ROBUST M-ESTIMATION BASED MATRIX COMPLETION

Michael Muma, Wen-Jun Zeng, Abdelhak M. Zoubir

Signal Processing Group
Technische Universität Darmstadt
Merckstraße 25, 64283 Darmstadt, Germany
{muma, zoubir}@spg.tu-darmstadt.de, wenjzeng@gmail.com

ABSTRACT

Conventional approaches to matrix completion are sensitive to outliers and impulsive noise. This paper develops robust and computationally efficient M-estimation based matrix completion algorithms. By appropriately arranging the observed entries, and then applying alternating minimization, the robust matrix completion problem is converted into a set of regression M-estimation problems. Making use of differentiable loss functions, the proposed algorithm overcomes a weakness of the ℓ_p-loss ($p \leq 1$), which easily gets stuck in an inferior point. We prove that our algorithm converges to a stationary point of the nonconvex problem. Huber's joint M-estimate of regression and scale can be used as a robust starting point for Tukey's redescending M-estimator of regression based on an auxiliary scale. Numerical experiments on synthetic and real-world data demonstrate the superiority to state-of-the-art approaches.

***Index Terms*—** Low-rank factorization, matrix completion, M-estimation, robust statistics, image inpainting

1. INTRODUCTION

Matrix completion refers to recovering a low-rank matrix from a subset of its entries [1]–[5]. It has numerous applications in recommender systems, computer vision, image inpainting, biomedicine, and information retrieval [1]–[9]. Generally speaking, matrix completion approaches may be loosely grouped into three categories. The first is based on nuclear norm [10]–[12] or Schatten p-norm [13] minimization. The second is based on the hard thresholding, i.e., projection onto nonconvex rank constraint sets [15]–[18], The third uses direct matrix factorization to ensure low-rank [6, 14, 19, 20], and has the lowest computational complexity, making it the most attractive approach in the "big data" setting.

The occurrence of outliers and impulsive noise, e.g., ratings with frauds in recommender systems [1], and salt-and-pepper noise in image processing [14], is common in many practical applications [21, 22]. As conventional matrix completion methods are based on quadratic loss functions, their performance significantly degrades in the presence of impulsive noise or outliers. Therefore, ℓ_p-loss ($p \leq 1$) based schemes have been recently developed. However, there are two major drawbacks of the ℓ_p-loss based methods. First, although providing robustness in the face of outliers, they may be statistically inefficient with respect to additional background noise. Second, they easily get stuck at an inferior solution rather than a stationary point when alternating minimization is employed to solve the nonsmooth ℓ_p-minimization [23, 24].

The contribution of this paper is to propose robust and computationally efficient matrix completion approaches based on M-estimation and matrix factorization. By appropriately arranging the observed entries, and then applying alternating minimization, we cast the robust matrix completion problem into a set of regression M-estimation problems. Since our approach uses differentiable, and statistically efficient loss functions, such as, Huber's and Tukey's, we overcome the above mentioned drawbacks of the ℓ_p-loss ($p \leq 1$). We prove that our proposed algorithm converges to a stationary point of the nonconvex problem, and provide an expression for the computational complexity for Huber's M-estimation based approach.

The remainder of the paper is organized as follows. Section 2 details the proposed approaches. Section 3 contains numerical experiments with synthetic and real data. Finally, Section 4 concludes the paper.

2. ROBUST M-ESTIMATION BASED APPROACH

The observed matrix $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$ is modeled as

$$\boldsymbol{X} = \boldsymbol{M} + \boldsymbol{S} + \boldsymbol{N}, \tag{1}$$

where $\boldsymbol{M}$ is a low-rank matrix of rank-r, $\boldsymbol{S}$ is an entry-wise sparse outlier matrix, and $\boldsymbol{N}$ represents the background noise. Our goal is to recover the low-rank component $\boldsymbol{M}$ from partially observed entries of $\boldsymbol{X}$ corrupted by noise and outliers.

