Matrix Completion with Columns in Union and Sums of Subspaces

Pengzhi Gao, Meng Wang, and Joe H. Chow
Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY, USA
{gaop, wangm7, chowj}@rpi.edu

Matthew Berger, and Lee M. Seversky
Information Directorate
Air Force Research Laboratory, Rome, NY, USA
{matthew.berger.1, Lee.Seversky}@us.af.mil

***Abstract*—Motivated by missing data recovery for power system monitoring, we study the problem of matrix completion when each column belongs to either one of a few low-dimensional subspaces or the sum of a subset of subspaces. This novel model "union and sums of subspaces" is a generalization of the model of a union of subspaces (UoS) studied under subspace clustering. It characterizes the low-dimensional nonlinear structure in high-dimensional data points. We propose a convex-optimization-based method to recover missing entries under this model and provide its theoretical guarantee. Numerical experiment on power system data is conducted to verify the effectiveness of the method.**

***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 [8], [34], machine learning [2], [3], remote sensing [32], system identification [25], power system monitoring [15], etc. Due to data correlations, a collection of data points often exhibits low-dimensional structures despite its high dimension. The low-dimensional structures have been utilized for various applications, such as dimensionality reduction and data compression [5], [11], [31], Internet topology estimation [13], and power system monitoring [15], [24]. Under this model, the missing data recovery problem can be formulated as a low-rank matrix completion problem. If an $n \times N$ $(n < N)$ real-valued matrix is rank r $(r \ll n)$, the matrix can be correctly recovered as long as $O(rN \log^2 N)$[1] randomly selected entries are observed ([6], [7], [17], [30]).

[12], [13], [18] considered matrix completion when the columns belong to a union of k r-dimensional subspaces in $\mathbb{R}^n$. Since kr can be greater than n, the matrix is no longer low-rank. Nevertheless, [13] shows that one can still recover the matrix from $O(rN \log^2 n)$ observed entries. Matrix completion under such model is equivalent to subspace clustering with missing data. Subspace clustering [29], [36] aims to learn the subspaces from data points that belong to the union of subspaces and assign each point correctly to its corresponding subspace. Numerous methods (e.g., [21], [23], [27], [33], [35], [37]) have been developed for subspace clustering in diverse fields such as disease clustering [26], user identification in recommender systems [38] and music analysis [20].

In some applications, the model of a union of subspaces is no longer accurate. In power system monitoring, each column represents a quantity like a bus voltage or a line current across time. Because a disturbance affects nearby quantities, and the changes of these quantities are highly correlated, the corresponding columns belong to the same subspace. When multiple disturbances happen, and the affected regions partially overlap, the corresponding column of a quantity affected by multiple disturbances belongs to the sum of multiple subspaces. We say that the columns in this case belong to the "union and sums" of subspaces and consider the missing data recovery and subspace identification under this model.

Our contributions are threefold. (1) We propose a model to characterize the low-dimensional nonlinear structure in high-dimensional data points. (2) We propose a missing data recovery method under such model and provide its theoretical guarantee. We show that $O(r^2 N \log^3 n)$ observations are sufficient to determine the missing points of every column of an $n \times N$ matrix with columns belonging to the "union and sums" of k r-dimensional subspaces. (3) Numerical experiment on power system data is conducted to verify the proposed method.

The rest of the paper is organized as follows. We summarize the problem formulation and propose our method and theoretical guarantee in Section II. We provide the skeleton-proof of the main theorem in Section III. Section IV records our numerical experiment. We conclude the paper in Section V.

II. Problem Formulation and Main Results

A. Problem formulation and assumptions

Consider k linearly independent subspaces in $\mathbb{R}^n$, denoted by $S_1, ..., S_k$. Each subspace is at most r-dimensional. Denote the sum of two subspaces[2] S_i and S_j $(i \neq j)$ as $S_{i,j}$. $x \in S_{i,j}$ if and only if there exist $y \in S_i$ and $z \in S_j$ such that $x = y + z$. Assume $S_i \subsetneq S_{i',j'}$, $\forall i', j' \neq i$. An example of union and sums of subspaces is shown in Fig. 1. The coherence of an r-dimensional subspace S_i is defined as

$$\mu(S_i) := \frac{n}{r} \max_j \|\mathcal{P}_{S_i} e_j\|_2^2, \tag{1}$$

[1] We use the big O notation $g(n) \in O(f(n))$ if as n goes to infinity, $g(n) \leq C \cdot f(n)$ eventually holds for some positive constant C.

[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.

where $\{e_j\}$ are the canonical unit vectors for $\mathbb{R}^n$, and $\mathcal{P}_{S_i}$ is the projection operator onto S_i. We have

$$\mathcal{P}_{S_i} := U^i((U^i)^T U^i)^{-1}(U^i)^T, \tag{2}$$

where U^i is the orthonormal column span of S_i. The coherence of $x \in \mathbb{R}^n$ is defined as $\mu(x) := n\|x\|_\infty^2/\|x\|_2^2$. Note that $\|x\|_\infty$ is the maximum absolute value of the entries of x.

