Dynamic Matrix Recovery from Partially Observed and Erroneous Measurements

Pengzhi Gao, Ren Wang, and *Meng Wang*

Department of Electrical, Computer & Systems Engineering

Rensselaer Polytechnic Institute

April 20, 2018

Outline

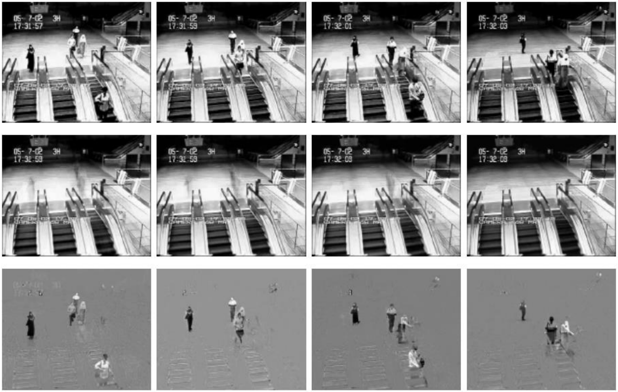
- Background
- Problem Formulation
- Theoretical Result
- Algorithm
- Simulations
- Conclusions & Future Work

Background

- Many practical datasets have intrinsic low-dimensional structures despite the high ambient dimension.
- The low dimensional structures have been extensively studied in problems like Low-Rank Matrix Completion (LRMC) and Robust Principal Component Analysis (RPCA).
- Example: well-known Netflix Prize Problem

Background

- Example: RPCA for Background Subtraction [Zou et al., 2014]



Background

- Most existing work assume that the low-dimensional structure does not change over time and consider one fixed low rank matrix. The temporal variation of the low-dimensional structure has not been much investigated.
- For example, users' preferences for the movies may change over time.

Existing Work on Time-varying Low-dimensional Models

- Parametric models like hidden Markov models [Mohammadiha et al., 2013], [Mysore et al., 2010] and autoregression models [Hall et al., 2015], [Mohammadiha et al., 2015] have been employed to model the temporal correlations and demonstrated encouraging numerical performance, but the theoretical study is very limited.
- RPCA with the weak temporal correlations was studied in [Zhan et al., 2016], and the theoretical analysis only holds when the temporal correlations of the data points are relatively weak.
- [Xu et al., 2016] proposed a model of a sequence of dynamically correlated matrices through slow varying subspaces. They assume that each matrix is low-rank, but the sparse errors are not considered in [Xu et al., 2016].

Our Contributions

- Our model generalizes from the one in [Xu et al., 2016] 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. We also propose a fast iterative method to solve the nonconvex problem approximately and show that the algorithm can always converge to a critical point, while [Xu et al., 2016] has no convergence guarantee for their algorithm.

Problem Formulation

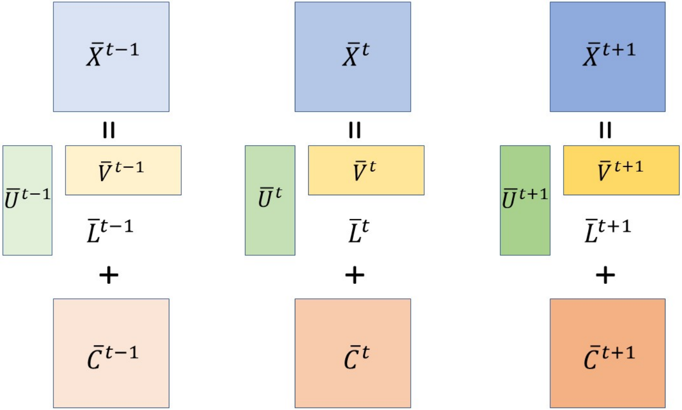
- Let $\bar{L}^t \in \mathbb{R}^{n_1 \times n_2}$ denote the actual data at time t, and let $\bar{C}^t \in \mathbb{R}^{n_1 \times n_2}$ denote the sparse additive errors in the measurements at time t.

- The temporal correlations are modeled as a sequence of low-rank matrices with correlated low-dimensional subspaces. Specifically, let $\bar{X}^t \in \mathbb{R}^{n_1 \times n_2}$ 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, $\bar{U}^t \in \mathbb{R}^{n_1 \times r}$, $\bar{V}^t \in \mathbb{R}^{n_2 \times r}$, and $\bar{C}^t$ has at most s nonzero entries.

Problem Formulation



Problem Formulation

- 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

$$y^t = \mathcal{P}_{\Omega^t}(\bar{X}^t + Z^t), \tag{2}$$

  where $\mathcal{P}_{\Omega^t} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{m^t}$ is a linear operator.

- The data recovery question is stated as follows. Given partially observed and corrupted measurements $\{y^t\}$ for $t = 1, ..., d$, can we recover the actual data $\bar{L}^d$?

Assumptions

- Assume $\|\bar{L}^t\|_\infty \leq \alpha$ and $\|\bar{C}^t\|_\infty \leq \alpha$ for some constant α.
- 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.

$$\bar{V}^t = \bar{V}^{t-1} + \epsilon^t, \ t = 2, ..., d, \tag{3}$$

where ϵ^t represents the perturbation noise in $\bar{V}^t$.

Proposed Approach

- We estimate $(\bar{L}^d, \bar{C}^d)$ by $(\hat{L}, \hat{C})$, where

$$(\hat{L}, \hat{C}) = \arg\min_{(L,C)} \frac{1}{2} \sum_{t=1}^{d} \omega_t \|\mathcal{P}_{\Omega^t}(L + C) - y^t\|_2^2, \quad \text{s.t. } L + C \in \mathcal{C}(r, s, \alpha), \tag{4}$$

and the feasible set $\mathcal{C}(r, s, \alpha)$ is defined as

$$\mathcal{C}(r, s, \alpha) := \{X \in \mathbb{R}^{n_1 \times n_2} : X = L + C, \|L\|_\infty \le \alpha, \|C\|_\infty \le \alpha, \operatorname{rank}(L) \le r, \sum_{ij} \mathbf{1}_{[C_{ij} \neq 0]} \le s\}. \tag{5}$$

- Problem (4) is nonconvex due to the nonconvexity of $\mathcal{C}(r, s, \alpha)$.

Theoretical Result

Theorem 1

If

$$m_0 \geq \frac{c_1 n_1 n_2 \log(n_1+n_2)(\sqrt{2\log(d(n_1+n_2)n_1 n_2)\sigma_{\max}^2}+2\alpha)^2}{5n_{\max}\sum_{t=1}^{d}\omega_t^2(\sigma_1^2+(d-t)\sigma_2^2)+2\alpha^2(\sqrt{2}n_{\max}+\frac{4s}{n_{\min}})}, \tag{6}$$

the estimator $(\hat{L}, \hat{C})$ *from (4) satisfies*

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

with probability at least $1-\frac{11}{n_1+n_2}-7dn_{\max}e^{-n_{\min}}$, *where* B_1 *and* B_2 *are coefficients depending on* n_1, n_2, m_0, *and* d.

Comments on Theoretical Result

Assume n_1 and n_2 are in the same order $O(n)$.

- When the measurements do not contain corruptions, i.e. $\bar{C}^t$'s are all zeros, [Xu et al., 2016] showed that if $m_0 \geq O(n(log(n))^2)$, we have

$$\|\hat{L} - L^d\|_F^2 / n_1 n_2 \leq \max(O(\sqrt{\log n / m_0}), O(n \log n / m_0)), \quad (7)$$

  when the feasible set is imposed by the rank constraint.

- If $s = O(n)$, i.e., the number of corrupted measurements per row is bounded, we have if $m_0 \geq O(n(log(n))^2)$,

$$\|\hat{L} + \hat{C} - L^d - C^d\|_F^2 / (n_1 n_2) \leq \max(B_1, B_2), \quad (8)$$

  where $B_1 = \max(O(\sqrt{\frac{\log n}{m_0}}), O(\frac{\log n}{n}))$ and $B_2 = O(\frac{n \log n}{m_0})$. Note that (8) diminishes to zero when n increases, and (8) is in the same order as the result in [Xu et al., 2016].

Comments on Theoretical Result

- If we choose $\omega_t = \frac{1}{d}$ for $t = 1, ..., d$, one can check that (6) becomes $m_0 \geq O(\frac{\log d}{d}) \times f(n)$, which implies that the required number of observations of each matrix is reduced by a factor of $O(\frac{\log d}{d})$ when d increases.

- 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.

Algorithm

- An algorithm to solve the non-convex optimization problem approximately.
- We factorize the low-rank matrix L into $L = UV^T$ with $U \in \mathbb{R}^{n_1 \times r}$ and $V \in \mathbb{R}^{n_2 \times r}$.
- In each iteration, we fix the current estimation of C and optimize over U and V by an approximate projected gradient method.
- We then fix U and V and update the estimation of C by a gradient descent method. A hard thresholding is applied to C afterwards by keeping s entries with the largest absolute values and setting others to zero.

