Robust Matrix Completion based on Factorization and Truncated-Quadratic Loss Function

Wang, Zhi-Yong; Li, Xiao Peng; So, Hing Cheung

Published in:
IEEE Transactions on Circuits and Systems for Video Technology

Published: 01/04/2023

Document Version:
Post-print, also known as Accepted Author Manuscript, Peer-reviewed or Author Final version

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1109/TCSVT.2022.3214583

Publication details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.
Robust Matrix Completion based on Factorization and Truncated-Quadratic Loss Function

Zhi-Yong Wang, Xiao Peng Li, Hing Cheung So*, Fellow, IEEE

Abstract

Robust matrix completion refers to recovering a low-rank matrix given a subset of the entries corrupted by gross errors, and has various applications since many real-world signals can be modeled as low-rank matrices. Most of the existing methods only perform well for noise-free data or those with zero-mean white Gaussian noise, and their performance will be degraded in the presence of outliers. In this paper, based on the factorization framework, we propose a novel robust matrix completion scheme via using the truncated-quadratic loss function, which is non-convex and non-smooth, and half-quadratic theory is adopted for its optimization. By introducing an auxiliary variable, half-quadratic optimization (HO) can transform the loss function into two tractable forms, that is, additive and multiplicative formulations. Block coordinate descent method is then exploited as their solver. Compared with the additive form, the multiplicative variant has lower computational cost since we attempt to take the observations contaminated by outliers as missing entries. Numerical simulations and experimental results based on image inpainting and hyperspectral image recovery demonstrate that our algorithms are superior to the state-of-the-art methods in terms of restoration accuracy and runtime. MATLAB code is available at https://github.com/bestzywang.

Index Terms

Robust low-rank matrix completion, outlier, non-convex and non-smooth loss function, block coordinate descent

I. INTRODUCTION

Low-rank matrix completion [1]–[3] is an important approach to recover the missing entries given an incomplete matrix. It has been widely used in numerous areas, e.g., hyperspectral remote sensing [4], [5], collaborative filtering [6], [7], image inpainting [8], [9], multi-task learning [10], system identification [11] and deep learning [12], because many real-world data have a low-rank or approximately low-rank structure. For instance, matrix completion can be used to restore the missing pixels of images since their main information often lies in a lower-dimensional subspace [13].

Z.-Y. Wang, X. P. Li and H. C. So are with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong, China. (E-mail: z.y.wang@my.cityu.edu.hk, x.p.li@my.cityu.edu.hk, hcso@ee.cityu.edu.hk). The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No. CityU 11207922].
Intuitively, constrained rank minimization [14] can be utilized to match an incomplete matrix, and recover its missing entries. Nevertheless, the model is NP-hard since the rank function is discrete. Nuclear norm as the convex relaxation of the rank function is suggested to practically address this problem, and the corresponding theoretical analysis can be found in [14], [15]. To solve nuclear norm minimization, schemes such as singular value thresholding (SVT) [16], fixed point continuation with approximating singular value decomposition (FPCA) [17], and accelerated proximal gradient with linesearch (APGL) [18], are proposed. However, full singular value decomposition (SVD) calculation is needed at each iteration, leading to a huge computational cost, particularly for large-size data.

To avoid performing full SVD, approaches based on reduced rank approximation or truncated SVD, including singular value projection (SVP) [19], normalized iterative hard thresholding (NIHT) [20], and alternating projection (AP) [21], are developed. Since truncated SVD is performed at each iteration, their computational cost can be greatly reduced especially when the assumed rank $r$ is much smaller than the matrix column/row length. In addition, low-rank matrix factorization based methods such as low-rank matrix fitting (LMaFit) [22] and alternating minimization for matrix completion (AltMinComplete) [23] have been proposed, whose idea is to replace the target matrix with the product of two rank-$r$ matrices. Moreover, Zhu et al. [24], [25] have analyzed that although the matrix completion problem based on factorization is non-convex, there are no spurious local minima and the global optimality can be attained under some mild conditions.

Conventionally, the derivation of matrix completion methods is based on the Euclidean space. The aforementioned schemes can work well under the assumption of data without noise or corrupted by white Gaussian noise. However, non-Gaussian gross errors occur in many practical scenarios [26], [27]. For example, the salt-and-pepper noise is a common impulsive disturbance in images [28]. The recovery performance of traditional matrix completion algorithms will be degraded when the observations contain outliers. Hence, it is necessary to develop recovery schemes to resist outliers. In this study, outliers and gross errors refer to sparse corruptions whose magnitudes are much bigger than those of their vicinity. In [29], the authors replace the $\ell_2$-norm with the $\ell_p$-norm ($0 < p < 2$) to achieve robustness, and have proposed two algorithms, namely, iterative $\ell_p$-regression and $\ell_p$-alternating direction method of multipliers ($\ell_p$-ADMM). The $\ell_p$-norm has also been applied in low-rank tensor completion and robust PCA in [8] and [30], respectively. Although they attain outstanding performance in the presence of outliers, there are two open questions. The first issue is to decide $p$ because its appropriate value varies with the intensity of impulsive noise. Second, when $0 < p < 1$, $\ell_p$-norm is non-convex and non-smooth, which results in a challenging optimization problem. Furthermore, a new surrogate of $\ell_0$-'norm’ is devised to achieve low-rank recovery and combat outliers, but it introduces two additional user-defined parameters [31] and it is vital to choose their proper values. Two new robust loss functions are put forth in [32] to reduce the impact of outliers on the recovery performance, and they use a distributed optimization framework to tackle matrix completion. Huber function [33] as a common M-estimator is also applied on matrix completion to resist against gross errors and majorization minimization (MM) algorithm is used to jointly estimate its parameter vector and scale. Moreover, robust matrix factorization based on MM (RMF-MM) [34] is developed. The authors replace the $\ell_2$-norm with the $\ell_1$-norm to attain robustness, find a convex surrogate to relax the non-smooth and non-convex objective function, and then employ the MM technique as the solver. Recently, He et al. [37] adopt the non-convex Welsch cost function to handle outliers, and propose two robust
matrix completion algorithms, i.e., half-quadratic power factorization (HQ-PF) and half-quadratic alternating steepest descent (HQ-ASD). Furthermore, Bayesian methods, including variational Bayesian matrix factorization based on $\ell_1$-norm (VBMFL1) [35] and Bayesian estimator for noisy matrix completion (BENMC) [36], are suggested.

Truncated-quadratic loss function is non-convex and non-smooth, and has been utilized as the regularization term for edge-detection and denoising [38], [39], [43]. However, it is exploited as a non-convex loss function to resist outliers in this paper. Research results [32], [33], [37], [40] have shown that compared with the convex loss functions such as $\ell_1$-norm and Huber function, whose influence functions are not redescending, the non-convex penalty function can suppress large outliers well. Intuitively, truncated-quadratic function as the cost function is more robust than the $\ell_p$-norm and Huber function especially for large outliers, because the corresponding weights in the former will be assigned to zero. Besides, it is more preferable that loss functions only assign smaller weights for outliers, while keeping the weights of normal data fixed. Compared with other non-convex loss functions [41], i.e., Cauchy and Welsch, which down-weigh all observations including uncorrupted data, the truncated-quadratic function is better to combat outliers since it only down-weighs the entries contaminated by outliers. However, its non-smoothness and non-convexity create challenges in the optimization process.