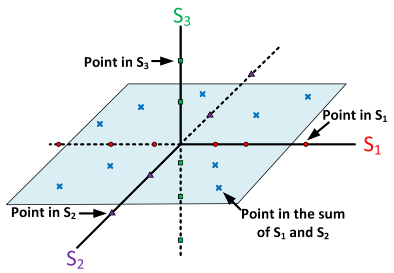

Fig. 1: Lines S_1, S_2 and S_3 are three one-dimensional subspaces in $\mathbb{R}^3$. x-marks stand for the points in $S_{1,2}$.

Consider a real-valued $n \times N$ matrix X. The l_2-norm of each column is at most one. We say X is in the "union and sums" of $\{S_1, ..., S_k\}$ if the columns of X belong to the union of S_i and $S_{i,j}$ for $i, j \in \{1, ..., k\}$. p_i denotes the fraction of columns belong to S_i. p_{ij} denotes the fraction of columns belong to S_{ij}. Define $q_i = \sum_{j \neq i} p_{ij}$, $p_* = \min_i (p_i)$ and $q_* = \min_i (q_i)$. Given a vector $x \in \mathbb{R}^n$ and a positive constant ϵ, we define the ball centered at x_0 with radius ϵ as

$$B_{x_0,\epsilon} = \{x \in \mathbb{R}^n \mid \|x - x_0\|_2 \leq \epsilon\}. \tag{3}$$

Define $S_{i,\epsilon}$ as the subset of points in S_i that are at least ϵ distance away from other subspaces and their sums. For the partially observed columns x_1 and x_2, define the partial distance as $\sqrt{n/q}\|(x_1 - x_2)_\omega\|_2$, where ω is the common set of indices of size q where both x_1 and x_2 are observed. Given matrix M, $\|M\|_*$ denotes its Nuclear Norm, which is the sum of singular values. $\|M\|_{1,2}$ denotes the sum of the l_2-norms of its columns. We assume X satisfies the following assumptions.

A1. The coherences of S_i and $S_{i,j}$ are bounded above by μ_0. The coherence of each column is bounded above by μ_1. For any pair of columns, x_1 and x_2, the coherence of $x_1 - x_2$ is also bounded above by μ_1.

A2. Let $0 < \epsilon_0 < 1$. There exist constants $0 < v_0 \leq 1$ and $0 < v_1 < v_2 \leq 1$, depending on ϵ_0 such that: **(i)** The probability that a column selected uniformly at random belongs to S_{i,ϵ_0} is at least $v_0 p_i$. **(ii)** If x belongs to S_{i,ϵ_0}, the probability that a column selected uniformly at random belongs to both $B_{x,\epsilon_0/\sqrt{3}}$ and S_i is at least $v_0(\epsilon_0/\sqrt{3})^r p_i$, and the probability that it belongs to both B_{x,ϵ_0} and any of $S_{i,j}$'s ($j \neq i$) is between $v_1\epsilon_0^{2r} q_i$ and $v_2\epsilon_0^{2r} q_i$.

Disturbances happen in power systems constantly [22]. Each column of X represents the observations of a bus voltage or line current across time. One disturbance may affect nearby voltages and currents, and the corresponding columns belong to the same subspace, as discussed in [15]. When multiple disturbances happen, quantities affected by multiple overlapping disturbances belong to the sum of subspaces.

Subroutine 1 Local Neighbor Matrix Selection

Input: s_0, l_0, ζ_0, ϵ_0 and t_0.

(i) Choose s_0 columns uniformly at random as seeds and discard all with less than ζ_0 observations.

(ii) For each seed, find all columns with at least t_0 observations at locations observed in the seed, and randomly select $l_0 n$ columns from each such set.

(iii) Form local neighbor matrix by selecting the columns with partial distance less than $\epsilon_0/\sqrt{2}$ from each seed.

Return: s' ($s' \leq s_0$) neighbor matrices $\{M_\Omega^1, ..., M_\Omega^{s'}\}$.

Subroutine 2 Robust Subspace Completion for matrix M_Ω^i

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

$$\min_{L,C} \ \|L\|_* + \lambda_i\|C\|_{1,2} \qquad \text{s.t. } (L + C)_\Omega = M_\Omega^i \tag{4}$$

Compute the SVD of $L_i^* = U_i^* \Sigma_i^* V_i^{*T}$.

Return: The subspace S_i^*, represented by U_i^*.

Data losses often happen in power system due to communication congestion and device malfunction. To model the data losses, we assume each entry of X is observed independently with probability p_0. Ω denotes the set of observed entries. M_Ω denotes the observed entries in any submatrix M of X. Our goal is to recover $S_1, ..., S_k$ and the missing entries from X_Ω and classify each column to its corresponding subspace.

B. Missing data recovery and its theoretical guarantee

We first define some key quantities used in our method.

$$\delta_0 := cn^{-5}, \ \ s_0 := \lceil \frac{\log k + \log \frac{1}{\delta_0}}{(1 - \exp(-\frac{\eta_1 \mu_0^2 r^2 \log^3(4\beta_0 n)}{4v_0}))v_0 p_*} \rceil,$$

$$l_0 := \lceil \max(\frac{2}{v_0(\frac{\epsilon_0}{\sqrt{3}})^r p_*}, \frac{8\log(2s_0/\delta_0)}{nv_0(\frac{\epsilon_0}{\sqrt{3}})^r p_*}, \frac{12\log(2s_0/\delta_0)}{nv_1\epsilon_0^{2r} q_*}) \rceil,$$

$$\zeta_0 := \eta_1\mu_0^2 r^2 \log^3(4\beta_0 n)/v_0 \text{ and } t_0 := \lceil 8\mu_1^2 \log(2s_0 l_0 n/\delta_0) \rceil,$$

for some positive constant c and η_1. Note that $\beta_0 \geq 1$ denotes the ratio of the number of columns that belongs to exactly one subspace to n in the neighbor matrices, and $\beta_0 \leq l_0$.

We first pick s_0 columns, called seeds, uniformly at random from X_Ω and select $l_0 n$ nearest neighbors to form a local neighbor matrix M_Ω^i for each seed (Subroutine 1). Secondly, Subroutine 2 is applied to each M_Ω^i to estimate the subspace. Thirdly, $\{\hat{S}_1, ..., \hat{S}_k\}$ are identified from a set of candidate subspaces and used as the estimation of $\{S_1, ..., S_k\}$. Assign each partially observed column x_ω to its closest $\hat{S}_{ij}$ and use its closest point in $\hat{S}_{ij}$ as an estimate of x (Subroutine 3).

Our method differs from [13] mainly in Subroutine 2. One recovers S_i from a group of partially observed columns. Under the UoS model, all these points belong to S_i, and low-rank matrix completion methods can be applied to identify S_i, as described in [13]. Under our generalized model, however, part

Subroutine 3 Subspace Refinement and Erasure Estimation

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

For each partially observed column x_ω, find

$$(i^*, j^*) = \arg\min_{i,j}(\|x_\omega - \mathcal{P}_{\omega,\hat{S}_{i,j}} x_\omega\|_2), \quad (5)$$

Estimate x by $\hat{x} = \hat{U}^{i^*,j^*}((\hat{U}_\omega^{i^*,j^*})^T \hat{U}_\omega^{i^*,j^*})^{-1}(\hat{U}_\omega^{i^*,j^*})^T x_\omega$.

Return: $\{\hat{S}_1, ..., \hat{S}_k\}$ and $\hat{X}$.

of these points belong to $S_{i,j}$ for some j, which can be viewed as a matrix separation problem [9]. We use (4) to recover S_i. Its solution L^* represents a low-rank matrix containing points in the low-dimensional subspace, and C^* represents a column-sparse matrix containing points not in the subspace.

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

$$\underline{p} \geq 2\eta_1\mu_0^2 r^2 \log^3(4\beta_0 n)/(nv_0), \quad (6)$$

$$N \geq 2l_0 n(2s_0 l_0 n/\delta_0)^{-8\mu_1^2 \log \underline{p}}, \quad (7)$$

$$q_i \leq \frac{\eta_2 v_0}{12 v_2(\sqrt{3}\epsilon_0)^r} \frac{\underline{p}^2}{(1+\frac{2\mu_0 r}{\underline{p}\sqrt{n}})^2 \mu_0^3 r^3 \log^6(4\beta_0 n)} p_i, \ \forall i, \quad (8)$$

with probability at least $1-(3+s_0)\delta_0$, subspaces $\{S_1, ..., S_k\}$ are correctly recovered for some $\beta_0 \geq 1$ and positive constants η_1 and η_2. Furthermore, with probability at least $1-(4k(k-1)+2)\delta_0$, every fixed column of X is correctly recovered.

III. Skeleton-Proof of Theorem 1

The proof follows the same line as the proof of Theorem 2.1 in [13]. The skeleton-proof will be presented according to the steps of the method. Please refer to [14] for the proofs.

A. Selection of local neighbor matrix

Lemma 1. *Assume **A1** holds. If the number of seeds,*

$$s \geq \frac{\log k + \log \frac{1}{\delta_0}}{(1-\exp(-\frac{\eta_1 \mu_0^2 r^2 \log^3(4\beta_0 n)}{4v_0}))v_0 p_*}, \quad (9)$$

then with probability at least $1-\delta_0$, for all $i = 1, ..., k$, at least one seed is in S_{i,ϵ_0}, and each seed has at least ζ_0 observed entries for some $\beta_0 \geq 1$ and absolute constant η_1.

Lemma 1 guarantees that if enough seeds are selected, then with high probability at least one seed is in S_{i,ϵ_0}, $\forall i$, and has at least ζ_0 observed entries.

Lemma 2 (Lemma 3 of [13]). *Assume **A1** holds and let $y = x_1 - x_2$. Assume there is a common set of indices ω of size $q \leq n$ where both x_1 and x_2 are observed. Let y_ω denote the corresponding subset of y. Then for any $\delta > 0$, if $q \geq 8\mu_1^2 \log(2/\delta)$, then with probability at least $1-\delta$*

$$\|y\|_2^2/2 \leq n\|y_\omega\|_2^2/q \leq 3\|y\|_2^2/2. \quad (10)$$

We pick $t_0 := \lceil 8\mu_1^2 \log(2s_0 l_0 n/\delta_0) \rceil$, then $\delta = \delta_0/(s_0 l_0)$ in Lemma 2. From Lemma 2 and the union bound, we know that with probability at least $1-\delta_0$, (10) holds for all $s_0 l_0 n$ columns selected in Step (ii) for all s_0 neighbor matrices, where x_1 and x_2 represent a seed and a selected column for that seed respectively. 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). 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 $x \in S_{i,\epsilon_0}$, $T^1_{x,\epsilon_0/\sqrt{3}}$ denotes the number of columns belong to both $B_{x,\epsilon_0/\sqrt{3}}$ and S_i, and T^2_{x,ϵ_0} denotes the number of columns belong to both B_{x,ϵ_0} and any of $S_{i,j}$'s ($j \neq i$). In the neighbor matrix of x, at least $T^1_{x,\epsilon_0/\sqrt{3}}$ columns belong to S_i, and at most T^2_{x,ϵ_0} columns not in S_i.

Lemma 3. *Assume **A2** holds. If the number of columns selected for each seed, ln, such that,*

$$l \geq \max(\frac{2}{v_0(\frac{\epsilon_0}{\sqrt{3}})^r p_*}, \frac{8\log(2s/\delta_0)}{nv_0(\frac{\epsilon_0}{\sqrt{3}})^r p_*}, \frac{12\log(2s/\delta_0)}{nv_1\epsilon_0^{2r} q_*}), \quad (11)$$

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

$$T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} \geq n, \quad (12)$$

$$\text{and} \quad T^2_{x,\epsilon_0}/T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} \leq 3v_2(\sqrt{3}\epsilon_0)^r q_i/(v_0 p_i), \ \forall i, \quad (13)$$

for all the seeds x that belong to S_{i,ϵ_0} for any i.

It guarantees that the local neighbor matrix of seed x in S_{i,ϵ_0} has at least n columns in S_i and the number of columns not in S_i is upper bounded by (13) when l is sufficiently large.

Lemma 4. *If $N \geq 2l_0 n(2s_0 l_0 n/\delta_0)^{-8\mu_1^2 \log \underline{p}}$ and $\zeta_0 > t_0$, then Subroutine 1 satisfies the following, at least one seed x belongs to S_{i,ϵ_0} for each $i = 1, ..., k$, and (12) and (13) hold with probability at least $1-3\delta_0$.*

Lemma 4 says that if N is sufficiently large, at least one seed x is in S_i, $\forall i$. Its neighbor matrix has at least n columns in S_i, and the ratio of the number of columns not in S_i to the number of columns in S_i is at most $3v_2(\sqrt{3}\epsilon_0)^r q_i/(v_0 p_i)$.

B. Local subspace estimation

We here show that $\{S_1, ..., S_k\}$ can be correctly identified even when local neighbor matrices contain columns outside the subspace of interest.

Lemma 5. *Suppose there exists at least one seed x in S_i for every S_i, $i = 1, ..., k$ such that (12) holds, for every S_i, $i = 1, ..., k$, Assume $n \geq 32$. All neighbor matrices are observed uniformly at random with probability $p_0 \geq \underline{p}$. If*

$$\underline{p} \geq 2\eta_1\mu_0^2 r^2 \log^3(4\beta_0 n)/(nv_0), \quad (14)$$

$$\text{and} \quad T^2_{x,\epsilon_0}/T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} \leq \eta_2 \frac{p^2}{4(1+\frac{2\mu_0 r}{\underline{p}\sqrt{n}})^2 \mu_0^3 r^3 \log^6(4\beta_0 n)}, \tag{15}$$

for at least one seed x in S_i, $\forall i$, 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 absolute constants.

Please refer to [9] for the exact definition of λ_i. Augmented Lagrange Multiplier (ALM) method [9] can be utilized to facilitate efficient solution to Subroutine 2 with a quasilinear convergence speed. The condition (8) in Theorem 1 implies

$$\frac{3v_2(\sqrt{3}\epsilon_0)^r q_i}{v_0 p_i} \leq \eta_2 \frac{p^2}{4(1+\frac{2\mu_0 r}{\underline{p}\sqrt{n}})^2 \mu_0^3 r^3 \log^6(4\beta_0 n)}, \forall i. \tag{16}$$

By applying the union bound to lemma 4 and lemma 5, we can prove that $\{S_1, ..., S_k\}$ belongs to the output of Subroutine 2 with probability at least $(1 - (3 + s_0))\delta_0$.

For a seed not in S_{i,ϵ_0}, the recovered subspace of its neighbor matrix might not belong to $\{S_1, ..., S_k\}$. In this case, this subspace has rank greater than r 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.

