{P}_{\omega, S} = U_\omega ((U_\omega)^T U_\omega)^\dagger (U_\omega)^T, \tag{5}$$

where $\dagger$ denotes the Moore-Penrose pseudoinverse[3]. U_ω denotes the submatrix of U with row indices in ω. The partial distance of two partially observed columns $\mathbf{x}_1$ and $\mathbf{x}_2$ are defined as $\sqrt{n/q} \|(\mathbf{x}_1 - \mathbf{x}_2)_{\tilde{\omega}}\|_2$, where $\tilde{\omega}$ is the common set of indices of size q that both $\mathbf{x}_1$ and $\mathbf{x}_2$ are observed. The coherence of the vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\mu(\mathbf{x}) := n \|\mathbf{x}\|_\infty^2 / \|\mathbf{x}\|_2^2. \tag{6}$$

Given matrix M, $\|M\|_*$ denotes its Nuclear Norm, which is the sum of singular values. $\|M\|_2$ denotes the largest singular value. $\|M\|_{1,2}$ denotes the sum of the l_2-norms of its columns. $\|M\|_\infty$ denotes the largest absolute value of its entries.

We have three technical assumptions for the analysis.

A1 (Subspace coherence). $\|\mathbf{e}_j^T U^i\|_2 \leq \sqrt{\mu_1 r/n}$, $\forall i \in \{1, ..., k\}$, $\forall j \in \{1, ..., n\}$, for some $\mu_1 > 0$.

A2 (Subspace separation). $\|(U^i)^T U^j\|_2 \leq \theta$, $\forall i, j \in \{1, ..., k\}$ $(i \neq j)$, for some constant $\theta < 1/3$.

θ characterizes the separations of the subspaces. When S^i's are sufficiently separated, θ is close to zero. One can infer from **A1** and **A2**[4] that:

$$\mu(S^i) \leq \mu_1, \forall i \in \{1, ..., k\}, \tag{7}$$

[3]Given the singular value decomposition of $A = U \Sigma V^*$, we have $A^\dagger = V \Sigma^\dagger U^*$. For a rectangular diagonal matrix such as Σ, we compute the pseudoinverse by taking the reciprocal of each non-zero entry on the diagonal and transposing the matrix.

[4]**A1** and **A2** imply (A2) in [18]. We start with assumptions on U^i's because it is more straightforward to analyze the coherence properties this way.

$$\mu(S^{ij}) \leq \mu_1/(1-\theta), \forall i,j \in \{1,...,k\}, \quad (8)$$

$$\mu(\mathbf{x}_i - \mathbf{x}_j) \leq 4\mu_1 r/(1-3\theta), \forall i,j \in \{1,...,N\},^5 \quad (9)$$

$$\text{and } \|X\|_\infty \leq \sqrt{2\mu_1 r/n}. \quad (10)$$

For successful subspace estimation and data recovery, we need another assumption that each subspace S^i contains a fraction of points that are close to each other and sufficiently away from other subspaces. Our data recovery method essentially relies on these points to identify S^i. Since points in S^{ij}'s can be close to S^i and thus degrade the subspace estimation performance, an upper bound of the number of these points is necessary. The assumption is formally stated as follows.

A3 (Continuous distribution of the points).

Given any $\epsilon_0 \in (0, \sqrt{1-\frac{2\theta^2}{(1-\theta)^2}} - 2\sqrt{2}\epsilon_1)$, define

$$\sigma_1 := (\epsilon_0 - 2\sqrt{2}\epsilon_1)/\sqrt{3}, \text{ and } \sigma_2 := \epsilon_0 + 2\sqrt{2}\epsilon_1. \quad (11)$$

There exist constants v_0, v_1, v_2, and v_3 in $(0,1]$, depending on ϵ_0 and ϵ_1, such that:

(i) The percentage of points in $S^i_{\sigma_2}$ with respect to all the columns in X is at least $v_0 p_i$ (*enough points in S^i are not too close to other subspaces*).

(ii) Given $\mathbf{x}$ in $S^i_{\sigma_2}$, the percentage of points in both $B_{\mathbf{x},\sigma_1}$ and S^i with respect to all the columns in X is at least $v_1 \sigma_1^r p_i$ (*these points in S^i are sufficiently close*), and the percentage of points in both $B_{\mathbf{x},\sigma_2}$ and any of S^{ij}'s ($j \neq i$) but not S^i with respect to all the columns in X is between $v_2(\frac{\sigma_2}{1+\theta})^{2r} q_i$ and $v_3(\frac{\sigma_2}{1-\theta})^{2r} q_i$ (*the number of points in S^{ij} that are close to $\mathbf{x}$ is bounded*).

A3 holds if the columns are drawn from continuous distributions such as uniform distributions from the constrained balls in S^i's and S^{ij}'s for all i,j. If S^i's are sufficiently separated, i.e., θ is small, v_0 and v_1 will not be too close to 0. In Appendix B, we consider the special case that points are uniformly distributed in the corresponding balls of their respective subspaces (i.e., $\boldsymbol{\alpha}$ is uniformly distributed in the unit ball of $\mathbb{R}^r$, and $\boldsymbol{\beta}$ is uniformly distributed in the unit ball of $\mathbb{R}^{2r}$ for $\boldsymbol{\beta}^T = [\boldsymbol{\beta}_1^T \ \boldsymbol{\beta}_2^T]$) and compute these parameters explicitly.

To differentiate noiseless ($\epsilon_1 = 0$) and noisy cases, we use constants v_0, v_1, v_2, and v_3 for the former and v_0', v_1', v_2', and v_3' for the latter case in the analysis.

B. Motivation and connections to existing works

Our problem formulation is motivated by missing PMU data recovery in power system monitoring. PMUs provide synchronized voltage and current phasor measurements in the power system at a rate of 30 samples per second [39]. Data losses constantly happen due to the communication congestions or device malfunctions [48]. The data quality issue limits the real-time implementation of PMU data in power system monitoring. Exploiting the low-dimensionality in the high-dimensional PMU data, Ref. [21] the first time connected missing PMU data recovery with low-rank matrix completion. Since disturbances happen in power systems constantly [30], the rank of the PMU data matrix increases significantly when the system experiences multiple disturbances. Our proposed model characterizes the structures of the measurements under multiple disturbances. Each column of $\tilde{X}$ represents the observations of a bus voltage or a line current across time. One disturbance may affect nearby voltages and currents, and the corresponding columns belong to the same low-dimensional subspace. Quantities affected by multiple overlapping disturbances belong to the sum of subspaces. Similarly, our model also finds applications in environmental monitoring using sensor networks [7], where sensor measurements affected by the same event belong to the same subspace, and measurements affected by multiple events belong to the sum of subspaces.

Low-rank matrix completion theory guarantees that an $n \times N$ rank-r matrix can be correctly recovered provided that $O(rN\log^2 N)$ entries are observed. High-rank matrix completion extends the recovery to matrices with column vectors belonging to one of k r-dimensional subspaces in $\mathbb{R}^n$ (kr could be larger than n) [11], [18], [60], [61]. Only Ref. [18] provides the theoretical guarantee that if $O(rN\log^2 n)$ noiseless entries are observed, its recovery method completes the high-rank matrix accurately.

When some columns belong to the union and sums of the subspaces, one can generalize the method in [18] by treating each sum of subspaces as a different subspace. This direct generalization, however, significantly increases the number of subspaces ($k(k-1)/2$ subspaces in this case) and in turn increases the computational time and the required number of columns in X for data recovery, both of which scale at least linearly in the number of subspaces. Furthermore, the direct extension only succeeds under a strong assumption that a sufficient number of the points in the sum of the subspaces are clustered and away from all the other subspaces.

Instead of recovering a sum of some subspaces as a distinctive subspace, we focus on only identifying k atomic subspaces and treat the points in the sums of the subspaces as outliers. Different from [18], we solve a robust matrix completion problem to recover each of the k subspaces (Subroutine 2). The points in the sum of subspaces are then recovered through projections to multiple atomic subspaces. Since we only solve for k subspaces, both the computational time and the required number of columns are significantly reduced. The numerical superiority of our method is demonstrated in Section IV. Furthermore, we derive the theoretical bounds of the recovery error when the measurements are noiseless (Theorem 1) and noisy (Theorem 2), while the recovery performance is only analyzed for noiseless measurements in [18].

III. Missing data recovery method and the performance guarantee

A. Missing Data Recovery Method

We first summarize the key idea of our proposed missing data recovery method. The parameters used in the method will be specified in Sections III-B and III-C. First, we pick s_τ columns ($\tau = 0$ for noiseless case and $\tau = 1$ noisy case), referred to as seeds, uniformly at random from X_Ω and select $l_\tau n$ nearest neighbors to form a local neighbor

[5] Explanations see Appendix A.

Subroutine 1 Local Neighbor Matrix Selection

Input: s_τ, l_τ, ζ_τ, t_τ and ϵ_0 ($\tau = 0$ for noiseless formulation and $\tau = 1$ for noisy formulation).

(i) Choose s_τ columns uniformly at random as seeds and discard all with less than ζ_τ observations.

(ii) For each seed, find all columns with at least t_τ observations at locations observed in the seed, and randomly select $l_\tau n$ columns from each such set.

(iii) Form a candidate set by selecting the columns with partial distance less than $\epsilon_0/\sqrt{2}$ from each seed among the $l_\tau n$ columns selected in Step (ii).

(iv) If less than $2n$ columns are selected in Step (iii), form the local neighbor matrix directly. Otherwise, form the local neighbor matrix by randomly selecting $2n$ columns from the candidate set.

Return: s' ($s' \leq s_\tau$) neighbor matrices $\{M_\Omega^1, ..., M_\Omega^{s'}\}$.

Subroutine 2 Robust Subspace Estimation for M_Ω^i (Noiseless)

Input: Partially observed local neighbor matrix M_Ω^i, λ_i. Find (L^{i^*}, C^{i^*}), the optimum solution to the following optimization problem.

$$\min_{L,C} \ \|L\|_* + \lambda_i \|C\|_{1,2} \quad \text{s.t. } L_\Omega + C_\Omega = M_\Omega^i \tag{12}$$

Compute the SVD of $L^{i^*} = U^{i^*}\Sigma^{i^*}(V^{i^*})^T$.

Return: The subspace S^{i^*}, represented by U^{i^*}.

matrix M_Ω^i for each seed (Subroutine 1). We then estimate the corresponding subspace spanned by points in each M_Ω^i. A set of candidate subspaces are computed in this step (Subroutine 2 for noiseless case or Subroutine 4 for noisy case). Lastly, we refine the subspace estimation from the candidate subspaces, and the remaining subspaces after refinement, denoted by $\{\hat{S}^1, ..., \hat{S}^k\}$, are used as the estimation of $\{S^1, ..., S^k\}$. To recover the missing points, we assign each partially observed column $\mathbf{x}_\omega$ to its closest $\hat{S}^i$. If the closest distance is beyond a given threshold, which is set to be the expected noise level here, then assign $\mathbf{x}_\omega$ to the closet $\hat{S}^{ij}$. Estimate $\mathbf{x}$ by its closest point in $\hat{S}^i$ or $\hat{S}^{ij}$ (Subroutine 3).

Our method is built upon the method in [18] and differs from it mainly in the subspace estimation part. We discuss the differences in the noiseless case, and the discussion directly generalizes to the noisy case. The key of the method is to select enough seeds such that at least one seed $\mathbf{x}$ lies in $S_{\epsilon_0}^i$ for each i. One then recovers S^i based on a group of partially observed columns that are at most ϵ_0 away from $\mathbf{x}$. Under the UoS model, provided that enough seeds are selected, there always exists a good local neighbor matrix that only contains points in S^i for every i. Nuclear Norm Minimization is employed in [18] to recover S^i from the neighbor matrix with missing data. Under our model, however, no matter how large ϵ_0 is, there may always exist points in S^{ij} that are close to a point $\mathbf{x}$ in $S_{\epsilon_0}^i$. Therefore, some columns in the neighbor matrix M_Ω^i do not belong to S^i. We employ (12) to recover S^i, since it can correctly recover the low-dimensional subspace even when some columns do not belong to the same subspace [14].

One major contribution of this paper is the theoretical

Subroutine 3 Subspace Refinement & Missing Data Recovery

Input: s ($s \leq s'$) candidate subspaces $\{S^{1^*}, ..., S^{s^*}\}$. Partially observed matrix $\tilde{X}_\Omega$.

Sort $\{S^{1^*}, ..., S^{s^*}\}$ in the increasing order of the rank, denoted as $S^{(1)}, ..., S^{(s)}$. Define $\hat{S}^1 := S^{(1)}$, and set j to be one.

for $i = 2, ..., s$ **do**

if $S^{(i)}$ is not in the span of $\{\hat{S}^1, ..., \hat{S}^j\}$ and $j \leq k$ **then**

$j \leftarrow j + 1$ and $\hat{S}^j := S^{(i)}$.

end if

end for

Project each partially observed column $\mathbf{x}_\omega$ to its nearest subspace by computing

$$i^* = \arg\min_i (\|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^i}\mathbf{x}_\omega\|_2). \tag{13}$$

if $\|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^{i^*}}\mathbf{x}_\omega\|_2^2 < |\omega|\epsilon_1^2/(3n)$ **then**

$\mathbf{x}_\omega$ belongs to S^i. Estimate $\mathbf{x}$ by

$$\hat{\mathbf{x}} = \hat{U}^{i^*}((\hat{U}_\omega^{i^*})^T\hat{U}_\omega^{i^*})^\dagger(\hat{U}_\omega^{i^*})^T\mathbf{x}_\omega. \tag{14}$$

else

$\mathbf{x}_\omega$ belongs to the sum of two subspaces. Find

$$(i^*, j^*) = \arg\min_{i,j} (\|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^{ij}}\mathbf{x}_\omega\|_2). \tag{15}$$

Estimate $\mathbf{x}$ by $\hat{\mathbf{x}} = \hat{U}^{i^*j^*}((\hat{U}_\omega^{i^*j^*})^T\hat{U}_\omega^{i^*j^*})^\dagger(\hat{U}_\omega^{i^*j^*})^T\mathbf{x}_\omega$.

end if

Return: $\{\hat{S}^1, ..., \hat{S}^k\}$ and $\hat{X}$.

