DYNAMIC MATRIX RECOVERY FROM PARTIALLY OBSERVED AND ERRONEOUS MEASUREMENTS

Pengzhi Gao *Meng Wang*

Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY, 12180, USA

ABSTRACT

This paper studies the low-rank matrix recovery problem from partially lost and partially corrupted measurements. It shows both analytically and numerically that the recovery performance can be greatly enhanced if one further exploits the temporal correlations among a sequence of low-rank matrices. The matrix recovery problem is formulated as a non-convex optimization problem, and the recovery error is quantified analytically. A fast iterative algorithm is proposed to solve the non-convex problem, and every sequence generated by the algorithm converges to a critical point of the optimization problem. The method is numerically evaluated on the synthetic datasets.

***Index Terms*—** low-rank matrix, robust PCA, temporal correlation, matrix completion.

1. INTRODUCTION

Many practical datasets have intrinsic low-dimensional structures despite the high ambient dimension. Examples of low-dimensional structures include sparse signals [5] and low-rank matrices [7]. A signal is sparse if it only has a small fraction of nonzero entries. A matrix $M \in \mathbb{R}^{m\times n}$ is low-rank if its rank $r \ll m, n$. Motivated by applications like image and video processing [8, 21], remote sensing [20] and collaborative filtering [1], the low-dimensional structures have been extensively studied in recent years in problems like Low-Rank Matrix Completion (LRMC) [9] and Robust Principal Component Analysis (RPCA) [6].

LRMC aims to recover unobserved entries of a rank-r matrix M from partial observations. Given $M = \bar{L} + \bar{C}$, where $\bar{L}$ is low-rank matrix, and $\bar{C}$ is a sparse matrix with at most s nonzero entries, RPCA wants to separate $\bar{L}$ and $\bar{C}$ from M. In video processing, $\bar{L}$ represents slow-varying background, and $\bar{C}$ models moving objects. In power system monitoring, $\bar{L}$ models the spatial-temporal blocks of ground-truth data, and $\bar{C}$ models the erroneous measurements [10]. Both problems can be formulated as non-convex optimization problems, due to the non-convex rank and sparsity constraints. Solving convex relaxations of these problems by replacing the rank constraint with the nuclear norm [3, 7, 9] constraint and the sparsity constraint with the ℓ_1-norm constraint [5] are proven to return the original matrix under certain conditions [6, 7]. Since it is still time-consuming to solve large-scale convex problems, fast approximation methods for the original non-convex formulations have attracted much attention recently. Despite the numerical superiority, the theoretical analyses of non-convex methods are significantly lagging. Only a few recent work such as [4, 14, 19] proved the convergence of some non-convex approaches to the ground truth data if the starting points of the algorithms are smartly chosen.

Most existing work assume that the low-dimensional structure does not change over time and consider one fixed low-rank matrix. However, users' preference may change over time in collaborative filtering [22]. The background of images and videos also change over time. The temporal variation of the low-dimensional structure has not been much investigated. Parametric models like hidden Markov models [16, 18] and autoregression models [13, 17] have been applied to model the temporal correlations and demonstrated encouraging numerical performance, but the theoretical study is very limited. Moreover, the accuracy of the algorithms largely depends on the correct estimation of the model parameters. RPCA with the weak temporal correlations was studied in [23], and the theoretical analysis only holds when the temporal correlations of the data points are relatively weak. [22] proposes a model of a sequence of dynamically correlated low-rank matrices through slow-varying subspaces. It develops non-convex-optimization-based methods for both matrix sensing and matrix completion, and further characterizes the performance analytically as a function of correlations in the model. Each matrix is low-rank, and the sparse errors are not considered in [22].

This paper studies the problems of recovering low-rank matrices when the measurements are partially corrupted and partially lost. It models the correlations of low-rank matrices through time-varying subspaces. The model studied in this paper generalizes from the one in [22] by additionally modeling sparse errors in the measurements. To the best of our knowledge, this is the first analytical study of robust matrix recovery of temporally correlated matrices when the matrices contain partial corruptions and erasures at the same time. We formulate the matrix recovery problem as a non-convex optimization problem and characterize the recovery error theoretically. Our theoretical bound verifies the intuition that compared with recovering each low-rank matrix individually, one can enhance the recovery performance by further exploiting the temporal correlations in a sequence of low-rank matrices. We propose a fast iterative method to solve the non-convex problem approximately and show that the algorithm can converge to a critical point of the non-convex problem.

The rest of the paper is organized as follows. Section 2 introduces our model and problem formulation. Section 3 provides the analytical 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. PROBLEM FORMULATION AND MOTIVATION

Our problem formulation is built upon and generalizes the problem setup in [22]. Let $\bar{L}^t \in \mathbb{R}^{n_1\times n_2}$ denote the actual data at time t, and let $\bar{C}^t \in \mathbb{R}^{n_1\times n_2}$ 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 sub-

978-1-5386-4658-8/18/$31.00 ©2018 IEEE

spaces. Specifically, let $\bar{X} \in \mathbb{R}^{n_1 \times n_2}$ denote the measurements at time t,

$$\bar{X}^t = \bar{L}^t + \bar{C}^t = \bar{U}^t(\bar{V}^t)^T + \bar{C}^t, \tag{1}$$

where $\bar{L}^t$ has rank at most r, and $\bar{C}^t$ has at most s nonzero entries. Note that the low-rank constraint on $\bar{L}^t$ is imposed by factorizing $\bar{L}^t$ as $\bar{L}^t = \bar{U}^t(\bar{V}^t)^T$, where $\bar{U}^t \in \mathbb{R}^{n_1 \times r}$, and $\bar{V}^t \in \mathbb{R}^{n_2 \times r}$. We further assume $\|\bar{L}^t\|_\infty \leq \alpha$ and $\|\bar{C}^t\|_\infty \leq \alpha$ for some constant α. The temporal correlations can be modeled by correlations in $\bar{U}^t$'s and $\bar{V}^t$'s. Without loss of generality, we assume $\bar{V}^t$ changes over time, while $\bar{U}^t$ is fixed to be $\bar{U}$ such that we have $\bar{L}^t = \bar{U}(\bar{V}^t)^T$. For the sake of simplicity in our analysis, we consider a simple model on $\bar{V}^t$ as follows.

$$\bar{V}^t = \bar{V}^{t-1} + \epsilon^t, t = 2, ..., d, \tag{2}$$

where ϵ^t represents the perturbation noise in $\bar{V}^t$. Note that (2) is only used for analysis, while our proposed method does not need the information of ϵ^t. Let $Z^t \in \mathbb{R}^{n_1 \times n_2}$ represent the measurement noise. Ω^t is the set of observed entries in $\bar{X}^t$ with $|\Omega^t| = m^t$. The partial observed measurements can be presented by

$$Y^t = \mathcal{P}_{\Omega^t}(\bar{X}^t + Z^t). \tag{3}$$

To simplify our discussion, our goal is to recover the matrix at the most recent time-step. The data recovery question is stated as follows. Given partially observed and partially corrupted measurements $\{Y^t\}$ for $t = 1, .., d$, can we recover the actual data $\bar{L}^d$?

Our problem formulation is motivated by background subtraction in video analysis. Each column of $\bar{X}^t$ is a vectorized frame of a sequence of n_2 video frames. $\bar{L}^t$ models the slow-changing common background of the video, and $\bar{C}^t$ models the foreground features or bad measurements in the video.

We propose to estimate the unknown $(\bar{L}^d, \bar{C}^d)$ using a non-convex optimization approach. Note that the objective function is given by

$$F(X) = \frac{1}{2} \sum_{t=1}^{d} \omega_t \|\mathcal{P}_{\Omega^t}(X) - Y^t\|_F^2, \tag{4}$$

where $\{\omega_t\}$ for $t = 1, ..., d$ 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

$$\begin{aligned}(\hat{L}, \hat{C}) &= \text{argmin}_{(L,C)} \frac{1}{2} \sum_{t=1}^{d} \omega_t \|\mathcal{P}_{\Omega^t}(L + C) - Y^t\|_F^2 \\ \text{s.t. } &\|L\|_\infty \leq \alpha, \|C\|_\infty \leq \alpha, \text{rank}(L) \leq r, \sum_{ij} 1_{[C_{ij} \neq 0]} \leq s.\end{aligned} \tag{5}$$

3. DYNAMIC MATRIX COMPLETION WITH ERRONEOUS MEASUREMENTS

Although (5) is non-convex due to the non-convexity of the feasible set, we first analyze the recovery accuracy of the global minimizer of (5). we will propose an algorithm to solve (5) approximately in Section 4.

