Journal of Computational and Applied Mathematics 272 (2014) 430–448

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Variable projection for affinely structured low-rank approximation in weighted 2-norms

Konstantin Usevich *, Ivan Markovsky

Vrije Universiteit Brussel, Department ELEC, Pleinlaan 2, B-1050, Brussels, Belgium

ARTICLE INFO

Article history:
Received 29 October 2012
Received in revised form 7 March 2013

MSC:
15B99
15B05
41A29
49M30
65F30
65K05
65Y20

Keywords:
Structured low-rank approximation
Variable projection
Mosaic Hankel matrices
Weighted 2-norm
Fixed elements
Computational complexity

ABSTRACT

The structured low-rank approximation problem for general affine structures, weighted 2-norms and fixed elements is considered. The variable projection principle is used to reduce the dimensionality of the optimization problem. Algorithms for evaluation of the cost function, the gradient and an approximation of the Hessian are developed. For $m \times n$ mosaic Hankel matrices the algorithms have complexity $O(m^2 n)$.

1. Introduction

An *affine matrix structure* is an affine map from a *structure parameter space* $\mathbb{R}^{n_p}$ to a space of matrices $\mathbb{R}^{m\times n}$, defined by

$$\mathscr{S}(p) = S_0 + \sum_{i=1}^{n_p} p_k S_k, \tag{$\mathscr{S}$}$$

where $S_k \in \mathbb{R}^{m\times n}$. Without loss of generality, we can consider only the case $m \le n$. The *structured low-rank approximation* is the problem of finding the best low-rank structure-preserving approximation of a given data matrix [1,2].

Problem 1 (*Structured Low-Rank Approximation*)**.** Given an affine structure $\mathscr{S}$, data vector $p_D \in \mathbb{R}^{n_p}$, norm $\|\cdot\|$ and natural number $r < \min(m, n)$

$$\underset{\widehat{p}\in\mathbb{R}^{n_p}}{\text{minimize}}\ \|p_D - \widehat{p}\| \quad \text{subject to rank}\,\mathscr{S}(\widehat{p}) \le r. \tag{SLRA}$$

* Corresponding author. Tel.: +32 0 2 629 29 79; fax: +32 0 2 629 28 50.
E-mail addresses: konstantin.usevich@vub.ac.be (K. Usevich), Ivan.Markovsky@vub.ac.be (I. Markovsky).

0377-0427/$ – see front matter
http://dx.doi.org/10.1016/j.cam.2013.04.034

In this paper, we consider the case of a *weighted* 2-*norm*, given by

$$\|p\|_{\mathrm{W}}^2 := p^\top \mathrm{W} p, \quad \mathrm{W} \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{p}}, \tag{$\|\cdot\|_{\mathrm{W}}^2$}$$

where W is

- either a symmetric positive definite matrix,
- or a diagonal matrix

$$\mathrm{W} = \operatorname{diag}(w_1, \ldots, w_{n_\mathrm{p}}), \quad w_i \in (0; \infty], \tag{$w \to \mathrm{W}$}$$

where $\infty \cdot 0 = 0$ by convention. A *finiteness* constraint $\|p_\mathrm{D} - \widehat{p}\|_{\mathrm{W}}^2 < \infty$ is additionally imposed on (SLRA). Problem (SLRA) with the weighted norm ($\|\cdot\|_{\mathrm{W}}^2$) given by ($w \to \mathrm{W}$) is equivalent to (SLRA) with an *element-wise weighted* 2-norm

$$\|p\|_w^2 = \sum_{w_i \neq \infty} w_i p_i^2,$$

and a set of *fixed values* constraints

$$(p_\mathrm{D})_i = \widehat{p}_i \quad \text{for all } i \text{ with } w_i = \infty.$$

The structured low-rank approximation problem with the weighted 2-norm appears in signal processing, computer algebra, identification of dynamical systems, and other applications. We refer the reader to [1,2] for an overview. In this paper, we consider general affine structures ($\mathscr{S}$) and, in particular, structures that have the form

$$\mathscr{S}(p) = \Phi \mathscr{H}_{\mathbf{m},\mathbf{n}}(p), \tag{$\Phi\mathscr{H}_{\mathbf{m},\mathbf{n}}$}$$

where Φ is a full row rank matrix and $\mathscr{H}_{\mathbf{m},\mathbf{n}}$ is a mosaic Hankel [3] matrix structure.

Many data modeling problems can be reduced to (SLRA) with the structure ($\Phi\mathscr{H}_{\mathbf{m},\mathbf{n}}$) and weighted norm, defined by ($w \to \mathrm{W}$), see [1,4]. In data modeling, the number of rows m usually has the meaning of the model complexity and the number of columns n is of the same order as the number of data points [2]. Typically, the case $m \ll n$ is of interest, i.e. approximation of a large amount of data by a low-complexity model.

1.1. Mosaic Hankel structure

A *mosaic Hankel* matrix structure $\mathscr{H}_{\mathbf{m},\mathbf{n}}$ [3] is a map defined by two integer vectors

$$\mathbf{m} = \begin{bmatrix} m_1 & \cdots & m_q \end{bmatrix} \in \mathbb{N}^q \quad \text{and} \quad \mathbf{n} = \begin{bmatrix} n_1 & \cdots & n_N \end{bmatrix} \in \mathbb{N}^N \tag{$\mathbf{m},\mathbf{n}$}$$

as follows:

$$\mathscr{H}_{\mathbf{m},\mathbf{n}}(p) := \begin{bmatrix} \mathscr{H}_{m_1,n_1}(p^{(1,1)}) & \cdots & \mathscr{H}_{m_1,n_N}(p^{(1,N)}) \\ \vdots & & \vdots \\ \mathscr{H}_{m_q,n_1}(p^{(q,1)}) & \cdots & \mathscr{H}_{m_q,n_N}(p^{(q,N)}) \end{bmatrix}, \tag{$\mathscr{H}_{\mathbf{m},\mathbf{n}}$}$$

where

$$p = \operatorname{col}(p^{(1,1)}, \ldots, p^{(q,1)}, \ldots, p^{(1,N)}, \ldots, p^{(q,N)}), \quad p^{(k,l)} \in \mathbb{R}^{m_k + n_l - 1}, \tag{p}$$

is the partition of the parameter vector, and $\mathscr{H}_{m,n} : \mathbb{R}^{m+n-1} \to \mathbb{R}^{m \times n}$ is the *Hankel* structure

$$\mathscr{H}_{m,n}(p) := \begin{bmatrix} p_1 & p_2 & p_3 & \cdots & p_n \\ p_2 & p_3 & ⋰ & & p_{n+1} \\ p_3 & ⋰ & & & \vdots \\ \vdots & & & & p_{m+n-2} \\ p_m & p_{m+1} & \cdots & p_{m+n-2} & p_{m+n-1} \end{bmatrix}.$$

Note that the number of parameters for ($\mathscr{H}_{\mathbf{m},\mathbf{n}}$) is equal to

$$n_\mathrm{p} = N \sum_{k=1}^{q} m_k + q \sum_{l=1}^{N} n_l - Nq,$$

the number of columns is equal to $\sum_{k=1}^{q} m_k$, and the number of rows is $\sum_{l=1}^{N} n_l$.