Guaranteed Matrix Completion under Multiple Linear Transformations

Chao Li[†], Wei He[†], Longhao Yuan[‡†], Zhun Sun[†], Qibin Zhao[†§*]
[†]RIKEN Center for Advanced Intelligence Project, Japan
[‡]Saitama Institute of Technology, Japan
[§]School of Automation, Guangdong University of Technology, China
{chao.li, wei.he, longhao.yuan, zhun.sun, qibin.zhao}@riken.jp

Abstract

Low-rank matrix completion (LRMC) is a classical model in both computer vision (CV) and machine learning, and has been successfully applied to various real applications. In the recent CV tasks, the completion is usually employed on the variants of data, such as "non-local" or filtered, rather than their original forms. This fact makes that the theoretical analysis of the conventional LRMC is no longer suitable in these applications. To tackle this problem, we propose a more general framework for LRMC, in which the linear transformations of the data are taken into account. We rigorously prove the identifiability of the proposed model and show an upper bound of the reconstruction error. Furthermore, we derive an efficient completion algorithm by using augmented Lagrangian multipliers and the sketching trick. In the experiments, we apply the proposed method to the classical image inpainting problem and achieve the state-of-the-art results.

1. Introduction

Low-rank matrix completion (LRMC) is a classical model to utlize the low-rank structure of the matrix to recover the missing entries given a small number of observations [5], and has many applications in different fields [17]. Especially in computer vision (CV), LRMC has been widely applied to image/video inpainting [45], reflection removal [14] and occlusion removal [40] and so on. Although the black-box methods like generative adversarial nets (GANs) become popular recently and also achieve good performance in these tasks [30, 42], one of the advantage of LRMC is that the completion precision is theoretically guaranteed, which is of importance in practical applications.

*Q. Zhao is the corresponding author. This work is partially supported by JSPS KAKENHI (Grant No. 17K00326) and NSFC China (Grant No. 61773129).

One well known LRMC method is to perform nuclear norm minimization (NNM) on the incomplete matrix [32, 41]. Due to the fact that nuclear norm is the convex envelope of the matrix rank [32], solving NNM is equivalent to searching the optimal low-rank approximation of the observations. Furthermore, it has been proved that, under mild conditions, the reconstruction error of NNM is bounded by $\mathcal{O}\left(\sqrt{\frac{n(p+2)}{p}}\delta + 2\delta\right)$ where p denotes a constant *w.r.t.* the rank of the $n \times n$ matrix and δ denotes the strength of the noise. Although there exist many studies on more efficient algorithms for NNM [15, 34], NNM cannot give satisfactory performance when the matrix is of high rank or several whole rows or columns are missing, and such situations often happen in practice.

On the other hand, recent CV-driven studies show that we can obtain better performance than NNM if exploiting the low-rank structure on a "transformed" variant of the matrix. For example, in low-level CV tasks, the "non-local" methods are popularly used for image inpainting and denoising [6, 8], in which the observations are split into "non-local" groups, and the low-rank approximation is then applied to each group. Likewise, in the problem of image occlusion removal, the low-rank structure of the obscured part of the images is used to regularize the regression model [40], implying that the low-rank approximation is employed on the down-sampled data rather than the original ones in the applications. In addition, the low-rank structures of the filtered or transformed data are also studied for different problems [10, 26, 33, 43]. Although these methods have achieved the state-of-the-art performance in their fields, the existing theoretical results for the conventional LRMC are no longer suitable for them.

To fill the gap of the theoretical study on the aforementioned CV applications, we derive a new upper bound of the reconstruction error for matrix completion by imposing the linear transformations into the conventional NNM framework. In contrast to the conventional NNM model, we minimize a sum of nuclear norms of the *linearly transformed*

matrices. For different applications, we can specify the linear transformations as down-sampling or filtering, *etc.*. By choosing some specific but simple linear transformations (as shown in the experiment), the new framework can easily handle the row/column missing situations. In summary, this paper makes the following contributions:

- We rigorously prove the theoretical guarantee of the proposed model, and give an upper bound of the reconstruction error influenced by noise.
- We propose an efficient algorithm to minimize the new model, by which our model significantly outperforms the state-of-the-art methods for image inpainting.

1.1. Related works

We first review the existing approaches for low-rank matrix completion. One line of work focuses on low-rank decomposition of the incomplete matrix [20, 27, 38, 39]. In these methods, the observations are represented by the multiplication of latent factors. Although low-rank matrix decomposition is a non-convex problem, in recent studies it has been proved that the local minimum of the model is also the global minimum in general cases [11]. However, how to determine the optimal rank for the decomposition is still a tough work in practice. The second line of work is to directly estimate the optimal low-rank approximation of the observation [5, 15, 19, 24]. As mentioned in Section 1, NNM is the most well known method, in which the matrix nuclear norm is minimized to find the optimal low-rank approximation. Recent studies along this line attempt to find more suitable surrogate of the matrix rank than the nuclear norm [13] or to impose additional regularization items on the model [16]. In general, such modification can eliminate the bias brought by the matrix nuclear norm, but it also leads to the non-convex model and the completion performance lacking of theoretical guarantee. Furthermore, all the aforementioned LRMC methods only consider the low-rank structure of the original matrix, but our work is to explore the additional low-rank structures resulted by linear transformations, even if the original matrix might be high-rank.

Next, we review the studies for the theoretical properties of LRMC. The theory for LRMC can be seen as an extension of compressed sensing [3]. Based on the perspective of the restricted isometry property (RIP) [7, 12], many studies proved the sample complexity of LRMC, and looked for tighter bound of the essential number of the observations for exact recovery. Instead of RIP, the approaches taken in [2, 4, 31] are based on dual theory, by which the notion of *dual certificate* is proposed as a condition to ensure the exact recovery. However, all the works do not concern how linear transformations influence the exact recovery of LRMC. In this paper, we extend the existing theoretical studies based on the dual theory, but we impose the linear transformations in the proof. From the theoretical results, we reveal how the characteristic of the linear transformations influences the completion performance.

Besides matrix completion, there also exist some methods which solve the similar problem. One example is matrix sensing [29]. In contrast to LRMC, matrix sensing uses Gaussian measure to obtain the observations instead of down-sampling. As another example, tensor completion is recently well studied using different decomposition models [9, 21, 25]. Especially, the convex tensor decomposition (CTD), as an extension of NNM, exploits the multi-linear low-rank structure to find the optimal low-rank approximation of a tensor [35, 36]. In the next section, we will discuss the differences of our model with these methods, and show that CTD is a special case of our model.

