Appl. Comput. Harmon. Anal. 40 (2016) 600–609

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

SVD revisited: A new variational principle, compatible feature maps and nonlinear extensions

Johan A.K. Suykens

KU Leuven, ESAT-STADIUS, Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee), Belgium

A R T I C L E I N F O

Article history:
Received 18 May 2015
Accepted 18 September 2015
Available online 25 September 2015
Communicated by Charles K. Chui

Keywords:
Singular value decomposition (SVD)
Variational principle
Least squares support vector machines
Lagrange duality
Feature maps
Kernel methods
Nonlinearity

A B S T R A C T

In this letter a new variational principle to the matrix singular value decomposition (SVD) is proposed. It is formulated as a constrained optimization problem where two sets of constraints are expressed in terms of compatible feature maps, which are evaluated on data vectors that relate to the rows and columns of the given matrix. Provided that a compatibility condition holds the solution can be related to Lanczos' decomposition theorem. The method is further extended to nonlinear SVD, which is illustrated also on image examples.

1. Introduction

The Singular Value Decomposition (SVD) is a fundamental method in linear algebra [7,9,20], with numerous applications in various different fields. The historical overview paper by Stewart [19] explains the early history of this method, including e.g. contributions by Beltrami [3], Jordan [11,12], Eckart & Young [6] and Lanczos [14], and several others.

In this letter we propose an alternative variational principle to the SVD, which will be connected to Lanczos' decomposition theorem [14]. The formulation is given within the setting of least squares support vector machines [21,23], for which the primal problem consists of constraints related to the data points and the L_2 loss function is used in the objective function. In order to conceive the SVD within this setting, two data sets are first defined on the given data matrix. These relate to the rows and columns of the given matrix. On these two data sets compatible linear feature maps are applied then, for which the features are linearly combined. The objective function involves then the inner product of the weight vectors and the L_2 loss function parts.

E-mail address: johan.suykens@esat.kuleuven.be.

http://dx.doi.org/10.1016/j.acha.2015.09.004

Due to the fact that the given matrix in the SVD is typically non-square, an additional difficulty is that the weight vectors and feature maps should be made compatible in dimension, in order to be able to compare them. This leads to a compatibility condition that should hold, from which the compatibility matrix (or matrices, depending on the formulation) can be computed.

A major difference with the support vector machine method in [4,24], where a Mercer kernel is employed, is that in the main theorem proposed in this letter, no Mercer kernel is employed at the dual level. The reason is that one obtains inner products between the two different feature maps, instead of among the same feature maps. In fact this is to be expected, because the SVD formulations in integral equations make use of unsymmetric kernels, according to the early work by Schmidt [18].

The proposed formulation in this letter is related to the shifted eigenvalue problem [14], while least squares support vector machine formulations were previously given to various eigenvalue and generalized eigenvalue problems [1,2,15,21,22] arising in kernel principal component analysis [16,22], kernel spectral clustering [2,15] and kernel canonical correlation analysis [1,10,13,8,21]. A further possible extension to nonlinear SVD is proposed. It is not restricted to the use of Mercer kernels and reproducing kernels, which are commonly used in learning theory [5,17,24,25].

This letter is organized as follows. In Section 2 a number of aspects of SVD are introduced. In Section 3 a new variational principle to the SVD is proposed. In Section 4 possible extensions to nonlinear SVD are explained. A few illustrations are presented in Section 5.

2. Context and problem statement

The Singular Value Decomposition (SVD) [7] of a real-valued matrix $A \in \mathbb{R}^{N\times M}$ is given by

$$A = U\Sigma V^T \tag{1}$$

with orthonormal matrices $U = [u_1 \dots u_N] \in \mathbb{R}^{N\times N}$, $V = [v_1 \dots v_M] \in \mathbb{R}^{M\times M}$ satisfying $U^TU = I_N$, $V^TV = I_M$, and diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{N\times M}$ where $p = \min\{N, M\}$ and singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$. For a rank r matrix A one has the dyadic decomposition

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T. \tag{2}$$

The SVD is related to the following variational principle [3,11,12,19] which looks for the extrema of the bilinear form

$$f(u, v) = u^T A v \text{ subject to } \|u\|^2 = \|v\|^2 = 1. \tag{3}$$

The solutions are also obtained then from the eigenvalue decomposition

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix} \tag{4}$$

where $\lambda \in \{\pm\sigma_1, \pm\sigma_2, \dots, \pm\sigma_p, 0\}$ with multiplicity $M - N$ for the zero eigenvalue (assuming $M > N$ and non-zero σ_i values).

In this letter the following theorem by Lanczos [14] is of special importance.

Theorem 1 *(Decomposition Theorem, Lanczos (1958)). (See [14, pp. 671–672].) An arbitrary non-zero matrix A can be written as*