ROBUST RECOVRY OF HIGHLY CORRUPTED LOW RANK MATRIX

Tuersunjiang YIMAMU [1] and Toshio EISAKA [2]

[1] Dept. of Manufacturing Eng., Kitami Inst. of Tech
(Koencho 165, Kitami, Hokkaido 090-8507, Japan)
E-mail: tuteu@yahoo.com
[2] Professor, Div. of Information and Communication Eng., Kitami Inst. of Tech
(Koencho 165, Kitami, Hokkaido 090-8507, Japan)
E-mail: eisaka@mail.kitami-it.ac.jp

Usually, real-world observed data are sampled from low-dimensional subspaces with sample-specific corruptions (known as outliers) and random noises. The task of recovering low rank matrix from its noisy measurement plays a central role in data science. Our goal is to recover low rank component from highly corrupted data, and to correct the possible errors as well. In this paper, we propose a new method termed Low Rank Matrix Estimation (LRME) by augmented Lagrangian multiplier, which seeks the low rank component from highly corrupted observation. Our algorithm is to reduce not only the outliers, but also the random corruptions. Theoretical analysis on convergence and optimality, and experimental result on synthetic data are provided to demonstrate the efficacy of our proposed method.

Key Words: *Outlier Estimation, Low Rank Matrix Recovery, Matrix Factorization, Low Rank Matrix Completion*

1. INTRODUCTION

Recovering a low rank matrix from its noisy measurements plays a central role in computational science and large-scale data analysis. The superiority of Low Rank Matrix Recovery (LRMR) has been tested and applied in many research areas, such as covariance matrix estimation [1], [2], robust principal component analysis [3], [4], collaborative filtering [5], signal and image processing [6],[7], photometric stereo [8]. In those areas, the data to be analyzed often have high dimensionality, which brings great challenges to data analysis. Fortunately, the high dimensional data are observed to have low intrinsic dimension.

In order to recover the low rank component matrix$L \in \mathbb{R}^{m\times n}$ from the given observation matrix $Y \in \mathbb{R}^{m\times n}$ which highly corrupted by error $E \in \mathbb{R}^{m\times n}$. It is straightforward to consider the following regularized rank minimization problem :

$$\min_{L,E} \ rank(L)+\lambda\|E\|_p \ \ s.t. \ \ P_\Omega(Y) = P_\Omega(L+E) \qquad (1)$$

where $\|E\|_p$ indicates certain regularization strategy, such as squared Frobenius norm $\|\cdot\|_F^2$ used for modeling the random noise [9], whereas the l_0 norm adopted by [4] for modeling the outlier. The $\lambda > 0$ is a regularization parameter balancing the tolerance in training error. Furthermore, $p_\Omega(\cdot)$ is orthogonal projection operator onto the linear space of matrices on the support of binary valued $\Omega_{ij} \in \{0,1\}^{m\times n}$.

As one of the two pivotal factors in LRMR, the rank constraint in (1) makes the problem combinatorically hard [10], due to its discreteness. Other widely using low rank matrix approximation is Nuclear Norm Minimization (NNM) [4], [11]. The nuclear norm of a matrix L is denoted by $\|L\|_*$, i.e., the sum of its singular values. However, their applicability is often limited by the necessity of executing expensive Singular Value Decomposition (SVD) for multiple times [12].

Another widely using scheme is Matrix Factorization (MF) method, i.e., the Bilinear Factorization (BF) [13], [14]. The BF method aims to find two smaller low rank factor matrices $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{r\times n}$ whose product is equal to low rank matrix $L \in \mathbb{R}^{m\times n}$,e.g.,$L = UV$, the rank of approxi-

mation typically satisfies $r \ll \min(m,n)$. Although, BF method has computational superiority when the target rank is given, it is not necessarily the best choice when the target rank is unknown.

In the present paper, we design new method to solve LRMR problem with outlier. Unlike existing robust LRMR approaches, our method takes into account not only additive outlier or missing data, but also incomplete and random corrupted measurement. The key contributions of this paper are summarized as follows:

1. We used general Tikhonov type of regularization with Frobenius norm penalty to obtain stable approximate solutions to LRMR problems by means of a stabilizing functional.
2. We used maximum entropy term to model the discreteness problem and increase our prior knowledge (weights) on missing data.
3. Finally, we perform experiments on both synthetic and real data and demonstrate its correctness, efficiency, and effectiveness.

