

Recovery of Corrupted Low-Rank Matrices via Half-Quadratic based Nonconvex Minimization

Ran He¹, Zhenan Sun¹, Tieniu Tan¹ and Wei-Shi Zheng²

¹National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing 100190, China

{rhe, znsun, tnt}@nlpr.ia.ac.cn

² School of Information Science and Technology
Sun Yat-sen University, Guangzhou 510275, China

wszheng@ieee.org

Abstract

Recovering arbitrarily corrupted low-rank matrices arises in computer vision applications, including bioinformatic data analysis and visual tracking. The methods used involve minimizing a combination of nuclear norm and l^1 norm. We show that by replacing the l^1 norm on error items with nonconvex M-estimators, exact recovery of densely corrupted low-rank matrices is possible. The robustness of the proposed method is guaranteed by the M-estimator theory. The multiplicative form of half-quadratic optimization is used to simplify the nonconvex optimization problem so that it can be efficiently solved by iterative regularization scheme. Simulation results corroborate our claims and demonstrate the efficiency of our proposed method under tough conditions.

1. Introduction

Principal Component Analysis (PCA) [31] is a linear data transformation technique which plays an important role in the studies of image processing and machine learning. It assumes that the high-dimensional data reside in a low-dimensional linear subspace [9], and has been widely used for the representation of high dimensional data such as image data for appearance, shape, visual tracking, etc. Moreover, it is commonly used as a preprocessing step to project high-dimensional data into a low-dimensional subspace.

Consider a data set of samples $D = [d_1, \dots, d_n]$ where d_i is a variable in Euclidean space with dimensionality m , $U = [u_1, \dots, u_r] \in R^{m \times r}$ be a projection matrix whose columns constitute the bases of an r -dimensional subspace, and $V = [v_1, \dots, v_n] \in R^{r \times n}$ be the principal components that are projection coordinates under the projection matrix

U . From the viewpoint of Mean Square Error (MSE), PCA assumes that the data matrix D is generated by perturbing a matrix $A = UV \in R^{m \times n}$ whose columns reside in a subspace of dimension $r \ll \min(m, n)$, i.e., $D = A + E$, where A is a rank- r matrix and E is a matrix whose entries are i.i.d. Gaussian random variables [22]. In this setting, PCA can be formulated as the following constrained optimization problem [26]:

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

where $\|\cdot\|_F$ is the Frobenius norm.

However, PCA breaks down under outliers¹ because large errors will dominate the mean square error (MSE) [8, 18]. To solve this limitation, one strategy is to replace the MSE in PCA with a robust measure, i.e., Huber's M-estimator. Those M-estimators based PCA methods [8, 18, 27, 16, 11] iteratively reweight features (or samples) and then utilize the uncorrupted features (or samples) to compute a robust subspace.

Recently, Wright et al [37] shows that the robust PCA problem can be exactly solved via convex optimization that minimizes a combination of the nuclear norm and l^1 -norm. By assuming that the error matrix E has a sparse representation, robust PCA can be formulated as the following problem [37]:

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

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix (i.e., the sum of its singular values), $\|\cdot\|_0$ is the counting norm (i.e., the number of non-zero entries in the matrix), and λ is a positive constant. Since the problem (2) is NP-hard and cannot

¹In robust statistics, outliers are those data points that deviate significantly from the rest of the data [8].

be efficiently solved, one considers its convex relaxation instead [26],

$$\min_{A,E} \|A\|_* + \lambda \|E\|_1 \quad s.t. \quad D = A + E \quad (3)$$

where $\|\cdot\|_1$ represents the matrix 1-norm (i.e., the sum of absolute values of all entries of a matrix). Many methods [37, 9, 21, 22, 10, 23] have been developed to optimize (3).

Recent advance in l^0 - l^1 theory [4] shows that when the l^1 -norm in a underdetermined system is substituted with a concave semi-monotone increasing function, a sparse (with nonzeros) feasible solution is a unique global optimum if the solution is indeed sparse. In [5, 6, 7], the non-convex quasi l_p -norm (i.e., $\|\cdot\|_p^p$) approaches for $0 \leq p < 1$ are developed for recovery of sparse signal. Iteratively reweighted least squares (IRLS) approaches are used to solve the l_p -norm.

Considering that nonconvex M-estimators may deal with large and non-Gaussian noise better in real world problems [25][11][13], this paper addresses robust PCA problem from the view point of M-estimators, i.e.,

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

where $\Phi(E) = \sum_{i=1}^n \sum_{j=1}^m \phi(E_{ij})$ and $\phi(\cdot)$ is a robust non-convex M-estimator. Although there are many methods to solve (3), they often require that the functions in the objective function are convex. Hence those methods can not be directly used to solve (4). Facing this difficulty, we harness the multiplicative form of the half-quadratic (HQ) optimization² to solve the nonconvex problem in $\Phi(E)$ such that the problem can be iteratively solved by solving a number of unconstrained quadratic problems.

