Low Rank Matrix Recovery from Sparse Noise by $\ell_{2,1}$ Loss Function

Ji Li, Lina Zhao, Meiling Zhang and Xuke Hou
Department of Mathematics, College of Science
Beijing University of Chemical Technology
Bei Jing, China
e-mail: zhaoln@mail.buct.edu.cn

***Abstract*—In the last decades, Robust Principal Component Analysis (PCA) has been drawn much attention in the image processing, computer vision and machine learning communities and various robust PCA methods have been developed. This paper introduces a new generalized robust PCA with emphasizing on $\ell_{2,1}$ - $norm$ minimization on loss function. The $\ell_{2,1}$ - $norm$ instead of Frobenius norms based loss function is robust to outliers in data points. An efficient algorithm combine augmented Lagrange multiplier is develops. The experiments on both numerical simulated data and benchmark picture demonstrate that the proposed method outperforms the state-of-the-art because our method needs less iteration and more robust to outliers in data points.**

Keywords-robust PCA; $\ell_{2,1}$ - $norm$; loss function.

I. INTRODUCTION

Robust Principal Component Analysis considers an idealization of the robust PCA problem [1][4], in which the goal is to recover a low-rank matrix A from highly corrupted measurements D = A + E. The errors E can be arbitrary in magnitude, but are assumed to be sparsely supported, affecting only a fraction of the entries of D. This should be contrasted with the classical setting in which the matrix A is perturbed by small (but densely supported) noise. In that setting, classical PCA computed via the singular value decomposition, remains optimal if the noise is Gaussian. Here, on the other hand, even a small fraction of large errors can cause arbitrary corruption in PCA' s estimate of the low rank structure, A. When the elements of the matrix E is subject to independent and identically distributed Gaussian distribution, the available classical PCA to achieve optimum matrix A, namely solving the following optimization problem:

$$\min_{A,E} \|E\|_F \quad s.t. rank(A) \le r, D = A + E \tag{1}$$

For matrix D, using SVD can obtain the optimal solution. But when the E is a big noise spare matrix, PCA is no longer applicable. So the problem becomes a double objective optimization problem:

$$\min_{A,E} (rank(A), \|E\|_0) \quad s.t. D = A + E \tag{2}$$

By introducing a parameter λ , the dual-objective optimization problem converted to single-objective optimization problem:

$$\min_{A,E} rank(A) + \lambda \|E\|_0 \quad s.t. A + E = D \tag{3}$$

This problem is NP problem; therefore, the objective function of this problem needs to be relaxed. It can become that [5]:

$$\min_{A,E} \|A\|_* + \lambda \|E\|_{1,1} \quad s.t. A + E = D \tag{4}$$

There are some main algorithms for solving this problem, such as Iterative Thresholding (IT) [1], Accelerated Proximal Gradient (APG) [6] and Augmented Lagrange Multiplier (ALM) [7]. In this paper, we propose a new robust PCA with emphasizing on $\ell_{2,1}$ - $norm$ [8][9]minimization on loss function. Instead of using F - $norm$ based loss function that is sensitive to outliers, a $\ell_{2,1}$ - $norm$ based loss function is adopted in our work to remove outliers. To solve this new robust PCA, we propose an efficient algorithm.

II. ROBUST PCA WITH $\ell_{2,1}$ - $norm$

In this section, we introduce a novel robust PCA with $\ell_{2,1}$ - $norm$ based loss function, it instead of using F - $norm$ based loss function.