Subroutine 4 Robust Subspace Estimation for M_Ω^i (Noisy)

Input: Partially observed local noisy neighbor matrix M_Ω^i, coefficients $\hat{\lambda}_i$ and $\tilde{\lambda}_i$. Find (L^{i^*}, C^{i^*}), the optimum solution to the following optimization problem.

$$\begin{aligned} &\min_{L,C} \frac{1}{|\Omega|}\|(M^i - L - C)_\Omega\|_F^2 + \hat{\lambda}_i\|L\|_* + \tilde{\lambda}_i\|C\|_{1,2} \\ &\text{s.t. } \|L\|_\infty \leq \sqrt{2\mu_1 r/n} \text{ and } \|C\|_\infty \leq \sqrt{2\mu_1 r/n} \end{aligned} \tag{16}$$

Compute the SVD of $L^{i^*} = U^{i^*}\Sigma^{i^*}(V^{i^*})^T$.

Return: The subspace S^{i^*}, represented by U^{i^*}.

performance analysis of this data recovery method. Roughly speaking, we show that the recovery method can recover a significant amount of the missing entries even when many columns of the matrix lie in the sum of subspaces. The results are formally summarized in the Theorems 1 and 2.

B. Performance analysis of noiseless measurements

We first consider the scenario that there is no noise ($\epsilon_1 = 0$) in the measurements. We define

$$s_0 := \lceil \frac{\log k + \log\frac{1}{\delta_0}}{(1 - \exp(-\frac{\eta_1\mu_1^4 r^2 \log^3(4\beta_0 n)}{4}))v_0 p_*} \rceil, \tag{17}$$

$$l_0 := \lceil \max(\frac{2}{v_1(\frac{\epsilon_0}{\sqrt{3}})^r p_*}, \frac{8\log(2s_0/\delta_0)}{nv_1(\frac{\epsilon_0}{\sqrt{3}})^r p_*}, \frac{12\log(2s_0/\delta_0)}{nv_2(\frac{\epsilon_0}{1+\theta})^{2r} q_*}) \rceil, \tag{18}$$

$$\zeta_0 := \eta_1\mu_1^4 r^2 \log^3(4\beta_0 n), \text{ and } t_0 := \lceil \frac{128\mu_1^2 r^2}{(1-3\theta)^2} \log(\frac{2s_0 l_0 n}{\delta_0}) \rceil, \tag{19}$$

where $\delta_0 := cn^{-5}$, c and η_1 are positive constants. Constant $\beta_0 \in [1,2]$ denotes the maximum ratio of the number of columns in the neighbor matrices (Subroutine 1) to n.

Theorem 1 establishes the requirements on the data observation rate p_0, the number of seeds s_0, the number of columns of the neighbor matrix $l_0 n$, and the ratio q_i/p_i of the number of points in the sums of subspaces to the number of points in S_i. Intuitively, as long as the number of observed entries is not too small, the number of columns in X is not too small, and the number of columns in the sum of the subspaces is not too large, our recovery method can successfully estimate the subspaces and recover the missing points. Formally, we have

Theorem 1. *Suppose that $n \geq 32$ and each entry of X is observed independently with probability $p_0 \geq \underline{p}$. If*

$$\underline{p} \geq 2\eta_1 \mu_1^2 r^2 \log^3(4\beta_0 n)/n, \tag{20}$$

$$N \geq 2l_0 n(6s_0 l_0 n/\delta_0)^{-\frac{72\mu_1^2 r^2}{(1-5\theta)^2}\log \underline{p}}, \tag{21}$$

$$\frac{q_i}{p_i} \leq \frac{\eta_2 v_1 (1-\theta)^{2r}}{36 v_3 (\sqrt{3}\epsilon_0)^r} \frac{\underline{p}^2}{(1+\frac{2\mu_1 r}{\underline{p}\sqrt{n}})^2 \mu_1^3 r^3 \log^6(4\beta_0 n)}, \ \forall i, \tag{22}$$

for some η_1, $\eta_2 > 0$ and $\beta_0 \in [1,2]$, subspaces $\{S^1, ..., S^k\}$ are correctly recovered with probability at least $1-(6+s_0)\delta_0$. Furthermore, with probability at least $1-(4k(k-1)+2)\delta_0$, every fixed column of X is correctly recovered.

Note that μ_1, η_1, ϵ, v_0, v_1, v_2, v_3, θ, and β_0 are all constants. To understand the implication of Theorem 1, we ignore constants and analyze the requirements in terms of orders. Consider the most challenging data recovery case that $k = \Theta(n)$ and $p_* = \Theta(\frac{1}{n})$ such that X can be full-rank. From (17), the required number of seeds $s_0 = \Theta(n \log n)$. Since δ_0 is selected to be $\Theta(n^{-5})$, the probabilities of recovery success $1-(6+s_0)\delta_0$ and $1-(4k(k-1)+s_0+8)\delta_0$ in Theorem 1 tend to one when n increases. From (20), we have $\underline{p} = \Theta(r^2 \log^3 n/n)$. Then our method only requires $\Theta(r^2 N \log^3 n)$ entries to estimate the remaining ones. From (18), $l_0 = \Omega(n)$. Then the number of columns in each neighbor matrix is $\Omega(n^2)$. From (19), $\zeta_0 = \Theta(r^2 \log^3 n)$, and $t_0 = \Theta(r^2 \log(n^8 \log n))$.

From (21), $N = \Omega(n^2 (n^8 \log n)^{\log n - 3\log(\log n)})$, which is super-polynomial in n. The information-theoretical limit of the required number of columns to recover missing data is proven to be $O(krn)$ for the UoS model of k r-dimensional subspaces in $\mathbb{R}^n$ [40], [42]. The gap between our bound and the theoretical limit results from the proof artifacts inherited from [31]. Specifically, the partial distance is computed to estimate the actual distance between any two points. The analysis is based on sufficient but not necessary conditions for this estimation to be approximately accurate. To see this, consider the case that no missing points exist ($\underline{p} = 1$), N then reduces to $\Omega(n^{10} \log n)$, which is polynomial in n. Moreover, we observe from numerical experiments that the method performs well in practice even though these quantities are much less than the bounds provided.

Furthermore, (22) indicates our method can tolerate points in S^{ij}'s as long as

$$\frac{q_i}{p_i} = O(\frac{v_1 (1-\theta)^{2r} r^3 \log^6 n}{v_3 (\sqrt{3}\epsilon_0)^r n^3}), \forall i. \tag{23}$$

θ is small if the subspaces are sufficiently separated. v_1/v_3 is small for given ϵ_0 if under a certain distribution of the points, the ratio of the number of points in S^i to the number of points in S^{ij} and close to S^i is small. In this case, (23) indicates that a high ratio of points in S^{ij}'s can be tolerated.

(12) is convex and can be solved by off-the-shelf softwares such as CVX [22]. Other fast algorithms such as Augmented Lagrange Multiplier (ALM) method [31] also exist. The computational complexity of our method is dominated by solving (12) s_0 times for s_0 seeds. When $k = \Theta(n)$ and $p_* = \Theta(\frac{1}{n})$, we have $s_0 = \Theta(n \log n)$. In contrast, if we directly extend the method in [18] and treat each S^{ij} as a separate subspace, the required number of seeds is $\Theta(n^2 \log n)$. Thus, the computational complexity of the naive extension is significantly higher, in addition to its doubtable recovery performance.

C. Performance analysis of noisy measurements

When the observed measurement matrix contains noise ($\epsilon_1 \neq 0$), Subroutine 2 is replaced with Subroutine 4. We include the constraints on $\|L\|_\infty$ and $\|C\|_\infty$ because $\|X\|_\infty$ is bounded by $\sqrt{2\mu_1 r/n}$. If $\sqrt{2\mu_1 r/n}$ is unknown, we replace it with the largest absolute value of the observed entries in the matrix in (16). The parameters are defined as

$$s_1 := \lceil \frac{\log k + \log \frac{1}{\delta_1}}{(1-\exp(-\frac{\log((\beta+1)n)\max(\frac{\beta}{\alpha_1}, \frac{\log^2 n}{\alpha_2})}{\beta}))v_0' p_*} \rceil, \tag{24}$$

$$l_1 := \lceil \max(\frac{2}{v_1' \sigma_1^r p_*}, \frac{8\log(2s_1/\delta_1)}{n v_1' \sigma_1^r p_*}, \frac{12\log(2s_1/\delta_1)}{n v_2' (\frac{\sigma_2}{1+\theta})^{2r} q_*}) \rceil, \tag{25}$$

$$\zeta_1 := 4\log((\beta+1)n)\max(\frac{1}{\alpha_1}, \frac{\log^2 n}{\beta\alpha_2}), \tag{26}$$

$$\text{and} \quad t_1 := \lceil \frac{128\mu_1^2 r^2}{(1-3\theta)^2} \log(\frac{2s_1 l_1 n}{\delta_1}) \rceil, \tag{27}$$

where $\delta_1 := \frac{6}{(\beta+1)n}$, and constants $\alpha_1, \alpha_2 \geq 1$. The dimension of M_Ω^i is denoted by $n \times \beta_i n$. Let $\beta = \max_i \beta_i$, and $\beta \in [1,2]$.

Theorem 2 and Lemma 1 are counterparts of Theorem 1 when the measurements contain noise. The recovery is not exact due to the noise, but the error can be quantified as a function of the noise level and other parameters of the recovered method. Moreover, we show that the recovery error diminishes as n increases with appropriately chosen parameters. Formally, we have

Theorem 2. *Suppose that each entry of $\tilde{X}$ is observed independently with probability $p_0 \geq \underline{p}$. If*

$$\underline{p} \geq \frac{8\log((\beta+1)n)}{n}\max(\frac{1}{\alpha_1}, \frac{\log^2 n}{\beta\alpha_2}), \tag{28}$$

and

$$N \geq 2l_1 n(6s_1 l_1 n/\delta_1)^{-\frac{72\mu_1^2 r^2}{(1-5\theta)^2}\log \underline{p}}, \tag{29}$$

then with probability at least $1-(6+s_1)\delta_1$, it holds that

$$\frac{\|L^{i*} - L^i\|_F^2 + \|C^{i*} - C^i\|_F^2}{\beta_i n^2}$$

$$\leq \frac{2\mu_1 r\kappa}{n(1+\kappa)} + 3\alpha_3\phi^2\log((\beta_i+1)n)(1+\kappa)^2(\frac{2r}{np_0}+\frac{3\kappa}{1+\kappa}) \quad (30)$$

for all the neighbor matrices of the seeds with noiseless counterparts in $S^i_{\sigma_2}$, $i = 1, ..., k$, where

$$\kappa = 9v_3'(\frac{\sigma_2}{1-\theta})^{2r}q_i/(v_1'\sigma_1^r p_i), \quad (31)$$

L^{i*} and C^{i*} are the solutions returned by Subroutine 4, constants $\alpha_1 \geq 1$, $\alpha_2 \geq 1$, $\alpha_3 > 0$, $\beta \in [1, 2]$, and $\phi = \max(\frac{\epsilon_1}{\sqrt{n}}, \sqrt{2\mu_1\frac{r}{n}})$.

Theorem 2 establishes the error bound of the recovered matrix from partial and noisy measurements. The left-hand side of (30) represents the recovery error of a local neighbor matrix averaged over all the entries. Since $\|X\|_\infty = \sqrt{2\mu_1 r/n}$, the absolute value of each entry in X is $\Theta(\sqrt{r/n})$ on average. Then as long as the right hand side of (30) is $O(1/n)$, Theorem 2 indicates that the entry-wise relative recovery error diminishes to zero when n increases. The estimation error of S^i would in turn reduce to zero. Then with the same arguments as for the noiseless case (last paragraph of Appendix D (2)), we can conclude that Subroutine 3 returns $\{\hat{S}^1, ..., \hat{S}^k\}$ as the estimation of $\{S^1, ..., S^k\}$ with diminishing errors.

For example, consider the case that the number of subspaces $k = \Theta(n^t)$, and $p_* = \Theta(n^{-t})$ correspondingly for some constant t in $(0, 1)$. We ignore the constants and focus on the orders. μ_1, σ_1, σ_2, v_0', v_1', v_2', v_3', θ, α_1, α_2, and β are all constants. From (24)-(27), we have $s_1 = \Theta(n^t \log n)$, $l_1 = \Theta(n^t)$, $\zeta_1 = \Theta(\log^3 n)$, and $t_1 = \Theta(r^2 \log n)$. The probability of recovery success $1 - (6 + s_1)\delta_1$ goes to 1 as n increases to infinity. The right side of (28) is $O(\frac{\log^3 n}{n})$. If we choose p_0 as $O(r^2 \log^3 n/n)$ then $p_0 \geq \underline{p}$ is met, and the number of observed entries is $O(r^2 N \log^3 n)$, which is much less than nN. With this p_0, as long as

$$\frac{q_i}{p_i} \leq \frac{v_1'\sigma_1^r(1-\theta)^{2r}}{9v_3'\sigma_2^{2r} r \log^3 n}, \forall i, \quad (32)$$

from (31), $\kappa = O((r \log^3 n)^{-1})$. Then the right-hand side of (30) is $O((n \log^2 n)^{-1})$. Thus, the entry-wise relative recovery error decreases to zero at a rate of $1/\log n$.

The following lemma provides the guarantee of missing data recovery for a fixed column $\tilde{\mathbf{x}}_\omega$ of the matrix. If its noiseless counterpart $\mathbf{x}$ belongs to S^{ij}, we assume $\frac{\|\tilde{\mathbf{x}}-P_{\hat{S}^{ij}}\tilde{\mathbf{x}}\|_2^2}{\|\tilde{\mathbf{x}}-P_{\hat{S}^{i'j'}}\tilde{\mathbf{x}}\|_2^2} \leq \epsilon_3$ holds for some positive ϵ_3 for every $(i', j') \neq (i, j)$. Let $\rho_1\mu_1$ denote the coherence of subspaces $\hat{S}^{ij}$'s. Let μ_2 denote the coherence of $\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2$ for pairs of columns $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ in $\tilde{X}$. We assume both ρ_1 and μ_2 are constants.

