Robust Matrix Completion by Exploiting Dynamic Low-dimensional Structures

Ren Wang\textsuperscript{1,2}†, Pengzhi Gao\textsuperscript{1,3}† and Meng Wang\textsuperscript{1*}

Abstract
This paper studies the robust matrix completion problem for time-varying models. Leveraging the low-rank property and the temporal information of the data, we develop novel methods to recover the original data from partially observed and corrupted measurements. We show that the reconstruction performance can be improved if one further leverages the information of the sparse corruptions in addition to the temporal correlations among a sequence of matrices. The dynamic robust matrix completion problem is formulated as a nonconvex optimization problem, and the recovery error is quantified analytically and proved to decay in the same order as that of the state-of-the-art method when there is no corruption. A fast iterative algorithm with convergence guarantee to the stationary point is proposed to solve the nonconvex problem. Experiments on synthetic data and real video dataset demonstrate the effectiveness of our method.

Keywords: robust matrix completion; low-rank; temporal correlation; proximal alternative gradient descent; nonconvex optimization

1 Introduction
Many practical datasets have intrinsic low-dimensional structures despite the high ambient dimension. Motivated by applications like image and video processing [1, 2], remote sensing [3] and collaborative filtering [4], the low-dimensional structures have been extensively studied in problems like Low-Rank Matrix Completion (LRMC) [5, 6] and Robust Principal Component Analysis (RPCA) [7, 8]. LRMC aims to recover the unobserved entries of a rank-\(r\) matrix \(\bar{L}\) from the partial observations. Given some entries of matrix \(\bar{L}\), the goal of LRMC is to recover \(\bar{L}\) from those observed entries. In a more general setup of low-rank matrix recovery, the observations can be linear function of \(\bar{L}\) instead of partial entries [9]. RPCA decomposes a given data matrix into the sum of a low-rank matrix (true data) and a sparse matrix (outliers). Given \(M = \bar{L} + \bar{C}\), where \(\bar{L}\) is a low-rank matrix, and \(\bar{C}\) is a sparse matrix with at most \(s\) nonzero entries, RPCA aims to separate \(\bar{L}\) and \(\bar{C}\) from \(M\). For example, in power system monitoring, \(\bar{L}\) models the spatial-temporal blocks of ground-truth data, and \(\bar{C}\) models the erroneous measurements [10]. This problem can be formulated as nonconvex optimization problems, due to the rank and sparsity constraints. Solving convex relaxations of such problem by replacing the rank constraint with the nuclear norm [11, 12, 5] constraint and the sparsity constraint with the \(\ell_1\)-norm constraint [13] are proven to return the original matrix under certain conditions [7, 11]. Since the nonconvex formulation directly addresses the low-rank constraint, it is more effective than the convex relaxations and has attracted much attention recently. Despite the effectiveness, it is generally more challenging to prove the convergence for nonconvex problems, and it is even harder to prove the convergence to the global optimum. We list a few works that provide the convergence guarantees for some nonconvex approaches [14, 15, 16, 17, 18, 19, 20]. Except for one work that considers the robust PCA [15], all the rest of the listed works focus on the low-rank matrix completion. In this paper, we also consider the existence of sparse error matrices and guarantee the convergence of our algorithm to stationary points.

The existing work on LRMC and RPCA assume that the low-dimensional structure does not change over time and consider one fixed low-rank matrix. However, for example, users’ preference may change over time in collaborative filtering [21]. In this case, a more general model is to assume that the data lie in a low-dimensional subspace that gradually changes over time. The problem of tracking data points that lie in a gradually changing subspace and being robust
to the sparse outliers is referred to Dynamic RPCA [22, 23, 24, 25]. Ref. [22] studies the online version of RPCA where the sparsity pattern and the signal value of the sparse components change in a correlated manner over time. Ref. [23] proposes an online algorithm for recovering a time sequence of sparse vectors and a time sequence of dense vectors from their sum, where the dense vectors lie in a slowly changing low-dimensional subspace. Ref. [24] studies the online RPCA problem and proposes a practical algorithm with theoretical guarantee. RPCA with the weak temporal correlations was studied in [25], and the theoretical analysis only holds when the temporal correlations of the data points are relatively weak. These papers consider the temporal correlations in either the underlying subspace or the sparse part, but the problem of missing data is not addressed.

Some methods have been developed to track the low-dimensional structure with incomplete data, under the terminology of subspace tracking with missing data or Dynamic MC. Examples include Grassmannian Rank-One Update Subspace Estimation (GROUSE) [26], Parallel Estimation and Tracking by Recursive Least Squares (PETRELS) [27] when the low-dimensional is represented by a linear subspace, as well as Multiscale Online Union of Subsets Estimation (MOUSSE) [28] when the low-dimensional structure is represented by a union of subspaces. Ref. [21] proposes a model of a sequence of dynamically correlated matrices through slow-varying subspaces. Each matrix is assumed to be low-rank and contain erasures, but the sparse errors are not modeled in [21]. A detailed review of Dynamic RPCA and Dynamic Matrix Completion is provided in [29]. Another line of work is the tensor robust completion [30, 31, 32], which can also be applied on the missing and corrupted data scenario. However, robust tensor completion is far from being well-studied compared with robust matrix completion. This is because it is computationally intractable to determine the rank or its best convex approximation [33].

This paper studies the problems of recovering low-rank matrices when the measurements are partially corrupted and partially lost. It models the dynamic correlations of low-rank matrices through time-varying subspaces. Our model generalizes from the one in [21] by additionally modeling sparse errors in the measurements. To the best of our knowledge, this is the first analytical study of robust matrix completion of temporally correlated matrices with partially corrupted measurements. We formulate the problem as a nonconvex optimization problem and theoretically characterize the recovery error. Our analytical bound characterizes the recovery enhancement of exploiting the temporal correlations in a sequence of low-rank matrices over recovering each low-rank matrix individually. We propose a fast iterative method to solve the nonconvex problem approximately and show that the algorithm can always converge to a critical point.

The rest of the paper is organized as follows. Section 2 introduces our model and problem formulation. Section 3 provides the theoretical bound of the matrix recovery error. Section 4 proposes an iterative algorithm and shows its convergence result. Section 5 records the numerical results. Section 6 concludes the paper.

2 Model and Proposed Method

Our problem formulation is built upon and generalizes the problem setup in [21]. Let $\mathbf{L}^t \in \mathbb{R}^{n_1 \times n_2}$ denote the actual data at time $t$, and let $\tilde{C}^t \in \mathbb{R}^{n_1 \times n_2}$ denote the sparse additive errors in the measurements at time $t$. The temporal correlations are modeled as a sequence of low-rank matrices with correlated low-dimensional subspaces. Specifically, let $\tilde{X}^t \in \mathbb{R}^{n_1 \times n_2}$ denote the measurements at time $t$,

$$\tilde{X}^t = \tilde{L}^t + \tilde{C}^t = \tilde{U}^t (\tilde{V}^t)^T + \tilde{C}^t,$$

where $\tilde{L}^t$ has rank at most $r$, $\tilde{U}^t \in \mathbb{R}^{n_1 \times r}$, $\tilde{V}^t \in \mathbb{R}^{n_2 \times r}$, and $\tilde{C}^t$ has at most $s$ nonzero entries. We further assume $\|\tilde{L}^t\|_\infty \leq \alpha$ and $\|\tilde{C}^t\|_\infty \leq \alpha$ for some constant $\alpha$. The temporal correlations can be modeled by correlations in $\tilde{U}^t$'s and $\tilde{V}^t$'s. Without loss of generality, we assume $\tilde{V}^t$ changes over time, while $\tilde{U}^t$ is fixed to be $\tilde{U}$ such that we have

$$\tilde{L}^t = \tilde{U} (\tilde{V}^t)^T.$$

For the sake of simplicity in our analysis, we consider a simple model on $\tilde{V}^t$ as follows.

$$\tilde{V}^t = \tilde{V}^{t-1} + \epsilon^t, \quad t = 2, \ldots, d,$$

where $\epsilon^t$ represents the perturbation noise in $\tilde{V}^t$. Note that the temporal information (1) is only used in the theoretical analysis, while our proposed method does not need the information of $\epsilon^t$. Moreover, the factorization of a low-rank matrix is not unique. Our result holds as long as there exists one factorization such that (1) holds. $Z^t \in \mathbb{R}^{n_1 \times n_2}$ denotes the measurement noise. $\Omega^t$ is the set of observed entries in $\tilde{X}^t$ with $|\Omega^t| = m^t$. The partial observed measurements can be presented by

$$y^t = \mathcal{P}_{\Omega^t}(\tilde{X}^t + Z^t),$$

where $\mathcal{P}_{\Omega^t} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{m^t}$ is a linear operator. In this paper, we consider the case that operator $\mathcal{P}_{\Omega^t}$ is a uniform sampling ensemble (with replacement), defined as follows.
Definition 1  A linear operator \( \mathcal{P}_{\Omega} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{m_1} \) is a uniform sampling ensemble (with replacement), if we can express each entry of \( \mathcal{P}_{\Omega}(X) \) as \( \mathcal{P}_{\Omega}(X)_{i} = \langle P_{i}, X \rangle \), where all matrices \( P_{i} \) are i.i.d. uniformly distributed on the set

\[
X = \{ e_j(n_1)e_k^T(n_2), 1 \leq j \leq n_1, 1 \leq k \leq n_2 \},
\]

and \( e_j(n) \) are the canonical basis vectors in \( \mathbb{R}^n \).

The dynamic robust matrix completion problem is stated as follows. Given partially observed and corrupted measurements \( \{ y^t \} \) for \( t = 1, \ldots, d \), can we recover the actual data \( \bar{L}^d \)? Note that [21] does not consider \( \bar{C}^d \), and our problem reduces to the one in [21] when \( \bar{C}^t = 0 \) for \( t = 1, \ldots, d \).

We estimate the unknown \( (\bar{L}^d, \bar{C}^d) \) using a nonconvex optimization approach. The objective function is given by

\[
F(X) = \frac{1}{2} \sum_{t=1}^{d} \omega_t \| \mathcal{P}_{\Omega}(X) - y^t \|_2^2,
\]

where \( \omega_t \)'s are predetermined non-negative weights, and \( \sum_{t=1}^{d} \omega_t = 1 \). We estimate \( (\bar{L}^d, \bar{C}^d) \) by \( (\hat{L}, \hat{C}) \), where

\[
(\hat{L}, \hat{C}) = \arg \min_{(L,C)} \frac{1}{2} \sum_{t=1}^{d} \omega_t \| \mathcal{P}_{\Omega}(L + C) - y^t \|_2^2
\]

\[\text{s.t. } L + C \in \mathbb{C}(r, s, \alpha)\]

and the feasible set is defined as

\[
\mathbb{C}(r, s, \alpha) := \{ X \in \mathbb{R}^{n_1 \times n_2} : X = L + C, \| L \|_{\infty} \leq \alpha, \| C \|_{\infty} \leq \alpha, \text{rank}(L) \leq r, \sum_{ij} 1[C_{ij} \neq 0] \leq s \}.
\]

We consider the online setting where the objective is to estimate the data at the current time step with \( t = d \). The past observations with \( t < d \) are used to enhance the estimation of \( \bar{L}^d \) because \( \bar{L}^1, \ldots, \bar{L}^d \) are correlated. One alternative approach is to estimate matrices in all time instants by replacing the variable \( L \) with \( L^1, L^2, \ldots, L^d \). However, this alternative is more computationally expensive than (4) because the number of variables increases by a factor of \( d \). Similarly, in our proposed approach (4), we only estimate \( \bar{C}^d \) rather than all \( \bar{C}^1, \ldots, \bar{C}^d \) to reduce the computational complexity. We consider the error of estimating historical data with \( t < d \) using \( \mathcal{P}_{\Omega}(L + C) - y^t \) rather than \( \mathcal{P}_{\Omega}(L) - y^t \), because the former provides a better estimation numerically when the errors have temporal correlations, which often happen when sensors (e.g., cameras) are partially masked or damaged. In our theoretical analysis in Section III, we do not assume any correlations among sparse error matrices. Moreover, our approach does not need the information of \( \epsilon_t \) and applied to the cases that the temporal correlations in the low-rank matrices are unknown.

3 Results: Theoretical

Our theoretical bound is built upon and generalizes the result in [21]. The major difference from [21] is the extra effort to handle the nonzero \( \bar{C}^t \)'s.

3.1 Assumptions of Our Approach

We first define some coefficients.

\[
n_{\min} = \max(n_1, n_2) \] and \( n_{\max} = \min(n_1, n_2) \).