2. Preliminaries

2.1. Notation

Given a positive integer d, let $[d]$ denote the set of integers from 1 to d. We denote the scalars and matrices by Italic letters, *e.g.* a, K, and boldface capital letters, *e.g.* $\mathbf{X}$, respectively. For a matrix $\mathbf{X}$, let $\mathbf{X}(i,j)$ denote the i-th row and j-th column entry. Let $\|\cdot\|_2$ and $\|\cdot\|_F$ respectively denote the spectral norm the Frobenius norm. Let $\|\cdot\|_*$ denote the matrix nuclear norm, which equals the sum of the singular values. We denote the linear functions on matrix by calligraphic script letters, *e.g.* $\mathcal{Q}: \mathbb{R}^{m_1\times m_2} \to \mathbb{R}^{n_1\times n_2}$, and its conjugate by $\mathcal{Q}^\star$. If we define the basis for both input and output space of $\mathcal{Q}$, then the linear function can be represented by a 4-th order tensor, *e.g.* $\mathcal{Q} \in \mathbb{R}^{m_1\times m_2\times n_1\times n_2}$. Let $[\mathcal{Q}]_{\langle i\rangle}, i \in [3]$ denote unfolding the tensor $\mathcal{Q}$ along the first i-orders [28]. By using the tensor form, we denote the condition number of $\mathcal{Q}$ as $cond(\mathcal{Q})$, which is defined as

Definition 1 (Condition number)**.** *The condition number of a linear transformation with its tensor representation* $\mathcal{Q} \in \mathbb{R}^{m_1\times m_2\times n_1\times n_2}$ *is defined as the ratio of the largest and smallest non-zero singular values of* $[\mathcal{Q}]_{\langle 2\rangle}$.

The condition number in Definition 1 is used to measure how much error in the output results from an error in the input for the linear function $\mathcal{Q}$.

2.2. Nuclear norm minimization (NNM)

Assuming a perturbed low-rank matrix $\mathbf{Y} \in \mathbb{R}^{m_1\times m_2}$, the notion of NNM is formulized as:

$$\min_{\mathbf{X}\in\mathbb{R}^{m_1\times m_2}} \|\mathbf{X}\|_*, \quad s.t. \|\mathcal{P}_\Omega(\mathbf{X}) - \mathcal{P}_\Omega(\mathbf{Y})\|_F < \delta, \quad (1)$$

where $\mathbf{X}$ denotes the recovered matrix, $\mathcal{P}_\Omega(\cdot)$ is the down-sampling operation under the index set of the missing pattern Ω such that $\mathcal{P}_\Omega(\mathbf{Y})$ represents the observed entries.

Here we use the same notation for the linear functions and their tensor representation without ambiguity.

2.3. Generalization of LRMC

To leverage the additional low-rank structures of the "transformed" matrix as mentioned in Section 1, we propose a generalization of LRMC called Matrix Completion under Multiple linear Transformations (MCMT). In the model, we impose K linear transformations $\mathcal{Q}_i : \mathbb{R}^{m_1 \times m_2} \to \mathbb{R}^{n_1^{(i)} \times n_2^{(i)}}, i \in [K]$ and simultaneously consider the low-rank structures of the transformed matrices. Specifically, the model of MCMT is given by

$$\begin{aligned} &\min_{\mathbf{X} \in \mathbb{R}^{m_1 \times m_2}} \sum_{i \in [K]} \left\| \mathcal{Q}_i(\mathbf{X}) \right\|_*, \\ &s.t. \left\| \mathcal{P}_\Omega(\mathbf{X}) - \mathcal{P}_\Omega(\mathbf{Y}) \right\|_F < \delta. \end{aligned} \tag{2}$$

We can see that (2) degenerates (1) when $K = 1$ and $\mathcal{Q}_1$ is the identical function. But the difference from NNM is that MCMT seeks for the low-rank solution over the linear transformations rather than the matrix itself. It implies that, providing proper $\mathcal{Q}_i$'s, (2) can be used to complete the matrix that has a high-rank structure.

Comparison with matrix sensing. Matrix sensing is to recover the original matrix from the Gaussian measurements. The model is formalized as

$$\min_{\mathbf{X} \in \mathbb{R}^{m_1 \times m_2}} \left\| \mathbf{X} \right\|_*, \quad s.t. \left\| \mathcal{Q}(\mathbf{X}) - \mathcal{Q}(\mathbf{Y}) \right\|_F < \delta, \tag{3}$$

where the entries of $\mathcal{Q}$ follows the *i.i.d.* Gaussian distribution. Compared to (2), the model (3) only considers the linear transformation $\mathcal{Q}$ in the constraint term. Furthermore, similar to NNM, matrix sensing exploits the low-rank structure of the original matrix, while MCMT takes into account the additional low-rank structures under linear transformations.

Comparison with CTD. As mentioned in the related works, CTD is to seek for the approximation of a tensor with multi-linear low-rank structures. For a K-th order tensor and its perturbed variant $\mathcal{Y}$, CTD is given by [35]

$$\begin{aligned} &\min_{\mathcal{X} \in \mathbb{R}^{m_1 \times m_2}} \sum_{i \in [K]} \left\| [\mathcal{X}]_{(i)} \right\|_*, \\ &s.t. \left\| \mathcal{P}_\Omega(\mathcal{X}) - \mathcal{P}_\Omega(\mathcal{Y}) \right\|_F < \delta, \end{aligned} \tag{4}$$

where $[\mathcal{X}]_{(i)}$ denotes unfolding the tensor $\mathcal{X}$ along i-th order [18]. Due to the fact that the unfolding operations are linear functions, the model (4) is a special case of MCMT when $\mathcal{Q}_i(\cdot) = [\,\cdot\,]_{(i)}$. It is worthwhile to mention that tensor unfolding only rearrange the tensor into different shapes, but MCMT can exploit more general linear functions like re-sampling, rotation and filtering in the linear space to dig more structures of the matrix.

2.4. Examples of $\mathcal{Q}_i$ in MCMT

In MCMT, the linear transformations $\mathcal{Q}_i, \forall i$ can be used to formulate various operations in various CV applications. Below we show some examples.

Example 1 ("Non-local" image restoration)**.** *To exploit the non-local similarity of the images, the methods usually split the whole matrix into many "non-local groups", and each group is a concatenation of similar patches of the image. We can see that such grouping operation is mathematically a re-sampling (definitely linear) function from the image to the non-local groups. Therefore, each $\mathcal{Q}_i(\mathbf{X}), i \in [K]$ in (2) corresponds to K non-local groups, and solving (2) is to find the optimal low-rank approximation for each non-local group and then merge the approximations back to the global image.*

Example 2 (Occlusion removal)**.** *In the occlusion removal problem, the original image is partially covered by other objects, and the aim of this application is to recover the hidden part of the image. To solve this problem, a previous study [40] assumes that both the original image and the covered part have the low-rank structures. Under the framework of MCMT, we can specify $K = 2$, set $\mathcal{Q}_1$ to be the identical function to catch the low-rank structure of the image, and set $\mathcal{Q}_2$ as down-sampling to obtain the covered sub-image with the low-rank structures.*

Besides these examples, we can also specify $\mathcal{Q}_i$ as the 2-D wavelet filters to catch the short-term fluctuation of the image under multiple resolutions or even random reshuffling [22].