The robustness of the proposed method is guaranteed by the M-estimator theory [15] and maximum correntropy criterion (MCC) [25], which have shown that M-estimators have potential to control large outliers and non-Gaussian noise [25, 38, 13]. We show that by replacing the l^1 norm on corrupted items with nonconvex measures, exact recovery of densely corrupted low-rank matrices is possible. If MCC adaptation's distribution has the maximum at the origin [25], our method will have potential to recover arbitrarily corrupted low-rank matrix. Simulations on randomly generated matrices are run to corroborate our claims and demonstrate the efficiency of our proposed method under tough conditions.

The remainder of this paper is organized as follows. In Section 2, we briefly review maximum correntropy criterion for robust learning. In Section 3, we propose a half-quadratic algorithm to solve (4) and discuss its relationship

²Note that not all M-estimators have the additive form of the HQ, e.g., L1 M-estimator has no additive form. Since the additive form of correntropy is still unknown, we only discuss our method by using the multiplicative form.

with previous work. In Section 4, we compare the new algorithm with other exiting algorithms for low-rank matrix recovery. Finally, we conclude the paper in Section 5.

2. Maximum Correntropy Criterion (MCC)

Recently, the concept of correntropy is proposed in ITL [25] to process non-Gaussian noise [25] and impulsive noise [33]. Correntropy has shown its superiority in term of robustness in signal processing [25], feature extraction [38], subspace learning [11] and face recognition [13][12]. It is directly related to the Renyi's quadratic entropy [34], and is a local similarity measure between two arbitrary random variables A and D , defined by [25]:

$$V_\sigma(A, D) = E[k_\sigma(A, D)] = E[k_\sigma(A - D)] \quad (5)$$

where $k_\sigma(\cdot)$ is a kernel function that satisfies Mercer theory [36] and $E[\cdot]$ is the expectation operator. It takes advantage of kernel trick that nonlinearly maps the input space to a higher dimensional feature space. Different from conventional kernel methods, it works independently with pairwise samples. With a clear theoretic foundation, the correntropy is symmetric, positive, and bounded.

Based on (5), Liu et al. [25][24] further extended the sample based correntropy criterion for a general similarity measurement between any two discrete vectors. That is they introduced the correntropy induced metric (CIM) [25][24] for any two vectors $A = (a_1, \dots, a_m)^T$ and $D = (d_1, \dots, d_m)^T$ as follows:

$$CIM(A, D) = (g(0) - \frac{1}{m} \sum_{j=1}^m g(e_j))^{\frac{1}{2}} \quad (6)$$

where $g(\cdot)$ is the Gaussian function, and error e_j is defined as $e_j = a_j - d_j$. For adaptive systems, the below correntropy of error e_j :

$$\max_{\theta} \frac{1}{m} \sum_{j=1}^m g(e_j) = \frac{1}{m} \sum_{j=1}^m g(a_j - d_j) \quad (7)$$

is called the maximum correntropy criterion (MCC) [25], where θ is the parameter in the criterion.

MCC has a probabilistic meaning of maximizing the error probability density at the origin [24], and MCC adaptation is applicable in any noise environment when its distribution has the maximum at the origin [25]. Compared with the mean square error (MSE), a global metric, the correntropy is local. That means the correntropy value is mainly decided by the kernel function along the line $A = B$ [25].

Correntropy has a close relationship with m-estimators [15]. If we define $\rho(x) \triangleq 1 - \exp(-x)$, (6) is a robust formulation of Welsch m-estimator [25]. Furthermore, $\rho(x)$ satisfies $\lim_{|x| \rightarrow \infty} \rho'(x) = 0$, and thus it also belongs to the

so called redescending m-estimators [15], which has some special robustness properties[28]. A main merit of correntropy is that the kernel size controls all its properties. Due to the close relationship between m-estimation and methods of ITL, choose an appropriate kernel size [25] in the correntropy criterion becomes practical.

3. Nonconvex Minimization

In this section, we firstly propose a half-quadratic optimization based iterative regularization method to solve (4). Then we investigate possible robust M-estimators, and discuss the relationship between proposed method and previous work.

3.1. The Half-quadratic Approach

Substituting the equation constraint in (4) into its objective function, we have the following unconstrained problem³:

$$\min_{A,E} \|A\|_* + \lambda \Phi(D - A) \quad (8)$$

Generally, the robust M-estimator $\Phi(\cdot)$ is often nonconvex and hence is difficult to be directly optimized. Fortunately, the half-quadratic technique [29][13] has been developed to optimize those nonconvex cost-functions by alternately minimizing their augmented (resultant) cost-functions. Experiments show that half-quadratic methods are substantially faster (in terms of computational times) than gradient based methods [29].

