Recovering Low-Rank and Sparse Matrices via Robust Bilateral Factorization

Fanhua Shang[†], Yuanyuan Liu[‡], James Cheng[†], Hong Cheng[‡]
[†]Dept. of Computer Science and Engineering, The Chinese University of Hong Kong
{fhshang, jcheng}@cse.cuhk.edu.hk
[‡]Dept. of Systems Engineering and Engineering Management, The Chinese University of Hong Kong
{yyliu, hcheng}@se.cuhk.edu.hk

***Abstract*—Recovering low-rank and sparse matrices from partial, incomplete or corrupted observations is an important problem in many areas of science and engineering. In this paper, we propose a scalable robust bilateral factorization (RBF) method to recover both structured matrices from missing and grossly corrupted data such as robust matrix completion (RMC), or incomplete and grossly corrupted measurements such as compressive principal component pursuit (CPCP). With the unified framework, we first present two robust trace norm regularized bilateral factorization models for RMC and CPCP problems, which can achieve an orthogonal dictionary and a robust data representation, simultaneously. Then, we apply the alternating direction method of multipliers to efficiently solve the RMC problems. Finally, we provide the convergence analysis of our algorithm, and extend it to address general CPCP problems. Experimental results verified both the efficiency and effectiveness of our RBF method compared with the state-of-the-art methods.**

***Index Terms*—compressive principal component pursuit; robust matrix completion; RPCA; low-rank**

I. Introduction

In recent years, robust low-rank matrix factorization (RLRMF) problems have drawn a lot of attention from researchers in various research communities such as machine learning [1], [2], data mining [3], [4], signal/image processing [5], [6], and computer vision [7], [8]. In these areas, the dimension of the input data, such as digital photographs, surveillance videos, text and web documents, etc., is very high. This makes inference, learning, and recognition impractical due to the "curse of dimensionality". Even though the input data lies in a high dimensional space, in practice it is observed that the data has low intrinsic dimension which is often much smaller than the dimension of the ambient space, and the data points lie in or are close to low-dimensional structures [9].

For the high-dimensional data analysis, traditional matrix factorization (MF) methods, such as principal component analysis (PCA) and non-negative matrix factorization (NMF) are commonly used, mainly because they are simple to implement, can be solved efficiently, and are often effective in real-world applications such as latent semantic indexing, face recognition and text clustering. However, one of the main challenges faced by traditional MF methods is that the observed data is often contaminated by outliers and missing values [10], or is a small set of linear measurements [11]. To address these issues, many methods based on compressive sensing and rank minimization have been proposed. In principle, those methods aim to minimize a hybrid optimization problem involving both the l_1-norm and the trace norm (also called the nuclear norm) minimization. Moreover, it is shown that the l_1-norm and the trace norm as the convex surrogates for the l_0-norm and rank function are powerfully capable of inducing sparse and low-rank, respectively [12], [13].

In this paper, we are particularly interested in the trace norm regularized problem for RLRMF:

$$\min_{U,V} f(U,V) + \lambda\|Z\|_*, \qquad \text{s.t.}, \ Z = UV^T, \tag{1}$$

where $\lambda \geq 0$ is a regularization parameter, $\|Z\|_*$ is the trace norm of the matrix Z, i.e., the sum of its singular values, and $f(\cdot)$ denotes the loss function such as the l_2-norm or the l_1-norm loss functions. In the following, we will give a few examples of applications where the RLRMF is useful.

RMC: When the real measurements contain both outliers and missing data, the model (1) is a robust matrix completion (RMC) problem [14], where missing data and outliers are presented at arbitrary location in the measurement matrix. RMC has been used in a wide range of problems such as collaborative filtering [15], structure-from-motion [7], [16] and face reconstruction [8].

RPCA: When $f(\cdot)$ is the l_1-norm loss function, the model (1) is a low-rank and sparse matrix recovery problem from grossly corrupted data matrices, also called robust principal component analysis (RPCA) [17] or low-rank and sparse decomposition (LRSD) [18]. RPCA problem has been successfully applied in many important applications such as latent semantic indexing [19], video surveillance [12], [17], and image alignment [20].

In this paper, we aim to learn robust low-rank bilateral factorization to recover low-rank and sparse matrices from corrupted data or a small set of linear measurements. Unlike existing RLRMF methods, our approach not only takes into account the fact that the observation is contaminated by both additive outliers and missing values, but also can identify both low-rank and sparse noisy components from missing and grossly corrupted measurements, i.e., CPCP problems [11]. We present a unified robust bilateral factorization (RBF) framework for RMC and CPCP problems. We verify with convincing experimental results both the efficiency and effectiveness of our RBF method.

1550-4786/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDM.2014.80

The main contributions of this paper include:

1. We propose a unified RBF framework to learn both an orthogonality dictionary and a low-dimensional data representation with missing and grossly corrupted observations, which can simultaneously recover both low-rank and sparse matrices.

