Leveraging subspace information for low-rank matrix reconstruction

Wei Zhang [a],*, Taejoon Kim [b], Guojun Xiong [b], Shu-Hung Leung [a,c]

[a] *Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, China*
[b] *Department of Electrical Engineering and Computer Science, The University of Kansas, Lawrence, KS 66045, USA*
[c] *State Key Laboratory of Terahertz and Millimeter Waves, City University of Hong Kong, Hong Kong, China*

ARTICLE INFO

ABSTRACT

The problem of low-rank matrix reconstruction arises in various applications in communications and signal processing. The state of the art research largely focuses on the recovery techniques that utilize affine maps satisfying the restricted isometry property (RIP). However, the affine map design and reconstruction under a priori information, i.e., column or row subspace information, has not been thoroughly investigated. To this end, we present designs of affine maps and reconstruction algorithms that fully exploit the low-rank matrix subspace information. Compared to the randomly generated affine map, the proposed affine map design permits an enhanced reconstruction. In addition, we derive an optimal representation of low-rank matrices, which is exploited to optimize the rank and subspace of the estimate by adapting them to the noise level in order to achieve the minimum mean square error (MSE). Moreover, in the case when the subspace information is not a priori available, we propose a two-step algorithm, where, in the first step, it estimates the column subspace of a low-rank matrix, and in the second step, it exploits the estimated information to complete the reconstruction. The simulation results show that the proposed algorithm achieves robust performance with much lower complexity than existing reconstruction algorithms.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The size of matrices that appear in modern science and engineering is steadily getting large, which has required heavy computing power to timely process them. Fortunately, in many cases, data matrices have the low-rank property. Since the low-rank matrices have much lower degrees of freedom than the full-rank matrices, it is possible to recover them by taking limited observations [1,2]. The limited observation refers to the situation when the number of observations involved is much smaller than the dimension of the matrix.

The low-rank matrix reconstruction is to recover a large-dimensional low-rank matrix $\mathbf{L} \in \mathbb{R}^{M \times N}$ from an affine observation $\mathcal{A}(\mathbf{L})$, where $\mathcal{A}(\cdot) : \mathbb{R}^{M \times N} \mapsto \mathbb{R}^p$ [3], where p is the number of observations. Early work in low-rank matrix reconstruction focuses on non-adaptive approaches [3–9], where the affine map $\mathcal{A}(\cdot)$ is fixed during the reconstruction. To provide a reliability guarantee, the affine map $\mathcal{A}(\cdot)$ is designed to satisfy the restricted isometry property (RIP) [3–7]. Popularly studied non-adaptive methods are nuclear norm minimization (NNM) [3–5], matrix factorization

(MF) [6], and iterative thresholding [10,11]. With convex relaxation, NNM will achieve a reliable reconstruction when the RIP holds [3–5]. However, the main issue is that NNM does not scale with the problem dimension. Alternatively, other non-convex techniques [6,10,11] were investigated to address the complexity issue introduced by NNM. In particular, the MF in [6] is a block-coordinate descent method, where the low-rank matrix is expressed as the product of two low-dimensional matrices. In [10,11], the low-rank solution is guaranteed by taking singular value projection of each iteration result. Though these non-convex techniques have low complexity, they are still limited by an increased number of observations and many iterations to converge.

Recent work has studied adaptive low-rank matrix reconstruction techniques [12,13] that significantly reduce the computational complexity as well as enhance the reconstruction accuracy. Unlike the non-adaptive approaches [3–9] where they exploit the entire observations $\mathcal{A}(\mathbf{L})$ for reconstruction, the methods proposed in [12,13] divide the reconstruction task into two steps, where each step only needs to manipulate partial observations of $\mathcal{A}(\mathbf{L})$ to obtain partial information of $\mathbf{L}$. Considering that the dimension of partial observations is much smaller than the entire observations, the computational complexity can be reduced substantially. An interesting observation is that any partial information of matrix subspace provides a significant clue when adapting the affine map

to enhance the reconstruction accuracy. However, how to adapt the affine map to the available subspace information has not been thoroughly studied in the existing literature (e.g., [12,13]).

In this paper, we analyze and design low-rank matrix reconstruction algorithms that best exploit available subspace information of a low-rank matrix to enhance the reconstruction accuracy in two aspects: (1) we find the optimal representation of the low-rank matrix estimate and (2) we provide the optimal design of the affine map $\mathcal{A}(\cdot)$. By the optimal representation, we mean that the degrees of freedom of the subspace estimate, i.e., the rank of the estimate, is adapted to obtain the minimum mean squared error (MSE). By the optimal design of the affine map $\mathcal{A}(\cdot)$, we mean that we provide an optimality condition of $\mathcal{A}(\cdot)$ so the estimate achieves the minimum MSE. Compared with randomly generated, conventional affine map, the optimized $\mathcal{A}(\cdot)$ is shown to capture specific information of low-rank matrices.

In practice, the subspace information of a low-rank matrix must be estimated. Therefore, we propose a two-step method, where the affine map $\mathcal{A}(\cdot)$ is decomposed into two parts, $\mathcal{A}_1(\cdot)$ and $\mathcal{A}_2(\cdot)$. In the first step, the affine map $\mathcal{A}_1(\cdot)$ is designed to capture the sparse column subspace information of $\mathbf{L}$, and the subspace estimation accuracy is analyzed. For the second step, different from the existing works in [12,13] where the affine map $\mathcal{A}_2(\cdot)$ is fixed, the affine map $\mathcal{A}_2(\cdot)$ in the proposed algorithm is constructed based on the estimated column subspace information from the first step. By doing so, the observations $\mathcal{A}_2(\mathbf{L})$ collect the coefficient matrix that represents the optimal combining weights of the estimated column subspace. An important implication is that the proposed two-step method will not waste the number of observations by not sampling the orthogonal complement of the estimated column subspace. As demonstrated in our simulation studies, this has its own merit when estimating a large-scale, but substantially low-rank, matrix subject to a limited number of observations.

The paper is organized as follows. Section 2 presents the traditional techniques for low-rank matrix reconstruction and the motivation of our work. In Section 3, the low-rank matrix reconstruction under the column subspace information is investigated, where the optimization for the affine map and the minimal matrix representation is discussed. In Section 4, we introduce the proposed two-step low-rank matrix reconstruction technique, and analyze its error bound. The simulation results are illustrated and analyzed in Section 5. Finally, conclusions are drawn in Section 6.

Notations

A bold lower case letter $\mathbf{a}$ is a vector and a bold capital letter $\mathbf{A}$ is a matrix. $\mathbf{A}^T$, $\mathbf{A}^{-1}$, $\mathrm{tr}(\mathbf{A})$, $|\mathbf{A}|$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_*$, and $\|\mathbf{a}\|_2$ are, respectively, the transpose, inverse, trace, determinant, Frobenius norm, nuclear norm (i.e., the sum of the singular values of $\mathbf{A}$) of $\mathbf{A}$, and l_2-norm of $\mathbf{a}$. $[\mathbf{A}]_{:,i}$, $[\mathbf{A}]_{i,:}$, $[\mathbf{A}]_{i,j}$, $[\mathbf{a}]_i$ are, respectively, the ith column, ith row, ith row and jth column entry of $\mathbf{A}$, and ith entry of vector $\mathbf{a}$. $\mathrm{vec}(\mathbf{A})$ stacks the columns of $\mathbf{A}$ and form a long column vector. $\mathrm{diag}(\mathbf{A})$ extracts the diagonal entries of $\mathbf{A}$ to form a column vector. $\mathbf{I}_M \in \mathbb{R}^{M \times M}$ is the identity matrix. $\mathrm{col}(\mathbf{A})$ denotes the subspace spanned by the columns of matrix $\mathbf{A}$.

2. Low-rank matrix reconstruction

In this section, we define and review the low-rank matrix reconstruction problem, and provide a general motivation of our work.

2.1. Classical techniques: non-adaptive

Suppose a matrix observation model where the observations of $\mathbf{L} \in \mathbb{R}^{M \times N}$ are made through the affine map $\mathcal{A}(\cdot) : \mathbb{R}^{M \times N} \to \mathbb{R}^p$, where p is the number of observations and $\mathrm{rank}(\mathbf{L}) = r \ll$

$\min(M, N)$. Here, the ith element of the affine map $\mathcal{A}(\mathbf{L})$ is given by

$$[\mathcal{A}(\mathbf{L})]_i = \mathrm{tr}(\mathbf{X}_i^T \mathbf{L}), \quad i = 1, 2 \ldots p, \tag{1}$$

where $\mathbf{X}_i \in \mathbb{R}^{M \times N}$ is a measurement matrix associated with the ith entry of $\mathcal{A}(\mathbf{L})$. In practice, the observations are corrupted by noise

$$\mathbf{y} = \mathcal{A}(\mathbf{L}) + \mathbf{n}, \tag{2}$$

where $\mathbf{n} \in \mathbb{R}^p$ is the additive white Gaussian noise vector with zero mean and $\mathbf{C} \in \mathbb{R}^{p \times p}$ covariance, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$.

When the number of observations p is not sufficiently large, recovering the matrix $\mathbf{L}$ from $\mathbf{y}$ is not always possible. Nonetheless, when the $\mathbf{L}$ is low-rank, it is possible to reconstruct $\mathbf{L}$ within a certain accuracy [3].

