AN ALTERNATING MINIMIZATION METHOD FOR MATRIX COMPLETION PROBLEM

YUAN SHEN* AND XIN LIU†

Abstract. In this paper, we focus on solving matrix completion problem arising from applications in the fields of information theory, statistics, engineering, etc. However, the matrix completion problem involves nonconvex rank constraints which make this type of problem difficult to handle. Traditional approaches use a nuclear norm surrogate to replace the rank constraints. The relaxed model is convex, and hence can be solved by a bunch of existing algorithms. However, these algorithms need to compute the costly singular value decomposition (SVD) which makes them impractical for handling large-scale problems. We retain the rank constraints in the optimization model, and propose an alternating minimization method for solving it. The resulting algorithm does not need SVD computation, and shows satisfactory speed performance. As a nonconvex algorithm, the new algorithm has better theoretical property than competing algorithms.

1. Introduction. In this paper, we focus on the matrix completion (MC) problem which aims to find a lowest-rank matrix given a subset of its entries, that is,

$$\min_{W\in\mathbb{R}^{m\times n}} \ \mathrm{rank}(W), \mathrm{s.t.} W_{ij} = M_{ij},\ \forall (i,j)\in\Omega, \tag{1.1}$$

where $\mathrm{rank}(W)$ denotes the rank of W, and $M_{i,j}\in\mathbb{R}$ are given for $(i,j)\in\Omega\subset\{(i,j): 1\le i\le m, 1\le j\le n\}$.

The problem of minimizing the rank of a matrix arises in many applications, and can be formulated as the above model, for example, image restoration [42], control and systems theory, model reduction and minimum order control synthesis [29], recovering shape and motion from image streams [33, 39], data mining and pattern recognitions [14] and machine learning such as latent semantic indexing, collaborative prediction and low-dimensional embedding.

As is well known, the ℓ_1-norm minimization has been used to recover sparse signals in compressive sensing (see, for example, [5, 6, 7, 12]), and nuclear-norm minimization has been used to recover a low-rank matrix from a subset of its entries in matrix completion (see, for example, [2, 40]).

In particular, Candès and Recht in [4] proved that a given rank-r matrix M satisfying certain incoherence conditions can be recovered exactly by (1.2) with high probability from a subset Ω of uniformly sampled entries whose cardinality $|\Omega|$ is of the order $O(r(m+n)\log(m+n))$. For more refined theoretical results on matrix completion we refer the reader to [3, 8, 16, 20, 21, 35, 44].

Although problem (1.1) is generally NP-hard due to the combinational nature of the function $\mathrm{rank}(\cdot)$, it has been shown in [36, 4, 8] that, under some reasonable conditions, the solution of problem (1.1) can be found by solving a convex optimization problem:

$$\min_{W\in\mathbb{R}^{m\times n}} \ \|W\|_*, \mathrm{s.t.} W_{ij} = M_{ij},\ \forall (i,j)\in\Omega, \tag{1.2}$$

where the *nuclear* or *trace* norm $\|W\|_*$ is the summation of the singular values of W.

A prominent feature of the model (1.2) lies at its convexity. Several algorithms have been developed to solve it. One type of algorithms is based on a scheme called "iterative shrinkage" which has been successfully applied to ℓ_1 minimization (see, for example [13]), and to matrix completion [2, 30].

Various types of algorithms have been proposed to recover the solution of (1.1) based on solving (1.2). One method is the singular value thresholding (SVT) algorithm [2] using soft-thresholding operations on the singular

*School of Applied Mathematics, Nanjing University Of Finance & Economics, CHINA (ocsiban@126.com)Research supported by NSFC grants 11401295 and 11726618, and by Major Program of the National Social Science Foundation of China under Grant 12&ZD114 and by National Social Science Foundation of China under Grant 15BGL58 and by Qinglan Project of Jiangsu Province.

†State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, CHINA (liuxin@lsec.cc.ac.cn). Research supported in part by NSFC grants 11471325, 91530204 and 11622112, and the National Center for Mathematics and Interdisciplinary Sciences, CAS.

values of a certain matrix at each iteration. Another approach is the fixed-point shrinkage algorithm [30] which solves the regularized linear least problem:

$$\min_{W\in\mathbb{R}^{m\times n}} \ \mu\|W\|_* + \frac{1}{2}\|\mathcal{P}_\Omega(W-M)\|_{\mathrm{F}}^2, \tag{1.3}$$

where $\mathcal{P}_\Omega$ is the projection onto the subspace of sparse matrices with nonzeros restricted to the index subset Ω.

Extensions of this basic shrinkage scheme have been devised to solve model (1.2). Utilizing Nesterov's acceleration approach [34], an accelerated proximal gradient algorithm has been developed in [38] based on a fast iterative shrinkage-thresholding algorithm [1] for compressive sensing. The classical alternating direction augmented Lagrangian methods have been applied to solve (1.2) in [15, 41] and the closely related sparse and low-rank matrix decomposition in [43]. Other approaches include [22, 28, 32, 31, 11, 25].

A major drawback of the above methods lies at the expensive computational cost in every step. In fact, all of these algorithms bear the computational cost required by singular value decompositions (SVD) which becomes increasingly costly as the sizes of the underlying matrices increase. The most expensive computational task required by nuclear-norm minimization algorithms is the singular value decomposition (SVD) at each iteration, which becomes increasingly costly as the matrix dimensions grow. Even with the efficient partial SVD implementations which only calculate the dominated singular values and vectors, the scalability of nuclear-norm minimization algorithms are still limited by the computational complexity of SVD as matrix sizes and ranks both increases.

By introducing splitting variable, model (1.1) can be formulated into a rank-constrained model:

$$\begin{aligned} \min_{W,Z\in\mathbb{R}^{m\times n}} \quad & f(W,Z) := \|W-Z\|_{\mathrm{F}}^2, \\ \text{s.t.} \quad & W_{ij} = M_{ij},\ \forall (i,j)\in\Omega, \\ & \mathrm{rank}(Z)\le p, \end{aligned} \tag{1.4}$$

where $p < \min\{n,m\}$ is a given integer. As simple as the motivation is, a number of issues needs to be carefully addressed in order to make this approach work reliably and efficiently.

To improve the time efficiency of solving (1.2), alternative model without nuclear norm term was then proposed, *e.g.*, in [37]. The idea of these references is that the low-rank matrix Z was explicitly formulated into a factorization form rather than minimizing the nuclear norm of Z, hence avoiding SVD computation all together.

We note that the rank constraint $\mathrm{rank}(Z)\le p$ is not only nondifferentiable, but also discontinuous. To overcome this issue, we introduce the following equivalent model where the low-rank matrix Z is factorized into the product of two matrices:

$$\begin{aligned} \min_{W\in\mathbb{R}^{m\times n},X\in\mathbb{R}^{m\times p},Y\in\mathbb{R}^{n\times p}} \quad & f(W,X,Y) = \|W-XY^\top\|_{\mathrm{F}}^2, \\ \text{s.t.} \quad & W_{ij} = M_{ij},\ \forall (i,j)\in\Omega, \end{aligned} \tag{1.5}$$

The nonconvex model (1.4) has been transformed into a constrained model, then be solved by the classic ADMM in reference [37], *i.e.*, minimizing the augmented Lagrangian function w.r.t. X, Y, and W in a Gaussian-Seidel scheme and then update the multipliers. The resulting algorithm provides accelerations of computing speed up to multiple orders of magnitude on some difficult instances, compared with the algorithms based on nuclear norm minimization, such as SVT. However, a major disadvantage of the yielded algorithm lies at its limited theoretical property since there is no theory to guarantee the convergence of ADMM for nonconvex and multi-block problems.