2. We present two trace norm regularized RBF models for RMC and CPCP problems. With the orthogonality constraint applied to the dictionary component, we convert the proposed models into two smaller-scale matrix trace norm regularized problems.

3. Finally, we propose an efficient alternating direction method of multipliers (ADMM) to solve RMC problems with guaranteed local convergence and extend it to address CPCP problems with the linearization technique.

II. BACKGROUND

A low-rank structured matrix $Z \in \mathbb{R}^{m\times n}$ $(m \geq n)$ and a sparse one $E \in \mathbb{R}^{m\times n}$ could be recovered from highly corrupted measurements $D = \mathcal{P}_Q(X) \in \mathbb{R}^N$ via the following CPCP problem [11],

$$\min_{Z,\,E} \|E\|_1 + \lambda\|Z\|_*, \text{ s.t., } \mathcal{P}_Q(X) = \mathcal{P}_Q(Z+E), \quad (2)$$

where $\|\cdot\|_1$ denotes the l_1-norm, i.e., $\|E\|_1 = \Sigma_{ij}|e_{ij}|$, $Q \subseteq \mathbb{R}^{m\times n}$ is a linear subspace, and $\mathcal{P}_Q$ is the projection operator onto that subspace. The theoretical result in [11] states that a commensurately small number of measurements are sufficient to accurately recover the low-rank and sparse matrices with high probability. If Q is the entire space, the model (2) degenerates to the RPCA problem [12], [17].

When $\mathcal{P}_Q = \mathcal{P}_\Omega$, the model (2) is the following robust matrix completion (RMC) problem

$$\min_{Z,\,E} \|E\|_1 + \lambda\|Z\|_*, \text{ s.t., } \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(Z+E), \quad (3)$$

where Ω is the index set of observed entries and $\mathcal{P}_\Omega$ keeps the entries in Ω and zeros out others. Although both models (2) and (3) are convex optimization problems, and their algorithms converge to the globally optimal solution, they involve singular value decomposition (SVD) in each iteration and suffer from a high computational cost of $O(mn^2)$. To address this problem, we will propose a robust bilateral factorization method with missing and grossly corrupted observations.

III. OUR RBF FRAMEWORK

The high computational complexity of existing CPCP and RMC algorithms is caused by the singular value shrinkage operator, which involves the SVD of a matrix. Motivated by this, robust bilateral factorization (RBF) aims to find two smaller matrices $U \in \mathbb{R}^{m\times d}$ $(U^TU = I)$ and $V \in \mathbb{R}^{n\times d}$ whose product is equal to the matrix of low-rank $Z \in \mathbb{R}^{m\times n}$, i.e., $Z = UV^T$, where d is an upper bound for the rank of Z, i.e., $d \geq r = \text{rank}(Z)$.

A. RMC Model

From the optimization problem (3), we easily find the optimal solution $E_{\Omega^C} = \mathbf{0}$, where Ω^C is the complement of Ω, i.e., the index set of unobserved entries. Consequently, we have the following lemma.

Lemma 1 ([21]). *The RMC model (3) with the operator $\mathcal{P}_\Omega$ is equivalent to the following problem*

$$\begin{aligned}&\min_{Z,E} \|\mathcal{P}_\Omega(E)\|_1 + \lambda\|Z\|_*,\\ &\text{s.t., } \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(Z+E),\ \mathcal{P}_{\Omega^C}(E) = \mathbf{0}.\end{aligned} \quad (4)$$

For the incomplete and corrupted matrix X, our RBF model is to find two smaller matrices of low-rank, whose product approximates X, formulated as follows:

$$\begin{aligned}&\min_{U,V,E} \|\mathcal{P}_\Omega(E)\|_1 + \lambda\|UV^T\|_*,\\ &\text{s.t., } \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(UV^T+E).\end{aligned} \quad (5)$$

Lemma 2. *Let U and V be two matrices of compatible dimensions, where U has orthogonal columns, i.e., $U^TU = I$, then we have $\|UV^T\|_* = \|V\|_*$.*

By substituting $\|UV^T\|_* = \|V\|_*$ into (5), we obtain a much smaller-scale matrix trace norm minimization problem,

$$\begin{aligned}&\min_{U,V,E} \|\mathcal{P}_\Omega(E)\|_1 + \lambda\|V\|_*,\\ &\text{s.t., } \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(UV^T+E),\ U^TU = I.\end{aligned} \quad (6)$$

Theorem 1. *Suppose $(Z^*,\, E^*)$ is a solution of the problem (4) with rank$(Z^*) = r$, then there exists the solution $U_k \in \mathbb{R}^{m\times d}$, $V_k \in \mathbb{R}^{n\times d}$ and $E_k \in \mathbb{R}^{m\times n}$ to the problem (6) with $d \geq r$ and $(E_k)_{\Omega^C} = \mathbf{0}$, and $(U_kV_k^T,\, E_k)$ is also a solution to the problem (4).*

The detailed proof of this theorem can be found in [21].

