Robust Low-Rank Matrix Recovery as Mixed Integer Programming via ℓ_0-Norm Optimization

Zhang-Lei Shi, Xiao Peng Li, Weiguo Li, Tongjiang Yan, Jian Wang, *Senior Member, IEEE*, and Yaru Fu, *Member, IEEE*

***Abstract*—This letter focuses on the robust low-rank matrix recovery (RLRMR) in the presence of gross sparse outliers. Instead of using ℓ_1-norm to reduce or suppress the influence of anomalies, we aim to eliminate their impact. To this end, we model the RLRMR as a mixed integer programming (MIP) problem based on the ℓ_0-norm. Then, a block coordinate descent (BCD) algorithm is developed to iteratively solve the resultant MIP. At each iteration, the proposed approach first utilizes the ℓ_0-norm optimization theory to assign binary weights to all entries of the residual between the known and estimated matrices. With these binary weights, the optimization over the bilinear term is reduced to a weighted extension of the Frobenius norm. As a result, the optimization problem is decomposed into a group of row-wise and column-wise subproblems with closed-form solutions. Additionally, the convergence of the proposed algorithm is studied. Simulation results demonstrate that the proposed method is superior to five state-of-the-art RLRMR algorithms.**

***Index Terms*—Robust low-rank matrix recovery, mixed integer programming, binary optimization, ℓ_0-norm optimization.**

I. Introduction

LOW-RANK matrix recovery (LRMR) is a process of discovering the latent structure from the given high-dimensional data [1], [2], [3], [4]. The success of LRMR roots from that high-dimensional data should have a low-dimensional structure [1], [5], such as low-rank property [6], sparseness in some basis [7]. As a foundation in computer vision and machine learning, LRMR has attracted great interest recently in text classification [8], subspace segmentation [9], image denoising [10], and so on [11], [12], [13], [14], [15]. However, the observations in practical applications are often interfered with small dense noise and/or even gross sparse outliers. Therefore, noisy observations may ruin the recovery without proper technique against noises.

To address light dense noises, e.g., zero-mean Gaussian noises, it is natural for LRMR to seek solution in the Euclidean space [16]. That is, the LRMR is formulated as a least squares problem to minimize the squared Frobenius norm of residual between the recovered matrix and noisy observations. Under the Gaussian noise assumption, the singular value decomposition (SVD) provides the optimal recovery to the target matrix [17]. However, impulsive noises or even outliers with arbitrary magnitudes are unavoidable in practice [17], [18], [19], [20]. When data is polluted by anomalies, the Frobenius norm can be easily governed by errors caused by outliers [6]. Therefore, the restoration performance will become unsatisfactory or erroneous.

To achieve robustness against outliers, the ℓ_1-norm, namely, the sum of absolute values of matrix elements, is widely recommended [1], [18], [21], [22], [23]. The motivation is that only a small portion of the observations are corrupted by gross errors. That is, the noisy observations can be viewed as a sum of a low-rank target component and a sparse component [1], [18]. Then, a conceptual idea is to seek a low-rank matrix that approximates the target matrix, and the residual between the recovered and observation matrices is restricted to be sparse. Another reason is that, compared to the Frobenius norm, the ℓ_1-norm based objective function can suppress the impact of outliers instead of magnifying it [24]. Thus, one pipeline for robust LRMR (RLRMR) jointly minimizes the nuclear norm and ℓ_1-norm of the residual [1], [18]. However, for some cases such as image recovery, the rank can be estimated in advance, where the nuclear norm is not easy to incorporate the prior rank information [17]. In this case, an alternative is to minimize the ℓ_1-norm of the residual only, and the low-rank part is expressed as the factorization form [21], [22].

Despite the great success of ℓ_1-norm based methods, they are limited by drawbacks such as memory requirement and computation efficiency due to the nonsmoothness of ℓ_1-norm [25]. To overcome these limitations, weighted extension of the Frobenius norm is proposed [6], [16]. Specifically, the weighted algorithms aim at assigning small weights to outlier-polluted entries and large weights to clean entries of the residual between the known and recovered matrices. In this way, the objective function remains the smooth Frobenius norm, making the RLRMR model computationally attractive. However, the above-mentioned frameworks can only reduce or suppress the influence of outliers, rather than eliminate it entirely. Hence, anomalies will still degrade the recovery.