The NNM formulation of the low-rank matrix reconstruction problem [3–5,14] can be given by,

$$\min_{\mathbf{L}} \|\mathbf{L}\|_* + \tau \|\mathbf{y} - \mathcal{A}(\mathbf{L})\|_F^2, \tag{3}$$

where τ is a parameter which balances the values of two parts in the objective function. This problem is an unconstrained convex optimization program, and can be solved by using a proximal gradient method [15,16]. The main drawback is that it does not scale to large-dimensional matrices due to the high complexity of singular value decomposition (SVD) in each iteration [14,17]. Though the computational complexity can be resolved by using accelerated proximal gradient method and in-exact SVD [18,19], the number of required observations for NNM to recover a rank-r matrix $\mathbf{L} \in \mathbb{R}^{M \times N} (M \leq N)$ is still high, i.e., $\mathcal{O}(rN\log(N))$ [17,18]. When the number of available observations is less than this requirement, the reconstruction accuracy will deteriorate [9].

The development of MF [6,20,21] has provided an efficient technique to handle large-scale low-rank matrices. The MF problem for low-rank matrix reconstruction can be formulated as

$$\left(\mathbf{B}^\star, \mathbf{R}^\star\right) = \operatorname*{argmin}_{\mathbf{B} \in \mathbb{R}^{M \times r}, \mathbf{R} \in \mathbb{R}^{N \times r}} \left\|\mathbf{y} - \mathcal{A}(\mathbf{B}\mathbf{R}^T)\right\|_2^2. \tag{4}$$

Solving (4) results in $\widehat{\mathbf{L}} = \mathbf{B}^\star(\mathbf{R}^\star)^T$. The problem in (4) is non-convex, however, the alternating minimization with power factorization [21] can be used to obtain an effective sub-optimal solution. The complexity of the MF is much lower than that of NNM, and MF outperforms NNM when the matrix is substantially low-rank [6,21,22]. A critical drawback of MF is that it can not guarantee the convergence to the true low-rank matrix, which is largely dependent on the initializations of $\mathbf{B}$ and $\mathbf{R}$ [6]. Moreover, since each iteration of solving MF is a least square, it still requires much computational time when the dimension of $\mathbf{L}$ is very large.

2.2. Previous work: adaptive techniques

In order to reduce the computational complexity of low-rank matrix reconstruction, the works in [12,13] have proposed adaptive matrix sensing techniques. Specifically, in stead of manipulating the whole observations $\mathbf{y}$ in (2), they process the partial observations in two stages,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathcal{A}_1(\mathbf{L}) \\ \mathcal{A}_2(\mathbf{L}) \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{bmatrix}, \tag{5}$$

where $\mathbf{y}_1 \in \mathbb{R}^{p_1}$, $\mathbf{y}_2 \in \mathbb{R}^{p_2}$, and $p = p_1 + p_2$. In the first stage, $\mathbf{y}_1 = \mathcal{A}_1(\mathbf{L}) + \mathbf{n}_1$ is taken to estimate the column subspace of the $\mathbf{L}$, which we denote $\widehat{\mathbf{F}} \in \mathbb{R}^{M \times r}$. The second part $\mathbf{y}_2 = \mathcal{A}_2(\mathbf{L}) + \mathbf{n}_2$ is to estimate the coefficient matrix with respect to the estimated $\widehat{\mathbf{F}}$ in order to reconstruct $\mathbf{L}$. Each technique in [12,13] proposes a method of generating $\mathcal{A}_2(\cdot)$ in (5). In particular, in [12], the elements in $\mathcal{A}_2(\cdot)$ are drawn from i.i.d. Gaussian distributions. While

in [13], $\mathcal{A}_2(\cdot)$ consists of elements vectors to select m_2 ($> r$) rows of **L**.

Since $\mathbf{y}_1$ and $\mathbf{y}_2$ are of much smaller number than p, [12,13] provide the ways to utilize the subspace information to assist the reconstruction of **L**, and reduce the computational complexity as well. However, these techniques did not fully consider the design of affine maps ($\mathcal{A}_1(\cdot), \mathcal{A}_2(\cdot)$) and the optimal representation of the estimated low-rank matrix, which can potentially improve the reconstruction accuracy. Overall, how to adapt the affine maps to the available subspace information has not been thoroughly studied in the existing literature.

2.3. Motivation

The RIP is an important characterization that provides the performance guarantees for reliable low-rank matrix reconstruction. Because of this, it has been widely employed [3–6,14]. Under the RIP, many of the state of the art reconstruction methods [3–6,10–13] employ randomly generated affine maps and focus on the reconstruction algorithms. On the other hand, if one can optimize the affine map to lower reconstruction error of a specific algorithm, it is expected that this will reduce the observation overhead. When the subspace information of a low-rank matrix is partially known, it would be possible to adapt $\mathcal{A}(\cdot)$ to the known subspace by minimizing MSE.

3. Affine map design and minimal representation with known column subspace

In this section, we analyze the optimal design of the affine map and provide the optimal representation of the estimate given the column subspace information.

Denote the singular value decomposition (SVD) of **L** as

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \sum_{k=1}^{r} \lambda_k \mathbf{u}_k \mathbf{v}_k^T \tag{6}$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r] \in \mathbb{R}^{M \times r}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r] \in \mathbb{R}^{N \times r}$ are the left and right singular matrix, respectively. The $\mathbf{\Lambda} \in \text{diag}([\lambda_1, \lambda_2, \cdots, \lambda_r])$ is a diagonal matrix where the singular values λ_k are arranged in descending order.

3.1. Design of affine map

Denote $\mathbf{F} \in \mathbb{R}^{M \times r}$ as a semi-unitary matrix, such that col(**F**) is the column subspaces of **L**. It is worth noting that **F** is not necessarily same as **U** in (6). Given the **F**, the low-rank matrix **L** can be expressed as

$$\mathbf{L} = \mathbf{FQ}. \tag{7}$$

where $\mathbf{Q} \in \mathbb{R}^{r \times N}$ is a combining matrix. Given the representation in (7), we find the estimate of **L** in the form, $\widehat{\mathbf{L}} = \mathbf{F}\widehat{\mathbf{Q}}$. Thus, the problem of estimating **L** is equivalent to reconstruct $\widehat{\mathbf{Q}}$.

To begin with, we reshape the affine map $\mathcal{A}$ in (1) as $\mathbf{S} \in \mathbb{R}^{p \times MN}$,

$$\mathbf{S} = \left[\text{vec}(\mathbf{X}_1), \cdots, \text{vec}(\mathbf{X}_p) \right]^T. \tag{8}$$

Then the observation **y** in (2) can be rewritten as

$$\begin{aligned}
\mathbf{y} &= \mathbf{S}\,\text{vec}(\mathbf{L}) + \mathbf{n} \\
&= \mathbf{S}\,\text{vec}(\mathbf{FQ}) + \mathbf{n} \\
&= \mathbf{S}(\mathbf{I}_N \otimes \mathbf{F})\,\text{vec}(\mathbf{Q}) + \mathbf{n}.
\end{aligned} \tag{9}$$

To formulate a design criterion to optimize the affine map **S**, we impose a power constraint $\|\mathbf{S}\|_F^2 \leq P$ to **S**. In (9), we denote

$$\mathbf{A} = \mathbf{S}(\mathbf{I}_N \otimes \mathbf{F}), \tag{10}$$

where $\mathbf{A} \in \mathbb{R}^{p \times Nr}$. Given the observation model in (9), the efficient estimator[1] of **Q** yields [23]

$$\text{vec}(\widehat{\mathbf{Q}}) = (\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}^{-1}\mathbf{y}.$$

The MSE of $\widehat{\mathbf{L}}$ is then written as

$$\begin{aligned}
\mathbb{E}\left(\left\| \widehat{\mathbf{L}} - \mathbf{L} \right\|_F^2 \right) &= \mathbb{E}\left(\left\| \widehat{\mathbf{Q}} - \mathbf{Q} \right\|_F^2 \right) \\
&= \mathbb{E}\left(\left\| (\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}^{-1}\mathbf{n} \right\|_F^2 \right) \\
&= \text{tr}(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}.
\end{aligned} \tag{11}$$

Now the problem is optimizing the affine map **S** by minimizing the MSE in (11) subject to (10) and the power constraint and it can formally be written as

$$\min_{\mathbf{S}} \text{tr}(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}$$

$$\text{subject to} \quad \mathbf{A} = \mathbf{S}(\mathbf{I}_N \otimes \mathbf{F}), \ \|\mathbf{S}\|_F^2 \leq P. \tag{12}$$

The following lemma shows that the problem in (12) is equivalent to a more simple problem.

Lemma 1. *Denote the optimal solution to (12) as $\widehat{\mathbf{S}}$. Suppose the following problem*

$$\min_{\mathbf{A}} \text{tr}(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1} \text{ subject to } \text{tr}(\mathbf{A}^T\mathbf{A}) \leq P \tag{13}$$

has the minimal solution as $\widehat{\mathbf{A}}$. Then, the equality holds,

$$\widehat{\mathbf{S}} = \widehat{\mathbf{A}}(\mathbf{I}_N \otimes \mathbf{F})^T. \tag{14}$$

Proof. This is a direct consequence of the equality $(\mathbf{I}_N \otimes \mathbf{F})^T (\mathbf{I}_N \otimes \mathbf{F}) = \mathbf{I}_{Nr}$. □

Lemma 1 specifies that in order to obtain the optimal **S**, we can turn to solving the problem (13), and the optimal $\widehat{\mathbf{S}}$ can be calculated according to (14). The solution of (13) is summarized in the following lemma.