B. CPCP Model

For a small set of linear measurements $D \in \mathbb{R}^N$, our CPCP problem is to recover the low-rank and sparse matrices as follows,

$$\begin{aligned}&\min_{U,V,E} \|E\|_1 + \lambda\|V\|_*,\\ &\text{s.t., } \mathcal{P}_Q(X) = \mathcal{P}_Q(UV^T+E).\end{aligned} \quad (7)$$

Theorem 2. *Suppose $(Z^*,\, E^*)$ is a solution of the problem (2) with rank$(Z^*) = r$, then there exists the solution $U_k \in \mathbb{R}^{m\times d}$, $V_k \in \mathbb{R}^{n\times d}$ and $E_k \in \mathbb{R}^{m\times n}$ to the problem (7) with $d \geq r$, and $(U_kV_k^T,\, E_k)$ is also a solution to the problem (2).*

We omit the proof of this theorem since it is similar to that of Theorem 1. In the following, we will discuss how to solve the models (6) and (7). It is worth noting that the RPCA problem can be viewed as a special case of the RMC problem (6) when all entries of the corrupted matrix are directly observed. We will mainly develop an efficient solver based on alternating direction method of multipliers (ADMM) for solving the non-convex problem (6). Although our algorithm will produce different estimations of U and V, the estimation of UV^T is stable as guaranteed by Theorems 1 and 2, and the convexity of the problems (2) and (3).

IV. OPTIMIZATION ALGORITHM

In this section, we propose an efficient ADMM algorithm for solving the RMC problem (6), and then extend it for solving the CPCP problem (7). We provide the convergence analysis of our algorithm in Section V.

A. Formulation

For efficiently solving the RMC problem (6), we can assume without loss of generality that the unobserved data $X_{\Omega^C}=\mathbf{0}$, and E_{Ω^C} may be any values such that $\mathcal{P}_{\Omega^C}(X)=\mathcal{P}_{\Omega^C}(UV^T)+\mathcal{P}_{\Omega^C}(E)$. Consequently, the constraint with the linear projection operator $\mathcal{P}_\Omega$ in (6) is simplified into $X=UV^T+E$. We introduce the constraint $X=UV^T+E$ into (6), and obtain the following equivalent form:

$$\begin{aligned}&\min_{U,V,E}\ \|\mathcal{P}_\Omega(E)\|_1+\lambda\|V\|_*,\\ &\text{s.t., } X=UV^T+E,\ U^TU=I.\end{aligned}\tag{8}$$

The partial augmented Lagrangian function of (8) is

$$\begin{aligned}&\mathcal{L}_\alpha(U,V,E,Y)=\|\mathcal{P}_\Omega(E)\|_1+\lambda\|V\|_*\\ &+\langle Y,X-E-UV^T\rangle+\frac{\alpha}{2}\|X-E-UV^T\|_F^2,\end{aligned}\tag{9}$$

where $Y\in\mathbb{R}^{m\times n}$ is a matrix of Lagrange multipliers, and $\alpha>0$ is a penalty parameter.

B. Robust Bilateral Factorization Scheme

We will derive our scheme for solving the following sub-problems with respect to U, V and E, respectively,

$$\begin{aligned}U_{k+1}=&\arg\min_{U\in\mathbb{R}^{m\times d}}\mathcal{L}_{\alpha_k}(U,V_k,E_k,Y_k),\\ &\text{s.t., } U^TU=I,\end{aligned}\tag{10}$$

$$V_{k+1}=\arg\min_{V\in\mathbb{R}^{n\times d}}\mathcal{L}_{\alpha_k}(U_{k+1},V,E_k,Y_k),\tag{11}$$

$$E_{k+1}=\arg\min_{E\in\mathbb{R}^{m\times n}}\mathcal{L}_{\alpha_k}(U_{k+1},V_{k+1},E,Y_k).\tag{12}$$

1) Updating U: Fixing V and E at their latest values, and by removing the terms that do not depend on U and adding some proper terms that do not depend on U, the problem (10) with respect to U is reformulated as follows:

$$\min_U\ \|UV_k^T-P_k\|_F^2,\qquad \text{s.t., } U^TU=I,\tag{13}$$

where $P_k=X-E_k+Y_k/\alpha_k$. In fact, the optimal solution can be given by the SVD of the matrix P_kV_k as in [22]. To further speed-up the calculation, we introduce the idea in [23] that uses a QR decomposition instead of SVD. The resulting iteration step is formulated as follows:

$$U_{k+1}=Q,\qquad \text{QR}(P_kV_k)=QR,\tag{14}$$

where U_{k+1} is an orthogonal basis for the range space $\mathcal{R}(P_kV_k)$, i.e., $\mathcal{R}(U_{k+1})=\mathcal{R}(P_kV_k)$. Although U_{k+1} in (14) is not an optimal solution to (13), our iterative scheme and the one in [24] are equivalent to solve (13) and (15), and their equivalent analysis is provided in Section V. Moreover, the use of QR factorizations also makes our update scheme highly scalable on modern parallel architectures [25].