A. Preliminary

Throughout this paper, we use bold uppercase characters to denote matrices, bold lowercase characters to denote vectors. For any matrix A, a_i means the i-th column vector of A, a^i means the i-th row vector of A, A_{ij} denotes the (i,j)-element of A and Tr[A] is the trace of A if A is square. A^T denotes the transposed matrix of A. We define for $q \ge 1$, the

978-1-5090-2377-6/16/$31.00 ©2016 IEEE

ℓ_q-norm of a vector $a \in \Re^m$ as $\|a\|_q = (\sum_{i=1}^{m} |a_i|^q)^{1/q}$. We consider the Frobenius norm of a matrix $A \in \Re^{m\times n}$:

$$\|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2 = Tr[A^T A] \tag{5}$$

The $\ell_{2,1}$-norm for A is defined as [10]

$$\|A\|_{2,1} = \sum_{i=1}^{m}\sqrt{\sum_{j=1}^{n} A_{ij}^2} = 2Tr[A^T DA] \tag{6}$$

where D is a diagonal matrix with $D_{ii} = \dfrac{1}{2\|a^i\|_2}$.

The general method of augmented Lagrange multipliers is introduced for solving constrained optimization problems of the kind [7]:

$$\begin{aligned} &\min f(X) \\ &\textit{subject to} \;\; \mathrm{h(X)} = 0 \end{aligned} \tag{7}$$

where $f:\Re^n \to \Re$ and $h:\Re^n \to \Re^m$. One may define the augmented lagrangian function:

$$L(X,Y,u) = f(X) + \langle Y, h(X)\rangle + \frac{\mu}{2}\|h(X)\|_F^2 \tag{8}$$

where μ is a positive scalar, and then the optimization problem can be solved via the method of augmented Lagrange multipliers.

For the RPCA problem, we may apply the augmented Lagrange multiplier method by identifying:

$$X = (A,E), f(X) = \|A\|_* + \lambda\|E\|_1, \textit{and} \;\; h(X) = D - A - E. \tag{9}$$

Then the Lagrangian function is:

$$L(A,E,Y,\mu) = \|A\|_* + \lambda\|E\|_1 + \langle Y, D-A-E\rangle + \frac{\mu}{2}\|D-A-E\|_F^2 \tag{10}$$

In this paper, $\ell_{2,1}$-norm instead of using F-norm based loss function that is sensitive to outliers:

$$L(A,E,Y,\mu) = \|A\|_* + \lambda\|E\|_1 + \langle Y, D-A-E\rangle + \frac{\mu}{2}\|D-A-E\|_{2,1} \tag{11}$$

B. Formulation

With $\ell_{2,1}$-norm, we can formulate the robust PCA problem as follows:

$$L(A,E,Y,\mu) = \|A\|_* + \lambda\|E\|_1 + \langle Y, D-A-E\rangle + \frac{\mu}{2}\|D-A-E\|_{2,1}$$

From the preliminary, (11) can rewrite as follows:

$$\begin{aligned} L(A,E,Y,\mu) = {} & \|A\|_* + \lambda\|E\|_1 + \mu Tr[(D-A-E)^T(\mathrm{diag}(\frac{1}{2\|(D-A-E)^i\|_2}))(D-A-E)] \\ & + Tr[Y^T(D-A-E)] \end{aligned} \tag{12}$$

Then, from [11] we have that when update A:

$$\begin{aligned} \frac{\partial L(A,E,Y,\mu)}{\partial A} &= \frac{\partial\|A\|_*}{\partial A} - 2\mu G(D-A-E) - Y \\ &= \frac{G^{-1}}{2\mu}\frac{\partial\|A\|_*}{\partial A} - D + A + E - \frac{G^{-1}}{2\mu}Y = 0 \\ &= \frac{G^{-1}}{2\mu}\frac{\partial\|A\|_*}{\partial A} + A - (D - E + \frac{G^{-1}}{2\mu}Y) = 0 \end{aligned} \tag{13}$$

where

$$G = diag(\frac{1}{2\|(D-A-E)^i\|_2}).$$

$$A_{k+1}^{j+1} = D_{\frac{G^{-1}}{2\mu_k}}(D - E_{k+1}^{j+1} + \frac{G^{-1}}{2\mu_k}Y_k)$$