Definition 2  A rank-\( r \) matrix \( X \in \mathbb{R}^{n_1 \times n_2} \) with SVD \( X = U\Sigma V^T \) is incoherent with parameter \( \mu \) if

\[
\| U_{j} \|_2 \leq \sqrt{r/n_1} \text{ for any } i = 1, \ldots, n_1,
\]

and

\[
\| V_{j} \|_2 \leq \sqrt{r/n_2} \text{ for any } j = 1, \ldots, n_2.
\]

For the simplicity of the theoretical analysis, we assume the number of observations \( |\Omega^t| \)'s are the same for \( t = 1, \ldots, d \), and define \( n_0 = |\Omega^t| \). Note that the locations of the observations in \( \Omega^t \)'s are different for different \( t \). Let

\[
p = n_0/(n_1n_2)
\]

denote the fraction of observed entries.

Suppose that we are given measurements as in (2) where all \( \mathcal{P}_{\Omega^t} \)'s are uniform sampling ensembles. Assume that

- (a) \( \bar{L}^t \) evolves according to (1), has rank at most \( r \), and is incoherent with parameter \( \mu_0 \).
- (b) \( \bar{C}^t \) has at most \( s \) nonzero entries. \( \| \bar{L}^t \|_{\infty} \leq \alpha \) and \( \| \bar{C}^t \|_{\infty} \leq \alpha \).
- (c) Measurement noise \( Z^t \) is i.i.d. \( \mathcal{N}(0, \sigma_2^2) \) for \( 1 \leq t \leq d \), and the perturbation noise \( \epsilon^t \) is i.i.d. \( \mathcal{N}(0, \sigma_1^2) \) for \( 2 \leq t \leq d \).

3.2 Theoretical Recovery Guarantee of Our Approach

We first define some coefficients that will be used in Theorem 1.

\[
\sigma_{\max} = \max_{t=1, \ldots, d} \omega_t^2 \left( \frac{\mu_0^2 r}{n_1} (d-t) \sigma_2^2 + \sigma_1^2 \right),
\]

\[
\sigma_{\sum} = \sum_{t=1}^{d} \omega_t^2 (\sigma_1^2 + (d-t) \sigma_2^2),
\]
If (Simplified) Theorem 1 lower bound.

Theorem 1 establishes the recovery error for the dynamic robust matrix completion problem. With some tedious calculations, one can check that

\[ Q_1 = \alpha^2 \max \left( \sqrt{c_1 \frac{\log(n_1 + n_2)}{m_0}}, \frac{c_2 \log(n_1 + n_2)}{n_1} \right), \]

\[ Q_2 = \frac{\alpha^2}{m_0} \left( c_3 n_{\max} \log(n_1 + n_2) \sum_{t=1}^d \omega_t^2 + c_4 n_1 + c_5 \sqrt{\frac{2skr}{n_{\min}}} \right) + c_6 \sqrt{\frac{\log(n_1 + n_2)}{m_0}} \]

\[ + \frac{c_7 \sum_{t=1}^d \omega_t^2 m_0}{\alpha m_0} + \frac{s \alpha}{n_1 n_2} + \frac{\sqrt{2 \alpha m_0 \log(n_1 + n_2)}}{m_0}, \]

and

\[ \kappa = c_6 n_{\max} \sigma_{\text{sum}} + c_7 \alpha^2 \rho^2 s + c_8 \alpha^2 \left( n_{\max} + \frac{2 \sqrt{2} s}{n_{\min}} \right), \]

where \( c_1, c_2, c_3, c_4, c_5, c_6, c_7, \) and \( c_8 \) are constants.

We provide below our simplified theorem with order-wise bound. The complete version can be viewed in Appendix B. For brevity, we assume \( n_1 \) and \( n_2 \) are in the same order \( O(n) \) and we choose the weights \( \omega_t = \frac{1}{t} \) for \( t = 1, ..., d \) to simplify the order of the observation lower bound.

(Simplified) Theorem 1 If \( m_0 \geq O \left( \frac{n(\log(n))^2 \log(d)}{d} \right) \), then the estimator \((\hat{L}, \hat{C})\) from (4) satisfies

\[ \frac{1}{n_1 n_2} \left( \| \hat{L} - L^d \|^2_F + \| \hat{C} - C^d \|^2_F \right) \leq \max(Q_1, Q_2, O \left( \frac{s}{n_2^2} \right)) \]

and

\[ \frac{1}{n_1 n_2} \| \hat{L} - L^d \|^2_F \leq \max(Q_1, Q_2, O \left( \frac{s}{n_2^2} \right)) \]

with probability at least \( 1 - \frac{1}{2n^2} - 7d n e^{-n} \).

Analysis of the additive errors. If \( s = O(n) \), i.e., the number of corrupted measurements per row is bounded, from the Theorem, we have if \( m_0 \geq O(\min(n(\log(n))^2, n_{\max}) \cdot q_6 \cdot \log(d)) \),

\[ \frac{1}{n_1 n_2} \left( \| \hat{L} - L^d \|^2_F + \| \hat{C} - C^d \|^2_F \right) \leq O \left( \frac{n \log n}{m_0} \right) \]

and

\[ \| \hat{L} - L^d \|^2_F \leq O \left( \frac{n \log n}{m_0} \right) \]

hold with probability at least \( 1 - \frac{1}{2n^2} - 7d n e^{-n} \). Note that (5) and (6) diminish to zero when \( n \) increases.

When the measurements do not contain corruptions, i.e. \( \hat{C}^d \)'s are all zeros, [21] showed that if \( m_0 \geq O \left( \frac{n(\log(n))^2}{n^2}, \frac{n \log n}{m_0} \right) \), with probability at least \( 1 - \frac{5}{2n^2} - 5d n e^{-n} \), we have

\[ \| \hat{L} - L^d \|^2_F \leq O \left( \frac{n \log n}{m_0} \right) \]

when \( r = O(1) \). Here

\[ Q_1 = \max(O \left( \frac{\log n}{m_0} \right), O \left( \frac{n \log n}{n} \right)) \]

and \( Q_2 \geq Q_1 \). Note that the probability of success in Theorem 1 is in the same order as that in [21] as well.

The role of \( m_0 \). One can check that \( Q_1 \) and \( Q_2 \) are in the order of \( O \left( \frac{1}{m_0} \right) \), which means the recovery error is reduced with the increasing number of observations in the measurements.

The role of \( d \). If we choose the weight \( \omega_t = \frac{1}{t} \) for \( t = 1, ..., d \), one can check that the required number of observations of each matrix is reduced by a factor of \( O \left( \frac{\log d}{d} \right) \) when \( d \) increases. One can also check that two terms in \( Q_2 \) decrease with the increasing of \( d \), which means the recover error reduces by exploiting the temporal dynamic in the low rank matrices.

The optimal weights. Note that if the number of corrupted measurements per row is bounded, the upper bound in Theorem 1 is dominated by the following term

\[ \alpha^2 c_3 \sum_{t=1}^d \omega_t^2 + c_6 \sigma_{\text{sum}} = \sum_{t=1}^d \omega_t^2 (\alpha^2 c_3 + c_6 \sigma_t^2) + c_6 (d-t) \sigma_t^2. \]

By minimizing this term with respect to \( \omega_t \) for \( t = 1, ..., d \) and subject to \( \sum_t \omega_t = 1, \omega_t \geq 0 \), we can obtain
the optimal selection of $\omega_t$. The analytical solution $\omega_t^\star$ is shown as follows

$$
\omega_t^\star = \left(\sum_{i=1}^{d} \frac{\alpha^2 c_3 + c_6 (d-t) \sigma_2^2}{\alpha^2 c_3 + c_6 (d-t) \sigma_2^2} \right)^{-1}.
$$

Note that $\sigma_2$ affects the perturbation noises. $V^t$’s are the same when $\sigma_2 = 0$. In this case, $L^t$’s are also the same. Therefore, $\sigma_2 = 0$ suggests that we choose the same weights for all time instants, i.e., $\omega_t = 1/d$. In contrast, $\sigma_2 > 0$ suggests that we shall choose larger $\omega_t$ when $t$ increases. In addition, the similarities among $L^t$’s become smaller when $\sigma_2$ increases, which suggests that we should choose larger $\omega_t$ for larger $t$. Please refer to Appendix for the proof of Theorem 1.

4 Algorithm for Dynamic Matrix Completion with Erroneous Measurements

Since (4) is nonconvex, one option is to solve its convex relaxation instead by replacing the rank constraint with the nuclear norm constraint and the sparsity constraint with the $\ell_1$-norm constraint. The relaxed problem is shown as follows:

$$
\min_{L,C} \frac{1}{2} \sum_{t=1}^{d} \frac{1}{\omega_t |\Omega^t|} \|P_{\Omega^t}(L + C) - y^t \|_2^2 + \lambda_1 \|L\|_* + \lambda_2 \|C\|_1
$$

s.t. $\|L\|_\infty \leq \alpha_1$ and $\|C\|_\infty \leq \alpha_2$, \hspace{1cm} (8)

where $\|L\|_*$ is the sum of all singular values of $L$, and $\|C\|_1 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |C_{ij}|$. Note that (8) reduces to the Robust Matrix Completion (RMC) problem [34] when $d = 1$. We also remark that setting the same $\alpha$ for $\|L\|_\infty$ and $\|C\|_\infty$ can simplify the presentation of the upper bound in the theoretical analysis. In the algorithm, one can set $\|L\|_\infty \leq \alpha_1$, $\|C\|_\infty \leq \alpha_2$ with different constants $\alpha_1$ and $\alpha_2$.

Since solving the convex relaxation problem (8) is not effective, we propose an algorithm to solve the nonconvex problem (4) directly. The low-rank matrix $L$ is factorized into two matrices $U \in \mathbb{R}^{n_1 \times r}$ and $V \in \mathbb{R}^{n_2 \times r}$ such that $L = UV^T$. We move the constraint $L = UV^T$ into a penalty function $\frac{\rho}{2} \|L - UV^T\|_F^2$ in the objective function. The revised problem of (4) is written as follows:

$$(\hat{U}, \hat{V}, \hat{L}, \hat{C}) = \arg \min_{(U,V,L,C)} \hat{F}(L,C) + \frac{\rho}{2} \|L - UV^T\|_F^2,$$

s.t. $L + C \in \mathbb{C}(r,s,\alpha_1, \alpha_2)$, \hspace{1cm} (9)

where $\mathbb{C}(r,s,\alpha_1, \alpha_2)$ is a replacement of $\mathbb{C}(r,s,\alpha)$ in (4) by replacing $\|L\|_\infty \leq \alpha$ and $\|C\|_\infty \leq \alpha$ with $\|L\|_\infty \leq \alpha_1$ and $\|C\|_\infty \leq \alpha_2$. We then have

$$
\hat{F}(L,C) = \frac{1}{2} \sum_{t=1}^{d} \omega_t \|P_{\Omega^t}(L + C) - y^t \|_2^2.
$$

We define $H(U,V,L,C) = \hat{F}(L,C) + \frac{\rho}{2} \|L - UV^T\|_F^2$, as our revised objective function. The solution of (9) is the same as that of (4) when $\rho$ approaches infinity. However, if $\rho$ is too large, the step size is very small, and that affects the convergence rate. One practical solution is to dynamically change $\rho$ from small to large with a fixed multiplier $\gamma > 1$ and use the result from the previous step as the initialization in the current step [35]. We remark that the decomposition $L = UV^T$ is a natural way to replace the rank inequality constraint [36, 20]. With only the low-rank constraint, methods like alternating least squares and gradient descent on Grassmann manifold can be implemented to solve the problem [36, 20]. To address the infinity norm constraint on $L$, we still keep the variable $L$ and leverage the proximal alternating gradient descent method, which provides straightforward solutions to the infinity norm constraint.