2. PROBLEM FORMULATION

A low rank matrix $L \in \mathbb{R}^{m\times n}$ can be recovered from noisy measurements $Y \in \mathbb{R}^{m\times n}$ via the following nuclear norm minimization model:

$$\min_{L} \ \lambda\|L\|_* + \tfrac{1}{2}\|\Omega\odot(Y-L)\|_F^2 \qquad (2)$$

where $\odot$ is Hadamard product, the Ω is index set of observed entries, the $\lambda \geq 0$ is regularization parameter controlling the nuclear norm of the minimizer. In large data case, to mitigate the computational pressure, Theorem 1 is a best choice between NNM and bilinear factorization:

Theorem 1: For any matrix $L \in \mathbb{R}^{m\times n}$, the following relationship holds [10]:

$$\|L\|_* = \min_{U,V} \frac{1}{2}\|U\|_F^2 + \frac{1}{2}\|V\|_F^2 \qquad s.t. \quad L = UV$$

if $\mathrm{rank}(L) \leq \min(\mathrm{m},\mathrm{n})$, then the minimum solution above is at a factor decomposition $L=UV$, where $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{r\times n}$.

As a result, applying Theorem 1 on (2), we obtain the following Tikhonov type of regularization which impose smoothness on the solution for ill conditioned problems.

$$\min_{U,V} \ \tfrac{\lambda}{2}\|\Gamma_U U\|_F^2 + \tfrac{\lambda}{2}\|V\Gamma_V\|_F^2 + \tfrac{1}{2}\|\Omega\odot(Y-UV)\|_F^2 \quad (3)$$

For the purpose of increase applicability of our algorithm, we use Generalized Tikhonov regularization, and our problem becomes the form:

$$\begin{aligned} \min_{U,V} \ & \frac{\lambda}{2}\|\Gamma_U U\|_F^2 + \frac{\lambda}{2}\|V\Gamma_V\|_F^2 + \frac{1}{2}\|\Omega\odot(Y-UV)\|_F^2 \\ s.t. \ & z_U{}^T\Gamma_U z_U > 0 \ \ and \ \ z_V{}^T\Gamma_V z_V > 0; \\ & \forall \ z_U \in \Gamma_U \ \ and \ \ z_V \in \Gamma_V \end{aligned} \qquad (4)$$

where Γ_U and Γ_V are used to enforce certain application dependent characteristics of the solution. If Γ is equal to identity matrix I, (3) and (4) are same problem in standard form, however, these matrices are typically $\Gamma \neq I$ or discrete approximations to some derivative operators. Compared to directly minimizing (2), (4) inherits some advantages and avoids overfitting when the target rank larger than ground truth rank.

In practice, we may combine several regularization terms to satisfy specific constraints together. To enforce sparsity, we can use L_1-norm regularization terms and combine them. It is worth noting that L_1-norm also used for modeling the outlier.

$$\begin{aligned} \min_{U,V} \ & \tfrac{\lambda}{2}\|\Gamma_U U\|_F^2 + \tfrac{\lambda}{2}\|V\Gamma_V\|_F^2 + \tfrac{1}{2}\|\Omega\odot(Y-UV)\|_F^2 \\ & +\beta_U\|U\|_1 + \beta_V\|V\|_1 \\ s.t. \ & z_U\Gamma_U z_U{}^T > 0 \ \ and \ \ z_V\Gamma_V z_V{}^T > 0; \\ & \forall \ z_U \in \Gamma_U \ \ and \ \ z_V \in \Gamma_V \end{aligned} \qquad (5)$$

where $\beta_U \geq 0$ and $\beta_V \geq 0$ parameters controlling the degree of sparsity of the matrices.

In many situations, the data are polluted by, besides random noises, sample-specific corruptions, which may prevent our algorithm efficacy. To reduce the negative effect of such pollution, we used real valued weight matrix, such as: $(W \ \ and \ \ \bar{W}) \in [0,1]^{m\times n}$. Real valued weight matrices are required to tell which elements are polluted by random noises W and which by sample-specific corruptions $\bar{W}$. We employ maximum entropy term to combine with real valued weight matrices as probability of current state of knowledge. The definition of entropy is $-\sum_{i=1}^{k} p_i log p_i$ with $\sum_{i=1}^{k} p_i = 1$. Maximizing the entropy is equivalent to minimizing its negative log function. Finally, our problem can be formulated as follows:

$$\begin{aligned} \min_{U,V} \ & \frac{\lambda}{2}\|\Gamma_U U\|_F^2 + \frac{\lambda}{2}\|V\Gamma_V\|_F^2 + \frac{1}{2}\|W\odot(Y-UV)\|_F^2 \\ & +\beta_U\|U\|_1 + \beta_V\|V\|_1 + \gamma\sum_{ij}(w_{ij}log w_{ij} + \bar{w}_{ij}log\bar{w}_{ij}) \\ s.t. \ & W + \bar{W} = 1; \ \ W \ \ and \ \ \bar{W} \ \in [0,1]^{m\times n} \\ & z_U\Gamma_U z_U{}^T > 0 \ \ and \ \ z_V\Gamma_V z_V{}^T > 0; \\ & \forall \ z_U \in \Gamma_U \ \ and \ \ z_V \in \Gamma_V \end{aligned} \qquad (6)$$

where the $\gamma \geq 0$ is non-negative coefficient and 1

represents an all-one matrix with comparable size. w_{ij} and $\overline{w}_{ij}$ are represent (ij)-th elemet of W, $\overline{W}$, respectively. The support Ω is replaced by a weight matrix W containing both the given support and the estimated outlier support.

3. METHOTHOLOGY

(1) Optimization

Recently, it has been shown in the literature [15], [16] that alternating direction method of multipliers (ADMM) is very efficient for some convex or nonconvex programming problem. To apply ADMM on our problem, the objective function is required to be separable. Hence, we introduce the constraint $L = UV$ into (6), and obtain the following equivalent problem formulation:

$$\begin{aligned} \min_{U,V} \ & \frac{\lambda}{2}\|\Gamma_U U\|_F^2 + \frac{\lambda}{2}\|V\Gamma_V\|_F^2 + \frac{1}{2}\|W\odot(Y-L)\|_F^2 \\ & + \beta_U\|U\|_1 \\ & + \beta_V\|V\|_1 + \gamma\sum_{ij}(w_{ij}logw_{ij} + \overline{w}_{ij}log\overline{w}_{ij}) \\ s.t. \ \ & L = UV,\ W + \overline{W} = 1;\ W\ and\ \overline{W} \in [0,1]^{m\times n};\ z_U\Gamma_U{z_U}^T > 0\ and\ z_V\Gamma_V{z_V}^T > 0; \\ & \forall\ z_U \in \Gamma_U\ and\ z_V \in \Gamma_V \end{aligned} \quad (7)$$

The Augmented Lagrangian Multiplier of (7) is defined as:

$$\begin{aligned} & \mathcal{L}_{\{W+\overline{W}=1;W,\overline{W}\in[0,1]^{m\times n}\}}(U,V,L,W,M) := \\ & \frac{\lambda}{2}\|\Gamma_U U\|_F^2 + \frac{\lambda}{2}\|V\Gamma_V\|_F^2 + \frac{1}{2}\|W\odot(Y-L)\|_F^2 \\ & + \beta_U\|U\|_1 + \beta_V\|V\|_1 + \gamma\sum_{ij}(w_{ij}logw_{ij} + \overline{w}_{ij}log\overline{w}_{ij}) \\ & + \frac{\alpha}{2}\|L-UV\|_F^2 + \langle M, L-UV\rangle \end{aligned} \quad (8)$$

where $\langle\cdot\rangle$ denotes the inner product between two matrices of equal size, the $\alpha > 0$ is positive penalty parameter, and $M \in \mathbb{R}^{m\times n}$ is a Lagrangian multipliers. Notice that the constraints on $W, \overline{W}$ and Γ are enforced as hard constraints. The solver updates the variables in an iterative manner. We will derive our scheme for solving the following subproblems with respect to U, V, L and $W, \overline{W}$, respectively:

$$\begin{aligned} V^{(t+1)} &\leftarrow \left(\lambda\Gamma_V^T\Gamma_V + \alpha^{(t)}U^{(t)T}U^{(t)}\right)^{-1}(\alpha U^{(t)T}L^{(t)} \\ &\quad + U^{(t)T}M^{(t)} - \frac{\beta}{2}) \\ U^{(t+1)} &\leftarrow [(\alpha^{(t)}L^{(t)}V^{(t+1)T} - \frac{\beta}{2} + M^{(t)}V^{(t+1)T}) \\ &\quad \left(\lambda\Gamma_U\Gamma_U^T + \alpha^{(t)}V^{(t+1)}V^{(t+1)T}\right)^{-1}] \\ L^{(t+1)} &\leftarrow \frac{W^{(t)}\odot Y + \alpha^{(t)}U^{(t+1)}V^{(t+1)} - M^{(t)}}{W^{(t)} + \alpha^{(t)}I} \end{aligned} \quad (9)$$

We update U, V, L via equating the derivatives of (8) with respect to U, V and L to zero respectively with the other variables fixed. The division in updating L is elementwise.