3. Identifiablity

One of the advantage of LRMC is that the completion performance is theoretically guaranteed. In this section, we theoretically analyze the reconstruction error for MCMT, and reveal what conditions $\mathcal{Q}_i, \forall i$ should satisfy for exact recovery. In the rest of this section, we first establish an upper bound of MCMT under a single linear transformation, *i.e.* $K = 1$. After that, we extend the results to the case of multiple transformations.

3.1. Single linear transformation

Assume that $\mathbf{M}_0 \in \mathbb{R}^{m_1 \times m_2}$ denotes the "true" matrix that we want to recover, and its rank equals R. The perturbed variant of $\mathbf{M}_0$ is generated by $\mathbf{Y} = \mathbf{M}_0 + \mathbf{H}$ where the entries of $\mathbf{H}$ obey the *i.i.d.* Gaussian distribution, *i.e.* $\mathbf{H}(i,j) \sim N(0, \sigma^2)$ for all $i \in [m_1], j \in [m_2]$. With the single linear transformation, we simplify (2) as

$$\min_{\mathbf{X} \in \mathbb{R}^{m_1 \times m_2}} \left\| \mathcal{Q}(\mathbf{X}) \right\|_* \quad s.t. \left\| \mathcal{P}_\Omega(\mathbf{X}) - \mathcal{P}_\Omega(\mathbf{Y}) \right\|_F \leq \delta, \tag{5}$$

where the subscript of $\mathcal{Q} \in \mathbb{R}^{m_1 \times m_2 \times n_1 \times n_2}$ is removed for brevity. Let $\mathcal{Q}(\mathbf{M}_0) = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ be the truncated singular value decomposition (SVD), in which only the singular vectors and non-zero singular values

are kept. Furthermore, we define a linear space $\mathbb{T} = \left\{\mathbf{U}\mathbf{X}^\top + \mathbf{Y}\mathbf{V}^\top | \forall \mathbf{X} \in \mathbb{R}^{n_1 \times R}, \mathbf{Y} \in \mathbb{R}^{n_2 \times R}\right\}$, which reflects the properties of the neighborhood around $\mathbf{M}_0$. Let $\mathbb{T}^\perp$ denote the orthogonal complement to $\mathbb{T}$. Based on the dual theory, we define the dual certificate for the unique solution to (5) as follow:

Definition 2 (Dual certificate)**.** *A matrix $\mathbf{\Lambda} \in \mathbb{R}^{m_1 \times m_2}$ is defined as a dual certificate of (5), if $\mathcal{P}_\Omega(\mathbf{\Lambda}) = \mathbf{\Lambda}$ and $\mathbf{\Lambda}$ can be decomposed as*

$$\mathbf{\Lambda} = \mathcal{Q}^\star \left(\mathbf{U}\mathbf{V}^\top + \mathbf{R}_{\mathbf{\Lambda}}\right), \tag{6}$$

where $\mathbf{R}_{\mathbf{\Lambda}} = \mathcal{P}_{\mathbb{T}^\perp}(\mathbf{\Lambda})$, $\mathcal{P}_{\mathbb{T}^\perp}$ denotes the projection to $\mathbb{T}^\perp$ and $\|\mathbf{R}_{\mathbf{\Lambda}}\|_2 \leq 1$.

The existence of the dual certificate was proved as a critical condition to ensure the exact recovery in the previous work [4]. The following lemma shows how Definition 2 certifies that the $\mathbf{M}_0$ is one of the optimal solutions to (5).

Lemma 1. *Assume that there exists a dual certificate $\mathbf{\Lambda}$ and the perturbation $\mathbf{H}$ obeys $\mathcal{P}_\Omega(\mathbf{H}) = 0$. Then the inequality*

$$\begin{aligned} &\|\mathcal{Q}(\mathbf{M}_0 + \mathbf{H})\|_* \\ &\geq \|\mathcal{Q}(\mathbf{M}_0)\|_* + (1 - \|\mathbf{R}_{\mathbf{\Lambda}}\|_2)\, \|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{Q}(\mathbf{H})\|_* \end{aligned} \tag{7}$$

holds.

The proof is given in the supplementary material.

Lemma 1 reflects how the perturbation $\mathbf{H}$ changes the objective function in (5), and any perturbation around $\mathbf{M}_0$ cannot decrease the objective function. It means that $\mathbf{M}_0$ is one of the solutions to (5). Furthermore, Lemma 1 degenerates to Lemma 4 in [4] if we specify $\mathcal{Q}$ to be the identity.

However, the imposed linear transformation $\mathcal{Q}$ leads to a new problem compared to the conventional studies. Lemma 1 cannot guarantee that $\mathbf{M}_0$ is the unique solution to (5). It can be seen from the second term of the right side of (7) that the perturbation $\mathbf{H}$ could vanish in the null space of $\mathcal{Q}$. To prove the theoretical guarantee for (5), we further make the following assumption to restrict the relation of the null space of $\mathcal{Q}$ and $\mathcal{P}_\Omega$:

Assumption 1. *Let $\mathbb{N}_\mathcal{Q}$ and $\mathbb{N}_\Omega$ denote the null spaces of the linear transformations $\mathcal{Q}$ and $\mathcal{P}_\Omega$, respectively. We assume that the relation $\mathbb{N}_\mathcal{Q} \cap \mathbb{N}_\Omega = \{0\}$ obeys.*

It can be inferred from Assumption 1 that $\mathcal{Q}(\mathbf{H}) \neq 0$ for all the perturbation $\mathbf{H}$ with $\mathcal{P}_\Omega(\mathbf{H}) = 0$. Assumption 1 can be also considered as the strong convexity assumption of (5), which is illustrated by the following proposition:

Proposition 1. *Assume that the null space of $\mathcal{P}_\Omega$ is not trivial, i.e. $\mathbb{N}_\Omega \neq \{0\}$, and $\mathbf{M}_0$ is the unique solution to (5), then Assumption 1 should be satisfied.*

Proof. For contradiction, assume that Assumption 1 is not satisfied, then there exists a non-zero perturbation $\mathbf{H} \in \mathbb{N}_\Omega$ on the missing entries such that $\mathcal{Q}(\mathbf{H}) = 0$ holds. Thus the objective function

$$\|\mathcal{Q}(\mathbf{M}_0 + \mathbf{H})\|_* = \|\mathcal{Q}(\mathbf{M}_0) + \mathcal{Q}(\mathbf{H})\|_* = \|\mathcal{Q}(\mathbf{M}_0)\|_*. \tag{8}$$

It implies that $\hat{\mathbf{M}} = \mathbf{M}_0 + \mathbf{H} \neq \mathbf{M}_0$ is also the solution of (5), which violates the assumption in the proposition. □

The strong convexity of (5) guarantees the uniqueness of the solution. With the dual certificate and Assumption 1, we propose the main result of our paper in the case of single linear transformation.