In each iteration, we apply proximal alternating gradient descent to update the estimations of $U$, $V$, $L$ and $C$.

$$
U_{i+1} \in \text{prox}_{\rho}(U_i - \tau \nabla \hat{F}(U_i, V_i, L_i, C_i)),
$$

$$
V_{i+1} \in \text{prox}_{\rho}(V_i - \tau \nabla \hat{F}(U_i, V_i, L_i, C_i)),
$$

$$
L_{i+1} \in \text{prox}_{\rho}^{\rho}(L_i - \tau \nabla \hat{F}(U_{i+1}, V_{i+1}, L_i, C_i)), \hspace{1cm} (10)
$$

and

$$
C_{i+1} \in \text{prox}_{\rho}^{\rho}(C_i - \tau \nabla \hat{F}(U_{i+1}, V_{i+1}, L_{i+1}, C_i)), \hspace{1cm} (11)
$$

where the proximal map associated to $\sigma$ is defined as:

$$
\text{prox}_{\rho}^{\sigma}(x) = \arg \min_{u} \{\sigma(u) + \frac{1}{2}\|u - x\|_F^2\},
$$

and functions corresponding to $\sigma$ in (10) and (11) are indicator functions as follows.

$$
J(L) = \begin{cases} +\infty & \text{if } \|L\|_\infty > \alpha_1, \\ 0 & \text{otherwise}. \end{cases} \hspace{1cm} (12)
$$
Similarly, (13) can be obtained by setting
\[ Q(C) = \begin{cases} +\infty & \text{if } \|C\|_\infty > \alpha_2, \\ 0 & \text{otherwise}. \end{cases} \]
\[ K(C) = \begin{cases} +\infty, & \text{if } \sum_{i,j} 1_{\{C_{ij} \neq 0\}} > s, \\ 0, & \text{otherwise}. \end{cases} \]

The details of our algorithm are summarized in Algorithm 1. (12) is obtained by setting
\[ (L_{i+1})_{kj} = \begin{cases} \alpha_1, & \text{if } (L_{i+1})_{kj} > \alpha_1, \\ -\alpha_1, & \text{if } (L_{i+1})_{kj} < -\alpha_1, \end{cases} \]

where \((L_{i+1})_{kj}\) denotes the entry in row \(k\) and column \(j\) of \(L_{i+1}\). This corresponds to lines 5 - 8 of Algorithm 1. Similarly, (13) can be obtained by setting
\[ (C_{i+1})_{kj} = \begin{cases} \alpha_2, & \text{if } (C_{i+1})_{kj} > \alpha_2, \\ -\alpha_2, & \text{if } (C_{i+1})_{kj} < -\alpha_2. \end{cases} \]

(14) is equivalent to only keeping \(s\) entries of \(C\) with the largest absolute values. Operations of \(C\) correspond to lines 6-10 of Algorithm 1.

The variables \(L, U, V, C\) and \(L, C\) are initialized as follows: We set \((L_0)_{ij} = 0\) if \((i, j) \notin \Omega^d\), and \((L_0)_{ij} = X_{ij}\) if \((i, j) \in \Omega^d\). Then we perform truncated singular value decomposition with the \(r\) largest singular values on \(L_0\). Let \(U_0, \Sigma_{L_0}, V_0^T\) denote the rank-\(r\) approximation of \(L_0\). Then we initialize \(V_0 = \sum_{L_0}^{1/2} V_0\) and \(U_0 = U_0 \sum_{L_0}^{1/2}\). \(C_0\) is set to be a zero matrix.

The step sizes \(\tau_U, \tau_V, \tau_L,\) and \(\tau_C\) in Algorithm 1 are selected by calculating Lipschitz constants in each iteration. More specifically,
\[ \tau_{U,V,L,C} = \frac{1}{\beta_i^{U,V,L,C}}, \]
where \(\beta\) is a constant larger or equal to one and \(\beta_i^{U,V,L,C}\) is the Lipschitz constant for variables \(U, V, L,\) and \(C\) computed in each iteration:\[^1\]
\[ l_i^U = \rho\|V_i^T V_i\|_F, \quad l_i^V = \rho\|U_i^{T+1} U_{i+1}\|_F, \]
\[ l_i^L = \rho + 1, \quad \text{and } l_i^C = 1. \]

Similar to the assumptions made on the Nonnegative Matrix Factorization in Ref. [38] and the assumptions on Nonnegative Tensor Factorization in Ref. [39], we assume that the sequences \(\{U_i\}, \{V_i\}\) are bounded and \(\inf_{i \in \mathbb{N}} \{\|U_i\|_F, \|V_i\|_F\} > 0\). As suggested by Ref. [38], the second boundedness assumption can be easily dropped by replacing \(l_i^U, l_i^V\) with \(\max\{\rho\|V_i^T V_i\|_F, \nu\}, \max\{\rho\|U_i^{T+1} U_i\|_F, \nu\}\), where \(\nu\) is a small positive constant. We fixed a constant \(\delta\), and the algorithm terminated if
\[ \|U_{i+1} - U_i\|_F + \|V_{i+1} - V_i\|_F + \|L_{i+1} - L_i\|_F + \|C_{i+1} - C_i\|_F < \delta. \]

**Algorithm 1** Algorithm for dynamic robust matrix completion

**Input:** Partially observed matrix \(Y^t \in \mathbb{R}^{n_1 \times n_2}\) for \(t = 1, \ldots, d\), initialization matrices \(L_0 \in \mathbb{R}^{n_1 \times n_2}, U_0 \in \mathbb{R}^{n_1 \times \tau}, V_0 \in \mathbb{R}^{n_2 \times \tau}\) and zero matrix \(C_0 \in \mathbb{R}^{n_1 \times n_2},\) parameters \(r, \alpha_1, \alpha_2, \beta, \rho, \gamma, \delta,\) and \(s\), weights \(\omega_i\) for \(i = 1, \ldots, d\).

1. for \(i = 0, 1, 2, \ldots \), until stopping criteria is met do
2. \(U_{i+1} = U_i - \tau_U \nabla_U H(U_i, V_i, L_i, C_i).\)
3. \(V_{i+1} = V_i - \tau_V \nabla_V H(U_{i+1}, V_i, L_i, C_i).\)
4. \(L_{i+1} = L_i - \tau_L \nabla_L H(U_{i+1}, V_{i+1}, L_i, C_i).\)
5. if \((L_{i+1})_{kj} > \alpha_1\) then
6. Set \((L_{i+1})_{kj} = \alpha_1.\)
7. else if \((L_{i+1})_{kj} < -\alpha_1\) then
8. Set \((L_{i+1})_{kj} = -\alpha_1.\)
9. end if
10. \(C_{i+1} = C_i - \tau_C \nabla_C H(U_{i+1}, V_{i+1}, L_{i+1}, C_i).\)
11. if \((C_{i+1})_{kj} > \alpha_2\) then
12. Set \((C_{i+1})_{kj} = \alpha_2.\)
13. else if \((C_{i+1})_{kj} < -\alpha_2\) then
14. Set \((C_{i+1})_{kj} = -\alpha_2.\)
15. end if
16. if \(\sum_{kj} 1_{\{(C_{i+1})_{kj} \neq 0\}} > s\) then
17. \(C_{i+1}\) only keeps \(s\) entries with the largest absolute values. Other nonzero entries are set to be zero.
18. end if
19. \(\rho = \gamma \rho.\)
20. end for
21. Return: \(L = UV^T\) and \(C.\)

We next show that every sequence generated by our algorithm will converge to a critical point of (9). Our algorithm is a special case of Proximal Alternating Linearized Minimization (PALM) algorithms. The convergence of PALM algorithms have been proved in [38]. One can check that \(\nabla_U H, \nabla_V H, \nabla_L H\) and \(\nabla_C H\) are Lipschitz continuous, and \(H(U, V, L, C) + J(L) + Q(C) + K(C)\) satisfies the Kurdyka-Lojasiewicz (KL) property based on Proposition 3 in Ref. [38]. Then our algorithm converges to a critical point of (9) from every initial point. See Appendix G for more details. According to Theorem 2.9 in Ref. [39], if the initial point is sufficiently close to the optimal point, our algorithm converges at least sublinearly. We also remark that the complexity of our algorithm in each iteration is in the order of \(O(n_1 n_2 \tau^2).\)

**5 Results: Numerical Experiments**

We test the performance of Algorithm 1 on both synthetic dataset and the actual video datasets. The re-
covery performance is measured by the relative recovery error $\| \hat{L}^d - \hat{L}^d \|_F / \| L^d \|_F$, where $\hat{L}$ represents the actual data, and $\hat{L}$ represents the recovered data. The corruption rate denotes the fraction of nonzero entries in $C^d$. In the simulations on synthetic datasets, we generate the nonzero entries of the sparse matrix $C^d$ i.i.d. from the Gaussian distribution $\mathcal{N}(0, 2.5^2)$. In the simulations on the actual video dataset, we generate the nonzero entries of the sparse matrix $C^d$ i.i.d. from the Gaussian distribution $\mathcal{N}(0, 1)$. We consider two types of corruptions: (1) uniformly chosen locations for all corrupted entries and (2) uniformly chosen locations for corrupted blocks, and each block is a square shape with corruptions. The corruptions follow type (1) if not otherwise specified. The average erasure rate is the percentage of the missing entries. We set $\beta = 1$, $\delta = 10^{-6}$, $\gamma = 1.01$ and set the starting point of $\rho$ to be 0.1. The simulations are run in MATLAB on a computer with 3.4 GHz Intel Core i7. All results are averaged over 100 runs.

5.1 Performance on synthetic data

5.1.1 Experimental set-up

We first compare Algorithm 1 with the convex relaxation (8) and solve (8) by CVX [40]. We set $n_1 = 50$, $n_2 = 50$, $r = 5$ and construct $L^d \in \mathbb{R}^{n_1 \times n_2}$ as $L^d = U(V^1)^T$ where $U \in \mathbb{R}^{n_1 \times r}$ and $V^t \in \mathbb{R}^{n_2 \times r}$. $U$ and $V^1$ are matrices with i.i.d. entries drawn from standard Gaussian distribution. For all $t \geq 2$, $V^t = V^{t-1} + \epsilon^t$, where matrix $\epsilon^t$ is drawn from Gaussian distribution $\mathcal{N}(0, \sigma^2_2)$. Noise matrix $Z^t$ is drawn from Gaussian distribution $\mathcal{N}(0, \sigma^2_t)$. The weights $\{\omega_t\}$ in Algorithm 1 are set to be $(\frac{1}{2}, \ldots, \frac{1}{2})$. We first set $s = 500$, $\sigma_1 = 0.01$, and $\sigma_2 = 0.03$. One can check that the variance of each entry of $L$ is equal to 5 when $d = 1$, and the variance increases to 5.018 when $d = 5$. Then we choose $\alpha_1$ to be 10, which is two times of the variance when $d = 1$. According to Monte Carlo simulations, the probability of $\| \hat{L} \|_\infty$ being less than 10 in our setup is larger than 99% when $d \leq 5$. From the property of the Gaussian distribution, we choose $\alpha_2 = 7.5$ since $P(\| C^d \|_\infty \leq 7.5) > 99.7\%$. We set $\lambda_1 = 0.001$, $\lambda_2 = 0.00015$ in (8).

5.1.2 Experimental results

Fig. 1 shows the recovery performance of the convex relaxation (8) and Algorithm 1 with different $d$. We can see that Algorithm 1 performs better than the convex relaxation under the same $d$, and the performance of Algorithm 1 improves when $d$ increases. When $d = 5$, the average running time of Algorithm 1 is 0.35 seconds, while the convex relaxation using CVX needs 300 seconds. Even though CVX is not designed for fast speed, the comparison still shows that the improvement of our algorithm in the speed.

We study the impact of $\rho$ on the number of iterations. If we set $\rho = 5$ and $\gamma = 1$, i.e., $\rho$ does not change, the average number of iterations is 8000 when the average erasure rate equals 0.8. As a comparison, the average number of iterations is 1300 if $\rho$ starts with 0.1 and $\gamma = 1.01$. We also test other fixed values of $\rho$ from 5 to 3000. The smallest number of iterations is still seven times larger than the number of iterations by updating $\rho$ in our setup. Thus, dynamically adjusting $\rho$ indeed speeds up the convergence.

We then set $d = 3$ and keep the other simulation setup the same. Fig. 2 shows the comparison of Algorithm 1 with the locally weighted matrix smoothing (LOWEMS) proposed in [21], and the convex relaxation (8). LOWEMS alternatively minimizes $U$ and $V$ in the objective function of equation (13) in [21] over $U$ and $V$. This method does not consider the corruptions, and the performance degrades when a significant fraction of measurements are corrupted. We compare the three methods under the cases where the corruption rates are 20% and 30%. We can see that Algorithm 1 performs better than the other two methods.

We also vary the perturbation noise $\epsilon$ by increasing $\sigma_2$ from 0.05 to 0.5. We fix $d = 3$, $s = 500$,
and the average erasure rate as 0.5 in this experiment. We set \((\omega_1, \omega_2, \omega_3)\) to four different groups of values: (a) \((0, 0, 1)\) (b) \((0.33, 0.33, 0.33)\) (c) \((0.1, 0.3, 0.6)\) (d) \((0.05, 0.15, 0.8)\). Here \((0, 0, 1)\) is the baseline since no history is considered in this case. The relative recovery error against \(\sigma_2\) is shown in Fig. 3. One can find that the relative recovery error in case (b) increases when the perturbation noise increases, and exceeds the baseline after \(\sigma_2 = 0.25\). This means that placing equal weights on measurements at all time degrades the performance if the perturbation is significant. One can improve the performance by setting a large weight on the current incident and a small weight on the history. The recovery error is the smallest in case (d). We will leave the selection of optimal weights as future work.