Our theoretical bound is built upon and generalizes the result in [22]. We first define some coefficients which appear in the bound. Define

$$n_{\max} = \max(n_1, n_2), \tag{6}$$

and

$$n_{\min} = \min(n_1, n_2). \tag{7}$$

Definition 3.1. A rank-r matrix $X \in \mathbb{R}^{n_1 \times n_2}$ with SVD $X = U\Sigma V^T$ is incoherent with parameter μ if

$$\|U_{:i}\|_2 \leq \mu \sqrt{\frac{r}{n_1}} \text{ for any } i = 1, ..., n_1 \tag{8}$$

and

$$\|V_{:j}\|_2 \leq \mu \sqrt{\frac{r}{n_2}} \text{ for any } j = 1, ..., n_2, \tag{9}$$

i.e., the subspaces spanned by the columns of U and V are both μ-incoherent.

Note that $U \in \mathbb{R}^{n_1 \times n_1}$ and $V \in \mathbb{R}^{n_2 \times n_2}$ in Definition 3.1 are different from $\bar{U}^t \in \mathbb{R}^{n_1 \times r}$ and $\bar{V}^t \in \mathbb{R}^{n_2 \times r}$ in (1). U and V here are orthogonal matrices, however there is no orthogonality requirement on $\bar{U}^t$ and $\bar{V}^t$.

We also assume the operator $\mathcal{P}_{\Omega^t}$ is a uniform sampling ensemble with replacement, which means all sensing matrices in the operator are i.i.d. uniformly distributed on the set

$$\mathcal{X} = \{e_j(n_1)e_k^T(n_2), 1 \leq j \leq n_1, 1 \leq k \leq n_2\}, \tag{10}$$

where $e_j(n)$ are the canonical basis vectors in $\mathbb{R}^n$. For the simplicity of analysis, we assume $|\Omega^t|$'s are the same for $t = 1, ..., d$ and set $m_0 = |\Omega^t|, \forall t$. Note that the locations of observations in Ω^t's are different. Let $p = m_0/(n_1 n_2)$ denote the fraction of sampled entries.

Suppose that we are given measurements as in (3) where all $\mathcal{P}_{\Omega^t}$'s are uniform sampling ensembles. Assume that $\bar{L}^t$ evolves according to (2), has rank at most r, and is incoherent with parameter μ_0, $\bar{C}^t$ has at most s nonzero entries. $\|\bar{L}^t\|_\infty \leq \alpha$ and $\|\bar{C}^t\|_\infty \leq \alpha$. Further assume that the measurement noise Z^t is i.i.d. $\mathcal{N}(0, \sigma_1^2)$ for $1 \leq t \leq d$ and that the perturbation noise ϵ^t is i.i.d. $\mathcal{N}(0, \sigma_2^2)$ for $2 \leq t \leq d$. The recovery guarantee is as follows.

Theorem 1. If $m_0 \geq$

$$\frac{c_1 n_1 n_2 \log(n_1 + n_2)(\sqrt{2\log(d(n_1 + n_2)n_1 n_2)\sigma_{\max}^2} + 2\alpha)^2}{5n_{\max} \sum_{t=1}^{d} \omega_t^2(\sigma_1^2 + (d - t)\sigma_2^2) + 2\alpha^2(\sqrt{2}n_{\max} + \frac{4s}{n_{\min}})}, \tag{11}$$

then the estimator $(\hat{L}, \hat{C})$ from (5) satisfies

$$\frac{1}{n_1 n_2}\|\hat{L} + \hat{C} - \bar{L}^d - \bar{C}^d\|_F^2 \leq \max(B_1, B_2), \tag{12}$$

with probability at least $1 - \frac{11}{n_1 + n_2} - 7dn_{\max}\exp(-n_{\min})$, where

$$\sigma_{\max}^2 = \max_t \omega_t^2 (\frac{\mu_0^2 r}{n_1}\sigma_2^2(d - t) + \sigma_1^2), \tag{13}$$

$$B_1 = 16\alpha^2 \max(\sqrt{c_2 \frac{\log(n_1 + n_2)}{m_0 \log(6/5)}}, \frac{\log(n_1 + n_2)}{2n_1 \log(6/5)}), \tag{14}$$