Theorem 1 (Error bound for a single $\mathcal{Q}$)**.** *With Assumption 1, and suppose the additional assumptions:*

i) There exists a dual certificate obeying $\|\mathbf{R}_{\mathbf{\Lambda}}\|_2 < 1$;

ii) $\exists p > 0, s.t. \mathcal{P}_\mathbb{T}\mathcal{Q}\mathcal{P}_\Omega\mathcal{Q}^\star\mathcal{P}_\mathbb{T} \succeq p\mathcal{I}$

iii) The product $[\mathcal{Q}]_{\langle 2\rangle} \cdot [\mathcal{Q}]^\star_{\langle 2\rangle}$ is a diagonal matrix.

Then the solution of (5) $\hat{\mathbf{M}}$ obeys

$$\begin{aligned} &\|\hat{\mathbf{M}} - \mathbf{M}_0\|_F \\ &\leq 2\delta \cdot \frac{cond(\mathcal{Q})}{1 - \|\mathbf{R}_{\mathbf{\Lambda}}\|_2} \sqrt{\frac{\min\{n_1, n_2\}(p + \|\,[\mathcal{Q}]_{\langle 2\rangle}\,\|_2^2)}{p}}. \end{aligned} \tag{9}$$

The proof is given in the supplemental material.

First, We consider the four assumptions given in Theorem 1. As mentioned above, Assumption 1 and *(i)* guarantee the uniqueness of the solution to (1). However, it is not ensured that the unique solution $\hat{\mathbf{M}}$ equals the "true" matrix $\mathbf{M}_0$ because of the existence of projection $\mathcal{P}_{\mathbb{T}^\perp}$ in (7). To constraint the structure of $\mathcal{P}_{\mathbb{T}^\perp}$, we impose the assumption *(ii)*, which implies that a sequential product of the functions $\mathcal{P}_\mathbb{T}$, $\mathcal{Q}$ and $\mathcal{P}_\Omega$ is positive definite, and the constant p can be considered as the "incoherence" level among these functions. The assumption *(iii)* further constraints the rows of $[\mathcal{Q}]_{\langle 2\rangle}$ to be orthogonal to each other, which is useful for the proof procedure of the theorem.

Second, we consider the upper bound given in (9). We can find that the reconstruction error of (5) is upper bounded by not only the noise strength δ but also the properties of the linear transformation $\mathcal{Q}$. If $\mathcal{Q}$ is a well-posed linear function, then the upper bound from (9) tends to be zero when decreasing δ. It implies that the missing entries can be exactly recovered if there is no noise.

3.2. Extension to multiple $\mathcal{Q}$'s

Compared to the case using a single $\mathcal{Q}$, the model (2) exploits the multiple low-rank structures under different linear transformations. To bound the reconstruction error of (2), we construct a block diagonal tensor $\widetilde{\mathcal{Q}}$ by which we have $\|\widetilde{\mathcal{Q}}(\mathbf{X})\|_* = \sum_{i\in[K]} \|\mathcal{Q}_i(\mathbf{X})\|_*$. Mathematically, the entries of the constructed 4th-order tensor $\widetilde{\mathcal{Q}} \in \mathbb{R}^{m_1\times m_2\times \prod_{i\in[K]} n_1^{(i)}\times \prod_{j\in[K]} n_2^{(j)}}$ obey the following equation:

$$\widetilde{\mathcal{Q}}(:,:,\prod_{k\in[i-1]} N_1^{(i)} : \prod_{k\in[i]} N_1^{(i)}, \prod_{k\in[i-1]} N_2^{(i)} : \prod_{k\in[i]} N_2^{(i)}) = \mathcal{Q}_i,\ i\in[K], \tag{10}$$

where we use the Matlab syntax to denote a sub-block of a tensor. Using $\widetilde{\mathcal{Q}}$, it is easy to find that $\widetilde{\mathcal{Q}}(\mathbf{X})$ is a block diagonal matrix of which each block equals $\mathcal{Q}_i(\mathbf{X})$. Due to the block diagonal structure of $\widetilde{\mathcal{Q}}(\mathbf{X})$, the following equation holds

$$\|\widetilde{\mathcal{Q}}(\mathbf{X})\|_* = \sum_{i\in[K]} \|\mathcal{Q}_i(\mathbf{X})\|_*, \quad \forall \mathbf{X}\in\mathbb{R}^{m_1\times m_2}. \tag{11}$$

Using (11), the model (2) under multiple linear transformations $\mathcal{Q}_i, i\in[K]$ can be converted to the model (5) under a single transformation $\widetilde{\mathcal{Q}}$. Hence Theorem 1 can be directly extended to bound the reconstruction of (2).

Corollary 1 (Error bound for multiple $\mathcal{Q}_i$'s)**.** *Let $\widetilde{\mathcal{Q}}$ be a concatenation of $\mathcal{Q}_i, i\in[K]$ by using (10), and $\widetilde{\mathcal{Q}}$ obeys the assumptions given in Theorem 1. Then the reconstruction error of the solution to (2) obeys*

$$\|\hat{\mathbf{M}}-\mathbf{M}_0\|_F \le 2\delta\cdot\frac{cond(\widetilde{\mathcal{Q}})}{1-\|\mathbf{R}_{\mathbf{\Lambda}}\|_2}\sqrt{\frac{\min\{n_1,n_2\}(p+\|\left[\widetilde{\mathcal{Q}}\right]_{\langle 2\rangle}\|_2^2)}{p}}. \tag{12}$$

Although MCMT under multiple $\mathcal{Q}_i$'s can be converted to the case under the single $\widetilde{\mathcal{Q}}$, we find that imposing more linear transformations could result in easier conditions to exactly recover the missing entries. This statement can be partially supported by the following proposition:

Proposition 2. *Let $\mathbb{N}_{\mathcal{Q}_i}$ and $\mathbb{N}_{\widetilde{\mathcal{Q}}}$ be the null space of $\mathcal{Q}_i$, $i\in[K]$ and $\widetilde{\mathcal{Q}}$, respectively. Then it yields that $\mathbb{N}_{\widetilde{\mathcal{Q}}} = \bigcap_{i\in[K]} \mathbb{N}_{\mathcal{Q}_i}$.*

The proof is trivial. It can be inferred from Proposition 2 that the null space of $\widetilde{\mathcal{Q}}$ is "smaller" than the one of any single linear transformation $\mathcal{Q}_i$. It means that Assumption 1 can be still held by $\widetilde{\mathcal{Q}}$ even if each $\mathcal{Q}_i$ cannot satisfy Assumption 1. Therefore, it might be easier for MCMT to exactly reconstruct the missing entries if we impose more $\mathcal{Q}_i$'s.

