Applied Mathematics and Computation 314 (2017) 133–141

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

The two-stage iteration algorithms based on the shortest distance for low-rank matrix completion☆

Rui-Ping Wen[a,*], Li-Xia Liu[b]

[a] Higher Education Key Laboratory of Engineering and Scientific Computing in Shanxi Province, Taiyuan Normal University, Taiyuan 030012, Shanxi Province, PR China
[b] School of Mathematics Sciences, Taiyuan University of Technology, Jinzhong, 030600, Shanxi Province, PR China

ARTICLE INFO

MSC:
65F50
65F10
15A83

Keywords:
Two-stage iteration
Alternating steepest descent algorithm
Low-rank matrix completion
Distance

ABSTRACT

Despite matrix completion requiring the global solution of a non-convex objective, there are many computational efficient algorithms which are effective for a broad class of matrices. Based on these algorithms for matrix completion with given rank problem, we propose a class of two-stage iteration algorithms for general matrix completion in this paper. The inner iteration is the scaled alternating steepest descent algorithm for the fixed-rank matrix completion problem presented by Tanner and Wei (2016), the outer iteration is used two iteration criterions: the gradient norm and the distance between the feasible part with the corresponding part of reconstructed low-rank matrix. The feasibility of the two-stage algorithms are proved. Finally, the numerical experiments show the two-stage algorithms with shorting the distance are more effective than other algorithms.

1. Introduction

From the pioneering work on low-rank approximation by Fazel [10] as well as on matrix completion by Candès and Recht [7], there has been a lot of study (see [1–26] and references therein) both from theoretical and algorithmic aspects on the problem of recovering a low-rank matrix from partial entries-also known as matrix completion. The problem occurs in many areas of engineering and applied science such as model reduction [18], machine learning [1,2], control [20], pattern recognition [9], imaging inpainting [3] and computer vision [23] and so on. There is a rapidly growing interest for this issue. Explicitly seeking the lowest rank matrix consistent with the known entries is mathematically expressed as:

$$\begin{aligned} &\min_{Z\in\Re^{m\times n}} \ \operatorname{rank}(Z) \\ &\text{subject to} \ \ P_\Omega(Z) = P_\Omega(Z^0), \end{aligned} \tag{1}$$

where the matrix $Z^0 \in \Re^{m\times n}$ is the underlying matrix to be reconstructed, Ω is a random subset of indices for the known entries, and P_Ω is the associated sampling orthogonal projection operator which acquires only the entries indexed by $\Omega \subset \{1,2,\ldots,m\}\times\{1,2,\ldots,n\}$.

Since a manifold of rank r matrices can be factorized into a bi-linear form $Z = XY$ where $X \in \Re^{m\times r}$ and $Y \in \Re^{r\times n}$, a few algorithms have been presented to solve (1) while the rank r was known or can be estimated [18,19]. The general

☆ This work is supported by NSF of China (11371275) and NSF of Shanxi Province (201601D011004).
* Corresponding author.
E-mail address: wenrp@163.com (R.-P. Wen).

http://dx.doi.org/10.1016/j.amc.2017.07.024
0096-3003/

problem (1), however, is non-convex and is NP-hard [12] due to the rank objective. Vandereycken [24] applied the Riemannian optimization to the problem by minimizing the least square distance on the sampling set over the Riemannian manifold of matrix Z^0. Then a Riemannian geometry method and a Riemannian trust-region method were given by Mishra et al. [21] and Boumal et al. [5], respectively. Evidently, the computation of a gradient is expensive in method, and then several methods resulted in alternating optimization and XY were raised [8,15,22].

On the other hand, Candés and Rechat [7] replaced the rank objective in (1) with its convex relaxation, the nuclear norm $\|Z\|_*$ which is the sum of all singular values of matrix Z, that is

$$\min_{Z\in\Re^{m\times n}} \|Z\|_* \quad \text{subject to} \quad P_\Omega(Z) = P_\Omega(Z^0). \tag{2}$$

Alternative to the convex optimization, there have been many algorithms which are designed to attempt to solve for the global minimum of (1) directly; many of them are adaptations of algorithms for compressed sensing, such as the hard thresholding algorithms [4,14,16], the singular value thresholding (SVT) method as well as its variants [6,13,25]. However, the computations of a partial singular value decomposition (SVD) were required at each iteration in the most direct implementation of these algorithms. The computational cost of computing the SVD has complexity of $O(n^3)$ when the rank r and matrix-size n are proportional, resulting in computing the SVD to be the dominant computational cost at each iteration and then limits their applicability for large n. In addition, Lin et al. [17] proposed an augmented Lagrange multiplier (ALM) method which performs better both in theory and algorithms than the others that with a Q-linear convergence speed globally.

