Mechanical Systems and Signal Processing 95 (2017) 77–89

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Low-rank approximation pursuit for matrix completion ☆

An-Bao Xu *, Dongxiu Xie

College of Mathematics and Econometrics, Hunan University, Changsha 410082, People's Republic of China
School of Science, Beijing Information Science and Technology University, Beijing 100192, People's Republic of China

ARTICLE INFO

Article history:
Received 13 October 2016
Received in revised form 5 February 2017
Accepted 17 March 2017

Keywords:
Low rank approximation
Matrix completion
Randomized algorithm
Rank minimization
Matching pursuit

ABSTRACT

We consider the matrix completion problem that aims to construct a low rank matrix X that approximates a given large matrix Y from partially known sample data in Y. In this paper we introduce an efficient greedy algorithm for such matrix completions. The greedy algorithm generalizes the orthogonal rank-one matrix pursuit method (OR1MP) by creating $s \geqslant 1$ candidates per iteration by low-rank matrix approximation. Due to selecting $s \geqslant 1$ candidates in each iteration step, our approach uses fewer iterations than OR1MP to achieve the same results. Our algorithm is a randomized low-rank approximation method which makes it computationally inexpensive. The algorithm comes in two forms, the standard one which uses the Lanzcos algorithm to find partial SVDs, and another that uses a randomized approach for this part of its work. The storage complexity of this algorithm can be reduced by using an weight updating rule as an economic version algorithm. We prove that all our algorithms are linearly convergent. Numerical experiments on image reconstruction and recommendation problems are included that illustrate the accuracy and efficiency of our algorithms.

1. Introduction

Low-rank matrix completions have a wide range of applications, including recommender systems [10], quantum state tomography [1,6], linear system identification and control [13], and angle estimation in antenna arrays [18], for example. Motivated by applications, low rank matrix completions have recently gained increased attention. Our basic matrix completion model involves a large data matrix $Y_{m,n}$, a set Ω of row and column indices $\{(O_r, O_c)\}$ with $1 \leqslant O_r \leqslant m$ and $1 \leqslant O_c \leqslant n$ and $|\Omega| \ll m \cdot n$ that indicates the row and column indices of a subset of entries of Y. With this setup we try to find a low rank completion matrix $X_{m,n}$ that has the same entries as Y in the positions of Ω and arbitrary entries in its complementary positions chosen appropriately so that X has the minimal possible rank. This problem can be formalized as follows.

$$\min_{X \in \mathbb{R}^{m \times n}} rank(X) \quad \text{such that} \quad P_\Omega(X) = P_\Omega(Y), \tag{1.1}$$

where Ω denotes the set of all index pairs of chosen Y entries that X shares with Y. Here P_Ω is the orthogonal projector onto the span of matrices with zeros at all positions not in Ω. The purpose of this matrix completion from given Y and Ω is to construct a low rank matrix X from the partially observed matrix $P_\Omega(Y)$ of Y. By design this matrix completion relies on fewer than the $m \cdot n$ measurements in Y and can be used reconstruct the signal in Y accurately enough in much simpler form,

☆ Research supported by the National Natural Science Foundation of China (11371075, 61473325).
* Corresponding author at: College of Mathematics and Econometrics, Hunan University, Changsha 410082, People's Republic of China.
E-mail addresses: xuanbao777@163.com (A.-B. Xu), xkris@tom.com (D. Xie).

http://dx.doi.org/10.1016/j.ymssp.2017.03.024
0888-3270/

bringing significant benefits for many applications. It is well known that the matrix rank function $f(X) = rank(X)$ is a not convex and therefore rank optimization is difficult in general. The nuclear norm has been advocated as a convex surrogate function for rank, see [13,3]. The nuclear norm $||X||_*$ of a matrix X is defined as the sum of its singular values. And a relaxed convex formulation for problem (1.1) in terms of the nuclear norm is as follows.

$$\min_{X\in\mathbb{R}^{m\times n}} ||X||_* \quad \text{such that} \quad P_\Omega(X) = P_\Omega(Y).$$

This constrained nuclear norm minimization problem has been well studied. See for example the seminal work of [13,3]. There are several convex relaxation-based algorithms to solve this problem and variants: singular value projection (SVP) [9], singular value thresholding (SVT) [2], Jaggi's fast algorithm for trace norm constraint (JS) [8], the spectral regularization algorithm (SoftImpute) [12], low rank matrix fitting (LMaFit) [17], a boosting-type accelerated matrix-norm penalized solver (Boost) [19]. Unfortunately the high computational complexity of these convexity based methods make them impractical for many application.