To develop an efficient algorithm for MCMT, it would be better to consider the following optimization model, in which the constraint in (2) is absorbed into the objective function such that solving MCMT becomes a unconstrained problem.

$$\min_{\mathbf{X}\in\mathbb{R}^{m_1\times m_2}} \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{X})-\mathcal{P}_\Omega(\mathbf{Y})\|_F^2 + \sum_{i\in[K]}\lambda_i\|\mathcal{Q}_i(\mathbf{X})\|_*. \tag{13}$$

Besides the constraint term, we also impose a set of tuning parameters $\lambda_i, i\in[K]$ for each $\mathcal{Q}_i$ to trade-off the influences on the objective function by different linear transformations. To analyze the performance of (13), we assume $\lambda=\lambda_i, \forall i$ for brevity. The error bound of (13) is derived by the following theorem:

Theorem 2. *With the assumptions in Theorem 1 for the concatenation $\widetilde{\mathcal{Q}}$, and further assume that the tuning parameter satisfies $\lambda > \|\mathcal{P}_\Omega(\mathbf{H})\|_2/\sqrt{\min\{m_1,m_2\}}$. Then the reconstruction error of (13) is bounded by*

$$\|\hat{\mathbf{M}}-\mathbf{M}_0\|_F \le 8\lambda\left(\min\{m_1,m_2\}+\sum_{i\in[K]}\sqrt{\min\{n_1^{(i)},n_2^{(i)}\}}\|[\mathcal{Q}_i]_{\langle 2\rangle}\|_2\right)\cdot\frac{cond(\widetilde{\mathcal{Q}})\cdot\min\{\prod n_1^{(i)},\prod n_2^{(i)}\}(p+\|\left[\widetilde{\mathcal{Q}}\right]_{<2>}\|_2)}{p(1-\|\mathbf{R}_{\mathbf{\Lambda}}\|_2)^2}. \tag{14}$$

Note that λ can be an arbitrarily small number if the strength of the perturbation matrix $\mathbf{H}$ is weak enough. Therefore the performance of (13) is theoretically guaranteed under above conditions.

4. Algorithm

In this section, we present an efficient algorithm to solve the optimization problem (13).

Due to the convexity of the model, we use the augmented Lagrangian multipliers to search the minimum of the model. Inspired by the existing tensor nuclear norm minimization methods [25], we also impose auxiliary variables to get a simple algorithm. Specifically, we rewrite (13) as:

$$\min_{\mathbf{X},\mathbf{W}_i,i\in[K]} \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{X})-\mathcal{P}_\Omega(\mathbf{Y})\|_F^2 + \sum_{i\in[K]}\lambda_i\|\mathbf{W}_i\|_*, \quad s.t.\ \mathbf{W}_i=\mathcal{Q}_i(\mathbf{X}),\ i\in[K] \tag{15}$$

https://www.mathworks.com/products/matlab.html

where $\mathbf{W}_i, i \in [K]$ denote the auxiliary matrices to decompose the objective function of (13) into several parts with independent variables. But it is easy to find that (15) is with the same solution to (13). The corresponding augmented Lagrangian function of (15) is given by

$$\begin{aligned} &L(\mathbf{X}, \mathbf{W}_i, \mathbf{Z}_i, \forall i \in [K], \sigma) \\ =&\frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{X}) - \mathcal{P}_\Omega(\mathbf{Y})\|_F^2 + \sum_{i\in[K]} \lambda_i \|\mathbf{W}_i\|_* \\ &+ \sum_{i\in[K]} \langle \mathbf{Z}_i, \mathcal{Q}_i(\mathbf{X}) - \mathbf{W}_i \rangle + \frac{\sigma}{2} \sum_{i\in[K]} \|\mathcal{Q}_i(\mathbf{X}) - \mathbf{W}_i\|_F^2 \end{aligned}, \tag{16}$$

where $\mathbf{Z}_i$ denotes the Lagrangian multipliers for all $i \in [K]$ and $\sigma > 0$. In the algorithm, we alternatively update each variable in (16) until convergence, and the updating order is given by

$$1)\, \mathbf{W}_i, i \in [K] \;\Rightarrow\; 2)\, \mathbf{X} \;\Rightarrow\; 3)\, \mathbf{Z}_i, i \in [K] \;\Rightarrow\; 4)\, \sigma.$$

Updating $\mathbf{W}_i$. We treat all variables except $\mathbf{W}_i$ as constants during updating. In this case, (16) can be rewritten as

$$\begin{aligned} &\hat{L}(\mathbf{W}_i) = \\ &\lambda_i \|\mathbf{W}_i\|_* + \langle \mathbf{Z}_i, \mathcal{Q}_i(\mathbf{X}) - \mathbf{W}_i \rangle + \frac{\sigma}{2}\|\mathcal{Q}(\mathbf{X}) - \mathbf{W}_i\|_F^2, \end{aligned} \tag{17}$$

where we omit the constant terms. It is known from [1] that minimizing (17) has a closed-form solution, which is given by

$$\mathbf{W}_i^+ := D_{\frac{\lambda_i}{\sigma}} \left(\mathbf{Q}_i(\mathbf{X}) + \frac{1}{\sigma}\mathbf{Z}_i \right), \tag{18}$$

where $D_{\frac{\lambda_i}{\sigma}}(\cdot)$ denotes the soft-thresholding operation [23]. If $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ is the SVD of $\mathbf{X}$, then $\mathbf{D}_\lambda(\mathbf{X}) = \mathbf{U}\bar{\mathbf{D}}\mathbf{V}^\top$, where the entries $\bar{\mathbf{D}}(i,j), \forall i, j$ satisfies

$$\bar{\mathbf{D}}(i,j) = \begin{cases} \mathbf{D}(i,j) - \lambda & \mathbf{D}(i,j) > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{19}$$

Updating $\mathbf{X}$. Likewise, we treat all the variables except $\mathbf{X}$ as the constants during updating. In this case, (16) is rewritten as

$$\begin{aligned} \hat{L}(\mathbf{X}) = &\frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{X}) - \mathcal{P}_\Omega(\mathbf{Y})\|_F^2 \\ &+ \sum_{i\in[K]} \langle \mathbf{Z}_i, \mathcal{Q}_i(\mathbf{X}) - \mathbf{W}_i \rangle + \frac{\sigma}{2} \sum_{i\in[K]} \|\mathcal{Q}_i(\mathbf{X}) - \mathbf{W}_i\|_F^2 \end{aligned}. \tag{20}$$

We can see that solving (20) is a least squares problem, and its solutions is given by

$$\begin{aligned} \mathbf{X}^+ := &\left(\mathcal{P}_\Omega^\star \mathcal{P}_\Omega + \sigma \sum_{i\in[K]} \mathcal{Q}_i^\star \mathcal{Q}_i \right)^{-1} \\ &\cdot \left[\left(\mathcal{P}_\Omega^\star \mathcal{P}_\Omega\right) \mathbf{Y} - \sum_{i\in[K]} \mathcal{Q}_i^\star \left(\mathbf{Z}_i - \mathbf{W}_i \right) \right]. \end{aligned} \tag{21}$$

Updating $\mathbf{Z}_i$ and σ. We update the Lagrangian multipliers $\mathbf{Z}_i, i \in [K]$ and σ by using

$$\mathbf{Z}_i^+ := \mathbf{Z}_i + \sigma(\mathcal{Q}_i(\mathbf{X}) - \mathbf{W}_i) \tag{22}$$

and

$$\sigma^+ := \rho\sigma, \tag{23}$$

where $\rho > 1$ is a constant. The details of the algorithm are given by Algorithm 1.