5.2 Performance on actual video dataset 1

5.2.1 Experimental set-up

In this section, we test our method on a noisy video dataset, in which a man slowly moves his head. The first row of Fig. 4 shows three different frames of the video. The second row shows the frames with 15% corruptions and 15% missing entries, and the third row shows the frames with 20% block corruptions and 15% missing entries. The video dataset with only a few moving objects usually changes slowly. We remark that the perturbation model on \(V\) and \(L\) simplifies the analyses but is not required to implement our recovery method. Our method applies to datasets with smooth changes, and this testing dataset satisfies this property.

We choose 180 frames, and reduce each frame to a vector in \(\mathbb{R}^{270}\). Then each \(\hat{L}^t(t = 1, 2, 3)\) contains 60 consecutive frames. We set \((\omega_1, \ldots, \omega_d)\) to \((1), (0.2, 0.8), (0.05, 0.15, 0.8)\) when \(d = 1, 2, 3\), respectively. We set \(\alpha_1 = 1\) since all the values are normalized, and the largest pixel value is transformed to 1. We set \(\alpha_2 = 3\) because \(P(\|\hat{C}\|_{\infty} \leq 3) > 99.7\%\). \(r\) is set to be 5.

5.2.2 Experimental results

Under different corruption rates (5%, 10%, 15%, 20%), we run our algorithm with different \(d\). Fig. 5 shows that \(d > 1\) outperforms \(d = 1\). Moreover, performances of \(d = 3\) and \(d = 2\) are close when the erasure rate is small, while \(d = 3\) is the best when the average erasure rate increases. We also test our algorithm on frames with 20% corruptions in blocks. As shown in Fig. 6, increasing \(d\) helps decreasing the recovery error when the average erasure rate increases. The recovery error curves are similar to the curves (20% corruption rate) in Fig. 5, indicating that our algorithm is not sensitive to corruption types.

We then set \(d = 2\) and keep the other simulation setup the same. Fig. 7 shows the recovery performance of our method with different corruption rates. Performance of our method improves when the corruption rate decreases.

We also show the box-plot-diagram of relative recovery error when the average erase rate changes in Fig. 8. The tops and bottoms of each "box" are the 25th and 75th percentiles of the samples, respectively. The results show that our method has small variances.
5.3 Performance on actual video dataset 2

5.3.1 Experimental set-up

We test our method on a public video dataset, in which an anchor is broadcasting news (The dataset can be downloaded from this website [41]). The dataset contains 300 frames with each frame size 720 × 486. We downsize the video into a 3168 × 300 matrix. Then each \( \bar{L}^t \) \((t = 1, 2, 3) \) contains 100 consecutive frames. \( r \) is set to be 12. The corrupted entries are set to be 20% of the total number of entries in \( \bar{L}^d \). The other parameters are the same as in Section 5.2.

5.3.2 Experimental results

In Fig. 9, we show the performance of Algorithm 1 given \( d = 1, 2, 3 \). Again, one can see that the recovery error decreases when \( d \) increases. We compare Algorithm 1 to the Tensor Robust Principal Component Analysis (TRPCA) [32] which extends the matrix robust principal component analysis to the tensor case. We also compare to the convex relaxation (8) when \( d = 3 \). We can see that Algorithm 1 outperforms the convex relaxation but is slightly worse than TRPCA. We remark that TRPCA needs to estimate the whole dataset and thus is far more time-consuming than Algorithm 1. Fig. 10 shows the visualization of the recovered frames.

6 Conclusion and Discussion

This paper develops a data recovery method from partially observed and partially corrupted measurements with time-varying low dimensional structures. Exploiting the temporal correlations in the low dimensional structures, we show that the recovery error of our proposed method diminishes as the problem size increases, and the error decays in the same order as that of the state-of-the-art data recovery method with uncorrupted measurements. A proximal algorithm with convergence guarantee to stationary points is developed. Although the proposed theorem is built on the global optimum and there is a gap between the theorem and our algorithm, the algorithm is proved effective on both the synthetic dataset and the actual video datasets. One future work is to study the global convergence of the algorithm.

Appendix A: Notations

The notations in the proofs are summarized in Table 1.
Appendix C: Proof sketch of Theorem 1

The proof of Theorem 1 follows the same framework of the proof of Theorem 3.8 in [21]. With the additional sparse matrix, our proof is more involved than the one in [21].

Given the estimator \( \hat{X} = \hat{L} + \hat{C} \) from (4), we define the recovery error to be

\[
\Delta^d = \hat{X} - \bar{X} = \hat{L} + \hat{C} - \bar{L} - \bar{C}^d.
\]

The goal will be to provide bounds on \( \| \bar{L} - \hat{L}^d \|_F^2 + \| \bar{C} - \hat{C}^d \|_F^2 \) and \( \| \hat{L} - \hat{L}^d \|_F^2 \) respectively. In order to simplify the proof, we first define

\[
\Sigma_R = \sum_{t=1}^{d} \sum_{i=1}^{m_0} \omega_t \gamma_i^t P_i^t,
\]

where \( \gamma_i^t \) is the Rademacher variable, \( P_i^t \in \mathcal{X} \), and

\[
\mathcal{X} = \{ e_j(n_1)e_k^T(n_2) \mid 1 \leq j \leq n_1, 1 \leq k \leq n_2 \},
\]

where \( e_j(n) \) are the canonical basis vectors in \( \mathbb{R}^n \). Our theoretical analysis builds on the following inequality.

**Lemma 1** The estimator \( \hat{X} \) by (4) satisfies

\[
\sum_{t=1}^{d} \omega_t \| P_{1^t} (\Delta^d) \|_F^2 \leq 2\sqrt{2} \| \Sigma_R \|_F^2 + 2\sqrt{2} \| \Sigma_R \|_F^2
\]

\[
+ 8 \alpha \| \Sigma_R \|_F^2 + \frac{8m_0}{n_1 n_2} \alpha^2,
\]

with probability at least \( 1 - \frac{2}{n_1 n_2} \).

Lemma 1 is a counterpart of Proposition 3.1 in [21]. Since we take the sparse matrix into account, Lemma 1 is a probabilistic inequality while Proposition 3.1 in [21] is a deterministic inequality. Please refer to Appendix D for the proof of Lemma 1.

The remaining work is to lower bound the LHS of (15) and upper bound the RHS of (15). In Lemma 2, we will provide an upper bound of the term \( \| \sum_{t=1}^{d} \omega_t P_{1^t} [P_{1^t} (\hat{X}^d) - y^t] \|_2 \), which will eventually upper bound the RHS of (15). In Lemma 3, we will provide a lower bound of the term \( \sum_{t=1}^{d} \omega_t \| P_{1^t} (\hat{X}^d) \|_2^2 \), which contains \( \| \Delta^d \|_F^2 \) and \( \mathbb{E} (\| \Sigma_R \|_2^2) \). Lemma 6 in Appendix H provides an upper bound of \( \mathbb{E} (\| \Sigma_R \|_2^2) \). We will then get an upper bound of the term

\[
\| \Delta^d \|_F^2 = \| \hat{L} + \hat{C} - \hat{L}^d - \hat{C}^d \|_F^2.
\]
by combining Lemma 1, Lemma 2, Lemma 3, and Lemma 6. Assume we have
\[ \| \hat{L} + \hat{C} - \hat{L}^d - \hat{C}^d \|_F^2 \leq U. \]
Because \( \| C \|_\infty \leq \alpha \) and \( \sum_{ij} 1 \{C_{ij} \neq 0\} \leq s \), we then have
\[ \| \hat{C} - \hat{C}^d \|_F^2 \leq 4\alpha^2 s. \]
According to the triangle inequality, we have
\[ \| \hat{L} - \hat{L}^d \|_F \leq \| \hat{L} + \hat{C} - \hat{L}^d - \hat{C}^d \|_F + \| \hat{C} - \hat{C}^d \|_F. \]
We have
\[ \| \hat{L} - \hat{L}^d \|_F^2 \leq (\sqrt{U} + 2\sqrt{s})^2. \]
We then get the upper bounds on \( \| \hat{L} - \hat{L}^d \|_F^2, \| \hat{C} - \hat{C}^d \|_F^2 \), and \( \| \hat{L} - \hat{L}^d \|_F^2 \).

**Lemma 2** If \( m_0 \geq \frac{32 n_1 n_2 \log(n_1 + n_2)(\sqrt{2 \log(d(n_1 + n_2)n_1 n_2)} + 2\alpha^2)}{5n_{\max} \sum_{i=1}^d \omega_i^2 (\sigma_i^2 + (d-t)\sigma_i^2) + 2\alpha^2(\sqrt{2n_{\max} + 4s})}, \)
we have
\[ \left\| \sum_{i=1}^d \omega_i P_{\Omega^t}[X^d] - y' \right\|^2 \leq 16\alpha^2 d^2 s + m_0 \log(n_1 + n_2) \frac{\log(n_1 + n_2)}{n_1 n_2} \]
\[ + \frac{256n_{\max}}{\sum_{i=1}^d \omega_i^2 (\sigma_i^2 + (d-t)\sigma_i^2) + 192\alpha^2(\sqrt{2n_{\max} + 4s})} \]
hold with probability at least \( 1 - \frac{3}{n_1 + n_2} - 7dn_{\max} \cdot \exp(-n_{\min}) \).

The proof of Lemma 2 follows the same line as the part of proof of Theorem 3.8 after Lemma E.2. in [21]. With the additional sparse matrix, our proof is more involved than that in [21]. Please refer to Appendix F for the proof of Lemma 3.

**Lemma 3** Suppose all \( P_{\Omega^t} \)'s are fixed uniform sampling ensembles. For all \( X \in \mathcal{E}(r, s) \), we have
\[ \frac{p}{2} \| \Delta^d \|_F^2 \leq \frac{8\sqrt{rs} \log(n_1 + n_2)}{m_0} (\mathbb{E}(\| \Sigma_R \|_2))^2 + 8 \sqrt{rs} \mathbb{E}(\| \Sigma_R \|_2 + \frac{3456}{5} d n_1 64\alpha^2. \]

Combining Lemma 1 and (16) yields
\[ \frac{p}{2} \| \Delta^d \|_F^2 \leq \frac{8\sqrt{rs} \log(n_1 + n_2)}{m_0} (\mathbb{E}(\| \Sigma_R \|_2))^2 + 8 \sqrt{rs} \mathbb{E}(\| \Sigma_R \|_2 + \frac{3456}{5} d n_1 64\alpha^2. \]

With some tedious calculations, we then have
\[ \frac{p}{2} \| \Delta^d \|_F^2 \leq \frac{8\sqrt{rs} \log(n_1 + n_2)}{m_0} (\mathbb{E}(\| \Sigma_R \|_2))^2 + 8 \sqrt{rs} \mathbb{E}(\| \Sigma_R \|_2 + \frac{3456}{5} d n_1 64\alpha^2. \]

The proof of Lemma 3 follows the similar line as the proof of Lemma E.1. in [21]. With the additional sparse matrix, our proof is more involved than that in [21]. Please refer to Appendix F for the proof of Lemma 3. Note that
\[ \| \Delta^d \|_F^2 \leq \| \hat{L} - \hat{L}^d \|_F + \| \hat{C} - \hat{C}^d \|_F \]
\[ \leq \| \hat{L} \|_F + \| \hat{C} \|_F + \| \hat{L}^d \|_F + \| \hat{C}^d \|_F \leq 4\alpha. \]

We here consider the following two cases. Case I. \( \frac{ds}{s^2} \notin \mathcal{E}(2r, 2s) \). Following the definition of \( \mathcal{E}(2r, 2s) \), we have
\[ \| \Delta^d \|_F^2 \leq 16\alpha^2 n_1 n_2 m_0(\sqrt{\frac{\log(n_1 + n_2)}{25n_0 \log(6/5)}} - \frac{3456}{5} d n_1 64\alpha^2. \]

With probability at least \( 1 - \frac{6}{n_1 + n_2} \).
The second part of the inequality (Q2) in Theorem 1 follows from combining Lemma 2, Lemma 6 and (17).