In our approach, the outlier-robust "norm" of $\boldsymbol{X}$ is defined

The work of M. Muma is supported by the 'Athene Young Investigator Programme' of Technische Universität Darmstadt.

as

$$\|\boldsymbol{X}\|_{\sigma,c} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \rho\left(\frac{x_{ij}}{\sigma}\right), \tag{2}$$

where $\sigma > 0$ is the scale parameter, x_{ij} is the (i, j)th entry of $\boldsymbol{X}$, and $\rho(\cdot)$ is a differentiable loss function. In this paper, we consider Huber's loss function

$$\rho_{\text{hub}}(x) = \begin{cases} \frac{1}{2}x^2, & |x| \le c \\ c|x| - \frac{1}{2}c^2, & |x| > c \end{cases} \tag{3}$$

and Tukey's loss function

$$\rho_{\text{tuk}}(x) = \begin{cases} \frac{1}{2}x^2 - \frac{x^4}{2c^2} + \frac{x^6}{6c^4}, & |x| \le c \\ \frac{c^2}{6}, & |x| > c \end{cases} \tag{4}$$

for which the associated tuning parameter c trades off the efficiency and robustness. For example, $c = 1.345$ and $c = 4.685$, respectively, yield 95% asymptotic relative efficiency for Huber's loss and Tukey's loss functions. Unlike the ℓ_1-loss, Huber's and Tukey's losses are differentiable. Further, Huber's loss is convex while Tukey's loss is nonconvex.

To describe missing data, the row-column indices of the partially observed entries are collected in the set $\Omega \subset \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$. We use $\boldsymbol{X}_\Omega \in \mathbb{R}^{n_1 \times n_2}$ to denote the projection of the $\boldsymbol{X}$ onto Ω. As a result, we have $[\boldsymbol{X}_\Omega]_{ij} = 0$ if $(i, j) \notin \Omega$ and $[\boldsymbol{X}_\Omega]_{ij} = x_{ij}$ if $(i, j) \in \Omega$.

Popular methods for matrix completion solve the problem of Schatten p-norm regularization [13]:

$$\min_{\boldsymbol{M}} \|(\boldsymbol{M})_\Omega - \boldsymbol{X}_\Omega\|_{\text{F}}^2 + \gamma\|\boldsymbol{M}\|_{S_p}^p. \tag{5}$$

Here, $\gamma > 0$ is the regularization factor and $\|\boldsymbol{M}\|_{S_p}$ refers to the Schatten p-norm that promotes rank-sparsity. When $p = 1$, $\|\boldsymbol{M}\|_{S_p}$ reduces to the nuclear norm [4], [10]–[12], denoted by $\|\boldsymbol{M}\|_{\text{nuc}}$. Methods that are based on (5) require computing the singular value decomposition (SVD) of an $n_1 \times n_2$ matrix at each iteration, resulting in a high complexity. To reduce complexity, we use direct matrix factorization $\widehat{\boldsymbol{M}} = \boldsymbol{U}\boldsymbol{V}$ to make the estimate $\widehat{\boldsymbol{M}}$ low-rank, where $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{V} \in \mathbb{R}^{r \times n_2}$. In the presence of outliers, our robust M-estimation based matrix completion is expressed as:

$$\min_{\boldsymbol{U},\boldsymbol{V}} \|(\boldsymbol{U}\boldsymbol{V})_\Omega - \boldsymbol{X}_\Omega\|_{\sigma,c}. \tag{6}$$

Herein, the scale parameter σ is unknown and is estimated jointly with $(\boldsymbol{U}, \boldsymbol{V})$, whereas c is set in advance, and is considered as a constant.