Algorithm 1 Matrix completion under multiple linear transformations (MCMT)

Input: Observation $\mathcal{P}_\Omega(\mathbf{Y})$ and its corresponding downsampling projection $\mathcal{P}_\Omega$. The linear transformations $\mathcal{Q}_i$, and tuning parameters $\lambda_i > 0, i \in [K]$

Output: Reconstructed matrix $\mathbf{X}$.

1: Initialize $\mathbf{X}^0$ by $\mathcal{P}_\Omega(\mathbf{X}^0) = \mathcal{P}_\Omega(\mathbf{Y})$, and fill unobserved entries by zero. Let $\mathbf{W}_i^0$ equal $\mathcal{Q}_i(\mathbf{X}^0)$ for all $i \in [K]$. Initialize $\mathbf{Z}_i^0 = sgn(\mathbf{W}_i^0)$, where $sgn(\cdot)$ denotes elementwisely choosing the sign of the entries. Let $\sigma = 1, \rho = 1.01$.
2: **Repeat**
3: Update $\mathbf{W}_i, \forall i$ by (18)
4: Update $\mathbf{X}$ by (21)
5: Update $\mathbf{Z}_i, \forall i$ by (22)
6: Update σ by (23)
7: **Until** convergence

4.1. The sketching trick

We can see one computational bottleneck in Algorithm 1 is the numerous SVD operations when updating $\mathbf{W}_i$. To speed up our method, we utilize the sketching trick to reduce the computational complexity. Specifically, we generate the Gaussian random projection matrix $\mathbf{P}_i \in \mathbb{R}^{n_2^{(i)} \times l^{(i)}}$ *w.r.t.* each $\mathcal{Q}_i$. By using the random projections, we estimate the left singular vectors of $\mathcal{Q}_i(\mathbf{X})$ by using QR decomposition, *i.e.* $[\hat{\mathbf{U}}_i, \sim] = qr\left(\mathcal{Q}_i\left(\mathbf{X}\right) \cdot \mathbf{P}_i\right)$. When obtaining $\hat{\mathbf{U}}_i$'s, the SVD operation on $\mathcal{Q}_i(\mathbf{X})$ in Algorithm 1 can be replaced by $\mathcal{Q}_i(\mathbf{X}) = \hat{\mathbf{U}}_i \mathbf{D}_{tmp} \mathbf{V}_{tmp}^\top$, where $\mathbf{D}_{tmp}$ and $\mathbf{V}_{tmp}$ respectively denote the matrices which contain the singular values and right singular vectors of $\hat{\mathbf{U}}_i^\top \mathcal{Q}_i(\mathbf{X})$. By using the sketching trick, the computational complexity in the SVD procudure can be reduced from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ if $n = n_1^{(i)} = n_2^{(i)}$ and $l^{(i)} \ll n$.

Table 1. Performance comparison for image inpainting by using different methods. In the experiment, we utilize PSNR(dB) to quantify the reconstruction. Furthermore, we consider various missing patterns, including Uniformly random missing (U), Row missing (R), Column missing (C), and their combinations.

Missing pattern	Observation ratio	FBCP	SPC	TRALS	FaLRTC	MCMT	NNM	MCMT (sketching)
U	0.1	19.18 ± 1.06	17.68 ±1.01	17.70 ±1.68	11.50 ±1.24	**22.08±1.58**	14.29±1.47	22.06±1.57
	0.3	22.62 ±1.27	21.05 ±0.99	23.03 ± 1.34	18.25 ±1.05	26.39±2.10	22.05±1.53	**26.78±2.10**
	0.5	25.06 ± 1.41	23.36 ± 1.02	24.93 ±1.33	21.93 ±0.99	**29.03±2.25**	25.61±1.53	28.75±2.30
	0.7	27.78 ± 1.46	26.05 ±1.04	27.29 ±1.33	26.27 ±1.18	**31.20±2.28**	28.48±1.46	30.69±2.39
	0.9	32.88 ±1.47	31.15 ±1.06	32.13 ±1.33	**33.43±1.52**	33.22±2.24	31.40±1.19	32.37±2.42
R	0.1	1.01 ±7.73	12.37 ± 1.80	-	5.59 ±1.14	**17.51±1.45**	6.23±0.96	**17.51±1.45**
	0.3	12.70 ±3.23	18.17 ± 1.00	-	9.86 ±0.98	**22.38±1.75**	9.10±0.95	**22.38±1.75**
	0.5	19.86±1.96	21.15 ± 0.85	11.95 ±5.45	16.14 ±1.01	**25.54±1.80**	10.67±1.78	25.49±1.79
	0.7	25.21 ± 1.86	24.47 ±0.82	24.14 ±2.32	21.60 ±0.89	**28.67±2.22**	9.68±1.84	28.47±2.22
	0.9	31.36 ± 1.48	30.02 ±0.95	30.72 ±1.82	29.22 ±1.15	**31.97±2.25**	12.55±1.93	31.78±2.33
C	0.1	2.92 ±3.35	2.27 ± 1.86	-	5.58 ±1.14	**17.30±1.37**	7.74±0.86	**17.30±1.37**
	0.3	12.70 ±3.29	18.14±1.20	-	9.88 ±1.10	22.13±1.33	5.08±1.68	**22.13±1.32**
	0.5	19.74 ±1.50	21.20 ± 1.01	15.20 ±2.64	16.23 ±1.12	**25.40±1.42**	0.38±1.45	25.36±1.42
	0.7	24.86 ± 1.20	24.45 ±0.94	24.12 ±1.42	21.59 ±1.11	**28.42±1.74**	-0.80±1.43	28.26±1.76
	0.9	31.00 ± 1.57	29.93 ±1.00	30.19 ±1.60	28.94 ±1.08	**32.03±2.16**	2.42±1.59	31.52±2.27
RC	0.1	6.73 ± 1.24	5.20 ±1.11	-	4.91 ±1.15	**15.01±1.23**	5.18±0.99	**15.01±1.23**
	0.3	13.93 ±1.84	15.44 ± 1.22	-	6.45 ±1.10	**19.69±1.24**	7.45±0.64	**19.69±1.24**
	0.5	18.14 ±1.13	18.52 ± 0.96	9.24±5.94	12.43 ±0.97	**22.94±1.54**	9.97±1.74	22.93±1.54
	0.7	22.35 ± 1.17	21.59 ±0.85	21.52 ±1.79	18.43 ±0.97	**26.30±1.76**	2.08±1.66	26.24±1.76
	0.9	28.74 ± 1.34	26.88 ±0.86	27.51 ±1.55	25.84 ±0.95	**30.61±2.18**	1.43± 1.64	30.27±2.22
URC	0.1	4.85 ± 1.15	4.84 ±1.15	-	4.84 ±1.15	12.68±1.24	5.15±0.99	**12.68±1.23**
	0.3	14.89 ± 1.61	13.01 ± 1.30	-	5.21 ±1.14	**18.11±0.94**	5.63±0.96	18.10±0.93
	0.5	17.94 ± 0.99	17.31 ±1.04	11.59 ±5.94	10.18 ±0.99	21.74±1.23	7.87±0.74	**21.75±1.23**
	0.7	21.57 ± 0.97	20.38 ±0.92	20.67±1.57	16.92 ±1.02	**25.31±1.70**	14.74±0.84	25.27±1.70
	0.9	27.09 ± 1.34	25.56 ±0.89	26.42 ± 1.26	24.56 ±1.01	**29.90±2.07**	13.90±1.83	29.59±2.10

