Applied Mathematics and Computation 258 (2015) 537–544

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A new gradient projection method for matrix completion ☆

Rui-Ping Wen *, Xi-Hong Yan

Higher Education Key Laboratory of Engineering and Scientific Computing in Shanxi Province, Taiyuan Normal University, Taiyuan 030012, Shanxi Province, PR China

ARTICLE INFO

Keywords:
Matrix completion
Gradient
Projection
Feasible sequence
Convergence

ABSTRACT

In this paper, a new gradient projection method is proposed, which generates a feasible matrix sequences. The decent property of this method is proved. Based on the decent property, the convergence of the new method is discussed. Moreover, a sufficient and necessary condition for the optimal matrix is obtained. Finally, numerical experiments show the new method is effective in precision.

1. Introduction

The matrix completion problem occurs in many areas of engineering and applied science such as machine learning [1,2], control [8] and computer vision[12]. There is a rapidly growing interest for this issue. As extension of the matrix completion problem, Candés and Recht [4], Recht et al. [9] consider the following optimization model:

$$\begin{aligned} &\text{minimize} \quad \|X\|_* \\ &\text{subject to} \quad X_{ij} = M_{ij}, \ (i,j) \in \Omega, \end{aligned} \tag{1.1}$$

where the functional $\|X\|_*$ is the nuclear norm of the matrix M, the unknown matrix $M \in \mathbb{R}^{n\times n}$ of rank r is square, and that one has available m sampled entries $\{M_{ij} : (i,j) \in \Omega\}$ with Ω is a random subset of cardinality m.

For the convex optimization, because minimizing the nuclear norm both provably recovers the lowest-rank matrix subject to constraints (see [9] for details) and gives generally good empirical results in a variety of situations, it is understandably of great interest to develop numerical methods for solving (1.1). In [4], this optimization model was solved by exploiting one of the most advanced semidefinite programming solvers, namely, SDPT3 [10]. This solver is based on interior-point methods and problematic when the size of the matrix is large because they need to solve huge systems of linear equations to compute the Newton direction. Cai e.g. [3] presented the singular value thresholding (SVT) method for approximately solving the nuclear norm minimization problem (1.1). This method is a simple and efficient first-order matrix completion method to recover the missing values when the original data matrix is of low rank, and is especially well suited for problems of very large sizes in which the solution has low rank. However, SVT is computationally expensive when the size of the data matrix is large, which significantly limits its applicability. Next, Hu et al. [6] give an accelerated singular values thresholding method (ASVT) by solving the dual problems. Combettes and Wajs [5] introduced the proximal forward–backward splitting, Toh and Yun [11] proposed some working implementations, Tseng [13] proposed the accelerated proximal gradient methods. Lin et al. [7] proposed the augmented Lagrange multiplier (ALM) methods. But these methods have a disadvantage of the

☆ This work is supported by NSF of China (11371275) and NSF of Shanxi Province (2014011006-1).
* Corresponding author.
E-mail address: wenrp@163.com (R.-P. Wen).

http://dx.doi.org/10.1016/j.amc.2015.02.041
0096-3003/

penalty function: transformed the constraint optimization into the unconstraint optimization problem by penalty, which should result in a matrix sequence $\{X_k\}$ generated by these methods is not feasible. So that the accepted solution is not feasible. Even though the interior point methods [4,10] generated a feasible matrix sequence, but the computation is large enough when the size of the matrix is large, because they need to solve the huge systems of linear equations to compute the Newton direction. These motivated us come up with the gradient projection method for matrix completion. First, we decrease the nuclear norm by using the singular value thresholding method. Second, we project the matrix to a feasible point, and generate a feasible matrix sequence.

Here are some notations and preliminaries. 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. $\|X\|_2, \|X\|_*, \|X\|_F$ denote 2-norm, the nuclear norm and F-norm, respectively. We denote by $\langle X,Y\rangle = \mathrm{trace}(X^*,Y)$ the inner product between two matrices ($\|X\|_F^2 = \langle X,X\rangle$). The Cauchy–Schwartz inequality gives $\langle X,Y\rangle \leqslant \|X\|_F \cdot \|Y\|_F$ and it is well known that $\langle X,Y\rangle \leqslant \|X\|_2 \cdot \|Y\|_*$ [4,13].

Let 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.

The rest of the paper is organized as follows. After we provide a brief review of the standard SVT and the ALM methods, a new gradient projection method is proposed in Section 2. Some properties and convergence of the new method are discussed in Section 3. Finally, numerical experiments are shown and comparison to other methods in Section 4.

2. Methods

First of this section, for completeness as well as purpose of comparison, we briefly review and summarize other two methods for solving the matrix completion problem (1.1).

2.1. The method of singular value thresholding (the SVT method)

The "shrinkage", $D_\tau(X)$ is soft-thresholding operator as follows,

$$D_\tau(X) = U\Sigma_\tau V^T, \tag{2.1}$$

$$\begin{cases} \Sigma_\tau = \mathrm{diag}(\max(\sigma_i - \tau, 0)), \\ X = U\Sigma V^T, \\ \Sigma = \mathrm{diag}(\sigma_i), \end{cases} \tag{2.2}$$

The singular value thresholding method is in [3] solves a convex optimization (1.1):

Method 2.1. Given $\Omega \subset \{(i,j), i = 1,2,\ldots,m;\ j = 1,2,\ldots,n\},\ P_\Omega(M)$ of the matrix $M \in \mathbb{R}^{m\times n}$.

Fix $\tau > 0$ and a sequence $\{\delta_k\}$ of positive step sizes. Starting with Y_0, inductively define for $k = 1,2,\ldots,$

1. $X_k = D_\tau(Y_{k-1})$.
2. $Y_k = Y_{k-1} + \delta_k P_\Omega(M - X_k)$.
3. If $\|P_\Omega(X_k - M)\|_F < \varepsilon\|P_\Omega(M)\|_F$, stop; otherwise, $k \leftarrow k+1$, go to 1.

where $0 < \delta_k < 2, X \in \mathbb{R}^{m\times n}$ is a matrix of rank r, and U, V are respectively $m \times r$ and $n \times r$ matrices with orthonormal columns, and the singular values σ_i are positive.

2.2. The method of augmented lagrange multipliers (the ALM method)

The augmented Lagrange multipliers method is in [7] solves a convex optimization (1.1). The problem (1.1) admits the following equivalent form,

$$\begin{aligned} &\text{minimize } \|X\|_* \\ &\text{subject to } X + E = M,\ \pi_\Omega(E) = 0, \end{aligned} \tag{2.3}$$

where $\pi_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is a linear operator that keeps the entries in Ω unchanged and sets those outside Ω (i.e., in $\overline{\Omega}$) zeros. As E will compensate for the unknown entries of M, the unknown entries of M are simply set as zeros. Then the partial augmented Lagrange function is

$$L(X,E,Y,\mu) = \|X\|_* + \langle Y, M - X - E\rangle + \frac{\mu}{2}\|M - X - E\|_F^2.$$

The augmented Lagrange multipliers method is described in the following: