Commun Nonlinear Sci Numer Simulat 21 (2015) 70–88

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

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

Stochastic algorithms for solving structured low-rank matrix approximation problems

J.W. Gillard *, A.A. Zhigljavsky

Cardiff School of Mathematics, Cardiff University, United Kingdom

ARTICLE INFO

Article history:
Available online 2 October 2014

Keywords:
Structured low rank approximation
Hankel matrix
Global optimization

ABSTRACT

In this paper, we investigate the complexity of the numerical construction of the Hankel structured low-rank approximation (HSLRA) problem, and develop a family of algorithms to solve this problem. Briefly, HSLRA is the problem of finding the closest (in some predefined norm) rank r approximation of a given Hankel matrix, which is also of Hankel structure. We demonstrate that finding optimal solutions of this problem is very hard. For example, we argue that if HSLRA is considered as a problem of estimating parameters of damped sinusoids, then the associated optimization problem is basically unsolvable. We discuss what is known as the orthogonality condition, which solutions to the HSLRA problem should satisfy, and describe how any approximation may be corrected to achieve this orthogonality. Unlike many other methods described in the literature the family of algorithms we propose has the property of guaranteed convergence.

1. Introduction

1.1. Statement of the problem

Let $L, \ K$ and r be given positive integers such that $1 \leqslant r < L \leqslant K$. Denote the set of all real-valued $L \times K$ matrices by $\mathbb{R}^{L\times K}$. Let $\mathcal{M}_r = \mathcal{M}_r^{L\times K} \subset \mathbb{R}^{L\times K}$ be the subset of $\mathbb{R}^{L\times K}$ containing all matrices with rank $\leqslant r$, and $\mathcal{H} = \mathcal{H}^{L\times K} \subset \mathbb{R}^{L\times K}$ be the subset of $\mathbb{R}^{L\times K}$ containing matrices of some known structure. The set of structured $L \times K$ matrices of rank $\leqslant r$ is $\mathcal{A} = \mathcal{M}_r \cap \mathcal{H}$.

Assume we are given a matrix $\mathbf{X}_* \in \mathcal{H}$. The problem of structured low rank approximation (SLRA) is:

$$f(\mathbf{X}) \to \min_{\mathbf{X}\in\mathcal{A}} \tag{1}$$

where $f(\mathbf{X}) = \rho^2(\mathbf{X}, \mathbf{X}_*)$ is a squared distance on $\mathbb{R}^{L\times K} \times \mathbb{R}^{L\times K}$.

In this paper we only consider the case where $\mathcal{H}$ is the set of Hankel matrices and thus refer to (1) as HSLRA. Recall that a matrix $\mathbf{X} = (x_{lk}) \in \mathbb{R}^{L\times K}$ is called Hankel if $x_{lk} = \text{const}$ for all pairs (l, k) such that $l + k = \text{const}$; that is, all elements on the anti-diagonals of $\mathbf{X}$ are equal. There is a one-to-one correspondence between $L \times K$ Hankel matrices and vectors of size $N = L + K - 1$. For a vector $Y = (y_1, \ldots, y_N)^T$, the matrix $\mathbf{X} = \mathbb{H}(Y) = (x_{lk}) \in \mathbb{R}^{L\times K}$ with elements $x_{lk} = y_{l+k-1}$ is Hankel and vise-versa: for any matrix $\mathbf{X} \in \mathcal{H}$, we may define $Y = \mathbb{H}^{-1}(\mathbf{X})$ so that $\mathbf{X} = \mathbb{H}(Y)$.

We consider the distances ρ defined by the semi-norms

$$||\mathbf{A}||_W^2 = \text{tr}\,\mathbf{AWA}^T \quad (\text{so that } f(\mathbf{X}) = \text{tr}(\mathbf{X} - \mathbf{X}_*)\mathbf{W}(\mathbf{X} - \mathbf{X}_*)^T), \tag{2}$$

* Corresponding author.
E-mail addresses: GillardJW@Cardiff.ac.uk (J.W. Gillard), ZhigljavskyAA@Cardiff.ac.uk (A.A. Zhigljavsky).

http://dx.doi.org/10.1016/j.cnsns.2014.08.023
1007-5704/

where $\mathbf{A} \in \mathbb{R}^{L \times K}$ and $\mathbf{W}$ is a symmetric non-negative definite negative matrix of size $K \times K$. Moreover, in our main application the weight matrix $\mathbf{W}$ is diagonal:

$$\mathbf{W} = \mathrm{diag}(w_1, \ldots, w_K), \tag{3}$$

where $w_1, \ldots, w_K$ are some positive numbers.