Appendix D: Proof of Lemma 1

Proof Define

\[
x = \text{vec}(X) \in \mathbb{R}^{n_1n_2}, \quad l = \text{vec}(L) \in \mathbb{R}^{n_1n_2}
\]

\[
c = \text{vec}(C) \in \mathbb{R}^{n_1n_2}, \quad \hat{F}(x) = F(X).
\]

Since (3) is continuous in \(X\) and the set \(C(r, s, \alpha)\) is compact, (3) achieves a minimizer at some point \(\hat{X} \in C(r, s, \alpha)\). Since \(\hat{X}\) is a minimizer of the constrained problem, then for all matrices \(X \in C(r, s, \alpha)\) we have the following inequality

\[
\hat{F}(\hat{x}) - \hat{F}(x) \leq 0.
\]

By the second-order Taylor’s theorem, we expand \(\hat{F}(x)\) around \(\bar{x}^d = \text{vec}(\bar{X}^d)\),

\[
\hat{F}(x) = \hat{F}(\bar{x}^d) + \left\langle \nabla \hat{F}(\bar{x}^d), x - \bar{x}^d \right\rangle + \frac{1}{2} \left\langle \nabla^2 \hat{F}(\bar{x}^d)(x - \bar{x}^d), x - \bar{x}^d \right\rangle,
\]

where \(\hat{x} = a\bar{x}^d + (1 - a)x\) for some \(a \in [0, 1]\). Plugging (19) with \(x = \hat{x}\) into (18) we obtain

\[
\left\langle \nabla \hat{F}(\bar{x}^d), \hat{x} - \bar{x}^d \right\rangle + \frac{1}{2} \left\langle \nabla^2 \hat{F}(\bar{x}^d)(\hat{x} - \bar{x}^d), \hat{x} - \bar{x}^d \right\rangle \leq 0.
\]

(20)

With some tedious calculations, we have the following expression for the gradient of \(\hat{F}(x)\):

\[
\nabla \hat{F}(x) = \text{vec}(\sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* (\hat{P}_{\Omega_t} (X) - y')).
\]

(21)

Based on the above gradient and the definitions of \(\hat{P}_{\Omega_t}\) and \(P_{\Omega_t}\), we have, for any \(B \in \mathbb{R}^{n_1 \times n_2}\),

\[
\nabla^2 \hat{F}(x) b = \text{vec}\left(\sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* (\hat{P}_{\Omega_t} (B))\right).
\]

(22)

when \(b = \text{vec}(B)\). Note that the second term in (20) is

\[
\frac{1}{2} \left\langle \nabla^2 \hat{F}(\bar{x}^d)(\hat{x} - \bar{x}^d), \hat{x} - \bar{x}^d \right\rangle
\]

\[
= \frac{1}{2} \left\langle \sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* (\hat{P}_{\Omega_t} (\Delta)) (\Delta), \Delta \right\rangle
\]

\[
= \frac{1}{2} \sum_{t=1}^{d} \omega_t \left\langle \hat{P}_{\Omega_t} (\Delta), \hat{P}_{\Omega_t} (\Delta) \right\rangle = \frac{1}{2} \sum_{t=1}^{d} \omega_t \|\hat{P}_{\Omega_t} (\Delta)\|^2.
\]

(23)

Based on (21) and (22), the absolute value of first term in (20) can be bounded as

\[
|\left\langle \nabla \hat{F}(\bar{x}^d), \hat{x} - \bar{x}^d \right\rangle|
\]

\[
= |\left\langle \sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* [\hat{P}_{\Omega_t} (\bar{X}^d) - y'], \hat{L} - \hat{L}^d + \hat{C} - \hat{C}^d \right\rangle|
\]

\[
\leq \|\sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* [\hat{P}_{\Omega_t} (\bar{X}^d) - y']\|_2 \|\hat{L} - \hat{L}^d\|_F
\]

\[
+ \left| \left\langle \sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* [\hat{P}_{\Omega_t} (\bar{X}^d) - y'], \hat{C} - \hat{C}^d \right\rangle \right|
\]

\[
\leq \sqrt{2r} \|\sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* [\hat{P}_{\Omega_t} (\bar{X}^d) - y']\|_2 \|\hat{X} - \hat{X}^d\|_F
\]

\[
+ \sqrt{2r} \|\sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* [\hat{P}_{\Omega_t} (\bar{X}^d) - y']\|_2 \|\hat{C} - \hat{C}^d\|_F
\]

where the first inequality comes from the fact \(|\langle A, B \rangle| \leq \|A\|_* \|B\|_F\). The second inequality holds because \(\|A\|_* \leq \sqrt{r} \|A\|_F\) if \(A\) is a rank-\(r\) matrix. The above analysis follow the similar line of the proof of Proposition 3.1 in [21]. The following proof take care of the additional sparse corruptions in our problem formulation. We will use Mcdiarmid’s Inequality (see Lemma 7 in Appendix H) to bound the sparse term. Note that

\[
| \left\langle \sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* [\hat{P}_{\Omega_t} (\bar{X}^d) - y'], \hat{C} - \hat{C}^d \right\rangle |
\]

\[
= | \left\langle \sum_{t=1}^{d} \omega_t \hat{P}_{\Omega_t}^* \hat{P}_{\Omega_t} \left[ \hat{U} \left( \sum_{s=t+1}^{d} \epsilon^s \right)^T + \hat{C}^d - \hat{C}^d + Z^t \right], \hat{C} - \hat{C}^d \right\rangle |
\]
where each entry of the random matrix $Z^t \in \mathbb{R}^{n_1 \times n_2}$ is i.i.d. Gaussian distributed with variance $\sigma_1^2$. Consider the function

$$f(J_{11}S_{11}, \ldots, J_{n_1 n_2}S_{n_1 n_2}) = \langle J, S \rangle,$$

where

$$J = \sum_{t=1}^{d} \omega_t P_{t}^* P_{t} [\hat{U} (\sum_{s=t+1}^{d} e^s)^T + \hat{C}^d - \hat{C}^d + Z^t],$$

and

$$S = \hat{C} - \hat{C}^d.$$

$$\mathbb{E}(P_{t}^* P_{t} [\hat{U} (\sum_{s=t+1}^{d} e^s)^T + Z^t]) = 0$$ since $Z^t$ and $e^s$ are all from zero mean Gaussian distributions. Then $\mathbb{E}(f(J_{11}S_{11}, \ldots, J_{n_1 n_2}S_{n_1 n_2}))$ can be bounded by

$$\frac{4m_0 s}{n_1 n_2} \alpha^2 \leq \mathbb{E}(f(J_{11}S_{11}, \ldots, J_{n_1 n_2}S_{n_1 n_2})) \leq \frac{4m_0 s}{n_1 n_2} \alpha^2,$$

where $s$ comes from the sparsity upper bound of $S$. Entries of $JS$ are independent with respect to uniform sampling $P_{t}^* P_{t}$ in $J$. $\hat{X}^d - y^t$ is fixed when we apply the McDiarmid’s inequality below. McDiarmid’s inequality bounds the distance between a function and its expectation when the function contains independent random variables (see Lemma 7 in Appendix H). Since $\mathbb{E}(|\langle J, S \rangle|)$ is bounded by $\frac{4m_0 s}{n_1 n_2} \alpha^2$, we can further bound the function $\mathbb{E}(\langle J, S \rangle)$ using McDiarmid’s Inequality. Note that $S$ has $s(s \leq 2s)$ nonzero entries. We have

$$P(|\langle J, S \rangle| \leq \frac{4m_0 s}{n_1 n_2} \alpha^2 + \xi) \geq 1 - 2 \exp(-\frac{\xi^2}{16\alpha^2 s ||J||^2_{\infty}}).$$

By selecting $\xi = 4\alpha ||J||_{\infty} \sqrt{s \log(n_1 + n_2)}$, we then have

$$|\langle \sum_{t=1}^{d} \omega_t P_{t}^* [P_{t} (\hat{X}^d) - y^t], \hat{C} - \hat{C}^d \rangle| \leq \frac{4m_0 s}{n_1 n_2} \alpha^2 + 4\alpha ||J||_{\infty} \sqrt{s \log(n_1 + n_2)}.$$

holds with probability at least $1 - \frac{2}{n_1 + n_2}$. We then have

$$\mathbb{E}||\sum_{t=1}^{d} \omega_t P_{t}^* [P_{t} (\hat{X}^d) - y^t], \hat{C} - \hat{C}^d \rangle||_{\infty} \leq \frac{4m_0 s}{n_1 n_2} \alpha^2 + 4\alpha ||J||_{\infty} \sqrt{s \log(n_1 + n_2)}.$$

$$\mathbb{E}||\sum_{t=1}^{d} \omega_t P_{t}^* [P_{t} (\hat{X}^d) - y^t], \hat{C} - \hat{C}^d \rangle||_{2} \leq \frac{4m_0 s}{n_1 n_2} \alpha^2 + 4\alpha ||J||_{\infty} \sqrt{s \log(n_1 + n_2)}.$$

The result follows from combining (23) and (24). \qed

Appendix E: Proof of Lemma 2

We here upper bound

$$||J||_{2}^2 := \sum_{t=1}^{d} \omega_t ||P_{t} (\hat{X}^d) - y^t||_{2}^2.$$ We first rewrite $J$ as

$$J = \sum_{t=1}^{d} \omega_t P_{t}^* P_{t} [\hat{U} (\sum_{s=t+1}^{d} e^s)^T + \hat{C}^d - \hat{C}^d + Z^t],$$

where each entry of the random matrix $Z^t \in \mathbb{R}^{n_1 \times n_2}$ is i.i.d. Gaussian distributed with variance $\sigma_1^2$. Define

$$Y^t = \hat{U} (\sum_{s=t+1}^{d} e^s)^T, \quad S^t = \hat{C}^d - \hat{C}^d,$$

and

$$F^t = \hat{F}^t + S^t = Y^t + Z^t + S^t.$$
We here introduce the term $S^t$ to take care of the sparse part in our problem formulation. The proof of Lemma 2 follows the similar line of that in [21]. But we need to pay more attention on bounding $S^t$ in the following analysis. We here introduce a random matrix $G^t \in \mathbb{R}^{n_1 \times n_2}$ that has exactly one non-zero entry:

$$G^t = \omega_t n_1 n_2 F^t_{ij} E_{ij},$$

where $E_{ij}$ is the canonical basis of the matrices. We also introduce the following random matrix $H^t$, which is the average of $m_0$ independent copies of $G^t$:

$$H^t = \frac{1}{m_0} \sum_{t=1}^{m_0} G^t_i,$$

where each $G^t_i$ is an independent copy of $G^t$. $J$ then can be decomposed as sum of independent random matrices:

$$J = \frac{m_0}{n_1 n_2} \sum_{t=1}^{d} H^t.$$

We have that

$$\mathbb{E}(G^t) = \mathbb{E}(H^t) = \omega_t F^t, \quad \mathbb{E}(J) = \frac{m_0}{n_1 n_2} \sum_{t=1}^{d} \omega_t F^t.$$  

Lemma 5 in Appendix H describes the spectral norm deviation of a sum of uncentered random matrices from its mean value. We then will apply Lemma 5 to the sum of $dm_0$ independent random matrices.

$$\sum_{t=1}^{d} H^t = \frac{1}{m_0} \sum_{t=1}^{m_0} \sum_{k=1}^{d} G^t_k.$$  

Note that for given $t$ and $k$,

$$\|G^t_k - \mathbb{E}(G^t_k)\|_2 \leq \|G^t_k\|_2 + \|\mathbb{E}(G^t_k)\|_2 \\
\leq \|G^t_k\|_2 + \mathbb{E}(\|G^t_k\|_2) \leq 2\|G^t_k\|_2,$$

where the first inequality applies the triangle inequality, and the second one applies Jensen’s inequality. Also note that

$$0 \leq \sum_{t=1}^{d} \sum_k \mathbb{E} \left[ (G^t_k - \mathbb{E}(G^t_k))(G^t_k - \mathbb{E}(G^t_k))^T \right] \\
= \sum_{t=1}^{d} \sum_k \mathbb{E} (G^t_k (G^t_k)^T) - (\mathbb{E}(G^t_k))(\mathbb{E}(G^t_k))^T \\
\leq \sum_{t=1}^{d} \sum_k \mathbb{E} (G^t_k (G^t_k)^T) \\
= m_0 \sum_{t=1}^{d} \mathbb{E} (G^t (G^t)^T).$$

The third inequality holds because $(\mathbb{E}(G^t_k))(\mathbb{E}(G^t_k))^T$ is positive semidefinite; the last equality uses the fact that for a fixed $t$, $G^t_k$ are random matrices following identical distributions independently for all $1 \leq k \leq m_0$. We then have

$$\rho_0 \left( \sum_{t=1}^{d} H^t \right) \\
\leq \frac{1}{m_0} \max \left\{ \sum_{t=1}^{d} \mathbb{E} \left[ (G^t_k)^T \right], \sum_{t=1}^{d} \mathbb{E} \left[ (G^t_k) G^t_k \right] \right\}.$$