In this study, to avoid full SVD calculation and resist gross errors, the factorization idea and truncated-quadratic cost function are jointly leveraged for robust matrix completion. Moreover, since it is difficult for MM technique to find the surrogate function [42], half-quadratic optimization (HO) theory [43], [44] is utilized. In order to seek the equivalent form of the original optimization problem, a new auxiliary term is introduced in the HO formulation to transform the non-convex and non-smooth problem into a tractable quadratic optimization problem, which permits to derive an efficient algorithm based on alternating minimization between the matrix entries and auxiliary variable. It is worth pointing out that from the HO theory, the truncated-quadratic function can be converted into two different optimization forms, namely, the additive and multiplicative formulations. Accordingly, two algorithms are devised, called HO based on additive form for truncated-quadratic function (HOAT) and HO based on multiplicative form for truncated-quadratic function (HOMT). In summary, the main contributions of our paper are as follows:

1) We derive the dual function and minimizer function of the truncated-quadratic function based on the HO theory.
2) We devise two robust factorization based matrix completion algorithms, namely, HOAT and HOMT, to resist gross errors. In particular, the observations detected as outliers are considered as missing entries in HOMT.
3) Experiments demonstrate that compared to the competing methods, the proposed algorithms exhibit better recovery performance for different real-world data including gray-scale and hyperspectral images, in the presence of outliers and Gaussian noise.

The remainder of this paper is organized as follows. In Section II, notations are provided, and matrix completion methods as well as HO theory are reviewed. The two developed approaches and their theoretical analysis, are presented in Section III. Section IV discusses the numerical simulation results based on synthetic data and real images. Finally, conclusions are drawn in Section V.
II. Preliminaries

In this section, notations, related works and HO theory are reviewed.

A. Notations

Italic, bold lower-case and bold upper-case letters denote scalars, vectors and matrices, respectively. The $i$th entry of a vector $a$ is represented by $a_i$, and the $(i,j)$ element of a matrix $A$ is denoted by $A_{ij}$. In particular, a matrix of zeros and ones are signified by $0$ and $1$, respectively. Besides, $\Omega \subset \{1, \cdots, m\} \times \{1, \cdots, n\}$ is used to denote the index set of the observed entries of $A$, $\Omega^c$ is the complement of $\Omega$, and $(\cdot)_\Omega$ is a projection operator, defined as:

$$[A_\Omega]_{ij} = \begin{cases} A_{ij}, & \text{if } (i,j) \in \Omega \\ 0, & \text{if } (i,j) \in \Omega^c \end{cases}$$

In addition, $\|a\|_2 = \sqrt{a^T a}$ is the $\ell_2$-norm of $a \in \mathbb{R}^m$, $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$ is the Frobenius norm of $A \in \mathbb{R}^{m \times n}$, and $\text{vec}(A)$ is a vector generated via vectorizing $A$. Moreover, $A_j$; and $A_{i,j}$ represent the $j$th row and $j$th column of $A$, respectively. Furthermore, $|a|$ stands for the absolute value of $a$, while $|\mathcal{I}|$ is the cardinality of the set $\mathcal{I}$. Finally, the pseudo-inverse operator and transpose operator are denoted by $(\cdot)^+$ and $(\cdot)^T$, respectively.

B. Related Work

1) Matrix Completion: To obtain a low-rank matrix $M$ to match the incomplete matrix $X_\Omega$, a direct formulation is [14]:

$$\min_M \text{rank}(M), \text{ s.t. } M_\Omega = X_\Omega$$

(1)

However, (1) is a NP-hard optimization problem. Thereby, nuclear norm [15] is suggested, leading to

$$\min_M \|M\|_*, \text{ s.t. } M_\Omega = X_\Omega$$

(2)

where the nuclear norm $\|M\|_*$ is the sum of singular values of $M$. Nevertheless, it requires performing full SVD at each iteration, which limits its application on large-size data. To avoid full SVD calculation, matrix factorization [22] is applied, resulting in:

$$\min_{U,V} \|X_\Omega - (UV)_\Omega\|_F^2$$

(3)

where $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{r \times n}$ are the two small size matrices with rank $r \ll \min(m,n)$. Apparently, $M = UV$ has low rank. However, its performance will be degraded in the presence of outliers.

On the other hand, in order to achieve robustness, a modified model [29] based on (3) via employing $\ell_p$-norm with $0 < p < 2$, is proposed:

$$\min_{U,V} \|X_\Omega - (UV)_\Omega\|_p^p$$

(4)

For the entry $(i,j) \in \Omega$ corrupted by outliers, its residual error $e = X_{i,j} - M_{i,j}$ in the $\ell_p$-norm space is less than the error in the Frobenius norm space, i.e., $|e|^p < |e|^2$ with $0 < p < 2$. Thus, (4) can reduce the impact of outliers, and the $\ell_p$-norm is more capable to resist gross errors than the Frobenius norm.
In addition, robust M-estimation based matrix completion method [33] is developed, formulated as:

\[ \min_{U,V} \| X_\Omega - (UV)_{\Omega} \|_{\sigma,c} \]  

where \( \| X_\Omega - (UV)_{\Omega} \|_{\sigma,c} = \sum_{(i,j) \in \Omega} \rho\left( \frac{X_{i,j} - (UV)_{i,j}}{\sigma} \right) \) with \((i,j) \in \Omega\), and \( \rho(\cdot) \) is the Huber function:

\[ \rho(x) = \begin{cases} \frac{1}{2} x^2, & |x| \leq c \\ c|x| - \frac{1}{2} c^2, & |x| > c \end{cases} \]  

However, Huber function is a ‘monotone M-estimator’ [49], namely, large outliers still affect its robustness, but the effect remains bounded.

2) **Half-Quadratic Theory:** HO is an efficient tool for convex and non-convex optimization based on the conjugate function theory. Let \( \phi(x) \) denote a loss function that satisfies the conditions of HO theory [43]–[45], and \( \phi(x) \) can be reformulated as the following model:

\[ \phi(x) = \inf_y Q(x, y) + \varphi(y) \]  

where \( y \) is an introduced auxiliary variable, \( \varphi(y) \) is the dual function of \( \phi(x) \), and \( Q(x, y) \) is a quadratic function of \( x \), which has two forms:

\[ Q(x, y) = \frac{(x-y)^2}{2}, \quad y \in \mathbb{R} \]  

and

\[ Q(x, y) = \frac{y \cdot x^2}{2}, \quad y \in \mathbb{R}_+ \]  

Accordingly, \( \phi(x) \) has two expressions, that is, the additive and multiplicative forms:

\[ \phi(x) = \inf_y \frac{(x-y)^2}{2} + \varphi(y) \]  

\[ \phi(x) = \inf_y \frac{y \cdot x^2}{2} + \varphi(y) \]  

In addition, for a fixed \( x \) in (10) and (11), there exists the corresponding minimizer function \( \delta_A(x) \) and \( \delta_M(x) \):

\[ \delta_A(x) := \arg \inf_y \frac{(x-y)^2}{2} + \varphi(y) \]  

\[ \delta_M(x) := \arg \inf_y \frac{y \cdot x^2}{2} + \varphi(y). \]  

Moreover, if \( \delta_A(x) \) and \( \delta_M(x) \) exist, the minimization of loss function \( \phi(x) \) is equal to:

\[ \min_{x,y} Q(x, y) + \varphi(y) \]  

which can be solved by alternating minimization. In other words, for a given \( x \), the optimal solution to \( y \) is obtained via (12) or (13). While given \( y \), (14) can also be easily solved since it is a quadratic optimization problem.
III. PROPOSED ALGORITHMS

In this section, we first derive the minimizer functions of the additive and multiplicative forms according to the definition of HO theory, and their corresponding robust matrix completion methods are then devised. In particular, the convergence and computational complexity of the developed algorithms are analyzed.