Lemma 1. *For any $\delta_1' < 1$, if p_0 satisfies*

$$p_0 \geq \frac{4r\rho_1\mu_1(1+b)^2}{n(1-c)((1-a)-(1+a)\epsilon_3)} \quad (33)$$

for some $\epsilon_3 > 0$, where $a = \sqrt{\frac{4\mu_2^2}{np_0}\log(\frac{1}{\delta_1'})}$, $b = \sqrt{2\mu_2\log(\frac{1}{\delta_1'})}$, and $c = \sqrt{\frac{32r\rho_1\mu_1}{3np_0}\log(\frac{4r}{\delta_1'})}$, then with probability at least $1-(2(k-1)(k^2-2k+4)+2)\delta_1'$, if the column of interest $\tilde{\mathbf{x}}$ is closest to $\hat{S}^i$, for any $i' \neq i$,

$$\|\tilde{\mathbf{x}}_\omega - \mathcal{P}_{\omega,\hat{S}^{ij}}\tilde{\mathbf{x}}_\omega\|_2 < \|\tilde{\mathbf{x}}_\omega - \mathcal{P}_{\omega,\hat{S}^{i'j'}}\tilde{\mathbf{x}}_\omega\|_2, \ \forall j, j'; \quad (34)$$

if $\tilde{\mathbf{x}}$ is closest to $\hat{S}^{ij}$, for any $(i', j') \neq (i, j)$,

$$\|\tilde{\mathbf{x}}_\omega - \mathcal{P}_{\omega,\hat{S}^{ij}}\tilde{\mathbf{x}}_\omega\|_2 < \|\tilde{\mathbf{x}}_\omega - \mathcal{P}_{\omega,\hat{S}^{i'j'}}\tilde{\mathbf{x}}_\omega\|_2. \quad (35)$$

Pick $\delta_1' = n^{-4}$, then $1-(2(k-1)(k^2-2k+4)+2)\delta_1'$ goes to one when n increases. Lemma 1 guarantees that each column of $\tilde{X}_\Omega$ is assigned to its closest subspace with high probability. When $p_0 = \Theta(r \log^2 n/n)$, we have $a = \Theta(\sqrt{1/(r \log n)})$, $b = \Theta(\sqrt{\log n})$, and $c = \Theta(\sqrt{1/(\log n)})$. Then as long as $1/(1-\epsilon_3) = O(\log n)$, the right hand side of (33) should be $O(r \log^2 n/n)$. Then (33) is met when n is sufficiently large.

IV. Simulation

We explore the performance of our method on both synthetic data and simulated PMU data. We use Augmented Lagrange Multiplier (ALM) method [31] to solve Subroutine 2 and use CVX [22] to solve Subroutine 4. We use power system toolbox (PST) [15] to generate the simulated PMU data based on the linear model of IEEE 39-bus New England Power System [4]. The recovery performance is measured by the relative recovery error $\|X - X_{\text{rec}}\|_F/\|X\|_F$, where matrix X represents the actual data, and X_{rec} represents the recovered data matrix. The average erasure rate p_{avg} is the percentage of missing entries. All results are averaged over 100 runs.

We compare the performance with two missing data recovery methods: a low-rank matrix completion method called Singular Value Thresholding (SVT) and a high-rank matrix completion method in [18] denoted by HRMC in this section. We treat S^{ij} as a separate subspace when implementing HRMC. We also compare the performance with four subspace clustering methods: GSSC [41], SSC [61], Ksub [6], and rMiCUSaL [60].

A. Performance on synthetic data

1) Noiseless formulation: The key parameters of X are chosen as follows: $n = 100$, $N = 5000$, $k = 10$, and $r = 10$. Note that matrix X is full rank. We generate the i^{th} subspace S^i by the random orthonormal basis matrix $U^i \in \mathbb{R}^{n\times r}$. For each column $\mathbf{x} \in S^i$, we generate a random Gaussian vector $\boldsymbol{\alpha} \in \mathbb{R}^r$. We then set $\mathbf{x} = U^i\boldsymbol{\alpha}$. For each column $\mathbf{x} \in S^{ij}$, we generate random Gaussian vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^r$ and set $\mathbf{x} = U^i\boldsymbol{\alpha}_1 + U^j\boldsymbol{\alpha}_2$. We set $p_i = (1 - p_c)/k$ and $p_{ij} = 2p_c/(k(k-1))$, $\forall i \neq j$.

We simulate the observed measurement matrix according to (1) with $\xi = 0$. We delete some entries of X uniformly at random and vary the average erasure rate p_{avg}. We apply Subroutine 2 to obtain the corresponding subspace of each local neighbor matrix. We use 100 seeds to implement HRMC and our method. λ in (12) is set to be 1.

Fig. 2 shows the recovery performance of SVT, HRMC, and our method for different p_c. p_c does not affect the rank of X. Thus, SVT has the similar performance for different p_c. The recovery performance of our method is generally better than SVT and HRMC when p_{avg} is below 0.4. For instance, when 15% of the entries are lost, and 10% of the columns belong to the sums of subspaces, the relative recovery error of our method is within 0.02, while the recovery errors of SVT and HCMC are above 0.26, resulting in failed recoveries.

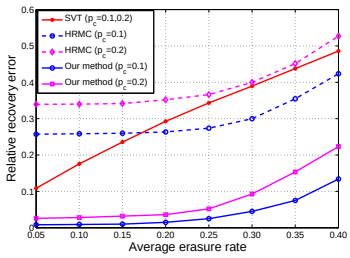

Fig. 2: Relative recovery errors of SVT, HRMC and our method for various values of p_c.

Fig. 3 shows the recovery performance of these k subspaces by our method, measured by the average separation of every pair of recovered and original subspaces. We define

$$\psi_{\text{avg}} = \frac{1}{k}\sum_{i=1}^{k} \|(U^i)^T \hat{U}^i\|_2, \tag{36}$$

where $\hat{U}^i$ denotes the reconstructed orthonormal basis of S^i. ψ_{avg} belongs to $[0, 1]$, and a larger value indicates better recovery. One can see from Fig. 3 that our recovered subspaces are exact when the average erasure rate is below 0.25.

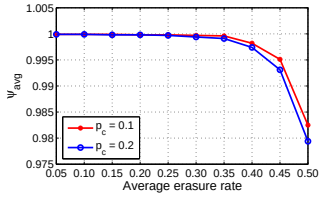

Fig. 3: ψ_{avg} according to various average erasure rates by our method.

We then change the values of r and N to generate different X. We fix both p_{avg} and p_c to be 0.2. We select the same parameters for HRMC and our method as mentioned above. Fig. 4 shows the recovery performance of SVT, HRMC, and our method. SVT performs similarly since N does not affect the rank of X. When N is sufficiently large, our method clearly outperforms SVT and HRMC.

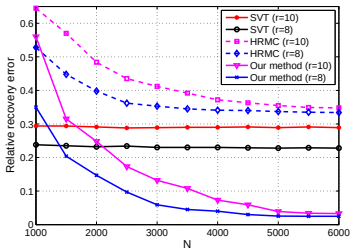

Fig. 4: Relative recovery errors of SVT, HRMC and our method for various values of r and N ($p_{\text{avg}} = 0.2$, $p_c = 0.2$).

We study the data recovery performance with respect to the separations among subspaces. The parameters of X are chosen as follows: $n = 50$, $N = 2500$, $k = 10$, and $r = 5$. Using the same ideas as in [60], we control the subspace separation as follows. We independently generate k orthonormal bases $\{\tilde{U}^i \in \mathbb{R}^{n\times r}\}_{i=1}^k$ from random Gaussian distributions. Let $U^1 = \tilde{U}^1$. For all $i \geq 2$, let $W^i = U^{i-1} + t\tilde{U}^i$, let S^i be the column span of W^i, and let U^i be the orthonormal basis of S^i. Then the parameter t controls the separation among subspaces, and the separation increases as t increases. To measure the separation, We define

$$\theta_{\max} = \max_{i\neq j} \|(U^i)^T U^j\|_2, \text{ and } \theta_{\min} = \min_{i\neq j} \|(U^i)^T U^j\|_2, \forall i, j. \tag{37}$$

Fig. 5 shows $\theta_{\max}$ and $\theta_{\min}$ when t varies. Smaller values of $\theta_{\max}$ and $\theta_{\min}$ correspond to a larger separation. Fig. 6 shows the relative recovery error of our method with $p_{\text{avg}} = 0.15$ when t increases. The recovery error first increases when t increases from 0. That is because in this region, subspaces are close to each other, and a neighbor matrix contains points from other subspaces inevitably. Thus, the recovery performance degrades when the subspaces separate more. When t further increases, the subspaces are separated enough such that the number of points in the neighbor matrix belonging to other subspaces decrease, then the recovery performance improves.

Here, points in the sums of the subspaces are spread over all the possible combinations, i.e., $p_{ij} = 2p_c/(k(k-1)), \forall i \neq j$. We also obtain similar results when some of the p_{ij}'s are zero and do not report these results due to the page limit.

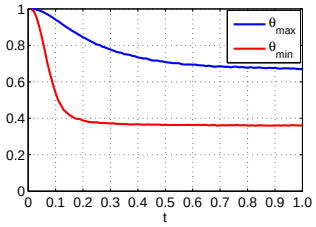

Fig. 5: $\theta_{\max}$ and $\theta_{\min}$ according to various t.

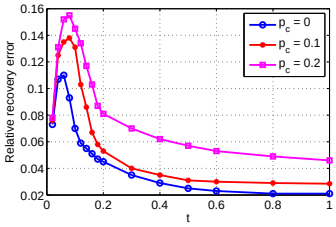

Fig. 6: Relative recovery error of our method with $p_{\text{avg}} = 0.15$ when the separation among subspaces increases.

We then compare the data recovery performance with SVT, HRMC, GSSC [41], SSC [61], Ksub [6], and rMiCUSaL [60]. The key parameters of X are chosen as follows: $n = 50$, $N = 2500$, $k = 10$, and $r = 5$. We follow the same line of the simulation setup of Fig. 2 to generate the matrix. We use 100 seeds to implement HRMC and our method. Due to the high computational time of SSC and GSSC, the results of SSC and GSSC are only averaged over 10 times. Fig. 7 shows

the reconstruction performance of SVT, HRMC, SSC, GSSC, Ksub, and rMiCUSaL when $p_c = 0$. Here our method reduces to HRMC when $p_c = 0$, so we only show HRMC in Fig. 7. Ksub, rMiCUSaL, SSC, and GSSC have smaller recovery errors than our method. However, the convergence of Ksub and rMiCUSaL greatly depends on the initialization of the subspaces. In our simulation, we use random initialization for rMiCUSaL and probabilistic farthest insertion [37] for Ksub. rMiCUSaL converges with probability around 0.85, partially because it is specially designed to cluster subspaces that are close to each other. Ksub converges with probability around 0.8. SSC and GSSC have much higher computational cost than our method. For instance, the computational time of GSSC is about fifty times that of our method in this setup. Moreover, the computational time of SSC and GSSC increases significantly when the problem size increases.

Fig. 8 shows the recovery performance of these methods when $p_c = 0.2$, and our method has the best performance. In order to handle outliers in the sums of the subspaces, we extend SSC by combining it with Robust PCA (with missing data). Subroutine 2 is applied on each cluster for data recovery after the clustering is obtained by SSC. The other methods do not consider the points in the sum.

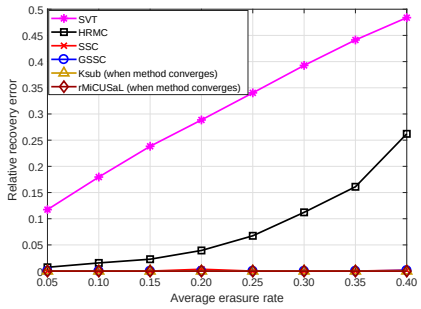

Fig. 7: Relative recovery errors of SVT, HRMC, SSC, GSSC, Ksub, and rMiCUSaL when $p_c = 0$.

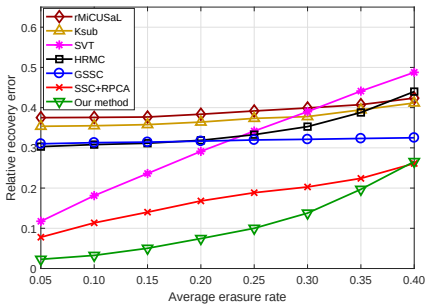

Fig. 8: Relative recovery errors of SVT, HRMC, SSC, GSSC, Ksub, rMiCUSaL, and our method when $p_c = 0.2$.

2) Noisy formulation: The key parameters of X are chosen as follows: $n = 50$, $N = 2500$, $k = 10$, and $r = 5$. We follow the same line of the simulation setup of Fig. 2 to generate the matrix. We generate matrix $\xi \in \mathbb{R}^{n \times N}$ with independent Gaussian $\mathcal{N}(0, \sigma^2)$ entries and simulate the observed measurement matrix $\tilde{X}$ according to (1). p_c and p_{avg} are chosen to be 0.2 and 0.1 respectively. We apply Subroutine 4 to obtain the corresponding subspace of each local neighbor matrix. We use 50 seeds to implement HRMC and our method. $\hat{\lambda}_i$ and $\tilde{\lambda}_i$ in (16) are set to be 0.005 and 0.003 respectively. Fig. 9 shows our method outperforms SVT and HRMC under different noise level σ.

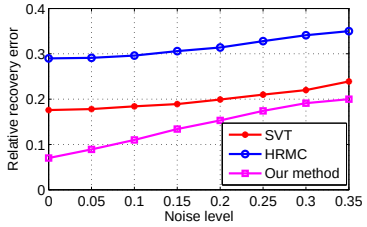

Fig. 9: Relative recovery errors of SVT, HRMC, and our method for different noise level σ ($p_{\text{avg}} = 0.1$, $p_c = 0.2$).

B. Performance on simulated PMU data

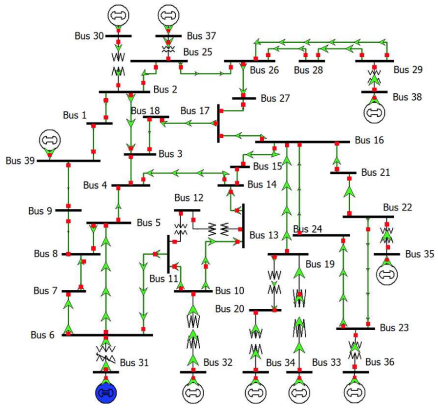

Fig. 10: IEEE 39 New England Power System [26][38]. We assume that sixteen PMUs are installed at bus 2, 4, 6, 8, 10, 12, 16, 18, 20, 22, 26, 33, 36, 37, 38 and 39.