In this letter, we aim to eliminate the impact of outliers by formulating the RLRMR as a mixed integer programming (MIP)

Manuscript received 26 April 2023; revised 23 June 2023; accepted 23 June 2023. Date of publication 2 August 2023; date of current version 9 August 2023. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 22CX06039A and in part by Research Matching Grant under Grant CP/2022/2.1. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lu Gan. *(Corresponding author: Jian Wang.)*

Zhang-Lei Shi, Weiguo Li, Tongjiang Yan, and Jian Wang are with the College of Science, China University of Petroleum (East China), Qingdao 266580, China (e-mail: zlshi@upc.edu.cn; liwg@upc.edu.cn; yantoji@163.com; wangjiannl@upc.edu.cn).

Xiao Peng Li is with the State Key Laboratory of Radio Frequency Heterogeneous Integration, Shenzhen University, Shenzhen 518060, China (e-mail: x.p.li.frank@gmail.com).

Yaru Fu is with the Department of Electronic Engineering and Computer Science, Hong Kong Metropolitan University, Hong Kong SAR 999077, China (e-mail: yfu@hkmu.edu.hk).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LSP.2023.3301244, provided by the authors.

Digital Object Identifier 10.1109/LSP.2023.3301244

problem via ℓ_0-norm optimization. To address the resultant nonconvex and discontinuous problem, we develop a block coordinate descent (BCD) algorithm [26] consisting of three steps. At each iteration, the binary weights (0-1) for the entries of the residual between the recovered and known matrices are first analytically determined based on ℓ_0-norm optimization. Then the optimization over the bilinear term becomes weighted version of the Frobenius norm, which can be divided into two sets of row-wise and column-wise subproblems. For each row-wise or column-wise subproblem, its closed-form solution is derived.

II. Problem Formulation

Given a data matrix $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$, we can approximate $\boldsymbol{Y}$ as the sum of two components: a low-rank part $\boldsymbol{L}$ and a perturbation part $\boldsymbol{O}$. Here, the rank of $\boldsymbol{L}$, i.e., $\text{rank}(\boldsymbol{L}) = r$, satisfies $r \ll \min\{m, n\}$. Given $\boldsymbol{Y}$ and r, the matrix recovery task aims to recover the low-rank component $\boldsymbol{L}$. Since $\boldsymbol{L}$ is low-rank, we can estimate $\boldsymbol{U} \in \mathbb{R}^{m\times r}$ and $\boldsymbol{V} \in \mathbb{R}^{r\times n}$ such that $\boldsymbol{L} = \boldsymbol{UV}$. When the elements of $\boldsymbol{O}$ follow zero-mean Gaussian distribution, the optimal recovery can be obtained by minimizing the following least squares problem:

$$\min_{\boldsymbol{U},\boldsymbol{V}} \|\boldsymbol{Y} - \boldsymbol{UV}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} \left(Y_{ij} - \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{v}_j\right)^2, \tag{1}$$

where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]^{\mathrm{T}}$, $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]$, and $\|\cdot\|_F$ is the Frobenius norm of a matrix. With the Frobenius norm, the resultant $\boldsymbol{U}$ and $\boldsymbol{V}$ are the maximum likelihood estimation, which can be obtained via singular value decomposition [17].

However, $\boldsymbol{O}$ might be contaminated by outliers in practical situations. Hence, one can exploit the following minimization problem for robust matrix recovery [21], [22]:

$$\min_{\boldsymbol{U},\boldsymbol{V}} \|\boldsymbol{Y} - \boldsymbol{UV}\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n} |Y_{ij} - \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{v}_j|, \tag{2}$$

where $\|\cdot\|_1$ denotes ℓ_1-norm, i.e., the sum of absolute values of matrix entries. Nevertheless, the ℓ_1-norm based model is limited by drawbacks such as memory requirement and computational efficiency [25]. To overcome these issues, one can adopt the following weighting-based model [6], [16]:

$$\min_{\boldsymbol{U},\boldsymbol{V},\boldsymbol{W}} \sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j}\left(Y_{ij} - \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{v}_j\right)^2, \tag{3}$$

where the outlier contaminated entries are with small $w_{i,j}$'s.

III. Algorithm Development

Inspired by model (3), we propose utilizing binary values to perform hard weighting in RLRMR, aiming at eliminating the influence of outliers. To do so, we formulate RLRMR as a MIP problem via ℓ_0-norm optimization, given by

$$\min_{\boldsymbol{U},\boldsymbol{V},\overline{\boldsymbol{W}},\boldsymbol{W}} \quad \|\boldsymbol{W} \odot (\boldsymbol{Y} - \boldsymbol{UV})\|_F^2 \tag{4a}$$

$$\text{s.t.} \quad \overline{\boldsymbol{W}} + \boldsymbol{W} = \boldsymbol{1} \tag{4b}$$