In general, for multi-block problem, it is well known that the ADMM does not necessarily produce a convergent sequence to the optimal solution, see [10] for counterexamples. However, it has been shown that, the convergence can still be derived under some additional assumptions, *e.g.*, see [18, 17, 26]. The nonconvexity is another issue which is difficult to overcome, not only for matrix completion problem, but also for general optimization problem. However, due to the time efficiency, ADMM for nonconvex problems has attracted researchers' attention, *e.g.*, see [9]. Researches on deriving its convergence result can be found in, such as [19]. However, its theory is still limited.

To solve the nonconvex constrained problem (1.5), a natural way is alternatively minimizing variables W and variables (X, Y) in turn. In fact, this alternating minimization technique has been successfully applied to a bunch of application problems such as clustering [24], sparse PCA [45], nonnegative matrix factorization [23] etc. We note that in (1.5), the W related subproblem has closed-form solution. The (X, Y) related subproblem is actually the singular value decomposition, which can be solved by any existent eigensolver. Therefore, the efficiency of the solver for (X, Y) related subproblem can largely determine the efficiency of the whole algorithm. Recently, the authors in [27] introduced a low-rank product minimization model which is equivalent to the trace minimization model for eigenpair calculation. They also proposed a new Gauss-Newton method entitled SLRP which has good convergence result and reliable numerical performance. Hence, the above block coordinate descent scheme is practically tractable.

Based on the above idea, we propose a new algorithm for solving the matrix completion problem. The contribution of this paper is:

- Propose a new algorithm for a general form of nonconvex matrix completion problem;
- Construct global convergence results of the proposed algorithm to local optimal point with mild assumptions;
- Carry out extensive numerical experiments and confirm the time efficiency of the proposed algorithm.

The remaining part of this paper is organized as follows. In Section 2, we will present our algorithm framework and analyze its theoretical property. In Section 3, we will compare our algorithm with some state-of-the-art algorithms via extensive numerical experiments, and report the experimental results. Conclusions will be drawn in the last section.

2. New method. For a suitably chosen p, problem (1.1) can be formulated as such a more general form

$$\begin{aligned} \min_{W,Z\in\mathbb{R}^{m\times n}} \quad & \|W-Z\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & \mathcal{A}(W)=b, \\ & \mathrm{rank}(Z)\le p. \end{aligned} \tag{2.1}$$

Here, $b\in\mathbb{R}^l$, $\mathcal{A}:\mathbb{R}^{m\times n}\mapsto\mathbb{R}^l$ with $l<mn$ and let $\mathcal{A}^\top$ be its adjoint operator. Since the linear constraints can be orthogonalized, we can assume that $\mathcal{A}\mathcal{A}^\top=I$ without loss of generality. It is easy to verify that model (1.4) is a special case of (2.1), and the underlying $\mathcal{A}$ satisfies $\mathcal{A}\mathcal{A}^\top=I$. We define the matrix version of restricted isometry property (RIP) as follows (see [36])

DEFINITION 2.1. *Let $\mathcal{A}:\mathbb{R}^{m\times n}\mapsto\mathbb{R}^l$ be a linear map. For every integer p with $1\le p\le\min\{m,n\}$, define the p-restricted isometry constant to be the smallest number $\delta_p(\mathcal{A})$ such that*

$$(1-\delta_p(\mathcal{A}))\|Z\|_{\mathrm{F}}\le\|\mathcal{A}(Z)\|_{\mathrm{F}}\le(1+\delta_p(\mathcal{A}))\|Z\|_{\mathrm{F}} \tag{2.2}$$

holds for all matrices Z of rank at most p.

For convenience, we denote $\mathbf{P}_{\mathcal{A}}:=\mathcal{A}^\top(\mathcal{A}\mathcal{A}^\top)^{-1}\mathcal{A}=\mathcal{A}^\top\mathcal{A}$ and $\mathbf{P}_{\mathcal{A}}^\top=I-\mathbf{P}_{\mathcal{A}}$ as the projectors to the range space and the null space of $\mathcal{A}$, respectively. Throughout this paper, we assume $0\le\delta_p(\mathcal{A})<1$. It then directly follows

Definition 2.1 and the orthogonality of $\mathcal{A}$ that

$$(1-\delta_p(\mathcal{A}))\|Z\|_{\mathrm{F}} \leq ||\mathbf{P}_{\mathcal{A}}(Z)||_{\mathrm{F}}, \quad \text{for all } \mathrm{rank}(Z) \leq p. \tag{2.3}$$

Starting from $Z^0 \in \mathbb{R}^{m\times n}$, an alternating minimization scheme for solving problem (2.1) gives

$$W^k = \underset{W\in\mathbb{R}^{m\times n}}{\mathrm{argmin}} \ \|W - Z^{k-1}\|_{\mathrm{F}}^2, \mathrm{s.t.}\mathcal{A}(W) = b, \tag{2.4}$$

$$Z^k = \underset{Z\in\mathbb{R}^{m\times n}}{\mathrm{argmin}} \ \|W^k - Z\|_{\mathrm{F}}^2, \mathrm{s.t.rank}(Z) \leq p. \tag{2.5}$$

The optimal solution of (2.4) is

$$W^k \ = \ Z^{k-1} + \mathcal{A}^\top \left(\mathcal{A}\mathcal{A}^\top\right)^{-1}\left(b - \mathcal{A}(Z^{k-1})\right) \ = \ \mathcal{A}^\top(b) + \mathbf{P}_{\mathcal{A}}^\top(Z^{k-1}). \tag{2.6}$$

The optimal solution of (2.5) is determined by the singular value decomposition (SVD) of W^k:

$$Z^k = \sum_{i=1}^{p} \sigma_i u_i v_i^\top, \tag{2.7}$$

where $U_p = [u_1, \ldots, u_p] \in \mathbb{R}^{m\times p}$ and $V = [v_1, \ldots, v_p] \in \mathbb{R}^{n\times p}$ are orthogonal matrices, $\sigma_1, \ldots, \sigma_p$ are the p-largest singular values of W^k. In fact, the SVD is not necessary, due to the following proposition.

PROPOSITION 2.2. *Let X^k be a global solution of the following problem (entitled with SLRP [27] problem)*

$$\underset{X\in\mathbb{R}^{m\times p}}{\mathrm{argmin}} \quad \|W^k(W^k)^\top - XX^\top\|_{\mathrm{F}}^2, \tag{2.8}$$

and Y^k is defined by

$$Y^k = (W^k)^\top X^k \left((X^k)^\top X^k\right)^{-1}. \tag{2.9}$$

Then $Z^k := X^k(Y^k)^\top$ solves (2.5)*.*

Proof. Problem (2.5) can be reformulated as

$$\min_{X\in\mathbb{R}^{m\times p}, Y\in\mathbb{R}^{n\times p}} \quad \|W^k - XY^\top\|_{\mathrm{F}}^2.$$

Let (X^k, Y^k) be any global minimizer of (2). It can be easily verified that $Z^k := X^k(Y^k)^\top$ solves (2.5). On the other hand, we have

$$\|W^k - XY^\top\|_{\mathrm{F}}^2 \ = \ ||X(Y - Y(X))^\top||_{\mathrm{F}}^2 + ||(I - \mathbf{P}_{\mathbf{X}})W||_{\mathrm{F}}^2, \tag{2.10}$$

where $Y(X) := (W^k)^\top X \left(X^\top X\right)^{-1}$ and the projector to the range space of X is denoted by $\mathbf{P}_{\mathbf{X}} := X(X^\top X)^{-1}X^\top$. It directly follows from (2.10) that $Y^k = Y(X^k)$ and X^k should be a global minimizer of

$$\min_{X\in\mathbb{R}^{m\times p}} ||(I - \mathbf{P}_{\mathbf{X}})W||_{\mathrm{F}}^2. \tag{2.11}$$