A. Robust Matrix Completion via Additive Form

The truncated-quadratic function has the form of [38]:

\[
\phi(x) = \begin{cases} 
  \frac{x^2}{2}, & |x|<e \\
  \frac{e^2}{2}, & |x| \geq e 
\end{cases}
\]  

(15)

Fig. 1 plots the truncated-quadratic function and typical loss functions. It is seen that the truncated-quadratic function expresses a wish to suppress big outliers, compared with other functions. However, it is non-convex and non-smooth, which results in an intractable optimization problem. In this study, HO theory is exploited to transform (15) into (10) via adding an auxiliary variable \( y \). According to Section II-B2, we obtain the minimizer function of the additive form of (15) first, i.e.,

\[
\delta_A(x) = \begin{cases} 
  0, & |x|<e \\
  x, & |x| \geq e 
\end{cases}
\]  

(16)

whose derivation is shown in Appendix A.

Motivated by (4) and (5), the truncated-quadratic function is adopted to achieve robustness in the presence of gross errors, formulated as:

\[
\min_{U,V} \|X_{\Omega} - (UV)_{\Omega}\|_\phi 
\]  

(17)

Since \( \| \cdot \|_\phi \) is an entry-wise operator, (17) can be reformulated as:

\[
\min_{U,V} \sum_{(i,j) \in \Omega} \phi \left( X_{i,j} - (UV)_{i,j} \right) 
\]  

(18)
Based on the additive form of HO theory in (10), (18) becomes:

$$\min_{U,V,N} \sum_{(i,j) \in \Omega} \frac{1}{2} \left( X_{i,j} - (UV)_{i,j} - N_{i,j} \right)^2 + \varphi(N_{i,j})$$  \hspace{1cm} (19)$$

where $N_{i,j}$ is the auxiliary variable to model the outlier component of noise, and is determined by the minimizer function $\delta_A(\cdot)$ in (16). Besides, $\varphi(N_{i,j})$ is the regularization term, whose expression is provided in Appendix A. We denote $\varphi(N) = \sum_{(i,j)} \varphi(N_{i,j})$, and then (19) can be rewritten as the matrix form:

$$\min_{U,V,N} \frac{1}{2} \| X_{\Omega} - (UV)_{\Omega} - N_{\Omega} \|_F^2 + \varphi(N_{\Omega})$$  \hspace{1cm} (20)$$

It is worth mentioning that (20) has three variables without constraints, and block coordinate descent (BCD) [46], [47] is adopted to find its solutions, whose idea is to optimize one factor while fixing the remaining two factors. Thus, (20) can be decomposed as the following three subproblems:

$$U^{k+1} = \arg \min_U \frac{1}{2} \| X_{\Omega} - (UV^k)_{\Omega} - N^k_{\Omega} \|_F^2$$  \hspace{1cm} (21)$$

$$V^{k+1} = \arg \min_V \frac{1}{2} \| X_{\Omega} - (U^{k+1}V)_{\Omega} - N^k_{\Omega} \|_F^2$$  \hspace{1cm} (22)$$

$$N^{k+1} = \arg \min_N \frac{1}{2} \| X_{\Omega} - (U^{k+1}V^{k+1})_{\Omega} - N_{\Omega} \|_F^2 + \varphi(N_{\Omega})$$  \hspace{1cm} (23)$$

It is easy to know that given $V^k$ and $N^k$, (21) can be decoupled as $m$ subproblems with respect to (w.r.t.) $u_i$ for $i = 1, \ldots, m$:

$$u_i^{k+1} = \arg \min_{u_i} \frac{1}{2} \| X_{i,J_i} - u_i^T V_{J_i}^k - N_{i,J_i}^k \|_2^2$$  \hspace{1cm} (24)$$

where $u_i^T$ represents the $i$th row of $U$, i.e.,

$$U = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix}$$  \hspace{1cm} (25)$$

and $J_i$ is denoted as the set of column indices of the observed entries in the $i$th row of $X$ with $\sum_{i=1}^m |J_i| = |\Omega|$.

Herein, we provide an example to determine $J_i$. If $X_{i,:} = [x_{i,1} \ 0 \ 0 \ x_{i,5}]$, then $J_i = \{1, 3, 5\}$, and $|J_i| = 3$. Besides, $V_{J_i}^k \in \mathbb{R}^{r \times |J_i|}$ contains $|J_i|$ columns indexed by $J_i$, namely,

$$V_{J_i}^k = \begin{bmatrix} V_{i,:J_{i1}} & V_{i,:J_{i2}} & \cdots & V_{i,:J_{i|J_i|}} \end{bmatrix}$$  \hspace{1cm} (26)$$

where $J_{ij}$ stands for the $j$th entry of $J_i$ with $j = 1, \ldots, |J_i|$. Similarly,

$$X_{i,J_i} = \begin{bmatrix} X_{i,J_{i1}} & X_{i,J_{i2}} & \cdots & X_{i,J_{i|J_i|}} \end{bmatrix}$$  \hspace{1cm} (27)$$

and

$$N_{i,J_i}^k = \begin{bmatrix} N_{i,J_{i1}}^k & N_{i,J_{i2}}^k & \cdots & N_{i,J_{i|J_i|}}^k \end{bmatrix}$$  \hspace{1cm} (28)$$

Thus, (24) is rewritten as:

$$u_i^{k+1} = \arg \min_{u_i} \frac{1}{2} \| D_{i,J_i}^k - u_i^T V_{J_i}^k \|_2^2$$  \hspace{1cm} (29)$$
where $D^k = X - N^k$. It is clear that (29) is a least squares problem, whose closed-form solution is:

$$u_i^{k+1} = \left( (V^k_{J_i})^T \right)^\dagger(D_i^{k,J_i})^T$$

and its computational complexity is $O(|J_i|r^2)$.

Problem (22) has a similar structure to (21), and hence can be solved in an analogous manner. Given $U^{k+1}$ and $N^k$, it is easy to observe that (22) can be decoupled w.r.t. $v_j$ for $j = 1, \cdots, n$, leading to:

$$v_j^{k+1} = \arg\min_{v_j} \frac{1}{2} \left\| D_i^{k,J_i} - U_i^{k+1} v_j \right\|^2_2$$

where $v_j$ is the $j$th column of $V$, $J_i$ is the set of row indices of the observed entries in the $j$th column of $X$, and $\sum_{j=1}^n |J_i| = |\Omega|$. Apparently, (31) is also a least squares problem, and its closed-form solution is:

$$v_j^{k+1} = \left( U_i^{k+1} \right)^\dagger_i D_i^{k,J_i}$$

whose computational complexity is $O(|J_i|r^2)$.