Another competitive algorithm is based on the greedy strategy, including atomic decomposition for minimum rank approximation (ADMiRA) [11], the greedy efficient component optimization (GECO) [14], and orthogonal rank-one matrix pursuit (OR1MP) [16]. These approaches have received significant attention due to their low complexity and simple implementation.

Our work builds on the OR1MP algorithm [16], which is in turn based on the OMP algorithm [15]. We generalize OR1MP in the low-rank approximation pursuit algorithm (LRAP) that solves the matrix completion problem (1.1). Differing from the OR1MP algorithm, the LRAP algorithm selects multi-candidates and adds them to the basis set by the best rank-s approximations in each iteration step. It is the procedure of selecting multi-candidates with OR1MP and further decreasing the computational complexity. Note that for $s = 1$, LRAP is identical to OR1MP. In the standard LRAP algorithm we fully update the weights (or the coefficients) for all rank-1 matrices in the current basis set after each iteration. However, the full weight updating of the standard algorithm involves all rank-1 matrices in the current basis set, i.e., $k \cdot s \cdot |\Omega|$ elements after the kth iteration. To overcome this drawback, we also adopt an economic weight updating rule in the more economical version ELRAP of our algorithm. ELRAP reduces both time and storage needs for LRAP. Finally we prove that our both algorithms achieve linear convergence.

The main contributions of our paper are:

- We propose a computationally more efficient greedy algorithm for the matrix completion (1.1), which extends the orthogonal rank-one matrix pursuit from selecting just one candidate per iteration step to multiple candidates that are added to the basis set. We further reduce the storage complexity of our basic algorithm by using an economic weight updating rule. We show that both versions of our algorithm achieve linear convergence.
- We count the number of floating-point operations of our LRAP algorithm and of its more economic version ELRAP in order to show that our algorithms scale well to large problems.
- To verify the efficiency of our algorithm, we compare our LRAP and ELRAP algorithms with three state-of-the-art matrix completion algorithms on large-scale data sets, such as Jester[1] and MovieLens.[2]

This paper is organized as follows: Additional notations are introduced below. Then we present our the LRAP algorithm and its economic ELRAP version in Section 2. In Section 3, we extend our algorithms to deal with the matrix sensing problem and prove the algorithms' linear convergence rate. Empirical numerical test evaluations are presented in Section 4 that verify the efficiency of our algorithms.

By $vec(X)$ we denote the column vector derived from a matrix X by concatenating all its columns in one column. We define $\dot{\mathbf{x}}$ as the vector obtained by concatenating all entries with row and column indices in Ω for X, i.e., the column vector $vec(P_\Omega(X))$. For two compatibly sized matrices X and Y, their Frobenius inner product and their matrix norm are defined as $\langle Y, X\rangle = trace(X^T Y)$ and $\|X\| = \sqrt{\langle X, X\rangle}$, respectively. We denote $P_\Omega(X)$ by X_Ω and define $\langle X, Y\rangle_\Omega = \langle X_\Omega, X_\Omega\rangle$ and $\|X\|_\Omega = \sqrt{\langle X, X\rangle_\Omega}$.

2. Low-rank approximation pursuit and an economic version

The singular value decomposition (SVD) of a matrix $X \in \mathbb{R}^{m\times n}$ factors it as $U\Sigma V^T$, where $U_{m,m}$ and $V_{n,n}$ are unitary matrix, and $\Sigma \in \mathbb{R}^{m\times n}$ is a rectangular diagonal matrix with non-negative decreasing diagonal entries. Since U and V both have mutually orthonormal columns, $X = U\Sigma V^T$ can be rewritten as $X = M(\boldsymbol{\theta}) = \sum_{i=I}\theta_i M_i$, where θ_i are diagonal entries of Σ, the vector $\boldsymbol{\theta}$ is the vector of θ_i, and each M_i is an $m \times n$ rank-one matrix generated by the rows of U and V. Therefore the original low rank approximation problem (1.1) can be rewritten as:

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_0 \quad \text{such that} \quad P_\Omega(M(\boldsymbol{\theta})) = P_\Omega(Y),$$

[1] http://eigentaste.berkeley.edu/dataset/.

[2] https://grouplens.org/datasets/movielens/.