2) Updating V: Fixing U and E, the optimization problem (11) with respect to V can be rewritten as follows:

$$\min_V\ \frac{\alpha_k}{2}\|U_{k+1}V^T-P_k\|_F^2+\lambda\|V\|_*.\tag{15}$$

To solve the problem (15), we first introduce the following soft-thresholding operator $\mathcal{S}_\tau$ and the singular value thresholding (SVT) operator.

$$\mathcal{S}_\tau(A_{ij}):=\begin{cases}A_{ij}-\tau, & A_{ij}>\tau,\\ A_{ij}+\tau, & A_{ij}<-\tau,\\ 0, & \text{otherwise.}\end{cases}$$

Definition 1. *For any given matrix $M\in\mathbb{R}^{n\times d}$ whose rank is r, and $\mu\geq 0$, the* SVT *operator is defined as follows:*

$$\text{SVT}_\mu(M)=\overline{U}\text{diag}(\mathcal{S}_\mu(\sigma))\overline{V}^T,$$

where $\overline{U}\in\mathbb{R}^{n\times r}$, $\overline{V}\in\mathbb{R}^{d\times r}$ and $\sigma=(\sigma_1,\ldots,\sigma_r)^T\in\mathbb{R}^{r\times 1}$ are obtained by SVD of M, i.e., $M=\overline{U}\,\text{diag}(\sigma)\,\overline{V}$.

Theorem 3. *The trace norm minimization problem (15) has a closed-form solution given by:*

$$V_{k+1}=\text{SVT}_{\lambda/\alpha_k}(P_k^TU_{k+1}).\tag{16}$$

Proof. The first-order optimality condition for (15) is

$$0\in\lambda\partial\|V^T\|_*+\alpha_kU_{k+1}^T(U_{k+1}V^T-P_k),$$

where $\partial\|\cdot\|_*$ is the set of subgradients of the trace norm. Since $U_{k+1}^TU_{k+1}=I$, the optimality condition of (15) is rewritten as follows:

$$0\in\lambda\partial\|V\|_*+\alpha_k(V-P_k^TU_{k+1}).\tag{17}$$

(17) is also the optimality condition for the following problem,

$$\min_V\ \frac{\alpha_k}{2}\|V-P_k^TU_{k+1}\|_F^2+\lambda\|V\|_*.\tag{18}$$

According to Theorem 2.1 in [26], then the optimal solution of (18) is given by (16). □

3) Updating E: Fixing U and V, E is updated by solving

$$\min_E\ \|\mathcal{P}_\Omega(E)\|_1+\frac{\alpha_k}{2}\|E+U_{k+1}V_{k+1}^T-X-Y_k/\alpha_k\|_F^2.\tag{19}$$

The optimal solution E_{k+1} can be obtained by solving the following two subproblems with respect to E_Ω and E_{Ω^C}:

$$\min_{E_\Omega}\ \|E\|_1+\frac{\alpha_k}{2}\|E+U_{k+1}V_{k+1}^T-X-Y_k/\alpha_k\|_F^2,\tag{20}$$

$$\min_{E_{\Omega^C}}\ \|E+U_{k+1}V_{k+1}^T-X-Y_k/\alpha_k\|_F^2.\tag{21}$$

By the operator $\mathcal{S}_\tau$ and letting $\tau=1/\alpha_k$, the closed-form solution to the problem (20) is given by

$$(E_{k+1})_\Omega=\mathcal{S}_\tau((X-U_{k+1}V_{k+1}^T+Y_k/\alpha_k)_\Omega).\tag{22}$$

We can easily obtain the closed-form solution to (21)

$$(E_{k+1})_{\Omega^C}=(X-U_{k+1}V_{k+1}^T+Y_k/\alpha_k)_{\Omega^C}.\tag{23}$$

Summarizing the analysis above, we arrive at an ADMM algorithm to solve the RMC problem (6), as outlined in **Algorithm 1**. Our algorithm is essentially a Gauss-Seidel-type scheme of ADMM, and the update strategy of the Jacobi version of ADMM is easily implemented, well suited for parallel and distributed computing and hence is particularly attractive for solving large-scale problems [27]. In addition, E_{Ω^C} should be set to $\mathbf{0}$ for the expected output E. This algorithm can also be accelerated by adaptively changing α. An efficient strategy [28] is to let $\alpha=\alpha_0$ (the initialization in Algorithm 1) and increase α_k iteratively by $\alpha_{k+1}=\rho\alpha_k$, where $\rho\in(1.0,1.2]$ in general and α_0 is a small constant. Algorithm 1 is easily used to solve the RPCA problem, where all entries of the corrupted matrix are directly observed.