Lemma 2. *Suppose the power-constrained MSE minimization problem in (13), i.e.,*

$$\min_{\mathbf{A}} \text{tr}(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1} \text{ subject to } \text{tr}(\mathbf{A}^T\mathbf{A}) \leq P. \tag{15}$$

*Denote the SVD of **A** and **C** as $\mathbf{A} = \mathbf{U}_A \mathbf{\Sigma}_A \mathbf{V}_A^T$ with $\mathbf{U}_A \in \mathbb{R}^{p \times Nr}$, $\mathbf{\Sigma}_A \in \mathbb{R}^{Nr \times Nr}$, $\mathbf{V}_A \in \mathbb{R}^{Nr \times Nr}$ and $\mathbf{C} = \mathbf{U}_C \mathbf{\Sigma}_C \mathbf{U}_C^T$ with $\mathbf{U}_C \in \mathbb{R}^{p \times p}$, $\mathbf{\Sigma}_C \in \mathbb{R}^{p \times p}$, respectively.*

Then, the optimal solution to (15) has the following form,

$$\mathbf{U}_A = [\mathbf{U}_C]_{p-Nr+1:p},$$

$$\mathbf{D} = [\mathbf{\Sigma}_C]_{p-Nr+1:p, p-Nr+1:p},$$

$$\mu = \text{tr}\left(\mathbf{D}^{\frac{1}{2}}\right)^2 \Big/ P^2,$$

$$\mathbf{\Sigma}_A^2 = \mu^{-\frac{1}{2}} \mathbf{D}^{\frac{1}{2}}, \mathbf{V}_A^T \mathbf{V}_A = \mathbf{I}_{Nr}. \tag{16}$$

In addition, the minimal MSE is given by

$$\frac{1}{P}\left(\sum_{k=p-Nr+1}^{p} \lambda_{c,k}^{\frac{1}{2}} \right)^2, \tag{17}$$

*where $\lambda_{c,k}$ is the kth largest eigenvalue of **C**. After $\widehat{\mathbf{A}}$ is designed based on (16), the $\widehat{\mathbf{S}}$ is obtained by (14).*

Proof. The problem (15) is a standard differentiable convex optimization problem. This can be optimally solved by applying Karush-Kuhn-Tucker (KKT) conditions [24]. We omit the details because it is a standard procedure. □

---

[1] We call an estimator is efficient, if it is unbiased and attains the Cramer-Rao lower bound [23].

3.2. Optimal representation of an estimator

In the previous section, we assume that the estimate of $\mathbf{L}$ has the form $\widehat{\mathbf{L}} = \mathbf{F}\widehat{\mathbf{Q}}$. In this subsection, we investigate the optimal representation of the low-rank matrix estimate given the noisy affine map in (2), which can achieve the minimum MSE.

The following lemma shows that the estimation error will not increase if we project any estimation result onto the column subspace of the low-rank matrix.

Lemma 3. *Suppose that $\mathbf{L} \in \mathbb{R}^{M \times N}$ is a low-rank matrix and has a decomposition $\mathbf{L} = \mathbf{FQ}$, where $\mathbf{F} \in \mathbb{R}^{M \times r}$ with $\mathbf{F}^T\mathbf{F} = \mathbf{I}_r$. Given the observation $\mathbf{y} = \mathcal{A}(\mathbf{L}) + \mathbf{n}$ with $\mathbb{E}(\mathbf{nn}^T) = \mathbf{C}$, the following inequality holds*

$$\left\| \tilde{\mathbf{L}} - \mathbf{L} \right\|_F^2 \geq \left\| \mathbf{FF}^T\tilde{\mathbf{L}} - \mathbf{L} \right\|_F^2, \tag{18}$$

where $\tilde{\mathbf{L}}$ is the arbitrary estimate of $\mathbf{L}$.

Proof. It is worth noting that $\mathbf{FF}^T$ denotes the projection matrix for the column subspace of $\mathbf{L}$. Let us first define $\mathbf{P_F} = \mathbf{FF}^T$, $\mathbf{P_{F_\perp}} = \mathbf{I}_M - \mathbf{P_F}$, then the estimation error of $\tilde{\mathbf{L}}$ is given by

$$\begin{aligned}
\| \tilde{\mathbf{L}} - \mathbf{L} \|_F^2 &= \| \mathbf{P_F}\tilde{\mathbf{L}} + \mathbf{P_{F_\perp}}\tilde{\mathbf{L}} - \mathbf{L} \|_F^2 \\
&= \| \mathbf{P_F}\tilde{\mathbf{L}} - \mathbf{L} \|_F^2 + \| \mathbf{P_{F_\perp}}\tilde{\mathbf{L}} \|_F^2 \\
&\geq \| \mathbf{P_F}\tilde{\mathbf{L}} - \mathbf{L} \|_F^2.
\end{aligned}$$

This concludes the proof. □

The lemma concludes that the optimal estimation should have the form $\widehat{\mathbf{L}} = \mathbf{P_F}\tilde{\mathbf{L}}$ in order to further minimize the estimation error. In other words, the column subspace of $\widehat{\mathbf{L}}$ should lie on the column subspace of $\mathbf{L}$, $\mathrm{col}(\mathbf{F}_1) \subseteq \mathrm{col}(\mathbf{F})$ where we denote the column subspace of estimate $\widehat{\mathbf{L}}$ as $\mathrm{col}(\mathbf{F}_1)$ and $\mathbf{F}_1 \in \mathbb{R}^{M \times d}$ with $\mathbf{F}_1^T\mathbf{F}_1 = \mathbf{I}_d$, $d \leq r$.

Given $\mathbf{F}_1$, we rewrite the true low-rank matrix in (7) as

$$\mathbf{L} = \mathbf{FQ} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} \tag{19}$$

where $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2]$ is semi unitary $\mathbf{F}^T\mathbf{F} = \mathbf{I}_r$ and $\mathbf{Q}_1 \in \mathbb{R}^{d \times N}$, and $\mathbf{Q}_2 \in \mathbb{R}^{(r-d) \times N}$. With the column subspace of $\widehat{\mathbf{L}}$ being $\mathrm{col}(\mathbf{F}_1)$, we obtains the estimate as

$$\widehat{\mathbf{L}} = \mathbf{F}_1\widehat{\mathbf{Q}}_1, \tag{20}$$

where $\widehat{\mathbf{Q}}_1 \in \mathbb{R}^{d \times N}$. Then, the observation $\mathbf{y}$ in (2) satisfies the following

$$\begin{aligned}
\mathbf{y} &= \mathcal{A}(\mathbf{F}_1\mathbf{Q}_1) + \mathcal{A}(\mathbf{L} - \mathbf{F}_1\mathbf{Q}_1) + \mathbf{n} \\
&= \mathcal{A}(\mathbf{F}_1\mathbf{Q}_1) + \tilde{\mathbf{n}} \\
&= \mathbf{S}(\mathbf{I} \otimes \mathbf{F}_1)\,\mathrm{vec}(\mathbf{Q}_1) + \tilde{\mathbf{n}} \\
&= \mathbf{A}_1\,\mathrm{vec}(\mathbf{Q}_1) + \tilde{\mathbf{n}},
\end{aligned} \tag{21}$$

where $\tilde{\mathbf{n}} = \mathcal{A}(\mathbf{L} - \mathbf{F}_1\mathbf{Q}_1) + \mathbf{n}$, and $\mathbf{A}_1 = \mathbf{S}(\mathbf{I} \otimes \mathbf{F}_1) \in \mathbb{R}^{p \times dN}$. The effective noise covariance can be calculated as

$$\mathbf{C}_1 = \mathbb{E}(\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T) = \mathbf{A}_1\,\mathrm{vec}(\mathbf{Q}_1)\,\mathrm{vec}(\mathbf{Q}_1)^T\mathbf{A}_1^T + \mathbf{C}. \tag{22}$$

Given the observation (21), the efficient estimator of $\mathbf{Q}_1$ yields

$$\mathrm{vec}(\widehat{\mathbf{Q}}_1) = (\mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{A}_1)^{-1}\mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{y}. \tag{23}$$

The MSE of $\widehat{\mathbf{L}}$ is given by

$$\begin{aligned}
\mathbb{E}\left(\left\| \widehat{\mathbf{L}} - \mathbf{L} \right\|_F^2 \right) &= \mathbb{E}\left(\left\| \mathbf{F}_1\widehat{\mathbf{Q}}_1 - \mathbf{L} \right\|_F^2 \right) \\
&= \mathbb{E}\left(\left\| \widehat{\mathbf{Q}}_1 - \mathbf{Q}_1 \right\|_F^2 \right) + \|\mathbf{L}\|_F^2 - \|\mathbf{Q}_1\|_F^2 \\
&= \mathrm{tr}(\mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{A}_1)^{-1} + \|\mathbf{L}\|_F^2 - \|\mathbf{Q}_1\|_F^2.
\end{aligned} \tag{24}$$

In order to minimize the MSE above, we solve the following optimization problem in terms of the affine map,

$$\min_{\mathbf{S},\mathbf{F}_1} \mathrm{tr}(\mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{A}_1)^{-1} + \|\mathbf{L}\|_F^2 - \|\mathbf{Q}_1\|_F^2$$

subject to $\mathbf{A}_1 = \mathbf{S}(\mathbf{I} \otimes \mathbf{F}_1)$, $\|\mathbf{S}\|_F^2 \leq P$. (25)

The following Lemma analyzes optimality conditions for $\mathbf{F}_1$ and $\mathbf{S}$ to achieves the minimal MSE.