$$\|\overline{\boldsymbol{W}}\|_0 \le pmn \tag{4c}$$

$$\boldsymbol{W} \text{ and } \overline{\boldsymbol{W}} \in \{0,1\}^{m\times n}, \tag{4d}$$

where $\odot$ denotes the element-wise multiplication operator, $\boldsymbol{1} \in \mathbb{R}^{m\times n}$ is the matrix of all ones, and p is the percentage of outliers. The binary values 0 and 1 in $\boldsymbol{W}$ indicate elements contaminated by outliers and clean elements respectively, while they mean the contrary in $\overline{\boldsymbol{W}}$. Besides, $\|\overline{\boldsymbol{W}}\|_0 \le pmn$ is used for imposing sparsity on outliers. Hereafter, we let $s = pmn$ for brevity. We name the proposed model (4) as **RMR-MIP**.

Solving problem (4) is challenging since it is nonlinear and nonconvex with combinatorial constraint. To make (4) more trackable, one solution is to relax the value range of $\boldsymbol{W}$ and $\overline{\boldsymbol{W}}$ from binary $\{0,1\}^{m\times n}$ into real-valued $[0,1]^{m\times n}$. Then one can add ℓ_1-norm regularization over $\overline{\boldsymbol{W}}$ into the objective of (4) to obtain sparseness [6]. However, such a relaxation and regularization-based formulation is not able to eliminate influence of outliers. The reason is that, ℓ_1-norm regularization will reduce the magnitudes of all entries in $\overline{\boldsymbol{W}}$ [27], [28], resulting in $\overline{\boldsymbol{W}} \in [0,1)^{m\times n}$. Then, according to (4b), it holds that $\boldsymbol{W} \in (0,1]^{m\times n}$, meaning that the outlier-contaminated elements are assigned with nonzero weights. Besides, the sparseness of outliers is controlled by regularization parameter which is crucial for recovery results, but difficult to choose appropriately in practice.

Concerning the above, we directly address (4) without relaxing the binary matrices. Define $\mathcal{J}(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W}) = \|\boldsymbol{W} \odot (\boldsymbol{Y} - \boldsymbol{UV})\|_F^2$, a feasible solution of $\mathcal{J}(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W})$ can be calculated by BCD [26] algorithm:

$$\boldsymbol{W}^k = \arg\min_{\boldsymbol{W}} \mathcal{J}(\boldsymbol{U}^{k-1}, \boldsymbol{V}^{k-1}, \boldsymbol{W}), \tag{5}$$

$$\boldsymbol{U}^k = \arg\min_{\boldsymbol{U}} \mathcal{J}(\boldsymbol{U}, \boldsymbol{V}^{k-1}, \boldsymbol{W}^k), \tag{6}$$

$$\boldsymbol{V}^k = \arg\min_{\boldsymbol{V}} \mathcal{J}(\boldsymbol{U}^k, \boldsymbol{V}, \boldsymbol{W}^k). \tag{7}$$

With $\boldsymbol{W}^k$, $\overline{\boldsymbol{W}}^k$ is given by $\overline{\boldsymbol{W}}^k = \boldsymbol{1} - \boldsymbol{W}^k$. The BCD iteratively finds the optimal solution to one of involved variables with other ones fixed. It is clear that the accurate estimation of $\boldsymbol{W}^k$ is the core for our algorithm. Correct estimation will completely avoid the influence of outliers, while incorrect estimation will result in poor recovery or even ruin the recovery. In the next subsection, we will show that our algorithm analytically updates $\boldsymbol{W}^k$ based on an indicator function.

A. Development of the BCD Optimizer for RMR-MIP

In this subsection, we elaborate on the BCD optimizer for RMR-MIP, including $\boldsymbol{W}$-update, $\boldsymbol{V}$-update, and $\boldsymbol{U}$-update.

1) Development of $\boldsymbol{W}$-update: According to (4b)–(4d), it holds that $\|\boldsymbol{W}\|_0 \ge (mn - s)$. With $\boldsymbol{U}^{k-1}$ and $\boldsymbol{V}^{k-1}$ fixed as the estimation of $(k-1)$-th iteration and let $\boldsymbol{\Delta} = \boldsymbol{Y} - \boldsymbol{U}^{k-1}\boldsymbol{V}^{k-1}$, the subproblem for updating $\boldsymbol{W}$ is simplified as:

$$\min_{\boldsymbol{W}} \|\boldsymbol{W} \odot \boldsymbol{\Delta}\|_F^2 \ \text{ s.t. } \|\boldsymbol{W}\|_0 \ge mn - s. \tag{8}$$