Algorithm 1 Solving RMC problem (6) via ADMM

Input: $\mathcal{P}_\Omega(X)$, λ and tol.

Initialize: $U_0 = \text{eye}(m,d)$, $V_0 = \mathbf{0}$, $Y_0 = \mathbf{0}$, $\alpha_0 = 10^{-4}$, $\alpha_{max} = 10^{10}$, and $\rho = 1.2$.

while not converged **do**
  Update U_{k+1} by (14);
  Update V_{k+1} by (16);
  Update E_{k+1} by (22) and (23);
  Update Y_{k+1} by
    $Y_{k+1} = Y_k + \alpha_k(X - U_{k+1}V_{k+1}^T - E_{k+1})$;
  Update α_{k+1} by $\alpha_{k+1} = \min(\rho\alpha_k,\, \alpha_{max})$;
  Check the convergence condition,
    $\|X - U_{k+1}V_{k+1}^T - E_{k+1}\|_F < \text{tol}$;
end while

Output: U, V and E, where $\mathcal{P}_{\Omega^C}(E)$ is set to $\mathbf{0}$.

C. Extension for CPCP

Algorithm 1 can be extended to solve the CPCP problem (7) with the complex operator $\mathcal{P}_Q$, as outlined in **Algorithm 2**, which is to optimize the following augmented Lagrange function,

$$\begin{aligned}\mathcal{F}_\alpha(U,V,E,Y) =& \lambda\|V\|_* + \|E\|_1 \\ +\langle Y, D - \mathcal{P}_Q(E+UV^T)\rangle +& \frac{\alpha}{2}\|D - \mathcal{P}_Q(E+UV^T)\|^2.\end{aligned} \tag{24}$$

We minimize $\mathcal{F}_\alpha$ with respect to (U,V,E) by using a recently proposed linearization technique [29], which solves such problems where $\mathcal{P}_Q$ is not the identity operator. Specifically, for updating U and V, let $T = UV^T$ and $g(T) = \frac{\alpha_k}{2}\|D - \mathcal{P}_Q(E_k + T) + Y_k/\alpha_k\|^2$, then $g(T)$ is approximated by

$$g(T) \approx g(T_k) + \langle \nabla g(T_k),\, T - T_k\rangle + \tau\|T - T_k\|^2, \tag{25}$$

where $\nabla g(T_k) = \alpha_k \mathcal{P}_Q^\star(\mathcal{P}_Q(T_k + E_k) - D - Y_k/\alpha_k)$, $\mathcal{P}_Q^\star$ is the adjoint operator of $\mathcal{P}_Q$, and τ is chosen as $\tau = 1/\|\mathcal{P}_Q^\star\mathcal{P}_Q\|_2$ as in [29], and $\|\cdot\|_2$ the spectral norm of a matrix, i.e., the largest singular value of a matrix.

Similarly, for updating E, let $T_{k+1} = U_{k+1}V_{k+1}^T$ and $h(E) = \frac{\alpha_k}{2}\|D - \mathcal{P}_Q(E + T_{k+1}) + Y_k/\alpha_k\|^2$, then $h(E)$ is approximated by

$$h(E) \approx h(E_k) + \langle \nabla h(E_k),\, E - E_k\rangle + \tau\|E - E_k\|^2, \tag{26}$$

where $\nabla h(E_k) = \alpha_k \mathcal{P}_Q^\star(\mathcal{P}_Q(E_k + T_{k+1}) - D - Y_k/\alpha_k)$.

V. Theoretical Analysis

In this section, we will present several theoretical properties of Algorithm 1. First, we provide the equivalent relationship analysis for our iterative scheme and the one in [24], as shown by the following theorem.

Theorem 4. *Let (U_k^*, V_k^*, E_k^*) be the solution of the subproblems (10), (11) and (12) at the k-th iteration, respectively, $Y_k^* = Y_{k-1}^* + \alpha_{k-1}(X - U_k^*(V_k^*)^T - E_k^*)$, and (U_k, V_k, E_k, Y_k) be generated by Algorithm 1 at the k-th iteration ($k = 1, \ldots, T$). Then*

1) *$\exists O_k \in \mathcal{O} = \{M \in \mathbb{R}^{d\times d} | M^T M = I\}$ such that $U_k^* = U_k O_k$ and $V_k^* = V_k O_k$.*
2) *$U_k^*(V_k^*)^T = U_k V_k^T$, $\|V_k^*\|_* = \|V_k\|_*$, $E_k^* = E_k$, and $Y_k^* = Y_k$.*