We simulate the PMU data in the IEEE 39-bus New England Power System (Fig. 10). We assume sixteen PMUs are installed in the power system. Each PMU measures the voltage phasor at the corresponding bus and the current phasors on its incident lines at a rate of thirty samples per second. The sixteen PMUs measure fifty-seven voltage and current phasors in total. We first consider the scenario that no disturbance happens in the power system. We add the noise to the ground-truth data with Gaussian distribution $\mathcal{N}(0, \sigma^2)$. We compare several methods, including a simple interpolation method that uses the most recent observed value in the same channel as an estimate of the missing point. Fig. 11 (a) shows the current magnitudes of the noisy PMU data when $\sigma = 0.05$. We delete some measurements uniformly at random in each channel and test the recovery performance of SVT, HRMC, interpolation method, and our method. Fig. 11 (b) shows the relative recovery errors of SVT, HRMC, interpolation

method, and our method on the noisy PMU data. We can see that our method achieves the similar performance as HRMC and performs generally better than SVT and the interpolation method when p_{avg} is less than 0.3.

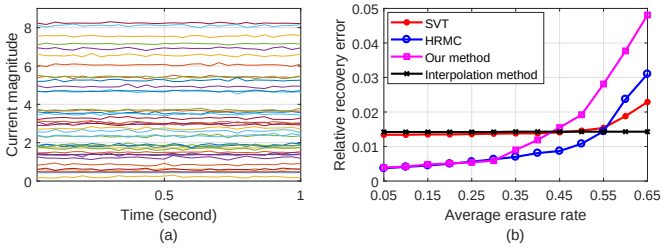

Fig. 11: (a) Current magnitudes of the noisy PMU data when $\sigma = 0.05$; (b) Relative recovery errors of SVT, HRMC, interpolation method, and our method on the simulated PMU data when $\sigma = 0.05$.

We then consider the scenario that multiple disturbances happen in the power system. Assume two events happen consecutively due to the sudden drop of the active power generations of two generators. Generator 32 drops at $t = 0.07$ s, and generator 33 drops at $t = 0.6$ s. Fig. 12 shows the active power of generators 32 and 33 and the voltage magnitudes at bus 2, 6 and 12. Let the complex matrix X in $\mathbb{C}^{30\times 57}$ contain the PMU measurements in one second. Phasors are represented by complex numbers in the rectangular form. Each column corresponds to the measurements of one PMU channel across time. Each row corresponds to the measurements at the same sampling instant.

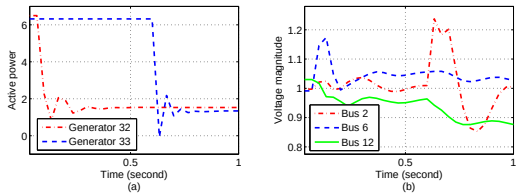

Fig. 12: (a) Active power of generator 32 and 33; (b) Voltage magnitudes at bus 2, 6 and 12.

We preprocess the PMU data by subtracting the mean for each channel. We delete some measurements uniformly at random for each channel and test the recovery performance of SVT, HRMC, and our method. As shown in Fig. 13, our method outperforms SVT and HRMC when p_{avg} is below 0.2.

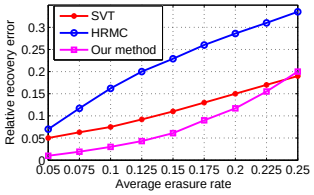

Fig. 13: Relative recovery errors of SVT, HRMC, and our method on the simulated PMU data.

V. Conclusion

We study the missing data recovery problem when the high-dimensional signals exhibit nonlinear low-dimensional structures. We have proposed a new model called "union and sums of subspaces" that extends the existing UoS model. This model characterizes the nonlinear structures while maintaining the model simplicity. We have developed missing data recovery methods under this model and analyzed the performance theoretically and numerically. Motivated by power system monitoring, our generic model and method can potentially contribute to other applications like sensor network applications and remote sensing. One future direction is to develop new models to describe highly nonlinear high-dimensional datasets. Missing data recovery for streaming data with nonlinear structures also pose additional interesting open problems.

Acknowledgment

We thank Prof. Waheed U. Bajwa at Rutgers, The State University of New Jersey for sharing the codes of rMiCUSaL [60] and his implementation of Ksub [6]. We thank Dr. Daniel Pimentel-Alarcón at the University of Wisconsin-Madison for sharing the codes of GSSC [41] and his implementation of SSC [61]. This research is supported in part by the ERC Program of NSF and DoE under the supplement to NSF Award EEC-1041877 and the CURENT Industry Partnership Program, NYSERDA Grant #36653, and NSF Grant #1508875.

References

[1] "The Netflix Prize," http://netflixprize.com/, June 2006.

[2] Y. Amit, M. Fink, N. Srebro, and S. Ullman, "Uncovering shared structures in multiclass classification," in *Proc. the 24th international conference on Machine learning*. ACM, 2007, pp. 17–24.

[3] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," *Advances in neural information processing systems*, vol. 19, p. 41, 2007.

[4] T. Athay, R. Podmore, and S. Virmani, "A practical method for the direct analysis of transient stability," *IEEE Trans. Power App. Syst.*, no. 2, pp. 573–584, 1979.

[5] L. Balzano, B. Recht, and R. Nowak, "High-dimensional matched subspace detection when data are missing," in *Proc. IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2010, pp. 1638–1642.

[6] L. Balzano, A. Szlam, B. Recht, and R. Nowak, "k-subspaces with missing data," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, 2012, pp. 612–615.

[7] T. Banerjee, B. Xie, and D. P. Agrawal, "Fault tolerant multiple event detection in a wireless sensor network," *Journal of parallel and distributed computing*, vol. 68, no. 9, pp. 1222–1234, 2008.

[8] R. G. Baraniuk, V. Cevher, and M. B. Wakin, "Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 959–971, 2010.

[9] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[10] E. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053 –2080, 2010.

[11] E. Candès, L. Mackey, and M. Soltanolkotabi, "From subspace clustering to full-rank matrix completion," *unpublished abstract*, 2014.

[12] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[13] P. Chen and D. Suter, "Recovering the missing components in a large noisy low-rank matrix: Application to SFM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1051–1063, 2004.

[14] Y. Chen, H. Xu, C. Caramanis, and S. Sanghavi, "Robust matrix completion with corrupted columns," *arXiv preprint arXiv:1102.2254*, 2011.

[15] J. H. Chow and K. W. Cheung, "A toolbox for power system dynamics and control engineering education and research," *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1559–1564, 1992.

[16] M. Elad, R. Goldenberg, and R. Kimmel, "Low bit-rate compression of facial images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2379–2383, 2007.

[17] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, 2013.

[18] B. Eriksson, L. Balzano, and R. Nowak, "High-rank matrix completion," in *Proc. International Conference on Artificial Intelligence and Statistics*, 2012, pp. 373–381.

[19] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, 2002.

[20] P. Gao, M. Wang, J. H. Chow, M. Berger, and L. M. Seversky, "Matrix completion with columns in union and sums of subspaces," in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP) 2015*, 2015, pp. 785–789.

[21] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, "Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements," *IEEE Trans. Power Systems*, vol. 31, no. 2, pp. 1006 – 1013, 2016.

[22] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/, Oct. 2010.

[23] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[24] R. Heckel and H. Bölcskei, "Robust subspace clustering via thresholding," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6320–6342, 2015.

[25] F. J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," *Geophysical Journal International*, vol. 173, no. 1, pp. 233–248, 2008.

[26] IEEE 10 Generator 39 Bus System, http://sys.elec.kitami-it.ac.jp/ueda/demo/WebPF/39-New-England.pdf.

[27] B. Jafarpour, V. K. Goyal, D. B. McLaughlin, and W. T. Freeman, "Transform-domain sparsity regularization for inverse problems in geosciences," *Geophysics*, vol. 74, no. 5, pp. R69–R83, 2009.

[28] O. Klopp, K. Lounici, and A. B. Tsybakov, "Robust matrix completion," *arXiv preprint arXiv:1412.8132*, 2014.

[29] G. Lerman, T. Zhang *et al.*, "Robust recovery of multiple subspaces by geometric lp minimization," *The Annals of Statistics*, vol. 39, no. 5, pp. 2686–2715, 2011.

[30] W.-M. Lin, C.-H. Wu, C.-H. Lin, and F.-S. Cheng, "Detection and classification of multiple power-quality disturbances with wavelet multiclass svm," *IEEE Trans. Power Del.*, vol. 23, no. 4, pp. 2575–2582, 2008.

[31] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[32] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, 2013.

[33] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.

[34] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, 2012.

[35] S. Mika, B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel pca and de-noising in feature spaces." in *Proc. Advances in Neural Information Processing Systems (NIPS)*, vol. 4, no. 5, 1998, p. 7.

[36] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 2, pp. 849–856, 2002.

[37] R. Ostrovsky, Y. Rabani, L. J. Schulman, and C. Swamy, "The effectiveness of lloyd-type methods for the k-means problem," in *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, 2006, pp. 165–176.

[38] A. Pai, *Energy function analysis for power system stability*. Springer, 1989.

[39] A. Phadke and J. Thorp, *Synchronized phasor measurements and their applications*. Springer, 2008.

[40] D. Pimentel, R. Nowak, and L. Balzano, "On the sample complexity of subspace clustering with missing data," in *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, 2014, pp. 280–283.

[41] D. Pimentel-Alarcón, L. Balzano, R. Marcia, R. Nowak, and R. Willett, "Group-sparse subspace clustering with missing data," in *Statistical Signal Processing Workshop (SSP)*. IEEE, 2016, pp. 1–5.

[42] D. L. Pimentel-Alarcón and R. D. Nowak, "The information-theoretic requirements of subspace clustering with missing data," in *Proc. International Conference on Machine Learning*, 2016, pp. 802–810.

[43] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1832–1845, 2010.

[44] B. Recht, "A simpler approach to matrix completion," *The Journal of Machine Learning Research*, pp. 3413–3430, 2011.

[45] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[46] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[47] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Proc. International Conferences on Artificial Neural Networks (ICANN)*, 1997, pp. 583–588.

[48] A. Silverstein and J. E. Dagle, "Successes and challenges for synchrophasor technology: An update from the north american synchrophasor initiative," in *Proc. Hawaii International Conference on System Science (HICSS)*, 2012, pp. 2091–2095.

[49] M. Soltanolkotabi and E. J. Candès, "A geometric analysis of subspace clustering with outliers," *The Annals of Statistics*, vol. 40, no. 4, pp. 2195–2238, 2012.

[50] M. Soltanolkotabi, E. Elhamifar, and E. J. Candés, "Robust subspace clustering," *The Annals of Statistics*, vol. 42, no. 2, pp. 669–699, 2014.

[51] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 8, pp. 831–836, 1996.

[52] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[53] P. Tseng, "Nearest q-flat to m points," *Journal of Optimization Theory and Applications*, vol. 105, no. 1, pp. 249–252, 2000.

[54] M. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[55] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, 2005.

[56] M. Wang, J. H. Chow, P. Gao, X. T. Jiang, Y. Xia, S. G. Ghiocel, B. Fardanesh, G. Stefopolous, Y. Kokai, and N. Saito, "A low-rank matrix approach for the analysis of large amounts of power system synchrophasor data," in *Proc. IEEE Hawaii International Conference on System Sciences (HICSS)*, 2015, pp. 2637–2644.

[57] Y.-X. Wang and H. Xu, "Noisy sparse subspace clustering," in *Proceedings of The 30th International Conference on Machine Learning*, 2013, pp. 89–97.

[58] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.

[59] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[60] T. Wu and W. U. Bajwa, "Learning the nonlinear geometry of high-dimensional data: Models and algorithms," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6229–6244, 2015.

[61] C. Yang, D. Robinson, and R. Vidal, "Sparse subspace clustering with missing entries," in *Proceedings of The 32nd International Conference on Machine Learning*, 2015, pp. 2463–2472.

APPENDIX

A. An upper bound of $\mu(\mathbf{x}_1 - \mathbf{x}_2)$ from *A1* and *A2*

First suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ belong to S^{ij} and $S^{i'j'}$ $(i, j \neq i', j')$ respectively. We have $\mathbf{x}_1 = U^{ij}\boldsymbol{\beta}$, and $\mathbf{x}_2 = U^{i'j'}\boldsymbol{\beta}'$, where $U^{ij} = [U^i \;\; U^j]$, $U^{i'j'} = [U^{i'} \;\; U^{j'}]$, $\boldsymbol{\beta}^T = [\boldsymbol{\alpha}_1^T \;\; \boldsymbol{\alpha}_2^T]$, and $\boldsymbol{\beta'}^T = [\boldsymbol{\alpha}_3^T \;\; \boldsymbol{\alpha}_4^T]$. From the definition, we have

$$\mu(\mathbf{x}_1 - \mathbf{x}_2) = n\|U^{ij}\boldsymbol{\beta} - U^{i'j'}\boldsymbol{\beta}'\|_\infty^2 / \|U^{ij}\boldsymbol{\beta} - U^{i'j'}\boldsymbol{\beta}'\|_2^2. \quad (38)$$

From **A1**, we have

$$n\|U^{ij}\boldsymbol{\beta}-U^{i'j'}\boldsymbol{\beta}'\|_\infty^2 \le n(\|U^{ij}\boldsymbol{\beta}\|_\infty+\|U^{i'j'}\boldsymbol{\beta}'\|_\infty)^2$$
$$\le n(\sqrt{\frac{2\mu_1 r}{n}}\|\boldsymbol{\beta}\|_2+\sqrt{\frac{2\mu_1 r}{n}}\|\boldsymbol{\beta}'\|_2)^2 = 2\mu_1 r(\|\boldsymbol{\beta}\|_2+\|\boldsymbol{\beta}'\|_2)^2. \quad (39)$$

From **A2** and the Cauchy-Schwarz inequality, we have

$$\|\boldsymbol{\alpha}_1\|_2^2+\|\boldsymbol{\alpha}_2\|_2^2+2\boldsymbol{\alpha}_1^T(U^i)^TU^j\boldsymbol{\alpha}_2 \ge (1-\theta)\|\boldsymbol{\beta}\|_2^2, \quad (40)$$

$$\|\boldsymbol{\alpha}_3\|_2^2+\|\boldsymbol{\alpha}_4\|_2^2+2\boldsymbol{\alpha}_3^T(U^{i'})^TU^{j'}\boldsymbol{\alpha}_4 \ge (1-\theta)\|\boldsymbol{\beta}'\|_2^2, \quad (41)$$

$$\text{and } 2\boldsymbol{\alpha}_1^T(U^i)^TU^{i'}\boldsymbol{\alpha}_3+2\boldsymbol{\alpha}_2^T(U^j)^TU^{i'}\boldsymbol{\alpha}_3+2\boldsymbol{\alpha}_1^T(U^i)^TU^{j'}\boldsymbol{\alpha}_4$$
$$+2\boldsymbol{\alpha}_2^T(U^j)^TU^{j'}\boldsymbol{\alpha}_4 \le 2\theta(\|\boldsymbol{\beta}\|_2^2+\|\boldsymbol{\beta}'\|_2^2). \quad (42)$$

Combine (40),(41), and (42), we have

$$\|U^{ij}\boldsymbol{\beta}-U^{i'j'}\boldsymbol{\beta}'\|_2^2 \ge (1-3\theta)(\|\boldsymbol{\beta}\|_2^2+\|\boldsymbol{\beta}'\|_2^2). \quad (43)$$

Combine (38), (39), and (43), we then have $\mu(\mathbf{x}_1-\mathbf{x}_2) \le \frac{4\mu_1 r}{1-3\theta}$. The analyses when $\mathbf{x}_1$ and $\mathbf{x}_2$ are in other subspaces are similar, leading to smaller upper bounds on $\mu(\mathbf{x}_1-\mathbf{x}_2)$.

*B. Parameters in **A3** with uniform distributions*

Let $\mathbf{x}_0 = U^i\boldsymbol{\alpha}$ denote a column in S^i with $\boldsymbol{\alpha}$ uniformly distributed in the unit ball of $\mathbb{R}^r$. Let $\mathbf{x}' = U^{i'j'}\boldsymbol{\beta}$ denote a column in $S^{i'j'}$ with $\boldsymbol{\beta}$ uniformly distributed in the unit ball of $\mathbb{R}^{2r}$, $\forall i', j' \neq i$, where $U^{i'j'} = [U^{i'}\ U^{j'}]$, and $\boldsymbol{\beta}^T = [\boldsymbol{\beta}_1^T\ \boldsymbol{\beta}_2^T]$. Then,

$$\|\mathbf{x}_0-\mathbf{x}'\|_2^2 = \|U^i\boldsymbol{\alpha}-U^{i'j'}\boldsymbol{\beta}\|_2^2$$
$$\ge \|\boldsymbol{\alpha}\|_2^2+\|\boldsymbol{\beta}\|_2^2-2\theta\|\boldsymbol{\alpha}\|_2(\|\boldsymbol{\beta}_1\|_2+\|\boldsymbol{\beta}_2\|_2)-2\theta\|\boldsymbol{\beta}_1\|_2\|\boldsymbol{\beta}_2\|_2$$
$$\ge \|\boldsymbol{\alpha}\|_2^2+(1-\theta)\|\boldsymbol{\beta}\|_2^2-2\sqrt{2}\theta\|\boldsymbol{\alpha}\|_2\|\boldsymbol{\beta}\|_2. \quad (44)$$

Given $\mathbf{x}_0$, the right hand side of (44) achieves the minimum value $(1-\frac{2\theta^2}{(1-\theta)^2})\|\boldsymbol{\alpha}\|_2^2$ (positive since $\theta < 1/3$) when $\|\boldsymbol{\beta}\|_2 = \sqrt{2}\theta\|\boldsymbol{\alpha}\|_2/(1-\theta)$. Thus, if

$$\|\boldsymbol{\alpha}\|_2 \ge \epsilon_0/\sqrt{1-2\theta^2/(1-\theta)^2},$$

we have $\|\mathbf{x}_0-\mathbf{x}'\|_2 \ge \epsilon_0$ for any $\mathbf{x}'$ in $S^{i'j'}$. When $\epsilon_0 \le \sqrt{1-2\theta^2/(1-\theta)^2}$, since $\boldsymbol{\alpha}$ is uniformly distribution, the fraction of the columns in S^i that are ϵ_0 away from $S^{i'j'}$ is at least $1-(\epsilon_0/\sqrt{1-2\theta^2/(1-\theta)^2})^r$, which is v_0.

Given $\mathbf{x}_1 = U^i\boldsymbol{\alpha}_1$ in $S^i_{\epsilon_0}$, $B_{\mathbf{x}_1,\epsilon_0/\sqrt{3}}$ only contains points in S^i and S^{ij}'s. Let $\mathbf{x}_2 = U^i\boldsymbol{\alpha}_2$ denote a column in S^i with $\boldsymbol{\alpha}_2$ uniformly distributed in the unit ball of $\mathbb{R}^r$. Since $\|\mathbf{x}_1-\mathbf{x}_2\|_2 = \|\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2\|_2$, then $\mathbf{x}_2$ belongs to $B_{\mathbf{x}_1,\epsilon_0/\sqrt{3}}$ if $\|\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2\|_2 \le \epsilon_0/\sqrt{3}$ holds. Since $\boldsymbol{\alpha}_2$ is uniformly distributed in the unit ball of $\mathbb{R}^r$, the fraction of the columns in S^i that also belong to $B_{\mathbf{x}_1,\epsilon_0/\sqrt{3}}$ is at least $v_1(\epsilon_0/\sqrt{3})^r$ for some constant $v_1 \in (0,1]$. In fact $v_1 = 1$ if $\|\boldsymbol{\alpha}_1\| \in [\epsilon_0/\sqrt{1-2\theta^2/(1-\theta)^2}, 1-\epsilon_0/\sqrt{3}]$.

Let $\mathbf{x}_3 = U^i\boldsymbol{\alpha}_3+U^j\boldsymbol{\alpha}_4$ denote a column in S^{ij}, where $\|\boldsymbol{\alpha}_3\|_2^2+\|\boldsymbol{\alpha}_4\|_2^2 \le 1$. We have

$$\|\mathbf{x}_1-\mathbf{x}_3\|_2^2 = \|U^i\boldsymbol{\alpha}_1-(U^i\boldsymbol{\alpha}_3+U^j\boldsymbol{\alpha}_4)\|_2^2$$
$$\ge \|\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_3\|_2^2+\|\boldsymbol{\alpha}_4\|_2^2-2\theta\|\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_3\|_2\|\boldsymbol{\alpha}_4\|_2$$
$$\ge (1-\theta)(\|\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_3\|_2^2+\|\boldsymbol{\alpha}_4\|_2^2) = (1-\theta)\|\boldsymbol{\beta}'-\tilde{\boldsymbol{\beta}}\|_2^2, \quad (45)$$

where $\boldsymbol{\beta}'^T = [\boldsymbol{\alpha}_1^T\ \mathbf{0}]$ and $\tilde{\boldsymbol{\beta}}^T = [\boldsymbol{\alpha}_3^T\ \boldsymbol{\alpha}_4^T]$. We also have

$$\|\mathbf{x}_1-\mathbf{x}_3\|_2^2 = \|U^i\boldsymbol{\alpha}_1-(U^i\boldsymbol{\alpha}_3+U^j\boldsymbol{\alpha}_4)\|_2^2$$
$$\le \|\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_3\|_2^2+\|\boldsymbol{\alpha}_4\|_2^2+2\theta\|\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_3\|_2\|\boldsymbol{\alpha}_4\|_2$$
$$\le (1+\theta)(\|\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_3\|_2^2+\|\boldsymbol{\alpha}_4\|_2^2) = (1+\theta)\|\boldsymbol{\beta}'-\tilde{\boldsymbol{\beta}}\|_2^2. \quad (46)$$

From (45), $\|\mathbf{x}_1-\mathbf{x}_3\| > \epsilon_0$ if $\|\boldsymbol{\beta}'-\tilde{\boldsymbol{\beta}}\| \ge \epsilon_0/\sqrt{1-\theta}$. From (46), $\|\mathbf{x}_1-\mathbf{x}_3\| > \epsilon_0$ if $\|\boldsymbol{\beta}'-\tilde{\boldsymbol{\beta}}\| < \epsilon_0/\sqrt{1+\theta}$. Since $\tilde{\boldsymbol{\beta}}$ is uniformly distributed in the unit ball of $\mathbb{R}^{2r}$, the fraction of columns in S^{ij} but not S^i that also belong to $B_{\mathbf{x},\epsilon_0}$ is between $v_2(\epsilon_0/(1+\theta))^{2r}$ and $v_3(\epsilon_0/(1-\theta))^{2r}$ for some constants v_2, $v_3 \in (0,1]$. In fact, if $\frac{\epsilon_0}{\sqrt{1-2\theta^2/(1-\theta)^2}}+\frac{\epsilon_0}{\sqrt{1-\theta}} \le 1$, and $\|\boldsymbol{\alpha}_1\| \in [\epsilon_0/\sqrt{1-2\theta^2/(1-\theta)^2}, 1-\epsilon_0/\sqrt{1-\theta}]$, we have $v_2 = 1$ and $v_3 = 1$.

C. Proof of Lemma 1

Proof. Let $\{\hat{S}^1,...,\hat{S}^k\}$ denote the collection of k approximated reconstructed subspaces of dimension at most r. Consider a column $\tilde{\mathbf{x}}$ in matrix $\tilde{X}$. Let the entries of $\tilde{\mathbf{x}}$ be sampled uniformly with replacement. Let ω denote the set of indices of the observed entries in $\tilde{\mathbf{x}}$, and let $m = |\omega|$.

Lemma 2 (Theorem 1 of [5])**.** *Pick any $\delta > 0$ and $m > \frac{8}{3}r\mu(S)\log(\frac{2r}{\delta})$. Then with probability at least $1-4\delta$*[6]*,*

$$\frac{m(1-a)-r\mu(S)\frac{(1+b)^2}{(1-c)}}{n}\|\tilde{\mathbf{x}}-\mathcal{P}_S\tilde{\mathbf{x}}\|_2^2 \le \|\tilde{\mathbf{x}}_\omega-\mathcal{P}_{\omega,S}\tilde{\mathbf{x}}_\omega\|_2^2,$$

$$\textit{and } \|\tilde{\mathbf{x}}_\omega-\mathcal{P}_{\omega,S}\tilde{\mathbf{x}}_\omega\|_2^2 \le (1+a)\frac{m}{n}\|\tilde{\mathbf{x}}-\mathcal{P}_S\tilde{\mathbf{x}}\|_2^2, \quad (47)$$

where $a = \sqrt{\frac{2\mu(\tilde{\mathbf{x}}-\mathcal{P}_S\tilde{\mathbf{x}})^2}{m}\log(\frac{1}{\delta})}$, $b = \sqrt{2\mu(\tilde{\mathbf{x}}-\mathcal{P}_S\tilde{\mathbf{x}})\log(\frac{1}{\delta})}$, $c = \sqrt{\frac{8r\mu(S)}{3m}\log(\frac{2r}{\delta})}$, and r is the rank of S.

We here consider the case that the column $\tilde{\mathbf{x}}$ is closest to $\hat{S}^{ij}$. Let $\delta = \delta_1'$. From Lemma 2, if $m > 16r\rho_1\mu_1\log(4r/\delta_1')/3$,

$$\frac{\|\tilde{\mathbf{x}}-\mathcal{P}_{\hat{S}^{ij}}\tilde{\mathbf{x}}\|_2^2}{\|\tilde{\mathbf{x}}-\mathcal{P}_{\hat{S}^{i'j'}}\tilde{\mathbf{x}}\|_2^2} < \frac{m(1-a)-2r\rho_1\mu_1\frac{(1+b)^2}{(1-c)}}{(1+a)m} \quad (48)$$

holds for any $(i',j') \neq (i,j)$. By the union bound, we have $\|\tilde{\mathbf{x}}_\omega-\mathcal{P}_{\omega,\hat{S}^{ij}}\tilde{\mathbf{x}}_\omega\|_2 < \|\tilde{\mathbf{x}}_\omega-\mathcal{P}_{\omega,\hat{S}^{i'j'}}\tilde{\mathbf{x}}_\omega\|_2$ holds for any $(i',j') \neq (i,j)$ with probability at least $1-2k(k-1)\delta_1'$. We then require $m > \frac{2r\rho_1\mu_1(1+b)^2}{(1-c)((1-a)-(1+a)\epsilon_3)}$, which means we require

$$m > \max(\frac{16}{3}r\rho_1\mu_1\log(\frac{4r}{\delta_1'}), \frac{2r\rho_1\mu_1(1+b)^2}{(1-c)((1-a)-(1+a)\epsilon_3)}). \quad (49)$$

By the Chernoff's bound, we have

$$\mathcal{P}(m \le np_0/2) \le \exp(-np_0/8) < \delta_1', \quad (50)$$

and note that

$$np_0/2 \ge 4\log((\beta+1)n)\max(\frac{\beta}{\alpha_1}, \frac{\log^2 n}{\alpha_2})/\beta. \quad (51)$$

[6]The nonexistence of $P_{\omega,\hat{S}^{ij}}$ is taken into account in the failure probability.

There is an $O(\log^2(n))$ gap between the right hand side of (51) and $\frac{16}{3}r\rho_1\mu_1\log(\frac{4r}{\delta_1'})$. Therefore, the former is larger when n is sufficiently large. From (33), we have

$$\frac{np_0}{2} > \frac{2r\rho_1\mu_1(1+b)^2}{(1-c)((1-a)-(1+a)\epsilon_3)}. \tag{52}$$

Then, (49) is met with probability at least $1-\delta_1'$ by combining (50) to (52). By the union bound, we have that if the column $\tilde{\mathbf{x}}$ is closest to $\hat{S}^{ij}$, then

$$\|\tilde{\mathbf{x}}_\omega - \mathcal{P}_{\omega,\hat{S}^{ij}}\tilde{\mathbf{x}}_\omega\|_2 < \|\tilde{\mathbf{x}}_\omega - \mathcal{P}_{\omega,\hat{S}^{i'j'}}\tilde{\mathbf{x}}_\omega\|_2 \tag{53}$$

holds for any $(i',j') \neq (i,j)$ with probability at least $1-(2k(k-1)+1)\delta_1'$. The analysis for the case that $\tilde{\mathbf{x}}$ is closest to $\hat{S}^i$ is similar. If $\tilde{\mathbf{x}}$ is closest to $\hat{S}^i$, we have that

$$\|\tilde{\mathbf{x}}_\omega - \mathcal{P}_{\omega,\hat{S}^{ij}}\tilde{\mathbf{x}}_\omega\|_2 < \|\tilde{\mathbf{x}}_\omega - \mathcal{P}_{\omega,\hat{S}^{i'j'}}\tilde{\mathbf{x}}_\omega\|_2 \tag{54}$$

holds for $\forall i' \neq i$ with probability at least $1-(2(k-1)((k-1)(k-2)+2)+1)\delta_1'$. By the union bound, Lemma 1 holds with probability at least $1-(2(k-1)(k^2-2k+4)+2)\delta_1'$. □

D. Skeleton-Proof of Theorem 1

The proof of Theorem 1 follows the same line as the proof of Theorem 2.1 in [18]. The skeleton-proof will be presented according to the key steps of our recovery method. Please refer to Appendix E, F, G, H, and I for the proofs of Lemma 3, 5, 6, 7, and 8.