Lemma 4. *The optimal $\mathbf{F}_1 \in \mathbb{R}^{M \times d}$ and $\mathbf{S}$ that achieve the minimal MSE in (25) satisfy the following conditions:*

(1) *The optimal rank of the estimate satisfies*

$$d_{opt} = \underset{d}{\arg\min}\left(\frac{1}{P}\left(\sum_{k=p-Nd+1}^{p} \lambda_{c_1,k}^{\frac{1}{2}} \right)^2 + \sum_{k=d+1}^{r} \lambda_k^2 \right), \tag{26}$$

where $\lambda_{c_1,k}$ is the kth largest singular value of $\mathbf{C}_1$, and λ_k is the kth largest singular value of $\mathbf{L}$.

(2) *$\mathbf{F}_1$ spans the subspace of the dominant d_{opt} left singular vectors of $\mathbf{L}$.*

(3) *The optimal $\mathbf{S}$ is $\widehat{\mathbf{S}} = \widehat{\mathbf{A}}_1(\mathbf{I} \otimes \mathbf{F})^T$, where $\widehat{\mathbf{A}}_1 \in \mathbb{R}^{p \times dN}$ is the solution of (15) in Lemma 2 by substituting $\mathbf{C}$ with $\mathbf{C}_1$ in (22).*

Proof. Since there are two optimization variables: $\mathbf{F}_1$ and $\mathbf{S}$, we will solve the whole optimization problem in two steps, where these two steps can be described by using the problem

$$\min_{\mathbf{F}_1}\left(\min_{\mathbf{S}} \mathrm{tr}(\mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{A}_1)^{-1} + \|\mathbf{L}\|_F^2 - \|\mathbf{Q}_1\|_F^2 \right)$$

subject to $\mathbf{A}_1 = \mathbf{S}(\mathbf{I} \otimes \mathbf{F}_1)$, $\|\mathbf{S}\|_F^2 \leq P$. (27)

The first step is to design $\mathbf{S}$ for any $\mathbf{F}_1$. The second step is to optimize $\mathbf{F}_1$ that minimizes the whole objective function.

First of all, for any arbitrary $\mathbf{F}_1$, the value $(\|\mathbf{L}\|_F^2 - \|\mathbf{Q}_1\|_F^2)$ is constant due to $\mathbf{Q}_1 = \mathbf{F}_1^T\mathbf{L}$, simplifying (27) to

$$\min_{\mathbf{S}} \mathrm{tr}(\mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{A}_1)^{-1}$$

subject to $\mathbf{A}_1 = \mathbf{S}(\mathbf{I} \otimes \mathbf{F}_1)$, $\|\mathbf{S}\|_F^2 \leq P$. (28)

The objective value $\mathrm{tr}(\mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{A}_1)^{-1}$ attains the minimum when $\mathbf{S}$ is calculated according to Lemmas 1 and 2, yielding

$$\frac{1}{P}\left(\sum_{k=p-Nd+1}^{p} \lambda_{c_1,k}^{\frac{1}{2}} \right)^2.$$

Now the problem in (27) is over the $\mathbf{F}_1$,

$$\min_{\mathbf{F}_1} \frac{1}{P}\left(\sum_{k=p-Nd+1}^{p} \lambda_{c_1,k}^{\frac{1}{2}} \right)^2 + \|\mathbf{L}\|_F^2 - \|\mathbf{Q}_1\|_F^2. \tag{29}$$

It is worth noting that the first part of (29) only relies on the dimension of $\mathbf{F}_1$, i.e., d. In other words, the value of the first part is a constant for any fixed d. However, for any given d, the value of the second part (i.e., $\|\mathbf{L}\|_F^2 - \|\mathbf{Q}_1\|_F^2$) can be minimized when $\mathbf{F}_1$ spans the subspace of the dominant d left singular vectors of $\mathbf{L}$. Then, we obtain the MSE in terms of d as

$$\mathrm{MSE}(d) = \frac{1}{P}\left(\sum_{k=p-Nd+1}^{p} \lambda_{c_1,k}^{\frac{1}{2}} \right)^2 + \sum_{k=d+1}^{r} \lambda_k^2. \tag{30}$$

To minimize (30), the d should be chosen by

$$d_{opt} = \min_{d} \mathrm{MSE}(d),$$
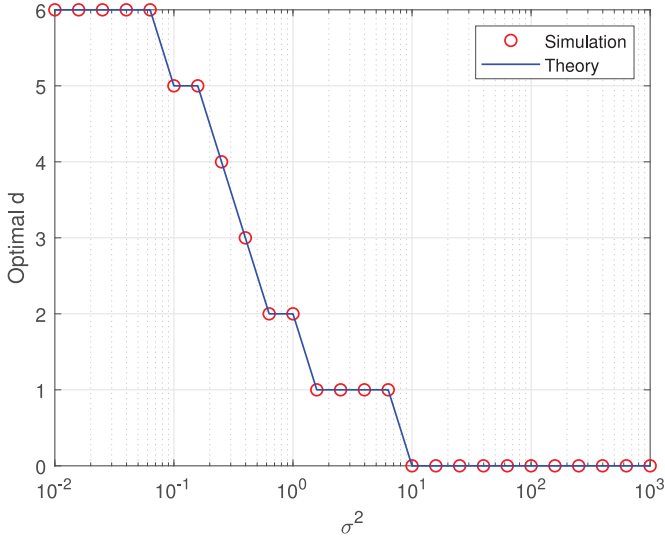
which concludes the proof. □

Fig. 1. Optimal d vs. Noise levels ($M = 20, N = 50, r = 6$).

In order to understand the Lemma 4 in detail, we analyze and illustrate the scenario when the noise is i.i.d. Gaussian such as $\mathbb{E}(\mathbf{nn}^T) = \sigma^2 \mathbf{I}_p$. In this case, the MSE over d is expressed as

$$\frac{d^2 N^2}{P} \sigma^2 + \sum_{k=d+1}^{r} \lambda_k^2, \tag{31}$$

Because λ_k is monotonically decreasing with k, the optimal d that minimizes (31) should satisfy the following

$$d_{opt} = \underset{d=0,1\cdots,r}{\operatorname{argmin}} \left(\frac{d^2 N^2}{P} \sigma^2 + \sum_{k=d+1}^{r} \lambda_k^2 \right). \tag{32}$$

Observing (32), the value of first part is increasing over d, while the value of second part is decreasing over d. This means that when the noise level is high, such as a large σ^2, we will have a small d_{opt} for fixed P and $\{\lambda_k\}_{k=1}^r$. When the noise level is low, the value of d_{opt} will approach r, which is the rank of matrix $\mathbf{L}$.

In Fig. 1, we plot the optimal values of d which achieve the smallest MSEs for different noise levels σ^2. The line of theoretical optimal d is calculated according to (32). It is clear that the theoretical d_{opt} well matches with the simulation results. Specifically, when the noise level is low, the d_{opt} is equal to the rank of matrix, i.e., r. When the noise level is high, d_{opt} approaches to zero, which is consistent with our analysis.

4. Two-step low-rank matrix reconstruction

As we discussed in the previous section, when the column subspace of low-rank matrix is known in advance, the affine map can be attained as we analyzed in Section 3. In practice, however, the column subspace of $\mathbf{L}$ is often not known as a priori, in which we have to first estimate the column subspace and then using this estimated priori information to finish the estimation of the low-rank matrix.

In this section, we propose a two-step method to reconstruct $\mathbf{L}$, which consists of column subspace learning and coefficient matrix learning, where we assume the powers for these two steps are P_1 and P_2 with $P_1 + P_2 = P$.

4.1. Column subspace learning

For a matrix $\mathbf{L} \in \mathbb{R}^{M \times N}$ with rank r, the column subspace of $\mathbf{L}$, i.e., col($\mathbf{L}$), can be expressed by the a semi-unitary matrix $\mathbf{U}_p \in$

$\mathbb{R}^{M \times r}$ such that col($\mathbf{L}$) = col($\mathbf{U}_p$). Here, we denote $\mathbf{U}_p$ as the column subspace matrix of $\mathbf{L}$. It is worth noting that the column subspace matrix of a given $\mathbf{L}$ is not unique. In particular, the left singular vectors of $\mathbf{L}$ is a candidate of column subspace matrix for $\mathbf{L}$.