To solve (6), an alternating minimization strategy is applied. To be more specific, at the $(k + 1)$th iteration ($k = 0, 1, 2, \ldots$), $\boldsymbol{V}$ and $\boldsymbol{U}$ are alternately minimized:

$$\boldsymbol{V}^{k+1} = \arg\min_{\boldsymbol{V}} \|(\boldsymbol{U}^k\boldsymbol{V})_\Omega - \boldsymbol{X}_\Omega\|_{\sigma,c} \tag{7}$$

$$\boldsymbol{U}^{k+1} = \arg\min_{\boldsymbol{U}} \|(\boldsymbol{U}\boldsymbol{V}^{k+1})_\Omega - \boldsymbol{X}_\Omega\|_{\sigma,c}. \tag{8}$$

Note that (8) and (7) have the same structure. Thus, we only discuss how to solve (7), because (8) is solved analogously. With the following definitions:

- $\boldsymbol{u}_i^\top \in \mathbb{R}^{1\times r}$, and $\boldsymbol{v}_j \in \mathbb{R}^r$, respectively denote the ith row of $\boldsymbol{U}$ and jth column of $\boldsymbol{V}$, where $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$
- $\mathcal{I}_j = \{j_1, \ldots, j_{|\mathcal{I}_j|}\} \subseteq \{1, \ldots, n_1\}$, represents the set containing the row indices for the jth column, $j = 1, \ldots, n_2$, of Ω, where, $\sum_{j=1}^{n_2} |\mathcal{I}_j| = |\Omega|$, and, in general, $|I_j| > r$
- $\boldsymbol{U}^k_{\mathcal{I}_j} \in \mathbb{R}^{|\mathcal{I}_j|\times r}$ is the matrix containing the $|\mathcal{I}_j|$ rows of $\boldsymbol{U}^k$ that are indexed by $\mathcal{I}_j$
- $\boldsymbol{b}_{\mathcal{I}_j} = [x_{j_1 j}, \cdots, x_{j_{|\mathcal{I}_j|} j}]^\top \in \mathbb{R}^{|\mathcal{I}_j|}$,

Eq. (7) can be rewritten as a series of robust linear regression problems that can be solved in parallel, i.e.,

$$\min_{\boldsymbol{v}_j} \left\{ f_\sigma(\boldsymbol{v}_j) \triangleq \|\boldsymbol{U}^k_{\mathcal{I}_j}\boldsymbol{v}_j - \boldsymbol{b}_{\mathcal{I}_j}\|_{\sigma,c} \right\},\ j = 1, \ldots, n_2. \tag{9}$$

Since σ is unknown, and a preliminary estimate is difficult to obtain, we propose to solve (9) using Huber's joint M-estimation of regression and scale approach

$$\min_{\boldsymbol{v}_j,\sigma>0} L_{\text{hub}}(\boldsymbol{v}_j, \sigma) \triangleq \sigma \sum_{i\in\mathcal{I}_j} \rho_{\text{hub}}\left(\frac{x_{ij} - (\boldsymbol{u}_i^\top)^k\boldsymbol{v}_j}{\sigma}\right) + |\mathcal{I}_j|(\alpha\sigma) \tag{10}$$

where α is a fixed scaling factor to obtain Fisher consistency of the scale estimate $\widehat{\sigma}$. Eq. (10) is jointly convex in $(\boldsymbol{v}_j, \sigma)$. Therefore, the global minimizer $(\widehat{\boldsymbol{v}}_j, \widehat{\sigma})$ is a stationary point of Huber's criterion, and a solution can be found by solving the M-estimating equations obtained by setting the gradient of $L_{\text{hub}}(\boldsymbol{v}_j, \sigma)$, with respect to its arguments, to zero. Since the complexity for solving (10) is $\mathcal{O}(|\mathcal{I}_j|r^2)$, the per-iteration complexity of the matrix completion is $\mathcal{O}(|\Omega|r^2)$ due to $\sum_{j=1}^{n_2} |\mathcal{I}_j| = |\Omega|$. The Huber's M-estimation based matrix completion is summarized in Algorithm 1.

As mentioned in Section 1, there is no theoretical guarantee that the alternating ℓ_p-minimization method [14] converges to a stationary point. In contrast, Huber's M-estimator, as computed in Algorithm 1, converges to a stationary point. This is stated in the following theorem.