Algorithm 2 Solving CPCP problem (7) via ADMM

Input: $D \in \mathbb{R}^N$, $\mathcal{P}_Q$, λ and tol.

while not converged **do**
  Update U_{k+1} by
    $U_{k+1} = Q,\ \text{QR}((U_k V_k^T - \nabla g(U_k V_k^T)/\tau)V_k) = QR$;
  Update V_{k+1} by
    $V_{k+1}^T = \text{SVT}_{\lambda/\alpha_k}(U_{k+1}^T(U_k V_k^T - \nabla g(U_k V_k^T)/\tau))$;
  Update E_{k+1} by $E_{k+1} = \mathcal{S}_{1/\alpha_k}(E_k - \nabla h(E_k)/\tau)$;
  Update Y_{k+1} by
    $Y_{k+1} = Y_k + \alpha_k(D - \mathcal{P}_Q(U_{k+1}V_{k+1}^T + E_{k+1}))$;
  Update α_{k+1} by $\alpha_{k+1} = \min(\rho\alpha_k,\, \alpha_{max})$;
  Check the convergence condition,
    $(\|T_{k+1} - T_k\|_F^2 + \|E_{k+1} - E_k\|_F^2)/(\|T_k\|_F^2 + \|E_k\|_F^2) <$ tol;
end while

Output: U, V and E.

Remark: The proof of this theorem can be found in [21]. Since the Lagrange function (9) is determined by the product UV^T, V, E and Y, the different values of U and V are essentially equivalent as long as the same product UV^T and $\|V\|_* = \|V^*\|_*$. Meanwhile, our scheme replaces SVD by the QR decomposition, and can avoid the SVD computation for solving the optimization problem with the orthogonal constraint.

A. Convergence Analysis

The convergence of our derived ADMM algorithm is guaranteed, as shown in the following theorem.

Theorem 5. *Let (U_k, V_k, E_k) be a sequence generated by Algorithm 1, then we have the following conclusions:*

1) *(U_k, V_k, E_k) approaches to a feasible solution, i.e., $lim_{k\to\infty}\|X - U_k V_k^T - E_k\|_F = 0$.*
2) *Both sequences $U_k V_k^T$ and E_k are Cauchy sequences.*
3) *(U_k, V_k, E_k) converges to a KKT point of the problem (8).*

The proof of this theorem can be found in [21].

B. Complexity Analysis

We also analyze the time complexity of our RBF algorithm. For the RMC problem (6), the main running time of our RBF algorithm is consumed by performing SVD on the small matrix of size $n \times d$, the QR decomposition of the matrix $P_k V_k$, and some matrix multiplications. In (16), the time complexity of performing SVD is $O(d^2 n)$. The time complexity of QR decomposition and matrix multiplications is $O(d^2 m + mnd)$. The total time complexity of our RBF algorithm for solving the problem (6) is $O(t(d^2 n + d^2 m + mnd))$ (usually $d \ll n \leq m$), where t is the number of iterations.

VI. Experimental Evaluation

We now evaluate the effectiveness and efficiency of our RBF method for RMC and CPCP problems, such as text removal and face reconstruction. We ran experiments on an Intel(R)

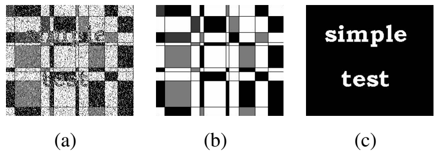

Fig. 1. Images used in text removal experiments: (a) Input image; (b) Original image; (c) Outlier mask.

Core (TM) i5-4570 (3.20 GHz) PC running Windows 7 with 8GB main memory.

A. Text Removal

We first conduct an experiment on artificially generated data, whose goal is to remove some generated text from an image. The ground-truth image for the data matrix is of size 256×222 with rank equal to 10. We then add to the image a short phase in text form which plays the role of outliers. Fig. 1 shows the image together with the clean image and outliers mask. For fairness, we set the rank of all the algorithms to 20, which is two times the true rank of the underlying matrix. The input data are generated by setting 30% of the randomly selected pixels of the image as missing entries. We compare our RBF method with the state-of-the-art methods, including PCP [12], SpaRCS[1] [30], RegL1[2] [7] and BF-ALM [31]. We set the regularization parameter $\lambda = \sqrt{\max(m, n)}$ for RegL1 and RBF, and the stopping tolerance $\text{tol} = 10^{-4}$ for all algorithms in this section.