Finally, given $U^{k+1}$ and $V^{k+1}$, (23) is equal to:

$$N^{k+1} = \arg\min_N \frac{1}{2} \left\| R^{k+1} - N_\Omega \right\|^2_F + \varphi(N_\Omega)$$

where $R^{k+1} = X - U^{k+1}V^{k+1}$ represents the residual information in the $(k+1)$th iteration. In accordance to the HO theory, the solution to $N^{k+1}$ is:

$$N^{k+1} = \delta_A(R^{k+1})$$

As shown in (16), the minimizer function $\delta_A(x)$ depends on the parameter $e$ which controls robustness. That is to say, any entry in $R_\Omega$ bigger than $e$, will be considered as outliers. For a smaller $e$, (19) will reduce the impact of all the outliers, but also mistakenly regard the ‘normal’ entries as outliers. While for a larger $e$, it is possible that some outliers remain in the data, leading to the recovery performance degradation. For ‘normal’ entries, they refer to the elements in the data corrupted by zero-mean Gaussian noise, which are not the gross errors. Analogous to [48], we take the following strategy to choose the value of $e$:

$$e^k = \min \{ \xi \sigma^k, e^{k-1} \}$$

where $\xi > 0$ is a constant and $\sigma^k$ is the robust normalized median absolute deviation of the vectorized $R_\Omega$, i.e.,

$$\sigma^k = 1.4815 \times \text{Med}(|\text{vec}(R^k_\Omega)| - \text{Med}(\text{vec}(R^k_\Omega)))$$

with $\text{Med}(\cdot)$ being the sample median operator [49]. Besides, in order to achieve reliable recovery, (3) is first handled to provide a good initialization for the calculation of (21), although it is vulnerable to outliers, which is a similar strategy adopted in [37]. Actually, the solutions to $U$ and $V$ in (3) are same as those of (21) and (22), respectively, when $N = 0$ at each iteration. Hence, the initialized estimates of $U$ and $V$ in (3) are computed as:

$$u_i^{k+1} = \left( (V^k_{J_i})^T \right)^\dagger_i X_i^{k,J_i}, \; i = 1, \cdots, m$$

$$v_j^{k+1} = \left( U_i^{k+1} \right)^\dagger_i X_i^{k,J_i}, \; j = 1, \cdots, n$$
Algorithm 1 summarizes the detailed optimization procedure. As stated in [29], the convergence rate to finding the solutions to (37) and (38) is fast and it is not necessary to obtain the optimal solutions since (37) and (38) are only used to provide initialization for (30) and (34), not the final results. Thus, a small number of $I_p$ is enough, specifically, $I_p = 3$ is employed in our algorithm. In addition, Algorithm 1 will be terminated when its iteration number exceeds $I_m$ and/or $\kappa = \|X_\Omega - M_{\Omega}^{k} - N^k\|_F/\|X_\Omega\|_F$ is less than the pre-set threshold value $\eta > 0$. Note also that, the calculation of $u_i$ with $i = 1, \cdots, m$ in (30) and $v_j$ with $j = 1, \cdots, n$ in (32) are independent, and hence distributed and parallel realizations can be exploited to shorten computational time. Finally, the convergence of Algorithm 1 is analyzed in the following proposition. We first define $\mathcal{L}(U,V,N) = \frac{1}{2} \|X_\Omega - (UV)_\Omega - N\|_F^2 + \phi(N)$ as the objective function.

**Proposition 1.** The sequence $\{\mathcal{L}(U^k, V^k, N^k)\}$ produced by Algorithm 1 is convergent.

**Proof:** Since $e^k$ is updated at each iteration, we provide the proof via discussing the following two cases. First, when updating $e^k \rightarrow e^{k+1}$, $\phi(N)$ in $\mathcal{L}(U,V,N)$ is the only component related to $e$, and to make the development clearer, we write $\phi(N,e) = \phi(N)$. Besides, $\phi(N,e)$ is separable, namely, $\phi(N,e) = \sum_{i,j} \phi(N_{i,j}, e)$. Then, we have $\frac{\partial \phi(N_{i,j}, e)}{\partial e} > 0$ via taking the partial derivative of $\phi(N_{i,j}, e)$ w.r.t. $e$:

$$
\frac{\partial \phi(N_{i,j}, e)}{\partial e} = \begin{cases} |N_{i,j}|, & |N_{i,j}| < e \\ e, & |N_{i,j}| \geq e \end{cases}
$$

and we obtain:

$$
\frac{\partial \mathcal{L}}{\partial e} = \sum_{i,j} \frac{\partial \phi(N_{i,j}, e)}{\partial e} = \sum_{i,j} \frac{\partial \phi(N_{i,j}, e)}{\partial e} > 0
$$

Thus, $\mathcal{L}(U,V,N)$ increases monotonically w.r.t. $e$, and the updating rule in (35) makes $e$ non-increasing. Therefore, $\{\mathcal{L}(U^k, V^k, N^k)\}$ is non-increasing when updating $e^k \rightarrow e^{k+1}$.

Second, after obtaining $e^{k+1}$, since alternating minimization is used at each step in Algorithm 1, we have:

$$
\begin{align*}
\mathcal{L}(U^{k+1}, V^{k+1}, N^{k+1}) &= \min_V \frac{1}{2} \|X_\Omega - (U^{k+1}V)_\Omega - N^{k+1}\|_F^2 + \phi(N^{k+1}) \\
&\leq \min_U \frac{1}{2} \|X_\Omega - (UV)^k_\Omega - N^{k+1}\|_F^2 + \phi(N^{k+1}) \\
&\leq \min_N \frac{1}{2} \|X_\Omega - (U^kV)^k_\Omega - N\|_F^2 + \phi(N) \\
&\leq \frac{1}{2} \|X_\Omega - (U^kV)^k_\Omega - N\|_F^2 + \phi(N) \\
&= \mathcal{L}(U^k, V^k, N^k)
\end{align*}
$$

which means that $\mathcal{L}(U,V,N)$ is non-increasing. Since $\mathcal{L}(U,V,N)$ is lower bounded by 0, thus the sequence $\{\mathcal{L}(U^k, V^k, N^k)\}$ generated by Algorithm 1 is convergent.■
Algorithm 1 HOAT

Input: Incomplete matrix $X_\Omega$, index set $\Omega$, maximum allowable iteration numbers $I_p$ and $I_m$, and tolerance parameter $\eta$

Initialize: Generate a standard Gaussian matrix $V^0 \in \mathbb{R}^{r \times n}$, and determine $\{J_i\}_{i=1}^{m}$ and $\{J_j\}_{j=1}^{n}$ according to $\Omega$.

for $p = 1, 2, \cdots, I_p$ do
  // Fix $V^{p-1}$, optimize $U$
  Find $U^p$ according to (37)
  // Fix $U^p$, optimize $V$
  Find $V^p$ according to (38)
end for

Set $U^0 = U^p$ and $V^0 = V^p$

for $k = 1, 2, \cdots, I_m$ do
  Calculate $e^k$ according to (35)
  // Fix $U^{k-1}$ and $V^{k-1}$, optimize $N$
  Find $N^k$ according to (34)
  // Fix $V^{k-1}$ and $N^k$, optimize $U$
  Find $U^k$ according to (30)
  // Fix $U^k$ and $N^k$, optimize $V$
  Find $V^k$ according to (32)
  Stop, if a termination condition is satisfied.
end for

Output: $M = U^kV^k$.