C. Matched subspace assignment

Lemma 6. *Assume **A1** holds, and p_0 satisfies*

$$p_0 \geq \underline{p} \geq 2\eta_1 \mu_0^2 r^2 \log^3(4\beta_0 n)/(nv_0). \tag{17}$$

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

$$0 = \|x_\omega - \mathcal{P}_{\omega,\hat{S}_{i,j}} x_\omega\|_2 < \|x_\omega - \mathcal{P}_{\omega,\hat{S}_{i',j'}} x_\omega\|_2, \ \forall j, j'; \tag{18}$$

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

$$0 = \|x_\omega - \mathcal{P}_{\omega,\hat{S}_{i,j}} x_\omega\|_2 < \|x_\omega - \mathcal{P}_{\omega,\hat{S}_{i',j'}} x_\omega\|_2. \tag{19}$$

Lemma 6 shows that each column x_ω is assigned to the right subspace by (5). Since the dimension of a subspace is at most $2r$, and the number of observations in x_ω is greater than $2r$, then x can be correctly recovered.

IV. Simulation

We explore the performance of our method on simulated PMU data. We compare it with one low-rank matrix completion method Singular Value Thresholding (SVT) and the high-rank matrix completion method [13] denoted by HRMC. In order to implement HRMC, $S_{i,j}$ is treated as a separate subspace. We use CVX [16] to solve Subroutine 2. The recovery performance is measured by the relative recovery error $\|X - X_{\text{rec}}\|_F/\|X\|_F$, where X denotes the actual data matrix, and X_{rec} denotes the recovered data matrix. The average erasure rate p_{avg} is the percentage of missing entries.

We use power system toolbox (PST) [10] to simulate the PMU data based on the linear model of IEEE 39 New England Power System [19][28]. 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. Each PMU measures the voltage phasor and the current phasors at a rate of thirty samples per second. The sixteen PMUs measure fifty-seven voltage and current phasors.

We consider the scenario that the outputs of two generators have sharp decrease due to sudden decrease of mechanical torques (Fig. 2). One disturbance happens around $t = 0.1$ s, when the output of generator 32 drops. The other disturbance happens around $t = 1.7$ s, when the output of generator 33 drops. Let the matrix X in $\mathbb{C}^{90\times 57}$ contain the PMU data in three seconds. Each column corresponds to the measurements of one PMU channel across time. Each row corresponds to the PMU measurements at the same sampling instant.

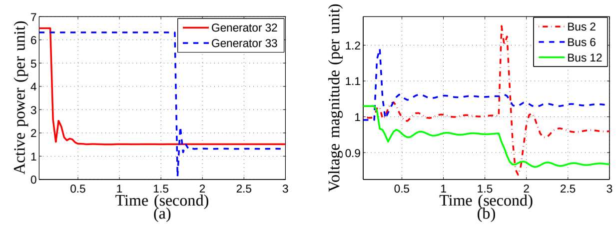

Fig. 2: (a) Active power of generator 32 and 33; (b) Voltage magnitudes at bus 2, 6 and 12.

We randomly delete some simulated PMU measurements. Fig. 3 shows the relative recovery error of SVT, HRMC and our method. The performance of our method is generally better than SVT and HRMC when p_{avg} is below 0.5.

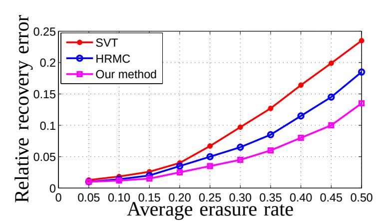

Fig. 3: Relative recovery errors of SVT, HRMC and our method on simulated PMU data.

V. Conclusion

We explore the matrix completion problem when each column belongs to either one of the low-dimensional subspaces or the sum of a subset of the subspaces. We propose a convex-optimization-based method to reconstruct the missing entries of such matrix and provide its theoretical guarantee. Although motivated by power system monitoring, our results can be applied to other scenarios like Internet topology estimation and recommender system. We are currently developing online methods for missing data recovery under this model.

Acknowledgement

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, and in part by NYSERDA Grants #36653 and #28815.

References

[1] "The netflix prize," June 2006. [Online]. Available: http://www.netflixprize.com/

[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] 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.

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

[6] 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.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

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

[14] P. Gao, M. Wang, and J. H. Chow, "Matrix completion with columns in union and sums of subspaces," 2015. [Online]. Available: http://ecse.rpi.edu/~wang/pub/GlobalSIP2015.pdf

[15] P. Gao, M. Wang, S. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, "Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements," *accepted to IEEE Trans. Power Systems*, 2015.

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

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

[18] A. Gruber and Y. Weiss, "Multibody factorization with uncertainty and missing data using the em algorithm," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2004, pp. I–707.

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

[20] Y. P. C. Kotropoulos and G. R. Arce, "l1-graph based music structure analysis," in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2011.

[21] 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.

[22] 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.

[23] 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.

[24] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 612–621, 2014.

[25] 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.

[26] B. McWilliams and G. Montana, "Subspace clustering of high-dimensional data: a predictive approach," *Data Mining and Knowledge Discovery*, vol. 28, no. 3, pp. 736–772, 2014.

[27] 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.

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

[29] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: a review," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 90–105, 2004.

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

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

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

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

[34] 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.

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

[36] R. Vidal, "A tutorial on subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 52–68, 2010.

[37] 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.

[38] A. Zhang, N. Fawaz, S. Ioannidis, and A. Montanari, "Guess who rated this movie: Identifying users through subspace clustering," *arXiv preprint arXiv:1208.1544*, 2012.