Set

$$\rho_0 := \max \left\{ \sum_{t=1}^{d} \mathbb{E} \left[ (G^t_k)^T \right], \sum_{t=1}^{d} \mathbb{E} \left[ (G^t_k) G^t_k \right] \right\}.$$

The remaining work is to uniformly upper bound $\|G^t_k\|_2$ for all $1 \leq t \leq d$ and $1 \leq k \leq m_0$. We have for all $1 \leq t \leq d$ and $1 \leq k \leq m_0$

$$\|G^t_k\|_2 \leq \max_{i,j,t} \omega_t |n_1 n_2 F^t_{ij} E_{ij}| = n_1 n_2 \max_{i,j,t} |F^t_{ij}|,$$

$$\leq n_1 n_2 \max_{i,j,t} \omega_t (|\hat{F}^t_{ij}| + |S_{ij}|).$$

Since $\hat{L}^t$ is coherent with parameter $\mu_0$ from assumption (a) in Section 3, we have

$$\|\hat{U}_{ii}\|_2 \leq \mu_0 \sqrt{r/n_1} \text{ for any } i = 1, ..., n_1.$$  

We then have

$$\text{Var}(\hat{F}^t_{ij}) \leq \frac{\mu_0^2 r}{n_1} \sigma^2_1 (d - t) + \sigma^2_1.$$  

Let

$$\sigma^2_{\text{max}} = \max_t \omega_t^2 \left( \frac{\mu_0^2 r}{n_1} \sigma^2_1 (d - t) + \sigma^2_1 \right).$$

By the tail probability of Gaussian random variables and the standard union bound, for all $1 \leq t \leq d$ and $1 \leq k \leq m_0$, we have

$$\mathbb{P}(\|G^t_k\|_2 \leq L) \geq 1 - 2/(n_1 + n_2),$$

where

$$L := n_1 n_2 (2 \log(d(n_1 + n_2)n_1 n_2) \sigma^2_{\text{max}} + 2\alpha).$$

We have

$$\mathbb{E}(G^t (G^t)^T) = \omega_t^2 n_1^2 n_2^2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (F^t_{ij})^2 E_{ij} E_{ij}^T \frac{1}{n_1 n_2}$$

$$= \omega_t^2 n_1 n_2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (F^t_{ij})^2 E_{ii}.$$
Similarly, we have
\[ \mathbb{E}((G_t)^T G_t) = \omega_i^2 n_1 n_2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (F_{ij}^t)^2 E_{ij}. \]

Then \( \rho_0 = \)
\[ n_1 n_2 \max \{ \max_i \sum_{t=1}^{d} \sum_{j=1}^{n_2} \omega_i^2 (F_{ij}^t)^2, \max_j \sum_{t=1}^{d} \sum_{i=1}^{n_1} \omega_i^2 (F_{ij}^t)^2 \}. \]

Let
\[ a_i = \sum_{t=1}^{d} \sum_{j=1}^{n_2} \omega_i^2 (F_{ij}^t)^2, \]
and
\[ b_j = \sum_{t=1}^{d} \sum_{i=1}^{n_1} \omega_i^2 (F_{ij}^t)^2. \]

Note that
\[ a_i = \sum_{t=1}^{d} \omega_i^2 \sum_{j=1}^{n_2} (Y_{ij}^t + S_{ij}^t + Z_{ij}^t)^2 \]
\[ \leq 3 \sum_{t=1}^{d} \omega_i^2 \sum_{j=1}^{n_2} [(Y_{ij}^t)^2 + (S_{ij}^t)^2 + (Z_{ij}^t)^2]. \]

Based on the analysis in [21], we have
\[ \mathbb{P}(\max_i \sum_{t=1}^{d} \sum_{j=1}^{n_2} (Z_{ij}^t)^2 \leq 5n_2 \sum_{t=1}^{d} \omega_i^2 \sigma_i^2) \geq 1 - dn_1 e^{-n_2}, \]
\[ \mathbb{P}(\max_j \sum_{t=1}^{d} \sum_{i=1}^{n_1} (Z_{ij}^t)^2 \leq 5n_1 \sum_{t=1}^{d} \omega_i^2 \sigma_1^2) \geq 1 - dn_2 e^{-n_1}, \]
\[ \mathbb{P}(\max_i \sum_{t=1}^{d} \sum_{j=1}^{n_2} (Y_{ij}^t)^2 \leq 5n_2 \sum_{t=1}^{d} \omega_i^2 \sigma_1^2 \sum_{t=1}^{d} (d-t) \omega_i^2 \sigma_1^2) \geq 1 - dn_1 e^{-n_2}, \]
and
\[ \mathbb{P}(\max_j \sum_{t=1}^{d} \sum_{i=1}^{n_1} (Y_{ij}^t)^2 \leq 5n_1 \sum_{t=1}^{d} (d-t) \omega_i^2 \sigma_1^2) \geq 1 - dn_2 e^{-n_1}. \]

Since we involve the sparse corruptions in our problem formulation, we need to use Hoeffding’s Inequality and union bound to bound the term \( S \), as computed in the following equations (25) and (26). By Hoeffding’s Inequality and the union bound, we have
\[ \mathbb{P}(\max_i \sum_{t=1}^{d} \sum_{j=1}^{n_2} (S_{ij}^t)^2 \leq 2\alpha^2 (\sqrt{2}\sigma_1 + \frac{4\alpha}{n_1})) \geq 1 - dn_1 e^{-n_2}, \]
and
\[ \mathbb{P}(\max_j \sum_{t=1}^{d} \sum_{i=1}^{n_1} (S_{ij}^t)^2 \leq 2\alpha^2 (\sqrt{2}\sigma_2 + \frac{4\alpha}{n_2})) \geq 1 - dn_2 e^{-n_1}. \]

We then have
\[ \mathbb{P}(\max_i a_i \leq 15n_2 \sum_{t=1}^{d} \omega_i^2 (\sigma_1^2 + \frac{\mu_0^2 \sigma_1}{n_1} (d-t) \sigma_2^2) + 12\alpha^2 (\frac{\sqrt{5}}{2}n_2 + \frac{2s}{n_1}) \geq 1 - 3dn_1 \exp(-n_2), \]
and
\[ \mathbb{P}(\max_j b_j \leq 15n_1 \sum_{t=1}^{d} \omega_i^2 (\sigma_1^2 + (d-t) \sigma_2^2) + 12\alpha^2 (\frac{\sqrt{5}}{2}n_1 + \frac{2s}{n_2}) \geq 1 - 3dn_2 \exp(-n_1). \]

Note that \( 1 \leq \mu_0 \leq \sqrt{n_1/r} \), so \( \mu_0^2 \leq 1 \). Now we are ready to bound \( \rho_0 \) by combining (27) and (28):
\[ \mathbb{P}(\rho_0 \leq n_1 n_2 (15n_1 \max \sum_{t=1}^{d} \omega_i^2 (\sigma_1^2 + (d-t) \sigma_2^2) + 12\alpha^2 (\frac{\sqrt{5}}{2}n_2 + \frac{2s}{n_1}) \geq 1 - 6dn_1 \exp(-n_2). \]

Now by Lemma 5, we have
\[ \mathbb{P}(\| \sum_{t=1}^{d} H^t - \sum_{t=1}^{d} \omega_i F_t \|_2 \geq s \) \leq (n_1 + n_2) \exp(\frac{-m_0 s^2}{2}) \frac{1}{3}. \]

If we let \( s = \sqrt{8 \log(n_1 + n_2) \nu} \) and substitute this into the above matrix Bernstein inequality we obtain
\[ \mathbb{P}(\| \sum_{t=1}^{d} H^t - \sum_{t=1}^{d} \omega_i F_t \|_2 \leq \sqrt{8 \log(n_1 + n_2) \nu} \leq \frac{1}{n_1 + n_2}. \]
A hidden condition when the above inequality holds is that \( \nu \) dominates the denominator of the exponential term. The remaining work is to have sufficiently large \( m_0 \) to guarantee that \( \nu \) dominates the denominator of the exponential, which follows

\[
\nu \geq (2/3) \sqrt{8 \log(n_1 + n_2) \nu} \frac{m_0}{m_0}.
\]

The above inequality immediately implies that \( m_0 \geq \frac{\sqrt[3] 3 n_1 n_2 \log(n_1 + n_2)(\sqrt 2 \log(d(n_1 + n_2)n_1n_2)\sigma_{\max}^2 + 2\alpha)^2}{5n_{\text{max}} \sum_{t=1}^d \omega_t^2 (\sigma_t^2 + (d - t)\sigma_t^2) + 2\alpha^2 (\sqrt 2 n_{\text{max}} + \frac{4s}{n_{\text{min}}})}
\]

The remaining work is to bound \( \| \sum_{t=1}^d \omega_t F^t \|_2 \). Note that

\[
\| \sum_{t=1}^d \omega_t F^t \|_2 \leq \| \sum_{t=1}^d \omega_t \hat F^t \|_2 + \| \sum_{t=1}^d \omega_t S^t \|_2.
\]

In [21], we only have the first term. We here need to bound the second term because of the sparse corruptions involved. First we note that each entry of \( \hat F \) is Gaussian and the variance is not greater than \( \sigma_t^2 + (d - t)\sigma_t^2 \). Then, according to results on bounds for the spectral norm of i.i.d. Gaussian ensemble, we have

\[
\mathbb{P}(\| \sum_{t=1}^d \omega_t \hat F^t \|_2 \leq 2 \sqrt{\sum_{t=1}^d \omega_t^2 (\sigma_t^2 + (d - t)\sigma_t^2) \sqrt{n_{\text{max}}}} \geq 1 - c_1 \exp(-c_2 n_{\text{max}}),
\]

where \( c_1, c_2 \) are absolute positive constants. Note that \( c_1 \exp(-c_2 n_{\text{max}}) \ll dn_{\text{max}} \exp(-n_{\text{min}}) \). We also have

\[
\| \sum_{t=1}^d \omega_t S^t \|_2 \leq \| \sum_{t=1}^d \omega_t S^t \|_2 \leq 2\alpha \sqrt{s}.
\]

We then have

\[
\mathbb{P}(\| \sum_{t=1}^d \omega_t F^t \|_2 \leq 2 \sqrt{\sum_{t=1}^d \omega_t^2 (\sigma_t^2 + (d - t)\sigma_t^2) \sqrt{n_{\text{max}}}} + 2\alpha \sqrt{s} \geq 1 - c_1 \exp(-c_2 n_{\text{max}}).
\]

Now we are ready to bound \( \| J \|_2^2 \). With probability at least \( 1 - \frac{3}{n_1 + n_2} - 7d n_{\text{max}} \exp(-n_{\text{min}}) \) we have

\[
\| J \|_2^2 \leq p^2 (\| \sum_{t=1}^d \omega_t F^t \|_2 + \sqrt{8 \log(n_1 + n_2) \nu})^2 \\
\leq 2p^2 (\| \sum_{t=1}^d \omega_t F^t \|_2 + \sqrt{8 \log(n_1 + n_2) \nu})^2 \\
\leq 2p^2 (8n_{\text{max}} \sum_{t=1}^d \omega_t^2 (\sigma_t^2 + (d - t)\sigma_t^2) + 8\alpha^2 s + 8n_1 n_2 \log(n_1 + n_2) (15n_{\text{max}} \sum_{t=1}^d \omega_t^2 (\sigma_t^2 + (d - t)\sigma_t^2) \\
+ 12\alpha^2 (\sqrt 2 n_{\text{max}} + \frac{2s}{n_{\text{min}}})) \\
\leq m_0 \log(n_1 + n_2) (25n_{\text{max}} \sum_{t=1}^d \omega_t^2 (\sigma_t^2 + (d - t)\sigma_t^2) \\
+ 192\alpha^2 (\sqrt 2 n_{\text{max}} + \frac{2s}{n_{\text{min}}})) + 16\alpha^2 p^2 s.
\]

**Appendix F: Proof of Lemma 3**

The proof of Lemma 3 follows the same line of the proof of Lemma E.1 in [21]. The major modification is that we need to involve the sparse term. Lemma 4 is the key step of proving Lemma 3, and will be included in the following proof.