Let Ω be the index set to indicate the nonzero elements of $\boldsymbol{W}$ and Ω^C be the index set to indicate the zero elements of $\boldsymbol{W}$. The objective value $f(\boldsymbol{W})$ of (8) becomes:

$$f(\boldsymbol{W}) = \sum_{(i,j)\in\Omega} W_{i,j}\Delta_{i,j}^2 + \sum_{(i',j')\in\Omega^C} W_{i',j'}\Delta_{i',j'}^2. \tag{9}$$

Since $\boldsymbol{W}$ is a binary matrix, $W_{i',j'} = 0$ for $(i', j') \in \Omega^C$. Hence, we have

$$f(\boldsymbol{W}) = \sum_{(i,j)\in\Omega} \Delta_{i,j}^2. \tag{10}$$

It is easy to observe that the objective $f(\boldsymbol{W})$, i.e., $\|\boldsymbol{W} \odot \boldsymbol{\Delta}\|_2^2$, increases as the number of nonzero elements of $\boldsymbol{W}$ increases. That is, the increase of $\|\boldsymbol{W}\|_0$ will lead to the growth of the objective function. Hence, to minimize $f(\boldsymbol{W})$, $\|\boldsymbol{W}\|_0$ should be $(mn - s)$, and the index set Ω should contain the indices of the $(mn - s)$ smallest (in absolute value) components of $\boldsymbol{\Delta}$. Correspondingly, the index set Ω^C should contain the indices of the s largest (in absolute value) components of $\boldsymbol{\Delta}$.

In other words,

$$\boldsymbol{W}^k = \mathcal{I}_s(\boldsymbol{\Delta}), \tag{11}$$

where $\mathcal{I}_s$ is an elementwise indicator function:

$$\mathcal{I}_s(\Delta_{i,j}) = \begin{cases} 0, & \text{if } |\Delta_{i,j}| > \delta, \\ 1, & \text{if } |\Delta_{i,j}| \le \delta. \end{cases} \tag{12}$$

In (12), δ is the $(mn - s)$-th smallest element (in absolute value) of $\boldsymbol{\Delta}$.

2) Development of $\boldsymbol{U}$-update: To solve (6), we fix $\boldsymbol{W}^k$ and $\boldsymbol{V}^{k-1}$ in (4). Thus, the $\boldsymbol{U}$-update is reduced to

$$\boldsymbol{U}^k = \arg\min_{\boldsymbol{U}} \|\boldsymbol{W}^k \odot (\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{V}^{k-1})\|_F^2, \tag{13}$$

which can be decomposed into a set of row-wise problems w.r.t. $\boldsymbol{u}_i^{\mathrm{T}}$. We resolve the $\boldsymbol{U}$-update one by one as follows:

$$(\boldsymbol{u}_i^{\mathrm{T}})^k = \arg\min_{\boldsymbol{u}_i} \|(\boldsymbol{b}_i - \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{V}^{k-1}) \odot \boldsymbol{\Lambda}_i^k\|_2^2, \tag{14}$$

where $\boldsymbol{b}_i$ is the i-th row of $\boldsymbol{Y}$, and $\boldsymbol{\Lambda}_i^k$ is a diagonal matrix composed by the i-th row of $\boldsymbol{W}^k$. Taking the derivative w.r.t $\boldsymbol{u}_i^{\mathrm{T}}$ and setting it to $\boldsymbol{0}$, we get the closed form solution:

$$(\boldsymbol{u}_i^{\mathrm{T}})^k = (\boldsymbol{b}_i\boldsymbol{\Lambda}_i^k)(\boldsymbol{V}^{k-1}\boldsymbol{\Lambda}_i^k)^\dagger, \tag{15}$$

where $(\cdot)^\dagger$ is the Moore-Penrose inverse of a matrix.

3) Development of $\boldsymbol{V}$-update: Similarly, fixing $\boldsymbol{U}^k$ and $\boldsymbol{W}^k$ in (4), we update $\boldsymbol{V}$ at k-th iteration by solving

$$\boldsymbol{V}^k = \arg\min_{\boldsymbol{V}} \|\boldsymbol{W}^k \odot (\boldsymbol{Y} - \boldsymbol{U}^k\boldsymbol{V})\|_F^2. \tag{16}$$