B. Robust Matrix Completion via Multiplicative Form

For the multiplicative form of truncated-quadratic function in (11), we derive its minimizer function as (see Appendix B):

$$\delta_M(x) = \begin{cases} 
1, & |x| < e \\
0, & |x| \geq e 
\end{cases} \quad (42)$$

According to (11), (18) is equal to:

$$\min_{U, V, W} \sum_{(i,j) \in \Omega} \frac{1}{2} W_{i,j} \left( X_{i,j} - (UV)_{i,j} \right)^2 + \varphi(W_{i,j}) \quad (43)$$

where $W_{i,j}$ is a new auxiliary variable to assign the weights to different entries, and its value depends on the minimizer function $\delta_M(\cdot)$ in (42). In addition, we define $\varphi(W) = \sum_{(i,j)} \varphi(W_{i,j})$, then (43) can be rewritten as:

$$\min_{U, V, W} \frac{1}{2} \left\| \sqrt{W_{\Omega}} \circ (X_{\Omega} - (UV)_{\Omega}) \right\|_F^2 + \varphi(W_{\Omega}) \quad (44)$$
where \( \sqrt{\cdot} \) is the element-wise square root operator, and \( \circ \) denotes the entry-wise product. In particular, \( W_{\Omega^c} = 0 \).

BCD is also used to seek the solutions to (44), and thus we can decompose (44) into three subproblems:

\[
U^{k+1} = \arg \min_U \frac{1}{2} \left\| \sqrt{W_{\Omega}} \circ (X_{\Omega} - (UV^k)_{\Omega}) \right\|_F^2
\]

(45)  

\[
V^{k+1} = \arg \min_V \frac{1}{2} \left\| \sqrt{W_{\Omega}} \circ (X_{\Omega} - (U^{k+1}V)_{\Omega}) \right\|_F^2
\]

(46)  

\[
W^{k+1} = \arg \min_W \frac{1}{2} \left\| \sqrt{W_{\Omega}} \circ (X_{\Omega} - (U^{k+1}V^{k+1})_{\Omega}) \right\|_F^2 + \varphi(W_{\Omega})
\]

(47)

Note that (45) and (46) are weighted least squares problems, whose computational complexity is high since their weights are updated via (47) at each iteration. However, it is easy to observe that the entry of \( W_k \) is either ‘0’ or ‘1’, thus we can set \( \Omega^k = W^k \) as the current observation index set of \( X \), that is to say, the observations contaminated by outliers are taken as missing entries. Accordingly, (45), (46) and (47) are reformulated as:

\[
U^{k+1} = \arg \min_U \frac{1}{2} \left\| X_{\tilde{\Omega}^k} - (UV^k)_{\tilde{\Omega}^k} \right\|_F^2
\]

(48)  

\[
V^{k+1} = \arg \min_V \frac{1}{2} \left\| X_{\tilde{\Omega}^k} - (U^{k+1}V)_{\tilde{\Omega}^k} \right\|_F^2
\]

(49)  

\[
\tilde{\Omega}^{k+1} := \arg \min_W \frac{1}{2} \left\| \sqrt{W_{\Omega}} \circ (X_{\Omega} - (U^{k+1}V^{k+1})_{\Omega}) \right\|_F^2 + \varphi(W_{\Omega})
\]

(50)

To solve (48)-(50), we first recall that \( \tilde{\mathcal{J}}^k_i \) and \( \tilde{\mathcal{J}}^k_j \) are the set of column indices of the observed entries in the \( i \)th row of \( X_{\tilde{\Omega}^k} \) and the set of row indices of the observed entries in the \( j \)th column of \( X_{\tilde{\Omega}^k} \) at the \( k \)th iteration, respectively. Problem (48) can be decomposed as the following \( m \) subproblems:

\[
u_{i}^{k+1} = \arg \min_{u_i} \frac{1}{2} \left\| X_{i,\tilde{\mathcal{J}}^k_i} - u_i^T \mathbf{v}_{\tilde{\mathcal{J}}^k_i} \right\|_2^2
\]

(51)  

where \( i = 1, \cdots, m \), and the solution is:

\[
u_{i}^{k+1} = \left( (\mathbf{v}_{\tilde{\mathcal{J}}^k_i})^T \right)^T X_{i,\tilde{\mathcal{J}}^k_i}
\]

(52)

whose computational complexity is \( O(|\tilde{\mathcal{J}}^k_i|^2) \).

Similarly, (49) can be separated into \( n \) independent subproblems:

\[
u_{j}^{k+1} = \arg \min_{v_j} \frac{1}{2} \left\| X_{\tilde{\mathcal{J}}^k_j, j} - U_{\tilde{\mathcal{J}}^k_j} \mathbf{v}_j \right\|_2^2
\]

(53)  

where \( j = 1, \cdots, n \), and its solution is:

\[
u_{j}^{k+1} = \left( U_{\tilde{\mathcal{J}}^k_j}^T \right)^T X_{\tilde{\mathcal{J}}^k_j, j}
\]

(54)

whose computational complexity is \( O(|\tilde{\mathcal{J}}^k_j|^2) \). Given \( U^{k+1} \) and \( V^{k+1} \), (50) can be rewritten as:

\[
W^{k+1} = \arg \min_W \frac{1}{2} \left\| \sqrt{W_{\Omega}} \circ (R_{\Omega^c}^{k+1}) \right\|_F^2 + \varphi(W_{\Omega})
\]

(55)

According to (13), its solution is:

\[
\tilde{\Omega}^{k+1} = W^{k+1} = \delta_{M} (R_{\Omega^c}^{k+1})
\]

(56)
Algorithm 2 summarizes the matrix completion procedure based on HOMT. We take the same strategy as Algorithm 1 to terminate Algorithm 2 and determine its parameters including \( I_p \), \( I_m \), \( \eta \) and \( e \). In addition, we use the solution to (3) as the initialization for (48) and (50). It is worth noting that the processes of finding the solutions to different \( u_i \) in (52) for \( i = 1, \cdots, m \) and \( v_j \) in (54) for \( j = 1, \cdots, n \), are independent of each other, (56) is entry-wise operator, and hence those steps can be calculated in parallel to reduce the computational time.

Furthermore, we denote \( C(U, V, W) = \frac{1}{2} \| \sqrt{W} \odot (X - (UV)) \|_F^2 + \varphi(W) \), and the convergence analysis of Algorithm 2 is provided in Proposition 2.

Algorithm 2 HOMT

**Input:** Incomplete matrix \( X_\Omega \), index set \( \Omega \), maximum allowable iteration numbers \( I_p \) and \( I_m \), and tolerance parameter \( \eta \)

**Initialize:** Generate a standard Gaussian matrix \( V^0 \in \mathbb{R}^{r \times n} \), and find \( \{ J_i \}_{i=1}^m \) and \( \{ I_j \}_{j=1}^n \) according to \( \Omega \).

for \( p = 1, 2, \cdots, I_p \) do
// Fix \( V^{p-1} \), optimize \( U \)
Find \( U^p \) according to (37)
// Fix \( U^p \), optimize \( V \)
Find \( V^p \) according to (38)
end for

Set \( U^0 = U^p \) and \( V^0 = V^p \)

for \( k = 1, 2, \cdots, I_m \) do
Calculate \( e^k \) according to (35)
// Fix \( U^{k-1} \) and \( V^{k-1} \), optimize \( \tilde{\Omega} \)
Find \( \tilde{\Omega}^k \) according to (56)
Update \( \tilde{J}_i^k \) and \( \tilde{I}_j^k \) according to \( \tilde{\Omega}^k \)
// Fix \( V^{k-1} \) and \( \tilde{\Omega}^k \), optimize \( U \)
Find \( U^k \) according to (52)
// Fix \( U^k \) and \( \tilde{\Omega}^k \), optimize \( V \)
Find \( V^k \) according to (54)
Stop, if a termination condition is satisfied.
end for

**Output:** \( M = U^k V^k \).

**Proposition 2.** The sequence \( \{ C(U^k, V^k, W^k) \} \) generated by Algorithm 2 converges.

**Proof:** Similar to Proposition 1, since \( \varphi(W) \) in \( C(U, V, W) \) is the only component related to \( e \) and is separable,
we first obtain:
\[
\frac{\partial \phi(W_{i,j}, e)}{\partial e} = \begin{cases} 
(1 - W_{i,j})e, & 0 \leq W_{i,j} < 1 \\
0, & y \geq 1 
\end{cases}
\]
and then we have:
\[
\frac{\partial \mathcal{L}}{\partial e} = \frac{\partial}{\partial e} \sum_i \sum_j \phi(N_{i,j}, e) = \sum_i \sum_j \frac{\partial \phi(N_{i,j}, e)}{\partial e} \geq 0
\]
Thus, for the non-increasing $e^{k+1}$ in (35), $C(U, V, W)$ is non-increasing.