When update E:

$$\begin{aligned} \frac{\partial L(A,E,Y,\mu)}{\partial E} &= \gamma\frac{\partial\|E\|_1}{\partial E} - 2\mu G(D-A-E) - Y \\ &= \gamma\frac{G^{-1}}{2\mu}\frac{\partial\|E\|_1}{\partial E} - D + A + E - \frac{G^{-1}}{2\mu}Y = 0 \\ &= \gamma\frac{G^{-1}}{2\mu}\frac{\partial\|E\|_1}{\partial E} + E - (D - A + \frac{G^{-1}}{2\mu}Y) = 0 \end{aligned} \tag{14}$$

$$E_{k+1}^{j+1} = S_{\frac{G^{-1}\gamma}{2\mu_k}}(D - A_{k+1}^{j+1} + \frac{G^{-1}}{2\mu_k}Y_k)$$

C. Algorithm

Input: Observation matrix $\mathrm{D} \in \Re^{m\times n}, \lambda$.

$1: Y_0^* = \mathrm{sgn}(D)/J(\mathrm{sgn}(D)); \mu_0 > 0; \rho > 1; k = 0.$

$2:$ *while* not converged do

$3:$ //Lines 4-13 solve $(\mathrm{A}_{k+1}^*, \mathrm{E}_{k+1}^*) = \arg\min_{A,E} L(A,E,Y_k^*,\mu_k)$.

$4: A_{k+1}^0 = A_k^*, E_{k+1}^0 = E_k^*, j = 0;$

$5:$ *while* not converged do

$6: G = diag(\dfrac{1}{2\|(D - A_{k+1}^j - E_{k+1}^j)^i\|_2});$

$7:$ //Lines 8-9 solve $A_{k+1}^{j+1} = \arg\min_A L(A, E_{k+1}^j, Y_k^*, \mu_k)$.

$8: (U,S,V) = svd(D - E_{k+1}^j + \dfrac{G^{-1}}{2\mu_k}Y_k^*);$

$9: A_{k+1}^{j+1} = US_{\frac{G^{-1}}{2\mu_k}}[S]V^T;$

$10:$ // *Line* 11 *solve* $E_{k+1}^{j+1} = \arg\min_E L(A_{k+1}^{j+1}, E, Y_k^*, \mu_k)$.

$11: E_{k+1}^{j+1} = S_{\lambda\frac{G^{-1}}{2\mu_k}}[D - A_{k+1}^{j+1} + \dfrac{G^{-1}}{2\mu_k}Y_k^*];$

$12: j \leftarrow j+1.$

$13:$ *end while*

$14: Y_{k+1}^* = Y_k^* + \mu_k(D - A_{k+1}^* - E_{k+1}^*).$

$15:$ *Update* μ_k *to* μ_{k+1}.

$16: k \leftarrow k+1.$

$17:$ *end while*

Output: (A_k^*, E_k^*).

For the Algorithm any accumulation point (A^*, E^*) of (A_k^*, E_k^*) is an optimal solution to the RPCA problem and the convergence rate is at least $O(\frac{G_k}{\mu_k^2})$,Where

$$G_k = \sum_{i=1}^{n} \frac{1}{2\left\|(D - A_{k+1}^* - E_{k+1}^*)^i\right\|_2} / n \cdot$$

In the sense that

$$\left|\|A_k^*\|_* + \lambda\|E_k^*\|_1 - f^*\right| = O(\frac{G_k}{\mu_k^2}) \tag{15}$$

where f^* is the optimal value of the RPCA problem.

Proof By

$$L(A_{k+1}^*, E_{k+1}^*, Y_k^*, \mu_k) = \min_{A,E} L(A, E, Y_k^*, \mu_k) \le \min_{A+E=D} (A, E, Y_k^*, \mu_k) \tag{16}$$
$$= \min_{A+E=D} (\|A\|_* + \lambda\|E\|_1) = f^*$$

We have

$$\|A_{k+1}^*\|_* + \lambda\|E_{k+1}^*\|_1 \le L(A_{k+1}^*, E_{k+1}^*, Y_k^*, \mu_k) - \frac{G_k}{\mu_k^2}(\|Y_{k+1}^*\|_F^2 - \|Y_k^*\|_F^2) \tag{17}$$
$$\le f^* - \frac{G_k}{\mu_k^2}(\|Y_{k+1}^*\|_F^2 - \|Y_k^*\|_F^2)$$