Appendix

A. Proof of Lemma 1

Proof. From (6), we have that the expected number of observed entries per column is at least

$$\zeta = 2\eta_1 \mu_0^2 r^2 \log^3(4\beta_0 n)/v_0. \tag{20}$$

Let $\hat{\zeta}$ denote the number of observed entries in a column selected uniformly at random. By Chernoff's bound, we have

$$\mathcal{P}(\hat{\zeta} \le \zeta/2) \le \exp(-\zeta/8). \tag{21}$$

The probability that a randomly selected column belongs to S_{i,ϵ_0} and has $\zeta/2$ or more observed entries is at least v_0', where

$$v_0' := (1-\exp(-\zeta/8))v_0 p_i. \tag{22}$$

Consider a set of s randomly selected columns. The probability that this set does not contain a column from S_{i,ϵ_0} with at least $\zeta/2$ observed entries is less than $(1-v_0')^s$. Then from union bound, the probability that this set does not contain at least one column from S_{i,ϵ_0} with $\zeta/2$ or more observed entries, for any $i = 1, ..., k$, is less than

$$\sum_{i=1}^{k}(1-(1-\exp(-\zeta/8))v_0 p_i)^s. \tag{23}$$

Choose

$$s = \frac{\log k + \log\frac{1}{\delta_0}}{\log\frac{1/((1-\exp(-\zeta/8))v_0 p_*)}{1/((1-\exp(-\zeta/8))v_0 p_*)-1}} \tag{24}$$

such that

$$\delta_0 = k(1-(1-\exp(-\zeta/8))v_0 p_*)^s \quad (25)$$

holds. Lemma 1 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 $x > 1$. □

B. Proof of Lemma 3

Proof. The probability that a column selected uniformly at random from X belongs to both $B_{x,\frac{\epsilon_0}{\sqrt{3}}}$ and S_i is at least $v_0(\epsilon_0/\sqrt{3})^r p_i$. Therefore the expectation of $T^1_{x,\frac{\epsilon_0}{\sqrt{3}}}$ satisfies

$$E(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}}) \geq lnv_0(\epsilon_0/\sqrt{3})^r p_i. \quad (26)$$

The probability that a column selected uniformly at random from X belongs to both B_{x,ϵ_0} and the sum of subspaces is between $v_1\epsilon_0^{2r}q_i$ and $v_2\epsilon_0^{2r}q_i$. Therefore the expected number of points is

$$lnv_1\epsilon_0^{2r}q_i \leq E(T^2_{x,\epsilon_0}) \leq lnv_2\epsilon_0^{2r}q_i. \quad (27)$$

By Chernoff's bound, we have

$$P(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} < E(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}})/2) \leq \exp(-E(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}})/8). \quad (28)$$

Note that from (26) we have

$$P(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} < lnv_0(\epsilon_0/\sqrt{3})^r p_i/2) \leq P(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} < E(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}})/2), \quad (29)$$

$$\text{and } \exp(-E(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}})/8) \leq \exp(-lnv_0(\epsilon_0/\sqrt{3})^r p_i/8). \quad (30)$$

Combining (28), (29) and (30), we have

$$P(T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} < lnv_0(\epsilon_0/\sqrt{3})^r p_i/2) \leq \exp(-lnv_0(\epsilon_0/\sqrt{3})^r p_i/8). \quad (31)$$

Similarly, from Chernoff's bound we have

$$P(T^2_{x,\epsilon_0} > 3E(T^2_{x,\epsilon_0})/2) \leq \exp(-E(T^2_{x,\epsilon_0})/12). \quad (32)$$

From (27) we have

$$P(T^2_{x,\epsilon_0} > 3lnv_2\epsilon_0^{2r}q_i/2) \leq P(T^2_{x,\epsilon_0} > 3E(T^2_{x,\epsilon_0})/2), \quad (33)$$

$$\text{and } \exp(-E(T^2_{x,\epsilon_0})/12) \leq \exp(-lnv_1\epsilon_0^{2r}q_i/12). \quad (34)$$

Combining (32), (33) and (34), We have

$$P(T^2_{x,\epsilon_0} > 3lnv_2\epsilon_0^{2r}q_i/2) \leq \exp(-lnv_1\epsilon_0^{2r}q_i/12). \quad (35)$$

We choose l sufficiently large such that

$$lnv_0(\epsilon_0/\sqrt{3})^r p_i/2 \geq n. \quad (36)$$

From (31) and (35), the probability that (13) does not hold is no greater than $\mathcal{P}_1 + (1-\mathcal{P}_1)\mathcal{P}_2$, where

$$\mathcal{P}_1 := \exp(-\frac{lnv_0(\epsilon_0/\sqrt{3})^r p_i}{8}), \text{ and } \mathcal{P}_2 := \exp(-\frac{lnv_1\epsilon_0^{2r}q_i}{12}). \quad (37)$$