Moreover, when $e^k$ has been updated, since BCD is utilized to find the solution to (44), hence we have:
\[
C(U^{k+1}, V^{k+1}, W^{k+1})
= \min_{V} \left\{ \frac{1}{2} \left\| \sqrt{W^{k+1}} \odot \left( X - (U^{k+1}V) \right) \right\|_F^2 + \phi(W^{k+1}) \right\}
\leq \min_{U} \left\{ \frac{1}{2} \left\| \sqrt{W^{k+1}} \odot \left( X - (UV^k) \right) \right\|_F^2 + \phi(W^{k+1}) \right\}
\leq \min_{W} \left\{ \frac{1}{2} \left\| \sqrt{W^k} \odot \left( X - (U^kV^k) \right) \right\|_F^2 + \phi(W^k) \right\}
= C(U^k, V^k, W^k)
\]
implying that $C(U, V, W)$ is non-increasing with a lower bound, thus the sequence $\{C(U^k, V^k, W^k)\}$ produced via Algorithm 2 converges.

\[\square\]

C. Computational Complexity

Although both Algorithms 1 and 2 introduce an auxiliary variable and utilize BCD to solve their corresponding optimization problems, they have different computational costs. For Algorithm 1, the complexity of least squares contains the cost of solving (21) by (30) and the cost of solving (22) by (32). The resultant complexity is $\sum_{i=1}^m O(|I_i|r^2) + \sum_{j=1}^n O(|I_j|r^2) = O\left(2|\Omega|r^2\right)$ at each iteration. Besides, the cost of solving (23) is $O(|\Omega|)$. Assuming that Algorithm 1 requires $K_1$ iterations to converge, its total computational complexity is $O\left(K_1(2r^2 + 1)|\Omega|\right)$. Similarly, for Algorithm 2, the complexity of least squares is $\sum_{i=1}^m O(|\tilde{I}_i|n^2 + \sum_{j=1}^n O(|\tilde{I}_j|n^2) = O\left(2|\tilde{\Omega}|n^2\right)$ in the $k$th iteration, which is not bigger than $O\left(2|\Omega|r^2\right)$ due to $|\tilde{\Omega}| \leq |\Omega|$ at each iteration, while the cost of solving (47) is $O\left(|\Omega|\right)$. Accordingly, the total computational complexity of Algorithm 2 is $\sum_{k=1}^{K_2} O\left(2|\tilde{\Omega}|n^2 + |\Omega|\right)$, where $K_2$ is the iteration number for convergence. As analyzed in [45], the iteration number of the multiplicative form is less than that of the additive form, and hence the running time of the multiplicative form will be less than that of the additive form.

IV. RESULTS

In this section, compared with the state-of-the-art robust matrix completion methods, namely, $\ell_1$-ADMM [29], RMC-Huber [33], RMF-MM [34], VBMFL$_3$ [35] and HQ-PF [37], we perform experiments on synthetic data, real
gray-scale and hyperspectral images, to show the effectiveness of the two proposed algorithms. For the parameters in the competing algorithms, the recommended setting is adopted, and if it is not available, we select the best parameters through experiments. In addition, we set $I_m = 50$ and $\eta = 1 \times 10^{-6}$ for the developed methods. All the experiments are conducted on a computer with 3.2 GHz CPU and 16 GB memory.

A. Results of Synthetic Random Data

The experimental strategy in [29], [32], [37] is adopted. We first generate two random matrices $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{r \times n}$, whose entries satisfy the standard Gaussian distribution, and then construct the synthetic matrix as $X = UV$. For convenience, we set $m = n$, and $r = m/50$. Unless stated otherwise, $m = 400$, and the observation ratio is 50%. The incomplete matrix $X_\Omega$ is perturbed via Gaussian mixture model (GMM). The probability density function of GMM is:

$$p_v(v) = \frac{1 - c}{\sqrt{2\pi}\sigma_1} \exp \left( -\frac{v^2}{2\sigma_1^2} \right) + \frac{c}{\sqrt{2\pi}\sigma_2} \exp \left( -\frac{v^2}{2\sigma_2^2} \right)$$

(60)

where $\sigma_1^2$ and $\sigma_2^2$ are variances with $\sigma_1^2 \ll \sigma_2^2$. Besides, $c$ controls the ratio of outliers, namely, a large value of $c$ gives rise to higher ratio of outliers. In our experiments, to model gross errors, we set $\sigma_2^2 = 100\sigma_1^2$ and $c = 0.1$. Thus, the signal-to-noise ratio (SNR) is given by:

$$\text{SNR} = \frac{\|X_\Omega\|_F^2}{|\Omega|((1-c)\sigma_1^2+c\sigma_2^2)}$$

(61)

In addition, we set SNR = 10 dB unless stated otherwise. To evaluate the recovery performance, the root mean square error (RMSE) is employed, defined as:

$$\text{RMSE} = \frac{\|X - M\|_F}{\sqrt{m \times n}}$$

(62)

which is calculated based on 100 independent runs. It is easy to understand that a smaller value of RMSE implies a better recovery performance.

Fig. 2 plots RMSE versus iteration number by the seven methods, namely, HOAT, HOMT, VBMFL$_1$, RMC-Huber, $\ell_1$-ADMM, RMF-MM and HQ-PF, at SNR = 10 dB. It is observed that the recovery accuracy of our algorithms and HQ-PF is higher than that of the remaining schemes. It is because the penalty functions of the former are non-convex, while those of the latter are convex. Compared with $\ell_2$-norm, although $\ell_1$-norm and Huber function can reduce the influence of outliers, their ability to resist outliers is restricted especially for large outliers. Besides, the reason why HOAT and HOMT are superior to HQ-PF is that the latter changes the weights of ‘normal’ data, even though they all utilize the dual function and Frobenius norm to resist outliers and suppress Gaussian noise, respectively. Compared with the competing methods, the proposed algorithms not only resist gross errors, but also do not change the weights of ‘normal’ data.