It can be verified that X^* is a global solution of (2.11) if and only if the columns of X^* spans the eigenvectors associated with the p largest eigenvalues of $WW^\top$. Namely, any global minimizer of (2.8) belongs to the solution set

of (2.11). This completes the proof. □

In summary, the alternating minimization approach can be described as follows

Algorithm 1: Alternating Minimization for Low Rank Decomposition

1 Input $b \in \mathbb{R}^l$, $\mathcal{A} : \mathbb{R}^{m\times n} \mapsto \mathbb{R}^l$ satisfying $\mathcal{A}\mathcal{A}^\top = I$ where $l < mn$.
2 Initialize an $m \times n$ matrix $Z^0 := 0$ and set $k := 1$.
3 **while** *not "converged"* **do**
4 Compute $W^k = \mathcal{A}^\top(b) + \mathbf{P}_{\mathcal{A}}^\top(Z^{k-1})$.
5 Solve SLRP problem (2.8):
6 i) to stationary point $X^k = \mathrm{SLRP}_{\mathrm{ST}}(W^k(W^k)^\top)$;
7 ii) to global minimizer $X^k = \mathrm{SLRP}_{\mathrm{GM}}(W^k(W^k)^\top)$.
8 Calculate Y^k by (2.9).
9 Update $Z^k = X^k(Y^k)^\top = \mathbf{P}_{\mathbf{X}}W^k$.
10 Increment k and continue.

In order to accelerate the convergence of the alternating minimization, one can apply an extension to the variable Z by replacing the step 9 of Algorithm 1 with

$$Z^k = (1-\alpha)W^k + \alpha\mathbf{P}_{\mathbf{X}}W^k. \tag{2.12}$$

In fact, this idea is similar to that of SOR approach. For convenience, the theoretical analysis will be only based on the case that $\alpha = 1$. The cases of $\alpha \in (0, 1)$ can be proved along the similar line, though more notationally involved and tedious.

Before analyzing the convergence properties of Algorithm 1, we give the optimality condition for the following equivalent smoothing model of (2.1)

$$\min_{\substack{W\in\mathbb{R}^{m\times n}\\ X\in\mathbb{R}^{m\times p}, Y\in\mathbb{R}^{n\times p}}} \quad f(W,X,Y) := \|W - XY^\top\|_{\mathrm{F}}^2 \tag{2.13}$$

$$\text{s.t.} \qquad \mathcal{A}(W) = b. \tag{2.14}$$

PROPOSITION 2.3. *Let $(\hat{W}, \hat{X}, \hat{Y})$ be a stationary point of* (2.13)-(2.14)*, then the following first-order optimality conditions hold*

$$\begin{array}{rcl} \hat{W} - \hat{X}\hat{Y}^\top & = & \mathcal{A}^\top(\lambda); \\ \hat{X}^\top(\hat{W} - \hat{X}\hat{Y}^\top) & = & 0; \\ (\hat{W} - \hat{X}\hat{Y}^\top)\hat{Y} & = & 0; \\ \mathcal{A}(\hat{W}) & = & b, \end{array} \tag{2.15}$$

where $\lambda \in \mathbb{R}^l$ is the Lagrangian multiplier of the equality constraint (2.14)*.*

LEMMA 2.4. *Suppose $\delta_p(\mathcal{A}) \in [0, 1)$. Let $\{(W^k, X^k, Y^k)\}$ be the iterative sequence generated by Algorithm 1 with SLRP subproblem* (2.8) *being solved to stationarity, then $\{(W^k, X^k, Y^k)\}$ is bounded and consequently has accumulation point.*

Proof. It is clear that the iterates generated by Algorithm 1 have monotonically non-increasing function values.

Namely, $f(W^k, Z^k) \leq f(W^{k-1}, Z^{k-1})$ for any $k = 1, 2, ...$ Hence we have

$$f(W^{k+1}, Z^k) \ \leq \ f(W^k, Z^k) \leq \cdots \ \leq \ f(W^1, Z^1) \leq f(W^1, Z^0). \tag{2.16}$$

According to step 4 in Algorithm 1, we have

$$\begin{aligned} f(W^{k+1}, Z^k) = ||W^{k+1} - Z^k||_{\mathrm{F}}^2 \ &= \ ||\mathcal{A}^\top(b) + \mathbf{P}_{\mathcal{A}}^\top(Z^k) - Z^k||_{\mathrm{F}}^2 \\ &= ||\mathcal{A}^\top(b) - \mathbf{P}_{\mathcal{A}}(Z^k)||_{\mathrm{F}}^2 \ \geq \ (||\mathbf{P}_{\mathcal{A}}(Z^k)||_{\mathrm{F}} - ||\mathcal{A}^\top(b)||_{\mathrm{F}})^2. \end{aligned}$$

Combining (2.16) with (2.17) we obtain

$$(||\mathbf{P}_{\mathcal{A}}(Z^k)||_{\mathrm{F}} - ||\mathcal{A}^\top(b)||_{\mathrm{F}})^2 \leq f(W^{k+1}, Z^k) \ \leq \ f(W^1, Z^0) = ||\mathcal{A}^\top(b)||_{\mathrm{F}}^2,$$

which implies

$$||\mathbf{P}_{\mathcal{A}}(Z^k)||_{\mathrm{F}} \leq 2||\mathcal{A}^\top(b)||_{\mathrm{F}}.$$

Then it follows from the RIP assumption that

$$||Z^k||_{\mathrm{F}} \leq \frac{1}{1 - \delta_p(\mathcal{A})}||\mathbf{P}_{\mathcal{A}}(Z^k)||_{\mathrm{F}} \leq \frac{2}{1 - \delta_p(\mathcal{A})}||\mathcal{A}^\top(b)||_{\mathrm{F}}. \tag{2.17}$$

Moreover,

$$\begin{aligned} ||W^k||_{\mathrm{F}}^2 = ||\mathcal{A}^\top(b) + \mathbf{P}_{\mathcal{A}}^\top(Z^k)||_{\mathrm{F}}^2 \ &\leq \ 2(||\mathcal{A}^\top(b)||_{\mathrm{F}}^2 + ||\mathbf{P}_{\mathcal{A}}^\top(Z^k)||_{\mathrm{F}}^2) \\ \leq 2(||\mathcal{A}^\top(b)||_{\mathrm{F}}^2 + \delta_p^2(\mathcal{A})||Z^k||_{\mathrm{F}}^2) \ &\leq \ \frac{2(1 - 2\delta_p(\mathcal{A}) + 5\delta_p^2(\mathcal{A}))}{(1 - \delta_p(\mathcal{A}))^2}||\mathcal{A}^\top(b)||_{\mathrm{F}}^2. \end{aligned}$$

Here the second last inequality holds due to the orthogonality of $\mathcal{A}$ and the RIP assumption. The last inequality is due to relationship (2.17). Hence, we complete the proof. □

LEMMA 2.5. *Let* $\{(W^k, X^k, Y^k)\}$ *be the iterative sequence generated by Algorithm 1 with SLRP subproblem* (2.8) *being solved to stationarity, and suppose* $(\hat{W}, \hat{X}, \hat{Y})$ *is an accumulation point of* $\{(W^k, X^k, Y^k)\}$. *Then* $(\hat{W}, \hat{X}, \hat{Y})$ *satisfies the first-order optimality conditions* (2.15).