We are picking out the terms related to $W, \overline{W}$ and the result in the following optimization problem:

$$\begin{aligned} \min_{W,\overline{W}} & \frac{1}{2}\left\|W\odot\left(Y - L^{(t+1)}\right)\right\|_F^2 \\ & + \gamma\sum_{ij}(w_{ij}logw_{ij} + \overline{w}_{ij}log\overline{w}_{ij}) \\ s.t. & W + \overline{W} = 1;\ W\ and\ \overline{W} \in [0,1]^{m\times n} \end{aligned} \quad (10)$$

This problem is also separable, thus, we can solve this problem without loss of generality. The Lagrange Multiplier gives the following Lagrange function:

$$\begin{aligned} & Q(w_i, \overline{w}_i, \lambda_i) := w_{ij}[Y - L^{(t+1)}]_{ij}^2 + \\ & \gamma\left(w_{ij}logw_{ij} + \ \overline{w}_{ij}log\overline{w}_{ij}\right) + \lambda_i(w_{ij} + \overline{w}_{ij} - 1) \end{aligned} \quad (11)$$

where subscript ij denotes that (i,j)-th element, the λ_i is Lagrange multiplier. Taking the derivative of $Q(w_i, \overline{w}_i, \lambda_i)$ to $w_i, \overline{w}_i$ and λ_i respectively and setting them to zero. We obtain the following optimal solution of w_i:

$$w_{ij}^{(t+1)} \leftarrow \frac{1}{1+\exp\{[(Y-L^{(t+1)})_{ij}^2/2]/\gamma\}} \quad (12)$$

which is standard sigmoid function form. Its complementary is equal to $\overline{w}_{ij}^{(t+1)}$.

$$\overline{w}_{ij}^{(t+1)} \leftarrow 1 - w_{ij}^{(t+1)} \quad (13)$$

We updated the Lagrange multipliers M and α via:

$$M^{t+1} \leftarrow M^t + \alpha^t(L^{(t+1)} - U^{(t+1)}V^{(t+1)}) \quad (14)$$

$$\alpha^{t+1} \leftarrow \alpha^t\rho,\ \rho > 1 \quad (15)$$

It is worth noting that the Tikhonov matrices Γ_U and Γ_V requires no iteration. We used $\Gamma_U \leftarrow \varepsilon_U I$ and $\Gamma_V \leftarrow \varepsilon_V I$ to setting them. ε_U and ε_V are non-negative parameters.

For clarity, the customized solver to (6) is outlined in Algorithm 1. The procedure should not be terminated until the equality constraint $L = UV$ is satisfied up to a given tolerance, that is $\|L - UV\|_F \le \phi\|Y\|_F$. We chose the tolerance factor ϕ is 1e-7 in all our experiment.

(2) Complexity Analysis

The computational cost of the existing algorithms for nuclear norm regularization is dominated by the

computation of SVD in each iteration. The complexity of SVD for a matrix of size $M \times N$ is $O(MN\ \min(MN))$. In contrast, our algorithm does not need to compute SVD, and its computational cost is mainly spent on matrix multiplication for updating U, V and wight matrix W with complexity $O(NlogN)$. Our algorithms have much lower computational complexity than the existing algorithms. Furthermore, our algorithms converge very fast, much more reducing the computational time.

Algorithm 1: solving LRMR problem (6)

Input: $Y \in \mathbb{R}^{m\times n}$, $\Omega \in \{0,1\}^{m\times n}$, rank r, and parameters (α, β, γ)

Output: $U^*, V^* or\ L^*$ and W^*

Initialize: $U^{(0)} \in \mathbb{R}^{m\times r}$ and $V^{(0)} \in \mathbb{R}^{r\times n}$, $L^{(0)} \in \mathbb{R}^{m\times n}$ $W^{(0)} \in \mathbb{R}^{m\times n} \leftarrow \Omega \odot 1$ are initialized randomly; $t \leftarrow 0$

1. **While** not converged **do:**
2. Update $V^{(t+1)}$, $U^{(t+1)}$, $L^{(t+1)}$ via (9)
3. **For** $\forall(ij)$ **do:**
4. Update $w_{ij}^{(t+1)}$ via (12)
5. **End For**
6. Update the multiplier M^{t+1}, α^{t+1} via (14),(15)
7. $t \leftarrow t + 1$
8. **End while**

4. EXPERIMENTA L VERIFICATION

In this section, we perform extensive experiments on LRMR to evaluate the effectiveness and efficiency of our method. We compare our method with state-of-the-art LRMR methods. The compared methods include ROUTE [17], Active [18], AIS-Impute [19], IRNN [20], RegL1[23], PCP [4], PRMF [21], Unify [22]. The codes of which are downloaded from the author's websites. Their settings follow the suggestions by the author or the given parameters.