$$\begin{aligned}B_2 = &\frac{256\alpha^2}{m_0}(176ec_3^2 r n_{\max} \log(n_1 + n_2) \sum_{t=1}^{d} \omega_t^2 + \frac{3456}{5} n_1 \\ &+ 8c_3\sqrt{rs}\sqrt{\frac{2e\log(n_1 + n_2)\sum_{t=1}^d \omega_t^2 m_0}{n_{\min}}}) + \frac{16\alpha\sqrt{2s\kappa r}}{m_0} \\ &+ \frac{32r}{m_0}\log(n_1 + n_2)\kappa + \frac{32\alpha^2}{n_1 n_2} + \frac{32\alpha}{m_0}\sqrt{\kappa s \log(n_1 + n_2)},\end{aligned} \tag{15}$$

$$\kappa = 256 n_{\max} \sum_{t=1}^{d} \omega_t^2 (\sigma_1^2 + (d-t)\sigma_2^2) + 16\alpha^2 p^2 s$$
$$+192\alpha^2 (\frac{\sqrt{2}}{2} n_{\max} + \frac{2s}{n_{\min}}), \tag{16}$$

and c_1, c_2 and c_3 are constants.

The proof of Theorem 1 is skipped due to the page limit. Please refer to [11] for the proofs. Assume n_1 and n_2 are in the same order $O(n)$. When the measurements do not contain corruptions, i.e., $\bar{C}^t = 0$ for $t = 1, ..., d$, Ref. [22] showed that if $m_0 \geq O(n(\log(n))^2)$, it holds that

$$\frac{\|\hat{L} - L^d\|_F^2}{n_1 n_2} \leq \max(O(\sqrt{\frac{\log n}{m_0}}), O(\frac{n \log n}{m_0})), \tag{17}$$

when the feasible set is imposed by rank constraint. One can check that our theoretical result reduces to a result comparable to (17) when we set $s = 0$ in Theorem 1.

Theorem 1 establishes the recovery error when the measurements are partially corrupted and partially observed. If $s = O(n)$, i.e., the number of corrupted measurements per row is bounded, we then have if $m_0 \geq O(n(\log(n))^2)$,

$$\frac{\|\hat{L} + \hat{C} - L^d - C^d\|_F^2}{n_1 n_2} \leq \max(B_1, B_2), \tag{18}$$

where

$$B_1 = \max(O(\sqrt{\frac{\log n}{m_0}}), O(\frac{\log n}{n})), \tag{19}$$

and

$$B_2 = O(\frac{n \log n}{m_0}). \tag{20}$$

Note that (18) diminishes to zero when n increases, and (18) is in the same order as the result in [22], where $\bar{C}^t$'s are all zeros. If we choose the weight $\omega_t = \frac{1}{d}$ for $t = 1, ..., d$, one can check that the right hand side of (11) is in the order of $O(\frac{\log d}{d})$, which implies that the required number of observations of each matrix is reduced when d increases, by a factor of $O(\frac{\log d}{d})$. One can also check that two terms in B_2 decrease with the increasing of d, which means the recover error reduces by exploiting the temporal dynamic in the low rank matrices.

4. ALGORITHM FOR DYNAMIC MATRIX COMPLETION WITH ERRONEOUS MEASUREMENTS

We have discussed the theoretical guarantee of dynamic robust matrix completion in Section 3. We here propose an algorithm to solve the non-convex optimization problem (5) approximately. We use the matrix factorization technique, where 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 $L = UV^T$. In each iteration, we apply alternating gradient descent to update the estimations of U, V and C. We use the hard thresholding technique to refine the matrix C by only keeping s entries with the largest absolute values. Define

$$\mathcal{P}_\alpha^1(\begin{bmatrix} U \\ V \end{bmatrix}) = \begin{cases} \sqrt{\frac{\alpha}{\|UV^T\|_\infty}} \begin{bmatrix} U \\ V \end{bmatrix}, & \text{if } \|UV^T\|_\infty > \alpha, \\ \begin{bmatrix} U \\ V \end{bmatrix}, & \text{if } \|UV^T\|_\infty \leq \alpha, \end{cases} \tag{21}$$

and

$$\mathcal{P}_\alpha^2(C) = \begin{cases} \frac{\alpha}{\|C\|_\infty} C, & \text{if } \|C\|_\infty > \alpha, \\ C, & \text{if } \|C\|_\infty \leq \alpha. \end{cases} \tag{22}$$