By the boundedness of $\{Y_k^*\}$, we see that

$$\|A_{k+1}^*\|_* + \lambda\|E_{k+1}^*\|_1 \le f^* + O(\frac{G_k}{\mu_k^2}) \tag{18}$$

By letting $k \to +\infty$, we have that

$$\|A^*\|_* + \lambda\|E^*\|_1 \le f^* \tag{19}$$

As $D - A_{k+1}^* - E_{k+1}^* = \mu_k^{-1}(Y_{k+1}^* - Y_k^*)$, by the boundedness of Y_k^* and letting $k \to \infty$ we see that

$$A^* + E^* = D \tag{20}$$

Therefore, (A^*, E^*) is an optimal solution to the RPCA problem.

III. Experimental Results

In this section, we test the proposed Robust PCA with $\ell_{2,1}$-$norm$ Method, Augmented Lagrange Multiplier (ALM) Method and Accelerated Proximal Gradient (APG) Method for some examples and report the numerical results.

A. Simulation Conditions.

We use randomly generated square matrices for our simulations. We denote the true solution by the ordered pair $(A_0, E_0) \in \Re^{m\times m} \times \Re^{m\times m}$. We generate the rank-r (r=25) matrix A_0 as a product UV^T , where U and V are independent $m \times r$ matrices whose elements are i.i.d. Gaussian random variables with zero mean and unit variance. We generate E_0 as a sparse matrix whose support is chosen uniformly at random, and whose non-zero entries are i.i.d. uniformly in the interval [−500, 500].Randomly generating Gaussian dense small noise matrix N_0 whose elements are i.i.d. The matrix $D = A_0 + E_0 + \sigma N_0$ ($\sigma = 0.001$) is the input to the algorithm, and ($\hat{A}, \hat{E}$) denotes the output.

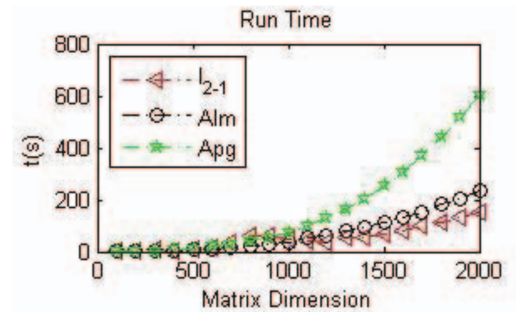

Figure 1. The run time of the three algorithms

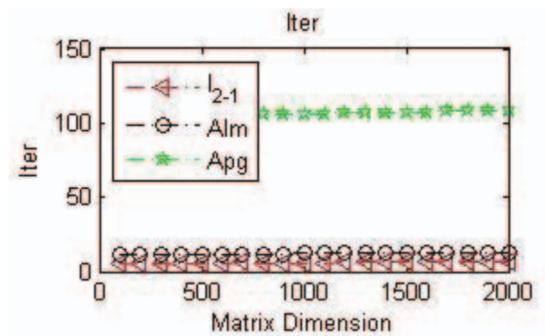

Figure 2. The iteration times of the three algorithms

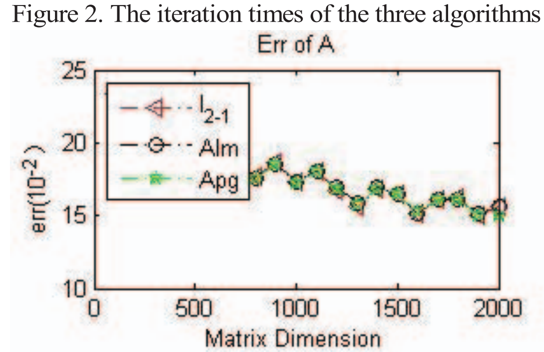

Figure 3. The err($\frac{\|\hat{A} - A_0\|_F}{\|A_0\|_F}$) of the three algorithms

From Fig. 1, Fig. 2, Fig. 3, Fig. 4, we have that the $\ell_{2,1}$-$norm$ method is more faster than APG and ALM, and their err of recovered A are almost same.