Theorem *The sequence generated by Algorithm 1, i.e., $\{\boldsymbol{U}^k, \boldsymbol{V}^k\}$, converges to a stationary point of the nonconvex problem of* (6).

Proof. We first prove that the function

$$f_\sigma(\boldsymbol{z}) = \|\boldsymbol{U}^k_{\mathcal{I}_j}\boldsymbol{z} - \boldsymbol{b}_{\mathcal{I}_j}\|_{\sigma,c}$$

with $\boldsymbol{z} \in \mathbb{R}^r$, whose form is the same as the objective function in (9), is *strictly* convex. Define the following affine transformation $\mathbb{R}^r \to \mathbb{R}^{|\mathcal{I}_j|}$ by:

$$\boldsymbol{r}(\boldsymbol{z}) = \boldsymbol{U}^k_{\mathcal{I}_j}\boldsymbol{z} - \boldsymbol{b}_{\mathcal{I}_j}. \tag{11}$$

Algorithm 1 Robust Matrix Completion via Huber's M-Estimation

Input: $\boldsymbol{X}_\Omega$, Ω, and rank r

Initialize: Randomly initialize $\boldsymbol{U}^0 \in \mathbb{R}^{n_1 \times r}$

Determine $\{\mathcal{I}_j\}_{j=1}^{n_2}$ and $\{\mathcal{J}_i\}_{i=1}^{n_1}$ according to Ω.

for $k = 0, 1, \cdots$ **do**

// Fix $\boldsymbol{U}^k$, optimize $\boldsymbol{V}$

$$\boldsymbol{v}_j^{k+1} = \arg\min_{\boldsymbol{v}_j, \sigma} \left\{ \sigma \sum_{i \in \mathcal{I}_j} \rho_{\text{hub}} \left(\frac{x_{ij} - (\boldsymbol{u}_i^\top)^k \boldsymbol{v}_j}{\sigma} \right) + |\mathcal{I}_j|(\alpha\sigma) \right\}$$

for all $j = 1, 2, \cdots, n_2$.

// Fix $\boldsymbol{V}^{k+1}$, optimize $\boldsymbol{U}$

$$(\boldsymbol{u}_i^\top)^{k+1} = \arg\min_{\boldsymbol{u}_i^\top, \sigma} \left\{ \sigma \sum_{j \in \mathcal{J}_i} \rho_{\text{hub}} \left(\frac{x_{ij} - \boldsymbol{u}_i^\top \boldsymbol{v}_j^{k+1}}{\sigma} \right) + |\mathcal{J}_i|(\alpha\sigma) \right\}$$

for all $i = 1, 2, \cdots, n_1$.

Stop if a termination condition is satisfied.

end for

Output: $\widehat{\boldsymbol{M}} = \boldsymbol{U}^{k+1}\boldsymbol{V}^{k+1}$

Consider the function

$$g_\sigma(\boldsymbol{r}) = \|\boldsymbol{r}\|_{\sigma,c} = \sum_{i=1}^{|\mathcal{I}_j|} \rho_{\text{hub}} \left(\frac{r_i}{\sigma} \right)$$

with r_i being the ith entry of $\boldsymbol{r}$. Since the univariate Huber's loss function $\rho_{\text{hub}}(r_i/\sigma)$ is strictly convex with respect to $r_i \in \mathbb{R}$, the sum of strictly convex functions, i.e., $g_\sigma(\boldsymbol{r})$, is also strictly convex. Then, $\forall\, \boldsymbol{r}_1, \boldsymbol{r}_2 \in \mathbb{R}^{|\mathcal{I}_j|}$, $\boldsymbol{r}_1 \neq \boldsymbol{r}_2$, and $\alpha \in (0,1)$, we have

$$g_\sigma(\alpha\boldsymbol{r}_1 + (1-\alpha)\boldsymbol{r}_2) < \alpha g_\sigma(\boldsymbol{r}_1) + (1-\alpha) g_\sigma(\boldsymbol{r}_2). \quad (12)$$