**Proof Set**

\[
\mathcal{F} = \frac{44rn_1 n_2}{m_0} (\| \Sigma \|_2^2) + 4\sqrt{r} \mathbb{E}(\| \Sigma \|_2^2) + 3456 \frac{1}{5} n_1.
\]

We will show that the probability of the following event is small:

\[
\mathcal{B} = \{ \exists X \in \mathcal{E}(r,s) \text{ such that} \| \sum_{t=1}^d \omega_t \|_2 X \|_2^2 > p\| X \|_2^2 + \mathcal{F} \}.
\]

Note that \( \mathcal{B} \) contains the complement of the event in Lemma 3. We use a peeling argument to bound the probability of \( \mathcal{B} \). Let

\[
\nu = \max(\sqrt{\frac{186624 \log(n_1 + n_2) \log(n_1 + n_2)}{25m_0 \log(6/5) + 2n_1 \log(6/5)}}, \frac{3}{n_1 n_2}),
\]

and \( \alpha = 6/5 \). For \( l \in \mathcal{N} \) let

\[
S_l = \{ X \in \mathcal{E}(r,s) : \nu^{l-1} \leq \frac{1}{n_1 n_2} \| X \|_2^2 \leq \nu^l \}.
\]
If event $\mathcal{B}$ holds for some $X \in \mathcal{E}(r,s)$, it must be that $X$ belongs to some $S_t$ and

$$\left| \sum_{i=1}^{d} \omega_i \| \mathcal{P}_{\Omega^i}(X) \|_2^2 - p \| X \|_F^2 \right| > \frac{p}{2} \| X \|_F^2 + \mathcal{F} \geq \frac{5}{12} \alpha' \nu m_0 + \mathcal{F}. \quad (29)$$

For $T > \nu$ consider the set

$$\mathcal{E}(r,s,T) = \{ X \in \mathcal{E}(r,s) : \| X \|_F^2 \leq n_1 n_2 T \}$$

and the event

$$\mathcal{B}_t = \{ \exists X \in \mathcal{E}(r,s,\alpha't) \text{ such that } \left| \sum_{i=1}^{d} \omega_i \| \mathcal{P}_{\Omega^i}(X) \|_2^2 - p \| X \|_F^2 \right| > \frac{5}{12} \alpha' \nu m_0 + \mathcal{F} \}. \quad (30)$$

Note that $X \in S_t$ implies that $X \in \mathcal{E}(r,s,\alpha't)$. Then (29) implies that $\mathcal{B}_t$ holds and $\mathcal{B} \subset \cup \mathcal{B}_t$. Thus, it is sufficient to bound the probability of the event $\mathcal{B}_t$. Let

$$H_T = \sup_{X \in \mathcal{E}(r,s,T)} \left| \sum_{i=1}^{d} \omega_i \| \mathcal{P}_{\Omega^i}(X) \|_2^2 - p \| X \|_F^2 \right|.$$ 

Such bound is given by the following Lemma 4.

**Lemma 4** Suppose all $\mathcal{P}_{\Omega^i}$’s are fixed uniform sampling ensembles. Then

$$\mathbb{P}(H_T > \frac{5}{12} m_0 T + \mathcal{F}) \leq \exp(-c_5 m_0 T^2) + 2 \exp(-2n_1 T),$$

where $c_5 = \frac{1}{8} (\frac{1}{18} - \frac{5}{12})^2$.

**Proof** The proof follows the similar line of the proof of Lemma F.1 in [21]. We will use McDiarmid’s Inequality to bound the sparse term introduced by our problem formulation. By Massart’s concentration inequality, we have

$$\mathbb{P}(H_T \geq \mathbb{E}(H_T) + \frac{1}{18} \frac{5}{12} m_0 T) \leq \exp(-c_5 m_0 T^2),$$

where $c_5 = \frac{1}{8} (\frac{1}{18} - \frac{5}{12})^2$. Next we bound the expectation $\mathbb{E}(H_T)$. Using a symmetrization argument we have

$$\mathbb{E}(H_T) \leq 2 \mathbb{E} \left( \sup_{X \in \mathcal{E}(r,s,T)} \left| \sum_{i=1}^{d} \omega_i \gamma_i \sum_{i=1}^{m_0} \langle \mathcal{P}_{\Omega^i}, X \rangle \right| \right),$$

where $\gamma_i$ is a Rademacher variable (independent on both $i$ and $t$). The assumptions $\| L \|_F \leq \frac{1}{2}$ and $\| C \|_\infty \leq \frac{1}{2}$ imply that $\| X \|_\infty \leq 1$. We then have

$$\mathbb{E}(H_T) \leq 8 \mathbb{E}( \sup_{X \in \mathcal{E}(r,s,T)} \left| \sum_{i=1}^{d} \omega_i \gamma_i \sum_{i=1}^{m_0} \langle \mathcal{P}_{\Omega^i}, X \rangle \right|) \leq \frac{1}{2}.$$ 

where $f = \sum_{i=1}^{d} \sum_{i=1}^{m_0} \omega_i \gamma_i P_{\Omega^i}$. One can easily obtain the upper bound on (30) in [21] when there are no corruptions. If corruptions exist, we need to use McDiarmid’s Inequality to bound the sparse term, as computed in the following analysis.

Note that $\langle \Sigma_R, X \rangle = \langle \Sigma_R, L + C \rangle \leq \langle \Sigma_R, L \rangle + \langle \Sigma_R, C \rangle \leq \| \Sigma_R \|_2 \| L \|_* + \langle \Sigma_R, C \rangle$. Since $X \in \mathcal{E}(r,s,T)$, we have

$$\| L \|_* \leq \sqrt{r} \| L \|_F \leq \sqrt{ \frac{r}{n_1 n_2 T} + \frac{1}{2} \sqrt{r} \frac{\nu}{s} \| C \|_F}.$$ 

We have

$$\langle \Sigma_R, C \rangle = \langle \sum_{i=1}^{d} \sum_{i=1}^{m_0} \omega_i \gamma_i \mathcal{P}_{\Omega^i}, C \rangle = \sum_{i=1}^{d} \sum_{i=1}^{m_0} \langle \omega_i \gamma_i \mathcal{P}_{\Omega^i}, C \rangle.$$ 

Note that $\langle \omega_i \gamma_i \mathcal{P}_{\Omega^i}, C \rangle$’s are independent variables, and $\mathbb{E}(\langle \omega_i \gamma_i \mathcal{P}_{\Omega^i}, C \rangle) = 0$. By McDiarmid’s Inequality, we have

$$\mathbb{P}(\| \langle \Sigma_R, C \rangle \| \geq \xi) \leq 2 \exp(-\frac{2 \xi^2}{m_0}).$$

By selecting $\xi = \sqrt{m_0 n_1 T}$, we then have

$$\| \langle \Sigma_R, C \rangle \| \leq \sqrt{m_0 n_1 T}$$

hold with probability at least $1 - 2e^{-2n_1 T}$. We then have with probability at least $1 - 2e^{-2n_1 T}$,

$$\mathbb{E}(H_T) \leq 8 \sqrt{r n_1 n_2 T} \| \Sigma_R \|_2 + 4 \sqrt{r} s \mathbb{E}(\| \Sigma_R \|_2) + 8 \sqrt{m_0 n_1 T}.$$


Note that we have
\[
\begin{align*}
&\frac{1}{18} \left( \frac{5}{12} m_0 T + 8 \sqrt{r n_1 n_2 T} \| \Sigma_R \|_2 \right) + 8 \sqrt{m_0 n_1 T} \\
&+ 4 \sqrt{r s E} \| \Sigma_R \|_2 \\
\leq &\frac{1}{18} \left( \frac{5}{12} m_0 T + \frac{8}{5} \frac{5}{9} m_0 T + \frac{44 r n_1 n_2 (E \| \Sigma_R \|_2)^2}{m_0} \right) \\
&+ \frac{1}{18} \frac{5}{12} m_0 T + \frac{3456}{5} n_1 + 4 \sqrt{r s E} \| \Sigma_R \|_2 \\
= &\frac{5}{12} m_0 T + \frac{44 r n_1 n_2 (E \| \Sigma_R \|_2)^2}{m_0} + \frac{3456}{5} n_1 \\
&+ 4 \sqrt{r s E} \| \Sigma_R \|_2.
\end{align*}
\]

We then complete the proof. \( \square \)

The above lemma implies that
\[
P(\mathcal{B}_i) \leq \exp(-c_5 m_0 \alpha^{2t} \nu^2) + 2 \exp(-2 n_1 \alpha^t \nu).
\]

By applying union bound, we have
\[
P(\mathcal{B}) \leq \sum_{i=1}^\infty P(\mathcal{B}_i)
\leq \sum_{i=1}^\infty \exp(-c_5 m_0 \alpha^{2t} \nu^2) + \sum_{i=1}^\infty 2 \exp(-2 n_1 \alpha^t \nu)
\leq \sum_{i=1}^\infty \exp(-2 c_5 m_0 \log(\alpha) \nu^2 t)
\]
\[
+ \sum_{i=1}^\infty 2 \exp(-2 n_1 \log(\alpha) \nu t),
\]

where the last equality uses the bound \( e^x \geq x \). Then, substituting
\[
\nu = \max\left( \sqrt{\frac{\log(n_1 + n_2)}{2c_5 m_0 \log(6/5)}}, \frac{\log(n_1 + n_2)}{2n_1 \log(6/5)} \right)
\]
into the above summation we obtain
\[
P(\mathcal{B}) \leq \frac{6}{n_1 + n_2}.
\]

We then complete the proof. \( \square \)

**Appendix G: Convergence analysis of Algorithm 1**

To prove the convergence of our algorithm, we first prove the Lipschitz continuous property of the first-order partial derivatives, which plays a key role in proving the nonincrease property of the function value.

We here calculate the first-order partial derivatives of \( H(U, V, L, C) \):

\[
\nabla_U H(U, V, L, C) = -\rho(L - UV^T)V,
\]
\[
\nabla_V H(U, V, L, C) = -\rho(L - UV^T)^T U,
\]
\[
\nabla_L H(U, V, L, C) = \sum_{i=1}^d \omega_i P^*_{\Omega_i} (P_{\Omega_i} (L + C) - y_i^l)^T + \rho(L - UV^T),
\]
\[
\nabla_C H(U, V, L, C) = \sum_{i=1}^d \omega_i P^*_{\Omega_i} (P_{\Omega_i} (L + C) - y_i^l).
\]

With the other three variables fixed, we have
\[
||\nabla_U H(U_1, V_i, L_i, C_i) - \nabla_U H(U_2, V_i, L_i, C_i)||_F \\
= ||-\rho(L_i - U_1 V_i^T) V_i + \rho(L_i - U_2 V_i^T) V_i||_F \\
\leq \rho \|V_i^T V_i\|_2 \|U_1 - U_2\|_F \\
\leq \rho \|V_i^T V_i\|_F \|U_1 - U_2\|_F \\
\leq \lambda_i \|U_1 - U_2\|_F \\
= \frac{1}{\beta TV} \|U_1 - U_2\|_F,
\]

where (a) follows from the inequality \( \| \cdot \|_2 \leq \| \cdot \|_F \). Under the assumptions of boundedness, \( \nabla_U H \) is Lipschitz continuous, and \( \lambda_i \) is bounded away from zero.

Similarly, \( \nabla_V H \) is Lipschitz continuous, and \( \lambda_i \) is bounded away from zero.
\[ \| \nabla \nabla \|_F \| U_{i+1}, V_{i+1}, L, C_i \|_F \]
\[ - \nabla \nabla \|_F \| U_{i+1}, V_{i+1}, L, C_i \|_F \]
\[ = \| \sum_{t=1}^d \omega_t (P_{\Omega^t}^* (P_{\Omega^t} (L_1 - L_2))) + \rho(L_1 - L_2) \|_F \]
\[ \leq \sum_{t=1}^d \omega_t \| P_{\Omega^t}^* \|_F \| P_{\Omega^t} (L_1 - L_2) \|_F + \rho \| L_1 - L_2 \|_F \]
\[ \leq \sum_{t=1}^d \omega_t \| L_1 - L_2 \|_F + \rho \| L_1 - L_2 \|_F \]
\[ = (\rho + 1) \| L_1 - L_2 \|_F \]
\[ = \| L_1 - L_2 \|_F \]
\[ = \frac{1}{\beta^{2} \| L_1 - L_2 \|_F} \]

where (b) follows the fact that \( \sum_{i=1}^d \omega_t = 1 \).

\[ \| \nabla C \|_F \| U_{i+1}, V_{i+1}, L, C_i \|_F \]
\[ - \nabla \nabla \|_F \| U_{i+1}, V_{i+1}, L, C_i \|_F \]
\[ = \| \sum_{t=1}^d \omega_t (P_{\Omega^t}^* (P_{\Omega^t} (C_1 - C_2))) \|_F \]
\[ \leq \sum_{t=1}^d \omega_t \| P_{\Omega^t}^* \|_F \| P_{\Omega^t} (C_1 - C_2) \|_F \]
\[ \leq \sum_{t=1}^d \omega_t \| C_1 - C_2 \|_F \]
\[ = \| C_1 - C_2 \|_F = \| C_1 - C_2 \|_F \]
\[ = \frac{1}{\beta^{2} \| C_1 - C_2 \|_F} \]

We already show that the partial derivatives of \( H(U, V, L, C) \) are Lipschitz continuous. In the light of Ref. [38], we separate the main steps to achieve the convergence result into two major parts and outline them as follows. The first part guarantees that the sequence generated by the algorithm has convergent subsequence and the set of the accumulations points is a subset of the critical points. Let \( \Phi = H(U, V, L, C) + J(L) + Q(C) + K(C) \). Given the Lipschitz continuous property, we can then prove that

\[ \gamma_1 \| z_i - z_{i+1} \| \leq \| \Phi(z_i) - \Phi(z_{i+1}) \|, \forall i = 0, 1, 2, \ldots \]

where \( \{ z_i \} \) is the sequence generated by our algorithm, and \( \gamma_1 \) is a positive constant. This result shows that the function value is nonincreasing. With the boundness assumption of the sequence, we can then show that

\[ \gamma_2 \| \omega_i \| \leq \| z_i - z_{i+1} \|, \omega_i \in \partial \Phi(z_i), \forall i = 0, 1, 2, \ldots \]

where \( \gamma_2 \) is another positive constant. The conclusions can be drawn from the above results.