B. Photo Example

From Fig. 5 we have that; the $\ell_{2,1}$ - $norm$ method is faster than APG and ALM. In addition, from the decomposition of the sparse part (E) of the picture, we can see that it is clearer than the other two algorithms, so we can know that the $\ell_{2,1}$ - $norm$ method is more robust than other two.

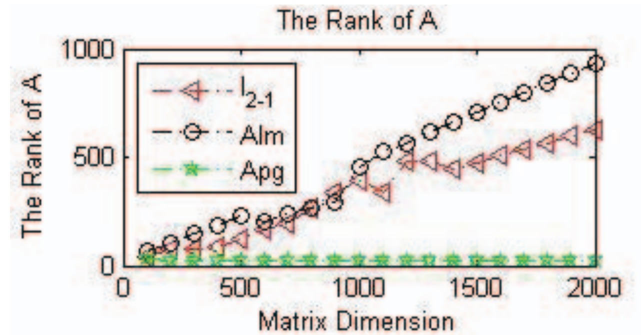

Figure 4. The rank($\hat{A}$) of the three algorithms

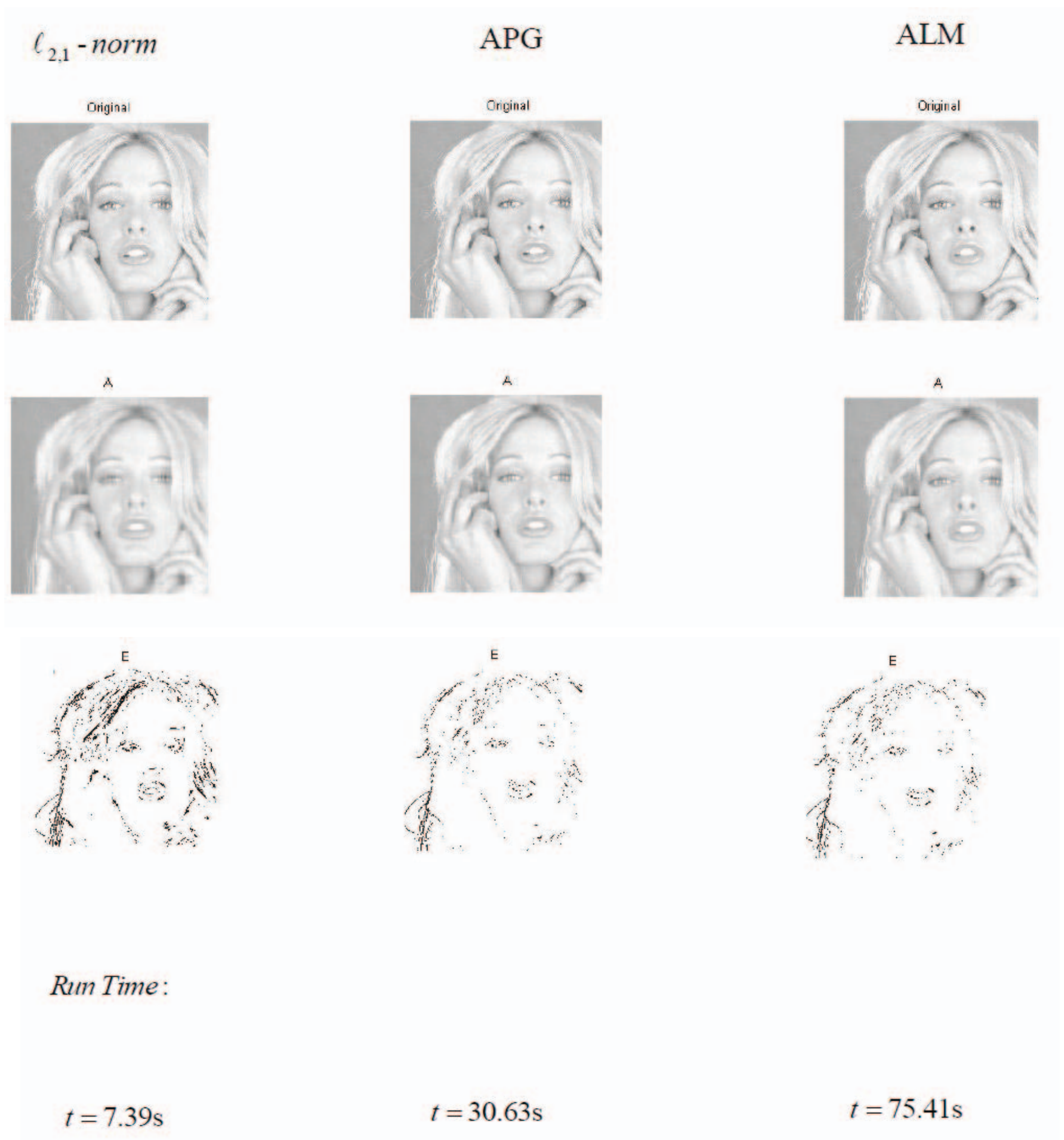

Figure 5. An Experimental Comparative Study on Three Algorithms

IV. CONCLUSIONS

In this paper, we propose a new robust PCA with emphasizing on $\ell_{2,1}$ - *norm* minimization on loss function. The $\ell_{2,1}$ - *norm* based loss function is robust to outliers in data points. An efficient algorithm is introduced. Compared with traditional method, our method needs less iteration and more robust to outliers in data points and it run fast than APG and ALM.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No.11301021 and 11571031).

REFERENCES

[1] John Wright, Yigang Peng and Yi Ma. Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization. Advances in Neural Information Processing Systems, 2009，87(4), 20:3 – 20:56.

[2] Emmanuel J. Cand`es, Xiaodong Li, Yi Ma and John Wright. Robust principal component analysis. Journal of the Acm, 2000，1(1), 1-73.

[3] Huan Xu, Constantine Caramanis, Member and Sujay Sanghavi. Robust PCA via Outlier Pursuit [J]. Information Theory IEEE Transactions on, 2010, 58(5):3047 - 3064.