Denote $\mathbf{L}_1 \in \mathbb{R}^{M \times m}$ as the sub-matrix of $\mathbf{L}$, which is generated by selecting m columns of $\mathbf{L}$. If there are r columns in $\mathbf{L}_1$ are linearly independent, we will have col($\mathbf{L}$) = col($\mathbf{L}_1$). In general, however, we can not guarantee that subspace equality between $\mathbf{L}_1$ and $\mathbf{L}$. Fortunately, according to Rahmani and Atia [13, Lemma 2], if we randomly select m columns of $\mathbf{L}$ as $\mathbf{L}_1$, and $m(\geq r)$ is large enough, the selected columns of matrix $\mathbf{L}_1$ span the column subspace of $\mathbf{L}$ with high probability.

In more detail, we let $\mathbf{Y}_1 \in \mathbb{R}^{M \times m}$ be the observations under noise, i.e.,

$$\mathbf{Y}_1 = \mathbf{L}\mathbf{Z}_1 + \sqrt{mM/P_1}\mathbf{W}_1 = \mathbf{L}_1 + \mathbf{W}_s, \tag{33}$$

where $\mathbf{Z}_1 \in \mathbb{R}^{N \times m}$ is the column sampling matrix, which randomly selects m columns from $\mathbf{L}$, and $\mathbf{W}_1 \in \mathbb{R}^{M \times m}$ is reshaped the noise matrix with $[\mathbf{W}_1]_{i,j} \sim \mathcal{N}(0, \sigma^2)$. Note that the number of observations to obtain $\mathbf{Y}_1$ in the first step is $p_1 = mM$. The remaining question is to obtain the column subspace of $\mathbf{L}_1$ from the observations $\mathbf{Y}_1$.

Considering that $\mathbf{L}_1$ is also a low-rank matrix when $m \gg r$, it is intuitive to estimate $\mathbf{L}_s$ by solving the following nuclear norm and ℓ-norm minimization problem [13],

$$\min_{\mathbf{L}_1} \|\mathbf{L}_1\|_* + \varepsilon\|\mathbf{Y}_1 - \mathbf{L}_1\|_\ell, \tag{34}$$

where ℓ denote different norm operation determined by the type of noise $\mathbf{W}_s$ in (34), and ε is the trade-off parameter.

However, the column subspace learning approach above requires $m \gg r$, which needs mM observations in total. Alternatively, we put a more straightforward but rather robust method. Considering the observations $\mathbf{Y}_1 = \mathbf{L}_1 + \mathbf{W}_s$, when the elements in $\mathbf{W}_s$ are sufficiently small, the column subspace of $\mathbf{L}_1$ can be approximated by the column subspace of $\mathbf{Y}_1$. Suppose the singular value decomposition of $\mathbf{L}_1$ is given by,

$$\mathbf{L}_1 = \mathbf{U}_p \mathbf{\Sigma}_p \mathbf{V}_p^T, \tag{35}$$

where the left singular vector matrix $\mathbf{U}_p \in \mathbb{R}^{M \times r}$, the right singular vector matrix $\mathbf{V}_p \in \mathbb{R}^{N \times r}$, and $\mathbf{\Sigma}_p$ is a diagonal matrix where the singular values are sorted in a descending order.

Because of the noise matrix $\mathbf{W}_s$, the matrix $\mathbf{Y}_1$ has full rank with probability one, and the singular value decomposition of $\mathbf{Y}_1$ can be expressed as,

$$\mathbf{Y}_1 = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^T = \begin{pmatrix} \widehat{\mathbf{U}}_p & \widehat{\mathbf{U}}_o \end{pmatrix} \begin{pmatrix} \widehat{\mathbf{\Sigma}}_p & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{\Sigma}}_o \end{pmatrix} \begin{pmatrix} \widehat{\mathbf{V}}_p^T \\ \widehat{\mathbf{V}}_o^T \end{pmatrix} \tag{36}$$

where $\widehat{\mathbf{U}}_p \in \mathbb{R}^{M \times r}$, $\widehat{\mathbf{V}}_p \in \mathbb{R}^{N \times r}$, $\widehat{\mathbf{U}}_o \in \mathbb{R}^{M \times (m-r)}$, $\widehat{\mathbf{V}}_o \in \mathbb{R}^{N \times (m-r)}$. The diagonal matrices $\widehat{\mathbf{\Sigma}}_p$ and $\widehat{\mathbf{\Sigma}}_o$ include singular values in a descending order.

When the elements in error matrix $\mathbf{W}_s$ are small compared with the $\mathbf{L}_1$, we can treat col($\widehat{\mathbf{U}}_p$) as the estimation for the column subspace of $\mathbf{L}$, i.e., col($\mathbf{U}_p$) in (35). In the following sections, we will discuss the subspace estimation accuracy in more details.

Remark 1. When we have no idea about the rank of the matrix $\mathbf{L}$, we can estimate it through the observations $\mathbf{Y}_1$ in the first step. When the matrix $\mathbf{L}_1$ is zero, the matrix $\mathbf{Y}_1$ only contains noise. Moreover, the singular values of $\mathbf{Y}_1$ are supported on $\left[\sigma\sqrt{M}(1 - \sqrt{m/M}), \sigma\sqrt{M}(1 + \sqrt{m/M})\right]$ as $M \to \infty$ [25]. It inspires us to use the threshold $\tau = \sigma\sqrt{M}(1 + \sqrt{m/M})$ to determine the rank estimation, namely

$$\hat{r} = \max\{i : \hat{\lambda}_i \geq \tau\}, \tag{37}$$

where $\hat{\lambda}_i$ is the ith largest singular value of $\mathbf{Y}_1$. Moreover, when all the singular values of $\mathbf{Y}_1$ is smaller that τ, we just let $\hat{r} = 0$. In the simulation part, we show that the similar performance can be achieved by using the estimation of rank $\hat{r}$.

4.2. Coefficient matrix learning

Given the expression in (35) and the fact that the column subspace of $\mathbf{L}$ is the same as that of $\mathbf{L}_1$, the low-rank matrix $\mathbf{L}$ can be represented as $\mathbf{L} = \mathbf{U}_p\mathbf{Q}$, where $\mathbf{Q} \in \mathbb{R}^{r \times N}$ is the coefficient matrix associated with the column subspace matrix $\mathbf{U}_p$.

Remark 2. For the coefficient matrix $\mathbf{Q}$, we can rewrite $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2]$, where $\mathbf{Q}_1 \in \mathbb{R}^{r \times m}$ and $\mathbf{Q}_2 \in \mathbb{R}^{r \times (N-m)}$ are the coefficient matrices of the m columns of $\mathbf{L}$ in the first step and the remaining $(N-m)$ columns, respectively. Therefore, the matrix $\mathbf{L}$ can be expressed as

$$\mathbf{L} = \mathbf{U}_p[\mathbf{Q}_1, \mathbf{Q}_2]\mathbf{P} = [\mathbf{L}_1, \mathbf{L}_2]\mathbf{P}, \tag{38}$$

where $\mathbf{P} \in \mathbb{R}^{N \times N}$ is the permutation matrix known in advance. Since we have already observe the m columns of $\mathbf{L}$, we can get the estimation of $\mathbf{Q}_1$ from the first step as

$$\widehat{\mathbf{Q}}_1 = \widehat{\mathbf{U}}_p^T\mathbf{Y}_1 = \widehat{\mathbf{U}}_p^T\mathbf{L}_1 + \sqrt{mM/P_1}\widehat{\mathbf{U}}_p^T\mathbf{W}_1. \tag{39}$$

Therefore, we only focus on the estimation for $\mathbf{Q}_2$ in the second step.

By denoting the affine map in the second step is $\mathcal{A}_2(\cdot)$ and p_2 is the number of observations in the second step, the observation $\mathbf{y}_2 \in \mathbb{R}^{p_2}$ in the second step can be expressed as

$$\begin{aligned} \mathbf{y}_2 &= \mathcal{A}_2(\mathbf{U}_p\mathbf{Q}_2) + \mathbf{n}_2 \\ &= \mathbf{S}_2(\mathbf{I} \otimes \mathbf{U}_p)\,\mathrm{vec}(\mathbf{Q}_2) + \mathbf{n}_2, \end{aligned} \tag{40}$$

where $\mathbf{n}_2 \in \mathbb{R}^{p_2}$ with $\mathbb{E}(\mathbf{n}_2\mathbf{n}_2^T) = \mathbf{C}_2$, and $\mathbf{S}_2 \in \mathbb{R}^{p_2 \times MN}$ is generated by reforming $\mathcal{A}_2(\cdot)$. In the following part, we will discuss how to obtain the estimation for $\mathbf{Q}_2$ and the design of the affine map $\mathcal{A}_2(\cdot)$.