According to conjugate function theory and HQ [29, 38, 13], we have that for a fixed e_{ij} , the following equation holds,

$$\phi(e_{ij}) = \min_{p_{ij}} Q(e_{ij}, p_{ij}) + \varphi(p_{ij}) \quad (9)$$

where $\varphi(\cdot)$ is the conjugate function of $\phi(\cdot)$, p_{ij} is the auxiliary variable of HQ, and $Q(e_{ij}, p_{ij})$ is the multiplicative form of HQ. The variable p_{ij} is determined by the minimizer function $\delta(\cdot)$ in HQ and is only relative to a specific function $\phi(\cdot)$ (See Table 1 for two specific functions and their minimizer functions). $Q(e_{ij}, p_{ij})$ has the following form [29],

$$Q(e_{ij}, p_{ij}) = \frac{1}{2} p_{ij} e_{ij}^2 \quad (10)$$

Substituting (9) and (10) into (8), we have the augmented cost-function of (8), i.e.,

$$J(A, E, P) = \min_{A,E,P} \|A\|_* + \frac{\lambda}{2} \|P \otimes (D - A)\|_F^2 + \sum_{i=1}^m \sum_{j=1}^n \varphi(p_{ij}) \quad (11)$$

³If we substitute $\Phi(\cdot)$ with Welsch M-estimator, (8) is actually based on MCC.

where P is the auxiliary variable matrix, $P_{ij} = \sqrt{p_{ij}}$, $\varphi(\cdot)$ is the conjugate function of $\phi(\cdot)$, and $P \otimes (D - A) \doteq \sum_{i=1}^m \sum_{j=1}^n P_{ij} (D_{ij} - A_{ij})$.

The cost function in (11) can be optimized by the following alternate minimization scheme:

$$P_{ij}^k = \sqrt{\delta(D_{ij} - A_{ij})} \quad (12)$$

$$A^{k+1} = \min_A \|A\|_* + \lambda \|P^k \otimes (D - A)\|_F^2 \quad (13)$$

where k is iteration number, and $\delta(\cdot)$ is a minimizer function. The above alternate minimization scheme can be viewed as the l^1 iterative regularization scheme [37]. In the first step, we compute the auxiliary variable matrix P according to the conjugate function theory [3, 29]; then in the second step, we solve a unconstrained minimization problem. As in [37], the solution of subproblem in (13) can be iteratively computed by the singular value thresholding operator [37]:

$$S_\tau[t] = \begin{cases} 0 & |t| \leq \tau \\ t - \tau \text{sign}(t) & |t| > \tau \end{cases} \quad (14)$$

Many methods have been developed to iteratively compute the unconstrained sub-problem in (13). We follow the approach of Singular Value Thresholding (SVT) algorithm [37] and summarize the optimal procedure in Algorithm 1. Let $Y = P \otimes (D - A)$, we compute Y step by step to obtain a smooth and continuous updating instead of computing Y_k as $D - A_k$. The learning rate θ is used to control the updating of Y_k . The θ and λ are set as suggested in [37].

Algorithm 1: Robust PCA via Half-quadratic Singular Value Thresholding (HQ-SVT)

- Input:** Observation matrix D , τ and θ .
Output: $A \leftarrow A^{k+1}$, $E \leftarrow D - A$
1: $Y^0 \leftarrow \mathbf{0}$, $A^0 \leftarrow \mathbf{0}$, $k \leftarrow 1$, and $\lambda \leftarrow 1$.
2: **while** "not converged" **do**
3: $(U, S, V) \leftarrow \text{svd}(Y^{k-1})$.
4: $A^k \leftarrow US_\tau[S]V^T$.
5: Update P^k according to (12).
6: $Y_k = Y_{k-1} + \theta(D - A_k)$.
7: $Y_k = Y_k \otimes P^k$.
8: **end while**
-

Since the M-estimator $\phi(\cdot)$ is often non-convex, Algorithm 1 can only find a local minimal solution of (8). The stopping criterion for the sequence of iterations of Algorithm 1 is that the variation of A_k between two iterations is smaller than a predefined precision. During each iteration, Algorithm 1 alternately minimizes the objective until it converges [29].