[4] Emmanuel J. Cand`es and Terence Tao. The Power of Convex Relaxation: Near-Optimal Matrix Completion [J]. IEEE Transactions on Information Theory, 2009, 56(5):2053 - 2080.

[5] Emmanuel J. Cand`es and Benjamin Recht. Exact matrix completion via convex optimization [J]. Communications of the Acm, 2012, 55(6):111-119.

[6] Zhouchen Lin, Arvind Ganesh, John Wright, Leqin Wu, Minming Chen and Yi Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. Journal of the Marine Biological Association of the 2015, UK, 56(3), 707-722.

[7] Lin, Zhouchen, Chen, Minming and Ma, Yi. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices [J]. Arxiv Preprint Arxiv: 1009, 2010.

[8] Chris Ding., Ding Zhou, Xiaofeng He and Hongyuan Zha. R1-PCA: Rotational Invariant L1-norm Principal Component Analysis for Robust Subspace Factorization. International Conference on Machine Learning, 2006, (Vol.2006, pp.281-288). ACM.

[9] Feiping Nie, Heng Huang, Xiao Cai and Chris Ding. Efficient and Robust Feature Selection via Joint $l_{2,1}$-norm Minimization. Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada. (Pp.1813-1821).

[10] Zechao Li, Jing Liu and Jinhui Tang. Robust structured subspace learning for data representation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(10), 2085-98.

[11] Jian-Feng Cai，Emmanuel J. Cand`es and Zuowei Shen. A Singular Value Thresholding Algorithm for Matrix Completion. [J]. Siam Journal on Optimization, 2010, 20(4):1956-1982.