Note that $\mathcal{P}_\alpha^1(\cdot)$ and $\mathcal{P}_\alpha^2(\cdot)$ are introduced for the infinity norm constraints on L and C. The details of our algorithm is summarized in Algorithm 1. Note that the step size τ in Algorithm 1 is selected via a backtracking line search using Armijo's rule as follows. We fix a parameter $\beta \in (0, 1)$ and start with $\tau = 1$. We update τ by $\beta\tau$ until it holds that

$$F(X - \tau \nabla F(X)) \leq F(X) - \frac{\tau}{2} \|\nabla F(X)\|_2^2. \tag{23}$$

Algorithm 1 Algorithm for dynamic matrix completion

Input: Partially observed matrix $Y^t \in \mathbb{R}^{n_1 \times n_2}$ for $t = 1, ..., d$, initialization matrices $U_0 \in \mathbb{R}^{n_1 \times r}$, $V_0 \in \mathbb{R}^{n_2 \times r}$ and zero matrix $C_0 \in \mathbb{R}^{n_1 \times n_2}$, parameters r, α, β, and s, weights ω_t for $t = 1, ..., d$.

1 **for** $i = 0, 1, 2, ...,$ until convergence **do**
2 $\quad U_{i+1} = U_i - \tau \nabla_U F(U_i V_i^T + C_i)$, where τ is selected via a backtracking line search using Armijo's rule with parameter β.
3 $\quad V_{i+1} = V_i - \tau \nabla_V F(U_{i+1} V_i^T + C_i)$, where τ is selected via a backtracking line search using Armijo's rule with parameter β.
4 $\quad \begin{bmatrix} U_{i+1} \\ V_{i+1} \end{bmatrix} = \mathcal{P}_\alpha^1(\begin{bmatrix} U_{i+1} \\ V_{i+1} \end{bmatrix})$.
5 $\quad C_{i+1} = C_i - \tau \nabla_C F(U_{i+1} V_{i+1}{}^T + C_i)$, where τ is selected via a backtracking line search using Armijo's rule with parameter β.
6 $\quad C_{i+1} = \mathcal{P}_\alpha^2(C_{i+1})$.
7 $\quad$ **if** $\sum_{i'j'} \mathbf{1}_{[C_{i+1 i'j'} \neq 0]} > s$ **then**
8 $\qquad C_{i+1}$ only keeps s entries with the largest absolute values. Other nonzero entries are set to be zero.
9 $\quad$ **end if**
10 **end for**
11 **Return:** $\hat{L} = \hat{U}\hat{V}^T$ and $\hat{C}$.

We next discuss the convergence analysis of Algorithm 1. Using the idea similar to that in [12], we will show that if we drop the constant constraint α on the infinity norms of the matrices L and C in both the problem formulation (5) and the algorithm (lines 4 and 6 of Algorithm 1), every sequence generated by the resulting simplified algorithm will converge to a critical point of the simplified non-convex optimization problem. Dropping the α constraint simplifies the algorithm and is practical for the cases when α is unknown. Lines 2, 3 and 5 of Algorithm 1 can be equivalently written in the form of proximal regularization as follows.

$$U_{i+1} \in \text{prox}(U_i - \tau \nabla_U F(U_i V_i^T + C_i)), \tag{24}$$

$$V_{i+1} \in \text{prox}(V_i - \tau \nabla_V F(U_{i+1} V_i^T + C_i)), \tag{25}$$

$$C_{i+1} \in \text{prox}(C_i - \tau \nabla_C F(U_{i+1} V_{i+1}^T + C_i)). \tag{26}$$

The proximal map is defined as:

$$\text{prox}(B_i - \tau \nabla_B F(B_i)) :=$$
$$\arg\min_B \{\langle B - B_i, \nabla_B F(B_i)\rangle + \frac{1}{2\tau} \|B - B_i\|_F^2 + K(C)\}, \tag{27}$$

where

$$K(C) = \begin{cases} 0, & \text{if } \sum_{ij} \mathbf{1}_{[C_{ij} \neq 0]} \leq s, \\ +\infty, & \text{if } \sum_{ij} \mathbf{1}_{[C_{ij} \neq 0]} > s. \end{cases} \quad (28)$$

$K(C)$ in (27) can be achieved by keeping s entries with the largest absolute values and setting others to be zero, which corresponds to lines 7-8 of Algorithm 1.