Table 1. Minimizer function δ relevant to multiplicative form of HQ for potential M-estimator ϕ . The dashline in the figures in the first row is drawn by l^1 -norm. α in L1-L2 M-estimator is a constant.

| M-estimator | L1-L2 [29] | Welsch [38, 13] |
|---------------------------------|---------------------------|------------------------------|
| Potential function $\phi(t)$ | $\sqrt{\alpha + t^2} - 1$ | $1 - \exp(-\frac{t^2}{0.5})$ |
| Multiplicative form $\delta(t)$ | $1/\sqrt{\alpha + t^2}$ | $\exp(-\frac{t^2}{0.5})$ |

3.2. M-estimators

In statistics, one popular robust technique is the so-called M-estimators [15], which are obtained as the minima of sums of functions of the data and have more than 30 years of history. They have been widely used in machine learning and computer vision for robust learning. In regression models, IRLS approach is often used to solve M-estimators by iteratively solving a number of weighted least square problems. Another common used technique to solve M-estimators is the half-quadratic optimization [29]. By the multiplicative and the additive half-quadratic reformulation of M-estimator cost-function, the original problem is solved by the alternate minimization of augmented cost-function.

Table 1 shows two selected M-estimators and their corresponding minimizer functions of multiplicative HQ form. It has been known that l^1 norm also belongs to Huber's M-estimators. We can observe from Table 1 that L1-L2 M-estimator and Welsch M-estimator seem to give less punishment to large outliers than l^1 M-estimator. The curves of L1-L2 and Welsch M-estimator are always lower than that of l^1 M-estimator. It is interesting to observe that the curve of Welsch M-estimator is close to that of l^1 M-estimator between 0 and 1. That is, when the magnitude of outlier is small, l^1 M-estimator and Welsch M-estimator may obtain similar robustness.

In information theoretic learning (ITL), Welsch M-estimator induced cost function is also called Maximum Correntropy Criterion (MCC) [25]. The MCC further gives an information theoretic foundation of robustness of adaptive systems. It has a probabilistic meaning of maximizing the error probability density at the origin [24], and MCC adaptation is applicable in any noise environment when its distribution has the maximum at the origin [25]. When outliers are different from the uncorrupted data in a low-

rank matrix, MCC adaptation only focuses on those e_j that are close to zero, and will detect outliers. Hence, different from previous theory that M-estimators very low breakdown points, MCC has potential to make recovery of arbitrarily corrupted low-rank matrix possible [25][13][14].

3.3. Relation to previous work

Robust subspace learning [18] for linear models has been widely used for shape representation, tracking, etc. Different approaches have been explored in the literature to make the learning more robust. To alleviate the problem of occlusion, modular eigenspaces are proposed in [32]. In [30], eigenwindow method is presented to recognize partially occluded objects. These methods based on "eigenwindow" partially alleviate the problems of occlusion but cannot solve them entirely [20]. In [2], a conventional robust M-estimator is used for computing the coefficients of subspace by substituting the MSE by a robust one. In [18, 8, 17], M-estimators based robust principal component analysis methods are developed to learn robust components and coefficients.

The similarity of our HQ-SVT and previous robust subspace learning methods is that they are all based on robust M-estimators. However, HQ-SVT more focuses on accurately recovering a corrupted low-rank matrix. It combines the nuclear norm of a matrix into original M-estimators based objectives and uses an iterative regularization way to automatically determine an appropriate matrix rank.

Another variation of robust PCA is also called the exact recovery of arbitrarily corrupted low-rank matrix. Based on the l^0 - l^1 equivalence theory [4], many methods [37, 9, 21, 22, 26] have been developed to efficiently estimate the sparse error matrix E . Wright et al. [37] shows that if the error matrix E is sufficiently sparse (relative to the rank of A), one can exactly recover the corrupted matrix A .

The similarity of our HQ-SVT and those low-rank matrix recovery methods is that they all harness iterative regularization and shrink operation to estimate eigenvalues. However, different from those methods, HQ-SVT is based on the M-estimator theory and hence there is no special assumption on error matrix E . Simulation results show that HQ-SVT can exactly recover a corrupted low-rank matrix even if the noise is non-sparse.

4. Experiment

In this section, numerical simulations are run to evaluate the recovery ability of the proposed method and to compare it with the singular value thresholding (SVT) algorithm [37], the accelerated proximal gradient (APG) algorithm [35], and the augmented Lagrange multipliers (ALM) algorithm [21] under different levels of corruptions. All algorithms were implemented in MATLAB on an AMD Quad-

| rank(A_0) | HQ-SVT1 | HQ-SVT2 | SVT | APG | ALM |
|---------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 5 | 2.1×10^{-4} | 3.7×10^{-4} | 4.6×10^{-5} | 1.9×10^{-5} | 2.3×10^{-7} |
| 10 | 2.3×10^{-4} | 4.0×10^{-4} | 4.1×10^{-5} | 1.7×10^{-5} | 3.6×10^{-7} |
| 15 | 2.5×10^{-4} | 4.3×10^{-4} | 5.3×10^{-5} | 1.5×10^{-5} | 2.5×10^{-7} |
| 20 | 2.5×10^{-4} | 4.1×10^{-4} | 5.2×10^{-5} | 1.3×10^{-5} | 3.5×10^{-7} |

Table 2. Average reconstruction errors ($\frac{\|A-A_0\|_F}{\|A_0\|_F}$) of the four algorithms on low rank matrices ($m = 400$ and $\frac{\|E_0\|_0}{m^2} = 0.05$). For all algorithms, the rank is correctly estimated.