Proof. First, we note that $f(W^k, X^k, Y^k)$ is monotonically decreasing and bounded below. Therefore, there exists f^* so that

$$\lim_{k \to \infty} f(W^k, X^k, Y^k) = f^*. \tag{2.18}$$

Without loss of generality, we assume that $(W^{k_j}, X^{k_j}, Y^{k_j})$ converges to $(\hat{W}, \hat{X}, \hat{Y})$ with $j \to \infty$. According to step 4 of Algorithm 1, we have

$$W^{k_j+1} \ = \ \mathcal{A}^\top(b) + \mathbf{P}_{\mathcal{A}}^\top(X^{k_j}(Y^{k_j})^\top) \ = \ \mathbf{P}_{\mathcal{A}}W^{k_j} + (I - \mathbf{P}_{\mathcal{A}})(X^{k_j}(Y^{k_j})^\top).$$

Namely, we have

$$W^{k_j+1} - X^{k_j}(Y^{k_j})^\top = \mathbf{P}_{\mathcal{A}}(W^{k_j} - X^{k_j}(Y^{k_j})^\top). \tag{2.19}$$

On the other hand, it holds that

$$(W^{k_j} - X^{k_j}(Y^{k_j})^\top) - (W^{k_j+1} - X^{k_j}(Y^{k_j})^\top) = -(I - \mathbf{P}_{\mathcal{A}})(W^{k_j} - X^{k_j}(Y^{k_j})^\top).$$

Together with (2.19), we obtain

$$\begin{aligned} & f(W^{k_j}, X^{k_j}, Y^{k_j}) - f(W^{k_j+1}, X^{k_j}, Y^{k_j}) \\ &= ||W^{k_j} - X^{k_j}(Y^{k_j})^\top||_{\mathrm{F}}^2 - ||W^{k_j+1} - X^{k_j}(Y^{k_j})^\top||_{\mathrm{F}}^2 \\ &= ||(I - \mathbf{P}_{\mathcal{A}})(W^{k_j} - X^{k_j}(Y^{k_j})^\top||_{\mathrm{F}}^2. \end{aligned} \tag{2.20}$$

Take the limit $j \to \infty$ of both sides of (2.20), we obtain

$$0 = ||(I - \mathbf{P}_{\mathcal{A}})(\hat{W} - \hat{X}\hat{Y}^\top)||_{\mathrm{F}}^2.$$

Namely,

$$\hat{W} = \mathbf{P}_{\mathcal{A}}(\hat{W}) + \mathbf{P}_{\mathcal{A}}^\top(\hat{X}\hat{Y}^\top). \tag{2.21}$$

By assuming $\lambda = \mathcal{A}(\hat{W} - \hat{X}\hat{Y}^\top)$, we obtain the first condition of (2.15). The fourth condition of (2.15) directly holds after taking $\mathcal{A}$ from both sides of (2.21).

On the other hand, pre-multiplying $\left((X^{k_j})^\top X^{k_j}\right)$ to the both sides of the Y updating formula (2.9), we obtain

$$\left((X^{k_j})^\top X^{k_j}\right)(Y^{k_j})^\top = (X^{k_j})^\top W^{k_j}, \tag{2.22}$$

which gives the second condition of (2.21) after taking the limit from both sides.

Finally, since X^{k_j} is a stationary point of the SLRP problem (2.8), we have

$$W^{k_j}(W^{k_j})^\top X^{k_j} = X^{k_j}((X^{k_j})^\top X^{k_j}), \tag{2.23}$$

which implies

$$((X^{k_j})^\top X^{k_j})^{-1}(X^{k_j})^\top W^{k_j}(W^{k_j})^\top X^{k_j} = (X^{k_j})^\top X^{k_j}. \tag{2.24}$$

Substituting relationship (2.24) into (2.23), we obtain

$$W^{k_j}Y^{k_j} = X^{k_j}((Y^{k_j})^\top Y^{k_j}), \tag{2.25}$$

which finally implies the third condition of (2.15) after taking the limit from the both sides. Now, we recall Proposition 2.3 and complete the proof. □

Next, we consider to obtain stronger convergence results by assuming that we solve the SLRP subproblem (2.8) in Algorithm 1 to global optimality.

LEMMA 2.6. *Suppose $\sigma_p(W)$ and $\sigma_{p+1}(W)$ are the p-th and $(p+1)$-st largest singular values of matrix $W \in \mathbb{R}^{m\times n}$, respectively. Let (X, Y) be a global minimizer of unconstrained problem* (2.13)*. Then, we have*

$$||W - (X + \Delta_X)(Y + \Delta_Y)^\top||_{\mathrm{F}}^2 - ||W - XY^\top||_{\mathrm{F}}^2$$

$$\geq (1-\sigma_{p+1}(W)/\sigma_p(W))\cdot||\Delta_Z||_{\rm F}^2+o(||\Delta_Z||_{\rm F}^2), \tag{2.26}$$

holds for any $\Delta_X\in\mathbb{R}^{m\times p}$ *and* $\Delta_Y\in\mathbb{R}^{n\times p}$. *Here* $\Delta_Z:=\Delta_XY^\top+X\Delta_Y^\top+\Delta_X\Delta_Y^\top$.

Proof. Firstly, we have

$$||W-(X+\Delta_X)(Y+\Delta_Y)^\top||_{\rm F}^2-||W-XY^\top||_{\rm F}^2\ \geq\ ||\Delta_Z||_{\rm F}^2-2\langle W-XY^\top,\Delta_Z\rangle. \tag{2.27}$$

On the other hand, let columns of $\tilde{X}$ and $\tilde{Y}$ are the left and right singular vectors, associated with $\sigma_1(W),...,\sigma_p(W)$, of W. According to the global optimality of (X,Y), we obtain $X=\tilde{X}Q_X$ and $Y=\tilde{Y}Q_Y$ and $Q_XQ_Y^\top=\Sigma_p$, where Σ_p is a $p\times p$ diagonal matrix with $\sigma_1(W),...,\sigma_p(W)$ in its diagonal. Denoting $\tilde{\Delta}_X=\Delta_XQ_Y^\top\Sigma_p^{-1/2}$, $\tilde{\Delta}_Y=\Delta_YQ_X^\top\Sigma_p^{-1/2}$, we have

$$\begin{aligned}||\Delta_Z||_{\rm F}^2&=||X\Delta_Y^\top+\Delta_X^\top Y^\top||_{\rm F}^2+o(||\Delta_Z||_{\rm F}^2)\\&=||\tilde{X}\Sigma_p^{1/2}\tilde{\Delta}_Y^\top||_{\rm F}^2+||\tilde{\Delta}_X^\top\Sigma_p^{1/2}\tilde{Y}^\top||_{\rm F}^2+o(||\Delta_Z||_{\rm F}^2)\\&\geq\sigma_p(W)(||\tilde{\Delta}_Y||_{\rm F}^2+||\tilde{\Delta}_X||_{\rm F}^2)+o(||\Delta_Z||_{\rm F}^2).\end{aligned} \tag{2.28}$$

On the other hand, by using the first-order optimality, we have

$$\begin{aligned}&\langle W-XY^\top,\Delta_Z\rangle=\langle W-XY^\top,\Delta_X\Delta_Y^\top\rangle\\&=\langle W-XY^\top,\tilde{\Delta}_X\tilde{\Delta}_Y^\top\rangle\ =\ \langle W-XY^\top,\mathbf{P}_{\mathbf{X}}^\top(\tilde{\Delta}_X)\mathbf{P}_{\mathbf{Y}}^\top(\tilde{\Delta}_Y)^\top\rangle\\&=\langle W,\mathbf{P}_{\mathbf{X}}^\top(\tilde{\Delta}_X)\mathbf{P}_{\mathbf{Y}}^\top(\tilde{\Delta}_Y)^\top\rangle\ \leq\ \sigma_{p+1}(W)||\mathbf{P}_{\mathbf{X}}^\top(\tilde{\Delta}_X)||_{\rm F}||\mathbf{P}_{\mathbf{Y}}^\top(\tilde{\Delta}_Y)||_{\rm F}\\&\leq\sigma_{p+1}(W)||\tilde{\Delta}_X||_{\rm F}||\tilde{\Delta}_Y||_{\rm F}\ \leq\ \sigma_{p+1}(W)(||\tilde{\Delta}_X||_{\rm F}^2+||\tilde{\Delta}_Y||_{\rm F}^2)/2.\end{aligned} \tag{2.29}$$

Substituting (2.28) and (2.29) into (2.27), we complete the proof. □

LEMMA 2.7. *Let* $\{(W^k,X^k,Y^k)\}$ *be the iterative sequence generated by Algorithm 1 with SLRP subproblem* (2.8) *being solved to global optimality, and suppose* $(\hat{W},\hat{X},\hat{Y})$ *is an accumulation point of* $\{(W^k,X^k,Y^k)\}$. *Suppose* $\delta_{2p}(\mathcal{A})\in[0,1)$ *and*

$$1-\sigma_{p+1}(\hat{W})/\sigma_p(\hat{W})>\delta_{2p}(\mathcal{A}). \tag{2.30}$$

Then $(\hat{W},\hat{X},\hat{Y})$ *is a local optimizer of the decomposition model* (2.13)-(2.14).