Let $\forall\, \boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^r$, whose affine transform applying (11), respectively, is $\boldsymbol{r}_1 = \boldsymbol{U}_{\mathcal{I}_j}^k \boldsymbol{z}_1 - \boldsymbol{b}_{\mathcal{I}_j}$ and $\boldsymbol{r}_2 = \boldsymbol{U}_{\mathcal{I}_j}^k \boldsymbol{z}_2 - \boldsymbol{b}_{\mathcal{I}_j}$. We first need to prove that: if $\boldsymbol{z}_1 \neq \boldsymbol{z}_2$, then $\boldsymbol{r}_1 \neq \boldsymbol{r}_2$. Due to $|\mathcal{I}_j| > r$, $\boldsymbol{U}_{\mathcal{I}_j}^k \in \mathbb{R}^{|\mathcal{I}_j| \times r}$ for all $j = 1, 2, \cdots, n_2$, are full column rank. Then, only $\boldsymbol{0}$ is in the null space of $\boldsymbol{U}_{\mathcal{I}_j}^k$. That is, we have $\boldsymbol{U}_{\mathcal{I}_j}^k \boldsymbol{z} = \boldsymbol{0} \Leftrightarrow \boldsymbol{z} = \boldsymbol{0}$ and $\boldsymbol{U}_{\mathcal{I}_j}^k \boldsymbol{z} \neq \boldsymbol{0} \Leftrightarrow \boldsymbol{z} \neq \boldsymbol{0}$. Thus, if $\boldsymbol{z}_1 \neq \boldsymbol{z}_2$, then $\boldsymbol{r}_1 \neq \boldsymbol{r}_2$ due to $\boldsymbol{r}_1 - \boldsymbol{r}_2 = \boldsymbol{U}_{\mathcal{I}_j}^k (\boldsymbol{z}_1 - \boldsymbol{z}_2)$.

For $\forall\, \boldsymbol{z}_1 \neq \boldsymbol{z}_2$ and $\alpha \in (0,1)$, it follows that

$$\begin{aligned}
& f_\sigma(\alpha\boldsymbol{z}_1 + (1-\alpha)\boldsymbol{z}_2) \\
&= \|\boldsymbol{U}_{\mathcal{I}_j}^k(\alpha\boldsymbol{z}_1 + (1-\alpha)\boldsymbol{z}_2) - \boldsymbol{b}_{\mathcal{I}_j}\|_{\sigma,c} \\
&= \|\alpha(\boldsymbol{U}_{\mathcal{I}_j}^k\boldsymbol{z}_1 - \boldsymbol{b}_{\mathcal{I}_j}) + (1-\alpha)(\boldsymbol{U}_{\mathcal{I}_j}^k\boldsymbol{z}_2 - \boldsymbol{b}_{\mathcal{I}_j})\|_{\sigma,c} \\
&= \|\alpha\boldsymbol{r}_1 + (1-\alpha)\boldsymbol{r}_2\|_{\sigma,c} \\
&= g_\sigma(\alpha\boldsymbol{r}_1 + (1-\alpha)\boldsymbol{r}_2) \\
&< \alpha g_\sigma(\boldsymbol{r}_1) + (1-\alpha) g_\sigma(\boldsymbol{r}_2) \qquad (13) \\
&= \alpha f_\sigma(\boldsymbol{z}_1) + (1-\alpha) f_\sigma(\boldsymbol{z}_2). \qquad (14)
\end{aligned}$$

Note that when deriving (13), we have used the important fact that $\boldsymbol{r}_1 \neq \boldsymbol{r}_2$ due to $\boldsymbol{z}_1 \neq \boldsymbol{z}_2$ and the property of strict convexity of $g_\sigma(\boldsymbol{r})$ in (12).

Since $f_\sigma(\boldsymbol{z})$ is strictly convex, the minimization problems of (9) have unique solution. That is, all optimal $\{\boldsymbol{v}_j\}_{j=1}^{n_2}$ are uniquely determined, and it follows that the optimal solution of (7) is unique. Clearly, the optimal solution of (8) is also unique since (7) and (8) have the same structure. Algorithm 1 is in fact a block coordinate descent method with two blocks $\boldsymbol{U}$ and $\boldsymbol{V}$. We have proved that the minimizers of each block are unique. In addition, the objective function of (6) is continuously differentiable. According to Proposition 2.7.1 in [25], the sequence generated by Algorithm 1, i.e., $\{\boldsymbol{U}^k, \boldsymbol{V}^k\}$, converges to a stationary point. □

Remark: Although the Huber's loss function has been employed for matrix completion very recently [1], our algorithm is different from that of [1]. The estimator of [1] is based on the nuclear norm regularization:

$$\min_{\boldsymbol{M}} \sum_{(i,j)\in\Omega} \rho_{\text{hub}} \left(\frac{[\boldsymbol{M}]_{ij} - x_{ij}}{\sigma} \right) + \gamma \|\boldsymbol{M}\|_{\text{nuc}} \qquad (15)$$

which requires SVD at each iteration and has a high complexity. Our approach is based on matrix factorization and more computationally efficient.

The estimate obtained from Algorithm 1 can be used as a robust starting point for Tukey's M-estimator, which solves the nonconvex optimization problem

$$\min_{\boldsymbol{v}_j} L_{\text{tuk}}(\boldsymbol{v}_j, \sigma) \triangleq \sum_{i \in \mathcal{I}_j} \rho_{\text{tuk}} \left(\frac{x_{ij} - (\boldsymbol{u}_i^\top)^k \boldsymbol{v}_j}{\sigma_{\text{hub}}} \right) \qquad (16)$$

using an iteratively reweighted least-squares (IRWLS) algorithm. For further details on Huber's joint M-estimation of regression and scale approach, and on M-estimation of regression with an auxiliary scale estimate, see Chapter 2 of [22].

3. SIMULATION RESULTS

In all simulations, we use the MATLAB functions `hubreg`, and `Mreg` of the RobustSP toolbox [22], and set $c = 1.345$ for Huber's estimator while $c = 1.481$ for Tukey's estimator.

Results for Synthetic Random Data: We set $n_1 = 150$, $n_2 = 300$, and $r = 10$. The proposed methods are compared with SVT [12], SVP [15], AP [17], WNNM [9], RPCA [27], and VBMFL_1 [28]. The matrix is generated by $\boldsymbol{X} = \boldsymbol{X}_1\boldsymbol{X}_2$ where $\boldsymbol{X}_1 \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{r \times n_2}$ are Gaussian random matrices. Then, impulsive noise is added, which follows the Gaussian mixture model (GMM). The GMM noise contains two components. The component with smaller variance models the background noise $\boldsymbol{N}$ while the one with larger variance models the sparse outlier $\boldsymbol{S}$. Fig. 1 shows the normalized root mean square error (RMSE) versus signal-to-noise ratio (SNR) where 45% random entries of $\boldsymbol{X}$ are observed. The SNR is defined as $\text{SNR} = \|\boldsymbol{X}_\Omega\|_\text{F}^2/(|\Omega|\sigma_n^2)$ with σ_n^2 denoting the noise variance. The RMSEs are the average of 100 independent trials. Fig. 2 plots the normalized RMSE versus percentage of observations at $\text{SNR} = 9$ dB.

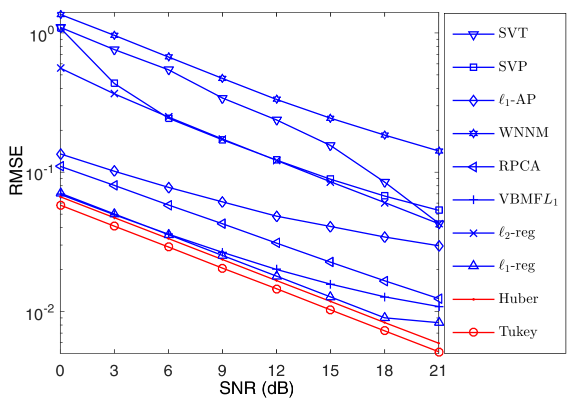

Fig. 1. RMSE versus SNR in GMM noise.

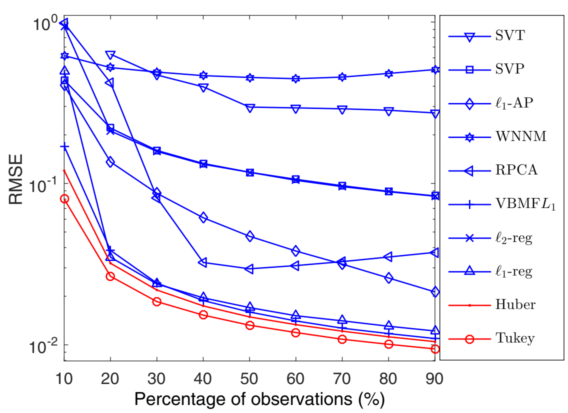

Fig. 2. RMSE versus percentage of observations.

The simulations show that the ℓ_2-loss based SVT, SVP, WNNM and ℓ_2-regression are not robust to impulsive noise, and that the proposed Huber and Tukey M-estimators outperform the state-of-the-art robust schemes, i.e., ℓ_1-AP, ℓ_1-regression, RPCA and VBMFL_1 in terms of robustness and estimation accuracy.

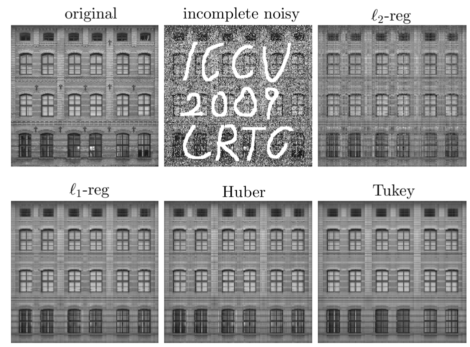

Fig. 3. Image inpainting in salt-and-pepper noise.

Image Inpainting in Salt-and-Pepper Noise: We apply robust matrix completion to image inpainting. A building image is adopted [14]. The missing entries correspond to “ICCV”, “2009”, and “LRTC”. The observed entries are contaminated by salt-and-pepper noise. Fig 3 shows the original image, and the noisy version with missing and recovered values. The rank is set as $r = 6$ for all methods. The peak signal-to-noise ratio (PSNR)

$$\text{PSNR} = 255^2/\text{MSE} \tag{17}$$

is taken as the performance measure, where 255 is the peak value of a gray-scale image and $\text{MSE} = \|\widehat{\boldsymbol{M}} - \boldsymbol{X}\|_\text{F}^2/(n_1 n_2)$, with $\widehat{\boldsymbol{M}}$ being the estimate of a matrix completion method. The PSNR of the noisy image with missing values without any processing is taken as the baseline. The PSNRs (in dB) of the baseline, ℓ_2-regression, ℓ_1-regression, Huber and Tukey based methods are 10.83, 19.18, 21.61, 23.29, and 23.71, respectively, at $\text{SNR} = 6$ dB. It is observed the proposed M-estimators have the best recovery performance.

4. CONCLUSION

A robust and computationally efficient M-estimation based matrix completion approach, using Huber’s and Tukey’s M-estimators has been proposed. Through alternating minimization, the matrix completion problem is decomposed into a series of regression M-estimation problems. The approach outperforms the existing ℓ_p-minimization method when outlier and background noise coexist. We prove that our algorithm converges to a stationary point of the nonconvex problem, i.e., does not get stuck at an inferior solution. The per-iteration complexity of M-estimation based matrix completion using Huber’s loss is $\mathcal{O}(|\Omega|r^2)$, which makes it attractive tool for the “big data” setting. Numerical experiments illustrate the superiority of the proposed approach.