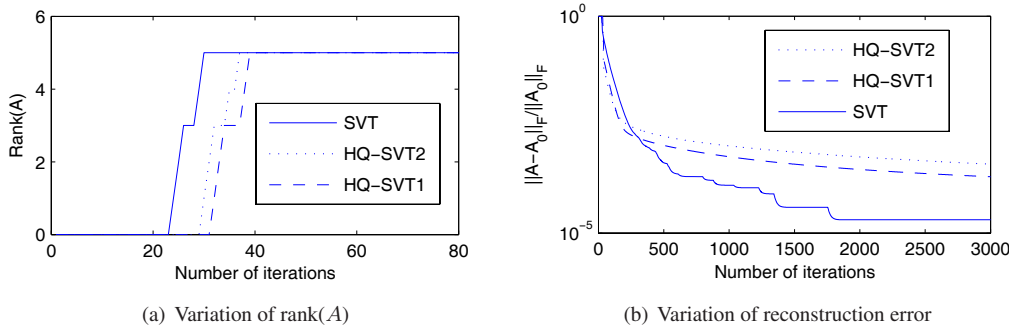


Figure 1. Convergence of singular value thresholding algorithms.

Core 1.80GHz Windows XP machine with 2GB memory. The SVT, APG and ALM matlab codes are from [26] and their parameters are default as suggested in [26].

4.1. Simulation Conditions

To fairly and quantitatively evaluate different methods, we randomly generated matrices as suggested in [37][35][21]. Without loss of generality, we suppose to simplify that the unknown matrix $A \in R^{m \times m}$ is square [37]. The ordered pair $(A_0, E_0) \in R^{m \times m} \times R^{m \times m}$ denotes the true solution. And the observation matrix $D = A_0 + E_0$ is the input to all algorithms, and the ordered pair (A, E) denotes the output. The matrix A_0 is generated as a product UV^T according to the random orthogonal model of rank r [37]. The matrix U and V are independent $m \times r$ matrices whose elements are i.i.d. Gaussian random variables with zero mean and unit variance. The error matrix E_0 is generated as a matrix whose zero elements are chosen uniformly at random and non-zero elements are i.i.d. uniformly in the interval $[-500, 500]$. The distributions of A_0 and E_0 are identical to those used in [37][35][21]. All of these simulations are averaged over 20 runs.

We set threshold $\tau = 10,000$, step size $\theta = 0.9$, and the maximum number of iterations to 3000 for both SVT and HQ-SVT. The singular value decomposition (SVD) is implemented by PROPACK⁴, which uses the iterative Lanczos algorithm to compute the SVD directly. We denote the HQ-SVT based on Welsch M-estimator by HQ-SVT1 and denote the HQ-SVT based on L1-L2 M-estimator by HQ-

SVT2.

4.2. Recovery of Low Rank Matrix

In this subsection, we evaluate the recovery scalability of HQ-SVT algorithm on corrupted low-rank matrix. We set $m = 400$ and $\frac{\|E_0\|_0}{m^2} = 0.05$. Table 2 shows the simulation results on different values of rank(A_0). All five algorithms can accurately find the true rank of low rank matrix, and hence we do not list rank(A) of different algorithms in Table 2. We can observe that ALM algorithm achieves the lowest reconstruction errors. The reconstruction errors of HQ-SVT are larger than those of three other compared algorithms. That is, HQ-SVT seems to need more iterations to achieve higher accuracy.

Fig. 1 further shows the convergence of three singular value thresholding algorithms. We can observe that the three algorithms can find the true rank of low rank matrix less than 40 iterations. The variation of reconstruction error of HQ-SVT is continuous and small after 100 iterations. Algorithm based on Welsch M-estimator can converge more quickly than that based on L1-L2 M-estimator.

Although the convergence speed of HQ-SVT is slower than other methods, those simulations demonstrate that instead of using l^1 -norm to model noise, algorithms based on traditional M-estimators can also accurately recovery a corrupted low-rank matrix when error matrix E_0 is sparse. In robust statistics, outliers are those data points that are far away from other data points. If outliers are sparse as compared with other data points, M-estimator based algorithms can naturally obtain a sparse solution in terms of outliers.