Proof. According to Lemma 2.5, $(\hat{W},\hat{X},\hat{Y})$ shall satisfy the first-order optimality conditions (2.15). Moreover, it is not difficult to verify that $(\hat{X},\hat{Y})$ is the global solution of (2.13) with $\hat{W}$.

Suppose that $(\hat{W}+\Delta_W,\hat{X}+\Delta_X,\hat{Y}+\Delta_Y)$ is in a neighborhood of $(\hat{W},\hat{X},\hat{Y})$. Let

$$\tilde{\Delta}_W=-\hat{W}+\mathcal{A}^\top b+\mathbf{P}_{\mathcal{A}}^\top((\hat{X}+\Delta_X)(\hat{Y}+\Delta_Y)^\top).$$

Then, we have

$$\tilde{\Delta}_W=-\mathbf{P}_{\mathcal{A}}^\top(\hat{W}-(\hat{X}+\Delta_X)(\hat{Y}+\Delta_Y)^\top)\ =\ \mathbf{P}_{\mathcal{A}}^\top(\Delta_Z). \tag{2.31}$$

It is clear that

$$f(\hat{W}+\Delta_W,\hat{X}+\Delta_X,\hat{Y}+\Delta_Y)-f(\hat{W},\hat{X},\hat{Y})$$

$$
\begin{aligned}
&\geq f(\hat{W}+\tilde{\Delta}_W, \hat{X}+\Delta_X, \hat{Y}+\Delta_Y) - f(\hat{W},\hat{X},\hat{Y}) \\
&= ||\tilde{\Delta}_W - \Delta_Z||_{\mathrm{F}}^2 + 2\langle W - XY^\top, \tilde{\Delta}_W - \Delta_Z\rangle \\
&= ||\Delta_Z||_{\mathrm{F}}^2 - 2\langle W - XY^\top, \mathbf{P}_{\mathcal{A}}(\Delta_Z)\rangle - ||\mathbf{P}_{\mathcal{A}}^\top(\Delta_Z)||_{\mathrm{F}}^2.
\end{aligned}
$$

The last equality uses the relationship (2.31). By using the stationarity of $(\hat{W},\hat{X},\hat{Y})$ and Lemma 2.6, we further have

$$
\begin{aligned}
& f(\hat{W}+\Delta_W, \hat{X}+\Delta_X, \hat{Y}+\Delta_Y) - f(\hat{W},\hat{X},\hat{Y}) \\
&\geq ||\Delta_Z||_{\mathrm{F}}^2 - 2\langle W - XY^\top, \mathbf{P}_{\mathcal{A}}(\Delta_Z)\rangle - ||\mathbf{P}_{\mathcal{A}}^\top(\Delta_Z)||_{\mathrm{F}}^2 \\
&= ||\Delta_Z||_{\mathrm{F}}^2 - 2\langle W - XY^\top, \Delta_Z\rangle - ||\mathbf{P}_{\mathcal{A}}^\top(\Delta_Z)||_{\mathrm{F}}^2 \\
&\geq (\sigma_p(\hat{W}) - \sigma_{p+1}(\hat{W}))\cdot ||\Delta_Z||_{\mathrm{F}}^2 - ||\mathbf{P}_{\mathcal{A}}^\top(\Delta_Z)||_{\mathrm{F}}^2 + o(||\Delta_Z||_{\mathrm{F}}^2) \\
&\geq (\sigma_p(\hat{W}) - \sigma_{p+1}(\hat{W}))\cdot ||\Delta_Z||_{\mathrm{F}}^2 - \delta_{2p}||\Delta_Z||_{\mathrm{F}}^2 + o(||\Delta_Z||_{\mathrm{F}}^2) > 0.
\end{aligned}
$$

Here the second last inequality use the fact that $\mathrm{rank}(\Delta_Z) \leq 2p$ and the RIP assumption. We complete the proof. □

3. Numerical Experiments. To investigate the performance of our proposed algorithm, in this section, we do some numerical experiments on matrix completion problem (1.1) with synthetic data, though our algorithm was designed for more general model (2.1).

3.1. Experimental Setup. We first describe the setup of our numerical experiments, including the generation of test problems, the solvers used for comparison, and how to compare the test algorithms.

We evaluate the proposed algorithm by comparing its performance with some most efficient algorithms. All codes used in our experiments were written in MATLAB (though with some tasks done in C or Fortran), and the experiments were performed in MATLAB version R2015b on a PC with Intel Core i7 1.73Ghz and 6GB memory. For solving matrix completion problem, there have been a bunch of efficient algorithms/solvers such as SVT[1], FPCA[2], Optspace[3], LMaFit[4], and etc. These algorithms can be categorized into two groups: SVD-based algorithms and SVD-free algorithms. The SVD-based algorithms such as SVT are able to produce a satisfactory result, but it is relatively slow in terms of running time due to the presence of expensive SVD computation. In contrary, the SVD-free algorithm such as LMaFit can obtain competitive result with much less computing time, however, it can only guarantee the convergence to a stationary point. For simplicity, we only include SVT and LMaFit in our experiments. Note that although our proposed algorithm is designed for solving a more general form of MC problem (2.1), but we still only consider the standard MC problem (1.4) since SVT and LMaFit can only handle (1.4). The parameter settings of SVT and LMaFit are left to their default values. In our algorithm, the setting of extension parameter α can be critical to the performance of our algorithm according to our limited numerical experience, we manually tuned it to optimize the performance of our algorithm. With some certain problem settings, we may violate the constraint $\alpha \in (0, 1]$ (i.e., $\alpha > 1$) to achieve better speed performance although the convergence may not be guaranteed. For LMaFit and our algorithm which need to input the correct rank before implementation, we always input the correct rank for these two algorithms.

The original low-rank matrix $L^* \in R^{m\times n}$ with rank k was constructed by (see also [40]): let $L^* = U * V^T$ where $U \in R^{m\times k}$ and $V \in R^{n\times k}$ were generated randomly with i.i.d. standard Gaussian entries and then to build a rank-k matrix. The indices subset Ω of p entries were sampled uniformly at random by the Matlab notation "temp =

[1] http://svt.stanford.edu/
[2] http://www.columbia.edu/~sm2756/FPCA.htm
[3] http://www.stanford.edu/~raghuram/optspace/papers.html
[4] http://lmafit.blogs.rice.edu/

randperm(m*n); Omega = temp(1:p);", while the ratio $p/(m*n) \in (0,1)$ between the number of measurements p and the number of total entries is entitled "sampling ratio".