Our algorithm is a special case of Proximal Alternating Linearized Minimization (PALM) algorithms. The convergence of PALM algorithms have been proved in [2]. Based on Proposition 3 in [2], if we can show that $\nabla F(X)$ is Lipschitz continuous, and $F(UV^T + C) + K(C)$ satisfies the Kurdyka-Lojasiewicz (KL) property, then our algorithm converges to a critical point of (5) from every initial point. The proofs of Lipschitz continuity of $\nabla F(X)$ and the KL property of $F(UV^T + C) + K(C)$ are skipped due to the page limit. Please refer to [11] for the proofs. In practice, if α is known, we observe that the numerical performance can be improved by imposing the constraint on the infinity norm, although we cannot extend the convergence analysis to this case yet.

5. SIMULATION

We test the performance of Algorithm 1 on synthetic dataset in this section. The recovery performance is measured by the relative recovery error $\|\hat{L} - \bar{L}\|_F / \|\bar{L}\|_F$, where matrix $\bar{L}$ represents the actual data, and $\hat{L}$ represents the recovered data. The corruption rate denotes the fraction of nonzero entries in $\bar{C}$. The average erasure rate is the percentage of missing entries.

We compare our method with one convex method [15] for robust matrix completion (RMC), which solves the following convex problem

$$\min_{L,C} \frac{1}{|\Omega^d|} \|\mathcal{P}_{\Omega^d}(L + C) - Y^d\|_2^2 + \lambda_1 \|L\|_* + \lambda_2 \|C\|_1 \quad (29)$$
$$\text{s.t. } \|L\|_\infty \leq \alpha \text{ and } \|C\|_\infty \leq \alpha.$$

We set λ_1 and λ_2 to be 0.001 and 0.00015 respectively in the convex RMC method. The weights $\{\omega_t\}$ in our method are set to be $(\frac{1}{d}, ..., \frac{1}{d})$. All results are averaged over 100 runs.

We set $n_1 = 50$, $n_2 = 50$, $r = 5$ and construct $\bar{L}^t \in \mathbb{R}^{n_1 \times n_2}$ as $\bar{L}^t = \bar{U}(\bar{V}^t)^T$ where $\bar{U} \in \mathbb{R}^{n_1 \times r}$ and $\bar{V}^t \in \mathbb{R}^{n_2 \times r}$. $\bar{U}$ and $\bar{V}^1$ are matrices with i.i.d. entries drawn from standard Gaussian distribution. For all $t \geq 2$, let $\bar{V}^t = \bar{V}^{t-1} + \epsilon^t$, where matrix ϵ^t is drawn from Gaussian distribution $\mathcal{N}(0, \sigma_2^2)$. Noise matrix Z^t is drawn from Gaussian distribution $\mathcal{N}(0, \sigma_1^2)$. We first set $s = 500$, $\sigma_1 = 0.01$, and $\sigma_2 = 0.03$. Fig. 1 shows the recovery performance of the convex RMC method and our method with different d. We can see that our method performs generally better than the convex RMC method, and the performance of our method improves when d increases.

We then set $d = 3$ and keep the other simulation setup the same. Fig. 2 shows the recovery performance of the convex RMC method and our method according to different corruption rate. We can see that our method performs generally better than the convex RMC method, and the performance of our method improves when the corruption rate decreases.

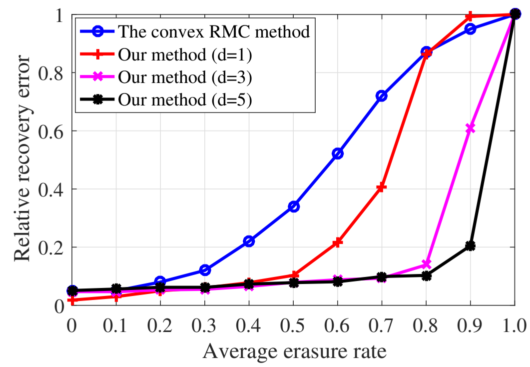

Fig. 1. Relative recovery error of the convex RMC method and our method according to different d.