Again, the $\boldsymbol{V}$-update can be decomposed into a group of column-wise problems w.r.t $\boldsymbol{v}_j$. For each column $\boldsymbol{v}_j$, we have

$$\boldsymbol{v}_j^k = \arg\min_{\boldsymbol{v}_j} \|\boldsymbol{\Upsilon}_j^k \odot (\boldsymbol{d}_j - \boldsymbol{U}^k\boldsymbol{v}_j)\|_2^2, \tag{17}$$

where $\boldsymbol{d}_j$ is the j-th column of $\boldsymbol{Y}$, and $\boldsymbol{\Upsilon}_j^k$ is a diagonal matrix composed by the j-th column of $\boldsymbol{W}^k$. Hence, the analytic solution is obtained by:

$$\boldsymbol{v}_j^k = (\boldsymbol{\Upsilon}_j^k\boldsymbol{U}^k)^\dagger(\boldsymbol{\Upsilon}_j^k\boldsymbol{d}_j). \tag{18}$$

Since our formulation in (4) is highly nonconvex and discontinuous, powerfactorization [29] is adopted to initialize $\boldsymbol{U}$ and $\boldsymbol{V}$. In detail, we optimize the unweighted matrix recovery formulation in (1) for several steps, and employ the resultant $\boldsymbol{U}$ and $\boldsymbol{V}$ as our $\boldsymbol{U}^0$ and $\boldsymbol{V}^0$. Moreover, we propose initializing $\boldsymbol{W}$ by $\boldsymbol{W}^1 = \mathcal{I}_s(\boldsymbol{Y})$, rather than $\boldsymbol{W}^1 = \mathcal{I}_s(\boldsymbol{Y} - \boldsymbol{U}^0\boldsymbol{V}^0)$. The reason is that, in the presence of large portion of outliers, $\boldsymbol{U}^0$ and $\boldsymbol{V}^0$ become too poor to support updating $\boldsymbol{W}^1$. On the other hand, updating $\boldsymbol{W}^1$ directly from noisy observations is intuitive in the sense that data with large magnitudes is more likely to be outliers [14]. For instance, for image data $\boldsymbol{Y}$, if an area is overexposed, the pixel values are close to 255. It should also be noticed that pixel values of underexposed area are close to 0 for image data. Then according to (11), these outliers cannot be detected directly by our algorithm. To make RMR-MIP adoptable for image recovery, we can recover $\boldsymbol{Y} - 127.5$ instead of $\boldsymbol{Y}$ as in deep learning [30]. Algorithm 1 has summarized the proposed BCD-based optimizer for RMR-MIP.

Algorithm 1: RMR-MIP (BCD).

Input: $\boldsymbol{Y}$, p, r, and k_{max}
  Initialize: $s = pmn$, $\boldsymbol{U}^0$, $\boldsymbol{V}^0$, and $\boldsymbol{W}^1 = \mathcal{I}_s(\boldsymbol{Y})$.
  while *not converged* **do**
    for $i = 1, 2, \cdots, m$ **do**
      Calculate $(\boldsymbol{u}_i^{\mathrm{T}})^k$ according to (15).
    end
    for $j = 1, 2, \cdots, n$ **do**
      Calculate $\boldsymbol{v}_j^k$ according to (18).
    end
    Update $\boldsymbol{W}^{k+1}$ according to (11).
  end
Output: Optimal $\boldsymbol{U}^*$, $\boldsymbol{V}^*$

The convergence of the proposed BCD for RMR-MIP is studied in the following proposition. And we present the convergence results in supplementary material.

Proposition 1: Let $\mathcal{J}^k$ be the function value $\mathcal{J}(\boldsymbol{U}^k, \boldsymbol{V}^k, \boldsymbol{W}^k)$ at k-th iteration. Then the sequence $\{\mathcal{J}^k\}_{k=1,2,\cdots}$ generated by the proposed BCD algorithm converges.

Proof: We calculate

$$\begin{aligned} \mathcal{J}^k - \mathcal{J}^{k-1} &= \mathcal{J}(\boldsymbol{U}^k, \boldsymbol{V}^k, \boldsymbol{W}^k) - \mathcal{J}(\boldsymbol{U}^k, \boldsymbol{V}^{k-1}, \boldsymbol{W}^k) \\ &\quad + \mathcal{J}(\boldsymbol{U}^k, \boldsymbol{V}^{k-1}, \boldsymbol{W}^k) - \mathcal{J}(\boldsymbol{U}^{k-1}, \boldsymbol{V}^{k-1}, \boldsymbol{W}^k) \\ &\quad + \mathcal{J}(\boldsymbol{U}^{k-1}, \boldsymbol{V}^{k-1}, \boldsymbol{W}^k) - \mathcal{J}(\boldsymbol{U}^{k-1}, \boldsymbol{V}^{k-1}, \boldsymbol{W}^{k-1}). \end{aligned} \tag{19}$$