⁴<http://sun.stanford.edu/~rmunk/PROPACK/>

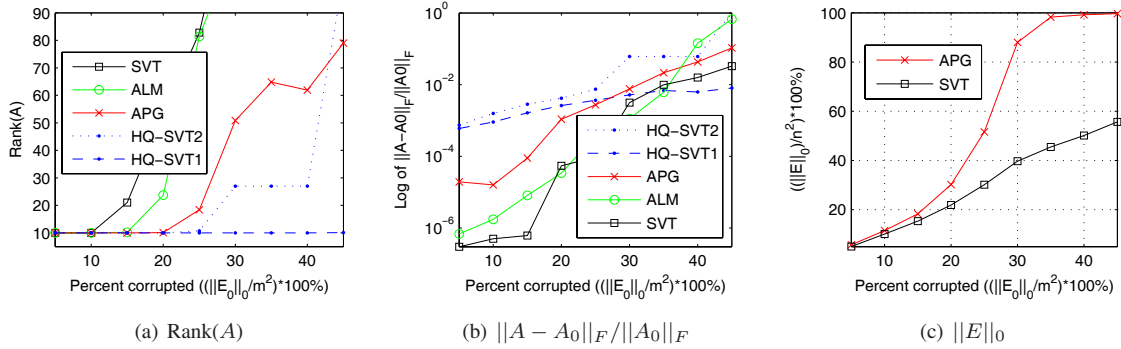


Figure 2. Comparison of the four algorithms on different levels of corruptions. All of these results on each level of corruption are averaged over 20 runs. (a) The average rank of recovered low-rank matrix A . (b) The average reconstruction error. (c) The average l^0 norm of E estimated by different methods.

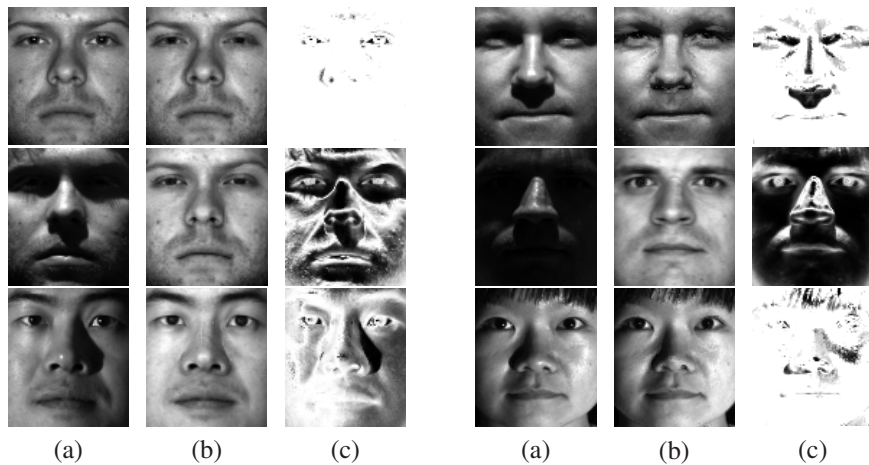


Figure 3. Removing shadows from face images. (a) Original images of a face under different illuminations from the Extended Yale B database. (b) Reconstructed images by our algorithm. (c) The auxiliary variable matrix P returned by our algorithm. The white-value denotes large weight and black-value denotes small weight.

4.3. Different Level of Corruptions

In practical application scenarios, corruptions of a low-rank matrix are often unknown. Hence we varied the percentage of non-zero elements in E_0 from 5% to 45% to evaluate robustness of different algorithms. We set $m = 200$ and $r = 10$. Fig. 2 shows the average simulation results.

From Fig. 2(a), we can observe that SVT, ALM, and APG algorithms can accurately estimate the true rank $r = 10$ under 10%, 15%, and 20% corruptions respectively. When the error matrix E_0 is non-sparse, the rank of output matrix A tends to be larger as the level of corruption is increased. That is, SVT, ALM, and APG algorithms fail to find the true rank of a large corrupted matrix. From Fig. 2(b), we can further learn that the average reconstruction errors of the three methods increase rapidly when the level of corruption is larger than 15%. Fig. 2(c) further shows the results of $\|E_0\|_0$ of SVT and APG. When the level of corruption is larger than 20%, the two algorithms begin to fail to correctly estimate the true errors in E . More elements of

E are non-zero and estimated as noise. This may be due to the fact that both two algorithms assume that noise matrix is sparse.

Compared with three other algorithms, HQ-SVT algorithms obtain more stable results on rank(A) and reconstruction error. When the low-rank matrix is 40% corrupted, HQ-SVT1 can also find the true rank $r = 10$ and the variation of its reconstruction error is small. When the level of corruption is larger than 35%, HQ-SVT achieves the lowest reconstruction errors. For L1-L2 M-estimator, HQ-SVT2 can still work under 25% corruption. This stability may benefit from the robustness of Welsch M-estimator and L1-L2 M-estimator.