Given the estimated $\widehat{\mathbf{U}}_p$ from the first step, we assume the expression of $\widehat{\mathbf{L}}_2$ as $\widehat{\mathbf{L}}_2 = \widehat{\mathbf{U}}_p\widehat{\mathbf{Q}}_2$, where $\widehat{\mathbf{Q}}_2 \in \mathbb{R}^{r \times (N-m)}$. Based on the analysis in Section 3, under the observations $\mathbf{y}_2$ in (40), the efficient estimation of $\widehat{\mathbf{Q}}_2$ is given by

$$\mathrm{vec}(\widehat{\mathbf{Q}}_2) = (\widehat{\mathbf{A}}_2^T\mathbf{C}_2^{-1}\widehat{\mathbf{A}}_2)^{-1}\widehat{\mathbf{A}}_2^T\mathbf{C}_2^{-1}\mathbf{y}_2, \tag{41}$$

where $\widehat{\mathbf{A}}_2 = \mathbf{S}_2(\mathbf{I} \otimes \widehat{\mathbf{U}}_p) \in \mathbf{R}^{p_2 \times r(N-m)}$. Here, in the second step, we assume the difference between $\mathrm{col}(\widehat{\mathbf{U}}_p)$ and $\mathrm{col}(\mathbf{U}_p)$ is small. Ideally, when $\mathrm{col}(\widehat{\mathbf{U}}_p) = \mathrm{col}(\mathbf{U}_p)$, the affine map $\mathbf{S}_2$ with the minimal MSE can be obtained through the optimality condition in (16). Specifically, when $\mathbb{E}(\mathbf{n}\mathbf{n}^T) = \sigma^2\mathbf{I}_{p_2}$, the optimal $\mathbf{S}_2$ associated with affine map $\mathcal{A}_2(\cdot)$ is given by

$$\mathbf{S}_2 = \sqrt{\frac{P_2}{r(N-m)}}\mathbf{I}_{r(N-m)}\big(\mathbf{I} \otimes \widehat{\mathbf{U}}_p\big)^T. \tag{42}$$

Though the design of the affine map above utilize the approximation $\mathrm{col}(\widehat{\mathbf{U}}_p) \approx \mathrm{col}(\mathbf{U}_p)$, in the simulation part, we verify that this approximation is still helpful to improve the reconstruction accuracy. Given the affine map in the design in (42), for the coefficient matrix $\widehat{\mathbf{Q}}_2$, we will have the following estimation by substituting (42) into (41),

$$\widehat{\mathbf{Q}}_2 = \widehat{\mathbf{U}}_p^T\mathbf{L}_2 + \sqrt{r(N-m)/P_2}\mathbf{W}_2 = \widehat{\mathbf{U}}_p^T\mathbf{L}_2 + \mathbf{W}_r, \tag{43}$$

where $\mathbf{W}_2 \in \mathbb{R}^{r \times (N-m)}$ is the noise matrix which is reshaped from $\mathbf{n}_2$ in (40). The details of the proposed two-step method is stated in Algorithm 1. The estimation result is given by $\widehat{\mathbf{L}} = \widehat{\mathbf{U}}_p[\widehat{\mathbf{Q}}_1, \widehat{\mathbf{Q}}_2]\mathbf{P}$, where $\mathbf{P}$ is the permutation matrix defined in (38).

Algorithm 1 Two-step method for reconstruction of low-rank matrices.

1: Input: The data matrix $\mathbf{L} \in \mathbb{R}^{M \times N}$ and the power P.
2: Initialization: The column selecting matrix $\mathbf{Z} \in \mathbb{R}^{N \times m}$, powers for two steps are P_1 and P_2 with $P = P_1 + P_2$, and the permutation matrix $\mathbf{P}$.
3: Column subspace learning:

- Sample the columns of $\mathbf{L}$ by $\mathbf{Y}_1 = \mathbf{LZ} + \sqrt{mM/P_1}\mathbf{W}_1 = \mathbf{L}_1 + \mathbf{W}_s$ with noise matrix $\mathbf{W}_1$.
- Calculate the column subspace basis matrix $\widehat{\mathbf{U}}_p$ through (36) and $\widehat{\mathbf{Q}}_1$ in (39).

4: Calculate the coefficient matrix:

- Let the reshaped affine map of $\mathcal{A}_2(\cdot)$ be $\mathbf{S}_2 = \sqrt{\frac{P_2}{r(N-m)}}\mathbf{I}_{r(N-m)}\big(\mathbf{I} \otimes \widehat{\mathbf{U}}_p\big)^T$ obtain the observations through $\mathbf{y}_2 = \mathcal{A}_2(\mathbf{L}_2) + \mathbf{n}_2$.
- Calculate the estimation of coefficient matrix $\widehat{\mathbf{Q}}_2$ through (43).

5: Obtain the estimate of the LR matrix: $\widehat{\mathbf{L}} = \widehat{\mathbf{U}}_p[\widehat{\mathbf{Q}}_1, \widehat{\mathbf{Q}}_2]\mathbf{P}$.
6: Output: Estimate result $\widehat{\mathbf{L}}$.

4.3. Estimation error analysis

Before evaluating the estimation error, the following Lemma provides the accuracy of the estimated column subspace $\mathrm{col}(\widehat{\mathbf{U}}_p)$ compared to $\mathrm{col}(\mathbf{U}_p)$. Denote $\mathbf{U}_{p\perp} \in \mathbb{R}^{M \times (N-r)}$ as the matrix which satisfies $\mathbf{U}_{p\perp}\mathbf{U}_{p\perp}^T = \mathbf{I}_M - \mathbf{U}_p\mathbf{U}_p^T$. We introduce the

$$\eta = \big\|\widehat{\mathbf{U}}_{p\perp}^T\mathbf{U}_p\big\|_2 = \big\|\widehat{\mathbf{U}}_p^T\mathbf{U}_{p\perp}\big\|_2 \tag{44}$$

as the subspace distance defined in [26]. The value of η is ranged from 0 to 1. When $\mathrm{col}(\widehat{\mathbf{U}}_p) = \mathrm{col}(\mathbf{U}_p)$, we can easily verify $\eta = 0$.

Lemma 5. *Assume the singular value gap $\delta = \lambda_r - \widehat{\lambda}_{r+1}$, where λ_r is the rth largest singular value of $\mathbf{L}_1 \in \mathbb{R}^{M \times m}$, and $\widehat{\lambda}_{r+1}$ is the $(r+1)$th largest singular value of $\mathbf{Y}_1 \in \mathbb{R}^{M \times m}$ with $\mathbf{Y}_1 = \mathbf{L}_1 + \mathbf{W}_s$. According to the Wedin's Theorem [27,28], the following bound holds*

$$\begin{aligned} \big\|\widehat{\mathbf{U}}_p^T\mathbf{U}_{p\perp}\big\|_2 &\leq \frac{\max\{\|\mathbf{W}_s\mathbf{V}_p\|_2, \|\mathbf{W}_s^T\mathbf{U}_p\|_2\}}{\delta} \\ &\leq \frac{\|\mathbf{W}_s\|_2}{\delta} \\ &= \sqrt{\frac{mM}{P_1}}\,\frac{\|\mathbf{W}_1\|_2}{\delta}, \end{aligned} \tag{45}$$

where $\mathbf{Y}_1 = \widehat{\mathbf{U}}_p\widehat{\mathbf{\Sigma}}_p\widehat{\mathbf{V}}_p^T$ and $\mathbf{L}_1 = \mathbf{U}_p\mathbf{\Sigma}_p\mathbf{V}_p^T$ are the SVD of $\mathbf{Y}_1$ and $\mathbf{L}_1$, respectively.

Given the subspace estimation accuracy in (45), it is of interest of evaluate estimation error between $\mathbf{L}$ and $\widehat{\mathbf{L}}$, which is given in the following lemma.

Lemma 6. *Suppose the low-rank matrix $\mathbf{L} \in \mathbb{R}^{M \times N}$, and the estimation result $\widehat{\mathbf{L}}$ is obtained through Algorithm 1. The estimation error is given by*

$$\begin{aligned} \big\|\mathbf{L} - \widehat{\mathbf{L}}\big\|_F &\leq \sqrt{\frac{mM}{P_1}}\bigg(\|\mathbf{W}_1\|_F + \frac{\|\mathbf{W}_1\|_2\|\mathbf{L}\|_F}{\delta}\bigg) \\ &\quad + \sqrt{\frac{r(N-m)}{P_2}}\|\mathbf{W}_2\|_F, \end{aligned} \tag{46}$$

where δ is defined in same way as Lemma 5.

Proof. Given the expression of $\widehat{\mathbf{Q}}$, we formulate the estimation error in the following form,

$$
\begin{aligned}
\| \mathbf{L} - \widehat{\mathbf{U}}_p\widehat{\mathbf{Q}}\|_F &\leq \| \widehat{\mathbf{U}}_p\widehat{\mathbf{U}}_p^T\mathbf{W}_s + \mathbf{W}_r\|_F + \| \mathbf{L} - \widehat{\mathbf{U}}_p\widehat{\mathbf{U}}_p^T\mathbf{L}\|_F \\
&\leq \| \widehat{\mathbf{U}}_p\widehat{\mathbf{U}}_p^T\mathbf{W}_s\|_F + \| \mathbf{W}_r\|_F + \| \mathbf{L} - \widehat{\mathbf{U}}_p\widehat{\mathbf{U}}_p^T\mathbf{L}\|_F \\
&\leq \| \mathbf{W}_s\|_F + \| \mathbf{W}_r\|_F + \| \left(\mathbf{I} - \widehat{\mathbf{U}}_p\widehat{\mathbf{U}}_p^T\right)\mathbf{U}_p\mathbf{U}_p^T\|_2 \| \mathbf{L}\|_F \\
&= \| \mathbf{W}_s\|_F + \| \mathbf{W}_r\|_F + \| \left(\widehat{\mathbf{U}}_{p\perp}\widehat{\mathbf{U}}_{p\perp}^T\right)\mathbf{U}_p\mathbf{U}_p^T\|_2 \| \mathbf{L}\|_F \\
&\leq \| \mathbf{W}_s\|_F + \| \mathbf{W}_r\|_F + \| \widehat{\mathbf{U}}_{p\perp}^T\mathbf{U}_p\|_2 \| \mathbf{L}\|_F, \qquad (47)
\end{aligned}
$$

where $\widehat{\mathbf{U}}_{p\perp}\widehat{\mathbf{U}}_{p\perp}^T$ denotes the projection matrix onto the complementary subspace spanned by $\widehat{\mathbf{U}}_p$, and $\mathbf{U}_p\mathbf{U}_p^T$ is the projection matrix onto the subspace spanned by $\mathbf{U}_p$.