We generate a synthetic low-rank matrix Y as a product of two matrices, such that $Y = UV$. The rank of matrix satisfies $r \leq \min(m, n)$. Both matrices U and V are randomly created by independent identically distributed (i.i.d.) samples from the Gaussian Distribution $\mathcal{N}(0,1)$. Then we corrupted the entries via replacing a fraction outlier ratio (s) of Y with errors drawn from a uniform distribution over [-25, 25] at random, and the rest entries are polluted by Gaussian noise $\mathcal{N}(0, \sigma^2)$. We have used root mean square error (RMSE) and mean absolute error (MAE) to measure the recovery performance. In all the experiments, we fix the noise level $\sigma = 0.1$ and set the dimension of the matrix larger than the ground-truth matrix rank. We repeat the experiments 20 times and calculate the average accuracy and CPU time.

We first test RMSE when the outlier ratio s changes from 0.1 to 0.7 for all algorithms. In this test we fix the data size m=n=500, rank=10 estimation, each graph generated by averaging 20 independent runs. From Fig. 1, we can see that our Low Rank Matrix Estimation (LRME) method consistently outperforms the compared algorithms in RMSE.

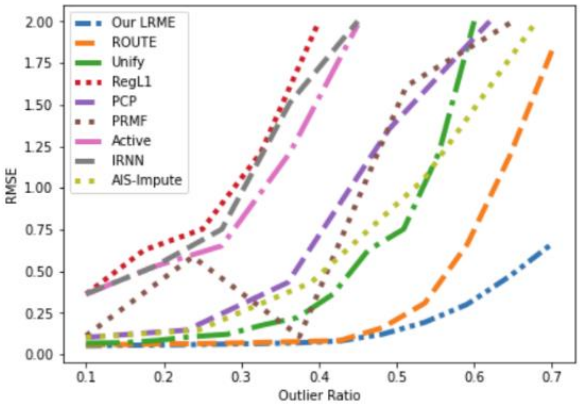

Fig.1 Performance comparison between RMSE and different Outlier ratios *s*.

In Fig.2, we compare the tolerance to outlier between ROUTE and our LRME algorithm. We fix the size of data m=n=600 and test the tendency by changing outlier ratio s [0-0.7] and rank r [20, 40, 60].

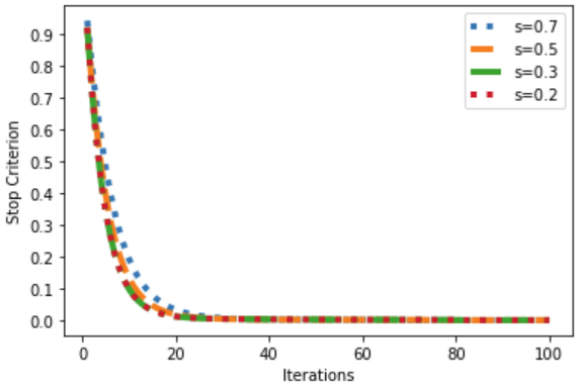

Fig.2 Convergence speed in the different outlier ratios s.

Fig. 3 shows that at the beginning, the RMSE of our LRME algorithm smaller than ROUTE method. When the outlier ratio increases, the margin between two algorithm enlarges in the case of r=60 and s=0.7, our LRME algorithm also stays lower RMSE than ROUTE method.

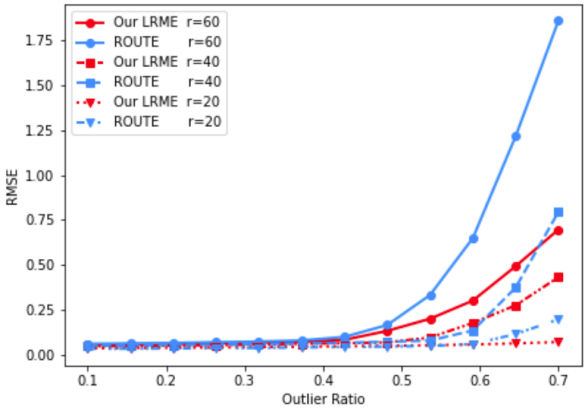

Fig.3 Performance comparison between Our LRME and ROUTE for different outlier ratio and ranks.