- We also show that every sequence generated by the proximal algorithm converges to a critical point of the optimization problem.

Simulation on Synthetic Data

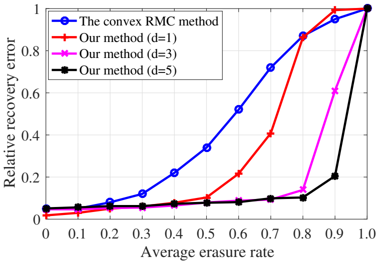
- We compare our method with one convex method [Klopp et al., 2017] for robust matrix completion (RMC), which solves the following convex problem:

$$\min_{L,C} \frac{1}{|\Omega^d|} \|\mathcal{P}_{\Omega^d}(L + C) - y^d\|_2^2 + \lambda_1 \|L\|_* + \lambda_2 \|C\|_1 \quad \text{s.t. } \|L\|_\infty \leq \alpha \text{ and } \|C\|_\infty \leq \alpha. \tag{9}$$

- The weights $\{\omega_t\}$ in our method are set to be $(\frac{1}{d}, ..., \frac{1}{d})$.
- The recovery performance is measured by the relative recovery error $\|\hat{L} - \bar{L}\|_F / \|\bar{L}\|_F$.

Simulation on Synthetic Data

- Set $n_1 = 50$, $n_2 = 50$, $r = 5$. 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 $\bar{V} \in \mathbb{R}^{n_2 \times r}$ are matrices with i.i.d. entries drawn from standard Gaussian distribution.
- For all $t \geq 2$, $\bar{V}^t = \bar{V}^{t-1} + \epsilon^t$, where matrix ϵ^t is drawn from Gaussian distribution $\mathcal{N}(0, \sigma_2^2)$. Noise matrix Z^t is drawn from Gaussian distribution $\mathcal{N}(0, \sigma_1^2)$.
- Set $s = 500$, $\sigma_1 = 0.01$, and $\sigma_2 = 0.03$.



Simulation on Synthetic Data

- Set $d = 3$ and keep the other simulation setup the same.

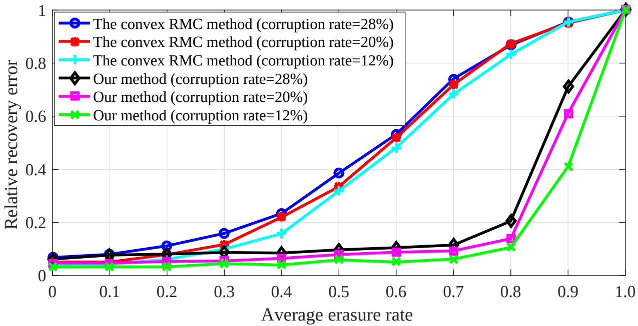


Figure: Relative recovery error of the convex RMC method and our method according to different corruption rate ($d = 3$ in our method).

Simulation on Actual Video Dataset

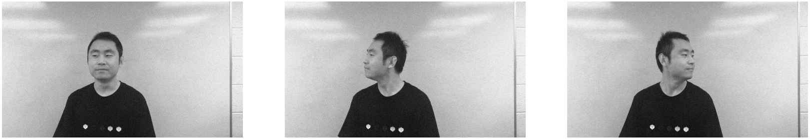


Figure: Video frames only including the first RGB value with the additional Gaussian noise drawn from $\mathcal{N}(0, 0.04^2)$.

- Choose 180 frames and construct a 2073600×180 matrix, where each columns represents a frame and 60 columns make up the matrix $\bar{L}^t$.
- Form $\bar{L}^1$, $\bar{L}^2$, and $\bar{L}^3$ by reducing the dimension to $\bar{L}^t \in \mathbb{R}^{270 \times 60}$ for $t = 1, 2, 3$. Set $r = 5$.

Simulation on Actual Video Dataset

- Under different corruption rates (5%, 10%, 15%, 20%), we run our algorithm with different d.