1) Selection of local neighbor matrix:

Lemma 3. *Assume **A3** holds. If the number of seeds,*

$$s \geq \frac{\log k + \log\frac{1}{\delta_0}}{(1-\exp(-\frac{\eta_1\mu_1^2r^2\log^3(4\beta_0 n)}{4}))v_0p_*}, \tag{55}$$

then with probability at least $1-\delta_0$, for all $i=1,...,k$, at least one seed is in $S^i_{\epsilon_0}$, and each seed has at least ζ_0 observed entries for some constants $\beta_0 \geq 1$ and η_1.

Lemma 3 guarantees that if enough seeds are selected, then with high probability at least one seed is in $S^i_{\epsilon_0}$, $\forall i$, and has at least ζ_0 observed entries.

Lemma 4 (Lemma 3 of [18])**.** *Assume **A1** and **A2** hold. Let $\mathbf{y} = \mathbf{x}_1 - \mathbf{x}_2$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are two columns of X. Assume there is a common set of indices ω of size $q \leq n$ where both $\mathbf{x}_1$ and $\mathbf{x}_2$ are observed. Let $\mathbf{y}_\omega$ denote the corresponding subset of $\mathbf{y}$. Then for any $\delta > 0$, if $q \geq \frac{128\mu_1^2r^2}{(1-3\theta)^2}\log(\frac{2}{\delta})$, then with probability at least $1-\delta$,*

$$\frac{1}{2}\|\mathbf{y}\|_2^2 \leq \frac{n}{q}\|\mathbf{y}_\omega\|_2^2 \leq \frac{3}{2}\|\mathbf{y}\|_2^2. \tag{56}$$

We pick $t_0 := \frac{128\mu_1^2r^2}{(1-3\theta)^2}\log(\frac{2s_0l_0n}{\delta_0})$, then $\delta = \delta_0/(s_0l_0n)$ in Lemma 4. From Lemma 4 and the union bound, with probability at least $1-\delta_0$, (56) holds for all s_0l_0n columns selected in Step (ii) for all s_0 neighbor matrices, where $\mathbf{x}_1$ and $\mathbf{x}_2$ represent a seed and a selected column for that seed respectively. It has two implications. First, every point within $\frac{\epsilon_0}{\sqrt{3}}$ of the seed has partial distance within $\frac{\epsilon_0}{\sqrt{2}}$ of the seed and thus, would be selected in Step (iii). Second, every point selected in Step (iii), which means its partial distance is within $\frac{\epsilon_0}{\sqrt{2}}$ of the seed, is within ϵ_0 away from the seed.

For each seed $\mathbf{x} \in S^i_{\epsilon_0}$, let $\hat{T}^1_{\mathbf{x}}$ and $\hat{T}^2_{\mathbf{x}}$ denote the number of columns that are in S^i and not in S^i, respectively, in its neighbor matrix selected in Step (iv). Then we have,

Lemma 5. *Assume **A3** holds. If the number of columns selected for each seed in Step (ii), ln, satisfies that*

$$l \geq \max(\frac{2}{v_1(\frac{\epsilon_0}{\sqrt{3}})^r p_*}, \frac{8\log(2s/\delta_0)}{nv_1(\frac{\epsilon_0}{\sqrt{3}})^r p_*}, \frac{12\log(2s/\delta_0)}{nv_2(\frac{\epsilon_0}{1+\theta})^{2r}q_*}), \tag{57}$$

then with probability at least $1-4\delta_0$, it holds that

$$\hat{T}^1_{\mathbf{x}} \geq n, \tag{58}$$

$$\text{and} \quad \hat{T}^2_{\mathbf{x}}/\hat{T}^1_{\mathbf{x}} \leq 9v_3(\sqrt{3}\epsilon_0)^r q_i/(v_1(1-\theta)^{2r}p_i), \tag{59}$$

for the local neighbor matrices of all the seeds $\mathbf{x}$ in $S^i_{\epsilon_0}$, $\forall i$.

Lemma 5 guarantees that the local neighbor matrix of seed $\mathbf{x}$ in $S^i_{\epsilon_0}$ has at least n columns in S^i and the number of columns not in S^i is upper bounded by (59) when l is sufficiently large.

Lemma 6. *If $N \geq 2l_0n(2s_0l_0n/\delta_0)^{-\frac{128\mu_1^2r^2}{(1-3\theta)^2}\log\underline{p}}$ and $\zeta_0 > t_0$, then Subroutine 1 satisfies the following, at least one seed $\mathbf{x}$ belongs to $S^i_{\epsilon_0}$ for each $i=1,...,k$, and (58) and (59) hold with probability at least $1-6\delta_0$.*

Lemma 6 says that if N is sufficiently large, at least one seed $\mathbf{x}$ is in S^i, $\forall i$. (58) and (59) hold for its neighbor matrix.

2) Local subspace estimation: [14] provides the theoretical guarantee of low-rank matrix completion when some columns do not belong to the same subspace. Here, building on [14] and applying the technique in [18], we show that $\{S^1,...,S^k\}$ can be correctly identified even when local neighbor matrices contain columns outside the subspace of interest.

Lemma 7. *Suppose there exists at least one seed $\mathbf{x}$ in S^i such that (58) holds, for every $i=1,...,k$. Assume $n \geq 32$. All neighbor matrices are observed uniformly at random with probability $p_0 \geq \underline{p}$. If*

$$\underline{p} \geq 2\eta_1\mu_1^2r^2\log^3(4\beta_0 n)/n, \tag{60}$$

$$\text{and} \quad \hat{T}^2_{\mathbf{x}}/\hat{T}^1_{\mathbf{x}} \leq \eta_2\frac{\underline{p}^2}{4(1+\frac{2\mu_1 r}{\underline{p}\sqrt{n}})^2\mu_1^3r^3\log^6(4\beta_0 n)}, \tag{61}$$

for at least one seed $\mathbf{x}$ in S^i for $i=1,...,k$, then with probability at least $1-s_0\delta_0$, $\{S^1,...,S^k\}$ belong to the candidate subspaces $\{S^{1^*},...,S^{s'^*}\}$ returned by Subroutine 2 with properly chosen λ_i's, where η_1 and η_2 are the constants.

Please refer to [14] for the exact definition of λ_i. Note that condition (22) in Theorem 1 and (59) implies (61). By applying the union bound to lemma 6 and lemma 7, we can prove that $\{S^1,...,S^k\}$ belongs to the output of Subroutine 2 with probability at least $(1-(6+s_0))\delta_0$.

For a seed not in $S^i_{\epsilon_0}$ for any i, the recovered subspace of this neighbor matrix might not belong to $\{S^1,...,S^k\}$. In this case, the recovered subspace has rank greater than r, as argued in [18], and will be deleted in the refinement step in

Subroutine 3. Therefore, $\{\hat{S}^1, ..., \hat{S}^k\}$ is the correct estimation of $\{S^1, ..., S^k\}$ with high probability.

3) Matched subspace assignment: Finally, we show that each column $\mathbf{x}$ is assigned to its correct subspace and correctly recovered from $\mathbf{x}_\omega$.

Lemma 8. *Assume **A1** and **A2** holds, and p_0 satisfies*

$$p_0 \geq \underline{p} \geq 2\eta_1 \mu_1^2 r^2 \log^3(4\beta_0 n)/n. \tag{62}$$

Then with probability at least $1 - (4k(k-1)+2)\delta_0$, if the column of interest $\mathbf{x} \in \hat{S}^i$, for any $i' \neq i$,

$$0 = \|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^{ij}}\mathbf{x}_\omega\|_2 < \|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^{i'j'}}\mathbf{x}_\omega\|_2, \ \forall j, j'; \tag{63}$$

if the column of interest $\mathbf{x} \in \hat{S}^{ij}$, for any $(i', j') \neq (i, j)$,

$$0 = \|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^{ij}}\mathbf{x}_\omega\|_2 < \|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^{i'j'}}\mathbf{x}_\omega\|_2. \tag{64}$$

Lemma 8 shows that each column $\mathbf{x}_\omega$ is assigned to the right subspace by (15).

E. Proof of Lemma 3

Proof. From (20), we have that the expected number of observed entries per column is at least $\zeta = 2\eta_1\mu_1^2 r^2 \log^3(4\beta_0 n)$. Let $\hat{\zeta}$ denote the number of observed entries in a column selected uniformly at random. By the Chernoff's bound, we have $\mathcal{P}(\hat{\zeta} \leq \zeta/2) \leq \exp(-\zeta/8)$. The probability that a randomly selected column belongs to $S^i_{\epsilon_0}$ and has $\zeta/2$ or more observed entries is at least v_0', where $v_0' := (1 - \exp(-\zeta/8))v_0 p_i$. Consider a set of s randomly selected columns. The probability that this set does not contain a column from $S^i_{\epsilon_0}$ with at least $\zeta/2$ observed entries is less than $(1-v_0')^s$. Then from the union bound, the probability that this set does not contain at least one column from $S^i_{\epsilon_0}$ with $\zeta/2$ or more observed entries, for any $i = 1, ..., k$, is less than $\sum_{i=1}^k (1-(1-\exp(-\zeta/8))v_0 p_i)^s$. Choose

$$s = \frac{\log k + \log \frac{1}{\delta_0}}{\log \frac{1/((1-\exp(-\zeta/8))v_0 p_*)}{1/((1-\exp(-\zeta/8))v_0 p_*)-1}} \tag{65}$$

such that $\delta_0 = k(1-(1-\exp(-\zeta/8))v_0 p_*)^s$ holds. Lemma 3 then holds with probability at least $1 - \delta_0$. The result follows by noting that $\log(\frac{x}{x-1}) \geq \frac{1}{x}$, for $x > 1$. □

F. Proof of Lemma 5

Proof. For each seed $\mathbf{x} \in S^i_{\epsilon_0}$, among the $l_0 n$ columns selected by Step (ii), let $T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}}$ denote the number of columns that belong to both $B_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}}$ and S^i, and let $T^2_{\mathbf{x},\epsilon_0}$ denote the number of columns that belong to both $B_{\mathbf{x},\epsilon_0}$ and any of S^{ij}'s (for every $j \neq i$) but not S^i. The probability that a column selected uniformly at random from X belongs to both $B_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}}$ and S^i is at least $v_1(\epsilon_0/\sqrt{3})^r p_i$. Therefore, the expectation of $T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}}$ satisfies

$$E(T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}}) \geq l_0 n v_1(\epsilon_0/\sqrt{3})^r p_i. \tag{66}$$

The probability that a column selected uniformly at random from X belongs to both $B_{\mathbf{x},\epsilon_0}$ and the sum of subspaces but not S^i is between $v_2(\epsilon_0/(1+\theta))^{2r} q_i$ and $v_3(\epsilon_0/(1-\theta))^{2r} q_i$. Therefore, the expected number of points satisfies

$$l_0 n v_2(\frac{\epsilon_0}{1+\theta})^{2r} q_i \leq E(T^2_{\mathbf{x},\epsilon_0}) \leq l_0 n v_3(\frac{\epsilon_0}{1-\theta})^{2r} q_i. \tag{67}$$

By the Chernoff's bound, we have

$$P(T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}} < E(T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}})/2) \leq \exp(-E(T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}})/8). \tag{68}$$

Combining (66) and (68), we have

$$P(T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}} < l_0 n v_1(\epsilon_0/\sqrt{3})^r p_i/2) \leq \exp(-l_0 n v_1(\epsilon_0/\sqrt{3})^r p_i/8). \tag{69}$$

Similarly, from the Chernoff's bound we have

$$P(T^2_{\mathbf{x},\epsilon_0} > 3E(T^2_{\mathbf{x},\epsilon_0})/2) \leq \exp(-E(T^2_{\mathbf{x},\epsilon_0})/12). \tag{70}$$

Combining (67) and (70), we have

$$P(T^2_{\mathbf{x},\epsilon_0} > 3l_0 n v_3(\frac{\epsilon_0}{1-\theta})^{2r} q_i/2) \leq \exp(-l_0 n v_2(\frac{\epsilon_0}{1+\theta})^{2r} q_i/12). \tag{71}$$

We choose l_0 sufficiently large such that

$$l_0 n v_1(\epsilon_0/\sqrt{3})^r p_i/2 \geq n. \tag{72}$$

From (69) and (71), the probability that (59) does not hold is no greater than $P_1 + (1 - P_1)P_2$, where $P_1 := \exp(-\frac{l_0 n v_1(\epsilon_0/\sqrt{3})^r p_i}{8})$, and $P_2 := \exp(-\frac{l_0 n v_2(\epsilon_0/(1+\theta))^{2r} q_i}{12})$. We choose l_0 sufficiently large such that

$$P_1 \leq \delta_0/(2s), \text{ and } P_2 \leq \delta_0/(2s) \tag{73}$$

hold. The conditions in (72) and (73) lead to the requirement

$$l_0 \geq \max(\frac{2}{v_1(\frac{\epsilon_0}{\sqrt{3}})^r p_i}, \frac{8\log(2s/\delta_0)}{n v_1(\frac{\epsilon_0}{\sqrt{3}})^r p_i}, \frac{12\log(2s/\delta_0)}{n v_2(\frac{\epsilon_0}{1+\theta})^{2r} q_i}). \tag{74}$$

Then combining (69), (71), (72), and (73), and taking the union bound over s subspaces, we have $T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}} \geq n$ and $T^2_{\mathbf{x},\epsilon_0}/T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}} \leq 3v_3(\sqrt{3}\epsilon_0)^r q_i/(v_1(1-\theta)^{2r} p_i)$ hold for any i with probability at least $1 - \delta_0$.