According to (5) to (7), all these terms at the right side of above equation are non-positive. That is, $\mathcal{J}^k - \mathcal{J}^{k-1} \le 0$. Hence, the sequence $\{\mathcal{J}^k\}_{k=1,2,\cdots}$ is non-increasing. In addition, since $\mathcal{J}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}) \ge 0$, $\mathcal{J}^k$ is bounded below. We can conclude that $\{\mathcal{J}^k\}_{k=1,2,\cdots}$ is convergent. ■

IV. Experiments

To evaluate the performance of RMR-MIP, we conduct experiments on synthetic data and several images. We compare the proposed algorithm with five state-of-the-art algorithms, including ROUTE [6], RegL_1 [25], RMF-MM [17], CWM [23], and RPCA-HQF [4]. To measure the recovery performance, mean squared error (MSE) is utilized for synthetic data, while peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are used for images.

A. Synthetic Data

In this subsection, we fix $m = 300$, $n = 200$, and $r = 5$. The ground truth matrix $\boldsymbol{L}$ is generated by $\boldsymbol{L} = \boldsymbol{U}_0\boldsymbol{V}_0$, where

TABLE I
AVERAGE MSE RESULTS BY DIFFERENT ALGORITHMS UNDER GAUSSIAN MIXTURE NOISE AT VARIOUS SNR LEVELS

Method	SNR=-9dB	SNR=-6dB	SNR=-3dB	SNR=0dB	SNR=3dB	SNR=6dB	SNR=9dB
RegL1 [25]	1.544×10^{-3}	7.582×10^{-4}	3.770×10^{-4}	1.883×10^{-4}	9.403×10^{-5}	4.703×10^{-5}	2.357×10^{-5}
ROUTE [6]	4.035×10^{-3}	1.310×10^{-3}	3.757×10^{-4}	1.307×10^{-4}	6.226×10^{-5}	3.674×10^{-5}	2.578×10^{-5}
CWM [23]	1.513×10^{-3}	7.408×10^{-4}	3.691×10^{-4}	1.843×10^{-4}	9.204×10^{-5}	4.612×10^{-5}	2.313×10^{-5}
RPCA-HQF [4]	1.108×10^{-3}	5.002×10^{-4}	2.484×10^{-4}	1.239×10^{-4}	6.200×10^{-5}	3.133×10^{-5}	2.220×10^{-5}
RMR-MM [17]	1.552×10^{-3}	7.620×10^{-4}	3.789×10^{-4}	1.892×10^{-4}	9.455×10^{-5}	4.729×10^{-5}	2.371×10^{-5}
RMR-MIP	$\mathbf{9.988 \times 10^{-4}}$	$\mathbf{4.937 \times 10^{-4}}$	$\mathbf{2.457 \times 10^{-4}}$	$\mathbf{1.228 \times 10^{-4}}$	$\mathbf{6.154 \times 10^{-5}}$	$\mathbf{3.077 \times 10^{-5}}$	$\mathbf{1.543 \times 10^{-5}}$

TABLE II
AVERAGE MSE RESULTS BY DIFFERENT ALGORITHMS UNDER UNIFORM NOISE AT VARIOUS OUTLIER PERCENTAGES

Method	p_0=10%	p_0=20%	p_0=30%	p_0=40%	p_0=50%	p_0=60%	p_0=70%
RegL1 [25]	2.323×10^{-1}	2.411×10^{-1}	3.673×10^{-1}	5.931×10^{-1}	9.137×10^{-1}	1.723	3.486
ROUTE [6]	1.404×10^{-3}	1.557×10^{-3}	1.752×10^{-3}	2.017×10^{-3}	2.363×10^{-3}	2.891×10^{-3}	$\mathbf{5.297 \times 10^{-3}}$
CWM [23]	8.087×10^{-4}	1.067×10^{-3}	1.478×10^{-3}	2.258×10^{-3}	2.000×10^{-1}	1.642	2.829
RPCA-HQF [4]	7.223×10^{-2}	2.832×10^{-1}	2.593×10^{-1}	5.756×10^{-1}	3.234	8.904	7.345
RMR-MM [17]	3.588×10^{-3}	1.104×10^{-3}	5.878×10^{-2}	9.017×10^{-1}	4.781	3.257	4.638
RMR-MIP	$\mathbf{6.682 \times 10^{-4}}$	$\mathbf{7.903 \times 10^{-4}}$	$\mathbf{9.647 \times 10^{-4}}$	$\mathbf{1.232 \times 10^{-3}}$	$\mathbf{1.6784 \times 10^{-3}}$	$\mathbf{2.778 \times 10^{-3}}$	9.153×10^{-3}

$\boldsymbol{U}_0 \in \mathbb{R}^{m\times r}$ and $\boldsymbol{V}_0 \in \mathbb{R}^{r\times m}$. For the perturbation matrix $\boldsymbol{O}$, we consider two kinds of outliers. The first one is the impulsive noise modeled by Gaussian mixture model (GMM) [4], [10]. To thoroughly show the robustness against outliers, we vary the signal-to-noise (SNR) values in a wide range from -9 dB to 9 dB. At each SNR value, we run the experiments 100 trials. Under GMM, we find empirically that $p = 8\%$ is good enough for our RMR-MIP (Please see supplementary material).