1.2. Background

The aim of low-rank approximation methods is to approximate a matrix containing observed data, by a matrix of pre-specified lower rank r. The rank of the matrix containing the original data can be viewed as the order of complexity required to fit to the data exactly, and a matrix of lower complexity (lower rank) 'close' to the original matrix is often required. A further requirement is that if the original matrix of the observed data is of a particular structure, then the approximation should also have this structure. An example is the HSLRA problem as defined in the previous section.

HSLRA is a very important problem with applications in a number of different areas. In addition to the clear connection with time series analysis and signal processing, HSLRA has been extensively used in system identification (modeling dynamical systems) [1], in speech and audio processing [2], in modal and spectral analysis [3] and image processing [4]. Some discussion on the relationship of HSLRA with some well known subspace-based methods of time series analysis and signal processing is given in [5]. Similar structures used in (1) include Toeplitz, block Hankel and block Toeplitz structures. In image processing, there is much use of Hankel-block Hankel structures. Further details, references and specific applications of SLRA are provided in [6–8].

1.3. Notation

The following list contains the main notation used in this paper.

N, L, K, r	Positive integers with $1 \leqslant r < L \leqslant K < N$, $N = L + K - 1$
$\mathbb{R}^{L \times K}$	Set of $L \times K$ matrices
$\mathbb{R}^N$	Set of vectors of length N
$\mathcal{H}^{L \times K}$ or $\mathcal{H}$	Set of $L \times K$ Hankel matrices
$\mathcal{M}_r$	Set of $L \times K$ matrices of rank r
$\mathcal{A} = \mathcal{M}_r \cap \mathcal{H}$	Set of $L \times K$ Hankel matrices of rank r
$Y = (y_1, \ldots, y_N)^T$	Vector in $\mathbb{R}^N$
$\mathbb{H}(Y)$	Hankel matrix in $\mathcal{H}^{L \times K}$ associated with vector $Y \in \mathbb{R}^N$
$\mathbf{X}_* \in \mathcal{H}^{L \times K}$	Given matrix
$Y_* = (y_{1*}, \ldots, y_{N*})^T$	Vector in $\mathbb{R}^N$ such that $\mathbb{H}(Y_*) = \mathbf{X}_*$ (vector of observed values)
$\pi_{\mathcal{H}}(\mathbf{X})$	Projection of the matrix $\mathbf{X} \in \mathbb{R}^{L \times K}$ onto the set $\mathcal{H}$
$\pi^{(r)}(\mathbf{X})$	Projection of a matrix $\mathbf{X} \in \mathbb{R}^{L \times K}$ onto the set $\mathcal{M}_r$
$\mathbf{I}_p$	Identity matrix of size $p \times p$.

1.4. Structure of the paper and the main results

The structure of the paper is as follows. In Section 2 we formally define the HSLRA problem (1) as an optimization problem in the space of matrices and introduce a generic norm defining the objective function $f(\cdot)$. In Section 2 we also describe projection operators to $\mathcal{H}$ and especially $\mathcal{M}_r$ that are used throughout the majority of the algorithms introduced in this paper, in the process of solving the HSLRA problem. In Section 3 we study the relations between different norms which define the objective function in two different setups. In Section 3, we also discuss some computational aspects for dealing with infinite and infinitesimal numbers. In Section 4 we study the so-called orthogonality condition which optimal solutions of (1) should satisfy, and describe how an approximation may be corrected to achieve this orthogonality. Section 5 considers some algorithms for solving the HSLRA problem represented as optimization problems in the set of Hankel matrices $\mathcal{H}$. We start with formulating a well-known algorithm based on alternating projections to the spaces $\mathcal{M}_r$ and $\mathcal{H}$, and call this AP. This is followed by an introduction of an improved version of this algorithm which we call 'Orthogonalized Alternating Projections' (OAP). In Section 5.2 we introduce a family of algorithms which incorporate randomization, backtracking, evolution and selection. The algorithms described in this section have guaranteed convergence to the optimum, unlike all other methods described in the literature. The main algorithm introduced and studied in the paper is called APBR (which in an abbreviation for 'Alternating Projections with Backtracking and Randomization'). Examples provided in Section 6 show that APBR significantly outperforms AP, as well as some other methods. In (A), we consider the HSLRA problem (1) by associating matrices $\mathbf{X} \in \mathcal{A}$ with vectors whose elements can be represented as sums of damped sinusoids; this approach is popular in the signal processing literature. We demonstrate that the resulting objective function can be very complex which means that the associated optimization problems are basically unsolvable. Section 7 concludes the paper.