5. REFERENCES

[1] A. Elsener and S. van de Geer, "Robust low-rank matrix estimation," *Ann. Stat.,* vol. 46, no. 6, pp. 3481–3509, 2018.

[2] T. Schäck, M. Muma, and A. M. Zoubir, "Estimation of blood pressure, pulse wave velocity and vascular age index from a single photoplethysmographic signal," *IEEE J. Biomed. Health Inf.,* (submitted), 2018.

[3] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE J. Sel. Top. Signal Process.,* vol. 10, no. 4, pp. 608–622, Jun. 2016.

[4] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE,* vol. 98, no. 6, pp. 925–936, Jun. 2010.

[5] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inf. Theory,* vol. 56, no. 6, pp. 2980–2998, Jun. 2010.

[6] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer,* vol. 42, no. 8, pp. 30–37, 2009.

[7] Y. Zhou, D. Wilkinson, R. Schreiber, and R. Pan, "Large-scale parallel collaborative filtering for the Netflix prize," in *Proc. 4th Int Conf. Algo. Inf. Manag.,* LNCS 5034, pp. 337–348, Springer, 2008.

[8] A. Eriksson and A. van den Hengel, "Efficient computation of robust weighted low-rank matrix approximations using the L_1 norm," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. 34, no. 9, pp. 208–220, Sep. 2012.

[9] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vision,* vol. 121, no. 2, pp. 183–208, Jan. 2017.

[10] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory,* vol. 56, no. 5, pp. 2053–2080, May 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, 2010.

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

[13] F. Nie, H. Wang, H. Huang, and C. Ding, "Joint Schatten p-norm and ℓ_p-norm robust matrix completion for missing value recovery," *Knowl. Inf. Syst.*, vol. 42, no. 3, pp. 525–544, Mar. 2015.

[14] W.-J. Zeng and H. C. So, "Outlier-robust matrix completion via ℓ_p-minimization," *IEEE Trans. Signal Process.,* vol. 66, no. 5, pp. 1125–1140, Mar. 2018.

[15] P. Jain, R. Meka, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Adv. Neural Inf. Process. Syst. (NIPS),* pp. 937–945, 2010.

[16] J. Tanner and K. Wei, "Normalized iterative hard thresholding for matrix completion," *SIAM J. Sci. Comput.,* vol. 35, no. 5, pp. S104–S125, Oct. 2013.

[17] X. Jiang, Z. Zhong, X. Liu, and H. C. So, "Robust matrix completion via alternating projection," *IEEE Sig. Process. Lett.,* vol. 24, no. 5, pp. 579–583, May 2017.

[18] M. -J. Lai, and A. Varghese, "On convergence of the alternating projection method for matrix completion and sparse recovery problems, arXiv:1711.02151 [math.OC], Nov. 2017.

[19] M. Hardt, "Understanding alternating minimization for matrix completion," in *Proc. 55th IEEE Annual Sym. Found. Comput. Science (FOCS),* pp. 651–660, Philadelphia, PA, Oct. 2014.

[20] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via nonconvex factorization," *IEEE Trans. Inf. Theory,* vol. 62, no. 11, pp. 6535–6579, Nov. 2016.

[21] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.,* vol. 29, no. 4, pp. 61–80, Jul. 2012.

[22] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing.* Cambridge University Press, Cambridge, UK, 2018.

[23] S. J. Wright, "Coordinate descent algorithms," *Math. Program., Ser. A*, vol. 151, no. 1, pp. 3–34, Jun. 2015.

[24] M. J. D. Powell, "On search directions for minimization algorithms," *Math. Program.,* vol. 4, pp. 193–201, 1973.

[25] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, Belmont, 1999.

[26] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming (4th Edition).* NY: Springer, 2016.

[27] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM,* vol. 58, no. 3, Article 11, May 2011.

[28] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and Y. Yan, "L_1-norm low-rank matrix factorization by variational Bayesian method," *IEEE Trans. Neural Netw. Learn. Syst.,* vol. 26, no. 4, pp. 825–839, Apr. 2015.