Therefore, combining (47) and (45), we can have the error in (46). This concludes the proof. □

Note that the bound in (46) will hold for any possible noise. Compared to the MSE when the column subspace is exact, the MSE in (46) includes one more part, such that $\sqrt{\frac{mM}{P_1}}\frac{\|\mathbf{W}_1\|_2\|\mathbf{L}\|_F}{\delta}$, which comes from the error of estimate of column subspace. Using the proposed two-step method, the necessary number of samples is

$$
p = mM + r(N - m). \qquad (48)
$$

In particular, if we set the $m = r$, the number of observations is just equal to the degrees of freedom of esimated low-rank matrix.

Remark 3. The complexity of the two-step method mainly comes from SVD computation $\mathbf{Y}_1$, i.e., $\mathcal{O}(Mm^2)$. Thus, when m is just several times of the rank r, the computational complexity of the proposed two-step method is in a small order.

5. Simulation results

In this section, we simulate the performance of the proposed two-step matrix reconstruction method compared to two non-adaptive methods, i.e., MF [6] and NNM [3], and one adaptive method, i.e., subspace pursuit (SP) approach [13]. The simulation parameters are $M = 20, N = 50, r = 6$. The normalized MSE (NMSE) is defined as

$$
\mathrm{NMSE} = \mathbb{E}\left(\left\|\widehat{\mathbf{L}} - \mathbf{L}\right\|_F^2 / \|\mathbf{L}\|_F^2\right).
$$

Here, we assume the power associate with the two steps are $P_1 = P_2 = MN$. For the simulated methods, each point of the curve is plotted by averaging the NMSE of 1000 trails.

5.1. The affect of number of observations

First of all, we illustrate the effect of number of observations on the reconstruction accuracy in Fig. 2. We let the number of column m for the first step is ranged in $\{r, 1.5r, 2r, 2.5r\} = \{6, 9, 12, 15\}$. Accordingly, the total number of observations are $mM + r(N - m)$. In general, when more columns are utilized for the first step, more accurate columns subspace can be obtained. As we can see, when the number of columns for the first step is $1.5r$, the robust performance can be achieved. This means that when we let $m = 1.5r$ in the first step, we can acquire a robust estimation for the subspace information.

5.2. The characteristic of designed affine map

In this subsection, we compare the affine map of the proposed two-step method with the randomly generated affine map. As far as we know, it is NP-hard to compute the RIP constant for the given affine map. Therefore, we turn to compute the alternative constant to RIP, namely, averaged mutual coherence defined in
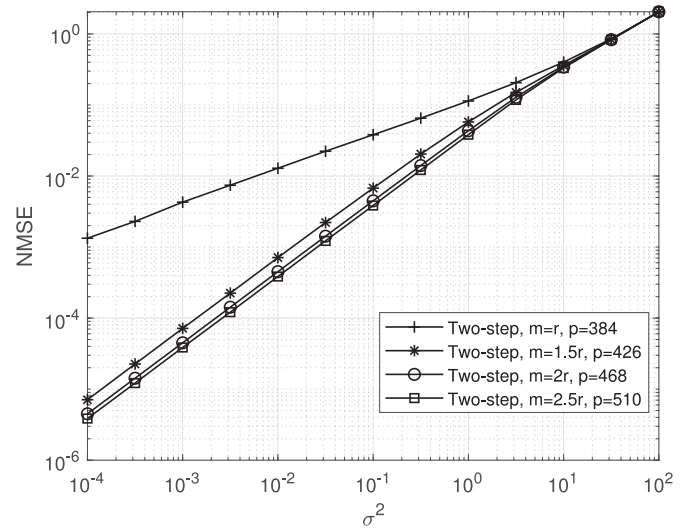

Fig. 2. NMSE vs. Noise levels with differen number of observations($M = 20, N = 50, r = 6$).
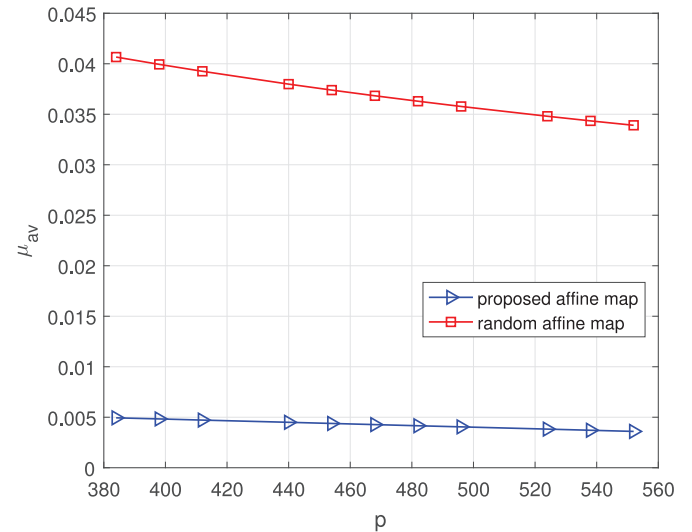

Fig. 3. Averaged mutual coherence vs. Number of observations ($M = 20, N = 50$).

[29], which evaluates the averaged coherence of the columns of affine map matrix $\mathbf{S} \in \mathbb{R}^{p\times MN}$. Specifically, letting $\bar{\mathbf{S}} \in \mathbb{R}^{p\times MN}$ be the matrix which normalizes the columns of $\mathbf{S}$, the averaged mutual coherence of $\mathbf{S}$ is given by

$$
\mu_{av}(\mathbf{S}) = \frac{\sum\limits_{1\leq i, j\leq MN, i\neq j} \mathrm{tr}([\bar{\mathbf{S}}]_{:,i}^T[\bar{\mathbf{S}}]_{:,j})}{M^2N^2 - p}.
$$

It has been shown that the reconstruction accuracy is related to the value of μ_{av} [29,30]. In Fig. 3, we evaluate the averaged mutual coherence of proposed affine map compared to that of gaussian random affine map, i.e., $[\mathbf{S}]_{i,j} \sim \mathcal{N}(0, 1), \forall i, j$. As can be seen from Fig. 3, the proposed affine map design achieves the lower averaged mutual coherence compared to the randomly generated affine map. Therefore, it could be expected that reconstruction accuracy achieved by the proposed affine map will be higher than that of randomly generated affine map.

5.3. The accuracy comparisons with benchmarks

In Fig. 4, we compare the performance of proposed two-step method with NNM, MF and SP approaches. We set the number
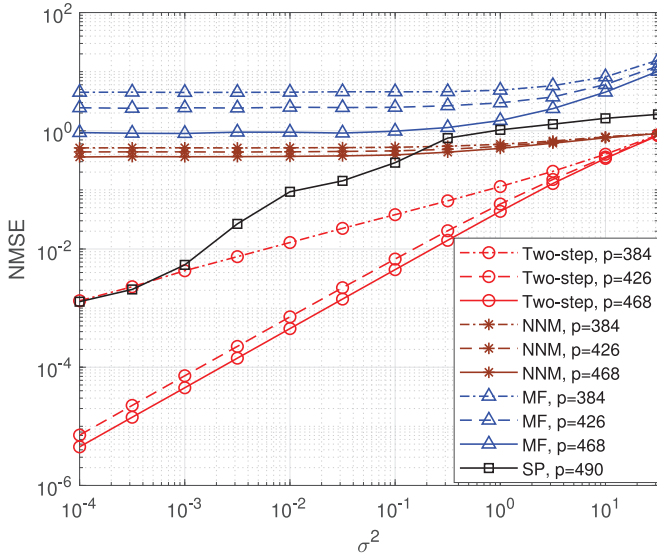
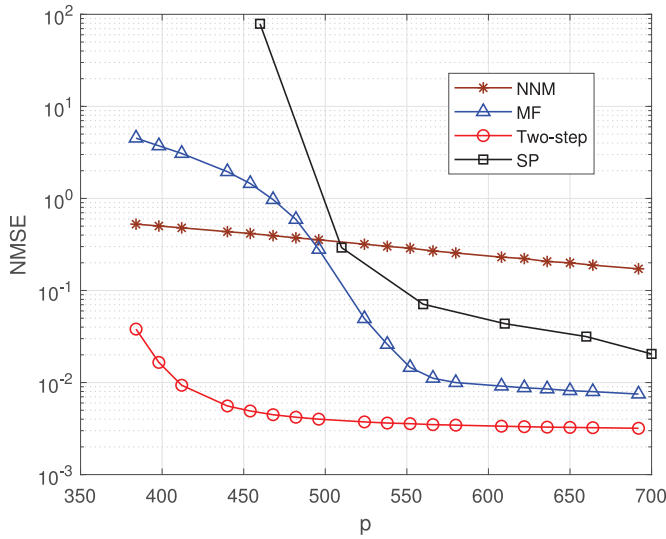**Fig. 4.** NMSE vs. Noise levels ($M = 20, N = 50, r = 6$).

