Available online at www.sciencedirect.com
ScienceDirect
automatica
Automatica 44 (2008) 891–909
www.elsevier.com/locate/automatica

Survey paper

Structured low-rank approximation and its applications ☆

Ivan Markovsky

School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK

Received 7 November 2006; received in revised form 2 August 2007; accepted 5 September 2007
Available online 26 December 2007

Abstract

Fitting data by a bounded complexity linear model is equivalent to low-rank approximation of a matrix constructed from the data. The data matrix being Hankel structured is equivalent to the existence of a linear time-invariant system that fits the data and the rank constraint is related to a bound on the model complexity. In the special case of fitting by a static model, the data matrix and its low-rank approximation are unstructured.

We outline applications in system theory (approximate realization, model reduction, output error, and errors-in-variables identification), signal processing (harmonic retrieval, sum-of-damped exponentials, and finite impulse response modeling), and computer algebra (approximate common divisor). Algorithms based on heuristics and local optimization methods are presented. Generalizations of the low-rank approximation problem result from different approximation criteria (e.g., weighted norm) and constraints on the data matrix (e.g., nonnegativity). Related problems are rank minimization and structured pseudospectra.

Keywords: Low-rank approximation; Total least squares; System identification; Errors-in-variables modeling; Behaviors

1. Introduction

Fitting linear models to data can be achieved, both conceptually and algorithmically, by solving a system of equations $AX = B$, where the matrices A and B are constructed from the given data and the matrix X parameterizes the model to be found. In this classical approach, the main tools are the least squares method and its variations—data least squares (Degroat & Dowling, 1991), total least squares (TLS) (Golub & Van Loan, 1980), structured TLS (De Moor, 1993), robust least squares (Chandrasekaran, Gu, & Sayed, 1998), etc. The least squares method and its variations are mainly motivated by their applications for data fitting, but they invariably consider solving approximately an overdetermined system of equations.

In this paper we show that a number of linear data fitting problems are equivalent to the abstract problem of approximating a matrix D constructed from the data by a low-rank matrix. Partitioning the data matrix into matrices $A \in \mathbb{R}^{N\times \mathtt{m}}$ and $B \in \mathbb{R}^{N\times \mathtt{p}}$ and solving approximately the system $AX = B$ is a way to achieve rank-$\mathtt{m}$ or less approximation. The converse implication, however, is not true, because $[A \;\; B]$ having rank-$\mathtt{m}$ or less does not imply the existence of X, such that $AX = B$. This lack of equivalence between the original low-rank approximation problem and the $AX = B$ problem motivates what is called nongeneric TLS problem (Van Huffel & Vandewalle, 1991), whose theory is more complicated than the one of the generic problem and is difficult to solve numerically.

Alternative approaches for achieving a low-rank approximation are to impose that the data matrix has

1. at least $\mathtt{p} := \text{coldim}(B)$ dimensional nullspace, or
2. at most $\mathtt{m} := \text{coldim}(A)$ dimensional column space.

Parameterizing the nullspace and the column space by sets of basis vectors, the alternative approaches are:

1. *kernel representation*: there is a full rank matrix $R \in \mathbb{R}^{\mathtt{p}\times(\mathtt{m}+\mathtt{p})}$, such that $[A \;\; B]R^\top = 0$, and
2. *image representation*: there are matrices $P \in \mathbb{R}^{(\mathtt{m}+\mathtt{p})\times \mathtt{m}}$ and $L \in \mathbb{R}^{\mathtt{m}\times N}$, such that $[A \;\; B]^\top = PL$.

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Wolfgang Scherrer under the direction of Editor Manfred Morari.

E-mail address: im@ecs.soton.ac.uk.

0005-1098/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2007.09.011

The approaches using kernel and image representations are equivalent to the original low-rank approximation problem. Next we illustrate the use of $AX = B$, kernel, and image representations on the most simple data fitting problem—line fitting.

1.1. Line fitting example

Given a set of points $\{d_1, \dots, d_N\} \subset \mathbb{R}^2$ in the plane, the aim of the line fitting problem is to find a line passing through the origin that "best" matches the given points. The classical approach for line fitting is to define $\operatorname{col}(a_i, b_i) := d_i$ and solve approximately the overdetermined system

$$\operatorname{col}(a_1, \dots, a_N)x = \operatorname{col}(b_1, \dots, b_N) \tag{1}$$

by the least squares method. Let x_{ls} be the least squares approximate solution to (1). Then the least squares fitting line is

$$\mathscr{B}_{\text{ls}} := \{d = \operatorname{col}(a, b) \in \mathbb{R}^2 \mid a x_{\text{ls}} = b\}. \tag{2}$$

Geometrically, $\mathscr{B}_{\text{ls}}$ minimizes the sum of the squared vertical distances from the data points to the fitting line.