After the first part, the global convergence to a critical point has not been proved yet since the sequence itself is not a Cauchy sequence. We then prove the Kurdyka-Lojasiewicz (KL) property of \( \Phi \), which does not depend on the structure of the specifically chosen algorithm. Since \( H(U, V, L, C) \) is differentiable everywhere, the function is real analytic, therefore \( H(U, V, L, C) \) has the KL property according to Section 2.2 of [39]. Note that \( J(L), Q(C), \) and \( K(C) \) are all indicator functions of semi-algebraic sets [38]. Thus, \( \Phi \) satisfies the KL property. From [38], we can conclude that \( \{ z_i \} \) is a Cauchy sequence. All the properties above guarantee that our algorithm can converge globally to a critical point of (9).

**Appendix H: Additional auxiliary Lemmas**

**Lemma 5** [Corollary 6.1.2 in [42]] Consider a finite sequence \( \{ S_k \} \) of independent random matrices with common dimension \( n_1 \times n_2 \). Assume that each matrix has uniformly bounded deviation from its mean:

\[ \| S_k - E(S_k) \|_F \leq L \quad \text{for each index } k. \]

Consider the sum

\[ Z = \sum_k S_k. \]

Let \( \rho(Z) \) denote the matrix variance statistic of the sum:

\[ \rho(Z) = \max \{ \| E[(Z - E(Z))(Z - E(Z))^T] \|_2, \]
\[ \| E[(Z - E(Z))^T (Z - E(Z))] \|_2 \}
\[ = \max \{ \| \sum_k E[(S_k - E(S_k))(S_k - E(S_k))^T] \|_2 \}
\[ \| \sum_k E[(S_k - E(S_k))^T (S_k - E(S_k))] \|_2 \}. \]

Then for all \( \kappa \geq 0 \),

\[ P(\| Z - E(Z) \|_2 \geq \kappa) \leq (n_1 + n_2) \exp(\frac{-\kappa^2}{\rho(Z) + L \kappa/2^3}). \]

**Lemma 6** (Lemma E.2 in [21]) Suppose all \( \mathcal{P}_{\Omega^t} \)’s are fixed uniform sampling ensembles. For

\[ m_0 \geq Dn_{\text{min}} \log(n_1 + n_2) \phi(\omega), \]

where \( \phi(\omega) = \omega_{\text{max}}^2 / \sum_{t=1}^d \omega_t^2 \), there exists an absolute positive constant \( c \) such that

\[ E(\| \Sigma_k \|_2) \leq c \frac{2e \log(n_1 + n_2) \sum_{t=1}^d \omega_t^2 m_0}{n_{\text{min}}}. \]
Lemma 7 (McDiarmid’s Inequality) Let $X_1, \ldots, X_m$ be independent random variables all taking values in the set $\mathcal{X}$. Further, let $f : \mathcal{X}^m \rightarrow \mathbb{R}$ be a function of $X_1, \ldots, X_m$ that satisfies $\forall i, \forall x_1, \ldots, x_m, x'_i \in \mathcal{X}$,

$$|f(x_1, \ldots, x_i, \ldots, x_m) - f(x_1, \ldots, x'_i, \ldots, x_m)| \leq c_i.$$

Then for all $\epsilon > 0$,

$$P(f - \mathbb{E}(f) \geq \epsilon) \leq \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^{m} c_i^2}\right). \quad (31)$$

Abbreviations

LRMC: Low-rank matrix completion; RPCA: Robust principal component analysis; RMC: Robust matrix completion; PALM: Proximal alternating linearized minimization; LOWEMS: Locally weighted matrix smoothing; TRPCA: Tensor robust principal component analysis

Acknowledgements

This research is supported in part by ARO W911NF-17-1-0047, NSF #1508875 and the ERC Program of NSF and DoE under the supplement to NSF Award EEC-1041877 and the CURENT Industry Partnership Program.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Informed consent was obtained from all authors included in the study.

Author’s contributions

Ren and Pengzhi draft the manuscript. Meng revised the paper. Ren and Pengzhi conceived the method, and the experiments. Ren and Pengzhi revised the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The public video dataset analysed during the current study are available in https://media.xiph.org/video/derf/

Author details

1 Dept. of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, US. 2 University of Michigan, Ann Arbor, MI, US. 3 Baidu Research.

References

1. Veganzoni, M.A., Simões, M., Liciardi, G., Yokoya, N., Bioucas-Dias, J.M., Chanussot, J.: Hyperspectral super-resolution of locally low rank images from complementary multisource data. IEEE Transactions on Image Processing 25(1), 274–288 (2016)
2. Dao, M., Nguyen, L., Tran, T.D.: Temporal rate up-conversion of synthetic aperture radar via low-rank matrix recovery. In: Proc. IEEE International Conference on Image Processing, pp. 2358–2362 (2013)
3. Sumarsono, A., Du, Q.: Low-rank subspace representation for estimating the number of signal subspaces in hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing 53(11), 6286–6292 (2015)
4. The Netflix Prize. http://netflixprize.com/ (2006)
5. Fazel, M.: Matrix rank minimization with applications. PhD thesis, Stanford University (2002)
6. Eftekhar, A., Yang, D., Wakin, M.B.: Weighted matrix completion and recovery with prior subspace information. arXiv:1612.01720 (2018)
7. Candes, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM (JACM) 58(3), 11 (2011)
8. Tanner, J., Thompson, A., Vary S.: Matrix rigidity and the ill-posedness of robust pca and matrix completion. SIAM Journal on Mathematics of Data Science 1(3), 537–554 (2019)
9. Davenport, M.A., Romberg, J.: An overview of low-rank matrix recovery from incomplete observations. IEEE J. Sel. Topics Signal Process. 10(4), 608–622 (2016)
10. Gao, P., Wang, M., Chow, J.H., Ghioel, S.G., Fardanesh, B., Stefopoulos, G., Razanousky, M.P.: Identification of successive “unobservable” cyber data attacks in power systems. IEEE Trans. Signal Process. 64(21), 5557–5570 (2016)
11. Candes, E.J., Recht, B.: Exact matrix completion via convex optimization. Foundations of Comput. Math. 9(6), 717–772 (2009)
12. Dai, J.-F., Candes, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optimiz. 20(4), 1956–1982 (2010)
13. Candes, E., Tao, T.: Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inf. Theory 52(12), 5406–5425 (2006)
14. Jain, P., Netrapalli, P., Sanghavi, S.: Low-rank matrix completion using alternating minimization. In: Proc. 45th Annu. ACM Symp. Theory of Comput., pp. 665–674 (2013)
15. Netrapalli, P., Niranjan, U., Sanghavi, S., Anandkumar, A., Jain, P.: Non-convex robust PCA. In: Adv. Neural Inf. Process. Syst., pp. 1107–1115 (2014)
16. Cai, J.-F., Shen, Z.: Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank Hankel matrix completion. Applied and Computational Harmonic Analysis 46(1), 94–121 (2017)
17. Jain, P., Meka, R., Dhillon, I.S.: Guaranteed rank minimization via singular value projection. In: Advances in Neural Information Processing Systems, pp. 937–945 (2010)
18. Lee, K., Bresler, Y.: Admira: Atomic decomposition for minimum rank approximation. IEEE Transactions on Information Theory 56(6), 4402–4416 (2010)
19. Keshavan, R.H., Montanari, A., Oh, S.: Matrix completion from a few entries. IEEE transactions on information theory 56(6), 2980–2998 (2010)
20. Dai, W., Milenkovic, O., Kerman, E.: Subspace evolution and transfer (set) for low-rank matrix completion. IEEE Transactions on Signal Processing 59(7), 3120–3132 (2011)
21. Xu, L., Davenport, M.: Dynamic matrix recovery from incomplete observations under an exact low-rank constraint. In: Advances in Neural Information Processing Systems 29, pp. 3585–3593 (2016)
22. Qiu, C., Vaswani, N.: Real-time robust principal components’ pursuit. In: Allerton Conf. on Communication, Control, and Computing (2010)
23. Guo, H., Qiu, C., Vaswani, N.: An online algorithm for separating sparse and low-dimensional signal sequences from their sum. IEEE Trans. Signal Processing. 62(16), 4284–4297 (2014)
24. Qiu, C., Vaswani, N., Lois, B., Hegden, L.: Recursive robust PCA or recursive sparse recovery in large but structured noise. IEEE Transactions on Information Theory 60(8), 5007–5039 (2014)
25. Zhan, J., Lois, B., Guo, H., Vaswani, N.: Online (and offline) robust PCA: Novel algorithms and performance guarantees. In: Artificial Intelligence and Statistics, pp. 1488–1496 (2016)
26. Balzano, L., Nowak, R., Recht, B.: Online identification and tracking of subspaces from highly incomplete information. In: Proc. Allerton Conf. Communication, Control, and Computing, pp. 704–711 (2010)
27. Chi, Y., Eldar, Y.C., Calderbank, R.: PETRELS: parallel subspace estimation and tracking by recursive least squares from partial observations. IEEE Trans. Signal Processing. 61(23), 5947–5959 (2013)
28. Xie, Y., Huang, J., Willett, R.: Change-point detection for high-dimensional time series with missing data. IEEE J. Sel. Topics Signal Process. 7(1) (2013)
29. Vaswani, N., Narayanamurthi, P.: Static and dynamic robust PCA and matrix completion: a review. arXiv:1803.00651 (2018)
30. Yang, Y., Feng, Y., Suykens, J.A.: Robust low-rank tensor recovery with regularized redescending m-estimator. IEEE transactions on neural networks and learning systems 27(9), 1933–1946 (2015)
31. Goldfarb, D., Qin, Z.: Robust low-rank tensor recovery: Models and algorithms. SIAM Journal on Matrix Analysis and Applications 35(1), 225–253 (2014)
32. Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., Yan, S.: Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5249–5257 (2016)
33. Song, G., Ng, M.K., Zhang, X.: Robust tensor completion using...
transformed tensor svd. arXiv preprint arXiv:1907.01113 (2019)
34. Klopp, O., Lounici, K., Tsybakov, A.B.: Robust matrix completion. Probability Theory and Related Fields 169(1), 523–564 (2017)
35. Snyman, J.A., Stander, N., Roux, W.J.: A dynamic penalty function method for the solution of structural optimization problems. Appl. Math. Model. 18(8), 453–460 (1994)
36. Haldar, J.P., Hernando, D.: Rank-constrained solutions to linear matrix equations using powerfactorization. IEEE Signal Processing Letters 16(7), 584–587 (2009)
37. Gao, P., Wang, M.: Dynamic matrix recovery from partially observed and erroneous measurements. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2018)
38. Bolte, J., Shoham, S., Marc, T.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. 146(1-2), 459–494 (2014)
39. Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM J. Imag. Sci. 6(3), 1758–1789 (2013)
40. Grant, M., Boyd, S.: CVX: Matlab Software for Disciplined Convex Programming, version 1.21. http://cvxr.com/ (2010)
41. Foundation, T.X.O.: Xiph.org Video Test Media. https://media.xiph.org/video/derf/ (accessed Nov. 10, 2020)
42. Tropp, J.A.: An introduction to matrix concentration inequalities. Foundations and Trends in Machine Learning 8(1-2), 1–230 (2015)