From the paragraph after Lemma 4, we know that the columns in the candidate set selected by Step (iii) belong to either S^i or any of S^{ij}'s (for every $j \neq i$) but not S^i. The number of columns in S^i is at least $T^1_{\mathbf{x},\frac{\epsilon_0}{\sqrt{3}}}$. The number of columns not in S^i is at most $T^2_{\mathbf{x},\epsilon_0}$. If the size of the candidate set is less than $2n$, (58) and (59) follow immediately.

If the candidate set has more than $2n$ columns, one could apply the Chernoff's bound and the union bound similarly and verify that the number of columns in the candidate sets are all $O(n)$ with probability at least $1 - \delta_0$. Also note that condition (22) in Theorem 1 implies

$$\frac{9v_3(\sqrt{3}\epsilon_0)^r q_i}{v_1(1-\theta)^{2r} p_i} \leq \eta_2 \frac{\underline{p}^2}{4(1+\frac{2\mu_1 r}{\underline{p}\sqrt{n}})^2 \mu_1^3 r^3 \log^6(4\beta_0 n)}, \forall i, \tag{75}$$

where the right hand side of (75) is much smaller than one. Step (iv) samples $2n$ columns uniformly from a candidate set with $O(n)$ columns. Applying the Chernoff's bound and the union bound, one can verify that when n is sufficiently large, with probability at least $1 - 2\delta_0$,

$$\hat{T}^1_{\mathbf{x}} \geq n, \text{ and } \hat{T}^2_{\mathbf{x}}/\hat{T}^1_{\mathbf{x}} \leq 9v_3(\sqrt{3}\epsilon_0)^r q_i/(v_1(1-\theta)^{2r} p_i), \tag{76}$$

hold for all the seeds $\mathbf{x}$ that belong to $S^i_{\epsilon_0}$ for any i. By taking the union bound, the claim follows. □

G. Proof of Lemma 6

Proof. Let γ_0 denote the probability that a column selected uniformly at random has at least t_0 observed indices in common with a fixed seed. Note that each seed has at least ζ_0 observed entries. Then we have $\gamma_0 \geq \sum_{j=t_0}^{\zeta_0} \binom{\zeta_0}{j} \underline{p}^j (1-\underline{p})^{\zeta_0 - j}$. Let $\tilde{n}$ denote the number of columns with t_0 or more observed indices in common with a fixed seed. By the Chernoff's bound, we have $P(\tilde{n} \leq \gamma_0 N/2) \leq \exp(-\gamma_0 N/8)$. Suppose $N \geq 2 l_0 \gamma_0^{-1} n$, then with probability at least $1 - s_0 \exp(-l_0 n/4)$ for each seed, there exist at least $l_0 n$ columns that have at least t_0 observed indices in common with the seed. Note that $s_0 \exp(-l_0 n/4)$ tends to zero exponentially in n, and

$$\gamma_0 \geq \underline{p}^{t_0} = \exp(t_0 \log \underline{p}) \geq (2 s_0 l_0 n/\delta_0)^{\frac{128\mu_1^2 r^2}{(1-3\theta)^2} \log \underline{p}}. \quad (77)$$

Then we have $N \geq 2 l_0 n (2 s_0 l_0 n/\delta_0)^{-\frac{128\mu_1^2 r^2}{(1-3\theta)^2} \log \underline{p}}$. By taking the union bound on Lemma 3, the arguments after Lemma 4 and Lemma 5, and Lemma 5, Lemma 6 follows. □

H. Proof of Lemma 7

Proof. Given a matrix $M \in \mathbb{R}^{n \times t}$, which can be decomposed as $M = L_0 + C_0$. The rank of L_0 is r. C_0 represents the column corruptions. Let γ denote the fraction of corrupted columns in M. ρ denotes the percentage of the observed entries in the non-corrupted columns. β_0 is the ratio of the number of non-corrupted columns to n.

Lemma 9 (Theorem 1 of [14]). *Consider an $n \times t$ matrix M and row and column spaces with coherences bounded above by some constant μ_1. Suppose M is uniformly random sampled, $n \geq 32$, $\beta_0 \geq 1$, $r \leq \overline{r}$, $\gamma \leq \overline{\gamma}$ and $\rho \geq \underline{\rho}$. If $(\overline{r}, \overline{\gamma}, \underline{\rho})$ satisfies*

$$\underline{\rho} \geq \eta_1 \mu_1^2 \overline{r}^2 \log^3(4\beta_0 n)/n \quad (78)$$

and

$$\frac{\overline{\gamma}}{1-\overline{\gamma}} \leq \eta_2 \frac{\underline{\rho}^2}{(1 + \frac{\mu_1 \overline{r}}{\underline{\rho}\sqrt{n}})^2 \mu_1^3 \overline{r}^3 \log^6(4\beta_0 n)}, \quad (79)$$

where η_1 and η_2 are absolute constants, then with probability at least $1 - cn^{-5}$ for constant $c > 0$, it holds that $\mathcal{P}_{\mathcal{I}_0^c}(L^*) = L_0$, $\mathcal{P}_{U_0}(L^*) = L^*$, and $\mathcal{I}^* = \mathcal{I}_0$, where (L^*, C^*) is the solution of the convex program

$$\min_{L,C} \|L\|_* + \lambda \|C\|_{1,2} \quad s.t.\ (L+C)_\Omega = M_\Omega \quad (80)$$

with $\lambda = \sqrt{\underline{\rho}/(\overline{\gamma}\mu_1 \overline{r} t \log^2(4\beta_0 n))}/48$.

Because the neighbors for each seed are identified by firstly selecting the columns that have at least t_0 overlaps with the observations of each seed, the partially observed local neighbor matrices are not sampled uniformly at random. We here apply the thinning process introduced in [18] to address this issue. We summarize the thinning process as follows.

Define a Bernoulli random variable y, which is '1' with probability ρ and '0' with probability $1-\rho$. Define a random variable z, which takes the values in $\{0, ..., t_0 - 1\}$ with the probability density $\mathcal{P}(z = i) = \binom{t}{i} p_0^i (1-p_0)^{t-i}/(1-\rho)$. For each column, draw an independent sample of y. If the sample is 1, keep the column. Otherwise, draw an independent sample of z. Select a random subset of size z from the observed entries in the support of the seed and discard the remainder. It is shown in [18] that after the thinning procedure, the resulting sampling is equivalent in distribution to uniform sampling with ρ. We then can apply the theoretical result of Lemma 9 to our problem setup.

Assume (60) and (61) hold, we here will show that $\{S^1, ...S^k\}$ belong to the candidate subspaces $\{S^{1^*}, ..., S^{s^*}\}$ returned by Subroutine 2 with probability at least $1 - s_0 cn^{-5}$ for constant $c > 0$. Given a seed $\mathbf{x}$ in S^i and its neighbor matrix, we treat the columns in S^i as non-corrupted columns and the columns not in S^i as the corrupted columns. Since (60), then $\beta_0 \geq 1$. We will show (78) and (79) are met for this neighbor matrix, and then Lemma 7 can be applied.

The random number of entries observed in the $n \times \beta_0 n$ non-corrupted matrix is $\hat{m} \sim \text{Binomial}(\underline{p}, \beta_0 n^2)$. By the Chernoff's bound, we have

$$\mathcal{P}(\hat{m} \leq \beta_0 n^2 \underline{p}/2) \leq \exp(-\beta_0 n^2 \underline{p}/8). \quad (81)$$

From (60), we have

$$\beta_0 n^2 \underline{p}/2 \geq \eta_1 \mu_1^2 r^2 \beta_0 n \log^3(4\beta_0 n). \quad (82)$$

Combining (81) and (82) and applying the union bound, we have with probability at least $1 - s_0 \exp(-\beta_0 n^2 \underline{p}/8)$ in all of the s_0 neighbor matrices, it holds that

$$\underline{\rho} = \hat{m}/(\beta_0 n^2) \geq \underline{p}/2 \geq \eta_1 \mu_1^2 r^2 \log^3(4\beta_0 n)/n. \quad (83)$$

Then (78) is met. Let γ denote the percentage of corrupted columns in the neighbor matrix. From the discussion after Lemma 4 and Lemma 5, we know $\gamma/(1-\gamma) \leq \hat{T}_{\mathbf{x}}^2/\hat{T}_{\mathbf{x}}^1$. Since $n^2 \underline{p} > r^2 n \log^3 n$ and $s_0 = O(n \log n)$, then $s_0 \exp(-\beta_0 n^2 \underline{p}/8)$ tends to zero exponentially in n. Since the right side of (61) increases as $\underline{p}$ increases, and $\underline{\rho} \geq \underline{p}/2$, then (79) is met. Then, Lemma 7 holds with probability at least $1 - s_0 \delta_0$, where δ_0 is cn^{-5} for some constant $c > 0$.

□

I. Proof of Lemma 8

Proof. Consider a column vector $\mathbf{x}$ in $\hat{S}^{ij}$. Let ω denote the set of indices of the observed entries in $\mathbf{x}$, and let $m = |\omega|$. From (8), (9), and Lemma 2, let $\delta = \delta_0$, if

$$m > 16 r \frac{\mu_1}{1-\theta} \log(4r/\delta_0)/3 \quad (84)$$

holds, we have that $\|\mathbf{x}_\omega - P_{\omega, \hat{S}^{ij}} \mathbf{x}_\omega\|_2^2 = 0$ with probability at least $1 - 4\delta_0$. If (84) and

$$0 < \frac{m(1-a) - 2r \frac{\mu_1}{1-\theta} \frac{(1+b)^2}{(1-c)}}{n} \|\mathbf{x} - P_{\hat{S}^{i'j'}} \mathbf{x}\|_2^2 \quad (85)$$

hold for any $(i', j') \neq (i, j)$, where $a = \sqrt{\frac{32\mu_1^2 r^2}{(1-3\theta)^2 m} \log(\frac{1}{\delta_0})}$, $b = \sqrt{\frac{8\mu_1 r}{1-3\theta} \log(\frac{1}{\delta_0})}$, $c = \sqrt{\frac{16 r \mu_1}{3(1-\theta)m} \log(\frac{4r}{\delta_0})}$, then by the union bound, we have $\|\mathbf{x}_\omega - \mathcal{P}_{\omega, \hat{S}^{ij}} \mathbf{x}_\omega\|_2 < \|\mathbf{x}_\omega - \mathcal{P}_{\omega, \hat{S}^{i'j'}} \mathbf{x}_\omega\|_2$ holds for any $(i', j') \neq (i, j)$ with probability at least $1 - 2k(k-1)\delta_0$. Then in order to let (84) and (85) hold in our setup, we must have

$$m > \max(\frac{16}{3} r \frac{\mu_1}{1-\theta} \log(\frac{4r}{\delta_0}), \frac{2r\mu_1(1+b)^2}{(1-\theta)(1-c)(1-a)}). \quad (86)$$

By the Chernoff's bound, we have

$$\mathcal{P}(m \leq np_0/2) \leq \exp(-np_0/8) < \delta_0, \tag{87}$$

and note that

$$np_0/2 \geq \eta_1 \mu_1^2 r^2 \log^3(4\beta_0 n). \tag{88}$$

We only need to show the right side of (88) is larger than the right side of (86). For the first term in the max of (86), there is an $O(r\log^2(n))$ gap between $\eta_1\mu_1^2 r^2 \log^3(4\beta_0 n)$ and $\frac{16}{3(1-\theta)} r\mu_1 \log(\frac{4r}{\delta_0})$. Therefore, the former is larger when n is sufficiently large. Note that $a = \Theta(1/\log(n))$ and $c = \Theta(1/\log(n))$. For the second term in the max of (86), a sufficiently large n can be found such that both $a \leq 1/2$ and $c \leq 1/2$. We then have

$$\frac{(1+b)^2}{(1-c)(1-a)} \leq 4(1+b)^2 \leq 8b^2 = \frac{64\mu_1 r}{1-3\theta}\log(\frac{1}{\delta_0}) \tag{89}$$

if b is greater than 2.5. Note that there is an $O(r\log^2(n))$ gap between $\eta_1\mu_1^2 r^2\log^3(4\beta_0 n)$ and $\frac{64\mu_1 r}{1-3\theta}\log(\frac{1}{\delta_0})$. Therefore, the former is larger when n is sufficiently large. From (87), we have that the inequality (86) holds with probability $1-\delta_0$.

By the union bound, we have that if the column $\mathbf{x} \in \hat{S}^{ij}$,

$$0 = \|\mathbf{x}_\omega - P_{\omega,\hat{S}^{ij}}\mathbf{x}_\omega\|_2 < \|\mathbf{x}_\omega - P_{\omega,\hat{S}^{i'j'}}\mathbf{x}_\omega\|_2 \tag{90}$$

holds for any $(i',j') \neq (i,j)$ with probability at least $1-(2k(k-1)+1)\delta_0$. The analysis for the case that the column $\mathbf{x}$ belongs to exact one subspace is similar with the analysis above. If $\mathbf{x} \in \hat{S}^i$, we have that

$$0 = \|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^{ij}}\mathbf{x}_\omega\|_2 < \|\mathbf{x}_\omega - \mathcal{P}_{\omega,\hat{S}^{i'j'}}\mathbf{x}_\omega\|_2 \tag{91}$$

holds for any $i' \neq i$ with probability at least $1-(2k(k-1)+1)\delta_0$. Then by the union bound, we have lemma 8 hold with probability at least $1-(4k(k-1)+2)\delta_0$. □

J. Skeleton-Proof of Theorem 2

The proof of Theorem 2 follows the same line as the proof of Theorem 1 with some modifications to address the noise. Please refer to Appendix K, L, M, and N for the proofs of Lemmas 10, 11, 12, and 13.

1) Selection of local neighbor matrix: Lemma 10 is the counterpart of Lemma 3.