We choose l sufficiently large such that

$$\exp(-lnv_0(\epsilon_0/\sqrt{3})^r p_i/8) \leq \delta_0/(2s), \quad (38)$$

$$\text{and } \exp(-lnv_1\epsilon_0^{2r}q_i/12) \leq \delta_0/(2s) \quad (39)$$

hold. The conditions in (36), (38) and (39) lead to the requirement

$$l \geq \max(\frac{2}{v_0(\frac{\epsilon_0}{\sqrt{3}})^r p_i}, \frac{8\log(2s/\delta_0)}{nv_0(\frac{\epsilon_0}{\sqrt{3}})^r p_i}, \frac{12\log(2s/\delta_0)}{nv_1\epsilon_0^{2r}q_i}). \quad (40)$$

Then combining (31), (35), (36), (38) and (39), we have $T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} \geq n$ and

$$T^2_{x,\epsilon_0}/T^1_{x,\frac{\epsilon_0}{\sqrt{3}}} \leq 3v_2(\sqrt{3}\epsilon_0)^r q_i/(v_0 p_i). \quad (41)$$

hold for a given i with probability at least $1-\delta_0/s$. Finally, by taking the union bound over s subspaces, the claim follows. □

C. Proof of Lemma 4

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

$$\gamma_0 \geq \sum_{j=t_0}^{\zeta_0} \binom{\zeta_0}{j} \underline{p}^j (1-\underline{p})^{\zeta_0 - j}. \quad (42)$$

Let $\tilde{n}$ denote the number of columns with t_0 or more observed indices in common with a fixed seed. By Chernoff's bound, we have

$$\mathcal{P}(\tilde{n} \leq \gamma_0 N/2) \leq \exp(-\gamma_0 N/8). \quad (43)$$

Suppose $N \geq 2l_0\gamma_0^{-1}n$, 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

$$\gamma_0 \geq \underline{p}^{t_0} = \exp(t_0 \log \underline{p}) \geq (2s_0 l_0 n/\delta_0)^{8\mu_1^2 \log \underline{p}}. \quad (44)$$

Then we have

$$N \geq 2l_0 n(2s_0 l_0 n/\delta_0)^{-8\mu_1^2 \log \underline{p}}. \quad (45)$$

By taking the union bound on Lemma 1, the arguments after Lemma 2, and Lemma 3, Lemma 4 follows. □

D. Proof of Lemma 5

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. C_0 has γt non-zero columns. 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 7 (Theorem 1 of [9])**.** *Consider an $n \times t$ matrix M and row and column spaces with coherences bounded above by some constant μ_0. 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*

$$\underline{\rho} \geq \eta_1\mu_0^2\overline{r}^2\log^3(4\beta_0 n)/n \quad (46)$$

$$\text{and } \frac{\overline{\gamma}}{1-\overline{\gamma}} \leq \eta_2 \frac{\underline{\rho}^2}{(1+\frac{\mu_0\overline{r}}{\underline{\rho}\sqrt{n}})^2\mu_0^3\overline{r}^3\log^6(4\beta_0 n)}, \quad (47)$$

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

$$\begin{aligned} \min_{L,C} \quad & \|L\|_* + \lambda \|C\|_{1,2} \\ \textit{subject to} \quad & (L+C)_\Omega = M_\Omega \end{aligned} \tag{48}$$

with $\lambda = \sqrt{\underline{\rho}/(\overline{\gamma}\mu_0 \overline{r} t \log^2(4\beta_0 n))}/48$.

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 [13] 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

$$\mathcal{P}(Z = i) = \binom{t}{i} p_0^i (1-p_0)^{t-i}/(1-\rho). \tag{49}$$

For each column, draw an independent sample of Y. If the sample is 1, keep the column. Otherwise, draw an independent sample of Z, denoted by 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 [13] 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 7 to our problem setup.

Assume (14) and (15) 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 $1 - s_0 c n^{-5}$ for constant $c > 0$. Given a seed 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 (14), then $\beta_0 \geq 1$. We will show (46) and (47) are met for this neighbor matrix, and then Lemma 5 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 Chernoff's bound, we have

$$\mathcal{P}(\hat{m} \leq \beta_0 n^2 \underline{p}/2) \leq \exp(-\beta_0 n^2 \underline{p}/8). \tag{50}$$

From (14), we have

$$\beta_0 n^2 \underline{p}/2 \geq \eta_1 \mu_0^2 r^2 \beta_0 n \log^3(4\beta_0 n)/v_0. \tag{51}$$

Combining (50) and (51) 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

$$\underline{\rho} = \hat{m}/(\beta_0 n^2) \geq \underline{p}/2 \geq \eta_1 \mu_0^2 r^2 \log^3(4\beta_0 n)/(v_0 n). \tag{52}$$

Then (46) is met. Let γ denote the percentage of corrupted columns in the neighbor matrix. From the discussion after Lemma 2, we know

$$\gamma/(1-\gamma) \leq T_{x,\epsilon_0}^2 / T_{x,\frac{\epsilon_0}{\sqrt{3}}}^1. \tag{53}$$

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 (15) increases as $\underline{p}$ increases, and $\underline{\rho} \geq \underline{p}/2$, then (47) is met. Then, Lemma 5 holds with probability at least $1 - s_0\delta_0$, where δ_0 is cn^{-5} for some constant $c > 0$. □

E. Proof of Lemma 6

Proof. Consider a column x in matrix X. Let ω denote the set of indices of the observed entries in x, and let $m = |\omega|$.