We initialize our algorithm with Matlab notation "X = eye(m,k); Z = zeros(m,n);", and terminate it when either a maximal number of iterations $maxit$ is reached, or the relative error of L between the recovered matrix and the exact one on Ω is smaller than some prescribed tolerance tol, *i.e.*,

$$\text{err}_\Omega(L) := \frac{\|\mathcal{P}_\Omega(L - L^*)\|_{\text{F}}^2}{\|\mathcal{P}_\Omega(L^*)\|_{\text{F}}^2} < tol. \tag{3.1}$$

To unify the stopping rules of the test algorithms, the codes of SVT and LMaFit are modified by the present authors such that they exactly follow the stopping rule of our algorithm. The performance indices of the involved algorithms include three quantities: the number of iterations denoted by "iter", the CPU time (in seconds) denoted by "time", and the relative error on Ω denoted by "err$_\Omega(L)$", Unless otherwise specified, in the whole experimental section, we always set $maxit = 200$ and $tol = 1e-13$, and the results of "iter" and "time" shown are always the average results on 10 randomly generated problems.

3.2. Experiments on synthetic data sets. We first investigate the speed performance of test algorithms under several different settings of rank(L^*) and sampling ratio. The results are reported in Tables 3.1.

Table 3.1 Results of speed performance test for synthetic data with m=n=2000

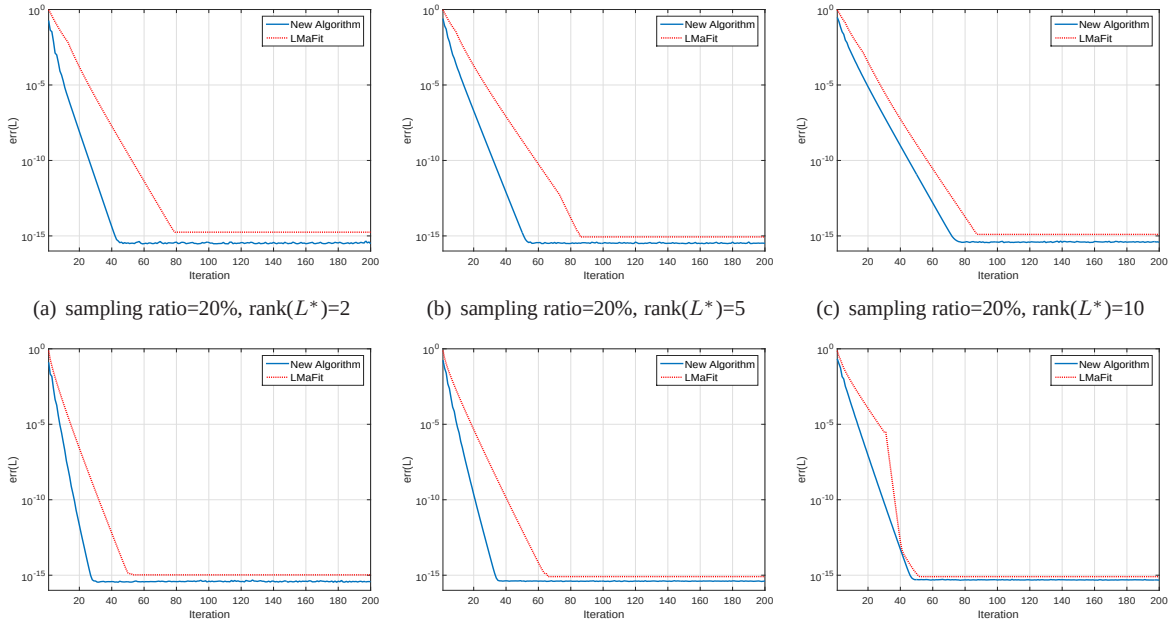
Problem settings		SVT			LMaFit			New Algorithm		
sampling ratio	rank(L^*)	err$_\Omega(L)$	iter	time	err$_\Omega(L)$	iter	time	err$_\Omega(L)$	iter	time
20%	5	8.947e-13	126.0	10.288	8.396e-13	67.7	1.746	8.880e-13	42.7	1.218
	10	1.183e-7	190.2	22.593	9.005e-13	66.8	1.892	8.525e-13	57.3	1.821
40%	5	8.494e-13	87.0	12.286	7.022e-13	33.3	1.376	6.705e-13	24.2	1.100
	10	8.702e-13	101.7	19.319	7.169e-13	39.2	1.722	7.001e-13	29.9	1.468
60%	5	8.470e-13	72.3	14.429	6.895e-13	33.2	1.821	5.106e-13	18.0	1.096
	10	8.465e-13	80.4	18.717	7.573e-13	39.1	2.237	5.467e-13	21.0	1.334
80%	5	8.765e-13	62.3	15.112	6.556e-13	23.1	1.547	3.881e-13	15.0	1.100
	10	8.180e-13	67.5	19.641	6.421e-13	25.1	1.801	4.308e-13	16.7	1.322

It is easy to observe from Table 3.1 that all competing algorithms can obtain prescribed tolerance under all settings (only except SVT when sampling ratio=20%, rank(L^*)=10). We observe that our algorithm is faster than LMaFit and SVT in terms of computing time for synthetic problem. In terms of computing time, when sampling ratio=20%, our algorithm is around 5%-40% faster than LMaFit, and 7-11 times faster than SVT; when sampling ratio=40% or 80%, our algorithm is around 20% faster than LMaFit, and 10-12 times faster than SVT; when sampling ratio=60%, our algorithm is around 70% faster than LMaFit, and 12-13 times faster than SVT. As SVT is much slower than our algorithm and LMaFit due to the SVD computation, it will not be included in the rest part of the experiments.

To examine the convergence behavior of the test algorithms more precisely, in terms of iteration numbers rather than of running time, we ran the test algorithms on 4 random problems of sizes m=n=2000 for fixed 200 iterations and recorded the progress of relative errors at all iterations. The random problems were generated with 12 different problem settings of sampling ratio and rank(L^*). The results on these problems are given in Figure 1, where the relative errors are plotted against the iteration numbers.

The results shown in Figure 1 coincide with the results shown in Table 3.1. Concretely, we see that our algorithm usually converges faster than LMaFit, and the advantage can be more obvious when the sampling ratio is higher. In addition, our algorithm can always obtain a result with better accuracy (around half order of magnitude lower in terms of relative error) than that obtained by LMaFit.

We have observed that the sampling ratio has strong influence on the performance of the test algorithms. To investigate the influence of sampling ratio on the speed performance, with fixed m=n=2000 and four different settings of rank(L^*), we let the sampling ratio vary between 10% and 95% with interval 5%. The results of computing time



(a) sampling ratio=20%, rank(L^*)=2 (b) sampling ratio=20%, rank(L^*)=5 (c) sampling ratio=20%, rank(L^*)=10

(d) sampling ratio=60%, rank(L^*)=10 (e) sampling ratio=60%, rank(L^*)=20 (f) sampling ratio=60%, rank(L^*)=40

FIG. 1. *Relative error vs iteration number,* m=n=*2000*

v.s. sampling ratios are plotted in Figure 2.