The recovery results obtained by all these methods are shown in Fig. 2, where the error of low-rank component recovery (i.e., Error $:= \|M - Z\|_F / \|M\|_F$, where M and Z represent the ground-truth image matrix and the recovered image matrix, respectively) and the outlier detection accuracy (the score Area Under the receiver operating characteristic Curve, AUC) are also reported. For outlier detection, it can be observed that our RBF method significantly outperforms the other methods. As far as low-rank matrix recovery is concerned, these RMC methods including SpaRCS, RegL1 and RBF, outperform PCP, not only visually but also quantitatively. Overall, RBF significantly outperforms PCP, RegL1 and SpaRCS in terms of both low-rank matrix recovery and sparse outlier identification. In terms of efficiency, RBF is tens of times faster than the other methods, where the running time of PCP, SpaRCS, RegL1, BF-ALM and RBF is 36.25sec, 15.68sec, 26.85sec, 6.36sec and 0.87sec, respectively.

B. Face Reconstruction

We also test our RBF method for the face reconstruction problems with the incomplete and corrupted face data or a small set of linear measurements D as in [11], respectively. The face database used here is a part of Extended Yale Face Database B [32] with large corruptions. The face images can often be decomposed as a low-rank part, capturing the face appearances under different illuminations, and a sparse

[1] http://www.ece.rice.edu/~aew2/sparcs.html

[2] https://sites.google.com/site/yinqiangzheng/

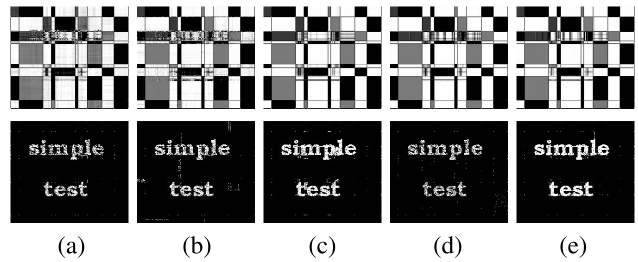

Fig. 2. Text removal results. The first row shows the foreground masks and the second row shows the recovered background images: (a) PCP (Error: 0.2516; AUC: 0.8558); (b) SpaRCS (Error: 0.2416; AUC: 0.8665); (c) RegL1 (Error: 0.2291; AUC: 0.8792); (d) BF-ALM (Error: 0.2435; AUC: 0.8568); (e) RBF (Error: 0.1844; AUC: 0.9227).

Fig. 3. Face recovery results by these algorithms. From left column to right column: Input corrupted images (black pixels denote missing entries), original images, reconstruction results by CWM (1830.18sec), RegL1 (2416.85sec) and RBF (52.73sec), respectively.

component, representing varying illumination conditions and heavily "shadows". The resolution of all images is 192×168 and the pixel values are normalized to [0, 1], then the pixel values are used to form data vectors of dimension 32,256. The input data are generated by setting 40% of the randomly selected pixels of each image as missing entries.

Fig. 3 shows some original and reconstructed images by CWM[3] [8], RegL1 and RBF, where the average computational time of all these algorithms on each people's faces is reported. It can be seen that RBF not only performs better than the other methods visually, but is also more than 30 times faster, and effectively eliminates the heavy noise and "shadows" and simultaneously completes the missing entries. In other words, RBF can achieve the latent features underlying the original images regardless of the observed data corrupted by outliers and missing values.

Moreover, we implement a more challenging problem to recover face images from incomplete line measurements. Considering the computational burden of the projection operator $\mathcal{P}_Q$, we resize the original images into 42×48 and normalize the raw pixel values to form data vectors of dimension 2016. Following [11], the input data is $\mathcal{P}_Q(X)$, where Q is a sub-

[3] http://www4.comp.polyu.edu.hk/~cslzhang/papers.htm

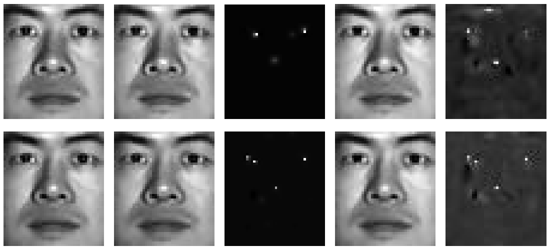

Fig. 4. Face reconstruction results by CPCP and RBF, where the first column shows the original images, the second and third columns show the low-rank and sparse components obtained by CPCP, while the last two columns show the low-rank and sparse components obtained by RBF.

space generated randomly with the dimension $0.75mn$. Fig. 4 illustrates some reconstructed images by CPCP [11] and RBF, respectively. It is clear that both CPCP and RBF effectively remove "shadows" from faces images and simultaneously successfully recover both low-rank and sparse components from the reduced measurements.

VII. CONCLUSIONS

In this paper, we proposed a robust bilateral factorization (RBF) framework for RMC and CPCP problems. Unlike existing RLRMF methods, RBF can not only address large-scale RMC problems, but also low-rank and sparse matrix recovery problems with incomplete or corrupted observations. To this end, we first presented two smaller-scale matrix trace norm regularized models for RMC and CPCP problems. Then we developed an efficient ADMM algorithm to solve both RMC and RPCA problems, and analyzed the convergence of our algorithm. Finally, we extended our algorithm to address CPCP problems. The extensive experimental results on real-world data sets demonstrated the superior performance of our RBF method in comparison with the state-of-the-art methods in terms of both efficiency and effectiveness.