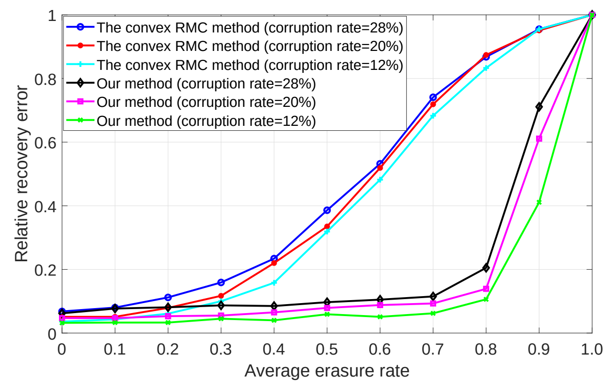

Fig. 2. Relative recovery error of the convex RMC method and our method according to different corruption rate ($d = 3$ in our method).

6. CONCLUSION AND DISCUSSIONS

This paper develops a data recovery method from partially observed and partially corrupted measurements with time-varying low dimensional structure. This work extends the existing research of robust principle component analysis to the recovery of a sequence of time-varying low-dimensional structures when part of the measurements are corrupted. 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 is developed and numerically evaluated on synthetic dataset. One future work is to test our method on the practical datasets in applications such as background subtraction in video processing. We will also study the convergence of the algorithm when the infinity norm constraint is considered.

7. ACKNOWLEDGMENT

This research is supported in part by ARO W911NF-17-1-0407, NSF #1508875 and the ERC Program of NSF and DoE under the supplement to NSF Award EEC-1041877 and the CURENT Industry Partnership Program.

8. REFERENCES

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

[2] J. Bolte, S. Shoham, and T. Marc, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.

[3] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[4] J.-F. Cai, T. Wang, and K. Wei, "Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion," *Applied and Computational Harmonic Analysis*, 2017.

[5] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[6] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

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

[8] M. Dao, L. Nguyen, and T. D. Tran, "Temporal rate upconversion of synthetic aperture radar via low-rank matrix recovery," in *Proc. IEEE International Conference on Image Processing*, 2013, pp. 2358–2362.

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

[10] P. Gao, M. Wang, J. H. Chow, S. G. Ghiocel, B. Fardanesh, G. Stefopoulos, and M. P. Razanousky, "Identification of successive "unobservable" cyber data attacks in power systems." *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5557–5570, 2016.

[11] P. Gao, R. Wang, and M. Wang, "Robust matrix completion by exploiting dynamic low-dimensional structures," 2018. [Online]. Available: http://ecse.rpi.edu/~wang/pub/SPL18.pdf

[12] P. Gao, R. Wang, M. Wang, and J. H. Chow, "Low-rank matrix recovery from noisy, quantized and erroneous measurements," *submitted to IEEE Transactions on Signal Processing*, 2017.

[13] E. C. Hall and R. M. Willett, "Online convex optimization in dynamic environments," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 647–662, 2015.

[14] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, 2013, pp. 665–674.

[15] O. Klopp, K. Lounici, and A. B. Tsybakov, "Robust matrix completion," *Probability Theory and Related Fields*, vol. 169, no. 1, pp. 523–564, Oct. 2017.

[16] N. Mohammadiha and A. Leijon, "Nonnegative hmm for babble noise derived from speech hmm: Application to speech enhancement," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 5, pp. 998–1011, 2013.

[17] N. Mohammadiha, P. Smaragdis, G. Panahandeh, and S. Doclo., "A state-space approach to dynamic nonnegative matrix factorization," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 949–959, 2015.

[18] G. J. Mysore, P. Smaragdis, and B. Raj, "Non-negative hidden markov modeling of audio with application to source separation," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2010, pp. 140–148.

[19] P. Netrapalli, U. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain, "Non-convex robust pca," in *Advances in Neural Information Processing Systems*, 2014, pp. 1107–1115.

[20] A. Sumarsono and Q. Du, "Low-rank subspace representation for estimating the number of signal subspaces in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 11, pp. 6286–6292, 2015.

[21] M. A. Veganzones, M. Simões, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral superresolution of locally low rank images from complementary multisource data," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, 2016.

[22] L. Xu and M. Davenport, "Dynamic matrix recovery from incomplete observations under an exact low-rank constraint," in *Advances in Neural Information Processing Systems*, 2016, pp. 3585–3593.

[23] J. Zhan, B. Lois, H. Guo, and N. Vaswani, "Online (and offline) robust pca: Novel algorithms and performance guarantees," in *Artificial Intelligence and Statistics*, 2016, pp. 1488–1496.