The left plot in Fig. 1 shows a particular example with $N = 10$ data points. The data points $d_1, \dots, d_{10}$ are the circles in the figure, the fit $\mathscr{B}_{\text{ls}}$ is the solid line, and the fitting errors $e := a x_{\text{ls}} - b$ are the dashed lines. Visually we expect the best fit to be the vertical axis, so minimizing vertical distances is not appropriate in this example.

Note that by solving (1), we treat the a_i (the first components of the d_i) differently from the b_i (the second components): b_i is assumed to be a *function* of a_i. This is an arbitrary choice; we can as well fit the data by solving approximately the system

$$\operatorname{col}(a_1, \dots, a_N) = \operatorname{col}(b_1, \dots, b_N)x, \tag{3}$$

in which case a_i is assumed to be a function of b_i. Let x'_{ls} be the least squares approximate solution to (3). It gives the fitting line

$$\mathscr{B}'_{\text{ls}} := \{d = \operatorname{col}(a, b) \in \mathbb{R}^2 \mid a = b x'_{\text{ls}}\}, \tag{4}$$

which minimizes the sum of the squared horizontal distances (see the right plot in Fig. 1). The line $\mathscr{B}'_{\text{ls}}$ happens to achieve the desired fit in the example. This shows that

> in the classical method for data fitting, i.e., solving approximately a linear system of equations in the least squares sense, the choice of the model representation determines the fitting criterion.

This feature of the classical method is undesirable: it is more natural for a user of a data fitting method to specify a desired fitting criterion instead of a model representation that implicitly corresponds to that criterion.

TLS is an alternative to least squares for approximately solving an overdetermined system of equations. In terms of data fitting, the TLS method minimizes the sum of the squared orthogonal distances from the data points to the fitting line. Using the system of equations (1), line fitting by the TLS

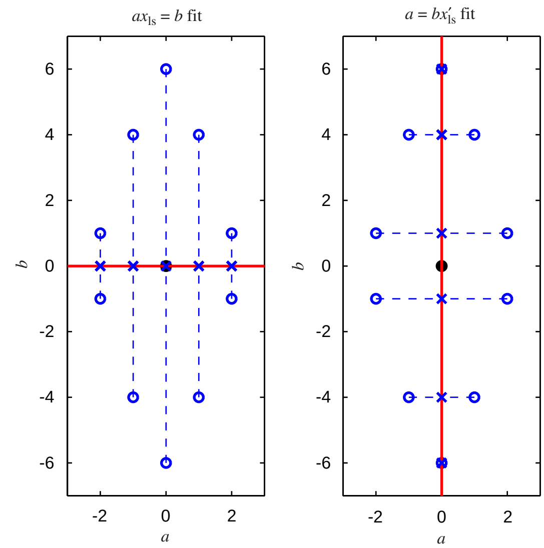

Fig. 1. The data and the least squares fits, minimizing vertical (left) and horizontal (right) distances.

method leads to the problem

$$\min_{\substack{x \in \mathbb{R} \\ \widehat{a}_1, \dots, \widehat{a}_N \in \mathbb{R} \\ \widehat{b}_1, \dots, \widehat{b}_N \in \mathbb{R}}} \sum_{i=1}^N \left\| d_i - \begin{bmatrix} \widehat{a}_i \\ \widehat{b}_i \end{bmatrix} \right\|_2^2 \quad \text{s.t.} \quad \widehat{a}_i x = \widehat{b}_i \quad \text{for } i = 1, \dots, N. \tag{5}$$

However, for the given data it has no solution. Informally, the TLS solution is $x_{\text{tls}} = \infty$, which corresponds to a fit by a vertical line. However, formally

> the TLS problem (5) has no solution for the data in the example and therefore does not give a fitting line.

By using (1) to define the TLS line fitting problem, we restrict the fitting line to be a graph of a function $ax = b$ for some $x \in \mathbb{R}$. Thus, we a priori exclude the vertical line as a possible solution. In the example, the line minimizing the sum of the squared orthogonal distances happens to be the vertical line. For this reason, x_{tls} does not exist.

Any line $\mathscr{B}$ passing through the origin can be represented as an image and a kernel, i.e., there exist matrices $P \in \mathbb{R}^{2 \times 1}$ and $R \in \mathbb{R}^{1 \times 2}$, such that

$$\mathscr{B} = \operatorname{image}(P) := \{d = Pl \in \mathbb{R}^2 \mid l \in \mathbb{R}\}$$

and

$$\mathscr{B} = \ker(R) := \{d \in \mathbb{R}^2 \mid Rd = 0\}.$$

Using an image representation, the problem of minimizing the sum of the orthogonal distances is

$$\min_{\substack{P \in \mathbb{R}^{2 \times 1} \\ l_1, \dots, l_N \in \mathbb{R}}} \sum_{i=1}^N \|d_i - \widehat{d}_i\|_2^2 \quad \text{s.t.} \quad \widehat{d}_i = P l_i \quad \text{for } i = 1, \dots, N. \tag{6}$$