5. Experiments

In this section, we employ the proposed method on the classical grayscale image inpainting problem to demonstrate the effectiveness of our method.

Although there exist many methods proposed for the image inpainting problem, it is still a challenging problem to complete a grayscale image with missing entries. Compared to the RGB images, there is no additional similarity from RGB channels in the grayscale images, which can be utilized to improve the completion performance.

In the experiment, we choose 12 benchmark images (256×256) to evaluate the performance of the proposed method. To generate the test dataset, we employ 5 different missing patterns to remove the entries including Uniformly random missing (U), random Row-missing (R), random Column-missing (C) and their combinations (RC and URC). In addition, we set 5 different observation ratios $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ for each type of the missing pattern. To achieve more reliable experimental results, we *i.i.d.* generate 20 samples for each image with the given missing pattern and observation ratio.

The experimental results are shown in Table 1 and Fig. 1, in which the mean and the standard deviation of the PSNR is calculated using all images and samples under each missing pattern and observation ratio. For comparison, we also employ the current state-of-the-art methods in the experiment, including three tensor decomposition based methods (FBCP [46], FaLRTC [25] and TRALS [37]), a spatial smoothness based method (SPC) [44], and NNM as the baseline.

For our method, we specify the number of the linear transformation to be equal to $K = 1$ and $\lambda = 1$ for brevity, and simply set the transformation $\mathcal{Q}$ to be a two dimensional (2D) differential filtering by using the taps

$$\mathbf{T} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{24}$$

and the tensor representation of $\mathcal{Q}$ is given by the circular form of the 2D filter $\mathbf{T}$. Furthermore, we also consider using sketching to speed up our method, by which the dimension is decreased to 100 after random projection (the original dimension equals 256).

As shown in Table 1, MCMT outperforms other methods in all cases. Especially when the observation ratio is low, the performance gap between our method and others is significant. For example, the PSNR of MCMT is averagely 4dB higher than the state-or-the-art methods for all the missing pattern when the observation ratio equals 0.5. Furthermore, the performance of MCMT is about 5dB higher than the baseline method NNM in the uniformly ran-

The benchmark images include cameraman, house, jetplane, lake, lena, livingroom, mandril, peppers, pirate, walkbridge and blonde. All these images are shown in the supplementary material.

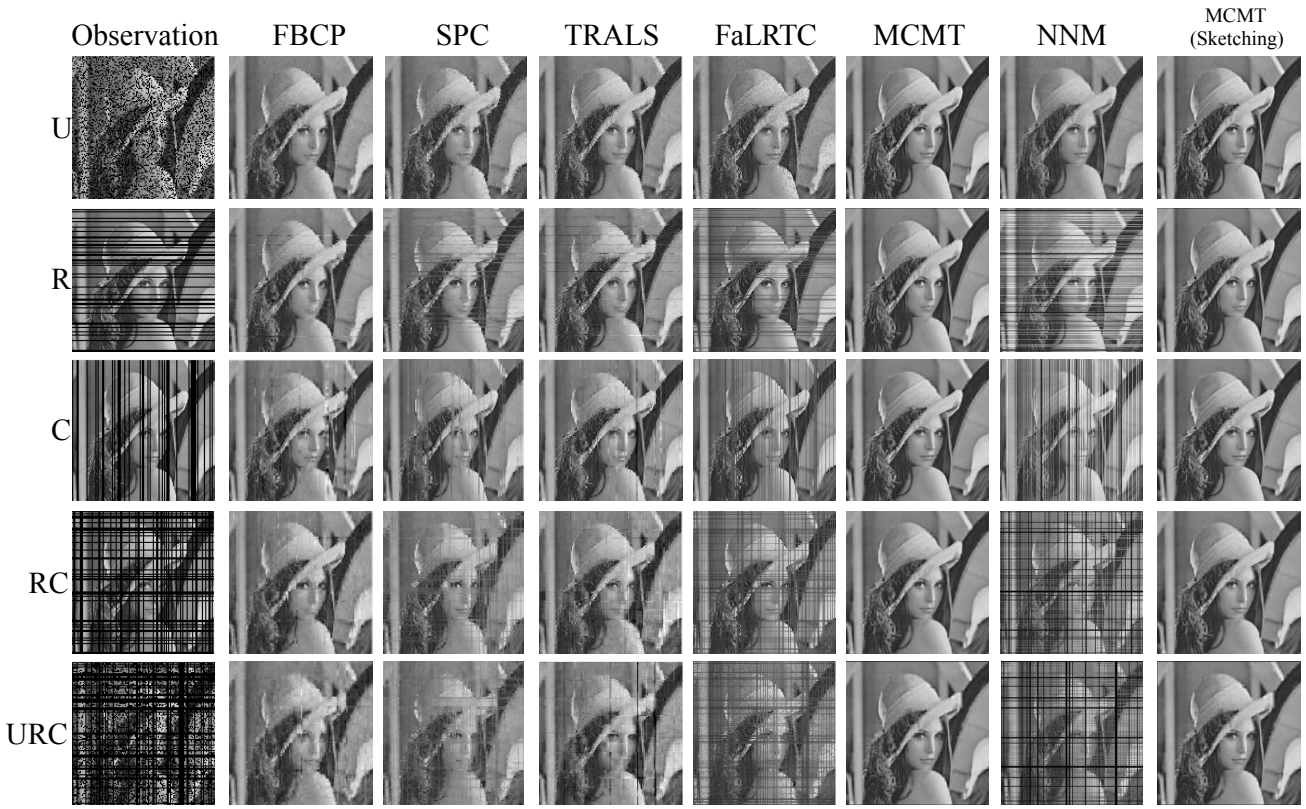

Figure 1. Examples of the completion results by using different methods, in which the rows correspond different missing patterns with the obervation ratio equals 0.7, and the columns correspond different methods. It can be easily seen that MCMT and its sketching version obtain higher qualified results.

dom missing pattern (U), and has more than 10dB performance improvement in other missing patterns. Comparing the performance between MCMT and its sketching version, we can find the sketching trick does not decease the performance of MCMT with a appropriate projection dimensions.