VIII. ACKNOWLEDGMENTS

We thank the reviewers for giving us many constructive comments, with which we have significantly improved our paper. This research is supported in part by SHIAE Grant No. 8115048, MSRA Grant No. 6903555, GRF No. 411211, and CUHK direct grant Nos. 4055015 and 4055017.

REFERENCES

[1] K. Mitra, S. Sheorey, and R. Chellappa, "Large-scale matrix factorization with missing data under additional constraints," in *NIPS*, pp. 1651–1659, 2011.

[2] N. Gillis and F. Glineur, "Low-rank matrix approximation with weights or missing data is NP-hard," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1149–1165, 2011.

[3] R. Gemulla, P. Haas, E. Nijkamp, and Y. Sismanis, "Large-scale matrix factorization with distributed stochastic gradient descent," in *KDD*, pp. 69–77, 2011.

[4] Y. Liu, L. Jiao, and F. Shang, "An efficient matrix factorization based low-rank representation for subspace clustering," *Pattern Recogn.*, vol. 46, no. 1, pp. 284–292, 2013.

[5] R. Ma, N. Barzigar, A. Roozgard, and S. Cheng, "Decomposition approach for low-rank matrix completion and its applications," *IEEE Trans. Singal Proc.*, vol. 62, no. 7, pp. 1671–1683, 2014.

[6] F. Shang, Y. Liu, J. Cheng, and H. Cheng, "Robust principal component analysis with missing data," in *CIKM*, 2014.

[7] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, "Practical low-rank matrix approximation under robust L1-norm," in *CVPR*, pp. 1410–1417, 2012.

[8] D. Meng, Z. Xu, L. Zhang, and J. Zhao, "A cyclic weighted median method for L1 low-rank matrix factorization with missing entries," in *AAAI*, 2013.

[9] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, 2013.

[10] P. Favaro, R. Vidal, and A. Ravichandran, "A closed form solution to robust subspace estimation and clustering," in *CVPR*, pp. 1801–1807, 2011.

[11] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Inform. Infer.*, vol. 2, pp. 32–68, 2013.

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

[13] M. Tao and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM J. Optim.*, vol. 21, no. 1, pp. 57–81, 2011.

[14] Y. Chen, H. Xu, C. Caramanis, and S. Sanghavi, "Robust matrix completion and corrupted columns," in *ICML*, pp. 873–880, 2011.

[15] F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding, "Robust matrix completion via joint Schatten p-norm and lp-norm minimization," in *ICDM*, pp. 566–574, 2012.

[16] A. Eriksson and A. van den Hengel, "Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L1 norm," in *CVPR*, pp. 771–778, 2010.

[17] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," in *NIPS*, pp. 2080–2088, 2009.

[18] X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," *Pac. J. Optim.*, vol. 9, no. 1, pp. 167–180, 2013.

[19] K. Min, Z. Zhang, J. Wright, and Y. Ma, "Decomposition background topics from keywords by principal component pursuit," in *CIKM*, pp. 269–278, 2010.

[20] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, 2012.

[21] F. Shang, Y. Liu, H. Tong, J. Cheng, and H. Cheng, "Structured low-rank matrix factorization with missing and grossly corrupted observations," *CoRR abs/1409.1062*, 2014.

[22] H. Nick, "Matrix procrustes problems," 1995.

[23] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Math. Prog. Comp.*, vol. 4, no. 4, pp. 333–361, 2012.

[24] G. Liu and S. Yan, "Active subspace: toward scalable low-rank learning," *Neur. Comp.*, vol. 24, no. 12, pp. 3371–3394, 2012.

[25] H. Avron, S. Kale, S. Kasiviswanathan, and V. Sindhwani, "Efficient and practical stochastic subgradient descent for nuclear norm regularization," in *ICML*, 2012.

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

[27] F. Shang, Y. Liu, and J. Cheng, "Generalized higher-order tensor decomposition via parallel ADMM," in *AAAI*, pp. 1279–1285, 2014.

[28] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *NIPS*, pp. 612–620, 2011.

[29] J. Yang and X. Yuan, "Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization," *Math. Comp.*, vol. 82, pp. 301–329, 2013.

[30] A. Waters, A. Sankaranarayanan, and R. Baraniuk, "SpaRCS: Recovering low-rank and sparse matrices from compressive measurements," in *NIPS*, pp. 1089–1097, 2011.

[31] R. Cabral, F. Torre, J. Costeira, and A. Bernardino, "Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition," in *ICCV*, pp. 2488–2495, 2013.

[32] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, 2005.