4.4. Removing Shadows

In past decades, face recognition have received a lot of attention in computer vision and pattern recognition. It has been demonstrated that variation of many face images under variable lighting can be effectively modeled by low dimensional linear spaces [1, 19]. Under certain ideal-

Table 3. Comparison of different algorithms for face recognition. A lower average error rate means a better reconstruction ability.

| Method | Error rate | Std | Min | Max |
|---------|------------|------|-------|-------|
| PCA | 13.8% | 1.7% | 11.2% | 15.8% |
| APG | 12.5% | 1.9% | 9.2% | 15.1% |
| HQ-SVT1 | 12.0% | 1.7% | 8.9% | 14.5% |
| HQ-SVT2 | 12.1% | 1.8% | 9.1% | 14.2% |

ized circumstances, aligned facial images of the same subject approximately exist in a nine dimensional linear subspace [37]. However, face images would suffer from self-shadowing or saturations in brightness under directional illumination [37].

In this subsection, we validate the reconstruction ability of HQ-SVT on the Extended Yale B database [19]. The Welsch M-estimator is used. A total of 31 different illuminations are used for each person. All facial images are cropped and aligned to 96×84 pixels according to two eyes' positions. The matrix D contains well-aligned training images of a person's face under various illumination conditions. Fig. 3 shows the results of our algorithm on images from subsets 1-5. We can observe that the proposed method can remove the shadows around the nose and eye region. The auxiliary variable p_{ij} corresponding to the shadows and saturations receives a small value in Fig.3 (c).

From Fig. 3, we also observe that our method can recover a facial image even under large corruption. Although the corruption occupies nearly 90%, HQ-SVT can still utilize the left 10% pixels to perform learning. According to MCC, HQ-SVT only focuses on those pixels corresponding to white-value in Fig.3 (c). As a result, it can recover the corrupted image.

To quantitatively evaluate the reconstruction ability of different methods, we make use of them as a preprocessing step and then perform classification in the subspaces computed by learned projection matrix \hat{U} . We randomly divided 31 different face images of one subject into two subset. The one for training contains 27 face images per subject, and the one for testing contains 4 face images per subject. We performed low-rank matrix recovery algorithms on the training set and projected all data into a subspace by projection matrix \hat{U} . Then nearest-neighbor classifier is used as classifier, and average error rates of different methods are averaged over 20 runs. A lower average error rate means a better reconstruction ability.

Table 3 shows the statistical results of the compared fourth methods. In face recognition, the PCA based face recognition method is also called Eigenfaces. Although PCA can deal with small Gaussian noise in face images, it fails to deal with the errors incurred by shadows and specularities. Hence it obtains the highest average error rate. The

error rates of the three low-rank methods are 90%, 87%, and 88% of that of the PCA method respectively. That is the three methods can efficiently deal with shadows and specularities. These experimental results suggest that low-rank matrix recovery algorithms are efficient preprocessing tools for face recognition.

5. Conclusion and Future work

This paper investigates the recovery of corrupted low-rank matrix via non-convex minimization, and introduces a novel algorithm, namely, the half-quadratic based singular value thresholding algorithm. Different from those methods that assume error matrix E has a sparse representation, the proposed method is based on robust M-estimator. Simulation results demonstrate that instead of using l^1 -norm to model noise, singular value thresholding algorithms based on traditional M-estimators can also accurately recovery a corrupted low-rank matrix when the minimum nuclear-norm solution is also the lowest-rank solution. Moreover, they seem to have the ability to accurately recover a matrix even under 40% corruption. The maximum correntropy criterion gives a theoretical guaranty of this robustness under large corruption.

A half-quadratic based alternate minimization scheme is also investigated for low-rank matrix recovery. Future work is to investigate more M-estimators (such as Huber M-estimator) for accurate low-rank matrix recovery and to develop more efficient approaches to accelerate convergence under HQ optimization.

Acknowledgment

This work was supported in part by Research Foundation for the Doctoral Program of the Ministry of Education of China (Grant No. 20100041120009), 985 Project in Sun Yat-sen University (Grant No. 35000-3181305), NSF-Guangdong (Grant No. U0835005), National Natural Science Foundation of China (Grant No. 60736018, 61075024), Program of International Science and Technology Cooperation in China (Grant No. 2010DFB14110).