The second type is the uniform noise [1], [6]. Following [6], we replace a fraction p_0 of matrix $\boldsymbol{L}$ with outliers drawn from a uniform distribution over $[-20, 20]$ at random, and the rest entries are contaminated by Gaussian noise $\mathcal{N}(0, 0.01)$. The fraction p_0 ranges from 10% to 70%. We repeat the experiments 100 times at each p_0 and set $p = p_0 + 5\%$ for our RMR-MIP.

The averaged MSE results under GMM and uniform noise are reported in Tables I and II, respectively. It is seen that the proposed method is superior to its competitors. One exception is that, the RMP-MIP provides the second best recovery result at $p_0 = 70\%$ under uniform noise. The reason is that, in this case, our RMR-MIP only utilizes 25% entries for recovery, which is too little to include enough information.

B. Image Data

In this subsection, we also utilize two kinds of outliers to corrupt the image data $\boldsymbol{L}$, including salt-and-pepper noise and text mask. Besides, images are also polluted by Gaussian noise $\mathcal{N}(0, 0.004)$. For all our experiments, we let $r = 8$.

In the first test, we add salt-and-pepper noise into $\boldsymbol{L}$ by the built-in function in MATLAB. In detail, we obtain the corrupted image by "$\boldsymbol{Y}$ = imnoise ($\boldsymbol{L}$,'salt & pepper',γ)", where γ is the normalized noise intensity given by $\gamma = 1/\text{SNR}$. We set SNR = 3 dB and $p = 35\%$ for our RMR-MIP. The simulation results under salt-and-pepper noise are depicted in Fig. 1. It is observed that all algorithms except RPCA-HQF provide satisfactory recovery. RMR-MIP has the highest PSNR of 27.28 dB and SSIM of 0.78404.

In the second test, the image is contaminated by adding several letters into it. For our RMR-MIP, $p = 15\%$. Fig. 2 shows the recovery results and the estimated $\boldsymbol{W}$ by the proposed RMR-MIP. According to the estimated $\boldsymbol{W}$, the proposed RMR-MIP can classify outliers by assigning zero weights to them. As a result, the proposed RMR-MIP gives out the highest PSNR of 26.92 dB and SSIM of 0.78373. On the contrary, all its competitors fail to eliminate the impact of outliers, since black scratches still exist in their recovered images. Please see more results for image data in the supplementary material.

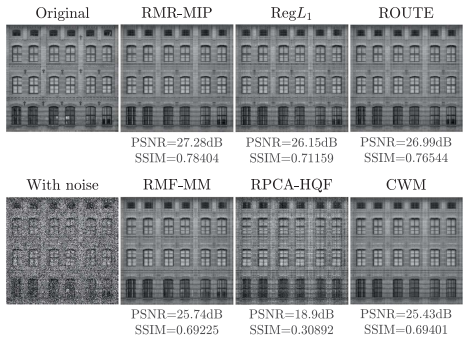

Fig. 1. Noisy image and recovered images in salt-and-pepper noise.

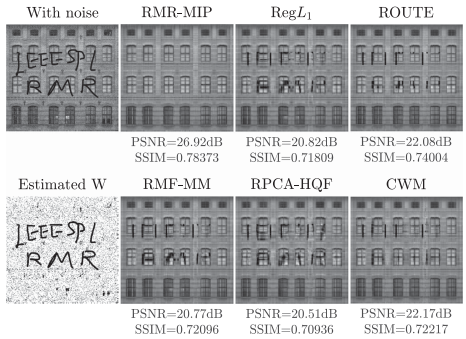

Fig. 2. Noisy image with a text mask, recovered images, and the estimated $\boldsymbol{W}$.

V. CONCLUSION

Many RLRMR algorithms aim at reducing or suppressing the influence of outliers by utilizing ℓ_1-norm or the weighting-based idea. However, the recovery results are still degraded by anomalies. To eliminate the impact of outliers, this letter formulates the RLRMR as a MIP problem via ℓ_0-norm optimization. In our formulation, the outlier contaminated entries are with zero weights, and thus the error caused by outliers will be zeroed out. Simulation results on both synthetic and image data have demonstrated the superiority of our algorithm over five state-of-the-art algorithms.