Lemma 8 (Theorem 1 of [4]). *Pick any $\delta > 0$ and $m > \frac{16}{3} r \rho_1 \mu_0 \log(\frac{4r}{\delta})$. Then with probability at least $1 - 4\delta$,*

$$\frac{m(1-a) - 2r\rho_1\mu_0 \frac{(1+b)^2}{(1-c)}}{n} \|\tilde{x} - P_{\hat{S}_{i,j}} \tilde{x}\|_2^2 \leq \|\tilde{x}_\omega - P_{\omega, \hat{S}_{i,j}} \tilde{x}_\omega\|_2^2, \tag{54}$$

$$\textit{and } \|\tilde{x}_\omega - P_{\omega, \hat{S}_{i,j}} \tilde{x}_\omega\|_2^2 \leq (1+a)\frac{m}{n} \|\tilde{x} - P_{\hat{S}_{i,j}} \tilde{x}\|_2^2, \tag{55}$$

where $a = \sqrt{2\mu_2^2 \log(1/\delta)/m}$, $b = \sqrt{2\mu_2 \log(1/\delta)}$, and $c = \sqrt{16 r \rho_1 \mu_0 \log(4r/\delta)/(3m)}$.

We here consider the case that the column x belongs to $\hat{S}_{i,j}$. Let $\delta = \delta_0$ in Lemma 8. We then have that $\|x_\omega - P_{\omega, \hat{S}_{i,j}} x_\omega\|_2^2 = 0$ with probability at least $1 - 4\delta_0$. If

$$m > 16 r \mu_0 \log(4r/\delta_0)/3 \tag{56}$$

$$\text{and } 0 < \frac{m(1-a) - 2r\mu_0 \frac{(1+b)^2}{(1-c)}}{n} \|x - P_{\hat{S}_{i',j'}} x\|_2^2 \tag{57}$$

hold for any $(i', j') \neq (i, j)$, where $a = \sqrt{\frac{2\mu_1^2}{m} \log(\frac{1}{\delta})}$, $b = \sqrt{2\mu_1 \log(\frac{1}{\delta})}$ and $c = \sqrt{\frac{16 r \mu_0}{3m} \log(\frac{4r}{\delta})}$, then by union bound, we have

$$\|x_\omega - \mathcal{P}_{\omega, \hat{S}_{i,j}} x_\omega\|_2 < \|x_\omega - \mathcal{P}_{\omega, \hat{S}_{i',j'}} x_\omega\|_2 \tag{58}$$

hold for any $(i', j') \neq (i, j)$ with probability at least $1 - 2k(k-1)\delta_0$. Then in order to let (56) and (57) hold in our setup, we must have

$$m > \max(\frac{16}{3} r \mu_0 \log(\frac{4r}{\delta_0}), \frac{2r\mu_0(1+b)^2}{(1-c)(1-a)}). \tag{59}$$

By Chernoff's bound, we have

$$\mathcal{P}(m \leq np_0/2) \leq \exp(-np_0/8) < \delta_0, \tag{60}$$

and note that

$$np_0/2 \geq \eta_1 \mu_0^2 r^2 \log^3(4\beta_0 n)/v_0. \tag{61}$$

We only need to show the right side of (61) is larger than the right side of (59). For the first term in the max of (59), there is an $O(r \log^2(n))$ gap between $\eta_1 \mu_0^2 r^2 \log^3(4\beta_0 n)/v_0$ and $\frac{16}{3} r \mu_0 \log(\frac{4r}{\delta_0})$. 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 (59), a sufficiently large n can be found such that both $a \leq 1/2$ and $c \leq 1/2$. We then have

$$\frac{(1+b)^2}{(1-c)(1-a)} \leq 4(1+b)^2 \leq 8b^2 = 16\mu_1 \log(\frac{1}{\delta_0}) \tag{62}$$

if b is greater than 2.5. Note that there is an $O(r\log^2(n))$ gap between $\eta_1\mu_0^2 r^2\log^3(4\beta_0 n)/v_0$ and $16\mu_1\log(\frac{1}{\delta_0})$. Therefore, the former is larger when n is sufficiently large. From (60), we have that the inequality (59) holds with probability $1-\delta_0$.

By union bound, we have that if the column $x\in\hat{S}_{i,j}$,

$$0=\|x_\omega-P_{\omega,\hat{S}_{i,j}}x_\omega\|_2<\|x_\omega-P_{\omega,\hat{S}_{i',j'}}x_\omega\|_2, \tag{63}$$

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 x belongs to exact one subspace is similar with the analysis above. If $x\in\hat{S}_i$, we have that

$$0=\|x_\omega-\mathcal{P}_{\omega,\hat{S}_{i,j}}x_\omega\|_2<\|x_\omega-\mathcal{P}_{\omega,\hat{S}_{i',j'}}x_\omega\|_2, \tag{64}$$

holds for any $i'\neq i$ with probability at least $1-(2k(k-1)+1)\delta_0$. Then by union bound, we have lemma 6 hold with probability at least $1-(4k(k-1)+2)\delta_0$. □