Fig. 3 depicts the RMSE versus percentage of observations. We easily see that the HOAT, HOMT and HQ-PF have better recovery performance for all observation percentages. It is because they adopt non-convex cost functions to resist gross errors, and researches in [33], [40] show that non-convex penalty functions have a higher capability to resist big outliers than convex penalty functions. The impact of SNR on different algorithms is plotted in Fig. 4. We see that the RMSEs of all the algorithms decrease as SNR increases, and our algorithms perform the best.
Then, the relationship between the RMSE and matrix column $m$ is investigated. As can be seen from Fig. 5, our algorithms have the best recovery performance for all $m$, although the RMSEs of all the methods increase with $m$. Compared with HOAT, HOMT and HQ-PF, large outliers still affect the recovery accuracy of VBMFL$_1$, RMC-Huber, $\ell_1$-ADMM and RMF-MM, since they utilize convex functions, i.e., $\ell_1$-norm and Huber function, as the penalties. In addition, we present the average runtime of different approaches in Fig. 6. It is easy to see that the computational cost of HOMT is the lowest among all the algorithms, while RMF-MM is the most time-consuming, which is approximately two orders of magnitude slower than HOAT and HOMT. Besides, the two proposed algorithms are at least an order of magnitude faster than HQ-PF though they all use non-convex cost functions and alternating minimization to find the solutions. In particular, the running time of HOAT is longer than that of HOMT, which is consistent with our analysis in Section III-C.

Fig. 7 depicts the selection of $\xi$ in (35). We see that the two proposed approaches can work well in a wide range of $\xi$. In particular, when $\xi$ is very small, our algorithms have larger RMSEs, because a smaller $\xi$ will cause
them wrongly take the ‘normal’ values as outliers, leading to a large degradation in recovery performance. While a much bigger value of $\xi$ also gives rise to performance degradation since it may consider outliers as ‘normal’ values. When $2.5 \leq \xi \leq 4$, HOAT and HOMT have a higher recovery accuracy, and a larger value of $\xi$ may contain more useful details if all the outliers are excluded. Besides, it is observed that the selection of $\xi$ is not sensitive to the matrix size $m$.

Finally, since our algorithms require knowing the rank information, it is necessary to provide a method to estimate its value. Similar to [29], cross-validation is used. In our experiments, the matrix column length is $m = 400$ and the true rank is set to 8. Furthermore, we randomly choose 95% observed entries as the training set and the remaining 5% are considered as the test data. Fig. 8 plots the RMSE versus rank $r$. We can see that all the methods can correctly estimate the matrix rank via cross-validation. It is worth noting that when the estimated rank is less than the true rank, it will bring about a big recovery error.
B. Application on Image Inpainting

In some cases, the received images are incomplete because of the damage of photosensitive devices or the occlusion of other objects. Besides, the images may be contaminated by Gaussian noise and impulsive noise during the process of wireless transmission. In this section, we apply the robust matrix completion algorithms on image inpainting, and the images in [50], [51] are tested. To model outliers, a mixture of zero-mean Gaussian noise and impulsive noise is adopted. The standard deviation of zero-mean Gaussian noise is set to be 0.02, and the impulsive noise is generated by the built-in command of ‘imnoise(I, ‘salt & pepper’, ρ)’ in MATLAB, where I is the incomplete matrix, and ρ is the normalized noise intensity. The relationship between ρ and SNR is ρ = 1/SNR and SNR = 10 dB. Moreover, to measure the recovery performance of different algorithms, two evaluation indices, namely, peak SNR (PSNR) and structural similarity index measure (SSIM), are employed, and we directly use the built-in commands ‘psnr (recovered, original)’ and ‘ssim (recovered, original)’ in MATLAB for their calculation, respectively. In particular, all the algorithms use the same rank, and ξ = 4 in the proposed algorithms. Furthermore,
we investigate the impact of three kinds of masks [50], that is, random mask, text mask and block mask, on robust matrix completion, and all the methods are tested on 10 dB Gaussian noise, and the mixture noise for each mask.

We first address a relatively easy robust matrix completion problem, where the missing entries of an image are uniformly distributed. The image called Scenery in the top left of Fig. 9, is used, whose dimensions are $300 \times 300$. Then, 20% of pixels of the image are randomly removed and two kinds of noise, i.e., Gaussian noise and the mixture noise, are added to the incomplete image in order to generate two different corrupted images. Table I tabulates the average recovery performance and running time for the two noise types. To further evaluate the recovery of different algorithms, their restored images in the mixture noise are shown in Fig. 9. It is observed that compared to VBMFL$_1$, RMC-Huber, $\ell_1$-ADMM and RMF-MM, the non-convex penalty algorithms, i.e., HOAT, HOMT and HQ-PF, achieve better reconstruction performance. Besides, although HQ-PF has a comparable recovery accuracy with HOAT and HOMT, it requires longer runtime. In particular, HOAT and HOMT outperform the competing algorithms in Gaussian noise because when there are no outliers, the penalty function used in the former is equal to $\ell_2$-norm for a bigger value of $\epsilon$, compared with the remaining algorithms.

**TABLE I: Performance comparison for random mask: average PSNR, SSIM and running time.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Gaussian noise</th>
<th>Mixture noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR  SSIM  Runtime</td>
<td>PSNR  SSIM  Runtime</td>
</tr>
<tr>
<td>HOAT</td>
<td>23.250 0.3944 1.099</td>
<td>28.144 0.7504 3.196</td>
</tr>
<tr>
<td>HOMT</td>
<td><strong>23.254</strong> 0.3949 0.938</td>
<td>28.036 0.7497 2.672</td>
</tr>
<tr>
<td>VBMFL$_1$ [35]</td>
<td>23.157 0.3886 3.191</td>
<td>27.862 0.7390 12.466</td>
</tr>
<tr>
<td>RMC-Huber [33]</td>
<td>22.039 0.3264 9.926</td>
<td>27.817 0.7356 38.693</td>
</tr>
<tr>
<td>$\ell_1$-ADMM [29]</td>
<td>21.459 0.2959 21.799</td>
<td>26.720 0.6839 163.228</td>
</tr>
<tr>
<td>RMF-MM [34]</td>
<td>21.576 0.3006 94.867</td>
<td>27.703 0.7270 289.060</td>
</tr>
<tr>
<td>HQ-PF [37]</td>
<td>23.158 0.3321 29.999</td>
<td>28.128 0.7352 67.798</td>
</tr>
</tbody>
</table>
The second is the text mask where the part covered by text in the image is considered as missing entries, which is more common than the first scenario, but its recovery is more difficult because the concealed regions may contain important details. As shown in Fig. 10, the Windows image is employed and corrupted by the text ‘Matrix’ ‘Completion’ as the incomplete image, whose dimensions are $349 \times 366$. Again, the same intensity of Gaussian noise and mixture noise are used to contaminate the incomplete image. Fig. 10 shows the recovered images by the seven approaches in the presence of the mixture noise. Table II shows their average PSNR, SSIM and running time for the two noise models. We easily observe that the HOAT and HOMT are superior to the remaining methods in the presence of Gaussian noise and mixture noise. Note that the competing algorithms cannot restore the light on the wall in Fig. 10. In addition, HOMT is comparable with HOAT in terms of PSNR and SSIM, but the former needs less complexity.

Block mask is also difficult to tackle since the resultant missing pattern is not random as well. The Stripe image, whose dimensions are $272 \times 271$, is covered by different shaped blocks and corrupted via the Gaussian noise or mixture noise. Fig. 11 depicts the recovered images in the presence of the mixture noise, and Table III shows their average recovery performance and runtime. Again, compared with the VBMFL$_1$, RMC-Huber, $\ell_1$-ADMM, RMF-MM and HQ-PF, our algorithms are more robust to Gaussian noise and mixture noise, producing the best recovery performance in terms of PSNR as well as SSIM, and requiring the least computational time.