The performance improvement by MCMT can be expected because the existence of $\mathcal{Q}$ can enhance the incoherence between low-rank structure of the (transformed) matrix and its missing pattern. For example, NNM fails to recover the whole row/column missing because minimizing the rank of the original matrix tends to fill the missing rows/columns with zeros, which is definitely incorrect for completion. However, imposing the linear transformations can make the method take into account the dependence of the entries in different rows and columns to recover the missing entries.

6. Conclusion

Although there exist lots of methods for the data completion problem, matrix completion is still a basic but important issue for the further development of more sophisticated methods like tensor completion. Furthermore, it is also a challenging task to develop a theoretical framework to analyze the performance of the completion methods. In this paper, inspired by the existing studies on the CV applications, we proposed a new framework to leverage the additional low-rank structures of the data for the completion problem. In contrast to the conventional matrix completion methods, we impose multiple linear transformations into the model. As the theoretical result, we rigorously prove an upper bound for the reconstruction error of our method, and it implies that our model can get theoretically guaranteed performance under some conditions. Besides the theoretical works, we also developed an efficient algorithm by using the augmented Lagrangian multipliers. The experimental results show that the proposed method significantly improves the performance for image inpainting compared with the state-of-the-art methods.

References

[1] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. 6

[2] E. Candes and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Informa-*

tion Theory, 56(5):2053–2080, 2010. 2

[3] E. J. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346(9-10):589–592, 2008. 2

[4] E. J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010. 2, 4

[5] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009. 1, 2

[6] Q. Cheng, H. Shen, L. Zhang, and P. Li. Inpainting for remotely sensed images with a multichannel nonlocal total variation model. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1):175–187, 2014. 1

[7] M. A. Davenport and J. Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622, 2016. 2

[8] W. Dong, G. Shi, and X. Li. Nonlocal image restoration with bilateral variance estimation: a low-rank approach. *IEEE transactions on image processing*, 22(2):700–711, 2013. 1

[9] S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011. 2

[10] R. S. Ganti, L. Balzano, and R. Willett. Matrix completion under monotonic single index models. In *Advances in Neural Information Processing Systems*, pages 1873–1881, 2015. 1

[11] R. Ge, J. D. Lee, and T. Ma. Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems*, pages 2973–2981, 2016. 2

[12] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011. 2

[13] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 2

[14] B.-J. Han and J.-Y. Sim. Reflection removal using low-rank matrix completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5438–5446, 2017. 1

[15] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674. ACM, 2013. 1, 2

[16] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst. Matrix completion on graphs. *arXiv preprint arXiv:1408.1717*, 2014. 2

[17] Z. Kang, C. Peng, and Q. Cheng. Top-n recommender system via matrix completion. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016. 1

[18] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009. 3

[19] V. Koltchinskii, K. Lounici, A. B. Tsybakov, et al. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329, 2011. 2

[20] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008. 2

[21] D. Kressner, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014. 2

[22] C. Li, M. E. Khan, S. Xie, and Q. Zhao. Low-rank tensor decomposition via multiple reshaping and reordering operations. *arXiv preprint arXiv:1805.08465*, 2018. 3

[23] C. Li, Q. Zhao, J. Li, A. Cichocki, and L. Guo. Multi-tensor completion with common structures. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015. 6

[24] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010. 2

[25] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–220, 2013. 2, 5, 7

[26] Q. Liu, S. Li, J. Xiao, and M. Zhang. Multi-filters guided low-rank tensor coding for image inpainting. *Signal Processing: Image Communication*, 2018. 1

[27] H. Ma, H. Yang, M. R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 931–940. ACM, 2008. 2

[28] I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011. 2

[29] D. Park, A. Kyrillidis, C. Caramanis, and S. Sanghavi. Non-square matrix sensing without spurious local minima via the burer-monteiro approach. *arXiv preprint arXiv:1609.03240*, 2016. 2

[30] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016. 1

[31] B. Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(Dec):3413–3430, 2011. 2

[32] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010. 1

[33] L. Sun, B. Jeon, Y. Zheng, and Z. Wu. Hyperspectral image restoration using low-rank representation on spectral difference image. *IEEE Geoscience and Remote Sensing Letters*, 14(7):1151–1155, 2017. 1

[34] R. Sun and Z.-Q. Luo. Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory*, 62(11):6535–6579, 2016. 1

[35] R. Tomioka, K. Hayashi, and H. Kashima. On the extension of trace norm to tensors. In *NIPS Workshop on Tensors, Kernels, and Machine Learning*, page 7, 2010. 2, 3

[36] R. Tomioka and T. Suzuki. Convex tensor decomposition via structured schatten norm regularization. In *Advances in neural information processing systems*, pages 1331–1339, 2013. 2

[37] W. Wang, V. Aggarwal, and S. Aeron. Efficient low rank tensor ring completion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5697–5705, 2017. 7

[38] Z. Wen, W. Yin, and Y. Zhang. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation*, 4(4):333–361, 2012. 2

[39] Y. Xu, R. Hao, W. Yin, and Z. Su. Parallel matrix factorization for low-rank tensor completion. *arXiv preprint arXiv:1312.1254*, 2013. 2

[40] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu. Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE transactions on pattern analysis and machine intelligence*, 39(1):156–171, 2017. 1, 3

[41] J. Yang and X. Yuan. Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization. *Mathematics of computation*, 82(281):301–329, 2013. 1

[42] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5485–5493, 2017. 1

[43] T. Yokota, B. Erem, S. Guler, S. K. Warfield, and H. Hontani. Missing slice recovery for tensors using a low-rank model in embedded space. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8251–8259, 2018. 1

[44] T. Yokota, Q. Zhao, and A. Cichocki. Smooth parafac decomposition for tensor completion. *IEEE Transactions on Signal Processing*, 64(20):5423–5436, 2016. 7

[45] D. Zhang, Y. Hu, J. Ye, X. Li, and X. He. Matrix completion by truncated nuclear norm regularization. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2192–2199. IEEE, 2012. 1

[46] Q. Zhao, L. Zhang, and A. Cichocki. Bayesian cp factorization of incomplete tensors with automatic rank determination. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1751–1763, 2015. 7