Fig. 6. NMSE vs. Noise levels ($M = 20, N = 50, r = 6$).

Fig. 5. NMSE vs. Number of observations ($M = 20, N = 50, r = 6, \sigma^2 = 0.1$).

ber of observations. Moreover, the proposed two-step method outperforms the others. When the number of observations is $p = 384$, which is equal to the degrees of freedom of the simulated low-rank matrix, i.e., $(50 + 20 - 6) \times 6 = 384$, the proposed two-step method can achieve a robust estimation. This validates the fact the proposed two-step method only requires the number of observations that is approximately equal to the degrees of freedom of low-rank matrix.

5.4. The reconstruction accuracy by using estimated rank information

In this simulation part, we simulate the reconstruction performance when we utilize the estimated rank information instead of true rank information. As we can see in Fig. 6, by using the estimated rank $\hat{r}$ in (37), the similar performance can be achieved compared to the scenario when we utilize the true rank information. Therefore, the rank estimation method in (37) is robust for the low-rank matrix reconstruction under the noisy cases.

6. Conclusion

In this paper, we investigate the low-rank matrix reconstruction when the subspace information is known. Under the subspace information, the optimal representation of low-rank matrix is analyzed in order to obtain minimal MSE. In the case that no subspace information is aware, the proposed two-step method can handle this practical scenario. The first step will obtain the column subspace of the low-rank matrix, and the second step will get the remaining coefficient information of the low-rank matrix. By using the observations approximately equal to degrees of freedom of the low-rank matrix, the simulation results show that the proposed two-step method experiences robust performance compared to the existing reconstruction methods.

Conflict of interest

None.

of observations for proposed two-step method, NNM and MF selected from {384, 426, 468}. Since the SP method requires more observations in order to achieve valid reconstruction, we let the number of observations for SP be $p = 490$. As can been seen that NNM, MF and proposed two-step method can benefit from the increasing of number of observations. It is clear that the proposed two-step method outperforms the NNM, MF and SP. Moreover, the performance gap is larger when the noise level is low. In particular, when the number of observations is $p = 426$, the proposed two-step method has a clear performance gap compared to the benchmarks. This is because the performance of NNM, MF and SP will be restricted by the number of observations. If the number of observations is not sufficient, i.e., $p = 426$, the accuracy of NNM, MF and SP will be saturated when the noise level is low. Different from the saturated phenomenon of benchmarks, for proposed two-step method, more accurate reconstruction can be obtained if the noise level is decreasing furthermore.

In Fig. 5, we evaluate the performance of proposed two-step method under different observations compared to the MF, NNM, and SP. The noise level is set as $\sigma^2 = 0.1$. As we can see, all the simulated methods will benefits from the increasing of the num-
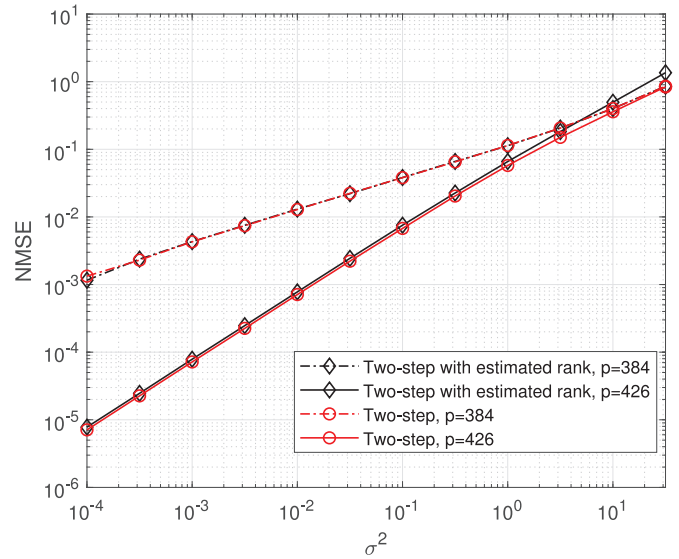
References

[1] M.E. Davies, Y.C. Eldar, Rank awareness in joint sparse recovery, IEEE Trans. Inf. Theory 58 (2) (2012) 1135–1146, doi:10.1109/TIT.2011.2173722.

[2] Z. Liu, L. Vandenberghe, Interior-point method for nuclear norm approximation with application to system identification, SIAM J. Matrix Anal. Appl. 31 (3) (2009) 1235–1256.

[3] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (3) (2010) 471–501.

[4] Y. Chen, Y. Chi, A.J. Goldsmith, Exact and stable covariance estimation from quadratic sampling via convex programming, IEEE Trans. Inf. Theory 61 (7) (2015) 4034–4059.

[5] T.T. Cai, A. Zhang, Rop: matrix recovery via rank-one projections, Ann. Stat. 43 (1) (2015) 102–138.

[6] P. Jain, P. Netrapalli, S. Sanghavi, Low-rank matrix completion using alternating minimization, in: Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing, ACM, 2013, pp. 665–674.

[7] E.J. Candes, Y. Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, IEEE Trans. Inf. Theory 57 (4) (2011) 2342–2359.

[8] D. Achlioptas, Database-friendly random projections: Johnson–Lindenstrauss with binary coins, J.Comput. Syst. Sci. 66 (4) (2003) 671–687.

[9] W. Zhang, T. Kim, D.J. Love, E. Perrins, Leveraging the restricted isometry property: improved low-rank subspace decomposition for hybrid millimeter-wave systems, IEEE Trans. Commun. 66 (11) (2018) 5814–5827, doi:10.1109/TCOMM.2018.2854779.

[10] P. Jain, R. Meka, I.S. Dhillon, Guaranteed rank minimization via singular value projection, in: Advances in Neural Information Processing Systems, 2010, pp. 937–945.

[11] J. Tanner, K. Wei, Normalized iterative hard thresholding for matrix completion, SIAM J. Sci. Comput. 35 (5) (2013) S104–S125.

[12] S. Biswas, S. Dasgupta, R. Mudumbai, M. Jacob, Subspace aware recovery of low rank and jointly sparse signals, IEEE Trans. Comput. Imaging 3 (1) (2017) 22–35, doi:10.1109/TCI.2016.2628352.

[13] M. Rahmani, G.K. Atia, High dimensional low rank plus sparse matrix decomposition, IEEE Trans. Signal Process. 65 (8) (2017) 2004–2019, doi:10.1109/TSP.2017.2649482.

[14] M.A. Davenport, J. Romberg, An overview of low-rank matrix recovery from incomplete observations, IEEE J. Sel. Top. Signal Process. 10 (4) (2016) 608–622, doi:10.1109/JSTSP.2016.2539100.

[15] N. Parikh, S. Boyd, et al., Proximal algorithms, Found. Trends Optim. 1 (3) (2014) 127–239.

[16] P.L. Combettes, J.-C. Pesquet, Proximal splitting methods in signal processing, in: Fixed-point Algorithms for Inverse Problems in Science and Engineering, Springer, 2011, pp. 185–212.

[17] J.-F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. 20 (4) (2010) 1956–1982.

[18] Q. Yao, J.T. Kwok, Accelerated and inexact soft-impute for large-scale matrix and tensor completion, IEEE Trans. Knowl. Data Eng. (2018) 1, doi:10.1109/TKDE.2018.2867533.

[19] Q. Yao, J.T. Kwok, T. Wang, T. Liu, Large-scale low-rank matrix learning with nonconvex regularizers, IEEE Trans. Pattern Anal. Mach.Intell. (2018) 1, doi:10.1109/TPAMI.2018.2858249.

[20] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (8) (2009).

[21] J. Haldar, D. Hernando, Rank-constrained solutions to linear matrix equations using powerfactorization, IEEE Signal Process. Lett. 16 (7) (2009) 584–587.

[22] W. Zhang, T. Kim, D.J. Love, Sparse subspace decomposition for millimeter wave mimo channel estimation, in: 2016 IEEE Global Communications Conference (GLOBECOM), 2016, pp. 1–6.

[23] S.M. Kay, Fundamentals of Statistical Signal Processing: Practical Algorithm Development, 3, Pearson Education, 2013.

[24] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[25] G.W. Anderson, A. Guionnet, O. Zeitouni, An Introduction to Random Matrices, 118, 2010.

[26] G.H. Golub, C.F. Van Loan, Matrix Computations, 3, JHU Press, 2012.

[27] R.-C. Li, Relative perturbation theory: Ii. Eigenspace and singular subspace variations, SIAM J. Matrix Anal. Appl. 20 (2) (1998) 471–492.

[28] M. Stewart, Perturbation of the svd in the presence of small singular values, Linear Algebra Appl. 419 (1) (2006) 53–77.

[29] M. Elad, Optimized projections for compressed sensing, IEEE Trans. Signal Process. 55 (12) (2007) 5695–5702, doi:10.1109/TSP.2007.900760.

[30] G. Li, Z. Zhu, D. Yang, L. Chang, H. Bai, On projection matrix optimization for compressive sensing systems, IEEE Trans. Signal Process. 61 (11) (2013) 2887–2898, doi:10.1109/TSP.2013.2253776.