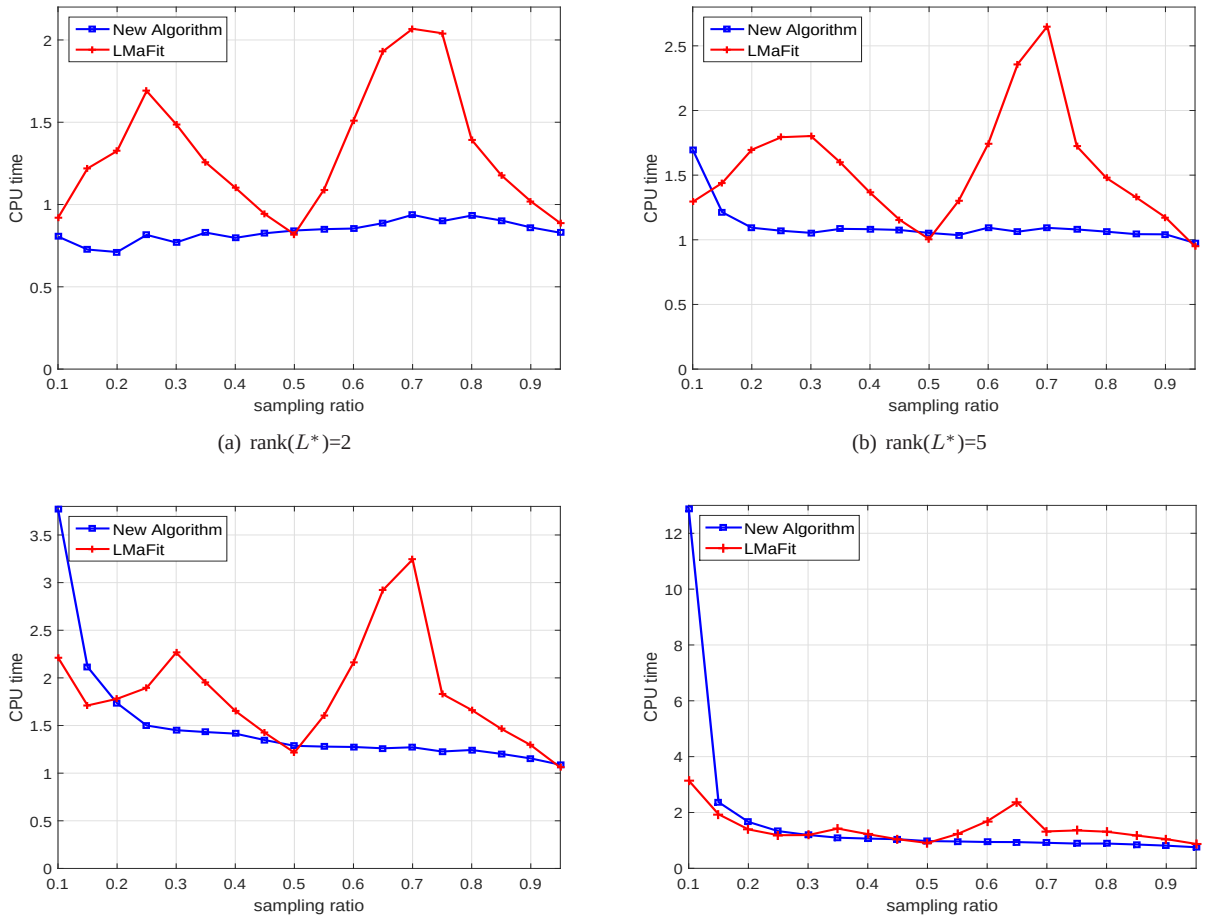
We see from Figure 2 that our algorithm usually takes less computing time, and this is especially true when rank(L^*) is small. In addition, very small sampling ratio (such as 10%) seems to be less preferable for our algorithm, *e.g.*, the last subfigure shows that our algorithm costs around 3 times more of CPU time than LMaFit to obtain the prescribed tolerance when the sampling ratio is 0.1. It is also interesting to observe that the speed performance of LMaFit is vulnerable to the change of sampling ratio, *e.g.*, its computing time is doubled when the sampling ratio increases from 50% to 70%. In contrary, our algorithm performs more stably as the sampling ratio changes.

To examine the numerical behaviour of the test algorithms as dimension increases, with three different settings of sampling ratio and two different settings of rank(L^*), we let the dimension (m, n) increase from 200 to 4000 with interval 200. The results of computing time v.s. dimension are plotted in Figure 3.

We observe from Figure 3 that when the sampling ratio is 30% or 70%, our algorithm takes less computing time, and the advantage in computing time is retained as the dimension increases. We also observe that when the sampling ratio is 50%, LMaFit is always better than our algorithm, and the performance gap is almost unchanged as the dimension increases. This result reflects the fact that, with fixed sampling ratio, the performance gap between our algorithm and LMaFit is unchanged as dimension increases, i.e., the speed performance of our algorithm has satisfactory dimensional scalability.

Via preliminary observation from the above experimental results, we conclude that for solving matrix completion problem with synthetic data, the new algorithm demonstrates the following three advantages:

- better time efficiency: it outperforms not only the SVD-based algorithms (such as SVT), but also the SVD-free algorithm (such as LMaFit), hence it is practically efficient.
- more stable performance: the performance of our algorithm is less vulnerable to the change of problem setting.
- satisfactory dimensional scalability of speed performance: the performance gap between our algorithm and LMaFit is almost unchanged as the dimension increases.



(a) rank(L^*)=2 (b) rank(L^*)=5

(c) rank(L^*)=10 (d) rank(L^*)=20

FIG. 2. *Computing time vs sampling ratio, m=n=2000.*

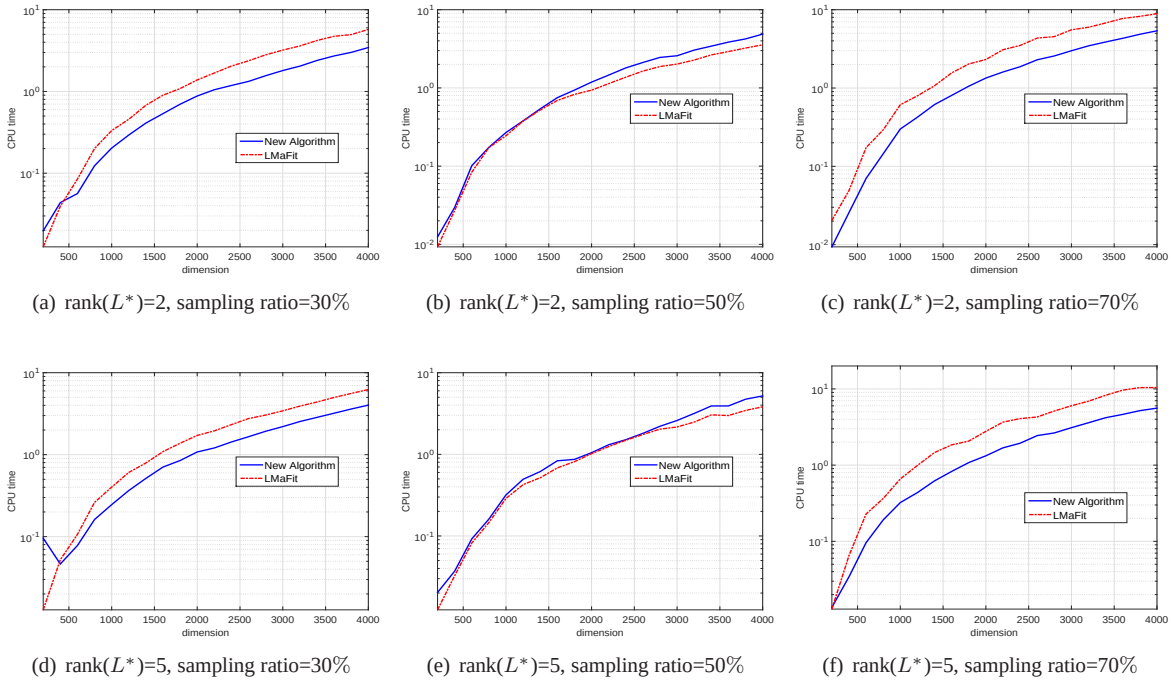
4. Conclusion. In this paper, we propose a new algorithm for solving a more general form of the nonconvex matrix completion problem. We also construct global convergence results of the proposed algorithm to stationary point. Additionally, with mild assumptions, we can guarantee the proposed algorithm converge to a local optimal point while some competing algorithms can only converge to a stationary point.