C. Application on Hyperspectral Imaging

Hyperspectral imaging (HSI) plays an important role in many applications including earth climate, agriculture and urban planning [4], [5], [52]. However, it may suffer from annoying performance degradation during the acquisition
process such as missing data and being contaminated by Gaussian noise and impulsive noise, due to photon effects and calibration error. Hence, there is a need to improve the HSI quality prior to its subsequent processing. In this section, Indian Pines [52] and Japser Ridge datasets are used, whose dimensions are $145 \times 145$ pixels per band with 200 bands and $100 \times 100$ pixels per band with 198 bands, respectively. Then, we reconstruct a matrix $X \in \mathbb{R}^{21025 \times 200} / \mathbb{R}^{10000 \times 198}$, whose columns comprise vectorized bands of the Indian Pines/Japser Ridge data. Furthermore, 40% of pixels in $X$ are randomly missing. 10 dB impulsive noise produced by GMM is added to the incomplete data, and $\xi = 3$ is chosen in the proposed algorithms. Fig. 12 displays the restoration results of

![Image](http://lesun.weebly.com/hyperspectral-data-set.html)

**TABLE II: Performance comparison for text mask: average PSNR, SSIM and running time.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Gaussian noise</th>
<th>Mixture noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR SSIM Runtime</td>
<td>PSNR SSIM Runtime</td>
</tr>
<tr>
<td>HOAT</td>
<td>24.202 0.6520 4.6316</td>
<td>28.116 0.8523 1.129</td>
</tr>
<tr>
<td>HOMT</td>
<td>24.200 0.6520 4.139</td>
<td>28.083 0.8522 0.729</td>
</tr>
<tr>
<td>VBMFL [35]</td>
<td>23.920 0.6387 6.663</td>
<td>27.747 0.8477 5.889</td>
</tr>
<tr>
<td>RMC-Huber [33]</td>
<td>23.064 0.5933 16.632</td>
<td>27.804 0.8469 19.456</td>
</tr>
<tr>
<td>$\ell_1$-ADMM [29]</td>
<td>22.323 0.5509 44.026</td>
<td>26.597 0.8073 27.982</td>
</tr>
<tr>
<td>RMF-MM [34]</td>
<td>22.532 0.5658 389.137</td>
<td>27.670 0.8435 148.132</td>
</tr>
<tr>
<td>HQ-PF [37]</td>
<td>24.142 0.6024 127.571</td>
<td>28.045 0.8449 40.394</td>
</tr>
</tbody>
</table>
Fig. 11: Image recovery results with block mask by different algorithms

TABLE III: Performance comparison for block mask: average PSNR, SSIM and running time.

<table>
<thead>
<tr>
<th>Method</th>
<th>Gaussian noise</th>
<th>Mixture noise</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>SSIM</td>
<td>Runtime</td>
<td>PSNR</td>
</tr>
<tr>
<td>HOAT</td>
<td>25.043</td>
<td>0.7442</td>
<td>1.032</td>
<td>29.150</td>
</tr>
<tr>
<td>HOMT</td>
<td>25.040</td>
<td>0.7441</td>
<td>0.948</td>
<td>29.175</td>
</tr>
<tr>
<td>VBMFL [35]</td>
<td>24.429</td>
<td>0.7101</td>
<td>1.993</td>
<td>28.980</td>
</tr>
<tr>
<td>RMC-Huber [33]</td>
<td>24.025</td>
<td>0.6958</td>
<td>14.090</td>
<td>28.859</td>
</tr>
<tr>
<td>$\ell_1$-ADMM [29]</td>
<td>23.206</td>
<td>0.6521</td>
<td>45.016</td>
<td>27.380</td>
</tr>
<tr>
<td>RMF-MM [34]</td>
<td>23.366</td>
<td>0.6637</td>
<td>77.555</td>
<td>28.837</td>
</tr>
<tr>
<td>HQ-PF [37]</td>
<td>25.018</td>
<td>0.7060</td>
<td>37.762</td>
<td>29.070</td>
</tr>
</tbody>
</table>

different methods for the Indian Pines data. It is seen that compared with VBMFL, $\ell_1$-ADMM, and RMF-MM, the four methods, namely, HOAT, HOMT, RMC-Huber and HQ-PF, achieve better recovery results and our algorithms outperform the RMC-Huber and HQ-PF in terms of PSNR and SSIM. Moreover, the recovered images for Jasper Ridge data are shown in Fig. 13. Again, the HOAT and HOMT are superior to the competing algorithms due to higher PSNR and SSIM.

V. CONCLUSION

In this paper, the truncated-quadratic function, which is non-convex and non-smooth, is employed to resist outliers and HO theory is adopted to convert the loss function into two forms, that is, additive and multiplicative models, via
introducing an auxiliary variable, giving rise to two easy-to-solve optimization problems. Accordingly, we propose two robust matrix completion algorithms, called HOAT and HOMT. In particular, the elements contaminated by outliers in HOMT are considered as the missing entries, which can reduce its computational cost. Besides, BCD is utilized to find their solutions, and we show that the computational complexity of the additive form is higher than that of the multiplicative form, implying that the latter is preferred, if possible. Furthermore, extensive experimental results based on synthetic and real-world data demonstrate the superiority of the HOAT and HOMT over the competing algorithms in terms of RMSE or PSNR/SSIM, and running time.

REFERENCES

Fig. 13: HSI restoration results for the 80th band.


APPENDIX A

DERIVATION OF MINIMIZER FUNCTION OF ADDITIVE FORM

Based on (10), we obtain the dual function of (15) [43]:

$$\phi(y) = \sup_x -\frac{(x - y)^2}{2} + \phi(x)$$  \hspace{2cm} (63)

which is determined by $\phi(x)$. Substituting (15) in (63) yields:

$$\phi(y) = \begin{cases} 
\frac{-|y|e^2}{2} + \frac{e^2}{2}, & |y| < e \\
\frac{e^2}{2}, & |y| \geq e
\end{cases}$$  \hspace{2cm} (64)

Then, plugging (64) into (10) leads to:

$$\inf_y \frac{(x - y)^2}{2} + \phi(y) = \begin{cases} 
x^2, & |x| < e \\
\frac{x^2}{2}, & |x| \geq e
\end{cases}$$  \hspace{2cm} (65)

We denote the minimizer function associated with (65) as $\delta_A(x)$:

$$\delta_A(x) = \arg \inf_y \frac{(x - y)^2}{2} + \phi(y) = \begin{cases} 
0, & |x| < e \\
x, & |x| \geq e
\end{cases}$$  \hspace{2cm} (66)

APPENDIX B

DERIVATION OF MINIMIZER FUNCTION OF MULTIPLICATIVE FORM

First, the dual function of (15) for multiplicative form is [44]:

$$\phi(y) = \sup_x -\frac{yx^2}{2} + \phi(x)$$  \hspace{2cm} (67)

Then, substituting (15) into (67), we have:

$$\phi(y) = \begin{cases} 
\frac{1-y}{2}e^2, & 0 \leq y < 1 \\
0, & y \geq 1
\end{cases}$$  \hspace{2cm} (68)

Hence,

$$\inf_y \frac{yx^2}{2} + \phi(y) = \begin{cases} 
\frac{x^2}{2}, & |x| < e \\
\frac{e^2}{2}, & |x| \geq e
\end{cases}$$  \hspace{2cm} (69)
The minimizer function $\delta_M(x)$ related to (69) is:

$$\delta_M(x) : = \begin{cases} 
1, & |x| < e \\
0, & |x| \geq e 
\end{cases}$$

(70)