REFERENCES

[1] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 1–37, May 2011.

[2] H. Zhang et al., "Generalized nonconvex nonsmooth low-rank matrix recovery framework with feasible algorithm designs and convergence analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 23, 2022, doi: 10.1109/TNNLS.2022.3183970.

[3] P. Magron and C. Févotte, "A majorization-minimization algorithm for nonnegative binary matrix factorization," *IEEE Signal Process. Lett.*, vol. 29, pp. 1526–1530, 2022.

[4] Z.-Y. Wang, X. P. Li, H. C. So, and Z. Liu, "Robust PCA via non-convex half-quadratic regularization," *Signal Process.*, vol. 204, Mar. 2023, Art. no. 108816.

[5] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.

[6] X. Guo and Z. Lin, "Low-rank matrix recovery via robust outlier estimation," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5316–5327, Nov. 2018.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.

[8] A. Acharya, R. Goel, A. Metallinou, and I. Dhillon, "Online embedding compression for text classification using low rank matrix factorization," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 6196–6203.

[9] H. Zhang, J. Yang, F. Shang, C. Gong, and Z. Zhang, "LRR for subspace segmentation via tractable schatten-p norm minimization and factorization," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1722–1734, May 2019.

[10] X. P. Li, Q. Liu, and H. C. So, "Rank-one matrix approximation with ℓ_p-norm for image inpainting," *IEEE Signal Process. Lett.*, vol. 27, pp. 680–684, 2020.

[11] C.-F. Chen, C.-P. Wei, and Y.-C. F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2618–2625.

[12] F. Nie, Z. Hu, and X. Li, "Matrix completion based on non-convex low-rank approximation," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2378–2388, May 2019.

[13] V. Garg, A. Pages-Zamora, and I. Santamaria, "Order estimation via matrix completion for multi-switch antenna selection," *IEEE Signal Process. Lett.*, vol. 28, pp. 2063–2067, 2021.

[14] X. P. Li, Z.-L. Shi, Q. Liu, and H. C. So, "Fast robust matrix completion via entry-wise ℓ_0-norm minimization," *IEEE Trans. Cybern.*, early access, Dec. 06, 2022, doi: 10.1109/TCYB.2022.3224070.

[15] A. Parekh and I. W. Selesnick, "Enhanced low-rank matrix approximation," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 493–497, Apr. 2016.

[16] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 720–727.

[17] Z. Lin, C. Xu, and H. Zha, "Robust matrix factorization by majorization minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 208–220, Jan. 2018.

[18] J. Wright, A. Ganesh, S. R. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 289–298.

[19] Z. Shi, H. Wang, C. S. Leung, and H. C. So, "Robust MIMO radar target localization based on Lagrange programming neural network," *Signal Process.*, vol. 174, Apr. 2020, Art. no. 107574.

[20] W. Xiong and H. C. So, "TOA-based localization with NLOS mitigation via robust multidimensional similarity analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1334–1338, Sep. 2019.

[21] G. Lerman and T. Maunu, "An overview of robust subspace recovery," *Proc. IEEE*, vol. 106, no. 8, pp. 1380–1410, Aug. 2018.

[22] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.

[23] D. Meng, Z. Xu, L. Zhang, and J. Zhao, "A cyclic weighted median method for L_1 low-rank matrix factorization with missing entries," in *Proc. AAAI Conf. Artif. Intell.*, 2013, pp. 704–710.

[24] X. P. Li, Z.-Y. Wang, Z.-L Shi, H. C. So, and N. D. Sidiropoulos, "Robust tensor completion via capped Frobenius norm," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 23, 2023, doi: 10.1109/TNNLS.2023.3236415.

[25] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, "Practical low-rank matrix approximation under robust L_1-norm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1410–1417.

[26] A. Beck and L. Tetruashvili, "On the convergence of block coordinate descent type methods," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2037–2060, Jan. 2013.

[27] T. Zhang, "Adaptive forward-backward greedy algorithm for learning sparse representations," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4689–4708, Jul. 2011.

[28] V. Lebedev and V. Lempitsky, "Fast convnets using group-wise brain damage," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2554–2564.

[29] J. P. Haldar and D. Hernando, "Rank-constrained solutions to linear matrix equations using powerfactorization," *IEEE Signal Process. Lett.*, vol. 16, no. 7, pp. 584–587, Jul. 2009.

[30] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 212–220.