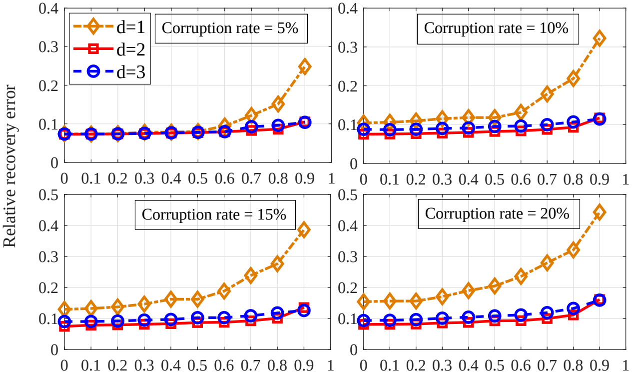


Figure: Relative recovery error of our method according to different d.

Simulation on Actual Video Dataset

- Set $d = 2$ and keep the other simulation setup the same.

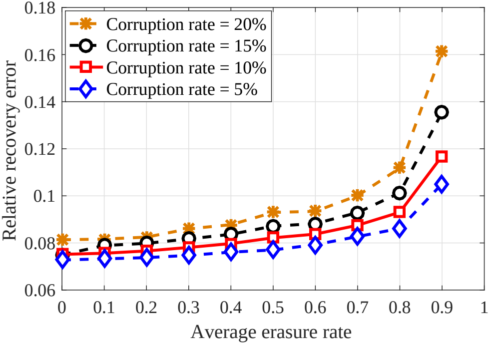


Figure: Relative recovery error according to different corruption rate ($d = 2$).

Conclusions & Future Work

- We study the dynamic matrix recovery problem from partially observed and erroneous measurements.
- The dynamic matrix recovery problem is formulated as a non-convex optimization problem.
- 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 guanrantee is proposed to solve the non-convex problem approximately.
- One future direction is to study the global convergence of the algorithm.

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.

Q & A

References

Pengzhi Gao, Ren Wang, and Meng Wang (2018)
Robust Matrix Completion by Exploiting Dynamic Low-dimensional Structures [Online]
Available: http://ecse.rpi.edu/∼wang/pub/SPL18.pdf.

Liangbei Xu, and Mark A. Davenport (2016)
Dynamic Matrix Recovery from Incomplete Observations under An Exact Low-rank Constraint
30th Conference on Neural Information Processing Systems (NIPS 2016).

Olga Klopp, Karim Lounici, and Alexandre B. Tsybakov (2017)
Robust Matrix Completion
Probability Theory and Related Fields 169(1 – 2), 523 – 564.

References

Nasser Mohammadiha, and Arne Leijon (2013)
Nonnegative HMM for Babble Noise Derived From Speech HMM: Application to Speech Enhancement
IEEE Transactions on Audio, Speech, and Language Processing 21 (5), 998 – 1011.

Gautham J. Mysore, Paris Smaragdis, and Bhiksha Raj (2010)
Non-negative Hidden Markov Modeling of Audio with Application to Source Separation
International Conference on Latent Variable Analysis and Signal Separation.

Eric C. Hall, and Alexandre B. Tsybakov (2015)
Online Convex Optimization in Dynamic Environments
IEEE J. Sel. Topics Signal Process. 9(4), 647 – 662.

References

Nasser Mohammadiha, Paris Smaragdis, and Ghazaleh Panahandeh (2015)
A State-Space Approach to Dynamic Nonnegative Matrix Factorization
IEEE Trans. Signal Process. 63 (4), 949 – 959.

Jinchun Zhan, Brian Lois, Han Guo, and Namrata Vaswani (2016)
Online (and Offline) Robust PCA: Novel Algorithms and Performance Guarantees
19th International Conference on Artificial Intelligence and Statistics.

Xiaowei Zou, Can Yang, Hongyu Zhao, and Wenchuan Yu (2014)
Low-Rank Modeling and Its Applications in Image Analysis
ACM Computing Surveys (CSUR) 47 (2).

Zhouchen Lin (2016)
A Review on Low-rank Models in Data Analysis
Big Data & Information Analytics 1(2/3): 139-161.

Theoretical Result

$$\sigma_{\max}^2 = \max_t \omega_t^2 (\frac{\mu_0^2 r}{n_1} \sigma_2^2 (d-t) + \sigma_1^2), B_1 = 16\alpha^2 \max(\sqrt{c_2 \frac{\log(n_1+n_2)}{m_0 \log(6/5)}}, \frac{\log(n_1+n_2)}{2n_1 \log(6/5)}),$$

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

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

and c_1, c_2 and c_3 are constants.