Lemma 10. *Assume **A3** holds. If the number of seeds,*

$$s \geq \frac{\log k + \log\frac{1}{\delta_1}}{(1-\exp(-\frac{\log((\beta+1)n)\max(\frac{\beta}{\alpha_1},\frac{\log^2 n}{\alpha_2})}{\beta}))v_0' p_*}, \tag{92}$$

then with probability at least $1-\delta_1$, for each $i = 1,...,k$, at least one seed has its noiseless counterpart in $S^i_{\sigma_2}$, and each seed has at least ζ_1 observed entries for some constants $\alpha_1 \geq 1$ and $\alpha_2 \geq 1$.

We pick $t_1 := \frac{128\mu_1^2 r^2}{(1-3\theta)^2}\log(\frac{2s_1 l_1 n}{\delta_1})$. With the same arguments after Lemma 4, with probability at least $1-\delta_1$, (56) holds for all columns selected in Step (ii) for all neighbor matrices with the seeds that their noiseless counterpart are in $S^i_{\sigma_2}$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ represent the noiseless counterparts of a seed and a selected column for that seed respectively. Then, given a seed $\tilde{\mathbf{x}}$ with its noiseless counterpart $\mathbf{x}$, if a column $\mathbf{x}'$ in X satisfies $\|\mathbf{x}-\mathbf{x}'\|_2 \leq \frac{\epsilon_0 - 2\sqrt{2}\epsilon_1}{\sqrt{3}}$, then from (56), we have $\sqrt{\frac{n}{q}}\|(\mathbf{x}-\mathbf{x}')_\omega\|_2 \leq \frac{\epsilon_0}{\sqrt{2}} - 2\epsilon_1$. Since each entry of ξ is uniformly distributed in $[-\frac{\epsilon_1}{\sqrt{n}}, \frac{\epsilon_1}{\sqrt{n}}]$, from triangle inequality, we have that $\sqrt{\frac{n}{q}}\|(\tilde{\mathbf{x}}-\tilde{\mathbf{x}}')_\omega\|_2 \leq \frac{\epsilon_0}{\sqrt{2}}$. Thus, $\tilde{\mathbf{x}}'$ can be selected in step (iii) of Subroutine 1. Conversely, if $\tilde{\mathbf{x}}'$ is selected in Subroutine 1, we have $\sqrt{\frac{n}{q}}\|(\mathbf{x}-\mathbf{x}')_\omega\|_2 \leq \frac{\epsilon_0}{\sqrt{2}} + 2\epsilon_1$. Then from (56), we have $\|\mathbf{x}-\mathbf{x}'\|_2 \leq \epsilon_0 + 2\sqrt{2}\epsilon_1$. For each seed $\tilde{\mathbf{x}}$ with its noiseless counterpart $\mathbf{x}$ in $S^i_{\sigma_2}$, let $\hat{T}^1_{\tilde{\mathbf{x}}}$ and $\hat{T}^2_{\tilde{\mathbf{x}}}$ denote the number of columns with noiseless counterparts in S^i and not in S^i, respectively, for its local neighbor matrix. Lemmas 11 and 12 are counterparts of Lemmas 5 and 6 with a change of parameters.

Lemma 11. *Assume **A3** holds. If the number of columns selected for each seed in Step (ii), ln, satisfies*

$$l \geq \max(\frac{2}{v_1'\sigma_1^r p_*}, \frac{8\log(2s/\delta_1)}{nv_1'\sigma_1^r p_*}, \frac{12\log(2s/\delta_1)}{nv_2'(\frac{\sigma_2}{1+\theta})^{2r}q_*}), \tag{93}$$

then with probability at least $1-4\delta_1$, it holds that

$$\hat{T}^1_{\tilde{\mathbf{x}}} \geq n, \tag{94}$$

$$\text{and} \quad \hat{T}^2_{\tilde{\mathbf{x}}}/\hat{T}^1_{\tilde{\mathbf{x}}} \leq 9v_3'(\frac{\sigma_2}{1-\theta})^{2r}q_i/(v_1'\sigma_1^r p_i), \forall i, \tag{95}$$

for the local neighbor matrices of all the seeds with their noiseless counterparts $\mathbf{x}$ in $S^i_{\sigma_2}$ for any i.

Lemma 12. *If $N \geq 2l_1 n(2s_1 l_1 n/\delta_1)^{-\frac{128\mu_1^2 r^2}{(1-3\theta)^2}\log\underline{p}}$ and $\zeta_1 > t_1$, then Subroutine 1 satisfies the following, with probability at least $1-6\delta_1$, at least one seed has its noiseless counterpart $\mathbf{x}$ in $S^i_{\sigma_2}$, for $i = 1,...,k$, and (94) and (95) hold.*

2) Local subspace estimation: [28] provides the theoretical guarantee of low-rank matrix completion in the presence of noise when some columns do not belong to the same subspace. Building on [28] and applying the technique in [18], we show that $\{S^1,...,S^k\}$ can be recovered when the noisy local neighbor matrices contain columns outside the subspace.

Lemma 13. *Suppose there exists at least one seed with its counterpart $\mathbf{x}$ in S^i for $i = 1,...,k$. Denote the dimension of its neighbor matrix by $n \times \beta_i n$. All neighbor matrices are observed uniformly at random with probability $p_0 \geq \underline{p}$, if*

$$\underline{p} \geq 8\log((\beta_i+1)n)\max(\frac{\beta_i}{\alpha_1}, \frac{\log^2 n}{\alpha_2})/(\beta_i n), \tag{96}$$

then with probability at least $1-s_1\delta_1$, we have

$$\begin{aligned}\frac{\|L^{i^*}-L^i\|_F^2 + \|C^{i^*}-C^i\|_F^2}{\beta_i n^2} \leq \frac{2\mu_1 r\hat{T}^2_{\tilde{\mathbf{x}}}}{n(\hat{T}^1_{\tilde{\mathbf{x}}}+\hat{T}^2_{\tilde{\mathbf{x}}})} + \\ 3\alpha_3\phi^2\log((\beta_i+1)n)(1+\frac{\hat{T}^2_{\tilde{\mathbf{x}}}}{\hat{T}^1_{\tilde{\mathbf{x}}}})^2(\frac{2r}{np_0} + \frac{3\hat{T}^2_{\tilde{\mathbf{x}}}}{\hat{T}^1_{\tilde{\mathbf{x}}}+\hat{T}^2_{\tilde{\mathbf{x}}}})\end{aligned} \tag{97}$$

with properly chosen $\hat{\lambda}_i$'s and $\tilde{\lambda}_i$'s, where L^{i^*} and C^{i^*} are the solutions returned by Subroutine 4, and constants $\alpha_1 \geq 1$, $\alpha_2 \geq 1$, $\alpha_3 > 0$, and $\phi = \max(\frac{\epsilon_1}{\sqrt{n}}, \sqrt{2\mu_1\frac{r}{n}})$.

Please refer to Lemma 14 for the exact definition of $\hat{\lambda}_i$'s and $\tilde{\lambda}_i$'s used in Subroutine 4. By applying the union bound to lemma 12 and lemma 13, we then can prove that Theorem 2 holds with probability at least $(1-(6+s_1))\delta_1$.

K. Proof of Lemma 10

Proof. From (28), we have the expected number of observed entries per column is at least $8\log((\beta+1)n)\max(\frac{\beta}{\alpha_1}, \frac{\log^2 n}{\alpha_2})/\beta$. The rest part of the proof follows the same line as the proof of Lemma 3. □

L. Proof of Lemma 11

Proof. For each seed $\tilde{\mathbf{x}}$ with its noiseless counterpart $\mathbf{x}$ in $S^i_{\sigma_2}$, among the ln columns selected by Step (ii), let $T^1_{\tilde{\mathbf{x}},\sigma_1}$ denote the number of columns with noiseless counterparts in both $B_{\mathbf{x},\sigma_1}$ and S^i, and let $T^2_{\tilde{\mathbf{x}},\sigma_2}$ denote the number of columns with its noiseless counterparts in both $B_{\mathbf{x},\sigma_2}$ and any S^{ij} but not S^i. The probability that a randomly selected column has its noiseless counterpart in both $B_{\mathbf{x},\sigma_1}$ and S^i is at least $v'_1\sigma_1^r p_i$. Therefore, the expectation of $T^1_{\tilde{\mathbf{x}},\sigma_1}$ is at least $lnv'_1\sigma_1^r p_i$. The probability that a randomly selected column has its noiseless counterpart in both $B_{\mathbf{x},\sigma_1}$ and any of S^{ij} but not S^i is between $v'_2(\frac{\sigma_2}{1+\theta})^{2r}q_i$ and $v'_3(\frac{\sigma_2}{1-\theta})^{2r}q_i$. The expectation of $T^2_{\tilde{\mathbf{x}},\sigma_2}$ is then between $lnv'_2(\frac{\sigma_2}{1+\theta})^{2r}q_i$ and $lnv'_3(\frac{\sigma_2}{1-\theta})^{2r}q_i$. The rest part of the proof follows the same line as the proof of Lemma 5. □

M. Proof of Lemma 12

Proof. The proof follows the same line as the proof of Lemma 6 and takes the union bound over Lemma 10, Lemma 11, and the arguments in between. □

N. Proof of Lemma 13

Proof. Given a matrix $M = L_0 + C_0$. The rank of L_0 is r. C_0 represents the column corruptions. γ denotes the fraction of corrupted columns in M. The entries of noise matrix N are uniformly distributed in $[-\frac{\epsilon_1}{\sqrt{n}}, \frac{\epsilon_1}{\sqrt{n}}]$. ψ_1 denotes the total number of the observed entries. ψ_2 denotes the number of the observed entries in the non-corrupted columns.

Lemma 14 (Corollary 7 of [28])**.** *Consider an $n \times t$ noisy matrix $\tilde{M}$ which is uniformly random sampled. Assume the noise variables are $\frac{\epsilon_1}{\sqrt{n}}$-subgaussian, $\|L_0\|_\infty \le \sqrt{2\mu_1 r/n}$ and $\|C_0\|_\infty \le \sqrt{2\mu_1 r/n}$. Let $\phi = \max(\epsilon_1/\sqrt{n}, \sqrt{2\mu_1 r/n})$. Let $(\hat{L}, \hat{C})$ be the solution of the convex program*

$$\begin{aligned} &\min_{L,C} \|(\tilde{M}-L-C)_\Omega\|_F^2/|\Omega| + \lambda_1\|L\|_* + \lambda_2\|C\|_{1,2} \\ &\text{s.t.} \|L\|_\infty \le \sqrt{2\mu_1 r/n} \text{ and } \|C\|_\infty \le \sqrt{2\mu_1 r/n} \end{aligned} \tag{98}$$

with $\lambda_1 = c_1\phi\sqrt{\frac{\alpha_2\log(n+t)}{\psi_1 n}}$, and $\lambda_2 = \alpha_1 c_1\phi\sqrt{\frac{\log(n+t)}{\psi_1 t}}$, where constants $c_1 > 0$, $\alpha_1 \ge 1$, and $\alpha_2 \ge 1$. If we have $\psi_2 \ge 2\log(n+t)\max(t/\alpha_1, n\log^2 n/\alpha_2)$, then with probability at least $1-6/(n+t)$,

$$\begin{aligned} &(\|\hat{L}-L_0\|_F^2 + \|\hat{C}-C_0\|_F^2)/(nt) \\ &\le \alpha_3\phi^2\log(n+t)\frac{\psi_1}{\psi_2}\left(\frac{rt}{\psi_2} + \frac{\psi_1-\psi_2}{\psi_2}\right) + \frac{2\mu_1 r\gamma}{n}, \end{aligned} \tag{99}$$

where $\alpha_3 > 0$ can only depend on α_1 and α_2.

The noise variables N_{ij} are $\frac{\epsilon_1}{\sqrt{n}}$-subgaussian since they are uniformly distributed in $[-\frac{\epsilon_1}{\sqrt{n}}, \frac{\epsilon_1}{\sqrt{n}}]$. Given a seed with its counterpart $\mathbf{x}$ in S^i and its neighbor matrix, we here also need to apply the thinning process summarized in the proof of Lemma 7. We then can apply the theoretical result of Lemma 14 to our problem setup. We treat the columns with their counterparts in S^i as non-corrupted columns and those not in S^i as the corrupted columns. The matrix has dimension $n \times \beta_i n$, and the percentage of corrupted columns is denoted by γ_i. From the discussion after Lemma 10 and Lemma 11, we know $\gamma_i/(1-\gamma_i) \le \hat{T}^2_{\mathbf{x}}/\hat{T}^1_{\mathbf{x}}$. We assume $\gamma_i \le 1/2$ here. One can check the discussion after Theorem 2 that this is implied by the requirements to obtain diminishing recovery error of S^i. The random number of entries observed in the $n \times \beta_i n/2$ non-corrupted matrix is $\psi_2 \sim$ Binomial$(p_0, \beta_i n^2/2)$. By the Chernoff's bound, we have $\mathcal{P}(\psi_2 \le \beta_i n^2 \underline{p}/4) \le \exp(-\beta_i n^2 p_0/16)$. From (96), we have $\frac{\beta_i n^2 p_0}{4} \ge 2\log((\beta_i+1)n)\max(\frac{\beta_i n}{\alpha_1}, \frac{n\log^2 n}{\alpha_2})$. Applying the union bound, we have

$$\psi_2 \ge 2\log((\beta_i+1)n)\max(\beta_i n/\alpha_1, n\log^2 n/\alpha_2) \tag{100}$$

in each of the s_1 neighbor matrices with probability at least $1-s_1\exp(-\beta_i n^2 p_0/16)$. Note that $s_1\exp(-\beta_i n^2 p_0/16)$ tends to zero exponentially in n. By the Chernoff's bound we have

$$\begin{aligned} &P(\psi_1 > 3\beta_i n^2 p_0/2) \le \exp(-\beta_i n^2 p_0/12), \\ &P(\psi_2 < (1-\gamma_i)\beta_i n^2 p_0/2) \le \exp(-(1-\gamma_i)\beta_i n^2 p_0/8), \\ &\text{and } P(\psi_1 - \psi_2 > 3\gamma_i\beta_i n^2 p_0/2) \le \exp(-\gamma_i\beta_i n^2 p_0/12). \end{aligned} \tag{101}$$

Note that $\exp(-\beta_i n^2 p_0/12)$ and $\exp(-\beta_i n^2 p_0/8)$ tend to zero exponentially in n. We then have (97) holds with probability at least $1-s_1\delta_1$ for all s_1 neighbor matrices. □