5. CONCLUSION

In this paper, we recommended a new method for low rank matrix recovering problem. We can achieve a better control of the target rank of the low rank component, even when the observed data is limited and highly corrupted. In experimental analysis on both syndetic and real dataset, we further showed that our recommended method can easily construct LRMR problem. Our method is much less time demanding and can converge quickly in a few iterations. The generality and the effectiveness of our method are supported through numerous and extensive experiment, and the results verified that our LRMR algorithm is an effective and efficient approach for low rank matrix recovery, especially for large-scale problems.

6. REFERENCES

1) Y. Chi, Y. Lu, Y. Chen "Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview," *IEEE Trans. Signal process.*, vol. 67, no. 20, pp. 5239–5269, Oct. 2019.
2) Y. Chen, Y. Chi, A. J. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *IEEE Trans. Inform. Theory.*, vol. 61, no. 7, pp. 4034–4059, Mar. 2015.
3) N. Tsagkarakis, P.P. Markopoulos,G. Sklivanitis, D.A. Pados, "L1-norm principal-component analysis of complex data," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3256–3267, Jun. 2018.
4) E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM.*, vol. 58, no. 1, pp. 1–37, 2009.
5) Q. Zhang and H. Wang, "Collaborative filtering with generalized Laplacian constraint via overlapping decomposition," *Proc. IJCAI.*, pp. 2329–2335, Jul. 2016.
6) H. Wang, Y. Li, Y. Cen and Z. He, "Multi-Matrices Low-Rank Decomposition With Structural Smoothness for Image Denoising," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 2, pp. 349–361, Feb. 2020.
7) H. Zhang ,W. He , L. Zhang ,H. Shen ,Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE TGRS.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
8) R. Basri, D. W. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *Int. J. Comput. Vis.*, vol. 72, no. 3, pp. 239–257, May 2007.
9) E. Candès, Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no.6, pp. 925–936, Jun. 2010.
10) R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *J. Mach. Learn. Res.*, vol. 99, pp. 2287–2322, Aug. 2010.
11) B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Aug.2010.
12) X. Guo and Z. Lin, "ROUTE: Robust outlier estimation for low rank matrix recovery," *Proc. IJCAI.*, pp. 1746–1752, 2017.
13) F. Shang, Y. Liu, H. Tong, J. Cheng, "Robust bilinear factorization with missing and grossly corrupted observations," *Inform. Sci.*, vol. 307, pp. 53–72, Jun. 2015.
14) B. Lakshminarayanan, G. Bouchard, C. Archambeau, "Robust Bayesian matrix factorisation," *Proc. AISTATS.*, vol. 15, pp. 425–433, Apr. 2011.
15) J. Yang, X. Yuan, "Linearized augmented lagrangian and alternating direction methods for nuclear nor minimization," *Math. Prog. Comput.*, vol. 4 pp. 333–361, 2012.
16) S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends. mach. Learn.*, vol. 3 no.1 pp. 1-122, 2011.
17) X. Guo, Z. Lin, "Low rank matrix recovery via robust outlier estimation," *IEEE Trans. Image Process.*, vol. 27, no. 11, Nov., 2018.
18) C. J. Hsieh and P. Olsen, "Nuclear norm minimization via active subspace selection," *Proc. Int. Conf. Mach. Learn.*, vol. 32 pp. 575–583,2014.
19) Q. Yao and J. T. Kwok, "Accelerated inexact soft-impute for fast largescale matrix completion," *Proc. 24th Int. Joint Conf. Artif. Intell.*, pp. 4002–4008, Jul. 2015.
20) C. Lu, J. Tang, S. Yan, and Z. Lin, "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 829–839, Feb. 2016.
21) N. Wang, T. Yao, J. Wang, and D. Yeung, "A probabilistic approach to robust matrix factorization," *Proc. ECCV.*, pp. 126–139, 2012.
22) R. Cabral, F. De la Torre, J. Costeira, and A. Bernardino, "Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition," *Proc. ICCV.*, pp. 2488–2495, Dec. 2013.
23) Y.zheng, G. Liu, S. Sugimoto, S. Yan, M. Okutomi, "Practical low-rank matrix approximation under robust L1-norm," *Proc. CVPR.*, pp. 1410–1417,2012