References

- [1] R. Basri and D. Jacobs. Lambertian reflection and linear subspaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(3):218–233, 2003.
- [2] M. Black and A. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, 26(1):63–84, 1998.
- [3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
- [4] A. M. Bruckstein, M. Elad, and M. Zibulevskyy. On the uniqueness of nonnegative sparse solutions to underdeter-

- mined systems of equations. *IEEE Trans. on Information Theory*, 54(11):4813–4820, 2008.
- [5] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 14(10):707–710, 2007.
- [6] R. Chartrand. Nonconvex compressed sensing and error correction. In *EEE International Conference on Acoustics, Speech and Signal Processing*, 2007.
- [7] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *EEE International Conference on Acoustics, Speech and Signal Processing*, 2008.
- [8] C. Ding, D. Zhou, X. He, and H. Zha. R1-pca: rotational invariant l1-norm principal component analysis for robust subspace factorization. In *International Conference on Machine Learning*, June 2006.
- [9] A. Ganesh, Z. Lin, J. Wright, L. Wu, M. Chen, and Y. Ma. Fast algorithms for recovering a corrupted low-rank matrix. In *International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2008.
- [10] A. Ganesh, J. Wright, X. Li, Emmanuel, J. Candes, and Y. Ma. Dense error correction for low-rank matrices via principal component pursuit. In *CoRR abs/1001.2362*, 2010.
- [11] R. He, B.-G. Hu, X.-T. Yuan, and W.-S. Zheng. Principal component analysis based on non-parametric maximum entropy. *Neurocomputing*, 73(10-12):1840–1852, 2010.
- [12] R. He, B. G. Hu, W. S. Zheng, and Y. Q. Guo. Two-stage sparse representation for robust recognition on large-scale database. In *Twenty- Fourth AAAI Conference on Artificial Intelligence*, 2010.
- [13] R. He, W.-S. Zheng, and B.-G. Hu. Maximum correntropy criterion for robust face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence, in press*, 2011.
- [14] R. He, W.-S. Zheng, B.-G. Hu, and X.-W. Kong. A regularized correntropy framework for robust pattern recognition. *Neural Computation, accepted*, 2011.
- [15] P. Huber. *Robust statistics*. Wiley, 1981.
- [16] J. Iglesias, M. de Bruijne, M. Loog, F. Lauze, and M. Nielsen. A family of principal component analyses for dealing with outliers. In *Medical Image Computing & Computer-Assisted Intervention*, pages 178–185, 2007.
- [17] N. Kwak. Principal component analysis based on l1-norm maximization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(9):1672–1677, 2008.
- [18] F. D. la Torre and M. Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1-3):117–142, 2003.
- [19] K.-C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.
- [20] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.
- [21] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *UIUC Technical Report UILU-ENG-09-2215*, 2009.
- [22] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *UIUC Technical Report UILU-ENG-09-2214*, 2009.
- [23] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. In *CoRR abs/1010.2955*, 2010.
- [24] W. Liu, P. Pokharel, and J. Principe. error entropy, correntropy and m-estimation. In *machine learning for signal processing*, 2006.
- [25] W. Liu, P. P. Pokharel, and J. C. Principe. Correntropy: Properties and applications in non-gaussian signal processing. *IEEE Trans. on Signal Processing*, 55(11):5286–5298, 2007.
- [26] Y. Ma. Low-rank matrix recovery and completion via convex optimization. <http://perception.csl.illinois.edu/matrix-rank/introduction.html>.
- [27] R. A. MARONNA. Principal components and orthogonal regression based on robust scales. *Technometrics*, 47:264–273, 2005.
- [28] I. Mizera and C. Muller. Breakdown points of cauchy regression-scale estimators. *Statistics and Probability Letters*, 57(1):79–89, 2002.
- [29] M. Nikolova and M. K. NG. Analysis of half-quadratic minimization methods for signal and image recovery. *Society for Industrial and Applied Mathematics*, 27(3):937–966, 2005.
- [30] K. Ohba and K. Ikeuchi. Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9:1043–1047, 1997.
- [31] K. Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science, Sixth Series*, 2:559–572, 1901.
- [32] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. Technical Report 245, 1994.
- [33] P. P. Pokharel, W. Liu, and J. C. Principe. A low complexity robust detector in impulsive noise. *Signal Processing*, 89(10):1902–1909, 2009.
- [34] J. C. Principe, D. Xu, and J. W. Fisher. Information-theoretic learning. In S. Haykin, editor, *Unsupervised Adaptive Filtering, Volume 1: Blind-Source Separation*. Wiley, 2000.
- [35] K.-C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *preprint*, 2009.
- [36] V. Vapnik. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [37] J. Wright, A. Ganesh, S. Rao, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. *Journal of the ACM*, 4:1–44, 2009.
- [38] X. Yuan and B.-G. Hu. Robust feature extraction via information theoretic learning. In *International Conference on Machine Learning*, 2009.