Recently, based on the simple factorization $Z = XY$ where $X \in \Re^{m\times r}$ and $Y \in \Re^{r\times n}$, mentioned above, rather than solving (1), algorithms are designed to compute the non-convex problem

$$\min_{X,Y} f(X,Y) \tag{3}$$

with $f(X,Y) := \frac{1}{2}\|P_\Omega(Z^0) - P_\Omega(XY)\|_F^2$. In fact, the model (3) replaced the rank objective in (1) with the distance between a matrix and an r-dimensional manifold for the fixed-rank problem. Algorithms for the solution of (3) with the distance objective usually follow an alternating minimization scheme, with PowerFactorization [11] and LMaFit [26] two representatives. Tanner and Wei [22] proposed an alternating steepest descent (ASD), and a scaled variant (ScaledASD) methods in 2016. In so doing ASD and ScaledASD [22] are able to recover matrices of substantially higher rank than can LMaFit [26]. However, the rank of most of completing is unknown such that we have to estimate it in advance or approximate it from a lower rank until satisfying $P_\Omega(Z^0) = P_\Omega(XY)$. In this study, we first define the distance between a matrix and the r-dimensional manifold and then come up with a class of two-stage iteration algorithms for the case that the rank r was unknown. The rank is increased either one-by-one until the optimal rank r is obtained for a lower rank (say, be estimated) model or by combining l jumping-space with one-by-one until the optimal rank r is obtained for a larger rank (say, be estimated) model. The inner iteration finds the matrix which is up to the shortest or approximation shortest distance and the outer iteration finds the optimal r-dimensional manifold with two kinds criterion. The convergence theory of the new algorithms are studied.

The rest of the paper is organized as follows. A class of two-stage iteration algorithms for the case that the rank r was unknown is proposed in Section 2. The convergence of the new algorithms are discussed in Section 3. The numerical experiments are shown and comparison to algorithms in Section 4. Finally, we end the paper with a concluding remark in Section 5.

Here are some necessary notations and preliminaries. $\Re^{m\times n}$ is used to denote the $m \times n$ real matrix set, and $\Re^n$ the n-dimensional real vector set. X^T denotes the transpose of the matrix or vector X. The Frobenius norm is denoted by $\|X\|_F$. For a matrix $X = (x_1, x_2, \ldots, x_n) \in \Re^{m\times n}$, $\dim(X)$ is always used to represent dimensions of the manifold of fixed-rank matrix X and $\mathrm{rank}(X)$ represents the rank of a matrix X. Let $\Omega \subset \{1,2,\ldots,m\} \times \{1,2,\ldots,n\}$ denote the indices of the observed entries of the matrix X, $\bar{\Omega}$ denote the indices of the missing entries. Then P_Ω be the orthogonal projection operator on the span of matrices vanishing outside of Ω. So that the (i,j)th component of $P_\Omega(X)$ is equal to X_{ij} when $(i,j) \in \Omega$, and zero otherwise. Also, $\mathcal{Z}_r = \{Z \in \Re^{m\times n} : \mathrm{rank}(Z) = r\}$ stands for the r-dimensional manifold of fixed-rank matrices.

The one-to-one correspondence between a matrix and its projection enables us to devise a notation of distance between a matrix and an r-dimensional manifold as follows.

Definition 1.1. For a matrix $Y \in \Re^{m\times n}$,

$$d(Y_r) = \min \|Y - Z\|_F^2 \tag{4}$$

is called *distance* between a matrix Y and an r-dimensional manifold $\mathcal{Z}_r$. That is essentially the *distance* between matrix Y and its projection onto the r-dimensional manifold $\mathcal{Z}_r$.

We trivially give the distance between a feasible matrix and its projection onto the r-dimensional manifold for introducing the new iteration algorithms.

For $P_\Omega(Y) = P_\Omega(Z^0)$, we called $d(Y_r) = \min\limits_{\dim(Z)=r} \|Y - Z\|_F^2$ as a distance between a feasible matrix Y and an r-dimensional manifold $\mathcal{Z}_r$.

Evidently, it should be noted that $\min d(Y_r) > 0$ if $r < \min \mathrm{rank}(Z)$ and $\min d(Y_r) = 0$ if $r = \min \mathrm{rank}(Z)$. To obtain the $\min d(Y_r)$, some algorithms are presented by combining the model (1) and (4).