We performed extensive computational experiments on matrix completion problem. Preliminary experimental results have confirmed the motivating premise for our approach that avoiding SVD-related calculations can lead to a very efficient algorithm for solving matrix completion problem. Its advantage is threefold, on one hand, it is much more efficient than those SVD-based algorithms such as SVT, and posses competitive performance against LMaFit which is also SVD free; on the other hand, it has better performance stability as the problem setting changes; additionally, it is almost parameter free, which makes it very easy to implement.

In all, the new algorithm is a simple and efficient algorithm. It could be improved by any existing technique, such as Nesterov acceleration technique. However, these are left to the future works.

Acknowledgment. The authors would like to thank Dr. Caihua Chen from Nanjing University who has provided me with valuable comments in the writing of this paper.

REFERENCES



(a) rank(L^*)=2, sampling ratio=30% (b) rank(L^*)=2, sampling ratio=50% (c) rank(L^*)=2, sampling ratio=70%

(d) rank(L^*)=5, sampling ratio=30% (e) rank(L^*)=5, sampling ratio=50% (f) rank(L^*)=5, sampling ratio=70%

FIG. 3. *Computing time vs dimension.*

[1] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[2] J. CAI, E. J. CANDÈS, AND Z. SHEN, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim., 20 (2010), pp. 1956–1982.

[3] E. J. CANDÈS AND Y. PLAN, *Matrix completion with noise*, Proceedings of the IEEE, 98 (2010), pp. 925 –936.

[4] E. J. CANDÈS AND B. RECHT, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.

[5] E. J. CANDÈS AND J. ROMBERG, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. Comput. Math., 6 (2006), pp. 227–254.

[6] E. J. CANDÈS AND T. ROMBERG, J.AND TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[7] E. J. CANDÈS AND T. TAO, *Near optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.

[8] E. J. CANDÈS AND T. TAO, *The power of convex relaxation: near-optimal matrix completion*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2053–2080.

[9] R. CHARTRAND, *Nonconvex splitting for regularized low-rank + sparse decomposition*, 60 (2012), pp. 5810–5819.

[10] C. CHEN, B. HE, Y. YE, AND X. YUAN, *The direct extension of admm for multi-block convex minimization problems is not necessarily convergent*, Math. Program., 155 (2016), pp. 57–79.

[11] W. DAI AND O. MILENKOVIC, *Set: an algorithm for consistent matrix completion*, CoRR, abs/0909.2705 (2009).

[12] D. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[13] M. ELAD, *Why simple shrinkage is still relevant for redundant representations?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5559–5569.

[14] L. ELDÉN, *Matrix Methods in Data Mining and Pattern Recognition (Fundamentals of Algorithms)*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2007.

[15] D. GOLDFARB, S. MA, AND Z. WEN, *Solving low-rank matrix completion problems efficiently*, in Proceedings of the 47th annual Allerton conference on Communication, control, and computing, Allerton'09, 2009, pp. 1013–1020.

[16] D. GROSS, *Recovering low-rank matrices from few coefficients in any basis*, tech. report, Leibniz University, 2009.

[17] D. HAN AND X. YUAN, *A note on the alternating direction method of multipliers*, J. Optim. Theory Appl., 155 (2012), pp. 227–238.

[18] M. HONG AND Z.-Q. LUO, *On the linear convergence of the alternating direction method of multipliers*. preprint, `http://arxiv.org/abs/1208.3922`, 2012.

[19] M. HONG, Z.-Q. LUO, AND M. RAZAVIYAYN, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*. preprint, `http://arxiv.org/abs/1410.1390`, 2014.

[20] R. H. KESHAVAN, A. MONTANARI, AND S. OH, *Matrix completion from a few entries*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2980–

2998.

[21] ———, *Matrix completion from noisy entries*, J. Mach. Learn. Res., 99 (2010), pp. 2057–2078.

[22] R. H. KESHAVAN AND S. OH, *A gradient descent algorithm on the grassman manifold for matrix completion*, tech. report, Dept. of Electrical Engineering, Stanford University, 2009.

[23] H. KIM AND H. PARK, *Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 713–730.

[24] J. KIM AND H. PARK, *Sparse nonnegative matrix factorization for clustering*, tech. report, Georgia Institute of Technology, 2008. Technical Report GT-CSE-08-01.

[25] K. LEE AND Y. BRESLER, *Admira: atomic decomposition for minimum rank approximation*, IEEE Trans. Inf. Theor., 56 (2010), pp. 4402–4416.

[26] M. LI, D. SUN, AND K.-C. TOH, *A convergent 3-block semi-proximal admm for convex minimization problems with one strongly convex block*, Asia Pac. J. Oper. Res., (2015).

[27] X. LIU, Z. WEN, AND Y. ZHANG, *An efficient gauss–newton algorithm for symmetric low-rank product matrix approximations*, SIAM J. Optim., 25 (2015), pp. 1571–1608.

[28] Y. LIU, D. SUN, AND K.-C. TOH, *An implementable proximal point algorithmic framework for nuclear norm minimization*, Math. Program., (2011), pp. 1–38.

[29] Z. LIU AND L. VANDENBERGHE, *Interior-point method for nuclear norm approximation with application to system identification*, SIAM J. Matrix Anal. A., 31 (2009), pp. 1235–1256.

[30] S. MA, D. GOLDFARB, AND L. CHEN, *Fixed point and bregman iterative methods for matrix rank minimization*, Math. Program., 128 (2009), pp. 321–353.

[31] R. MAZUMDER, T. HASTIE, AND R. TIBSHIRANI, *Regularization methods for learning incomplete matrices*, tech. report, Stanford University, 2009.

[32] R. MEKA, P. JAIN, AND I. S. DHILLON, *Guaranteed rank minimization via singular value projection*, CoRR, abs/0909.5457 (2009).

[33] T. MORITA AND T. KANADE, *A sequential factorization method for recovering shape and motion from image streams*, IEEE Trans. Pattern Anal. Mach. Intell., 19 (1997), pp. 858–867.

[34] Y. NESTEROV, *Introductory lectures on convex optimization. 87: xviii+236, 2004. A basic course.*

[35] B. RECHT, *A simpler approach to matrix completion*, J. Mach. Learn. Res., 12 (2011), pp. 3413–3430.

[36] B. RECHT, M. FAZEL, AND P. A. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.

[37] Y. SHEN, Z. WEN, AND Y. ZHANG, *Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization*, Optim. Methods Softw., 29 (2014), pp. 239–263.

[38] K.-C. TOH AND S. YUN, *An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems*, Pac. J. Optim., 6 (2010), pp. 615–640.

[39] C. TOMASI AND T. KANADE, *Shape and motion from image streams under orthography: a factorization method*, Int. J. Comput. Vision, 9 (1992), pp. 137–154.

[40] Z. WEN, W. YIN, AND Y. ZHANG, *Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm*, Math. Program. Comput., 4 (2012), pp. 333–361.

[41] J. YANG AND X. YUAN, *An inexact alternating direction method for trace norm regularized least squares problem*, tech. report, Dept. of Mathematics, Nanjing University, 2010.

[42] J. YANG, Y. ZHANG, AND W. YIN, *An efficient tvl1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM J. Sci. Comput., 31 (2008), pp. 2842–2865.

[43] X. YUAN AND J. YANG, *Sparse and low-rank matrix decomposition via alternating direction method*, Pac. J. Optim., 9 (2013), pp. 167–180.

[44] Z. ZHU, A. M.-C. SO, AND Y. YE, *Fast and near–optimal matrix completion via randomized basis pursuit*, tech. report, Stanford University, 2009.

[45] H. ZOU, T. HASTIE, AND R. TIBSHIRANI, *Sparse principal component analysis*, J. Comput. Graph. Stat., 15 (2006), pp. 265–286.